Introduction

Zhixiang ZHANG

University of Macau

zhixzhang@um.edu.mo

January 6, 2025

Table of Contents

1. Multivariate techniques
2. Organization of data
3. Distance

What is Multivariate Data Analysis

- The expression "multivariate analysis" is used to describe analyses of data that are multivariate in the sense that numerous variables (features) are obtained for each individual or unit studied.
- In a typical survey 30 to 100 questions are asked of each respondent. In describing the financial status of a company, an investor may wish to examine five to ten measures of the company's performance.

Goals of Multivariate Statistical Techniques

The main aim of multivariate data analysis is to understand relationship between several variables and their relevance to the problem being studied.

Multivariate methods are well-suited for achieving the objectives of scientific investigations, which include the following:

- Estimation and hypothesis testing
- Prediction
- Data reduction
- Grouping
- Investigation of the dependence among variables

Estimation and Hypothesis testing

- Estimating average housing characteristics in different neighborhoods by calculating the average values of multiple characteristics, such as price, size, age, and distance to shopping centers for each area. These multivariate means provide a snapshot of what constitutes a 'typical' property in each neighborhood.
- Experimental data on several variables were used to see whether the nature of the instructions makes any difference in perceived risks, as quantified by test scores.

Confidence region?

rent μ_1

home value μ_2

we can find $\begin{cases} [a,b]: P(\mu_1 \in [a,b]) = 0.95 = p_1 \\ [c,d]: P(\mu_2 \in [c,d]) = 0.95 = p_2 \end{cases}$

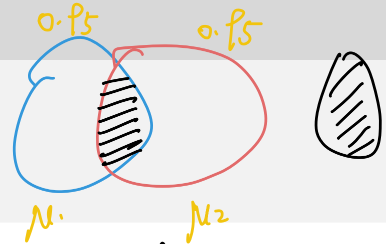
$[\bar{x}_1 - z_{\frac{\alpha}{2}} sd(\bar{x}_1), \bar{x}_1 + z_{\frac{\alpha}{2}} sd(\bar{x}_1)]$

or. $P\left(\begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix} \in R\right) = 0.95$

$\Updownarrow$?

intersect

$P(\mu_1 \in [a,b], \mu_2 \in [c,d]) = 0.95$



≤ 0.95

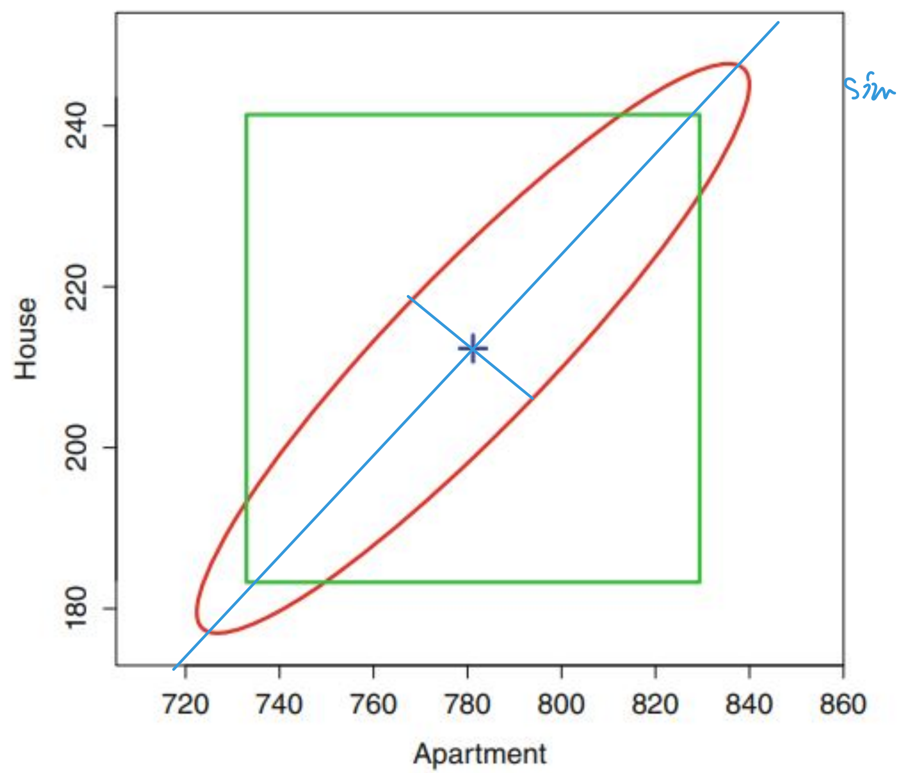
$\min[p_1, p_2] = 0.95$

State	Median apartment rent in \$	Median home value in \$1000
AK	949	237.8
AL	631	121.5
AR	606	105.7
⋮	⋮	⋮
WV	528	95.9
WY	636	188.2

Figure 1: Costs of living in each of the 50 states

- We can create 95 % confidence intervals for either the mean rent or the mean housing prices.
- How to create a 95% joint (simultaneous) confidence region of both the mean rent and housing prices?

Confidence region?



positive direction for relation.

Similarly to PCA.

orthogonal to represent variance.

Figure 2: The green box is formed from the marginal 95% confidence intervals. The sample averages are indicated in the center. The red ellipsoid is a joint 95% confidence region for housing prices and monthly apartment rents.

Prediction

- By taking multiple factors such as location, property characteristics, and market trends into consideration, one can predict a reasonable housing price.
- Credit card companies analyze multiple customer variables, including payment history, spending patterns, and financial background, to create a single risk score that helps predict the likelihood of payment defaults.

Data reduction

less variables to represent original data.

for easier intepretation

dimension reduction

- Athletics records from many nations were used to develop an index of performance for athletes.
- In personality research, factor analysis demonstrates data reduction by taking a large set of observable variables (questionnaire responses) and reducing them to core underlying factors.

personality is a very complex thing

Data Reduction

The Program for International Student Assessment (PISA) is a triennial survey of academic achievements of 15-year-old students in each of 70 different countries.

		Reading subscales						
Nation	Overall reading	Access retrieve	Integrate interpret	Reflect eval.	Continuous	Noncontin.	Math	Science
China:								
Shanghai	556	549	558	557	564	539	600	575
Korea	539	542	541	542	538	542	546	538
Finland	536	532	538	536	535	535	541	554
⋮		⋮		⋮		⋮		⋮
Peru	370	364	371	368	374	356	365	369
Azerbaijan	362	361	373	335	362	351	431	373
Kyrgyzstan	314	299	327	300	319	293	331	330

Source: OECD PISA 2009 database

Figure 3: Reading and other academic scores from OECD nations

How much is gained by providing the different subscales for reading?

Data Reduction

Figure 4: Matrix scatterplot of reading and academic scores of 15-year-old students in 2009 in OECT nations

Data Reduction

- All of the measurements are highly correlated with each other. Any one academic measurement in the data could stand as a good representation for most of the others.
- There is a large amount of data simplification that can be performed for this data without significant loss of information.

Grouping

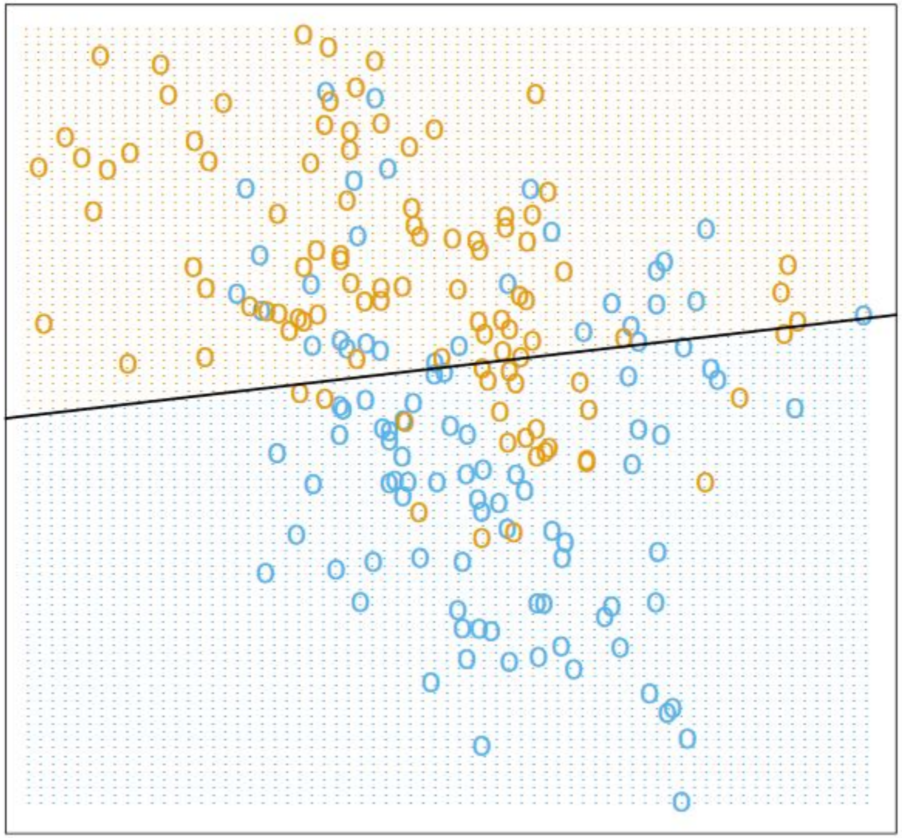
- Financial institutions and credit card companies can examine various variables such as transaction amount, location, and customer behavior patterns to prevent fraudulent transactions.
- Companies often analyze customer data to segment their target market into distinct groups based on demographics, preferences, or buying behavior

Grouping

A **classification** example in two dimensions. Blue and orange represent two different classes. labelled : supervised

i	X_i	Y_i	e.g.
1	x_1	y_1	1
2	x_2	y_2	2
$\vdots$	$\vdots$	$\vdots$	
n	x_n	y_n	1
Given	X_{n+1}	Y_{n+1}	

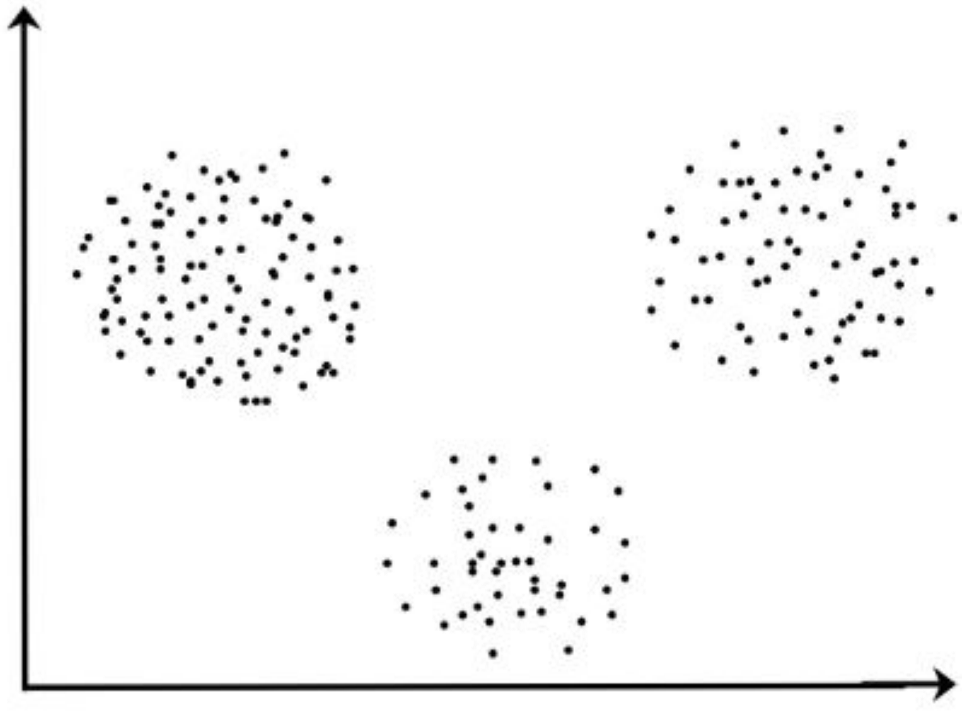
classification :



if there are 4 variables or more
it is impossible to draw the graph,
then multivariate analysis is useful here.

Grouping

A **clustering** example in two dimensions.



What if the data have more variables (features)?

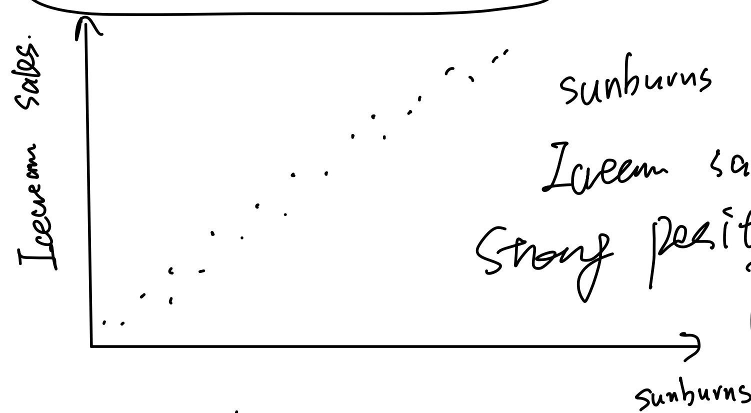
Investigation of the dependence among variables

might have relations.

- Analyze the dependence between sleeping patterns and health status.
- Association of a person's willingness to take risks and their social or economic traits.

Challenges of Multivariate Analysis

- The need to understand the relationships between many variables makes multivariate analysis an inherently difficult subject. Often, the human mind is overwhelmed by the sheer bulk of the data.
- Additionally, more mathematics is required to derive multivariate statistical techniques for making inferences than in a univariate setting.



sunburns doesn't cause Icecream sales, but have Strong positive correlation.

have "correlation" but not deterministic.

one aspect.

another aspect

deterministic: "Causality"

Some crucial tools

main tool we use is correlation.

e.g. confidence interval
type I II.

Tools for Multivariate Data Analysis in this course

- Multivariate inference Statistical inference for multivariate.
- Linear combinations of variables
- Eigenstructure of a matrix linear algebra.
- A little large-sample theory central limit theorem.

Multivariate Analysis Techniques

Multivariate Analysis Techniques

- Hotelling's T^2
- MANOVA
- Multiple Linear Regression (MATH2009)
- Principal Component Analysis PCA
- Factor Analysis FA
- Discriminant (& Classification) Analysis
- Cluster Analysis
- Canonical Correlation Analysis time permits.

Multivariate Analysis Techniques

- One technique can be utilized to achieve more than one aforementioned goal.
- e.g. Discriminant analysis is related to both grouping and prediction.
- The main aim of this course is to grasp these techniques.

Two general classes of MVA techniques

These techniques can be divided into two classes

feature and response.

- Supervised techniques: One (or more) variables are dependent variables, to be explained or predicted by other variables e.g. Multiple linear regression, Discriminant analysis

  label and features

- Unsupervised techniques: Look at the relationships among variables, e.g. principal component analysis, cluster analysis, factor analysis clustering: unsupervised only X for features.

Table of Contents

1 Multivariate techniques

2 Organization of data

3 Distance

Multivariate data representation

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

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

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

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

Matrix form

Or we can display these data as a rectangular matrix, called $\mathbf{X}$, of n rows and p columns:

tall matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix}$$

$n = 10, 20, \cdots, 100$

$p = 10$

Summary Statistics

The sample mean can be computed from the n measurements on each of the p variables, so that, in general, there will be p sample means:

$$\bar{x}_k = \frac{1}{n}\sum_{j=1}^{n} x_{jk} \quad k = 1, 2, \ldots, p$$

A measure of spread is provided by the sample variance, defined for n measurements on the k-th variable as

$$s_k^2 = \frac{1}{n-1}\sum_{j=1}^{n} \left(x_{jk} - \bar{x}_k\right)^2 \quad k = 1, 2, \ldots, p$$

Summary Statistics

- The sample covariance

$$s_{ik} = \frac{1}{n-1}\sum_{j=1}^{n}\left(x_{ji}-\bar{x}_i\right)\left(x_{jk}-\bar{x}_k\right) \quad i = 1, 2, \ldots, p, k = 1, 2, \ldots, p$$

  measures the association between the i th and k th variables. We note that the covariance reduces to the sample variance when $i = k$. Moreover, $s_{ik} = s_{ki}$ for all i and k.
- The sample correlation coefficient for the i th and k th variables is defined as

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}} = \frac{\sum_{j=1}^{n}\left(x_{ji}-\bar{x}_i\right)\left(x_{jk}-\bar{x}_k\right)}{\sqrt{\sum_{j=1}^{n}\left(x_{ji}-\bar{x}_i\right)^2}\sqrt{\sum_{j=1}^{n}\left(x_{jk}-\bar{x}_k\right)^2}}$$

  for $i = 1, 2, \ldots, p$ and $k = 1, 2, \ldots, p$. Note $r_{ik} = r_{ki}$ for all i and k. This measure of the linear association between two variables does not depend on the units of measurement.

Matrix representation

Gather a sample: $\mathbf{x}_1, .., \mathbf{x}_n$. Make these the rows of a data matrix $\mathbf{X}$:

$$\begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1p} \end{bmatrix} \quad \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2p} \end{bmatrix}$$

$$\mathbf{X}_{n\times p} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} := \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

From this we can easily represent the sample mean vector and sample covariance matrix: define $\mathbf{1}_n = (1, \ldots, 1)^T : n \times 1$, then

$$\bar{\mathbf{x}}_{p\times 1} = \frac{1}{n}\sum_i \mathbf{x}_i = \frac{1}{n}\mathbf{X}^T\mathbf{1}_n = \frac{1}{n}\underbrace{\begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix}}_{p\times n}\underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{n\times 1} = \frac{1}{n}X^T\underline{1}$$

$$\mathbf{S}_{p\times p} = \frac{1}{n-1}\sum_i (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = \frac{1}{n-1}\left(\sum_i \mathbf{x}_i\mathbf{x}_i^\top - n\bar{\mathbf{x}}\bar{\mathbf{x}}^\top\right)$$

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

$\mathbf{R}_{p\times p} = \mathbf{D}_s^{-1/2}\mathbf{S}\mathbf{D}_s^{-1/2}$. where $\mathbf{D}_s$ is the diagonal matrix of $\mathbf{S}$.

Table of Contents

1 Multivariate techniques

2 Organization of data

3 Distance

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ & & \vdots & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}$$

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})^T$$

$$= \frac{1}{n-1}\left(\sum x_i x_i^T - \sum x_i \bar{x}^T - \sum \bar{x} x_i^T + \sum_{i=1}^{n} \bar{x}\bar{x}^T\right)$$

$\sum x_i = n\bar{x}$

$$= \frac{1}{n-1}\left(\sum x_i x_i^T - n\bar{x}\bar{x}^T - n\bar{x}\bar{x}^T + n\bar{x}\bar{x}^T\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i x_i^T - n\bar{x}\bar{x}^T\right)$$

verified

$$\frac{1}{n-1}X^T\left(I - \frac{1}{n}11^T\right)X$$

$$= \frac{1}{n-1}X^TX - \frac{1}{n-1}\underbrace{\frac{1}{n}X^T1}_{=\bar{x}}\underbrace{1^TX}_{=n\bar{x}^T}$$

$$= \frac{1}{n-1}(x_1, x_2 \cdots x_n)\begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix} - \frac{n}{n-1}\bar{x}\bar{x}^T$$

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}$$

$C = I_n - \frac{1}{n} 1 1^T$

the sample mean of each column of CX is zero!

$$CX = \begin{pmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \cdots & x_{1p}-\bar{x}_p \\ x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & & x_{2p}-\bar{x}_p \\ \vdots & \vdots & & \vdots \\ x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & & x_{np}-\bar{x}_p \end{pmatrix}$$

$C = I_n - \frac{1}{n} 1 1^T$

$$S_{p\times p} = \left(X^T \right)\left(I_n - \tfrac{1}{n}11^T\right)\left(X \right)$$

$$\left(I_n - \tfrac{1}{n}11^T\right) = \left(I_n - \tfrac{1}{n}11^T\right)\left(I_n - \tfrac{1}{n}11^T\right)$$

$$X - \frac{1}{n}\begin{pmatrix} | \\ 1 \\ | \end{pmatrix}\underbrace{1^T X_{n\times p}}_{1\times p} = X - \frac{1}{n}\begin{pmatrix} | \\ 1 \\ | \end{pmatrix}(n\bar{x}_1, n\bar{x}_2, \ldots, n\bar{x}_p) = X - \begin{pmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_p \\ \bar{x}_1 & \bar{x}_2 & & \vdots \\ \vdots & \vdots & & \\ \bar{x}_1 & \bar{x}_2 & & \bar{x}_p \end{pmatrix}$$

$$[1\ 1\ \cdots\ 1]\begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} = [x_1' \ \cdots\ x_p'] = [n\bar{x}_1 \ \cdots\ n\bar{x}_p]$$

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}$$

$C = I_n - \frac{1}{n} 1 1^T \quad C^2 = C$

the sample mean of each column of CX is zero!

$$CX = \begin{pmatrix} (x_1-\bar{x})^T \\ (x_2-\bar{x})^T \\ \vdots \\ (x_n-\bar{x})^T \end{pmatrix}$$

$$S = \frac{1}{n-1}\sum (x_i-\bar{x})(x_i-\bar{x})^T$$

$$= \frac{1}{n-1}(x_1-\bar{x}, \ldots, x_n-\bar{x})\begin{pmatrix} (x_1-\bar{x})^T \\ (x_2-\bar{x})^T \\ \vdots \\ (x_n-\bar{x})^T \end{pmatrix}$$

$$= \frac{1}{n-1} X^T \underbrace{C^T C}_{C} X$$

$$X^T \left(I_n - \tfrac{1}{n}11^T\right) X$$

$$\left(I_n - \tfrac{1}{n}11^T\right) = \left(I_n - \tfrac{1}{n}11^T\right)\left(I_n - \tfrac{1}{n}11^T\right)$$

$x_1 \ldots x_p$

$$\begin{pmatrix} a_1 & & 0 \\ & a_2 & \\ 0 & & a_n \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} & \\ b_{21} & b_{22} & \\ & & \ddots \end{pmatrix}\begin{pmatrix} c_1 & & \\ & \ddots & \\ & & c_n \end{pmatrix} \qquad C = I_n - \tfrac{1}{n}11^T \quad C^2 = C$$

$$= \begin{pmatrix} a_1 b_{11} & a_1 b_{12} & \cdots \\ a_2 b_{21} & a_2 b_{22} & \end{pmatrix}$$

the sample mean of each column of CX is zero!

$S \qquad r_{ij} = \dfrac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}$

$$\begin{pmatrix} r_{11} & r_{12} & & r_{1p} \\ r_{21} & r_{22} & & \\ & & \ddots & \\ r_{p1} & & & r_{pp} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{s_{11}}} & & \\ & \frac{1}{\sqrt{s_{22}}} & \\ & & \ddots \\ & & & \frac{1}{\sqrt{s_{pp}}} \end{pmatrix}\begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & & \\ & & \ddots & \\ s_{p1} & & & s_{pp} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{s_{11}}} & & \\ & \frac{1}{\sqrt{s_{22}}} & \\ & & \ddots \\ & & & \frac{1}{\sqrt{s_{pp}}} \end{pmatrix}$$

Euclidean distance

- Most multivariate techniques are based upon the simple concept of distance.
- Straight-line, or Euclidean, distance should be familiar. If we consider the point $P = (x_1, x_2)$ in the plane, the straight-line distance, $d(O, P)$, from P to the origin $O = (0, 0)$ is

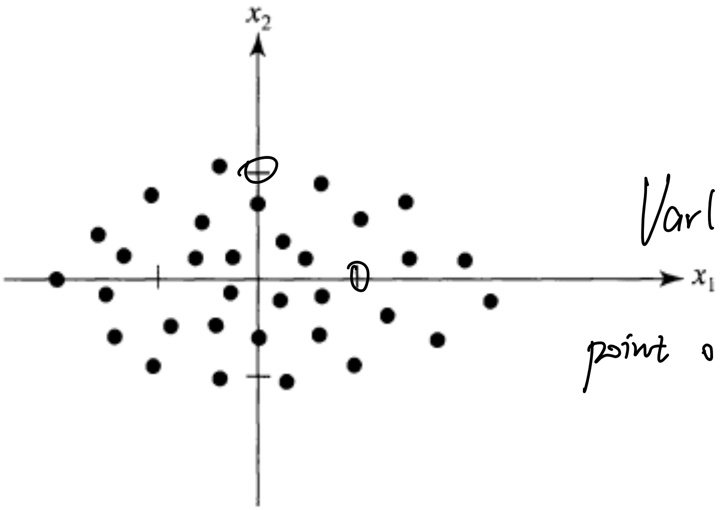
$$d(O, P) = \sqrt{x_1^2 + x_2^2}.$$

- In general, the straight-line distance between two arbitrary points P and Q with coordinates $P = (x_1, x_2, \dots, x_p)$ and $Q = (y_1, y_2, \dots, y_p)$ is given by

$$d(P, Q) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2}.$$

Statistical distance

- When the fluctuation of measurements of two variables is of different scale, it is desirable to weight the variable with greater variability less heavily than the other.
- To illustrate, suppose we have n pairs of measurements on two variables x_1 and x_2 and the variability in the x_1 measurements is larger than the variability in the x_2 measurements.



$Var(X_1) > Var(X_2)$

point on X_1 is closer.

Figure 5: A scatter plot with greater variability in the x_1 direction.

Statistical distance

- Glancing at Figure 5, we see that values which are a given deviation from the origin in the x_1 direction are not as "surprising" or "unusual" as are values equidistant from the origin in the x_2 direction.
- It seems reasonable, then, to weight an x_2 coordinate more heavily than an x_1 coordinate of the same value when computing the "distance" to the origin.
- One way to proceed is to divide each coordinate by the sample standard deviation. Therefore, we have the "standardized" coordinates $x_1^* = x_1/\sqrt{s_{11}}$ and $x_2^* = x_2/\sqrt{s_{22}}$. After taking the differences in variability into account, we determine distance using the standard Euclidean formula.

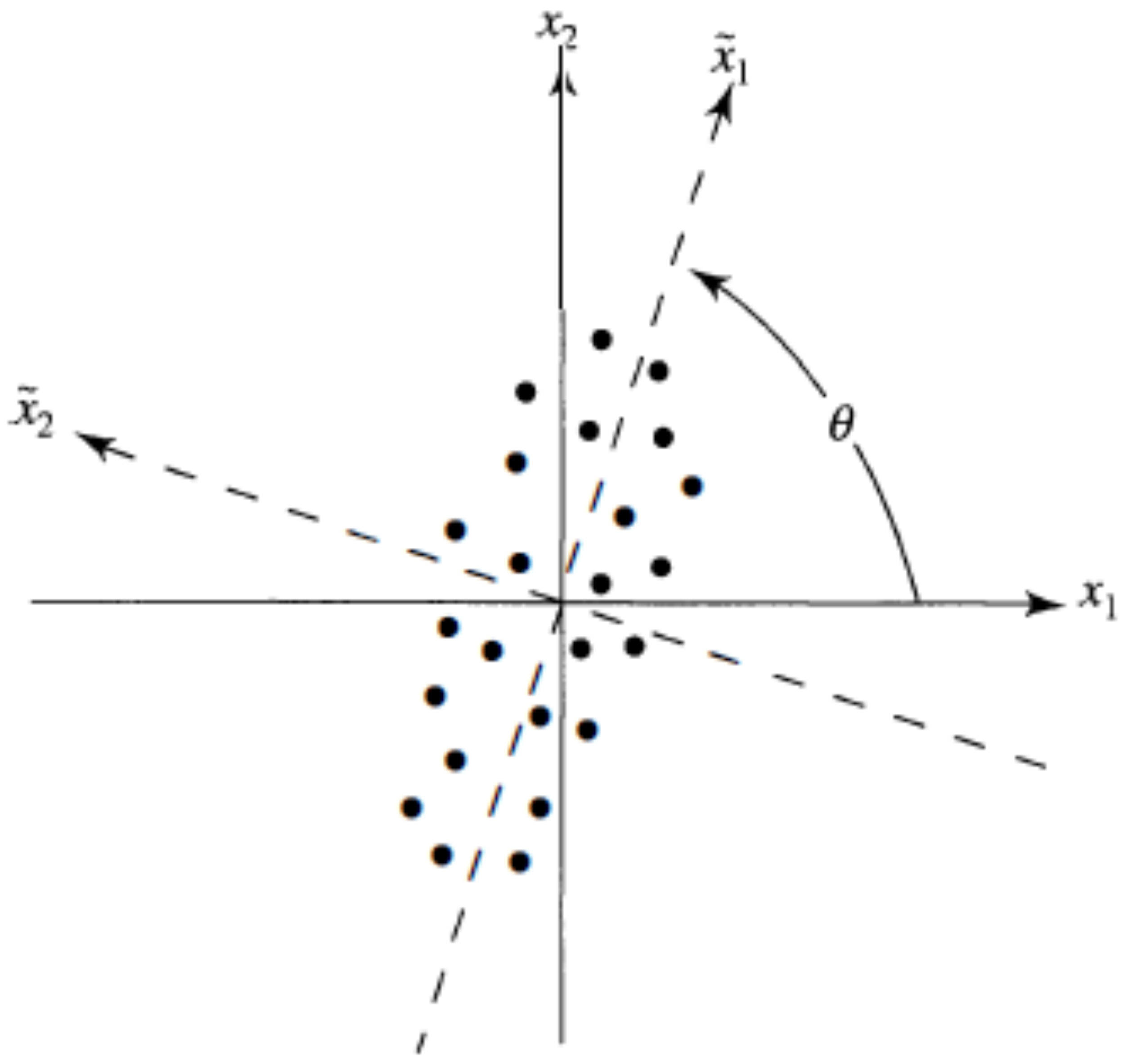
Statistical distance

Thus, a statistical distance of the point $P = (x_1, x_2)$ from the origin $O = (0, 0)$ can be computed from its standardized coordinates $x_1^* = x_1/\sqrt{s_{11}}$ and $x_2^* = x_2/\sqrt{s_{22}}$ as

$$
\begin{aligned}
d(O, P) &= \sqrt{(x_1^*)^2 + (x_2^*)^2} \\
&= \sqrt{\left(\frac{x_1}{\sqrt{s_{11}}}\right)^2 + \left(\frac{x_2}{\sqrt{s_{22}}}\right)^2} = \sqrt{\frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}}}
\end{aligned}
$$

Statistical distance

- The statistical distance considered before still does not include most of the important cases we shall encounter, because of the assumption of independent coordinates.
- The scatter plot in figure below depicts a two-dimensional situation in which the x_1 measurements do not vary independently of the x_2 measurements.


x_2
$\tilde{x}_1$
$\tilde{x}_2$
θ
x_1

Statistical distance

- Consider the new coordinate system

$$\tilde{x}_1 = x_1 \cos(\theta) + x_2 \sin(\theta)$$
$$\tilde{x}_2 = -x_1 \sin(\theta) + x_2 \cos(\theta)$$

- Define the distance from the point $P = (\tilde{x}_1, \tilde{x}_2)$ to the origin $O = (0, 0)$ as

$$d(O, P) = \sqrt{\frac{\tilde{x}_1^2}{\tilde{s}_{11}} + \frac{\tilde{x}_2^2}{\tilde{s}_{22}}}$$

where $\tilde{s}_{11}$, $\tilde{s}_{22}$ denote the sample variances of the $\tilde{x}_1$, $\tilde{x}_2$ measurements.

- The distance from $P = (\tilde{x}_1, \tilde{x}_2)$ to the origin $O = (0, 0)$ can be written in terms of the original coordinates x_1 and x_2 of P as $d(O, P) = \sqrt{a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2}$. Correlated of x_1 and x_2.

Mahalanobis Distance

The above intuition leads to the Mahalanobis Distance defined below:

- For a vector $\mathbf{x}$ and a distribution Q with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the *Mahalanobis distance* is:

$$d_M(\mathbf{x}, \boldsymbol{\mu}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

- Given two points $\boldsymbol{x}$ and $\boldsymbol{y}$, the Mahalanobis distance between them with respect to Q is

$$d_M(\boldsymbol{x}, \boldsymbol{y}; Q) = \sqrt{(\boldsymbol{x} - \boldsymbol{y})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{y})}$$

which means that $d_M(\boldsymbol{x}, Q) = d_M(\boldsymbol{x}, \boldsymbol{\mu}; Q)$