GÖSTERGE DEĞİŞKENLERLE REGRESYON

Doç. Dr. Durmuş DÜNDAR
İktisat Fak. Öğretim Üyesi

1. *GÖSTERGE DEĞİŞKENLERİNİN KULLANIMI*

1.1. Genel Açıklamalar

Regresyon denklemlerinde kullanılan değişkenler, bazı sürekli aralıkların dışında değerler olabilirler. Ayrıca gözlem olması gerekli olmayan, fakat iki veya daha çok farklı değer olabilen bağımsız değerler kullanmak isteyebiliriz. Örneğin bazı lise mezunları üniversiteye gider, bazıları gitmez. Bu konuda bir hesap tutulmak amacıyla, gösterge bir değişken kullanılabilir. Eğer lise mezunu üniversiteye girmişse onu 1 ile, girmemişse 0 ile belirterek gösterge değişken olarak kullanabiliriz.

Gösterge değişkenler özellikle nitel veriler söz konusu olduğunda yararlı olmaktadır.

Burada gösterge değişkenlerin, regresyon kavramında nasıl kullanıldığını göstermeye çalışacağız.

Gösterge değişkenlerin, regresyon analizlerinde belirli bir kategorideki gözlemlerin, bir gurup paremetreyle ilgili iken, diğer kategorideki gözlemlerin başka bir paremetre gurubuyla ilgili olduğunun gösterilmesinde nasıl kullanıldığını da açıklamaya çalışacağız.

Sözgelimi bir firmanın imalatta 2 tip üretim süreci kullandığım varsayalım. Her bir süreçten elde edilen çıktının farklı beklenen değerleri olmasına karşın, normal dağıldığını fakat varyanslarınmda eşit olduğunu kabul edelim.

Üretim sürecini bir regresyon denklemi şeklinde gösterebiliriz.

$$Y_i = \beta_1 + \beta_2 X_i + \epsilon$$

Burada Y_i, i inci girdi sürecine ilişkin çıktığı göstermektedir. X_i ise aşağıdaki gibi tanımlanmış gösterge bir değişkendir.

$$X_i \begin{cases} = 1 & \text{eğer çıktı A makinasından elde edilmişse} \\ = 0 & \text{eğer çıktı B makinasından elde edilmişse} \end{cases}$$

Bu basit örnekte, regresyon doğrusu B makinasına ilişkin beklenen çıktıyı ölçer, aynı zamanda eğim de A makinasından B makinasına değişiklikle ilgili çıktılar arasındaki farkı gösterir.

Bu

$$Y_i = \beta_1 + \beta_2 X_i + \epsilon$$

denkleminin her iki tarafının

$$X_i = 1 \quad \text{ve} \quad X_i = 0$$

için beklenen değeri alınarak görülebilir.

$$1) \quad X_i = 0 \qquad E(Y_i) = \beta_1$$

$$X_i = 1 \qquad E(Y_i) = \beta_1 + \beta_2$$

$\beta_2 = 0$ hipotezinin testi, bu durumda A ve B makinalarının çıktıları arasında fark olup olmadığının testi demek olacaktır. (Burada regresyon katsayıları gerçek, En Küçük Kareler yöntemiyle yapılan tahminlerin de, B makinasının ortalama çıktısı ile, A ve B makinalarının ortalama çıktıları arasındaki fark olduğunu belirtelim).

İkiden fazla farklı nitel değerle ilgilendiğinde de gösterge değişken kullanılabilir.

Örneğin, yukarıdaki örnekte, 3 farklı üretim süreci kullanılsa idi; bu üç ayrı çıktının aynı olup olmadığı araştırılmak istendiğinde, bu amaçla 2 gösterge değişkene ihtiyaç duyulacaktır. (Genelde N farklı durum için N−1 gösterge değişken yeterli olacaktır).

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

modeli şöyle tanımlanmış olsun :

$$X_{2i} \begin{cases} = 1 & \text{eğer çıktı A makinasmda üretilmişse} \\ = 0 & \text{eğer çıktı diğer makinalarda üretilmişse} \end{cases}$$

$$X_{3i} \begin{cases} = 1 & \text{eğer çıktı B makinasmda üretilmişse} \\ = 0 & \text{eğer çıktı diğer makianlarda üretilmişse} \end{cases}$$

Böylece gösterge değişkenlerin aldığı değerlere göre aşağıdaki kombinasyonlarla, 3 üretim süreci de gösterilebilecektir.

	X_2	X_3
A makinası	1	0
B makinası	0	1
C makinası	0	0

Bu üç durum için de beklenen değerler alınırsa, regresyon sonuçları yorumlanabilir.

1- $X_{2i} = 1$ $X_{3i} = 0$ $E(Y_i) = \alpha_1 + \alpha_2 + \alpha_3$

2- $X_{2i} = 0$ $X_{3i} = 1$ $E(Y_i) = \alpha_1 + \alpha_3$

3- $X_{2i} =$ $X_{3i} =$ $E(Y_i) = \alpha_1$

Regresyon denklemindeki β_1 katsayısı C makinasıyla ilgili çıktının beklenen değerini gösterir. 1. eğim katsayısı (β_2) C makinasından A makinasma değişiklikle ilgili çıktılar arasmdaki farkı, 2. eğim (β_3) C makinasmdanB makinasma bir değişiklik yapılırsa çıktıdaki ortalama değişmeyi ölçer.

$\beta_2 = 0$ hipotezinin testi: A ve C makinalarıyla üretim süreçleri arasında bir fark olup olmadığının testi olacaktır. Aynı zamanda t testi ile β_3 katsayısı üzerine bir test yapılabilir.

Burada :

I — 3 değişik üretim süreci 2 gösterge değişken tarafından gösterilmektedir. Böyle bir olayda tek bir gösterge değişken kullanıp buna üç farklı değer vermek gösterge değişken tekniğine uygun değildir. (Sözgelimi A makinası = 2, B makinası = 1, C makinası = 0 gibi).

II — Gösterge değişken süreçleri, 3 tane iki durumlu değişkenle gösterilmemelidir. Yani X_1, X_2 ve X_3 gibi X_4; C makinası kullanıldığında 1 değerini, diğer zamanlar ise 0 değeri olacak. Bu durumda X_4 değişkeninin kullanımı ek bir bilgi getirmediği gibi en küçük kareler tahmincilerini elde ederken denklemlere ek ve bağımsız olmayan bir denklem eklemek demektir. Bu nedenle de regresyon katsayılarının en küçük kareler tahmincilerinin elde edilmesi olanaksızlaşır. Bu modelde tam bir lineerliğin mevcut olduğu hatırlanırsa daha iyi anlaşılır. Lineerlik mevcuttur, çünkü;

Her i gözlemi için $X_{4i} = 1 - X_{2i} - X_{3i}$ dir.

III — A ve B makinalarının çıktıları arasında herhangi bir fark olup olmadığı test edilmek istenirse: Böyle bir testi F testi sağlar, red hipotezi (A ve B makinalarının çıktıları arasında fark yoktur) regresyon katsayıları β_2 ve β_3 eşittir şeklinde kurulur. Bununla birlikte araştırma regresyon denklemini yeniden yazarak, standart regresyon çıktısına uygulanan bazı t istatistiklerini kullanarak yapılabilir.

Regresyon modeli aşağıdaki gibi yeniden yazılır,

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 (X_{3i} + X_{2i}) + \varepsilon_i$$

3 durum göz önüne alınır.

1) $X_{2i} = 1$ $X_{3i} = 0$ $E(Y_i) = \beta_1 + \beta_2$

2) $X_{2i} = 0$ $X_{3i} = 1$ $E(Y_i) = \beta_1 + \beta_3$

3) $X_{2i} = 0$ $X_{3i} = 0$ $E(Y_i) = \beta_1$

İstenen hipotezin testi de böylece $\alpha_2 = 0$ red hipotezinin testiyle sağlanır.

IV — Gösterge değişken sonuçlarının,

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon \quad \text{ve}$$

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

modelleri için tarif edildiği gibi yorumlanması, farkın (β_1 aynı zamanda C makinasmm etkisiyle çakışmaktadır.

Gerçekte böyle bir karşılaştırmaya görek yoktur. Belli bazı durumlarda ilk sabit terimi (β_1) ihmal ederek 3 ayrı makina işlemini göstermek için 3 ayrı gösterge değişken kullanılabilir.

Gösterge değişkenler kavramı, daha genel durumlara, yani bağımsız değişkenlerin bazılarının sürekli, diğerlerinin gösterge olduğu durumlara genişletilebilir.

Klasik bir örnek toplam tüketim fonksiyonudur. Savaş zamanı tüketim tarzı barış zamanmdakinden farklı olacaktır. Basit bir toplam tüketim fonksiyonunda 5 farklı durum ayırılabilir. Burada tüketim fonksiyonunun harcanabilir gelir tarafından tayin edildiği ve gecikme (lag) lerin dahil edilmediği varsayımı yapılmıştır.

1. Durum

$$C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t$$

Savaş ve barış zamanı tüketim tarzları bu modelde her açıdan aynı varsayılmıştır. Eğer baştan böyle bir bilgi mevcutsa, gösterge değişken kullanımına gerek yoktur.

2. Durum

$$C_t = \beta_1 + \beta_2 Y_t + \alpha D_t + \varepsilon_t$$

Burada, eğer savaş zamanı ise $D_t = 1$ değilse $D_t = 0$ dır.

$$E(C_t) = \beta_1 + \beta_2 E(Y_t) \quad \text{Barış zamanı}$$

$$E(C_t) = (\beta_1 + \alpha) + \beta_2 E(Y_t) \quad \text{Savaş zamanıdır.}$$

Görüldüğü gibi ikinci durumda tüketim fonksiyonu savaş sırasında yalnızca regresyon denklemindeki α katsayısı değişmekte, eğimi aynı kalmaktadır. Değişikliğin istatistiksel olarak anlamlılığının testi $\alpha = 0$ hipotezinin testiyle yapılır.

3. Durum

$$C_t = \beta_1 + \beta_2 Y_t + \gamma(D_t Y_t) + \varepsilon_t$$

$$E(C_t) = \beta_1 + \beta_2 E(Y_t) \quad \text{Barış zamanı}$$

$$E(C_t) = \beta_1 + (\beta_2 + \gamma)(E(Y_t)) \quad \text{Savaş zamanıdır.}$$

3. durumda savaş zamanı α katsayısı sabit kaldığı, eğimin değiştiği varsayımına dayanmaktadır. Başka bir deyişle, 3. durumdaki model savaşın toplam marjinal tüketim eğiliminde bir değişikliğe neden olduğunu varyasar. Bu değişikliğin anlamlılık testi, $D_t Y_t$ nin katsaylsı $\gamma = 0$ hipotezinin testi ile sağlanır.

4. Durum

$$C_t = \beta_1 + \beta_2 Y_t + \alpha D_t + \gamma(D_t Y_t) + \varepsilon_t$$

Bu durum hem eğimin hemde α katsayısının değişmesinin mümkün kılındığı modeldir. (Model tek bir denklemle ifade edilirken savaş ve barış yıllarında hata teriminin varyansı aynı kabul edilmiştir) Gelirin en küçük kareler tahminleri, tahmini regresyon katsayılarının dağılımının ve regresyon standart hatasının bireysel tahminleridir.

5. Durum

$$C_t = \beta_1^* + \beta_2^* Y_t + \varepsilon_t^* \quad \text{Savaş yıllarında}$$

$$C_t = \beta_1' + \beta_2' Y_t + \varepsilon_t' \quad \text{barış yıllarında}$$

Bu modelde, hata varyansmın savaş ve barış yıllarında değişikliğe uğrayacağı kabul edilmiştir. 5. durum iki farklı regresyon ile bunlara ilişkin standart hataların iki farklı tahminini içerir.

4. ve 5. durumlar için tahmini regresyon katsayılarının esas olarak eşdeğer olup olmadığı test edilebilir.

$$\hat{\beta}_1' = \hat{\beta}_1 \ , \ \hat{\beta}_1^* = \hat{\beta}_1 + \hat{\alpha} \ , \ \hat{\beta}_2' = \hat{\beta}_2 , \ \beta_2^* = \beta_2 + \hat{\gamma}$$

Model 4 yada model 5 in seçimi veriyi test etmeden önce tüm yıllarda hata varyansmm sabit olup olmadığına bağlıdır.

1.2. Faktörler ve Düzeyler :

Nitel değişkenler olarak meslek ve öğrenim durumunun yatırım davranışı üzerindeki etkisini incelemek istersek, her değişkenin her durumunun yatırımla birlikteliğinin ne olduğunu araştırmamız gerekecektir. Örneğin bir araştırma sonucunda aşağıdaki tablonun elde edildiğini kabul edelim.

TABLO : 1.1.

Yatırım Verilerinde Belirtilmiş Kişi Sayıları (Öğrenim ve Mesleklerine Göre Sınıflandırılmış)

Meslek	Liseden terk	Lise mezunu	Yüksek okul mezunu
İşçi	14	8	7
Sanatkâr	10	—	—
Serbest meslek	—	17	22
Küçük üretici	3	9	10

Bir kişinin sanatkâr olmasının yatırım üzerindeki etkisinin, küçük üretici olmasının yatırım üzerindeki etkisinin ne olduğunu görmek istiyoruz. Başka bir deyişle bu iki özelliğin etkileri arasındaki farkları, verilerimizi rastlantısal bir örnek olarak düşünüp Ana Kütle için incelemek istiyoruz.

Değişkenlerin ölçülmemesi ve buna bağlı olarak gelişigüzel oluşu kategorileri hakkında karar verirken bizi «faktör» ve «düzey» terimleriyle karşı karşıya getiriyor. Böylece meslek bir faktör, öğrenim bir diğer faktör oluyor. Her faktörün (değişkenin) içinde bölündüğü kategoriler o faktörün düzeyleri olarak adlandırılıyor.

«Değişken» yerine faktör kullanılması faktör denilenlerin sayısal değerlerle tam olarak ölçülememesini vurguluyor; «değişken» kelimesi ise ölçülebilenler için kullanılıyor. Bu durumda yukarıda ele aldığımız örnekte değişken olarak yalnız yatırım kalıyor. Araştırmanın diğer elemaniarı, her biri çeşitli düzeylerdeki faktörler oluyor. «Düzeyler» terimi faktörü guruplamanm, değerlerine bir yükleme olmaksızın, sadece gelişigüzel bölünmeler olduğunu vurguluyor.

Değişkenleri guruplama yerine faktör düzeyleri terimleriyle düşünmek belirlenmiş kodlar kullanmadan gelen pek çok zorluğun üstesinden gelir. Nitel bir değişkeni guruplarken sayıları önemli değildir, faktörün düzeyleri olarak düşünülebilir. Faktörün düzeylerinin etkileri arasındaki farkların tahmini gösterge (0,1) değişkenleriyle yapılan regresyonla elde edilebilir.

1.3. Regresyon :

Amacımız her faktörün düzeylerinin yatırım üzerindeki etkilerini ele almaktır. Sadece öğrenim durumunun yatırım üzerindeki etkisini tahmin ederek başlayalım. Bunu yapabilmek için de X_1, X_2 ve X_3 bağımsız değişkenleriyle bir regresyon denklemi kuruyoruz.

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + e_i \qquad (1.1.)$$

Bu denklem de,

Y_i : yatırımı
b_0 : sabit ve
e_i : regresyon analizinde bulunan hata terimidir.

Bağımsız değişken X'lere karşı gelenler de tanımlanacak olan b_1, b_2 ve b_3 regresyon katsayılarıdır. X'lerin tanımlanacağı durumda, b'ler de, öğrenim faktörünün düzeylerinin yatırım üzerindeki etkileri arasındaki farkların tahminine yönelen terimlere döneceklerdir.

X'leri tanımlarken, her birinin bir ve yalnızca bir öğrenim düzeyinde olabileceğini belirtelim. Hangi düzeyde ise, karşılık gelen X, 1 değerini ve diğer tüm X'ler o kişi için 0 değerini alsın.

Lise mezunu, öğrenim faktörünün ikinci düzeyindedir. Onun için $X_{i2} = 1$, $X_{i1} = 0$ ve $X_{i3} = 0$ değerlerini alacaktır. Bu yolla, nümerik değerler (0'iar ve 1'ler) verideki her kişi için, bütün X'lere verilebilir. Çünkü her X değeri, kişiye karşılık gelen öğrenim düzeyinde ise 1, diğer durumlarda 0'dır. X' (0,1) değişkenleri olarak tanımlanmıştır. Önceden tanımlandığı anlamda gerçek değişkenler olmayıp; «gösterge» değişkenler olarak adlandırılırlar. Bu durumda da bilinen regresyon işlemleri ilgili sonuçlarıyla tamamlanabilir.

Örnek : Aşağıdaki yatırımla ilgili verilere sahip olduğumuzu varsayalım.

3 kişi Liseden terk
2 kişi Lise mezunu
1 kişi Yüksek okul mezunu

Bu 6 gözlemin (yatırım indisleri) aşağıdaki gibidir.

TABLO : 1.2.

Öğrenim durumu	Yatırım indisi
Liseden terk	y_{11}, y_{12}, y_{13}
Lise mezunu	y_{21}, y_{22}
Yüksek okul mezunu	y_{31}

Y_{ij}, j inci kişinin i inci öğrenim düzeyinde olduğunu gösteriyor.

Bir yerine iki indis kullanılması hariç,

$$e_{ij} = y_{ij} - E(y_{ij})$$

eşitliğiyle, regresyonda olduğu gibi, gözlemleri aşağıdaki gibi yazabiliriz.

$$y_{11} = b_0 + b_1(1) + b_2(0) + b_3(0) + e_{11}$$

$$y_{12} = b_0 + b_1(1) + b_2(0) + b_3(0) + e_{12}$$

$$y_{13} = b_0 + b_1(1) + b_2(0) + b_3(0) + e_{13}$$

$$y_{21} = b_0 + b_1(0) + b_2(1) + b_3(0) + e_{21}$$

$$y_{22} = b_0 + b_1(0) + b_2(1) + b_3(0) + e_{22}$$

$$y_{31} = b_0 + b_1(0) + b_2(0) + b_3(1) + e_{31}$$

Parantez içindeki 1 ve 0'lar (0,1) gösterge değişkeninin değerleridir. Modelimiz denklemler aşağıdaki gibi yazılırsa daha açık görülebilir.

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{31} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{31} \end{bmatrix} \qquad (1.2)$$

$$y = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{31} \end{bmatrix}, \quad e = \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{31} \end{bmatrix}, \quad b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \text{ ve } \quad x = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \qquad (1.3)$$

şeklinde yazılarak denklemler bilinen,

$$y = X\hat{b} + e$$

formuna getirilir. Bu denklemdeki e terimlerinin özellikleri, tam olarak regresyondaki gibidir. Yani,

$$e \sim (0, \sigma^2 I),$$

Formül 2.4'e uygulanan en Küçük Kareler Yöntemi, normal denklemleri sağlar.

$y = Xb + e$ denkleminde e, bileşenleri, gözlem değerleri ile tahmin değerleri arasındaki farklardan (kalıntı) oluşan tek sütunlu bir matris veya vektördür.

Kalıntıların karelerin toplamını veren,

$$e'e = (e_1 \; e_2 \; \ldots \; e_n)\begin{pmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{pmatrix} = \Sigma e_i^2$$

ifadesini 1.4 no.'lu formülden yararlanarak aşağıdaki biçimde gösterebiliriz.

$$e'e = (y-X\hat{b})'(y-X\hat{b})$$

$$= y'y - 2\hat{b}'X'y + \hat{b}'(X'X)\hat{b}, \quad \hat{b}'X'y = y'X'\hat{b}$$

skaler olduğundan,

$$= y'y - \hat{b}'(X'X)\hat{b}$$

En Küçük Kareler yöntemi ile tahmin yapmak için yukardaki denklemin b ye göre türevi alınır ve sıfıra eşitlenirse,

$$\frac{\partial(e'e)}{\partial\hat{b}} = -2X'y + 2X'X\hat{b} = 0$$

$X'X\hat{b} = X'y$ elde edilir.

Bununla beraber, X (1.3.) nolu denklemde görüldüğü gibi sütun ranka sahip değildir. Son üç sütunun toplamı, birinci sütuna eşittir. Böylece model «rankı tam olmayan model» olarak tanımlanır. Özelliği, X in rankının sütün sayısına eşit olmaması ve bunun önemli sonucu $(X'X)^{-1}$ in varolmadığı ve dolayısıyla,

$X'X\hat{b} = X'y$ eşitliğinin

$\hat{b} = (X'X)^{-1} X'y$ olarak çözülemiyeceğidir.

Buna rağmen X'X'in genelleştirilmiş tersi kullanılarak çözümler bululunabilir. Bu çözümlere değinmeden, bir başka örnek verelim ve lineer modellerin başka görünüşlerini ele alalım.

Örnek :

Her yıl tarım alanında sayısız deneyler yapılmakta ve bilim adamları çeşitli türlere uygulanan farklı gübrelerin ürün üzerindeki etkilerini araştırmaktadırlar. 6 değişik tarlada iki çeşit gübre kombinasyonu ile test edilmiş 3 çeşitten oluşan bir örneğimiz olduğunu varsayalım. Deneyin, çeşitlerinin gösterildiği satırlardaki tarlaların büyümesiyle bağlanması zorunlu değilse de datayı Tablo 1.3 deki gibi canlandırmak uygundur.

6 ÜRÜNE AİT TABLO

TABLO 1.3

	Uygulama	
Çeşit	1	2
1	y_{111}, y_{112}	y_{121}
2	y_{211}	y_{221}
3	y_{311}	

Tablodaki y_{ijk} gibi girişler k'mcı alandaki, i inci çeşidin j inci uygulamaya ait olan ürünü göstermektedir.

Biz bunları 5 tane (0,1) gösteren değişkeni ve 3 çeşit, 2 uygulamaya karşılık olan 5 regresyon katsayısı kullanarak yazabiliriz. 3 çeşit için, regresyon katsayıları α_1, α_2 ve a_3 ile uygulamalar içinde β_1 ve β_2 ile gösterilirse, Regresyon denklemindeki daha önce b_0 ile gösterilen sabit terim μ olarak yazılırsa, böylece katsayıları vektörü b';

$$\hat{b}' = \{\hat{\mu}\ \hat{\alpha}_1\ \hat{\alpha}_2\ \hat{\alpha}_3\ \hat{\beta}_1\ \hat{\beta}_2\} \text{ olur.}$$

(Bu notasyon çeşitler için, regresyon katsayılarıyla (α lar) uygulamalar için olanlar (β'lar) arasında açıkça ayrım yapar ve b'nin elemanları olarak b'ler kullanılmasına karşıt olarak, açıklık sağlayan çift indis kullanmaktan sakınılmış olur).

Bu notasyonla y_{ijk} için regresyon denklemi;

$$y_{ijk} = \hat{\mu} + \hat{\alpha}_1 x_{ijk,1} + \hat{\alpha}_2 x_{ijk,2} + \hat{\alpha}_3 x_{ijk,3} + \hat{\beta}_1 x^{*}_{ijk,1} + \hat{\beta}_2 x_{ijk,2} + e_{ij}$$

olur.

X ve X*'lar (0,1) gösterge değişkenleridir.

1 inci çeşit ve 2 inci uygulama için yapılan gözlemde;

$$x_{121,1}=1 \qquad x_{121,2}=0 \qquad \text{ve} \qquad x_{121,3}=0$$

ve

$$x^*_{121,1}=0 \qquad x^*_{121,2}=1 \quad \text{olacaktır.}$$

Tablodaki ürünler için regresyon denklemleri,

$$\begin{aligned}
y_{111} &= \hat{\mu} + \hat{\alpha}_1(1) + \hat{\alpha}_2(0) + \hat{\alpha}_3(0) + \hat{\beta}_1(1) + \hat{\beta}_2(0) + e_{111} \\
y_{112} &= \hat{\mu} + \hat{\alpha}_1(1) + \hat{\alpha}_2(0) + \hat{\alpha}_3(0) + \hat{\beta}_1(1) + \hat{\beta}_2(0) + e_{112} \\
y_{121} &= \hat{\mu} + \hat{\alpha}_1(1) + \hat{\alpha}_2(0) + \hat{\alpha}_3(0) + \hat{\beta}_1(0) + \hat{\beta}_2(1) + e_{121} \\
y_{211} &= \hat{\mu} + \hat{\alpha}_1(0) + \hat{\alpha}_2(1) + \hat{\alpha}_3(0) + \hat{\beta}_1(1) + \hat{\beta}_2(0) + e_{211} \\
y_{221} &= \hat{\mu} + \hat{\alpha}_1(0) + \hat{\alpha}_2(1) + \hat{\alpha}_3(0) + \hat{\beta}_1(0) + \hat{\beta}_2(1) + e_{221} \\
y_{311} &= \hat{\mu} + \hat{\alpha}_1(0) + \hat{\alpha}_2(0) + \hat{\alpha}_3(1) + \hat{\beta}_1(1) + \hat{\beta}_2(0) + e_{311}
\end{aligned} \qquad (1.5)$$

şeklini alır. Gözlem vektörleri ve hata terimleri için y ve e kullanılarak denklemler;

$$y = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} + e. \qquad (1.6)$$

şeklini alır.

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \quad \text{ve } b = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

yazılır.

X tam ranka sahip değildir: 2,3 ve 4. sütunların toplamı, 1. sütuna eşittir. 1.6 nolu denklem ile daha önceki gibi $y = xb + e$, modelin denklemleri tam ranka sahip değildir. Genel olarak, tüm elemanları 0 ve 1 olan X matrisli bir «isabet matris»i «Tasarım matrisi», yada daha uygun bir deyimle bir «gösterge matris» tir. Çünkü elemanları arasındaki 1'lerin varlığı, datadaki modelin terimlerinin (μ, α'lar ve β lar) isabetini tanımlar.

2. *LİNEER MODELLERİN TANIMLANMASI*

2.1. Bir Yönlü Sınıflama:

Yukardaki örneğin (1.2) ve (1.6) mcı denklemleri gösterge değişkenlerle regresyon açısından geliştirilmiştir. (1.4) denklemlerini tekrar ele alalım. Tablo 1.2'de gösterilen, 3 farklı öğrenim düzeyindeki 6 kişinin yatırım indisleriyle ilgilidir. 1.2 nolu denklemin tekrar aşağıdaki gibi yazıldığını varsayalım.

$$\begin{aligned} y_{11} &= \hat{\mu} + b_1 + e_{11} \\ y_{12} &= \hat{\mu} + b_1 + e_{12} \\ y_{13} &= \hat{\mu} + b_1 + e_{13} \qquad (1.8) \\ y_{21} &= \hat{\mu} + b_2 + e_{21} \\ y_{22} &= \hat{\mu} + b_2 + e_{22} \\ y_{31} &= \hat{\mu} + b_3 + e_{31} \end{aligned}$$

x ler açıkça gösterilmemiştir. μ, b_0 yerine yazılmıştır. (1.8) nolu denklemlerin her birinde b'deki indisin y'nin ilk indisine tam olarak karşı geldiğini görüyoruz. Yani b_1 ; y_{11}, y_{12} ve y_{13} de

b_2 ise y_{21} ve y_{11} de bulunuyor.

(1.8) nolu denklemlerdeki her denklem i ve j nin datadaki çeşitli değerleri için aşağıdaki gibi yazılabilir.

$$y_{ij} = \hat{\mu} + b_i + e_{ij} \qquad (1.9) \qquad (1.9)$$

Bu durumda, i=1, 2, 3 dür, ve i inci sınıftaki j'nin üst limiti, i inci sınıftaki gözlem sayısıdır. Bunu n_i ile göstererek $j=1, 2, \ldots n_i$ $n_1=3$, $n_2=2$ $n_3=1$ elde ederiz.

Böylece 3 sınıf için (1.9) denklemlerini genel lineer modelin denklemleri olarak geliştirmiş oluyoruz. a smıf için i=1, 2, ... a uygulanır.

Özel değerleri (1.2) denkleminde gösterilmiş olan (1.9) denklemleri, lineer model denklemlerinin genel biçimi ise de, tam anlamıyla regresyon çerçevesinde geliştirilmiştir. Bununla beraber ne b'nin elemanlarına regresyon katsayıları olarak, ne de X'in 0 ve 1 lerine gösterge değişkenler olarak bakmaya gerek yoktur. b'nin elemanlarına kendi değerlerinin anlamı verilebilir ve X'in 0 ve 1 leri de faktör düzeylerinin «varlığı» ve «yokluğu» ile ilgilidir.

μ (1.8) denklemlerinin her birine girdiği için, yatırım indisinin genel kütle ortalaması olarak tanımlanır. Öğrenim durumuyla ilgili olmaksızın, tüm değerlerle ilgili ortalamayı temsil eder.

b'lere anlam vermek üzere b_1'i ele alalım. (1.8) denklemlerinde (veya eşdeğer olan (1.2) denklemlerinde) b_1 yalnızca öğrenim durumu 1 olanların (liseyi tamamlıyamayanlar) yatırım indisiyle ilgili olan denklemlerde görülmektedir. Yani y_{11}, y_{12} ve y_{13} de.

Benzer şekilde b_2 yalnız öğrenim durumları 2 olanlarla ilgili denklemlerde vardır. Yani y_{21} ve y_{22} b_3 de y_{31} için olan denklemde vardır.

Böylece b_1, öğrenim durumu 1 olanların yatırım üzerindeki etkisi olarak tanımlanır. Benzer tanımlama b_2 ve b_3 içinde uygulanır.

Genel olarak (1.9) denklemiyle ilgili olarak b_i, yatırım üzerinde öğrenim durumu i'ye bağlı olan etki olarak tanımlanır. (1.2), (1.8) ve (1.9) nolu denklemler örnek için kurulmuşlardır. Denklemlerde bulunan e_{ij} hata terimleri ε_{ij} lerin tahminidir.

ε_{ij} lerin dağılım özellikleri belirlenmeksizin lineer modelin tanımlanması tamamlanmamıştır. Bu iş genellikle regresyon analizindeki benzer özelliklerin ε_{ij} lere atfedilmesiyle yapılır. ε_{ij} :

$$\varepsilon_{ij} = y_{ij} - E(y_{ij}) \text{ olarak tanımlanır.}$$

ve böylece,

$$E(\varepsilon_{ij}) = 0 \;;$$

$$E(y_{ij}) = \mu + \beta_i$$

eşitliğini verir.

Her ε_{ij}'nin varyansı σ^2 olarak tanımlanır ve

tüm i ve j ler için.

$$V(\varepsilon_{ij}) = E\{\varepsilon_{ij} - E(\varepsilon_{ij})\}^2 = E(\varepsilon_{ij}^2) = \sigma^2$$

Farklı ε'lerin tüm çiftleri arasındaki kovaryanslar sıfır alınır;

i=i' ve j=j' oldukça covarvans, varyansa eşit olur.

$$\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i'j}) = 0$$

Böylece;

$$\mathrm{Var}(\varepsilon) = \sigma^2 I \qquad \text{olur.}$$

Tek yönlü sınıflama modelinin genel tanımlanması aşağıdaki gibi özetlenebilir. i inci sınıftaki j inci gözlem y_{ij} için modelin denklemi (2.9) nolu denklem,

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

μ : genel ortalama
b_i : y_{ij} üzerindeki i inci sınıfa bağlı olan etki
e_{ij} : $e \sim (0, \sigma^2 I)$ olarak
y_{ij}'ye mahsus rastlantısal hata terimidir.

a sınıf için $i = 1, 2, \ldots a$ ve

i. sınıf için $j = 1, 2, \ldots n_i$ olur.

Hipotez testi ve güven aralıkları ele alındığı zaman ε ların normal dağıldığı kabul edildiğinde,

Yani:

$$\varepsilon \sim N(0, \sigma^2 I) \quad \text{dır}$$

2.2. *İki Yönlü Sınıflama:*

(1.5) nolu denklemler önceki örnekteki (1.2) denklemleri gibi, daha açık olmayarak aşağıdaki gibi yeniden yazıldığını, varsayalım; böylece (1.5) nolu denklemler,

$$\begin{aligned}
y_{111} &= \mu + \alpha_1 + \beta_1 + \varepsilon_{111} \\
y_{112} &= \mu + \alpha_1 + \beta_1 + \varepsilon_{112} \\
y_{121} &= \mu + \alpha_1 + \beta_2 + \varepsilon_{121} \\
y_{211} &= \mu + \alpha_2 + \beta_1 + \varepsilon_{211} \\
y_{221} &= \mu + \alpha_2 + \beta_2 + \varepsilon_{221} \\
y_{311} &= \mu + \alpha_3 + \beta_1 + \varepsilon_{311}
\end{aligned} \qquad (1.10)$$

haline gelir.

Buradaki her denklem α ve β ların indisleri y'nin ilk iki indisine sırayla karşılık gelir :

α_1 ve β_1 y_{111} ve y_{112}'de bulunur.

α_2 ve β_1 y_{211}de

(1.10) nodaki her denklem aşağıdaki gibi yazılabilir :

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \qquad (1.11)$$

i, j k'nın datada aldığı değerler, bu örnekte

i=1, 2, 3 ve j=1, 2 k'nın üst limitiyle j inci uygulamanın tatbik edildiği i inci çeşidin gözlem sayısı olur. Bunu n_{ij} ile gösterirsek,

$k=1,2,\ldots n_{ij}$ ve $n_{11}=2$, $n_{12}=1$, $n_{21}=1$

$n_{22}=1$ $n_{31}=1$ ve $n_{32}=0$ olacaktır.

Böylece (1.11) denklemleri, çeşitli ve uygulamaları içeren genel lineer modelin denklemleridir.

Önceki bölümde ki tek yönlü sınıflandırmada olduğu gibi, b'nin elemanlarının (bu örnekte μ, α'lar ve β lar) regresyon katsayıları gibi görülmelerine gerek yoktur, kendi değerleriyle anlam verilebilir.

Önce μ yü ele alalım. Bütün ürün kütlesinin ortalaması olarak ele alınabilir. Uygulama veya çeşide bağlı olmaksızın bütünü kapsayan ortalamayı temsil eder. İkinci olarak α'lar (1.10) denklemlerinde {veya eşdeğer olan (1.5) denklemlerinde} α_1 yalnızca 1 inci çeşidin ürünleriyle ilgili olan denklemlerde vardır; yani y_{111}, y_{112} ve y_{121}. Benzer şekilde α_2, 2 inci çeşidin ürünleriyle ilgili denklemlerde görülür; y_{211} ve y_{221}. α_3 de aynı şekilde y_{311} için olan denklemdedir, bunun dışında da yer almaz. Böylece α_1, 1 inci çeşit olan alanların ürün üzerindeki etkisi olarak tanımlanır. Benzer tanımlama α_2 ve α_3 içinde uygulanır. Genel olarak α_i, üründeki 1 inci çeşide bağlı olan etki diye tanımlanır. Benzer şekilde β ları alalım :

β ların genel tanımlanması α lara benzer; her ikisi de ürün üzerinalır : y_{111}, y_{112}, y_{211} ve y_{311}.

β_2 ise; yalnızca 2 inci uygulama ile ilgili denklemlerdedir : y_{121} ve y_{221}. Böylece β_j ürünündeki j inci uygulamaya bağlı olan etki olarak tanımlanır.

β ların genel tanımlanması α lara benzer; her ikiside ürün üzerindeki etkilerdir, fakat α lar çeşide, β lar uygulamaya bağlı olan etkilerdir.

Modeldeki hata terimleri ε_{ijk}'ların önceki gibi aynı özelliklere sahip olduğu varsayılmıştır.

Yani ε, ε_{ijk} ların vektörü ise;

$\varepsilon \sim (0, \sigma^2 I)$ dır.

Hipotez testleri ve güven aralıkları için normallik ek varsayımı ile beraber ele alınır.

μ ve ε_{ijk} den ayrı olarak (1.11) denklemlerinin terimleri genel olarak α ve β faktörlerine ısnat edilebilecek, sadece iki faktör içindir. Denklemi (1.11) olan model iki yönlü sınıflama adı daha kesinlikle saptanmış ise de «İki Faktör Modeli» olarak adlandırılır. Genel tanımlanması aşağıdaki gibidir. α inci düzeyindeki k'mcı gözlemi olan y_{ijk} ve β faktörünün j inci düzeyi için modelin denklemi :

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

μ ; genel ortalama, α_i; α faktörünün,

i inci düzeyine bağlı olan y_{ijk} üzerindeki etkisi,

β_j ; faktörünün j inci düzeyine bağlı etki ve

ε_{ijk} ; y_{ijk} nın rastlantısal hata terimi, ki :

$\varepsilon \sim (0, \sigma^2 I)$

α faktörü a sayıda düzeyde $i = 1, 2, \ldots a$,

β faktörü de b sayıda düzeye $j = 1, 2, \ldots b$ sahip olduğu zaman, $k = 1, 2, \ldots n_{ij}$ için, (i, j) çiftlerinde veya alt sınıflarındaki n_{ij} gözlem için «kesişim» α faktörünün I inci düzeyi ve β faktörünün j inci düzeyidir.

Güven aralıkları ve hipotez testi için,

$\varepsilon \sim N(0, \sigma^2 I)$ dır.

Model için burada tanımlanan örnek tarımdan alınmıştır, fakat aynı tür bir model iki faktör içeren diğer durumlara da uygulanabilir. Tablo 1.1 deki yatırım üzerindeki öğrenim ve mesleklerin etkisi ile ilgili örnek için (1.11) denklemleri tarım örneğine eşdeğer olarak uygulanabilirdi. Böylece α_i, i inci meslek kategorisinin yatırım üzerindeki etkisi (meslek faktörünün i inci düzeyi) ve β_j öğrenim faktörünün j inci düzeyinin etkisi olurdu.

2.3. *Üç Yönlü Sınıflama :*

Tablo I'deki data için yerleşim bölgelerinin de her gözlem için kaydedildiğini kabul edelim. Böylece, yatırım üzerindeki meslek, öğrenim ve bölgelerin etkisi üzerindeki çalışma, denklemi;

$$y_{ijkh} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijkh} \tag{1.12}$$

olan bir model kullanılarak yapılabilir. Burada y_{ijkh} : j inci öğrenim durumundaki, i inci meslekten, k'ıncı bölgede yaşayan h'ıncı şahsın yatırım indisidir.

μ genel ortalama, α_i; yatırım üzerindeki i inci mesleğe bağlı etki ve γ_k da k'ıncı bölgeye bağlı etkidir. Her zaman olduğu gibi ε_{ijkh}, y_{ijkh}'a ait hata terimidir ve $\varepsilon \sim (0, \sigma^2 I)$ dır. Eğer datada a sayıda meslek düzeyi, b sayıda öğrenim düzeyi ve c sayıda bölge varsa;

$$i = 1, 2, \ldots a \; ; \qquad j = 1, 2, \ldots b \; ; \qquad k = 1, 2, \ldots c$$

olur. Datanın alt sınıflarındaki n_{ijk} gözlem için,

$h = 1, 2, \ldots n_{ijk}$ i'inci meslek, j'inci öğrenim düzeyi, k'ıncı bölge temsil edilmektedir.

Bu yapıdaki modellerin 4 yönlü ve daha yüksek derecede sınıflamalar için genişletilmesi açıkça anlaşılabilir.

2.4. *Temel Etkiler ve Karşılıklı Etkiler*

2.4.1. *Temel Etkiler :*

α, β ve γ ların her biri bir faktörün bir düzeyinin y üzerindeki etkilerini göstermektedir. Tablo 1.3 ün iki yönlü sınıflanmasında (1.11) nolu denklemdeki α_i çeşit faktörünün i'inci düzeyinin ürün üzerindeki etkisini gösterir, yani i'inci çeşidin etkisini gösterir.

Aynı denklemdeki β_j j'inci uygulamanın ürün üzerindeki etkisini gösterir. Bu yapıdaki faktörün tek bir düzeyi ile ilgili etkiler ***TEMEL ETKİLER*** olarak adlandırılır. Bu mantıklıdır. Ürün üzerindeki çeşit ve uygulamanın etkileri, temel ilgili olduğumuz etkendir. Dolayısıyla, modelin bunlara karşılık olan elemanları, modelin temel etkileri olarak adlandırılır.

Modelin denklemi yapısı gereği, y_{ijk}'mn beklenen değerini tahmin ederken, a_i etkisinin β_j etkisine ekleneceğini gösterir.

$$E(y_{ijk}) = \mu + \alpha_i + \beta_i \quad (1.13)$$

Bunun anlamı beklenen ürün üzerindeki i'inci çeşit ve j'inci uygulamanın toplam etkisinin, α_i ve β_j bireysel etkilerinin toplamı olarak ele alındığıdır. Bu nedenle etkiler toplamsal olarak tanımlanır. Modelin bir anlamı da beklenen ürün üzerindeki i'inci çeşidin etkisinin hangi uygulamanın kullanıldığını dikkate almadan hep aynı olarak ele almasıdır. Bütün uygulamalar için i'inci çeşidin etkisi α_i olarak varsayılmıştır, i'inci çeşitle j'inci uygulamanın birleştirilmiş etkisi (μ ortalamasının üzerindeki etki) $\alpha_i + \beta_j$ olarak alınmıştır.

μ, α ve β değerlerini aşağıdaki gibi kabul edelim,

$$\begin{array}{lll} \mu=4 & \alpha_1=1 & \beta_1=4 \\ & \alpha_2=3 & \beta_2=7 \\ & \alpha_3=2 & \end{array} \quad (1.14)$$

Formül 1.13'ün μ, α ve β elemanları için bu değerler tasvir amacıyla verilmiştir. Gözlenmiş değerler değildirler. Aslmda bu elemanlar hiçbir zaman gözlenemezler ve pratikte hiç bilinmezler. Yalnızca uygun datalardan tahmin edilebilen ana kütle değerleridirler. Buna rağmen lineer modellerin belirli yanlarını tasvir etmek amacıyla, sonuçlar grafik olarak çizilebilsin diye, bu elemanlara aritmetik değerler verilmiştir.

Örneğin :

Varsayılan bu değerlerle;

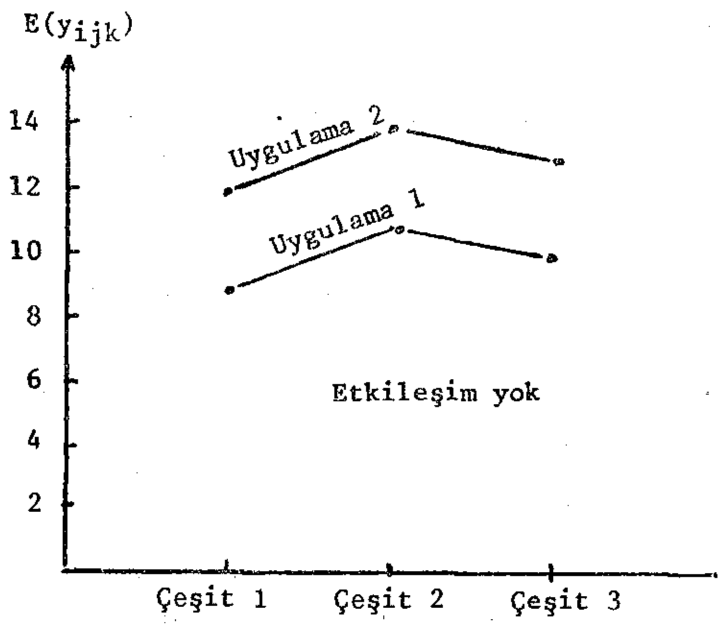
$$E(y_{ijk}) = \mu + \alpha_i + \beta_j = 4 + 1 + 4 = 9 \quad (1.15)$$

Bu E (y_{ijk}) nın veya y_{ijk} nm kendi gözlenmiş değeri değildir. Paremetrelerin (1.14) de verilen varsayılmış değerlerine dayanan E (y_{ijk}) nın varsayılan değeridir.

Önce (1.13) denkleminin i ve j'nin verilen değerlerine göre tüm k'lar için aynı olduğunu not edelim. k indisi sadece, (i, j) alt sınıflarındaki bireysel gözlemlerin ayırıcı (teşhis edicisi) olduğundan (1.13) denklemi, 0 alt sınıflardaki her gözlemin beklenen değerlerinin aynı olduğu anlamını verir. Böylece (1.15) nolu denkleme göre (1,1) gözündeki her gözlemin beklenen değeri 9'dur. Yani k'nın her $k=1, 2, \ldots n_{11}$ değeri için $E(y_{11k}) = 9$ olur. Bu yorumla (1.14) den türetilmiş diğer alt sınıflar için beklenen değerler Tablo 1.4'de gösterilmiş ve Şekil 1.1 de çizilmiştir.

TABLO : 1.4

	Uygulama	
Çeşit	1	2
1	$E(y_{11k})=4+1+4-1=8$	$E(y_{12k})=4+1+7+1=13$
2	$E(y_{21k})=4+3+4+0=11$	$E(y_{22k})=4+3+7-5=9$
3	$E(y_{31k})=4+2+4-2=8$	$E(y_{32k})=4+2+7-3=10$



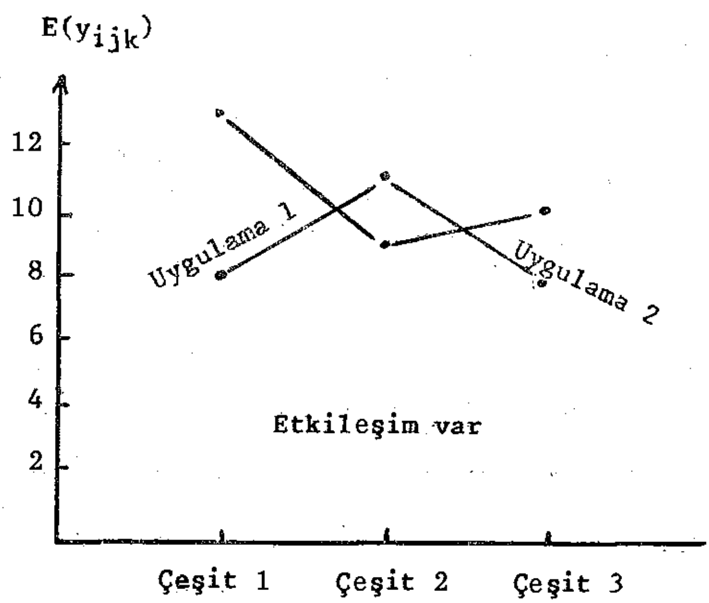
Şekil : 1.1

Şekil 1.1 de absis değişkeninin çeşit numarasının sürekli bir değişken olmadığını belirtelim E (y_{ijk}) doğrularının ortak değerleri, bir çeşitten diğerine (E (y_{ijk})'da sürekli bir değişmeyi göstermezler. Doğrular sadece değişmedeki trendi vurgulamak için gösterilmiştir. Benzer şekilde Şekil 1.2, 1.3 ve 1.4 de de kullanılmıştır. Şekillerde gösterilen noktalar (ordinatlar) E (y_{ijk}) nın değerleridir. y_{ijk} nın gözlem değerleri değildir.

(1.13) nolu denklem ile verilen modelin bu tipik örneğinde çeşitlerin etkisinin (uygulama ile ilgili olmaksızın)' aynı olduğu Şekil 4.1 den de açıkça bellidir. Her iki uygulama içinde, 2'inci çeşit, l'inci çeşitten 2 birim fazla beklenen ürüne sahiptir ve gene her iki uygulama için 3'üncü çeşit 2'inci çeşitten bir birim daha azdır.

2.4.2. *Karşılıklı Etkiler:*

Bir diğer tipik örnekte, beklenen ürünlerin Şekil 2.2 de gösterildiği gibi gerçekleştiğini kabul edelim.



Şekil: 1.2

Şekil 1.1 ile olan farklılık açıktır. Uygulama için çizilen doğrular paralel değildir. Bu çeşitlerin etkisinin, farklı uygulamalar için farklı sonuçlar verdiğini gösterir.

l'inci uygulamada 2'inci çeşit l'inci çeşitten 3 birim fazladır. (Beklenen üründe) Fakat, 2'inci uygulamada 2'inci çeşit 1'inci çeşitten 4 birim daha azdır. Bu ikinci *tipik* örnekte, çeşitler, hangi uygulamanın kullanıldığına bağlı olarak farklı şekilde etki etmektedirler. Çeşitlerin uygulamalarla «karşılıklı etkilendiğini» söyleriz. Her uygulama için aynı şekilde davranmamaları «karşılıklı etkilenme» deyimi ile ifade edilebilir.

Bu tartışma başka şekilde de ortaya konabilirdi. Şekil 1.4 de uygulamalar arasındaki fark her çeşit için aynıdır. Çeşitten çeşide değişmez ve bütün çeşitler için sabittir. Bu şekil 1.1 deki doğruların paralelliği ile kanıtlanır. Diğer yandan, Şekil 1.2 de paralelliğin yokluğu, uygulamalar arasındaki farkların çeşitten çeşide fark ettiğini gösterir. «Uygulama 1, Uygulama 2» olarak tanımlanan fark, üç çeşit için sırayla -5, $+2$, ve -2, şekil 4.1 için ise her çeşit için -3'dür. İki *tipik* örnek arasındaki bu fark, Şekil 1.3 ve Şekil 1.4 deki gibi çizildiğinde daha iyi gösterilir.

Şekil 2.3 deki paralel doğrular (Şekil 1.1 dekine karşılık olanlar) Tablo 1.4 deki ilk tipik örnek için bütün çeşitlerin uygulamaları arasında tekdüze bir fark göstermektedir. Fakat şekil 1.4 deki paralel olmayan doğrular, ikinci örnek için bütün çeşitlerde uygulamalar arasındaki farklar da bir benzerlik olmadığını göstermektedir.

Bu tartışmadan, Şekil 1.1 ve 1.3 de (Tablo 1.4) çeşitin beklenen ürün üzerindeki etkisinin bütün uygulamalar için aynı olduğu ve uygulamanın etkisinin de tüm çeşitler için aynı olduğu açıkça görülmektedir. Bu durum, Tablo 1.4 ve Şekil 1.1 ile 1.3'e temel olarak kullanılan (1.13) nolu denklemin formunda da açıktır. Fakat Şekil 2.2 ve 1.4'de uygulamanın etkisi tüm çeşitler için ve çeşitin etkisi, tüm uygulamalar için aynı değildir. Uygulamalar ve çeşitlerin karşılıklı etkileşmelerini izah edecek bazı ek etkiler vardır. Bu etkiler karşılıklı etkileşmeler olarak adlandırılır ve bir temel etkinin (çeşit) her düzeyinin, bir diğer temel etkinin her düzeyiyle karşılıklı etkileştiği durumu tmsil eder. Bu etkiler modelin denklemine bir başka terim ilavesiyle dikkate alınır. α etkisinin i'inci düzeyi ile, β etkisinin j'inci düzeyi arasındaki karşılıklı etki Y_{ij} ise; model denklemi;

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij} \qquad (1.16)$$

veya eşdeğer olarak,

$$y_{ijkh} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_{ij} + \hat{\gamma}_{ijk} + e_{ijkh} \qquad (1.17)$$

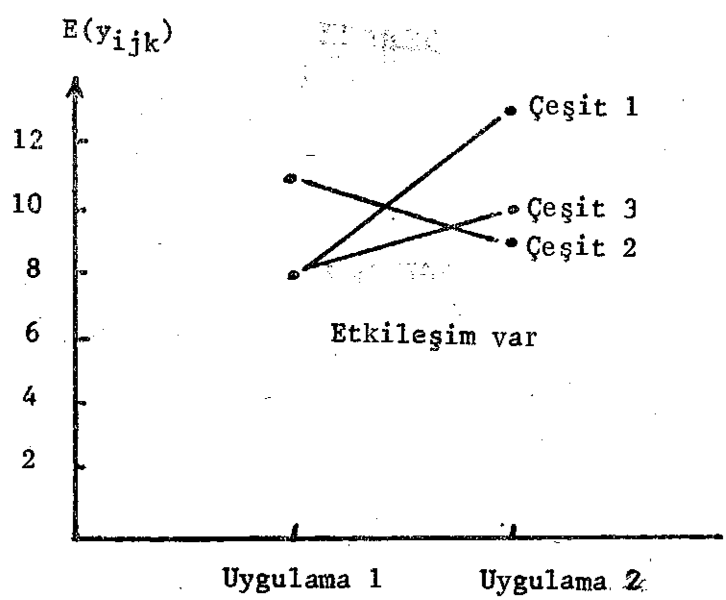
olur.

Diğer tüm elemanlar önceki aynı anlamlara sahiptirler.

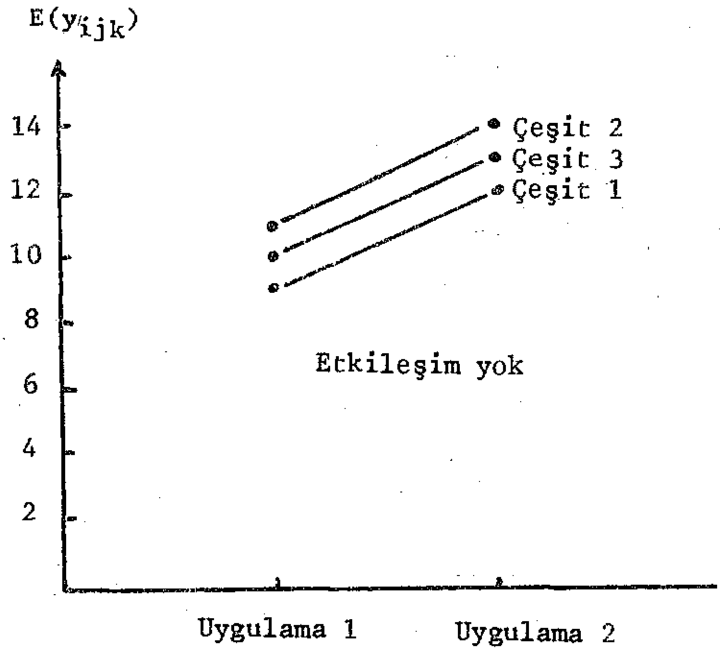
İkinci *tipik* örnek (Şekil 1.2 ve 1.4'deki) karşılıklı etkileşme γ_{ij}'nın aşağıdaki tipik değerleriyle beraber, μ, α ve β'lar için (1.14) nolu eşitliklerde verilen aynı *tipik* değerlere dayanır.

$$\begin{array}{ll} \gamma_{11} = -1 & \gamma_{21} = 1 \\ \gamma_{12} = 0 & \gamma_{22} = -5 \\ \gamma_{13} = -2 & \gamma_{31} = -3 \end{array} \qquad (1.18)$$

Bu yolla (1.16) nolu denklemlerden türetilen beklenen değerler Tablo 1.5 de gösterilmiş ve Şekil 1.2 ve 1.4'de çizilmiştir.



Şekil : 1.3



Şekil: 1.4

Karşılıklı etkileşme Modelinin Beklenen Değerleri :

(1.16) nolu denklemlerde yerine konmuş (1.14) ve 1.18) nolu eşitlikler.

TABLO: 1.5

Çeşit	1	2
1	$E(y_{11k})=4+1+4=9$	$E(y_{12k})=12$
2	$E(y_{21k})=11$	$E(y_{22k})=14$
3	$E(y_{31k})=10$	$E(y_{32k})=13$

Karşılıklı etkileşimin bu açıklamasının bütünüyle beklenen ürünlere dayanılarak yapıldığına dikkat edelim. Bu tarz modeller, ilgilenilen datanın tasvir edilen biçimde davrandığını düşündüğümüz zaman kullanılır. Fakat örnekte kullanılan basit sayılar datadan değildir, sadece modele örnek teşkil eder.

Karşılıklı etkileşim modeli (1.17) nolu denklemde tüm gözlerdeki $n_{ij} = 1$ olduğu zaman karşılıklı etkileşimin olmadığı modellerden {(1.11) nolu denklem} ayırdedilemez hale gelir. (1.17) nolu denklemin γ_{ij} ve ε_{ijk} terimleri, birlikte,

$$\gamma_{ij} + \varepsilon_{ijk} = \varepsilon_{ij}$$

olur ve (1.19) nolu denklem (1.13) nolu denkleme eşdeğer olan,

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_i$$

haline gelir.

(Burada bütün gözler için 1 değerini aldığından k indisine gerek yoktur).

Bunun anlamı, $n_{ij} = 1$ olduğu zaman, karşılıklı etkileşim olmayan modellere gelineceğidir.

Karşılıklı etkileşimin bu özet tartışmasının genelleştirilmesi açıktır. γ_{ij} iki faktör arasındaki etkileşimdir ve birinci derece karşılıklı etkileşim olarak bilinir. 3 faktör arasındaki etkileşime de ikinci derece etkileşim denir. 3, 4 ve daha yüksek dereceden etkileşimler benzer şekilde takip eder. Daha yüksek dereceler daha zor yorumlar gerektirir. Örneğin, 3 üncü derece etkileşim (4 temel etki arasındaki etkileşimi içeren), bir temel etki ile 2 inci derece etkileşim arasındaki etkileşim olarak, veya iki tane 2 inci derece etkileşim arasındaki etkileşim olarak yorumlanabilir.

2.4.2.1. *Notasyon :*

Etkileşimi yorumlamaya yardımcı olan, sık kullanılan bir notasyon γ_{ij} $(\alpha\beta)_{ij}$ sembolünü kullanmaya dayanır.

Bu sembol $(\alpha\beta)_{ij}$; α faktörünün i inci düzeyi ile β faktörünün j inci düzeyi arasındaki etkileşimidir. Bu sembol herhangi bir α ve β çarpımını göstermez. (a β_{ij} şeklinde parantezsiz yazıldığı zaman hariç). α ve β faktörlerinin düzeyleri arasındaki etkileşimi gösteren γ_{ij}'den, daha açıkça gösteren birleşik bir semboldür. Bu anlamı ile, yüksek derece bir etkileşimin, örneğin $(\alpha \beta \gamma \delta)_{hijk}$'nın yorumu şöyledir :

α_h ile $(\beta\gamma\delta)_{ijk}$ arasındaki etkileşim, veya

$(\alpha\beta)_{hi}$ ile $(\gamma\delta)_{jk}$ arasındaki etkileşim,

veya diğer yorumlardan biri. Bu notasyon modelin yazılışını da açıklar;

$$y_{i,jkm} = \mu + \alpha_i + \beta_j + \gamma_k + \tau_{ij} + \rho_{ik} + \theta_{jk} + \xi_{ijk} + \varepsilon_{ijkm}$$

modeli,

$$y_{ijkm} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkm}$$

modeli kadar açıklayıcı değildir.

Son olarak modelde karşılıklı etkileşimler olsa da, modelin derecesinin sahip olduğu temel etki faktörlerinin sayısıyla tanımlandığını belirtelim (1.11) no.lu denklem gibi. (1.16) no.lu denklem de iki yönlü sınıflama için bir denklemdir. Fakat (1.16) no.lu denklem karşılıklı etkileşimler içerdiği halde (1.11) no.lu denklem içermez.

2.5. *Küme ve Çapraz Sınıflamalar*

Tablo 1.3'deki örnekte her uygulama, her çeşide uygulanmıştır. Aslında, 3 çeşitle 2 uygulamanın kullanıldığı gözlemler yoktur. Fakat bu kombinasyonun uygunluğu data yokluğuyla engellenmemiştir. Bu yapıdaki durumlar «*Çapraz Sınıflama*» olarak tanımlanır. Bunun anlamı, her faktörün her düzeyinin, diğer her faktörün her düzeyi ile kombinasyonda kullanılabileceğidir. Bu yolla faktörler birbirlerine «çapraz» dırlar. (Çapraz kesişirler), «kesişmeleri» bu durumun alt sınıflarıdır veya gözleridir. Gözlerdeki data yokluğu, o gözün var olmadığını göstermez. Sadece data olmadığını belirtir. Çapraz sınıflamada gözlerin toplam sayısı, çeşitli faktörlerin düzeylerinin sayılarının çarpımıdır, yani iki yönlü sınıflamada, gözlerin hepsinde gözlem olmayabilir, s tanesinde olduğunu varsayalım. Böylece gözlemlerin toplam sayısı, bu s gözdeki gözlem sayılarının toplamıdır.

ÖRNEK :

Bir üniversitede öğrenci araştırmasının yapıldığını farzedelim. Sınıflarda daktilo bağlantılarıyla elde edilen yeni bir bilgisayar kullanma kolaylığının kullanılmasına gösterilen reaksiyon araştırılıyor olsun. Tüm birinci sınıf öğrencileri ilk sömestrede İngilizce veya jeoloji veya kimya dersleri almak zorunda olsunlar. (İkinci sömestrede de bunlardan bir diğerini).

Birinci sömestredeki tüm dersler bölümlere ayrılmıştır. Her bölümün farklı öğretmeni vardır ve bütün bölümlerde aynı sayıda öğrenci olması zorunluluğu yoktur. Araştırmada her öğrenciden elde edilen cevap; (1'den 10'a kadar bir ölçekle ölçülen), öğretmenin bilgisayar kullanması hakkındaki fikridir. Bu dataya dayanarak ilgilenilen sorular, öğretmenlerin bilgisayar kullanmasında fark olup olmadığı ve bilgisayar kullanılmasının öğretilen şeyler üzerinde etkisi olup olmadığıdır.

Bu durum için mümkün bir model, genel bir ortalama μ ve üç tip ders için de $\hat{a}_1$, $\hat{a}_2$ ve $\hat{a}_3$ temel etkilerini içermelidir. Her dersin bölümleri içinde terimler içermelidir. Her ders için 10 bölüm olduğunu ve aşağıdaki modeli kullanmaya çalıştığımızı farzedelim.

$$y_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_{ij} + e_{ijk}$$

$$i=1,2,3$$

$$j=1,2,\ldots 10$$

$$k=1,2,\ldots n_{ij}$$

(n_{ij}: i inci dersin j inci bölümündeki öğrenci sayısı)

j'inci bölümün etkisini temsil eder, j=1 için, İngilizce, Jeoloji ve Kimya derslerinin l'inci bölüm için etkisi olur. Bu anlamsızdır, çünkü bu üç bölüm, farklı gruplardaki öğrencilerin birleşimidir. 1 numaralı dersleri dışında ortak dersleri yoktur. Fakat tüm derslerdeki öğrencilerin bölümlerine raslantısal olarak dağıtıldıklarını varsayarsak, bu numaralama tamamiyle 1 no.lu bölümü tanımlama amacıyla olur. Üç bölüm hakkında ortak olanı göstermez; 5, 6 veya diğer herhangi bir no.lu üç bölüm hakkında ortak hiçbir şey olmasa da. Çeşit 1 üzerindeki l'inci uygulamanın, çeşit 2 ve 3 üzerindeki 1'inci uygulamayla aynı olduğu tarım örneğindeki uygulamalara benzemez. Bölümler de bu yolla ilişkili değildirler. Kendi dersleri içerisindeki niteleyicileridir. Dolayısıyla bunlara dersler içindeki bölümler olarak bakarız ve dersler içinde «*Kümelenme*» olarak tanımlarız. Bölümler küme faktörüdürler veya küme sınıflamasıdırlar. Bazen de hiyerarşik sınıflama olarak da söylenir.

Çapraz ve küme sınıflandırmalar arasındaki fark, daha önce açıklanan çeşit ve uygulama örneği açısından ve şimdi tartışılan dersler arası bölümler açısından Tablo 1.6'da gösterilmiştir.

TABLO 1.6

Çapraz ve Küme Sınıflamaların Sistematik Sunuluşu

Çapraz Sınıflama			Küme Sınıflama		
Uygulama			İngilizce	Ders Jeoloji	Kimya
Çeşit	1	2	Bölüm 1	Bölüm 1	Bölüm 1
1			Bölüm 2	Bölüm 2	Bölüm 2
2			Bölüm 3		Bölüm 3
3					Bölüm 4

Çapraz sınıflamada çeşit 1, uygulama 1 ve 2'nin kombinasyonu olarak kullanılmıştır. Her iki durumda da aynı çeşittir. Küme sınıflamada İngilizcenin l'inci bölümü, Jeolojinin l'inci bölümüyle hiçbir şekilde ilişkili değildir. Her ikisi arasındaki ortak olan tek şey, bir niteleyici olan li sayısıdır. Çapraz sınıflamada bir faktörün her düzeyi, bir diğer faktörün her düzeyi ile kombinasyon olarak kullamlmıştır. Fakat küme sınıflamada küme faktörün düzeyleri (bölümler) bir diğeri ile ilişkisizdir ve diğer faktörün bir düzeyi içinde kümelenmiştir. Ayrıca burada görüldüğü gibi, küme faktörün, bir diğer faktörün her düzeyi arasında, farklı sayıda düzeyleri de olabilir. (Farklı derslerde, farklı sayıda bölümler)

Modelin denklemi, j bölüm için $\hat{\beta}_j$ etkisini vererek, dersler içindeki bölümlerin kümelenmesi için açıklama yapar; öyle ki, $\hat{\beta}_{ij}$, i'inci ders içinde kümelenmiş j'inci bölüm için etki olur. $\hat{B}j$ j'inci bölümün yalnız başına tanımlanamayaçağı anlamına gelir. Yalnızca, bağımlı olduğu ders anlamında tanımlanabilir. Böylece model, y_{ijk}, i'inci dersin j'inci bölümündeki k'ıncı öğrencinin seçimi olmak üzere;

$$y_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_{ij} + e_{ijk} \qquad (1.19)$$

olarak ıtanımlanır.

k'nın limiti $k=1, 2, \ldots n_{ij}$'dir: i'inci dersin j'inci bölümünde n_{ij} öğrencisi vardır, $j=1, 2, \ldots b_i$ i'inci derste b_i bölüm vardır ve $i=1, 2, 3$. Tablo 1.7'de İngilizcede 3, Jeolojide 2 ve Kimyada 4 bölüm olan toplam 263 öğrenci için durum:

TABLO 1.7

Küme Sınıflaması

İngilizce (i=1) 3'üncü bölüm: $b_1=3$	Jeoloji (i=2) 2'nci bölüm: $b_2=2$	Kimya (i=3) 4'üncü bölüm: $b_3=4$
$n_{11}=28$	$n_{21}=31$	$n_{31}=27$
$n_{12}=27$	$n_{22}=29$	$n_{32}=32$
$n_{13}=30$		$n_{33}=29$
		$n_{34}=30$

Tablo 1.7'de gösterilen durum, 2 yönlü küme sınıflaması olarak adlandırılır: Dersler içinde bölümler. Bir öğrenciye iki farklı durum için fikrinin sorulduğunu düşünelim: Eğer y_{ijkh}, i'inci dersin j'inci bölümündeki k'inci öğrencinin 'inci cevabı (h=1 veya 2) ise; uygun model (1.20) no.lu denklem olur.

$$Y_{ijkh} = \widehat{\mu} + \widehat{\varphi_i} + \widehat{\beta_{ij}} + \widehat{\gamma_{ijk}} + e_{ijkh} \qquad (1.20)$$

Şimdi yalnızca dersler içinde kümelenmiş bölümler yok, fakat bölümler içinde kümelenmiş öğrenciler de var. Dersler içinde kümelenmiş bölümlerde olduğu gibi, bölümler içinde kümelenmiş öğrenciler içinde tamamen aynı nedenler var. Bunlar çapraz sınıflama olamazlar. Bu 3 yönlü küme sınıflamasına örnektir. Genel olarak, ele alınabilecek durumlarda, kümeleme derecesi için limit yoktur, kullanılışının kapsamı tümüyle verilere ve verilerin geldiği çevreye dayanır.

2.6. Bağımlı Değişkenin Gösterge Değişken Olarak Kullanılması

Bağımlı değişken sayısal olarak ifade edilemiyorsa bu faktörün düzeylerini de gösterge değişken olarak belirtebiliriz.

$$Y = X\beta + \varepsilon$$

eşitliği ile belirtilen lineer modelde, bağımlı değişken ikiden fazla düzeye ayrılıyorsa, her düzey için bir gösterge değişken tanımlanır. Her düzey için ayrı bir model uygulaması ile, iki düzeyli bağımlı değişken durumu gibi ele alınabilir.

Lineer model ile ilgili varsayıma göre hata terimlerinin beklenen değeri sıfırdır.

Buradan

$$E\ (Y) = X\beta \quad \text{yazılabilir.} \qquad (1.21)$$

Bağımlı değişken gösterge değişken olarak sıfır ve bir değerleri ile gösterildiğinde bu değişken Bernoulli raslantısal değişkeni olarak ele alınabilir. Y'nin sıfır ve bir olması olasılıkları

$$Y = 0 \quad \text{için} \quad P\ (Y = 0) = Q$$

ve

$$Y = 1 \quad \text{için} \quad P\ (y = 1) = P$$

şeklinde yazılabilir.

Y'nin beklenen değeri

$$E(Y) = \Sigma YP(Y) \qquad (1.22)$$

$$E(Y) = 0.P(Y=0) + 1.P(Y=1) \quad \text{den}$$

$$E(Y) = P \quad \text{olarak elde edilir.} \qquad (1.23)$$

(2.21) no.lu formülde E (Y)'yi yerine koyarsak

$$P = X\beta \qquad (1.24)$$

elde edilir. Bu bize bağımlı değişkenin seviyesi **X** ile belirlendiğinde Y'nin bir değerini alması olasılığını gösterir.

2.6.1. Hata Terimleri

Bağımlı değişken sıfır ve bir değerlerini aldığında hata terimleri de sadece sıfır ve bir değerlerini alabilir.

$$Y = X\beta + \varepsilon \qquad \text{denkleminden}$$

$$\varepsilon = Y - X\beta \qquad (1.25)$$

yazılabilir.

$$Y=1 \quad \text{için} \quad \varepsilon = 1-X\beta$$

$$Y=0 \quad \text{için} \quad \varepsilon = -X\beta \qquad \text{elde edilir.}$$

Bu durumda hataların normal dağılım göstermediği görülür.

2.6.2. Hata Varyansları

V (Y), Y'nin varyansmı gösterirse

$$V(Y) = E\{[Y - E(Y)]^2\}$$

eşitliği, raslantısal değişken için tanımlanan varyans eşitliğinden faydalanılarak yazılabilir.

$$(1.22) \text{ no.lu } E(Y) = \Sigma Y\, P(Y)$$

eşitliği dikkate alınırsa, Y'nin bir ve sıfır değerleri için

$$V(Y) = (1-P)^2 P + (0-P)^2 Q$$

eşitliği elde edilir.

$$\begin{aligned} V(Y) &= (1-P)^2 . P + P^2 Q \\ &= P\{(1-P)^2 + PQ\} \\ &= P\{(1-P)^2 + P(1-P)\} \\ &= P\{(1-P)^2 + P-P^2\} \\ &= P(1+P^2 - 2P + P-P^2) \\ &= P(1-P) \qquad (1.26) \end{aligned}$$

elde edilir.

$P = X\beta$ şeklindeki (1.24) eşitliği
$\varepsilon = Y - X\beta$ şeklindeki (1.25) eşitliklerini
birlikte ele alırsak

$$\varepsilon = Y - P$$

eşitliği elde edilir. Burada P bir sabittir. ε'nın varyansı Y ile aynıdır.

$$V(\varepsilon) = V(Y) = (X\beta)(1 - X\beta)$$

yazılabilir. ε'nin varyansı açıklayıcı değişkenin alacağı değerlere bağlıdır. Bu durumda değişen varyans durumundaki genelleştirilmiş En Küçük Kareler formüle uygulanır.

$$b = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y$$

eşitliği ile tahmin yapılması yoluna gidilir.

2.6.3. Tahminlerin Değeri

Bağımlı değişken bir faktörün düzeyleri olarak gösterge değişken olarak kullanıldığında modelden elde edilen tahminler olasılık değerlerini göstermektedir. (Formül 1.23).

Tahminlerin tanımlı olabilmesi için bunlar olasılık değerlerini gösterdiğinden sıfır ve bir değerleri arasında olması gerekir. Bu durumda bağımlı değişkenin tahminleri üzerinde aşağıdaki kısıtlamaların konması gerekir.

$$0 \leq E(Y) \leq 1$$

Lineer modelden elde edilen tahminlerin sıfırdan küçük ya da birden büyük çıkması durumunda

$$\hat{X\beta} < 0 \quad \text{ise} \quad \hat{X\beta} = 0$$
$$\hat{X\beta} > 1 \quad \text{ise} \quad \hat{X\beta} = 1$$

dönüşümü yapılarak, tahminlerin olasılık kavramına uygunluğu sağlanmalıdır.

$$E(Y_i) = \alpha + \beta X_i$$

denkleminde X sayısal bir değer olduğunda f_i (1) in olasılıklarmm farklı değerler olması beklenir. Böylece

$$E(Y_i) ;$$

açıklayıcı değişkenin belli bir değerine karşılık Y=1 olduğu zaman, oranın ölçülmesi olarak yorumlanabilir. Bu bir olasılık olduğundan;

$$0 \leq \alpha + \beta X_i \leq 1$$

olmasını gerektirir.

$$\varepsilon_i = Y_i - \alpha - \beta X_i$$

eşitliği yazılırsa, Y yalnızca 0 ve 1 değerlerini alabileceğinden X'in verilen bir değeri için hata terimi (ε_i) de yalnızca iki farklı değer olabilir.

$$(-\alpha-\beta X_i) \quad \text{ve} \quad (1-\alpha-\beta X_i)$$

Bu demektir ki ε_i normal dağılmamaktadır.

ε_i	$f(\varepsilon_i)$
$-\alpha-\beta X_i$	f
$1-\alpha-\beta X_i$	$1-f$
	1

ε_i yukarıdaki gibi tanımlanan kesikli bir dağılımdır. f ve 1-f olasılıkları $E(\varepsilon_i) = 0$ varsayımından yararlanarak belirlenebilir. Yani:

$$(-\alpha-\beta X_i) f + (1-\alpha-\beta X_i)(1-f) = 0$$

Buradan

$$f = 1 - \alpha - \beta X_i$$

elde edilir.

ε_i'nin varyansı da

$$E(\varepsilon_i^2) = (-\alpha-\beta X_i)^2(1-\alpha-\beta X_i) + (1-\alpha-\beta X_i)^2(\alpha+\beta X_i)$$
$$= (\alpha+\beta X_i)(1-\alpha-\beta X_i)$$
$$= E(Y_i)\{1-E(Y_i)\}$$

Buradan da, varyansı E (Y_i)'ye bağlı olduğu için ε_i'nin değişen varyanslı olduğu görülür.

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

modelinde bağımlı değişken Y_i'nin özel yapısı (0 ve 1 değerlerini alması) gereği tahmin ve öngörünün bazı problemleri ortaya çıkar.

i) Hata teriminin değişen varyanslı yapısı nedeniyle α ve β'nın En Küçük Kareler tahmin edicileri

$$\hat{\beta} = \Sigma a_i Y_i$$

Bu durumda. $\dfrac{1}{\sigma_i^2} = w_i$ alınırsa

$$\hat{\beta} = \frac{(\Sigma w_i)(\Sigma w_i X_i Y_i) - (\Sigma w_i X_i)(\Sigma w_i Y_i)}{(\Sigma w_i)(\Sigma w_i X_i^2) - (\Sigma w_i X_i)^2}$$

ve

$$\hat{\alpha} = \frac{\Sigma w_i Y_i}{\Sigma w_i} - \hat{\beta}\left[\frac{\Sigma w_i X_i}{\Sigma w_i}\right]$$

elde edilir.

Değişen varyans durumunda en iyi lineer sapmasız tahmin edici (BLUE) kullanılması en azından büyük örnekler için uygundur.

Bağımlı değişkenin gösterge değişken olduğu durumda varyansın tutarlı bir tahmin edicisini

$$\hat{\sigma}_i^2 = \hat{Y}_i(1-\hat{Y}_i)$$

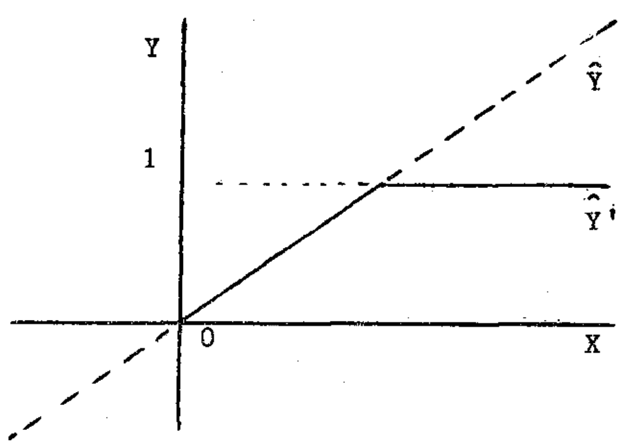
olarak elde ederiz.

i) İkinci problem α ve β tahmin edicilerinin dağılımı ile ilgilidir. α ve β'nın tahmin edicileri normal dağılmadığı için klasik anlamlılık testleri uygulanamaz. Ancak asimtotik durumda ($n=\infty$) problemi yoktur.

$$\{\lim_{n\to\infty} E(\hat{\beta})=\beta\}$$

iii) Üçüncü problem kestirimle ilgilidir. E (Y_i) bir olasılık olarak yorumlandığından E (Y_i) değerleri 1 ile 0 arasında düşmelidir.

$\hat{Y}$ değerleri bir doğru üzerinde bulunacağından (Şekil: 1.5) $-\infty$ ile $+\infty$ değerleri arasında bir değer olabilir. Bu durumda sayfa 55'de açıklanan dönüşüm uygulanır.



Şekil 1.5

2.6.4. Tetrakorik Korelasyon

«Tetrachoric» korelasyon açıklayıcı (X) ve bağımlı (Y) değişkenlerin her ikisinin de nitel değişken olarak kullanıldığı verilerde kullanılır. Veriler derecelenmiş niteliklerden elde edilemiyeceği için, iki değişken arasındaki korelasyonu tahmin etmenin tek yolu budur.

Gerekli varsayımları yaparak pek çok olay gösterge değişken ile ifade edilebilir.

Tetracorik korelasyonun hesabı r'nin pek çok kuvvetlerinden oluşan bir seriyi içerdiği için uzun ve karmaşık bir denklemdir.

TABLO 1.8

		1	0	TOPLAM	ORANLAR
Y	0	a	b	a+b	$\frac{a+b}{N} = p$
	1	c	d	c+d	$\frac{c+d}{N} = q$
		a+c	b+d	a+b+o+d=N	
TOPLAM ORANLAR		$\frac{a+c}{N} = p'$	$\frac{b+d}{N} = q'$		1

$$r_t + \frac{1}{2} r_t^2 zz' + \frac{1}{6} r_t^3 (z^2-1)(z'^2-1) + \frac{1}{24} r_t^4 z(z^2-3)z'(z'-3)$$

$$+ \frac{1}{120} r_t^5 (z^4-6z^2+3)(z'^4-6z'^2+3) + \frac{1}{720} r_t^6 z(z^4-10z^2+15)$$

$$z'(z'^4-10z'^2+15) + \frac{1}{5040} r_t^7 (z^6-15z^4+45z^2-15)$$

$$(z'^6-15z'^4+45z'^2-15) = \frac{ad-bc}{yy'N^2} \quad (1.27)$$

a, b, c, d Tablo 1.8'de gösterilen gözlerdeki frekanslardır.

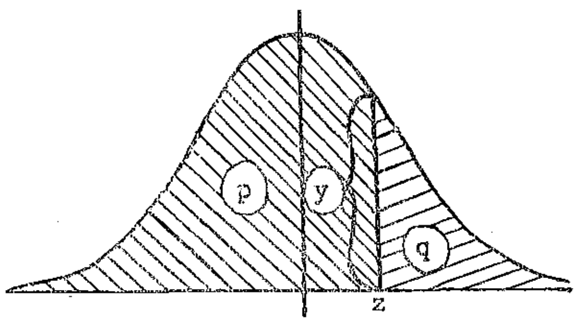
a_t lerde korelasyon katsayısının tetrakorik olduğunu göstermektedir.

p', q' X değişkenine ait, p ve q da Y değişkenine ait oranlardır. X ve Y değişkenlerinin dağılımı bu oranlarla ifade edilebilir.

Z standart normal dağılımı p ve q oranında bölen standart değerdir.

Benzer şekilde Z' de standart normal dağılımı p' ve q' oranında bölen standart değerdir.

y ve y'ler standart normal dağılımda Z ve Z' değerlerine karşılık olan ordinat değerleridir (Şekil 1.6).



Şekil 1.6

Yukarıdaki formülün çözümünün yaklaşık olarak yapılması en uygundur. Bunun için farklı metodlar kullanılmaktadır. Başlıca ikisini kısaca açıklayalım.

i) Cos pi formülü

$$r_{cos-pi} = \cos\left(\pi \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}\right)$$

şeklindedir.

pay ve paydayı $\sqrt{bc}$ ye bölersek

$$r_{cos-pi} = \cos\left(\frac{\pi}{1 + \sqrt{\frac{ad}{bc}}}\right)$$

Buradan sayısal değerler yerine konursa radyan cinsinden bir açı bulunur. Bu açının cos'ü aranan korelasyondur. Eğer b veya c, veya her ikisi birden 0 ise korelasyon +1 olacaktır. a veya d, veya her ikisi birden 0 ise bu defa da korelasyon −1 olacaktır. bc=ad ise açı 90°'dir.

r_{cos-pi}'nin bulunması için $\frac{ad}{bc}$ değerlerine göre hesaplanmış tablolar vardır.

ii) Grafik tahminleri
Çok sayıda tetrakorik r'nin hesaplanması gerektiğinde THURSTONE hesaplama diyagramları kullanılması önemli ölçüde kolaylık sağlar.

2.6.5. Tetrakorik r'nin st hatası

Bu test

i) N büyükken

ii) r büyükken

ili) İki kategoriye bölme ortalayanları yakından güvenilirdir.

Standart hatanın tahmininin tam formülü çok uzun olduğundan verilmeyecektir.

$r_t = 0$ iken bu formül basit duruma gelir ve

$$\sigma_{r_t} = \frac{\sqrt{pp'qq'}}{yy'\sqrt{N}}$$

şeklinde ifade edilir.

$$\frac{r_t}{\sigma_{r_t}}$$

oranının yaklaşık 2 değerinden büyük olması beklenir. Bu test için N'nin en az 200, tercihen 300 olması uygundur.

YARARLANILAN KAYNAKLAR

1) Davies M., Multiple Linear Regression Analysis With Adjustment For Class Differences Journal of the Econometric Society Vol. 56 No. 295, September 1961

2) Goldberger Arthur S., Econometric Theory, John Wiley & Sons, Inc. 1964 New York.

3) Green Richard D., Doll John P., Dummy Variables and Seasonality — A Curio The American Statistician, May 1974 Vol. 28 No. 2

4) Guilford J.P., Fundemental Statistics in Psychology and Education, Mc. Graw - Hill Book Company, Inc. 3 edition 1956, New York.

5) Işıkara Baki, Regresyon Yöntemleri ve Sorunları, İstanbul Üniversitesi İktisat Fakültesi Yayını. Yayın No: 358, İstanbul 1975

6) Johnston J., Econometric Methods, Mc. Graw - Hill 2. edition 1972, New York

7) Kmenta Jan, Elements of Econometrics, Macmillan Publishing Co. İnc. New York, 1971

8) Lebort Ludovic, Fenelon J.P., Statistique et Informatique Appliques, 3. baskı, Dunod, Paris, 1975

9) Morice E., Chartier F., Methode Statistique 2. Partie, Analyse Statistique, Paris, Imprimerie Nationale, 1954

10) Pindyck Robert S., Rubinfeld Daniel L., Econometric Models and Economic Forecatsts, Mc. Graw - Hill Book, Company, 1976, New York.

11) Searle S.R., Linear Models, John Wiley & Sons. Inc. New York.

12) Suits Daniel B., Use of Dummy Variables In Regression Equations, Journal of the Econometric Society Vol. 52 No. 280, December 1957

13) Tobin James, Estimation of Relationships For Limited Dependent Variables Journal of the Econometric Society, Vol. 26 No. 1, January, 1958

14) Waerden B., Vander, Statistique Mathématique Dunod, Paris 1967

15) Warner Stanley L., Multivariate Regression of Dummy Variates Under Normality Assumptions Journal of the Econometric Society Vol. 58, No. 304, December, 1963

16) Winer B.J., Statistical Principles in Experimental Design, Mc. Graw - Hill Book Company New York, 1971

17) Wonnacott T.H., Wonnacott R.J., Introductory Statistics for Business and Economics, John Wiley & Sons, Inc, New York, 1972

18) Yule Udny G., Kendall M.G., An Introduction To The Theory of Statistics, Charles Griftin & Company Limited, 1954; 14. Baskı.