4

ÖRNEKLEME HATASI

Örnekten elde edilen tahmin ve populasyon parametresi arasındaki farka örnekleme hatası denir. Örneğin; her biri 1$, 2$, 3$, 4$ ve 5$ a sahip N=5 öğrenciden oluşan bir populasyon olsun. Bu 5 öğrenci aynı zamanda popülasyon çerçevesini oluşturmaktadır. Bu N=5 öğrenci popülasyonu arasından n=2 hacimli bir örnek seçilsin. Örneğin X_1=1$ ve X_2=3$ a sahip iki öğrencinin seçilmiş olduğu varsayılsın. Not : Notasyon olarak örnek değerleri için küçük, populasyon için büyük harfler kullanılmıştır.

örnek ortalaması

$$\bar{x} = \frac{1}{n}(x_1 + x_2) = \frac{1}{2}(1+3) = 2\$ \text{ dır.}$$

populasyon ortalaması

$$\bar{X} = \frac{1}{N}(X_1 + X_2 + X_3 + X_4 + X_5) = \frac{1}{5}(1+2+3+4+5) = 3\$ \text{ dır.}$$

$\bar{x} = 2\$$, $\bar{X}$ 'nın bir tahminidir. Bu durumda $\bar{x}$ ile $\bar{X}$ arasındaki farka bu örnek için örnekleme hatası denir. Burada $\bar{x}$ - $\bar{X}$ =2-3=-1 $\bar{x}$'nın örnekleme hatasıdır.

Örnekleme hatası kavramı, sapmasızlık ve beklenen değer kavramları ile ilişkilidir. Örnek tahmin edicisi örneğin $\bar{x}$, populasyon parametresi $\bar{X}$ 'nın sapmasız bir tahmin edicisi olduğunda,

E($\bar{x}$)= $\bar{X}$ olur. Bu durumda **örnekleme hatası**, örnekten elde edilen tahmin ve tahmin edicinin beklenen değeri arasındaki fark haline gelir:

$$\text{Örnekleme hatası} = \bar{x} - \bar{X}$$
$$= \bar{x} - E(\bar{x})$$
$$\bar{X} = \bar{x} - (\text{örnekleme hatası})$$

olarak da yazılabilir.

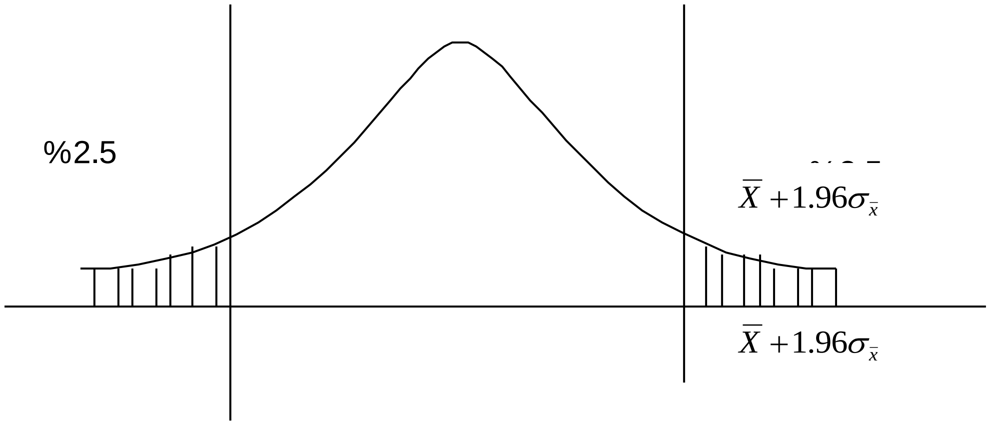
Genel olarak eğer $\hat{\theta}$, θ'nın sapmasız bir tahmin edicisi olduğunda

$$\text{Örnekleme hatası} = \hat{\theta} - \theta = \hat{\theta} - E(\hat{\theta})$$

olur. Böylece $\hat{\theta}$ 'nın örnekleme hatası tek örneğe ait sonuç ile populasyon parametresi arasındaki fark olarak yorumlanır.

DUYARLILIK

Örnek ortalamasının örnekleme dağılımı aşağıdaki gibi verilmiş olsun;



$\bar{X}-1.96\sigma_{\bar{x}}$

P($\bar{X}$ -1.96 $\sigma_{\bar{x}} < \bar{x} < \bar{X}+1.96\sigma_{\bar{x}}) = 0.95$

P(Z $_{\alpha/2}\sigma_{\bar{x}} < \bar{x}-\bar{X} < Z_{\alpha/2}\sigma_{\bar{x}}) = (1-\alpha)\%$

$\left|\bar{x}-\bar{X}\right|$: örnekleme hatası, $Z_{\alpha/2}\sigma_{\bar{x}}$: duyarlılık olarak ifade edilebilir

P(-1.96 $\sigma_{\bar{x}} < \bar{x}-\bar{X} < 1.96\sigma_{\bar{x}}) = 0.95$

P(-1.96< $\dfrac{\bar{x}-\bar{X}}{\sigma_{\bar{x}}}$ <1.96)=0.95

$P(Z_{\alpha/2} < Z < Z_{\alpha/2}) = (1-\alpha)\%$

Yukarıdaki ifade $\left|\bar{x}-\bar{X}\right|$’nın 1.96 $\sigma_{\bar{x}}$ ’dan daha az olma olasılığının 0.95 olduğunu ifade eder. Diğer bir deyişle $\bar{x}$ ’nın A ve B gibi iki çizgi arasına düşme olasılığının 0.95 olduğunu ifade eder. $\bar{x}$ A ve B arasında olduğu sürece fark $\left|\bar{x}-\bar{X}\right|$ < 1.96 $\sigma_{\bar{x}}$ ’dan az olacaktır.

P($\left|\bar{x}-\bar{X}\right|$<1.96 $\sigma_{\bar{x}}$)=0.95 şeklinde düzenlenebilir. Bu durumda $\left|\bar{x}-\bar{X}\right|$’nın alabileceği en büyük değer $\left|\bar{x}-\bar{X}\right|$=1.96 $\sigma_{\bar{x}}$ olur. Burada iki noktaya dikkat edilmelidir.

1. $\left|\bar{x}-\bar{X}\right|$ örnekleme hatasıdır.

2. 1.96 $\sigma_{\bar{x}}$ güven aralığının yarısıdır.

Burada örnekleme hatası $\left|\bar{x}-\bar{X}\right|$ farklı bir şekilde yorumlanmalıdır. Örnekleme hatası sadece verilen bir tek örnek için örnekleme hatasıdır. Duyarlılık ise örnekleme sebebiyle populasyon parametresi ve tahmin edici arasındaki değişmeyi gösterir . Bu da verilen bir güven katsayısıyla güven aralığının yarı uzunluğuna eşittir.

$\bar{x}-\bar{X}$, belli bir örnek hacmi (n) ve güvenilirlik için **tekrarlanan örneklemede** tahmin edici ve populasyon parametresi arasındaki maksimum değişimi ifade eder. $\bar{x}-\bar{X}$ bu anlamda dikkate alındığında, örnekleme hatası olarak değil tahmin edicinin duyarlılığı olarak dikkate alınır. Dolaysıyla, $\bar{x}-\bar{X}$, verilen herhangi bir örnek için değil, tesadüfi örnekleme yapıldığında en büyük değişmeyi gösterir. Örneğin, maksimum değişmenin 0.30$'dan az olmasına gereksinim duyulduğunda, duyarlılık 0.30 olacaktır. Bu da güven aralığının yarı uzunluğunun 0.30 $'dan az olacağını ifade eder.

NOT: Örnekleme hatası kavramı verilen bir örnek içindir. Duyarlılık ise güven aralığı ile ilgilidir ve ancak tekrarlanan örnekleme ile ilgili olarak tanımlanır.

GÜVENİLİRLİK

Duyarlılık=güvenilirlik standart hata*

$$d=z\,\sigma_{\bar{x}}$$

Duyarlılığa ilişkin %95 güvenilirlik sözkonusu olduğu zaman, bunun anlamı, duyarlılığı %95 güven katsayılı güven aralığının en fazla yarı uzunluğu kadar olduğudur. Örnek ortalaması ve %95 güvenilirlik için,

$\left|\bar{x}-\bar{X}\right|$ =1/2(güven aralığı) =1.96 $\sigma_{\bar{x}}$ olarak ifade edilebilir. Burada $\left|\bar{x}-\bar{X}\right|$ terimi duyarlılığı 1.96 da %95 güvenilirliği temsil eder. 1.96 ya da güvenilirlik katsayısı denir. Bu ilişki aşağıdaki şekilde gösterilebilir.

Duyarlılık=güvenilirlik standart hata*

$$d=z\,\sigma_{\bar{x}}$$

ÖRNEK HACMİ (n) VE DUYARLILIK (d)

Örnek hacmi arttıkça duyarlılık artmakta yani d küçülmektedir. Örnek ortalaması $\bar{x}$ kullanılarak bu ilişki açıklandığında:

d=z $\sigma_{\bar{x}}$

d =z $\dfrac{\sigma}{\sqrt{n}}$.

Burada d =$\left|\bar{x} - \bar{X}\right|$'dır. n büyüdükçe d küçülür ve tahmin edicinin duyarlılığı artar. Bunun genel anlamı, örnek hacmi büyüdükçe, örnek tahmin edicisi tahmin etmekte olduğu populasyon parametresine yaklaşır. Bu eşitlik ayrıca verilen bir duyarlılık için örnek hacmini elde etmede de kullanılabilir:

$$d = z\sigma_{\bar{x}} = z\frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = z\frac{\sigma}{d}$$

$n = \dfrac{z^2\sigma^2}{d^2}$ bu n değeri belirli bir duyarlılık için gerekli örnek hacmidir.

ÖRNEK: %95 güvenilirlikle $\bar{x}$ 'nın duyarlılığının 4 br'lik sınırlar içinde kalması arzu edildiğine ve populasyon standart sapmasının σ =18 br olduğu bilindiğine göre gerekli örnek hacmini elde ediniz.

d=$\left|\bar{x} - \bar{X}\right| \leq$ 4br   1-α=0,95   α=0,05   α/2=0,025

4=z $\dfrac{\sigma}{\sqrt{n}}$ =2 $\dfrac{18}{\sqrt{n}}$ 0,5-0,025=0,4750   Z(0,5-α/2)=1,96≈2

n=81 elde edilir.

ÖRNEK HACMİ (n) VE GÖRELİ TERİMLERDE DUYARLILIK (d')

Birçok durumda tahmin edicinin duyarlılığı, mutlak terimlerden çok göreli terimlerle ifade edilir. Örneğin belirli bir okuldaki öğrencilerin ortalama ağırlığını tahmin etmede duyarlılık $\overline{X}'$ nın ±7 br'lik sınırlar içinde olması yerine $\overline{X} \pm$%5'lik sınırlar içinde olması olarak ifade edilebilir.

Değişim katsayısı kullanılarak, duyarlılık, güvenilirlik ve standart hata arasındaki temel ilişki göreli terimlerle ifade edilebilir ve buradan da örnek hacmi ve duyarlılık arasındaki ilişki göreli terimler cinsinden ifade edilebilir.

Temel ilişki d=z$\sigma_{\bar{x}}$ şeklindedir. Dolayısıyla duyarlılık (d)

d= $\bar{x} - \overline{X}$ göreli terimlerle ($d^{'}$) ifade edilmek istenirse

$$d^{'} = \frac{\bar{x} - \overline{X}}{\overline{X}} = \frac{d}{\overline{X}}$$ şeklinde yazılabilir

$$d^{'} = \frac{d}{\overline{X}} = Z\frac{\sigma_{\bar{x}}}{\overline{X}} = Z\frac{\sigma_x / \sqrt{n}}{\overline{X}}$$

Açıklamalar : $C = \frac{\sigma_X}{\overline{X}}$

$$C(\bar{x}) = \frac{\sqrt{V(\bar{x})}}{E(\bar{x})} = \frac{\sigma_{\bar{x}}}{\overline{X}} = \frac{\sigma / \sqrt{n}}{\overline{X}} = \frac{\sigma}{\overline{X}\sqrt{n}} = \frac{C}{\sqrt{n}}$$

$$C(\hat{X}) = \frac{\sqrt{V(\hat{X})}}{E(\hat{X})}$$

$$V(\hat{X}) = V(N\bar{x}) = N^2 V(\bar{x}) = N^2 \frac{\sigma^2}{n}$$

$$C(\hat{X}) = \frac{N\sigma / \sqrt{n}}{X} = \frac{\sigma / \sqrt{n}}{\overline{X}} = \frac{C}{\sqrt{n}}$$ old için

$$C(\bar{x}) = C(\hat{X}) = \frac{C}{\sqrt{n}}$$ olduğuna göre,

$$d^{'} = Z\frac{C}{\sqrt{n}}$$

$$d^{'} = ZC(\bar{x})$$

$$n = \frac{(ZC)^2}{(d^{'})^2}$$ belirli bir göreli duyarlılık için gerekli örnek hacmi elde edilir,

bu yukarıdaki açıklamalara dayanılarak

$$n = \left(\frac{C}{C(\bar{x})} \right)^2$$ şeklinde de ifade edilebilir

ÖRNEK:Büyük bir grup öğrencinin ortalama ağırlığı tahmin edilmek istenmektedir. Tahmin edicinin göreli duyarlılığının, ortalamanın %5'i kadar olan sınırlar içinde olması istenmektedir. Güvenilirlik z=3 (hemen hemen kesin) ve C=0.1 olduğuna göre hangi büyüklükte bir örnek seçilmelidir?

$$n = \frac{(zC)^2}{(d')^2} = \frac{3^2 C^2}{0.05^2} = \frac{(3*0.1)^2}{0.05^2} = 36$$

HATA KARELER ORTALAMASI

HKO**,** Bir tahmin edici ile parametre değeri arasındaki farkın karesinin beklenen değeri *hata kareler ortalaması* olarak adlandırılır.

θ parametresinin θ tahmin edicisi için ;

H.K.O.= $E\left(\hat{\theta} - \theta\right)^2$

İle gösterilebilir. $E\left(\hat{\theta}\right)$bir kez eklenip çıkarılarak,

H.K.O= $E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right) + \left(E\left(\hat{\theta}\right) - \theta\right)\right)\right]^2$

H.K.O=E $\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2 + \left(E\left(\hat{\theta}\right) - \theta\right)^2$

=Var($\hat{\theta}$)+(sapma miktarı)^2

Sapmasız tahminciler için hata kareler ortalaması ile varyans aynı anlama gelir. Parametre değeri bilinmediğinden hata kareler ortalaması hesaplanamaz.

Tahmin edicilerin istenen özelliklerinden biri hata kareler ortalamasının küçük olmasıdır. Bazen bu terim *doğruluk* olarak adlandırılır. Sapmasız olan tahmin edicilerin doğruluk ölçütü varyastır. Bir tahmin edici ne kadar küçük varyanslı ise o kadar duyarlıdır denir.

Örnek: I, II, III olarak ifade edilen 3 farklı zarın 3 er kez atılışından elde edilen sonuçlar aşağıdaki gibi olsun. Bu farklı zar atılışlarından istatistiklei elde etmeye çalışırsak,
(Tek zar atılışında Populasyon Ortalaması:3.5 dur.)

	I	II	III
	5	5	6
	2	3	5
	1	5	1
ort	2.67	4.33	4
varyans	4,33	1,33	7
HKO	5,01	2,01	7,25
$sapma^2$	0,68................	0,68.................	0,25

$$HKO = \text{var}\, yans + sapma^2$$

	IV	V	VI
	2	5	1
	2	3	3
	5	2	4
ort.	3	3.33	2.67
varyans			
HKO			
$sapma^2$			

	VII	VIII	IX
	1	3	5
	6	3	1
	3	4	3
ort.	3.33	3.33	3
varyans			
HKO			
$sapma^2$			

ÖZET: Yukarıda 9 farklı örnekten elde edilen sonuçlar bulunmaktadır. Toplamda 6'nın 3'lü kombinasyonu kadar yani 20 farklı örnek elde edilebilir ve tüm bu sonuçlar kullanılarak tüm istatistiklerin beklenen değerleri elde edilebilir.