

ÖRNEKLEME YÖNTEMLERİ

Yrd. Doç. Dr. Emre ATILGAN

Örnekleme Nedir?

- İncelenen özellikleri yönünden genellemeler yapılması istenen bir ana kütlede, belli yöntemlerle örneklemin seçilmesi ve bu örneklemin incelenmesi sonucunda hesaplanan istatistiklerin genelleme amacıyla kullanılması işlemlerinden oluşan sürece **örnekleme** adı verilir.

Neden Örnekleme?

- ✓ Örneklemede çalışmak kitlede çalışmaktan daha kolaydır.
- ✓ Kitle üzerinde çalışmak çok daha masraflı olabilir.
- ✓ Çoğu durumda tüm kitleye ulaşmak mümkün değildir.
- ✓ Örneklem sonuçları daha doğru olabilir. Çünkü daha az sayıda kişi ile (örnek ile) çalışılacağından, araştırma daha özenli yürütülebilir.
- ✓ Eğer örneklem olasılıksal yöntemlerle seçiliyorsa, yapılan örnekleme hatasının kestirimini de bulmak mümkündür.

Neden Örnekleme?

- ✓ Örneklemede temel amaç seçilen örneklemin kitleyi temsil edebilecek özellikte olmasıdır.
- ✓ Örneklemin kitleyi tümüyle temsil etmesi beklenir. Ancak bu gerekli değildir.
- ✓ Örneklemin, kitleyi ulaşmak istediğimiz bilgide farklılık yaratabilecek etkenler yönünden temsil edebilecek özellikte olması yeterlidir.

Örnekleme Süreci



Örneklemede Yapılan Hatalar

1. Sistemik Hata: Hesaplanan örneklem istatistikleri, ana kütle parametresinden ya hep daha küçük, ya da hep daha büyük olarak belirleniyorsa, yani hesaplanan bu değerler ana kütle parametresinden her zaman tek bir yönde farklılık gösteriyor ise o araştırmada sistemik hatanın var olduğu söylenebilir. Sistemik hatanın nedenleri:

- ✓ Ana kütle nin tanımının doğru yapılmaması,
- ✓ Örnekleme birimi ile gözlem biriminin farklı olması,
- ✓ Örnekleme çerçevesinin eksik oluşturulmasından ya da aynı örnekleme biriminin örnekleme çerçevesinde birden fazla yer alması,
- ✓ Örnekleme birimlerinden veri elde edilmesi esnasında, yanlış anlamaya yol açan soruların bulunması,
- ✓ Örnekleme seçilen birimlerin bir kısmından bilgi sağlanamaması,
- ✓ Ana kütle nin yapısına uygun olmayan bir örnekleme yöntemi belirlenmesi

Örneklemede Yapılan Hatalar

2. Rassal Hata (Örnekleme Hatası): Örneklem istatistikleri, N sayıda birimden oluşan ana kütlede seçilen n hacimlik örneklemden elde edilen veriler kullanılarak hesaplandığı için, ana kütle parametresi ile örneklem istatistiği arasında belli miktarda bir fark ortaya çıkar. Bu farka örneklemede rassal hata ya da örnekleme hatası adı verilir.

Ör: örneklem standart sapması

Örnekleme Yöntemleri

Olasılıklı Örnekleme

- Basit Rastsal Örnekleme
- Sistemantik Örnekleme
- Tabakalı Örnekleme
- Küme Örnekleme

Olasılıksız Örnekleme

- Kolayda Örnekleme
- Kota Örnekleme
- Kartopu Örnekleme
- Karar Örnekleme

Olasılıklı Örnekleme Yöntemleri

Ana kütlede yer alan tüm örneklem birimlerinin belirli olasılıklarla örnekleme seçildiği yöntemlere olasılıklı ya da rassal örnekleme yöntemleri adı verilir. Bu yöntemlerde her bir örnekleme biriminin seçilme olasılığı belli bir değere eşittir.

- Bu yöntemlerde her bir örnekleme biriminin seçilme olasılığı belli bir değere eşittir.
- Maliyet, zaman, işgücü vb. gibi fiziki kaynaklar yeterli düzeyde ise, uygulamalarda her zaman olasılıklı örnekleme yöntemleri tercih edilmelidir. Çünkü olasılıklı örnekleme yöntemleri ile, seçilen örneklemin ana kütleği ne kadar temsil edebildiği tespit edilebilir. Seçilen örneklemin ana kütleği iyi temsil edebilmesi, dolaylı olarak rassal seçim yapma yoluyla sağlanır.

Basit Rassel Örneklem

- Ana kütlede yer alan birimlerin örneklem seçilme olasılıkları eşittir.
- BRÖ yönteminde seçilebilecek olası örneklem sayısı ${}_N C_n$ ile verilir.

SistematiK Örneklem

- Bu yöntemde öncelikle, ana kütlede yer alan her örneklem birimi için 1'den N'ye kadar birer sıra numarası verilir. N hacmindeki bu ana kütlede n birimlik bir örneklem seçebilmek için ilk k birimden rassel olarak bir birim seçilir. k sayısı; incelenecek ana kütledeki birim sayısının belirlenen örneklem hacmine bölünmesi sonucu elde edilir.
- $k = \frac{\text{Ana kütle hacmi}}{\text{Örneklem hacmi}} = \frac{N}{n}$

Ör: N=70, n=10 ise, k=7 olur. Örneklem alınacak ilk örnek 3. örnek ise örneklem girecek birim numaraları aşağıdaki gibi belirlenir:

3, 10, 17, 24, 31, 38, 45, 52, 59, 66

Tabakalı Örnekleme

- Bu yöntem, ana kütle için araştırmanın konusunu oluşturulan bir ya da daha fazla değişken bakımından çeşitli alt gruplara bölünmesi ve her bir alt gruptan ayrı ayrı örneklem çekilmesi şeklinde uygulanır.

Ör: bir markette satışa sunulan tüm gıda ürünleri incelenecek ana kütle oluşturulmuş ise, bu ana kütle, et ürünleri, süt ürünleri, bisküvi ve şekerleme, içecekler, hazır gıdalar, kuruyemişler vb. gibi tabakalara ayrılarak her tabakadan örneklem seçme yoluna gidilebilir.

- Ana kütlede oluşturulan bütün tabakalara BRÖ yönteminin uygulandığı tabakalı örnekleme türüne **Tabakalı Rassel Örnekleme (TRÖ)** adı verilir.

Küme Örneklemesi

- Örneklem birimlerinin birden çok ana kütle biriminden oluştuğu olasılıklı örnekleme Küme Örneklemesi (KÖ), her bir örneklem birimine de bir küme adı verilir.
 - Bu yöntemde ana kütle alt gruplara bölünür ve bu gruplar üzerinden örnekleme geçilerek örneklem giren grupların tamamı alınır. Çekilen kümelerin içerdiği ana kütle birimleri örnekleme oluşturacağı için, kümelerin en uygun şekilde elde edilmesi gerekir.
 - KÖ yöntemi BRÖ yöntemine kıyasla, araştırmanın maliyetini düşürmek ve zaman kısıtının etkisini azaltmak amacıyla uygulanır.
 - Küme verileri kendi içerisinde heterojendir (TÖ yönteminin tam tersi)
- Ör:** İlde ölçüğünde 5000 BRÖ yerine 5 mahalleden 1000 örnek seçmek.

Olasılıklı Olmayan Örneklem Yöntemleri

- Ana kütledeki bazı birimlerin seçilen örnekte yer alma şansının bulunmadığı, ya da seçilme olasılığının tam olarak belirlenemediği örnekleme yöntemlerine olasılıklı olmayan örnekleme yöntemleri adı verilir.
- Olasılıklı olmayan örnekleme yöntemlerine, genellikle olasılıklı örnekleme yöntemleri kullanmanın mümkün olmadığı durumlarda başvurulur (örneğin anakütle tam olarak bilinmediğinde ya da sayısı tam olarak tahmin edilemediğinde).
- Bu yöntemlerde, birimlerin seçimi rassal olmadığı için, örnekleme hataları belirlenemez. Böyle durumlarda seçilen örneklem ana kütleyle ilişkin olarak sınırlı bilgi içerir. Örneklem ile ana kütle arasındaki ilişkiye ait bilgiler sınırlı olduğundan, örneklemden elde edilen tahminlerin ana kütle için genellenmesi zorlaşır.

Kolayda Örneklem

- Kolayda örneklem temsil edebilirlikten çok, uygunluk ilkesine dayalı bir örneklem seçimi yöntemidir.

Ör: Sokaktan geçen insanlara anket uygulama.

Kota Örneklemesi

- Kota örneklemesi, tabakalı örneklemenin olasılıksal olmayan biçimidir. Tabakalı örneklemede olduğu gibi, kota örneklemesi de ana kütle içerisinde yaş, ırk, sosyal sınıf vb. gibi önemli farklılıklar bulunduğu bilindiğinde uygulanan bir yöntemdir.

Tabakalar Belirlenir → Tabaka Kotası Belirlenir → Kota sayısınca veri toplanır.

Kartopu Örneklemesi

- ✓ Bazı durumlarda, ana kütlelerin kimlerden oluşacağını bilmek mümkün olmaz.
- ✓ Kartopu örneklemede, ilk önce tanımlanan ana kütlede yer alan bir birim belirlenir. Bu birim seçilecek örnekleme ilk birimidir. Daha sonra, bu birime aynı ana kütlede yer alabilecek başka bir birimin olup olmadığı sorulur ve var ise bu birime gidilir. Bu da örnekleme yer alan ikinci birim olur.
- ✓ Süreç örneklemin tamamı elde edilinceye kadar devam eder.
- ✓ Çeşitli sebeplerden kimliklerini gizli tutmak isteyen birimlerden oluşan ana kütleler üzerinde inceleme yapılacağı zaman kullanılır.

Karar Örneklemesi

- ✓ Karar örneklemesi yönteminde, arařtırmacının kendi kişisel kararlarına göre örneklem birimleri belirlenir. Bu yöntemi uygulayan arařtırmacının ilgilenilen ana kütleyi çok iyi tanması gereklidir.
- ✓ konusunda uzman bir arařtırmacının kişisel kararları doğrultusunda ana kütleyi iyi temsil edebileceğini düşündüğü birimleri seçerek örneklem çerçevesi oluřturmasına karar örneklemesi adı verilir.
- ✓ Arařtırmanın konusu, ana kütlenin diğerlerine göre farklı özellikler gösteren birimlerine ilişkin bilgi toplama olduğunda da karar örneklemesi kullanılır.

Ör: Arařtırma zengin kişilerin tükettiğı gıda türlerine ilişkin özel bir arařtırma ise, örneklem alınacak birimler yalnızca arařtırmacının belirleyeceğı miktara göre zengin olan kişiler arasından seçilir.

ÖRNEKLEM BÜYÜKLÜĞÜNÜN (HACMİNİN) BELİRLENMESİ

- Uygun örneklem büyüklüğü kullanılacak istatistiksel yöntem temel alınarak hesaplanmalıdır.
- Temel ilke olarak örneklem büyüklüğü araştırmanın birincil hipotezi (ve dolayısıyla amacı) için hesaplanır. İdeal olarak ise, her önemli değişken için ayrı ayrı bir örneklem büyüklüğü hesaplanır ve en büyük tahmini yapacak olan seçilir.
- Uygulanabilecek her bir istatistiksel yöntem için farklı örneklem genişliği formülleri bulunmaktadır.

1- Kitle ortalamasının kestiriminde kullanılacak örneklem büyüklüğü formülü:

- ***Kitle büyüklüğü (N) bilinmiyorsa;***

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{d^2}$$

- ***Kitle büyüklüğü (N) biliniyorsa;***

$$n = \frac{N z_{\alpha/2}^2 \sigma^2}{d^2 (N - 1) + z_{\alpha/2}^2 \sigma^2}$$

$z_{\alpha/2}$: Belirlenen $\alpha/2$ yanılma düzeyine göre, z tablosundan bulunacak değer,

σ : Ana kütle standart sapması,

d : Hoşgörü miktarı (Örneklem ortalaması ile ana kütle ortalaması arasındaki hoşgörüleebilecek sapma miktarı)

Örnek: Bir araştırmacı yeni doğan bebeklerin ortalama doğum ağırlığını tahmin etmeye çalışıyor. Eğer % 95 güvenirlilikle bebek ağırlıklarını gerçek ortalamadan (kitle ortalaması) ± 250 gr sapma ile tahmin etmeye çalışıyorsa, örnekleme kaç yeni doğan alınmalıdır? (Kitle Standart Sapması 700gr. olarak kabul ediliyor)

$$z_{\alpha/2} = z_{0.025} = 1.96$$
$$\sigma^2 = 700^2$$
$$d = 250 \text{ gr}$$

$$n = \frac{1.96^2 \times 700^2}{250^2} = 30$$

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{d^2}$$

Eğer $N=60$ ise;

$$n = \frac{60 \times 1.96^2 \times 700^2}{250^2 (59) + 1.96^2 \times 700^2} = 20$$

$$n = \frac{N z_{\alpha/2}^2 \sigma^2}{d^2 (N - 1) + z_{\alpha/2}^2 \sigma^2}$$

2- Kitle oranının kestiriminde kullanılacak örneklem büyüklüğü formülü:

- *Kitle büyüklüğü (N) bilinmiyorsa;*

$$n = \frac{z_{\alpha/2}^2 P(1-P)}{d^2}$$

- *Kitle büyüklüğü (N) biliniyorsa;*

$$n = \frac{N z_{\alpha/2}^2 P(1-P)}{d^2 (N-1) + z_{\alpha/2}^2 P(1-P)}$$

$z_{\alpha/2}$: Belirlenen $\alpha/2$ yanılma düzeyine göre, z tablosundan bulunacak değer,

P: Ana küttelede incelenen olayın görülme oranı

d: Hoşgörü miktarı (Örneklem oranı ile ana kütle oranı arasındaki hoşgörüleebilecek sapma miktarı)

Örnek: Bir bölgede 50 yaş ve üstü yetişkinlerde 0.40 oranında olduğu tahmin edilen malnütrisyon (beslenme yetersizliği) sıklığını 0.04 hata ve 0.95 olasılıkla kestirebilmek için kullanılacak uygun örneklem büyüklüğü nedir?

$$z_{\alpha/2} = z_{0.025} = 1.96$$

$$P = 0.40$$

$$d = 0.04$$

$$n = \frac{1.96^2 (0.40)(0.60)}{0.04^2} = 577$$

$$n = \frac{z_{\alpha/2}^2 P(1-P)}{d^2}$$

Eğer bu bölgede 5000 yetişkin olduğu öngörülürse;

$$n = \frac{5000 \times 1.96^2 \times (0.40)(0.60)}{0.04^2 (4999) + 1.96^2 \times (0.40)(0.60)} = 517$$

$$n = \frac{N z_{\alpha/2}^2 P(1-P)}{d^2 (N-1) + z_{\alpha/2}^2 P(1-P)}$$

ÖRNEKLEME DAĞILIMLAR

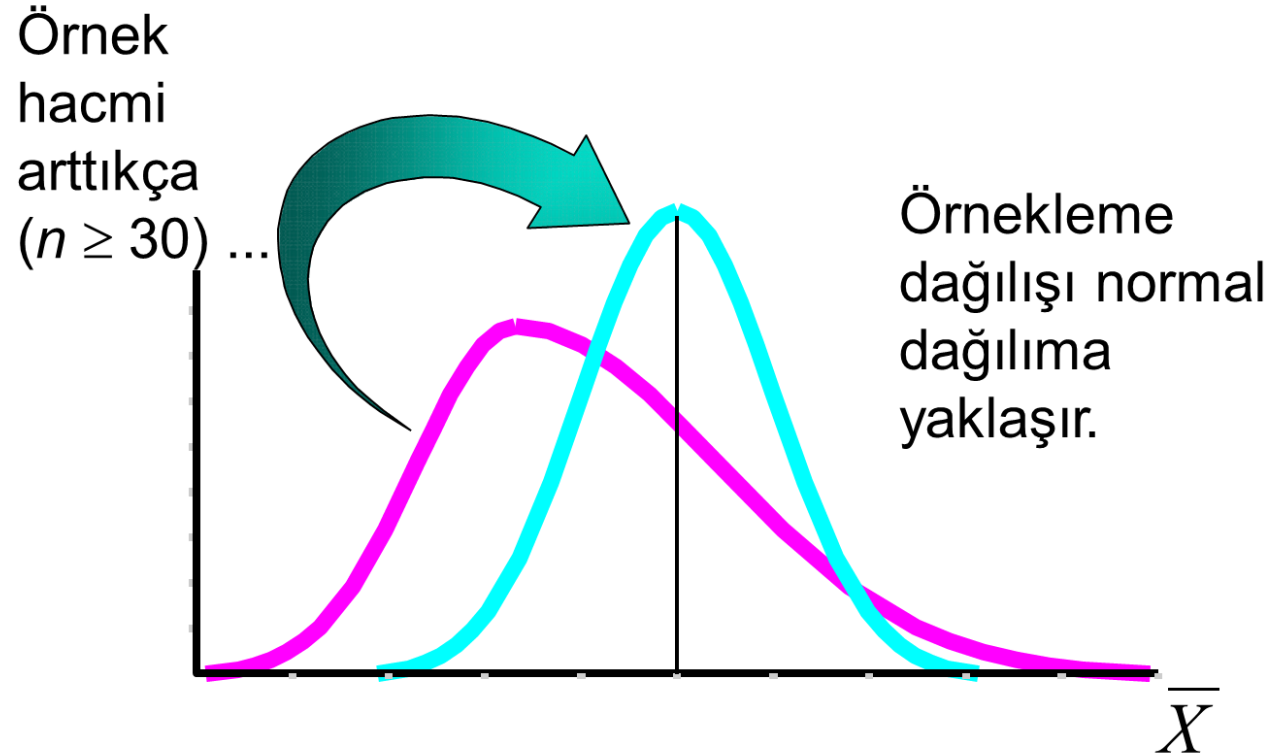
- Bir ana kütlede seçilebilecek tümü n birimlik olası bütün örneklemeler çekilip hepsinden ayrı ayrı istatistik hesaplandığında, söz konusu istatistiklerin dağılımına, hesaplanan istatistiğin örnekleme dağılımı adı verilir.
- Bir örneğin ortalaması hesaplanmışsa elde edilen \bar{X}_i dağılımı **ortalamların örnekleme dağılımını** verir.
- Her örnek için p oranları hesaplandığında **oranların örnek dağılımı** elde edilir.

ÖRNEKLEME DAĞILIMLAR

- Eğer örneklemeler eşit olasılıklı ve rassal olarak seçilmiş ise, örneklem ortalamalarının alacağı değerler μ ana kütle ortalamasından bir miktar farklılık gösterir. Bu farklılıklar örnekleme hatasından kaynaklanır.
- Ana kütlede n hacimlik olası bütün örneklemeler iadeli seçim yöntemi ile çekildiğinde, bir diğer deyişle seçilen birim yerine konularak seçim yapıldığında, belirlenen değişken için örneklem ortalamalarının dağılımının iki önemli özelliği bulunur:
 1. Örneklem ortalamalarının ortalaması ana kütle ortalamasına eşittir.
 2. Örneklem ortalamalarının standart sapmasının değeri, ana kütle standart sapmasından küçüktür ve ana kütle standart sapmasının örneklem hacminin kareköküne bölünmesiyle hesaplanır.

Merkezi Limit Teoremi

- Örneklem ortalamalarının örnekleme dağılımı, örneklem hacmi büyüdükçe ($n \geq 30$) Normal dağılıma yaklaşır.



Ortalamaların Örnekleme Dağılımı

- Ortalamaların örnekleme dağılımı anakütle ortalamasının iyi bir tahmincisidir.
- Her biri n hacimli çok sayıda örneğe ait ortalamaların gösterdiği dağılımın değişkenliği tek örneğin değişkenliğinden daha azdır.
- Standart sapma bir örneğin değişkenliği hakkında bilgi verirken ,
- Ortalamaların örnekleme dağılımınının değişkenliği standart hatayla gösterilir.

Ortalamaların Örneklem Dağılımı

- Ana kütle standart sapması bilindiğinde standart hata:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

- Standart z değerleri örneklem dağılımı

$$Z = \frac{\bar{X} - \mu_x}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu_x}{\sigma_x / \sqrt{n}}$$

Örnek: Bir arařtırmacı, annelerin ilk doğumlarını yaptıkları yaş ortalamasını tahmin etmek istiyor. Rasgele olarak 10 anne seçiyor:

Anne No	1	2	3	4	5	6	7	8	9	10
İlk doğum yaşı	24	20	26	19	20	23	28	22	18	25

Kitle ortalamasının nokta tahmini:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{225}{10} = 22.5$$

Örneklem standart sapması:

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{5159 - \frac{(225)^2}{10}}{9}} = 3.27$$

Ortalamanın tahmini standart hatası:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{3.27}{\sqrt{10}} = 1.04$$

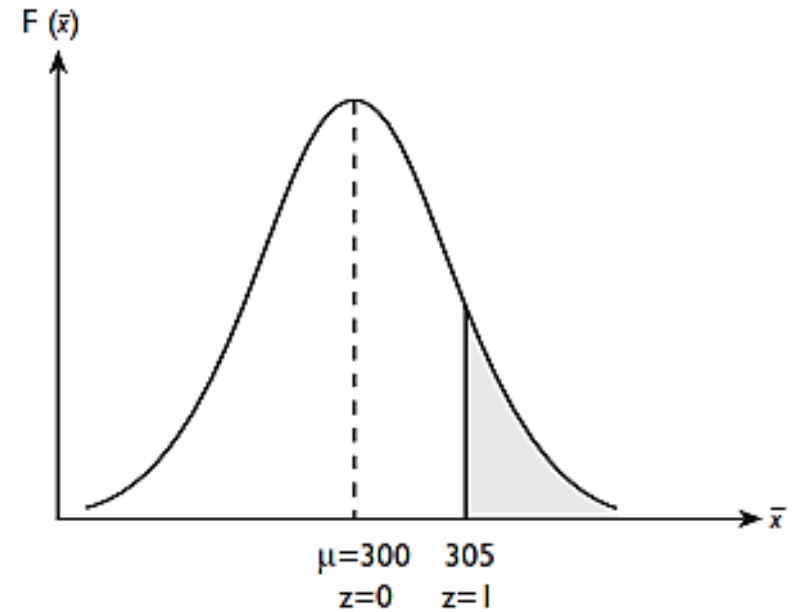
Örnek: Bir meşrubat otomat makinesinde hazırlanan meşrubat miktarı rassal değişkeninin aritmetik ortalamasının 300 ml. ve standart sapmasının 40 ml. olduğu bilinmektedir. Bu makinede hazırlanan 64 adet meşrubattan oluşan olası tüm örneklem için ortalama meşrubat miktarının en az 305 ml. olması olasılığı nedir?

- $\mu_{\bar{x}} = 300$

- $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{40}{\sqrt{64}} = 5$

- $Z = \frac{\bar{X} - \mu_x}{\sigma_{\bar{x}}} = \frac{305 - 300}{5} = 1$

- $P(\bar{X} > 305) = P(Z > 1) = 0,1587$



Oranların Örnekleme Dağılımı

- Oranların örnek dağılımının ortalaması anakütle oranına eşittir.

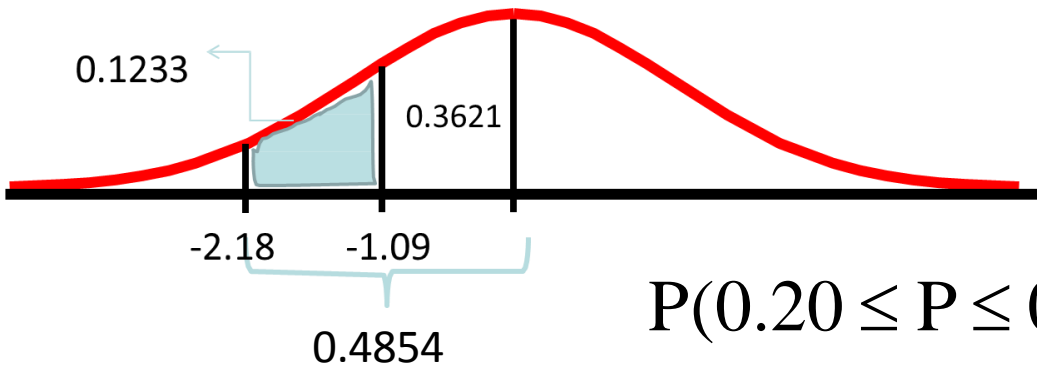
$$\sigma_p = \sqrt{\frac{P(1 - P)}{n}}$$

$$Z = \frac{p - P}{\sqrt{\frac{P(1 - P)}{n}}}$$

ÖRNEK: Bir alışveriş merkezinde 2000 YTL'den fazla alışveriş yapan müşterilerin %30'unun kredi kartı kullandığı tespit edilmiştir. 2000 YTL'den fazla alışveriş yapan 100 müşteri için;
a) oranların örneklem dağılımının standart hatası nedir?
b) 2000 YTL'den fazla alışveriş yapan 100 müşteriden %20 ile %25'inin kredi kartı kullanması olasılığını hesaplayınız

$$a) \sigma_p = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{0.30(1-0.30)}{100}} = 0.0458$$

$$b) Z = \frac{p-P}{\sqrt{\frac{P(1-P)}{n}}} \rightarrow Z_1 = \frac{0.20-0.30}{\sqrt{\frac{0.30(1-0.30)}{100}}} = -2.18 \text{ ve } Z_2 = \frac{0.25-0.30}{\sqrt{\frac{0.30(1-0.30)}{100}}} = -1.09$$



$$P(0.20 \leq P \leq 0.25) = P(-2.18 \leq Z \leq -1.09) = 0.4854 - 0.3621$$

$$P(0.20 \leq P \leq 0.25) = 0.1233$$