■ Five Questions to ask a DATA.
(i) What has happened? (Descriptive Studies)
(ii) What will happen? (Predictive Studies).
(iii) is it true or false? (Inferential Studies)
(iv) How it is working? (Pattern recognisation or Cluster Analysis)
(v) (Anomaly Detection) is it weird?

In general, most of the data-driven studies are conducted to understand the nature of various characteristics within a population. These studies encompass pure data analysis as well as statistical analysis of data obtained from the population. Insights derived from the analysis are used to explain the nature of these characteristics within the population and to inform policy development / making.
Any data-oriented study can be broadly classified into the following types:

1. Descriptive Analysis (these answer to a general questions such as "What has happened?")

2. Predictive Analysis (these answer to a general questions such as "What will happen?")

3. Inferential Studies (these answer to a general questions such as

    a) What is the value of _________? (estimation)
    b) Is the given statement true? (hypothesis testing)

4. Anomaly Detection (these answer to a general questions such as "is it wired ?")(eg. detecting fraudulent financial transaction)

5. Looking for patterns (these answer to a general questions such as "How is it organized?")

Generally, the data is provided or it has to be collected.
Types of Data Collection are:

1. Complete Enumeration (e.g., Census).

2. Sample Survey (taking a part of the population into the study).

3. Collecting Data from Secondary Sources (government publications, different websites).

In reality,data falls into two broad types, namely, unstructed data and structured data.

- Unstructed Data: data which are not expressed in a row-column format. Examples include audio clips, video clips, etc.

- Structured Data: data which are expressed in a row-column format. Examples Data of age , gender, income of 10 people

In practice, a structured dataset contains observations on several variables. For example, a dataset of socio-economic surveys may include values such as age, gender, monthly income, monthly expenditure, and educational quality of 1000 people. Typically, it looks like:

Sl.No.	Age	Gen	Monthly income	Monthly Expenditure	Qualification
1	24	F	20k	12k	HS
2	31	M	70k	50k	Graduation
.	.	.	.	.	.
.	.	.	.	.	.
1000	45	M	90k	75k	Post Graduation

It is clearly evident that the above dataset contains observations on 5 characteristics (variables) of the population. Hence, this is called a Multivariate Dataset. Therefore, any study out of the 5 types of studies mentioned above should include all the variables at a time. When the data analysis is carried out with more than 2 variables at a time, it is called "Multivariate Data Analysis." It may again be noted that studies on multivariate data will also fall into one or many of the above 5 types of studies.

When Multivariate Data Analysis (MDA) involves or takes into account the probability distribution of more than one variable or the joint probability distribution of more than one variable, then it is called Multivariate Statistical Data Analysis (MSDA) .

In 'Applied Multivariate Statistical Data Analysis' (MVSDA), we consider the following:

1. Dimensionality Reduction: Situations often arise when the number of variables in the dataset is quite large. Processing data with large dimensions becomes tedious and difficult to interpret. Therefore, there is no other way but to reduce the dimension of the dataset for effective analysis and interpretation of the data. The dimension of a dataset should be reduced in such a way that the information contained in the original dataset is not lost. We perform this task using Principal Component Analysis. Factor Analysis is also considered as a dimension reduction technique, although its primary usefulness is in explaining the covariance (correlation) structure of a group of variables.

2. Cluster Analysis: Here, the journey starts with the set of observations or objects and ends with obtaining one or more groups of objects where objects within a group are similar and objects between the groups are dissimilar. For example, grouping movies in a video repository.

3. Discriminant Analysis or Classification: In this type of study, we are given a set of multivariate observations, and we are also told from how many population groups they are coming from. Here, our task is to develop a discriminant rule which discriminates observations from different groups. Secondly, we use that rule to classify a new observation in the system to either of the known population groups or classes. For example, classification of emails as spam or non-spam.

4. Testing equality of Mean Vectors of Several Multivariate Populations: This is done using a technique called MANOVA (Multivariate Analysis of Variance). Few more inferential problems, such as estimating mean vectors and variance-covariance matrices, and different tests on those, is also undertaken here.

5. The studies also extend to finding correlations between two groups of variables; this is called 'Canonical Correlation Analysis.'

   NEEDS:

- Linear Algebra: To carry out the above studies, we should have a good understanding of vectors and matrices.
- Software: We would also use software like R, Python, and other open-source menu-driven solutions to demonstrate.

A typical multivariate data, say with p variables, can be represented as follows:

$$\text{The following matrix represents:} \longrightarrow \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

This matrix, also called a Data Matrix, has each column representing observations on a specific variable. The above matrix has the order $n \times p$, where there are n observations on each of the p variables.

$$\text{The same can also be represented as:} \longrightarrow \begin{pmatrix} x_{11} & x_{21} & \dots & x_{p1} \\ x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \dots & x_{pn} \end{pmatrix}_{n\times p}$$

It may be noted that each column represents observations on each variable on n individuals , and each row represents an observation of a population unit for all the variables.
Hence, we may call the rows above observational vectors.
Since we have to deal with data matrices and observational vectors, we should have a proper understanding of vector and matrix algebra.

Here is a quick recap:

- A vector is an ordered tuple of numbers. For example, (x, y) is a vector in 2-D plane where the first value indicates position on the horizontal axis and the second value indicates position on the vertical axis.
- Members of a vector are called its 'elements'. The maximum number of elements of a vector is its 'dimension'. That is, a p-component or p-dimensional vector will have p elements where the i-th element will indicate the position on the i-th axis.
- A p-component vector denoted by $\underline{x}^{p\times 1}$ can be represented as
  $\underline{x}^{p\times 1} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$ or
- The transpose of $x^{p\times 1}$ is denoted by x'. or $\underline{x}^{1\times p} = \begin{pmatrix} x_1 & x_2 & \cdots & x_p \end{pmatrix}_{1\times p}$
- Operations of Vectors: Let $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$ be a $p \times 1$ vector. If each element of x is

multiplied by a scalar c (say), then we get a new vector, say $\underline{y}$, where $\underline{y} = \begin{pmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_p \end{pmatrix}$.

Then y is called a scalar multiple of the vector $\underline{x}$

- Vector Addition: Addition of 2 vectors is defined if both vectors are of the same order/dimension.

$$\text{Let } \underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \text{ and } \underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}$$

then $\underline{x} + \underline{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_p + y_p \end{pmatrix} = \underline{z}$ where the elements of $\underline{z}$ are the sum of the corresponding elements of $\underline{x}$ and $\underline{y}$.

Similarly, $(\underline{x} - \underline{y})$ is defined as $\begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_p - y_p \end{pmatrix}$.

Similarly, subtraction of vectors is defined.

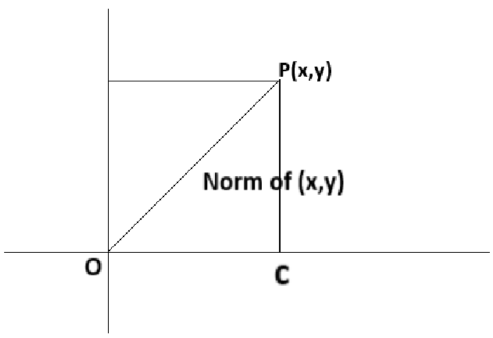
- Scalar Product: It is defined when two vectors are of the same order.

  Scalar product of two vectors when defined is the sum of products of corresponding elements of two vectors.

  That is, if $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$ and $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}$, the scalar product of x and y, defined as $x'y$ or $\underline{y}'x$, is defined as

$$\underline{x'y} = \sum_{i=1}^{p} x_i y_i$$

- Norm:



The distance of a vector from its origin:

$$(OP)^2 = (OC)^2 + (PC)^2$$
$$\Rightarrow \overline{OP} = \sqrt{(OC)^2 + (PC)^2} = \sqrt{x^2 + y^2}.$$

- Two vectors are said to be orthogonal if their scalar product is zero.
- Norm of a vector is denoted by $||x||$.A vector whose norm is unity is called a 'unit normed vector'.

- A set of vectors is said to be 'linearly independent' if none of the vectors can be linearly expressed by the remaining vectors or any one.
  That is, the set of vectors $x_1, x_2, \ldots, x_k$ will be called linearly independent if

$$l_1x_1 + l_2x_2 + \cdots + l_kx_k = 0$$

  implies

$$l_1 = l_2 = \ldots = l_k = 0 \quad \text{where} \quad l_i \quad \text{are scalers}$$

  For example, if $l_1x_1 + l_2y_2 = 0$, then $x_2 = -\frac{l_1}{l_2}x_1$.
  $(1,0,0), (0,1,0), (0,0,1)$ is a set of linearly independent vectors in $\mathbb{R}^3$.

Vector Space: A set of vectors closed under scalar multiplication and vector addition. Let V be a collection of vectors $\underline{a}_i, i = 1, 2, \ldots$. Then V will be called a vector space if:

1. $C\underline{a}_i \in V$ for all $\underline{a}_i \in V$, $i = 1, 2, \ldots$, where C is a scalar.
2. $\underline{a}_i + \underline{b}_i \in V$ for all $\underline{a}_i, \underline{b}_i \in V$, $i = 1, 2, \ldots$.

Example: $\mathbb{R}^2 : \{(x, y) : x, y \in \mathbb{R}\}$

Basis: A set of linearly independent vectors which span the vector space is called its basis.
The number of basis of a vector space may not be unique. However, the number of vectors in a basis of a vector space is unique.

Ex: $\underline{e_1}' = (1,0), \underline{e_2}'' = (0,1)$ is a basis of $\mathbb{R}^2$
$\left\{\underline{e_1}' = (1,0,0), \underline{e_2}' = (0,1,0), \underline{e_3}' = (0,0,1)\right\}$ is a basis of $\mathbb{R}^3$

Verify these examples (take any vector and express it using basis vectors).

Matrix and its Results:

Let $A = \begin{pmatrix} x \\ y \end{pmatrix}$ be any vector in the vector space $\mathbb{R}^2$.

Now, A can be expressed as:

$$x\underline{e}_1' + y\underline{e}_2'$$
Where $\underline{e}_1' = (1,0)$
$$\underline{e}_2' = (0,1)$$
Now, $x\underline{e}_1' + y\underline{e}_2'$
$$= x\begin{pmatrix}1\\0\end{pmatrix} + y\begin{pmatrix}0\\1\end{pmatrix}$$
$$= \begin{pmatrix}x\\0\end{pmatrix} + \begin{pmatrix}0\\y\end{pmatrix}$$
$$= \begin{pmatrix}x\\y\end{pmatrix} = A.$$

Matrices are rectangular arrays of numbers. The number of rows by the number of columns is called the order of the matrix.
Example:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \ddots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Here A is an $m \times n$ matrix, i.e., A has m rows and n columns.
Each row and each column is a vector.

Vector Space generated by the rows of a matrix, say A, is called the Row Space of A, denoted by $\mathscr{R}(A)$
Vector Space generated by the columns of a matrix, say A, is called the Column Space of A, denoted by $\mathscr{C}(A)$.
The number of linearly independent vectors in the row space of A is the row rank of A.
Similarly, the number of linearly independent vectors in the column space of A is the column rank of A.
It can be proved that for a matrix A, Row Rank = Column Rank.
The number of linearly independent vectors (row/column) in a matrix is called its rank.

Types of Matrices:

1. Square Matrix: The number of rows equals the number of columns.

2. Triangular Matrix: A square matrix where all the elements above or below the principal diagonal are zero. For example:

$$A = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \text{ (lower triangular)}$$

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & b_{23} \\ 0 & 0 & b_{33} \end{pmatrix} \text{ (upper triangular)}$$

3. Diagonal Matrix: A square matrix whose all elements other than the principal diagonal are zero.

$$A = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{12} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}$$

4. Identity Matrix: A diagonal matrix whose principal diagonal elements are unity or one.

5. Symmetric Matrix: A square matrix is said to be symmetric if its mirror image positional elements are the same.
The matrix A will be called a symmetric matrix if $a_{ij} = a_{ji}$.

6. Idempotent Matrix: A matrix B is said to be idempotent if $B^2 = B$.

7. Inverse of a Matrix: If the multiplication of two matrices in either order results in the identity matrix, then each one of the previous matrix are inverse of one another. That is, if A and B are two matrices such that $AB = BA = I$, then A is the inverse of B and vice versa. $A^{-1} = B$ and $B^{-1} = A$.

8. Orthogonal Matrix:

   - Transpose: The matrix obtained by transposing the rows and columns of a matrix is known as the Transpose of the Matrix.
     This means changing the rows to columns and columns to rows. Notation: $A^\top$ or A'

   A Matrix ' A ' is said to be orthogonal if $AA' = A'A = I$ (Square).

9. Determinant of a Matrix: A value obtained from the matrix which determines whether the inverse of a matrix exists or not.

   A matrix whose determinant is zero is called a Singular Matrix.

   Let A be a non-singular square matrix, then

$$A^{-1} = \frac{\text{Adj}(A)}{|A|}$$

where $\mathrm{Adj}(A)$ is the adjoint matrix of A,

which is the matrix obtained by transposing the cofactor of each element of A.

The cofactor of the element a_{ij} of A is $(-1)^{i+j}$ times the determinant of the matrix obtained from A by deleting its ith row and its jth column,

which is equal to $(-1)^{i+j}$ times the minor of a_{ij} of A.

■ System of Linear Equations:

$$A_{p\times p}x_{p\times 1} = b_{p\times 1} \cdots (i)$$

If $|A| \neq 0$, then $\underline{x} = A^{-1}\underline{b}$
where $\underline{x}$ is a vector of unknown quantities whose values can be found out using (i). Note that the solution is unique since A^{-1} is unique. The inverse of a matrix, if it exists, is unique. It may happen that the system of linear equations is of the form

$$A\underline{x} = \underline{b}$$

The system of linear equations will be called consistent if there exists at least one solution of the system of linear equations. A system of linear equations is consistent if

$$\mathrm{Rank}(A) = \mathrm{Rank}(A : b)$$

If a system of linear equations is not consistent, then it will have no solution.
If the coefficient matrix of the system of linear equations is of full rank, i.e., its inverse exists, then the system will have a unique solution.

If the coefficient matrix of the system of linear equations is not of full rank, then it will have an infinite number of solutions. If $\underline{b} = 0$, then the system is called a system of 'Homogeneous Equations'.

$$\begin{aligned}
&2x_1 + 3x_2 = 5\\
&4x_1 + 6x_2 = 10\\
&\Rightarrow \begin{pmatrix} 2 & 3\\ 4 & 6\end{pmatrix}\begin{pmatrix} x_1\\ x_2\end{pmatrix} = \begin{pmatrix} 5\\ 10\end{pmatrix}\\
&A\cdot x = b\\
&\text{where } |A| = 0 \quad \text{and} \quad \mathrm{Rank}(A) = \mathrm{Rank}\left(\begin{pmatrix} 2 & 3 & : & 5\\ 4 & 6 & : & 10\end{pmatrix}\right) = 1
\end{aligned}$$

So, the system of linear equations is consistent. Thus, it should have an infinite number of solutions.

$$\begin{aligned}
\begin{pmatrix} 2 & 3\\ 4 & 6\end{pmatrix}\begin{pmatrix} x_1\\ x_2\end{pmatrix} &= \begin{pmatrix} 5\\ 10\end{pmatrix}\\
\Rightarrow \begin{pmatrix} 2 & 3\\ 0 & 0\end{pmatrix}\begin{pmatrix} x_1\\ x_2\end{pmatrix} &= \begin{pmatrix} 5\\ 0\end{pmatrix}\\
\Rightarrow 2x_1 + 3x_2 &= 5\\
\Rightarrow x_1 &= \frac{5-3x_2}{2}\\
\text{Let } x_2 = k, x_1 &= \frac{1}{2}(5-3k)
\end{aligned}$$

Solution of the system of equations:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(5-3x) \\ x \end{pmatrix}$$

Now for an infinite choice of k, we have infinite solutions.

$$2x_1 + 3x_2 = 5$$
$$4x_1 + 5x_2 = 10$$

Check the consistency of the system of equations:

$$\begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 10 \end{pmatrix}$$
$$\text{Rank}(A) = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} = Z = \text{Rank}\left(\begin{pmatrix} 2 & 3 & : & 5 \\ 4 & 6 & : & 10 \end{pmatrix}\right)$$

So, the system of linear equations is consistent since it is of full rank. Therefore, it has a unique solution (its inverse exists).

$$\det \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} = 10 - 12 = -2, \qquad \det \begin{pmatrix} 3 & 5 \\ 5 & 10 \end{pmatrix} = 5.$$
$$\det \begin{pmatrix} 2 & 5 \\ 4 & 10 \end{pmatrix} = 20 - 20 = 0$$

Clearly, there are 2 nonzero minors of order 2 of the augmented matrix. Hence the rank of the augmented matrix is 2.

$$2x_1 + 3x_2 = 5$$
$$4x_1 + \lambda x_2 = \mu$$

For what value of λ and μ will the following system of linear equations have i) no solution, ii) a unique solution, iii) infinite solutions

■ Eigenvalues and Eigenvectors of a Matrix:
Let $A^{n \times n}$ be a square matrix. Then eigenvalues of A are the solutions of the system of equations

$$|A - \lambda I| = 0$$

Eigenvectors x_i of A corresponding to the eigenvalues $\lambda_i, i = 1, 2, \ldots$ of A satisfy $A\underline{x}_i = \lambda_i \underline{x}_i$.
Eigen vectors corresponding to distinct eigenvalues are linearly independent. Let $A = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix}$. To find its eigenvalues we need to solve

$$\left| \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = 0$$
$$\Rightarrow \begin{vmatrix} 2-\lambda & 3 \\ 4 & 5-\lambda \end{vmatrix} = 0$$
$$\Rightarrow (2-\lambda)(5-\lambda) - 12 = 0$$
$$\Rightarrow \lambda^2 - 7\lambda - 2 = 0$$
$$\Rightarrow \lambda = 7.25 \text{ or } \lambda = 0.25$$

$\therefore$ Eigen vectors corresponding to the eigen value 7.25 are

$$\begin{pmatrix} 2-7.25 & 3 \\ 4 & 5-7.25 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
$$\Rightarrow \quad -5.25x_1 + 3x_2 = 0$$
$$4x_1 - 2.25x_2 = 0$$

It may be noted that eigenvectors of a matrix can be converted to Orthogonal eigenvectors through Graham-Schmidt orthogonalization process.
Let A be a symmetric matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ and corresponding orthogonal eigenvectors $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_p$, then A can be expressed as:

$$\begin{aligned} A &= \lambda_1 \underline{e}_1 \underline{e}_1' + \lambda_2 \underline{e}_2 \underline{e}_2' + \ldots + \lambda_p \underline{e}_p \underline{e}_p' \\ &= \sum_{i=1}^{p} \lambda_i \underline{e}_i \underline{e}_i' \end{aligned}$$

This is also known as the spectral decomposition of the matrix A.

Take a matrix and find its spectral decomposition

$$\begin{aligned} & A = \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix} \\ & |A - \lambda I| = 0 \\ & \begin{vmatrix} 1-\lambda & 2 \\ 2 & -2-\lambda \end{vmatrix} = 0 \\ & \Rightarrow (1-\lambda)(-2-\lambda) - 4 = 0 \\ & \Rightarrow \lambda^2 + \lambda - 6 = 0 \\ & \Rightarrow \lambda = -3, \lambda = 2 \end{aligned}$$

For $\lambda = -3$, $Ax = \lambda x$:

$$\begin{aligned} & \Rightarrow \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -3\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ & \Rightarrow \begin{pmatrix} x_1 + 2x_2 \\ 2x_1 - 2x_2 \end{pmatrix} = \begin{pmatrix} -3x_1 \\ -3x_2 \end{pmatrix} \\ & \therefore x_1 + 2x_2 + 3x_1 = 0 \\ & \Rightarrow 4x_1 + 2x_2 = 0 \text{ and } 2x_1 - 2x_2 + 3x_2 = 0 \\ & \Leftrightarrow 2x_1 + x_2 = 0 \end{aligned}$$

For $\lambda = 2$, $Ax = \lambda x$:

$$\begin{aligned} & \Rightarrow \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \\ & \Rightarrow \begin{pmatrix} -x_1 + 2x_2 \\ 2x_1 - 4x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \therefore x_1 = 2, x_2 = 1 \end{aligned}$$

Now for $\lambda = -3$, norm of the eigen vectors is $\sqrt{1^2 + 2^2} = \sqrt{5}$

Hence an eigenvector for $\lambda = -3$ is $e_1' = \left(-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)$

$P'\lambda_P = \text{ding}\left(\lambda_1\lambda_2\ldots\lambda_p\right)$ where $\lambda_{3_1}\lambda_2\ldots\lambda_p$ are eigen values of A and P is the orthogonal matrix constructed by the corresponding eigenvectors $e_1, e_2, \ldots$ es the matrix A.

Similarly for $\lambda = 2$, norm of the eigen vector is $\sqrt{2^2+1^2} = \sqrt{5}$ $\therefore$ An eigenvector for $\lambda = 2$ is $\ell_2' = \left(1/\sqrt{5}, \frac{2}{\sqrt{5}}\right)'$ Hence a spectral decomposition of matrix A is

$$\begin{aligned}
\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} &= -3\begin{bmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}\begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \\
&+ 2\begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \\
&= -3\begin{bmatrix} \frac{1}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{4}{5} \end{bmatrix} + 2\begin{bmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{bmatrix}
\end{aligned}$$

■ **Spectral Decomposition, Eigenvalues, Eigenvectors.**

$$\text{Let } A = \begin{pmatrix} 2 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix}$$

Find the eigenvalues and eigenvectors of A.

$$\Rightarrow A - \lambda I = \left(\begin{array}{cc} 2-\lambda & \sqrt{2} \\ \sqrt{2} & 1-\lambda \end{array}\right), \text{ let } \lambda \text{ be an eigenvalue of matrix A by definition}$$

$\Rightarrow \det(A - \lambda I)$

A, then.

$= (2-\lambda)(1-\lambda) - 2$

$= \lambda^2 - 3\lambda + 2 - 2$

$= \lambda(\lambda - 3)$

$\Rightarrow$ if $\det(A - \lambda I) = 0 \Rightarrow \lambda = 0, 3$

$\lambda_1 = 0, \lambda_2 = 3$

For $\lambda_1 = 0$, let the eigenvector be $\underline{x} = \binom{x_1}{x_2}$

$$\begin{aligned}
&\Rightarrow \begin{pmatrix} 2-0 & \sqrt{2} \\ \sqrt{2} & 1-0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0 \\
&\Rightarrow 2x_1 + \sqrt{2}x_2 = 0 \\
&\Rightarrow \sqrt{2}x_1 + x_2 = 0 \Rightarrow x_1 = \frac{-x_2}{\sqrt{2}} \Rightarrow x_2 = -\sqrt{2}x_1
\end{aligned}$$

Also

$$\begin{aligned}
&A\underline{x} = \lambda_1\underline{x} \\
&\Rightarrow \begin{pmatrix} 2x_1 + \sqrt{2}x_2 \\ \sqrt{2}x_1 + x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
\end{aligned}$$

For $\lambda_2 = 3$, the eigenvector is $\underline{y} = \binom{y_1}{y_2}$

$$
\begin{aligned}
&A\underline{y} = \lambda_2 \underline{y} \\
&\Rightarrow \begin{pmatrix} 2y_1 + \sqrt{2}y_2 \\ \sqrt{2}y_1 + y_2 \end{pmatrix} = \begin{pmatrix} 3y_1 \\ 3y_2 \end{pmatrix} \\
&\Rightarrow y_1 = \sqrt{2}y_2 \quad \Rightarrow \quad y_1 = \sqrt{2}y_2 \\
&\sqrt{2y_1} = 2y_2 \quad \Rightarrow y_2 = \frac{y_1}{\sqrt{2}}
\end{aligned}
$$

Eigenvector for $\lambda_1 = 0 \Rightarrow \binom{x_1}{-\sqrt{2}x_1}, x_1 \in \mathbb{R}$

Eigenvector for $\lambda_2 = 3 \Rightarrow \begin{pmatrix} y_1 \\ y_1 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, y_1 \in \mathbb{R}$

based on the choice of $x_1 = a_1^{a_1}, b_1$, eigenvectors will vary.
This system of equations has infinitely many solutions. We can transform them to orthonormal eigenvectors by dividing with now.

$$
\begin{aligned}
\underline{e}_1 &= \begin{pmatrix} -\frac{1}{\sqrt{2}\sqrt{1.5}} \\ \frac{1}{\sqrt{1.5}} \end{pmatrix}, \underline{e}_2 = \begin{pmatrix} \frac{\sqrt{2}}{\sqrt{3}} \\ \frac{\sqrt{1}}{\sqrt{3}} \end{pmatrix} \\
\underline{e}_1' \underline{e}_2 &= \left(-\frac{\sqrt{2}}{3} + \frac{\sqrt{2}}{3} \right) = 0
\end{aligned}
$$

$\therefore e_1, e_2$ are orthonormal (for particular a_1, b_1)

$$\text{ie. } \|e_1\| = \|e_2\| = 1$$

Define

$$
\begin{aligned}
P &= \begin{pmatrix} e_1 & e_2 \end{pmatrix} \\
&= \begin{pmatrix} -\frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \\ \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
PAP' &= \begin{pmatrix} -\frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \\ \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} 2 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \\ \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 \\ \sqrt{6} & \sqrt{3} \end{pmatrix} \cdot \begin{pmatrix} -\frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \\ \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 \\ 0 & 3 \end{pmatrix}
\end{aligned}
$$

$$PAP' = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

$$\Rightarrow AP' = P' \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \text{ [by premultiplying } P' = P^{-1}\text{]}$$

$$\Rightarrow A = P' \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} P \text{ [post multiplying } P\text{]}$$

$$\Rightarrow A = \begin{pmatrix} \underline{e}'_1 \\ \underline{e}'_2 \end{pmatrix}_{2\times 1} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \underline{e}_1 & \underline{e}_2 \end{pmatrix}$$

$$\Rightarrow A = [\lambda_1 \underline{e}'_1 \underline{e}_1 + \lambda_2 \underline{e}'_2 \underline{e}_2] \quad \text{[Spectral Decomposition]}$$

A quadratic form is an algebraic expression of order two in two variables. A typical quadratic form can be expressed as $[a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2]$. The same may be expressed as follows in matrix notation:

$$\begin{array}{l} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \binom{x_1}{x_2} \\ a_{12} = a_{21} \text{ so, } \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{22} \\ a_{21} & a_{22} \end{pmatrix} \binom{x_1}{x_2} \end{array}$$

i.e., $x'Ax$ where x is the vector of variables and A is called the matrix (coefficient matrix) of the quadratic form. A is a symmetric matrix.

Example: Consider the quadratic form

$$\begin{array}{l} 4x_1^2 + 2x_1x_2 + 4x_2x_1 + 6x_2^2. \\ = 4x_1^2 + 6x_1x_2 + 6x_2^2 \end{array}$$

Matrix form $\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 4 & 3 \\ 3 & 6 \end{pmatrix} \binom{x_1}{x_2}$

It may be noted that the given quadratic form can be expressed in Matrix notation as follows as well

$$\begin{array}{l} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ 4 & 6 \end{pmatrix} \binom{x_1}{y_2} \Rightarrow \underline{x}' A \underline{x} \\ \Rightarrow \underline{x}' = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \\ A_1 = \begin{pmatrix} 4 & 2 \\ 4 & 6 \end{pmatrix} \\ \underline{x} = \binom{x_1}{x_2} \end{array}$$

However, it may be noted that,
A is a symmetric matrix.
here A is not symmetric.

Since the nature of a quadratic form entirely depends on its coefficient matrix and it is easier to work with a symmetric matrix, therefore, we shall always express a quadratic form through a symmetric matrix.

A quadratic form may be classified into the following classes based on the eigenvalues of matrix A:

1. Positive definite: $x'Ax > 0$. It happens when the coefficient matrix A is positive definite.

2. Positive semidefinite: $x'Ax \geqslant 0$.

3. Negative definite: $x'Ax < 0$.

4. Negative semidefinite: $x'Ax \leqslant 0$.

5. Indefinile: some +, some-
no of '+'eigen values - no of '-'eigen values. = Signature of the Coeff Matrix

A quadratic form $\underline{x}'A\underline{x}$ is said to be indefinite if the form is positive for some points $\underline{x}$ and negative for others.

A quadratic form $\underline{x}'\lambda\underline{x}$ is:

- Positive definite if it is positive (> 0) for every $\underline{x} \neq 0$.
- Positive semidefinite if it is non-negative (≥ 0) for every $\underline{x}$ and there exist points $\underline{x} \neq 0$ for which $\underline{x}'A\underline{x} = 0$.

If $\underline{x}'A\underline{x}$ is positive definite/semidefinite, then $\underline{x}'(-A)\underline{x}$ is negative definite/negative semidefinite.

A symmetric matrix A is often said to be positive definite, positive semidefinite, negative definite, etc., if the respective quadratic form $\underline{x}'A\underline{x}$ is positive definite, positive semidefinite, etc.

Example:

1. $F = 3x_1^2 + 5x_2^2$
   $F = 2x_1^2 + 3x_2^2 + x_3^2$
   $F = x_1^2$ are some examples of positive definite quadratic forms in 2,3,1 variables, respectively.

2. $F = 4x_1^2 + x_2^2 - 4x_1x_2 + 3x_3^2 = (2x_1 - x_2)^2 + 3x_3^2$ is a positive semidefinite quadratic form in 3 variables since it is never negative, and its value becomes zero if $x_2 = 2x_1$ and $x_3 = 0$.

(Homework) Find the associated coefficient matrix of the given quadratic forms in the above examples. Comment on the nature of the quadratic forms and the rank of the matrices.

Homework

(1)

$$\begin{aligned} F &= 3x_1^2 + 5x_2^2 \\ &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &\approx \underline{x}' A \underline{x} \\ A &= \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix} \text{ is a full rank matrix.} \end{aligned}$$

Hence, the determinant of A is non-zero.
All the eigenvalues of A are positive (3 and 5).
$\therefore$ the quadratic form is positive definite.

$$\begin{aligned} F &= 2x_1^2 + 3x_2^2 + x_3^2 \\ &= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad [\underline{x}' A \underline{x}] \end{aligned}$$

A is a full rank matrix.
All the eigenvalues are positive.
Therefore, it is positive definite.

(2)

$$\begin{aligned} F &= 4x_1^2 + x_2^2 - 4x_1x_2 + 3x_3^2 \\ &= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} 4 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \end{aligned}$$

Here, the coefficient matrix $A = \begin{pmatrix} 4 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ is a singular matrix since its rank is $\text{Rank}(A) = 2$.

$$[R_1 = -2R_2]$$

Hence, out of 3 eigenvalues of A, two are nonzero and the other one is zero.
Therefore, A is a positive semidefinite matrix, so is $x' A \underline{x}$.
* Maxima - minima of Quadratic form based on λ.
* Principal Components - choice.

$$\begin{aligned} &\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p \geqslant 0 \\ &V(Y_1) \geqslant V(Y_2) \geqslant \ldots \geqslant \text{Var}(Y_p) \qquad \begin{bmatrix} \text{we have formed } Y_1, \ldots, Y_p \\ \text{using } X_1, \ldots, X_n \end{bmatrix} \\ &\text{Var}(Y_1) = a_1' \Sigma a_1 \end{aligned}$$

Here Σ is the largest eigenvalue λ_1.
a_1 is a eigen vector using λ_1.

2nd Method.

$$\text{Var}(X_1) = \underline{a_1}' \Sigma \underline{a_1}$$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

Now, $\begin{pmatrix} a_{11} & a_{12} \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \binom{a_{11}}{a_{12}}$

We have to maximize $\mathrm{Var}(Y_2) \longrightarrow$ no feasible solution, unbounded solution. for feasible solution $a_i a_1 = 1$ i.e., norm vector.
Define $F = \underline{a_1}'\Sigma\underline{a_1} - \lambda(\underline{a_1}'\underline{a_1} - 1)$ (Lagrange Multiplier) Now,

$$\frac{\partial F_1}{\partial \underline{a_1}} = 0 \Rightarrow 2\Sigma\underline{a}_1 - 2\lambda\underline{a}_1 = 0$$
$$\Rightarrow (\Sigma - \lambda I)\underline{a_1} = 0 \cdots (*)$$

$\therefore (\Sigma - \lambda I)$ must be singular to have a nonzero solution of $(*)$

$$\therefore |\Sigma - \lambda I| = 0$$

From $(\Sigma - \lambda I)\underline{a}_1 = 0$
$\Rightarrow \Sigma\underline{a_1} = \lambda\underline{a_1}$
$\Rightarrow \underline{a_1}'\Sigma\underline{a_1} = \underline{a_1}'\lambda\underline{a_1}$
$\Rightarrow Var(Y_1) = \lambda \quad \text{as} \underline{a_1}'\underline{a_1} = 1$
$\Rightarrow$ largest eigenvalue of $|\Sigma - \lambda I| = 0$
for 2nd PC $\longrightarrow F_1^* = \underline{a}_2'\Sigma\underline{a}_2 - \lambda(\underline{a}_2'\underline{a}_2 - 1) - \mu(\underline{a}_2'\underline{a}_2)$

■ Principal Component Analysis:

$$V(\underline{X}) = \overset{p\times p}{\sum}$$

From $\underline{X}$ we came to $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}$ where Y_i's are linear combinations of X_i's.

$V(\underline{Y}) = \mathrm{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_p) = \Lambda$
where $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$

Again, we know that any symmetric matrix can be diagonalized. Similarly, the var-covariance matrix can also be diagonalized.

$A\Sigma A' = \Lambda$

where A is an orthogonal matrix constructed by orthonormal eigenvectors $(\underline{a}_1, \underline{a}_2 \ldots, \underline{a}_p)$ corresponding to the eigenvalues of Σ $(\lambda_1, \lambda_2, \ldots, \lambda_p)$.

Therefore, the total system variability can be expressed as:

$$\begin{aligned} &\text{Total Variance} = \mathrm{Var}(x_i) = \mathrm{tr}(\Sigma) \\ &= \mathrm{tr}(A'\Lambda A) \\ &= \mathrm{tr}(\Lambda AA') \quad [\mathrm{tr}(AB) = \mathrm{tr}(BA)] \\ &= \mathrm{tr}(\Lambda) \\ &= \sum_{i=1}^{p} \lambda_i \end{aligned}$$

Hence we can compute the proportion of system variability explained by each PC as follows:

$$\text{Variance explained by } 1^{\text{st}} \text{ PC} = \frac{\lambda_1}{\sum_{i=1}^{p} \lambda_i}$$

$$\text{Variance explained by } 2^{\text{nd}} \text{ PC} = \frac{\lambda_2}{\sum_{j=1}^{p} \lambda_j}$$

and in general, variance explained by the i-th PC is:

$$\left(\frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \right)$$

Since one of the major objectives of PCA is dimension reduction, then one can reduce the dimension, i.e, he/she may consider $m(\leq p)$ PC's to be retained for future analysis - using the cumulative proportion of total system variability explained by the retained PC 's. clearly the cumulative proportion of system variability explained by the first two PC 's is equal to $\frac{\lambda_1+\lambda_2}{\sum_{i=1}^{p} \lambda_i}$

similarly the cumulative proportion of system variability explained by first m PC 's $\Rightarrow \left(\frac{\lambda_1+\lambda_2+\ldots+\lambda_m}{\sum_{i=1}^{p} \lambda_i} \right)$.

Generally, if more than 80% of the total system variability is explained by the first m PC's, then m number of PCs, i.e., $Y_1, Y_2, \ldots, Y_m$, are retained out of the p PC's derived from the data.

It may also be noted that the satisfaction level of the cumulative proportion of the total system variability explained depends upto the analyzer, i.e. the quantity " 80% " is not a fixed value.

Example:
Consider the following var-cov matrix and find the PC's. Also, explain the proportion of variability explained by each PC and comment on the number of PCs to be retained if someone wishes to explain at least 80% of the total system variability.

$$\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

$\Rightarrow$ By looking at Σ, we can see

$$|\Sigma| = 2(5-4) = 2 \neq 0 \Rightarrow \Sigma \text{ is non-singular}$$

Let $\lambda_1, \lambda_2, \lambda_3$ be the eigenvalues of Σ, i.e., solve $|\Sigma - \lambda I| = 0$.

$$\Sigma - \lambda I = \begin{pmatrix} 1-\lambda & -2 & 0 \\ -2 & 5-\lambda & 0 \\ 0, & 0 & -2-\lambda \end{pmatrix}$$

$\Rightarrow |\Sigma - \lambda I| = 0$
$\Rightarrow (2-\lambda)[(1-\lambda)(5-\lambda) - 4] = 0$
$\Rightarrow (2-\lambda)[\lambda^2 - 6\lambda + 5 - 4] = 0$
$\Rightarrow (\lambda - 2)[\lambda^2 - 6\lambda + 1] = 0$
$\therefore \lambda = 2$
or

$$\lambda = \frac{6 \pm \sqrt{36-4}}{2}$$
$= 3 \pm 2\sqrt{2}$
$\Rightarrow \lambda_1 = 3 + 2\sqrt{2} = 5.828, \lambda_2 = 2, \lambda_3 = 3 - 2\sqrt{2} = 0.172$
and $\lambda_1 > \lambda_2 > \lambda_3$

$$\therefore \Sigma \underline{a}_1 = \lambda_1 \underline{a}_1$$
$$\underline{a}_i = \begin{pmatrix} a_{i1} \\ a_{i2} \\ a_{i3} \end{pmatrix}, \quad i = 1, 2, 3$$

$$\Rightarrow \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix} = \begin{pmatrix} 5.828a_{11} \\ 5.828a_{12} \\ 5.828a_{13} \end{pmatrix}$$
$\Rightarrow -4.828a_{11} - 2a_{12} = 0$
$\Rightarrow 2a_{11} - 0.828a_{12} = 0.$
$\Rightarrow a_{11} + 0.414a_{12} = 0$
$$a_{13} = 0$$
$\therefore$ if $a_{11} = 1 \Rightarrow a_{12} = -2.415$

$$\therefore \underline{a}_1 = \begin{pmatrix} 1 \\ -2.415 \\ 0 \end{pmatrix}; \quad \begin{aligned} \|a_1\| &= \sqrt{1^2 + (-2.415)^2} \\ &= 2.613 (Normed) \end{aligned}$$
$$\underline{e}_1 = \begin{pmatrix} 0.38 \\ -0.92 \\ 0 \end{pmatrix}$$
Hence the first PC is $Y_1 = 0.38X_1 - 0.92X_2$

For, $\lambda_2 = 2$

$$\begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{22} \\ a_{23} \end{pmatrix} = \begin{pmatrix} 2a_{21} \\ 2a_{22} \\ 2a_{23} \end{pmatrix}$$

$$\begin{aligned} &\Rightarrow a_{21} - 2a_{22} = 2a_{21} \\ &\Rightarrow a_{21} + 2a_{22} = 0 \\ &-2a_{21} + 5a_{22} = 2a_{22} \Rightarrow 2a_{21} = 3a_{22} \\ &-2a_{23} = 2a_{23} \end{aligned}$$

one such eigenvector is $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

Do the sames for the following var-cov matrix

$$\begin{pmatrix} 1 & 4 \\ 4 & 100 \end{pmatrix}$$

$$\begin{aligned} &(1-\lambda)(100-\lambda) - 16 = 0 \\ &\Rightarrow \lambda^2 - 101\lambda + 100 - 16 = 0 \\ &\Rightarrow \lambda^2 - 101\lambda + 84 = 0 \\ &\Rightarrow \lambda = \frac{101 \pm \sqrt{10201 - 336}}{2} \\ &\Rightarrow \lambda_1 = 100.16, \quad \lambda_2 = 0.84 \end{aligned}$$

$$\begin{pmatrix} 1 & 4 \\ 4 & 100 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 100.16a_1 \\ 100.16a_2 \end{pmatrix}$$

$$\Rightarrow \begin{matrix} a_1 + 4a_2 = 100.16a_1 \\ 4a_1 + 100a_2 = 100.16a_2 \end{matrix}$$

$$\begin{aligned} &\Rightarrow -99.16a_1 + 4a_2 = 0 \\ &\Rightarrow 4a_1 - 0.16a_2 = 0 \\ &\Rightarrow a_1 - 0.04a_2 = 0 \end{aligned}$$

Let $a_2 = 1$

$a_1 = 0.04$

$\|a\| = \sqrt{1^2 + 0.04^2}$

$\therefore \begin{pmatrix} 0.04 \\ 1 \end{pmatrix} = \underline{e}_1$

for $\lambda_2 = 0.84$

$$\begin{aligned} &b_1 + 4b_2 = 0.84b_1 \\ &4b_1 + 100b_2 = 0.84b_2 \\ &\therefore 0.16b_1 + 0.46b_2 = 0 \\ &\ \& \ 4b_1 + 99.16b_2 = 0 \\ &\Rightarrow b_2 = -0.04b_1 \end{aligned}$$

$\therefore \begin{pmatrix} 1 \\ -0.04 \end{pmatrix} = \underline{e}_2$

$$\boxed{y_1 = 0.04x_1 + 0.999x_2, y_2 = 0.999x_1 - 0.04x_2 \quad PC's}$$

Convert the above covariance matrix into correlation matrix

$$\Sigma = \begin{pmatrix} 1 & 4 \\ 4 & 100 \end{pmatrix} \Longrightarrow R = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$$

$$|R| = 1 - 0.4^2$$

$$= (1+0.4)(1-0.4) = \frac{1.4}{\lambda_1} \times \frac{0.6}{\lambda_2}$$

$$\therefore R\underline{a}_1 = \lambda_1 \underline{a}_1$$

$$\Rightarrow \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1.4a_1 \\ 1.4a_2 \end{pmatrix}$$

$$\Rightarrow a_1 + 0.4a_2 = 1.4a_1$$

$$0.4a_1 + a_2 = 1.4a_2$$

$$\Rightarrow 0.4a_2 = 0.4a_1 \Rightarrow a_1 = a_2$$

For a normal vector

$$a_1^2 + a_2^2 = 1$$

$$\Rightarrow a_1 = 0.707 = a_2$$

$$Y_1 = 0.707x_1 + 0.707x_2$$

$$R\underline{a}_2 = \lambda_2 \underline{a}_2$$

$$\Rightarrow \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0.6a_1 \\ 0.6a_2 \end{pmatrix}$$

$$\Rightarrow a_1 + 0.4a_2 = 0.6a_1$$

$$\Rightarrow a_1 + a_2 = 0$$

$$\text{Also} \quad a_1^2 + a_2^2 = 1$$

$$\Rightarrow a_1 = 0.707$$

$$\therefore a_2 = -0.707$$

$$\therefore Y_2 = 0.707x_1 - 0.707x_2$$

Correlation Matrix

Covariance Matrix of standardized variables.

Therefore, we see that principal components obtained from the covariance matrix and correlation matrix of the same data are not the same. It might also be noted that the correlation matrix is the covariance matrix of the standardized variables. For the above data, in the case of the covariance matrix, the first principal component explains $\frac{100.16}{101} = 0.992$ or 99.2% of the total system variability, i.e., it explains almost the entire system variability. This may also be pointed out that in the first PCs obtained from the Covariance matrix, the contribution of each variable differs significantly. For example, the first PC is dominated by the 2nd variable (X_2), and the 2nd PC is dominated by the 1st variable(X_1). However, if we look at the PC's obtained from the correlation matrix (X_1), the first PC explains $\frac{1.4}{2} = 0.7$ or 70% of the total system variability. More interestingly, in the PC's obtained from the correlation matrix, both variables X_1 and

X_2 contribute equally. Therefore, it may be concluded that in PC's obtained from the covariance matrix, variables with larger variance dominates, and in P'Cs obtained from R, dominance of the variables with larger numerical variance gets neutralized. Hence, finding out PC's from the correlation matrix is preferable. Standardization of variables is also recommended to deal with the units of different variables..

$\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ 0 & \rho\sigma^2 & \sigma^2 \end{pmatrix}$

be the variance- covariance matrix $-\frac{1}{\sqrt{2}} < \rho < \frac{1}{\sqrt{2}}$

Clearly, the correlation matrix

$$R = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix}$$

To find the PC's of R, the following equation is to be solved:

$$\begin{aligned}
&|R - \lambda I| = 0 \\
\Rightarrow &\begin{vmatrix} 1-\lambda & \rho & 0 \\ \rho & 1-\lambda & \rho \\ 0 & \rho & 1-\lambda \end{vmatrix} = 0 \\
\Rightarrow &(1-\lambda)\left[(1-\lambda)^2 - \rho^2\right] - \rho^2(1-\lambda) = 0 \\
\Rightarrow &(1-\lambda)\left[(1-\lambda)^2 - 2\rho^2\right] = 0 \\
\therefore &\lambda = 1 \text{ or } 1-\lambda = \pm\sqrt{2}\rho \\
\Rightarrow &\lambda = 1 \text{ or } \lambda = 1 \pm \sqrt{2}\rho
\end{aligned}$$

Therefore $\lambda_1 = 1$, $\lambda_2 = 1 + \sqrt{2}\rho$, $\lambda_3 = 1 - \sqrt{2}\rho$ are the eigenvalues.

let us try to find eigenvector $\underline{a}, \underline{b}, \underline{c}$ corresponding to $\lambda_1, \lambda_2, \lambda_3$

For, $\lambda_1 = 1$

$$\begin{aligned}
&\begin{pmatrix} 0 & \rho & 0 \\ \rho & 0 & \rho \\ 0 & \rho & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\
&\Rightarrow \rho a_2 = 0 \Rightarrow a_2 = 0 \\
&\Rightarrow \rho a_1 + \rho a_3 = 0 \Rightarrow a_1 + a_3 = 0 \Rightarrow a_1 = -a_3 \;\;\&\;\; a_1^2 + a_2^2 + a_3^2 = 0
\end{aligned}$$

$\therefore$ The normed eigenvector

$$e_1 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix},$$

For $\quad \lambda_2 = 1 + \sqrt{2}\rho$

$$\begin{pmatrix} -\sqrt{2}\rho & \rho & 0 \\ \rho & -\sqrt{2}\rho & \rho \\ 0 & \rho & -\sqrt{2}\rho \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow -\sqrt{2}\rho b_1 + \rho b_2 = 0 \Rightarrow b_2 = \sqrt{2} b_1$$

$$\Rightarrow \rho b_1 - \sqrt{2}\rho b_2 + \rho b_3 = 0,$$

$$\Rightarrow \rho b_2 = \sqrt{2}\rho b_3$$

$$\Rightarrow b_2 = \sqrt{2} b_3$$

$$\begin{aligned} \therefore b_1^2 + b_2^2 + b_3^2 &= 1 \\ \Rightarrow \frac{b_2^2}{2} + b_2^2 + \frac{b_2^2}{2} &= 1 \\ \Rightarrow b_2 &= \frac{1}{\sqrt{2}} \\ \therefore b_1 = \frac{1}{2}, \quad b_3 &= \frac{1}{2}. \\ \therefore \underline{e}_2 &= \begin{pmatrix} 1/2 \\ 1/\sqrt{2} \\ 1/2 \end{pmatrix} \end{aligned}$$

For $\lambda_3 = 1 - \sqrt{2}\rho$

$$\begin{pmatrix} \sqrt{2}\rho & \rho & 0 \\ \rho & \sqrt{2}\rho & \rho \\ 0 & \rho & \sqrt{2}\rho \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \sqrt{2}\rho c_1 + \rho c_2 = 0 \qquad \Rightarrow -\sqrt{2} c_1 = c_2$$

$$\rho c_1 + \sqrt{2}\rho c_2 + \rho c_3 = 0 \qquad \Rightarrow c_1 + c_3 = -\sqrt{2} c_2$$

$$\rho c_2 + \sqrt{2}\rho c_3 = 0 \qquad \Rightarrow c_2 = -\sqrt{2} c_3$$

and $\quad c_1^2 + c_2^2 + c_3^2 = 1$

$$\therefore \underline{e}_3 = \begin{pmatrix} -1/2 \\ 1/\sqrt{2} \\ 1/2 \end{pmatrix}$$

$\therefore \underline{e}_1, \underline{e}_2, \underline{e}_3$ are also mutually orthogonal.

Therefore the corresponding PC's are with original variables x_1, x_2 and x_3

$$
\begin{aligned}
Y_1 &= \frac{1}{\sqrt{2}}X_1 - \frac{1}{\sqrt{2}}X_3 \\
Y_2 &= \frac{1}{2}X_1 + \frac{1}{\sqrt{2}}X_2 + \frac{1}{2}X_3 \\
Y_3 &= -\frac{1}{2}X_1 + \frac{1}{\sqrt{2}}X_2 - \frac{1}{2}X_3
\end{aligned}
$$

Note that if $\rho > 0$, then the 1st PC would be Y_2. Hence the revised PC:

$$Y_{(1)} = \frac{1}{2}X_1 + \frac{1}{\sqrt{2}}X_2 + \frac{1}{2}X_3$$

Also, if $\rho < 0$, then the 1st PC would be Y_3. Hence the revised PC:

$$Y_{(1)} = -\frac{1}{2}X_1 + \frac{1}{\sqrt{2}}X_2 - \frac{1}{2}X_3$$

The above dispersion matrix can also be for a p-variate random vector, and in that case, we have the following form of the dispersion matrix:

$$\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{pmatrix}$$

If the above covariance matrix is converted into the correlation matrix, we would get the following:

$$
\begin{aligned}
P &= \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix} \\
|P| &= (1-\rho)^{p-1}[1+(p-1)\rho]
\end{aligned}
$$

To find PC's when we have the above correlation matrix,
find eigenvalues and eigenvectors.

$$
\begin{aligned}
&\lambda_1 = 1-p, \lambda_2 = 1-p, \quad \ldots, \quad \lambda_{p-1} = 1-p \\
&\lambda_p = 1+(p-1)\rho. \\
&R\underline{x} = \lambda_p \underline{x} \\
&\Rightarrow (R-\lambda_p I)\,\underline{x} = \underline{0}
\end{aligned}
$$

$$\Rightarrow \begin{pmatrix} 1-1-(p-1)\rho & \rho & \rho & \cdots & \rho \\ \rho & 1-1-(p-1)\rho & \rho & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho & \rho & \rho & \cdots & 1-1-(p-1)\rho \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} (1-p)\rho & \rho & \rho & \cdots & \rho \\ \rho & (1-p)\rho & \rho & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho & \rho & \rho & \cdots & (1-p)\rho \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}$$

For case of computation let p=3

$$\text{Then} \quad \begin{pmatrix} -2\rho & \rho & \rho \\ \rho & -2\rho & \rho \\ \rho & \rho & -2\rho \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{aligned} \Rightarrow \quad & x_2 + x_3 = 2x_1 \\ & x_1 + x_3 = 2x_2 \\ & x_1 + x_2 = 2x_3 \end{aligned}$$

$$\therefore x_1 = x_2 = x_3$$

Therefore, the required eigenvector is $k\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}'$. For $k = 1$, the normalized eigenvector is $\frac{1}{\sqrt{3}}(1,1,1)$.
Hence, generalizing this, we can say that the normalized eigenvectors for the eigenvalue $\lambda = 1 + (p-1)\rho$ are $\underline{e_1} = \left(\frac{1}{\sqrt{p}}, \frac{1}{\sqrt{p}}, \ldots, \frac{1}{\sqrt{p}}\right)$.

Now, for the eigenvalue $\lambda = 1 - \rho$, the eigenvector can be obtained from the equation,

$$\begin{pmatrix} 1-1+\rho & \rho & \cdots & \rho \\ \rho & 1-1+\rho & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1-1+\rho \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\begin{aligned} &\Rightarrow \rho x_1 + \rho x_2 + \cdots + \rho x_p = 0 \\ &\Rightarrow x_1 + x_2 + \ldots + x_p = 0. \end{aligned}$$

It may also be noted that $\lambda = 1 - p$ has multiplicity $(p-1)$ i.e. we need to find $(p-1)$ such orthonormal vectors satisfying the condition that the sum of all components of each vector is zero.
Hence, the following set of vectors would be a good choice:

$$\begin{aligned} \underline{e_2}' &= \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \ldots, 0\right)' \\ \underline{e_3}' &= \left(\frac{1}{\sqrt{2\times 3}}, \frac{1}{\sqrt{2\times 3}}, \frac{-2}{\sqrt{2\times 3}}, 0, \ldots, 0\right)' \\ &\vdots \\ \underline{e_i}' &= \left(\frac{1}{\sqrt{i(i-1)}}, \cdots, \frac{-(i-1)}{\sqrt{i(i-1)}}, 0, \ldots, 0\right)' \\ &\vdots \\ \underline{e_p}' &= \left(\frac{1}{\sqrt{p(p-1)}}, \cdots, \frac{-(p-1)}{\sqrt{p(p-1)}}\right)' \end{aligned}$$

Elmhirst transformation

Total system variability $=\operatorname{tr}(R) = P$. If $\rho > 0$, then $\lambda = 1 + (p-1)\rho$ is the largest eigenvalue. Hence, the corresponding orthonormal eigenvector $\underline{e_1}$ could serve as the vector of 1st PC:

$$Y_1 = \underline{e_1}'\underline{X} = \frac{1}{\sqrt{p}}X_1 + \frac{1}{\sqrt{p}}X_2 + \cdots + \frac{1}{\sqrt{p}}X_p.$$

The proportion of system variability explained by the 1st PC is $= \frac{1+(p-1)\rho}{p} = \rho + \left(\frac{1-\rho}{p}\right)$ Write down the other PC's for $\underline{e_2}', \underline{e_3}', \ldots, \underline{e_1}'$. Find system. Variability for the PC's. By borrowing the idea from Elmhirst transformations, we define the following PC's:

$$\begin{aligned}
Y_2 &= \frac{1}{\sqrt{2}}X_1 - \frac{1}{\sqrt{2}}X_2 \\
Y_3 &= \frac{1}{\sqrt{2\times 3}}X_1 + \frac{1}{\sqrt{2\times 3}}X_2 - \frac{2}{\sqrt{2\times 3}}X_3 \\
&\vdots \\
Y_p &= \frac{1}{\sqrt{p(p-1)}}X_1 + \frac{1}{\sqrt{p(p-1)}}X_2 + \cdots - \frac{(p-1)}{\sqrt{p(p-1)}}X_p
\end{aligned}$$

The proportion of system variance explained by rest of these $(p-1)$ PC's individually is $\left[\frac{1-p}{p}\right]$, which is less than the proportion explained by its PC by the amount p.

Let the Variance-Covariance Matrix be

$$\Sigma = \begin{pmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

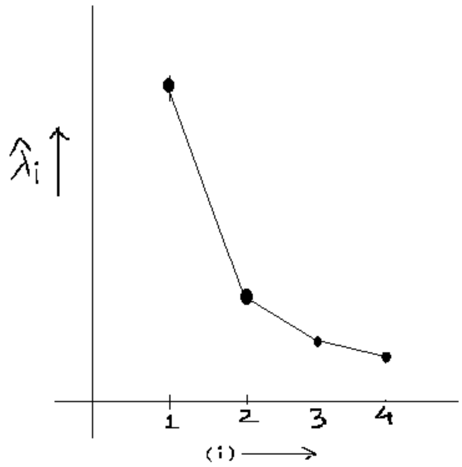
Find the PCs and interpret the results.

■ **The Number of PC's**

Though we compute p PC's $Y_1, Y_2, \ldots, Y_p$ corresponding to the p-component vector of variables $X_{p*1} = (X_1, X_2, \ldots, X_p)'$. But it is always a prime issue to decide on how many PC's should be retained to get the advantage of dimension reduction. Although the cumulative proportion of system variability explained by the first few PC's gives us a way to decide on the number of PC's to be retained. Generally, if the first K PC's explain more than 80% of the total system variability, then out of p PC's, $k(\leq p)$ PC's are retained. However, there is a graphical way to determine the number of PCs to be retained, known as Scree Plot.

A Scree plot of $\hat{\lambda}_{(i)}$ is a plot of the i^{th} sample eigenvalue vs. i. To determine the appropriate number of PC's, we look for an elbow (bend) in the scree plot.

The number of components (PC's) is taken to be the point at which the remaining eigenvalues are relatively small and all about the same size as the first.



An example of a scree plot with 4 PCs. In this figure, it may be noted that an elbow occurs near the 2 i.e., the eigenvalues after $\widehat{\lambda}_{(2)}$ are relatively small and about the same size. In this case, it appears, without any other evidence, that 2 sample PC's effectively summarize the total sample variance.

In reality, we do not have the population var-cor matrix; since we have samples, we have to work on the sample covariance matrix. Therefore, we generally get sample eigenvalues and corresponding eigenvectors from the sample covariance matrix. Obviously, as the sample changes, so do the sample eigenvalues and sample eigenvectors.

Hence, we need to know the large sample properties of $\hat{\lambda}_{(i)}$ and $\underline{\hat{e}}_{(i)}$ (sample eigenvector). Currently available results concerning large sample intervals for $\hat{\lambda}_{(i)}$ and $\underline{\hat{e}}_{(i)}$ assume that the observations $\underline{x_1}, \underline{x_2}, \ldots, \underline{x_n}$ are a random sample from a normal population. It must also be assumed that the (unknown) values of Σ are distinct and positive so that

$$\lambda_{(1)} > \lambda_{(2)} > \cdots > \lambda_{(p)} > 0 \quad \text{and} \quad (t\text{-rest} \quad \sigma_1^2 \neq \sigma_2^2) \text{ (unknown)}.$$

The one exception is the case where the number of equal eigenvalues is known. Usually, the conclusions for distinct eigenvalues are applied unless there is a strong reason to believe that Σ has a special structure that yields equal eigenvalues. Even when the normal assumption is violated, the confidence intervals obtained in this manner still provide some indication of the uncertainty in $\hat{\lambda}_i$ and $\underline{\hat{e}}_i$.

Anderson and Girshick have established the following sample distribution theory for the eigenvalues $\underline{\hat{\lambda}} = \left(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_p\right)$ and eigenvectors $\underline{\hat{e}}_1, \underline{\hat{e}}_2, \ldots, \underline{\hat{e}}_p$ of S.

(1) Let Λ be the diagonal matrix of eigenvalues $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_p$ of Σ. Then $\sqrt{n}(\underline{\hat{\lambda}} - \underline{\lambda})$ is approximately $N_p\left(\underset{\sim}{0}, 2\lambda^2\right)$.

(2) Let $\underline{E_i} = \lambda_i \sum_{\substack{k=1 \\ k \neq i}}^{p} \frac{\lambda_k}{(\lambda_k - \lambda_i)^2} \underline{e_k}\ \underline{e_k}'$. Then $\sqrt{n}(\underline{\hat{e_i}} - \underline{\hat{e_i}})$ is approximately $N_p(\underline{0}, E)$.

(3) Each $\hat{\lambda}_i$ is distributed independently of the elements of the associated $\underline{\hat{e_i}}$.