# 3. HAFTA

## Örneklem Temel Bileşenleri

Ortalama vektörü $\underline{\mu}$ ve varyans-kovaryans matrisi $\mathbf{\Sigma}$ olan $p$-boyutlu bir kitleden $\underline{X}' = (X_1, X_2, ..., X_p)$ rasgele vektörüne ilişkin alınan $n$ birimlik rasgele örneklem $\underline{X}_1, \underline{X}_2, ..., \underline{X}_n$ olsun. Bu rasgele örnekleme ilişkin gözlem değerleri $\underline{\mathrm{x}}_1, \underline{\mathrm{x}}_2, ..., \underline{\mathrm{x}}_n$ ‘ nin örneklem ortalama vektörü $\underline{\overline{\mathrm{x}}}_{px1}$, örneklem varyans-kovaryans matrisi $\mathbf{S}_{pxp}$ ve örneklem korelasyon matrisi $\mathbf{R}_{pxp}$ dir.

Herhangi bir lineer birleşimin $n$ tane değeri

$$\underline{l}'_1 \underline{\mathrm{x}}_j = l_{11} x_{1j} + l_{21} x_{2j} + ... + l_{p1} x_{pj} \quad ; \; j = 1, 2, ..., n$$

olmak üzere, bu lineer birleşimin örneklem ortalaması $\underline{l}'_1 \underline{\overline{\mathrm{x}}}$ ve örneklem varyansı $\underline{l}'_1 \mathbf{S} \underline{l}_1$ dir. Diğer bir lineer birleşim $\underline{l}'_2 \underline{\mathrm{x}}_j$ ise bu iki lineer birleşim için örneklem kovaryansı $\underline{l}'_1 \mathbf{S} \underline{l}_2$ dir. Kitle temel bileşenlerinin elde edilmesinde olduğu gibi, örneklem temel bileşenleri de katsayı vektörleri birim uzunluklu, varyans değerlerine göre büyükten küçüğe doğru sıralanmış ve ilişkisiz olarak elde edilmektedir.

Birinci örneklem temel bileşeni, $\underline{l}'_1 \underline{l}_1 = 1$ kısıtı altında $\underline{l}'_1 \underline{\mathrm{x}}_j$ ’ nin örneklem varyansını maksimum yapan $\underline{l}'_1 \underline{\mathrm{x}}_j$ lineer birleşimidir. İkinci örneklem temel bileşeni, $(\underline{l}'_1 \underline{\mathrm{x}}_j \; , \; \underline{l}'_2 \underline{\mathrm{x}}_j)$ çiftleri için örneklem kovaryansı sıfır ve $\underline{l}'_2 \underline{l}_2 = 1$ kısıtı altında $\underline{l}'_2 \underline{\mathrm{x}}_j$ ’ nin örneklem varyansını ikinci sırada maksimum yapan $\underline{l}'_2 \underline{\mathrm{x}}_j$ lineer birleşimidir. Böylece $i$ inci örneklem temel bileşeni, bütün $(\underline{l}'_i \underline{\mathrm{x}}_j \; , \; \underline{l}'_k \underline{\mathrm{x}}_j)$, $k < i$ çiftleri için örneklem kovaryansı sıfır ve $\underline{l}'_i \underline{l}_i = 1$ kısıtı altında $\underline{l}'_i \underline{\mathrm{x}}_j$ ’ nin örneklem varyansını $i$ inci sırada maksimum yapan $\underline{l}'_i \underline{\mathrm{x}}_j$ lineer birleşimidir.

Birinci temel bileşen $\underline{l}'_1 \mathbf{S} \underline{l}_1$ değerini maksimum yapar. Bu ifadenin maksimumu eğer $\underline{l}_1$ katsayı vektörü, $\mathbf{S}$ ’ nin birim özvektörü $\underline{\hat{e}}_1$ eşit alındığında, en büyük özdeğer $\hat{\lambda}_1$ ‘e eşitdir.

Böylece örneklem temel bileşenleri, eğer $\mathbf{S}=\{s_{ik}\}$, özdeğer ve birim özvektör çiftleri $(\hat{\lambda}_1,\underline{\hat{e}}_1),(\hat{\lambda}_2,\underline{\hat{e}}_2),...,(\hat{\lambda}_p,\underline{\hat{e}}_p)$ olan $pxp$ tipinde örneklem varyans-kovaryans matrisi ise, $i$ inci örneklem temel bileşeni

$$\begin{aligned}\hat{y}_i &= \underline{\hat{e}}_i{}'\,\underline{\mathrm{x}} \\ &= \hat{e}_{1i}x_1+\hat{e}_{2i}x_2+...+\hat{e}_{pi}x_p \quad ; \quad i=1,2,...,p\end{aligned}$$

biçimindedir, burada $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq ... \geq \hat{\lambda}_p \geq 0$ ve $\underline{\mathrm{x}}$, $\underline{X}$ rasgele vektörünün herhangi bir gözlemidir. Buradan

$\hat{y}_i$ 'nin örneklem varyans, $s^2_{\hat{y}_i}=\hat{\lambda}_i$ ; $i=1,2,...,p$

$(\hat{y}_i,\hat{y}_k)$ arasındaki örneklem kovaryansı, $s_{\hat{y}_i,\hat{y}_k}=0$ ; $i,k=1,2,...,p,\ \ i\neq k$

dır. Ayrıca, Toplam örneklem varyansı;

$$\begin{aligned}tr(\mathbf{S}) &= \sum_{i=1}^{p} s_{ii} \\ &= \hat{\lambda}_1+\hat{\lambda}_2+...+\hat{\lambda}_p\end{aligned}$$

ve

$$r_{\hat{y}_i,x_k}=\frac{\hat{e}_{ki}\sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}} \quad ; \quad i,k=1,2,...,p,\ \ i\neq k$$

dir.

$\hat{y}_1,\hat{y}_2,...,\hat{y}_p$ ile gösterilen örneklem temel bileşenleri $\mathbf{S}$'den veya $\mathbf{R}$' den elde edilebilir. Ancak $\mathbf{S}$'den veya $\mathbf{R}$' den elde edilen bileşenler aynı değildir. Başlangıçta hangi matrisin kullanılacağına açıklık getirilmelidir.

Örneklem temel bileşenleri normallik varsayımında da elde edilir. Yani $\underline{X}_j \sim N_p(\underline{\mu},\mathbf{\Sigma})$ ; $j=1,2,...,n$ olduğunda $\mathbf{\Sigma}$ varyans-kovaryans matrisinin en çok olabilirlik tahmin edicisi $\hat{\mathbf{\Sigma}}=\mathbf{S}_n$ ‘ den elde edilebilir. Bu durumda örneklem temel bileşenleri ilişkili kitle konturlarının en çok olabilirlik tahminleriyle incelenir. $\hat{\mathbf{\Sigma}}$' nın özdeğerleri $\left[(n-1)/n\right]\hat{\lambda}_i$ ve ilişkili birim özvektörleri $\underline{\hat{e}}_i$ lerdir, burada $(\hat{\lambda}_i,\underline{\hat{e}}_i)$ çifti $\mathbf{S}$' nin özdeğer ve ilişkili birim özvektörleridir. Böylece, $\mathbf{S}$ ve $\hat{\mathbf{\Sigma}}=\mathbf{S}_n$ her ikisi de aynı örneklem temel bileşeni $\underline{\hat{e}}_i{}'\,\underline{\mathrm{x}}$ verir ve

$\frac{\hat{\lambda}_i}{\hat{\lambda}_1 + \hat{\lambda}_2 + ... + \hat{\lambda}_p}$, $i = 1, 2, ..., p$ varyans oranı her ikisi için de aynıdır. Ayrıca $\mathbf{S}$ ve $\hat{\mathbf{\Sigma}} = \mathbf{S}_n$ ’ dan elde edilen örneklem korelasyon matrisi $\mathbf{R}$ aynı olduğundan, standartlaştırılmış değişkenler kullanıldığında, $\mathbf{S}$ veya $\hat{\mathbf{\Sigma}} = \mathbf{S}_n$ seçilmesi anlamsızdır.

**Örnek 4 :** Belli bir bölgedeki 14 ilçeye ilişkin 5 sosyo-ekonomik değişken üzerinden alınan veriler aşağıdaki gibidir:

| İlçeler | Toplam Nüfus(Bin) | Okul yaş ortancası | Toplam iş gücü(Bin) | Sağlık servis iş gücü(Yüz) | Aile gelir ortancası(10000) |
|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
| 1 | 5.935 | 14.2 | 2.265 | 2.27 | 2.91 |
| 2 | 1.523 | 13.1 | 0.597 | 0.75 | 2.62 |
| 3 | 2.599 | 12.7 | 1.237 | 1.11 | 1.72 |
| 4 | 4.009 | 15.2 | 1.649 | 0.81 | 3.02 |
| 5 | 4.687 | 14.7 | 2.312 | 2.5 | 2.22 |
| 6 | 8.044 | 15.6 | 3.641 | 4.51 | 2.36 |
| 7 | 2.766 | 13.3 | 1.244 | 1.03 | 1.97 |
| 8 | 6.538 | 17 | 2.618 | 2.39 | 1.85 |
| 9 | 6.451 | 12.9 | 3.147 | 5.52 | 2.01 |
| 10 | 3.314 | 12.2 | 1.606 | 2.18 | 1.82 |
| 11 | 3.777 | 13 | 2.119 | 2.83 | 1.8 |
| 12 | 1.530 | 13.8 | 0.798 | 0.84 | 4.25 |
| 13 | 2.768 | 13.6 | 1.336 | 1.75 | 2.64 |
| 14 | 6.585 | 14.9 | 2.769 | 1.91 | 3.17 |

Örneklem ortalama vektörü ve örneklam varyans-kovaryans matrisi aşağıdaki gibidir:

$$\underline{\mathrm{x}}' = [4,32 \quad 14,01 \quad 1,95 \quad 2,17 \quad 2,45]$$

$$S = \begin{bmatrix} 4,308 & 1,683 & 1,803 & 2,155 & -0,253 \\ & 1,768 & 0,588 & 0,177 & 0,176 \\ & & 0,801 & 1,065 & -0,158 \\ & & & 1,970 & -0,357 \\ & & & & 0,504 \end{bmatrix}$$

**a)** Temel bileşenleri elde ediniz.

**b)** Bir veya iki temel bileşen ile örneklem değişimi ifade edilebilir mi? Scree Plot çizip grafik üzerinden aynı soruyu cevaplandırınız.

**Çözüm 4 :** Burada örneklem varyans-kovaryans matrisi kullanıldı.

**a)** Örneklem Varyans-Kovaryans matrisinden elde edilen özdeğerler aşağıdaki tabloda verilmiştir :

| $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\hat{\lambda}_i$ | 6.931 | 1.786 | 0.390 | 0.230 | 0.014 |

Yukarıdaki özdeğerlerden yararlanılarak elde edilen birim özvektörler

| $\underline{\hat{e}}_1$ | $\underline{\hat{e}}_2$ | $\underline{\hat{e}}_3$ | $\underline{\hat{e}}_4$ | $\underline{\hat{e}}_5$ |
|---|---|---|---|---|
| 0.781 | -0.071 | 0.004 | 0.542 | -0.302 |
| 0.306 | -0.764 | -0.162 | -0.545 | -0.010 |
| 0.334 | 0.083 | 0.015 | 0.050 | 0.937 |
| 0.426 | 0.579 | 0.220 | -0.636 | -0.173 |
| -0.054 | -0.262 | 0.962 | -0.051 | 0.024 |

Birim özvektörler yardımıyla elde edilen örneklem temel bileşenleri aşağıdaki gibidir:

$$\hat{Y}_1 = 0.781X_1 + 0.306X_2 + 0.334X_3 + 0.426X_4 - 0.054X_5$$

$$\hat{Y}_2 = -0.071X_1 - 0.764X_2 + 0.083X_3 + 0.579X_4 - 0.262X_5$$

$$\hat{Y}_3 = 0.004X_1 - 0.162X_2 + 0.015X_3 + 0.220X_4 + 0.962X_5$$

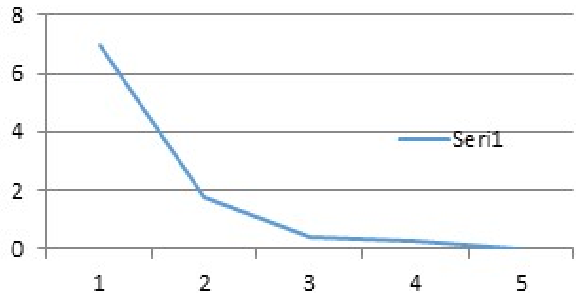
$$\hat{Y}_4 = 0.542X_1 - 0.545X_2 + 0.050X_3 - 0.636X_4 - 0.051X_5$$

$$\hat{Y}_5 = -0.302X_1 - 0.010X_2 + 0.937X_3 - 0.173X_4 + 0.024X_5$$

**b)**

| $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\hat{\lambda}_i$ | 6.931 | 1.786 | 0.390 | 0.230 | 0.014 |
| $y_i$ temel bileşeninin toplam değişimi açıklama oranı | $\frac{\hat{\lambda}_1}{\sum_{i=1}^{5}\hat{\lambda}_i} = \frac{6.931}{9.351} = \%74.1$ | $\frac{\hat{\lambda}_1}{\sum_{i=1}^{5}\hat{\lambda}_i} = \frac{1.786}{9.351} = \%19.1$ | $\%4.2$ | $\%2.5$ | $\%0.1$ |
| Kümülatif açıklama oranı | $\%74.1$ | $\%93.2$ | $\%97.4$ | $\%99.9$ | $\%100$ |

Birinci(1.) temel bileşen toplam varyansın $\%74.1$ 'ini açıklamaktadır. 1. ve 2. Temel bileşenleri aldığımızda toplam açıklama oranı 0.932 (oldukça yüksek) olduğundan iki temel bileşen ile örneklem değişimi ifade edilebilir. 5 değişken üzerinden alınan 14 gözlem, iki temel bileşen üzerinden 14 gözleme indirgenmiş olur. Aşağıdaki grafikte *x* ekseni temel bileşenleri ve *y* ekseni özdeğeleri göstermektedir. Şekilde de görüldüğü gibi temel bileşen sayısının 2 olduğunda çiz düzleşmeye başlamıştır. Buradan da temel bileşen sayısının 2 olduğuna karar verilebilir.




**Örnek 5 :** Aşağıda 24 erkek ve 24 dişi hasta kaplumbağanın üst kabuğuna ilişkin çeşitli ölçümler verilmiştir. Amaç kaplumbağaların üst kabuğunun şeklini ve büyüklüğünü belirlemektir. Kaplumbağaların kabuklarının uzunluklar, genişlik ve yükseklikleri aşağıdaki tabloda verilmiştir:

| **ERKEK KAPLUMBAĞA** | | | **DİŞİ KAPLUMBAĞA** | | |
|---|---|---|---|---|---|
| **uzunluk** | genişlik | yükseklik | uzunluk | genişlik | yükseklik |
| **93** | 74 | 37 | 98 | 81 | 38 |
| **94** | 78 | 35 | 103 | 84 | 38 |
| **96** | 80 | 35 | 103 | 86 | 42 |
| **101** | 84 | 39 | 105 | 86 | 42 |
| **102** | 85 | 38 | 109 | 88 | 44 |
| **103** | 81 | 37 | 123 | 92 | 50 |
| **104** | 83 | 39 | 123 | 95 | 46 |
| **106** | 83 | 39 | 133 | 99 | 51 |
| **107** | 82 | 38 | 133 | 102 | 51 |
| **112** | 89 | 40 | 133 | 102 | 51 |
| **113** | 88 | 40 | 134 | 100 | 48 |
| **114** | 86 | 40 | 136 | 102 | 49 |
| **116** | 90 | 43 | 138 | 99 | 51 |
| **117** | 90 | 41 | 138 | 99 | 51 |
| **117** | 91 | 41 | 141 | 105 | 53 |
| **119** | 93 | 41 | 147 | 108 | 57 |
| **120** | 89 | 40 | 149 | 107 | 55 |
| **120** | 93 | 44 | 153 | 107 | 56 |
| **121** | 95 | 42 | 155 | 115 | 63 |
| **125** | 93 | 45 | 155 | 117 | 60 |
| **127** | 96 | 45 | 158 | 115 | 62 |
| **128** | 95 | 45 | 159 | 118 | 63 |
| **131** | 95 | 46 | 162 | 124 | 61 |
| **135** | 106 | 47 | 177 | 132 | 67 |

Bu verilere doğal logaritmik dönüşüm uygulandığında;

erkek kaplumbağalar için örneklem ortalaması ve örneklem varyans-kovaryans matrisi

$$\underline{x}' = [4.7254 \quad 4.4776 \quad 3.7032]$$

$$S = 10^{-3}\begin{bmatrix} 11.072 & 8.019 & 8.160 \\ & 6.417 & 6.005 \\ & & 6.773 \end{bmatrix}$$

dişi kaplumbağalar için örneklem ortalaması ve örneklem varyans-kovaryans matrisi

$$\underline{x}' = [4.9007 \quad 4.6233 \quad 3.9403]$$

$$S = 10^{-2}\begin{bmatrix} 2.64 & 2.01 & 2,49 \\ & 1.62 & 1,94 \\ & & 2,49 \end{bmatrix}$$

**a)** Erkek kaplumbağalar için bir veya iki temel bileşen ile örneklem değişimi ifade edilebilir mi?

**b)** Dişi kaplumbağalar için bir veya iki temel bileşen ile örneklem değişimi ifade edilebilir mi? **(Ödev)**

**Çözüm 5 :**

**a)** Temel bileşenler için katsayılar
(Burada parantez içindekiler korelasyon katsayılarıdır. ) :

| | $\hat{\underline{e}}_1$ | $\hat{\underline{e}}_2$ | $\hat{\underline{e}}_3$ |
|---|---|---|---|
| ln (uzunluk) | 0.683 (0.99) | 0.162 | 0.712 |
| ln (genişlik) | 0.510 (0.97) | 0.591 | -0.624 |
| ln (yükseklik) | 0.522 (0.97) | -0.790 | -0.321 |
| $\hat{\lambda}_i$ | $24.31\times10^{-3}$ | $0.63\times10^{-3}$ | $0.38\times10^{-3}$ |
| $y_i$ temel bileşeninin toplam değişimi açıklama oranı | % 96 | % 98.5 | % 100 |

Birinci temel bileşen toplam varyansın % 96'sını açıklamaktadır. Örneklem değişiminin ifade edilebilmesi için bir Temel bileşen yeterlidir.

$$\begin{aligned} \hat{Y}_1 &= 0.683\ln(X_1) + 0.51\ln(X_2) + 0.522\ln(X_3) \\ &= \ln\left(X_1^{0.683} X_2^{0.51} X_3^{0.522}\right) \end{aligned}$$

olduğundan 1. Temel bileşen düzeltilmiş 3 boyutlu bir kutunun hacminin logaritmasının alınması görünümündedir. Örneğin üst kabuğun döndürülmüş hali için düzeltilmiş yükseklik

$$X_3^{0.522}$$

dir.