

Aykırı Değer Belirleme Testleri

Bu bölümde aykırı değerleri belirlemek için kullanılan istatistiksel testler anlatılacaktır.

Z Değerleri Tekniği

Standart Normal dağılımda hemen hemen her veri

$$z_i = \frac{x_i - \bar{x}}{s}$$

-3 ile 3 arasında değer alır. Burada, \bar{x} ortalamayı, s ise standart sapmayı ifade eder. Bu nedenle, standart normal dağılımda -3 ten küçük 3 ten büyük değerler aykırı değer olarak kabul edilir.

Örnek:

Grafiksel yöntemler ile aykırı değerlerin tespit edildiği veri setine bakacak olursak

$$\bar{x} = 180.45$$

$$s = 17.90$$

olmak üzere,

235 için,

$$z_1 = \frac{235 - 180.45}{17.90} = 3.0473$$

bulunur.

128 için,

$$z_2 = \frac{128 - 180.45}{17.90} = -2.9305$$

bulunur. $z_1 > 3$ olduğundan 235 aykırı değer iken, $-3 < z_2 < 3$ olduğundan 128 aykırı değer değildir.

Grubbs Testi

Grubbs testi normallik varsayımı altında aykırı değer tespit etmek için kullanılan bir testtir. Bir başka deyişle, Grubbs testini uygulamadan önce veri setinin normal dağılım varsayımını sağlaması gerekmektedir.

Grubbs testi için sıfır ve alternatif hipotezleri aşağıda verilmiştir.

H_0 : Veri setinde aykırı değer yoktur.

H_1 : Veri setinde en az bir aykırı değer vardır.

Test istatistiği,

$$G = \frac{\max|y_i - \bar{y}|}{s}$$

olarak ifade edilir. Burada, \bar{y} örneklem ortalamasını, s ise standart sapmayı gösterir. Eğer,

$$G > \frac{N-1}{\sqrt{N}} \sqrt{\frac{t^2_{(\alpha/2N, N-2)}}{N-2 + t^2_{(\alpha/2N, N-2)}}}$$

koşulu sağlanırsa H_0 hipotezi reddedilir. Burada, $t_{(\alpha/2N, N-2)}$, $N-2$ serbestlik derecesi ve $\frac{\alpha}{2N}$ anlam seviyeli t dağılımının tablo değeridir. N ise örneklemin hacmini gösterir.

Örnek: Aşağıdaki veri seti için Grubbs testini kullanarak veri setinde aykırı değer olup olmadığını belirleyiniz.

5.2	116.4	256.0	712.4
7.9	117.3	276.9	965.2
17.8	118.2	278.1	1576.4
32.3	127.5	302.5	2765.6
33.8	196.3	337.3	
41.7	201.8	432.1	
93.5	243.5	488.5	

$$\bar{y} = 389.768$$

$$s = 606.70$$

$N = 25$ ve $\alpha = 0.01$ olmak üzere,

$$G = \frac{|2765.6 - 389.768|}{606.70} = 3.916$$

$$\frac{(N-1)}{\sqrt{N}} \sqrt{\frac{t^2_{(\alpha/2N, N-2)}}{N-2 + t^2_{(\alpha/2N, N-2)}}} = 3.135$$

$3.916 > 3.135 \Rightarrow H_0$ reddedilir. Yani 2765.6 aykırı değerdir.

1576.4 için,

$$G = \frac{|1576.4 - 389.768|}{606.70} = 1.956$$

$1.956 < 3.135 \Rightarrow H_0$ reddedilemez. Yani 1576.4 aykırı değer değildir.

5.2 için,

$$G = \frac{|5.2 - 389.768|}{606.70} = 0.634$$

$0.634 < 3.135 \Rightarrow H_0$ reddedilemez. Yani 5.2 aykırı değer değildir.

Q Testi:

Q testi de aykırı değer belirleme yöntemlerinden biridir. Sadece bir veri üzerinde yoğunlaşılır. Sıfır ve alternatif hipotezler,

$H_0: x_i$ aykırı değer değildir

$H_1: x_i$ aykırı değerdir

olarak ifade edilir.

Test istatistiği,

$$Q_c = \frac{|x_i - x'_i \text{ ye en yakın gözlem değeri}|}{|max - min|}$$

olarak tanımlanır.

Q değeri için kritik değerler aşağıda gösterildiği gibidir.

n	%90	%95	%99
3	0.941	0.970	0.994
4	0.765	0.829	0.926
5	0.642	0.710	0.821
6	0.560	0.625	0.764
7	0.507	0.568	0.680
8	0.468	0.526	0.634

Eğer $Q_c > Q_{tablo} \Rightarrow H_0$ hipotezi reddedilir.

Örnek: Aşağıdaki veri seti için Q testini kullanarak veri setinde aykırı değer olup olmadığını belirleyiniz.

144.8, 177.2, 148.4, 143.2, 145.2, 147.1, 142.1, 143.4

Hipotezler

H_0 : 177.2 aykırı değer değildir

H_1 : 177.2 aykırı değerdir

olarak ifade edilir.

$x_i = 177.2$

x_i 'ye en yakın değer = 148.4

max=177.2

min=142.1

$$Q_c = \frac{177.2 - 148.4}{177.2 - 142.1} = 0.821$$

n=8 ve $\alpha = 0.10$ için $Q_{tablo} = 0.468$

$0.821 > 0.468$ olduğundan H_0 hipotezi reddedilir. Yani 177.2 aykırı değerdir.