IX. BÖLÜM

TAHMİN

IX.1 Giriş

Hemen hemen tüm araştırmalarda amaç; parametre ya da parametreler hakkında bilgiler elde etmektir. Parametrelerin hesaplanmasının olası olmadığı ve maliyet, zaman, fiziksel şartlar ve kalifiye eleman yetersizliği gibi nedenlerle hesaplanmalarının tercih edilmediği durumlarda ilgili kitle(ler)den tesadüfi örnekler seçilir. Örneklemenin amacı, örnek birimlerine ilişkin ortalama, oran ve varyans gibi istatistikler aracılığı ile parametreler hakkında bilgiler elde etmektir. Bunun iki yolu vardır. Birincisi istatistikleri parametrelere ilişkin hipotezlerin test edilmesinde delil olarak kullanmaktır. İkincisi ise bu istatistikleri ilgili parametrelerin tahmin edilmesinde kullanmaktır.

Bir parametre için iki tür tahminde bulunulabilir. Bunlarda birincisi nokta tahmini olarak bilinir ve parametrelerin tek bir değer olarak tahmin edilmesi anlamına gelir. Örneğin; μ parametresi için $\bar{X}$ istatistiği, π parametresi için P istatistiği ve σ^2 parametresi için S^2 istatistiği birer nokta tahmin edicidir.

Parametrelerin tahmin edilmesinde kullanılan istatistiklere tahmin edici denir. Örneğin; $\bar{X}$ istatistiğinin tahmin edicisi, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, P istatistiğinin tahmin edicisi, $P = \frac{X}{n}$ olup burada X ilgilenilen olayın oluşma sayısnı göstermektedir. Ayrıca S^2 istatistiğinin tahmin edicisi ise $S^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X})^2$ eşitliği ile verilir.

Bir istatistiğin beklenen değerinin (ortalamasının) parametresine eşit olmasına ise yansızlık (sapmasızlık yada tarafsızlık) adı verilir. Örneğin; $\bar{X}$, μ için, P, π için ve S^2 de σ^2 için yansız tahmin edicilerdir. Çünkü; $E(\bar{X}) = \mu$, $E(P) = \pi$ ve $E(S^2) = \sigma^2$'dir.

Tahminin ikinci yolu ise nokta tahminine göre daha kullanışlı olan, istatistiklerin delil olarak kullanılarak parametreler için aralıklar oluşturmaktır ve güven aralığı olarak bilinir. Bu tür tahmin, hipotez testi ile aynı analma gelir. Güven aralıkları genellikle iki yönlü olur; ancak tek yönlü güven aralıklar da tanımlanabilir.

IX.2 Nokta Ve Aralık Tahmini

IX.2.1 Tek Kitle Parametresi İçin Nokta ve Aralık Tahmini

IX.2.1.1 Kitle Oratalaması İçin Nokta ve Aralık Tahmini

a) Kitle Varansı Biliniyorken (σ^2 Biliniyor)

Kitle ortalaması (parametresi, μ ve μ'nün nokta tahmini ise $\bar{X} = \bar{x}$'dir. Kitle varyansı biliniyorken $\bar{X}$'nın varyansı, $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$'dir. Böylece $\bar{X}$'nın dağılımı, $\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$'dir. Buradan

μ'nün $100 \times (1-\alpha)\%$ güvenilirlikle güven aralığı tahmin edicisi, $P\{\bar{X} - Z_T\sigma_{\bar{X}} \leq \mu \leq \bar{X} + Z_T\sigma_{\bar{X}}\} = (1-\alpha)$ ile ve buna uygun tahmini ise $P\{\bar{x} - z_T\sigma_{\bar{x}} \leq \mu \leq \bar{x} + z_T\sigma_{\bar{x}}\} = (1-\alpha)$ eşitliği ile verilir.

♦ **Örnek:** OMÜ öğrencileri arasından tesadüfi olarak seçilen dokuz tanesinin hafta sonunda ev dışında harcadıkları süreler şöyle tespit edilmiştir. x_i(saat): 4, 5, 3, 8, 10, 7, 8, 4 ve 5. Öğrenciler kitlesine ilişkin, hafta sonunda ev dışında harcanan süreye göre dağılımının normal ve varyansı, $\sigma^2 = 4$ ise %99 güvenilirlikle kitle ortalaması, μ için bir güven aralığı oluşturarak sonucu yorumlayınız.

♦ **Çözüm**: $\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$ => μ parametresinin nukta tahmini, $\bar{x} = \frac{1}{9}\sum_{i=1}^{9} x_i = \frac{54}{9} =$ 6 saat bulunur. Kitle varyansı bilindiğinden $\bar{X}$'nın varyansı, $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = \frac{4}{9} \cong 0.444$ ve standart hatası ise $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{3} \cong 0.667$ bulunur. $\alpha = 0.01 => z_T = z_{\alpha/2} = z_{0.005} = 2.576$ (tablodan) bulunur. Öyle ise μ'nün %99 güvenilirlikle güven aralığı tahmin, $P\{\bar{x} - z_T\sigma_{\bar{x}} \leq \mu \leq \bar{x} + z_T\sigma_{\bar{x}}\} = (1-\alpha)$ eşitliğinden hareketle $P\{6 - 2.576(0.667) \leq \mu \leq 6 + 2.576(0.667\} = 0.99 => P\{4.282 \leq \mu \leq 7.718\} = 0.99$ bulunur. **Yorum:** $(4.282; 7.718)$ aralığının ÖMÜ öğrencileri kitlesine ilişkin μ parametresini kapsıyor olma olasılığı %99'dur.

b) Kitle Varansı Bilinmiyorken (σ^2 Bilinmiyor)

i) Büyük Örnek ($n \geq 30$)

$100 \times (1-\alpha)\%$ güven aralığı tahmin edicisi, $P\{\bar{X} - Z_T S_{\bar{X}} \leq \mu \leq \bar{X} + Z_T S_{\bar{X}}\} = (1-\alpha)$ ile ve buna uygun tahmini ise $P\{\bar{x} - z_T s_{\bar{x}} \leq \mu \leq \bar{x} + z_T s_{\bar{x}}\} = (1-\alpha)$ eşitliği ile verilir. Bu eşitlikte $S_{\bar{X}}^2 = \frac{S^2}{n}$ ve $S^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X})^2$ tahmin edicileri olup bunların tahminleri ise sırası ile $s_{\bar{X}}^2 = \frac{s^2}{n}$ ve $s^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2$ eşitlikleri ile hesaplanır.

ii) Küçük Örnek ($n < 30$)

$100 \times (1-\alpha)\%$ güven aralığı tahmin edicisi, $P\{\bar{X} - T_T S_{\bar{X}} \leq \mu \leq \bar{X} + T_T S_{\bar{X}}\} = (1-\alpha)$ ile ve buna uygun tahmini ise $P\{\bar{x} - t_T s_{\bar{x}} \leq \mu \leq \bar{x} + t_T s_{\bar{x}}\} = (1-\alpha)$ eşitliği ile verilir. Bu eşitlikte $S_{\bar{X}}^2 = \frac{S^2}{n}$ ve $S^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X})^2$ tahmin edicileri olup bunların tahminleri ise sırası ile $s_{\bar{X}}^2 = \frac{s^2}{n}$ ve $s^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2$ eşitlikleri ile hesaplanır. Ayrıca burada $t_T = t_{(n-1);\frac{\alpha}{2}}$ tablo değerini göstermektedir.

♦ **Örnek:** Belli bir hastalığın tedavisinde yeni bir yöntem geliştirilmiştir. Tesadüfi olarak seçilen 12 hasta sözü edilen yöntem ile tedavi edilerek iyileşinceye kadar geçen süreler şöyle tespit edilmiştir. x_i(gün): 10, 12, 8, 9, 9, 17, 9, 14, 3, 7, 12 ve 10. %95 güvenilirlikle hastalar kitlesinin ortalama iyileşme süresi, μ'ye ilişkin bir güve aralığı tahmin ediniz ve sonucu yorumlayınız.

♦ **Çözüm**: $\bar{x} = \frac{1}{12}\sum_{i=1}^{12} x_i = \frac{120}{12} = 10$ gün bulunur. Kitle varyansı bilinmediğinden $\bar{X}$'nın varyansı, $s_{\bar{x}}^2 = \frac{s^2}{n} \cong \frac{12.545}{12} \cong 1.045$ ve standart hatası ise $s_{\bar{x}} \cong \frac{3.542}{\sqrt{12}} \cong 1.022$ bulunur. Burada örneklem varyansının tahmini ise $s^2 = \frac{1}{(12-1)}\sum_{i=1}^{12}(x_i - \bar{x})^2 = \frac{138}{11} \cong 12.545$'dir. Böylece $\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right) => t_T = t_{(12-1); \frac{0.05}{2}} = t_{11;;0.025} = 2.201$ (tablodan) bulunur. $=> P\{\bar{x} - t_T s_{\bar{x}} \leq \mu \leq \bar{x} + t_T s_{\bar{x}}\} = (1-\alpha) => P\{10 - 2.201(1.022) \leq \mu \leq 10 + 2.201(1.022)\} = 0.95 => P\{7.751 \leq \mu \leq 12.249\} = 0.95$ bulunur. **Yorum:** %95 güvenilirlikle hastalar kitlesinin iyileşme süresine ilşkin ortalaması olan μ parametresi 7.751 ila 12.249 arasındadır. % 5 hata ile ise bu aralık dışındadır, denilebilir.

♦ **Ödev:** *A* marka elektrik ampüllerinden 10 tanesi tesadüfi olarak seçilerek aynı şartlar altında kullanılmışlardır. Ampüllerin kesilinceye kadar geçen süreleri şöyle tespit edilmiştir. x_i(yıl): 3, 3, 4, 4, 7, 6, 8, 5, 3 ve 7. %90 güvenilirlikle bu tip ampüllerin kitle ortalama dayanma süresi, μ'ye ilişkin bir güve aralığı tahmin ediniz ve sonucu yorumlayınız. (C: $\bar{x} = 5$ yıl; $s^2 = 3.556$, $s_{\bar{x}} = 0.596$ yıl; $t_T = t_{(10-1); \frac{0.10}{2}} = t_{9;;0.05} = 1.833$ olup % 90 G.A., (3.908; 6.092) bulunur.)

IX.2.1.2 Kitle Oranı İçin Nokta ve Aralık Tahmini

Kitle oranı, π için $100 \times (1-\alpha)\%$ güvenilirlikle bir güven aralığı tahmin edicisi, $P\{P - Z_T\sigma_P \leq \pi \leq P + Z_T\sigma_P\} = (1-\alpha)$ ve tahmini ise $P\{p - z_T\sigma_p \leq \pi \leq p + z_T\sigma_p\} = (1-\alpha)$ eşitliği ile verilir. Son eşitlikte $z_T = z_{\alpha/2}$ tablo değeri ve $\sigma_p = \sqrt{\frac{p(1-p)}{n}}$, p oranına ilişkin stadart hata tahminidir.

♦ **Örnek:** Bir süper markete reçel almak amacı ile gelen müşterilerden tesadüfi olarak seçilen 100 tanesinin 60 tanesi *A* marka reçelleri tercih ettiği belirlenmiştir. Reçel satın alanlarda *A* marka reçelleri tercih edenleri kitle oranı, π için % 95 güven düzeyinde güven aralığı tahmin ederek sonucu yorumlayınız.

♦ **Çözüm:** $p = \frac{x}{n} = \frac{60}{100} = 0.60$ bulunur. $\sigma_p^2 = \frac{p(1-p)}{n} = \frac{0.6(0.4)}{100} = 0{,}0024 => \sigma_p = \sqrt{0.0024} \cong 0.049$ ve $z_T = z_{\alpha/2} = z_{0.025} = 1.96$ (z tablosundan) bulunur. Böylece π için % 95 güven aralığı, $P\{p - z_T\sigma_p \leq \pi \leq p + z_T\sigma_p\} = (1-\alpha)$ eşitliği (ya da olasılık beyanı) yardımı ile tahmin edilir. $P\{0.6 - 1.96(0.049) \leq \pi \leq 0.6 + 1.96(0.049)\} = 0.95 =>$ (0.504;

0.696) bulunur. **Yorum:** % 95 güvenilirlikle reçel satın alan kitlesinde *A* marka reçelleri tercih edenleri oranı, π 0.504 ila 0.696 arasındadır, denilebilir.

IX.2.2 İki Kitle Parametresi Farkına İlişkin Güven Aralığı

IX.2.2.1 İki Kitle Oratalama Farkına İlişkin Güven Aralığı

a) Kitle Varyansları Biliniyorken (σ_1^2 ve σ_2^2 Biliniyor)

İki kitle ortalama farkı parametresi , $(\mu_1 - \mu_2)$ için $100 \times (1-\alpha)\%$ güvenilirlikle bir güven aralığı tahmin edicisi, $P\{(\bar{X}_1 - \bar{X}_2) - Z_T\sigma_{\bar{X}_1-\bar{X}_2} \leq (\mu_1 - \mu_2) \leq (\bar{X}_1 - \bar{X}_2) + Z_T\sigma_{\bar{X}_1-\bar{X}_2}\} = (1-\alpha)$ ve tahmini ise $P\{(\bar{x}_1 - \bar{x}_2) - z_T\sigma_{\bar{x}_1-\bar{x}_2} \leq (\mu_1 - \mu_2) \leq (\bar{x}_1 - \bar{x}_2) + z_T\sigma_{\bar{x}_1-\bar{x}_2}\} = (1-\alpha)$ eşitliği ile verilir. Son eşitlikte $z_T = z_{\alpha/2}$ tablo değeri ve $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, $(\bar{x}_1 - \bar{x}_2)$ ortalama farkına ilişkin stadart hata tahmini ve $\bar{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ji}, j = 1, 2$ olup her bir kitle için ortalamanın tahminidir.

♦ **Örnek:** Samsun'daki iki çocuklu ailelerin aylık mutfak harcamalarına ilişkin dağılımın normal ve varyansının, $\sigma_1^2 = 100$ olduğu bilnmektedir. Ayrıca Ordu'daki iki çocuklu ailelerin aylık mutfak harcamalarına ilişkin dağılımının da normal ve varyansının ise $\sigma_2^2 = 144$ olduğu bilinmektedir. Her iki ilden iki çocuğu olan aileler arasından sekizer tanesi tesadüfi olarak seçilerek aylık mutfak harcamalrı tespit edilmiştir. Veriler takipte verilen tablodaki gibidir.

Şehir	Aylık motfak harcaması (x_i; 100 TL)							
Samsun	3	4	2.5	2.8	3	5	4.5	6
Ordu	8	9	7	3	4.5	7	6	5.5

% 95 güvenilirlikle $(\mu_1 - \mu_2)$ parametresi için güven aralığı tahmin ediniz ve sonucu yorumlayınız.

♦ **Çözüm:** Samsun için ortalamanın tahmini, $\bar{x}_1 = \frac{1}{8}\sum_{i=1}^{8} x_{1i} = \frac{3080}{8} = 385\,TL$ ve ordu için ise $\bar{x}_2 = \frac{1}{8}\sum_{i=1}^{8} x_{2i} = \frac{5000}{8} = 625\,TL$ bulunur. Kitle varyansları bilindiğinden ortalama farkının standart hatası ise $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{100}{8} + \frac{144}{8}} \cong 5.523$ bulunur. $z_T = z_{\alpha/2} = z_{0.025} = 1.96$ (z tablosundan) bulunur. Böylece $(\mu_1 - \mu_2)$ parametresi için % 95 G. A., $P\{(\bar{x}_1 - \bar{x}_2) - z_T\sigma_{\bar{x}_1-\bar{x}_2} \leq (\mu_1 - \mu_2) \leq (\bar{x}_1 - \bar{x}_2) + z_T\sigma_{\bar{x}_1-\bar{x}_2}\} = (1-\alpha)$ eşitliği yardımı ile tahmin edilir. $P\{(385 - 600) - 1.96(5.523) \leq (\mu_1 - \mu_2) \leq (385 - 600) + 1.96(5.523)\} = 0.95$ => $P\{-250.825 \leq (\mu_1 - \mu_2) \leq -229.175\} = 0.95$ bulunur. Yorum: (-250.825; -229.175) aralığının $(\mu_1 - \mu_2)$ parametresini içermesi olasılığı % 95'dir. Yine aralığın değerleri negatif olduğundan iki şehir arasındaki ortalama mutfak harcamalarının farklı ve hatta ikincisinin (Ordu'da) daha fazla olduğu söylenebilir.

b) Kitle Varyansları Bilinmiyorken (σ_1^2 ve σ_2^2 Biliniyor)

i) Büyük Hacimli Örneklemlerde ($n_1 \geq 30$ ve $n_2 \geq 30$ iken)

İki kitle ortalama farkı parametresi , $(\mu_1 - \mu_2)$ için $100 \times (1-\alpha)\%$ güvenilirlikle bir güven aralığı tahmin edicisi, $P\{(\bar{X}_1 - \bar{X}_2) - Z_T S_{\bar{X}_1-\bar{X}_2} \leq (\mu_1 - \mu_2) \leq (\bar{X}_1 - \bar{X}_2) + Z_T S_{\bar{X}_1-\bar{X}_2}\} = (1-\alpha)$ ve tahmini ise $P\{(\bar{x}_1 - \bar{x}_2) - z_T s_{\bar{x}_1-\bar{x}_2} \leq (\mu_1 - \mu_2) \leq (\bar{x}_1 - \bar{x}_2) + z_T s_{\bar{x}_1-\bar{x}_2}\} = (1-\alpha)$ eşitliği ile verilir. Son eşitlikte $z_T = z_{\alpha/2}$ tablo değeri ve $s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, $(\bar{x}_1 - \bar{x}_2)$ ortalama farkına ilişkin stadart hata tahminidir. Yine aynı eşitlikte $\bar{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ji}$, $j = 1, 2$ olup her bir kitle için ortalamanın tahmini ve $s_j^2 = \frac{1}{(n_j-1)}\sum_{i=1}^{n_j}\left(x_{ji} - \bar{x}_x\right)^2$, $j = 1, 2$ ise her bir kitle için varyans tahminidir.

ii) Küçük Hacimli Örneklemlerde ($n_1 < 30$ ve $n_2 < 30$ iken)

İki kitle ortalama farkı parametresi , $(\mu_1 - \mu_2)$ için $100 \times (1-\alpha)\%$ güvenilirlikle bir güven aralığı tahmin edicisi, $P\{(\bar{X}_1 - \bar{X}_2) - T_T S_{\bar{X}_1-\bar{X}_2} \leq (\mu_1 - \mu_2) \leq (\bar{X}_1 - \bar{X}_2) + T_T S_{\bar{X}_1-\bar{X}_2}\} = (1-\alpha)$ ve tahmini ise $P\{(\bar{x}_1 - \bar{x}_2) - t_T s_{\bar{x}_1-\bar{x}_2} \leq (\mu_1 - \mu_2) \leq (\bar{x}_1 - \bar{x}_2) + t_T s_{\bar{x}_1-\bar{x}_2}\} = (1-\alpha)$ eşitliği ile verilir. Son eşitlikte $t_T = t_{(n_1+n_2-2);\alpha/2}$ tablo değeri ve $s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, $(\bar{x}_1 - \bar{x}_2)$ ortalama farkına ilişkin stadart hata tahminidir. Yine aynı eşitlikte $\bar{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ji}$, $j = 1, 2$ olup her bir kitle için ortalamanın tahmini ve $s_j^2 = \frac{1}{(n_j-1)}\sum_{i=1}^{n_j}\left(x_{ji} - \bar{x}_x\right)^2$, $j = 1, 2$ ise her bir kitle için varyans tahminidir.

♦ **Örnek:** Belli bir hastalığa yakalanmış hastalardan tesadüfi olarak seçilen sekiz tanesi cerrahi yöntem ile tedavi edilmişlerdir. Yine aynı hastalar içerisinden dokuz tanesi de tesadüfi olarak seçilerek ilaçla tedavi edilmişlerdir. İyileşme süreleri takipte verilen tablodaki gibi tespit edilmiştir.

Tedavi Yöntemi	İyileşme süresi (x_i; gün)								
Cerrahi	15	20	18	20	17	22	17	15	
İlaç	12	15	7	10	9	18	12	16	9

Her iki kitlenin iyileşme süresine göre dağılımının normal olduğu bilinmektedir. Kitle varyansları bilinmiyor; ancak eşit varsayımı altında $(\mu_1 - \mu_2)$ parametresi için % 95 güvenilirlikle bir güven aralığı tahmin ediniz ve sonucu yorumlayınız.

♦ **Çözüm:** $(\bar{x}_1 - \bar{x}_2)$, $s_{\bar{x}_1-\bar{x}_2}$ ve $t_T = t_{(n_1+n_2-2);\alpha/2}$ değerlerini bulalım. $\bar{x}_1 = \frac{1}{8}\sum_{i=1}^{8} x_{1i} = \frac{144}{8} = 18$ gün ve $\bar{x}_2 = \frac{1}{9}\sum_{i=1}^{9} x_{2i} = \frac{108}{9} = 12$ gün olup $(\bar{x}_1 - \bar{x}_2) = 18 - 12 = 6$ gün bulunur. $s_1^2 = \frac{1}{(8-1)}\sum_{i=1}^{8}(x_{1i} - \bar{x}_1)^2 = \frac{44}{7}$ ve $s_2^2 = \frac{1}{(9-1)}\sum_{i=1}^{9}(x_{2i} - \bar{x}_2)^2 = \frac{108}{8}$ olup $s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{44+108}{8+9-2}\left(\frac{1}{8}+\frac{1}{9}\right)} = \sqrt{\frac{152}{15}\left(\frac{17}{72}\right)} = \sqrt{\frac{2584}{1080}} \cong 1.547$ bulunur. $t_T = t_{(n_1+n_2-2);\alpha/2} = t_{15;0.025} = 2.131$ (t tablosundan) bulunur. Butün bu değerler aralık tahmin edicisinde değerlendirilirse $P\{6 - 2.131(1.547) \leq (\mu_1 - \mu_2) \leq 6 + 2.131(1.547)\} = 0.95$ $=> P\{6 - 3.296 \leq (\mu_1 - \mu_2) \leq 6 + 3.296\} = 0.95$ $=>$ $P\{2.704 \leq (\mu_1 - \mu_2) \leq 9.296\} = 0.95$ bulunur. Yorum: (2.704; 9.296) aralığının $(\mu_1 - \mu_2)$ parametresini içermesi olasılığı % 95'dir. % 95 gübenle tahmin edilen aralık sıfırı içermediğinden ve sınırları da pozitif olduğundan ilaç tedavi yöntemi cerrahi tedavi yöntemine göre daha kısa sürede iyileştirme yapmaktadır, denilebilir.