3

TAHMİNLEYİCİLERİN ÖZELLİKLERİ

İyi bir tahmin edici dağılımı tahmin edilecek populasyon parametresine yakın civarda yoğunlaşan bir tahmin edicidir.

İyi bir tahmin edici için gerekli kriterler:

1. Sapmasızlık
2. Tutarlılık
3. Etkinlik – minimum varyans

Sapmasızlık: Tahmin edici olarak kullanılan istatistiğin beklenen değeri, tahmin edilecek populasyon parametresine eşit olduğunda, tahmin edici **sapmasızdır** denir. Tahmin edicinin beklenen değeri, örnek dağılımının ortalamasıdır. Tahmin edicinin örnek uzayında her bir örnekte aldığı değerler ortalamasıdır.

$\hat{\theta}$ tahmin edicisinin beklenen değeri, θ parametresine eşit ise;

$$E(\hat{\theta}) = \theta$$

$\hat{\theta}$ tahmin edicisine sapmasızdır denir.

Tahmin edicinin beklenen değeri ile parametre arasındaki farka *sapma miktarı* denir.

S.M.= $E(\hat{\theta}) - \theta$

Örnekleme dağılımı bilinmeden sapma miktarı elde edilemez. Sapma örnekleme dışı hatalardan da oluşabilir. Örneğin, yetersiz bir örnekleme çerçevesinin kullanılması,örnek seçim sürecinin yanlış uygulanması, verilerin toplama ve değerlendirme işlemi için hatalı olması v.b nedenler sapmalı tahminlere yol açabilir.

Tutarlılık: Örnek genişliği verilen bir sayıdan daha büyük alındığında tahmin ile parametre arasındaki farkın düşünülebilen en küçük pozitif bir sayıdan daha küçük kalma olasılığı 1 ise, o tahmin tutarlıdır denir.

$\hat{\theta}$ tahmini için tutarlılık:

$$\lim_{n\to\infty} P\left(\left(\hat{\theta} - \theta\right) \prec E\right) = 1$$

Sonlu populasyon birimi içeren populasyondan yapılacak tahminlerde, eğer örnek genişliği populasyon genişliğine eşit olduğunda tahmin, parametre değerine eşit olduğunda o tahmin tutarlıdır denir.

Bir tahmin tutarlı ise geniş örnekler için, bir beklenen değere sahiptir. Tutarlı tahminler arasında tüm örnekler için beklenen değeri parametreye eşit olan seçilecektir.

Etkinlik: Sapmasız ve tutarlı tahmin ediciler içinde minimum varyansa sahip olan tahmin edici etkindir. Normal dağılıma sahip bir populasyonun olduğunu varsayalım ve populasyon ortalamasını tahmin etmek isteyelim. Bunun için sapmasız ve tutarlı bir tahmin edici olan örnek ortalaması $\bar{x}$ veya yine sapmasız ve tutarlı bir tahmin edici olan örnek medyanı $\tilde{x}$ da kullanılabilinir. Bu iki tahmin ediciden hangisinin tercih edileceği örnek ortalaması ya da örnek medyanından hangisinin örnekleme dağılımı populasyon parametresi etrafında daha yakın biçimde yoğunlaşıyorsa buna göre karar verilir. Bu her iki tahmin edicinin varyansları karşılaştırılarak belirlenebilir ve daha küçük varyansa sahip olan tercih edilir. Tutarlı tahmin ediciler için varyans küçüldükçe örnekleme dağılımı populasyon parametresi etrafında daha fazla yoğunlaşır. Büyük hacimli örnekler için $\bar{x}$ ve $\tilde{x}$ 'nın varyansları aşağıdaki şekildedir:

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

$$V(\tilde{x}) = \frac{\pi\sigma^2}{2n}$$

Dolayısıyla, belirli bir örnek hacmi için,

$$\frac{V(\bar{x})}{V(\tilde{x})} = \frac{2}{\pi} = 0.64$$

$V(\bar{x}) < V(\tilde{x})$ olduğundan bir tahmin edici olarak $\bar{x}$, $\tilde{x}$ 'ya tercih edilir. Belli bir örnek hacmi için, $\bar{x}$ 'nin örnekleme dağılımı $\tilde{x}$ 'nın örnekleme dağılımına göre $\bar{X}$ etrafında daha yoğun olduğundan, $\bar{x}$ 'nın $\tilde{x}$ 'dan daha etkin olduğu söylenebilir.

$V(\bar{x}) = V(\tilde{x})$*%64 dür. Her ikisi de aynı örnek hacmine sahipken $\bar{x}$ 'nın varyansı $\tilde{x}$ 'nın varyansının %64'ü kadardır.

Örnek hacimleri bakımından durum ele alınırsa örnek hacmi 100 olan örneklerden elde edilen medyan varyansı, örnek hacmi 64 olan örnek ortalaması varyansı ile hemen hemen aynıdır.

Özetle, $\hat{\theta}_1$ ve $\hat{\theta}_2$ gibi iki tahmin edici

$V(\hat{\theta}_1) < V(\hat{\theta}_2)$ ise $\hat{\theta}_1$'in $\hat{\theta}_2$'ye oransal etkinliği

$E_f = \dfrac{V(\hat{\theta}_1)}{V(\hat{\theta}_2)}$ olarak ifade edilebilir. Daha küçük varyanslı tahmin edicinin varyansı payda bulunduğundan $0 \leq E_f \leq 1$ olacaktır.

Varyans belli bir alt sınırdan küçük olamaz ve varyansı bu alt sınıra eşit bir tahmin edici bulunabilirse bu varyans en küçük varyans olur. Cramer Rao eşitsizliği ile varyans için böyle bir alt sınır değer elde edilebilir. Cramer Rao eşitsizliği $\hat{\theta}$'nin varyansının σ^2 / n den küçük olamayacağını söyler. $V(\hat{\theta}) \geq \sigma^2 / n$ dir.

ARALIK TAHMİNİ – GÜVEN ARALIĞI

Nokta tahmini için, en çok olabilirlik metodu, en küçük kareler metodu ve momentler metodu gibi metodlar bulunmaktadır. Aralık tahmini için ise genellikle güven aralığı kavramına başvurulur.

Örnekleme çalışmalarında kullanılan ana tahmin metodu aralık tahminidir. $\overline{X}$ (bilinmeyen) ortalamalı ve σ standart sapmalaı büyük bir populasyona sahip olunduğu varsayılsın. $\overline{X}$ tahmin edilmek istenmekte ve bu amaç için n hacimli bir tesadüfi örnek seçilmektedir. Merkezi limit teoreminden

Z=$\dfrac{\bar{x} - \overline{X}}{\sigma_{\bar{x}}}$ ifadesi asimtotik olarak 0 ortalama ve birim varyansa sahiptir. Normal dağılım tablosundan Z=1.96 %95'lik olasılığa karşılık gelir. Bu durumda, z bir tesadüfi değişken olduğu için geçerli bir olasılığı gösteren

P(-1.96< $\dfrac{\bar{x} - \overline{X}}{\sigma_{\bar{x}}}$ <1.96==0.95 ifadesi yazılabilir. Bu denklem

P($\bar{x} - 1.96\dfrac{\sigma}{\sqrt{n}} < \overline{X} < \bar{x} + 1.96\dfrac{\sigma}{\sqrt{n}}$)=0.95 olarak yeniden yazılabilir. Burada

$\bar{x} - 1.96\dfrac{\sigma}{\sqrt{n}}$ =a

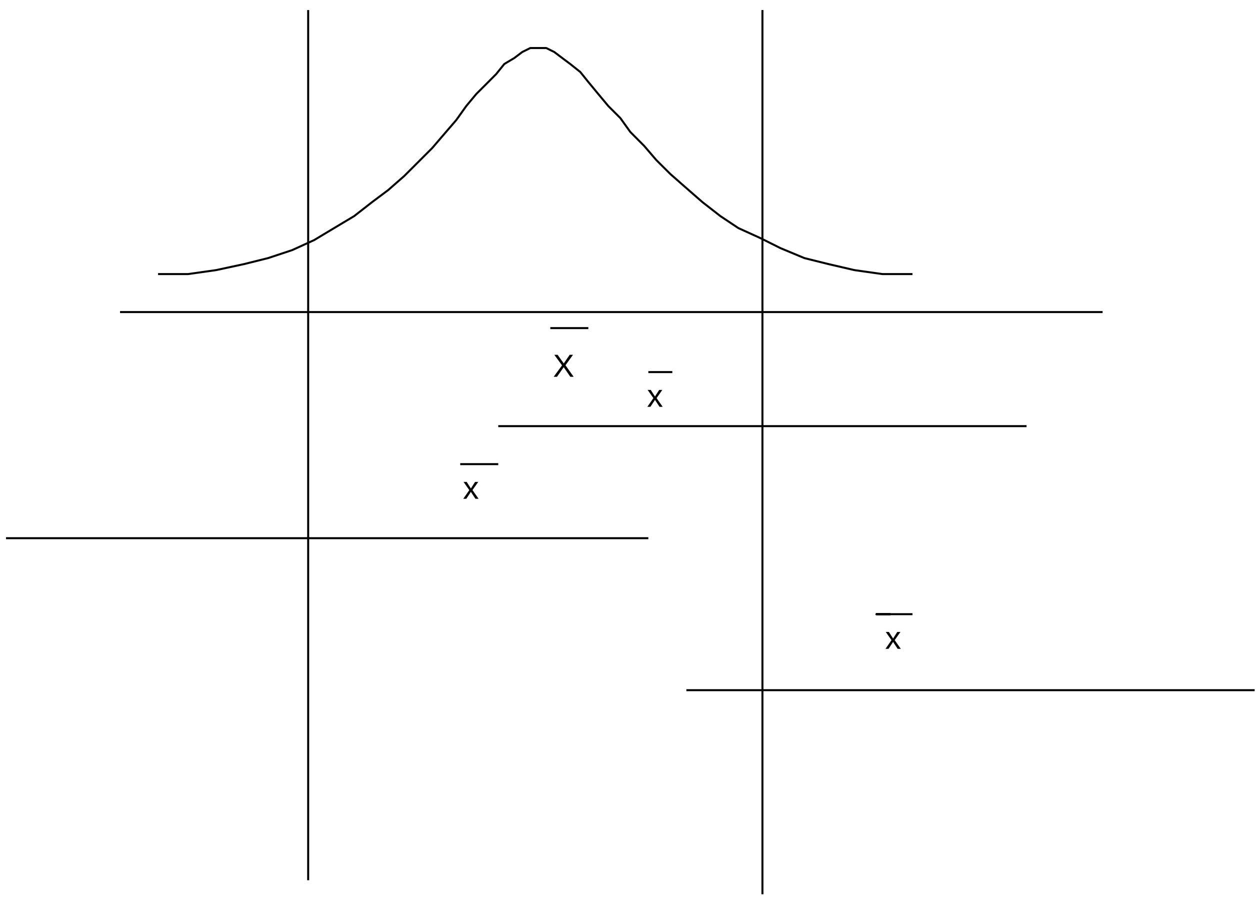
$\bar{x}+1.96\frac{\sigma}{\sqrt{n}}$ =b olarak tanımlandığında

P(a< $\bar{X}$ <b)=0.95 dır ve bu şu şekilde yorumlanabilinir:

Merkezi limit teoreminden, $\bar{x}$ 'nın asimtotik olarak, $\bar{X}$ ortalamalı ve σ^2/n varyanslı normal dağılıma sahip olunduğu bilinmektedir. Bu, diyagram olarak aşağıdaki şekil 3.1'de gösterilmektedir. Tesadüfi değişken $\bar{x}$ değişik değerler almaktadır. $\bar{x}$'nın aldığı değerler $\bar{x}_1, \bar{x}_2, \ldots$.... olarak tanımlansın. Örneğin $\bar{x}_1$ şekil 3.1 de verilen değeri alsın. Bu durumda aralık

$\bar{x}_1 - 1.96\frac{\sigma}{\sqrt{n}}$

$\bar{x}_1 + 1.96\frac{\sigma}{\sqrt{n}}$ olacaktır. Grafikten da görüldüğü gibi bu aralık $\bar{X}$'yı içerecektir.



Benzer şekilde diğer bir değer $\bar{x}_2$ için, aynı zamanda $\bar{X}$ yı içeren aralık

$\bar{x}_2 - 1.96\frac{\sigma}{\sqrt{n}}$ $\bar{x}_2 + 1.96\frac{\sigma}{\sqrt{n}}$

olacaktır. Fakat grafikte gösterildiği gibi $\bar{x}_3$

$\bar{x}_3 - 1.96\frac{\sigma}{\sqrt{n}}$ $\bar{x}_3 + 1.96\frac{\sigma}{\sqrt{n}}$ aralığını verir ve bu aralık $\bar{X}$ yı içermez. Grafiksel olarak görüldüğü gibi $\bar{x}_3$, $\bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$ değerlerinin dışına düşer. $\bar{X}$ 'nın, populasyonun gerçek parametresi olması koşulu ile $\bar{x}$ 'nın $\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$ aralığında olma olasılığı 0.95'tir. şekil 3.1'den görüldüğü gibi aralık, $\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}$, $\bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$ şeklinde oluşturulduğunda böyle 100 aralıktan 95 tanesinin $\bar{X}$ 'yı içermesi beklenir. Bu durum, $\bar{x}$ $\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}$ ile $\bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$ arasında kaldığı sürece geçerlidir. Bu sembolik olarak

$$P(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \bar{x} < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}})=0.95$$

Şeklinde gösterilebilir. Bunula birlikte bu ifade,

$P(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \bar{X} < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}})=0.95$ olarak yeniden yazılabilir. $\bar{x}$ 'ye sabit bir değer verilmediği ve bir tesadüfi değişken olarak dikkate alındığı sürece, sözkonusu ifade uygun bir olasılık ifadesidir.

ALIŞTIRMA: N=7 hacimli aşağıdaki gibi bir populasyon olduğunu varsayalım.

$X_1=1$, $X_2=2$, $X_3=3$,$X_4=4$, $X_5=5$ $X_6=6$, $X_7=7$

Bu populasyondan n=2 hacimli örnekler seçelim. Bu durumda $\binom{7}{2}$=21 tane mümkünörnek elde edilebilir ve dolayısıyla 21 örnek ortalaması vardır. Bunların listesi tablo 3.1 'de verilmektedir.

Tablo 3.1

	Örnek	$\bar{x}$	$\bar{x} - 1.64\sigma_{\bar{x}} < \bar{X} < \bar{x} + 1. + 64\sigma_{\bar{x}}$	
1	1.2	1.5	-0.6< $\bar{X}$ <3.6	1.5 ∓ 1.64($\sqrt{1.66}$)
2	1.3	2.0	-0.1< $\bar{X}$ <4.1	2.0 ∓ 1.64($\sqrt{1.66}$)
3	1.4	2.5	0.4< $\bar{X}$ <4.6	2.5 ∓ 1.64(1,29)
4	1.5	3.0	-0.1< $\bar{X}$ <4.1	
5	1.6	3.5	1.4< $\bar{X}$ <5.6	
6	1.7	4.0	1.9< $\bar{X}$ <6.1	
7	2.3	2.5	0.4< $\bar{X}$ <4.6	
8	2.4	3.0	0.9< $\bar{X}$ <5.1	
9	2.5	3.5	1.4< $\bar{X}$ <5.6	
10	2.6	4.0	1.6< $\bar{X}$ <6.1	
11	2.7	4.5	2.4< $\bar{X}$ <6.6	
12	3.4	3.5	1.4< $\bar{X}$ <5.1	
13	3.5	4.0	1.9< $\bar{X}$ <6.1	
14	3.6	4.5	2.4< $\bar{X}$ <6.6	
15	3.7	5.0	2.9< $\bar{X}$ <7.1	
16	4.5	4.5	2.4< $\bar{X}$ <6.6	
17	4.6	5.0	2.9< $\bar{X}$ <7.1	
18	4.7	5.5	3.4< $\bar{X}$ <7.6	
19	5.6	5.5	3.4< $\bar{X}$ <7.6	
20	5.7	6.0	3.9< $\bar{X}$ <8.1	
21	6.7	6.5	4.4< $\bar{X}$ <8.6	

$\bar{X}, \sigma$ ve $\sigma_{\bar{x}}$ hesaplanırsa

$\bar{X} = 4$, $\sigma^2 = 28/7 = 4$, $\sigma = 2$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\frac{N-n}{N-1} = \frac{4}{2}\frac{7-2}{7-1} = \frac{5}{3} = 1.666$$

$$\sigma_{\bar{x}} = \sqrt{1.666} = 1.29$$

z $\sigma_{\bar{x}}$ =1.64*1.29=2.1

z=1.64 olarak alındığı için güven katsayısı %90'dır. Bu, 100 aralıktan 90 tanesinin gerçek populasyon ortalaması $\overline{X}$'yı içermeyeceğinin beklendiğini ifade eder. Burada da ilk ve son sıradaki güven aralıkları populasyon ortalamasını kapsamaz.