



ISSN:1306-3111

e-Journal of New World Sciences Academy
2010, Volume: 5, Number: 2, Article Number: 1B0019

MEDICAL SCIENCES

Received: September 2009

Accepted: March 2010

Series : 1B

ISSN : 1308-7312

© 2010 www.newwsa.com

Hayal Boyacıođlu

Soner Duman

Ege University

hayal.boyacioglu@ege.edu.tr

Izmir-Turkey

**DENGELENMEMİŐ VERİLERDE VARYANS ANALİZİ (ANOVA) YÖNTEMİ VE NEFROLOJİK
VERİLERDE UYGULAMASI**

ÖZET

DengelenmemiŐ deneme düzenleriyle iki-yönlü varyans analizi çözümlenmesinde en iyi yöntemin belirlenmesine ilişkin çalıŐmalar 1934 yılında Yates ile baŐlamıŐtır. Bu çalıŐmada dengelenmemiŐ verilerde varyans analizi konusu ele alınmiŐ, buna ilişkin, Tip II (Fitting constants) ve Tip III (Weighted squares of means) yöntemleri incelenmiŐtir. ÇalıŐmada uygulama olarak 65 fareye ait veriler kullanılarak bir uygulama yapılmıŐtır.

Anahtar Kelimeler: Varyans Analizi, Fitting Sabitleri,
Anlamların Ađırlıklı Kareleri,
DengelenmemiŐ Veri, Nefroloji

**ANALYSIS OF VARIANCE (ANOVA) FOR UNBALANCED DATA AND APPLICATION IN
NEPHROLOGICAL DATA**

ABSTRACT

Methods for analyzing unbalanced factorial designs can be traced back to Yates (1934). In this study, variance analysis technique on unbalanced Data was examined, and in this sense, Type II (Fitting constants), and Type III (Weighted Squares Mean) methods were studied. The data of 65 rats is used to illustrate these methods.

Keywords: Analysis Of Variance, Fitting Constants,
Weighted Squares of Means, Unbalanced Data, Nephrology

1. GİRİŞ (INTRODUCTION)

A ve B gibi iki faktörlü dengelenmemiş deneme düzeni modeli,

$$X_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{ijk} \quad (1)$$

ile verilir, bu modelde;

μ : Genel ortalama

X_{ijk} = i'nci satır, j. Sütunda bulunan k'ıncı değer

$i=1\dots a$, $j=1\dots b$,

A_i = A faktörünün i'nci düzeyinin etkisi,

B_j = B faktörünün j'nci düzeyinin etkisi,

$(AB)_{ij}$ =A faktörünün i'nci düzeyi ile B faktörünün j'nci düzeyi bileşiminin ek etkisi

ε_{ijk} = Rasgele Hata, göstermektedir.

Böylece eşitsizlik,

Genel $KT \neq A\ KT + B\ KT + AB\ KT + Hata\ KT$ 'dir.

Ana etki ve etkileşim etkisi birbirinden bağımsız olamamaktadır. Hücreler arasındaki dengesizliğin yarattığı sorunları çözmek için literatürde üç tür yaklaşım önerilmektedir. Bu yaklaşımlar aşağıdaki gibi sıralanabilir.

- **Veri eklemesi ile hücrelerin dengelenmesi:** Bu yaklaşım pragmatik değildir ve rasgeleliğe aykırıdır.
- **Veri çıkarılması ile dengesizliğin giderilmesi:** Bu durumda, zaten zayıf olan testin gücü veri kaybedilerek daha da azalacaktır. Ayrıca, şimdiye kadar yapılan çalışmalarda hangi verinin deneyden çıkarılacağı konusunda yöntemler tanımlanmamıştır.
- **Dengesiz veri seti ile varyans analizi çözümü:** Yapılan çalışmalar, bu durum için en iyi çözüm yönteminin dengesiz veri seti ile çözümü yapmak olduğunu göstermiştir. Bu sebepten dolayı kareler toplamlarının hesaplanmasında çeşitli yöntemler geliştirilmiştir. Yöntemler, literatürde en genel anlamıyla Tip I, Tip II, Tip III, Tip IV adını alır. Bu çalışmada Tip II ve Tip III yöntemleri karşılaştırmalı olarak incelenmiştir [3, 4, 5, 6, 7, 8, 9, 10 ve 11].

- **Tip II Yöntemi**

Literatürde, "Yates's method of fitting constants" ve "Type II SS" yöntemi olarak adlandırılan bu yöntem ilk olarak Yates(1934) tarafından geliştirilmiştir(1,2). Bu yöntemde, her bir faktörün etki düzeyi(etkileşim terimi dışında), diğer faktöre göre düzeltilerek hesaplanmaktadır. Eğer deneme deseninde boş hücre varsa ve etkileşim önemsiz ise Tip II yöntemi varyans analizi çözümlemesinde en iyi sonucu vermektedir [10, 11, 12, 13 ve 14]. A ve B olarak tanımlanmış iki faktörlü dengelenmemiş bir deneme düzeninde model denklemi (1)'deki gibi oluşturulduğunda faktör ve etkileşim kareler toplamları aşağıdaki gibi tanımlanır.

$$KT_A = KT(A \setminus B)$$

$$KT_B = KT(B \setminus A)$$

$$KT_{AXB} = KT(AXB \setminus A, B)$$

A faktörüne ait kareler toplamları B faktörüne göre, B faktörüne ait kareler toplamları A faktörüne göre, etkileşim (AXB) kareler toplamları A ve B faktörüne göre düzeltilerek hesaplanmaktadır [10, 11, 12 ve 13].

- **Tip III Yöntemi**

Literatürde, "Yates's weighted squares of means" ve "Type III SS" şeklinde tanımlanan bu yöntem, deneme düzeninde boş hücre olmaması ve etkileşimin önemli olması durumunda, standart varyans analizi çözümlemesine en uygun yöntem olarak tanımlanmaktadır [10, 11, 12, 13 ve 14]. Tip III yönteminde, faktör kareler toplamları, diğer faktör ve etkileşim'e göre düzeltilerek hesaplanır. Modelde etkileşim önemli

ise, deneme düzeninde boş hücre olmaması koşulu ile Tip III yöntemi ile iki-yönlü varyans analizi çözümlenmesi en iyi sonucu vermektedir. Bu yöntemle kareler toplamları hesaplanırken faktörlerin hesaplamaya alınış sırası önemli değildir [9, 10, 11, 12, 13, 14 ve 15].

A ve B olarak tanımlanmış iki faktörlü dengelenmemiş bir deneme düzeninde(1)'deki gibi kurulmuş bir model denkleminde ait faktör ve etkileşim kareler toplamları, Tip III yönteminde aşağıdaki gibi tanımlanmaktadır.

$$KT_A = KT(A \setminus B, AXB)$$

$$KT_B = KT(B \setminus A, AXB)$$

$$KT_{AXB} = KT(AXB \setminus A, B)$$

A faktörüne ait kareler toplamları B faktörü ve etkileşim(AXB)'e göre, B faktörüne ait kareler toplamları A faktörü ve etkileşim(AXB)'e göre, etkileşim kareler toplamları A ve B faktörüne göre düzeltilerek hesaplanmaktadır [4, 6, 8, 13 ve 14].

2. ÇALIŞMANIN ÖNEMİ (RESEARCH SIGNIFICANCE)

DeneySEL çalışmaların, istatistiksel analizleri için önerilen tekniklerden pek çoğu dengelenmiş veriler, diğer bir ifade ile alt gruplardaki ölçüm sayılarının eşit ya da sabit oranlı olması durumuna dayalıdır. Ne var ki, alt gruptaki birey sayılarının çok farklı, kimi zaman bazı alt grupların boş olması durumu yaygın olarak karşılaşılan durumlardandır. Böyle durumlarda, yine de amaçlanan karşılaştırmaların yapılmasının ve kurulan hipotezlerin test edilmesinin mümkün olduğunu göstermek ve bunun yollarını belirlemek, araştırmacılara önemli çıkış yolları sağlayacaktır [1, 2 ve 3]. Çalışmanın amacı, en dengesiz koşullarda bile, anlamlı değerlendirmelerin yapılabileceğini göstermek suretiyle tıp alanındaki çalışmalara katkı sağlamaktır.

3. UYGULAMA (APPLICATION)

Bu çalışmada, Ege Üniversitesi Tıp Fakültesi İç Hastalıkları-Nefroloji Anabilim Dalında yapılmış 65 fareye uygulanmış deneysel araştırma sonuçları kullanılmıştır. Çalışmada etik kurul onayı alınmıştır. Fareler ağırlıklarına(gr) göre I. Grup(≤ 230), II.Grup (>230), olarak gruplandırılmış ve 7 farklı tedavi yöntemi uygulanmıştır. 2×7 'lik düzende bir deneme düzeni oluşturulmuştur. Veri setinde değişken olarak, Ultra Filtrasyon(UF), Glikoz, Sodyum(Na) ve Potasyum (K), değerleri yer almaktadır. İlk olarak değişkenlere ait tanımlayıcı istatistikler hesaplanmış daha sonra tüm değişkenler için Tip II ve Tip III yöntemlerine göre iki-yönlü varyans analizi SPSS 10.0 paket programı ile çözümlenmiştir. Çalışmada $p < 0.05$ olması anlamlı kabul edilmiştir.

UF, Glikoz, Sodyum, Potasyum için tanımlayıcı istatistikler Tablo 1'de sunulmuştur.

Tablo 1. Tanımlayıcı istatistikler
(Table 1. Descriptive statistics)

| | Ort. (Gr.I) | Ort. (Gr.II) | S.Sapma (Gr.I) | S.Sapma (Gr.II) | N (Gr.I) | N (Gr.II) |
|----------|----------------|-----------------|-------------------|--------------------|-------------|--------------|
| UF | 7.13 | 8.85 | 1.84 | 2.29 | 39 | 26 |
| Glikoz | 214.58 | 213.69 | 42.89 | 49.24 | 38 | 26 |
| Sodyum | 142.83 | 144.77 | 5.05 | 2.69 | 36 | 26 |
| Potasyum | 4.81 | 4.87 | 0.77 | 0.59 | 36 | 24 |

Yapılan varyans analizi sonucunda etkileşim iki yöntem ile aynı sonucu vermiş istatistiksel olarak anlamlı bulunmamıştır. UF(sd=6, F=0.882, P>0.05), Glikoz(sd=6, F=2.213, P>0.05), Sodyum(sd=6, F=0.331, P>0.05), Potasyum (sd=6, F=0.837, P>0.05) (Tablo 2).

Tüm parametreler için tedavi yöntemi sonuçları incelendiğinde, etkileşim önemsiz olduğu için, Tip II yöntemine göre çözümlene en iyi sonucu vermiştir. UF(sd=6, F=3.472, P<0.05), Glikoz(sd=6, F=4.241, P<0.05), Sodyum(sd=6, F=2.914, P<0.05), Potasyum(sd=6, F=4.112, P<0.05) (Tablo 2).

UF (7.82±2.19), Glikoz(214.22±45.19), Sodyum(143.65±4.30), Potasyum(4.83±0.69) değerlerinin ağırlık'a bağlı olarak farklılık göstermediği diğer bir ifade ile gruplar arasında istatistiksel olarak anlamlı bir fark olmadığı söylenebilir. UF(sd=1, F=3.292, P>0.05), Glikoz(sd=1, F=0.634, P>0.05), Sodyum(sd=1, F=0.996, P>0.05), Potasyum(sd=1, F=0.207, P>0.05) (Tablo 2).

Tablo 2. Varyans analizi sonuçları
(Table 2. Results of analysis of variance)

| Değişken | Değişim Kaynakları | sd | KO (Tip II) | KO (Tip III) | F (Tip II) | F (Tip III) | P (Tip II) | P (Tip III) |
|-------------------|--------------------|----|-------------|--------------|------------|-------------|------------|-------------|
| U.Flitrasyon (UF) | Yöntem | 6 | 11.692 | 8.898 | 3.472 | 2.642 | 0.006 | 0.026 |
| | Grup | 1 | 11.086 | 11.207 | 3.292 | 3.328 | 0.076 | 0.074 |
| | Etkileşim | 6 | 2.972 | 2.972 | 0.882 | 0.882 | 0.515 | 0.515 |
| | Hata | 51 | 3.368 | 3.368 | | | | |
| | Toplam | 65 | | | | | | |
| Glikoz | Yöntem | 6 | 6151.235 | 4310.058 | 4.241 | 2.972 | 0.002 | 0.015 |
| | Grup | 1 | 919.905 | 727.961 | 0.634 | 0.502 | 0.430 | 0.482 |
| | Etkileşim | 6 | 3210.112 | 3210.112 | 2.213 | 2.213 | 0.057 | 0.057 |
| | Hata | 50 | 1450.414 | 1450.414 | | | | |
| | Toplam | 64 | | | | | | |
| Sodyum (Na) | Yöntem | 6 | 46.287 | 31.824 | 2.914 | 2.004 | 0.017 | 0.084 |
| | Grup | 1 | 15.819 | 14.468 | 0.996 | 0.911 | 0.323 | 0.345 |
| | Etkileşim | 6 | 5.251 | 5.251 | 0.331 | 0.331 | 0.918 | 0.918 |
| | Hata | 48 | 15.883 | 15.883 | | | | |
| | Toplam | 62 | | | | | | |
| Potasyum (K) | Yöntem | 6 | 1.544 | 1.502 | 4.112 | 3.999 | 0.002 | 0.003 |
| | Grup | 1 | 0.077 | 0.009 | 0.207 | 0.024 | 0.651 | 0.877 |
| | Etkileşim | 6 | 0.315 | 0.315 | 0.837 | 0.837 | 0.547 | 0.547 |
| | Hata | 46 | 0.376 | 0.376 | | | | |
| | Toplam | 60 | | | | | | |

4. SONUÇLAR (CONCLUSIONS)

İki faktörlü dengelenmiş bir deneme deseninde, iki-yönlü varyans analizi çözümlenmesi yapıldığında, iki yöntem de aynı sonucu verecektir. Dengelenmemiş düzende ise, hata ve etkileşim kareler toplamlarının hesaplanmasında hangi yöntemle çözüm yapılırsa yapılsın bulunan sonuç değişmeyecektir.

NOT (NOTICE)

Bu çalışma, 20-22 Eylül 2005 tarihinde Bursa'da yapılan VIII. Ulusal Biyoistatistik Kongresinde sunulmuştur.

KAYNAKLAR (REFERENCES)

1. Boyacıoğlu, H. and Baskan, O., (2007). Unbalanced Factorial Design: Type I, II and III Sums of Squares, Online Journal of Bioinformatics; 8(1):45-55.
2. Boyacıoğlu, H. and Dural, B., (2009). The Statistical investigation on Posidonia oceanica (L.) Delile Meadows in Izmir-Turkey, Environmental Modeling and Assessment, 14(6), 769-772.

3. Driscoll, M. and Borrer, C.M., (2000). Sum of squares and Expected Mean Squares in SAS, *Quality and Reliability Engineering Int.* 16: 423-433.
4. Fujikoshi, Y., (1993). Two-way ANOVA Models with Unbalanced Data, *Journal of Discrete Mathematics*; 116:315-334
5. Graybill, F.A., (2000). *Theory and Application of the Linear Model*, Brooks/Cole Pub. Co. USA, 720p.
6. Herr, D.G., (1986). On the History of ANOVA in Unbalanced, Factorial Designs: The First 30 Years, *The American Statistician*; 40: 265-270.
7. Kirk, R.E., (1995). *Experimental Design: Procedures for the Behavioral Science*, Brooks/Cole Publishing Company, Newyork; 920p.
8. Langsrud, O., (2003). ANOVA for Unbalanced Data: Use Type II Instead of Type III Sum of Squares, *The Journal of Statistics and Computing*; 13, 163-167.
9. Lewsey, J.D., Gardiner, W.P., and Gettinby, G., (1997). A Study of Simple Unbalanced Factorial Designs that Use Type II and Type III Sums of Squares, *Communications in Statistics-Simulation and Computations*; 26:1315-1328.
10. Lewsey, J.D. and Gardiner, W.P., (2001). Gettinby, G., A Study of Type II and Type III Power for Testing Hypotheses from Unbalanced Factorial Designs, *Communications in Statistics-Simulation and Computations*; 30: 597-609.
11. Macnaughton, D.B., (1998). Which Sum of Squares Are Best In Unbalanced Analysis of Variance; *Joint Statistical Meetings in Boston*.
12. Shaw, R.G., and Mitchell-Olds, T., (1993). ANOVA for Unbalanced Data: An Overview, *Ecology*; 74: 1638-1645.
13. Spector, P.E., Voissem, N.H., and Cone, W.L., (1981). A Monte Carlo Study of Three Approaches to Nonorthogonal Analysis of Variance, *Journal of Applied Psychology*; 66, 535-540
14. Steel, G.D., Torrie, J.H., Dickey, D.A., (1997). *Principles and Procedures of Statistics a Biometrical Approach*, McGraw-Hill Series, USA; 666p.
15. Yates, F., (1934). The Analysis of Multiple Classification With Unequal Numbers In The Different Classes, *Journal of The American Statistical Assosiation*; 29, 51-56.