Uludağ Üniversitesi
İktisadi ve İdari Bilimler Fakültesi Dergisi
Uludağ Journal of Economy and Society
Cilt/Vol. XXXIV, Sayı/No. 1, 2015, pp. 235-242

VARYANS BİLEŞENLERİNİN NEGATİF KESTİRİLME OLASILIKLARI İÇİN BİR SİMÜLASYON ARAŞTIRMASI*

***Serdar KURT*[1]**
***Ahmet KAYA*[2]**

Özet

Varyans analizi, birçok alana uyarlanan bir istatistik analiz yöntemidir. Bu yöntem, ürün kalitesi geliştirmede, gıda kalite kontrol iyileştirmelerinde, tohum, bitki ve hayvan ıslahı araştırmalarında, optimizasyon çalışmalarında ve daha birçok alanda etkili ve verimli bir test etme aracıdır. Bu aracın en temel unsuru, varyans bileşenlerinin öngörümü için sıklıkla kullanılan yöntemlerden biri olan, ANOVA kestiricileri olarak bilinen ve kestirimleri VA (Varyans Analizi) tablosundaki kareler ortalamalarının beklenen değerlerinden hesaplayan yöntemdir. Varyans çözümlemesi için yapılan tüm varsayımlar geçerli olsa bile bu yöntem ile negatif varyans bileşeni kestirimi elde etme olasılığı bulunmaktadır. Bu çalışmada dengeli tek etkenli tam rastgele deney için $P(\hat{\sigma}_{\tau}^{2} < 0)$ *olarak tanımlanabilen bu olasılığın, her deneme ile yapılan tekrar sayısına (n) ve* σ^{2} *'ye bağlı olarak nasıl değiştiği, yapılan varsayımların sağlandığı bir kitleden çekilen örneklemler ile yapılacak bir simülasyon çalışması ile araştırılmaktadır.*

Anahtar Kelimeler: *Simülasyon, Negatif varyans bileşeni, Tahminleme.*

---

* Bu çalışma Çukurova Üniversitesi Ulusal Ekonometri ve İstatistik Sempozyumunda bildiri olarak sunulmuştur.

1 *Prof.Dr., Dokuz Eylül Üniversitesi, Fen Edebiyat Fakültesi İstatistik Bölümü, serdar.kurt@deu.edu.tr, Buca-İZMİR.*

2 *Doç.Dr., Ege Üniversitesi, Tire Kutsan Meslek Yüksekokulu, ahmet.kaya@ege.edu.tr, Tire-İZMİR.*

A Simulation Research For Negative Estimation Probabilities Of Variance Components

Abstract

Variance analysis (VA), which can be adaptive on many research areas, is an important statistical analysis method. This method is an effective and productive to put to test appliance for product quality improvement, food quality control enrichment, seed, plant, and animal culture refining, optimization, and other many areas. The best factor for VA tool is, known as ANOVA approximations which calculated from expected value of mean squares tables, a method the prediction of components of variance. Even all assumption viable for variance analysis, it is possible to take negative variance for variance components. In this study, the probability for experiment model for balanced unique factor exactly randomly defined as in the form of $P(\sigma_{\tau}^{2} < 0)$ *is depend on* n, *the number of the repetition, and* σ^2. *For that reason a simulation experiment will be planned and results about parameters have been investigating.*

Key Words: *Simulation, Negative variance component, Estimation.*

1. GİRİŞ

Varyans, istatistikte merkezi dağılım ya da dağılış ölçülerinin en çok bilineni ve kullanılanıdır. Değişimin ölçüsü olduğu için de bu değer olarak en az sıfır değeri alır. Dolaysıyla varyansın sıfırdan küçük bir değer alması beklenmez. Bu durumda varyansı meydana getiren bileşenlerin de sıfırdan küçük olmaması gerekir. Halbuki araştırmanın konusu; varyans bileşenlerinin negatif tahminlenme olasılığıdır. Bunun nedeni, varyans analizi çalışmalarında deneme desenlerinin yani çözümlemeye esas model deseninin yanlış kurgulanmasından kaynaklanmaktadır. Desenlerin ve etken düzeylerinin hatalı planlanması sonucunda pozitif sınırlar içinde tahminlenmesi gereken değişim ölçüleri, negatif olarak tahminlenebilmektedir. Varyans bileşenlerinin negatif tahminlenmesi durumunda model için öngörülen varsayımlar sağlanamaz, ayrıca elde edilen sonuçların güvenilirliği de tamamen ortadan kalkar.

Varyans bileşenleri çözümlemesi Eisenhart'ın 1947, Crum'un 1946 ve 1951 yıllarındaki çalışmaları ile başlamıştır *(Hendersen, 1953)*.

Daha sonra Henderson 1953 yılında yaptığı çalışmada kendi adıyla anılan Henderson 1, Henderson 2 ve Henderson 3 yöntemlerini önermiştir. Hartley ve Rao'nun 1967 yılındaki çalışmalarında en çok olabilirlik (Maximum Likelihood) yöntemini kullanarak varyans bileşenlerini benzetim yolu ile tahmin etmişlerdir. Daha sonra kısıtlanmış en çok olabilirlik (Restricted Maximum Likelihood) yöntemi olarak isimlendirilen, varyans bileşenlerini tahmin yöntemi ile ilgili pek çok araştırmacı çalışmalar yapmıştır *(Bek, 1992).*

Rao'nun 1970, 1971 ve 1972 yıllarında yaptığı çalışmalarda MINQUE (Minimum Norm Quadratic Unbiased Estimator) yöntemi önerilmiştir. Günümüze kadar önerilen bu yöntemlerin çeşitli özellikleri açısından karşılaştırılmaları ve deney düzenlerine uygulanmaları ile ilgili pek çok araştırma yapılmıştır.

Bu çalışmalardan farklı olarak, bu makalede dengeli tek etkenli tam rastgele deneye ilişkin model denklemi temel alınarak bir simülasyon çalışması planlanmıştır.

Buna göre; dengeli, tek etkenli tam rastgele bir deneye ilişkin model denklemi,

$$y_{ij} = \mu + \tau_j + \varepsilon_{ij} \qquad j = 1,2,...,k \qquad i = 1,2,...,n$$

Olarak verilebilir. Burada k deneme sayısı, n her deneme için yapılan tekrar sayısıdır. Buna göre, k sayıda etken düzeyi tüm olası deneyler arasından tamamen rastgele olarak seçilmiştir. Etken düzeyleri rastgele seçildiğinden model denklemi, rastgele etkili model olarak isimlendirilmektedir.

Bu model için τ_j ve ε_{ij} birer rastlantı değişkeni $\tau_j \approx N(0,\sigma_\tau^2)$ ve $y_{ij} \approx N(0,\sigma^2)$ ya da $y_{ij} \approx N(\mu_j,\sigma^2)$ $j = 1,2,...,k$, $i = 1,2,...,n$ varsayımlarının sağlanmakta olduğu kabul edilir. Burada σ_τ^2 ve σ^2 varyans bileşenleri olarak bilinir. Herhangi bir gözlemin varyansı, varyans bileşenlerinin toplamı olarak;

$Var(y_{ij}) = \sigma_\tau^2 + \sigma^2$ Biçiminde ifade edilir *(Montgomery, 2000).*

Bu tür bir deneyde araştırmacının temel amacı varyans bileşenlerinin kestirimlerini bulmaktır. Böylece araştırıcı bağımlı değişken üzerinde etkide bulunan çeşitli değişimlerin kaynaklarının toplam değişim üzerindeki ağırlıklarını görebilecektir.

Varyans bileşenlerinin varyans çözümleme tahmin edicileri,

$\hat{\sigma}^2 = HataKO$ ve $\hat{\sigma}_\tau^2 = \dfrac{DenemeKO - HataKO}{n}$ eşitlikleri ile elde edilir.

Bu tahmin ediciler minimum varyanslı ve yansız tahmin edicilerdir *(Lindman, 1991).*

Varyans analizi yöntemi ile bulunan varyans bileşenlerinden σ_τ^2 nin uygulamada bazen negatif olarak tahmin edilmesi mümkün olabilmektedir. Bu durumda akla gelen ilk şey, hesaplama hatası yapılmış olabileceğidir. Hesaplama hatası yapılmadığı konusunda emin olunduğunda ise model için yapılan varsayımların geçerli olup olmadıkları düşünülmektedir. Bu konuda da emin olunabilirse, kuramsal zorluklar nedeni ile negatif olan varyans bileşeninin sıfır olarak varsayılması yoluna gidilmektedir. Aksi halde, diğer varyans bileşeninin pozitif etkisini aşağı geçen bir durum söz konusu olabilecektir.

Model için yapılan tüm varsayımlar geçerli olsa bile σ_τ^2 nin negatif olasılıkları üzerinde Verdooren (1982) tarafından yapılan bir çalışmada bu olasılığın,

$$P(\sigma_\tau^2 < 0) = P\left[F_{1-1,1(n-1)} < \frac{1}{1 + n\dfrac{\sigma_\tau^2}{\sigma^2}} \right]$$

Eşitliği ile elde edilebileceği gösterilmiştir. Buna göre $P(\sigma_\tau^2 < 0)$ olasılığının deneme sayısı k ’ya, her deneme ile yapılan tekrar sayısı n ’e ve σ_τ^2 / σ^2 oranına bağlıdır *(Verdooren, 1982).*

Bu olasılık çeşitli k ve n değerleri ile σ_τ^2 / σ^2 oranı kullanılarak hesaplanmış ve tablolar haline getirilmiştir. Buna göre, $k = 5$ ve $n = 2, 3, 4, 5$ olduğunda σ_τ^2 / σ^2 oranı 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5 ve 4.0 olduğunda elde edilen olasılıklar Tablo-1’de verilmektedir *(Kurt ve Ege, 1998:4).*

Tablo 1: 5 Denemeli, Tek Etkenli Tam Rastgele Düzen için Elde Edilen Varyans Bileşenlerinin Negatif Elde Edilme Olasılıkları.

Oran	k=5			
	n=2	n=3	n=4	n=5
0.5	.2627	.1950	.1485	.1162
1.0	.1518	.0959	.0653	.0472
1.5	.0982	.0566	.0364	.0253
2.0	.0686	.0372	.0232	.0158
2.5	.0506	.0263	.0160	.0107
3.0	.0388	.0196	.0117	.0078
3.5	.0307	.0152	.0090	.0059
4.0	.0249	.0121	.0071	.0046

Bu tabloya göre, n ve $\sigma_{\tau}^{2}/\sigma^{2}$ ne olursa olsun $P(\sigma_{\tau}^{2}<0)$ olasılığının sıfırdan büyük olduğu ve bu olasılığın büyüyen n ve $\sigma_{\tau}^{2}/\sigma^{2}$ değerleri için küçüldüğü görülmektedir.

2. ÇALIŞMA

Burada istatistiksel model denklemi için yapılan tüm varsayımlar geçerli olsa bile $P(\sigma_{\tau}^{2}<0)$ olasılığının bulunması için bir simülasyon çalışması yapılmakta ve elde edilen sonuçlar Tablo-1'de bulunan kuramsal sonuçlarla karşılaştırılmaktadır. Simülasyon çalışması $k=5$ için $n=2, 3, 4, 5$ olduğunda $\sigma_{\tau}^{2}/\sigma^{2}$ oranı 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0 için elde edilmektedir. Simülasyon uygulamasında, sonuçlar 100.000 tekrarlı olarak planlanmış ve bu işlemler için MINITAB paket programı kullanılmıştır.

Simülasyon çalışması üç aşamada gerçekleştirilmiştir. Önce model varsayımlarını sağlayan kitle oluşturulmakta, sonra bu kitleden rastgele çekilen örneklemlerle varyans çözümlemesi yapılmakta ve nihayet varyans bileşenlerinin kestirimi elde edilerek $P(\sigma_{\tau}^{2}<0)$ olasılığı hesaplanmaktadır.

2.1. Birinci Aşama Simülasyon

$\sigma_{\tau}^{2}=144$ ve $\tau_{j}\approx N(0,\sigma_{\tau}^{2})$ varsayımlarını sağlayacak $y_{ij}\approx N(\mu_{j}, 288)$ olan her birinde 1000 gözlem bulunan 50 adet düzey için kitlenin oluşturulması:

```
1-a) Dosya: BEN1.MTB
SET C51
1:50
END
RANDOM 50 C100;
NORMAL 0,12.
LET C100=C100-MEAN(C100)
LET K11=0
LET K25=SQRT(288)
EXEC 'BEN10' 50
LET K11=0
EXEC 'BEN11' 50
```

1-b) Dosya: BEN10.MTB
```
LET K11=K11+1
LET K10=C100(K11)
RANDOM 1000 CK11;
NORMAL K10 K25.
```
1-c) Dosya: BEN11.MTB
```
LET K11=K11+1
LET K10=MEAN(CK11)
LET CK11=CK11-K10+C100(K11)
```

2.2. İkinci Aşama Simülasyon

Her defasında 50 olası düzey arasından rastgele 5 düzey seçmek ve bu düzeylerden n=5 olan rastgele örneklem çekerek varyans çözümlemesi yapmak ve $\hat{\sigma}^2$ ve $\hat{\sigma}_\tau^2$ değerlerini bulmak:

2-a) Dosya: BEN2.MTB
```
LET K1=5
LET K2=5
LET K20=0
EXEC ‘BEN3’ 100000
```
2-b) Dosya: BEN3.MTB
```
SAMPLE K1 C51 C52
SORT C52 C52
LET K11=C52(1)
LET K12=C52(2)
LET K13=C52(3)
LET K14=C52(4)
LET K15=C52(5)
SAMPLE K2 CK11 C101
SAMPLE K2 CK12 C102
SAMPLE K2 CK13 C103
SAMPLE K2 CK14 C104
SAMPLE K2 CK15 C105
STACK C101-C105 C55;
SUBACRIPTS C54.
ANOVA C55=C54;
RESIDUALS C56.
LET K10=SSQ(C56)
LET K13=(K1*K2-1)-(K1-1)
```

```
LET K30=K10/K13
LET K11=SSQ(C55)-SUM(C55))**2/(K1*K2)
LET K12=K11-K10
LET K31=K12/(K1-1)
LET K41=(K31-K30)/K2
LET K40=K30
LET C60(K20)=K40
LET C61(K20)=K41
PRINT K20
```

2.3. Üçüncü Aşama Simülasyon

$P(\sigma_\tau^2 < 0)$ olasılığını hesaplamak:

3-a) Dosya: BEN4.MTB

```
SORT C61 C60 C61 C60
LET C62=(C61>0)
LET K25=(100000-SUM(C62))/100000
PRINT K25
SAVE 'A55'
```

Burada $\hat{\sigma}^2$ ve $\hat{\sigma}_\tau^2$ kestirimleri C60 ve C61 kolonlarına yazılmakta, sonra C61 kolonuna göre küçükten büyüğe doğru sıralanarak bir isim altında saklanmaktadır. $n = 5$ ve $\sigma_\tau^2 / \sigma^2 = 0.5$ olduğunda elde edilen simülasyon sonuçları 'A55.MTW' isimli dosyaya yazılmaktadır. Bu şekilde hesaplanan simülasyon sonuçları Tablo-2'de verilmektedir. Bu sonuçlara göre de $P(\sigma_\tau^2 < 0)$ olasılığının her deneme ile yapılan tekrar sayısına (n) ve σ_τ^2 / σ^2 oranına bağlı olduğu görülmektedir.

Tablo 2: 5 Denemeli, $\sigma_\tau^2 = 144$ için Tek Etkenli Tam Rastgele Düzen için Elde Edilen Varyans Bileşenlerinin Negatif Elde Edilme Olasılıkları

Oran	σ^2	k=5			
		n=2	n=3	n=4	n=5
0.5	288	.2500	.1818	.1335	.1021
1.0	144	.1405	.0834	.0515	.0339
1.5	96	.0884	.0449	.0269	.0182
2.0	72	.0597	.0279	.0161	.0101
2.5	57	.0418	.0195	.0102	.0061
3.0	48	.0312	.0138	.0077	.0046
3.5	41	.0235	.0106	.0057	.0032
4.0	36	.0189	.0078	.0039	.0027

3. SONUÇ

Kuramsal olarak elde edilen olasılıklar ile simülasyon çalışması sonucunda elde edilen olasılıklar arasındaki farklar Tablo-3'te verilmiştir. Buna göre kuramsal sonuçlar her durumda daha büyük olasılıkları göstermekte, iki olasılık arasındaki fark, büyüyen σ_τ^2/σ^2 oranı için küçülmektedir. Bu farklar istatistiksel olarak anlamlı sayılamayacak kadar küçük değerler olarak elde edilmiştir. Burada yapılan simülasyon çalışmasında yaratılan kitlenin varsayımları sağladığı, benzetim çalışmasının da gerçek değerlere oldukça yakın değerler verdiği söylenebilir.

Tablo 3: Kuramsal Olasılıklar ile Simülasyon Sonuçları Arasındaki Farklar

Oran	σ^2	k=5			
		n=2	n=3	n=4	n=5
0.5	288	0.0127	0.0132	0.0150	0.0141
1.0	144	0.0113	0.0125	0.0138	0.0133
1.5	96	0.0098	0.0117	0.0096	0.0071
2.0	72	0.0090	0.0098	0.0071	0.0057
2.5	57	0.0088	0.0068	0.0058	0.0046
3.0	48	0.0076	0.0058	0.0040	0.0032
3.5	41	0.0072	0.0046	0.0033	0.0027
4.0	36	0.0060	0.0043	0.0032	0.0019

KAYNAKLAR

Bek Y., Kayaalp T., Cebeci Z.(1992), "Kısıtlanmış Maksimum Olabilirlik Yöntemi İle Varyans Unsurlarının Tahmini", Araştırma Sempozyumu'92 Bildirileri, Ankara.

Douglas C. Montgomery (2000), ***"Design and Analysis of Experiments",*** Third Edition, Wiley.

Harold R. Lindman (1991), "Analysis of Variance in Experimental Design", Springer-Verlag.

Hendersen C.R. (1953), "Estimation of Variance and Covariance Components", Biometrics 226-252.

Ruth Meyer and David Krueger (1988), ***"A Minitab Guide to Statistics",*** Prentice Hall.

Barbara F.Ryan and Brian L.Joiner (1994), ***"Minitab Handbook",*** Duxbury Press, 1994.

Verdooren L.R.(1982), "How large is the Probability for Estimation of a Variance Component to be Negative", Biometrics, 339-360.

Kurt S. Ege Ö. (1998), "Varyans Bileşenlerinin Negatif Kestirilme Olasılıkları", İstatistik Günleri Sempozyumu.