STATS 206
Applied Multivariate Analysis
Lecture 1: Introduction

Prathapasinghe Dharmawansa

Department of Statistics
Stanford University
Autumn 2013

Agenda

- Objectives
- Multivariate Analysis: Extension of Univariate Statistical Methods
- A broad view of Multivariate Analytical techniques
- Data organization and visualization
- *A brief introduction to vectors and matrices (a little digression)*
- Descriptive statistics
- Statistical distance
- Random vectors and matrices

Objectives

- To learn basic techniques for analyzing multi-dimensional data
- To study multivariate distributions (in particular Gaussian class distributions)
- To study various techniques used in
  - dimensionality reduction
  - sorting and grouping
  - prediction
  - hypothesis construction and testing
  - determining the dependence among variables

Multivariate Analysis

- Simultaneous measurements of many variables: in practical applications
  - variables related to cancer patient responses to radiotherapy
  - variables related to air-pollution
  - variables related to solar system
  - variables related to user behavior in a cellular network
- Need to understand the interactions/relationship among variables

Multivariate Statistical Analysis

Multi-dimensionality $\Longrightarrow$ high degree of analytical/numerical difficulty

We use Algebraic and Geometric methods $\Longrightarrow$ vectors/matrices and numerical computations

List of Topics

- Graphical representation of data and introduction to basic matrix algebra
- Multivariate Normal distribution
- Inferences about a mean vector
- Comparison of several multivariate means and MANOVA
- Multivariate linear regression
- Principle component analysis
- Factor analysis
- Canonical correlation analysis
- Discrimination and classification
- Clustering and multidimensional scaling

Data Organization

Measurements on variables/characteristics $\Longrightarrow$ Data

Assume one makes n measurements on p variables

	Variable 1	Variable 2	$\cdots$	Variable k	$\cdots$	Variable p
Item 1:	x_{11}	x_{12}	$\cdots$	x_{1k}	$\cdots$	x_{1p}
Item 2:	x_{21}	x_{22}	$\cdots$	x_{2k}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
Item j:	x_{j1}	x_{j2}	$\cdots$	x_{jk}	$\cdots$	x_{jp}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
Item n:	x_{n1}	x_{n2}	$\cdots$	x_{nk}	$\cdots$	x_{np}

$$\Longrightarrow \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots x_{np} \end{bmatrix} \quad \text{Rectangular array}$$

Data Visualization

"A picture is worth a thousand words"

- Powerful computer programs/display equipments enable graphical representation of data
- Types of graphs
  - one dimensional dot plot
  - box plot
  - 2D/3D scatter plots (limited to $3D$ visualization)
  - multiple 2D scatter plots
  - growth curves
  - Chernoff faces

Data Example
Scatter plots for paper-Quality measurements

- Data Example: (Paper quality measurements)
  - Paper: manufactured in continuous sheets a few feet wide
  - Due to the orientation of fibers, it has a different strength when measured in the direction produced by the machine from than when measured across (or at the right angles to the machine direction).
  - The following table shows:
    * x_1: density (g/cm^3)
    * x_2: strength (pounds) in the machine direction
    * x_3: strength (pounds) in the cross direction

Paper-Quality Data Example: Table – Part 1

Specimen	Density	Strength	
		Machine direction	Cross direction
1	0.801	121.41	70.42
2	0.824	127.7	72.47
3	0.841	129.2	78.2
4	0.816	131.8	74.89
5	0.84	135.1	71.21
6	0.842	131.5	78.39
7	0.82	126.7	69.02
8	0.802	115.1	73.1
9	0.828	130.8	79.28
10	0.819	124.6	76.48
11	0.826	118.31	70.25
12	0.802	114.2	72.88
13	0.81	120.3	68.23
14	0.802	115.7	68.12

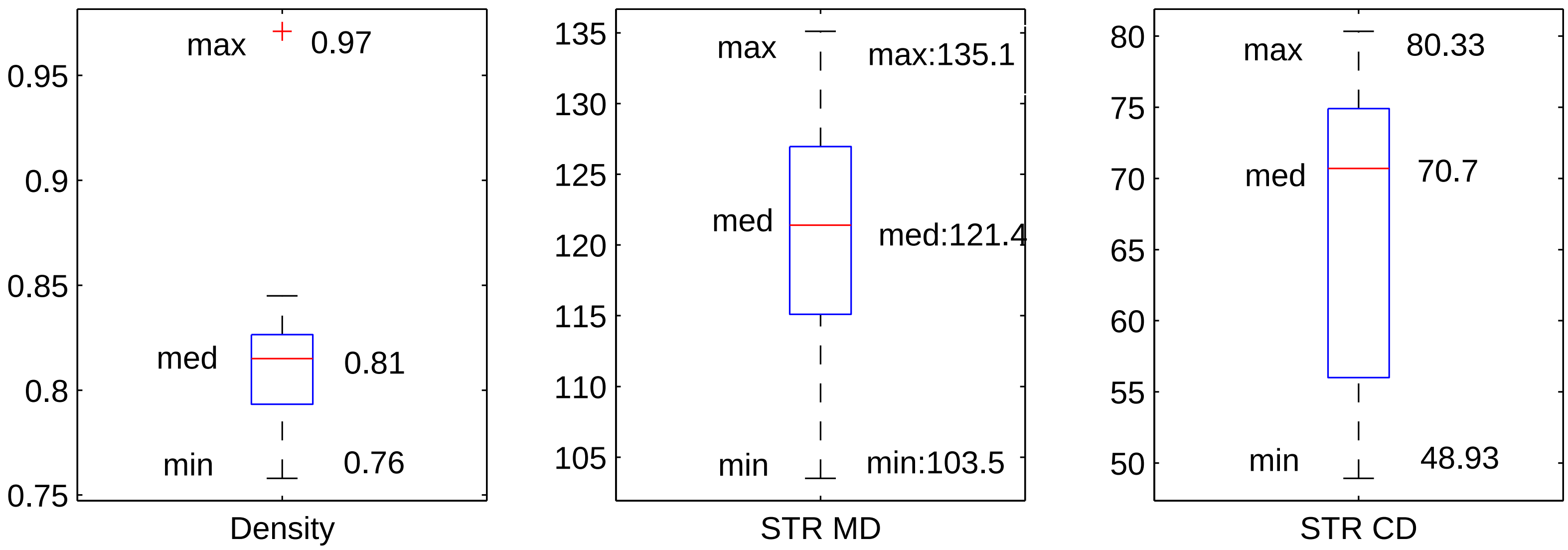
Paper-Quality Data Example: Table – Part 2

Specimen	Density	Strength	
		Machine direction	Cross direction
15	0.832	117.51	71.62
16	0.796	109.81	53.1
17	0.759	109.1	50.85
18	0.77	115.1	51.68
19	0.759	118.31	50.6
20	0.772	112.6	53.51
21	0.806	116.2	56.53
22	0.803	118	70.7
23	0.845	131	74.35
24	0.822	125.7	68.29
25	0.971	126.1	72.1
26	0.816	125.8	70.64
27	0.836	125.5	76.33
28	0.815	127.8	76.75

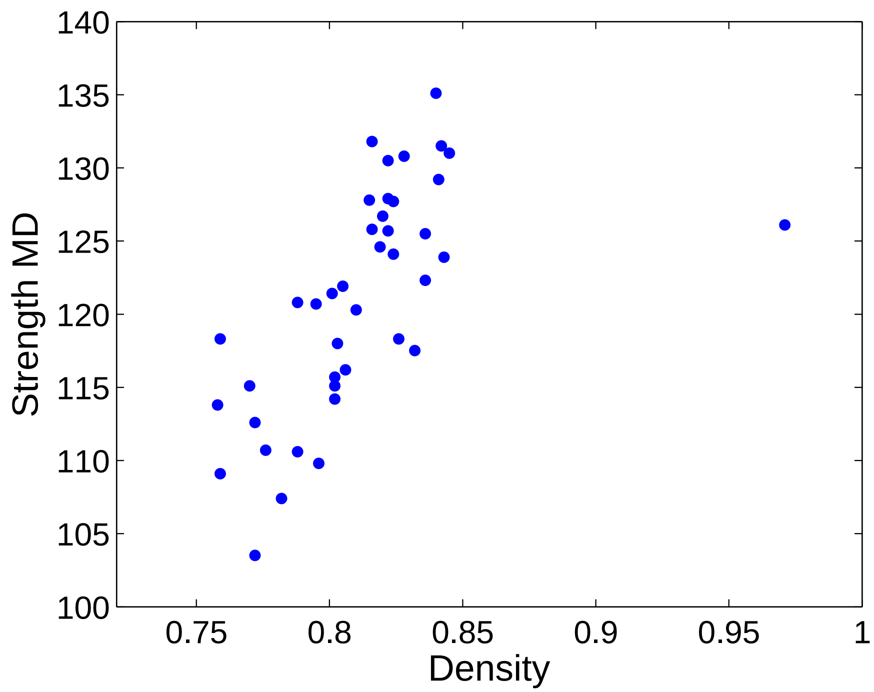
Paper-Quality Data Example: Table – Part 3

Specimen	Density	Strength	
		Machine direction	Cross direction
29	0.822	130.5	80.33
30	0.822	127.9	75.68
31	0.843	123.9	78.54
32	0.824	124.1	71.91
33	0.788	120.8	68.22
34	0.782	107.4	54.42
35	0.795	120.7	70.41
36	0.805	121.91	73.68
37	0.836	122.31	74.93
38	0.788	110.6	53.52
39	0.772	103.51	48.93
40	0.776	110.71	53.67
41	0.758	113.8	52.42

Box Plot(s) from Paper-Quality Data



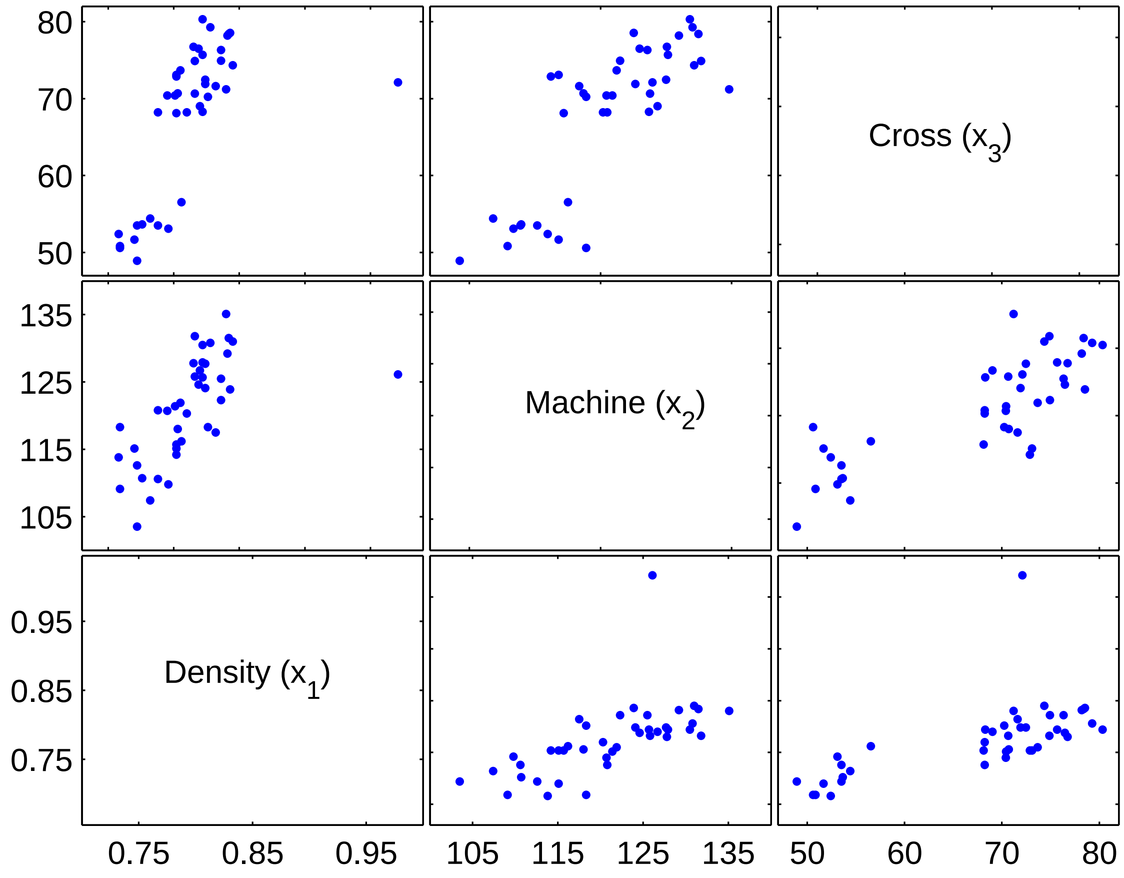
Single Scatter Plot: Strength MD vs Density



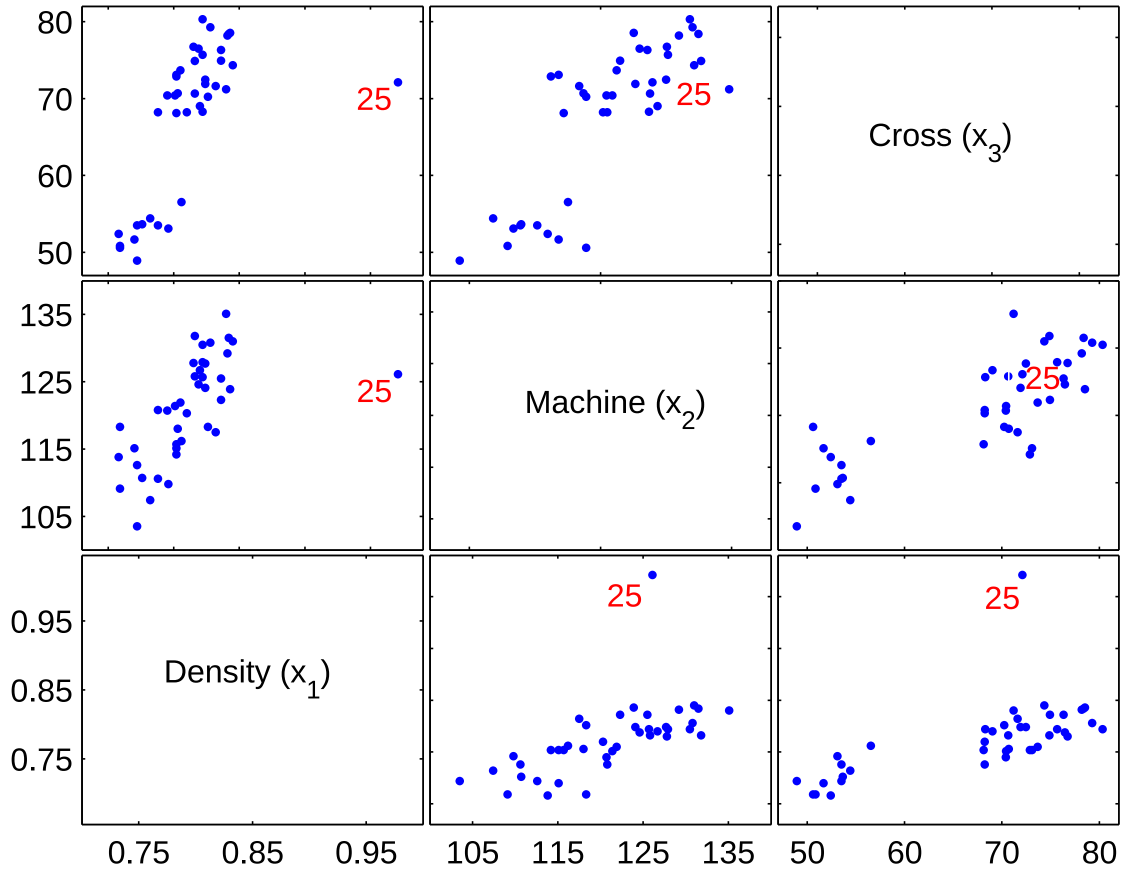
Such $\binom{3}{2} = 3$ plots are possible

Simultaneous visualization of all possible plots ?

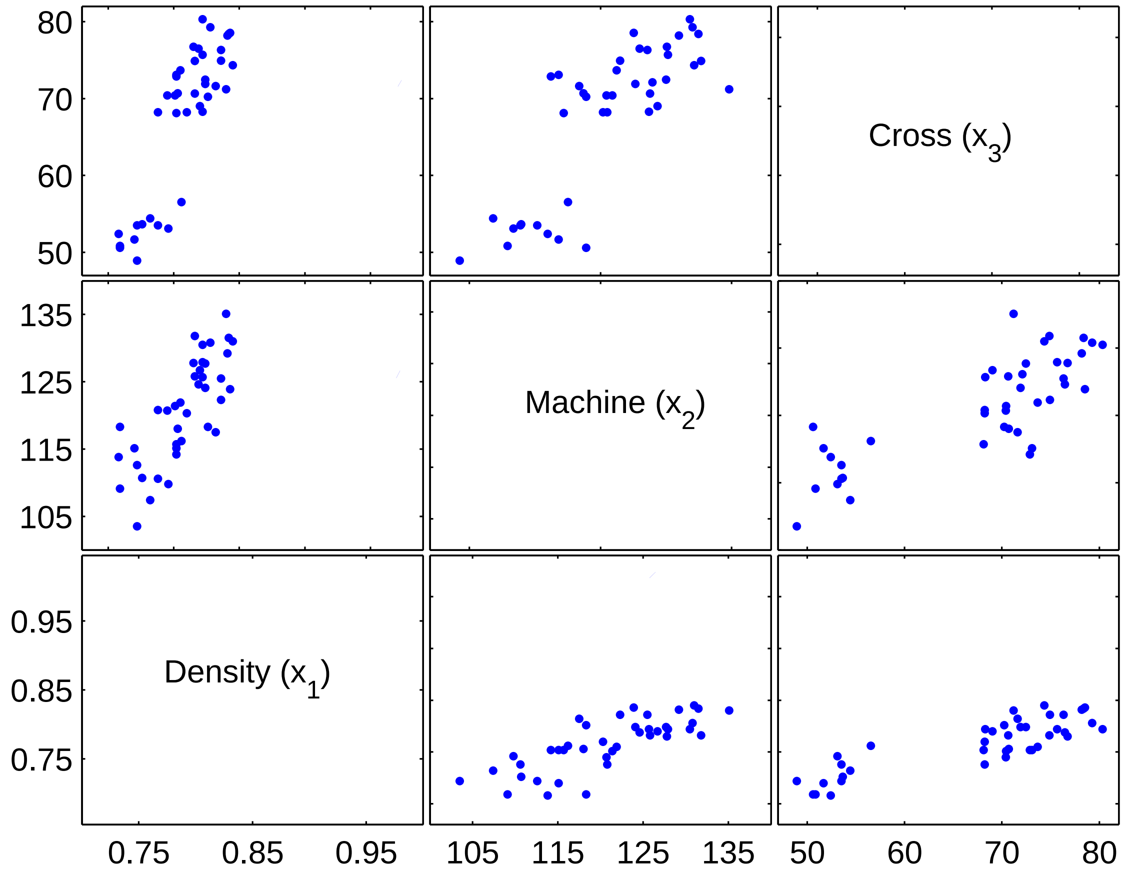
Multiple Scatter Plots from Paper-Quality Data



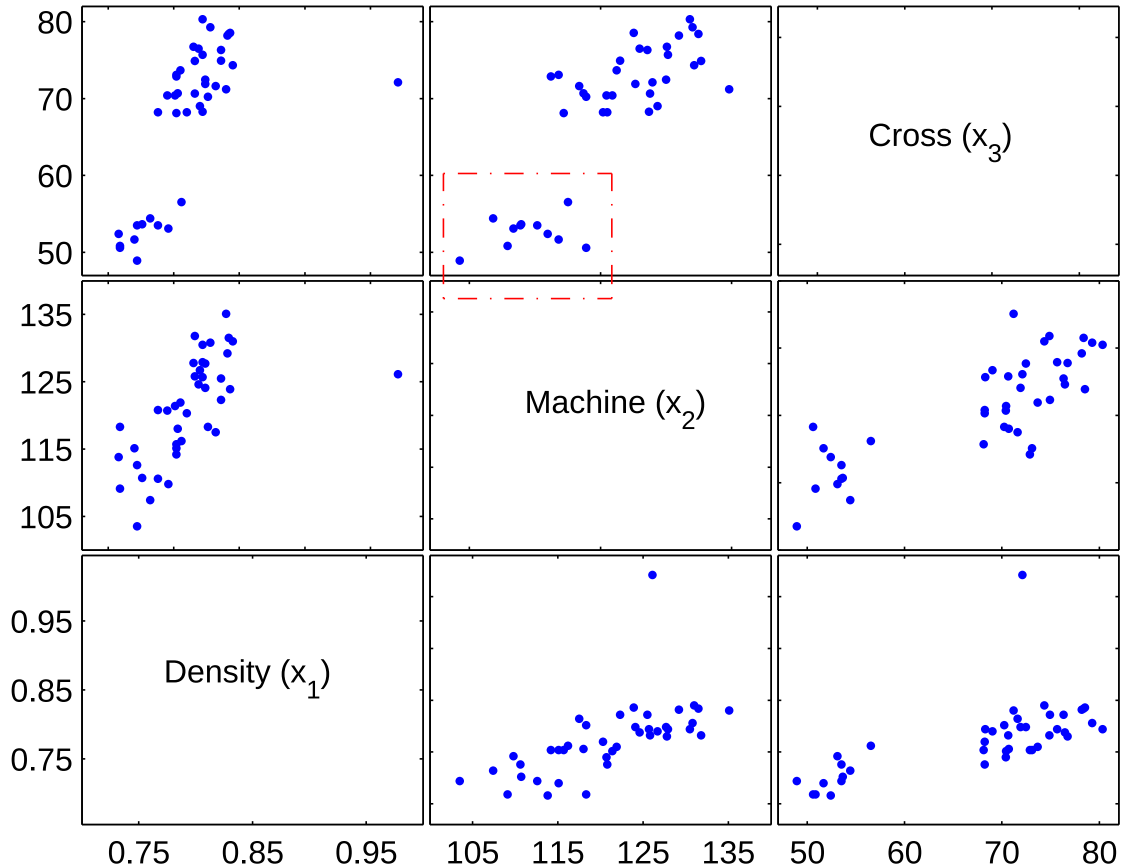
Papeer-Quality Data: Trivial Outlier Marked



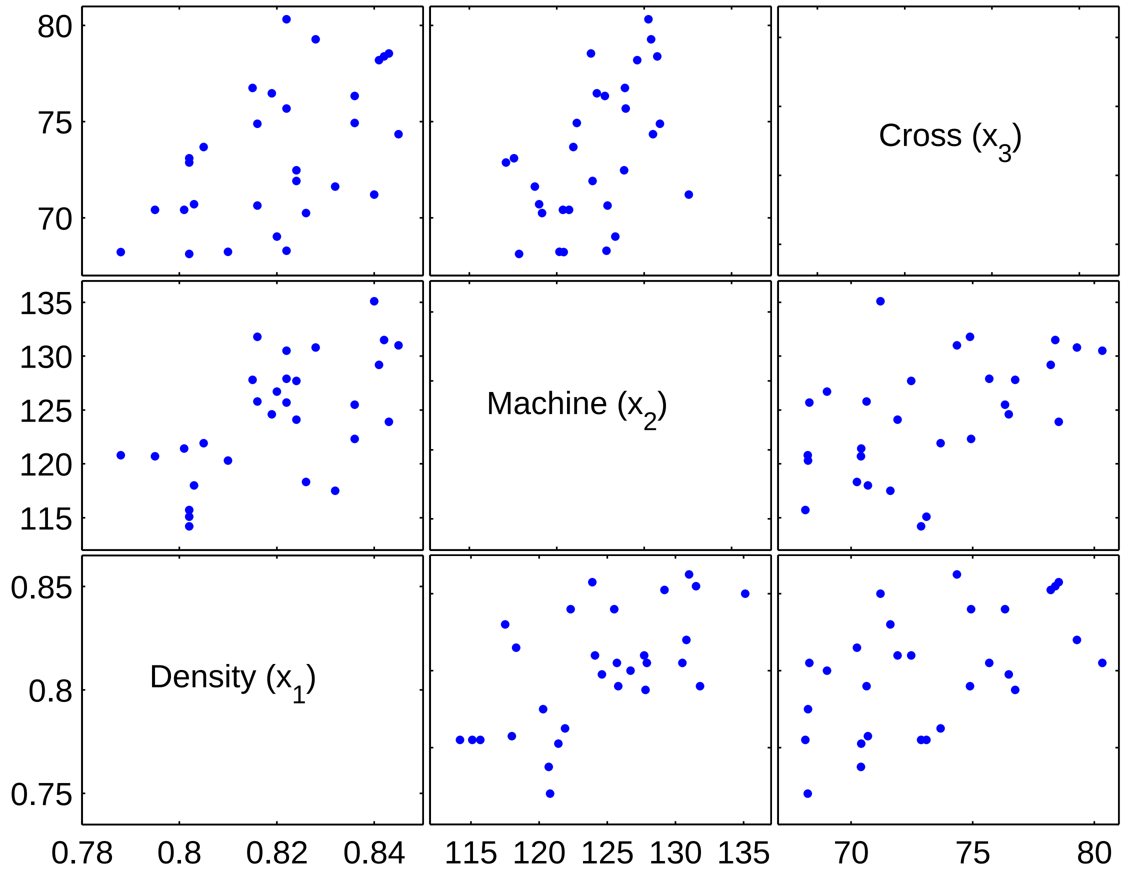
Paper-Quality Data: Outlier Removed



Paper-Quality Data: Brushing



Paper-Quality Data: After Cleaning



Data Example
Lizard Data (looking for a lower-D structure)

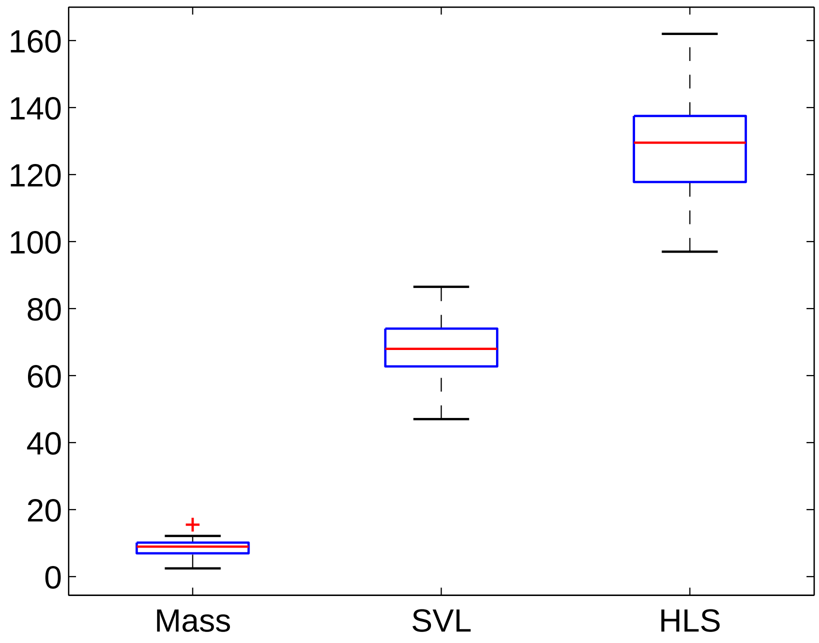
- Data Example: (Lizard Data)
  - $n = 25$ lizards measured
  - Parameters
    * Mass (weight) (g)
    * Snout-vent length (SVL) (mm)
    * Hind limb span (HLS) (mm)
  - 3-D data (see table on the next page)
  - Question: Is there any lower-dimension structure among the data?

Lizard Data Example: Table

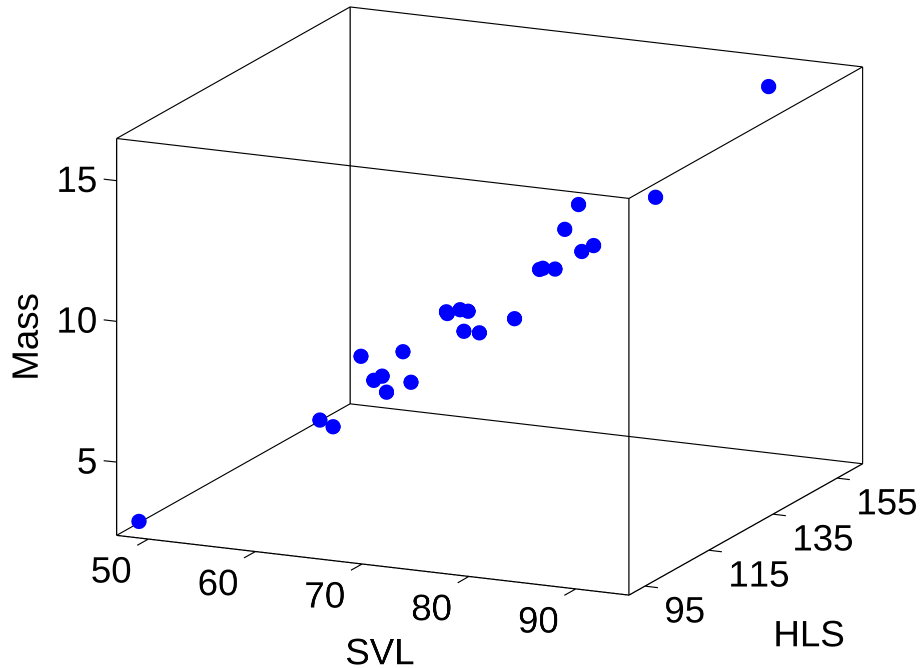
Table of lizard data:

Lizard	Mass	SVL	HLS		Lizard	Mass	SVL	HLS
1	5.526	59	113.5		14	10.067	73	136.5
2	10.401	75	142		15	10.091	73	135.5
3	9.213	69	124		16	10.888	77	139
4	8.953	67.5	125		17	7.61	61.5	118
5	7.063	62	129.5		18	7.733	66.5	133.5
6	6.61	62	123		19	12.015	79.5	150
7	11.273	74	140		20	10.049	74	137
8	2.447	47	97		21	5.149	59.5	116
9	15.493	86.5	162		22	9.158	68	123
10	9.004	69	126.5		23	12.132	75	141
11	8.199	70.5	136		24	6.978	66.5	117
12	6.601	64.5	116		25	6.89	63	117
13	7.622	67.5	135					

Box Plot from Lizard Data



3D Scatter Plot from Lizard Data



Most of the variation: variable determined by a single straight line through the point cloud

Agenda

- Objectives
- Multivariate Analysis: Extension of Univariate Statistical Methods
- A broad view of Multivariate Analytical techniques
- Data organization and visualization
- *A brief introduction to vectors and matrices (a little digression)*
- Descriptive statistics
- Statistical distance
- Random vectors and matrices

Vectors: basic operations

- $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^{n\times 1}$: $\quad \mathbf{x}_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{bmatrix}_{n\times 1} \quad \mathbf{x}_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{bmatrix}_{n\times 1}$

- Transpose: $\mathbf{x}_i' = [x_{1i}, x_{2i}, \ldots, x_{ni}]_{1\times n}$

- Addition, multiplication by a scalar c

$$\mathbf{x}_i + \mathbf{x}_j = \begin{bmatrix} x_{1i} + x_{1j} \\ x_{2i} + x_{2j} \\ \vdots \\ x_{ni} + x_{nj} \end{bmatrix} \qquad c\mathbf{x}_i = c \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{bmatrix} = \begin{bmatrix} cx_{1i} \\ cx_{2i} \\ \vdots \\ cx_{ni} \end{bmatrix}$$

Note: $\mathbf{0} = [0, 0, \ldots, 0]'_{1\times n}$

Vectors: vector length

- Length of a vector $\mathbf{x}_i$:

$$L_{\mathbf{x}_i} = \left(\sum_{k=1}^{n} x_{ki}^2 \right)^{1/2} = \sqrt{x_{1i}^2 + x_{2i}^2 + \ldots + x_{ni}^2}$$

- Unit vector: a vector with length 1

$$\mathbf{u}_{\mathbf{x}_i} = \frac{\mathbf{x}_i}{L_{\mathbf{x}_i}} \Longrightarrow L_{\mathbf{u}_{\mathbf{x}_i}} = 1$$

$\mathbf{u}_{\mathbf{x}_i}$: unit vector in the same direction as $\mathbf{x}_i$

Vectors: inner product

- Inner products of $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^{n\times 1}$

$$\mathbf{x}_i'\mathbf{x}_j = \sum_{k=1}^{n} x_{ki}x_{kj}$$

- Euclidean norm of a vector $\mathbf{x}_i$:

$$\|\mathbf{x}_i\| = \left(\sum_{k=1}^{n} x_{ki}^2\right)^{1/2} = (\mathbf{x}_i'\mathbf{x}_i)^{1/2} = L_{\mathbf{x}_i}$$

Note: $\mathbf{u}_{\mathbf{x}_i} = \dfrac{\mathbf{x}_i}{L_{\mathbf{x}_i}} = \dfrac{\mathbf{x}_i}{\|\mathbf{x}_i\|} \implies \|\mathbf{u}_{\mathbf{x}_i}\| = L_{\mathbf{u}_{\mathbf{x}_i}} = 1$

$\mathbf{u}_{\mathbf{x}_i}$: unit vector in the direction of $\mathbf{x}_i$

Vectors: angle between two vectors

- The angle θ between two vectors $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^{n\times 1}$ is specified by

$$\cos\theta = \frac{\mathbf{x}_i'\mathbf{x}_j}{L_{\mathbf{x}_i}L_{\mathbf{x}_j}} = \frac{\mathbf{x}_i'\mathbf{x}_j}{\|\mathbf{x}_i\|\|\mathbf{x}_j\|} = \frac{\sum_{k=1}^n x_{ki}x_{kj}}{\sqrt{\sum_{k=1}^n x_{ki}^2}\sqrt{\sum_{k=1}^n x_{kj}^2}}$$

  Note: $-1 \le \cos\theta \le 1, \quad 0 \le \theta \le \pi$ and

$$\mathbf{x}_i'\mathbf{x}_j = \|\mathbf{x}_i\|\|\mathbf{x}_j\|\cos\theta$$

- $\mathbf{x}_i$ and $\mathbf{x}_j$ are **orthogonal** or **perpendicular** (denoted as $\mathbf{x}_i \perp \mathbf{x}_j$) when

$$\mathbf{x}_i'\mathbf{x}_j = 0$$

Vectors: basis vectors

Any set of n linearly independent vectors

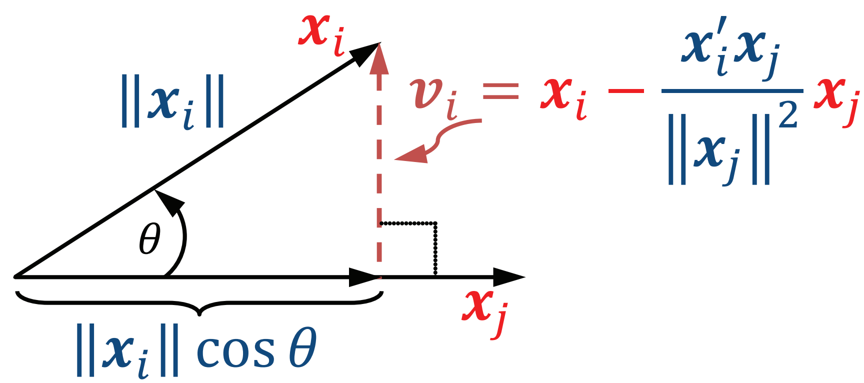
- Let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \in \mathbb{R}^{n\times 1}$

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \ldots \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Note that $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ form an Orthonormal basis. Therefore, for $\mathbf{x}_i \in \mathbb{R}^{n\times 1}$,

$$\mathbf{x}_i = x_{1i}\mathbf{e}_1 + x_{2i}\mathbf{e}_2 + \ldots + x_{ni}\mathbf{e}_n = \sum_{k=1}^{n} x_{ki}\mathbf{e}_k$$

Vectors: projection of $\mathbf{x}_i$ on $\mathbf{x}_j$



- Recall: $\mathbf{x}_i'\mathbf{x}_j = \|\mathbf{x}_i\| \|\mathbf{x}_j\| \cos\theta$

$$\underline{\text{Proj. of } \mathbf{x}_i \text{ on } \mathbf{x}_j} = \underbrace{(\|\mathbf{x}_i\| \cos\theta)}_{\text{length of proj.}} \cdot \underbrace{\frac{\mathbf{x}_j}{\|\mathbf{x}_j\|}}_{\text{unit vector}} = \frac{\mathbf{x}_i'\mathbf{x}_j}{\|\mathbf{x}_j\|^2} \cdot \mathbf{x}_j$$

- Perpendicular component: $\mathbf{v}_i \triangleq \mathbf{x}_i - \frac{\mathbf{x}_i'\mathbf{x}_j}{\|\mathbf{x}_j\|^2}\mathbf{x}_j \implies \mathbf{v}_i \perp \mathbf{x}_j$

  (used in *Gram-Schmidt Process*)

Matrices: basic operations

- $$\mathbf{A}_{n\times p} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_p \end{bmatrix}$$

  $\mathbf{A}_{n\times 1}$: column vector; $\mathbf{A}_{1\times 1}$: scalar; $\mathbf{A}_{1\times n}$: row vector

- Transpose: $$\mathbf{A}'_{p\times n} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1p} & a_{2p} & \dots & a_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_p \end{bmatrix}$$

Matrices: basic operations

- Summation of two matrices $\mathbf{A}_{n\times p}, \mathbf{B}_{n\times p}$

$$\mathbf{A}+\mathbf{B} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \ldots & a_{1p}+b_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}+b_{n1} & a_{n2}+b_{n2} & \ldots & a_{np}+b_{np} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{a}_1+\mathbf{b}_1 & \mathbf{a}_2+\mathbf{b}_2 & \ldots & \mathbf{a}_p+\mathbf{b}_p \end{bmatrix}$$

- Multiplication by a scalar c

$$c\mathbf{A}_{n\times p} = \begin{bmatrix} c\mathbf{a}_1 & c\mathbf{a}_2 & \ldots & c\mathbf{a}_p \end{bmatrix}$$

Matrices: basic operations

- Matrix multiplication: $\mathbf{C}_{n\times p} = \mathbf{A}_{n\times k}\mathbf{B}_{k\times p}$

$$\begin{bmatrix} \boxed{c_{11}} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{np} \end{bmatrix}$$

$$\boxed{\begin{array}{c} c_{ij} = \sum_{l=1}^{k} a_{il} b_{lj} \\ \text{(the inner product of} \\ i^{\text{th}} \text{ row of } \mathbf{A} \text{ and } j^{\text{th}} \text{ col. of } \mathbf{B}) \end{array}}$$

$$= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{kp} \end{bmatrix}$$

- Note: in general, $\mathbf{AB} \neq \mathbf{BA}$ (even if both exist)

Matrices: basic operations

- Matrix multiplication: $\mathbf{C}_{n\times p} = \mathbf{A}_{n\times k}\mathbf{B}_{k\times p}$. Alternatively,

$$\begin{aligned}
\mathbf{C} &= \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_p \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_k \end{bmatrix} \begin{bmatrix} \mathbf{b}_1^{\mathsf{R}} \\ \mathbf{b}_2^{\mathsf{R}} \\ \vdots \\ \mathbf{b}_k^{\mathsf{R}} \end{bmatrix} \\
&= \sum_{i=1}^{k} \mathbf{a}_i \mathbf{b}_i^{\mathsf{R}}
\end{aligned}$$

$\mathbf{a}_i$: the i^{th} column of $\mathbf{A}$; $\mathbf{b}_i^{\mathsf{R}}$: the i^{th} row of $\mathbf{B}$

- This decomposition is very important in multivariate analysis.

Matrices: basic operations

- Matrix multiplication example:

$$\begin{bmatrix} \boxed{\begin{matrix} 1 \\ 2 \end{matrix}} & \boxed{\begin{matrix} 3 \\ 4 \end{matrix}} & \boxed{\begin{matrix} 5 \\ 6 \end{matrix}} \end{bmatrix} \begin{bmatrix} \boxed{\begin{matrix} 1 & 3 \end{matrix}} \\ \boxed{\begin{matrix} -1 & 5 \end{matrix}} \\ \boxed{\begin{matrix} 1 & 7 \end{matrix}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 \cdot 1 + 3 \cdot (-1) + 5 \cdot 1 & 1 \cdot 3 + 3 \cdot 5 + 5 \cdot 7 \\ 2 \cdot 1 + 4 \cdot (-1) + 6 \cdot 1 & 2 \cdot 3 + 4 \cdot 5 + 6 \cdot 7 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 3 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \end{bmatrix} \begin{bmatrix} -1 & 5 \end{bmatrix} + \begin{bmatrix} 5 \\ 6 \end{bmatrix} \begin{bmatrix} 1 & 7 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 53 \\ 4 & 68 \end{bmatrix}$$

Matrices: basic operations

- Another matrix multiplication example:

$$[x \quad y]\begin{bmatrix} a & b & e \\ c & d & f \end{bmatrix}\begin{bmatrix} p \\ q \\ r \end{bmatrix} \overset{(1)}{=} \left[[x \quad y]\begin{bmatrix} a \\ c \end{bmatrix} \quad [x \quad y]\begin{bmatrix} b \\ d \end{bmatrix} \quad [x \quad y]\begin{bmatrix} e \\ f \end{bmatrix} \right]\begin{bmatrix} p \\ q \\ r \end{bmatrix}$$

$$\overset{(2)}{=} [x[a \;\; b \;\; e] + y[c \;\; d \;\; f]]\begin{bmatrix} p \\ q \\ r \end{bmatrix} = [ax + cy \quad bx + dy \quad ex + fy]\begin{bmatrix} p \\ q \\ r \end{bmatrix}$$

$$\overset{(3)}{=} [x \quad y]\left[\begin{bmatrix} a \\ c \end{bmatrix} p + \begin{bmatrix} b \\ d \end{bmatrix} q + \begin{bmatrix} e \\ f \end{bmatrix} r \right]$$

$$= p(ax + cy) + q(bx + dy) + r(ex + fy)$$

$\mathbf{x}'\mathbf{A}\mathbf{x}$: quadratic form (here $\mathbf{A}$: symmetric; to be defined soon); $\mathbf{x}'\mathbf{A}\mathbf{y}$: bilinear form

Square Matrices

- Symmetric matrices: $\mathbf{A}$ is said to be symmetric if

$$\mathbf{A} = \mathbf{A}' \qquad \text{or} \quad a_{ij} = a_{ji} \ \forall i, j$$

$$\text{Example :} \quad \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}' = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$

- Identity matrix $\mathbf{I}$: $\quad \mathbf{A}_{k\times k}\mathbf{I}_{k\times k} = \mathbf{I}_{k\times k}\mathbf{A}_{k\times k} = \mathbf{A}_{k\times k}, \quad \forall \mathbf{A}_{k\times k}$
- Note that $\mathbf{A} + \mathbf{A}'$ is a symmetric matrix

Inverse of a Square Matrix

- Inverse matrix: If there exists a matrix $\mathbf{B}$ such that

$$\mathbf{B}_{k\times k}\mathbf{A}_{k\times k} = \mathbf{A}_{k\times k}\mathbf{B}_{k\times k} = \mathbf{I}_{k\times k}$$

  then $\mathbf{B}$ is called the inverse of $\mathbf{A}$ (denoted by $\mathbf{A}^{-1}$).
- Condition for the existence of $\mathbf{A}^{-1}$
  - $\mathbf{A}$ non-singular, or
  - columns of $\mathbf{A}$ linearly independent
- Note that

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

- Also

$$(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$$

Example: Inverse Matrix

- Example of inverse matrix: $\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} -0.2 & 0.4 \\ 0.8 & -0.6 \end{bmatrix}$

$$\mathbf{AB} = \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} -0.2 & 0.4 \\ 0.8 & -0.6 \end{bmatrix} = \begin{bmatrix} -0.2 & 0.4 \\ 0.8 & -0.6 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix} = \mathbf{BA} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Longrightarrow \mathbf{AB} = \mathbf{BA} = \mathbf{I}, \quad \mathbf{A}^{-1} = \mathbf{B} \quad (\mathbf{B}^{-1} = \mathbf{A})$$

- Hard to find the inverse (typically)!! If $\mathbf{A}$ is diagonal ($a_{ii} \neq 0, \forall i$)

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{kk} \end{bmatrix} \text{ then } \mathbf{A}^{-1} = \begin{bmatrix} a_{11}^{-1} & 0 & \dots & 0 \\ 0 & a_{22}^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{kk}^{-1} \end{bmatrix}$$

Orthogonal Matrices

- Orthogonal matrices:

$$\mathbf{Q} \text{ orthogonal} \iff \mathbf{Q}\mathbf{Q}' = \mathbf{Q}'\mathbf{Q} = \mathbf{I} \iff \mathbf{Q}^{-1} = \mathbf{Q}'$$

- Let $\mathbf{q}_i'$ be the i^{th} row of orthogonal $\mathbf{Q}$.

$$\begin{bmatrix} \mathbf{q}_1' \\ \mathbf{q}_2' \\ \vdots \\ \mathbf{q}_k' \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \dots & \mathbf{q}_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$\implies \mathbf{q}_i'\mathbf{q}_i = 1, \qquad \mathbf{q}_i'\mathbf{q}_j = 0, \qquad \forall i, j \ (i \neq j)$$

Rows of orthogonal $\mathbf{Q}$: unit length, mutually perpendicular (orthogonal)

Spectral Decomposition of a Square Matrix

- Square matrix $\mathbf{A}$:

  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ $(\mathbf{x} \neq \mathbf{0}) \Longrightarrow \lambda$: an eigenvalue of $\mathbf{A}$ with corres. eigenvector $\mathbf{x}$

- Symmetric matrix $\mathbf{A}_{k\times k}$
  - $\mathbf{A}$ has k pairs of eigenvalues and eigenvectors $\{\lambda_i, \mathbf{e}_i\}_{i=1}^{k}$
  - $\mathbf{e}_i$: eigenvector normalized (assumed) and mutually orthogonal, i.e.,

$$\mathbf{e}_i'\mathbf{e}_j = \delta_{ij}, \quad \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

  - All eigenvalues: real

Spectral Decomposition: Example

- Example: let $\mathbf{A} = \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix}$.

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x} \to \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \to \begin{cases} x_1 - 5x_2 & = \lambda x_1 \\ -5x_1 + x_2 & = \lambda x_2 \end{cases}$$

$$\xrightarrow{\mathrm{x} \neq 0} (1-\lambda)^2 = 25$$

$$\to \begin{cases} \lambda_1 = 6 : x_1 - 5x_2 = 6x_1 & \boxed{x_1 = -x_2} \\ \lambda_2 = -4 : -5x_1 + x_2 = -4x_2 & \boxed{x_1 = x_2} \end{cases}$$

$$\to \left\{ \lambda_1 = 6, \mathbf{e}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \right\}, \left\{ \lambda_2 = -4, \mathbf{e}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}$$

Matrices: Symmetric Matrices

- Let $\mathbf{A}_{k\times k}$ be symmetric.

$$\begin{aligned}\mathbf{A}_{k\times k} &= \sum_{j=1}^{k} \lambda_j \mathbf{e}_j \mathbf{e}_j' \Longrightarrow \begin{cases} \text{positive definite (p.d.)}, & \lambda_j > 0\ \forall j \\ \text{positive semidef. (p.s.d.)}, & \lambda_j \geq 0\ \forall j \end{cases} \\ &= \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'\end{aligned}$$

where $\mathbf{P} = \begin{bmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_k\end{bmatrix}$, $\mathbf{e}_i'\mathbf{e}_j = \delta_{ij}$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_k \end{bmatrix} (\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k)$$

"Positive semi-definite" is also referred to as "nonnegative definite".

Matrices: Symmetric Matrices

- Let $\mathbf{A}_{k\times k}$ be symmetric

$$\mathbf{A}^{-1} = \sum_{j=1}^{k} \frac{1}{\lambda_j}\mathbf{e}_j\mathbf{e}_j' = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}'$$

$$\mathbf{A}^{-1/2} = \sum_{j=1}^{k} \sqrt{\lambda_j}\mathbf{e}_j\mathbf{e}_j' = \mathbf{P}\boldsymbol{\Lambda}^{-1/2}\mathbf{P}'$$

- $\mathbf{x}'\mathbf{A}\mathbf{x}$: quadratic form, important in multivariate analysis

$$\mathbf{A} = \begin{cases} \text{p.d.,} & \mathbf{x}'\mathbf{A}\mathbf{x} > 0, \forall \mathbf{x} \neq \mathbf{0} \\ \text{p.s.d.,} & \mathbf{x}'\mathbf{A}\mathbf{x} \geq 0, \forall \mathbf{x} \end{cases}$$

Matrices: Symmetric Matrices

- Example: (a positive definite matrix and quadratic form)

$$3x_1^2 + 2x_2^2 - 2\sqrt{2}x_1x_2 = \underbrace{\begin{bmatrix} x_1 & x_2 \end{bmatrix}}_{\mathbf{x}'} \underbrace{\begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}}$$

It can be shown that $\mathbf{A} = 4\mathbf{e}_1\mathbf{e}_1' + \mathbf{e}_2\mathbf{e}_2'$ (i.e., $\lambda_1 = 4, \lambda_2 = 1$)

$$\begin{aligned} 3x_1^2 + 2x_2^2 - 2\sqrt{2}x_1x_2 &= \mathbf{x}'\mathbf{A}\mathbf{x} = 4\mathbf{x}'\mathbf{e}_1\mathbf{e}_1'\mathbf{x} + \mathbf{x}'\mathbf{e}_2\mathbf{e}_2'\mathbf{x} \\ &= 4(\mathbf{x}'\mathbf{e}_1)^2 + (\mathbf{x}'\mathbf{e}_2)^2 \\ &> 0, \qquad \Longrightarrow \qquad \underline{\mathbf{A} \quad \text{positive definite}} \end{aligned}$$

Note: Last step: $\forall \mathbf{x} \neq \mathbf{0}$, $\mathbf{x}'\mathbf{e}_1$ and $\mathbf{x}'\mathbf{e}_2$ cannot be 0 simultaneously.

Trace of a Square Matrix

- Let $\mathbf{A}$ be a $k \times k$ square matrix

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^{k} a_{kk} = \sum_{i=1}^{k} \lambda_i$$

- For $\mathbf{A}, \mathbf{B}$ $k \times k$ square matrices
  - $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$
  - $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$
  - $\text{tr}(\mathbf{AA}') = \sum_{i=1}^{k} \sum_{j=1}^{K} a_{ij}^2$ (Frobenius norm squared)
  - $\text{tr}(\mathbf{QAQ}') = \text{tr}(\mathbf{A})$

Determinant of a Square Matrix

- Let $\mathbf{A}$ be a $k \times k$ square matrix

$$\det(\mathbf{A}) = \sum \mathsf{sgn}(\sigma) \prod_{i=1}^{k} a_{i\sigma_i}$$

- σ_i is a permutation of $\{1, 2, \cdots, k\}$
- For $\mathbf{A}, \mathbf{B}$ $k \times k$ square matrices
  - $\det(\mathbf{A}') = \det(\mathbf{A})$
  - $\det(c\mathbf{A}) = c^k \det(\mathbf{A})$
  - $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$
  - $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$
  - $\det(\mathbf{A}) = \frac{1}{\det(\mathbf{A}^{-1})}$ ($\mathbf{A}$ non-singular)
  - $\det(\mathbf{A}) = \prod_{i=1}^{k} \lambda_i$

Characteristic Equation

- Let $\mathbf{A}$ be a $k \times k$ square matrix and $\mathbf{I}$ be the $k \times k$ identity matrix

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0$$

- $\lambda_1, \lambda_2, \cdots, \lambda_k$ are the eigenvalues of $\mathbf{A}$
- In general, $\lambda_1, \lambda_2, \cdots, \lambda_k$ are complex numbers
- Complex values should occur in conjugate pairs

Agenda

- Objectives
- Multivariate Analysis: Extension of Univariate Statistical Methods
- A broad view of Multivariate Analytical techniques
- Data organization and visualization
- *A brief introduction to vectors and matrices (a little digression)*
- Descriptive statistics
- Statistical distance
- Random vectors and matrices

Descriptive Statistics (Summary Statistics)

p variables, n observations each

$$\mathbf{X}_{n\times p} = \begin{bmatrix} \boxed{\begin{matrix} x_{11} \\ \vdots \\ x_{n1} \end{matrix}} & \boxed{\begin{matrix} x_{12} \\ \vdots \\ x_{n2} \end{matrix}} & \begin{matrix} \cdots \\ \ddots \\ \cdots \end{matrix} & \boxed{\begin{matrix} x_{1p} \\ \vdots \\ x_{np} \end{matrix}} \end{bmatrix}$$

$$\begin{matrix} \overline{x}_1 & \overline{x}_2 & \cdots & \overline{x}_p \\ \frac{1}{n}\sum_{j=1}^n x_{j1} & \frac{1}{n}\sum_{j=1}^n x_{j2} & \cdots & \frac{1}{n}\sum_{j=1}^n x_{jp} \\ \frac{1}{n}\mathbf{x}_1'\mathbf{1} & \frac{1}{n}\mathbf{x}_2'\mathbf{1} & \cdots & \frac{1}{n}\mathbf{x}_p'\mathbf{1} \end{matrix}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \ldots & \mathbf{x}_p \end{bmatrix}$$

$$\mathbf{1} = \begin{bmatrix} 1 & 1 & \ldots & 1 \end{bmatrix}' \text{(all 1 vector)}$$

Descriptive Statistics

- Sample mean

$$\overline{x}_k = \frac{1}{n}\sum_{j=1}^{n} x_{jk} = \frac{1}{n}\mathbf{x}_k'\mathbf{1}$$

- Mean vector

$$\overline{\mathbf{x}} = \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \vdots \\ \overline{x}_p \end{bmatrix} = \frac{1}{n}\begin{bmatrix} \mathbf{x}_1'\mathbf{1} \\ \mathbf{x}_2'\mathbf{1} \\ \vdots \\ \mathbf{x}_p'\mathbf{1} \end{bmatrix} = \frac{1}{n}\begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_p' \end{bmatrix}\mathbf{1} = \frac{1}{n}\mathbf{X}'\mathbf{1}$$

Descriptive Statistics

- Sample covariance matrix $\mathbf{S}_{p\times p}$

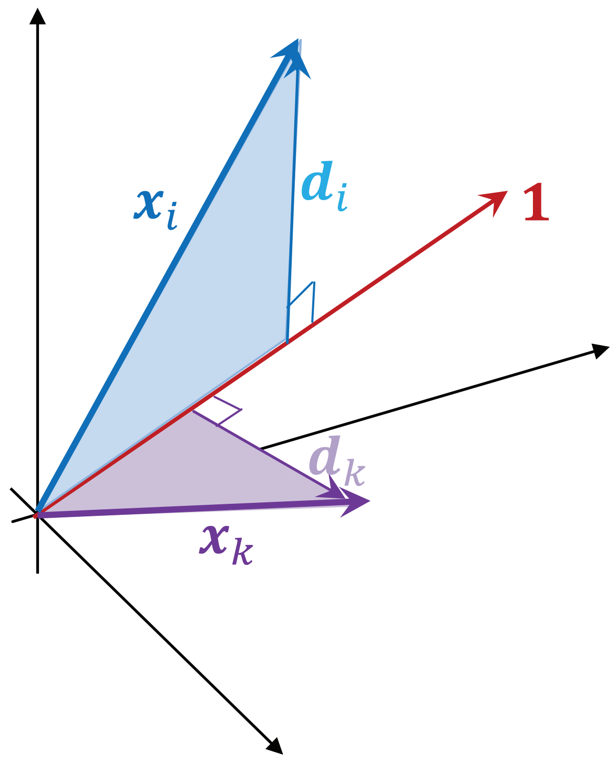
$$
\begin{aligned}
s_{ik} =& \frac{1}{n}\sum_{j=1}^{n}(x_{ji}-\overline{x}_i)(x_{jk}-\overline{x}_k), \quad i,k=1,2,\ldots,p \\
=& \frac{1}{n}(\underbrace{\mathbf{x}_i-\overline{x}_i\mathbf{1}}_{\mathbf{d}_i})'(\underbrace{\mathbf{x}_k-\overline{x}_k\mathbf{1}}_{\mathbf{d}_k}) = \frac{1}{n}\mathbf{d}_i'\mathbf{d}_k
\end{aligned}
$$

$$
\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}
$$

$\mathbf{S}$: symmetric, with $\dfrac{p(p+1)}{2}$ different elements

Descriptive Statistics

- Geometry of entries of $\mathbf{S}$: $s_{ik} = \frac{1}{n}\mathbf{d}_i'\mathbf{d}_k, \quad \mathbf{d}_i = \mathbf{x}_i - \overline{x}_i\mathbf{1}$



Descriptive Statistics: relation between $\mathbf{S}$ and $\mathbf{X}$

$$\mathbf{S} = \frac{1}{n}\begin{bmatrix} \mathbf{d}_1'\mathbf{d}_1 & \mathbf{d}_1'\mathbf{d}_2 & \dots & \mathbf{d}_1'\mathbf{d}_p \\ \mathbf{d}_2'\mathbf{d}_1 & \mathbf{d}_2'\mathbf{d}_2 & \dots & \mathbf{d}_2'\mathbf{d}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{d}_p'\mathbf{d}_1 & \mathbf{d}_p'\mathbf{d}_2 & \dots & \mathbf{d}_p'\mathbf{d}_p \end{bmatrix} = \frac{1}{n}\begin{bmatrix} \mathbf{d}_1' \\ \mathbf{d}_2' \\ \vdots \\ \mathbf{d}_p' \end{bmatrix}\begin{bmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \dots & \mathbf{d}_p \end{bmatrix}$$

$$= \frac{1}{n}\begin{bmatrix} \mathbf{x}_1' - \overline{x}_1\mathbf{1}' \\ \mathbf{x}_2' - \overline{x}_2\mathbf{1}' \\ \vdots \\ \mathbf{x}_p' - \overline{x}_p\mathbf{1}' \end{bmatrix}\begin{bmatrix} \mathbf{x}_1' - \overline{x}_1\mathbf{1}' \\ \mathbf{x}_2' - \overline{x}_2\mathbf{1}' \\ \vdots \\ \mathbf{x}_p' - \overline{x}_p\mathbf{1}' \end{bmatrix}' \qquad (\mathbf{d}_i = \mathbf{x}_i - \overline{x}_i\mathbf{1})$$

$$= \frac{1}{n}\begin{bmatrix} \mathbf{x}_1' - \mathbf{x}_1'\frac{\mathbf{1}\mathbf{1}'}{n} \\ \mathbf{x}_2' - \mathbf{x}_2'\frac{\mathbf{1}\mathbf{1}'}{n} \\ \vdots \\ \mathbf{x}_p' - \mathbf{x}_p'\frac{\mathbf{1}\mathbf{1}'}{n} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1' - \mathbf{x}_1'\frac{\mathbf{1}\mathbf{1}'}{n} \\ \mathbf{x}_2' - \mathbf{x}_2'\frac{\mathbf{1}\mathbf{1}'}{n} \\ \vdots \\ \mathbf{x}_p' - \mathbf{x}_p'\frac{\mathbf{1}\mathbf{1}'}{n} \end{bmatrix}' \left(\overline{x}_i = \frac{1}{n}\mathbf{x}_i'\mathbf{1} \Rightarrow \overline{x}_i\mathbf{1}' = \frac{1}{n}\mathbf{x}_i'\mathbf{1}\mathbf{1}'\right)$$

Descriptive Statistics: relation between $\mathbf{S}$ and $\mathbf{X}$

(Cont'd) $\mathbf{S} = \frac{1}{n} \begin{bmatrix} \mathbf{x}_1' - \mathbf{x}_1' \frac{\mathbf{1}\mathbf{1}'}{n} \\ \mathbf{x}_2' - \mathbf{x}_2' \frac{\mathbf{1}\mathbf{1}'}{n} \\ \vdots \\ \mathbf{x}_p' - \mathbf{x}_p' \frac{\mathbf{1}\mathbf{1}'}{n} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1' - \mathbf{x}_1' \frac{\mathbf{1}\mathbf{1}'}{n} \\ \mathbf{x}_2' - \mathbf{x}_2' \frac{\mathbf{1}\mathbf{1}'}{n} \\ \vdots \\ \mathbf{x}_p' - \mathbf{x}_p' \frac{\mathbf{1}\mathbf{1}'}{n} \end{bmatrix}'$

$$= \frac{1}{n} \left(\begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_p' \end{bmatrix} - \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_p' \end{bmatrix} \frac{\mathbf{1}\mathbf{1}'}{n} \right) \left(\begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_p' \end{bmatrix} - \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_p' \end{bmatrix} \frac{\mathbf{1}\mathbf{1}'}{n} \right)'$$

$$= \frac{1}{n} \left(\mathbf{X}' - \frac{1}{n} \mathbf{X}'\mathbf{1}\mathbf{1}' \right) \left(\mathbf{X}' - \frac{1}{n} \mathbf{X}'\mathbf{1}\mathbf{1}' \right)'$$

$$= \frac{1}{n} \mathbf{X}' \left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}' \right) \left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}' \right) \mathbf{X} = \frac{1}{n} \mathbf{X}' \left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}' \right) \mathbf{X}$$

Descriptive Statistics

- Sample variance: set $i = k$ in s_{ik}

$$s_{ii} = \frac{1}{n}\sum_{j=1}^{n}(x_{ji} - \overline{x}_i)^2 = \frac{1}{n}\mathbf{d}_i'\mathbf{d}_i$$

- Sample correlation (Pearson's product-moment correlation coefficient)

$$\begin{aligned} r_{ik} &= \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}} = \frac{\sum_{j=1}^{n}(x_{ji} - \overline{x}_i)(x_{jk} - \overline{x}_k)}{\sqrt{\sum_{j=1}^{n}(x_{ji} - \overline{x}_i)^2}\sqrt{\sum_{j=1}^{n}(x_{jk} - \overline{x}_k)^2}} \\ &= \frac{\mathbf{d}_i'\mathbf{d}_k}{\sqrt{\mathbf{d}_i'\mathbf{d}_i}\sqrt{\mathbf{d}_k'\mathbf{d}_k}} = \frac{\mathbf{d}_i'\mathbf{d}_k}{\|\mathbf{d}_i\|\|\mathbf{d}_k\|} \\ &= \cos\theta_{ik} \end{aligned}$$

Descriptive Statistics

- Sample correlation matrix $\mathbf{R}_{p\times p}$

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}} = \frac{\mathbf{d}_i'\mathbf{d}_k}{\|\mathbf{d}_i\|\|\mathbf{d}_k\|}, \qquad r_{ii} = 1\ \forall i$$

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix}$$

symmetric

Descriptive Statistics: relation between $\mathbf{S}$ and $\mathbf{R}$

$$\mathbf{D} \triangleq \begin{bmatrix} s_{11} & 0 & \dots & 0 \\ 0 & s_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_{pp} \end{bmatrix} \Longrightarrow \mathbf{D}^{-1/2} = \begin{bmatrix} \frac{1}{\sqrt{s_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{s_{pp}}} \end{bmatrix}$$

$$\mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2} = \begin{bmatrix} \frac{1}{\sqrt{s_{11}}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{s_{pp}}} \end{bmatrix} \begin{bmatrix} s_{11} & \dots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{p1} & \dots & s_{pp} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{s_{11}}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{s_{pp}}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{s_{11}}{\sqrt{s_{11}}\sqrt{s_{11}}} & \dots & \frac{s_{1p}}{\sqrt{s_{11}}\sqrt{s_{pp}}} \\ \vdots & \ddots & \vdots \\ \frac{s_{p1}}{\sqrt{s_{pp}}\sqrt{s_{11}}} & \dots & \frac{s_{pp}}{\sqrt{s_{pp}}\sqrt{s_{pp}}} \end{bmatrix} = \mathbf{R}$$

Linear Combination of Variables

$$\mathbf{y} = \mathbf{X}\mathbf{c} = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_p]\mathbf{c} = \sum_{j=1}^{p} c_j \mathbf{x}_j$$

- Sample mean:

$$\overline{y} = \frac{1}{n}\mathbf{y}'\mathbf{1} \overset{(1)}{=} \frac{1}{n}\left(\sum_{j=1}^{p} c_j \mathbf{x}_j'\right)\mathbf{1} = \sum_{j=1}^{p} c_j \frac{\mathbf{x}_j'\mathbf{1}}{n} = \sum_{j=1}^{p} c_j \overline{x}_j = \mathbf{c}'\overline{\mathbf{x}}$$

$$\overset{(2)}{=} \frac{1}{n}(\mathbf{X}\mathbf{c})'\mathbf{1} = \frac{1}{n}\mathbf{c}'\mathbf{X}'\mathbf{1} = \mathbf{c}'\frac{\mathbf{X}'\mathbf{1}}{n} = \mathbf{c}'\overline{\mathbf{x}}$$

Recall : $\overline{\mathbf{x}} = [\overline{x}_1\ \overline{x}_2\ \ldots\ \overline{x}_p]' = \frac{1}{n}\mathbf{X}'\mathbf{1}$ (mean vector of $\mathbf{X}$)

Linear Combination of Variables

$$\mathbf{y} = \mathbf{X}\mathbf{c} = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_p]\mathbf{c} = \sum_{j=1}^{p} c_j \mathbf{x}_j$$

- Sample variance:

$$\frac{1}{n}(\mathbf{y} - \overline{y}\mathbf{1})'(\mathbf{y} - \overline{y}\mathbf{1}) = \frac{1}{n}(\mathbf{c}'\mathbf{X}' - \mathbf{c}'\overline{\mathbf{x}}\mathbf{1}')(\mathbf{X}\mathbf{c} - \overline{\mathbf{x}}'\mathbf{c}\mathbf{1})$$

$$\text{Using } \overline{\mathbf{x}} = \frac{1}{n}\mathbf{X}'\mathbf{1} \Longrightarrow\ = \frac{1}{n}\left(\mathbf{c}'\mathbf{X}' - \frac{1}{n}\mathbf{c}'\mathbf{X}'\mathbf{1}\mathbf{1}'\right)\left(\mathbf{X}\mathbf{c} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X}\mathbf{c}\right)$$

$$= \frac{1}{n}\mathbf{c}'\mathbf{X}'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}\mathbf{c}$$

$$= \mathbf{c}'\left(\frac{1}{n}\mathbf{X}'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}\right)\mathbf{c}$$

$$= \mathbf{c}'\mathbf{S}\mathbf{c}$$

Linear Combination of Variables

$$\mathbf{y} = \mathbf{X}\mathbf{c}, \quad \mathbf{z} = \mathbf{X}\mathbf{b}$$

- Sample covariance:

$$\begin{aligned}
\text{Cov}\,(\mathbf{z}, \mathbf{y}) &= \frac{1}{n}(\mathbf{z} - \bar{z}\mathbf{1})'(\mathbf{y} - \bar{y}\mathbf{1}) \\
&= \frac{1}{n}\left(\mathbf{b}'\mathbf{X}' - \frac{1}{n}\mathbf{b}'\mathbf{X}'\mathbf{1}\mathbf{1}'\right)\left(\mathbf{X}\mathbf{c} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X}\mathbf{c}\right) \\
&= \frac{1}{n}\mathbf{b}'\mathbf{X}'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}\mathbf{c} \\
&= \mathbf{b}'\left(\frac{1}{n}\mathbf{X}'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}\right)\mathbf{c} \\
&= \mathbf{b}'\mathbf{S}\mathbf{c} = \mathbf{c}'\mathbf{S}\mathbf{b} \quad (\mathbf{S}: \underline{\text{symmetric}})
\end{aligned}$$

Linear Combination of Variables: Generalization

Given: $\mathbf{y}_1 = \mathbf{X}_{n\times p}\mathbf{c}_1 \quad \ldots \quad \mathbf{y}_q = \mathbf{X}_{n\times p}\mathbf{c}_q$

$$
\begin{aligned}
\mathbf{Y} &= \begin{bmatrix}\mathbf{y}_1 & \mathbf{y}_2 & \ldots & \mathbf{y}_q\end{bmatrix}_{n\times q} \\
&= \begin{bmatrix}\mathbf{X}\mathbf{c}_1 & \mathbf{X}\mathbf{c}_2 & \ldots & \mathbf{X}\mathbf{c}_q\end{bmatrix} \\
&= \mathbf{X}_{n\times p}\underbrace{\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \ldots & \mathbf{c}_q\end{bmatrix}}_{\mathbf{C}_{p\times q}} \\
&= \mathbf{X}\mathbf{C}
\end{aligned}
$$

$$\overline{\mathbf{y}} = \frac{\mathbf{Y}'\mathbf{1}}{n} = \frac{\mathbf{C}'\mathbf{X}\mathbf{1}}{n} = \mathbf{C}'\overline{\mathbf{x}} \quad \text{(Mean vector of } \mathbf{Y}\text{)}$$

$$\mathbf{S}_{\mathsf{Y}} = \frac{1}{n}(\mathbf{X}\mathbf{C})'(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}')(\mathbf{X}\mathbf{C}) = \mathbf{C}'\mathbf{S}\mathbf{C}$$

(Sample covariance matrix of $\mathbf{Y}$)

Agenda

- Objectives
- Multivariate Analysis: Extension of Univariate Statistical Methods
- A broad view of Multivariate Analytical techniques
- Data organization and visualization
- *A brief introduction to vectors and matrices (a little digression)*
- Descriptive statistics
- Statistical distance
- Random vectors and matrices

Random Vectors and Matrices

- Definitions:
  - Random vector: a vector whose elements are random variables
  - Random matrix: a matrix whose elements are random variables

- Random matrix $\mathbf{X}_{n\times p} = \{X_{ij}\} \Rightarrow$ the expected value of $\mathbf{X}$ ($\mathsf{E}(\mathbf{X})$):

$$\mathsf{E}(\mathbf{X}) = \begin{bmatrix} \mathsf{E}(X_{11}) & \mathsf{E}(X_{12}) & \dots & \mathsf{E}(X_{1p}) \\ \mathsf{E}(X_{21}) & \mathsf{E}(X_{22}) & \dots & \mathsf{E}(X_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{E}(X_{n1}) & \mathsf{E}(X_{n2}) & \dots & \mathsf{E}(X_{np}) \end{bmatrix}$$

$$\mathsf{E}(X_{ij}) = \begin{cases} \int_{-\infty}^{+\infty} x_{ij} f_{ij}(x_{ij}) dx_{ij}, & \text{continuous, p.d.f. } f_{ij}(x_{ij}) \\ \sum_{\text{all } x_{ij}} x_{ij} p_{ij}(x_{ij}), & \text{discrete, p.m.f. } p_{ij}(x_{ij}) \end{cases}$$

Random Vectors and Matrices

- Given random matrices $\mathbf{X}, \mathbf{Y}$ of the same dimension

  Basic operations:

$$\text{i)}\quad \mathsf{E}(\mathbf{X} + \mathbf{Y}) = \mathsf{E}(\mathbf{X}) + \mathsf{E}(\mathbf{Y})$$

$$\text{ii)}\quad \mathsf{E}(\mathbf{AXB}) = \mathbf{A}\mathsf{E}(\mathbf{X})\mathbf{B}$$

$$(\mathbf{A}, \mathbf{B} : \text{deterministic with proper sizes})$$

Random Vectors and Matrices

$$\mathbf{X}_{p\times 1} = \begin{bmatrix} X_1 & X_2 & \dots & X_p \end{bmatrix}' \quad (p \times 1 \text{ random vector})$$

- Marginal mean μ_i and variance σ_i^2

$$\mu_i = \mathsf{E}(X_i) = \begin{cases} \int_{-\infty}^{+\infty} x_i f_i(x_i) dx_i, & \text{continuous} \\ \sum_{\text{all } x_i} x_i p_i(x_i), & \text{discrete} \end{cases}$$

$$\sigma_i^2 = \mathsf{E}(X_i - \mu_i)^2 = \begin{cases} \int (x_i - \mu_i)^2 f_i(x_i) dx_i, & \text{continuous} \\ \sum (x_i - \mu_i)^2 p_i(x_i), & \text{discrete} \end{cases}$$

(X_i: p.d.f. $f_i(x_i)$ if continuous, p.m.f. $p_i(x_i)$ if discrete)

Random Vectors and Matrices

$$\mathbf{X}_{p\times 1} = \begin{bmatrix} X_1 & X_2 & \dots & X_p \end{bmatrix}' \quad (p \times 1 \text{ random vector})$$

- Covariance σ_{ik}

$$\sigma_{ik} = \mathsf{E}(X_i - \mu_i)(X_k - \mu_k)$$
$$= \begin{cases} \int\int (x_i - \mu_i)(x_k - \mu_k) f_{ik}(x_i, x_k) dx_i dx_k, & \text{continuous} \\ \sum_{\text{all } x_i} \sum_{\text{all } x_k} (x_i - \mu_i)(x_k - \mu_k) p_{ik}(x_i, x_k), & \text{discrete} \end{cases}$$
$$\sigma_{ii} = \sigma_i^2$$

(joint p.d.f $f_{ik}(x_i, x_k)$ if continuous; joint p.m.f. $p_{ik}(x_i, x_k)$ if discrete)

Random Vectors and Matrices

$$\mathbf{X}_{p\times 1} = \begin{bmatrix} X_1 & X_2 & \dots & X_p \end{bmatrix}' \quad (p \times 1 \text{ random vector})$$

- Population mean vector $\boldsymbol{\mu}$

$$\boldsymbol{\mu} = \mathsf{E}(\mathbf{X}) = \begin{bmatrix} \mathsf{E}(X_1) \\ \mathsf{E}(X_2) \\ \vdots \\ \mathsf{E}(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

Random Vectors and Matrices

$$\mathbf{X}_{p\times 1} = \begin{bmatrix} X_1 & X_2 & \ldots & X_p \end{bmatrix}' \quad (p \times 1 \text{ random vector})$$

- Population covariance matrix $\mathbf{\Sigma}$

$$\mathbf{\Sigma} = \mathsf{E}\left(\begin{bmatrix} X_1 - \mu_1 \\ \vdots \\ X_p - \mu_p \end{bmatrix} [X_1 - \mu_1 \ \ldots \ X_p - \mu_p] \right)$$

$$= \mathsf{E}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \ldots & \sigma_{pp} \end{bmatrix}$$

$\mathbf{\Sigma}$: symmetric, since $\sigma_{ik} = \sigma_{ki}$

Random Vectors and Matrices

$$\mathbf{X}_{p\times 1} = \begin{bmatrix} X_1 & X_2 & \dots & X_p \end{bmatrix}' \quad (p \times 1 \text{ random vector})$$

- Population correlation matrix $\boldsymbol{\rho}$

$$\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}} \quad \text{(Population correlation coefficient)}$$

$$\rho_{ii} = 1, \quad \rho_{ik} = \rho_{ki}$$

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{bmatrix} \quad \text{(symmetric)}$$

Random Vectors and Matrices

$$\mathbf{X}_{p\times 1} = \begin{bmatrix} X_1 & X_2 & \dots & X_p \end{bmatrix}' \quad (p \times 1 \text{ random vector})$$

- Let the $p \times p$ standard deviation matrix be

$$\mathbf{V}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{\sigma_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\sigma_{pp}} \end{bmatrix}$$

- It can be shown that

$$\mathbf{V}^{1/2}\boldsymbol{\rho}\mathbf{V}^{1/2} = \boldsymbol{\Sigma}, \qquad \left(\mathbf{V}^{1/2}\right)^{-1}\boldsymbol{\Sigma}\left(\mathbf{V}^{1/2}\right)^{-1} = \boldsymbol{\rho}$$

Random Vectors and Matrices

$$\mathbf{X}_{p\times 1} = \begin{bmatrix} X_1 & X_2 & \dots & X_p \end{bmatrix}' \quad (p \times 1 \text{ random vector})$$

- $X_1, \dots, X_p$ are mutually (statistically) independent if $\forall\ (x_1, x_2, \dots, x_p)$

$$P[X_1 \le x_1, X_2 \le x_2, \dots, X_p \le x_p] \\ = P[X_1 \le x_1]P[X_2 \le x_2] \dots P[X_p \le x_p]$$

  or, for continuous random variables, if $\forall\ (x_1, x_2, \dots, x_p)$

$$f_{12\dots p}(x_1, x_2, \dots, x_p) = f_1(x_1)f_2(x_2)\dots f_p(x_p)$$

- $X_i, X_k (i \neq k)$ independent $\implies \sigma_{ik} = 0$;
  however, in general, $\sigma_{ik} = 0 \not\Longrightarrow X_i, X_k$ independent!!

Linear Combination of Random Variables

$$Z = \mathbf{a}'\mathbf{X}, \quad \mathbf{a}: \text{deterministic}, \;\; \mathbf{X}_{p\times 1} = [X_1\ X_2\ \ldots\ X_p]'(\text{random})$$

- Mean

$$\mathsf{E}(Z) = \mathbf{a}'\mathsf{E}(\mathbf{X}) = \mathbf{a}'\boldsymbol{\mu}$$

- Variance

$$\begin{aligned}
\mathsf{Var}(Z) &= \mathsf{E}\left(Z - \mathsf{E}(Z)\right)^2 = \mathsf{E}\left(Z - \mathsf{E}(Z)\right)\left(Z - \mathsf{E}(Z)\right)' \\
&= \mathsf{E}\left(\mathbf{a}'\mathbf{X} - \mathbf{a}'\boldsymbol{\mu}\right)\left(\mathbf{a}'\mathbf{X} - \mathbf{a}'\boldsymbol{\mu}\right)' = \mathsf{E}\left\{\mathbf{a}'\left(\mathbf{X} - \boldsymbol{\mu}\right)\left(\mathbf{X} - \boldsymbol{\mu}\right)'\mathbf{a}\right\} \\
&= \mathbf{a}'\mathsf{E}\left\{\left(\mathbf{X} - \boldsymbol{\mu}\right)\left(\mathbf{X} - \boldsymbol{\mu}\right)'\right\}\mathbf{a} \\
&= \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}
\end{aligned}$$

Linear Combination of Random Variables

$\mathbf{Z}_{n\times 1} = \mathbf{A}_{n\times p}\mathbf{X}_{p\times 1}, \quad \mathbf{A}$: deterministic, $\mathbf{X} = [X_1\ X_2\ \ldots\ X_p]'$(random)

- Mean vector

$$\mathsf{E}(\mathbf{Z}) = \mathbf{A}\mathsf{E}(\mathbf{X}) = \mathbf{A}\boldsymbol{\mu}$$

- Covariance matrix

$$\begin{aligned}
\mathsf{Cov}(Z) &= \mathsf{E}\left(\mathbf{Z} - \mathsf{E}(\mathbf{Z})\right)\left(\mathbf{Z} - \mathsf{E}(\mathbf{Z})\right)' \\
&= \mathsf{E}\left(\mathbf{A}\mathbf{X} - \mathbf{A}\boldsymbol{\mu}\right)\left(\mathbf{A}\mathbf{X} - \mathbf{A}\boldsymbol{\mu}\right)' = \mathsf{E}\left\{\mathbf{A}\left(\mathbf{X} - \boldsymbol{\mu}\right)\left(\mathbf{X} - \boldsymbol{\mu}\right)'\mathbf{A}'\right\} \\
&= \mathbf{A}\mathsf{E}\left\{\left(\mathbf{X} - \boldsymbol{\mu}\right)\left(\mathbf{X} - \boldsymbol{\mu}\right)'\right\}\mathbf{A}' \\
&= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'
\end{aligned}$$

Random Samples

- n sets of measurements of p variables

$$\mathbf{X}_{n\times p} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \begin{array}{lll} \rightarrow & 1^{\text{st}} \text{ sample} & \mathbf{X}_1' \\ \rightarrow & 2^{\text{nd}} \text{ sample} & \mathbf{X}_2' \\ \vdots & \vdots & \vdots \\ \rightarrow & n^{\text{th}} \text{ sample} & \mathbf{X}_n' \end{array}$$

- If row vectors $\mathbf{X}_1', \mathbf{X}_2', \dots, \mathbf{X}_n'$:
  (1) independent, (2) sharing the joint p.d.f. $f(\mathbf{x}) = f(x_1, x_2, \dots, x_p)$
  then $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$: a random sample from $f(\mathbf{x})$.
- Joint p.d.f. of random sample $\mathbf{X}_1, \dots, \mathbf{X}_n$: $f(\mathbf{x}_1)f(\mathbf{x}_2)\dots f(\mathbf{x}_n)$
  $f(\mathbf{x}_j) = f(x_{j1}, x_{j2}, \dots, x_{jp}), j = 1, \dots, n$.

Random Samples

- n sets of measurements of p variables

$$\mathbf{X}_{n\times p} = \begin{bmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \\ \vdots \\ \mathbf{X}_n' \end{bmatrix}, \quad \mathbf{X}_j' = [X_{j1}, X_{j2}, \ldots, X_{jp}]$$

  - Measurements of p variables in a single trial,

    e.g., $\mathbf{X}_j' = [X_{j1}, X_{j2}, \ldots, X_{jp}]$, will generally be correlated.
  - However, Measurements from different trials must be independent.
  - Independence of measurements from trial to trial may not hold when the variables are likely to drift over time.

Random Samples

Let the $p \times 1$ vectors: $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from a joint distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

$$\overline{\mathbf{X}} \triangleq \frac{\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n}{n}, \quad \mathbf{S}_n \triangleq \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)'$$

1) $\mathsf{E}(\overline{\mathbf{X}}) = \dfrac{\mathsf{E}(\mathbf{X}_1 + \ldots + \mathbf{X}_n)}{n} = \dfrac{n\boldsymbol{\mu}}{n} = \boldsymbol{\mu}$ (Population mean vector)

2) $\mathsf{Cov}(\overline{\mathbf{X}}) = \dfrac{1}{n}\boldsymbol{\Sigma}$ (Population cov. matrix divided by sample size)

3) For the covariance matrix $\mathbf{S}_n$

$$\mathsf{E}\left(\mathbf{S}_n\right) = \frac{n-1}{n}\boldsymbol{\Sigma} \Longrightarrow \mathsf{E}\left(\frac{n}{n-1}\mathbf{S}_n\right) = \frac{n-1}{n}\boldsymbol{\Sigma}$$

2), 3) to be continued on the next page

Random Samples

$\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$: random sample from a joint distribution ($\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$)

$$\mathsf{Cov}(\overline{\mathbf{X}}) = \mathsf{E}\left\{\left[\overline{\mathbf{X}} - \mathsf{E}(\overline{\mathbf{X}})\right]\left[\overline{\mathbf{X}} - \mathsf{E}(\overline{\mathbf{X}})\right]'\right\} = \mathsf{E}\left\{\left(\overline{\mathbf{X}} - \boldsymbol{\mu}\right)\left(\overline{\mathbf{X}} - \boldsymbol{\mu}\right)'\right\}$$

$$= \mathsf{E}\left\{\left(\frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \boldsymbol{\mu})\right)\left(\frac{1}{n}\sum_{k=1}^{n}(\mathbf{X}_k - \boldsymbol{\mu})\right)'\right\}$$

$(\mathbf{X}_j, \mathbf{X}_k (j \neq k)$ independent $\rightarrow \mathsf{E}(\mathbf{X}_j - \boldsymbol{\mu})(\mathbf{X}_k - \boldsymbol{\mu})' = 0)$

$$= \frac{1}{n^2}\mathsf{E}\left\{\sum_{j=1}^{n}(\mathbf{X}_j - \boldsymbol{\mu})(\mathbf{X}_j - \boldsymbol{\mu})'\right\}$$

$$= \frac{n}{n^2}\boldsymbol{\Sigma} = \frac{1}{n}\boldsymbol{\Sigma}$$

Random Samples

$\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$: random sample from a joint distribution ($\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$)

$$\mathsf{E}(\mathbf{S}_n) = \frac{1}{n}\mathsf{E}\left\{\sum_{i=1}^{n}\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)'\right\}$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n}\mathsf{E}(\mathbf{X}_i\mathbf{X}_i') - \sum_{i=1}^{n}\mathsf{E}(\overline{\mathbf{X}}\mathbf{X}_i') - \sum_{i=1}^{n}\mathsf{E}(\mathbf{X}_i\overline{\mathbf{X}}') + n\mathsf{E}(\overline{\mathbf{X}}\overline{\mathbf{X}}')\right\}$$

$$\left(\sum \mathsf{E}(\overline{\mathbf{X}}\mathbf{X}_i') = \sum \mathsf{E}(\mathbf{X}_i\overline{\mathbf{X}}) = n\mathsf{E}(\overline{\mathbf{X}}\overline{\mathbf{X}}')\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\mathsf{E}(\mathbf{X}_i\mathbf{X}_i') - \mathsf{E}(\overline{\mathbf{X}}\overline{\mathbf{X}}')\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{(\boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\Sigma}) - (\boldsymbol{\mu}\boldsymbol{\mu}' + \frac{1}{n}\boldsymbol{\Sigma})\right\} = \frac{n-1}{n}\boldsymbol{\Sigma}$$

Random Samples

$\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$: random sample from a joint distribution ($\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$)

- From

$$\mathsf{E}(\mathbf{S}_n) = \frac{n-1}{n}\boldsymbol{\Sigma} \Longrightarrow \mathsf{E}\left(\frac{n}{n-1}\mathbf{S}_n\right) = \frac{n-1}{n}\boldsymbol{\Sigma}$$

  - $\mathbf{S}_n$: biased estimator of $\boldsymbol{\Sigma}$
  - $\frac{n}{n-1}\mathbf{S}_n$: unbiased estimator of $\boldsymbol{\Sigma}$