

VARYANS ANALİZİNDE KORELASYON YAKLAŞIMI

Yrd. Doç. Dr. Emel İMİR*

ÖZET

Bu çalışmada varyans analizinde korelasyon yaklaşımının kullanılması incelenmektedir. Ayrıca yaklaşımın kullanımını gösteren bir uygulama verilmektedir.

SUMMARY

In this paper, the analysis of variance is examined in term of the simple correlation. The usage of this procedure is demonstrated in an example.

1- GİRİŞ

R.A Fisher tarafından geliştirilen varyans analizi, deneysel verilerin çözümlenmesi için kullanılan bir istatistik tekniktir.

(*) Anadolu Üniversitesi Fen Edebiyat Fakültesi Öğretim Üyesi.

İstatistikte çok sayıdaki örnekleme birden karşılaştırmaya yarayan teknik olması nedeniyle varyans analizi geniş uygulama alanına sahiptir.

Varyans analizi tekniğiyle bir değişkenin toplam varyansı toplanabilen bileşenlere ayrılabilir. Toplam varyansın bu şekilde ayrıştırılmasıyla karşılıklı etkileşim yapan belirli değişkenlerin etkisinin ortadan kaldırılması için istatistiksel yöntemlerin kullanılabilmesi ve böylece deneyin duyarlılığının artırılması sağlanmış olur; varyans analizi bu görüşle tasarlanan bir deneyin analizinde uygulanan bir tekniktir.

Paremetrelerin kestirimi veya bir önsavın sınanmasında varyans analizinden yararlanılmaktadır. Çalışmamızda ise sayıları ikiden fazla işleyim ortalaması arasındaki farkın istatistiksel anlamlılığını sınamaya yarayan varyans analizine yer verilecektir. Gözlem değerleriyle grup ortalamaları arasındaki ilişkinin derecesi de görülebileceğinden varyans analizinde korelasyon yaklaşımı benimsenmiştir.

Korelasyon yaklaşımının varyans analizinde kullanılması grafikte gösterimi sağlamaktadır. Bu yaklaşım sayısal olmayan açıklayıcı değişkenin grup ortalaması kullanılarak grafikte gösterimine olanak vermektedir [1].

2- VARYANS ANALİZİ VE KORELASYON

Varyans analiziyle regresyon analizinin kavramsal açıdan aynı olduğu bilinmektedir. Her iki analiz tekniğinde de Y'deki toplam değişim iki toplanabilir bileşene ayrılmaktadır.

Varyans analizi tekniğine diğer bir yaklaşım ise basit korelasyon analizidir. Korelasyon ve varyans analizi arasındaki doğrudan bağıntılara zaman içinde literatürde yer verilmiştir. İlk olarak Beaton (1969) bu ilişkiyi kriter ölçeklemesi olarak tanımlamıştır ve Pedhazur (1982) aynı paralelde düşüncelerini geliştirmiştir [4].

Korelasyon analizi ile varyans analizi arasındaki ilişkinin araştırılması çalışmamızın konusunu oluşturacaktır. Bu yaklaşımın tek yönlü ve çift yönlü varyans analizinde kullanımına izleyen bölümde yer verilecektir.

2.1. Tek Yönlü Varyans Analizi ve Korelasyon

Tek Yönlü varyans analizinde k tane grup sözkonusu olduğunda her bir gözlem değeri ile grup ortalaması arasındaki basit korelasyon katsayısının karesi r_{yyk}^{2-}

$$r_{yyk}^{2-} = \frac{G.A.K.T}{G.K.T} \quad (1)$$

veya korelasyon oranı $\tilde{\eta}$ nin karesi $\tilde{\eta}^2$ 'ye eşittir [5].

k grup ortalamasının eşitliğinin sınanması F testiyle yapılmakta ve $\sqrt{1}$, $\sqrt{2}$ serbestlik derecesine bağlı olarak hesaplanan F istatistiği

$F = \frac{s_B^2}{s_W^2}$ dir. Korelasyon katsayısı veya korelasyon oranı yar-

dımıylada hesaplanabilen F istatistikleri ise

$$F = \left(\frac{\sqrt{1}}{\sqrt{2}} \right) \left(\frac{r^2}{1-r^2} \right) \quad \text{veya} \quad (2)$$

$$F = \left(\frac{\sqrt{1}}{\sqrt{2}} \right) \left(\frac{\hat{\eta}^2}{1-\hat{\eta}^2} \right) \quad (3)$$

şeklinde formüle edilmektedir [1].

$F = \frac{s_B^2}{s_W^2}$, örneklem değerinde açıklanan değişkendeki değişme-

lerin işleyimlerin ait olduğu açıklayıcı değişken tarafından açıklanan kısmının açıklanmayan (rasgele yanılğı) kısma oranı sözkonusudur.

$$F = \left(\frac{\sqrt{1}}{\sqrt{2}} \right) \left(\frac{r^2}{1-r^2} \right)$$

örneklem değeri ise açıklanan değişimin açıklanmayan değişime oranı şeklinde formüle edilmektedir.

Diğer taraftan

$$r_{yyk}^2 = \frac{\left[\sum \sum_{ij} (Y_{ij} - \bar{Y})(\bar{Y}_i - \bar{Y}) \right]^2}{\sum_{ij} (Y_{ij} - \bar{Y})^2 \sum_{ij} (\bar{Y}_i - \bar{Y})^2}$$

(4)

$$r_{yyk}^2 = \frac{\left[\sum_i (Y_i - N_i \bar{Y})(\bar{Y}_i - \bar{Y}) \right]^2}{\sum_{ij} (Y_{ij} - \bar{Y})^2 \sum_i N_i (\bar{Y}_i - \bar{Y})^2}$$

$$r_{yyk}^2 = \frac{\text{(Gruplararası kareler toplamı)}}{\text{(Gruplararası kareler toplamı) (Toplam Değişim)}}$$

Bu eşitlik korelasyon oranı (η^2) nın karesi olarak da ifade edilmektedir [2].

$$\eta^2 = \frac{s_k^2}{s^2}$$

s = Gözlem değerlerinin genel ortalamadan olan standart sapması
 s_k = Grup ortalamaları arasındaki farklılıktan doğan değişim

2.2. İki Yönlü Varyans Analizi ve Korelasyon

Deney tasarımında elde edilen veriler aynı anda iki ayrı niteliğe göre bölümlendirilmişse iki yönlü bir bölümlendirme sözkonusu olur. İki yönlü bölümlendirme ve "I_xJ" tablosunda etmenlerin düzeylerine göre r tane sıra ve k tane sütun yer almaktadır. Bölümlendirme tablosunun her bir gözesindeki gözlem sayısına göre çok sayıda iki yönlü bölümlendirme türü ortaya çıkmaktadır. Çalışmada gözelerde eşit sayıda gözlemin bulunduğu durumla ilgilenilecektir.

İki yönlü bölümlendirme sözkonusu olduğunda ve her bir gözede eşit sayıda gözlem olması durumunda doğrusal model

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{ijk} \quad (5)$$

şeklinde ifade edilmektedir.

(IxJ) iki etmenli düzenlemede varyans analizi için yapılacak F testiyle ilgili F istatistiği korelasyon katsayılarından yararlanarak

$$F_i = \left(\frac{\sqrt{1}}{\sqrt{2}} \right) \left(\frac{r_i^2}{1 - r_A^2 + r_B^2 + r_{AB}^2} \right) \quad (i = A, B) \quad (6)$$

şeklinde formüle edilmektedir.

Formülde;

r_A^2 ve r_B^2 ; sırasıyla A ve B esas etkilerle ilgili kareli korelasyon katsayıları; r_{AA}^2 ise (AxB) etkileşime ilişkin kareli korelasyon sayısıdır. (AB) etkileşime ilişkin korelasyon katsayısı, r_{AB}^2

$$r_{AB}^2 = r_{göze}^2 - r_A^2 - r_B^2 \quad (7)$$

olarak hesaplanmaktadır. $r_{göze}^2$ IJ gözelerine bağlı olan kareli korelasyondur.

$$r_A^2 = \frac{\text{Kareler Toplamı (A)}}{\text{Genel Kareler Toplamı}}$$

$$r_B^2 = \frac{\text{Kareler Toplamı (B)}}{\text{Genel Kareler Toplamı}}$$

$$r_{AB}^2 = \frac{\text{Kareler Toplamı (AB)}}{\text{Genel Kareler Toplamı}}$$

şeklinde hesaplandığında $r_{göze}^3$ değeri (7) nolu eşitlikten yararlanarak

$$r_{göze}^2 = r_{AB}^2 + r_A^2 + r_B^2 \quad (8)$$

formülüyle elde edilmektedir [3].

Etmenlerle ilgili kareli korelasyon katsayısı r_i^2 sıfırla bir arasında değer almaktadır. r_i^2 nin bire yakın değer alması gözlem değerleriyle grup ortalamaları arasındaki ilişkinin kuvvetli olduğunu göstermektedir. İlişkinin derecesinin yanında r_i^2 leri kullanarak hesaplanan F istatistiğinin değeri r_i^2 nin bire yaklaşmasıyla büyüyeceğinden işleyim ortalamaları arasındaki farkın istatistiksel olarak anlamlı olmasını sağlamaktadır. r_i^2 nin sıfıra yakın değer alması ise ortalamalar arasındaki farkın tesadüfen ortaya çıktığını, bu farkın istatistiksel olarak anlamlı olmadığını göstermektedir.

UYGULAMA

Varyans analizinde kareli korelasyon yaklaşımından yararlanılması hesaplamaların pratik olması yönünden bir üstünlük göstermektedir. Değişkenler arasındaki ilişkinin de bilinmesi önem kazandığında bu yaklaşım araştırmacıya pratik yarar sunmaktadır. Bu avantajların yanısıra; gözlemlerdeki gözlem sayılarının eşit olması durumunda grup toplamları grup ortalamalarının yerine kullanılabilmesi korelasyon yaklaşımıyla varyans analizinin diğer bir üstünlüğü olarak görülmektedir [1].

Çalışmamızda, A.Ü. Fen Edebiyat Fakültesi İstatistik bölümü öğrencilerinin başarıları incelenmek istenmiştir. Bu amaçla 1988 girişli istatistik bölümü öğrencilerinden 36 öğrencinin Betimsel İstatistik dersinden aldıkları dönem sonu başarı notları incelenmiştir. Farklı barınma yerlerinde kalan öğrencilerin ortalama başarıları arasında anlamlı bir farklılık olabileceği görüşüyle dönem sonu başarı notlarının barınma yerleriyle olan ilişkisi araştırılmıştır. İlgili veriler Tablo-1'de yer almıştır.

Öğrencilerin başarı düzeyiyle ilişkilendirilen sadece tek açıklayıcı değişken barınma yeri sözkonusu olduğundan tek yönlü varyans analizi yapılacaktır; açıklayıcı değişken için üç düzey dolaşısıyla 3 grup sözkonusudur. Başarı notları ile grup ortalamaları arasındaki basit korelasyon katsayısının karesi

$$r^2 = \frac{\left[\sum_k N_k (\bar{Y}_k - \bar{Y}) \right]^2}{\sum_{ki} (Y_{ki} - \bar{Y})^2 \sum_k N_k (\bar{Y}_k - \bar{Y})^2}$$

şeklinde ifade edilen formülden yararlanılarak

Tablo-1 : Öğrencilerin barınma yerine göre betimsel istatistik der-
si dönem sonu başarı notlarının dağılımı

Öğrencilerin Barınma Yerleri		
Ailesi Yanında Kalan	Evde arkadaşlarıyla Kalan	Yurtta Kalan
21	18	50
35	20	50
35	28	54
39	37	56
42	51	56
50	52	57
50	55	58
50	57	65
51	61	66
58	62	67
59	65	76
70	70	81

$$r_{yyk}^2 = \frac{2499606,533}{12287643,92}$$

$$r_{yyk}^2 = 0,2034243$$

olarak hesaplanmıştır. Başarı notları ile grup ortalamaları arasında kuvvetli olmayan bir ilişki görülmektedir. Korelasyon oranı η^2 nin değeri ise

$$\eta^2 = \frac{s_k^2}{s^2}$$

formülünden

$$\eta^2 = \frac{43,91706667}{215,888}$$

$$\eta^2 = 0,203425$$

elde edilmiştir. $r_{yyk}^2 = \hat{\eta}^2$ eşitliği regresyon yüzeyinin doğrusal olduğunu göstermektedir.

Öğrencilerin barınma yerleriyle başarı notları arasındaki bu ilişki k grup ortalamasının eşit olduğu varsayımının doğrudan F testiyle sınanmasını sağlamaktadır. $\sqrt{1}$ ve $\sqrt{2}$ serbestlik derecesine bağlı olarak hesaplanan F istatistiği

$$F = \left(\frac{\sqrt{1}}{\sqrt{2}} \right) \left(\frac{\hat{\eta}^2}{\hat{1-\eta}^2} \right)$$

formülü yardımıyla aşağıdaki şekilde hesaplanmıştır.

$$F = \left(\frac{33}{2} \right) \left(\frac{0,203425}{1-0,203425} \right)$$

$$F = 4,213 > F_{(2;33;0,95)} = 3,32$$

bulduğundan ortalamalar arasındaki farkın anlamlı olduğu görülmüş ve bu sonuçtan dolayı barınma yerleri farklılığının öğrencilerin başarıları üzerinde etkili olduğu kararına ulaşılmıştır.

Öğrencilerin başarıları üzerinde barınma yerinin yanısıra mezun oldukları lise türünün de etkisi olduğu düşüncesiyle iki yönlü bölümlendirme yapılmıştır. Tablo-2 de öğrencilerin başarı notları iki yönlü bölümlendirmeye göre verilmiştir.

Tablo-2 : Öğrencilerin barınma yeri ve lise eğitim türüne göre be-
timsel istatistik dersi başarı notlarının dağılımı

Barınma Yerleri

Mezun oldukları lise türü	Ailesi Yanında	Evre arkadaşlarıyla	Yurtta
Fen kolu	70,59,58	70,65,37	81,65,54
Edebiyat kolu	51,50,42	57,52,28	66,67,76
Klasik lise	50,50,35	51,55,62	50,50,58
Meslek lisesi	21,35,39	18,20,61	57,56,56

(IxJ) iki yönlü bölümlendirme esas alınarak gerçekleştirilen varyans analizi için yapılan F testiyle ilgili F istatistiği korelasyon katsayılarından yararlanarak

$$F_i = \left(\frac{\sqrt{1}}{\sqrt{2}} \right) \left(\frac{r_i^2}{1-r_A^2 + r_B^2 + r_{AB}^2} \right)$$

şeklindeki formülün kullanıldığı (2.2.) bölümünde açıklanmıştır. Formülde barınma yerleri (A) ve lise türleriyle (B) ilgili kareli korelasyon katsayıları hesaplanmış

$$r_A^2 = 0,19$$

$$r_B^2 = 0,27$$

bulunmuştur. (AxB) etkileşime ilişkin kareli korelasyon katsayısı r_{AB}^2

$$r_{AB}^2 = 0,1196$$

ve

$$r_{göze}^2 = r_{AB}^2 + r_B^2 - r_B^2$$

eşitliğinden

$$r_{göze}^2 = 0,5796$$

olarak hesaplanmıştır. Bu bulgular yardımıyla F testiyle ilgili F istatistikleri

$$F_A = \left(\frac{\sqrt{3}}{\sqrt{1}} \right) \left(\frac{r_A^2}{1-r_A^2 + r_B^2 + r_{AB}^2} \right)$$

$$F_A = 5.423 \text{ ve}$$

$$F_B = \left(\frac{\sqrt{3}}{\sqrt{2}} \right) \left(\frac{r_B^2}{1-r_A^2 + r_B^2 + r_{AB}^2} \right)$$

$$F_B = 5.13$$

olarak elde edilmiştir.

$$F_A = 5,423 > F_{2,24;0,95} = 3,40 \text{ ve}$$

$$F_B = 5,13 > F_{3,24;0,95} = 3,01$$

sonuçlarından dolayı ortalamalar arasındaki farkın anlamlı olduğu görülmüştür ve bu nedenle barınma yerlerinin ve mezun oldukları lise türlerinin farklılığı öğrencilerin başarılarında etkili olduğu sonucuna ulaşılmıştır.

$$F_{AB} = \left(\frac{\sqrt{3}}{\sqrt{4}} \right) \left(\frac{r_{AB}^2}{1-r_A^2 + r_B^2 + r_{AB}^2} \right)$$

$$F_{AB} = 1.137 > F_{12,24;0,95} = 2.18$$

olduğundan istatistiksel olarak anlamlı olmadığını yani etkileşimin olmadığını karar verilmiştir.

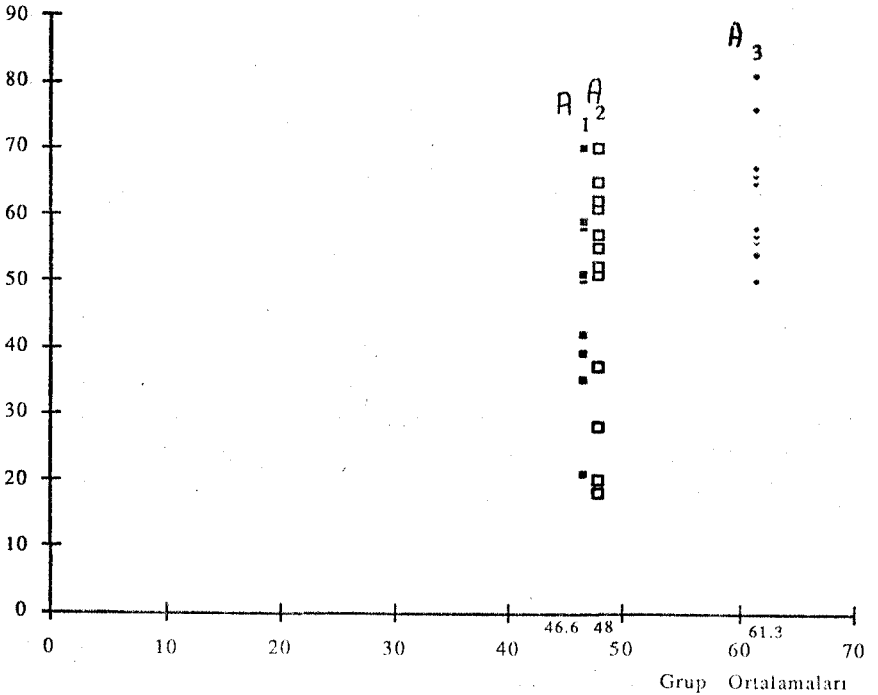
Varyans analizinde korelasyon yaklaşımının benimsenmesi durumunda, gözlem değerleriyle grup ortalamaları arasındaki doğrusal ilişki grafik yardımıyla gösterilebilir. Şekil-1 de 1988 girişli A.Ü. F.E. Fakültesi İst. Böl. öğrencilerinden 36 öğrencinin barınma yerleriyle başarıları arasındaki ilişki gösterilmeye çalışılmıştır. Şekilden görüldüğü gibi 3. kolonda yer alan en yüksek başarı notu 81 olup, yurttan kalan bir öğrencinin notudur. Beklentinin tersine yurttan kalan öğrenciler evde ailesi veya arkadaşlarıyla kalan öğrencilerden daha yüksek not aldıklarını Şekil-1 de görmek olasıdır. Evde ailesiyle olan öğrenciler ile evde arkadaşlarıyla kalan öğrencilerin ortalama başarı notları birbirine çok yakındır ve kareli korelasyonun $r_{yyk}^2 = 0,203$ küçük bir değer olması da bu nedendir. Eğer tüm grup ortalamaları eşit olsaydı tüm başarı notları tek bir değer üzerinde dizilirdi ve korelasyon katsayısı

ise sıfır olarak bulunurdu. Korelasyon katsayısının bir olabilmesi ise; her bir grup içindeki başarı notları aynı olurken aynı zamanda iki grup ortalamasının farklılığı durumunda olasıdır.

İki yönlü varyans analizinde korelasyon yaklaşımında, her göze ortalamasının gözlem değeriyle ilişkisi de grafik yardımıyla gösterilebilmektedir. Uygulama konumuz olan 36 istatistik öğrencisinin başarılarını etkileyen etmen olarak barınma yerlerinin yanında mezun oldukları lise türleri de düşünülmüş ve iki yönlü bölümlendirme yapılmıştır. Öğrencilerin başarı notları ile ilgili göze ortalamaları arasındaki ilişki Şekil-2 de gösterilmeye çalışılmıştır. Gözeler arası değişmeyi gösteren bu serpilme diyagramının bilgi verici özelliği azdır bu nedenle barınma yeri (A), mezun olunan lise türü (B) ve etkileşim (AB) ($Y_{ij} - \bar{Y}_j + \bar{Y}$) olarak tanımlanan) ayrı ayrı sırasıyla şekil-1,3 ve 4 de gösterilmiştir.

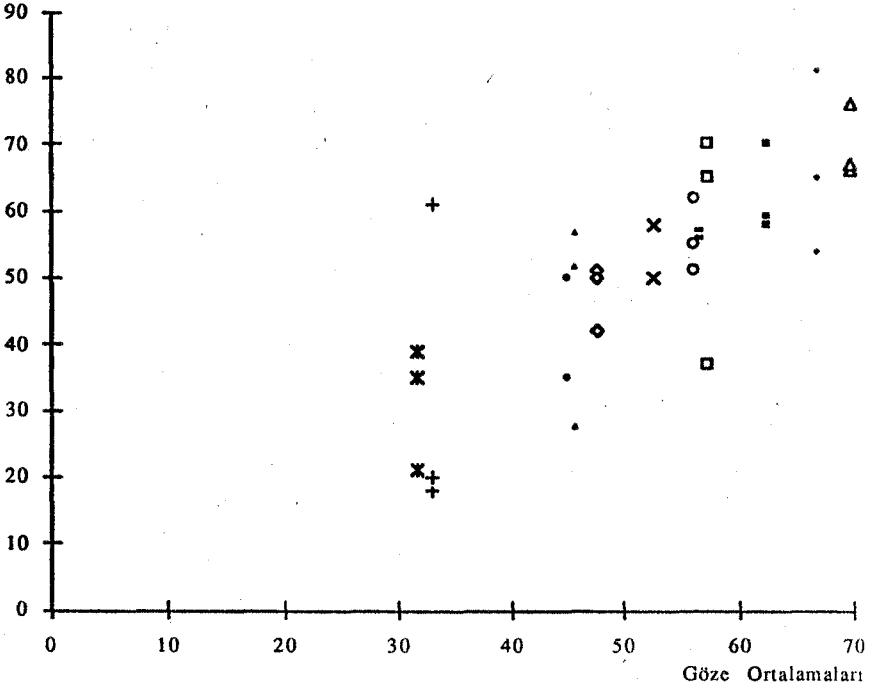
$r_B^2 = 0,27 > r_A^2 = 0,19 > r_{AB}^2 = 0,12$ sonucu şekiller yardımıyla da görülebilmektedir.

Dönem Sonu Başarı Notları



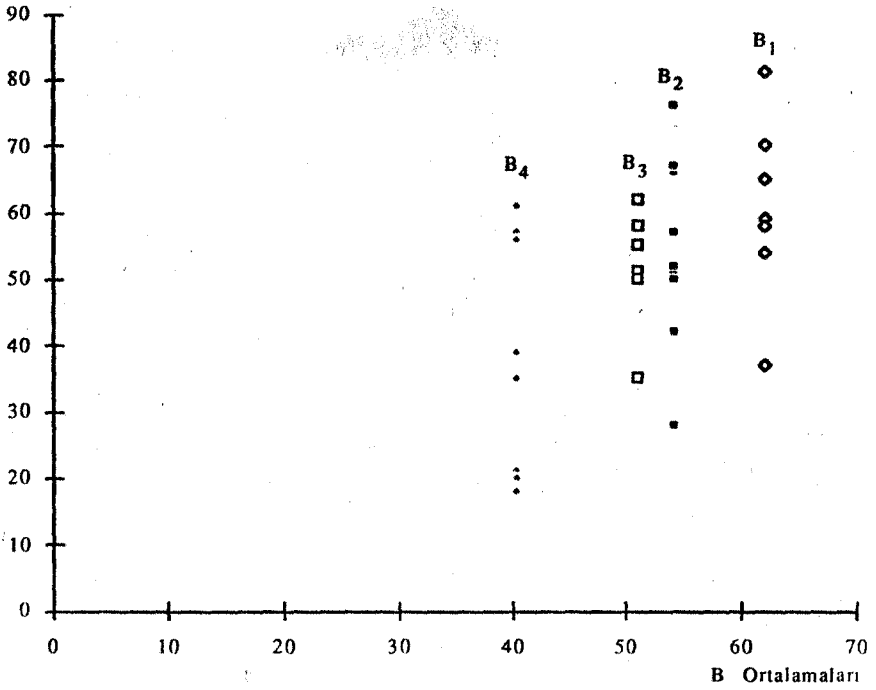
Şekil-1 : Barınma yerlerine göre öğrencilerin Dönem Sonu başarı notları ile grup ortalamaları arasındaki ilişki

Dönem Sonu Başarı Notları



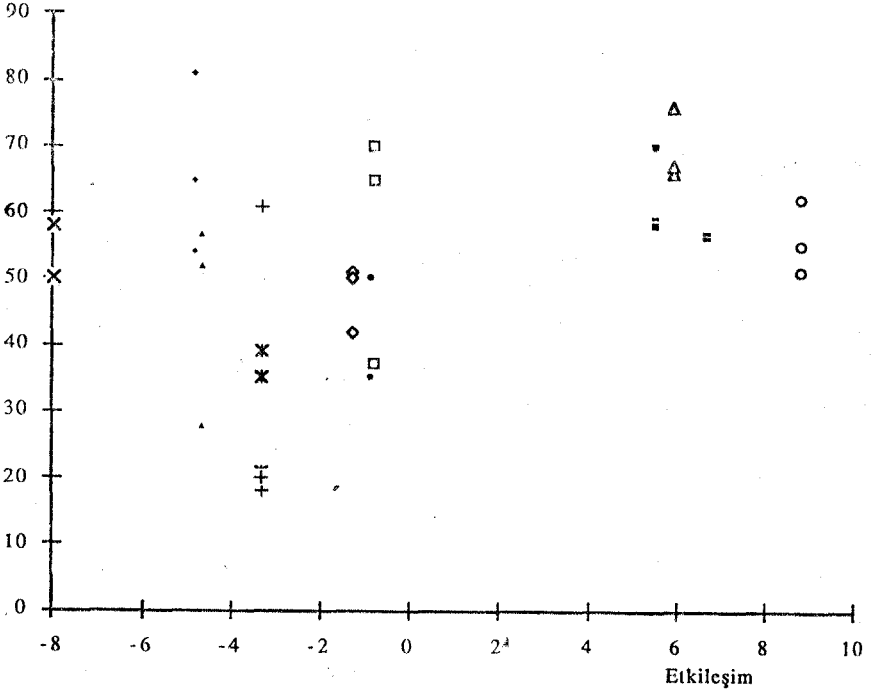
Şekil-2 : Barınma yerleri ve mezun oldukları lise türlerine göre öğrencilerin dönem sonu başarı notları ile göze ortalamaları arasındaki ilişki

Dönem Sonu Başarı Notları



Şekil-3 : Mezun oldukları lise türlerine göre öğrencilerin dönem sonu başarı notları ile grup ortalamaları arasındaki ilişki

Dönem Sonu Başarı Notları



Şekil - 4 -

SONUÇ:

İkiden fazla örneklem ortalamasının karşılaştırılmasında kullanılan tekniklerden birisi varyans analizi tekniğidir. Çalışmalarında geniş bir kullanım alanına sahip olan varyans analizi problemi-ne kareli korelasyon yaklaşımı uygulamasına yer verilmiştir.

1988 girişli A.Ü Fen Edebiyat Fakültesi öğrenci başarı farklılığını etkileyen etmen olarak barınma yeri düşünülmüştür. Ancak öğrencilerin gördükleri eğitim türü bakımından türdeş olmadıkları, dolayısıyla bu nedenle başarı farklılığı görülebileceği düşünülmüştür. Öğrenci başarısında barınma yeri türüyle birlikte mezun olunan lise türünün de etkileri olabileceği üzerinde durulmuş ve öğrenciler aldıkları eğitim türü bakımından homojen gruplarda bölüklerde toplanmıştır. Böylece iki yönlü bölümlendirme ortaya çıkmıştır.

Kareli korelasyon yaklaşımlarıyla yapılan varyans analizi sonucunda her iki açıklayıcı değişkenin öğrencinin başarısını gerçekten etkilediği görülmüştür. Kareli korelasyon katsayıları incelendiğinde ise barınma yerlerine (aralarında etki farklılığı araştırılan işleyimlere) göre öğrencinin lise eğitim türü ile (bölüklerle) başarı arasında daha kuvvetli ilişki olduğu sonucuna ulaşılmıştır.

KAYNAKLAR

- 1- JOEL, R. Levin, Ronald C. Serlin and Lisa Webne Behrman :
"Anaylsis of Variance Through Simple Correlation" The
American Statistician, February 1989, vol. 43,32-34, 1989.
- 2- KIRK, R. E.: Experimental Design (2 nded) Belmont CA: Brooks
Cole, 1982.
3. KLEINBAUM G. David, Kupper L.Lawrence and Muller E.Keith:
Applied Regression Analysis and Other Multivariable
Methods, The University of North Carolina at Chapel
Hill, Boston, 1988.
- 4- PEDHAZUR, E. S. : Multiple Regression in Behavioral Research
(2 nded) New York, Rinehart and Winston, 1982.
- 5- RODGERS, J.L. and Nice wander, W.A. : "Thirteen Ways to
Look at the Correlation Coefficient", The American Sta-
tistician, 42,59-66, 1988.