

**ÇAPRAZ TABLOLARIN REGRESYON YÖNTEMİ İLE GÖRSELLEŞTİRMESİ****H.Nazan ÇAĞLAR**

İstanbul Kültür Üniversitesi , İktisadi ve İdari Bilimler Fakültesi

ORCID No: 0000-0003-2152-8747

**ÖZET**

Sosyal Bilimler alanında , çapraz tabloların ( crosstab) yorumlanması önemli bir yere sahiptir. Veri analizlerinde iki boyutlu tabloların etkin biçimde özetlenmesi ve yorumlanması özellikle büyük boyutlu ise önemlidir. Tablonun  $m, n > 10$  ( m: satır , n: sütun ) olması durumunda okunması , anlaşılması ve yorumlanması zorlaşır. Bu noktada verileri görselleştirerek özetlemek kolaylık sağlar. Görsel nesnelerin kısa süre de karar vermede önemli olduğu bilinmektedir. Çapraz tablodaki verinin görselleştirilmesinde pek çok yöntem kullanılmaktadır.

Lokal polinomlar , parametrik olmayan regresyon modelleri için de önemli bir yere sahiptir. Bu çalışmada , kategorik verileri temsil eden tablolara çok değişkenli regresyon yöntemi uygulanmıştır. ( Seifert & Gasser , 1996) Model değişkenleri tablonun satır, sütun değişkeni olarak alınmıştır. Kurulan Model aracılığı ile tablonun hücre değerleri tekrar tahmin edilmiştir. Regresyon metodu polinomial olarak uygulanmıştır. Bulunan modelle crosstab'ın regresyon yüzey grafiği çizilmiştir. Çizilen yüzey grafiği çapraz tabloların görsel olarak daha iyi açıklanabildiğini göstermiştir.

Yazılan bir bilgisayar programı ile literatürden ( Simonoff ,1996 ) alınmış problem için metodun iki boyutlu tablolar için uygulaması açıklanmaktadır

**Anahtar Kelimeler :** polinomal regresyon , çapraz tablo , yüzey grafiği**1. LOKAL POLİNOMAL REGRESYON**

$( x_1, y_1 ) , \dots , ( x_n, y_n )$  gözlemleri için aşağıdaki fonksiyonel modeli düşünelim.

$$y_i = f ( x_i ) + \varepsilon_i$$

Burada  $x_i$  bağımsız değişken,  $y_i$  ler bağımlı değişken ve  $\varepsilon_i$  hata terimidir.

Taylor seri açılım yaklaşımı ile  $x_i$  noktası civarında  $f(x)$  fonksiyonunu seriye açalım,

$$f(x_i) = f(x) + f'(x)(x_i - x) + \dots + \frac{f''(x)(x_i - x)^2}{2!}$$

$f(x)$  fonksiyonunun bulunması lokal polinomal regresyon olarak bilinir. Aşağıdaki denklem en küçük kareler optimizasyonu ile çözülür. (Çağlar, 2004)

$$\min (\beta) \sum_{i=1}^n [y_i - \beta_0 - \beta_1(x_i - x) - \beta_2(x_i - x)^2 - \dots - \beta_p(x_i - x)^p]^2 \quad (1)$$

Lokal polinomların  $\beta_j$  tahmincileri yukarıda verilen toplamın minimize edilmesi sonucunda elde edilir.

$$\beta = (\beta_0, \dots, \beta_p)' \quad \text{tahmin edilecek parametrelerdir.}$$

$$X = \begin{bmatrix} 1 & (x_1 - x_0) & \dots & (x_1 - x_0)^p \\ \vdots & \ddots & \ddots & \vdots \\ 1 & (x_n - x_0) & \dots & (x_n - x_0)^p \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (2)$$

Olarak bulunur. Sonuç olarak Lokal polinomal yaklaşımda  $y$  tahmin değerleri

$$\hat{y} = \sum_{j=0}^p \hat{\beta}_j (x_i - x_0)^j, \quad i = 1, 2, \dots, n \quad (3)$$

$x_0$  : ortalama

formülünden hesaplanır.

## 2. UYGULAMA

147 kişinin Üniversiteden mezun oldukları yıllardan itibaren çalışma süreleri ve aldıkları ücretlere (para birimi) ait datalar Çizelge 1 de görülmektedir. Satırlar aylık ücret dilimlerini, sütunlar ise çalışma sürelerini göstermektedir.

Çapraz tablonun satır ve sütun numaraları X açıklayıcı değişken matrisini oluşturur. Tablonun hücre değerleri bağımlı değişken Y matrisini oluşturmaktadır.

Çizelge1. Ücretler ve yıllara ait datalar ( Simonoff ,1996 )

Ücret (pb)/yıllar	0-2	3-5	6-8	9-11	12-14	15-17	18-20	21-23	24-26	27-29
950-1150	5	0	1	0	0	0	0	0	0	0
1150-1350	2	1	0	0	0	0	0	2	0	0
1350-1550	5	1	3	2	0	0	1	0	0	0
1550-1750	5	5	2	1	0	1	0	1	0	0
1750-1950	9	9	5	0	2	2	1	1	1	0
1950-2150	3	5	2	1	2	0	1	0	0	1
2150-2350	0	1	4	3	2	1	3	0	2	1
2350-2550	0	0	4	0	1	2	2	0	0	1
2550-2750	0	0	2	2	0	5	1	2	1	1
2750-2950	0	0	1	0	0	1	4	0	2	3
2950-3150	1	0	1	0	1	1	1	0	1	0
3150-3750	0	0	0	0	0	0	5	0	0	1

Datalar için yöntemi uygulayalım :

$$y = \beta_0 + \beta_1(x_1 - \bar{x}_1) + \beta_2(x_2 - \bar{x}_2)^2 + \beta_3(x_1 - \bar{x}_1)^2 + \beta_4(x_2 - \bar{x}_2)^2 \quad (4)$$

(3) denklemindeki  $x_0$  lar  $x_1$  ve  $x_2$  ortalama değerleridir.

$$Y = [y_1 \ y_2 \ . \ . \ . \ y_{120}]'$$

Değerleri ise aşağıdaki vektörün transpozesi olarak alınır.

$$Y = \begin{pmatrix} 5010000000210000020051320010005521010100995022111035212010010143 \\ 21302100401220010022051211001001402310101110100000005001 \end{pmatrix}$$

**X** matrisinin genel formu aşağıda verilmektedir. Değerleri ise Ek1 de sunulmaktadır.

$$X = \begin{bmatrix} 1 & x_1 - \bar{x}_1 & x_2 - \bar{x}_2 & (x_1 - \bar{x}_1)^2 & (x_2 - \bar{x}_2)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Çizelge1 aşağıdaki şekilde kodlanır. (Çizelge 2)

#### Çizelge 2. Değişkenler

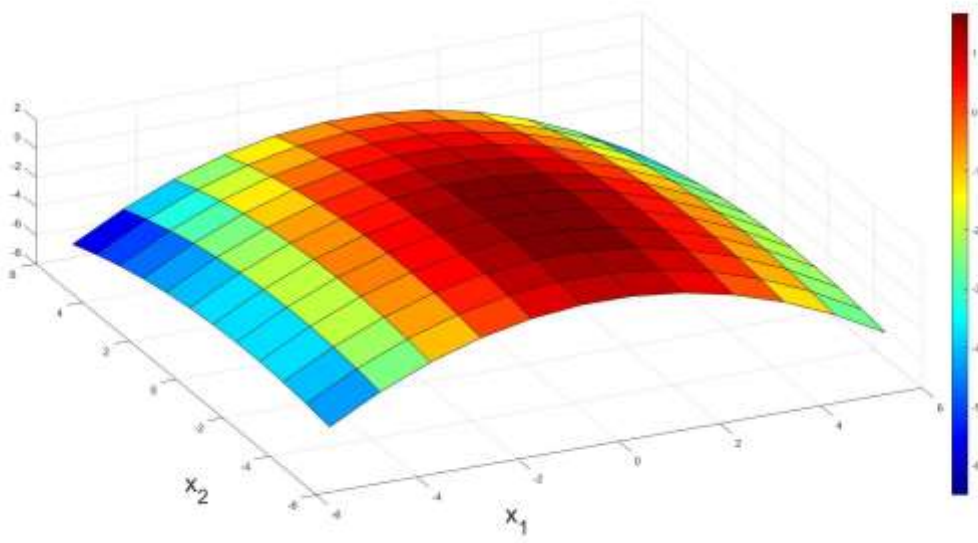
x1/x2	1	2	3	4	5	6	7	8	9	10
1	y1	y2	y3	y4	.	.	.	.	.	y10
2	y11	y12	.	.	.	.	.	.	.	.
3	.	.	.	.	.	.	.	.	.	.
4	.	.	.	.	.	.	.	.	.	.
5	.	.	.	.	.	.	.	.	.	.
6	.	.	.	.	.	.	.	.	.	.
7	.	.	.	.	.	.	.	.	.	.
8	.	.	.	.	.	.	.	.	.	.
9	.	.	.	.	.	.	.	.	.	.
10	.	.	.	.	.	.	.	.	.	.
11	.	.	.	.	.	.	.	.	.	.
12	.	.	.	.	.	.	.	.	.	y120

Veriler (2) denkleminde yerine yazılırsa  $\beta_j$  tahminçileri

$$\hat{\beta} = [1.5386 \quad -0.0213 \quad -0.1833 \quad -0.0453 \quad 0.0275]'$$
 olarak hesaplanır.

Çizelge 1 verilerine lokal polinomal regreseyon uygulanarak bulunan yoğunluğa bağlı farklı renklendirilmiş yüzey grafiği Görsel 1 de görülmektedir. Şekilden koyu kırmızı bölümler Çizelgedeki 1 de frekansı en yüksek olanlardır. Frekanslardaki yoğunluk düşüşünü mavi ye

doğru gözlemlemekteyiz.



**Görsel 1.** Verilere ait modelin yüzey grafiği

### 3. SONUÇ

Local polinomal regresyon yöntemi , mühendislik, istatistik ve matematik alanlarında yaygın kullanılmaktadır. Yaklaşım fonksiyonun bulunmasında en küçük kareler yöntemi uygulama kolaylığına sahip olmakla birlikte yöntemin kodları Matlab veya farklı bir progama dilinde kolaylıkla yazılabilir. Çalışmada Lokal polinomların teorisi anlatıldıktan sonra çapraz tablolarda tabloların görselleştirilmesinde ve yoğunluk alanlarının görsel olarak belirlenmesinde oldukça etkili bir yöntem olduğu görülmüştür. Yapılan çalışma iki boyutlu Kernel fonksiyonları (Linton et al., 2001 ) ( Yang & Yue , Nov.2014 kullanılarak geliştirilebilir.

### KAYNAKLAR

- [1] Çağlar N ,Çağlar H. , Stokastik trendin lokal polinomal yaklaşım ve üstel düzgünleştirme ile bulunmasında hata kareleri ortalamasınının karşılaştırılması. Öneri , 2004, 6(22), 301-304.
- [2] Linton O.,Mammen E., Nielsen J,Tanggaard C. , Yield curve estimation by Kernel smoothing methods, Journal of Econometrics, Nov 2001,105(1).
- [3] Seifert B. And Gasser T. , Finite-Sample variance of local polinomials : analysis and solutions, Journal of American Statistical Association, March 1996 , 91(433)..
- [4] Simonoff J.S. , *Smoothing Methods in Statistics*, Springer-Verlag New York, 1996.
- [5] Yang F.,Yue Z.,Kernel density estimation of three-parameter Weibull distribution with neural network and genetic algorithm. Applied Mathematics and computation, Nov.2014, 247,803-814.

Ek1:

X matrisi

1	-5,5	-4,5	30,25	20,25
1	-5,5	-3,5	30,25	12,25
1	-5,5	-2,5	30,25	6,25
1	-5,5	-1,5	30,25	2,25
1	-5,5	-0,5	30,25	0,25
1	-5,5	0,5	30,25	0,25
1	-5,5	1,5	30,25	2,25
1	-5,5	2,5	30,25	6,25
1	-5,5	3,5	30,25	12,25
1	-5,5	4,5	30,25	20,25
1	-4,5	-4,5	20,25	20,25
1	-4,5	-3,5	20,25	12,25
1	-4,5	-2,5	20,25	6,25
1	-4,5	-1,5	20,25	2,25
1	-4,5	-0,5	20,25	0,25
1	-4,5	0,5	20,25	0,25
1	-4,5	1,5	20,25	2,25
1	-4,5	2,5	20,25	6,25
1	-4,5	3,5	20,25	12,25
1	-4,5	4,5	20,25	20,25
1	-3,5	-4,5	12,25	20,25
1	-3,5	-3,5	12,25	12,25
1	-3,5	-2,5	12,25	6,25
1	-3,5	-1,5	12,25	2,25
1	-3,5	-0,5	12,25	0,25
1	-3,5	0,5	12,25	0,25
1	-3,5	1,5	12,25	2,25
1	-3,5	2,5	12,25	6,25
1	-3,5	3,5	12,25	12,25
1	-3,5	4,5	12,25	20,25
1	-2,5	-4,5	6,25	12,25
1	-2,5	-3,5	6,25	6,25
1	-2,5	-2,5	6,25	2,25
1	-2,5	-1,5	6,25	0,25
1	-2,5	-0,5	6,25	0,25
1	-2,5	0,5	6,25	0,25
1	-2,5	1,5	6,25	2,25
1	-2,5	2,5	6,25	6,25
1	-2,5	3,5	6,25	12,25
1	-2,5	4,5	6,25	20,25
1	-1,5	-4,5	2,25	20,25
1	-1,5	-3,5	2,25	12,25
1	-1,5	-2,5	2,25	6,25
1	-1,5	-1,5	2,25	2,25
1	-1,5	-0,5	2,25	0,25
1	-1,5	0,5	2,25	0,25
1	-1,5	1,5	2,25	2,25
1	-1,5	2,5	2,25	6,25
1	-1,5	3,5	2,25	12,25
1	-1,5	4,5	2,25	20,25
1	-0,5	-4,5	0,25	20,25
1	-0,5	-3,5	0,25	12,25
1	-0,5	-2,5	0,25	6,25
1	-0,5	-1,5	0,25	2,25
1	-0,5	-0,5	0,25	0,25
1	-0,5	0,5	0,25	0,25
1	-0,5	1,5	0,25	2,25
1	-0,5	2,5	0,25	6,25
1	-0,5	3,5	0,25	12,25
1	-0,5	4,5	0,25	20,25
1	0,5	-4,5	0,25	20,25
1	0,5	-3,5	0,25	12,25
1	0,5	-2,5	0,25	6,25
1	0,5	-1,5	0,25	2,25
1	0,5	-0,5	0,25	0,25
1	0,5	0,5	0,25	0,25
1	0,5	1,5	0,25	2,25
1	0,5	2,5	0,25	6,25
1	0,5	3,5	0,25	12,25
1	0,5	4,5	0,25	20,25
1	1,5	-4,5	2,25	20,25
1	1,5	-3,5	2,25	12,25
1	1,5	-2,5	2,25	6,25
1	1,5	-1,5	2,25	2,25
1	1,5	-0,5	2,25	0,25
1	1,5	0,5	2,25	0,25
1	1,5	1,5	2,25	2,25
1	1,5	2,5	2,25	6,25
1	1,5	3,5	2,25	12,25
1	1,5	4,5	2,25	20,25
1	2,5	-4,5	6,25	20,25
1	2,5	-3,5	6,25	12,25
1	2,5	-2,5	6,25	6,25
1	2,5	-1,5	6,25	2,25
1	2,5	-0,5	6,25	0,25
1	2,5	0,5	6,25	0,25
1	2,5	1,5	6,25	2,25
1	2,5	2,5	6,25	6,25
1	2,5	3,5	6,25	12,25
1	2,5	4,5	6,25	20,25
1	3,5	-4,5	12,25	20,25
1	3,5	-3,5	12,25	12,25
1	3,5	-2,5	12,25	6,25
1	3,5	-1,5	12,25	2,25
1	3,5	-0,5	12,25	0,25
1	3,5	0,5	12,25	0,25
1	3,5	1,5	12,25	2,25
1	3,5	2,5	12,25	6,25
1	3,5	3,5	12,25	12,25
1	3,5	4,5	12,25	20,25
1	4,5	-4,5	20,25	20,25
1	4,5	-3,5	20,25	12,25
1	4,5	-2,5	20,25	6,25
1	4,5	-1,5	20,25	2,25
1	4,5	-0,5	20,25	0,25
1	4,5	0,5	20,25	0,25
1	4,5	1,5	20,25	2,25
1	4,5	2,5	20,25	6,25
1	4,5	3,5	20,25	12,25
1	4,5	4,5	20,25	20,25
1	5,5	-4,5	30,25	20,25
1	5,5	-3,5	30,25	12,25
1	5,5	-2,5	30,25	6,25
1	5,5	-1,5	30,25	2,25
1	5,5	-0,5	30,25	0,25
1	5,5	0,5	30,25	0,25
1	5,5	1,5	30,25	2,25
1	5,5	2,5	30,25	6,25
1	5,5	3,5	30,25	12,25
1	5,5	4,5	30,25	20,25