Konum Tahmin Edicilerinin Karşılaştırılması

- Dağılımla ilgili varsayımlar uygulamada her zaman sağlanmayabilir.
- Bu durumlarda varsayılan dağılımın makul alternatifleri altında yüksek etkinliğe sahip tahmin ediciler kullanılır.
- Burada dağılımın simetrik olduğu varsayılacaktır.

Simetrik dağılımlar için,

ortalama = medyan = simetri merkezidir.

Dağılımlar genel olarak simetrik ve çarpık dağılımlar olarak ikiye ayrılmaktadır. Simetrik olmayan dağılımlarda ortalama değiştikçe varyans ta değişmektedir.

Tanım: $f(x;\theta,\lambda)$, X rastgele değişkenin olasılık yoğunluk fonksiyonu olsun. Eğer söz konusu olasılık yoğunluk fonksiyonu $(X-\theta)$' nın bir fonksiyonu şeklinde yazılabiliyorsa, θ parametresine konum (location) parametresi denir.

Tanım: (Sıra istatistikleri)
$X, X_2, \ldots, X_n$, rastgele değişkenler olmak üzere küçükten büyüğe doğru sıralandığında

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

sıra istatistikleri olarak adlandırılır. Burada, $X_{(i)}$ $i.$ sıra istatistiği olarak tanımlanır.

L Tahmin Edicileri (L Estimators)

L tahmin edicisi,

$$T = \sum_{i=1}^{n} a_i X_{(i)}$$

şeklinde tanımlanır.

Burada, $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ sıra istatistikleri, $a_1, a_2, \ldots, a_n$, $0 \leq a_i \leq 1, i = 1,2, \ldots, n$ ve $\sum a_i = 1$ koşullarını sağlayan reel sayılardır. $a_1, a_2, \ldots, a_n$ ağırlık olarak da tanımlanır.

L tahmin edicisi adı verilen tahmin ediciler sınıfı sıra istatistiklerinin doğrusal kombinasyonları olarak tanımlanır.

Örnek: Örneklem ortalaması $\bar{x}$ bu sınıfın içinde yer alır. Çünkü,

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1}{n}x_1 + \frac{1}{n}x_2 + \cdots + \frac{1}{n}x_n$$
$$= \frac{1}{n}x_{(1)} + \frac{1}{n}x_{(2)} + \cdots + \frac{1}{n}x_{(n)}$$
$$= \frac{\sum x_{(i)}}{n}$$

olarak ifade edilir.

Budanmış Ortalama (Trimmed Mean):

Budanmış ortalama, aykırı değerlerin etkisini yok etmek amacıyla kullanılan bir tahmin edicidir ve

$$T(\alpha) = \frac{1}{n(1-2\alpha)} \left\{ (1-r)\left[X_{(g+1)} + X_{(n-g)}\right] + \sum_{i=g+2}^{n-g-1} X_{(i)} \right\}$$

formülü yardımıyla hesaplanır.
Burada T(α)'nın, a_i $(i = 1,2, \ldots, n)$ ağırlıklarıyla

$$a_i = \begin{cases} 0 & eğer\ i \leq g\ ya\ da\ i \geq n+g+1 \\ \dfrac{(1-r)}{n(1-2\alpha)} & eğer\ i = g+1\ ya\ da\ i = i-g \\ \dfrac{1}{n(1-2\alpha)} & eğer\ g+2 \leq i \leq n-g-1 \end{cases}$$

bir L tahmin edicisi olduğu açıktır.

Burada, $g = [\alpha n]$ ve $r = \alpha n - g$ olarak tanımlanır. α verinin budanma yüzdesini gösterir.

Örnek: 20 hacimli bir örneklemin %20 budanmış ortalaması en küçük ve en büyük 4 gözlemin atılıp geriye kalan 12 gözlemin ortalaması alınarak bulunur.

$\alpha = 0.20 \;\Rightarrow g = [0.20(20)] = 4$ ve $r = 4 - 4 = 0$

$T(0.20) = \frac{1}{12}\left\{X_{(5)} + X_{(16)} + \sum_{i=6}^{15} X_{(i)}\right\}$

$T(0.20) = \frac{1}{12}\sum_{i=5}^{16} X_{(i)}$

- Örneklem ortalaması %0 budanmış ortalama olarak ifade edilir.
- Medyan %50 budanmış ortalama gibi ele alınabilir.

Örnek: Aşağıdaki veri setini kullanarak α=0, 0.05, 0.10, 0.20, 0.30 ve 0.40 için budanmış ortalamayı hesaplayınız.

30	25	31	23	33	-43	28	17	41	-3
27	21	25	26	26	32	25	28	31	18

-4	3									
-3										
-2										
-1										
-0	3									
0										
1	7	8								
2	1	3	5	5	5	6	6	7	8	8
3	0	1	1	2	3		26 medyan			
4	1									

$\bar{X} = T(0.0) = \frac{1}{20}\sum_{i=1}^{20} X_{(i)} = \frac{441}{20} = 22.05$

$T(0.05) = \frac{1}{18}\sum_{i=2}^{19} X_{(i)} = \frac{443}{18} = 24.61$

$T(0.10) = \frac{1}{16}\sum_{i=3}^{18} X_{(i)} = \frac{413}{16} = 25.81$ $\Rightarrow$ bundan sonra değişim çok değil.

$T(0.20) = \frac{1}{12}\sum_{i=5}^{16} X_{(i)} = \frac{315}{12} = 26.25$

$T(0.30) = \frac{1}{8}\sum_{i=7}^{14} X_{(i)} = \frac{210}{8} = 26.25$

$T(0.40) = \frac{1}{4}\sum_{i=9}^{12} X_{(i)} = \frac{104}{4} = 26$

- Bu sonuçlarla gövde yaprak gösterimini birlikte değerlendirirsek örneklem ortalaması $\bar{X}$ güvenilir değildir. Çünkü, veri setindeki aykırı değerden etkilenmektedir.
- Uygulamada en çok α=0.10 kullanılmaktadır. Çünkü bir veri setinde %5 ile %10 arasında aykırı değer bulunması beklenen bir durumdur.

Göreli Etkinlik (Relative Efficiency):

Medyanın $\bar{X}$ ‘ya göre göreli etkinliği

$$RE(Medyan|\bar{X}) = \frac{Var(Medyan)}{Var(\bar{X})}$$

olarak tanımlanır.

$RE(Medyan|\bar{X}) < 1 \Rightarrow$ Medyan daha etkindir.
$RE(Medyan|\bar{X}) > 1 \Rightarrow \bar{X}$ daha etkindir.

Bir tahmin edicinin robust olması herhangi bir özel dağılım için en etkin olmasına gerek duymadan bir dizi dağılım için etkin olmasını garanti eder.

Ortanın ortalaması (Midmean):

Sıra istatistiklerinin merkezinde yer alan %50’lik kısmın ortalamasına midmean adı verilir. Bu aynı zamanda %25’lik budanmış ortalamaya karşılık gelir.

Yukarıdaki veri seti için midmean;

$$Midmean = T(0.25) = \frac{1}{10}\sum_{i=6}^{15} X_{(i)} = \frac{263}{10} = 26.3$$

olarak hesaplanır.

Medyan:

Medyan sıra istatistiklerinin bir fonksiyonu olduğu için bir L tahmin edicisidir. Medyan n'in tek ya da çift olmasına göre ortadaki bir ya da iki gözlem haricindeki diğer tüm gözlemlere 0 ağırlık veren oldukça basit bir L tahmin edicisidir. Medyan

$$Medyan = \begin{cases} X_{\left(\frac{n+1}{2}\right)}, & n\ tek\ sayı\ ise \\ \dfrac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n+1}{2}\right)}}{2}, & n\ çift\ sayı\ ise \end{cases}$$

yardımıyla hesaplanır.

Yukarıdaki veri seti için medyan;

$$Medyan = \frac{X_{(10)} + X_{(11)}}{2} = 26$$

olarak hesaplanır.

Genişletilmiş Medyan (Broadened Medyan-BMED):

a) n tek ise;

Genişletilmiş medyan $5 \leq n \leq 12$ için ortada yer alan 3 sıra istatistiğinin ortalamasıdır.

$n \geq 13$ için ortada yer alan 5 sıra istatistiğinin ortalamasıdır.

b) n çift ise;

Genişletilmiş medyan $5 \leq n \leq 12$ için $1/6, 1/3, 1/3, 1/6$ ağırlıklarıyla ortada yer alan 4 sıra istatistiğinin ortalamasıdır.

$n \geq 13$ için 1/10 , 1/5 , 1/5 , 1/5 , 1/5 , 1/10 ağırlıklarıyla ortada yer alan 6 sıra istatistiğinin ortalamasıdır.

Örnek: n=9 ise

$$BMED = \frac{X(4) + X(5) + X(6)}{3}$$

olarak hesaplanır.

Örnek:

$BMED\ n = 5$ $\Rightarrow$ %20 budanmış ortalamaya eşittir.

$BMED\ n = 10$ $\Rightarrow$ %35 budanmış ortalamaya eşittir.

$BMED\ n = 15$ $\Rightarrow$ %33,3 budanmış ortalamaya eşittir.

$BMED\ n = 20$ $\Rightarrow$ %37,5 budanmış ortalamaya eşittir.

Yukarıdaki örnek için;

$$BMED = \frac{1}{10} X(8) + \frac{1}{5}\left(X(9) + X(10) + X(11) + X(12)\right) + \frac{1}{10} X(13)$$
$$= \frac{1}{10} 25 + \frac{1}{5}(25 + 26 + 26 + 27) + \frac{1}{10} 28 = 26.1$$

olarak bulunur.

Trimean (TRI):

F_L , F_U alt ve üst dörtlükler olmak üzere trimean TRI ile gösterilir ve aşağıdaki gibi hesaplanır.

$$TRI = \frac{1}{4}\left(F_L + 2M + F_U\right)$$

Örnek için;

$$F_L = \frac{1}{2}(21 + 23) = 22$$
$$F_U = \frac{1}{2}(30 + 31) = 30.5$$
$$M = 26$$
$$TRI = \frac{1}{4}(22 + 2 * 26 + 30.5) = 26.125$$

olarak bulunur.

Örnekteki veri için tahmin değerleri aşağıdaki tabloda özetlenmiştir.

Tahmin edici	**Tahmin**
T(0.0)	22.05
T(0.05)	24.61
T(0.10)	25.81
T(0.20)	26.25
T(0.30)	26.25
T(0.40)	26
Midmean	26.3
BMED	26.1
Medyan	26
Trimean	26.125

- En etkin olmayan tahmin edici T(0.0) dır, çünkü veri setinde aykırı değer vardır.
- Budanmış ortalamalar içinde T(0.10)'u kullanmak yeterlidir, çünkü T(0.10) dan sonra tahmin değerlerindeki değişim azalmaktadır.
- En iyi tahmin ediciyi belirlemek için tahmin edicilerin değişik dağılımlar altındaki etkinliklerini Monte Carlo simulasyonu yardımıyla karşılaştırmak gerekir.