

# ÇOKLU KARŞILAŞTIRMALARDA DAĞILIŞLA VE VARYANS YAPISIYLA İLGİLİ FARAZİYELERDEN AYRILIŞLARIN TEST ÖZELLİKLERİNE ETKİLERİ ÜZERİNDE BİR ARAŞTIRMA<sup>1</sup>

Mehmet Cudi OKUR<sup>2</sup>

## Ö Z E T

*Çoklu karşılaştırma testlerinden en fazla kullanılan LSD, Duncan, SNK ve Tukey testlerinin teorik özellikleri incelenerek temel faraziyelerdeki aksamaların bu testleri nasıl etkileyeceği incelenmiştir. Simulasyon teknikleriyle teşkil edilen deneyler yeter sayıda tekrarlanarak testlerin hata seviyeleri değişik popülasyonlar ve örnek genişlikleri esas alınarak tesbit edilmiştir. Normal şartlar için beklenen hata seviyeleri normal dağılış ve homojen varyans faraziyelerinin bozulduğu çeşitli durumlarda ortaya çıkan hata seviyeleriyle karşılaştırılarak hangi testlerin ne ölçüde etkilendiği incelenmiştir. Böylece, testlerin sağlamlık yönünden genel ve birbirlerine karşı durumları ortaya konmuştur.*

## GİRİŞ

Sabit modelin esas alındığı problemlerde varyans analizinin amacı muamelelerin popülasyon ortalamaları arasındaki sabit iliş-

kileri incelemektir. Bu amaç için genel sıfır hipotezlerine uygulanan F testleri yeterli olmayacağından varyans analizinin çoklu karşılaştırma testleriyle tamamlanması gerekmektedir. Çoklu karşılaştı-

- (1) Prof. Dr. Şaban Karataş, Prof. Dr. Fahrettin Tosun ve Prof. Dr. İbrahim Ak-söz'den müteşekkil jüri tarafından kabul edilen doktora tezinin özetidir.
- (2) Atatürk Üniversitesi Ziraat Fakültesi Zootečni Bölümü Dr. Asistanı.  
Dergi Komisyonuna Geliş Tarihi: 12.12.1974.

ma testleri varyans analizini tanımlayıcı teknikler olduklarından, kullandıkları test istatistikleri F testine de esas teşkil eden değerlerden hesaplanmaktadır. Bu bakımdan, testlerin sıhhatli sonuç verebilmeleri test istatistiklerinin doğru tahminlere dayanmalarıyla mümkündür. Özellikle plân seçimindeki ve matematik modellerle ilgili faraziyelerdeki aksamalar F testlerini olduğu gibi çoklu karşılaştırma testlerini de etkiliyerek araştırmacıyı yanlış sonuçlara götürebilir. Testlerin I. ve II. tip hata seviyeleri farklı olduğundan faraziye aksamalarının test özelliklerini nasıl değiştireceği üzerinde durulması gerekli bir problem olarak ortaya çıkmaktadır.

Bu çalışmanın amacı yaygın bir uygulama alanına sahip bulunan LSD, Duncan, Student—Newman—Keuls ve Tukey çoklu karşılaştırma testlerinin sağlanırlık problemlerini Monté Carlo örnekleme denemelerinden yararlanmak suretiyle ve I. tip hata nisbetlerini esas alarak incelemektir.

### Literatür Özeti

Çoklu karşılaştırmalar problemini sistematik olarak ilk defa Fisher (1935) ele alarak l.s.d. testini geliştirmiştir. Daha sonra Newman (1939) değişim genişliğini bir test kriteri olarak kullanmıştır. Ancak bu konuda asıl gelişmeler Duncan'ın 1947 ve Tukey'in 1949 da başlattıkları ça-

lışmalar sonunda ortaya çıkmıştır.

Ele alınan parametrik çoklu karşılaştırma testleri varyans analizi tablosundaki bazı değerleri kullandıklarından Eisenhart (1947) ın Model-I için belirlediği temel faraziyeler bu testler için de temel faraziye niteliğindedir. Yani çoklu karşılaştırmaların sıhhatli sonuç vermeleri için muamele ve çevre tesirlerinin toplanabilir, hataların bağımsız ve homojen olmaları ve normal dağılış göstermeleri gerekir. Bu faraziyelerden ayrışmalar, test özelliklerini değişik şekillerde etkiler. Cochran (1947) toplanabilirlik faraziyesinin tahmininde etkinlik ve önem varyansın  $\sigma^2$  yerine  $\sigma^2_e + \sigma^2_a$  gibi daha büyük birdeğeri ifade edeceğini ve bu varyansın ortalamalar arasındaki karşılaştırmalarda kullanılmasının hatalara yol açacağını belirtmektedir. Aynı yazar, hata varyansının homojen olmayışının da muamele tesirlerinin tahmininde etkinlik ve önem testlerinde hassasiyet kaybına yol açtığını göstermiştir. Böyle durumlarda, muamele karşılaştırmaları için standart hata grup varyansları yerine, hata kareler ortalamasından hesaplandığından test istatistikleri gerektiğinden daha büyük veya daha küçük olmaktadır. Scheffé (1959) örnek genişliklerinin eşit olması halinde bile gerçek I. tip hata seviyelerinin nominal seviyeden farklı olacağını göstermiştir.

Normal olmayışın testler üzerine etkileri çarpıklık ve basıklık

parametreleri dikkate alınarak incelenmektedir. Cochran (1947), normal olmayışın F ve iki yönlü t testlerinin önem seviyelerine fazla etki yapmadığını, buna karşılık tek yönlü t testlerinin durumdan daha fazla etkileneceğini ifade etmektedir. Hataların bağımsız olmayışı da gruplar ve müşahadeleri etkileyen bazı korelasyonlar sonucu ortaya çıkacağından nominal hata seviyeleri değişecektir Scheffé (1959).

İstatistik testlerde sağlamlık problemi I. ve II. tip hatalarla test güçlükleri esas alınarak incelenmektedir. Konu ile ilgili araştırmalar, Model-I. tipi F testiyle, t testinin faraziye aksaklıklarına karşı fazla hassas olmadıklarını göstermiştir. Scheffé (1959). Ele aldığımız çoklu karşılaştırma testlerinin faraziye aksaklıkları karşısındaki durumları bilinmemekle beraber, Miller (1966) bu testlerin Model-I. tipi F testi kadar sağlam olmadıklarını ancak durumun aşırı sonuçlar doğurmayacağını ifade etmektedir.

Çoklu karşılaştırma testlerinde esas alınan başlıca üç hata nisbeti kriteri yani karşılaştırma başına hata, deneme başına hata ve hatalı karar taşıyan denemelerin nisbeti Hartley (1955), Harter (1957), Federer (1961) tarafından teorik açıdan ele alınmıştır. Aynı nisbetler, çeşitli çoklu karşılaştırma testleri için Boardman ve Moffit (1971) ve Carmer ve Swanson (1971, 1973) tarafından örnekleme denemeleriyle elde edilmiştir. Testlerin hata ve

koruma seviyeleriyle, test güçleri de Duncan (1951, 1955) ve Lehmann (1957) tarafından incelenmiştir. Bu çalışmalar, çoklu karşılaştırma testlerinin farklı özelliklere sahip olduklarını hiçbir testin güç, hata seviyesi, koruma gibi kriterler yönünden optimum özellikleri taşımadığını göstermektedir.

## Materyal ve Metod

Çalışmamızda farklı yapılara sahip dağılışlardan örnekleme denemeleri teşkilinde simulasyon tekniklerinden faydalanılmıştır. Şans sayılarının elde edilmesinde, bilgisayarın kapasitesini de dikkate alan ve geçerliliği gerekli testlerle kontrol edilen özel bir program kullanılmıştır. Şans sayılarından normal dağılışa geçişte merkezi limit teoremi esas alınmıştır. n tane bağımsız uniform şans değişkeninin toplamının standardize edilmesiyle elde edilen,

$$X_n = \frac{U - n E(U_i)}{\sqrt{n \cdot \text{var}(U_i)}}$$

değişkenleri standart normal dağılış göstereceğinden, n = 12 alınarak,

$$X_n = U - 6$$

olmak üzere normal değişkenler bulunmuştur. Değişik varyanslara sahip populasyonlardan örnekleme yapmak için de,

$$\text{var}(X \sqrt{c}) = c \sigma^2$$

olduğundan varyansı c olan bir populasyondan gelen değerler,

$$X'_n = (U - 6) \sqrt{c}$$

ifadesinden hesaplanmıştır.

Poisson Dağılışı'na geçişte n tane üniform şans değişkeninin çarpımının ihtimal fonksiyonundan faydalanılmıştır.

$$\prod_{i=1}^n U_i < e^{-L}$$

eşitsizliğini sağlayan ilk n sayısı parametresi L olan bir poisson popülasyonundan gelen müşahede değeri olarak alınmıştır. Bu metodda farklı poisson popülasyonları tarif etmek için L yi değiştirmek yeterlidir.

Uniform dağılıştan gelen şans sayıları yukarıda belirtildiği şekilde elde edildikten ve diğer dağılışlara çevrildikten sonra,

$y_{ij} = \mu + d_i + e_{ij}$  matematik modeline uygun müşahedeler teşkil edilmiştir. Değişik grup sayıları ve örnek genişliklerine sahip denemelerde F testleri ve çoklu karşılaştırma testleri uygulanarak hata nisbetleri her durum için 1000 deneme üzerinden üç ayrı kritere göre tesbit edilmiştir. Normal durumlar için elde edilen hata seviyeleri normalliğin ve varyans homojenliğinin bozulduğu durumlarda test özelliklerini incelemeye esas alınmıştır.

## Araştırma Sonuçları

Temel faraziyelerin işlediği ve işlemediği çeşitli durumlarda testlerin hata seviyeleri incelenerek aşağıdaki sonuçlara ulaşılmıştır.

1. Çarpıklık ve basıklık parametreleri birden küçük olmak şartıyla normal dağılış faraziyesinin bozulması, değişim genişliğine dayanan Duncan, SNK ve Tukey testlerinde hatalı karar taşıyan denemelerin nispetine etki yapmaktadır. Ancak adı geçen testlerin karşılaştırma başına hata nispetleri normalliğin bozulmasıyla % 20 - % 30 oranında yükselmektedir. Çarpıklık ve basıklığın artışının bu etkileri daha da hızlandıracağı anlaşılmaktadır. Gruplardaki müşahede sayılarının artışı normal olmayışın etkilerini azaltmaktadır.

2. Varyans heterojenliği ele aldığımız şartlar altında hata nispetlerini normal dağılış faraziyesinin bozulmasından daha fazla etkilemektedir. Grup varyanslarının aşırı derecede farklı, örnek genişliklerinin yetersiz olduğu durumlarda özellikle Duncan, SNK ve Tukey testlerinin varyans heterojenliğine karşı sağlam olmadıkları görülmüştür. Müshade sayılarının artışı hernekadar hata seviyelerini düşürücü etki yapıyorsa da bu etkiler sınırlı kalmakta ve heterojen varyansın sebep olduğu hata artışları önemini korumaktadır.

3. Heterojen varyans testleri üzerinde hassasiyetleriyle ters

yönde etkiler yapmaktadır. Yani durumdan hassas testler daha az katı testler daha fazla etkilenmektedir.

4. Normal olmayış ve varyans heterojenliğiyle ilgili olarak ele aldığımız şartlar çerçevesinde çoklu karşılaştırma testleri faraziye aksaklıklarından F testini aşan ölçülerde etkilenmemektedir. Yani bu testler faraziyelerin bozulduğu hallerde en aşağı Model I. tipi F testi kadar sağlamdır.

Ulaşılan sonuçlar aşırı olmayan faraziye aksaklıklarının, eğer örnek genişlikleri de yeterli ise, çoklu karşılaştırma testlerinin özelliklerini fazla değiştirmeyeceğini ortaya koymaktadır.

## SUMMARY

A Research on the Effects of Nonnormality and Variance Heterogeneity on the Structure of Multiple Comparison Tests.

When the fixed model is used, the aim of the analysis of variance is to determine the differences between the population means. As the F test used for general null hypotheses is not sufficient, the analysis of variance has got to be completed by the multiple comparison tests. In order to detect the differences among the population means, some parametric and non parametric test procedures have been developed. In this research some properties of the parametric tests, such as L.S. d., Duncan, SNK and Tukey, were

investigated by using sampling experiments.

Multiple comparison tests are the main components of the analysis of variance. For this reason, the validity of the comparisons depends on the accuracy of the knowledge that is essential for the analysis of variance. Therefore the assumptions of the analysis of variance can also be extended for the multiple comparison tests. The occurrence of some departures from the underlying assumptions invalidates the multiple comparison tests as the F tests, and causes wrong decisions. The error rate and theoretical structure of multiple comparison procedures are different from each other. Therefore the levels of type I and type II errors and their powers are not the same. Because of the differences in the error rates, the decisions and the results obtained by the use of these tests are different.

In this research the sampling experiments obtained by using Monté-Carlo methods. The error rates of the multiple comparison tests were investigated under different conditions. The error rates obtained for the normal case have been taken as a criterion to compare the other cases where some departures from the underlying assumptions.

Taking two poisson distribution with the parameters  $L = 2$  and  $L = 3$ , the departures from the normality were considered.

Taking the samples of 5 groups, each with sample sizes of 3 and 7, type I error rates were estimated according to various error rate criteria. With the departures from the normality the error rates of the Duncan, SNK, and Tukey's methods were increased. This practically arises when the sample sizes are small.

Effects of the heterogeneity of the variances were also investigated for various combinations of the group variances. On each sampling, 4 and 5 groups each having 3 and 7 observations were chosen, Applying the analysis of variance and the multiple comparison tests, type I error rates were calculated it has been revealed that high heterogeneity of variance ended up with the increase of type I error rates of multiple comparison tests except L.S.D.. In the case of low heterogeneity, however, did not prove accordingly giving slightly decreasing error rates.

## LITERATURE

Boardman, T.J. and Moffit, D.R. 1971. Graphical Monté-Carlo Type I Error Rates for Multiple Comparison Procedures. *Biometrics*; 27, 738.

Carmer, S.G. and Swanson, M.R. 1973. Detection of Differences Between Means: A Monté-Carlo Study of Five Pairwise Multiple Comparison Proce-

dures. *Agronomy Journal*; 63, 940.

Carmer, S.G. and Swanson, M.R. 1973. An Evaluation of Ten Pairwise Multiple Comparison Procedures by Monté Carlo Methods. *Journal of American Stat. Assoc.*; 68, 66.

Cochran, W.G. 1947. Some Consequences When the Assumptions for the Analysis of Variance are not Satisfied. *Biometrics*; 3, 24.

Duncan, D.B. 1951. A Significance Test for Differences Between Ranked Treatments in an Analysis of Variance. *The Virginia Journal of Science*; 2, 171.

Duncan, D.B. 1955. Multiple Range and Multiple F Tests. *Biometrics*; 11, 1.

Eisenhart, C. 1947. The Assumptions Underlying the Analysis of Variance. *Biometrics*; 3, 1.

Federer, W.T. 1961. Experimental Error Rates. *Proc. Amer. Soc. Hort. Sci.*; 78, 605.

Fisher, R.A. 1935: *The Design of Experiments*. Oliver and Boyd, Edinburgh.

Harter, H.L. 1957. Error Rates and Sample Sizes for Range Tests in Multiple Comparisons. *Biometrics*; 13, 511.

Hartley, H.O. 1955. Some Recent Developments in Analysis of Variance. *Comm. Pure App. Math.*; 8, 47.

Lehmann, E.L. 1957. A Theory of Some Multiple Decision Problems, I. *Annals of Math. Stat.*; 28, 1.

Miller, R.G., Jr. 1966. *Simultaneous Statistical Inference*. McGraw-Hill Book Company, New York.

Newman, D. 1939. The Distribution of the Range in Samples

From a Normal Population, Expressed in Terms of an Independent Estimate of Standard Deviation. *Biometrika*; 31, 20.

Scheffé, H. 1959. *The Analysis of Variance*. John Wiley and Sons, Inc. New York.