BÖLÜM 8

POLİNOMSAL (ÇOK TERİMLİ) REGRESYON

GİRİŞ

- Önceden ele alınan birçok model (1) bağımlı ya da yanıt değişkeni ile bağımsız değişkenler arasındaki doğrusal ilişkiyi belirler ve (2) parametreler cinsinden doğrusaldır.

- Bağımsız değişkenlerden kaynaklı doğrusal ilişki her terimde tek bir değişkenin birinci dereceden fonksiyonu olarak ortaya çıkar.

- Hiçbir terim bağımsız değişkenlerin üssünü ya da çarpımlarını içermez.

- Böylece, yanıt değişken ortalamasının bir bağımsız değişkene göre değişimi, bu değişkenin bütün değerleri ve modeldeki diğer bağımsız değişkenler için sabittir.

- Parametreler cinsinden doğrusallık, bir parametrenin modeldeki her (toplamalı) terimde bağımsız değişkenlerin bir çarpımı olarak yer alması anlamına gelmektedir.

TEK DEĞİŞKENLİ POLİNOMLAR

- Bir bağımlı (yanıt) değişken ile bir bağımsız (girdi) değişken arasında varsayılan doğrusal ilişki sabit değişim oranı anlamına gelir ve gerçek ilişkiyi yeterince temsil etmeyebilir.

- Örneğin, kan akışındaki ilacın (uyuşturucu) yoğunluğu zamana göre doğrusal olmayabilir.

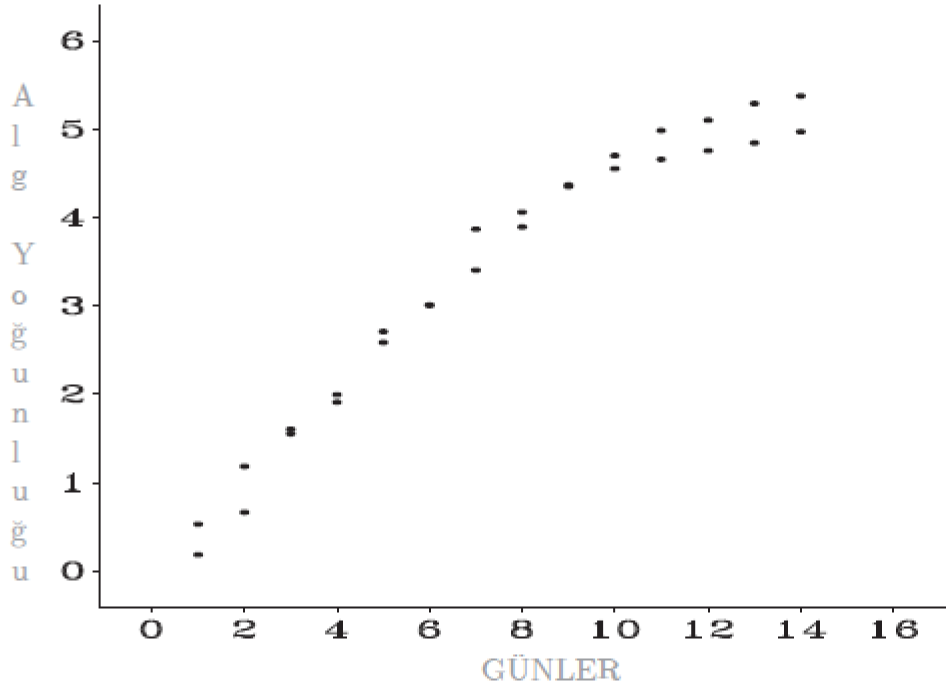
- Enflasyon endeksi ya da gayri safi milli hasıla gibi birçok ekonomik zaman serisi zamana göre doğrusal olmayan bir eğilim gösterir.

- Fırının sıcaklığının artmasıyla pasta pişirme zamanı azalmasına rağmen bu azalış doğrusal olmayabilir.

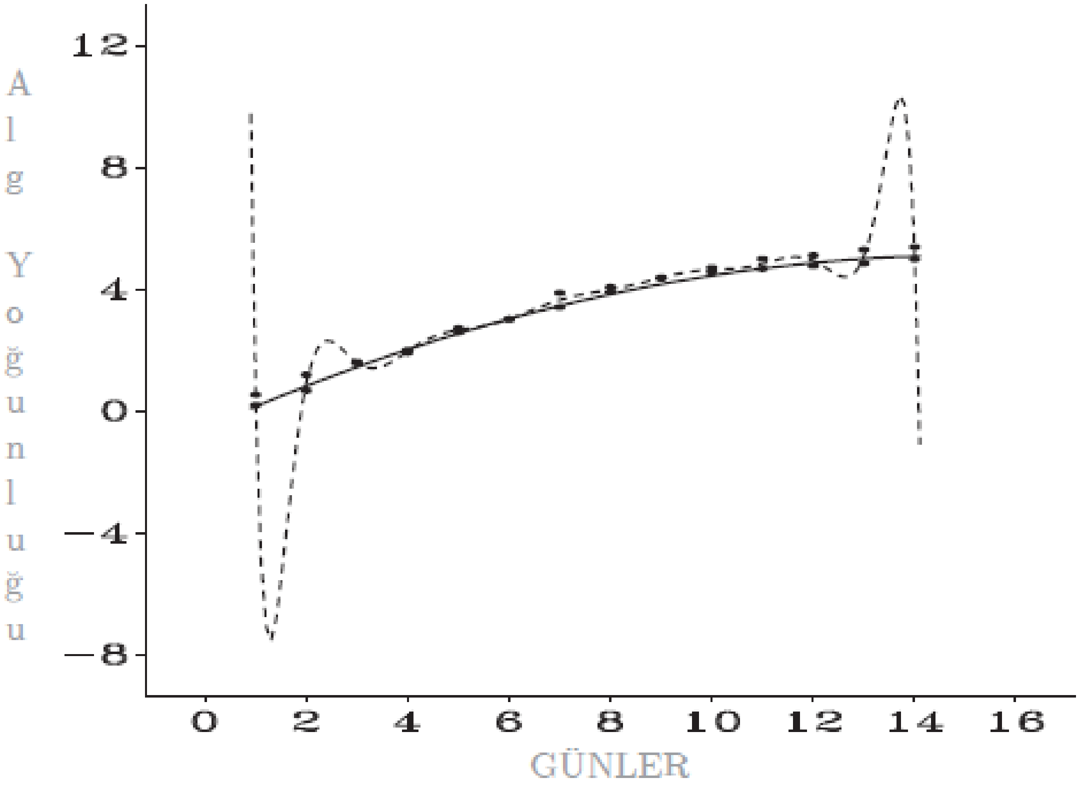
- Bütün bu örneklerde bağımlı değişken (Y) ortalamasının bağımsız değişkene (X) göre değişim oranı sabit olmayabilir.

TABLO 8.1. *Alg yoğunluğunun zaman içindeki ölçümleri.*

Gün	*Tekrar 1*	*Tekrar 2*	*Gün*	*Tekrar 1*	*Tekrar 2*
1	.530	.184	8	4.059	3.892
2	1.183	.664	9	4.349	4.367
3	1.603	1.553	10	4.699	4.551
4	1.994	1.910	11	4.983	4.656
5	2.708	2.585	12	5.100	4.754
6	3.006	3.009	13	5.288	4.842
7	3.867	3.403	14	5.374	4.969



ŞEKİL 8.1. *Alg yoğunluğuna karşı çalışma günleri.*



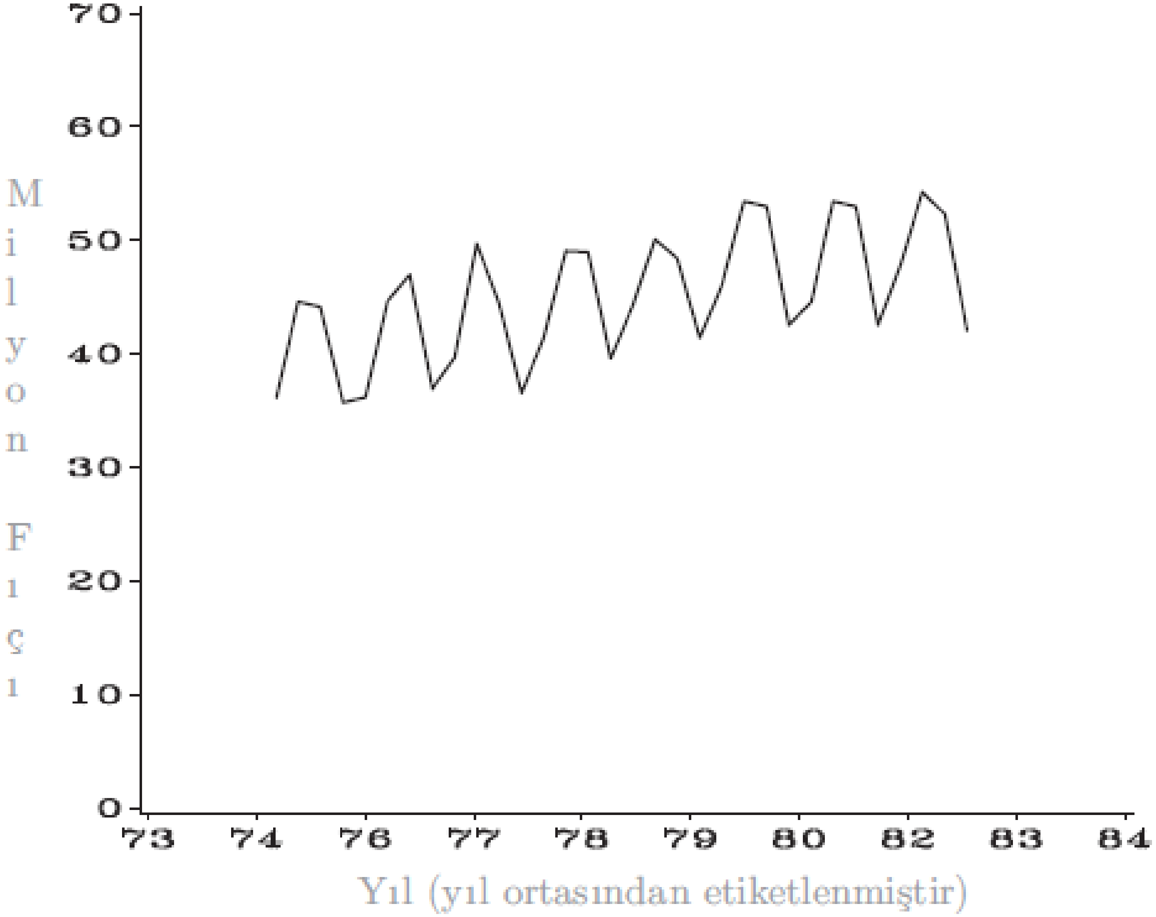
ŞEKİL 8.2. *Uydurulan karesel model (kesintisiz çizgi) ve uydurulan 13. derece polinom modelle birlikte Alg yoğunluk verisi.*

TABLO 8.2. *1975'in ilk çeyreğinden 1982'nin son çeyreğine kadar ABD'nin üç aylık bira üretimi (milyonlarca fıçı).*

Yıl	*Çeyrek*			
	I	*II*	*III*	*IV*
1975	36.14	44.60	44.15	35.72
1976	36.19	44.63	46.95	36.90
1977	39.66	49.72	44.49	36.54
1978	41.44	49.07	48.98	39.59
1979	44.29	50.09	48.42	41.39
1980	46.11	53.44	53.00	42.52
1981	44.61	55.18	52.24	41.66
1982	47.84	54.27	52.31	42.03

TRİGONOMETRİK REGRESYON MODELLERİ

Örnek 8.3'te olduğu gibi zamana göre (t) toplanan yanıt değişkeni ölçümlerine (Y_t) zaman serileri verileri denir. Örnek 8.3'te yer almamasına rağmen bu gibi veriler genellikle kendini her s zaman aralığında tekrarlayan periodik davranışlar sunarlar. Örneğin, Raleigh'in ortalama aylık sıcaklık verisinin kendisini yıllar içinde tekrarlaması beklenen periodik davranışlar sunabilir. Örneğin, bir yıl içindeki Ocak ayının ortalama sıcaklık değeri diğer yılın Ocak ayı değerine benzemesi beklenir; ya da bir yıl içindeki Şubat ayı sıcaklık değerinin diğer yılların Şubat ayı değerlerine benzemesi beklenir ve aynı durum diğer aylar için de geçerlidir. Ekonomik zaman serileri genellikle iş döngüsünü yansıtan peridik davranışlar sunar. Örneğin, kutlama kartlarının toplam aylık satışlarının yıllar içinde periodik olması beklenir; tıpkı toplam aylık parekende satışı ve konut başlangıçları gibi. $\sin(\omega t)$ ve $\cos(\omega t)$ gibi trigonometrik fonksiyonlar $2\pi/\omega$ **periyodu** ile zaman içinde periyodiktir. Yani, $\sin(\omega t)$ her $j = 1, 2, \ldots$ için $\sin[\omega(t + (2\pi/\omega)j)]$ ile aynıdır. Böylece, periyodik davranışlara sahip zaman serileri **trigonometrik regresyon modelleri** ile modellenebilirler.



ŞEKİL 8.3. *Zamana karşı ABD'nin üç aylık bira üretimi.*

YANIT EĞRİ MODELLEMESİ

Fonksiyon Yapısını Belirlerken Göz Önünde Bulundurulması Gerekenler

Modele dahil edilmesi gereken gerçeklik derecesi regresyon analizinin amacına bağlıdır. Regresyon modelinin belirli bir veri kümesindeki gözlemlenmiş ilişkileri özetlemek için basitçe kullanımı en zahmetsiz olan amaçtır. Bu amaçlı bir kullanımda, kendi başına model fonsiyon yapısıyla ya da diğer veri kümelerinin veya durumların kestirilmesiyle ilgilenilmemektedir. En zahmetli olanıysa; sistemdeki fiziksel, kimyasal ve biyolojik süreçleri tanımlamak için özelleşmiş matematiksel modellerin oluşturulmasıdır. Burada amaç, modelin bilgi birikimi el vereceği ölçüde gerçekçi olmasını sağlamaktır.

POLİNOM YANIT MODELLERİ

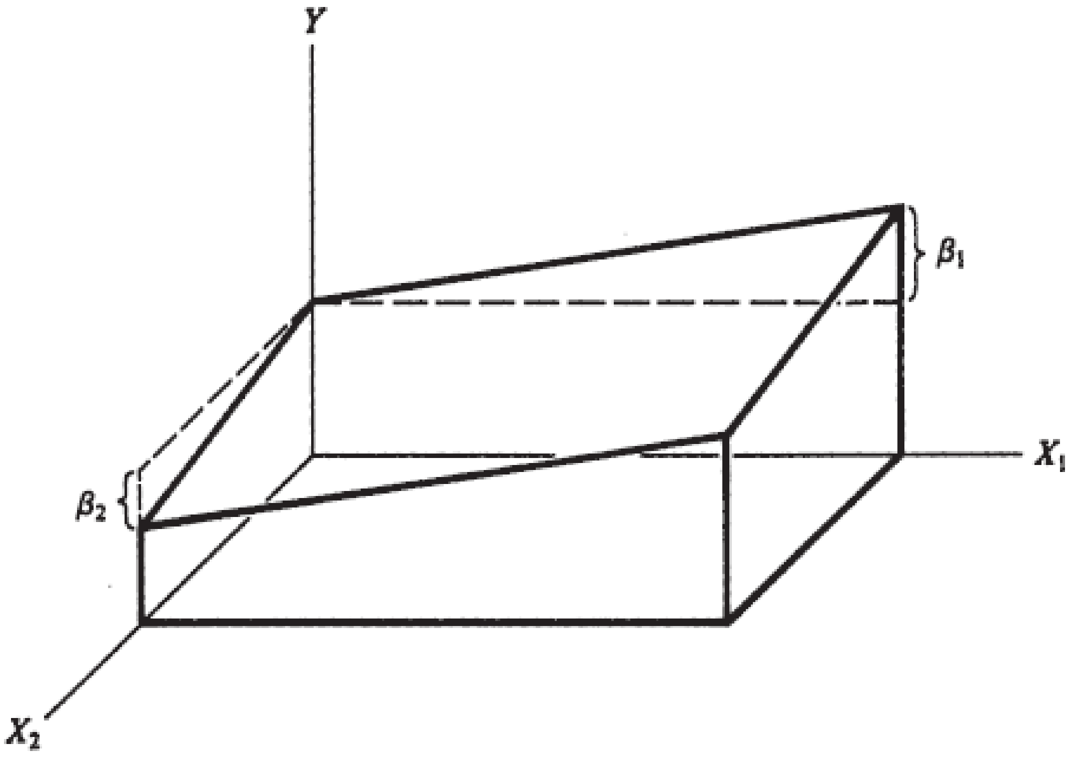
Önceden ele alınan modeller, her terimde sadece bir bağımsız değişkenin birinci üssünün yer aldığı birinci derece polinom modelleriydi. İki değişkenli birinci derece bir model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i. \quad (8.27)$$

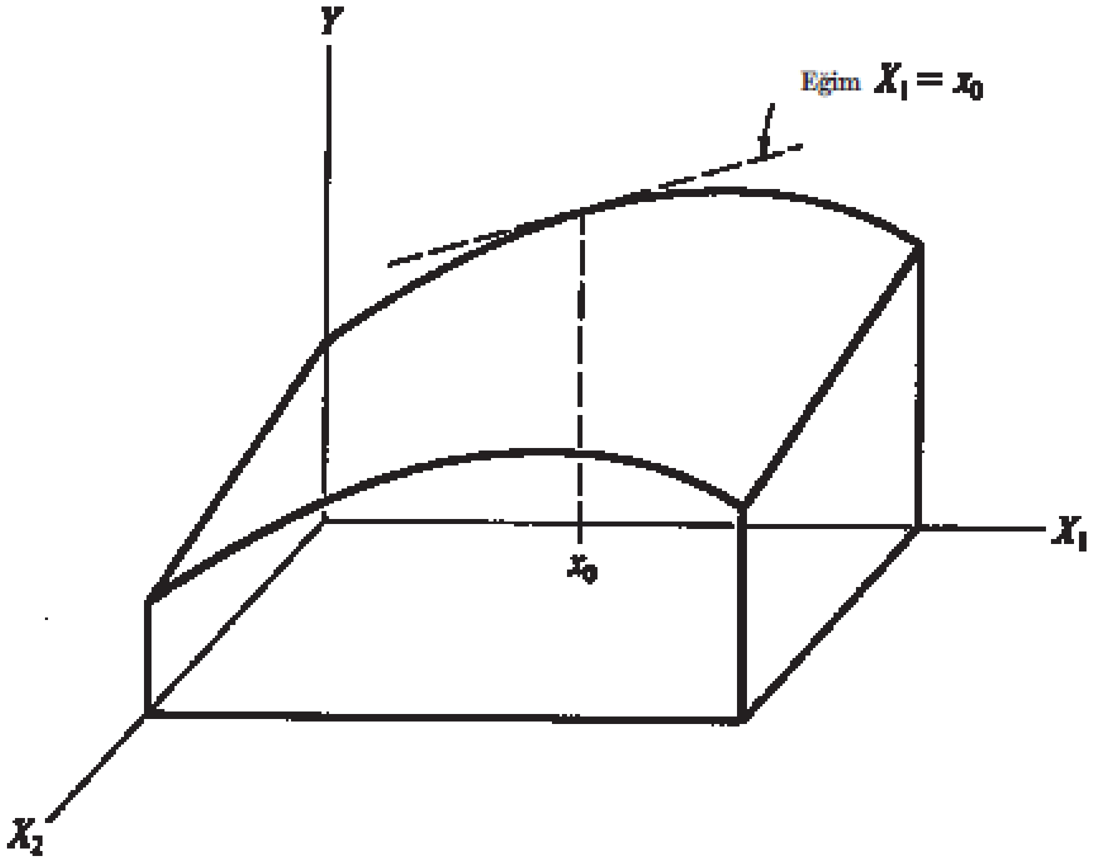
dir. İkinci derece bir polinom modeli, birinci derece terimlere ek olarak bağımsız değişkenlerin karelerini ve çarpımlarını içerir. İki değişkenli ikinci-derece bir polinom modeli

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_{11} X_{i1}^2 + \beta_{12} X_{i1} X_{i2} + \beta_{22} X_{i2}^2 + \epsilon_i. \quad (8.28)$$

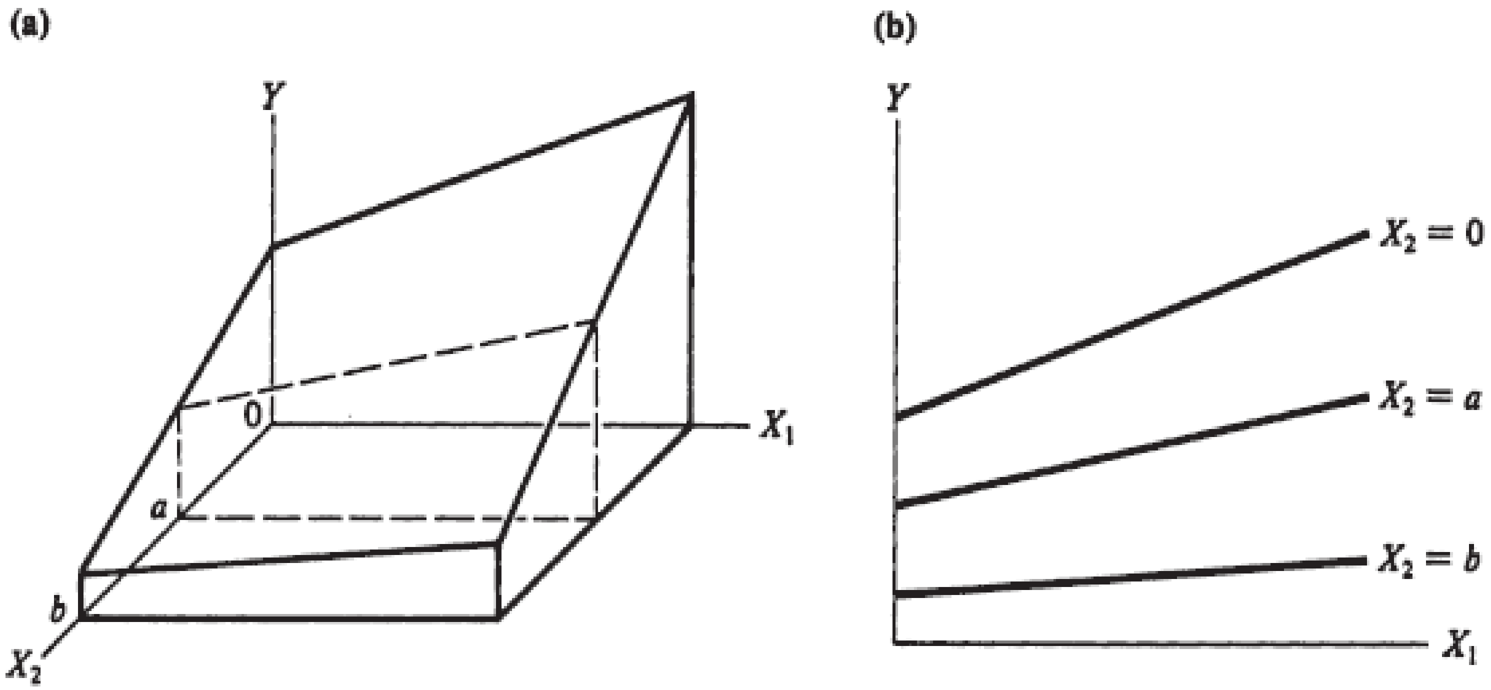
dir. Polinomdaki bir terimin **derecesi** (ya da sırası), terimde yer alan bağımsız değişkenlerin üslerinin *toplamıdır*. Bütün bir polinomun derecesi ise en yüksek dereceli terimin derecesidir. Bütün polinom modelleri, derecelerinden bağımsız olarak parametreler açısından doğrusaldır.



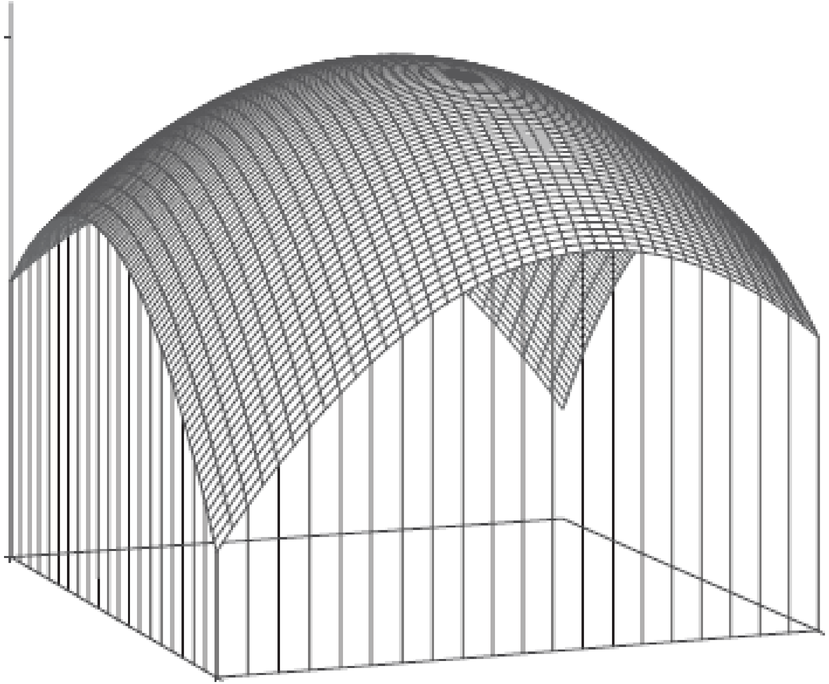
ŞEKİL 8.4. *Birinci derece iki değişkenli polinom yanıt yüzeyi.*



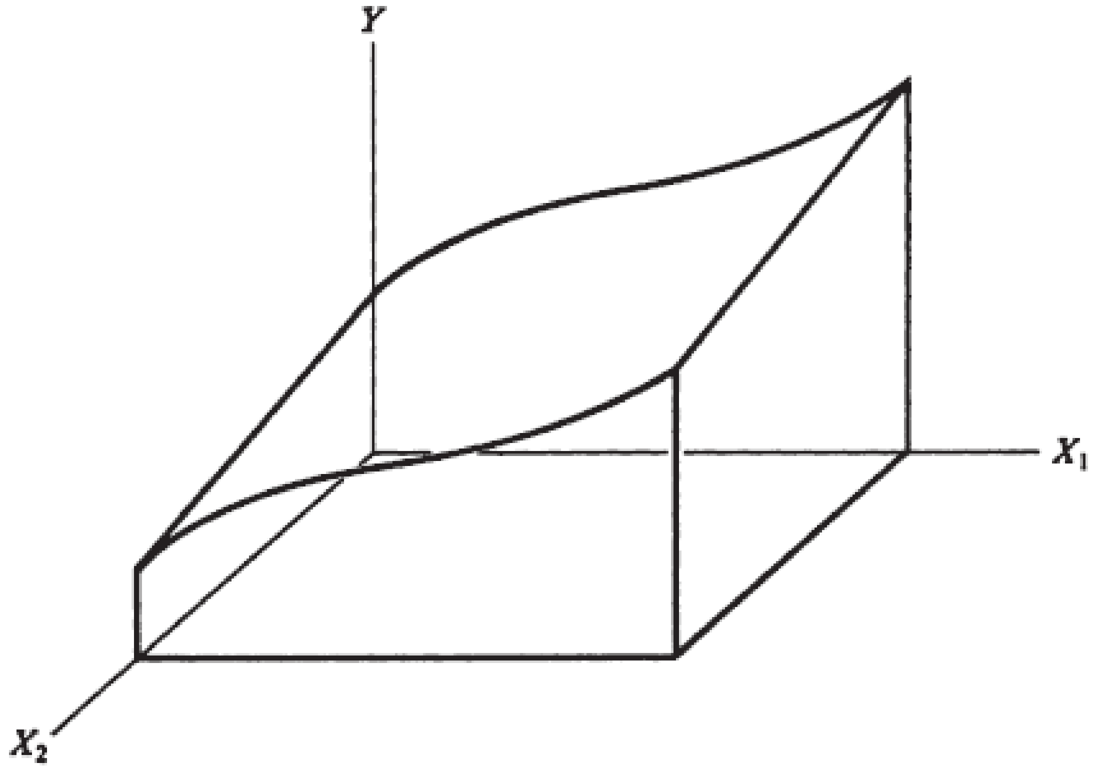
ŞEKİL 8.5. *İkinci derece* X_1 *ve birinci derece* X_2*'nin polinom yanıt yüzeyi.*



ŞEKİL 8.6. *İki değişkenli yanıt yüzeyi (a)etkileşimli ve (b) yüzeyin iki boyutlu bir temsili.*



ŞEKİL 8.7. *Maksimum noktasıyla birlikte iki değişkenli karesel bir yanıt yüzeyi.*



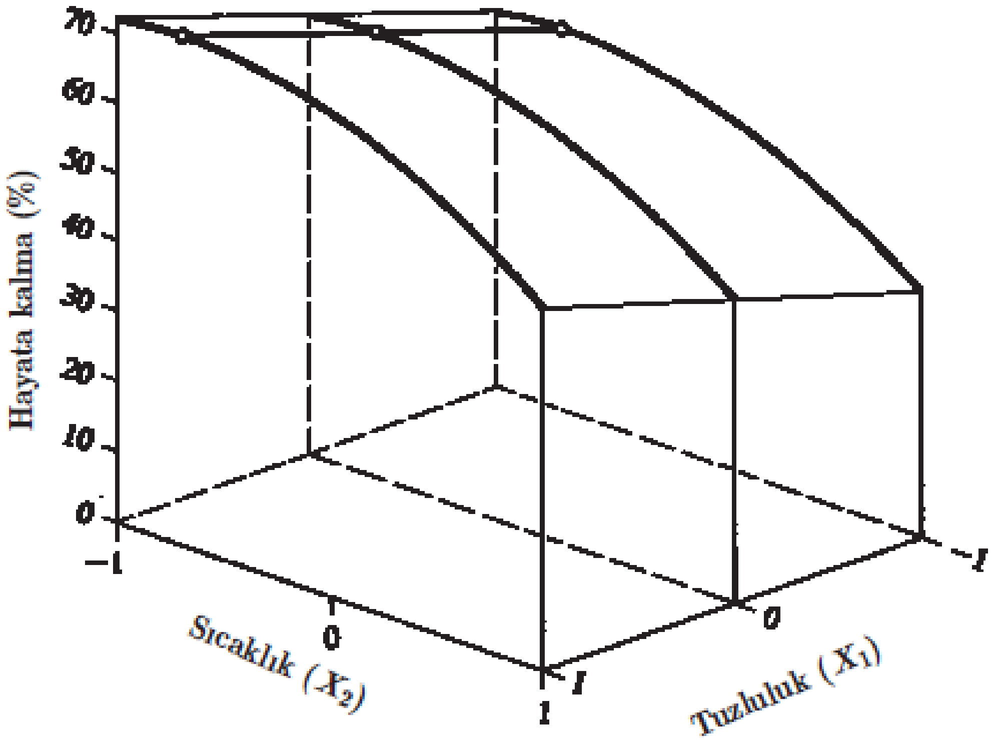
ŞEKİL 8.8. *X_1'in üçüncü derece terimiyle bir polinom yanıt yüzeyi.*

TABLO 8.3. *Tuzluluk* (X_1), *sıcaklık* (X_2) *ve çözünmüş oksijen* (X_3)*'nin işlem kombinasyonları ve 3 mg/l sodyum pentakloropenat'a maruz kalmanın ortanca ölüm zamanı [Alderdice (1963) verisi; izinle kullanılmıştır.].*

	Tuzluluk	*Sıcaklık*	*Oksijen*	*Ortanca Ölüm Zamanı*	
Deneme	X_1	X_2	X_3	*Tekrar 1*	*Tekrar 2*
1	−1	−1	−1	53	50
2	−1	−1	1	54	42
3	−1	1	−1	40	31
4	−1	1	1	37	28
5	1	−1	−1	84	57
6	1	−1	1	76	78
7	1	1	−1	40	49
8	1	1	1	50	54
9	0	0	0	50	50
10	1.215	0	0	61	76
11	−1.215	0	0	54	45
12	0	1.215	0	39	33
13	0	−1.215	0	67	54
14	0	0	1.215	44	45
15	0	0	−1.215	61	38
16	−1.2500	−1.8867	−.6350	46	
17	.8600	−2.2200	−.4250	66	
18	1.0000	−2.2400	−.3100	68	
19	2.1165	−2.4167	−.1450	75	
20	2.5825	−2.4900	−.0800	75	
21	3.2475	−2.6667	.0800	68	
22	1.1760	−1.3333	0	78	
23	1.4700	−1.6667	0	93	
24	1.7640	−2.0000	0	96	
25	2.0580	−2.3333	0	66	

TABLO 8.4. *Alderdice (1963) verisinindeki üç değşken için ikinci derece tam polinom modelinin kısmi regresyon katsayıları.*

Terim	$\widehat{\beta}_j$	$s(\widehat{\beta}_j)$	*Student* t^a *İstatistiği*
X_1	9.127	1.772	5.151
X_2	−9.852	1.855	−5.312
X_3	.263	1.862	.141
X_1^2	−1.260	1.464	−.861
X_2^2	−6.498	2.014	−3.225
X_3^2	−2.985	2.952	−1.011
X_1X_2	−.934	1.510	−.618
X_1X_3	2.242	2.150	1.042
X_2X_3	−.139	2.138	−.065



ŞEKİL 8.9. *3 mg/l sodyum pentakloropenate maruz kalan coho somonunun hayatta kalma süresini su sıcaklığı ve su tuzluluğu ile ilişkilendiren iki değişkenli yanıt yüzeyi. Çözünmüş oksijenin (X_3) önemli bir etkisi yoktu. [Alderdice(1963) verisi; izinle kullanılmıştır].*