Karadeniz Teknik Üniversitesi

Fen Fakültesi
İstatistik ve Bilgisayar Bilimleri Bölümü

Çok Değişkenli İstatistiksel Yöntemler

BÖLÜM 2 - Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler

Dr. Öğr. Üyesi Uğur ŞEVİK

Ders İçeriği

- Varyans – Kovaryans Matrisi, Korelasyon Matrisi
- Açıklayıcılık Katsayısı
- Saçılım Grafikleri
- Genelleştirilmiş Varyans
- Gruplara İlişkin Çözümlemeler

Ders Hedefleri

- Varyans – Kovaryans Matrisi, Korelasyon Matrisi
- Açıklayıcılık Katsayısı
- Saçılım Grafikleri
- Çok değişkenli analizdeki tanımlayıcı istatistiklerden biri olan Genelleştirilmiş Varyans'ın tanımı ve saplanması
- Çok değişkenli problemlerde benzer özellikli değişkenlerin gruplandırılması, bu grupların grup içi analizlerini yapılması ve gruplar arası ilişkilerin araştırılması

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

Çok değişkenli istatistiksel analizde genellikle n satır (gözlem) ve p sütundan (değişken) oluşan $n \times p$ boyutlu X veri matrisi ile ilgilenir. X'in elemanları x_{ij} ile gösterilirken sütun vektörleri (değişkenler) $x_1, x_2, \ldots, x_p$ ile gösterilir.

- DEĞİŞKENLER -

Gözlem - NO -	x_1	x_2	...	x_j	...	x_p
1	x_{11}	x_{12}	...	x_{1j}	...	x_{1p}
2	x_{21}	x_{22}	...	x_{2j}	...	x_{2p}
⋮	⋮	⋮				⋮
n	x_{n1}	x_{n2}	...	x_{nj}	...	x_{np}

- X veri Matrisi -

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

Tanımlayıcı istatistikler

çok değişkenli analizde değişkenler kümesini tanımlamak için ortalama vektörü ($\bar{x}$), kovaryans değeri, varyans-kov. matrisi, korelasyon katsayısı ve korelasyon matrisinden yararlanılır.

a) ortalama vektörü: çok değişkenli analizde tek bir ortalama değil p tane ortalama vektörü vardır. gösterimi ortalama vektörü yardımıyla yapılır.

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

$$\bar{X}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}$$ ile hesaplanır.

p tane değişkene ilişkin ortalama vektörü;

$$\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}_{p \times 1} \Rightarrow \bar{X}' = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_p \end{bmatrix}_{1 \times p}$$ ile gösterilir.

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

b) Varyans, Kovaryans ve Varyans-Kovaryans Matrisi : Herhangi bir değişkene ait varyans aşağıdaki gibi hesaplanır.

$$s_j^2 = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{n-1} = \frac{\sum_{i=1}^{n} x_{ij}^2 - \frac{\left(\sum_{i=1}^{n} x_{ij}\right)^2}{n}}{n-1}$$

$$= \frac{K.Tx_j}{n-1}$$

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

Burada paydaki ifadeye düzeltilmiş kareler toplamı denir. Yukarıdaki eşitlikte görüldüğü gibi varyans, ortalamaya göre düzeltilmiş değerlerin karelerinin toplanması (pay) ve serbestlik derecesine bölünmesi (payda) ile hesaplanır. Ortalamaya göre düzeltilmiş değerlerin kareleri toplamı kısaca kareler toplamı (KT) olarak adlandırılır.

Varyansın çok değişkenli analizdeki karşılığı ise $p \times p$ boyutlu simetrik varyans-kovaryans matrisidir. $S = (s_{ij})_{p \times p}$ ile gösterilir. Bu matriste s_{ii} yani $i=j$ elemanları değişkenlere ilişkin varyansı verirken, s_{ij} yani $i \neq j$ elemanları da kovaryans değerini verir.

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

$$S = \begin{bmatrix} var(X_1) & kov(X_1X_2) & \cdots & kov(X_1X_p) \\ kov(X_2X_1) & var(X_2) & \cdots & kov(X_2X_p) \\ \vdots & \vdots & & \vdots \\ kov(X_pX_1) & kov(X_pX_2) & \cdots & var(X_p) \end{bmatrix}$$

$$= \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix}$$

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

Ölçümle belirtilen (sürekli sayısal veri tipindeki) (örneğin; yaş, boy ...) iki değişken arasındaki doğrusal ilişki ölçülerinden biri olan kovaryans değerinin formülü X veri matrisindeki ilk iki değişken için aşağıdaki gibidir.

$$s_{12} = \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{n-1}$$

$$= \frac{\sum_{i=1}^{n} (x_{i1} \cdot x_{i2}) - \left(\sum_{i=1}^{n} x_{i1}\right)\left(\sum_{i=1}^{n} x_{i2}\right)}{n-1} = \frac{ÇT_{x_1 x_2}}{n-1}$$

Bu formülün payına ortalamaya göre düzeltilmiş çarpımlar toplamı (ÇT) denir.

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

Değişkenlerden birinin değeri artarken diğerinin değeri de artıyorsa, ya da biri azalıyorken diğeri de azalıyorsa iki değişken arasındaki kovaryans değeri pozitiftir. Tersi durumda ise negatif olur. Eğer değişkenler arasında belirgin bir ilişki yoksa kovaryans sıfıra yakın bir değer alır.

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

c) Korelasyon (İlişki) Katsayısı ve Korelasyon Matrisi: Kovaryans değeri gibi korelasyon katsayısı da (Pearson's Product Moment correlation coefficient) iki değişken arasındaki doğrusal ilişkinin ölçüsüdür ve "r" ile gösterilir. r incelenen değişkenlerin birimlerinden bağımsızdır. -1 ile +1 arasında değişir $(-1 \leq r \leq 1)$. r, -1 ve +1'e yaklaştıkça ilişkinin kuvveti artarken 0'a yaklaştıkça ilişkinin kuvveti azalır. -1 tam negatif ilişkiyi +1 ise tam pozitif ilişkiyi gösterir. Korelasyon katsayısı değişken arasındaki doğrusal ilişkiyi tanımladığı için doğrusal olmayan ilişkiler korelasyon katsayı ile temsil edilemez. X_1 ve X_2 gibi iki değişken için korelasyon katsayısı;

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

$$r_{x_1,x_2} = \frac{\sum_{i=1}^{n} x_{i1} x_{i2} - \dfrac{\sum_i x_{i1} \sum_i x_{i2}}{n}}{\sqrt{\left(\sum x_{i1}^2 - \dfrac{\left(\sum_i x_{i1}\right)^2}{n} \right)\left(\sum x_{i2}^2 - \dfrac{\left(\sum_i x_{i2}\right)^2}{n} \right)}}$$

$$= \frac{\text{ÇT}_{x_1 x_2}}{\sqrt{KT_{x_1} \, KT_2}} = \frac{\text{KOV}(x_1, x_2)}{\sqrt{\text{Var}(x_1)} \sqrt{\text{Var}(x_2)}}$$

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

Korelasyon Matrisi $R = (r_{ij})_{p \times p}$ ile gösterilir.

$$R = \begin{bmatrix} r_{x_1x_1} & r_{x_1x_2} & \cdots & r_{x_1x_p} \\ r_{x_2x_1} & r_{x_2x_2} & \cdots & r_{x_2x_p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{x_px_1} & r_{x_px_2} & \cdots & r_{x_px_p} \end{bmatrix} = \begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{21} & 1 & r_{23} & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & r_{p3} & \cdots & 1 \end{bmatrix}$$

Kovaryans matrisinden korelasyon matrisini elde edebiliriz.

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

Korelasyon Katsayısının Nitelendirilmesi

$\|r\|$	Açıklama
0,00 - 0,19	İlişki yok ya da önemsiz.
0,20 - 0,39	Zayıf (düşük) ilişki
0,40 - 0,59	Orta düzey ilişki
0,60 - 0,79	Dikkate değer bir ilişki
0,80 - 1,00	Yüksek ilişki

0,00 - 0,19	İlişki yok ya da önemsiz.
0,20 - 0,39	Zayıf ilişki
0,40 - 0,69	Orta düzey ilişki
0,70 - 0,89	Kuvvetli ilişki
0,80 - 1,00	Çok yüksek ilişki

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

– Korelasyona ilişkin varsayımlar –

X_i ve X_j gibi iki değişken için.

* X_i'nin her değeri için X_j değerlerinin normal dağılan bir alt kümesi vardır. Bu durum X_j için de geçerlidir.
* X_i ve X_j'nin birleşik dağılımı iki değişkenli normal dağılım gösterir.
* X_i ve X_j değerlerinin alt kümeleri eşit varyansa sahiptir.

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

* Açıklayıcılık Katsayısı *

(R^2)

İncelenen değişkenler bağımlı ve bağımsız değişken olarak tanımlanabiliyorsa açıklayıcılık katsayısı bağımlı değişkendeki değişimin yüzde kaçının bağımsız değişken tarafından açıklanabildiğini belirtir. İki değişken arasında doğrusal ilişki olması durumunda ilişki katsayısının karesi açıklayıcılık katsayısına (R^2) eşittir ($R^2=r^2$).

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

Örnek: Aşağıdaki tabloda verilen ve en az bir canlı doğum yapmış kadınlara ilişkin veri matrisini ele alarak ortalama vektörü, S, R ve değer matrisleri bulalım.

Kadın	X_1	X_2	X_3	X_4
1	13,0	1	20	40
2	14,0	1	25	50
3	12,5	1	40	40
4	12,0	2	22	60
5	12,5	2	33	60
6	12,0	3	35	55
7	11,0	4	21	50
8	10,0	5	25	70
9	10,0	6	42	50
10	12,0	4	30	75
11	10,5	3	35	60
12	10,0	4	28	70
13	11,0	5	25	65
14	8,0	6	40	45
15	11,5	2	33	65

X_1: Kadının hemoglobin düzeyi

X_2: canlı doğum sayısı

X_3: Kadının yaşı

X_4: Çocuk bakım bilgi puanı

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

Kadın	X_1	X_2	X_3	X_4
1	13,0	1	20	40
2	14,0	1	25	50
3	12,5	1	40	40
4	12,0	2	22	60
5	12,5	2	33	60
6	12,0	3	35	55
7	11,0	4	21	50
8	10,0	5	25	70
9	10,0	6	42	50
10	12,0	4	30	75
11	10,5	3	35	60
12	10,0	4	28	70
13	11,0	5	25	65
14	9,0	6	40	45
15	11,5	2	33	65
$\sum x$	171,0	48,000	454,000	855,000
$\sum x^2$	1875,00	203,000	14476,000	50425,000
$\bar{x}$	11,400	3,267	30,267	57,000
s^2	1,829	3,067	52,495	120,714
s	1,352	1,751	7,245	10,987

X_1: Kadının hemoglobin düzeyi

X_2: Canlı doğum sayısı

X_3: Kadının yaşı

X_4: Çocuk bakım bilgi puanı

$* \sum_i x_1 x_2 = 530$

$* \sum_i x_2 x_4 = 2870$

$* \sum_i x_2 x_3 = 1529$

$* \sum_i x_1 x_4 = 8697,5$

$* \sum_i x_1 x_3 = 5129,5$

$* \sum_i x_3 x_4 = 25655$

$$\bar{X} = \begin{bmatrix} 11,400 \\ 3,267 \\ 30,267 \\ 57,00 \end{bmatrix}$$

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

$$s_1^2 = var(x_1) = s_{11} = \frac{1875 - \frac{(171)^2}{15}}{14} = 1,829$$

$$\sum_i x_{i1} x_{i2} = (13).(1) + (14)(1) + \dots + (11,5).(2) = 530$$

$$Kov(x_1 x_2) = \frac{530 - \frac{(171).(48)}{15}}{15-1} = -2,043$$

S_i^2

Kov X_iX_j

$$S = \begin{bmatrix} 1,829 & -2,043 & -3,283 & -3,536 \\ -2,043 & 3,067 & 3,281 & 5,500 \\ -3,283 & 3,281 & 52,485 & -15,828 \\ -3,536 & 5,500 & -15,828 & 120,714 \end{bmatrix}$$

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

$$r_{21} = r_{x_1 x_2} = \frac{-2{,}043}{\sqrt{1{,}828}\,\sqrt{3{,}067}} = -0{,}863$$

$$R = \begin{bmatrix} 1 & -0{,}863 & -0{,}336 & -0{,}238 \\ -0{,}863 & 1 & 0{,}258 & 0{,}286 \\ -0{,}336 & 0{,}258 & 1 & -0{,}200 \\ -0{,}238 & 0{,}286 & -0{,}200 & 1 \end{bmatrix}$$

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

S ve R Matrisleri

Değişken	Ortalama vektör	X1	X2	X3	X4
X1	11,400	1,828	-2,043	-3,293	-3,536
X2	3,267	-0,863	3,067	3,281	5,500
X3	30,267	-0,336	0,259	52,455	-15,829
X4	57,000	-0,238	0,286	-0,200	120,714

Korelasyon Matrisi

Matrisi

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

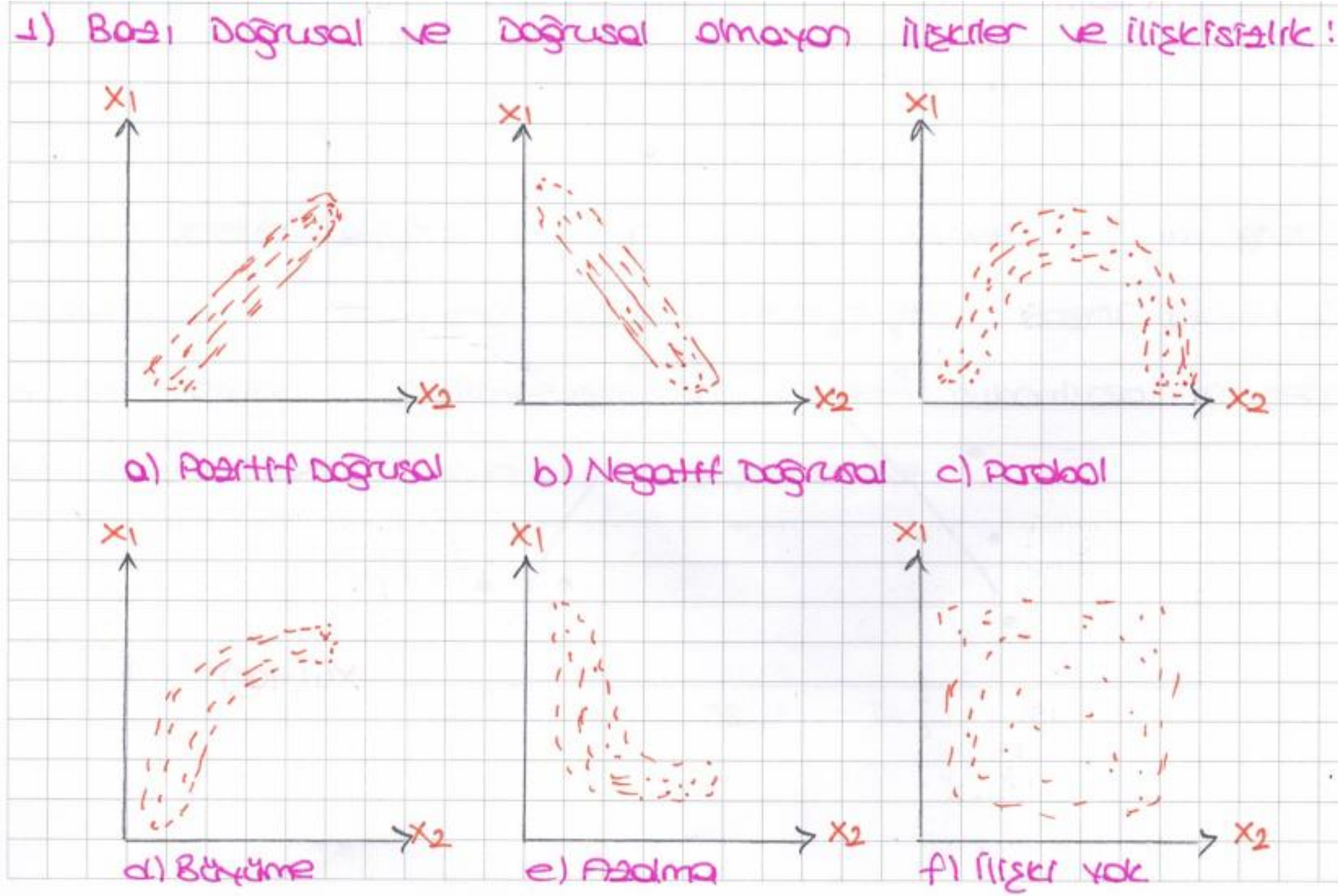
- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

- Saçılım Grafiklerinin İncelenmesi ve Önemi -

İki sayısal (sürekli, kesikli) değişken arasındaki ilişkilerin grafiklerle incelenmesi pearson korelasyon katsayısının doğru kullanımı ve ilişkinin modellenmesi açısından son derece önemlidir. Bu amaçlar için en iyi yöntem değişkenlerin saçılım grafiklerinin incelenmesidir.

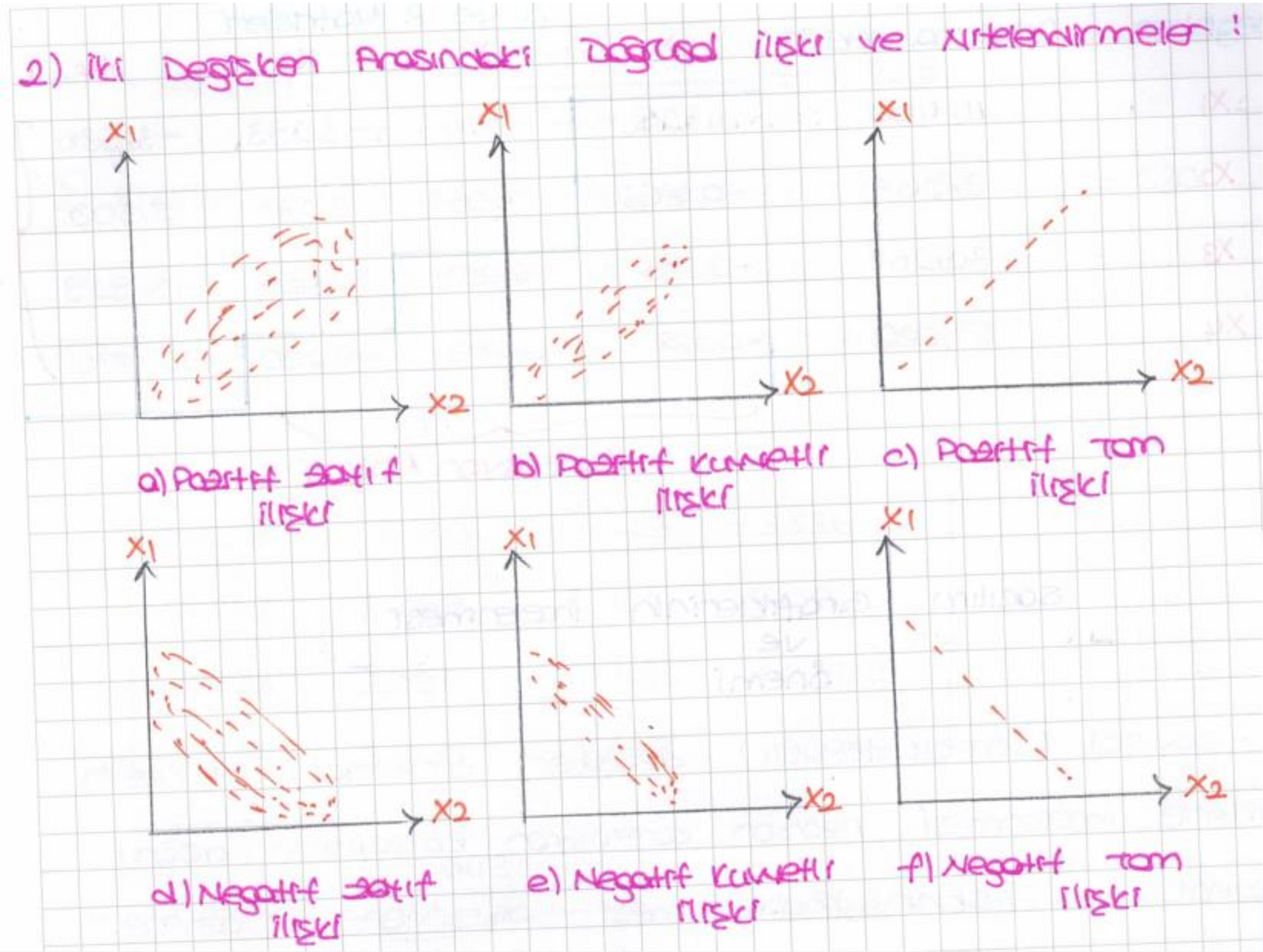
ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**



ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

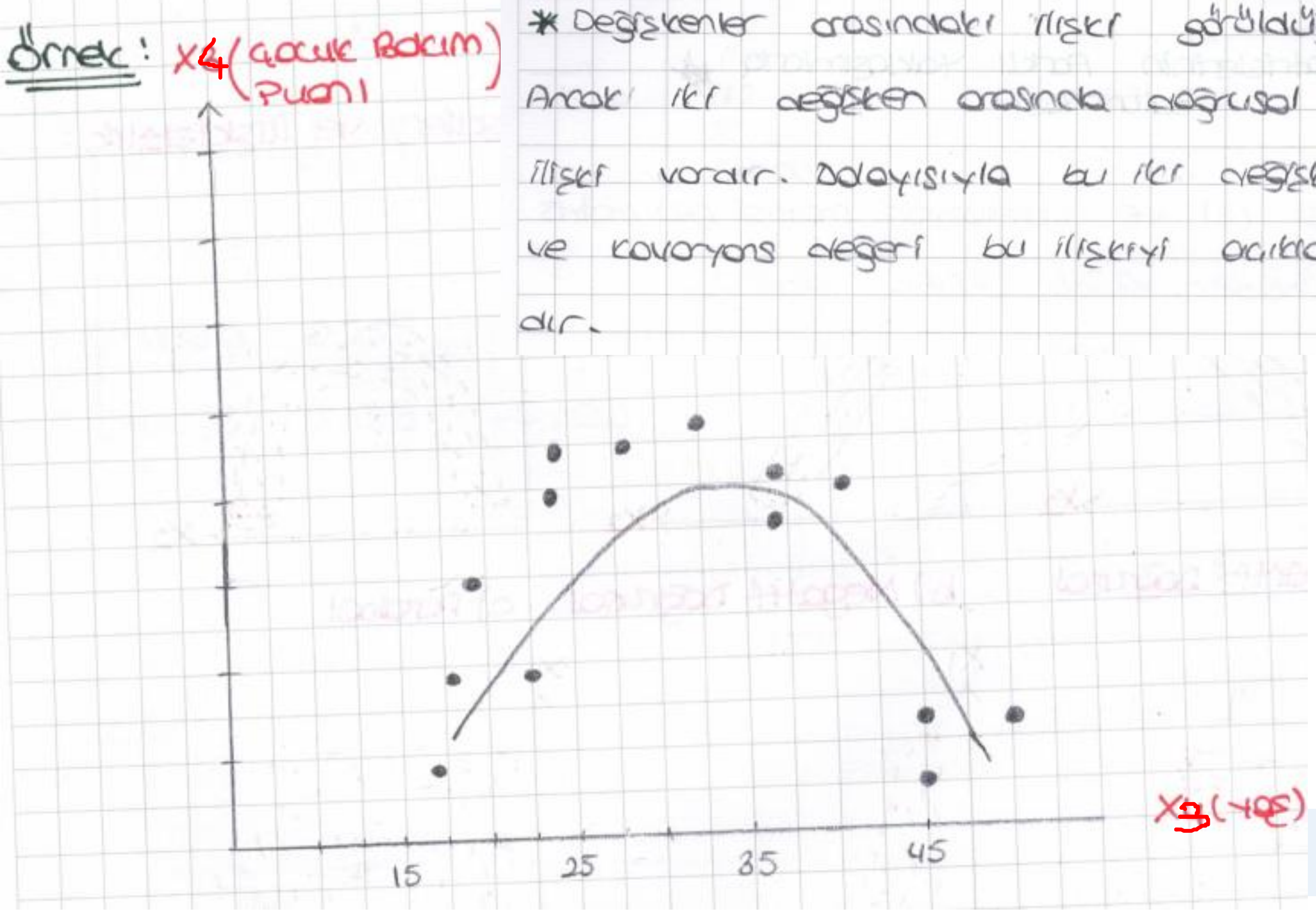
- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**



ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

* Değişkenler arasındaki ilişki görüldüğü üzere doğrusal değildir. Ancak iki değişken arasında doğrusal olmayan kuvvetli bir ilişki vardır. Dolayısıyla bu iki değişkene ilişkin R katsayısı ve kovaryans değeri bu ilişkiyi açıklamakta yetersiz kalmaktadır.



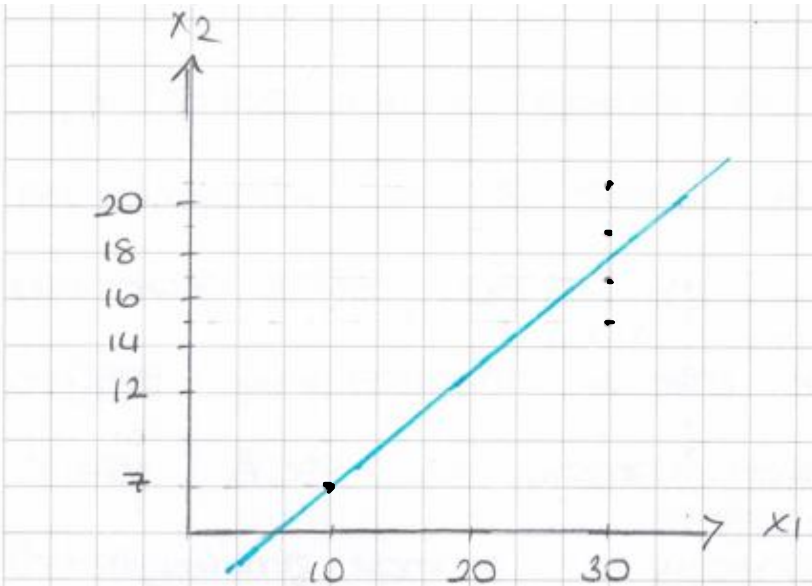
r_{34}= - 0,20

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

Örnek:

x_1	x_2
10	7
30	15
30	17
30	19
30	21



* Bu örnekte $r = 0{,}81$ bulunur. Ancak verilerin saçılım grafiği incelendiğinde bu kadar yüksek ilişkinin tek nedeni (10,7) gözleminin olduğu görülmektedir. Bu gözlem çıkarılırsa x_1 değişkeni sabit olacağı için r hesaplanamaz.

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

S ve R Matrislerinin Farklı Yaklaşımlarla Bulunması

Varyans-kovaryans matrisi (S) ve korelasyon matrisi (R) matris işlemleri yardımıyla bulunabilir.

$$S = (X'X - \bar{X}\bar{X}'n)/(n-1)$$

$$S = \frac{\text{Düzeltilmiş kareler ve çarpımlar toplamı matrisi}}{\text{Gözlem sayısı} - 1}$$

* $X'X$ matrisine kareler ve çarpımlar toplamı (KÇT) matrisi denir.

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

$$
X'X = KAT = \begin{bmatrix}
\sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} & \sum_{i=1}^{n} x_{i1}x_{i3} & \sum_{i=1}^{n} x_{i1}x_{i4} \\
\sum_{i=1}^{n} x_{i2}x_{i1} & \sum_{i=1}^{n} x_{i2}^2 & \sum_{i=1}^{n} x_{i2}x_{i3} & \sum_{i=1}^{n} x_{i2}x_{i4} \\
\sum_{i=1}^{n} x_{i3}x_{i1} & \sum_{i=1}^{n} x_{i3}x_{i2} & \sum_{i=1}^{n} x_{i3}^2 & \sum_{i=1}^{n} x_{i2}x_{i4} \\
\sum_{i=1}^{n} x_{i4}x_{i1} & \sum_{i=1}^{n} x_{i4}x_{i2} & \sum_{i=1}^{n} x_{i4}x_{i3} & \sum_{i=1}^{n} x_{i4}^2
\end{bmatrix}
$$

$$
KAT = \begin{bmatrix}
1975 & 530 & \text{---} & \text{---} \\
 & 203 & \text{---} & \text{---} \\
\text{sim.} & & & \\
 & & &
\end{bmatrix}
$$

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- **Çok Değişkenli Analizde Veri Matrisi ve Tanımlayıcı İstatistikler**

$$\bar{X}\bar{X}' = \begin{bmatrix} 11,4 \\ 3,267 \\ 30,267 \\ 57,000 \end{bmatrix} \begin{bmatrix} 11,4 & 3,267 & 30,267 & 57,000 \end{bmatrix}$$

$$= \begin{bmatrix} 129,96 & 37,24 & 345,04 & 649,8 \\ & 10,67 & 98,88 & 186,22 \\ & & 916,09 & 1725,22 \\ \text{sim.} & & & 3249 \end{bmatrix} \cdot n$$

Genelleştirilmiş Varyans

- pxp boyutlu bir S matrisi, p tane varyans ve p(p-1)/2 tane kovaryans değeri içerir. Eğer S matrisinin determinantı alınırsa bu sayısal değere genelleştirilmiş varyans denir ve |S| ile gösterilir.
- Önceki örnek için genelleştirilmiş varyans |S|=6712 bulunur.
- Genelleştirilmiş varyans S matrisine ilişkin öz değerler yardımıyla da hesaplanabilir.

$$|S| = \lambda_1 \times \lambda_2 \times \ldots\ldots \times \lambda_p$$

Önceki örnek için,

|S|= =124,491 x 49,731 x 3,577 x 0,303=6712 bulunur.

Gruplamalara ilişkin Çözümler

- Birçok çok değişkenli çözümleme (Diskriminant analizi, MANOVA, iki ve daha fazla grubun incelenmesini gerektirir.
- Korelasyon ve Kovaryans uygulaması örneğinde, veriyi ilk 7 gözlem kentsel, son 8 gözlem kırsal bölge kadınlarına ait veri grupları olarak ayıralım. Dolayısıyla veri 2 gruptan oluşacaktır.
- Araştırıcı, incelenen 4 değişken açısından her bir gruptaki kadınların birbirine ne derece benzer olduğunu ve de iki gruptaki kadınların incelenen değişkenler açısından ne derece farklı olduğunu belirlemek isteyebilir. Bu amaç için:
- 1. Veri, her bir grup içindeki benzerlikleri belirlemek amacıyla ayrı ayrı özetlenir ise bu işleme grup içi analiz (within-group analysis) adı verilir.
- 2. Veri, incelenen değişkenler açısından gruplar arasındaki farklılıkları belirlemek için özetlenir ise bu işleme gruplar arası analiz (between-group analysis) adı verilir.

Gruplamalara ilişkin Çözümler

- Önceki örnekteki ilk iki değişken (X_1 ve X_2) dikkate alınarak ve veriyi kentsel ve kırsal olarak iki gruba ayırarak elde edilen tanımlayıcı istatistikler aşağıdaki tabloda verilmiştir.
- Tabloda ortalamaya göre düzeltilmiş (**her bir gözlemin dağılım ortalamasından çıkartılarak düzeltildiği**) değerler de yer almakta ve hesaplamalarda bu değerlerden de yararlanılmaktadır.
- Tabloda görüldüğü gibi, ham verilerin varyansı ile ortalamaya göre düzeltilmiş verilerin varyansı aynıdır.

Kadın	X_1	X_2
1	13,0	1
2	14,0	1
3	12,5	1
4	12,0	2
5	12,5	2
6	12,0	3
7	11,0	4
8	10,0	5
9	10,0	6
10	12,0	4
11	10,5	3
12	10,0	4
13	11,0	5
14	9,0	6
15	11,5	2

X_1: Kadının hemoglobin düzeyi

X_2: canlı doğum sayısı

Gruplamalara ilişkin Çözümler

Tablo a . Ez Az Bir Çocuklu Kadınlara İlişkin Bulgular (Kentsel)

			Ortalamaya Göre Düzeltilmiş	
Sıra No	X_1	X2	X_1	X2
1	13,0	1	0,572	-1
2	14,0	1	1,572	-1
3	12,5	1	0,072	-1
4	12,0	2	-0,428	0
5	.12,5	2	0,072	0
6	12,0	3	-0,428	1
7	11,0	4	-1,428	2
$\overline{X}$	12,428	2	0,000	0,000
KT			5,214	8
S^2	0,869	1,333	0,869	1,333

Gruplamalara ilişkin Çözümler

Tablo b. Ez Az Bir Çocuklu Kadınlara İlişkin Bulgular (Kırsal)

			Ortalamaya Göre Düzeltilmiş	
Sıra No	X1	X2	X1	X2
8	10,0	5	-0,500	0,625
9	10,0	6	-0,500	1,625
10	12,0	4	1,500	-0,375
11	10,5	3	0,000	-1,375
12	10,0	4	-0,500	-0,375
13	11,0	5	0,500	0,625
14	9,0	6	-1,500	1,625
15	11,5	2	1,000	-2,375
$\bar{X}$	10,5	4,375	0,000	0,000
KT			6,500	13,875
S^2	0,929	1,982	0,929	1,982

A) Grup İçi Analiz

- Grup içi analiz için gerekli olabilen *ortak (pooled) grup içi düzeltilmiş* KÇT, *ortak varyans-kovaryans* ve *ortak R matrislerini* elde etmek amacıyla *gruplara göre düzeltilmiş KÇT*, *varyans-kovaryans* ve *korelasyon matrisleri* aşağıda verilmiştir.
- *Aşağıda verilen Düz.KÇT1 matrisindeki (1,1) elemanı olan 5.214 ve (2,2) elemanı olan 8,000 doğrudan Tablo a'dan alınırken (ki bunlar düzeltilmiş değerlerin kareleri toplamlarıdır), (1,2) ya da (2,1) elemanı olan -5,500 (ki bu düzeltilmiş değerlerin çarpımlar toplamıdır), (0,572)(-1)+… +(-1,428)(2) ile de bulunabilir.*

$$\sum x^2 - \left(\sum x\right)^2 / n \qquad \sum x_1 x_2 - \sum x_1 \sum x_2 / n$$

$$Düz.KÇT_1 = \begin{vmatrix} 5,214 & -5,500 \\ -5,500 & 8,000 \end{vmatrix} \quad S_1 = \begin{vmatrix} 0,869 & -0,917 \\ -0,917 & 1,333 \end{vmatrix} \quad R_1 = \begin{vmatrix} 1,000 & -0,852 \\ -0,852 & 1,000 \end{vmatrix}$$

$$Düz.KÇT_2 = \begin{vmatrix} 6,500 & -6,000 \\ -6,000 & 13,875 \end{vmatrix} \quad S_2 = \begin{vmatrix} 0,929 & -0,857 \\ -0,857 & 1,982 \end{vmatrix} \quad R_2 = \begin{vmatrix} 1,000 & -0,632 \\ -0,632 & 1,000 \end{vmatrix}$$

A) Grup İçi Analiz

- İki grubun grup içi düzeltilmiş KÇT matrislerini toplarsak, ortak (pooled) grup içi düz. KÇT matrisi elde edilir (sadece W ile de gösterilebilir):

- Ortak Grup İçi Düz. KÇT Matrisi = W = $\begin{bmatrix} 11{,}714 & -11{,}500 \\ -11{,}500 & 21{,}875 \end{bmatrix}$

- *Ortak varyans-kovaryans matrisi (Sw), ortak grup içi düz. KÇT matrisinin (W) serbestlik derecesi olan* $\sum_{i=1}^{k} n_i - k$ *'ya bölünmesi ile bulunabilir (burada k: grup sayısıdır). Örneğimiz için serbestlik derecesi (n1 + n2)-2 = 7+8-2=13 olarak bulunur ve Sw aşağıdaki gibi elde edilir.*

$$S_w = \begin{bmatrix} 0{,}901 & -0{,}885 \\ -0{,}885 & 1{,}683 \end{bmatrix}$$

- *Benzer şekilde grup içi ortak korelasyon matrisi (Rw), Sw yardımıyla aşağıdaki gibi bulunur.*

$$R_w = \begin{bmatrix} 1{,}000 & -0{,}718 \\ -0{,}718 & 1{,}000 \end{bmatrix}$$

$$\frac{\text{Kov}(x_1, x_2)}{\sqrt{\text{Var}(x_1)}\sqrt{\text{Var}(x_2)}}$$

B) Gruplar Arası Analiz

- Gruplar arası kareler toplamı, gruplardaki ortalamaların genel ortalamalardan ne derece farklı olduğu konusunda bilgi verir.

$$KT_j = \sum_{i=1}^{k} n_i(\bar{x}_{ji} - \bar{x}_{j.})^2$$

KT_j = j. değişken için gruplar arası KT

k = Grup sayısı

n_i = i. gruptaki gözlem sayısı

$\bar{x}_{ji}$ = i. gruptaki j. değişkenin ortalaması

$\bar{x}_{j.}$ = Toplam gözlem sayısı için j. değişkenin ortalaması

- Örneğimizdeki X1ve X2 değişkenleri için düzeltilmiş gruplar arası kareler toplamları,

$$KT_1 = 7(12,428-11,40)^2+8(10,5-11,40)^2=13,886$$

$$KT_2 = 7(2,000-3,267)^2+8(4,375-3,267)^2=21,058$$

B) Gruplar Arası Analiz

- *Bu matrisin çarpımlar toplamı (matrisin (1,2) ya da (2,1) elemanı) ise aşağıdaki gibi bulunabilir.*

$$ÇT_{12}=7(12,428-11,4)(2-3,267)+8(10,5-11,4)(4,375-3,267)=-17,100$$

- *Buradan, gruplar arası için düzeltilmiş KÇT matrisi, (B)*

$$\textit{Gruplar Arası İçin Düz. KÇT Matrisi} = SSCP_B = \begin{bmatrix} 13,886 & -17,100 \\ -17,100 & 21,058 \end{bmatrix}$$

- *W ve B matrislerinin toplanması ile toplam için düzeltilmiş kareler ve çarpımlar toplamı matrisi (T) elde edilir. Diğer bir deyişle, bu üç kareler toplamı arasında bir ilişki vardır. Bu ilişki,*
- *Düz. T = Düz. B + Düz. W ile verilir.*

$$\textit{Düz. } SSCP_T = \begin{bmatrix} 13,886 & -17,100 \\ -17,100 & 21,058 \end{bmatrix} + \begin{bmatrix} 11,714 & -11,500 \\ -11,500 & 21,875 \end{bmatrix} = \begin{bmatrix} 25,600 & -28,500 \\ -28,500 & 42,933 \end{bmatrix}$$

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

- Sunum hazırlanırken aşağıdaki kaynaktan yararlanılmıştır.