الهدف من تحليل البيانات هو تقليص واختزال
جداول ذات ابعاد متعددة (ذات ابعاد متعددة) من اجل
وضعها في كشف دقيق، ووضعها في جداول أقل واكتفاظ
لتسهيل عملية تصنيفها و تبويبها، مما يسهل عملية التفسير
والفهم والبحث عن العلاقات بين عديد المتغيرات وقياسها.

وسنتطرق إلى اوطرق تحليل قواعد البيانات
وهو التحليل باستخدام المركبات الاساسية

I* التحليل باستخدام المركبات الاساسية

استخدمت هذه الطريقة سنة 1901م، ولم يتم تحليل الاحصاء إلا سنة 1933 من طرف Harold Hotelling. وبرغم من ذلك لم يتم استعمال هذه الطريقة بشكل واسع لصعوبة التحاسب معها إلا في آواخر الثمانينيات عند ظهور جهاز الحاسوب والبرامج الاحصائية.

1) هدف التحليل باستخدام المركبات الاساسية: يكمن الهدف الرئيسي لطريقة التحليل باستخدام المركبات الاساسية اختزال وتقليص عدد كبير من المتغيرات إلى أقل عدد ممكن من المركبات وذلك دون تشويه أو تضييع للمعلومات الخاصة للمتغيرات، والحفاظ على أكبر كمية ممكنة.

Scanned with CamScanner

3/

1- مرحلة التحليل (ACP)

المعطيات المراد تحليلها ودراستها المأخوذة من الواقع لظاهرة معينة يتم وضعها في جدول والتي نأخذها على شكل مصفوفة ليتم التعامل معها وتحليلها ولتكن المصفوفة X التالية والممثلة كالتالي:

$$X_{n.p} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ x_{21} & \ddots & & \vdots & & \vdots \\ \vdots & & \ddots & & & \vdots \\ x_{i1} & \cdots & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & & & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & \cdots & \cdots & \cdots & x_{np} \end{pmatrix}$$

- المصفوفة X تمثل البيانات المراد تحليلها ودراستها

p = تمثل عدد المتغيرات وهو عدد الأعمدة للمصفوفة X

n = تمثل عدد الأفراد وهو عدد الأسطر للمصفوفة X

Scanned with CamScanner

3/

مثال: لتكن المصفوفة التالية

$$X = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

- عدد المتغيرات هنا هو 3
- عدد الافراد هو 2

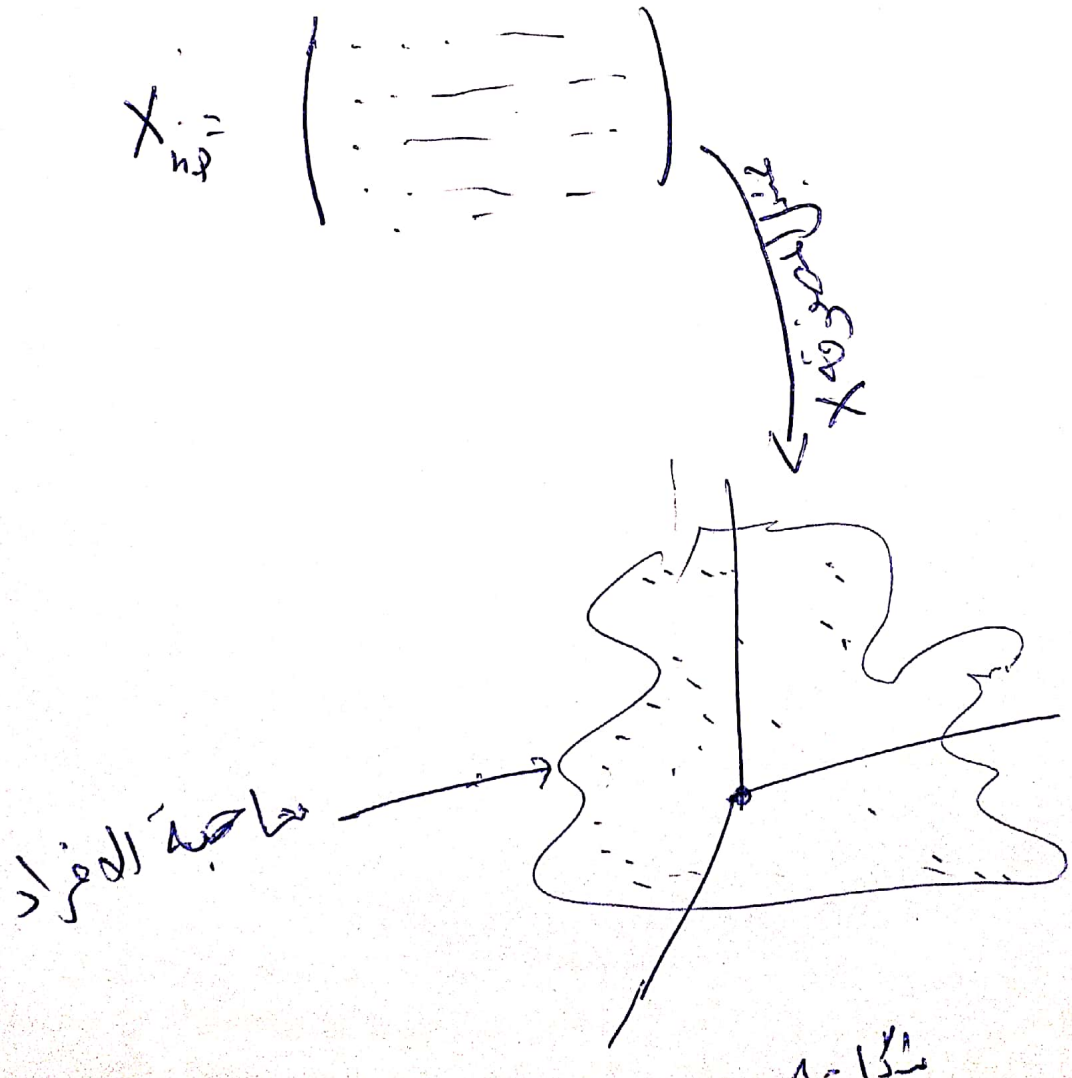
4/ إن التحليل بطريقة المركبات الرئيسية (الاساسية) يعتمد على وجهين وهما التحليل في $\mathbb{R}^p$ ونعني به التحليل في فضاء المتغيرات، و التحليل في $\mathbb{R}^n$ ونعني به التحليل في فضاء الافراد، وسنتطرق الى الطريقتين خلال المحاضرات.

Scanned with CamScanner

أولا = تحليل الأفراد في فضاء المتغيرات

في الدراسة التحليلية لبيانات تستلزم علينا تمثيل هاته البيانات في فضاء من أجل التحليل النفري لوجود ارتباط أو علاقات أو ما شابه ذلك، و بالتالي عند القيام بالتمثيل البياني نحصل على ما يلي:

$$X_{np} = \begin{pmatrix} \cdots & \cdots \\ \cdots & \cdots \\ \cdots & \cdots \\ \cdots & \cdots \end{pmatrix}$$



شكل - 1 -

Scanned with CamScanner

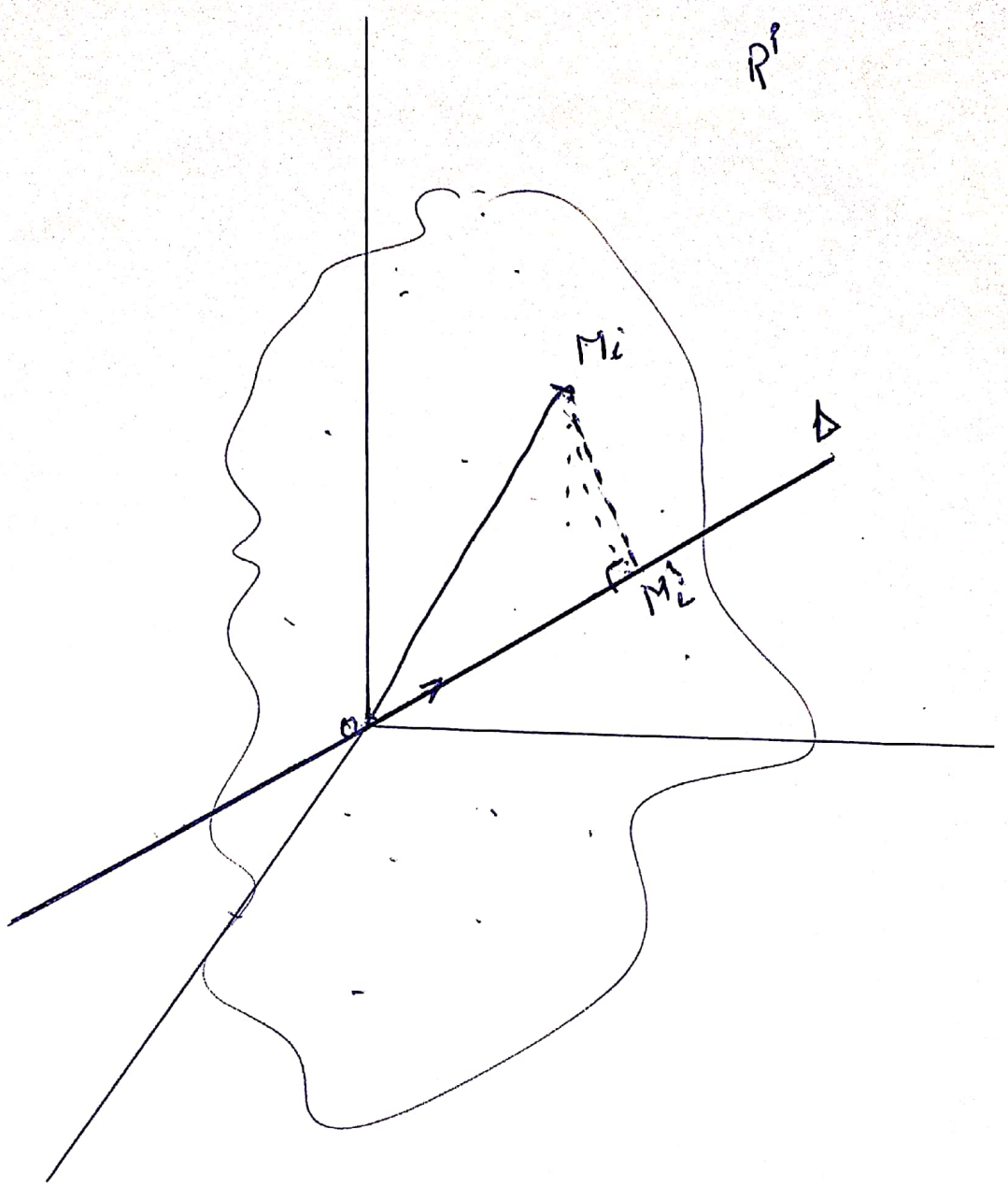
من المشاهد من الشكل ٥- ان عملية التمثيل المتوقعة
اعطتنا سحابة إستقاطات لكنها غير واضحة ولا تساعد
على تقديم اى تفسير او تحليل للبيانات.
هذا يدفعنا الى البحث عن فضاء شعاعي جزئي
يكون أقل درجة (اقل تعقيد) ويسمح لنا بإيجاد شكل
تمثيلي للبيانات بحيث تكون قابلة للفهم.

(1) البحث عن المحاور:

نفترض ان $\vec{u}$ هو احد محاور الفضاء الشعاع الجزئي
المراد إيجاده والذي تحربا طبيعة. حيث ان كل
سحابة الافراد يتم إسقاطها عليه. وبتالي
نحدها في الشكل الموالي

Scanned with CamScanner

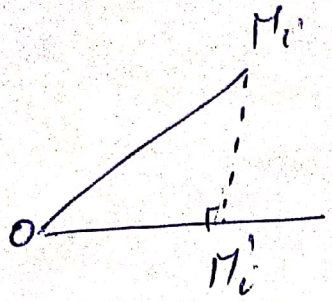
7/



من الملاحظ ان إسقاط النقطة M_i على المحور الجديد Δ تخلق مشكل وهو ان كمية المعلومات الخاصة سوف تنقص بصيغة اخرى عملية الاسقاط على المحور Δ نوعة تسبب ضياع لكمية المعلومات الخاصة للبيانات.

Scanned with CamScanner

8/



وذلك ناتج لان

$OM_i \geq OM'_i$

وذلك حسب نظرية فيثاغورث

وعليه من اجل الحفاظ على كمية المعلومات التي تحويها البيانات يجب ان نبحث عن محور Δ من شأنه ان يعظم مجموع مسافات الاسقاط هندسيا اي يكون $\sum OM'^2_i$ اكبر ما يمكن

ونكتب رياضيا $\max \Phi = \sum OM'^2_i$

مثال:

لتكن النقاط التالية $A:(1,3)$ ، $B(0,2)$ ، $C:(2,1)$

احسبا احداثيات اسقاط كل الاحور ذو الشعاع $u:(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$

* الحل

$$T_A = Au = (1,3)\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix} = 4/\sqrt{2}$$

$$T_B = Bu = (0,2)\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix} = 2/\sqrt{2}$$

$$T_C = Cu = (2,1)\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix} = 3/\sqrt{2}$$

Scanned with CamScanner