

Alternatif Regresyon Metotları

Bu bölümde, normal dağılım varsayımından sapmalara ve aykırı değerlere karşı robust olan alternatif regresyon metotları tanıtılacaktır.

1. Wald Metodu:

Adım 1. x_i değerleri küçükten büyüğe doğru sıralanır.

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

Adım 2. $x_{(i)}$ 'lere karşılık gelen y_i değerleri (concomitant olarak adlandırılır) belirlenir ve $x_{(i)}$ 'lerin karşılıklarına yazılır.

$$(x_{(i)}, y_{[i]}), i=1, 2, \dots, n$$

Adım 3. $(x_{(i)}, y_{[i]})$ verileri iki gruba ayrılır. Gruplar eşit hacimli olmalıdır. Bir başka deyişle, n çift sayı ve $m = \frac{n}{2}$ olmalıdır.

1. Grup: $(x_{(1)}, y_{[1]}), \dots, (x_{(m)}, y_{[m]})$
2. Grup: $(x_{(m+1)}, y_{[m+1]}), \dots, (x_{(n)}, y_{[n]})$

Wald metodunda, eğim

$$b_w = \frac{(y_{m+1} + \dots + y_n) - (y_1 + y_2 + \dots + y_m)}{(x_{m+1} + \dots + x_n) - (x_1 + x_2 + \dots + x_m)}$$

olarak ve kesim noktası,

$$a_w = \bar{y} - b_w \bar{x}$$

olarak tanımlanır. Burada,

$$\bar{x} = \frac{x_1 + \dots + x_n}{n} \quad \bar{y} = \frac{y_1 + \dots + y_n}{n}$$

olarak tanımlanır.

Wald metoduna dayalı olarak elde edilen regresyon denkleminin tahmini

$$\hat{y}_{w,i} = a_w + b_w x_i, \quad i = 1, 2, \dots, n$$

şeklinde ifade edilir.

Bartlett Metodu:

Bartlett metodu, Wald metoduna benzer bir yöntemdir.

Veriler eşit hacimli olmasına gerek olmayan Alt, Orta ve Üst grup olarak adlandırılan üç gruba bölünür.

Burada, $n = n_L + n_u + n_m$ dir ve

n_L : Alt gruptaki gözlem sayısı

n_u : Üst gruptaki gözlem sayısı ve

n_m : Orta gruptaki gözlem sayısı

olarak tanımlanır.

Regresyon denkleminin eğimi ve kesim noktası Bartlett yönteminde

$$b_B = \frac{\bar{y}_u - \bar{y}_L}{\bar{x}_u - \bar{x}_L} \text{ ve}$$

$$a_B = \bar{y} - b_B \bar{x}$$

olarak tanımlanır. Burada, regresyon doğrusu Alt ve Üst grupların ortalamaları olarak ifade edilen (\bar{x}_L, \bar{y}_L) ve (\bar{x}_u, \bar{y}_u) noktaları kullanılarak elde edilir. Bu ortalamalar,

$$\bar{x}_L = \frac{x_1 + x_2 + \dots + x_{n_L}}{n_L}$$

$$\bar{y}_L = \frac{y_1 + y_2 + \dots + y_{n_L}}{n_L}$$

$$\bar{x}_u = \frac{x_{n-n_u+1} + \dots + x_n}{n_u}$$

$$\bar{y}_u = \frac{y_{n-n_u+1} + \dots + y_n}{n_u}$$

eşitlikleri yardımıyla hesaplanır.

Bartlett metoduna dayalı olarak elde edilen regresyon denkleminin tahmini

$$\hat{y}_{B,i} = a_B + b_B x_i, \quad i = 1, 2, \dots, n$$

şeklinde ifade edilir.

Veriler aşağıda gösterildiği gibi 3 gruba bölünür.

$n=3k$	$n=3k+1$	$n=3k+2$
k	k	k+1
k	k+1	k
k	k	k+1

3) Theil Metodu (Median of Pairwise slopes)

Her bir veri çifti için eğimler

$$b_{ij} = \frac{y_j - y_i}{x_j - x_i}, \quad 1 \leq i < j \leq n$$

eşitliği yardımıyla hesaplanır.

Burada, $\frac{n(n-1)}{2}$ tane eğim (b_{ij}) değeri hesaplanır. Hesaplanan eğimlerin medyanı

$$b_T = \text{medyan}\{b_{ij}\}$$

bulunur ve bu medyan değeri kullanılarak kesim noktası

$$a_T = \text{medyan}\{y_i - b_T x_i\}$$

eşitliği kullanılarak elde edilir.

Sonuç olarak Theil metoduna dayalı olarak elde edilen regresyon denkleminin tahmini

$$\hat{y}_{T,i} = a_T + b_T x_i, \quad i = 1, 2, \dots, n$$

şeklinde ifade edilir.

Not: Bu yöntemde tüm x_i 'lerin birbirinden farklı olduğu varsayılır.

Örnek: Aşağıdaki veriler için,

a) Wald

b) Bartlett

c) Theil

metotlarını kullanarak regresyon denklemini tahmin ediniz.

X	Y
8	62
9	80
10	72
12	70
14	95
24	120
26	110
28	115
31	125
37	130

a) $n=10$

Veriler her biri 5 gözlem içeren 2 gruba bölünür. Buradan, regresyon denkleminin eğimi ve kesim noktası sırasıyla

$$b_w = \frac{(y_6 + \dots + y_{10}) - (y_1 + \dots + y_5)}{(x_6 + \dots + x_{10}) - (x_1 + \dots + x_5)} = \frac{600 - 379}{146 - 53} = 2.376$$

ve

$$\begin{aligned}
a_w &= \bar{y} - b_w \bar{x} \\
&= 97.9 - 2.376(19.9) \\
&= 50.618
\end{aligned}$$

olarak hesaplanır. Burada,

$$\begin{aligned}
\bar{y} &= \frac{\sum y_i}{n} = 97.9 \\
\bar{x} &= \frac{\sum x_i}{n} = 19.9
\end{aligned}$$

dır.

Regresyon denkleminin tahmini

$$\Rightarrow \hat{y}_{w,i} = 50.618 + 2.376x_i$$

olarak elde edilir.

b) $n=10$ ($3k+1$)

$k \rightarrow 3$

$k \rightarrow 4$ şeklinde veri seti üçe bölünür.

$k \rightarrow 3$

Üst ve Alt gruplar için gerekli ortalama değerleri

$$\bar{y}_u = \frac{\sum_{i=8}^{10} y_i}{3} = 123.33$$

$$\bar{y}_L = \frac{\sum_{i=1}^3 y_i}{3} = 71.33$$

$$\bar{x}_u = \frac{\sum_{i=8}^{10} x_i}{3} = 32$$

$$\bar{x}_L = \frac{\sum_{i=1}^3 x_i}{3} = 9$$

olarak elde edilir. Bu ortalamalar kullanılarak regresyon denkleminin eğimi ve kesim noktası sırasıyla

$$b_B = \frac{\bar{y}_u - \bar{y}_L}{\bar{x}_u - \bar{x}_L} = \frac{123.33 - 71.33}{32 - 9} = 2.26$$

ve

$$\begin{aligned} a_B &= \bar{y} - b_B \bar{x} \\ &= 97.9 - 2.26(19.9) \\ &= 52.926 \end{aligned}$$

olarak hesaplanır. Burada,

$$\begin{aligned} \bar{y} &= \frac{\sum y_i}{10} = 97.9 \\ \bar{x} &= \frac{\sum x_i}{10} = 19.9 \end{aligned}$$

dır. Regresyon denkleminin tahmini

$$\Rightarrow \hat{y}_{B,i} = 52.926 + 2.26x_i$$

olarak elde edilir.

c) $n=10$ tane gözlem olduğundan $\frac{n(n-1)}{2} = \frac{10 \cdot 9}{2} = 45$ tane eğim hesaplanır. Bu eğimler aşağıdaki tabloda gösterilmiştir.

j	b_{1j}	b_{2j}	b_{3j}	b_{4j}	b_{5j}	b_{6j}	b_{7j}	b_{8j}	b_{9j}
1									
2	18								
3	5	-8							
4	2	-3.33	-1						
5	5.5	3	5.75	12.5					
6	3.62	2.67	3.43	4.17	2.5				
7	2.66	1.76	2.37	2.86	1.25	-5			
8	2.65	1.84	2.39	2.81	1.43	-1.25	2.5		
9	2.74	2.04	2.52	2.89	1.76	0.71	3	3.33	
10	2.34	1.78	2.15	2.4	1.52	0.77	1.82	1.67	0.83

Bu eğimlerin medyanı

$$b_T = \text{medyan}\{b_{ij}\} = b_{(23)} = 2.39$$

regresyon denkleminin eğimidir. Bu medyan değeri kullanılarak kesim noktası

$$a_T = \text{medyan}\{y_i - b_T x_i\} = \text{medyan}\{a_i\} = \frac{a_{(5)} + a_{(6)}}{2} = 48.09$$

olarak hesaplanır. Burada, $a_i = y_i - b_T x_i$ deęerleri ařaęıdaki tabloda gsterildięi gibi elde edilmiřtir.

$y_i - b_T x_i$
42.88
58.49
48.10
41.32
61.54
62.64
47.86
48.08
50.91
41.57

Regresyon denkleminin tahmini

$$\hat{y}_{T,i} = 48.09 + 2.39x_i$$

olarak bulunur.