KURDISTAN REGIONAL GOVERNMENT

MINISTRY OF HIGHER EDUCATION & SCIENTIFIC RESEARCH

SALAHADDINUNIVERSITY – HAWLER (SUH)

COLLEGE OF ADM. & ECO.

STATISTICS & INFORMATICS

DEPARTMENT

Syllabus of

MULTIVARIATE STATISTICAL ANALYSIS

Diploma Students

by

Dr. Delshad Shaker Ismael Botani

E-mail: delshd.botani@su.edu.krd

2023 – 2024

Chapter One

Principles of Multivariate Statistical Analysis (MSA)

Principles of Multivariate Analysis

Why Multivariate Analysis:

Def.: MSA is a set of methods that deal with the simultaneous analysis of multiple outcome or response variables (Dependent Variables – DVs).

Multivariate analysis consists of a collection of methods that can be used when several measurements are made on each individual or object in one or more samples. We refer to the measurements as variables and to the individuals or objects as units (research units, sampling units, or experimental units) or observations. The below table contains some examples of multivariate analysis.

Examples of Multivariate Data

#	Units (individuals or objects)	Variables (measurements)
1	Students	Several exam scores in a single course
2	Students	Grades in mathematics, history, music, art, physics
3	People	Height, weight, percentage of body fat, resting heart rate
4	Skulls	Length, width, cranial capacity
5	Companies	Expenditures for advertising, labor, raw materials
6	Manufactured items	Various measurements to check on compliance with specifications
7	Applicants for bank loans	Income, education level, length of residence, savings account, current debt load
8	Segments of literature	Sentence length, frequency of usage of certain words and of style characteristics
9	Human hairs	Composition of various elements
10	Birds	Lengths of various bones

Multivariate analysis is concerned generally with two areas, **descriptive and inferential statistics.** In the descriptive field, we often obtain optimal linear combinations of variables. In the inferential area, many multivariate techniques are extensions of univariate procedures (Univariate Statistical Analysis – USA).

Why using MSA instead of USA

Multivariate analyses are conducted instead of univariate analyses for the following reasons:

1. With more than one DVs (say p), the use of p separate univariate tests inflates the **Type I error rate**, whereas a pertinent multivariate test preserves the significance level ($p>1$).
2. Univariate tests, no matter how many in number, ignore the interrelationships possible among the DVs, unlike multivariate analyses, and hence potentially waste important information contained in the available sample of data.

3. In many cases, the multivariate test is more powerful than a corresponding univariate test, because the former utilizes the information mentioned in the previous point 2. In such cases, we tend to trust MSA more when its results are at variance with those of USA.
4. Many multivariate tests involving means have as a by-product the construction of a linear combination of variables, which provides further information (in case of a significant outcome) about how the variables unite to reject the hypothesis.

Data Organization

We will use the notation to indicate the particular value of the k^{th} variable that is observed on the j^{th} item, or trial. That is,

$$x_{jk} = \text{measurement of the } k^{th} \text{ variable on the } j^{th} \text{ item.}$$

Consequently, n measurements on p variables can be displayed as follows:

	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}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
Item j:	x_{j1}	x_{j2}	$\cdots$	x_{jk}	$\cdots$	x_{jp}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
Item n:	x_{n1}	x_{n2}	$\cdots$	x_{nk}	$\cdots$	x_{np}

where,

p: number of numeric response variables (or characters) of being measured;

n: number of items (individuals, or experiment units) on which variables are being measured;

Or we can display these data as a rectangular array (Matrix), called **X**, of n rows and p columns:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$