Çıkarımsal İstatistik

Bir araştırma yapmanın amacı, çalışılan örneklemden kitleye genelleme yapabilmektir.

Çıkarsama yapma işlemi, olasılığa dayanan istatistiksel yöntemlerle yapılır.

Binom, poisson ya da Normal Dağılım gibi kuramsal (olasılıksal) dağılışlar yardımıyla, kitlede bir olayın görülme olasılıklarının nasıl elde edilebileceği konusu üzerinde önceki bölümlerde durulmuştu.

Bu bölümde

Örneklemden elde edilen bulgular yardımıyla çıkarsamalı istatistiğin iki ana konusu olan:

1) Evren hakkında kestirimde bulunma,

2) Hipotezleri test etmenin dayanağı olan

örneklem dağılışları **tanıtılacaktır.**

Kitleden Örneklem Çekmenin Nedenleri

1) Örneklemin incelenmesi kitlelere göre daha kısa sürede yapılır.
2) Örneklemin incelenmesi, kitlenin incelenmesinden daha ucuzdur.
3) Bazı durumlarda kitlenin incelenmesi olanaksız olabilir.
4) Örneklem sonuçları daha doğru olabilir. Çünkü daha az sayıda kişi ile (örnek ile) çalışılacağından, daha deneyimli insanlar daha özenli iş yapabilir.
5) Eğer örneklemimiz olasılıksal yöntemlerle seçiliyorsa, yapılan örnekleme hatasının kestirimini de bulmak mümkündür.

Gözlem değerlerinin dağılımından farklı olarak, bu gözlemlerin oluşturduğu örneklemlerden elde edilen (hesaplanan) istatistiklerin (ortalama, oran, varyans v.b.) dağılımları da önemlidir.

N genişliğinde bir kitleden n genişliğinden çekilebilecek bir çok örneklem vardır.

Eğer kitleden örneklem çekme işlemi yerine konulmadan yapılıyorsa n genişliğinde çekilebilecek örneklem sayısı

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$ **dir.**

Bu olası örneklemlerin her birinden bir istatistik (ortalama, oran, standart sapma v.b.) hesaplanabilir.

Örneklemlerden hesaplanan istatistiklerin dağılışına örneklem dağılışı denir.

N genişliğindeki kitleden n genişliğinde elde edilen tüm örneklemlerden birer ortalama hesaplanabilir ve bu örneklem ortalamalarının bir dağılımı elde edilebilir.

Buna ortalamanın örneklem dağılışı adı verilir.

Gözlem birimlerinin dağılımını tanımlamak için ortalama ya da standart sapma gibi ölçümler kullanılır.

Örneklem dağılımının özelliklerini tanımlamak için de örneklem dağılışının ortalamasını ve standart sapmasını kullanırız.

Merkezi eğilim ölçüsü olarak ortalama tek tek gözlem birimlerinin nerede odaklandığını gösterirken, örneklem dağılışının ortalaması da, örneklemlerden elde edilen ortalamaların nerede odaklandığını gösterir.

Yaygınlık ölçüsü olarak kullanılan standart sapma da, tek tek gözlem birimlerinin ortalamadan ne kadarlık bir ayrılış gösterdiğini tanımlarken, örneklem dağılışının standart sapması da her bir örneklemden elde edilen ortalamaların ne derece yaygınlık gösterdiğini tanımlar.

Aynı büyüklükteki örneklemlerden elde edilen ortalamalar ne kadar birbirine yakınsa (örneklemden örnekleme değişim ne kadar azsa) herhangi bir örneklem sonucu o kadar güvenilirdir ya da kesindir.

Eğer hesaplanan ortalamalar, bir örneklemden diğerine çok farklılık gösteriyorsa, çekilen herhangi bir örneklemden elde dilen ortalama (kestirim) o derece az güvenilir ya da kesindir.

Bu nedenle örneklem dağılışının standart sapması kesinliğin ya da hatanın bir ölçüsü olarak kullanılır.

Uygulamada hiçbir zaman olası tüm örneklemleri ya da bir kitleden bir çok örneklem çekmeyiz. İstatistik kuramı elimizdeki bir örneklemden yararlanarak örneklem dağılışının özelliklerini bulmamıza yardımcı olur.

Merkezi limit teoremi olarak adlandırılan teoreme göre örneklem ortalamalarının gösterdiği dağılım, normal dağılımdır.

Normal dağılımı tanımlayan parametreler, dağılımın ortalaması ve standart sapması olduğundan bu parametrelerin özelliklerinin bilinmesi gerekir.

Örnek: N=6 olan bir kitledeki gözlem değerleri aşağıdadır.

$x_1 = 5$ $x_2 = 9$ $x_3 = 4$ $x_4 = 1$ $x_5 = 7$ $x_6 = 6$

$\mu = 5{,}33$

$\sigma = 2{,}494$

Bu kitleden n=3 genişliğinde çekilebilecek

$$\binom{6}{3} = 20$$

Tane olası örneklem vardır. Bu örneklemlerin her birinden bir ortalama hesaplandığında, ortalamanın örneklem dağılımını elde ederiz.

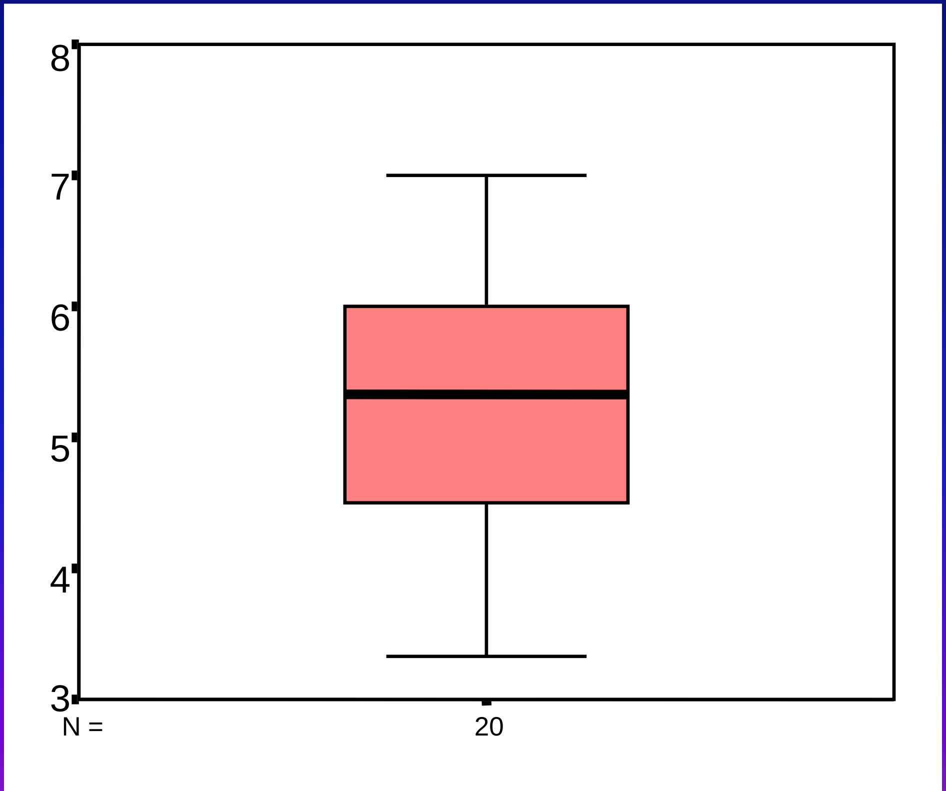
Olası Örneklem	Örneklemlerdeki Değerler	Örneklem Ortalamaları
1. $(x_1\ x_2\ x_3)$	5, 9, 4	$\bar{x}_1 = 6$
2. $(x_1\ x_2\ x_4)$	5, 9,1	$\bar{x}_2 = 5$
3. $(x_1\ x_2\ x_5)$	5, 9,7	$\bar{x}_3 = 7$
4. $(x_1\ x_2\ x_6)$	5, 9, 6	$\bar{x}_4 = 6{,}67$
5. $(x_2\ x_3\ x_4)$	5, 4, 1	$\bar{x}_5 = 3{,}33$
.	.	.
.	.	.
20. $(x_5\ x_6\ x_7)$	1, 7, 6	$\bar{x}_{20} = 4{,}67$

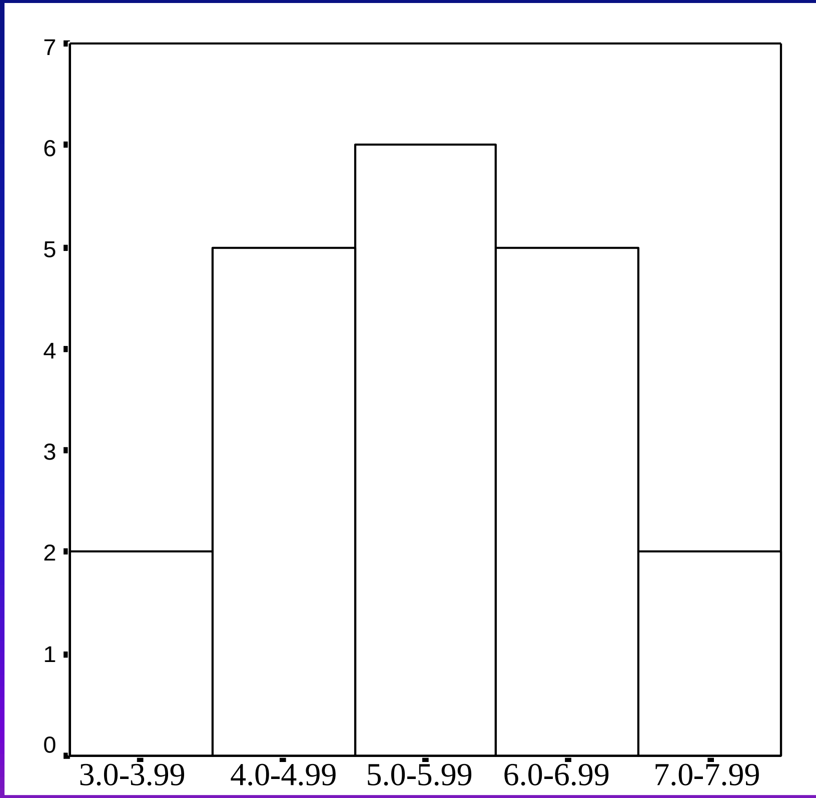
Örneklem No	Örneklemlerdeki Değerler			Örneklem Ortalamaları
1	5	9	4	6,00
2	5	9	1	5,00
3	5	9	7	7,00
4	5	9	6	6,67
5	5	4	1	3,33
6	5	4	7	5,33
7	5	4	6	5,00
8	5	1	7	4,33
9	5	1	6	4,00
10	5	7	6	6,00
11	9	4	1	4,67
12	9	4	7	6,67
13	9	4	6	6,33
14	9	1	7	5,67
15	9	1	6	5,33
16	9	7	6	7,33
17	4	1	7	4,00
18	4	1	6	3,67
19	4	7	6	5,67
20	1	7	6	4,67

Olası tüm örneklem ortalamalarının ortalaması alındığında

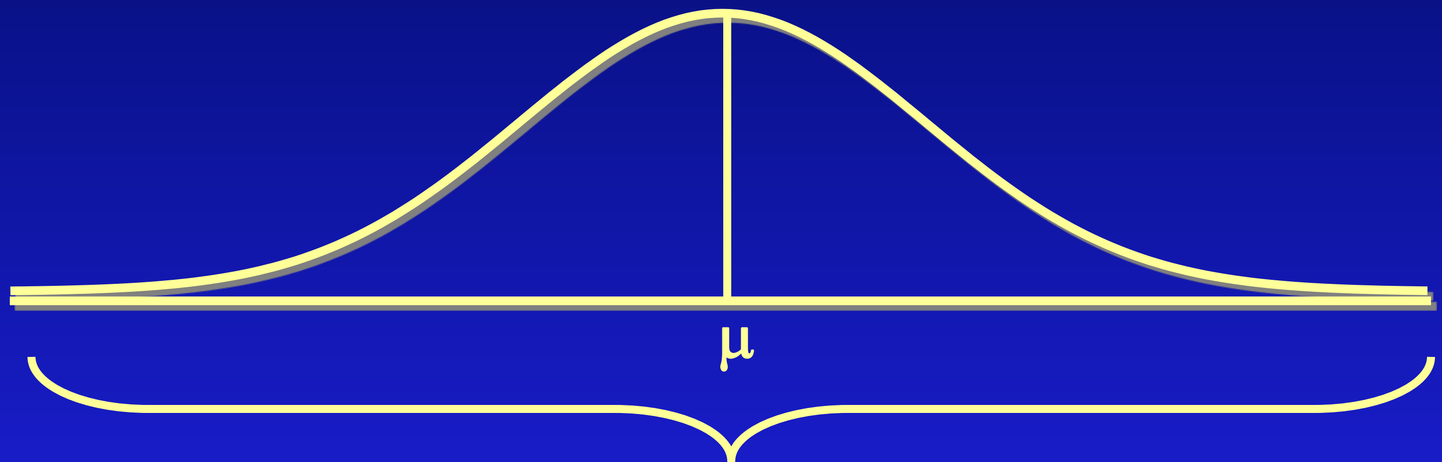
$$\mu_{\bar{x}} = 5{,}33$$

Olarak bulunur ve bu ortalama kitle ortalamasına eşittir. Bu ortalamaların dağılışı normal dağılım gösterir.



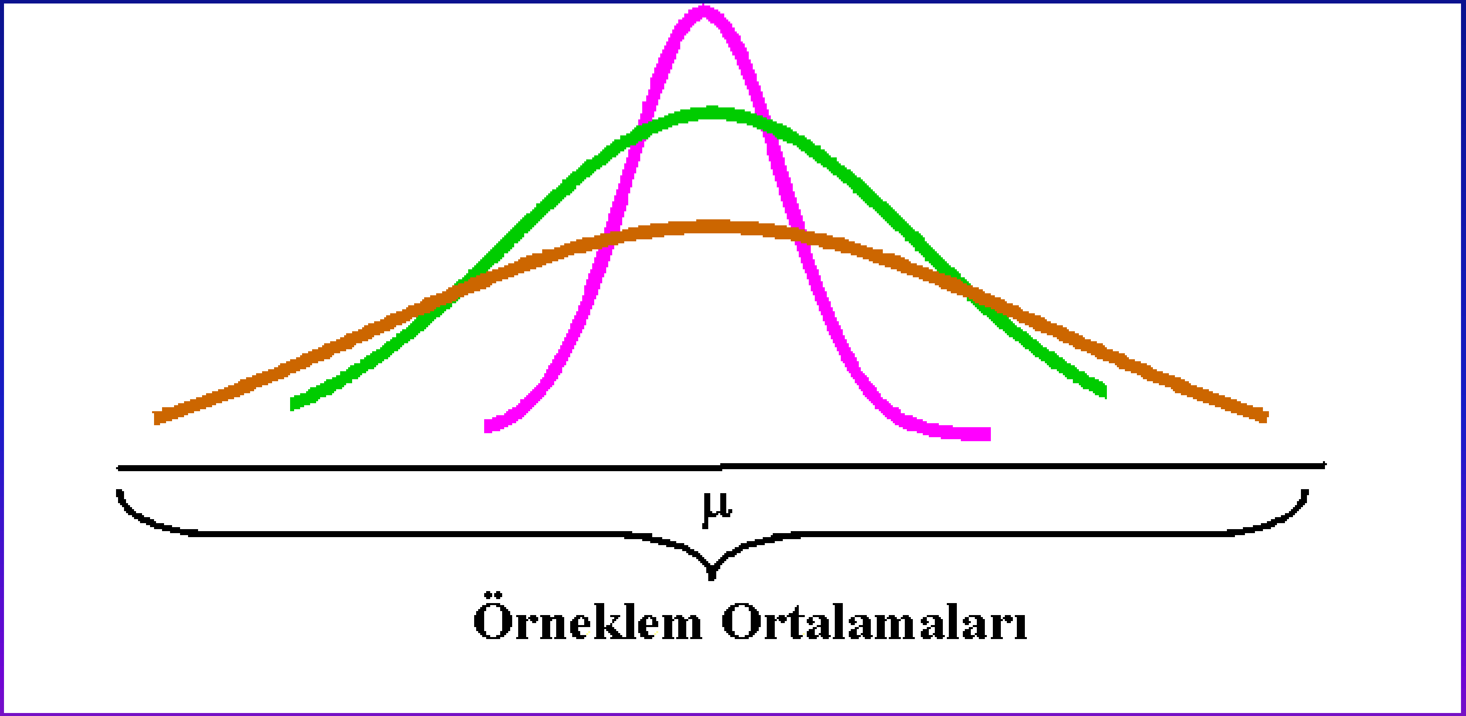


Bu bilgilere göre örneklem ortalamaları kitle ortalaması etrafında bir normal dağılım gösterir.



Örneklem Ortalamaları

Örneklem ortalamalarının standart sapması da örneklem ortalamalarının gerçekte bilmediğimiz kitle ortalaması etrafında nasıl bir dağılım gösterdiğini tanımlar. Örneklem ortalamaları, kitle ortalamasına çok yakın bir dağılım gösteriyorsa, bu ortalamaların dağılımının standart sapması küçük olacaktır.



Örneklem dağılışının değişkenliğini belirleyen iki parametre vardır.

1) **Kitle standart sapması (σ)**
2) **Örneklem genişliği (n)**

Kitledeki değişkenlik arttıkça (σ), örneklem dağılışının değişkenliği artar. Buna karşın, örneklem genişliğinin (n) büyümesi örneklem dağılışının standart sapmasını azaltır.

İncelenilen örnekte n=3 olduğunda örneklem ortalamaları 3,33 ile 7,33 arasında değişim göstermiştir.

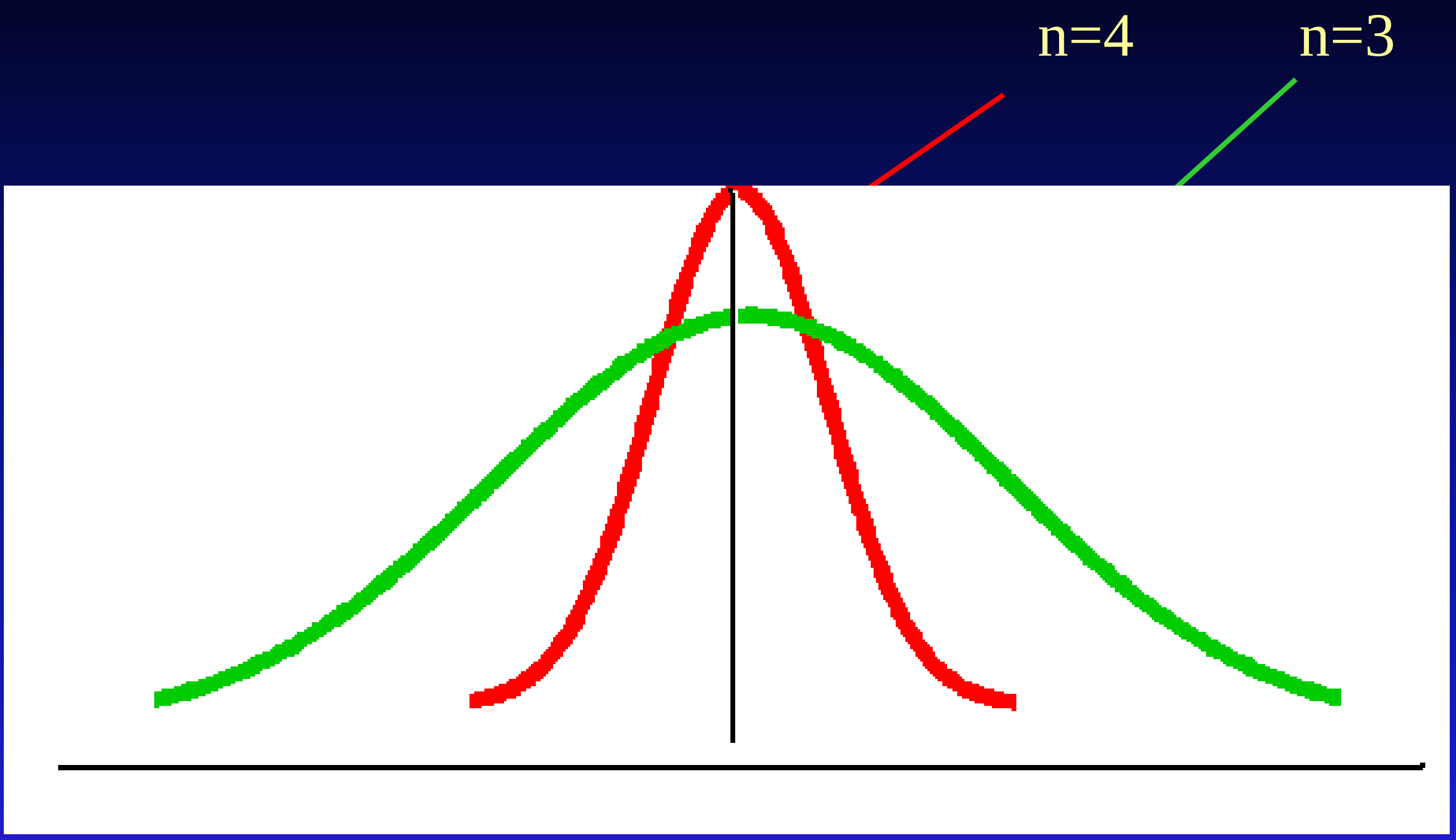
N=6 olan kitleden n=4 genişliğindeki tüm örneklemler incelendiğinde

çekilebilecek örneklem sayısı = $\binom{6}{4} = 15$ dir.

n=4 genişliğinde çekilen 15 örneklemden elde edilen ortalamalar 4 ile 6,75 arasında değişim göstermektedir.

n=3	**3,33 - 7,33**	**DA=4,00**
n=4	**4,00 - 6,75**	**DA=2,75**

Örneklem ortalamalarının dağılımının değişkenliği örneklem genişliği büyüyünce azalmıştır. Kitleden n=4 genişliğinde tek bir örneklem çekildiğinde elde edilen ortalama n=3 genişliğinde çekilen örneklemden elde dilen ortalamaya göre kitle ortalamasına daha yakın olma eğilimindedir.


n=4
n=3

Değişkenliğin ölçüsü olarak "varyans" kullanıldığında, örneklem ortalamalarının dağılımının varyansı, kitle varyansının örneklem genişliğine bölünmesi ile bulunur.

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

Varyans yerine daha yaygın kullanılan standart sapmayı kullanacak olursak, Örneklem ortalamalarının dağılımının standart sapması, kitle standart sapmasının örneklem genişliğinin kareköküne bölünmesi ile bulunacaktır.

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

Örneklem ortalamalarının gösterdiği dağılımın standart sapması ortalamanın standart hatası olarak adlandırılır.

Standart hata, tüm örneklem ortalamalarının kitle ortalaması etrafındaki dağılımını (yaygınlığını) gösterdiği için, örneklem ortalamasının kitle ortalamasını ne kadar kesinlikle kestirdiğinin bir ölçüsüdür.

N genişliğindeki bir örneklemden hesaplanan, örneklem standart sapması (s) kitle standart sapması (σ) nın bir nokta kestirimidir. Bu durumda standart hata

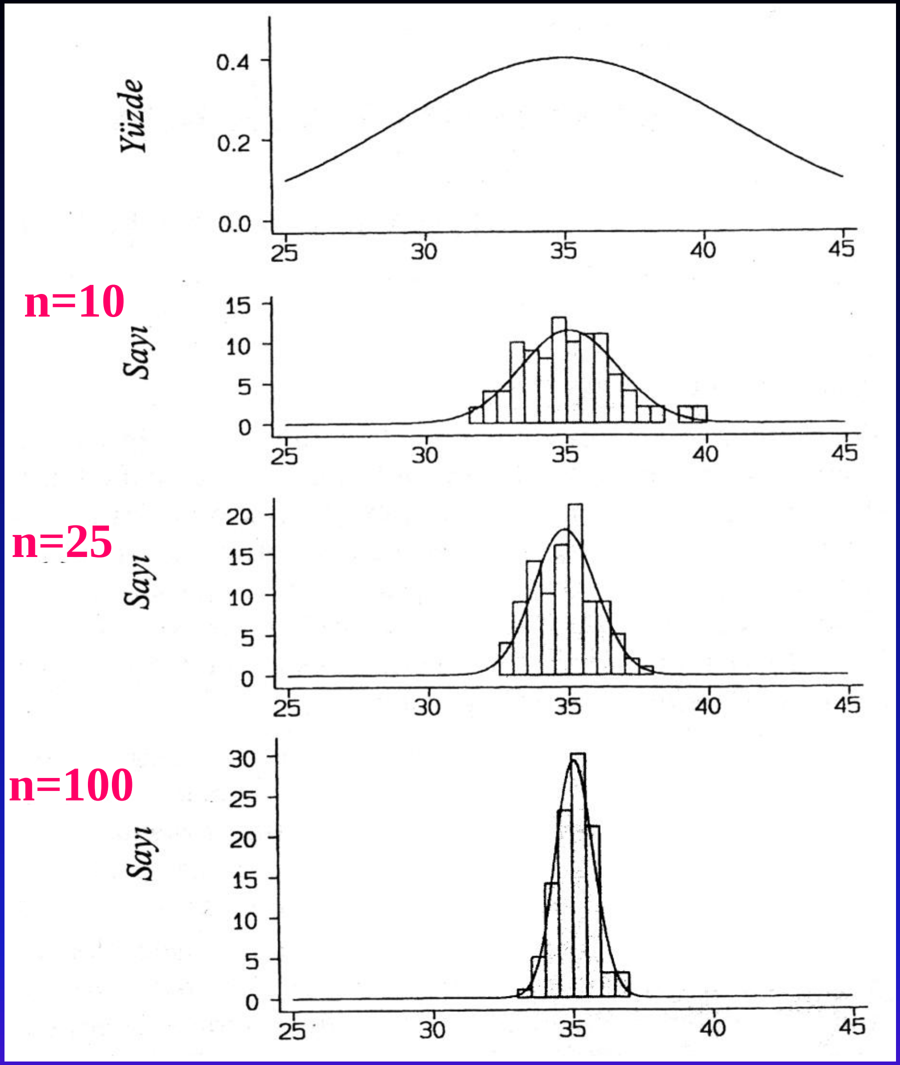
$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

İle kestirilir.

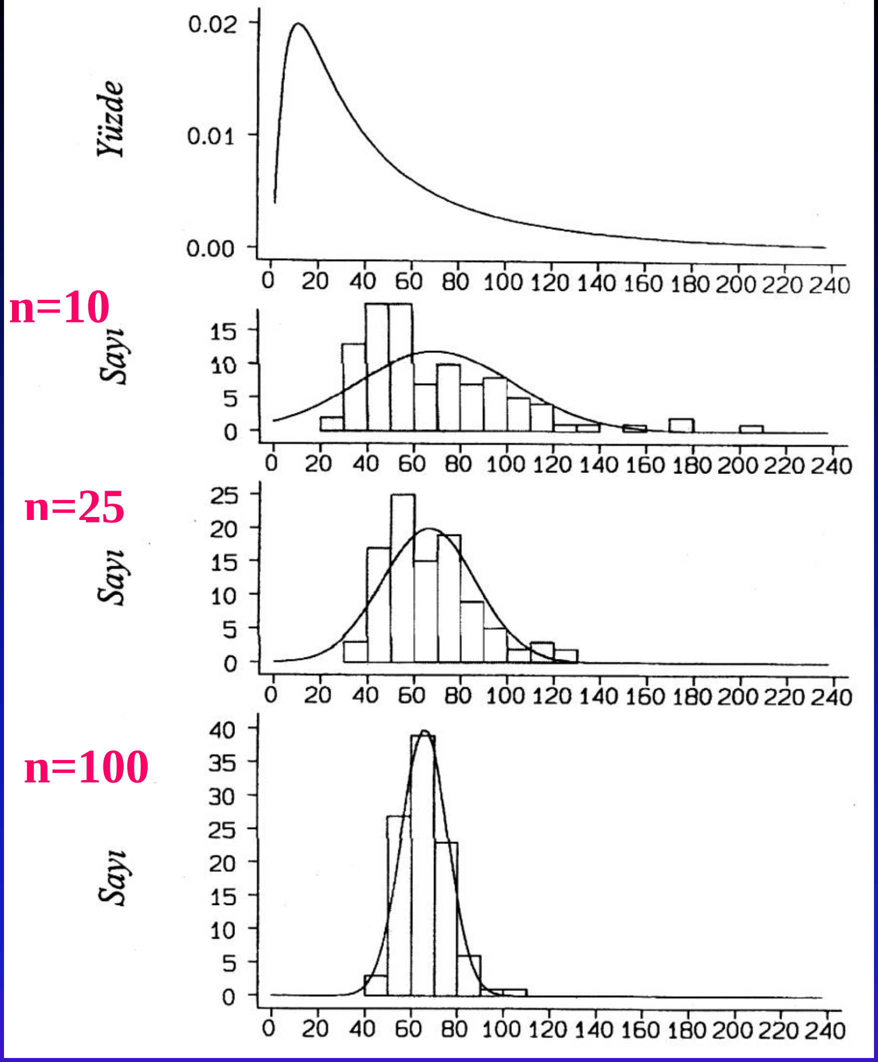
Eğer örneklemler normal dağılıma sahip bir kitleden çekiliyorsa, örneklem ortalamalarının dağılımı (ortalamanın örneklem dağılışı) da normaldir.

Eğer örneklemlerin çekildiği kitle normal dağılmıyorsa (örneklem normal dağılım göstermeyen bir kitleden çekiliyorsa), örneklem ortalamalarının dağılımı örneklem genişliği büyüdükçe normal dağılıma yaklaşır.

μ



Ortalaması 35 ve S. Sapması 6 Olan Bir Normal Dağılımdan Çekilen 10, 25 ve 100 gözlemli 100 Örnekleme İlişkin Ortalamaların Hipotetik Dağılımı



Ortalaması 64 mol mol Olan ve Normal Olmayan Bir Dağılımdan Çekilen 10,, 25 ve *100 Gözlemli 100 Örnekleme İlişkin Ortalamaların Hipotetik Dağılımı*

Kestirim

Pratikte kitle parametrelerinin doğrudan hesaplamak olanaklı değildir. Bunun yerine herhangi bir kitle parametresi, elde edilen örneklem istatistiğinden kestirilir.

İstatistik örneklemden örnekleme değişim gösterir.

Kestirim işleminde belirsizlik vardır.

Kitle parametrelerinin belirli bir güvenle içinde bulunduğu aralığın tanımlanması işlemine

***güven aralığı yöntemi* denir.**

Örnek : akut miyokard enfarktüs tanısı almış 100 erkekten elde edilen ortalama kolesterol düzeyi 240 mg/dl olarak bulunmuş olsun. Örneklemin çekildiği kitlenin ortalaması hakkında kestirim yapılmak istenebilir.

$$n = 100$$

$$\bar{x} = 240$$

Kitlede kolesterol düzeyi değerlerinin standart sapmasının 40 mg/dl olarak bilindiğini varsayalım.

$$\sigma = 40$$

Bu örnek için

$$\bar{x} \pm 1{,}96 \times \sigma_{\bar{x}}$$

$$240 \pm 1{,}96 \frac{40}{10} \Rightarrow 240 \pm 7{,}84$$

Bilinmeyen kitle ortalaması % 95 olasılıkla 232,16 ile 247,84 arasında yer almaktadır.

Uygulamada kitle standart sapması (σ) bilinmez ve örneklem standart sapması s ile kestirilir. σ yerine s'nin kullanımı ile

$\bar{x}$ 'da olduğu gibi s'nin de örneklemden örnekleme değişimi söz konusudur.

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Dağılımına dayandırılarak yazılan bu eşitlikte σ yerine s kullanıldığında:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

Elde edilir. Bu dağılıma t dağılışı adı verilir.

Uygulamada kitle standart sapmasını bilmediğimiz için bilinmeyen kitle ortalamasının güven sınırları aşağıdaki gibi belirlenir.

$$\bar{x} - t_{(n-1;\ \alpha/2)} \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{(n-1;\ \alpha/2)} \frac{s}{\sqrt{n}}$$

Bilinmeyen kitle oranı için güven sınırları aşağıdaki gibi belirlenir.

$$p - t_{(n;\ \alpha/2)}\sqrt{\frac{p(1-p)}{n}} \leq P \leq p + t_{(n;\ \alpha/2)}\sqrt{\frac{p(1-p)}{n}}$$