

# İSTATİSTİKSEL HATALAR VE ÖRNEKLEME HATASININ ÖLÇÜLMESİ

Yrd.Doç.Dr. Şehamet Bülbül (\*)

## 1.GİRİŞ

Herhangi bir konuda karar vermek veya tahmin yapabilmek için o konu ile ilgili birimler incelenerek gerekli bilgiler sağlanmalıdır. Bilgiler elde edilirken, konu ile ilgili bütün birimler incelenebileceği gibi, bunların arasından seçilecek bir grup birimde incelenebilir. Bunlardan ilkinde tam sayım, ikincisine de örnekleme adı verilir. Gerek tam sayım yönteminin, gerekse örnekleme yönteminin birbirine göre üstünlükleri mevcuttur.

İstatistiksel tahminde amaç, anakütlenin bilinmeyen bir parametresinin değerinin örnekten elde edilen bilgiler yardımı ile belirlenmesidir. Anakütleye ait bilgilerin örnekten elde edilen bilgilerle sağlanmasında da hatanın ortaya çıkması kaçınılmazdır. Çünkü anakütleyi oluşturan birimler temelde ortak özelliklere sahip olmakla birlikte, farklı özellikler de göstereceklerinden, örnekteki birimlerin anakütleyi yüzde yüz temsil etmeleri olanaksızdır. Böylece uygulamada birçok yararlar sağlayan örnekleme yöntemi beraberinde bir takım hatalarında ortaya çıkmasına neden olur. Ancak istatistiksel hatalar sadece örneklemeye başvurmanın sonucunda ortaya çıkmayıp, örneklemenin dışında meydana gelen hatalarda mevcuttur.

İstatistikte basit hesaplama hataları dışındaki hatalar iki grupta incelenebilir. Bunlar, tek yönde gelişen "Sistemik Hatalar" ve tamamen örneklemeden kaynaklanan "Tesadüfi Hatalar" veya "Örnekleme Hataları"dır.

## 2. SİSTEMATİK HATALAR (TEK YÖNLÜ HATALAR)

Bu tür hatalar örneklemeye başvurmanın bir sonucu olmayıp, tam sayım yönteminde de meydana gelen hatalardır. Bu nedenle sistemik hatalar incelenen birimlerin sayısı ile ilgili değildir. Sistemik hataların özelliği tek yönlü olmaları, yani daima aynı yönde ortaya çıkmalarıdır.

---

(\*) M.Ü.I.I.B.F. Ekonometri Bölümü

Tek yönlü hatalar örneklemeden kaynaklanabilecekleri gibi örnekleme dışı nedenlerden de ortaya çıkabilirler. Farklı nedenlerden meydana gelen sistematik hatalar, hataların nedenlerinin bulunup ortadan kaldırılması ile giderilebilirler.

Sistematik hataların nedenleri şöyle gruplandırılabilir:

- a- Anakütle parametresinin örnek değerinden tahmin edilmek istenmesi halinde, örnek seçiminde yanlış örnekleme metodunun kullanılması.
- b- Anakütle parametresi örnek değerinden hareketle tahmin edilir ise, yanlış tahmin yönteminin kullanılması.
- c- Anakütlenin iyi tanımlanamaması. Ana kütle oluşturacak birimlerin çerçevesinin iyi çizilememesi; yani hangi birimlerin anakütlerde yer alacağını iyi belirlenememesi.
- d- Birimlerin iyi tanımlanamaması, gözlenmek istenen birimlerle, gözlenen birimlerin farklı olması.
- e- Örnek seçiminde tesadüflüğün sağlanamamış olması. Tesadüflüğün iki koşulundan ilki anakütlerdeki her birime örneğe seçilmede eşit şans verilmesi, ikincisi ise birimlerin seçiminin birbirinden bağımsız olması veya bir birimin seçiminin başka bir biriminin seçimine bağlı olmaması.
- f- Örnekte yer alan bir kısım birimlerle ulaşım gücünün, zaman darlığı v.s. gibi nedenlerle ilişki kurulamayıp, bunlarla ilgili bilgi sağlanamaması.
- g- Soruların anlaşılır olmaması veya yanlış anlaşılması, doğru bilgi vermektan kaçınılması, sorumluluktan kaçınılması, çekingenlik v.s. gibi nedenlerle birimlerin bir bölümünden cevap alınmaması veya yanlış cevap alınması.
- h - Araştırmayı yapan grubun özellikle anketörlerin konuyu iyi kavrayamamış olması ve dolayısı ile yanlış cevaplamalara neden olmaları.

### 3. TESADÜFİ HATALAR ( İKİ YÖNLÜ HATALAR )

Tesadüfi hatalar, sistematik hataların aksine tamamen örnekleme başvurmanın bir sonucu olarak ortaya çıkan hatalardır. Tesadüfi hatalar iki yönde hareket eden hatalar olduklarından birbirlerinin etkilerini yok ederek, anakütle parametresinin gerçek değerinden tek yönde farklı çıkmasına neden olmazlar. Örneklemenin amacı, ana kütlede mümkün olan bütün örnekleri belirleyerek onları incelemek değildir. N birimlik bir anakütleden n birimden oluşan  $C_N^n$

kadar eşit sayıda birim içeren örnek seçilebilir. Mümkün bütün örneklerin seçilip incelenmesi, tam sayımdan farklı bir çözüm olmayacağından, örneklemenin amacı ile bağdaşmaz. Örneklemede amaç, anakütleden seçilecek bir örneğin incelenerek, elde edilen bilgilerle anakütle parametresinin tahmin edilmesidir.

Ancak aynı anakütleden aynı sayıda birim içeren bütün örnekler belirlenir ve her örnek için aynı değer hesaplanırsa, hesaplanan bu değerlerin birbirlerine eşit olmadığı görülür. Çünkü örneklerdeki birimler aynı birimler olmayacağından, bunlara ait hesaplanan aynı temsili değerler de birbirlerine eşit olmayacaklardır. Örneklerle ait bu değerler anakütle parametresi etrafında bir dağılım gösterirler. Yani bir bölümü anakütle parametresinden büyük, bir bölümü küçük olur. Dolayısı ile örnek değerleri anakütle parametresinden az veya çok bir takım sapmalar gösterirler.

Tesadüfi hataların tamamen ortadan kaldırılmaları mümkün değildir. Ancak bunlar iki yönlü hatalar olduklarından, örnek sayıları arttırıldıkça bu hatalar birbirlerini nötrleştirirler. Mümkün olan bütün örnekler düzenlenip, bunların belli bir istatistik değeri hesaplanıp ortalaması (beklenen değeri) alınırsa bu değer anakütlenin aynı parametresine eşit çıkacaktır.

Tesadüfi hatalar örnek değişkenliğinden kaynaklandıkları için istatistikte daima varlıkları kabul edilen ve istatistiksel olarak anlamlı olmayan hatalardır. Bu tür hataların kontrol edilmesi ve belirli olasılıklarla belirli sınırlar içinde tutulması mümkündür. Bunun için bu hataların ortalama bir ölçüsünü hesaplamak gerekir. Tesadüfi hataların ortalama ölçüsü, örnekleme dağılımının standart sapması veya kısaca standart hata diye adlandırılır.

#### 4. TESADÜFİ HATALARIN ÖLÇÜSÜ OLAN STANDART HATANIN ANLAMI

Standart hata, örnek değerlerinin (örnek istatistiği) anakütle parametresinden ne ölçüde bir sapma gösterdiklerinin ortak bir ölçüsüdür. Anakütle parametresinin bir örnek değerinden tahmin edilmesi standart hataya bağlıdır. Çünkü standart hatanın değeri anakütle parametresinin hassaslık ve güvenilirliğini etkiler.

Standart hata, tahmin edilecek anakütle parametresine, anakütlenin dağılımına, seçilecek örnekleme türüne göre farklı şekillerde hesaplanabilir. Örneğin anakütle ortalamasının, oranının, varyansının v.s. tahmininde standart hata farklı şekillerde hesaplanacaktır.

Burada standart hatanın farklı alanlarda nasıl hesaplandığını belirtmeyip, sadece anlamını açıklamaya çalışacağız.

N birimden oluşan bir anakütleden n birimlik k sayıda örnek düzenlediğimizi varsayalım. Bu durumda anakütle  $(X_1, X_2, X_3, X_4, \dots, X_N)$  birimlerinden oluşacaktır. k adet örneğin dağılımı ise, şöyle olacaktır.

$X_{1i}$	$X_{2i}$	$X_{3i} \dots \dots \dots X_{ki}$
$X_{11}$	$X_{21}$	$X_{31} \dots \dots \dots X_{k1}$
$X_{12}$	$X_{22}$	$X_{32} \dots \dots \dots X_{k2}$
$X_{13}$	$X_{23}$	$X_{33} \dots \dots \dots X_{k3}$
.	.	.
.	.	.
.	.	.
.	.	.
$X_{1n}$	$X_{2n}$	$X_{3n} \dots \dots \dots X_{kn}$

Her örneğin ortalaması  $(\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots, \bar{X}_k)$  örnekteki birimlerin toplamının birim sayısına bölümü ile,

$$\bar{X}_k = \frac{\sum X_{ki}}{n_k}$$

elde edilir.

Anakütle ortalaması  $\mu$  ise anakütle birimler toplamının anakütle birim sayısına bölümü ile,  $\mu = \frac{\sum X_i}{N}$  elde edilecektir.

Anakütle ortalaması örnek ortalaması yardımı ile tahmin edilmek istendiğinde, örneklemeye başvurulacaktır. Ancak aynı anakütleden seçilen ve aynı sayıda birimi içeren örneklerin ortalamaları örnek değişkenliği nedeni ile birbirlerine eşit çıkmayacaklarından

$(\bar{X}_1 \neq \bar{X}_2 \neq \bar{X}_3 \neq \dots \neq \bar{X}_k)$  bunlardan herhangi birini anakütle parametresine eşit kabul etmek  $\bar{X}_j = \mu$  güvenilir bir tahmin olmayacaktır. Örnek ortalamalarından bir bölümü  $\mu$ 'den küçük, bir bölümü büyük olacaktır. Ancak örneklerin sayısı artırılınca örnek ortalamalarının dağılımı normal dağılıma uygunluk gösterecektir. Böylece mümkün bütün örneklerin ortalamalarının ortalaması (beklenen değeri) anakütle ortalamasına eşit olacaktır.  $E(\bar{X}_j) = \mu$

Örnek ortalamalarının dağılımının standart sapması (standart hata); ortalamalar dağılımının ortalaması veya onun eşit olduğu anakütle parametresi yardımı ile şöyle hesaplanabilir.

$$\delta_{\bar{x}} = \sqrt{\frac{\sum (\bar{X}_i - \mu)^2}{k}} = \sqrt{\frac{\sum [\bar{X}_i - E(\bar{X})]^2}{k}}$$

Ancak standart hata, anakütle ortalaması veya örnek ortalamalarının beklenen değeri bilinmeyeceğinden, standart sapmalar yardımı ile farklı formüllerle de hesaplanabilecektir.

## 5. SONUÇ

İstatistikte hatalar iki grupta toplanır. Örneklemekten kaynaklanan ve iki yönlü olan tesadüfi hatalar, standart hata hesaplanarak, anakütle parametresinin örnek değeri yardımı ile yapılacak tahminine yansıtılabilir. Ortaya çıkış nedenleri tamamen örneklemekten kaynaklanmayan sistematik hatalar ise nedenlerinin bulunup, bu nedenlerin giderilmesi ile yok edilirler. Bu hatalar tek yönlü olduklarından anakütle parametresinin olması gerekenden büyük veya küçük tahmin edilmesine neden olurlar. Unutmamak gerekir ki istatistikte daima hata kavramının yeri vardır. Ancak bu hataları azaltmak ve belirli limitler içinde tutmakta daima mümkündür.

### YARARLANILAN KAYNAKLAR

- DANIEL, Wayne W.; TERRELL, James C., *Business Satatistics*, Boston, 1986.
- GÜRTAN, Kenan, *Istatistik ve Araştırma Metodları*, İstanbul, 1979.
- KÖKSAL, Bilge Aloba, *Istatistik Analiz Metodları*, Çağlayan Kitabevi, İstanbul, 1980.
- SHELDON, M.Ross, *Introduction To Probability and Statistics For Engineers and Scientists*, Singapore, 1987.
- YOĞURTUÇUGIL, M.Kemal, *Örneklemeye Yöntemler ve Uygulama*, İstanbul, 1976.