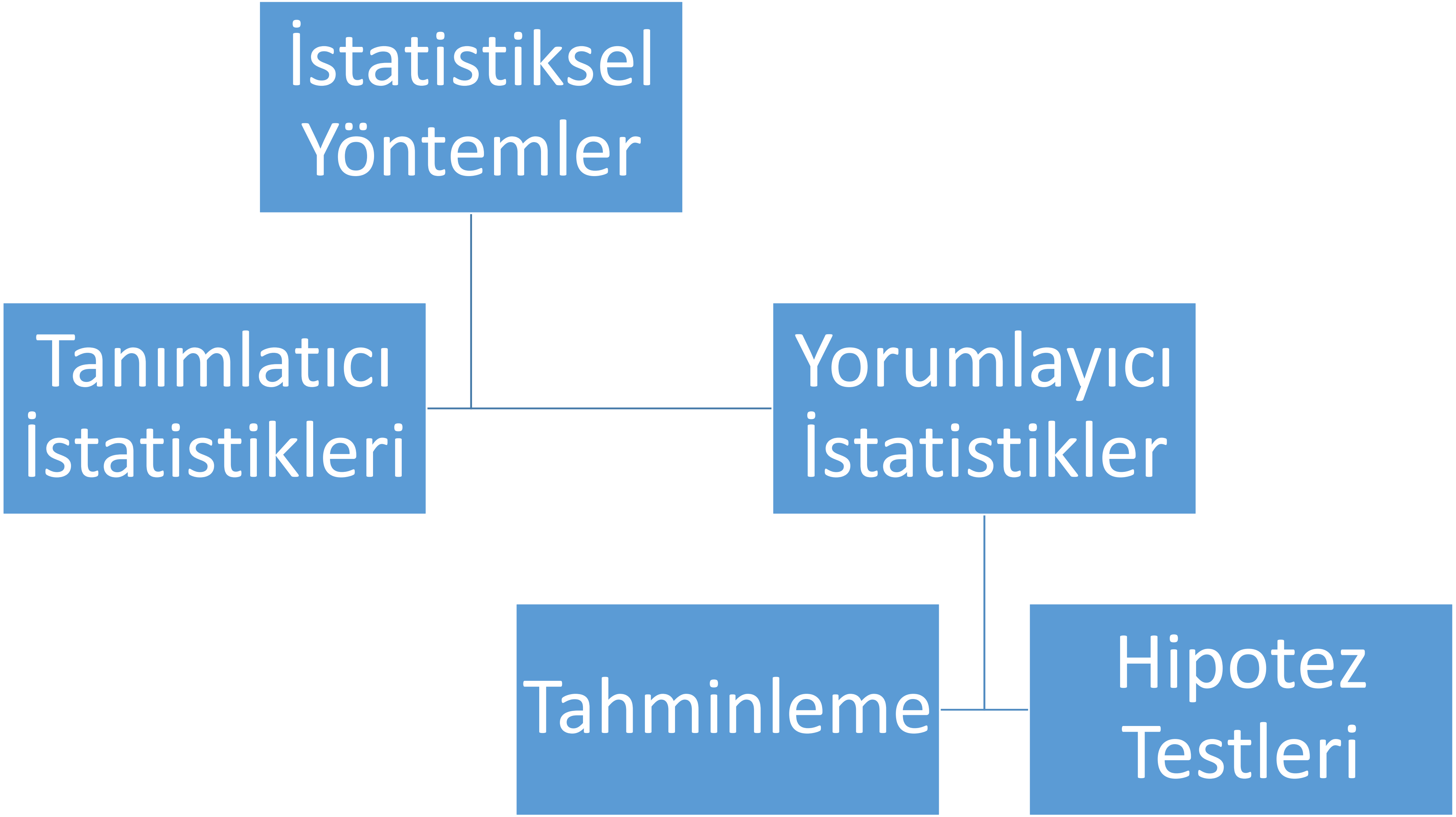
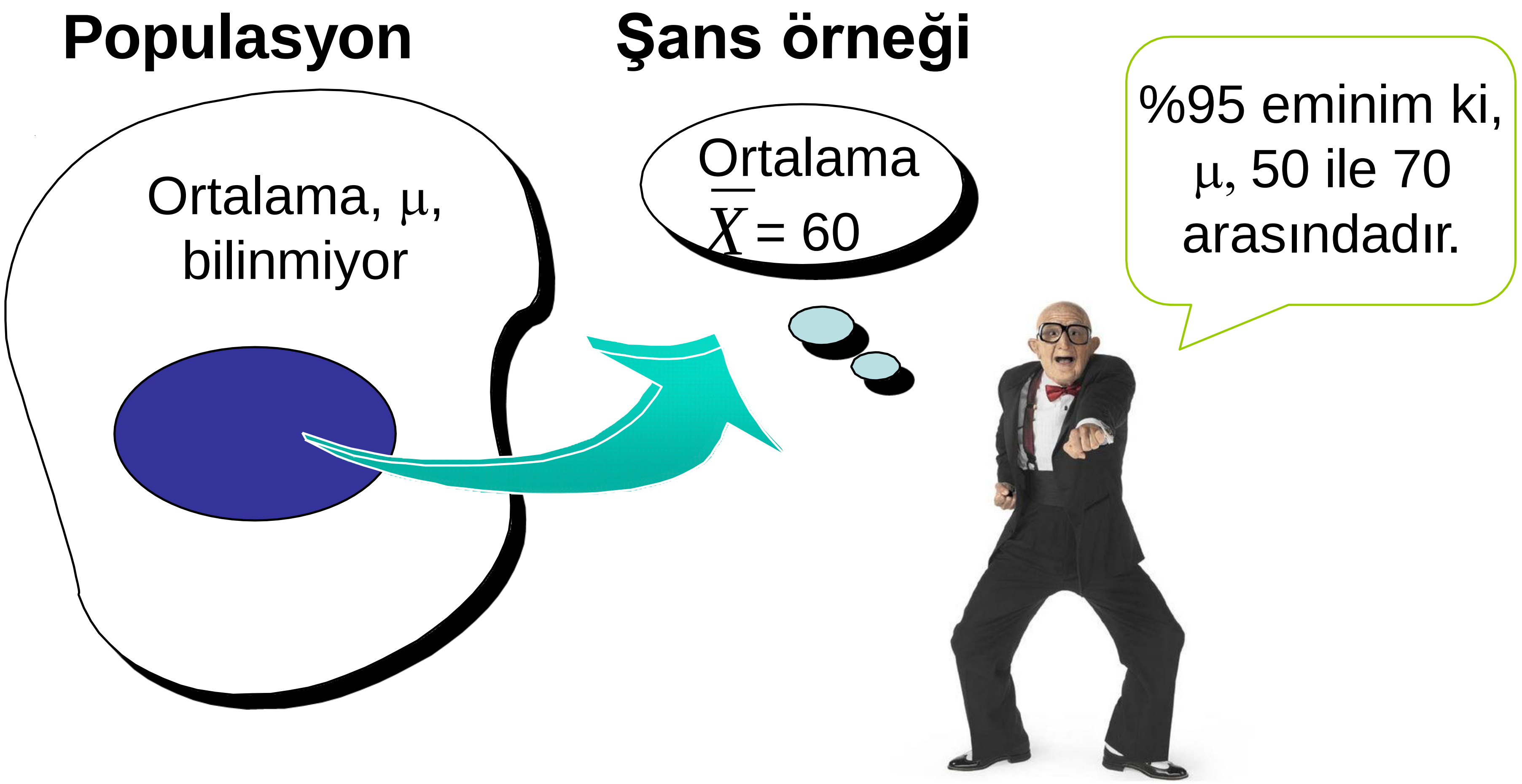
Nokta ve Aralık Tahminlemesi

Yrd. Doç . Dr. Emre ATILGAN


İstatistiksel Yöntemler
Tanımlatıcı İstatistikleri
Yorumlayıcı İstatistikler
Tahminleme
Hipotez Testleri

Tahmin Süreci Nedir?

Amacımız bilinmeyen popülasyon parametreleri tahmin yoluyla elde etmektir.



Tahmin

Nokta Tahmini

Populasyon parametresinin tek bir tahmin değerini verir

$$X = \hat{\mu}$$

$$s = \hat{\sigma}$$

$$p = \hat{P}$$

Aralık Tahmini

Populasyon parametresinin tahmin aralığını verir. Nokta tahmini kullanılarak hesaplanır.

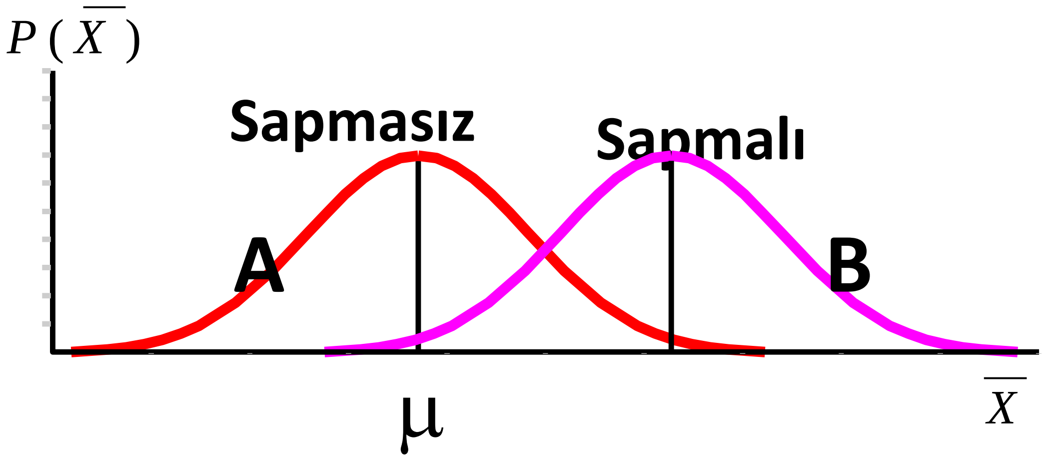
$$30 \leq \mu \leq 50$$

$$1.5 \leq \sigma^2 \leq 2.4$$

$$0.45 \leq P \leq 0.65$$

Tahminleyicilerin Özellikleri

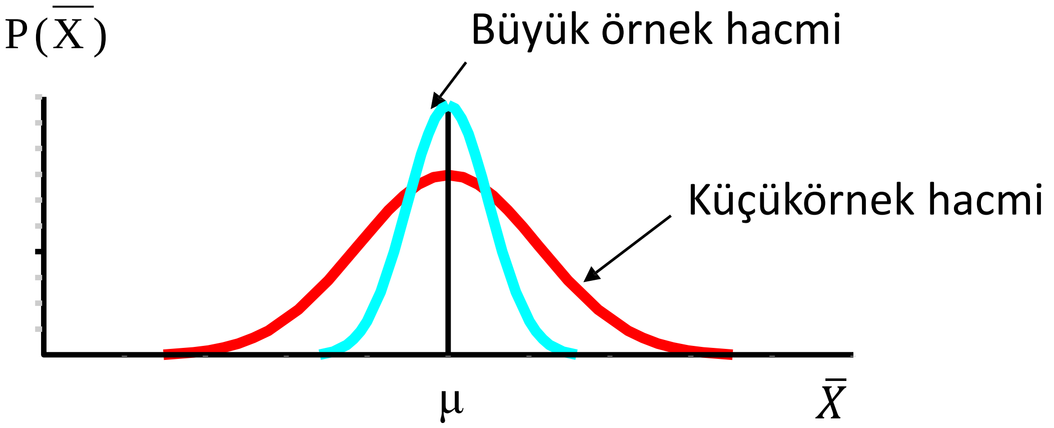
1. Sapmasızlık



$$E(\bar{X}) = \mu \rightarrow E(X) - \mu = 0$$

N birimlik aynı anakütleden farklı sayıda örneklem seçilebileceği için tahmin edicinin değeri de seçilen örnekleme göre değişmektedir. Bu durumda örneklem sayısı kadar elde edilen tahmin edici, bir rassal değişken olup, ortalaması ve varyansı olan bir olasılık dağılımına sahiptir. Bu dağılımın beklenen değerinin anakütle parametresine eşit olmasına, diğer bir ifadeyle bir istatistiğin beklenen değeri ile bilinmeyen anakütle parametresi arasındaki farkın sıfıra eşit olmasına *"sapmasızlık"* denir.

2. Tutarlılık (Kararlılık)

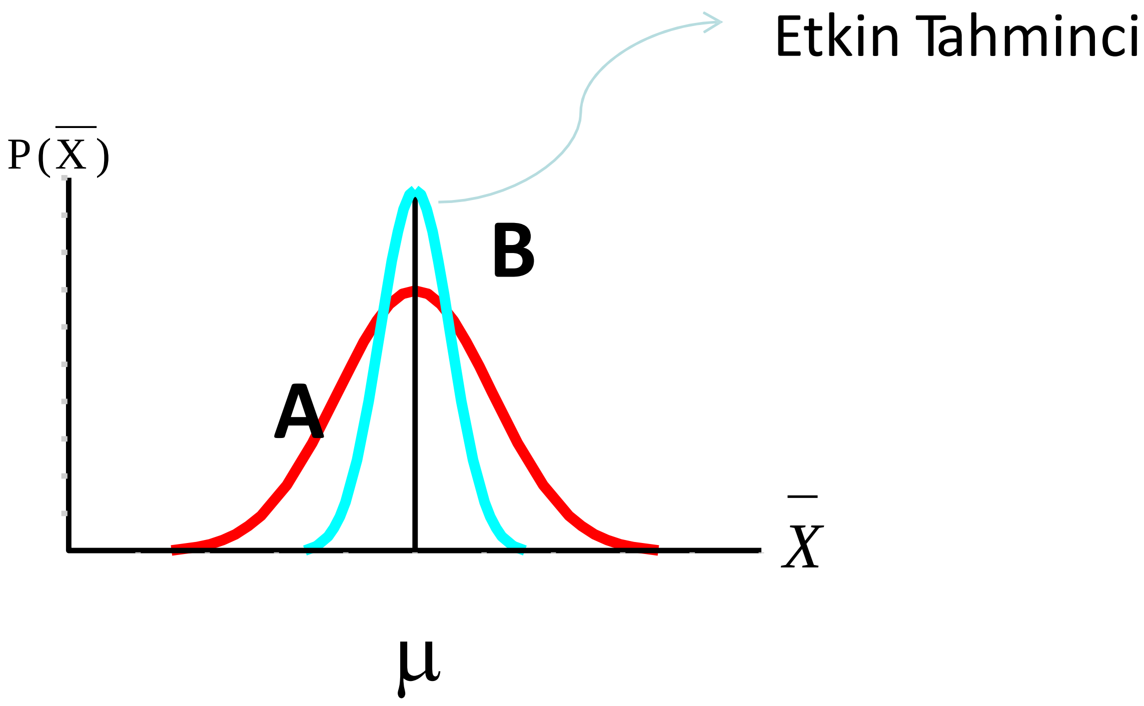


Örneklemdeki birim sayısı sonsuza doğru arttırıldığında, tahmin edicinin değerinin anakütle değerine yaklaşması ve n=N olması durumunda aralarındaki farkın sıfıra inmesi özelliğine ***"tutarlılık"*** denir.

$$\lim_{n\to\infty} P = \left(\left|\theta - \hat{\theta}\right| < \varepsilon\right) = 1$$

$\hat{\theta}$,θ'nın tutarlı tahmincisidir.

3. Etkinlik (En Küçük Varyans)



Birden fazla sapmasız ve tutarlı tahminci olması durumunda, bir tahmincinin varyansının, aynı anakütle parametresinin başka bir tahmincisinin varyansından daha küçük olması durumunda elde edilen tahmincilere ***"etkin"*** tahminci adı verilmektedir.

4. Yeterlilik

- Bir $\hat{\theta}$ tahmincisinin değeri belirlenirken seçilen örneklemdek tüm bilgilerden yararlanılıyor ise, bu $\hat{\theta}$ tahmincisi yeterli tahminci adını alır.
- Örneğin örneklem aritmetik ortalaması $\bar{X}$, örneklemdeki tüm değerler dikkate alınarak hesaplanır. Bu durumda $\bar{X}$ değeri, μ için yeterli bir tahmincidir.

Nokta Tahmini

Bir ana kütle parametresinin tek bir sayı olarak tahmininde kullanılan örneklem istatistiği değeridir.

	Ana Kütle Parametresi	Nokta Tahmini
Aritmetik Ortalama	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$
Varyans	$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n-1}$
Oran	$P = \frac{\sum_{i=1}^{N} x_i}{N}$	$p = \frac{\sum_{i=1}^{n} x_i}{n}$

Aralık Tahmini

- Ana kütle parametresinin içerisinde yer alacağı tahmin edilen ve belli bir güven düzeyine göre belirlenen sayısal değerler aralığına **Güven Aralığı** adı verilir.
- **Güven Düzeyi (1 - α):** Belli bir aralığın ana kütle parametresini içermesi olasılığına **güven düzeyi** denilir. Tipik değerler %99, %95, %90
- Bir θ parametresi için güven aralığı:

$\theta_{alt} \leq \theta \leq \theta_{\text{üst}}$ ve $P(\theta_{alt} \leq \theta \leq \theta_{\text{üst}}) = 1 - \alpha$

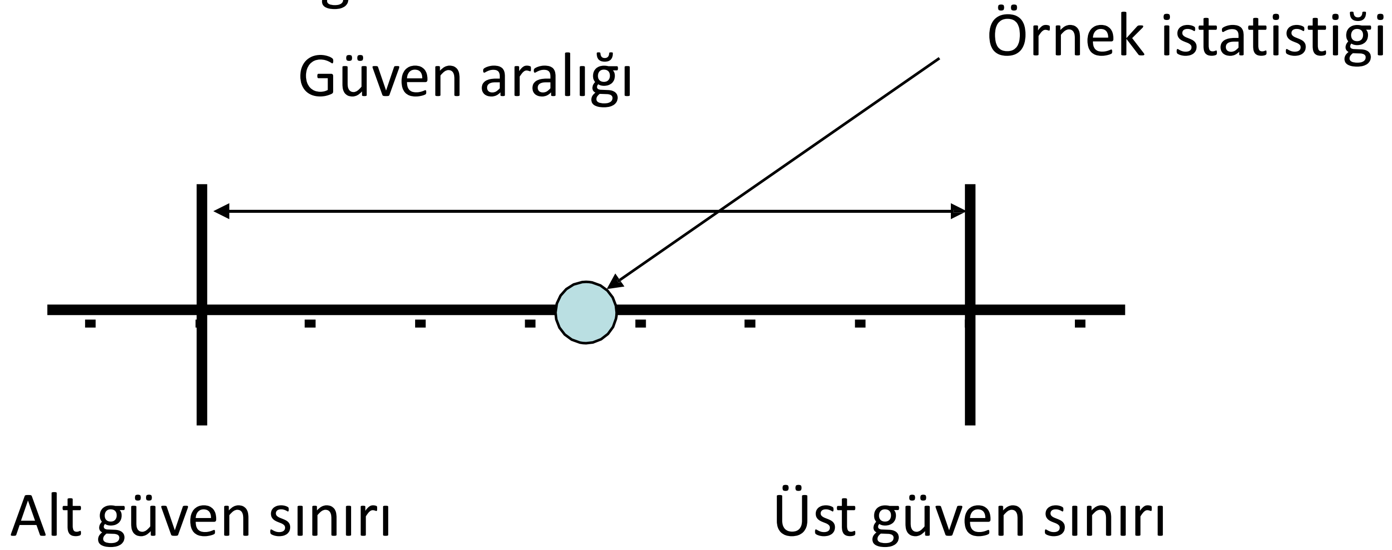
- $\alpha = 0.05$ ise güven düzeyi $1 - \alpha = 0.95$ olur.

Güven Aralığı Tahmini

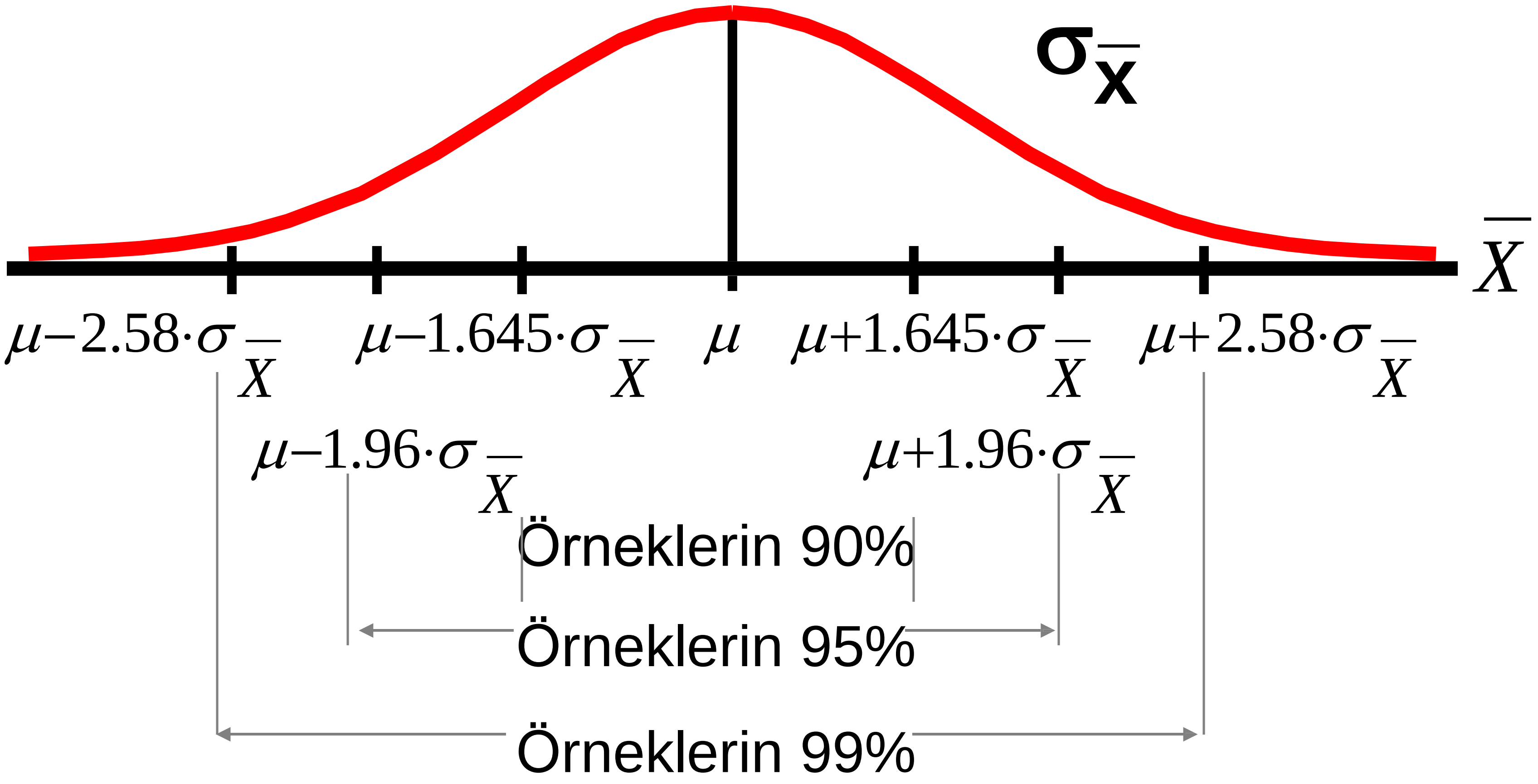
- ✰ Bir değer aralığı verir.
- ✰ Populasyon parametresine yakınlık hakkında bilgi verir.
- ✰ Olasılık terimleriyle ifade edilir.

Güven Aralığı Tahmininin Elemanları

Populasyon parametresinin aralık içinde bir yere düşmesinin olasılığı



Güven aralığı



Güven Seviyesi

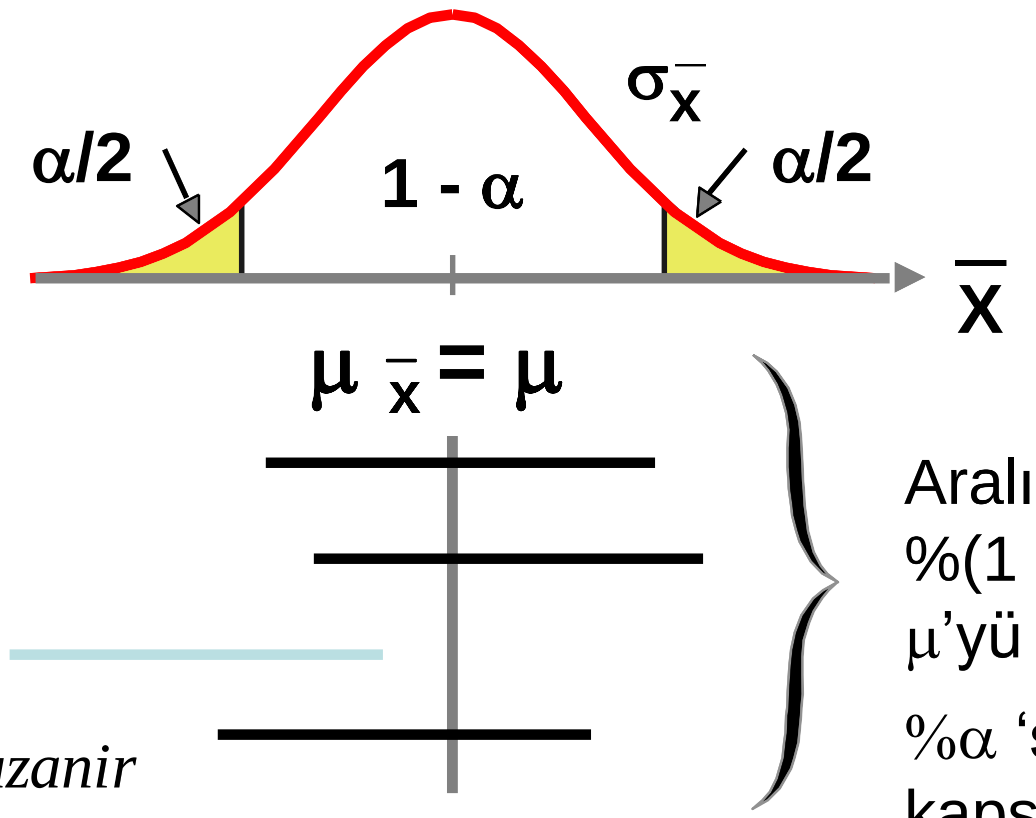
- Bilinmeyen populasyon parametresinin aralık içine düşme olasılığıdır.
- %(1 - α) = güven seviyesi

α : Parametrenin aralık içinde olmaması olasılığıdır.

- Tipik değerler %99, %95, %90

Aralıklar ve güven seviyesi

Ortalamanın örnekleme dağılımı



aralık

$\bar{X} - Z \cdot \sigma_{\bar{X}}$ 'dan

$\bar{X} + Z \cdot \sigma_{\bar{X}}$ 'a kadar uzanir

Aralıkların %(1 - α) 'ı μ'yü kapsar.

%α 'sı kapsamaz.

Çok sayıda aralık

Aralık genişliğini etkileyen faktörler

Aralık

$\bar{X} - Z \cdot \sigma_{\bar{X}}$ 'dan $\bar{X} + Z \cdot \sigma_{\bar{X}}$ 'ya uzanir.

- Verilerin yayılımı (σ)
- Örnek hacmi $\sigma_{\bar{X}} = \dfrac{s_x}{\sqrt{n}}$
- Güven seviyesi (1 - α)

Populasyon ortalamasının güven aralığının hesaplanması

Parametre= istatistik ±hata

(1) $\mu = \overline{X} \pm Hata$

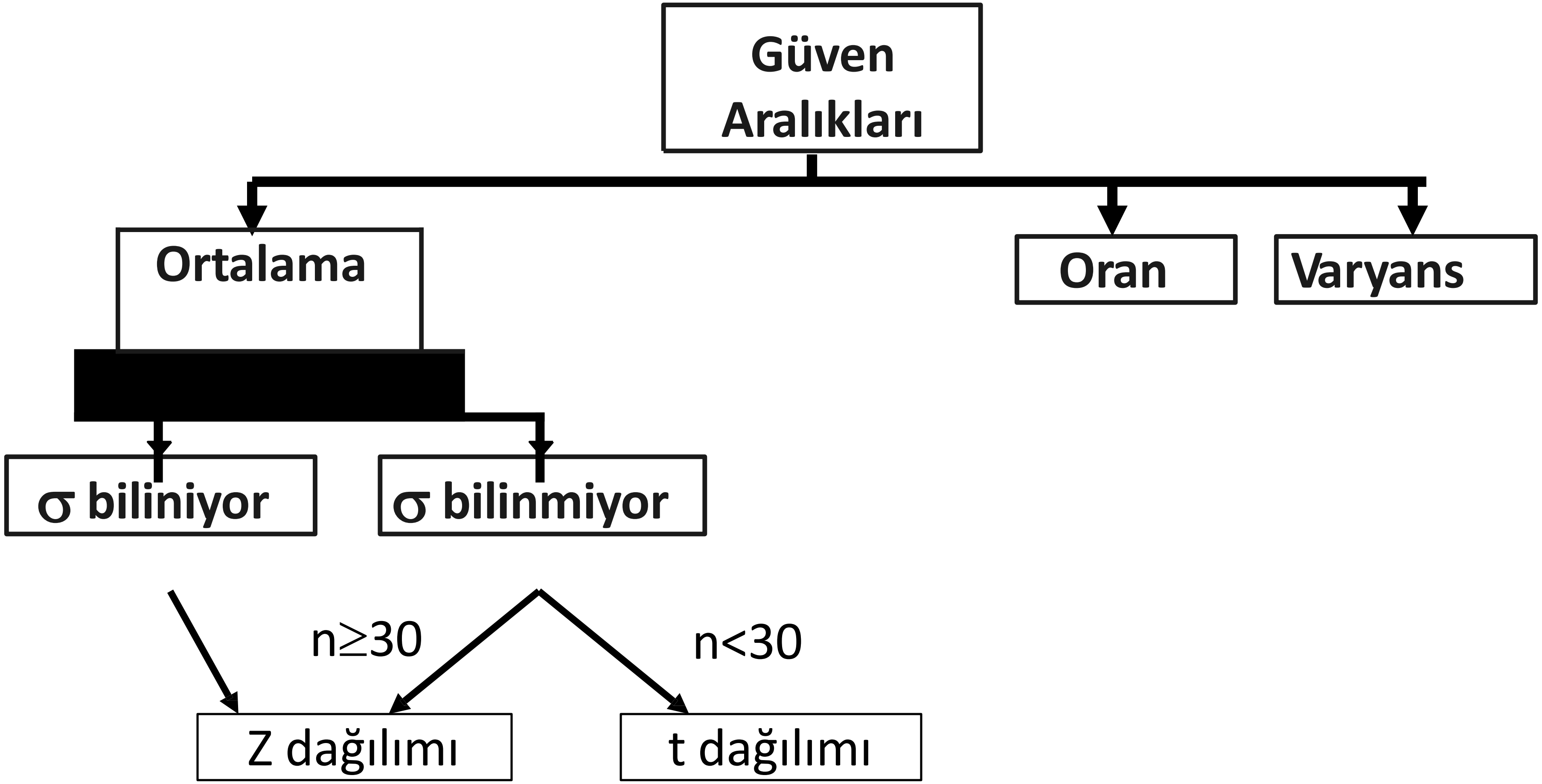
(2) $Hata = \overline{X} - \mu \quad yada \quad \overline{X} + \mu$

(3) $Z = \dfrac{\overline{X} - \mu}{\sigma_{\bar{x}}} = \dfrac{Hata}{\sigma_{\bar{x}}}$

(4) $Hata = Z\sigma_{\bar{x}}$

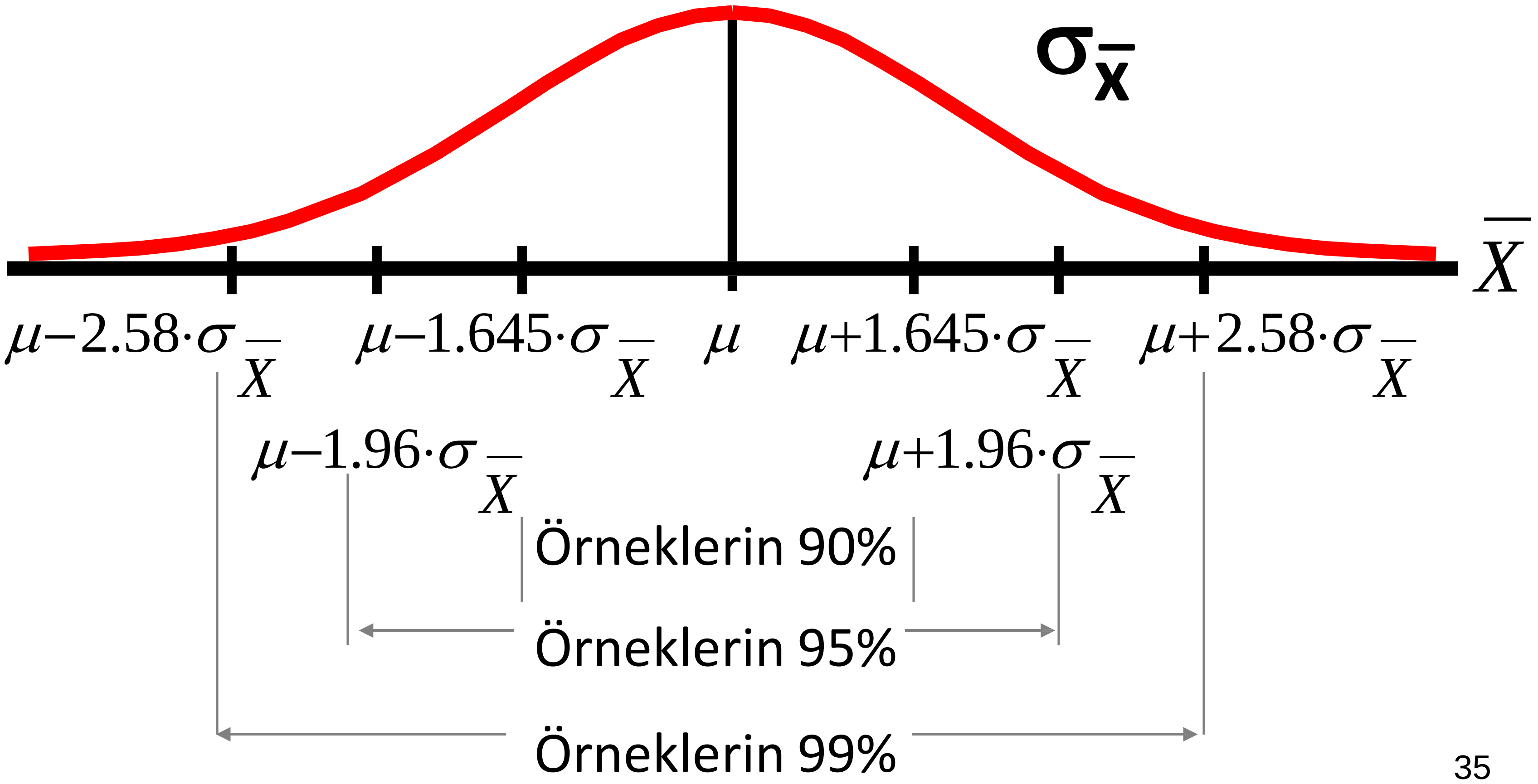
(5) $\mu = \overline{X} \pm Z\sigma_{\bar{x}}$

Güven Aralığı Tahminleri



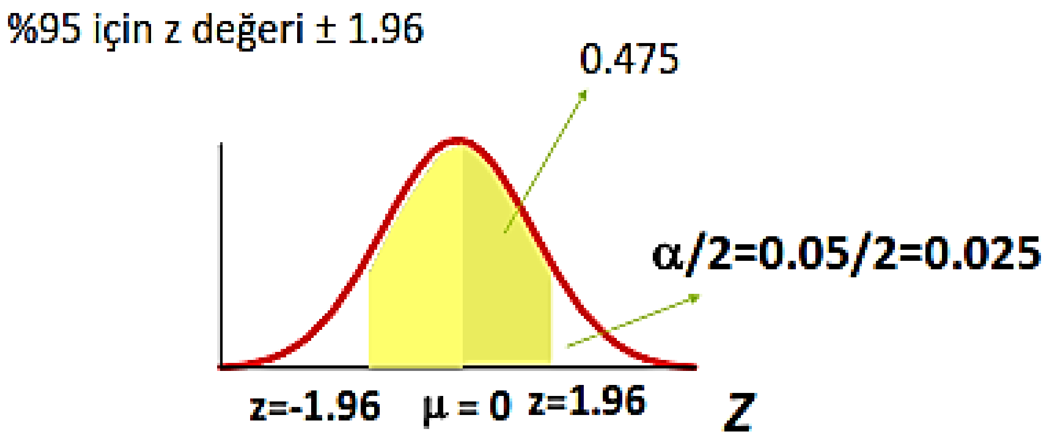
ORTALAMALAR İÇİN GÜVEN ARALIĞI

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} \le \mu_X \le \bar{X} + z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right) = 1-\alpha$$



ORTALAMALAR İÇİN GÜVEN ARALIĞI

Örnek: Bir fabrikada üretilen 100 mamulün ortalama ağırlığı 1040 gr standart sapması 25 gr bulunmuştur. Bu imalat prosesinde üretilen mamullerin ortalama ağırlığı %95 güvenle hangi aralıktadır?



$$P\left(\bar{X} - Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(1040 - 1.96\frac{25}{\sqrt{100}} \leq \mu \leq 1040 + 1.96\frac{25}{\sqrt{100}}\right) = 0.95$$

$$P(1035.1 \leq \mu \leq 1044.9) = 0.95$$

Örnek: n=25 hacimli bir örneklemde, ortalama $\bar{X} = 50$, popülasyon standart sapması $\sigma_x = 10$ ise μ için %95 güven aralığı nedir?

$$P\left(\bar{X} - Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(50 - 1.96\frac{10}{\sqrt{25}} \leq \mu \leq 50 + 1.96\frac{10}{\sqrt{25}}\right) = 0.95$$

$P(46.08 \leq \mu \leq 53.92) = 0.95$

Popülasyon standart sapması σ_x bilinmiyor ve $n \geq 30$ ise:

1. Varsayımlar:
   - POPULASYONUN standart sapması bilinmiyor
   - Populasyon Normal dağılımlıdır.
2. Merkezi limit teoremi kullanılarak Z Dağılımı kullanılır.
3. Güven aralığı tahmini:

Örneğin st.sapması

$$P\left(\bar{X} - Z_{\alpha/2}\frac{S_x}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2}\frac{S_x}{\sqrt{n}}\right) = 1 - \alpha$$

Örnek: Bir ampul şirketi yeni bir ampul geliştirerek piyasaya sürüyor. Üretim bandından 100 tanesi rassal olarak seçiliyor ve bunların standart sapması 140 saat, kulanım süreleri de ortalama olarak 1280 saat bulunuyor. α=0.05 için populasyon ortalamasının güven aralığını bulunuz.

$$P\left(\bar{X} - Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(1280 - 1.96\frac{140}{\sqrt{100}} \le \mu \le 1280 + 1.96\frac{140}{\sqrt{100}}\right) = 0.95$$

$P(1252.56 \le \mu \le 1307.44)$=0.95

Yorum: Şirketin ürettiği ampullerin ortalama ömrü, 0.95 olasılıkla 1252.56 ile 1307.44 saat arasındadır.

Popülasyon standart sapması σ_x bilinmiyor ve $n \leq 30$ ise:

$$P\left(\bar{X} - t_{\alpha/2,(\mathrm{n}-1)}\frac{s_x}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2,(\mathrm{n}-1)}\frac{s_x}{\sqrt{n}}\right) = 1 - \alpha$$

Örnek: Bir araştırmacı, 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:

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

Ortalamanın tahmini standart hatası:

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

Araştırmacı tahmininde %95 emin olmak isterse:

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

$$22.5 - 2.26(1.04) \leq \mu \leq 22.5 + 2.26(1.04)$$

$$20.15 \leq \mu \leq 24.85$$

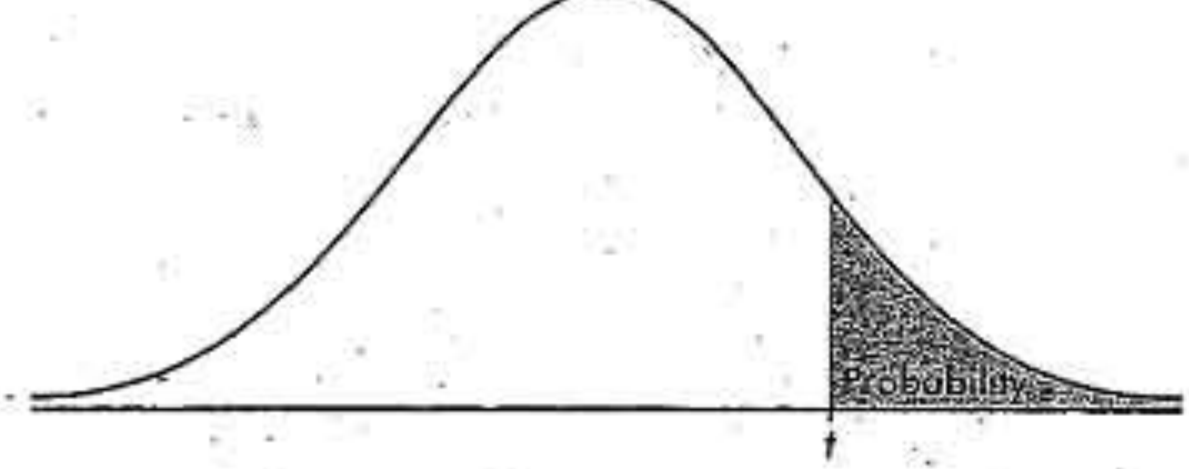


TABLE B: t-DISTRIBUTION CRITICAL VALUES

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587

Bir Oranın Güven Aralığı

- 1. Varsayımları
  - İki kategorik çıktı vardır.
  - Populasyon Binom dağılımı gösterir.
- 2. Güven aralığı tahmini:

$$P(\hat{p} - Z_{\alpha/2}.S_{\hat{p}} \le P \le \hat{p} + Z_{\alpha/2}.S_{\hat{p}}) = 1 - \alpha$$

$$\hat{p} = \frac{x}{n}$$

x → Özellikli birim sayısı

n → Örnek hacmi

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}.\hat{q}}{n}}$$

Bir Oranın Güven Aralığı

ÖRNEK:400 lise öğrencisinden oluşan bir örnekte 32 öğrenci üniversite sınavını kazanmıştır. Üniversite öğrencilerinin sınavı kazanma oranı için %95'lik güven aralığını bulunuz.

$$\hat{p} = {}^{32}/_{400} = 0.08$$

$$P\,(\hat{p} - Z_{\alpha/2}\,.\,S_{\hat{p}} \leq P \leq \hat{p} + Z_{\alpha/2}\,.\,S_{\hat{p}}) = 1 - \alpha$$

$$P\left(0.08 - 1.96\sqrt{\frac{0.08(1-0.08)}{400}} \leq P \leq 0.08 + 1.96\sqrt{\frac{0.08(1-0.08)}{400}}\right) = 0.95$$

$$P(0.053 \leq P \leq 0.107) = 0.95$$