

İstatistik

Regresyon Analizi

(11. Hafta)

Tekiner KAYA

DOĞRUSAL REGRESYON

Regresyon analizi, bilinen bulgulardan, bilinmeyen gelecekteki olaylarla ilgili tahminler yapılmasına izin verir. Regresyon, bağımlı ve bağımsız değişken(ler) arasındaki ilişkiyi ve doğrusal eğri kavramını kullanarak, bir tahmin eşitliği geliştirir. Değişkenler arasındaki ilişki belirlendikten sonra, bağımsız değişken(ler)in değeri bilindiğinde bağımlı değişkenin değeri tahmin edilebilir.

Bağımlı Değişken (y)

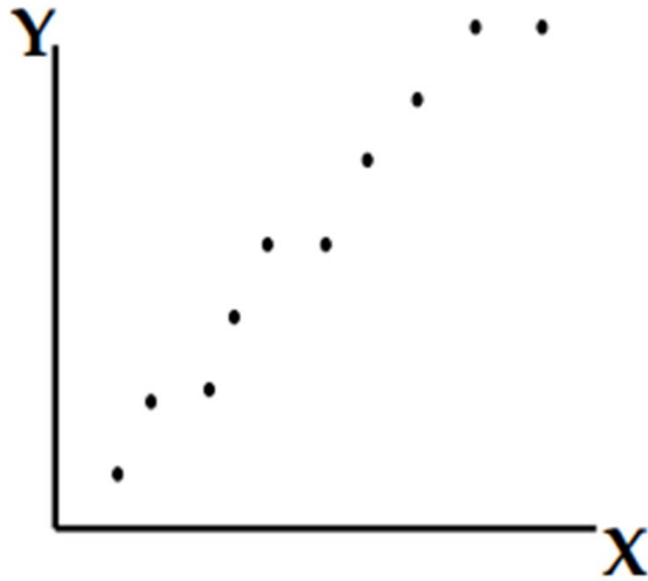
- Bağımlı değişken, regresyon modelinde açıklanan ya da tahmin edilen değişkendir. Bu değişkenin bağımsız değişken ile ilişkili olduğu varsayılır.

«Bağımsız değişkenin değişiminden etkilenen değişkendir»

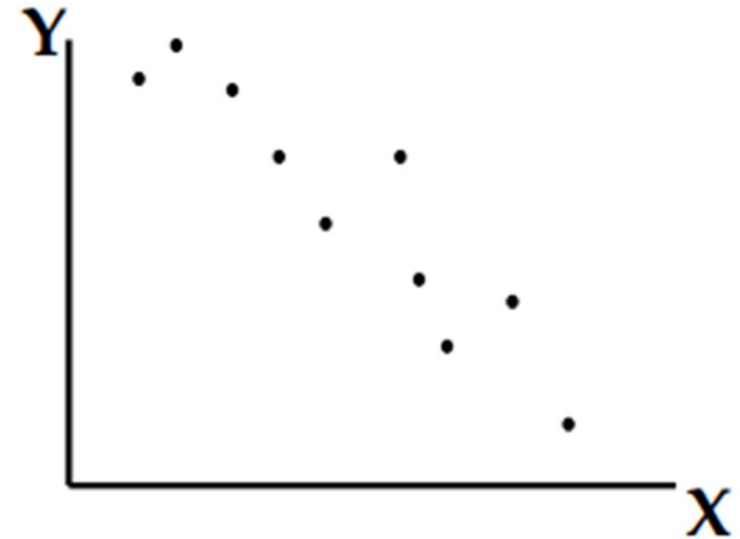
Bağımsız Değişken (x)

- Bağımsız değişken, regresyon modelinde açıklayıcı değişken olup; bağımlı değişkenin değerini tahmin etmek için kullanılır.
-

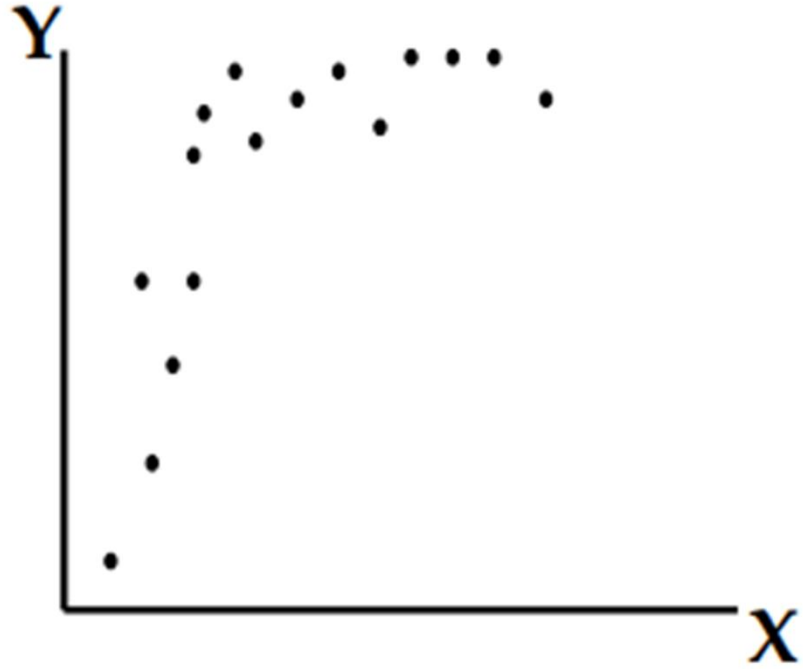
- Değişkenler arasında doğrusal ilişki olabileceği gibi, doğrusal olmayan bir ilişki de olabilir. Bu nedenle, saçılım grafiği yapılmadan (ilişki yok/doğrusal ilişki var/doğrusal olmayan ilişki var) regresyon analizine karar verilmemesi gerekir.
 - Bu bilgiler doğrultusunda, tek/çok değişkenli doğrusal regresyon analizlerinin yanı sıra, tek/çok değişkenli doğrusal olmayan regresyon analizleri de mevcuttur.
-



(a) (+) yönlü Doğrusal
İlişki



(b) (-) yönlü Doğrusal
İlişki



(c) Doğrusal Olmayan
İlişki



(d) İlişki Yok

Bağımlı değişken ile bağımsız değişken arasındaki doğrusal ilişkiyi açıklayan tek değişkenli regresyon modeli aşağıdaki gibidir:

$$y = a + bx$$

y = Bağımlı değişkenin değeri

a = Regresyon doğrusunun kesim değeri (Sabit değer)

b = Regresyon doğrusunun eğimi

x = Bağımsız değişkenin değeri

Regresyon Analizi

Doğrusal olduğu varsayılan genel ilişki : $Y = \alpha + \beta X + E$

n sayıda gözleme dayalı olarak tahmin edilir : $\hat{Y} = a + bX_i + e_i$
($i = 1, 2, 3, \dots, n$)

Talep “Y” ve reklam harcaması “X” arasındaki ilişkinin doğrusal olduğu varsayılır ve **reklama bağlı olarak talep tahmini** yapılacak ise;

$$\hat{Y} = a + bX_i + e_i$$

doğrusal modeldeki **geçmiş fiili veriler (gözlem değerleri)** kullanılır. Burada e_i **hata payı**, talebin reklam değişkeni tarafından açıklanamayan (*diğ.faktörler nedeniyle*) kısmını ifade etmektedir. **Gözlem değerlerine uyan en iyi doğruyu belirlemek için En Küçük Kareler Yöntemi** ile “a” ve “b” katsayıları tahmin edilir.

Regresyon Analizi

Satış Miktarı (10 ³ birim)	Reklam Harcaması (10 ⁶ TL)		
Y_i	X_i	$X_i Y_i$	X_i^2
18	200	3600	40000
10	86	860	7396
---	---	---	---
$\Sigma Y_i = 146$	$\Sigma X_i = 1503$	$\Sigma X_i Y_i = 27437$	$\Sigma X_i^2 = 287325$

$$\hat{Y} = a + bX_i$$

basit regresyon modeli aşağıdaki hali alır : $b = \left[\frac{n \Sigma X_i Y_i - \Sigma X_i \Sigma Y_i}{n \Sigma X_i^2 - (\Sigma X_i)^2} \right]$

$$b = \left[\frac{10 \cdot (27437) - 1503(146)}{10(287325) - (1503)^2} \right]$$

ve $b = 0.089$ olarak hesaplanır.

Regresyon Analizi

Benzer biçimde ; $a = \frac{\sum Y_i}{n} - \frac{b \sum X_i}{n}$

formülünden yararlanılarak “a” parametresi bulunur.

$$a = \frac{146}{10} - \frac{0.089(1503)}{10}$$

a = 1.22 olarak hesaplanır. Doğrusal tahmin modeli:

$$\hat{Y} = 1.22 + 0.089(400)$$

$$\hat{Y} = 36.82$$

Başka ifadeyle, **400milyon TL’lik reklam için 36.820 birimlik satış tahmin edilmektedir.**

Regresyon Analizi

Basit regresyon analizinde **“a” katsayısı** **“Y”** bağımlı değişkeninin **“X”** bağımsız değişkeninden bağımsız olarak alacağı değeri ve aynı zamanda **doğrunun “Y” dikey eksenini kestiği başlangıç noktasını gösterir.**

“b” katsayısı ise, “X” bağımsız değişkenindeki 1 birimlik değişimin “Y” bağımlı değişkeninde hangi oranda değişim oluşturduğunu gösterir. Ayrıca, modelin ifade ettiği **doğrunun da eğimine karşılık gelir.**

Regresyon Analizi

Eğer

$b = 0$ ise, Y ve X arasında ilişki yoktur.

$b < 0$ ise, Y ve X arasında ters (negatif) yönlü ilişki,

$b > 0$ ise, Y ve X arasında olumlu (pozitif) yönlü ilişki vardır.

Regresyon analizlerinde, tahmin yapılmadan önce modelin istatistiksel anlamlılık düzeyinin test edilmesi doğru tahmin işlemi için gerekliliktir.

Örnek:

Kardiyoloji kliniğine başvuran 20 yaş üzerindeki erkek hastalar üzerinde yapılan bir araştırmada, yaş ve kolesterol değişkeni arasındaki ilişkiden yola çıkılarak kurulan regresyon modeli aşağıdaki gibi elde edilmiştir:

$$Y = 151.44 + 1.428X$$

$$\text{Kolesterol} = 151.44 + 1.428 \times \text{Yaş}$$

Bu modele göre, yaştaki bir birimlik artışın, kolesterol değerinde 1.428 birimlik bir artışa neden olacağı, 20 yaşındaki bir erkeğin ($X=20$) kolesterol değerinin ise 180 olacağı söylenebilir.

Kurulan bu modele göre, 50 yaşında bir erkeğin kolesterol değerinin ne kadar olacağını tahmin edebiliriz

X=50 için :

$$Y = 151.44 + 1.428X$$
$$= 151.44 + 1.428 * 50 = 222.84$$

50 yaşında bir erkeğin kolesterol değerinin 222.84 olacağı söylenebilir.

Regresyon ilişki örnekleri

Örnek: Teneke ton balığı satışları, ton balığının fiyatına, diğer markaların fiyatlarına ve reklam kampanyalarına bağlıdır.

Örnek: Ev mutfak masrafı evin gelirine, kişi sayısına ve yaş dağılımına bağlıdır.

Örnek: Ev fiyatları evin büyüklüğüne, yerine ve şömüne, havuz vs. gibi House price depends upon its size, location, konforlara bağlıdır.

Regresyon Analizi

• Örnek 1

4 Basit Lineer Regresyon Doğrusu

- Bir araba satıcısı kilometre sayacı ile ikinci el araba satış fiyatı arasında bir ilişki bulmak istiyor.
- 100 araba rassal olarak seçiliyor ve veri kaydediliyor.
- Regresyon doğrusunu bulun.

Araba	Sayaç	Fiyat
1	37388	14636
2	44758	14122
3	45833	14016
4	30862	15590
5	31705	15568
6	34010	14718
7	...	...
8	...	...

Bağımsız değişken x

Bağımlı değişken y

Lineer regresyon doğrusunu tahmin edin

4 Basit Lineer Regresyon Doğrusu

- Çözüm

- Elle çözüm: İstatistikleri hesaplayın

$$\bar{x} = 36,009.45; \quad s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = 43,528,690$$

$$\bar{y} = 14,822.823; \quad \text{cov}(X, Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} = -2,712,511$$

$$n = 100.$$

$$b_1 = \frac{\text{cov}(X, Y)}{s_x^2} = \frac{-2,712,511}{43,528,690} = -.06232$$

$$b_0 = \bar{y} - b_1 \bar{x} = 14,822.82 - (-.06232)(36,009.45) = 17,067$$

$$\hat{y} = b_0 + b_1 x = 17,067 - .0623 x$$

Regresyon Analizi

Çözüm Yorumu ve Çıkarımlar

$$\hat{y} = b_0 + b_1x = 17,067 - .0623 x$$

Burada b_0 = sabit değerdir ($y=a+bx$ 'deki a ile aynıdır yani)

Aracın satış fiyatı kaç km de olduğuna göre değişmekte ise; ben x yerine almayı planladığım aracın km sini yazarsam ve denklemi hesaplarsam, aracın ortalama ne kadar edeceğini bulabilirim. Örneğin, 45.000 km deki bir aracın fiyatı;

$$Y = 17.067 - 0,0623 \times 45.000 = 14.263 \text{ dolar olacaktır.}$$
