

# VARYANS ANALİZİNİN FARAZİYELERİ VE BUNLARIN BOZULMASINDAN DOĞAN DURUMLAR(1)

Fatin SEZGİN(2)

## ÖZET

*Varyans analizi, müşahedelerin şans değişkenlerinin bir denemede gerçekleşmiş değerleri oluşu, tesiri kontrol edilebilen faktörler ve deneme hatasının toplanabilirliği, şans unsurlarının bağımsızlığı, eşit varyanslılığı ve normal dağılış göstermeleri faraziyelerine dayanır. Bunlardan biri veya birkaçının bozulması, sonuçların güvenirliliğini bozar. Durum, bazı müşahedelerin ihmali, hata varyansının parçalanması, serbestlik derecelerinde bir takım düzeltmeler, bazı özel analiz metodları yahut sağlam testlerin kullanılışı ve müşahedelerin transformasyonu ile düzeltilebilir. Denememizde çarpımlı bir modele sahip lognormal dağılış ve yanlış model seçimi ile ortaya çıkan durum, normallik faraziyesine uymayan Poisson dağılışı gibi faraziyelerdeki şartları taşımayan durumlar ele alınmış ortaya çıkan aksaklıklara ve düzeltici tedbirlere işaret edilmiştir.*

## Giriş

Varyans analizi çok yaygın kullanılış alanı bulan bir methoddur. Uygulamada bazı faraziyelerin ne dereceye kadar tuttuğu, bunun sonucu olarak da varılan hükümlerin hangi sınırlar arasında güvenilir olduğu üzerinde düşünölmeye değer bir konudur. Çalışmamızda varyans analizinin dayanakları ve bunların bozulmasının sonuçları belirtilmekte, durum Monté-Carlo si-

mülasyonlarıyla elde edilen sonuçlarla açıklığa kavuşturulmaktadır.

## Literatür Özeti

Fisher (1925) genel bir ortalama etrafındaki varyasyonu birtakım alt varyasyonlara parçalamıştır. Meselâ tam şansa bağılı deneme planında Genel K.T. = M.A.K.T. + Hata K.T. şeklinde ifade edilebilir. Buradan hesaplanacak kareler ortalamalarının yar-dımı ile muamelelerin ortalamaları veya

(1) Prof. Dr. Şaban Karataş, Prof. Dr. Orhan Düzgüneş ve Prof. Dr. Fahrettin Tosun'dan müteşekkil jüri tarafından kabul edilen doktora tezinin özetidir.

(2) Atatürk Üniversitesi Ziraat Fakültesi, Zootečni Bölümü Dr. Asistanı.

varyanslarına ait hipotezler kontrol edilebilir.

Eisenhart (1947) varyans analizinin başlıca iki gaye ile kullanıldığını belirtmektedir.

Model - I. Sabit Model: Muamele ortalamalarının sadece denemeye alınanları temsil ettiği, yani muamele popülasyonumuzun sadece denemeye girenlerden ibaret olduğu farzedilmektedir. Sonuçlar da sadece bu haller için yorumlanacak ve denenmeyen durumlara teşmil edilmeyecektir.

Model- II. Şansa bağlı model: Burada problem bileşik bir popülasyonda şansa bağlı varyans unsurlarının araştırılması ve tahminidir. Eldeki fertler türlü etkilerden ötürü varyasyon göstermektedirler. Gaye her kaynağın toplam varyasyondaki hissesini teşhis ve bunun önemini tesbit etmektir.

Bazen denemede faktörlerden bir kısmının sabit değerlerinin şansa bağlı seçilmesi haliyle karşılaşılabılır. O zaman karışık modeller ortaya çıkmış olur.

Faraziyeler her iki durum için de kısaca şöyle sıralanmaktadır: İki yönlü bir sınıflamada,

1)  $x_{ij}$  müşahedeleri bir şans değişkeni olan  $X_{ij}$  nin bir deneme sonunda gerçekleşmiş değerleridir. Deneme sonu-  
suz defa tekrarlınsaydı,  $x_{ij}$  değerlerinin bir  $\mu_{ij}$  ortalaması etrafında bir ihtimal fonksiyonuna bağlı olarak dağıldıklarını görecektik.

2) Toplanabilirlik faraziyesi:  $\mu_{ij}$  parametreleri,  $\mu_i$  ve  $\mu_j$  marjinal ortalamalarıyla genel ortalama  $\mu$  nün  $\mu_{ij} = \mu + (\mu_i - \mu) + (\mu_j - \mu)$  şeklinde toplanabilir bir fonksiyonudur.

3)  $X_{ij}$  şans değişkenleri eşit varyansa sahiptirler. Ayrıca

$$\text{Kov. } (X_{ij}, X_{i'j'}) = 0, (i'j' \neq ij)$$

dir. Eğer model birden fazla şans değişkeni ihtiva ediyorsa bunların da korelasyonsuz olmaları arzu edilir.

4)  $X_{ij}$  yi teşkil eden şans unsurları normal dağılışa sahiptirler. Böylece müşahedelerin linear fonksiyonlarına ait varyansları hesaplamak imkânı belirlenmektedir. Sıralar veya sütunlar arası kareler ortalamasının hata varyansına oranı iki  $\chi^2$  dağılışının oranı olduğundan F dağılışına sahip olacaktır. Bu faraziyelerin bozulması halinde meydana gelen durumu başlıca şu noktalarda toplayabiliriz :

1) Toplanabilirlik bozulunca II. tip hata artmaktadır (Eisenhart, 1947) Cochran (1947) bu bozukluğun varyans homojenitesini de etkilediğini ve bilgi kaybına yol açtığını belirtmiştir. Tukey (1949) ise bu tesiri elimine etmenin bir yolunu belirtmiştir.

2) Hata varyansının heterojenliği: Cochran (1947) bunun tahminde etkinlik kaybına sebep olduğunu belirtmiştir. Box (1954) I. tip hatanın yükseldiğini, Horsnell (1953) F testinin gücünün düştüğünü ortaya koymuştur.

3) Normallığın bozulması: Box ve Anderson (1955) F testinde I. tip hatanın arttığını belirtmekte, Scheffé (1959) şansa bağlı modelde basıklığın önemli II. tip hataya yol açtığını göstermektedir.

4) Bağımlılık: Unsurlar arasındaki korelasyonlar Cochran (1947) ve Fisher (1925) tarafından ele alınmıştır. Sonuç ortalamaların varyanslarının gerçekte olduğundan küçük veya büyük olarak hesaplanmasıdır.

Standart varyans analizi metodlarını geçerli kılmak ve sonuçları düzeltmek için şu tedbirler alınabilir :

- 1) Çok belirgin sapmalar gösteren müşahedeler atılır.
- 2) Hata varyansı kendi içinde homojen olan kısımlara parçalanır.
- 3) Serbestlik derecelerinde yapılan ayarlamalarla yüksek varyanslı gruba daha fazla müşahede tahsis edilir.
- 4) Toplanamazlık için 1 s.v. ayrılır.
- 5) Hatalar arasındaki korelasyonlar kovaryans analizi ile artırılır.
- 6) Faraziye bozukluğundan etkilenmeyen sağlam metodlar kullanılır.
- 7) Müşahedelerdeki özelliklere göre arzulanan yönde düzeltmeler sağlamak için transformasyonlara başvurulur. Transformasyonla linearite, normallik veya varyans homojenliği sağlanır.

### Materyal ve Metod

Çalışmamızda üniform şans rakamları computer yardımıyla elde edilmiş, merkezi limit teoreminden faydalanılarak bunların toplamından normal dağılışa sahip şans değişkenlerine varılmıştır. Poisson dağılışı için  $u_0, u_1, \dots, u_n$  gibi üniform dağılıştan gelen  $n+1$  değişkenin çarpımına ait ihtimal fonksiyonunun

$$f_n(x) = \frac{(-\ln x)^n}{n}$$

olmasından faydalanılmış,  $X < e^{-T}$  olduğu andaki  $n$  değeri müşahedemizi teşkil etmiştir. Lognormal dağılıştaki ise  $w_{ij} \sim N(0, 1)$  olmak üzere

$$y_{ij} = \alpha_i e^{w_{ij}}$$

bağıntısı kullanılmıştır.

1) Unsurların toplanabilir olmayışı: Bu maksatla, çarpımlı bir model ele alınmıştır. Elde edilen rakamlar lognormal dağılışı göstermektedirler. Ortalamalar arasında Hartley çoklu mukayeseleri büyük mutlak değere sahip ortalamalar arasında gereğinden fazla önemli fark vermekte, küçük ortalamalar arasındaki farkı ise belirtemektedir. Toplanamazlık, burada normallik ve hataların homojenitesi faraziyelerinin bozukluğu ile bir arada bulunmaktadır. Logaritmik transformasyonun düzeltici tesiri görülmüştür.

2) Normallik faraziyesinin bozulması : Müşahedelerin Poisson dağılışıya sahip olduğu bir tam şansa bağlı deneme örneklenmiştir. Lognormal dağılıştaki olduğu gibi burada da muamele ortalamaları arasındaki çoklu mukayeseler araştırıcıyı yanlış hükme yöneltmektedir. Bu Poisson dağılışı için her iki tip hatanın arttığını gösterir.

3) Hatalı bir modelin seçilmesi:  $11 \times 11$  bir Latin kare denemesi ele alınmıştır. Doğru model yanında, sıraların veya sütunların ihmali ile ortaya çıkan bloklar ve her ikisinin ihmali ile elde edilen tam şansa bağlı deneme planına göre analizler yapılmıştır.

Bazı varyansyon kaynaklarının ihmali hatayı büyütmede, sonuç olarak bilgi kaybına yol açmaktadır. Yanlış planın kullanılması ile aslında önemli olan bazı farklar önemsiz gibi görülmüştür. Uygulamada aksama daha da ciddi olabilir. Çünkü deneme planındaki bir indirgemede ihmal edilen faktörün geriye kalanlara ortogonal düşmesi şansı küçüktür.

*THE ASSUMPTIONS UNDERLYING  
THE ANALYSIS OF VARIANCE  
AND CONSEQUENCES OF DE-  
PARTURES FROM THEM.*

The purpose of this study was to describe the assumptions underlying the analysis of variance and indicate the effects of departures from these assumptions. Some methods to improve the analysis were mentioned.

The analysis of variance is used mainly to estimate fixed effects of treatments (Model-I) or to estimate some components of variance in a composite population (Model-II). In either case, the assumptions required for validity of inferences are as follows:

- 1) The observations must represent a random variable,
- 2) The treatment effects, the environmental effects and experimental errors must be additive,
- 3) The experimental errors must be independent and have equal variances,
- 4) The observations must be normally distributed.

The consequences of departures from these assumptions invalidate the final results obtained by an analysis of variance. The effects depend on the special assumptions and to the extent to which these are violated.

Some commonly used methods to make the analysis more valid in the case of violation of assumptions are :

- 1- Omission of some observations,
- 2- Subdivision of the error variance,
- 3- Some adjustments in the degrees of freedom,

4- Use of some refined techniques taking into account the violations (e.g. assignment of one degree of freedom for nonadditivity, elimination of the correlation between errors by an analysis of covariance),

- 5- Application of robust procedures,
- 6- Change of scale.

In order to demonstrate the effects of departures from the assumptions, three Monte-Carlo sampling experiments were performed. The results are as follows :

1) Nonadditivity of components: For this purpose a multiplicative model is adopted. The resulting data are coming from a Lognormal distribution. The multiple comparison of treatment means gives too many significant results for large means, whereas the significant differences between small valued means are masked. Here, nonadditivity is accompanied by the heterogeneity of errors and nonnormality. By taking logarithms one can improve the results.

2) Nonnormality: A completely randomised experiment is simulated, the observations having a Poisson distribution. As in the Lognormal distribution the multiple comparison between treatment means gives misleading results in raw data. This implies that both types of error are increased in the case of Poisson distribution.

3) Choosing a wrong model: The data are simulated from an 11x11 Latin square experiment and analysed by using the correct model or randomised blocks by choosing rows and columns as blocks, and by completely randomised design. The omission of some sources of variation leads to exaggeration of error

variance, ultimately causes loss of information. The significance of some differences of means are lost by using a wrong design. In practice the violation may

be more serious, because in reducing the experimental design, there is a great probability that the factor omitted will fall nonorthogonal to the remainings.

#### Literatür

Box, G.E.P., 1954. Some Theorems on Quadratic Forms Applied in the study of Analysis of Variance Problems: I. Effect of Inequality of Variance in the One-Way Classification, *Annals Math. Stat.*, Vol. 25, s. 290.

Box, G.E.P. and S.L. Andersen, 1955. Permutation Theory in the Derivation of Robust Criteria and the Study of Departures from Assumption, *J. Roy. Stat. Soc., Series B*, Vol. 17, s. 1

Cochran, W.G., 1947. Some Consequences when the Assumptions for the Analysis of Variance are not Satisfied, *Biometrics*, Vol, 3, No. 1, s. 22.

Eisenhart, C., 1947. The Assumptions Underlying Analysis of Variance, *Biometrics*, Vol. 3, No. 1, s. 1.

Fisher, R.A., 1925. *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.

Horsnell, G., 1953. The Effect of Unequal Group Variances on the F-test for the Homogeneity of Group Means, *Biometrika*, Vol. 40 s. 128.

Scheffé, H., 1959. *The Analysis of Variance* John Wiley and Sons, Inc., s. 225, 331.

Tukey, J.W., 1949. One Degree of Freedom for Non-Additivity, *Biometrics*, Vol. 5, s. 232.