BÖLÜM 10

# TAHMİN TEORİSİ

Daha önce tamsayım, örnek, örnekleme ve parametre kavramlarından sözedilmişti. Hem kısaca sayım olarak adlandırılan tam sayım, hem de örnekleme ile ulaşılmak istenen, anakütle parametrelerinin değerlerinin belirlenmesidir. Tam sayım sözkonusu olduğunda elde edilen değerler anakütle parametrelerinin gerçek değerleridir. Burada ancak ölçme, değerlendirme, işlem hataları gibi hatalar yapılmış olabileceği düşünülebilir. Anakütle parametrelerinin değerlerinin örnekleme ile belirlenmesi sözkonusu olduğunda tahmin olayı ile karşılaşılmaktadır.

Anakütle parametresi için örnekten bir tek değer tahmin edilmesi, nokta tahmini olarak adlandırılmaktadır. Daha sonra sözedileceği gibi bu tahmin örnekleme hatası içerebilir. Belirlenen bir olasılıkla örnekleme hatasının boyutları tahmin edilebilir ki, bu aralık tahminidir. Burada önce nokta tahmini, daha sonra ise aralık tahmini açıklanacaktır.

## 10.1. TAHMİN VE TAHMİNCİ

Birim sayısı N olan bir anakütleden birim sayısı n olan farklı örnekler alınabilir. Seçilebilecek farklı örnek sayısı, N ve n'in sayısal büyüklüğü yanında örneği oluşturacak birimlerin seçiminin iadeli veya iadesiz olarak yapılmasına da bağlıdır. Düzenlenebilecek farklı örnek sayısını k ile ifade edersek, iadeli seçim için,

$$k = N^n$$

ve iadesiz seçim için,

$$k = C_N^n$$

olacaktır.

Anakütle parametresinin belirlenmesi için uygulamada bu örneklerden biri seçilecektir. Seçilebilecek bu örneklerden bazılarının parametrelerinin değeri anakütle parametresine eşit, bazılarının ki anakütle parametresinden küçük ve bazılarının ki anakütle parametresinden büyük olacaktır. Bu nedenle örnekleme söz konusu olduğunda anakütle parametresinin örnekten tahmin edilmesi gerekmektedir. Anakütlenin örnekten tahmin edilen parametresi “örnek istatistiği” veya kısaca “istatistik” olarak adlandırılmaktadır.

Olaya mantıksal olarak bakılırsa, örnek birim sayısının artması ile daha doğru tahminlerin yapılması beklenecektir. Örnek birim sayısının anakütle birim sayısına oranı “örnekleme oranı” olarak adlandırılır.

$$\text{Örnekleme Oranı} = \frac{n}{N}$$

Örnekleme oranının artması yani 1’e doğru yaklaşması durumunda yapılacak tahminlerin daha doğru olması mümkün olabilir.

Daha doğru tahminler elde edilmesi anakütle birimlerinin homojenliği ile de ilgilidir. Homojenlik birimlerin değerlerinin birbirine yakınlığını ifade etmektedir. Anakütle ne kadar homojense, anakütleyi oluşturan birimlerin değerleri birbirine ne kadar yakınsa, yapılacak tahmin gerçeğe o kadar yakın olacaktır.

Tahminin doğruluğu, tahminde kullanılacak formüle de bağlıdır. Burada cevabı araştırılması gerekli soru en doğru tahminin hangi formülle yapılması gerektiğidir. En iyi sonucu anakütle parametresini hesapladığımız formülün mü yoksa farklı bir formülün mü vereceği belirlenmelidir. Örneğin, anakütle ortalaması veya varyansı daha önce verilen ve anakütle için kullanılan formüllerle mi, yoksa farklı formüllerle mi tahmin edilirse, daha doğru sonuç elde edilir. Diğer bir ifade ile anakütle ortalaması ve varyansı örnek ortalaması ve varyansına eşit olarak mı tahmin edilecektir.

Anakütle parametrelerinin tahmini için kullanılan formüllere tahminci adı verilmektedir. Yukarıda söz ettiğimiz soruların cevapları, ancak tahmincilerin özelliklerinin belirlenmesi ve tahmincilerde bu özelliklerin aranması ile açıklık kazanacaktır. Tahmincinin özellikleri daha sonra incelenecektir.

## 10.2. ÖRNEKLEME DAĞILIMI VE STANDART HATA

Açıklandığı gibi N birimli bir anakütleden n birimli k sayıda örnek düzenlenebilmektedir. Alınacak örnek ile tahmin edilecek anakütle parametresi $\theta$; kullanılacak tahminci $\hat{\theta}$ ise, seçilecek örneğe bağlı olarak k adet farklı tahmin yapılabilecektir.

| Örnek No | Tahmin |
|---|---|
| 1 | $\hat{\theta}_1$ |
| 2 | $\hat{\theta}_2$ |
| 3 | $\hat{\theta}_3$ |
| . | . |
| . | . |
| k | $\hat{\theta}_k$ |

$\hat{\theta}$ 'lerin beklenen değeri anakütle parametre θ'nın gerçek değerine eşit olacaktır.

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

$\hat{\theta}$ 'lerin standart sapması ise standart hata olarak adlandırılır. $\hat{\theta}$ 'lerin varyansı,

$$Var(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2$$

olacağından, bunun karekökü standart hataya eşittir. Örneğin, örnek ortalaması için standart hata,

$$Var(\overline{X}) = E(\overline{X} - \mu)^2$$

ifadesinin kareköküdür.

## 10.3. SAPMA, ÖRNEKLEME HATASI VE ORTALAMA HATA KARE

Anakütle parametresi θ ile $\hat{\theta}$ 'lerin beklenen değeri birbirine eşit olmayabilir. Bu durumda aradaki fark sapma olarak adlandırılır.

$$\text{Sapma} = E(\hat{\theta}) - \theta$$

Örnekleme hatası ise sapmadan farklı bir kavramdır ve tahminci ile anakütle parametresinin gerçek değeri arasındaki farktır.

$$\text{Örnekleme Hatası} = \hat{\theta} - \theta$$

Daha önce bir anakütleden düzenlenebilecek k sayıdaki farklı örneklerin bazılarında $\hat{\theta} = \theta$, bazılarında $\hat{\theta} < \theta$, bazılarında $\hat{\theta} > \theta$ olabileceğinden sözetmiştik. Örnekleme θ'yı belirlemek için yapıldığından uygulamada θ bilinmeyecektir. Bu nedenle örnekleme hatası da kesin olarak belirlenemeyecektir. Ancak örnekleme hatasının büyüklüğü belirlenecek bir olasılıkla tahmin edilebilecektir.

Açıkladığımız gibi farklı örneklerden tahmin edilecek $\hat{\theta}$ 'ler bir dağılım oluştururlar. Bu dağılım için varyansa benzer bir ölçü kullanılmaktadır. Ortalama hata kare olarak adlandırılan bu ölçü, $\hat{\theta}$ 'ler ile θ arasındaki farkların karelerinin beklenen değeridir. Bu tanıma göre ortalama hata kare,

$$OHK(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

olarak ifade edilir. Bu ifadeye $\pm E(\hat{\theta})$ eklenerek açıklanırsa,

$$OHK(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta}) - \theta]$$

bulunur. Eşitliğin sağ tarafındaki ilk ifade $\hat{\theta}$ 'in varyansına, ikinci ifade ise sapmanın karesine eşittir.

$$\text{OHK}(\hat{\theta}) = \text{Var}(\hat{\theta}) + (\text{sapma})^2$$

Görüldüğü gibi ortalama hata kare varyansa benzer, fakat farklı bir kavramdır. Varyans aritmetik ortalama etrafındaki dağılımın, ortalama hata kare ise, anakütle parametresinin gerçek değeri etrafındaki dağılımın ölçüsüdür. Ancak sapmanın sıfır olması durumunda,

$$\text{Var}(\hat{\theta}) = \text{OHK}(\hat{\theta})$$

olacaktır.

## 10.4. TAHMİNCİNİN ÖZELLİKLERİ

Anakütle parametresi $\theta$'nın tahmini için alınacak n birimli örnekten farklı tahminciler kullanılarak farklı tahminler yapılabilir. Burada önemli olan en iyi sonucu verecek tahmincinin kullanılmasıdır.

En iyi tahmini sağlayacak tahmincinin bazı özellikleri taşıması gerekmektedir. Bu özellikler şunlardır.

-Sapmasızlık

-Etkinlik

-Tutarlılık

-Yeterlilik

Sırası ile bu özellikleri açıklayalım.

### 10.4.1. Sapmasızlık

Sapma $\hat{\theta}$ 'in beklenen değeri ile $\theta$ arasındaki fark olarak tanımlanmıştı. Buna göre $\hat{\theta}$ 'in $\theta$'nın sapmasız tahmincisi olabilmesi için $E(\hat{\theta})$ ile $\theta$ arasındaki fark sıfır olmalıdır. Diğer bir ifade ile,

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

olmalıdır. $\hat{\theta}$ 'in beklenen değeri alındığında sonuç $\theta$, ise, $\hat{\theta}$ $\theta$'nın sapmasız tahmincisidir.

**ÖRNEK:** Anakütle ortalaması $\mu$'nün tahmini için örnek ortalaması $\overline{X}$ kullanıldığında, $\overline{X}$ 'nın $\mu$'nün sapmasız tahmincisi olup olmadığını inceleyelim.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

olduğundan, $\overline{X}$ 'nın beklenen değeri alınacaktır. Eğer $\overline{X}$ 'nın beklenen değeri μ'ye eşit ise, örnek ortalamasının anakütle ortalamasının sapmasız tahmincisi olduğunu söyleyebiliriz. Aksi halde anakütle ortalamasını tahmin etmek için farklı tahminciler veya farklı formüller kullanılacaktır.

$$E(\overline{X}) = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

$$= \frac{1}{n} nE(X_i) = E(X_i) = \mu$$

Örnek ortalaması anakütle ortalamasının sapmasız tahmincisidir.

### 10.4.2. Etkinlik

Anakütlenin bilinmeyen bir parametresinin tahmini için birden fazla sapmasız tahminci belirlenebilir. Bu durumda, bu tahmincilerden hangisinin kullanılması ile daha iyi tahmin yapılabileceğini belirlemek için tahmincilerin varyansları belirlenir. Kısaca etkinlik, sapmasız tahmincilerin varyansları ile ilgili bir kavramdır.

Bilindiği gibi, birden fazla serinin ortalamaları birbirine yakın veya eşitse dağılımların birbirinden farkını ortaya koymak için bazı ölçüler hesaplanır. Bu ölçülerden en çok kullanılanı ise varyans veya bunun karekökü olan standart sapmadır. Varyansı ve standart sapması küçük olan seride aritmetik ortalama etrafında daha yoğun bir dağılım olduğu anlaşılır. Yani, varyansı küçük olan seride seri birimleri, varyansı büyük olan seriye göre aritmetik ortalamaya daha yakın bir dağılım göstermişlerdir. Varyans ve standart sapma tek başına hesaplandığında fazla bir anlam ifade etmeyip, ancak başka varyans veya standart sapmalar ile karşılaştırıldığında anlam kazanmaktadır. Kısaca varyans nisbi bir ölçüdür.

Etkinlik kavramı ile ilgili olarak varyans ve standart sapma ile ilgili açıklamalarımıza benzer bir açıklama yapabiliriz. Burada konumuz anakütle parametresinin gerçek değerine daha yakın tahminler elde etmek olduğuna göre, bunun ölçüsü daha önce açıklanan ortalama hata karedir. Birden fazla sapmasız tahminci sözkonusu olduğunda, bunlardan ortalama hata karesi küçük olan tahminci anakütle parametresinin gerçek değeri etrafında daha yoğun bir dağılım gösterdiğinden, bu tahmincinin tercih edilmesi gerekmektedir. Ortalama hata karenin varyans ile sapmanın karesi toplamına eşit olduğu hatırlanırsa, sapmasız tahminciler sözkonusu olduğundan ortalama hata kareler varyanslara eşit olacaktır. bu nedenle varyansı küçük olan tahmincinin tercih edilmesi ile ortalama hata karesi küçük olan tahminci tercih edilmiş olacaktır.

Yaptığımız bu açıklamalar etkinlik için iki şart aranması gerektiğini ortaya koymaktadır.

1) $\hat{\theta}$, θ'nın sapmasız tahmincisi ise,

2) Var($\hat{\theta}$) ≤ Var($\hat{\theta}$) ise

$\hat{\theta}$, θ'nın etkin tahmincisidir. Burada $\hat{\theta}$, anakütle parametresi θ'nın $\hat{\theta}$'ten farklı diğer sapmasız tahmincilerini ifade etmektedir.

Etkinlik ile ilgili olarak buraya kadar yapılan açıklamalar, etkinliğin nisbi bir kavram olduğunu ortaya koymaktadır. Ancak birden fazla sapmasız tahminci olması durumunda etkinlikten sözedilebilir ve varyansı küçük olan tahmincinin daha etkin olduğu söylenir.

**ÖRNEK:** Anakütle ortalamasının tahmini için örnek aritmetik ortalaması ve örnek medyanının kullanıldığını düşünelim. Her ikisi de anakütle ortalamasının sapmasız tahmincisidir. Örnek ortalamalarının dağılımının varyansının,

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

olduğu daha önce gösterilmişti. Benzer şekilde örnek medyanlarının örnekleme dağılımlarının varyansı,

$$Var(M_e) = \frac{\pi\sigma^2}{2n}$$

olarak elde edilir. Bu tahmincilerden daha küçük varyansa sahip olan daha etkin olacağından varyansların karşılaştırılması gerekecektir.

$$Var(M_e) = Var(\overline{X})\frac{\pi}{2}$$

olduğundan $Var(\overline{X}) < Var(M_e)$'dır ve $\overline{X}$ daha etkin tahmincidir.

### 10.4.3. Tutarlılık

a küçük bir sayı olsun. Eğer n→ ∞ için,

$$P\left(\left|\hat{\theta} - \theta\right| \leq a\right) \to 1$$

ise $\hat{\theta}$, θ'nın tutarlı tahmincisidir. Burada ifade edilen örnek birim sayısının artması ile $\hat{\theta}$ ile θ arasında varolduğunu kabul ettiğimiz küçük farkın azalması olasılığının bire yaklaşmasıdır. Eğer örnek birim sayısının artması ile tahminci ile anakütle parametreleri birbirine yaklaşıyorsa tahminci tutarlıdır.

Örnek birim sayısının artması durumunda sapma ile örnekleme dağılımının varyansı küçülecektir. Böylece hata kare küçülmüş olacaktır. Eğer $n \to \infty$ ise ortalama hata kare sıfıra eşit olacaktır.

$$\lim_{n\to\infty} OHK(\hat{\theta}) = 0$$

Tahmincinin ortalama hata karesinin n sonsuza giderken limiti alınırsa, sonucun sıfır çıkması tahmincinin tutarlı olduğunu gösterecektir. Fakat n sonsuza giderken ortalama hata karesinin limiti sıfıra eşit olmayan tutarlı tahminciler de bulunabilir. Bu nedenle ortalama hata karenin n sonsuza giderken limitinin sıfır olması tahmincinin tutarlı olduğunu gösterirse de, limitin sıfırdan farklı olması tahmincinin tutarsız olduğunu göstermeyecektir.

### 10.4.4. Yeterlilik

Anakütle parametresi $\theta$'nın değerinin tahmini için kullanılan tahminci $\hat{\theta}$, örnekteki bilgilerin tümünü kullanıyorsa, yeterli tahmincidir. Bu özelliğe göre örnek anakütle ile ilgili bilgiler taşımaktadır. Bu bilgiler örnek birimlerinde veya bunların bütünün de bulunmaktadır. Tüm örnek birimlerini kullanmayan tahminciler yeterli değildir. Örneğin anakütle ortalaması $\mu$'nin tahmini için tahminci olarak örnek ortalaması $\overline{X}$, örnek modu ve örnek medyanı kullanılacak olsun. Örneğin tüm birimleri kullanıldığı için örnek ortalaması $\overline{X}$ yeterli tahmincidir. Fakat aynı tahmin için örneğin tüm birimleri kullanılmadan hesaplanan mod veya medyan yeterli olmayan tahmincidir.

## 10.5. TAHMİN YÖNTEMLERİ

Buraya kadar iyi bir tahmincinin taşıması gerekli özellikler incelenmiştir. Ayrıca, verilen örneklerle de tahmin formülleri herhangi bir yöntemle elde edilmeyip rastgele verilmiştir. Uygulamada farklı tahmin problemleri ile karşılaşılacak ve tahminci olarak yazılabilecek pek çok formülün sözkonusu özellikleri taşıyıp taşımadığını incelemek de çok zor olacaktır.

Tahmincilerin veya tahmin formüllerinin belirlenmesi için tahmin yöntemleri adı verilen yöntemlerden yararlanılır. Bu yöntemler daha önce söz edilen özellikleri veya bu özelliklerden bazılarını taşıyan tahmincilerin belirlenmesini sağlamaktadır.

### 10.5.1. Momentler Yöntemi

Adından da anlaşılacağı gibi bu yöntemde tahminciler momentler yardımı ile belirlenir. Örneğin anakütle aritmetik ortalaması tahmin edilecek ise örneğin orjine göre birinci momenti tahminci olarak kullanılır.

$$m_1 = \frac{\sum_{i=1}^{n} X_i}{n} = \mu$$

Aynı şekilde anakütle varyansı tahmin edilecekse örneğin ortalamaya göre ikinci momenti tahminci olarak kullanılır.

$$m_2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n} = \sigma^2$$

Momentler yöntemi ile elde edilen tahminciler genellikle tutarlı tahmincilerdir. Fakat bazıları bazı özellikleri taşımazlar. Karşılaşılan her olayda momentler belirlenemeyeceğinden bu durumda kullanılamazlar.

### 10.5.2. En Küçük Kareler Yöntemi

Yöntem, farkların kareleri toplamının minimize edilmesi esasına dayanır. Örneğin, bir X tesadüfi değişkeninin r. momentinin hesaplanacağını düşünelim. Burada r = 0,1,2,3,..... değerini alacaktır.

$$E(X^r) = m_r$$

$m_r$'nin en küçük kareler tahmincisini elde etmek için değişkenlerin r. kuvvetleri ile $m_r$ farklarının kareleri toplamı alınacak ve bu farkların kareleri toplamı minimize edilecektir.

$$T = \sum_{i=1}^{n} (X_i^r - m_r)^2$$

Aritmetik ortalama orjine göre birinci momenti olduğuna göre aritmetik ortalamanın tahmincisi için,

$$T = \sum_{i=1}^{n} (X_i - m_1)^2$$

minimize edilecektir. $m_1$ anakütle ortalaması olduğuna göre $m_1$ yerine $\mu$ kullanırsak, minimize edilecek toplam,

$$T = \sum_{i=1}^{n} (X_i - \mu)^2$$

olacaktır. Minimizasyon için T'nin $\mu$'ye göre birinci türevinin alınıp sıfıra eşitlenmesi gerekecektir. Elde edilen sonucun minimum veya maksimum olduğunun belirlenmesi için ise ikinci türevin işaretine bakılacaktır.

Aritmetik ortalama için birinci türevi alıp sıfıra eşitlersek,

$$\frac{dT}{d\mu} = \sum_{i=1}^{n} 2(X_i - \mu)(-1)$$

$$= \sum_{i=1}^{n} 2(X_i - \hat{\mu})(-1) = 0$$

görüldüğü gibi tahmin olduğunu belirtmek için sıfıra eşitlendiğinde μ yerine $\hat{\mu}$ sembolü kullanılmıştır.

$$-2\sum_{i=1}^{n}(X_i - \hat{\mu}) = 0$$

$$\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \hat{\mu} = 0$$

$$\sum_{i=1}^{n} X_i - n\hat{\mu} = 0$$

$$\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n}$$

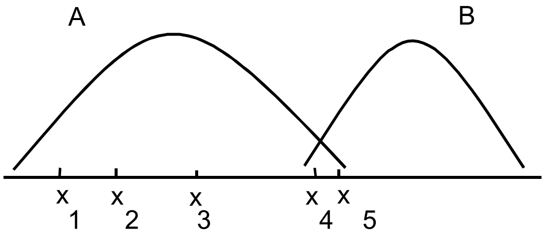
elde edilir. Yani, anakütle ortalamasının en küçük kareler yöntemi ile belirlenen tahmincisi örnek ortalamasıdır. İkinci türev alınırsa toplamın minimize edildiği görülür.

### 10.5.3. Maksimum Benzerlik Yöntemi

Maksimum benzerlik yönteminde herhangi bir örnek istatistiğinin ortaya çıkma olasılığı dikkate alınır. Farklı anakütlelerden alınacak örnekler de farklı olacağından, herhangi bir örneğin bazı anakütlelerden alınması olasılığı diğer bazı anakütlelerden alınması olasılığından daha yüksek olacaktır. Bir örnek diğer anakütlelerden çok alındığı anakütleye benzeyecektir.

Normal dağılmış bir anakütleyi düşünelim. Bu anakütleden alınacak n birimli örneğin birimleri ortalama etrafında dağılacaktır. Örnek ortalaması 8 ise, bu örneğin ortalaması 8 veya 8'e yakın anakütleden alınması olasılığı, ortalaması 6 veya 10 olan, yani 8'den farklı olan anakütlelerden alınması olasılığından daha fazladır. Örneğin, ortalaması 8 veya 8'e yakın anakütleye benzeme olasılığı daha fazladır.

Beş birimden oluşan ($X_1$, $X_2$, $X_3$, $X_4$, $X_5$) bir örnek alınmış olsun. A ve B gibi iki farklı anakütle düşünelim. Örnek birimleri şekilde görüldüğü gibi dağılıyorsa, bu örneğin A anakütlesinden alınması veya A anakütlesine benzemesi olasılığı diğer anakütleden alınması veya diğer anakütleye benzemesi olasılığından daha fazladır.




**Şekil 10.1.**

Dağılım parametreleri $\theta_1, \theta_2, ......, \theta_k$ ve olasılık dağılımı f(X) olan bir anakütleden n birimden oluşan ($X_1$, $X_2$, ....., $X_n$) bir örnek alındığında; $\theta_1, \theta_2, ......, \theta_k$ 'nın maksimum benzerlikler yöntemi ile elde edilecek tahmincileri ($X_1$, $X_2$, ....., $X_n$)'in maksimizasyonu ile belirlenecektir.

Örneğin beş kişilik bir örnek alınarak kişilere soru sorulduğunu ve bunlardan ikisinin soruya evet cevap verdiğini düşünelim. Bu durumda X tesadüfi değişkeni sadece iki değer alacak ve tek parametreli bir dağılım sözkonusu olacaktır.

f(E) = p

f(H) = 1-p

Örneğimizde örnek oranı,

$$p = \frac{2}{5} = 0.4$$

tür. İki evet, üç hayır sonucunu oranı ne olan anakütleden alma olasılığının daha yüksek olduğunu belirleyelim. Bunun için anakütle oranı 0;0.1;0.2;......;0.9;1 olarak alınacaktır. İki evet üç hayır sonucu bu olasılıklarla hesaplanacaktır.

| P | f(E,E,H,H,H) |
|---|---|
| 0 | (0x0x1x1x1) = 0 |
| 0.1 | (0.1x0.1x0.9x0.9x0.9)= 0.00729 |
| 0.2 | (0.2x0.2x0.8x0.8x0.8)= 0.02048 |
| 0.3 | (0.3x0.3x0.7x0.7x0.7)= 0.03087 |
| 0.4 | (0.4x0.4x0.6x0.6x0.6)= 0.03456 |
| 0.5 | (0.5x0.5x0.5x0.5x0.5)= 0.03125 |
| 0.6 | (0.6x0.6x0.4x0.4x0.4)= 0.02304 |
| 0.7 | (0.7x0.7x0.3x0.3x0.3)= 0.01323 |
| 0.8 | (0.8x0.8x0.2x0.2x0.2)= 0.00512 |
| 0.9 | (0.9x0.9x0.1x0.1x0.1)= 0.00081 |
| 1 | (1x1x0x0x0) = 0 |

Yukarıda verilen olasılıklar incelenirse, en yüksek olasılığın 0.03456 olduğu ve p=0.4 için hesaplandığı görülecektir. Bu sonuç sözkonusu örneğin oranı 0.4 olan anakütleden alınma olasılığının, oranı farklı olan anakütlelerden alınma olasılığından daha yüksek olduğunu göstermektedir. f(E,E,H,H,H) bir benzerlik fonksiyonudur ve 0.4 olasılığında maksimize olmuştur.

Maksimum benzerlik fonksiyonu örneğin bileşik olasılık dağılımının formülüne verilen addır ve $\ell$ ile gösterilir.

$$\ell = f(X_1, X_2, ....., X_n)$$

X tesadüfi değişkeni olasılık fonksiyonu f(X) olan bir tesadüfi değişken ise tesadüfi olarak seçilen n adet örnek birimi ($X_1, X_2, ....., X_n$), X ile aynı dağılıma sahip bağımsız tesadüfi değişkenlerdir. $X_1, X_2, ....., X_n$ bağımsız olduklarından benzerlik fonksiyonu,

$$\ell = f(X_1)f(X_2)f(X_3) .....f(X_n)$$

olarak yazılabilir. Örneğin birleşik olasılık formülü ile benzerlik fonksiyonu aynıdır. Birleşik olasılık dağılımında parametreler ($\theta_1, \theta_2, ......, \theta_k$) sabit, örnek birimleri ($X_1, X_2, ....., X_n$) değişkendir. Benzerlik fonksiyonunda ise parametreler değişken, belirli bir örnekten alındıkları için örnek birimleri sabittirler. Formüller aynı, fakat yorumları farklıdır.

Maksimum benzerlik yönteminde benzerlik fonksiyonu belirlenerek, bu fonksiyon parametreler için maksimize edilir. Maksimizasyon için benzerlik fonksiyonunun, fonksiyonda yer alan parametrelere göre birinci türevi alınarak bunlar sıfıra eşitlenir. Böylece istenen formüller elde edilmiş olur.

Elde edilen sonuçların maksimum veya minimum olduğunu belirlemek için ikinci türevin işaretine bakmak gerekir. Uygulamada benzerlik fonksiyonu $\ell$ yerine, bunun logaritmasının alınması kolaylık sağlayacaktır. Bir fonksiyonun kendisini veya logaritmasını maksimize etmek arasında bir fark olmayacaktır. Negatif değerler için logaritma alınması sözkonusu olamazsa da, burada birleşik olasılık fonksiyonları ile çalışıldığından, böyle bir problem ile karşılaşılmayacaktır. Logaritmik benzerlik fonksiyonunu $\ell$ yerine L ile ifade edeceğiz.

$$L = \ln \ell$$

**ÖRNEK:** Daha önce sonuçlarını sayısal olarak incelediğimiz örneği alarak tahminciyi maksimum benzerlik yöntemi ile belirleyelim. Örneğimizde X sadece iki değer alan bir tesadüfi değişkendi.

$$f(E) = p$$
$$f(H) = 1-p$$

olacaktır. Evet istenen cevap sayılarak bu durumda X=1, hayır istenmeyen cevap sayılarak bu durumda X=0 kabul edilirse, tesadüfi değişkenin olasılık fonksiyonu,

$$f(X) = p^x(1-p)^{1-x}$$

olacaktır. X'in 0 ve 1 değerleri için,

$$f(E) = (1-p)^{1-1}p^1 = p$$

$$f(H) = (1-p)^{1-0}p^0 = 1-p$$

bulunur. X tesadüfi değişkeninin beklenen değeri ise,

$$E(X) = \sum_{i=1}^{n} p_i X_i = 0(1-p)+1p = p$$

dir. Örneğimizde 5 birimden oluşan bir örnek alınmıştı. Burada örnek birim sayısını n olarak alalım. Benzerlik fonksiyonu,

$\ell = f(X_1)f(X_2)f(X_3)\,.....f(X_n)$

$$\ell = [p^{X_1}(1-p)^{1-X_1}][p^{X_2}(1-p)^{1-X_2}][p^{X_3}(1-p)^{1-X_3}]......[p^{X_n}(1-p)^{1-X_n}]$$

olacaktır. Çarpılacak ifadelerle tabanlar aynı olduğundan üstler toplanır.

$$\ell = p^{X_1+X_2+X_3+....+X_n}.(1-p)^{(1-X_1)+(1-X_2)+....+(1-X_n)}$$

$$\ell = p^{\sum_{i=1}^{n} X_i}.(1-p)^{n-\sum_{i=1}^{n} X_i}$$

Daha önce logaritmik fonksiyon ile çalışmanın kolaylığından sözetmiştik. Benzerlik fonksiyonunun logaritmasını alalım.

$$\mathrm{L} = \ln \ell = \left(\sum_{i=1}^{n} X_i\right)\ln p + \left(n - \sum_{i=1}^{n} X_i\right)\ln(1-p)$$

p'nin tahmincisi belirleneceğine göre L'nin p'ye göre kısmi türevi alınarak sıfıra eşitlenecektir.

$$\frac{dL}{dp} = \left(\sum_{i=1}^{n} X_i\right)\left[\frac{1}{p}\right] + \left(n - \sum_{i=1}^{n} X_i\right)\left[\frac{1}{1-p}\right](-1)$$

$$\left(\sum_{i=1}^{n} X_i\right)\left[\frac{1}{p}\right] + \left(n - \sum_{i=1}^{n} X_i\right)\left[\frac{1}{1-p}\right](-1) = 0$$

Paydaları eşitleyerek içler dışlar çarpımı yapalım.

$$\frac{\left(\sum_{i=1}^{n} X_i\right)(1-\hat{p}) + \left(n - \sum_{i=1}^{n} X_i\right)(-\hat{p})}{\hat{p}(1-\hat{p})} = 0$$

$$\left(\sum_{i=1}^{n} X_i\right)(1-\hat{p}) + \left(n - \sum_{i=1}^{n} X_i\right)(-\hat{p}) = 0$$

$$\sum_{i=1}^{n} X_i - \hat{p}\sum_{i=1}^{n} X_i - n\hat{p} + \hat{p}\sum_{i=1}^{n} X_i = 0$$

$$\sum_{i=1}^{n} X_i - n\hat{p} = 0$$

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Burada n olayın tekrar sayısı olduğundan örneği n'e bölünecektir. Evet sonucu 1 ile hayır sonucu 0 ile gösterildiğinden örneğimizde $\sum_{i=1}^{n} X_i = 2$ olacaktır. Bu durumda anakütle oranı,

$$\hat{p} = \frac{2}{5} = 0.4$$

olarak tahmin edilecektir.

### 10.5.4. En İyi Doğrusal Sapmasız Tahmin Yöntemi

Daha önce üç tahmin yönteminden söz ettik. Bu yöntemlerle elde edilen tahminciler daha önce açıklanan tahmincilerin özelliklerinden bazılarına sahipken, bazılarına sahip değildirler. Bu yöntemlerle istenen özelliklerin tümüne sahip tahminciler elde edilemez. Oysa en iyi doğrusal sapmasız tahmin yöntemi sözkonusu yöntemlerden farklıdır. Bu yöntem ile istenen özelliklere sahip olan tahminciler elde edilebilir.

Daha önce tahmincilerin özellikleri açıklanırken üçüncü küçük örnek özelliği olarak en iyi doğrusal sapmasız tahminci olma özelliği verilmiş ve tahminci $\hat{\theta}$'in anakütle parametresi θ'nın en iyi doğrusal sapmasız tahmincisi olması için,

- $\hat{\theta}$'in örnek birimlerinin doğrusal bir fonksiyonu olması,

- $\hat{\theta}$'in θ'nın sapmasız tahmincisi olması,

- $\hat{\theta}$'in varyansının diğer sapmasız tahmincilerin varyansına eşit veya onlardan küçük olması gerektiği belirtilmişti.

Burada en iyi doğrusal sapmasız tahminciler elde edilirken de tahmincilerin aynı özelliklere sahip olmaları istenecektir. Bu yöntem ile elde edilecek formüller yukarıda söz edilen özellikleri taşıyacaklardır. Bu formüller elde edilirken Lagrange Çarpanı Yöntemi kullanılmaktadır. Önce kısıtlamalar belirlenir ve daha sonra kısmi türevler alınarak tahminciler elde edilir.

## 10.6. ARALIK TAHMİNİ

Nokta tahmininde anakütlenin bilinmeyen parametresi θ için bir tek değer tahmin edilmektedir. Daha önce açıklandığı gibi anakütleden çok sayıda örnek alınabilir. Bu örneklerden elde edilecek tahminlerin bir kısmının değeri anakütle parametresinin değerine eşit, bir kısmınınki anakütle parametresinin değerinden büyük, bir kısmınınki ise anakütle parametresinin değerinden küçük olacaktır. Bu nedenle ele alınan örnekten ötürü tahmin ile anakütle parametresi gerçek değeri arasında bir fark olması olasılığı vardır. Örnekleme hatası olarak adlandırdığımız bu farkın belirlenen bir olasılık ile ne kadar olabileceğini tahmin etmek için aralık tahmini yapılır.

Aralık tahmininde belirlenen bir hata payı (anlam seviyesi ) α veya güven olasılığı (1-α) ile anakütle parametresi θ için alt ve üst güven sınırları oluşturulur. Oluşturulacak alt sınır a ve üst sınır b ile ifade edilirse,

$$P(a < \theta < b) = 1 - \alpha$$

olacaktır. θ'nın sınırlarını belirlenmesinde tahmin $\hat{\theta}$ kullanılır. $\hat{\theta}$'lerin oluşturduğu seti A($\hat{\theta}$) ile ifade edersek,

$$P\left[\theta \in A(\hat{\theta})\right] = 1 - \alpha \qquad \text{ve}$$

$$P\left[\theta \in \overline{A}(\hat{\theta})\right] = \alpha$$

yazabiliriz. Birinci ifade θ'nın $\hat{\theta}$'lerin oluşturduğu setin bir elemanı olması olasılığının (1-α) olduğunu; ikinci ifade ise $\hat{\theta}$'nın sözkonusu setin elemanı olmaması olasılığının α olduğunu ifade etmektedir.

Aralık tahmini ile amaçlanan örneklemeden doğan sapmanın belirlenen güven olasılığı ile ne kadar olacağının tahmin edilmesidir. Örneğin bir anakütlenin ortalaması tahmin edilirken nokta tahmini 5 olarak elde edilmiş olsun. 0.05 hata payı veya 0.95 güven olasılığı ile alt ve üst güven sınırları 4 ve 6 olarak hesaplanmış olsun. Bu sonuç anakütle parametresinin değerinin 0.95 olasılıkla 4-6 arası olduğunu; anakütle parametresinin değerinin 4'ten küçük veya 6'dan büyük olması olasılığının ise 0.05 olduğunu ifade etmektedir.

Dikkat edilirse güven aralığı simetrik olarak düzenlenmektedir. Bunun sebebi sapmanın tek yönlü olmayıp, çift yönlü olmasıdır. Bazı parametrelerin aralık tahmin ile ilgili açıklamalar aşağıda yapılacaktır.

Aralık tahmininin yapılabilmesi için ilgili parametrelerin örnekleme dağılımlarının bilinmesi gerekmektedir. Bu nedenle aralık tahmininden önce örnekleme dağılımları ile ilgili açıklamalar yapılacaktır.

### 10.6.1. Anakütle Ortalamasının Aralık Tahmini

Anakütle ortalamasının aralık tahmininin yapılabilmesi için ortalamaların örnekleme dağılımının bilinmesi gereklidir. Bu nedenle önce ortalamaların örnekleme dağılımı açıklanacaktır.

#### 10.6.1.1. Ortalamaların Örnekleme Dağılımı

Birim sayısı N ve ortalaması μ olan bir anakütleden alınacak n birimli örneğin ortalaması ( $\overline{X}$ ), μ'nin istenen özellikleri taşıyan tahmincisidir. Bu nedenle,

$$\mu = \overline{X}$$

olarak tahmin edilir.

Ortalaması μ ve standart sapması σ olan bir anakütleden tesadüfi olarak seçilecek n birimli tüm örneklerin ortalamalarının dağılımının ortalaması μ ve standart sapması $\frac{\sigma}{\sqrt{n}}$ 'dir. $\frac{\sigma}{\sqrt{n}}$ tahminlerin standart hatası olarak adlandırılır ve $\sigma_{\overline{X}}$ şeklinde gösterilir.

Ortalamaların standart hatası, anakütle varyansı ( $\sigma^2$ ) biliniyorsa şöyle tahmin edilecektir.

-iadeli seçimle örnek düzenleniyorsa,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \text{ dir.}$$

-iadesiz seçimle örnek düzenleniyorsa $\sigma_{\overline{X}}$, örnekleme oranına (n/N) bağlı olacaktır.

$$\frac{n}{N} < 0.05 \quad ise \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

$$\frac{n}{N} \geq 0.05 \quad ise \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \text{ dir.}$$

Burada $\sqrt{\frac{N-n}{N-1}}$ sonlu anakütle düzeltme faktörü veya bassel faktörüdür. Seçim iadesiz ise örnekleme oranına bakılmaksızın formülde sonlu anakütle düzeltme faktörünün yeralması gerekmektedir. Ancak,

$$\frac{n}{N} < 0.05 \quad ise \quad \sqrt{\frac{N-n}{N-1}} = 1$$

kabul edilmekte ve bu nedenle ihmal edilmektedir.

Örnek birim sayısı (n) büyük ise, örnek ortalamalarının teorik dağılımı, anakütle dağılımı ne olursa olsun normal dağılıma yaklaşır.

Örnek büyüklüğü için kriter n ≥30'dur. n ≥30 ise örnek ortalamalarının teorik dağılımı normal kabul edilebilir. Anakütleden alınan örnek ile örnek ortalaması arasında fark olabilir. Bu fark örnek birim sayısına (n) bağlı olacak, n arttıkça örnek ortalamasının, anakütle ortalamasına yakın olma olasılığı artacaktır.

**ÖRNEK:** μ=50 ve σ=6 olan ve normal dağıldığı kabul edilen bir anakütleden 36 birimli bir örnek alınacaktır. Örnek ortalamasının 48-52 arasında olması olasılığını hesaplayın. (N= çok sayıda)

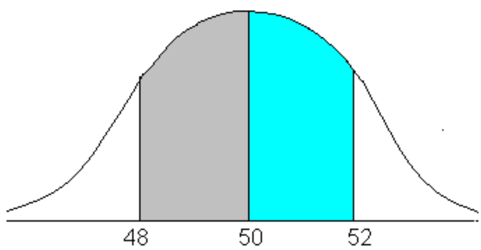
N büyük olduğundan n/N < 0.05 olacaktır. Örnek ortalamalarının dağılımı için standart değişken,

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \text{ dir.}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{36}} = \frac{6}{6} = 1$$

$$Z_1 = \frac{48-50}{1} = \frac{-2}{1} = -2 \rightarrow 0.4772$$

$$Z_2 = \frac{52-50}{1} = 2 \rightarrow 0.4772$$




$$P(48 < \overline{X} < 52) = 0.4772 + 0.4772 = 0.9544$$

n'in artarak 100 olduğunu varsayalım.

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{100}} = \frac{6}{10} = 0.6$$

$$Z_1 = \frac{48-50}{0.6} = \frac{-2}{0.6} = -3.33 \rightarrow 0.4996$$

$$Z_2 = \frac{52-50}{0.6} = 3.33 \rightarrow 0.4996$$

Bu durumda

$$P(48 < \overline{X} < 52) = 0.4996 + 0.4996 = 0.9992$$

olur. n'in artması ile olasılığın arttığı görülmektedir.

Buraya kadar yaptığımız açıklamalarda anakütle varyansının $(\sigma^2)$ bilindiği kabul edilmiştir. Anakütle varyansı bilinmiyorsa örnekten anakütle varyansı $\sigma^2$'nın tahmini gerekecektir. Anakütle varyansının sapmasız tahmincisi $S^2$'dir ve istenen diğer özellikleri de taşır.

$S^2$ veya S ,

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

veya

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

olarak tahmin edilir.

Örnek varyansı $\hat{\sigma}^2$ ise, $\hat{\sigma}^2$ ile $S^2$ arasında,

$$S^2 = \frac{n\hat{\sigma}^2}{n-1}$$

ilişkisi bulunacaktır. Buradan,

$$S = \hat{\sigma}\sqrt{\frac{n}{n-1}}$$

yazılabilir. Bu durumda,

$$\sigma_{\overline{X}} = \frac{S}{\sqrt{n}}$$

veya

$$\sigma_{\overline{X}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

olacaktır.

Ortalamaların örnekleme dağılımı ile ilgili özellikleri şöyle özetleyebiliriz.

1- Anakütle dağılımı normal ise ve anakütle varyansı biliniyorsa, örnek ortalamalarının dağılımı normaldir. Örnek birim sayısı önemli değildir. Bu durumda,

-iadeli seçim ve iadesiz seçimde n/N <0.05 ise

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

-iadesiz seçimde n/N ≥0.05 ise

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

olacaktır.

2- Anakütle dağılımı normal ise ve anakütle varyansı bilinmiyorsa, n ne olursa olsun örnek ortalamalarının dağılımı (n-1) serbestlik derecesi ile t-u dağılımıdır. n ≥30 ise t-dağılımı, normal dağılıma yaklaşacağından bu durumda dağılım normal kabul edilebilir.

Bu durumda σ bilinmediğinden S hesaplanarak birinci şıkta σ'ların yerini S alacaktır. Formüllerde bundan başka değişiklik olmayacaktır.

3- Anakütle dağılımı normal değil ise n ≥30 olmalıdır. Bu durumda n<30 için verilen formüller geçerli olmayacaktır.

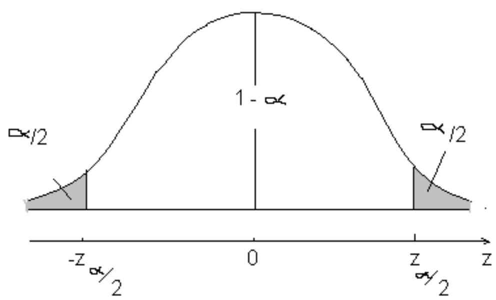
**10.6.1.2. Aralık Tahmini**

Örnek ortalamalarının dağılımı normal ise standart Z değişkeninin

$$Z = \frac{\overline{X}-\mu}{\sigma_{\overline{X}}}$$

değerleri normal dağılım tablosundan $Z_{\alpha/2}$ olarak elde edilecektir. Bu durumda,

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$




**Şekil 10.2**

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

$$-Z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma_{\overline{X}}} < +Z_{\alpha/2}$$

olacaktır. Tarafları $\sigma_{\overline{X}}$ ile çarparsak,

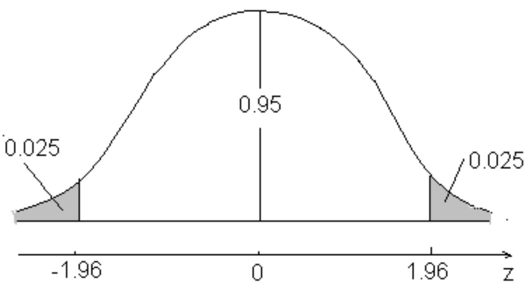
$$-Z_{\alpha/2}\sigma_{\overline{X}} < \overline{X}-\mu < +Z_{\alpha/2}\sigma_{\overline{X}}$$

$\overline{X}$ 'yı eşitsizliğin iki yanına alıp –1 ile çarparsak,

$$\overline{X}-Z_{\alpha/2}\sigma_{\overline{X}} < \mu < \overline{X}+Z_{\alpha/2}\sigma_{\overline{X}}$$

olarak anakütle ortalaması $\mu$'nün $\alpha$ hata payı veya 1-$\alpha$ güven olasılığı ile içinde bulunacağı aralığın alt ve üst sınırları bulunur. $\sigma_{\overline{X}}$ ise daha önce yapılan açıklamalara göre belirlenecektir.

**ÖRNEK:** Bir fabrikada üretilen A mamüllerinin boy uzunlukları araştırılmaktadır. Bu amaçla fabrikada üretilen çok sayıda mamülden 256 birimlik bir örnek alınmış ve örnek ortalaması 18 mm. bulunmuştur. Anakütle dağılımının normal ve varyansının 9 mm. olduğu bilindiğine göre, anakütle ortalamasını %95 olasılıkla (%5 hata payı ile) tahmin ediniz.



n=256 $\overline{X}=18$

$\sigma^2=9 \rightarrow \sigma=3$ $\alpha=0.05$

$$\overline{X}-Z_{\alpha/2}\sigma_{\overline{X}}<\mu<\overline{X}+Z_{\alpha/2}\sigma_{\overline{X}}$$

$$Z_{\alpha/2}\rightarrow 1-\alpha=1-0.05=0.95$$

0.95/2 =0.4750→ tablodan $Z_{\alpha/2}=1.96$

N→∞ n/N < 0.05 ve $\sigma_{\overline{X}}=\frac{\sigma}{\sqrt{n}}$

$$\overline{X}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$18-1.96\frac{3}{\sqrt{256}}<\mu<18+1.96\frac{3}{\sqrt{256}}$$

$$18-\frac{5.88}{16}<\mu<18+\frac{5.88}{16}$$

$$18-0.3675<\mu<18+0.3675$$

$$17.6325<\mu<18.3675$$

%95 olasılıkla üretilen mamüllerin boy uzunlukları (anakütle ortalaması) 17.6375 mm. ile 18.3675 mm. arasında olacaktır.

### 10.6.1.3. Küçük Örneklerde Anakütle Ortalamasının Aralık Tahmini

Anakütle varyansı bilinmiyorsa ve anakütle dağılımı normal dağılım ise veya normal dağılım kabul ediliyorsa, örnekleme dağılımı t dağılımı olacağından $Z_{\alpha/2}$ değerini belirlemek için normal dağılım yerine t dağılımı kullanılır.

n ≥30 ise, t dağılımının normal dağılıma yaklaştığı görülür. Bu nedenle, örnek ortalamaları dağılımının varyansının bilinmediği ve n <30 olduğu durumda örnekleme dağılımının t dağılımı olduğunu söyleyebiliriz. Bu durumda $Z_{\alpha/2}$ değerleri yerine t- dağılımı tablosundan yararlanarak $t_{\alpha/2}$ değerleri belirlenecektir. Aralık tahmini benzer şekilde,

$$P(-t_{\alpha/2}<t<t_{\alpha/2})=1-\alpha$$

$$-t_{\alpha/2}<\frac{\overline{X}-\mu}{\sigma_{\overline{X}}}<+t_{\alpha/2}$$

$$\overline{X}-t_{\alpha/2}\frac{\sigma_{\overline{X}}}{\sqrt{n}}<\mu<\overline{X}+t_{\alpha/2}\frac{\sigma_{\overline{X}}}{\sqrt{n}}$$

olarak yapılacaktır.

**ÖRNEK:** Normal dağıldığı bilinen bir anakütleden 16 birimli bir örnek alınmış ve örnek ortalaması $\overline{X}$ =35 ve S=4.6 bulunmuştur. 0.95 olasılıkla anakütle ortalamasını tahmin ediniz.

t dağılımı tablosundan,

$\alpha/2 = 0.05/2 = 0.025$ ve

serbestlik derecesi n-1 = 16-1=15 karşılığı $t_{\alpha/2} = 2.131$

$$\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$35 - 2.131\frac{4.6}{\sqrt{16}} < \mu < 35 + 2.131\frac{4.6}{\sqrt{16}}$$

$$32.55 < \mu < 37.45$$

olarak elde edilir.

### 10.6.2. Anakütle Oranının Tahmini

Anakütle ortalaması gibi anakütle oranında aralık tahmini yapılabilir. Bunun için oranların örnekleme dağılımının bilinmesi gerekmektedir.

#### 10.6.2.1. Oranların Örnekleme Dağılımı

Birim sayısı N olan anakütlenin oranını P ve bu anakütleden alınan n birimli örneğin oranını ise p ile ifade edelim. Anakütle oranı P, 0 veya 1'e yakın değilse ve anakütleden alınan örnek büyükse; örnek oranlarının dağılımının, ortalaması P ve standart sapması $\sigma_p = \sqrt{\frac{p.q}{n}}$ olan normal dağılıma yakındır.

Bu durumda standart değişken,

$$Z = \frac{p - P}{\sigma_p} = \frac{p - P}{\sqrt{\frac{pq}{n}}}$$

olacaktır. Anakütle oranının nokta tahmini ise P=p olarak yapılacaktır. Yani anakütle oranı P'nin tahmincisi, örnek oranı p'dir.

Anakütle oranının aralık tahmininde anakütle oranı P bilinmeyeceğinden $\sigma_p$'nin bilinmesi sözkonusu değildir. Seçimin iadeli ve iadesiz olmasına ve örnekleme oranına göre oranların standart hatası,

-iadeli seçim ve iadesiz seçimde n/N <0.05 ise

$$\sigma_p = \sqrt{\frac{pq}{n}}$$

-iadesiz seçimde n/N ≥0.05 ise

$$\sigma_p = \sqrt{\frac{pq}{n}}.\sqrt{\frac{N-n}{N-1}}$$

olacaktır.

Anakütle oranı bilinmediğinden $\sigma_p$ yerine örnekten tahmin edilecek $S_p$'nin kullanılması gerekecektir.

$$S_p^2 = \frac{n\sigma_p^2}{n-1} = \frac{n.\frac{pq}{n}}{n-1} = \frac{pq}{n-1}$$

ve

$$S_p = \sqrt{\frac{pq}{n-1}}$$

olarak tahmin edilecektir. Uygulamada n büyük ise $\sqrt{\frac{pq}{n}}$ ile $\sqrt{\frac{pq}{n-1}}$ değerleri birbirine çok yakın olacağından $\sigma_p \cong S_p$ kabul edilebilir. Verdiğimiz örneklerde de $\sigma_p \cong S_p$ kabul edilmiştir.

#### 10.6.2.2. Aralık Tahmini

Oranların aralık tahmininde de ortalamaların aralık tahmininde olduğu gibi standart değişken için,

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$

olacak şekilde anakütle oranının aralık tahmini yapılacaktır. Anakütle oranının aralık tahmini,

$$P(-Z_{\alpha/2} < \frac{p-P}{\sqrt{\sigma_p}} < +Z_{\alpha/2}) = 1-\alpha$$

$$p - Z_{\alpha/2}\sqrt{\sigma_p} < P < p + Z_{\alpha/2}\sqrt{\sigma_p}$$

olarak yapılır.

**ÖRNEK:** Bir işletmede çalışanların iş yerine zamanında gelip gelmedikleri araştırılmaktadır. Bu nedenle işletmede görev yapan 10.000 kişiden 200 kişilik bir örnek tesadüfi olarak seçilmiş ve çalışanların 0.67'sinin son 1 ay içinde görevlerine hiç geç kalmadıkları anlaşılmıştır. 0.95 olasılıkla tüm işletmede çalışanlardan iş yerine zamanında gelenlerin oranını tahmin ediniz.

$$Z_{\varepsilon/2} = 1.96$$

$$\frac{n}{N} = \frac{200}{10000} = 0.02 < 0.05 \rightarrow \sigma_p = \sqrt{\frac{pq}{n}}$$

$$p - Z_{\alpha/2}\sqrt{\frac{pq}{n}} < P < p + Z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

$$0{,}67 - 1{,}96\sqrt{\frac{0{,}67.0{,}33}{200}} < P < 0{,}67 + 1{,}96\sqrt{\frac{0{,}67.0{,}33}{200}}$$

$$0{,}67 - 1{,}96.0{,}033 < P < 0{,}67 + 1{,}67.0{,}033$$

$$0{,}67\text{-}0{,}064 < P < 0{,}67+0{,}064$$

$$0{,}606 < P < 0{,}734$$

0.95 olasılıkla tüm işletmede zamanında iş yerine gelenlerin oranı 0,606 ile 0,734 arasındadır.

### 10.6.3. Ortalamalar Arasındaki Farkların Aralık Tahmini

Farklı iki anakütlenin ortalamaları arasındaki farkın aralık tahmini yapılabilir. Bu nedenle ortalamalar arasındaki farkların örnekleme dağılımının bilinmesi gereklidir.

#### 10.6.3.1. Ortalama Farklarının Örnekleme Dağılımı

Birim sayıları $N_1$ ve $N_2$ ve ortalamaları $\mu_1$ ve $\mu_2$ olan iki anakütleden alınacak örneklerin birim sayıları $n_1$ ve $n_2$ ; ortalamaları $\overline{X}_1$ ve $\overline{X}_2$ olsun. Anakütle ortalamaları arasındaki fark ($\mu_1$ - $\mu_2$), örnek ortalamaları arasındaki fark ($\overline{X}_1$ - $\overline{X}_2$)'a eşit olarak tahmin edilir.

$$\mu_1 - \mu_2 = \overline{X}_1 - \overline{X}_2$$

Tahmin edilen bu fark örnekleme sapması içerdiğinden aralık tahmini yapılabilir. Ortalamalar arasındaki farkların dağılımı için standart değişken

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\overline{X}_1 - \overline{X}_2}}$$

dır.

Ortalamalar arasındaki farkların dağılımlarının standart sapması yani standart hata $(\sigma_{\overline{X}_1 - \overline{X}_2})$ şöyle belirlenir.

-anakütlenin standart sapmaları biliniyorsa,

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

-anakütlenin standart sapmaları bilinmiyorsa,

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

Bazı durumlarda anakütle varyansları birbirine eşit olabilir. $\sigma_1^2 = \sigma_2^2$ ise ve varyans biliniyorsa,

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sigma.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = \sigma\sqrt{\frac{n_1+n_2}{n_1 n_2}}$$

varyans bilinmiyorsa,

$$\sigma_{\bar{X}_1-\bar{X}_2} = S.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = S\sqrt{\frac{n_1+n_2}{n_1 n_2}}$$

olacaktır.

Örnek birim sayılarının ($n_1$+$n_2$ <120) olması durumunda anakütle varyansları-nın değeri bilinmiyor, fakat anakütle dağılımının normal olduğu ve varyansların birbirine eşit olduğu biliniyorsa,

$$S = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

olarak tahmin edilir. Bu durumda $n_1$ veya $n_2$'den biri veya ikisi de 30'dan küçük olacaktır. Dağılımın serbestlik derecesi ($n_1$+$n_2$-2)'dir. 30'dan büyük serbestlik dere-cesi için t- dağılımı normal dağılıma yaklaşacağından, bu durumda örnekleme dağı-lımı olarak kabul edilebilir. Yukarıda verilen formül özellikle aynı anakütleden iki örnek alınması durumunda geçerli olacaktır.

### 10.6.3.2. Aralık Tahmini

Ortalamaların ve oranların aralık tahmininde olduğu gibi, ortalama farklarının aralık tahmininde de benzer işlemler tekrarlanacaktır. Ortalama farklarının aralık tahmini,

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$

$$P(-Z_{\alpha/2} < \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{X}_1-\bar{X}_2}} < Z_{\alpha/2}) = 1-\alpha$$

$$(\bar{X}_1-\bar{X}_2) - Z_{\alpha/2}\sigma_{\bar{X}_1-\bar{X}_2} < (\mu_1-\mu_2) < (\bar{X}_1-\bar{X}_2) + Z_{\alpha/2}\sigma_{\bar{X}_1-\bar{X}_2}$$

olarak yapılacaktır.

**ÖRNEK:** İki farklı anakütleden alınan iki örnek ile ilgili bilgiler aşağıda verilmiştir.

$n_1 = 100 \qquad \overline{X}_1 = 42 \qquad S_1 = 6$

$n_2 = 120 \qquad \overline{X}_2 = 35 \qquad S_2 = 5$

0.05 hata payı ile bu örneklerin alındığı anakütlelerin ortalamaları arasındaki farkı tahmin ediniz.

$Z_{\varepsilon/2} = 1.96$

$$\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$(\overline{X}_1 - \overline{X}_2) - Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < (\mu_1 - \mu_2) < (\overline{X}_1 - \overline{X}_2) + Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$(42-35) - 1{,}96\sqrt{\frac{(6)^2}{100} + \frac{(5)^2}{120}} < (\mu_1 - \mu_2) < (42-35) + 1{,}96\sqrt{\frac{(6)^2}{100} + \frac{(5)^2}{120}}$$

$$7 - 1{,}96(0{,}7538) < (\mu_1 - \mu_2) < 7 + 1{,}96(0{,}7538)$$

$$7 - 1{,}4774 < (\mu_1 - \mu_2) < 7 + 1{,}4774$$

$$5{,}5226 < (\mu_1 - \mu_2) < 8{,}4774$$

### 10.6.4. Oranlar Arasındaki Farkların Aralık Tahmini

İki anakütle ortalamaları arasındaki farkın aralık tahmini gibi iki anakütlenin oranları arasındaki farkların da aralık tahmini yapılabilir. Ancak burada önce iki oran arasındaki farkın örnekleme dağılımı açıklanacaktır.

#### 10.6.4.1. Oran Farklarının Örnekleme Dağılımı

Birim sayıları $N_1$ ve $N_2$, oranları $P_1$ ve $P_2$ olan iki anakütleden alınacak örneklerin birim sayıları $n_1$ ve $n_2$, oranları $p_1$ ve $p_2$ olsun. Anakütle oranları arasındaki fark ($P_1$-$P_2$), örnek oranları arasındaki farka ($p_1$-$p_2$) eşit olarak tahmin edilir.

$P_1 - P_2 = p_1 - p_2$

Oranlar arasındaki farkların dağılımı için standart değişken,

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sigma_{p_1-p_2}}$$

olacaktır. Oran farklarının dağılımının standart hatası ise,

$$\sigma_{p_1-p_2} = \sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}$$

olarak belirlenir.

### 10.6.4.2. Aralık Tahmini

Oranların aralık tahmini formülü de standart değişkenin genel ifade de yerine konması ile elde edilecektir.

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$

$$P(-Z_{\alpha/2} < \frac{(p_1-p_2)-(P_1-P_2)}{\sigma_{p_1-p_2}} < Z_{\alpha/2}) = 1-\alpha$$

$$\text{P } (p_1-p_2) - Z_{\alpha/2}\sigma_{p_1-p_2} < (P_1-P_2) < (p_1-p_2) + Z_{\alpha/2}\sigma_{p_1-p_2}) = 1-\alpha$$

Yukarıda verilen formül ile iki anakütle oranının farkları belirlenen hata payı veya güven olasılığı ile tahmin edilmiş olacaktır.

**ÖRNEK:** İki farklı anakütleden alınan iki örnek ile ilgili bilgiler aşağıda verilmiştir.

$n_1 = 1250$ $\quad p_1 = 0{,}25$

$n_2 = 1500$ $\quad p_2 = 0{,}22$

0,95 güven olasılığı ile bu örneklerin alındığı anakütlelerin oranları arasındaki farkı tahmin ediniz.

$$Z_{\varepsilon/2} = 1.96$$

$q_1 = 1 - p_1 = 1 - 0{,}25 = 0{,}75$

$q_2 = 1 - p_2 = 1 - 0{,}22 = 0{,}78$

$$\sigma_{p_1-p_2} = \sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}$$

$$= \sqrt{\frac{(0{,}25)(0{,}75)}{1250}+\frac{(0{,}22)(0{,}78)}{1500}} = 0{,}01626$$

$$(p_1-p_2) - Z_{\alpha/2}\sigma_{p_1-p_2} < (P_1-P_2) < (p_1-p_2) + Z_{\alpha/2}\sigma_{p_1-p_2}$$

$(0{,}25-0{,}22)-1{,}96(0{,}01626) < (P_1-P_2) < (0{,}25-0{,}22)+1{,}96(0{,}01626)$

$0{,}03-0{,}032 < (P_1-P_2) < 0{,}03+0{,}032$

$-0{,}002 < (P_1-P_2) < 0{,}062$

### 10.6.5. Anakütle Varyansının Aralık Tahmini

Anakütle ortalaması ve oranı gibi anakütle varyansının da aralık tahmini yapılabilir. Bu durumda anakütle varyansının aralık tahmininde $S^2$'den yararlanılacaktır.

Ortalaması μ ve varyansı $\sigma^2$ olan bir anakütleden alınacak n birimli örneğin ortalaması $\overline{X}$ ve varyansı $S^2$ olsun. $\overline{X}$ ve $S^2$'nin bağımsız olması durumunda,

$$\frac{(n-1)S^2}{\sigma^2}$$

değişkeninin dağılımı (n-1) serbestlik dereceli ki-kare dağılımıdır. Bundan yararlanarak anakütle varyansının aralık tahmini,

$$P(\chi^2_{1-\alpha/2,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}) = 1-\alpha$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right) = 1-\alpha$$

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

olarak elde edilir.

**ÖRNEK:** Yeni piyasaya sürülen bir ürün nüfusları birbirine yakın 8 bölgede satışa sunulmuştur. Ürünün aylık satış ortalaması $\overline{X}$ =20500 ve S=1600 olarak hesaplanmıştır. 0,90 güven olasılığı ile anakütle varyansını ve standart sapmasını tahmin ediniz.

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

n =8

serbestlik derecesi = n-1=8-1=7

α=0,10

$$\chi^2_{0,10/2,7} = \chi^2_{0,05,7} = 14{,}067$$

$$\chi^2_{1-0,10/2,7} = \chi^2_{0,95,7} = 2{,}167$$

$$\frac{(8-1)(1600)^2}{14{,}067} < \sigma^2 < \frac{(8-1)(1600)^2}{2{,}167}$$

$$1273903{,}46 < \sigma^2 < 8269497$$

$$1128{,}67 < \sigma < 2875{,}67$$

Anakütle varyans ve standart sapmasının aralık tahmini büyük örnekler için örnekleme dağılımının normal dağılıma yaklaştırılması ile yapılabilir. Büyük örneklerde S'lerin dağılımının standart hatası ($\sigma_s$),

$$\sigma_s = \frac{\sigma}{\sqrt{2n}}$$

dir. Buna göre σ'ın güven aralığı ise,

$$S - Z_{\alpha/2}\frac{\sigma}{\sqrt{2n}} < \sigma < S + Z_{\alpha/2}\frac{\sigma}{\sqrt{2n}}$$

olur. Bu eşitsizlikten σ'ın güven aralık tahmini,

$$\frac{S}{1+\frac{Z_{\alpha/2}}{\sqrt{2n}}} < \sigma < \frac{S}{1-\frac{Z_{\alpha/2}}{\sqrt{2n}}}$$

olarak belirlenebilir.

**ÖRNEK:** 100 birimli bir örnekte S=12 olarak bulunmuştur. 0,05 hata payı ile anakütle standart sapmasını tahmin edin.

$$\frac{S}{1+\frac{Z_{\alpha/2}}{\sqrt{2n}}} < \sigma < \frac{S}{1-\frac{Z_{\alpha/2}}{\sqrt{2n}}}$$

$$\frac{12}{1+\frac{1,96}{\sqrt{2,100}}} < \sigma < \frac{12}{1-\frac{1,96}{\sqrt{2,100}}}$$

$$10,539 < \sigma < 13,930$$

### 10.6.6. Varyans Oranlarının Aralık Tahmini

İki anakütle ortalaması veya iki anakütle oranı arasındaki fark gibi, iki anakütle varyansının birbirine oranının aralık tahmini yapılabilir. Bu durumda varyans oranlarının örnekleme dağılımı F dağılımıdır.

Varyansları $\sigma_1^2$ ve $\sigma_2^2$ olan normal dağılmış iki anakütleden alınan $n_1$ ve $n_2$ birimli örneklerin varyansları $S_1^2$ ve $S_2^2$ olsun. Bu durumda $S^2$'lerin $\sigma^2$'lere oranlarının, birbirine oranının dağılımı ($n_1$-1) ve ($n_2$-1) serbestlik dereceli F dağılımıdır.

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

Varyans oranlarının aralık tahmini,

$$P(F_{1-\alpha/2,n_1-1,n_2-1} < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < F_{\alpha/2,n_1-1,n_2-1}) = 1-\alpha$$

olarak yapılacaktır. F tablo değerlerinin bulunmasında,

$$F_{1-\alpha/2,n_1-1,n_2-1} = \frac{1}{F_{\alpha/2,n_2-1,n_1-1}}$$

olduğundan bu ilişkiden yararlanılacaktır. Bu dönüşüm ile varyans oranlarının aralık tahmini formülü,

$$\frac{S_1^2}{S_2^2}.\frac{1}{F_{\alpha/2,n_1-1,n_2-1}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}.F_{\alpha/2,n_2-1,n_1-1}$$

olarak elde edilir.

**ÖRNEK:** A marka mallardan tesadüfi olarak 10 adet alınarak $\overline{X}_A = 1{,}5$ ve $S_A$=0,4 ; B marka mallardan tesadüfi olarak 11 adet alınarak $\overline{X}_B = 2{,}0$ ve $S_B$=0,6 olarak hesaplanmıştır. 0,98 güven olasılığı ile anakütle varyanslarının ve standart sapmalarının oranlarını tahmin ediniz.

$n_1$ =10

$n_2$ =11 $\alpha$=0,02

$$F_{0,02/2,10-1,11-1} = 4{,}94$$

$$F_{0,02/2,11-1,10-1} = 5{,}26$$

$$\frac{S_1^2}{S_2^2}.\frac{1}{F_{\alpha/2,n_1-1,n_2-1}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}.F_{\alpha/2,n_2-1,n_1-1}$$

$$\frac{(0{,}4)^2}{(0{,}6)^2}.\frac{1}{4{,}94} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{(0{,}4)^2}{(0{,}6)^2}.5{,}26$$

$$0{,}09 < \frac{\sigma_1^2}{\sigma_2^2} < 2{,}34$$

$$0{,}30 < \frac{\sigma_1}{\sigma_2} < 1{,}53$$

### 10.6.7. Optimum Örnek Büyüklüğünün Belirlenmesi

Örneklemenin özellikle maliyet ve zaman yönünden tasarruf sağladığını belirtmiştik. Örnek düzenlenirken örnek büyüklüğünün yeterli olması da aynı nedenlerden ötürü önemlidir. Örnek birim sayısı yetersiz ise, yapılan araştırmadan beklenen bilgiler sağlıklı olarak elde edilemeyeceği gibi, örnek birim sayısının fazla olması maliyetin artmasına ve zaman kaybına neden olacaktır.

Örneğe alınacak en az birim sayısını bazı şartlarla, matematiksel olarak belirleyebiliriz. Belirleyeceğimiz bu birim sayısını optimum örnek büyüklüğü olarak adlandıracağız.

#### 10.6.7.1. Anakütle Ortalamasının Tahmini İçin Optimum Örnek Büyüklüğünün Belirlenmesi

Anakütle ortalamasının aralık tahmini,

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

şeklinde yapılıyordu. Buradan,

$$-Z_{\alpha/2}\sigma_{\overline{X}} < \overline{X}-\mu < +Z_{\alpha/2}\sigma_{\overline{X}}$$

eşitliğini elde ederiz. $\overline{X}-\mu$ farkını E ile gösterirsek,

$$\mathrm{E} = \overline{X}-\mu$$

$$\mathrm{E} < Z_{\alpha/2}\sigma_{\overline{X}}$$

yazabiliriz. Burada hata miktarı E'nin en çok belirlenen değer kadar veya daha az olması için örnek büyüklüğünü belirleyebiliriz. Hatanın en çok E değeri kadar olması için,

$$\mathrm{E} = Z_{\alpha/2}\sigma_{\overline{X}}$$

olacaktır. $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$ olduğundan,

$$\mathrm{E} = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = \frac{Z_{\alpha/2}\sigma}{E}$$

$$n = \left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2$$

olacaktır. Bu eşitlikte n'in belirlenebilmesi için anakütle varyansının bilinmesi, hata payının ve hata miktarı E'nin belirlenmesi gerekmektedir. $\sigma^2$ bilinmiyor, fakat yaklaşık olarak tahmin edilebiliyorsa örnek büyüklüğü buna dayanarak da belirlenebilir. E anakütle ortalaması ile örnek ortalaması arasında olması istenen maksimum farkı belirtecektir.

**ÖRNEK:** σ =8 olan bir anakütleden n birimli bir örnek alınacaktır. Örnek ortalaması ile anakütle ortalaması arasındaki farkın 0,95 olasılıkla en çok 1,5 olması için n ne olmalıdır, hesaplayınız.

$$Z_{\varepsilon/2} = 1.96$$

$$n = \left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2$$

$$n = \left(\frac{1{,}96.8}{1{,}5}\right)^2 = 109{,}27 \cong 109 \; birim$$

n ≥109 birim alınırsa, $\overline{X}$ *ve* $\mu$ arasındaki fark 0,95 olasılıkla ±1,5 veya daha az olacaktır.

### 10.6.7.2.Anakütle Oranının Tahmini İçin Optimum Örnek Büyüklüğünün Belirlenmesi

Aynı şekilde p-P = E dersek,

$E = Z_{\alpha/2}\sigma_p$ ve

$$n = \frac{Z^2_{\alpha/2}pq}{E^2}$$

olacaktır.

Oranlarla ilgili örnek büyüklüğü belirlenirken, P bilinmediğinden tahmin edilmesi gerekecektir. p tahmin edilirken tahmin belirli bir aralıkta yapılıyorsa (%15-20 gibi) n'i belirlemek için tahmin edilen büyük oranın alınması iyi olacaktır. P tahmin edildikten sonra $Z_{\alpha/2}$ güven olasılığına ve E araştırmaya göre belirlenebilir.

**ÖRNEK:** Bir bölgede A malı kullanan ailelerin oranı araştırılmaktadır. Bölgede A malı kullanan ailelerin toplam ailelere oranının 0,40 olduğu tahmin edilmektedir. 0,90 olasılıkla örnek oranı ile anakütle oranı arasındaki farkın 0,04 olması için örneğe alınacak aile sayısını belirleyiniz.

$$Z_{\varepsilon/2} = 1.64$$

$$n = \frac{Z^2_{\alpha/2}pq}{E^2} = \frac{(1{,}64)^2(0{,}4)(0{,}6)}{(0{,}04)^2} = 403{,}44$$

$$n \cong 403$$

örnek en az 403 aileden oluşmalıdır.

## 10.7. ÇÖZÜMLÜ ÖRNEKLER

**ÖRNEK 1 :**$\sigma$=50 olan 5000 birimli bir anakütleden 400 birimden oluşan örnek iadesiz seçimle düzenlenerek, örnek ortalaması 4500 olarak bulunmuştur. 0,10 hata payı ile örnek ortalamasını tahmin ediniz.

**Çözüm :**

N =5000

n=400

$\sigma$=50 $\qquad$ $Z_{\alpha/2} \rightarrow 1-\alpha = 1-0,10 = 0,90$

$\overline{X} = 4500$ $\qquad$ $0,90/2 = 0,4500 \rightarrow$ tablodan $Z_{\alpha/2} = 1,64$

$\alpha$=0,10 $\qquad$ $\frac{n}{N} = \frac{400}{5000} = 0,08 > 0,05$ olduğundan

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$4500 - 1,64\frac{50}{\sqrt{400}}\sqrt{\frac{5000-400}{5000-1}} < \mu < 4500 + 1,64\frac{50}{\sqrt{400}}\sqrt{\frac{5000-400}{5000-1}}$$

$$4500 - 1,64.\frac{50}{20}.0,95 < \mu < 4500 + 1,64.\frac{50}{20}.0,95$$

$$4500 - 3,895 < \mu < 4500 + 3,895$$

$$4496,105 < \mu < 4503,895$$

0.10 hata payı ile örnek ortalaması bu aralık içinde olacaktır.

Not: Seçim iadeli veya örnekleme oranı 0,05 'ten küçük olsaydı, $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ alınacaktı.

**ÖRNEK 2:** Bir depoda bulunan 10000 mamülün ağırlıkları incelenmektedir. Bu mamüllerden 36 birimli bir örnek düzenlenmiş (iadesiz seçimle) ve $\overline{X}$ =20 gr. ve S=4 gr. bulunmuştur. 0,70 olasılıkla ambarda bulunan mamüllerin ortalamasını tahmin ediniz.

**Çözüm:**

$Z_{\alpha/2} \rightarrow 1-\alpha = 0,70/2 = 0,3500 \rightarrow$ tablodan $Z_{\alpha/2} = 1,04$

$\frac{n}{N} = \frac{36}{10000} = 0,0036 < 0,05 \rightarrow \qquad \sigma_{\overline{X}} = \frac{S}{\sqrt{n}}$

$$\overline{X} - Z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$20 - 1{,}04\frac{4}{\sqrt{36}} < \mu < 20 + 1{,}04\frac{4}{\sqrt{36}}$$

$$20 - 0{,}693 < \mu < 20 + 0{,}693$$

$$19{,}307 < \mu < 20{,}693$$

0,70 olasılıkla ambarda bulunan mamüllerin ortalaması 19,3 gr. ile 20,6 gr. arasında olacaktır.

**ÖRNEK 3:** 500 birimden oluşan bir topluluktan 49 birimlik bir örnek iadesiz seçimle alınmıştır. Örnek ortalaması $\overline{X}$ =98 ve örnek standart sapması $\hat{\sigma}$ =8 bulunduğuna göre, 0,01 hata payı ile anakütle ortalamasını tahmin ediniz.

**Çözüm:**

$Z_{\alpha/2} \rightarrow$ 1-$\alpha$ =1-0,01 =0,99

0,99/2 = 0,4950$\rightarrow$ tablodan $Z_{\alpha/2}$ =2,57

$\frac{n}{N} = \frac{49}{500} = 0{,}098 > 0{,}05$ olduğundan ve σ bilinmediğinden ,

$$\sigma_{\overline{X}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$S^2 = \frac{\Sigma(X_i - \overline{X})^2}{n-1} = \frac{n\hat{\sigma}^2}{n-1}$$

$$S = \sigma\sqrt{\frac{n}{n-1}} = 8\sqrt{\frac{49}{49-1}} = 8.1{,}01 = 8{,}08$$

$$\overline{X} - Z_{\alpha/2}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} < \mu < \overline{X} + Z_{\alpha/2}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\overline{X}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{8{,}08}{\sqrt{49}}\sqrt{\frac{500-49}{500-1}} = 1{,}096$$

$$98 - 2{,}57.1{,}096 < \mu < 98 + 2{,}57.1{,}096$$

$$98 - 2{,}81 < \mu < 98 + 2{,}81$$

$$95{,}19 < \mu < 100{,}81$$

0,01 hata payı veya 0,99 güven olasılığı ile anakütle ortalaması 95,19 ile 100,81 birim arasında olacaktır.

Aynı örnek iadeli seçimle düzenlenseydi veya n/N<0,05 olsaydı,

$$\sigma_{\overline{X}} = \frac{S}{\sqrt{n}} = \frac{8{,}08}{\sqrt{49}} = 1{,}15$$

$$98 - 2{,}57.1{,}15 < \mu < 98 + 2{,}57.1{,}15$$

$$95{,}05 < \mu < 100{,}95$$

olacaktı.

**ÖRNEK 4:** N=5000, n =49, $\sum_{i=1}^{49} X_i = 2940$ ve α=0,05 olduğuna göre μ'nün güven aralığını bulunuz. ( $\sigma$ =7)

**Çözüm:**

$Z_{\alpha/2}$ =2,57

$$\frac{n}{N} = \frac{49}{5000} < 0{,}05$$

$$\overline{X} = \frac{\Sigma X_i}{n} = \frac{2940}{49} = 60$$

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$60 - 1{,}96\frac{7}{\sqrt{49}} < \mu < 60 + 1{,}96\frac{7}{\sqrt{49}}$$

$$58{,}04 < \mu < 61{,}96$$

**ÖRNEK 5:** Bir önceki örnekte N=500 olduğuna göre anakütle ortalamasını tahmin ediniz.

$$\frac{n}{N} = \frac{49}{500} = 0{,}098 > 0{,}05$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{7}{\sqrt{49}}\sqrt{\frac{500-49}{500-1}} = 0{,}95$$

$$\overline{X} - Z_{\alpha/2}\sigma_{\overline{X}} < \mu < \overline{X} + Z_{\alpha/2}\sigma_{\overline{X}}$$

$$60 - 1{,}96.0{,}95 < \mu < 60 + 1{,}96.0{,}95$$

$$58{,}138 < \mu < 61{,}862$$

**ÖRNEK 6:** 6000 birimli bir anakütleden 64 birimli bir örnek iadesiz seçimle alınarak,

$$\sum_{i=1}^{64} X_i = 3200$$

$$\sum_{i=1}^{64}(X_i - \overline{X})^2 = 1575$$

bulunmuştur. Anakütle ortalamasını 0,99 olasılıkla tahmin ediniz.

**Çözüm:**

$Z_{\alpha/2} = 2,57$

$$\frac{n}{N} = \frac{64}{6000} < 0,05$$

$$\overline{X} = \frac{\Sigma X_i}{n} = \frac{3200}{64} = 50$$

$$S = \sqrt{\frac{(X_i - \overline{X})^2}{n-1}} = \sqrt{\frac{1575}{64-1}} = \sqrt{25} = 5$$

$$\overline{X} - Z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$50 - 2,57\frac{5}{\sqrt{64}} < \mu < 50 + 2,57\frac{5}{\sqrt{64}}$$

$$48,4 < \mu < 51,6$$

**ÖRNEK 7:** Bir anakütleden alınan 200 birimli örnek için şu değerler hesaplanmıştır.

$$\sum_{i=1}^{200} X_i = 1500$$

$$\sum_{i=1}^{200} X_i^2 = 12500$$

anakütle ortalamasını 0,80 güven olasılığı ile tahmin ediniz.

**Çözüm:**

$Z_{\alpha/2} \rightarrow$ 1-$\alpha$ = 0,80/2 = 0,4$\rightarrow$ tablodan $Z_{\alpha/2}$ =1,28

$$\frac{n}{N} < 0,05 \rightarrow \qquad \sigma_{\overline{X}} = \frac{S}{\sqrt{n}}$$

$$\overline{X} = \frac{\Sigma X_i}{n} = \frac{1500}{200} = 7,5$$

$$\hat{\sigma} = \sqrt{\frac{\Sigma X_i^2}{n} - (\overline{X})^2} = \sqrt{\frac{12500}{200} - (7,5)^2} = \sqrt{62,5 - 56,25}$$

$$= \sqrt{6,25} = 2,5$$

$$S = \hat{\sigma}\sqrt{\frac{n}{n-1}} = 2{,}5\sqrt{\frac{200}{200-1}} = 2{,}5.1{,}002 = 2{,}505$$

$$\overline{X} - Z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$7{,}5 - 1{,}28\frac{2{,}505}{\sqrt{200}} < \mu < 7{,}5 + 1{,}28\frac{2{,}505}{\sqrt{200}}$$

$$7{,}5 - 0{,}22 < \mu < 7{,}5 + 0{,}22$$

$$7{,}28 < \mu < 7{,}72$$

**ÖRNEK 8:** 1000 birimli bir anakütleden alınan örnek aşağıda verilmiştir. (Örnek iadesiz seçimle düzenlenmiştir.)

| $X_i$ | $f_i$ | $f_iX_i$ | $X_i - \overline{X}$ | $(X_i - \overline{X})^2$ | $f_i(X_i - \overline{X})^2$ |
|---|---|---|---|---|---|
| 2 | 11 | 22 | -2 | 4 | 44 |
| 3 | 26 | 78 | -1 | 1 | 26 |
| 4 | 45 | 180 | 0 | 0 | 0 |
| 5 | 30 | 150 | 1 | 1 | 30 |
| 6 | 9 | 54 | 2 | 4 | 36 |
| | 121 | 484 | | | 136 |

0,9544 güven olasılığı ile bu örneğin alındığı anakütle ortalamasını tahmin ediniz.

**Çözüm:**

$Z_{\alpha/2} \rightarrow 1\text{-}\alpha = 0{,}9544/2 = 0{,}4772 \rightarrow$ tablodan $Z_{\alpha/2} = 2$

$$\frac{n}{N} = \frac{121}{1000} = 0{,}121 > 0{,}05 \rightarrow \qquad \sigma_{\overline{X}} = \frac{S}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}$$

$$\overline{X} = \frac{\Sigma f_i X_i}{\Sigma f_i} = \frac{484}{121} = 4$$

$$S = \sqrt{\frac{\Sigma f_i (X_i - \overline{X})^2}{\Sigma f_i - 1}} = \sqrt{\frac{136}{121-1}} = \sqrt{1{,}133} = 1{,}064$$

$$\sigma_{\overline{X}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{1{,}064}{\sqrt{121}}\sqrt{\frac{1000-121}{121-1}} = 0{,}262$$

$$\overline{X} - Z_{\alpha/2}\sigma_{\overline{X}} < \mu < \overline{X} + Z_{\alpha/2}\sigma_{\overline{X}}$$

$$4 - 0{,}262 < \mu < 4 + 0{,}262$$

$$3{,}476 < \mu < 4{,}524$$

**ÖRNEK 9:** Normal dağıldığı bilinen ve varyansı 81 olan bir anakütleden 25 birimli bir örnek alınmıştır. Örnek ortalaması 150 bulunduğuna göre 0,90 olasılıkla anakütle ortalamasını tahmin ediniz.

**Çözüm:** Anakütle dağılımı normal olduğundan n <30 olsa da normal dağılım tablosundan yararlanılacaktır.

$Z_{\alpha/2} = 1{,}64$

$$\frac{n}{N} < 0{,}05 \rightarrow \qquad \sigma^2 = 81 \rightarrow \sigma = 9 \rightarrow \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$150 - 1{,}64\frac{9}{\sqrt{81}} < \mu < 150 + 1{,}64\frac{9}{\sqrt{81}}$$

$$148{,}36 < \mu < 151{,}64$$

**ÖRNEK 10:** A makinesinin ürettiği hatalı mamüllerin oranı araştırılmaktadır. Bu amaçla makinenin ürettiği 2000 mamülden 180 adedi tesadüfi olarak seçilerek, bunlardan 27 adedinin hatalı olduğu anlaşılmıştır. 0,01 hata payı ile makinenin ürettiği hatalı mamül oranını tahmin ediniz. (Seçim iadesizdir.)

**Çözüm:**

$Z_{\alpha/2} = 2{,}57$

$$\frac{n}{N} = \frac{180}{2000} = 0{,}09 > 0{,}05 \rightarrow \sigma_p = \sqrt{\frac{pq}{n}}.\sqrt{\frac{N-n}{N-1}}$$

$$p - Z_{\alpha/2}\sqrt{\frac{pq}{n}}.\sqrt{\frac{N-n}{N-1}} < P < p + Z_{\alpha/2}\sqrt{\frac{pq}{n}}.\sqrt{\frac{N-n}{N-1}}$$

$$\mathrm{P} = \frac{27}{180} = 0{,}15$$

$$\sigma_p = \sqrt{\frac{pq}{n}}.\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{0{,}15.0{,}85}{180}}.\sqrt{\frac{2000-180}{2000-1}} = 0{,}025$$

$$0{,}15 - 2{,}57.0{,}025 < \mathrm{P} < 0{,}15 + 2{,}57.0{,}025$$

$$0{,}085 < \mathrm{P} < 0{,}21$$

**ÖRNEK 11:** Varyanslarının birbirine eşit olduğu bilinen iki anakütleden alınan örnekler ile ilgili bilgiler aşağıda verilmiştir.

$$n_1 = 12 \qquad \overline{X}_1 = 158 \qquad S_1 = 10$$

$$n_2 = 15 \qquad \overline{X}_2 = 141 \qquad S_2 = 14$$

0,95 güven olasılığı ile ortalamalar arasındaki farkı tahmin ediniz.

**Çözüm:**

$n_1$ <30, $n_2$ <30 olduğundan $n_1$+$n_2$-2 serbestlik dereceli t dağılımı, $t_{0,05/2,12+15-2}$ = $t_{0,025,25}$=2,060

$$\sigma_1^2 = \sigma_2^2$$

olduğundan,

$$S = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{(12-1)(10)^2 + (15-1)(14)^2}{12+15-2}} = 12{,}4$$

$$\sigma_{\overline{X}_1-\overline{X}_2} = S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$= 12{,}4\sqrt{\frac{1}{12}+\frac{1}{15}} = 4{,}8025$$

$$(\overline{X}_1 - \overline{X}_2) - t_{\alpha/2}\sigma_{\overline{X}_1-\overline{X}_2} < (\mu_1 - \mu_2) < (\overline{X}_1 - \overline{X}_2) + t_{\alpha/2}\sigma_{\overline{X}_1-\overline{X}_2}$$

$$(158\text{-}141)\text{-}2{,}060(4{,}8025) < (\mu_1 - \mu_2) < (158\text{-}141)\text{+}2{,}060(4{,}8025)$$

$$7{,}11 < (\mu_1 - \mu_2) < 26{,}89$$