Örnekleme Dağılımı

Tanım: Bir istatistiğin olasılığına örnekleme dağılımı denir.
Örnek: Bir üniversitedeki öğrencilerin ortalama boy uzunluğu μ yü öğrenmek istediğimizde bir örnekleme dayanarak karar vereceğiz. Bu nedenle μ yü tam olarak bilemeyiz. Yapabileceğimiz en iyi şey n birimlik bir örneklemden μ yü tahmin etmektik.

Örnekleme Dağılımı

Kafkas Üniversitesinde ortalama bir öğrencinin boy ortalamasını belirlemeye çalışalım.
Bütün öğrencilerin boylarını ölçüp ortalama değerini hesaplamak mümkün değildir.
Rastgele seçilen 10 öğrencinin boylarını ölçüp kaydedelim.
Bu işlem beş kez tekrarlanıp aşağıda her bir tekrarın ortalama değeri belirtilmiştir.

Örneklem numarası	Boyların ortalamadeğeri (Örnek Ortalaması)
1	1.68
2	1.70
3	1.71
4	1.69
5	1.66

Tablodan görüleceği gibi her bir örnekleme kendisine ait ortalama değerine sahiptir ve birbirlerinden farklıdır.
Bu dağılma örnekleme dağılımı denir.

Örnekleme Dağılımı

Örnekden bir ana kütleyi tahmin etmek için kütle yerine örneklemin kullanılmasından oluşan hataya, yani $X - \mu$ farkına örnekleme hatası denir.

Merkezi Limit Teoremi

Ortalaması μ ve varyansı σ^2 olan herhangi bir anakütleden rastgele çekilen n örnek olsun.

$$Z = \frac{X - \mu}{\frac{\sigma}{\sqrt{n}}}$$

ile tanımlanan Z rastgele değişkeni n nin büyük değerleri için $(n > 30)$ yaklaşık olarak standart normal dağılıma sahiptir.
Bu teoremin bir sonucu olarak;
örnekteki birim sayısı yeterince büyük olduğunda $\overline{X} = N(\mu; \frac{\sigma^2}{n})$ ilişkisi anakütlenin dağılımına bakılmaksızın yazılabilmektedir.

Örnek

Örnek: Bir sanayii kuruluşunda çalışanların gündelikleri $\mu = 800.000$ TL ortalama ve $\sigma = 90.000$ TL standart sapmalıdır. Rastgele seçilen 81 işçinin gündeliklerinin ortalamasının 810.000 TL ve 825.000 TL arasında bulunması olasılığı nedir?

Çözüm: $\mu = 800.000$ ve $\sigma = 90.000$ verilmiştir. $\overline{X}$ in örnekleme dağılımı için $\mu_{\overline{X}} = \mu$ olduğundan $\mu_{\overline{X}} = 800.000$ ve ortalamanın standart hatası $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{90.000}{\sqrt{81}} = 10.000$

$n = 81 > 30$ olduğundan $\overline{X}$ yaklaşık olarak normal dağılıma sahip olur.O halde,

$$Z = \frac{X - \mu}{\frac{\sigma}{\sqrt{n}}}$$

değişkeni kullanılarak,

$z_1 = \frac{810.000-800.000}{10000} = 1$ ve $z_2 = \frac{825.000-800.000}{10000} = 2,5$

z değerlerini elde ederiz. O halde İstenen Olasılık,

$P(810.000 < \overline{X} < 825.000) = P(0 < Z < 2,5) - P(0 < Z < 1) =$

0,1525

İki Ortalama arasındaki fark ve toplam

Herhangi iki anakütleden rastgele çekilen n_1 ve n_2 büyüklükteki örneklerin toplamına (vefarkına) ait değerlerin ortalaması,anakütle ortalamalarının toplam(vefarkına),varyansları ise örnek varyanslarının toplamına eşittir.**Yani,**

$E(\overline{X}_1 \mp \overline{X}_2) = \mu_1 \mp \mu_2$

$V(\overline{X}_1 \mp \overline{X}_2) = \frac{\sigma_1^2}{n_1} \mp \frac{\sigma_2^2}{n_2}$

- Ortalamaların toplam ve farklarının dağılımı ya normaldir ya da yaklaşık olarak normaldir.

$$\overline{X}_1 \pm \overline{X}_2 \sim N\left(\mu_1 \pm \mu_2 ; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \Rightarrow Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- Ortalamaların toplam ve farklarının dağılımı ya normaldir ya da yaklaşık olarak normaldir.
- Bu ifadenin yazılışı ve kullanılacak **Z eşitliği**:

$$\bar{X}_1 \pm \bar{X}_2 \sim N\left(\mu_1 \pm \mu_2; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \Rightarrow Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Örnek:

Kablo üreticisi iki firmanın ürettikleri kabloların kopma mukavemetleri ortalamasının, sırasıyla 200kg/cm^2 ve 180kg/cm^2, standart sapmalarının 13.5kg/cm^2 ve 9 kg/cm^{2}olduğu belirtilmiştir. Bu iddianın doğru olup olmadığını test etmek isteyen tüketici bir firma ilk firmanın üretiminden rastgele 100 parça kablo, ikinci firmanın üretiminden rastgele 50 parça kablo almıştır. Üretici firmaların beyanatlarının doğru olduğu kabul edilirse; birinci ve ikinci firmanın kablolarının kopma mukavemetleri ortalamaları arasındaki farkın;

a) En fazla 17 kg/cm^2 çıkması olasılığı nedir?

b) En az 15 kg/cm^2 çıkması olasılığı nedir?

μ_1=200 kg/cm² μ_2=180 kg/cm² σ_1=13.5 kg/cm² σ_2=9 kg/cm² n_1=100 n_2=50

$$P((\bar{X}_1 - \bar{X}_2) < 17) = P\left(Z < \frac{17-20}{\sqrt{13.5^2/100 + 9^2/50}}\right) = P(Z < -1.61) = 0.0537$$

$$P((\bar{X}_1 - \bar{X}_2) > 15) = P\left(Z > \frac{15-20}{1.86}\right) = P(Z > -2.69) = 1 - P(Z < -2.69) = 0.9964$$

Tahmin Teorisi

Tahmin: Mevcut bilgi ve deneylere dayanarak olayın bütünü hakkında bir yargıya varmaktır.
Bu anlamda, anakütleden çekilen örnek verileri üzerinde istatistiksel yöntemler uygulanarak elde edilen sonuçlar anakütleye genelleştirilir.
Bu kapsamda uygulanan yöntemler karar teorisi içinde incelenir.
Karar teorisi:

1. Tahminler (örnek verilerinden hareketle parametre değerlerini tahmin etme)
2. Testler (örnek verilerinden hareketle tahmin edilen parametre değerleri hakkında karar verme)

Tahminler; nokta tahmini ve aralık tahmini olmak üzere iki başlık altında incelenmektedir.

Nokta Tahmini

Örneklerden hesaplanan ortalama ve varyans gibi istatistiksel değerler anakütle parametrelerinin nokta tahminleridir.

- Nokta tahminlerinin anakütle parametrelerine eşit olmaları beklenemez.

Nokta Tahmini

Örneklerden hesaplanan ortalama ve varyans gibi istatistiksel değerler anakütle parametrelerinin nokta tahminleridir.

- Nokta tahminlerinin anakütle parametrelerine eşit olmaları beklenemez.
- Belirli bir hata veya sapma her zaman için söz konusudur.

Anakütlenin Aritmetik Ortalamasının Nokta Tahmini

Örneklem oluşturulduktan sonra hesaplanan $\overline{X}$ ortalama değeri ana kütlenin ortalaması olan μ nün nokta tahminidir.
N birimlik bir ana kütleden n birimlik örnrklemler $\binom{N}{n}$ farklı biçimde seçilebileceğinden $\binom{N}{n}$ tane nokta tahmini söz konusudur.

Aralık Tahmini

- Nokta tahmininin belirli bir hata payı ile anakütle parametresine yakınsama derecesinin tespit edilmesi oldukça önemlidir.

Aralık Tahmini

- Nokta tahmininin belirli bir hata payı ile anakütle parametresine yakınsama derecesinin tespit edilmesi oldukça önemlidir.
- Nokta tahminini kullanarak anakütle parametresini belirli bir olasılıkla içinde bulunduracağı alt ve üst sınırları gösteren güven sınırları veya güven aralığı tahminleri yapılmaktadır.

Aralık Tahmini

- Nokta tahmininin belirli bir hata payı ile anakütle parametresine yakınsama derecesinin tespit edilmesi oldukça önemlidir.
- Nokta tahminini kullanarak anakütle parametresini belirli bir olasılıkla içinde bulunduracağı alt ve üst sınırları gösteren güven sınırları veya güven aralığı tahminleri yapılmaktadır.
- Tahminde yapılabilecek hata seviyesi α ile gösterilirse, $1-\alpha$ tahminin doğruluk seviyesini (güven düzeyini) gösterilebilir.

Aralık Tahmini

- Nokta tahmininin belirli bir hata payı ile anakütle parametresine yakınsama derecesinin tespit edilmesi oldukça önemlidir.
- Nokta tahminini kullanarak anakütle parametresini belirli bir olasılıkla içinde bulunduracağı alt ve üst sınırları gösteren güven sınırları veya güven aralığı tahminleri yapılmaktadır.
- Tahminde yapılabilecek hata seviyesi α ile gösterilirse, $1-\alpha$ tahminin doğruluk seviyesini (güven düzeyini) gösterilebilir.
- $1-\alpha$ ile gösterilen güven düzeyi için genellikle %99 veya %95, nadiren de %90 değerleri (bu durumda hata payları, seçilen güven düzeyine bağlı olarak, sırasıyla %1, %5 ve %10 olacaktır) esas alınmaktadır.

Hata terimi normal dağılım eğrisinin her iki ucunda eşit $\alpha/2$ olarak yer alır. $\alpha/2$ 'lik hata seviyesine karşı gelen tablo değeri $Z_{\alpha/2}$, ilgili dağılımın standart hatası ile çarpılarak aralığın
alt ve üst sınırlarını belirlemede kullanılacak olan hata terimi belirlenmiş olur.

Gerek tahminlerde, gerekse hipotez testlerinde işlemler parametre (anakütleye ait gösterge) ve tahminin (örneğe ait gösterge) dağılım biçimine göre yürütülür. Özellikle normal dağılım gösteren verilerden elde edilen tahminlerde:

- Z dağılım

kullanılmaktadır.

Gerek tahminlerde, gerekse hipotez testlerinde işlemler parametre (anakütleye ait gösterge) ve tahminin (örneğe ait gösterge) dağılım biçimine göre yürütülür. Özellikle normal dağılım gösteren verilerden elde edilen tahminlerde:

- Z dağılım
- t dağılım

kullanılmaktadır.

Gerek tahminlerde, gerekse hipotez testlerinde işlemler parametre (anakütleye ait gösterge) ve tahminin (örneğe ait gösterge) dağılım biçimine göre yürütülür. Özellikle normal dağılım gösteren verilerden elde edilen tahminlerde:

- Z dağılım
- t dağılım
- ki kare dağılım

kullanılmaktadır.

Güven aralığı tahminleri

Güven aralıkları ve hipotez testlerinde kullanılacak dağılım:

- ilgilenilen parametreye ait anakütle varyansının bilinip bilinmemesine

Kullanılacak dağılım aşağıdaki ilkelere göre belirlenir:

Güven aralığı tahminleri

Güven aralıkları ve hipotez testlerinde kullanılacak dağılım:

- ilgilenilen parametreye ait anakütle varyansının bilinip bilinmemesine
- örnek büyüklüğüne bağlı olarak belirlenmektedir.

Kullanılacak dağılım aşağıdaki ilkelere göre belirlenir:

Güven aralıkları ve hipotez testlerinde kullanılacak dağılım:

- ilgilenilen parametreye ait anakütle varyansının bilinip bilinmemesine
- örnek büyüklüğüne bağlı olarak belirlenmektedir.

Kullanılacak dağılım aşağıdaki ilkelere göre belirlenir:

- Anakütle varyansı σ^2 biliniyorsa Z dağılımı

Güven aralığı tahminleri

Güven aralıkları ve hipotez testlerinde kullanılacak dağılım:

- ilgilenilen parametreye ait anakütle varyansının bilinip bilinmemesine
- örnek büyüklüğüne bağlı olarak belirlenmektedir.

Kullanılacak dağılım aşağıdaki ilkelere göre belirlenir:

- Anakütle varyansı σ^2 biliniyorsa Z dağılımı
- Anakütle varyansı σ^2 bilinmiyorsa n$\geq$30 ise Z dağılımı n$<$30 ise t dağılımı

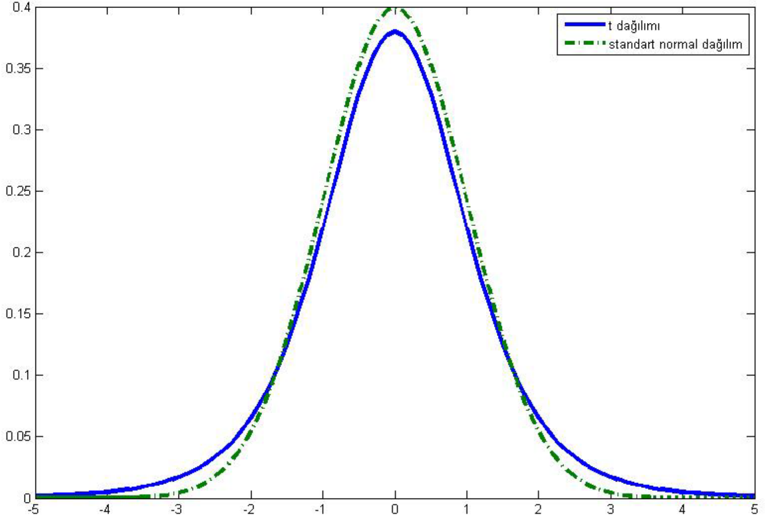
Güven aralığı tahminleri

Genellikle, büyük örneklere ($n \geq 30$) ait örnek dağılımlarının yaklaşık olarak normal dağılım gösterdiği ve n büyüdükçe normale daha fazla yakınsadığı bilinmektedir.
Küçük örneklere ($n < 30$) ait örnek dağılımları normal dağılımdan uzaklaşmaktadır. Bu uzaklaşma n küçüldükçe daha da fazlalaşmaktadır.
Bu nedenle, büyük örnekler için Z dağılımı kullanılırken, küçük örnekler için Z dağılımı yerine t (student) dağılımını kullanmak gerekmektedir.

t dağılımı

Küçük örnek istaistiklerinin gösterdiği dağılım normal dağılım eğrisi gibi simetrik bir görünümde olmakla birlikte,normal dağılım eğrisine göre daha basık ve yayvan biçimdedir.Yayvanlıktan dolayı t dağılımı eğrisinin kuyrukları altında kalan alan Z dağılımına göre daha büyüktür.

t dağılımı


t dağılımı
standart normal dağılım
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-5
-4
-3
-2
-1
0
1
2
3
4
5

t dağılımı

- Küçük örnekler için Z tablosu yerine farklı örnek büyüklükleri ve önem seviyeleri esas alınarak hesaplanan t tabloları kullanılır.

t dağılımı

- Küçük örnekler için Z tablosu yerine farklı örnek büyüklükleri ve önem seviyeleri esas alınarak hesaplanan t tabloları kullanılır.
- n$\geq$30   için t tablosu değeri Z tablosu değerine çok yaklaşır.

t dağılımı

- Küçük örnekler için Z tablosu yerine farklı örnek büyüklükleri ve önem seviyeleri esas alınarak hesaplanan t tabloları kullanılır.
- $n \geq 30$ için t tablosu değeri Z tablosu değerine çok yaklaşır.
- Bu sebeple $n \geq 30$ olan örneklerde t tablosu yerine Z tablosu kullanılmalıdır.

Ortalamanın güven aralığı

Anakütle Varyansı Biliniyorsa

- Bir örnekten elde edilen $\overline{X}$, anakütle ortalaması μ nün nokta tahminidir.

$$P(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}) = 1 - \alpha \longrightarrow Z = \frac{X-\mu}{\frac{\sigma}{\sqrt{n}}} \longrightarrow \sigma_x = \frac{\sigma}{\sqrt{n}} \longrightarrow Z = \frac{X-\mu}{\sigma_x}$$
$$P(-Z_{\frac{\alpha}{2}} < \frac{X-\mu}{\sigma_x} < Z_{\frac{\alpha}{2}}) = 1 - \alpha$$

Ortalamanın güven aralığı

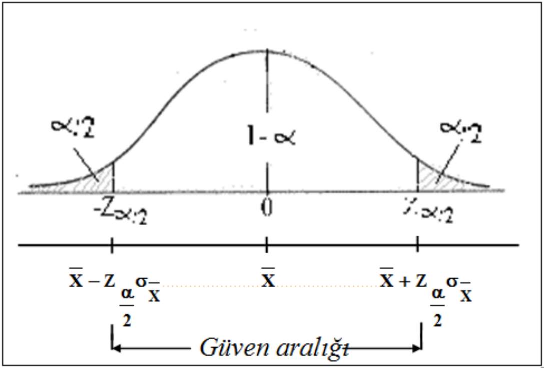
Anakütle Varyansı Biliniyorsa

- Bir örnekten elde edilen $\overline{X}$, anakütle ortalaması μ nün nokta tahminidir.
- Nokta tahmininin anakütle değerine eşit olması beklenemez. Bunun için anakütle ortalaması μ 'yü içinde bulunduracak $1-\alpha$ güven düzeyindeki aralık tahmini aşağıdaki gibi yapılır:

$$P(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}) = 1-\alpha \longrightarrow Z = \frac{X-\mu}{\frac{\sigma}{\sqrt{n}}} \longrightarrow \sigma_x = \frac{\sigma}{\sqrt{n}} \longrightarrow Z = \frac{X-\mu}{\sigma_x}$$

$$P(-Z_{\frac{\alpha}{2}} < \frac{X-\mu}{\sigma_x} < Z_{\frac{\alpha}{2}}) = 1-\alpha$$

$P(\overline{X} - Z_{\frac{\alpha}{2}}\sigma_{\overline{X}} < \mu < \overline{X} + Z_{\frac{\alpha}{2}}\sigma_{\overline{X}}) = 1 - \alpha$
ifadesi elde edilir ve ortalamanın güven aralığı olarak adlandırılır.



Ortalamanın güven aralığı

Örnek: Bir tezgahta üretilen parçaların dış çaplarının standart sapması σ=2.4 cm'dir. Tezgahın üretiminden rastgele seçilen 16 parçanın dış çap ortalaması 3.2 cm olarak bulunmuştur. %5 hata (%95 güven) seviyesinde anakütle ortalamasının güven aralığını tahmin ediniz.

Çözüm:

σ=2.4 cm n=16 parça, $\overline{X}$ =3.2 cm, 1-α=0.95 ise α=0.05 $\alpha/2$=0.025

Z tablosundan $Z_{\frac{\alpha}{2}} = Z_{\frac{0.05}{2}} = Z_{0.025} = 1.96$ değeri alınır.

$$P(\overline{X} - Z_{\frac{\alpha}{2}}\sigma_{\overline{X}} < \mu < \overline{X} + Z_{\frac{\alpha}{2}}\sigma_{\overline{X}}) = 1 - \alpha$$

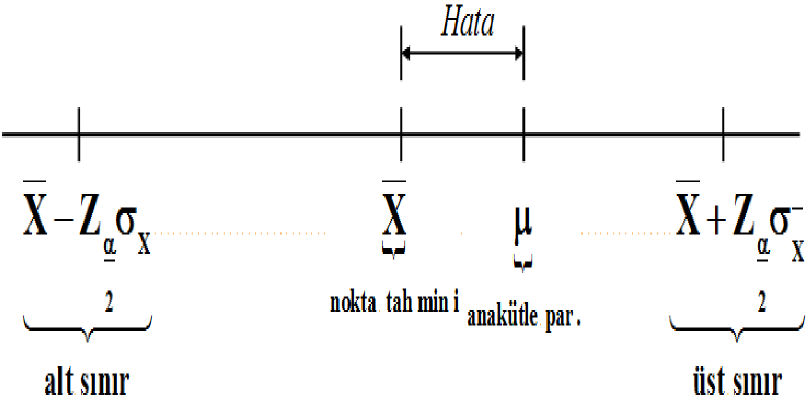
$$3.2 - 1.96\sqrt{\frac{(2.4)^2}{16}} < \mu < 3.2 + 1.96\sqrt{\frac{(2.4)^2}{16}}$$

$$2.024 < \mu < 4.376$$

Örnek Büyüklüğü

Az sayıda örneğin incelenmesi ile ulaşılan nokta tahmininin anakütle parametresine eşit olması beklenemez.
Belirli bir sapma her zaman için söz konusu olabilir.
Sapmanın büyüklüğü anakütle parametresi (örneğin μ) ile nokta tahmini (örneğin $\overline{X}$) arasındaki fark kadar olacaktır.
Sözü edilen fark büyük ise (örneğin μ nün değeri$\overline{X}$'a dayalı olarak oluşturulan güven aralığının sınırlarına yakın ise) hata miktarı maksimum düzeyde olacaktır.

Örnek Büyüklüğü



Örnek Büyüklüğü

Örnek hacmi (n) arttırılarak μ nün tahmininde yapılan hata miktarı azaltılabilir.
Bu amaçla ortalamanın güven aralığı oluşturulurken yapılabilecek hatanın belirlenen bir değerden az olması için alınması gereken örnek sayısı aşağıdaki formül yardımıyla belirlenebilir:
$n = \left(\frac{Z_{\alpha/2}\sigma}{d}\right)$ ve $d = \overline{X} - \mu$

Anakütle Varyansı Bilinmiyorsa

- Anakütle varyansının bilinmediği, fakat örnek hacminin 30 veya daha büyük olduğu (n$\geq$30) durumlarda örnek varyansı kullanılarak Z dağılımı yardımıyla güven aralığı

oluşturulur.

geliştirilen t dağılımı yardımıyla güven aralığı oluşturulur.

Ortalamanın güven aralığı

Anakütle Varyansı Bilinmiyorsa

- Anakütle varyansının bilinmediği, fakat örnek hacminin 30 veya daha büyük olduğu (n$\geq$30) durumlarda örnek varyansı kullanılarak Z dağılımı yardımıyla güven aralığı

oluşturulur.

- Anakütle varyansının bilinmediği durumlarda örnek hacmi 30 dan küçük (n$<$30) ise küçük örnek teorisine göre

geliştirilen t dağılımı yardımıyla güven aralığı oluşturulur.

Ortalamanın güven aralığı

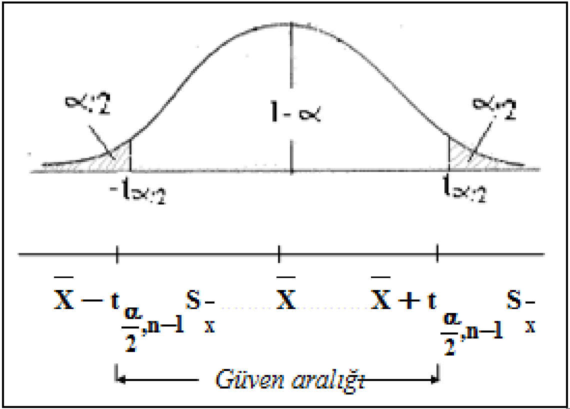
Anakütle varyansının bilinmediği ve n<30 olduğu durumlarda anakütle ortalaması μ yü içinde bulunduracak $1-\alpha$ güven düzeyindeki aralık tahmini aşağıdaki gibi yapılır:

$P(-t_{\frac{\alpha}{2},n-1} < t < t_{\frac{\alpha}{2},n-1}) = 1-\alpha \longrightarrow t = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \longrightarrow S_x = \frac{S}{\sqrt{n}} \longrightarrow Z = \frac{X-\mu}{S_x}$

$P(-t_{\frac{\alpha}{2},n-1} < \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} < t_{\frac{\alpha}{2},n-1}) = 1-\alpha$

$P(\overline{X} - t_{\frac{\alpha}{2},n-1} S_{\overline{X}} < \mu < \overline{X} + t_{\frac{\alpha}{2},n-1} S_{\overline{X}}) = 1-\alpha$

Ortalamanın güven aralığı



Ortalamanın güven aralığı

Örnek: Bir işyerinde çalışan işçilerin boylarına göre tezgah yüksekliklerinin ayarlanması amacıyla bir araştırma yürütülmüştür. Farklı bölümlerden rasgele 25 işçi seçilmiş ve boyları ölçülmüştür. İşçilerin boyları ortalaması 1.72 m ve varyansı 0.18 olarak belirlendiğine göre %99 güven (%1 hata) seviyesinde anakütle ortalamasının güven sınırlarını tahmin ediniz.

SD=n-1=25-1=24 $\quad 1-\alpha=0.99 \rightarrow \alpha=0.01 \quad \alpha/2=0.005 \quad t_{\alpha/2,n-1}=t_{0.005;24}=2.797$

$$\underbrace{1.72-2.797\cdot\sqrt{\frac{0.18}{25}}}_{\text{aralığın alt sınırı}} < \mu < \underbrace{1.72+2.797\cdot\sqrt{\frac{0.18}{25}}}_{\text{aralığın üst sınırı}} \Rightarrow 1.48 < \mu < 1.96$$

t-tablosunun okunuşu

İstenen hata (yani α veya $\alpha/2$) düzeyinin değeri tablonun yatay eksenindeki Pr kısmına işaretlenir. Serbestlik derecesi (yani SD=n-1) değeri düşey sütundaki SD kısmına işaretlenir.
Yatay ve düşey eksenlerde işaretlenen değerlerin kesiştiği hücrede bulunan değer aranan t tablosu olasılık değeridir.
Bu probleme ait $\alpha/2$=0.005 değeri yatay eksene, SD=24 değeri düşey eksene işaretlenir ve tablodan ilgili olasılık:
$t_{\alpha/2,n-1} = t_{0.005;24} = 2.797$

SD \ Pr	0.25	0.10	0.05	0.025	0.01	**0.005**	0.001
1							
2							
3							
4							
...							
...							
...							
24						2.797	
...							
...							
30							
...							
120							
∞							