

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO <sup>2ej</sup>  
7

FACULTAD DE CIENCIAS



GEOMETRIA DE LA ESTADISTICA

T E S I S  
QUE PARA OBTENER EL TITULO DE:  
LICENCIADO EN ACTUARIA  
P R E S E N T A:  
JOSEFINA CALVA MARQUEZ



Universidad Nacional  
Autónoma de México

UNAM



## **UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso**

### **DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# INDICE

## INTRODUCCION. 4

## CAPITULO 1. 5

### 1.1 ESPACIO VECTORIAL N-DIMENSIONAL 5

1.1.1 Longitud de un vector 6

1.1.2 Producto Punto 6

1.1.3 Ortogonalidad de Vectores 7

1.1.4 Proyección Ortogonal 7

### 1.2 ESPACIO EUCLIDEANO TRIDIMENSIONAL. 9

1.2.1 Angulo entre rectas 13

1.2.2 Producto Vectorial 15

1.2.3 Triple producto escalar 15

1.2.4 Independencia Lineal de Vectores 18

1.2.5 Normal a un plano 19

1.2.6 Bases 20

### 1.3 ESPACIO EUCLIDEANO N-DIMENSIONAL 21

### 1.4 VARIETADES 22

### 1.5 RECIPROCIDAD 25

### 1.6 ANGULOS 26

### 1.7 TRANSFORMACIONES 39

1.7.1 Transformación ortogonal 39

1.7.2 Transformación por coordenadas polares. 45

1.8 HIPERPLANDS TANGENTES 52

1.9 FORMAS CUADRATICAS 55

1.9.1 Reducción de una Forma Cuadrática Real a una Forma Diagonal. 55

1.10 FIGURAS GEOMETRICAS SIMPLS  
N-DIMENSIONALES 60

1.11 CONTENIDO 63

1.11.1 Contenido de un intervalo  
n-dimensional 63

1.11.2 Contenido de una Hiper-esfera 63

1.11.3 Contenido de una Hiper-elipsoide 65

1.11.4 Contenido de un Hiper-prisma 65

1.11.5 Contenido de un cilindro de orden n 66

1.11.6 Contenido de un Paralelo topó 66

1.11.7 Contenido de una Hiper-pirámide 67

1.11.8 Contenido de un Simplejo 68

1.12 RADIO DE CURVATURA. 72

CAPITULO 2. 73

2.1 INTERPRETACION GEOMETRICA DEL MODE-  
LO DE REGRESION LINEAL BIVARIADO 73

2.2 INTERPRETACION GEOMETRICA DEL MODE-  
LO DE REGRESION LINEAL SIMPLE. 76

2.3 REPRESENTACION GEOMETRICA DE LA  
MEDIA. 79

2.4 REPRESENTACION GEOMETRICA DE LA  
VARIANZA. 80

- 2.5 COLINEALIDAD EN EL MODELO DE REGRESION LINEAL SIMPLE. 81
- 2.6 PRUEBAS DE HIPOTESIS 82
  - 2.6.1 Analisis de los Residuales bajo la Hipotesis alternativa. 84
  - 2.6.2 Analisis de los Residuales bajo la Hipotesis nula. 85
- 2.7 DISTRIBUCION NORMAL 86
- 2.8 DISTRIBUCION  $t$  DE STUDENT 91
- 2.9 DISTRIBUCION DE LA MEDIA MUESTRAL DE UNA POBLACION UNIFORME. 93
- 2.10 COEFICIENTE DE CORRELACION EN LA DISTRIBUCION NORMAL BIVARIADA. 98
- 2.11 DISTRIBUCION WISHART. 102
- 2.12 CORRELACIONES 105
- 2.13 CORRELACION MULTIPLE 109
- 2.14 CORRELACIONES CANONICAS 113
- 2.15 MEDIDA DE INFORMACION 114

CONCLUSIONES 118

REFERENCIAS 119

## INTRODUCCION

El saber que determinada función nos representa una recta, una curva, o un cuerpo en el espacio, nos ayuda a comprender el comportamiento de dicha función; lo mismo sucede cuando un planteamiento estadístico es representado geoméricamente, la asimilación del concepto resulta ser más fácil.

Así pues, la presente Tesis tiene como objeto analizar algunos conceptos estadísticos, cuyas interpretaciones geométricas, espero, sean de gran ayuda para aquellas personas que estén o hayan llevado al menos un curso de Estadística, así como una propuesta de apoyo a dicho curso.

La Tesis consta únicamente de dos capítulos, en el primero se plantean algunos conceptos de Geometría y Cálculo Diferencial e Integral necesarios para poder comprender la interpretación geométrica planteada en el segundo capítulo, donde los conceptos estadísticos se manejan en forma gradual, es decir, primero se verá el modelo de regresión lineal bivariado, el modelo de regresión lineal simple, la interpretación de la media, la varianza, etc., hasta finalizar con el concepto de medida de información.

## CAPITULO 1

En este capítulo se plantearán algunos conceptos necesarios para poder interpretar geométicamente la Estadística, en los casos donde se juzgue conveniente se contará con gráficas que ayuden un poco a percibir con facilidad la idea de los planteamientos.

## 1.1

## ESPACIO VECTORIAL N-DIMENSIONAL

## Definición:

El Espacio Vectorial  $n$ -dimensional  $V_n$  es el conjunto de todas las  $n$ -adas de números reales a las que denotaremos por  $\bar{x} = (x_1, \dots, x_n)$ ,  $x_i \in \mathbb{R}$ ,  $(i=1, \dots, n)$  y llamaremos vectores, donde se cumplen las siguientes propiedades para las relaciones de igualdad y las operaciones de adición y de multiplicación por un número real:

a) Igualdad de vectores. Si  $\bar{x} = (x_1, \dots, x_n)$  y  $\bar{y} = (y_1, \dots, y_n)$  son vectores en  $V_n$ , entonces  $\bar{x} = \bar{y}$  si  $x_i = y_i \quad \forall i=1, \dots, n$ .

b) Adición de vectores. Si  $\bar{x} = (x_1, \dots, x_n)$  y  $\bar{y} = (y_1, \dots, y_n)$  son vectores en  $V_n$ , entonces  $\bar{x} + \bar{y} = (x_1 + y_1, \dots, x_n + y_n)$ .

c) Multiplicación de un vector por un número real. Si  $\bar{x} = (x_1, \dots, x_n)$  es un vector en  $V_n$  y  $r$  es un número real, entonces  $r\bar{x} = (rx_1, \dots, rx_n)$ .

En seguida se ilustra la adición de vectores y la multiplicación de un vector por un escalar.

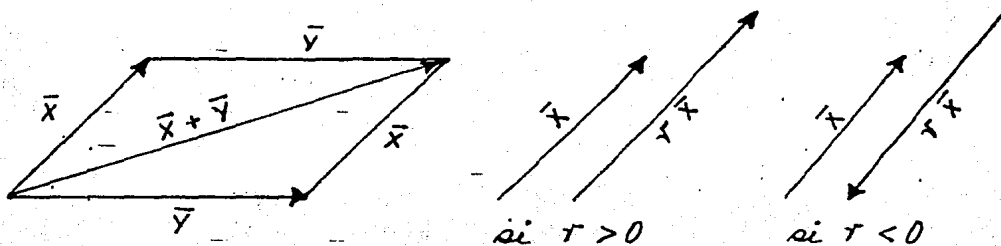


fig. 1.1

### 1.1.1

#### Longitud de un vector

Definición:

La longitud o norma de un vector  $\bar{x} = (x_1, \dots, x_n) \in V_n$ , denotada por  $\|\bar{x}\|$ , tiene como definición:

$$\|\bar{x}\| = \sqrt{x_1^2 + \dots + x_n^2} = \left[ \sum_{k=1}^n x_k^2 \right]^{1/2}$$

### 1.1.2

#### Producto Punto

Definición:

El Producto Punto o Producto Escalar  $\bar{x} \cdot \bar{y}$ , de dos vectores  $\bar{x}, \bar{y} \in V_n$ , donde  $\bar{x} = (x_1, \dots, x_n)$  y  $\bar{y} = (y_1, \dots, y_n)$ , está definido por:

$$\bar{x} \cdot \bar{y} = \sum_{k=1}^n x_k y_k = x_1 y_1 + \dots + x_n y_n$$



## 1.1.3

## Ortogonalidad de Vectores

Definición:

Un vector  $\bar{x}$  se dice que es ortogonal a un vector  $\bar{y}$  si  $\|\bar{x} + \bar{y}\| = \|\bar{x} - \bar{y}\|$ .

Geométicamente esto se expresa de la siguiente forma:

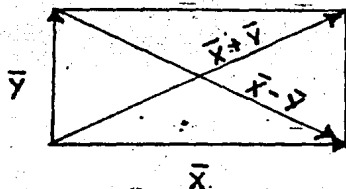


fig. 1.2

La definición de ortogonalidad es equivalente a afirmar que la diferencia de los cuadrados de las longitudes de las diagonales  $\bar{x} + \bar{y}$  y  $\bar{x} - \bar{y}$  del paralelogramo de lados  $\bar{x}$  y  $\bar{y}$  es cero, es decir:

$$\|\bar{x} + \bar{y}\|^2 - \|\bar{x} - \bar{y}\|^2 = 0$$

Como

$$\begin{aligned} \|\bar{x} + \bar{y}\|^2 - \|\bar{x} - \bar{y}\|^2 &= \sum_{k=1}^n (x_k + y_k)^2 - \sum_{k=1}^n (x_k - y_k)^2 \\ &= \sum_{k=1}^n (x_k^2 + 2x_k y_k + y_k^2 - x_k^2 + 2x_k y_k - y_k^2) \\ &= 4 \sum_{k=1}^n x_k y_k \end{aligned}$$

## 1.1.4

## Proyección Ortogonal

Definición:

Sean  $\bar{x}, \bar{y} \in V_n$ , con  $\bar{y} \neq \bar{0}$ . La Proyección

Ortogonal de  $\bar{x}$  sobre  $\bar{y}$ , denotada por  $P_{\bar{y}} \bar{x}$ , es el vector

$$P_{\bar{y}} \bar{x} = \frac{\bar{x} \cdot \bar{y}}{\|\bar{y}\|^2} \bar{y} = \frac{\bar{x} \cdot \bar{y}}{\|\bar{y}\|} \frac{\bar{y}}{\|\bar{y}\|}$$

Como el vector  $\frac{\bar{y}}{\|\bar{y}\|}$  es un vector unitario en la dirección de  $\bar{y}$ , el número  $\frac{\bar{x} \cdot \bar{y}}{\|\bar{y}\|}$  es la longitud de  $P_{\bar{y}} \bar{x}$ .

Este número se llama componente de  $\bar{x}$  en la dirección de  $\bar{y}$ , denotado por  $\text{Comp}_{\bar{y}} \bar{x}$ .

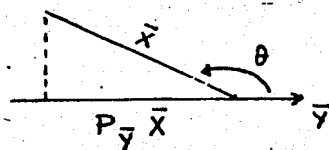
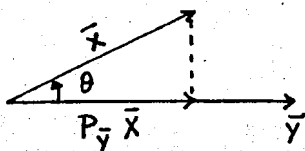


fig. 1.3

Debido a que un manejo de los conceptos referentes al Espacio Euclideo, tomando como base un espacio vectorial tridimensional  $V_3$  facilita un mejor entendimiento de su generalización, en esta sección se definirá lo que es una recta, un producto vectorial, la independencia lineal de vectores, etc., en  $V_3$ .

Definición:

El espacio euclideo tridimensional, denotado por  $\mathbb{R}^3$ , es el espacio vectorial tridimensional  $V_3$ , donde:

1) los elementos  $P = (x, y, z)$  de  $V_3$  son llamados puntos de  $\mathbb{R}^3$  (ver fig. 1.4)

2) un conjunto  $\mathcal{I}$  de puntos de  $\mathbb{R}^3$  es una recta si hay un punto  $P_0 \in \mathbb{R}^3$  y un vector no nulo  $a \in V_3$   $\rightarrow$

$$\mathcal{I} = \{P_0 + ta \mid t \in \mathbb{R}\} \quad (\text{ver fig. 1.5})$$

3) un conjunto  $\mathcal{P}$  de puntos de  $\mathbb{R}^3$  es un plano si hay un punto  $P_0 \in \mathbb{R}^3$  y dos vectores no paralelos  $a$  y  $b$  de  $V_3$  tales que

$$\mathcal{P} = \{P_0 + ua + vb \mid u, v \in \mathbb{R}\}$$

(ver fig. 1.6)

4) la distancia denotada por  $d(P_1, P_2)$ , del punto  $P_1 = (x_1, y_1, z_1)$  al punto  $P_2 = (x_2, y_2, z_2)$  es la longitud del vector  $P_2 - P_1$ , es decir

$$d(P_1, P_2) = \|P_2 - P_1\| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Las ecuaciones

$$P = P_0 + t a$$

$$y \quad P = P_0 + u a + v b$$

se llaman ecuaciones vectoriales de la recta y el plano, respectivamente, y las ecuaciones correspondientes entre los componentes

$$x = x_0 + t a_1, \quad y = y_0 + t a_2, \quad z = z_0 + t a_3$$

$$y \quad x = x_0 + u a_1 + v b_1, \quad y = y_0 + u a_2 + v b_2, \quad z = z_0 + u a_3 + v b_3$$

se llaman ecuaciones paramétricas de la recta y el plano respectivamente.

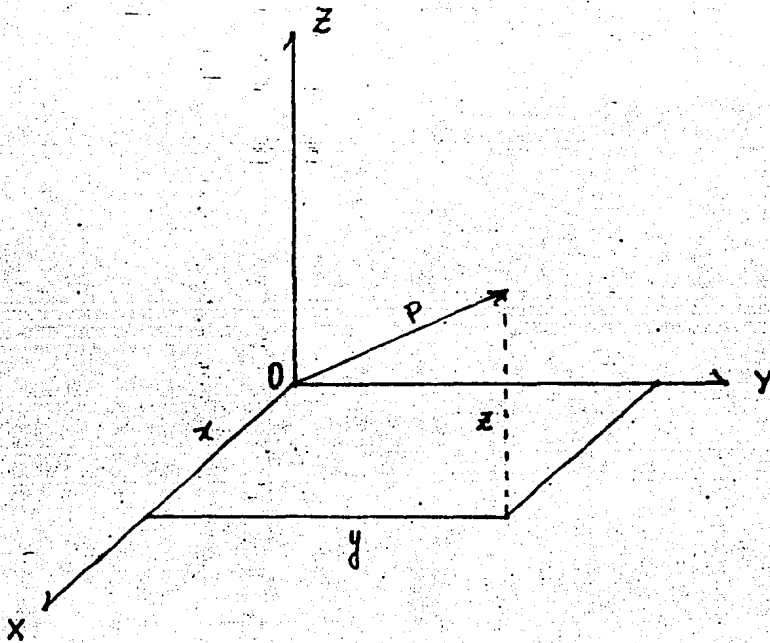


fig. 1.4

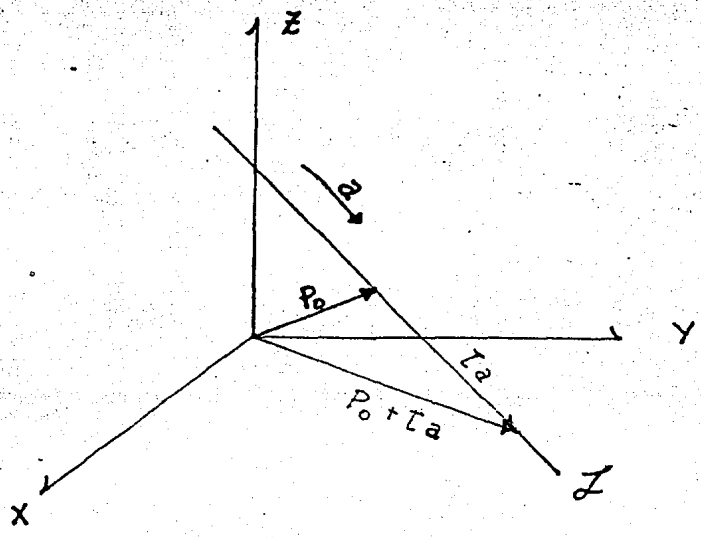
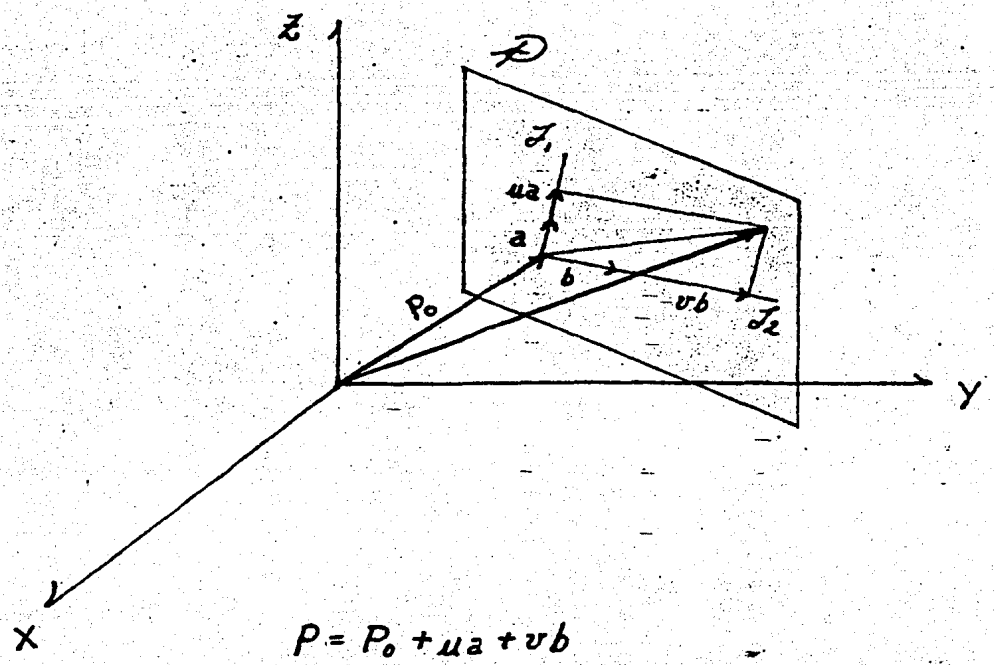


fig. 1.5



$$P = P_0 + ua + vb$$

fig. 1.6

Ejemplo:

Determinar el plano que pasa por los puntos  $P_0 = (1, 0, 0)$ ,  $P_1 = (0, 1, 0)$  y  $P_2 = (0, 0, 1)$ .

Definamos  $a = P_1 - P_0$ ,  $b = P_2 - P_0$  y

$$\mathcal{P} = \{ P_0 + ua + vb \mid u, v \in \mathbb{R} \} \\ = \{ P_0 + u(P_1 - P_0) + v(P_2 - P_0) \} \quad \text{----- (1)}$$

$\mathcal{P}$  es un plano que contiene a  $P_0$ ,  $P_1$  y  $P_2$  ya que  $a = (-1, 1, 0)$  y  $b = (-1, 0, 1)$ .

$P_0$  resulta de hacer  $u = 0$ ,  $v = 0$ ;  $P_1$  resulta de hacer  $u = 1$ ,  $v = 0$ ; y  $P_2$  de hacer  $u = 0$ ,  $v = 1$ .

Substituyendo los valores de  $a$ ,  $b$  y  $P_0$  en (1) obtenemos:

$$\mathcal{P} = \{ (1, 0, 0) + u(-1, 1, 0) + v(-1, 0, 1) \} \\ = \{ (1 - u - v, u, v) \}$$

que representa un plano que pasa por los puntos  $(1, 0, 0)$ ,  $(0, 1, 0)$ , y  $(0, 0, 1)$ .

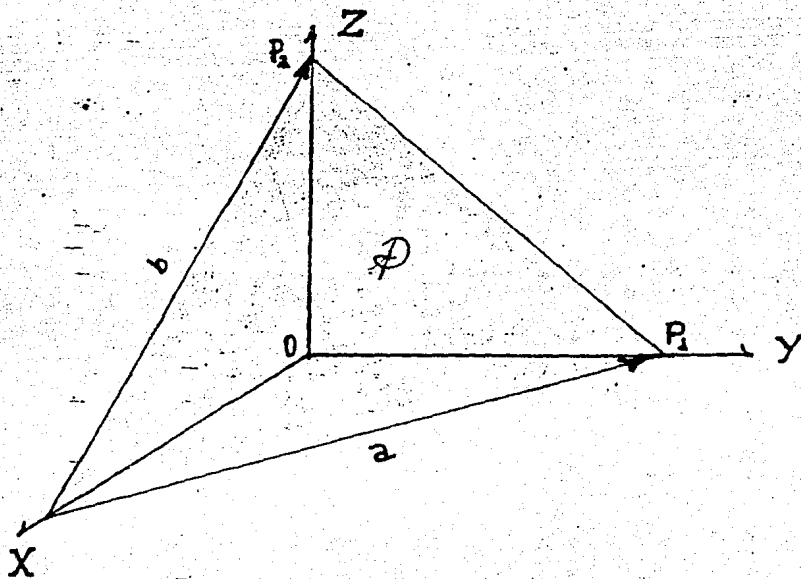


fig. 1.7

1.2.1

## Ángulo entre rectas

Definición:

$\theta$  es un ángulo entre las rectas  $\mathcal{L}_1$  y  $\mathcal{L}_2$  si para ciertos vectores no nulos  $a$  y  $b$ ,  $\mathcal{L}_1 = \{P_1 + sa\}$ ,  $\mathcal{L}_2 = \{P_2 + tb\}$ , y  $\theta$  es el ángulo entre  $a$  y  $b$ .

Ejemplo:

Encontrar un ángulo entre las dos rectas:

$$\mathcal{L}_1 = \{(1, 3, -2) + t(3, -6, 9)\} \text{ y } \mathcal{L}_2 = \{(2, 1, 7) + s(1, -3, 4)\}$$

Sean  $a = (3, -6, 9)$  y  $b = (1, -3, 4)$ . Si  $\theta$  es el ángulo entre  $a$  y  $b$ , entonces

$$\cos \theta = \frac{a \cdot b}{\|a\| \|b\|} = \frac{3 + 18 + 36}{3\sqrt{14} \sqrt{26}} = \frac{57}{6\sqrt{91}} = 0.99587$$

Expresando los vectores de  $V_3$  en términos de los vectores unitarios:

$$i = (1, 0, 0), \quad j = (0, 1, 0), \quad k = (0, 0, 1)$$

estos vectores apuntan en las direcciones positivas de los ejes  $X$ ,  $Y$  y  $Z$ , respectivamente. Así para cualquier vector  $a \in V_3$ , tendremos:

$$\begin{aligned} a = (a_1, a_2, a_3) &= (a_1, 0, 0) + (0, a_2, 0) + (0, 0, a_3) \\ &= a_1(1, 0, 0) + a_2(0, 1, 0) + a_3(0, 0, 1) \end{aligned}$$

o bien

$$a = a_1 i + a_2 j + a_3 k$$

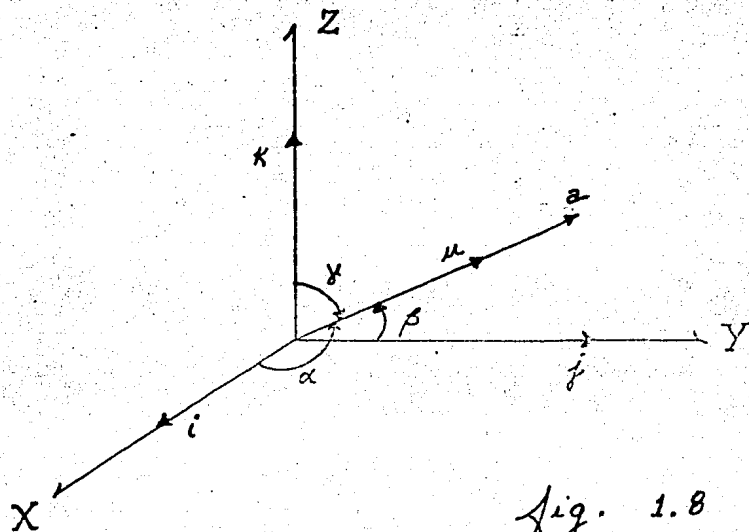


fig. 1.8

Sea  $a = (l, m, n)$  un vector no nulo paralelo a una recta  $\mathcal{L}$ . Los números  $l, m, n$  se llaman números directores de la recta  $\mathcal{L}$ . Sea  $\alpha$  el ángulo entre  $i$  y  $a$ ;  $\beta$  el ángulo entre  $j$  y  $a$ , y  $\gamma$  el ángulo entre  $k$  y  $a$  (ver fig. 1.8). Los ángulos  $\alpha, \beta$  y  $\gamma$  se llaman ángulos directores de  $\mathcal{L}$ , y  $\cos \alpha, \cos \beta, \cos \gamma$  se llaman cosenos directores de  $\mathcal{L}$ . Si  $u$  es el vector unitario en la dirección de  $a$ :

$$u = \frac{a}{\|a\|} = \frac{(l, m, n)}{\sqrt{l^2 + m^2 + n^2}} = (\cos \alpha, \cos \beta, \cos \gamma)$$

donde:

$$\cos \alpha = i \cdot u, \quad \cos \beta = j \cdot u, \quad \cos \gamma = k \cdot u$$

$$u = (\cos \alpha, \cos \beta, \cos \gamma)$$

$$\text{y} \quad \cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1$$



1.2.2

## Producto Vectorial

Definición:

El producto vectorial de dos vectores  $a = (a_1, a_2, a_3)$  y  $b = (b_1, b_2, b_3)$  de  $V_3$ , denotado por  $a \times b$ , lo que llamaremos " $a$  cruz  $b$ " es el vector definido por:

$$a \times b = (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1)$$

Geométicamente la norma de  $a \times b$ ,  $\|a \times b\|$ , donde  $a, b$  son vectores de  $V_2$ , nos representa el área del paralelogramo de lados  $a$  y  $b$ , ya que  $\|a \times b\| = \|a\| \|b\| \sin \theta$ , ver [1].

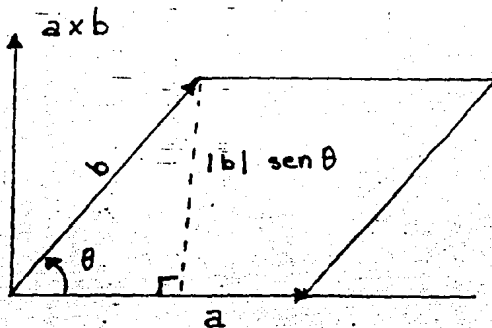


fig. 1.9

1.2.3

## Triple Producto Escalar

Dados 3 vectores cualesquiera  $a, b$  y  $c$  en  $V_3$ , se puede formar el producto escalar de  $a$  con  $b \times c$ , ya que  $b \times c$  es un vector, a este producto se le llamará triple producto escalar.

Definición:

Dados 3 vectores  $a, b, c \in V_3$ , el triple producto escalar de  $a, b, c$ , denotado por  $[a b c]$ , se define por:  $[a b c] = a \cdot (b \times c)$ .

Para obtener una interpretación geométrica de una terna arbitraria de vectores  $a, b, c$ , construimos  $a, b$  y  $c$  con el mismo punto inicial  $P_0$  y denominemos  $\mathcal{P}$  al plano que pasa por  $P_0$  determinado por  $b$  y  $c$ .

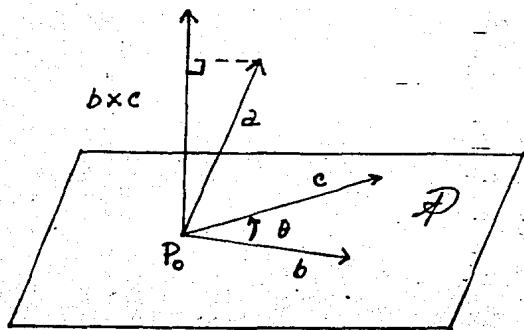


fig. 1.10

Tendremos que  $[a b c]$  representa el volumen del paralelepípedo de lados  $a, b$  y  $c$ , ya que el volumen es el área de la base por la altura. La base es un paralelogramo de lados  $b$  y  $c$ , de aquí que su área se represente por  $\|b \times c\|$  y la altura estará dada por la componente de  $a$  en la dirección de  $b \times c$  (ver sec. 1.1.4) denotada por  $\text{Comp}_{(b \times c)} a$ , de aquí que:

$$[a b c] = a \cdot (b \times c) = \|b \times c\| \text{Comp}_{(b \times c)} a$$

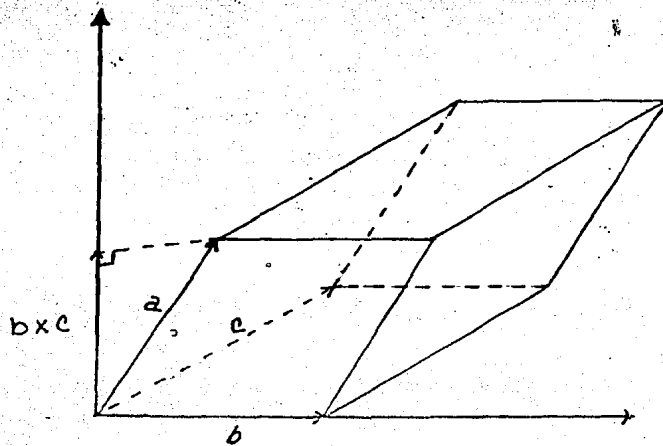


fig. 1.11

Ejemplo:

Encontrar el volumen del paralelepípedo de lados  $a = (2, 3, -1)$ ,  $b = (3, -7, 5)$  y  $c = (1, -5, 2)$ .

$$\begin{aligned}
 [a \ b \ c] &= a \cdot (b \times c) = a \cdot (-7(2) - 5(-5), 5(1) - 3(2), \\
 &\quad 3(-5) + 7(1)) \\
 &= a \cdot (11, -1, -8) \\
 &= (2, 3, -1) \cdot (11, -1, -8) \\
 &= 22 - 3 + 8 = 27
 \end{aligned}$$

lo cual equivale a resolver el siguiente determinante:

$$\begin{vmatrix} 2 & 3 & -1 \\ 3 & -7 & 5 \\ 1 & -5 & 2 \end{vmatrix} = 2 \begin{vmatrix} -7 & 5 \\ -5 & 2 \end{vmatrix} - 3 \begin{vmatrix} 3 & -1 \\ -5 & 2 \end{vmatrix} + 1 \begin{vmatrix} 3 & -1 \\ -7 & 5 \end{vmatrix}$$

$$\begin{aligned}
 &= 2(-14 + 25) - 3(6 - 5) + 1(15 - 7) \\
 &= 2(11) - 3 + 8 = 22 - 3 + 8 \\
 &= 27
 \end{aligned}$$

1.2.4

## Independencia Lineal de Vectores

Definición:

Un conjunto  $\{a_1, \dots, a_k\}$  de  $k$  vectores de  $V_n$  se dice que es linealmente independiente si:

$$r_1 a_1 + \dots + r_k a_k = 0 \quad (r_i \in \mathbb{R})$$

implica  $r_1 = \dots = r_k = 0$ . Si  $\{a_1, \dots, a_k\}$  no es linealmente independiente se dice que es linealmente dependiente.

Un conjunto de  $k$  vectores  $\{a_1, \dots, a_k\}$  es, pues, linealmente dependiente si y sólo si hay  $k$  números reales  $r_1, \dots, r_k$  no todos iguales a cero, tales que

$$r_1 a_1 + \dots + r_k a_k = 0$$

La expresión  $r_1 a_1 + \dots + r_k a_k$  donde  $r_1, \dots, r_k \in \mathbb{R}$  se dice que es una combinación lineal de los vectores  $a_1, \dots, a_k$ .

La dependencia lineal de dos vectores es equivalente a que los dos vectores sean paralelos.

Teorema:

Tres vectores  $a, b, c \in V_3$  son linealmente dependientes si y sólo si  $[a \ b \ c] = a \cdot (b \times c) = 0$

Ejemplo:

Verificar que los vectores  $a = (1, 2, -2)$ ,  $b = (0, 3, 1)$  y  $c = (-1, 1, 3)$  son linealmente dependientes.

$$[a \ b \ c] = \begin{vmatrix} 1 & 2 & -2 \\ 0 & 3 & 1 \\ -1 & 1 & 3 \end{vmatrix} = 1 \begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix} - 1 \begin{vmatrix} 2 & -2 \\ 3 & 1 \end{vmatrix} = 9 - 1 - 2 - 6 = 0$$

$\therefore a, b$  y  $c$  son linealmente dependientes.

La dependencia lineal de tres vectores es equivalente a que los 3 vectores sean paralelos (a un mismo plano).

1.2.5

Normal a un plano

Consideremos el plano

$$\mathcal{P} = \{ P_0 + ua + vb \mid u, v \in \mathbb{R} \}$$

$\mathcal{P}$  es el plano que pasa por  $P_0$  determinado por el par de vectores no paralelos  $a$  y  $b$ . Cualquier vector no nulo ortogonal a ambos  $a$  y  $b$  se llama vector normal a  $\mathcal{P}$ . Así pues  $a \times b$  es un vector normal a  $\mathcal{P}$  y toda otra normal es paralela a  $a \times b$ .

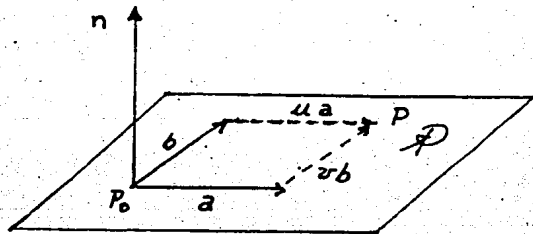


fig. 1.12.

Teorema:

Si  $n$  es una normal al plano

$$\mathcal{P} = \{ P_0 + ua + vb \mid u, v \in \mathbb{R} \}$$

entonces

$$\mathcal{P} = \{ P \mid n \cdot (P - P_0) = 0 \}$$

y  $\mathcal{P}$  es el único plano que pasa por  $P_0$  con normal  $n$ .  
La ecuación

$$n \cdot (P - P_0) = 0$$

recibe el nombre de ecuación vectorial del plano  $\mathcal{P}$ .

1.2.6

## Bases

Teorema:

Si  $a, b, c \in V_3$  son linealmente independientes, entonces para cada punto  $P \in \mathbb{R}^3$  existen números reales únicos  $u, v, t$  tales que:

$$P = ua + vb + tc$$

Definición:

Un conjunto  $\{a_1, \dots, a_k\}$  de vectores en  $V_n$  se dice que es una base de  $V_n$  si:

- i)  $\{a_1, \dots, a_k\}$  es linealmente independiente
- ii) todo vector de  $V_n$  puede expresarse como una combinación lineal de  $a_1, \dots, a_k$ .

Si todo vector de  $V_n$  puede expresarse como una combinación lineal de  $a_1, \dots, a_k$  entonces el conjunto  $\{a_1, \dots, a_k\}$  se dice que genera a  $V_n$ .

## 1.3

## ESPACIO EUCLIDEANO N-DIMENSIONAL

Definición:

El espacio euclideo  $n$ -dimensional denotado por  $\mathbb{R}^n$ , es el espacio vectorial  $n$ -dimensional  $V_n$ , donde:

i) los elementos  $x = (x_1, x_2, \dots, x_n)$  de  $V_n$  son puntos en  $\mathbb{R}^n$ .

ii) un conjunto  $P$  de puntos de  $\mathbb{R}^n$  se denomina plano  $k$ -dimensional o hiperplano en  $\mathbb{R}^n$  ( $k=1, \dots, n-1$ ), si hay un punto  $P_0 \in \mathbb{R}^n$  y  $k$  vectores linealmente independientes  $a_1, \dots, a_k \in V_n$  tales que:

$$P = \{P_0 + t_1 a_1 + \dots + t_k a_k \mid t_1, \dots, t_k \in \mathbb{R}\}$$

iii) la distancia,  $d(P, Q)$ , desde el punto  $P = (x_1, \dots, x_n)$  al punto  $Q = (y_1, \dots, y_n)$  en  $\mathbb{R}^n$  es la longitud del vector  $Q - P$ , es decir:

$$d(P, Q) = \|Q - P\| = [(y_1 - x_1)^2 + \dots + (y_n - x_n)^2]^{1/2}$$

Así pues un plano unidimensional

$$L = \{P_0 + t a \mid t \in \mathbb{R}\} \quad (a \neq 0)$$

es conocido como recta.

La ecuación

$$P = P_0 + t_1 a_1 + \dots + t_k a_k$$

es la ecuación vectorial de un plano, y las ecuaciones de las componentes, son ecuaciones paramétricas del plano.

## 1.4

## VARIEDAD

Todos los conceptos anteriormente expuestos nos ayudarán a comprender mejor la generalización que de ellos se maneja posteriormente. En esta sección se introducirá el concepto de Variedad, el cual viene a ser de hecho la manera de nombrar en forma generalizada a objetos geométricos tales como la recta, el plano, etc.

Definición:

Una Variedad Lineal  $H$  de  $k$  dimensiones en  $\mathbb{R}^n$  es el conjunto  $H$  de todos los puntos tales que:

$$P = P_0 + t_1 w_1 + \dots + t_k w_k \quad (2)$$

$$t_1, \dots, t_k \in \mathbb{R}$$

donde  $P_0$  es un punto dado de  $\mathbb{R}^n$  y  $w_1, \dots, w_k$  son  $k$  vectores linealmente independientes dados en  $V_n$ .

Para  $k=1$ , la variedad se conoce como una recta; para  $k=2$ , la variedad se conoce como un plano.

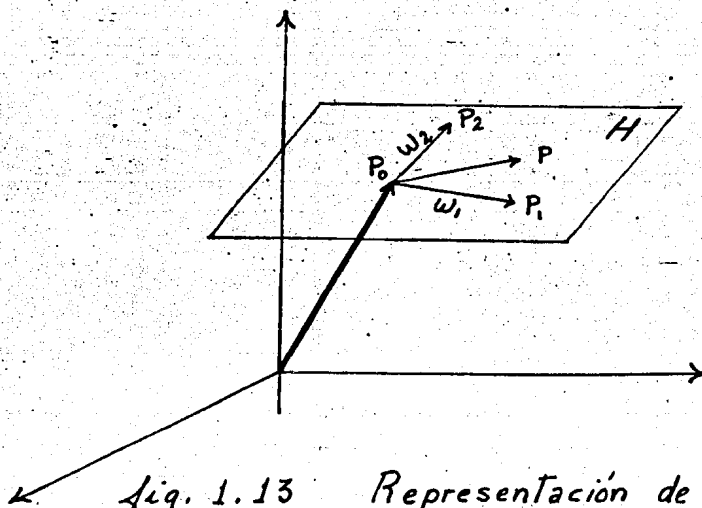


fig. 1.13

Representación de una variedad lineal de 2 dimensiones en  $\mathbb{R}^3$ .



Para  $k=n$ ,  $w_1, \dots, w_n$  constituyen una base en  $V_n$  y por lo tanto, la variedad lineal coincide con todo el espacio  $\mathbb{R}^n$ .

La ecuación (2) también se puede expresar de la siguiente forma:

$$P - P_0 = t_1 w_1 + \dots + t_k w_k$$

Si se igualan los componentes de los vectores, la ecuación (2) se transformará en un conjunto de  $n$  ecuaciones paramétricas.

$$x_1 = x_{10} + t_1 w_{11} + \dots + t_k w_{1k}$$

$$\vdots$$

$$x_n = x_{n0} + t_1 w_{n1} + \dots + t_k w_{nk}$$

Estas ecuaciones se pueden representar en la siguiente forma matricial

$$\bar{x} = \bar{x}_0 + W \bar{t}$$

donde:

$$\bar{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \bar{x}_0 = \begin{pmatrix} x_{10} \\ \vdots \\ x_{n0} \end{pmatrix}, \quad \bar{t} = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \text{ y } W = \begin{pmatrix} w_{11} & \dots & w_{1k} \\ \vdots & & \vdots \\ w_{n1} & \dots & w_{nk} \end{pmatrix}$$

Ejemplos:

1) En  $\mathbb{R}^n$ ,  $P - P_0 = t w_1$  es la ecuación vectorial de una recta  $L$ . Los puntos de  $L$ , son todos los puntos de la forma  $(t, 0, \dots, 0)$  con  $t \in \mathbb{R}$ . Decimos que  $L$  es el eje  $x_1$  en  $\mathbb{R}^n$ . Del mismo modo,  $P - P_0 = t w_2$  representa al eje  $x_2$ , y así sucesivamente.

2) En  $\mathbb{R}^n$ ,  $P - O = t_1 w_1 + t_2 w_2$  es la ecuación vectorial de un plano  $H$  que pasa por el origen  $O = (0, \dots, 0)$ . Los puntos de  $H$  son precisamente los puntos  $(x_1, x_2, 0, \dots, 0)$  en  $\mathbb{R}^n$ . Si  $H$  es el plano coordenado  $x_1, x_2$  de  $\mathbb{R}^n$ , observamos que, como un objeto geométrico,  $H$  es lo mismo

que el plano  $x_1, x_2$  de la geometría analítica en el plano, es decir,  $H$  es lo mismo que  $\mathbb{R}^2$

Si se generalizan los ejemplos 1) y 2) se tendrá que la ecuación:

$$P = \tau_1 w_1 + \dots + \tau_k w_k$$

es la ecuación de la variedad lineal coordenada,  $x_1, \dots, x_k$  formada por todos los puntos de la forma  $(x_1, x_2, \dots, x_k, 0, \dots, 0)$  y que, como un objeto geométrico, esta variedad es lo mismo que  $\mathbb{R}^k$ .

3) En el segundo capítulo se plantean conceptos cuya representación geométrica se basa en el concepto de las hiper-euadráticas, por lo cual es importante mencionarlo como ejemplo de una variedad.

Una hiper-euadrática es una variedad definida por una constante  $C$ , donde  $Q$  es una forma cuadrática de  $x_1, \dots, x_n$ .

La forma general de una hiper-euadrática es:

$$\sum_{j,k=1}^n a_{jk} x_j x_k = C$$

Si  $C=0$  tendremos un hiper-cono con vértice en el origen.

Si  $C \geq 0$  y  $j=k$  tendremos una hiper-esfera representada por:

$$\sum_{j=1}^n x_j^2 = C$$

1.5

RECIPROCIDAD

En geometría existe una cierta relación de reciprocidad; por ejemplo, dos puntos determinan una línea, dos líneas determinan un punto.

Si se considera un hiperplano expresado como:  
 $l_1 x_1 + \dots + l_n x_n = \text{constante} \quad \text{--- (3)}$

el planteamiento siguiente se refiere a una relación de reciprocidad: dadas las  $x$ 's, las  $l$ 's obedecen a una relación lineal, similarmente dadas las  $l$ 's, las  $x$ 's obedecen una relación lineal.

La simetría bilineal de (3) determina una relación uno a uno entre una familia de puntos determinados por las  $x$ 's y una familia de hiperplanos determinados por las  $l$ 's.

Si  $n$  puntos en  $p$  dimensiones están dados por el arreglo:

$$\begin{array}{cccc}
 x_{11} & \dots & \dots & x_{1n} \\
 x_{21} & & & \\
 \vdots & & & \\
 x_{p1} & \dots & \dots & x_{pn}
 \end{array}$$

analizando su forma transpuesta

$$\begin{array}{cccc}
 x_{11} & \dots & \dots & x_{p1} \\
 \vdots & & & \\
 x_{1n} & \dots & \dots & x_{pn}
 \end{array}$$

se puede hablar entonces de  $p$  puntos en  $n$  dimensiones.

1.6

## ANGULOS

En cualquier dimensión, la intersección de dos líneas rectas determina un plano que las contiene, y el ángulo formado entre ellas es único.

En el caso en que no se intersecten, se puede seleccionar un punto sobre alguna de ellas y dibujar una línea paralela a la otra recta. El ángulo entre las rectas que se intersectan es independiente de donde se escoja el punto.

Teorema:

La recta que pasa por el punto  $s = (x_1, \dots, x_n)$  y cuyos números directores son  $[l_1, l_2, \dots, l_n]$ , tiene por ecuaciones:

$$x_1 - x_1 = k l_1, \quad x_2 - x_2 = k l_2, \quad \dots, \quad x_n - x_n = k l_n$$

en donde  $k$  es una constante diferente de cero.

Si los números directores  $[l_1, \dots, l_n]$  son todos diferentes de cero, estas ecuaciones pueden escribirse en la forma simétrica:

$$\frac{x_1 - x_1}{l_1} = \frac{x_2 - x_2}{l_2} = \dots = \frac{x_n - x_n}{l_n}$$

La ecuación de la recta paralela que pasa por el origen 0 es:

$$\frac{x_1}{l_1} = \frac{x_2}{l_2} = \dots = \frac{x_n}{l_n}$$

Si las constantes  $l_i$  se multiplican por una misma constante, la recta será la misma.

Por lo tanto se puede pedir sin pérdida de generalidad, que las  $l_i$  cumplan con la condición de normalización  $\sum_{i=1}^n l_i^2 = 1$

Consideremos ahora un plano  $H$  generado por las  $n$ -adas de números reales cuya primera coordenada es igual a la constante  $a_1$ , es decir  $x_1 = a_1$ . Este plano intersecta a la recta que pasa por el origen, en el punto  $P$  cuyas coordenadas se obtienen de la siguiente forma:

$$\frac{x_1}{l_1} = \frac{a_1}{l_1} = \frac{x_2}{l_2} \Rightarrow \begin{aligned} x_1 &= a_1 \\ x_2 &= \frac{a_1 l_2}{l_1} \\ &\vdots \\ x_n &= \frac{a_1 l_n}{l_1} \end{aligned}$$

Así pues el punto  $P$  de intersección estará dado por:

$$\left( a_1, \frac{a_1 l_2}{l_1}, \frac{a_1 l_3}{l_1}, \dots, \frac{a_1 l_n}{l_1} \right)$$

La distancia del origen a este punto se calculará de la siguiente forma:

$$\overline{OP} = \left[ a_1^2 + \left( \frac{l_2 a_1}{l_1} \right)^2 + \dots + \left( \frac{l_n a_1}{l_1} \right)^2 \right]^{1/2}$$

$$\Rightarrow \overline{OP}^2 = a_1^2 + \left( \frac{l_2 a_1}{l_1} \right)^2 + \dots + \left( \frac{l_n a_1}{l_1} \right)^2$$

$$\Rightarrow \overline{OP}^2 = \frac{a_1^2}{l_1^2} \quad \text{ya que } \sum_{i=1}^n l_i^2 = 1$$

$$\therefore l_1 = \frac{a_1}{\overline{OP}}$$

donde la expresión  $a_1 / \overline{OP}$  representa el coseno del ángulo entre la recta y el eje  $x_1$ , ver fig. 1.14.

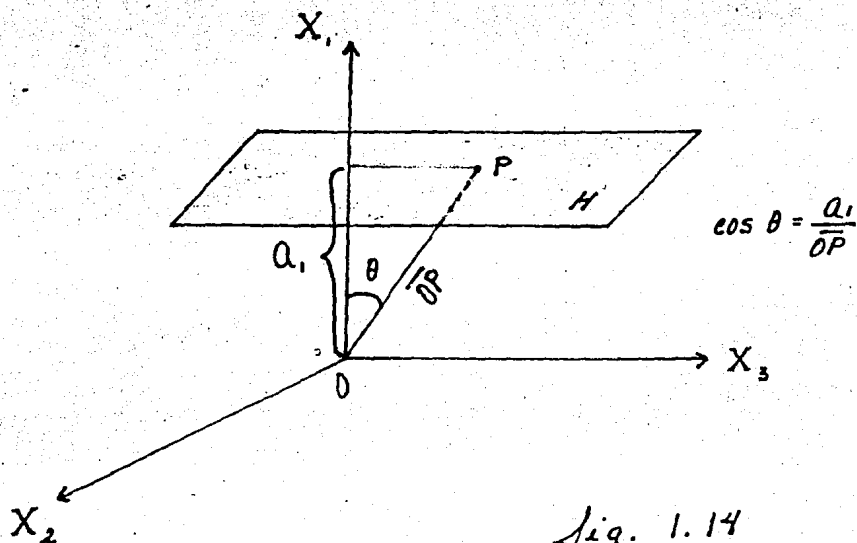


fig. 1.14

En forma similar se obtendrá que cada  $l_i$  representa el coseno del ángulo entre la recta y el eje  $x_i$ , (ver sección 1.2.1).

La distancia entre la proyección ortogonal de un punto  $s = (x_1, \dots, x_n)$ , sobre la línea  $\frac{x_1}{l_1} = \frac{x_2}{l_2} = \dots = \frac{x_n}{l_n}$ , y el origen, está dada por  $[\sum (l_i x_i)^2]^{1/2}$ ; y la distancia entre el punto de proyección y la recta es:

$$[\sum x_i^2 - \sum (l_i x_i)^2]^{1/2}$$

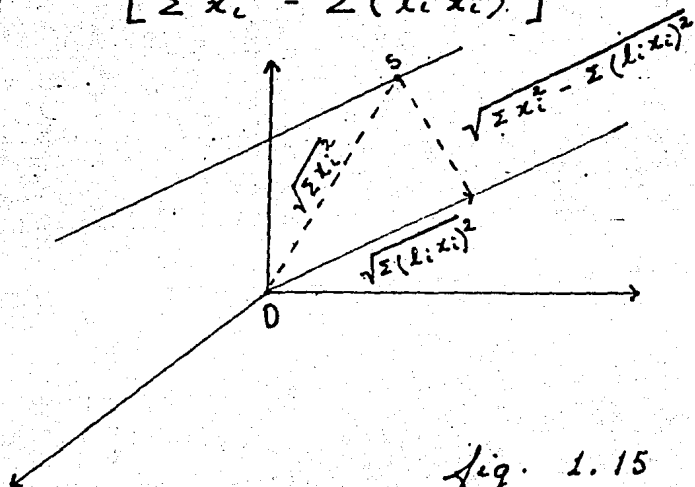


fig. 1.15

Consideremos otra recta cuyos cosenos directores son  $l'_i$ : La proyección de  $\overline{OP}$  sobre esta recta es  $\overline{OP} \cos \varphi$  donde  $\varphi$  es el ángulo entre las dos rectas.

La proyección de la recta con cosenos directores  $l_i$ , representada por  $l$ , sobre el eje  $x_i$ , es  $\overline{OP} l_i$  y la proyección de  $\overline{OP}$  sobre la recta con cosenos directores  $l'_i$ , representada por  $l'$ , se expresa a través de la suma de los términos  $\overline{OP} l_i l'_i$ . Por lo tanto:

$$\cos \varphi = \sum_{i=1}^n l_i l'_i$$

Esta fórmula corresponde a una generalización de los casos de 2 y 3 dimensiones.

Las rectas serán ortogonales si se cumple que:

$$\sum l l' = 0$$

Si dos líneas forman un ángulo  $\theta$  entre ellas, y estas se proyectan sobre un plano  $H$  arbitrario. Denotemos por  $\alpha$  y  $\beta$  a los ángulos entre las normales y las líneas, respectivamente. Entonces el ángulo entre las proyecciones, denotado por  $\varphi$ , estará dado por:

$$\cos \varphi = \frac{\cos \theta - \cos \alpha \cos \beta}{\sqrt{(1 - \cos^2 \alpha)(1 - \cos^2 \beta)}}$$

Tomemos en cuenta un hiperplano  $H$ , el cual sin pérdida de generalidad, podemos suponer que pasa por el origen, la ecuación de este plano la representaremos por:

$$\sum_{i=1}^n a_i x_i = 0$$

y consideremos además una recta que pasa por el origen cuya ecuación es:

$$\frac{x_1}{l_1} = \dots = \frac{x_n}{l_n} = \rho$$

La recta estará contenida en el hiperplano  $H$ , si y solo si:

$$\sum a_i p_i = \rho \sum a_i l_i = 0$$

De aquí que la recta que tenga cosenos directores proporcionales a las  $a$ 's, será ortogonal a cualquier recta en el hiperplano  $H$ .

La recta que es ortogonal a cualquier recta del hiperplano  $H$ , recibe el nombre de normal, ver sección 1.2.5.

Un hiperplano cualquiera, se puede representar en términos de su normal, de la siguiente forma:

$$\sum l_i x_i = p$$

donde las  $l$ 's son los cosenos directores de la normal, y  $p$  es la longitud de la perpendicular que va del origen al hiperplano.

Por definición tenemos que un ángulo entre dos planos, es un ángulo entre sus normales. Se sabe que: dos hiperplanos que se intersectan, tienen más de un ángulo entre ellos.

Si se fijara un plano y además un punto sobre él, se puede variar infinidad de veces otro plano sobre este punto.

Más aún, si se fijara un ángulo con el primer plano, no se podría determinar el segundo plano completamente.

Ejemplo

El ángulo entre los planos

$$4x + 3y + z = 0$$

$$x + y - z = 15$$

se calcula de la siguiente forma:



Las normales a dichos planos, son:

$$n_1 = (4, 3, 1) \text{ y } n_2 = (1, 1, -1)$$

si  $\theta$  es un ángulo entre los dos planos, entonces:

$$\cos \theta = \frac{n_1 \cdot n_2}{\|n_1\| \|n_2\|} = \frac{4 \cdot 1 + 3 \cdot 1 + 1 \cdot (-1)}{\sqrt{4^2 + 3^2 + 1^2} \sqrt{1^2 + 1^2 + 1^2}} = \frac{6}{\sqrt{26} \sqrt{3}} = 0.6794$$

de donde  $\theta = 47^\circ 12' \text{ o } 312^\circ 48'$

En forma general tendremos que, el ángulo entre

$$\sum l_i x_i = p$$

y

$$\sum l'_i x_i = p'$$

está dado por:  $\arccos(\sum l_i l'_i)$

ya que:

$$\frac{\sum l_i l'_i}{\sum l_i^2 \cdot \sum (l'_i)^2} = \frac{\sum l_i l'_i}{1 \cdot 1} = \sum l_i l'_i$$

En seguida vemos cómo calcular los cosenos directores entre espacios.

Sea  $\mathbb{R}^{n-p}$  un sub-espacio definido por  $p$  ecuaciones de la forma:  $a_{i1} x_1 + a_{i2} x_2 + \dots + a_{in} x_n = 0$   
 $i = 1, 2, \dots, p$

el cual en forma matricial se expresa como:

$$A \underline{x} = 0$$

donde  $A$  es una matriz de  $p \times n$  y  $\underline{x}$  un vector de  $n \times 1$ , por lo tanto la expresión de una recta sobre el plano es:

$$A L = 0 \quad \text{--- (4)}$$

$$\text{sujeta a } L' L = 1 \quad \text{--- (5)}$$

donde  $L$  es la matriz de cosenos directores  $l_i$ .

En forma similar consideremos un segundo sub-espacio  $\mathbb{R}^{n-9}$  determinado por 9 ecuaciones, con coeficientes expresados a través de la matriz  $B$ , la expresión de una recta sobre este plano está dada por:

$$BM = 0 \quad \text{--- (6)}$$

$$M'M = 1 \quad \text{--- (7)}$$

donde  $M$  es la matriz de cosenos directores  $m_i$ , los cuales determinan una recta arbitraria.

Así pues el ángulo entre una recta y otra es el ángulo  $\psi$  cuyo coseno es:

$$L'M = M'L = \cos \psi = R$$

para encontrar los valores de  $R$  que satisfacen las condiciones (4), (5), (6) y (7) se utilizan los multiplicadores de Lagrange  $\lambda_1, \lambda_2, \alpha_1, \alpha_2$ ; donde  $\lambda_1$  es un vector de  $(1 \times 6)$  y  $\lambda_2$  es un vector de  $(1 \times 9)$ .

Haciendo uso de los multiplicadores obtendremos la siguiente expresión:

$$f = L'M - \lambda_1 AL - \lambda_2 BM - \alpha_1 (LL' - 1) - \alpha_2 (M'M - 1)$$

entonces:

$$\frac{\partial f}{\partial l_i} = M' - \lambda_1 A - 2\alpha_1 L' = 0 \quad \text{--- (8)}$$

$$\frac{\partial f}{\partial m_i} = L' - \lambda_2 B - 2\alpha_2 M' = 0 \quad \text{--- (9)}$$

Si multiplicamos la expresión (8) por la derecha, por  $L$ , tendremos:

$$M'L - \lambda_1 AL - 2\alpha_1 L'L = 0$$

como queremos encontrar el valor de  $R$  que satisfaga las condiciones (4), (5), (6) y (7), entonces:

$$M'L - \lambda_1 \underbrace{AL}_0 - 2\alpha_1 \underbrace{L'L}_1 = 0$$

$$\Rightarrow M'L - 2\alpha_1 = 0$$

$$\Rightarrow M'L = 2\alpha_1 = R$$

en forma similar si se multiplica (9) por  $M$ , tendremos:

$$L'M - \lambda_2 \underbrace{BM}_0 - 2\alpha_2 \underbrace{M'M}_1 = 0$$

$$\Rightarrow L'M - 2\alpha_2 = 0$$

$$\Rightarrow L'M = 2\alpha_2 = R$$

Tomando en cuenta la expresión (8), obtendremos:

$$-\lambda_1 A = 2\alpha_1 L' - M'$$

y dado que:

$$M'L = 2\alpha_1 = R$$

entonces:

$$-\lambda_1 A = RL' - M' \quad \text{----- (10)}$$

Igualmente de la expresión (9) obtendremos:

$$-\lambda_2 B = 2\alpha_2 M' - L'$$

y dado que:

$$L'M = 2\alpha_2 = R$$

entonces:

$$-\lambda_2 B = RM' - L' \quad \text{----- (11)}$$

Si se multiplica (10) por  $A'$  o  $B'$  se tendrá:

$$-\lambda_1 AA' = RL'A - M'A$$

$$\Rightarrow -\lambda_1 AA' = -M'A \quad \text{----- (12)}$$

ya que  $L'A' = (AL)' = 0$

y multiplicando por  $B'$  tendremos:

$$-\lambda_1 AB' = RL'B - M'B'$$

$$\Rightarrow -\lambda_1 AB' = RL'B \quad \text{----- (13)}$$

ya que  $M'B' = (BM)' = 0$ .

De la misma forma si se multiplica (11) por  $B'$  se tendrá:

$$-\lambda_2 BB' = RMB' - L'B'$$

$$\Rightarrow -\lambda_2 BB' = -L'B' \quad \text{--- (14)}$$

y multiplicando por  $A'$

$$-\lambda_2 BA' = RM'A' - L'A'$$

$$\Rightarrow -\lambda_2 BA' = RM'A' \quad \text{--- (15)}$$

Utilizando las expresiones (12) y (15), tendremos

$$R\lambda_1 AA' + \lambda_2 BA' = 0 \quad \text{--- (16)}$$

ya que

$$-R\lambda_1 AA' - \lambda_2 BA' = -RM'A' + RM'A' = 0$$

Utilizando de igual forma operaciones algebraicas, obtendremos de (13) y (14) la expresión siguiente:

$$-\lambda_1 AB' + R\lambda_2 BB' = 0 \quad \text{--- (17)}$$

Redefinamos las expresiones (16) y (17) como sigue:

$$R\lambda_1 U + \lambda_2 V = 0$$

$$\lambda_1 V' + R\lambda_2 W = 0$$

donde:

$$U = AA' \quad (\text{matriz simétrica de } p \times p)$$

$$V = BA' \quad (\text{matriz de } q \times p)$$

$$V' = (BA')' \quad (\text{matriz de } p \times q)$$

$$y \quad W = BB' \quad (\text{matriz simétrica de } q \times q)$$

el determinante de este sistema de ecuaciones es:

$$\begin{vmatrix} RU' & V \\ V' & RW' \end{vmatrix} = \begin{vmatrix} RU & V' \\ V & RW \end{vmatrix} = R^2 UW - VV' = 0 \quad \text{--- (18)}$$

el cual se puede expresar también como:

$$\begin{array}{r} R^2 U \quad V' \\ V \quad W \end{array} = R^2 U W - V V' = 0$$

otra forma de expresarlo se obtiene con el siguiente desarrollo:

Multipliquemos  $R \lambda_1 U + \lambda_2 V = 0$  por  $R$  y  
 $\lambda_1 V' + R \lambda_2 W = 0$  por  $W^{-1} V$   
 restando ambas expresiones

$$\begin{array}{r} R^2 \lambda_1 U + R \lambda_2 V = 0 \\ \lambda_1 V' W^{-1} V + R \lambda_2 W W^{-1} V = 0 \end{array}$$

---


$$R^2 \lambda_1 U - \lambda_1 V' W^{-1} V = 0$$

$$\Rightarrow (R^2 U - V' W^{-1} V) \lambda_1 = 0 \quad \text{----- (19)}$$

de aquí que el determinante en (18) se pueda expresar como:

$$|R^2 U - V' W^{-1} V| = 0 \quad \text{----- (20)}$$

Para el desarrollo anterior se suponen las siguientes condiciones:

1.- Se supone que  $U$  y  $W$  tienen inversa, con lo cual se impide que las matrices  $A$  y  $B$  sean degeneradas, si esto sucede los espacios  $\mathbb{R}^{n-p}$  y  $\mathbb{R}^{n-q}$  serán de dimensión muy grande.

2.- Se ha supuesto que  $p \leq q$ .

Si  $p > q$  únicamente se invierte el papel de los dos espacios, el más pequeño entre los números  $p, q$  determina el número de ángulos mínimos no nulos entre los espacios.

3. Las rectas que corresponden a los ángulos mínimos, son las rectas ortogonales a los espacios respectivos.

Como  $V'W^{-1}V$  y  $U$  son matrices simétricas, también lo será  $U^{-1}V'W^{-1}V = K$ .

De (19) tendremos que:

$$\begin{aligned} R^2 U \lambda_1 &= V'W^{-1}V \lambda_1 \\ \Rightarrow R^2 \lambda_1 &= U^{-1}V'W^{-1}V \lambda_1 \\ \Rightarrow R^2 \lambda_1 &= K \lambda_1 \end{aligned} \quad \text{--- (21)}$$

Supongamos que  $R$  y  $S$  son dos raíces diferentes de (20) con vectores correspondientes  $\lambda_1, P_1$ , entonces

$$S^2 P_1 = K P_1 \quad \text{--- (22)}$$

multiplicando (21) y (22) por  $P_1'$  y  $\lambda_1'$  respectivamente:

$$\begin{aligned} P_1' K \lambda_1 &= R^2 P_1' \lambda_1 \\ \lambda_1' K P_1 &= S^2 \lambda_1' P_1 \end{aligned}$$

debido a que  $K$  es simétrica se cumple la siguiente igualdad:

$$\begin{aligned} P_1' K \lambda_1 &= \lambda_1' K P_1 \\ \Rightarrow (R^2 - S^2) P_1' \lambda_1 &= 0 \quad \text{como } R \neq S \\ \Rightarrow P_1' \text{ y } \lambda_1 &\text{ son vectores ortogonales.} \end{aligned}$$

Ejemplo:

Supongamos que tenemos 2 planos en  $\mathbb{R}^4$  determinados por:

$$\begin{aligned} x_1 + 7x_2 + x_3 &= 0 \\ x_4 &= 0 \end{aligned}$$

y

$$\begin{aligned} x_1 + x_2 &= 0 \\ x_3 + x_4 &= 0 \end{aligned}$$

donde:

$$A = \begin{pmatrix} 1 & 7 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

$$AA' = \begin{pmatrix} 1 & 7 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}_{2 \times 4} \begin{pmatrix} 1 & 0 \\ 7 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}_{4 \times 2} = \begin{pmatrix} 1+49+1 & 0 \\ 0 & 1 \end{pmatrix}_{2 \times 2} = \begin{pmatrix} 51 & 0 \\ 0 & 1 \end{pmatrix}$$

$$BB' = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1+1 & 0 \\ 0 & 1+1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$BA' = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 7 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 0 \\ 1 & 1 \end{pmatrix}$$

como:

$$\begin{aligned} U &= AA' \\ V &= BA' \\ V' &= (BA')' = AB' \\ W &= BB' \end{aligned}$$

entonces

$$\begin{vmatrix} R^2 U & V' \\ V & W \end{vmatrix} = \begin{vmatrix} 51R^2 & 0 & 8 & 1 \\ 0 & R^2 & 0 & 1 \\ 8 & 0 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{vmatrix} = 0$$

$$= 51R^2 \begin{vmatrix} R^2 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{vmatrix} + 8 \begin{vmatrix} 0 & 8 & 1 \\ R^2 & 0 & 1 \\ 1 & 0 & 2 \end{vmatrix} - 1 \begin{vmatrix} 0 & 8 & 1 \\ 0 & R^2 & 0 \\ 0 & 2 & 0 \end{vmatrix}$$

$$= 51R^2 (4R^2 - 2) + 8(-16R^2 + 8) - 1(2R^2)$$

$$= 4(51)R^4 - 2(51)R^2 - 128R^2 + 64 - 2R^2$$

$$= 4(51)R^4 - 102R^2 - 128R^2 + 64 - 2R^2$$

$$= 4(51)R^4 - 232R^2 + 64 = 0$$

$$\Rightarrow 51R^4 - 58R^2 + 16 = (17R^2 - 8)(3R^2 - 2) = 0$$

$$\Rightarrow R = \pm \sqrt{8/17} \quad \text{o} \quad \pm \sqrt{2/3}$$

otra forma de calcular los cosenos directores entre planos es la siguiente:

$$V'W^{-1}V = \begin{pmatrix} 8 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 8 & 0 \\ 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 8/2 & 1/2 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 8 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 65/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

$$R^2U = \begin{pmatrix} 51R^2 & 0 \\ 0 & R^2 \end{pmatrix}$$

$$R^2U - V'W^{-1}V = \begin{pmatrix} 51R^2 - 65/2 & -1/2 \\ -1/2 & R^2 - 1/2 \end{pmatrix}$$

de aquí que:

$$|R^2U - V'W^{-1}V| = \begin{vmatrix} 51R^2 - 65/2 & -1/2 \\ -1/2 & R^2 - 1/2 \end{vmatrix} = 0$$

$$= (51R^2 - 65/2)(R^2 - 1/2) - (-1/2)(-1/2)$$

$$= 51R^4 - \frac{51R}{2} - \frac{65}{2}R^2 + \frac{65}{4} - \frac{1}{4}$$

$$= 51R^4 - 58R^2 + 16$$

$$\Rightarrow (17R^2 - 8)(3R^2 - 2) = 0$$

$$\Rightarrow R = \pm \sqrt{8/17} \quad \text{o} \quad \pm \sqrt{2/3}$$



## 1.7

## TRANSFORMACIONES

## 1.7.1

## Transformación Ortogonal

Definición:

Una transformación lineal cuya matriz de transformación es  $T$  de un espacio vectorial Euclideo  $\mathbb{R}^n$ , es ortogonal si preserva la longitud de cada vector  $\xi$ , de manera que  $\|\xi T\| = \|\xi\|$ .

Ejemplo:

Se determinarán las transformaciones ortogonales  $Y = XL$  del plano Euclideo.

Si las transformaciones

$$(1, 0) \begin{pmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{pmatrix} = (l_{11}, l_{12}) \quad (0, 1) \begin{pmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{pmatrix} = (l_{21}, l_{22})$$

de los vectores unitarios  $(1, 0)$  y  $(0, 1)$  tienen longitud 1, para que se tenga una transformación ortogonal  $L$  tiene que ser ortogonal, es decir,  $L'L = I$  (donde  $L'$  representa la transpuesta de la matriz  $L$ ).

Que los vectores de transformación tengan longitud unitaria significa que:

$$\|(l_{11}, l_{12})\|^2 = l_{11}^2 + l_{12}^2 = 1 \quad \text{--- (23)}$$

$$\|(l_{21}, l_{22})\|^2 = l_{21}^2 + l_{22}^2 = 1$$

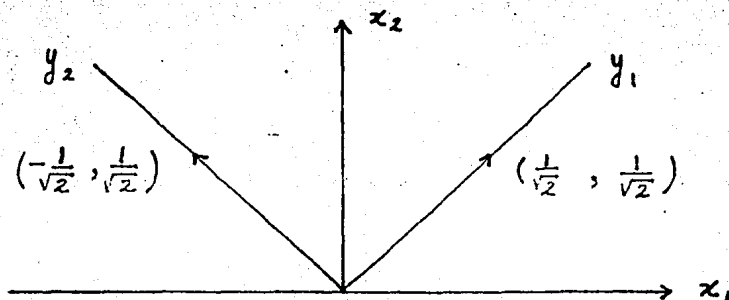


fig. 1.16 Transformación ortogonal en 2 dimensiones.

Para el vector  $(1, 1)$  su correspondiente vector de transformación será  $(l_{11} + l_{21}, l_{12} + l_{22})$  de longitud  $\sqrt{2}$ , por lo tanto:

$$(l_{11} + l_{21})^2 + (l_{12} + l_{22})^2 = 2$$

Utilizando el resultado (23) tendremos que:

$$(l_{11}, l_{12}) \cdot (l_{21}, l_{22}) = l_{11}l_{21} + l_{12}l_{22} = 0 \quad \text{--- (24)}$$

Asociando un ángulo  $\theta$  al resultado (23), se obtendrán las siguientes igualdades:

$$\begin{aligned} \cos \theta &= l_{11} \\ \sin \theta &= l_{12} \end{aligned}$$

Del resultado (24) se deduce que  $\theta = l_{12}/l_{11} = -l_{21}/l_{22}$  y regresando a (23) tendremos que  $l_{22} = \pm \cos \theta$ ,  $l_{21} = \pm \sin \theta$ .

Nota 1:

Si se considera la Transformación

$$y_i = \sum_{j=1}^n l_{ij} x_j + a_i \quad i = 1, 2, \dots, n$$

$a_i$  representa una traslación y los coeficientes  $l_{ij}$  una rotación.

Los 2 cambios de signo en  $l_{22}$  y  $l_{21}$  dan exactamente las 2 matrices:

$$\begin{pmatrix} \cos \theta & \operatorname{sen} \theta \\ -\operatorname{sen} \theta & \cos \theta \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} \cos \theta & \operatorname{sen} \theta \\ \operatorname{sen} \theta & -\cos \theta \end{pmatrix} \quad \text{----- (25)}$$

estas matrices representan una transformación por medio de un ángulo  $\theta$ , la primera matriz es una transformación en el sentido de rotación de las manecillas de un reloj y la segunda en sentido contrario.

Por lo tanto cada transformación ortogonal del plano es una rotación, que puede obtenerse con el producto de una matriz, la matriz asociada a la transformación ortogonal.

**Teorema:**

Una matriz real de  $n \times n$  representa una transformación ortogonal si y solo si  $LL' = I$  (esto es,  $L$  es una matriz ortogonal).

Si  $L$  es una matriz ortogonal entonces

$$\begin{aligned} \sum_{j=1}^n l_{ij} l_{kj} &= 1 \quad \text{si } i=k \\ \sum_{j=1}^n l_{ij} l_{ks} &= 0 \quad \text{si } i \neq k \end{aligned} \quad \text{----- (26)}$$

Estas ecuaciones son exactamente iguales a las encontradas en (23) y (24) para una matriz de  $2 \times 2$ .

Si denotamos a  $L_i$  como el  $i$ -ésimo renglón de la matriz  $L$  y a  $L'_i$  como su transpuesto, entonces (26) se podrá escribir de la siguiente forma matricial:

$$\begin{aligned} L_i L_i &= 1 \\ L_i L'_j &= 0 \quad i \neq j \end{aligned} \quad \text{----- (27)}$$

Las ecuaciones en (27) pueden expresarse como:

$$L_i L_j = \delta_{ij}$$

donde  $\delta_{ij}$  representa el elemento en el  $i$ -ésimo renglón y la  $j$ -ésima columna de la matriz identidad  $I = \|\delta_{ij}\|$  con entradas en la diagonal  $\delta_{ii} = 1$  y sus entradas restantes iguales a 0, ( $\delta_{ij}$  se llamada la delta de Kronecker).

Entre las transformaciones ortogonales que son más utilizadas está la Transformación de Helmert.

Esta transformación consiste en expresar las  $y$ 's en términos de las  $x$ 's de la siguiente forma:

$$y_1 = \frac{1}{\sqrt{2}} (x_1 - x_2)$$

$$y_2 = \frac{1}{\sqrt{6}} (x_1 + x_2 - 2x_3)$$

$$y_3 = \frac{1}{\sqrt{12}} (x_1 + x_2 + x_3 - 3x_4)$$

$$\vdots$$

$$y_{n-1} = \frac{1}{\sqrt{n(n-1)}} \{x_1 + x_2 + \dots + x_{n-1} - (n-1)x_n\}$$

$$y_n = \frac{1}{\sqrt{n}} (x_1 + x_2 + \dots + x_n)$$

Su matriz de transformación correspondiente es:

$$L = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \cdot & \cdot & \cdot & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \cdot & \cdot & \cdot & 0 \\ \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{3}{\sqrt{12}} & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -\frac{1}{\sqrt{n(n-1)}} & -\frac{1}{\sqrt{n(n-1)}} & \cdot & \cdot & \cdot & \cdot & \frac{n-1}{\sqrt{n(n-1)}} \end{bmatrix}$$

Despejando las  $x$ 's obtenemos:

$$\begin{aligned} x_1 &= 1/\sqrt{2} y_1 + 1/\sqrt{6} y_2 + \dots + 1/\sqrt{n(n-1)} y_{n-1} + 1/\sqrt{n} y_n \\ x_2 &= 1/\sqrt{2} y_1 + 1/\sqrt{6} y_2 + \dots + 1/\sqrt{n(n-1)} y_{n-1} + 1/\sqrt{n} y_n \\ x_3 &= -2/\sqrt{6} y_2 + \dots + 1/\sqrt{n(n-1)} y_{n-1} + 1/\sqrt{n} y_n \\ &\vdots \\ x_{n-1} &= -\frac{n-2}{\sqrt{(n-1)(n-2)}} y_{n-2} + 1/\sqrt{n(n-1)} y_{n-1} + 1/\sqrt{n} y_n \\ x_n &= -1/\sqrt{n(n-1)} y_{n-1} + 1/\sqrt{n} y_n \end{aligned}$$

En base a estas expresiones se puede ver que:

$$\sum_{i=1}^n x_i = \frac{n}{\sqrt{n}} y_n$$

$$\Rightarrow \bar{x} = \frac{\sum x_i}{n} = \frac{1}{\sqrt{n}} y_n$$

por lo tanto las variables  $x_1 - \bar{x}$ ,  $x_2 - \bar{x}$ ,  $\dots$ ,  $x_{n-1} - \bar{x}$ , dependerán solo de  $y_1, y_2, \dots, y_{n-1}$ .

Consideremos una muestra  $x_1, \dots, x_n$  de una distribución normal con media  $\mu$  y varianza  $\sigma^2$ .

Si aplicamos la transformación de Helmert, el término  $y_n$  será igual a la media multiplicada por una constante, y los términos anteriores  $y_1, \dots, y_{n-1}$  expresarán simples diferencias a través de las  $x$ 's, por lo que será fácil calcular cualquier diferencia con respecto a las  $x$ 's conociendo los valores  $y_1, \dots, y_{n-1}$ .

Si tomamos en cuenta que:

$$y_n^2 = n \bar{x}^2 \quad \text{--- (28)}$$

y además que debido a la ortogonalidad de la transformación podemos obtener la siguiente expresión:

$$\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2 \quad \text{--- (29)}$$

entonces la media muestral  $\bar{x} = y_n / \sqrt{n}$  será una estadística independiente de la varianza muestral:

$$S_x^2 = (n-1)^{-1} \sum (x_i - \bar{x})^2$$

ya que restando las expresiones (28) y (29), la varianza muestral quedará expresada en términos de la  $\sum_{i=1}^n y_i$ .

$$\sum_{i=1}^n y_i^2 - y_n^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

$$\Rightarrow \sum_{i=1}^{n-1} y_i^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\therefore S_x^2 = (n-1)^{-1} (y_1^2 + \dots + y_{n-1}^2)$$

Para el caso bivariado se tendrá que, si se aplica a la forma:

$$\sum_{i=1}^n x_i^2 - 2\rho \sum_{i=1}^n x_i u_i + \sum_{i=1}^n u_i^2$$

$$= \sum (x - \bar{x})^2 - 2\rho \sum (x - \bar{x})(u - \bar{u}) + \sum (u - \bar{u})^2 + n\{x^2 - 2\rho\bar{x}\bar{u} + \bar{u}^2\}$$

una transformación de Helmerl de  $x$  a  $y$  y de  $u$  a  $v$ , el resultado será la suma de dos componentes, una forma cuadrática en  $y_1, \dots, y_{n-1}, v_1, \dots, v_{n-1}$  y otra en  $y_n$  y  $v_n$ .

Recordar que:

$$\begin{array}{l} y_n = \bar{x} \\ v_n = \bar{u} \end{array}$$

Nota 2:

$$\begin{aligned} \sum (x_i - \bar{x})^2 - n\bar{x}^2 &= \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2n\bar{x} \left( \frac{\sum x_i}{n} \right) + \sum_{i=1}^n \bar{x}^2 + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + 2n\bar{x}^2 = \sum_{i=1}^n x_i^2 \end{aligned}$$

En forma similar al caso univariado se tendrá que las medias  $\bar{x}$ ,  $\bar{u}$  se distribuyen independientemente de una forma cuadrática en  $x_i - \bar{x}$  y  $u_i - \bar{u}$ .

1.7.2

### Transformación por Coordenadas Polares.

En el caso de 2 dimensiones,  $x$  y  $y$  se expresan en función de  $\theta$  y  $r$  como sigue:

$$x = r \cos \theta$$

$$y = r \sin \theta$$

con  $x^2 + y^2 = r^2$ .

Las condiciones para que la transformación del plano  $r\theta$  al plano  $xy$  sea uno a uno, son que:

$$r > 0$$

$$\theta_0 \leq \theta < \theta_0 + 2\pi$$

El Jacobiano de esta transformación es:

$$J(r, \theta) = \frac{\partial(x, y)}{\partial(r, \theta)} = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix}$$

$$= \begin{vmatrix} \cos \theta & \sin \theta \\ -r \sin \theta & r \cos \theta \end{vmatrix} = r(\cos^2 \theta + \sin^2 \theta) = r$$

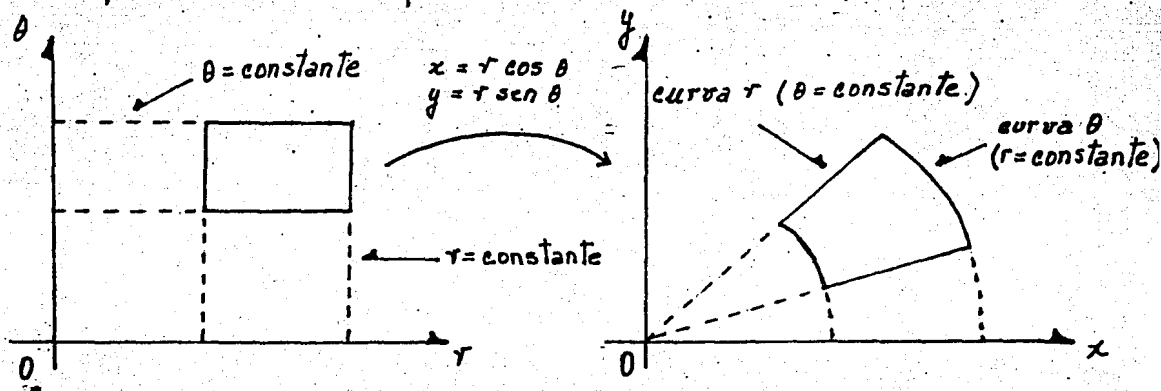


fig. 1.17

Las curvas  $r = \text{constante}$  son rectas desde el origen y las curvas  $\theta = \text{constante}$  son círculos concéntricos con el origen, al aplicar la transformación.

Para obtener el área del plano transformado en el plano  $x, y$ , hay que multiplicar el área del rectángulo en el plano  $r, \theta$ , por el Jacobiano.

De aquí que el Jacobiano puede considerarse como un "factor de ampliación" de áreas.

Con el caso de 3 dimensiones,  $x_1, x_2$  y  $x_3$  quedarán expresados de la siguiente forma:

$$\begin{aligned} x_1 &= r \cos \theta_1 \cos \theta_2 & 0 \leq r &\leq \infty \\ x_2 &= r \cos \theta_1 \sin \theta_2 & -\frac{1}{2}\pi \leq \theta_1 &\leq \frac{1}{2}\pi \\ x_3 &= r \sin \theta_1 & 0 \leq \theta_2 &\leq 2\pi \end{aligned}$$

con  $\sum_{i=1}^3 x_i^2 = r^2$ .

El Jacobiano de esta transformación es:

$$J = \frac{\partial(x_1, x_2, x_3)}{\partial(r, \theta_1, \theta_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_2}{\partial r} & \frac{\partial x_3}{\partial r} \\ \frac{\partial x_1}{\partial \theta_1} & \frac{\partial x_2}{\partial \theta_1} & \frac{\partial x_3}{\partial \theta_1} \\ \frac{\partial x_1}{\partial \theta_2} & \frac{\partial x_2}{\partial \theta_2} & \frac{\partial x_3}{\partial \theta_2} \end{vmatrix}$$

$$= \begin{vmatrix} \cos \theta_1 \cos \theta_2 & \cos \theta_1 \sin \theta_2 & \sin \theta_1 \\ -r \sin \theta_1 \cos \theta_2 & -r \sin \theta_1 \sin \theta_2 & r \cos \theta_1 \\ -r \cos \theta_1 \sin \theta_2 & r \cos \theta_1 \cos \theta_2 & 0 \end{vmatrix}$$

$$= r^2 \cos \theta_1$$



Geométicamente esta transformación en  $\mathbb{R}^3$ , representa esferas concéntricas en el espacio  $x_1, x_2, x_3$ .

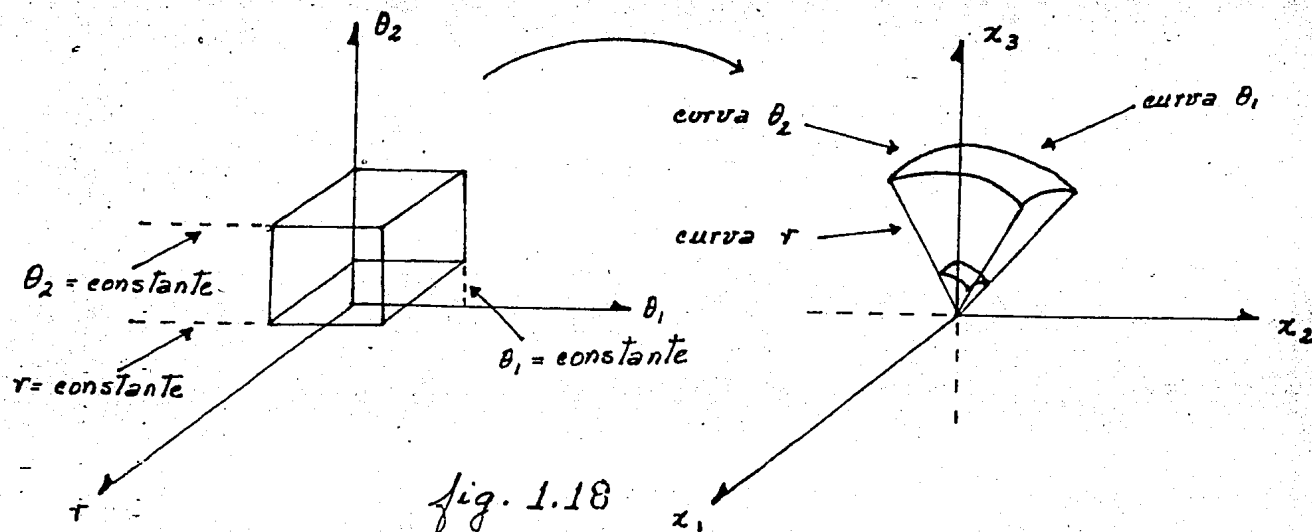


fig. 1.18

Para el caso de  $n$  dimensiones, consideramos:

$$\begin{aligned}
 x_1 &= r c_1 c_2 \cdots c_{n-2} c_{n-1} & 0 \leq \theta_{n-1} \leq 2\pi \\
 x_2 &= r c_1 c_2 \cdots c_{n-2} s_{n-1} & -\frac{1}{2}\pi \leq \theta_{n-2} \leq \frac{1}{2}\pi \\
 x_3 &= r c_1 c_2 \cdots s_{n-2} & -\frac{1}{2}\pi \leq \theta_{n-3} \leq \frac{1}{2}\pi \\
 &\vdots & \vdots \\
 x_j &= r c_1 \cdots c_{n-j} s_{n-j+1} & \vdots \\
 &\vdots & \vdots \\
 x_n &= r s_1 & -\frac{1}{2}\pi \leq \theta_1 \leq \frac{1}{2}\pi
 \end{aligned}$$

$$\text{con } \sum_{i=1}^n x_i^2 = r^2$$

$$\begin{aligned}
 \text{donde } c_i &= \cos \theta_i \\
 s_i &= \sin \theta_i \quad i=1, n-1
 \end{aligned}$$

El Jacobiano de esta transformación es:

$$J = r^{n-1} \begin{vmatrix} c_1 & \cdot & \cdot & \cdot & c_{n-1} & c_1 \cdot \cdot \cdot c_{n-2} s_{n-1} & \cdot \cdot \cdot & s_1 \\ -s_1 c_2 & \cdot & \cdot & \cdot & c_{n-1} & -s_1 \cdot \cdot \cdot c_{n-2} s_{n-1} & \cdot \cdot \cdot & c_1 \\ -c_1 s_2 & \cdot & \cdot & \cdot & c_{n-1} & -c_1 s_2 \cdot \cdot \cdot c_{n-2} s_{n-1} & \cdot \cdot \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -c_1 c_2 & \cdot & \cdot & \cdot & s_{n-1} & c_1 c_2 \cdot \cdot \cdot c_{n-1} & \cdot \cdot \cdot & 0 \end{vmatrix}$$

substituyendo los términos de senos y cosenos por tangente =  $t$  tendremos:

$$= r^{n-1} c_1^{n-1} c_2^{n-2} \cdot \cdot \cdot c_{n-1} s_2 \cdot \cdot \cdot s_{n-1} \begin{vmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 \\ -t_1 & -t_1 & -t_1 & \cdot & \cdot & \cdot & 1/t_1 \\ -t_2 & -t_2 & -t_2 & \cdot & \cdot & \cdot & 0 \\ \vdots & \vdots & \vdots & \cdot & \cdot & \cdot & \cdot \\ \vdots & \vdots & \vdots & \cdot & \cdot & \cdot & \cdot \\ -t_{n-1} & 1/t_{n-1} & 0 & \cdot & \cdot & \cdot & 0 \end{vmatrix}$$

Si se nota cada columna de su columna precedente, la expresión para el Jacobiano será:

$$J = r^{n-1} c_1^{n-2} c_2^{n-3} \cdot \cdot \cdot c_{n-2}$$

Geométricamente, a través de la transformación obtendremos hipersferas concéntricas en el espacio  $x_1, x_2, \dots, x_n$ .

Antes de presentar un ejemplo, veamos algo sobre cambio de variables en una integral doble.

En dos dimensiones una integral doble de la forma:

$$\iint_S f(x, y) dx dy$$

extendida a una región  $S$  en el plano  $xy$  se transforma en otra integral doble  $\iint_T f[x(u, v), y(u, v)] |J(u, v)| du dv$ ,

extendida a una nueva región  $T$  del plano  $uv$ .

Así pues la fórmula para la transformación de integrales dobles:

$$\iint_S f(x, y) dx dy = \iint_T f[X(u, v), Y(u, v)] |J(u, v)| du dv$$

es válida si  $X, Y, U, V$  son continuas, la aplicación de  $T$  en  $S$  es uno a uno, y el jacobiano  $J(u, v)$  es distinto de cero.

Las funciones  $X$  y  $Y$  relacionan  $x, y$  con  $u, v$  de la siguiente forma:  $x = X(u, v)$   $y = Y(u, v)$   
y las funciones  $U$  y  $V$  son las funciones inversas de  $X$  y  $Y$ .

Además consideremos el siguiente teorema:

**Teorema:**

Supongamos que  $(X, Y)$  es una variable aleatoria bidimensional continua con función de probabilidad conjunta  $f$ .

Sea  $u = U(x, y)$  y  $v = V(x, y)$  y supongamos que las funciones  $U$  y  $V$  satisfacen las condiciones siguientes:

- Las ecuaciones  $u = U(x, y)$  y  $v = V(x, y)$  se pueden resolver únicamente para  $x, y$  en función de  $u, v$ , digamos  $x = X(u, v)$  y  $y = Y(u, v)$ .
- Las derivadas parciales  $\partial X / \partial u$ ,  $\partial X / \partial v$ ,  $\partial Y / \partial u$  y  $\partial Y / \partial v$  existen y son continuas.

Luego la función de probabilidad conjunta de  $(u, v)$ , digamos  $K(u, v)$ , está dada por la expresión siguiente:

$$K(u, v) = f[X(u, v), Y(u, v)] |J(u, v)|$$

en donde  $J(u, v)$  es el siguiente determinante:

$$J(u, v) = \begin{vmatrix} \frac{\partial X}{\partial u} & \frac{\partial Y}{\partial u} \\ \frac{\partial X}{\partial v} & \frac{\partial Y}{\partial v} \end{vmatrix}$$

Ejemplo:

Tomemos como ejemplo una muestra de  $n$  variables aleatorias con distribución normal, media cero y varianza unitaria.

Considerando el Teorema y la aplicación sobre cambio de variables en una integral doble, vistos anteriormente, la expresión de la función de distribución siguiente

$$F = \frac{1}{\sqrt{2\pi}} \int \dots \int \exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2\right) dx_1 \dots dx_n$$

se transforma en:

$$F = \frac{1}{\sqrt{2\pi}} \int \dots \int \exp\left(-\frac{1}{2} r^2\right) r^{n-1} dr c_1^{n-2} c_2^{n-3} \dots c_{n-2} d\theta_1 \dots d\theta_{n-1} \dots (30)$$

De donde vemos que la distribución de  $r$ , es independiente de la distribución de los ángulos  $\theta_1, \dots, \theta_{n-1}$ , por lo tanto su función de distribución se expresa de la forma siguiente:

$$F = \frac{1}{\pi^{\frac{(n+1)}{2}} 2^{n/2}} \int \exp\left(-\frac{1}{2} r^2\right) r^{n-1} dr$$

la cual es una función de distribución correspondiente a una densidad  $\chi^2_{(n+1)}$ .

Otra consecuencia de la independencia de  $r$  y los ángulos  $\theta$  es la siguiente:

Supongamos que se tienen dos formas algebraicas homogéneas y del mismo grado en las  $x$ 's, digamos  $f_1, f_2$ , su razón  $t = f_1/f_2$  es, entonces, de grado cero en  $x$ .

Como  $r$  está en función de las  $x$ 's,  $t$  será independiente de  $r$ , y en virtud de (30), la esperanza de cualquier potencia de  $t$  es independiente de  $r$ .

Si  $f_2$  es una función de  $r$ , los momentos de  $t$  (y por tanto la distribución de  $t$ ) son independientes de  $r$ .

En el caso normal, con  $t$  y  $f_2$  independientes:

$$E(f_1^k) = E(t^k f_2^k) = E(t^k) E(f_2^k)$$

$$\Rightarrow E(t^k) = \frac{E(f_1^k)}{E(f_2^k)}$$

se obtiene una fórmula que nos permite calcular los momentos de la razón  $t$ .

## 1.8

## HIPERPLANOS TANGENTES

Utilizando el concepto de Variedad visto en la sección 1.4, se planteará en esta sección el concepto de hiperplanos tangentes.

Si tenemos  $n$  ecuaciones

$$x_i = f_i(t_1, \dots, t_k) \quad \text{--- (31)}$$

$i = 1, \dots, n; k < n$

podemos interpretar a  $t_1, \dots, t_k$  como parámetros y a las propias ecuaciones como ecuaciones paramétricas de una variedad de dimensión  $k$  en  $\mathbb{R}^n$ , donde las  $f_i$  son de clase  $C^1$ .

Dado que las  $f_i$  son de clase  $C^1$ , entonces podemos considerar diferenciales:

$$\begin{aligned} \partial x_1 &= \frac{\partial f_1}{\partial t_1} dt_1 + \dots + \frac{\partial f_1}{\partial t_k} dt_k \\ &\vdots \\ \partial x_n &= \frac{\partial f_n}{\partial t_1} dt_1 + \dots + \frac{\partial f_n}{\partial t_k} dt_k \end{aligned}$$

las cuales describen a la transformación lineal, cuya imagen es la variedad lineal de  $k$  dimensiones, tangente a la variedad de dimensión  $k$  (31) en un cierto punto.

Así pues si  $f(x_1, \dots, x_n) = 0$  determina una variedad  $H$  de  $n-1$  dimensiones y  $(x_1, \dots, x_n)$  es un punto sobre dicha variedad, entonces por el Teorema de Taylor, ver [2], tendremos que:

$$f(x_1, \dots, x_n) = f(x_1, \dots, x_n) + \sum \left( \frac{\partial f}{\partial x_i} \right)_{x_i} (x_i - x_i)$$

+ términos de orden superior en  $(x_i - x_i)$

En donde  $\sum_{i=1}^n (x_i - x_i) \left( \frac{\partial f}{\partial x_i} \right)_{x_i} = 0$ , esta ecuación que representa un hiperplano será tangente a  $H$  en el punto  $x$ .

La ecuación de la recta normal a  $H$  que pasa por el punto  $x$  estará dada por:

$$\frac{x_1 - x_1}{\left( \frac{\partial f}{\partial x_1} \right)_{x_1}} = \frac{x_2 - x_2}{\left( \frac{\partial f}{\partial x_2} \right)_{x_2}} = \dots = \frac{x_n - x_n}{\left( \frac{\partial f}{\partial x_n} \right)_{x_n}}$$

Veamos ahora qué sucede en  $\mathbb{R}^3$ , utilizando el análisis anterior.

Sea  $x_0 = (x_0, y_0, z_0)$  un punto sobre una variedad de dimensión 2, la cual representa una superficie que denotaremos como:

$$S = \{ (x, y, z) \mid z = f(x, y) \}$$

El plano tangente a  $S$  en  $x_0$  tendrá la ecuación:

$$(x - x_0) \left( \frac{\partial f}{\partial x} \right)_{x_0} + (y - y_0) \left( \frac{\partial f}{\partial y} \right)_{y_0} - (z - z_0) = 0$$

donde las derivadas parciales están evaluadas en el punto  $(x_0, y_0)$ .

Como  $\nabla f(x_0, y_0, z_0) = \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, -1 \right)$ , el gradiente de  $f$  no puede ser cero y, por tanto, existe un plano tangente en todo punto de  $S$ .

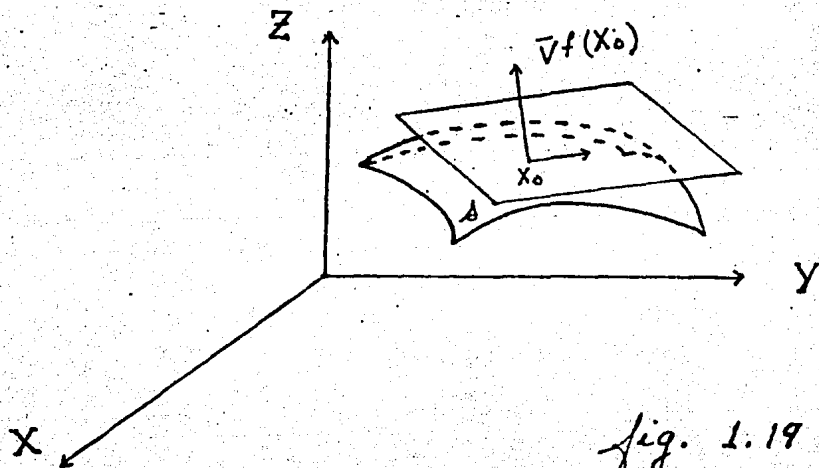


fig. 1.19

Ejemplo:

Determinaremos el plano tangente en el punto  $(1, 2, 6)$  a la superficie  $A$  de ecuación  $z = 2x^2 + y^2$ .

Haciendo  $f(x, y) = 2x^2 + y^2$  tendremos:

$$\left(\frac{\partial f}{\partial x}\right) = 4x \qquad \left(\frac{\partial f}{\partial y}\right) = 2y$$

$$\left(\frac{\partial f}{\partial x}\right)_{x_0} = 4 \qquad \left(\frac{\partial f}{\partial y}\right)_{y_0} = 4$$

Por lo tanto la ecuación del plano tangente queda expresada como:  $4(x-1) + 4(y-2) - (z-6) = 0$ .

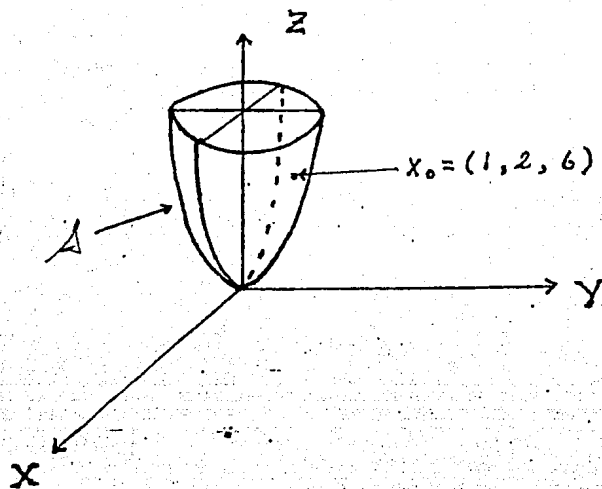


fig. 1.20 Gráfica de la superficie  $A$ .



## 1.9

## FORMAS CUADRATICAS

## Definición

Sea  $V$  un espacio euclideo real con una base ortonormal  $(e_1, \dots, e_n)$ , y sea  $A = (a_{ij})$  una matriz  $n \times n$  de escalares. La función de valores escalares  $Q$  definida en cada elemento  $x = \sum_{i=1}^n x_i e_i$  de  $V$  por la doble sumatoria:

$$Q(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \quad \text{-----} \quad (32).$$

se denomina la forma cuadrática asociada a  $A$ .

Si  $A$  es una matriz diagonal, entonces  $a_{ij} = 0$  si  $i \neq j$  de manera que la expresión (32) contendrá únicamente cuadrados y se podrá expresar de la siguiente forma:

$$Q(x) = \sum_{i=1}^n a_{ii} x_i^2$$

En este caso la forma cuadrática se llama forma diagonal.

La forma (32) se puede representar en forma matricial de la siguiente manera:

$$Q(x) = \underline{x}' A \underline{x}$$

donde  $A$  es la matriz con elementos  $(a_{ij})$  y  $\underline{x}$  es el vector columna  $(x_1, x_2, \dots, x_n)$ .

## 1.9.1

Reducción de una Forma Cuadrática real a una forma diagonal.

## Teorema:

Sea  $\underline{x}' A \underline{x}$  una forma cuadrática asociada a una ma-

triz simétrica real  $A$ , y sea  $L$  una matriz diagonal  
 $\Lambda = LAL'$ . Tenemos entonces:

$$\underline{X}'AX = \underline{Y}'\Lambda\underline{Y} = \sum_{i=1}^n \lambda_i y_i^2$$

donde  $\underline{Y}$  es el vector columna  $\underline{Y} = L\underline{X}$  y  $\lambda_1, \dots, \lambda_n$  son los valores propios de  $A$ .

Demostración:

Dado que  $L$  es ortogonal entonces  $L^{-1} = L'$  por consiguiente si  $\underline{Y} = L\underline{X}$  esto implica que  $\underline{X} = L'\underline{Y}$ , de esto obtenemos que:

$$\begin{aligned} X'AX &= (L'Y)'A(L'Y) \\ &= Y'LA L'Y \\ &= Y'\Lambda Y \end{aligned}$$

De esta forma la transformación lineal  $\underline{Y} = L\underline{X}$  reduce la forma cuadrática  $\underline{X}'AX$  a la forma diagonal  $\underline{Y}'\Lambda\underline{Y}$ .

Si el  $i$ -ésimo vector columna de  $L$  se expresa como  $l_i$  y dado que  $\Lambda$  es una matriz diagonal, entonces:

$$Al_i - \lambda_i l_i = 0$$

$$\therefore |A - \lambda_i I| = 0$$

esta ecuación es de grado  $n$  y sus raíces son los  $n$  valores de  $\lambda$  en  $\Lambda$ , para toda  $i$ .

Para el análisis anterior se deben de tomar en cuenta las siguientes consideraciones:

- a) Antes de obtener la forma canónica, es más conveniente quitar los términos lineales a través de una transformación por el origen.

La forma canónica que se obtendrá será la siguiente:

$$\sum_{i=1}^n \lambda_i y_i^2 = \text{constante}$$

- b) Si  $A$  es una matriz simétrica, entonces todas las raíces  $\lambda$  son reales.
- c) Si  $X'AX$  es una forma cuadrática positiva definida, entonces todas las raíces  $\lambda$  son positivas, por otro lado la forma canónica  $Y'AY$  deberá ser positiva o cero.
- d) Cuando algunas  $\lambda$ 's son cero, se está hablando de un caso degenerado.
- e) Si las  $\lambda$ 's son iguales la transformación no es única. Por ejemplo si  $\lambda_1$  y  $\lambda_2$  son iguales,  $l_1$  y  $l_2$  serán indeterminadas hasta que alguna condición sea impuesta.

### Ejemplo

Veamos cómo se puede reducir la forma cuadrática

$$Q(x) = 2x_1^2 + 4x_1x_2 + 5x_2^2$$

a una forma diagonal.

Si expresamos  $Q(x)$  en forma matricial, i.e.  $Q(x) = \bar{x}'A\bar{x}$ , entonces:

$$A = \begin{vmatrix} 2 & 2 \\ 2 & 5 \end{vmatrix}$$

Esta matriz simétrica tiene los valores propios  $\lambda_1 = 1$ ,  $\lambda_2 = 6$  y un conjunto ortonormal de vectores propios  $u_1$ ,  $u_2$ , siendo  $u_1 = t(2, -1)$ ,  $u_2 = t(1, 2)$ , donde  $t = 1/\sqrt{5}$ .

Una matriz diagonalizante ortogonal es:

$$C = t \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$$

La forma diagonal correspondiente es:

$$Y'AY = \lambda_1 y_1^2 + \lambda_2 y_2^2$$

$$= y_1^2 + 6y_2^2$$

La transformación lineal  $Y = LX$  representa una rotación que aplica la base  $i, j$  en la nueva  $u_1, u_2$ .

La ecuación  $y_1^2 + 6y_2^2 = C$ , es la ecuación cartesiana de una elipse si  $C > 0$ .

La interpretación geométrica de este ejemplo se da en la siguiente representación gráfica.

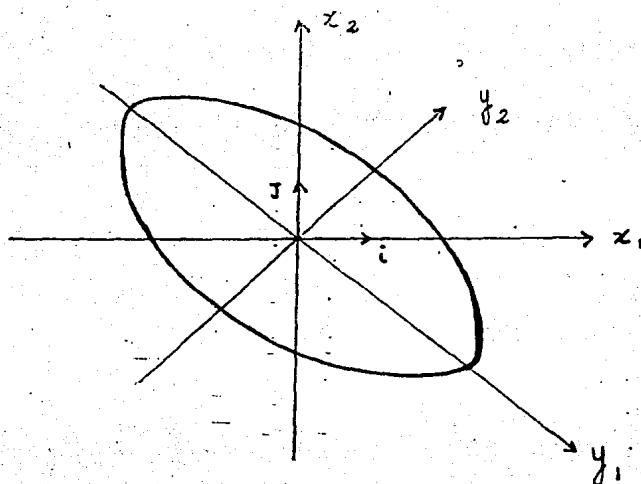


fig. 1.21

Así pues  $2x_1^2 + 4x_1x_2 + 5x_2^2 = C$  representa la misma elipse pero en el sistema coordenado original.

Consideremos la ecuación  $\sum_{i=1}^n \lambda_i y_i^2 = C$ , si todas las  $\lambda_i$  son positivas obtendremos una hiper-elipsoide, si algunas  $\lambda_i$  son negativas una hiper-hiperboloide y una "simetría circular" cuando las  $\lambda_i$  son iguales a cero.

Tomemos en cuenta una transformación de la siguiente forma  $z_i = y_i \sqrt{\lambda_i}$ , si  $\sum_{i=1}^n z_i^2 = C$ , esta ecuación corresponde a la de una hiper-esfera.  $z_i = y_i \sqrt{\lambda_i}$  será invariante bajo cualquier rotación de ejes.

Si se tiene 2 formas cuadráticas

$$\sum_{i,j}^n a_{ij} X_i X_j = Q_1$$

$$\sum_{i,j}^n b_{ij} X_i X_j = Q_2$$

existe una transformación lineal de las coordenadas la cual las reduce simultáneamente a una forma canónica.

Supongamos que esa transformación está dada por:

$$X = C'Y$$

entonces las formas cuadráticas se transforman en:

$$Y' C A C' Y = Q_1$$

$$Y' C B C' Y = Q_2 \quad \text{--- (33)}$$

donde

$$C A C' = \Lambda$$

$$C B C' = \Omega$$

son matrices diagonales, entonces (33) queda expresado como:

$$Y' \Lambda Y = Q_1$$

$$Y' \Omega Y = Q_2$$

si expresamos a  $V$  como  $\Lambda \Omega^{-1}$ , esta será también una matriz diagonal.

$$\begin{aligned} V &= \Lambda \Omega^{-1} = C A C' (C B C')^{-1} \\ &= C A C' (C')^{-1} B^{-1} C^{-1} \\ &= C A B^{-1} C^{-1} \end{aligned}$$

Sea  $D = A B^{-1}$  entonces:

$$\begin{aligned} V &= C D C^{-1} \Rightarrow V C = C D \\ &\Rightarrow C D - V C = 0 \\ &\Rightarrow (D - V) C = 0 \end{aligned}$$

Dadas  $A, B, D$  podemos encontrar las raíces en  $V$  resolviendo la siguiente ecuación:

$$\det (A B^{-1} - V I) = 0$$

la cual es equivalente a:

$$\det (A - V B) = 0$$

1.10

## FIGURAS GEOMETRICAS SIMPLS N-DIMENSIONALES.

Un politopo en  $\mathbb{R}^n$  es el interior de una figura acotada por un conjunto no vacío de hiperplanos que se intersectan entre sí.

Por ejemplo, como se puede observar en la figura 1.22, el conjunto de planos que acotan al cubo en los bordes se intersectarán en líneas y en los vértices en puntos.

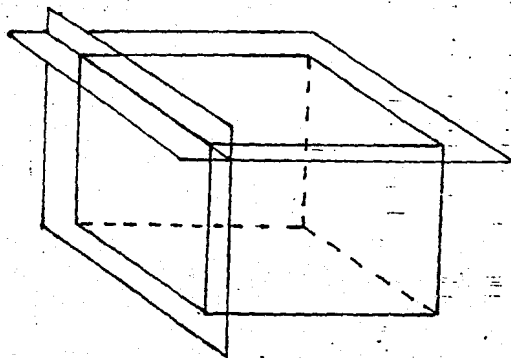


fig. 1.22

Los hiperplanos que acotan a una figura en  $\mathbb{R}^n$ , reciben el nombre de caras, y el menor número de hiperplanos que pueden encerrar un espacio y formar un politopo, es  $n+1$ .

A un politopo en  $\mathbb{R}^n$  acotado por  $n+1$  hiperplanos se le llama simplejo, tenemos como ejemplos un triángulo en 2 dimensiones y un tetraedro en 3 dimensiones, como lo muestra la figura 1.23, el tetraedro tiene 4 caras, 6 lados y 4 vértices.

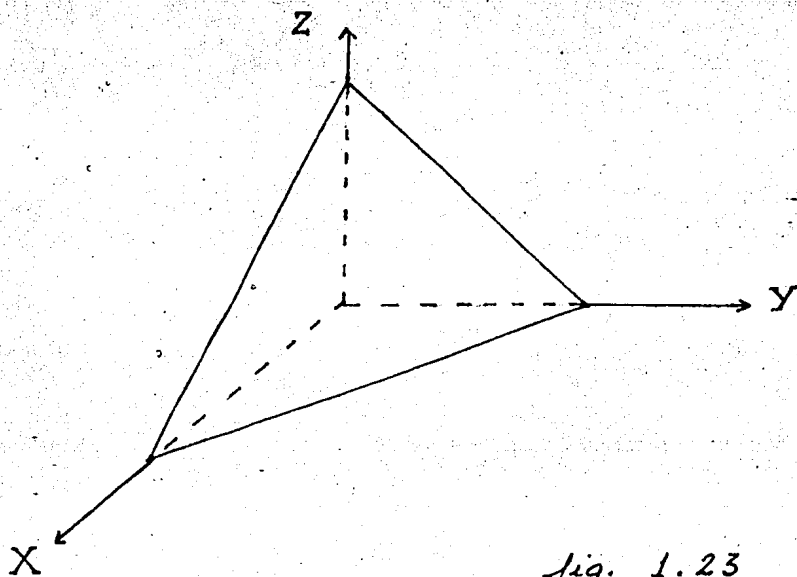


fig. 1.23

Ejemplos

2) Un simplejo en  $\mathbb{R}^4$ , tendrá como caras 5 espacios de dimensión 3, contará con 10 hiperplanos de dimensión 2, 10 hiperplanos de dimensión 1 y 5 hiperplanos de dimensión 0.

La regla general para calcular el número de hiperplanos de dimensión  $n-1, n-2, n-3, \dots$ , es que este número estará dado por términos sucesivos del desarrollo del binomio.

$$(1+1)^{n+1} = 2^{n+1} = 2^m = 1^m + \frac{m}{1} 1 \cdot 1^{m-1} + \frac{m(m-1)}{1 \cdot 2} 1^2 \cdot 1^{m-2} + \dots + 1^m$$

omitiendo el primero y último término que son iguales a 1, ver [3].

Por consiguiente en  $\mathbb{R}^4$  tendremos el siguiente desarrollo:

$$\begin{aligned} 2^{n+1} &= 2^5 = 1^5 + 5 + \frac{5(5-1)}{1 \cdot 2} + \frac{5(5-1)(5-2)}{1 \cdot 2 \cdot 3} + \frac{5(5-1)(5-2)(5-3)}{1 \cdot 2 \cdot 3 \cdot 4} \\ &+ \frac{5(5-1)(5-2)(5-3)(5-4)}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} \\ &= 1 + 5 + 10 + 10 + 5 + 1 \end{aligned}$$

Donde el número 5 corresponde al número de hiperplanos de dimensión  $n-1=3$ ; el 10 al número de hiperplanos de dimensión  $n-2=2$ , los cuales se encuentran en pares a razón de  $\binom{5}{2}=10$ , el siguiente término corresponde al número de hiperplanos de dimensión  $n-3=1$ , estos se encontrarán en tercias a razón de  $\binom{5}{3}=10$  y por último tenemos el número de hiperplanos de dimensión 0, que se encuentran en conjuntos de cuatro  $\binom{5}{4}=5$  y que corresponden a los 0-vértices.

b) Otro caso particular de los politopos, es un paraleloto, el cual es la generalización en  $\mathbb{R}^n$  de un paralelogramo. El paraleloto se encontrará acotado por pares de hiperplanos paralelos.

En  $\mathbb{R}^3$  tendremos un paralelepípedo rectangular limitado por 6 rectángulos iguales dos a dos.

c) Si desplazamos un paralelogramo a lo largo de una línea como se muestra en la siguiente figura.

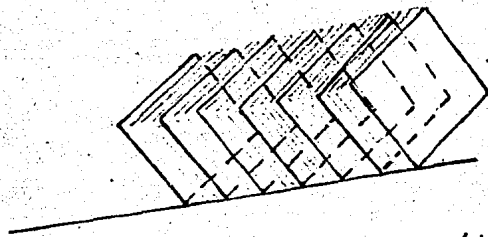


fig. 1.24

se genera un paralelepípedo en  $\mathbb{R}^3$ . Este proceso se puede generalizar para un espacio de  $n$  dimensiones, así tendremos que, un paraleloto que se encuentra en  $\mathbb{R}^{n-1}$ , si se desplaza a lo largo de una línea recta, generará un paraleloto en  $\mathbb{R}^n$ .



1.11

## CONTENIDO

El concepto de volumen se puede extender a ciertas clases de conjuntos (denominados conjuntos medibles) de  $\mathbb{R}^n$  de tal forma que si  $S$  es medible, su volumen es igual a la integral de la función constante igual a 1 en  $S$ .

Una forma de nombrar el volumen en  $\mathbb{R}^n$  es utilizando la palabra Contenido. Así tendremos que si  $C(S)$  representa el contenido de  $S$ , éste quedará expresado como:

$$C(S) = \int_S \dots \int dx_1 \dots dx_n$$

ver [4]

1.11.1

Contenido de un intervalo  $n$ -dimensional.

Si  $S$  es un intervalo en  $\mathbb{R}^n$ , por ejemplo:  
 $S = [a_1, b_1] \times \dots \times [a_n, b_n]$ , la integral múltiple para el cálculo de  $C(S)$  es el producto de  $n$  integrales uni-dimensionales.

$$C(S) = \int_{a_1}^{b_1} dx_1 \dots \int_{a_n}^{b_n} dx_n = (b_1 - a_1) \dots (b_n - a_n)$$

1.11.2

Contenido de una Hiper-esfera.

Consideremos la hiper-esfera  $S_n(a)$  de radio  $a$  determinada por:

$$S_n(a) = \{(x_1, \dots, x_n) \mid x_1^2 + \dots + x_n^2 \leq a^2\}$$

y sea

$$C_n(a) = \int_{S_n(a)} \dots \int dx_1 \dots dx_n$$

el contenido de  $S_n(a)$ .

Se puede demostrar, ver [4], que  $C_n(a) = \frac{\pi^{n/2}}{\Gamma(\frac{1}{2}n+1)} a^n$

y debido a que  $\Gamma(t+1) = t \Gamma(t)$ , entonces:

$$C_n(a) = 2 \frac{a^n}{n} \cdot \frac{\pi^{n/2}}{\Gamma(\frac{1}{2}n)}$$

Si  $n=1$ , como  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ , entonces:

$$C_1(a) = 2a \frac{\pi^{1/2}}{\sqrt{\pi}} = 2a$$

que es la longitud del intervalo  $[-a, a]$ .

Si  $n=2$ , como  $\Gamma(1) = 1$ , entonces:

$$C_2(a) = 2 \frac{a^2}{2} \pi = \pi a^2$$

que es el área de un círculo.

Si  $n=3$ , como  $\Gamma(\frac{3}{2}) = \frac{1}{2} \sqrt{\pi}$ , entonces:

$$C_3(a) = \frac{2a^3}{3} \cdot \frac{\pi^{3/2}}{\frac{1}{2} \sqrt{\pi}} = \frac{4\pi a^3}{3}$$

que es el volumen de una esfera.

Si se diferencia con respecto a  $a$ , la fórmula para obtener el contenido de la hiper-esfera, obtendremos el contenido de la superficie o área de la superficie de la hiper-esfera, el cual se denotará por  $CS$ :

$$CS = \frac{2a^{n-1} \pi^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)}$$

Por lo tanto si  $n=2 \Rightarrow CS = 2\pi a =$  circunferencia

si  $n=3 \Rightarrow CS = 4\pi a^2 =$  sup. de una esfera.

## 1.11.3

## Contenido de una Hiper-elipsoide

La ecuación de una hiper-elipsoide es:

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \dots + \frac{x_n^2}{a_n^2} = 1 \quad \text{--- (35)}$$

donde  $a_1, \dots, a_n$  representan los semi-ejes de la hiper-elipsoide.

Si se utiliza la siguiente transformación  $x_i = a_i y_i$ , el contenido de la región determinada por la ecuación (35), será igual a  $a_1 \dots a_n$  veces el contenido de la esfera unitaria.

Por lo tanto el contenido de la hiper-elipsoide se calculará por medio de la siguiente ecuación:

$$C = \frac{2a_1 \dots a_n}{n} \frac{\pi^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)}$$

## 1.11.4

## Contenido de un Hiper-prisma.

Un hiper-prisma es una figura generada por un polítopo de contenido  $C$  en  $\mathbb{R}^{n-1}$ , el cual se desplaza sobre una línea recta, que forma un ángulo  $\theta$  con respecto a  $\mathbb{R}^{n-1}$ .

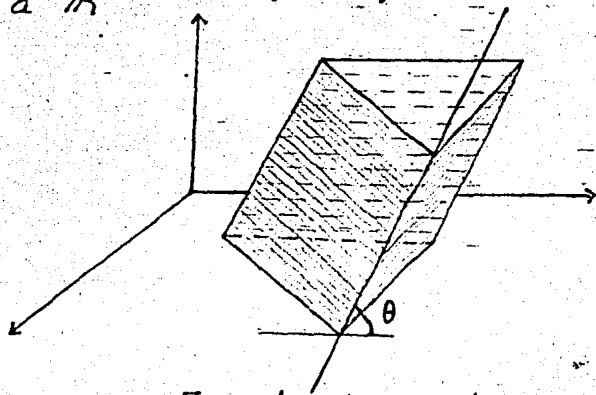


fig. 1.25 Ejemplo de un hiper-prisma en  $\mathbb{R}^3$

Si el politopo se desplaza una longitud  $h$ , entonces el contenido del hiper-prisma será  $C \cdot h \sin \theta$ .

### 1.11.5

Contenido de un cilindro de orden  $n$ .

Si en lugar de desplazar un politopo a lo largo de una línea recta, como en el punto anterior, se desplaza una hiper-esfera, se generará un cilindro de orden  $n$  cuyo contenido es:

$$C = \frac{2h a^{n-1} \pi^{\frac{1}{2}(n-1)}}{\Gamma[\frac{1}{2}(n-1)]}$$

### 1.11.6

Contenido de un Paralelotopo

Sean  $x_{ij}$ ,  $i, j = 1, 2, \dots, n$ , las coordenadas de los puntos  $P_1, \dots, P_n$  que se encuentran en  $\mathbb{R}^n$  y los cuales junto con el origen  $O$ , definen al paralelotopo.

Consideremos ahora la transformación

$$X = AY \quad \text{--- (36)}$$

el jacobiano asociado a dicha transformación será  $|A|$ . La matriz  $Y$  es ortogonal y la distancia entre el origen y cada uno de los puntos es unitaria.

Por lo tanto el contenido del paralelotopo será  $|A|$ , el cual solo resta expresarlo en términos de las  $x$ 's.

Pero de la expresión (36), tendremos que si las  $Y$ 's son unitarias, la matriz  $Y$  será la matriz identidad y  $X = A$ .

Por lo tanto el contenido del paralelotopo se obtendrá a través del determinante  $|X|$ .

Si el paraleloto po está definido por  $(n+1)$  puntos, excluyendo el origen, podemos trasladar el origen a alguno de los puntos, digamos  $x_{i,n+1}$ , de esta forma las coordenadas de los puntos restantes quedarán expresadas por  $x_{ij} - x_{i,n+1}$ .

Así pues el contenido del paraleloto po se obtendrá a través del determinante.

$$|x_{ij} - x_{i,n+1}| = \begin{vmatrix} x_{11} - x_{1,n+1} & x_{12} - x_{1,n+1} & \dots & x_{1n} - x_{1,n+1} \\ x_{21} - x_{2,n+1} & x_{22} - x_{2,n+1} & \dots & x_{2n} - x_{2,n+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - x_{n,n+1} & x_{n2} - x_{n,n+1} & \dots & x_{nn} - x_{n,n+1} \end{vmatrix}$$

$$= \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1n} & x_{1,n+1} \\ x_{21} & x_{22} & \dots & x_{2n} & x_{2,n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} & x_{n,n+1} \\ | & | & & | & | \end{vmatrix}$$

1.11.7

### Contenido de una Hiper-pirámide

La unión de los vértices de un polítopo en  $\mathbb{R}^{n-1}$  a un punto que no está en este espacio, genera una hiper-pirámide. Si el contenido de la base es  $C$ , y la altura es  $h$ , entonces, si un hiperplano paralelo a la base, corta a la pirámide generando de esta forma un polítopo de contenido:

$$\frac{C x^{n-1}}{h^{n-1}}$$

el contenido de la hiper-pirámide será:

$$\int_0^h \frac{x_n^{n-1}}{h^{n-1}} C dx_n = \frac{C}{h^{n-1}} \int_0^h x_n^{n-1} dx_n$$

$$= \frac{C}{h^{n-1}} \frac{x_n^n}{n} \Big|_0^h = \frac{Ch}{n}$$

1.11.8

### Contenido de un Simplejo

Ya se había comentado anteriormente, en la sección 1.10, que un simplejo es la generalización de un tetraedro, el cual es una pirámide con base triangular.

Así pues el simplejo es una hiper-pirámide con un punto  $P_n$  como vértice.

Si  $C_{n-1}$  es el contenido de la base, por el desarrollo de la sección 1.11.7, el contenido del simplejo será:

$$\frac{C_{n-1} h}{n}$$

Sea  $D_n$  el contenido del simplejo que corresponde a la transformación  $X = AY$  y con uno de sus vértices en  $P_n$ , entonces:

$$D_n = \frac{1}{n} D_{n-1}$$

por repetición del proceso se obtendrá:

$$D_n = \frac{1}{n} D_{n-1} = \frac{1}{n} \left( \frac{1}{n-1} D_{n-2} \right) = \frac{1}{n} \left( \frac{1}{n-1} \left( \frac{1}{n-2} D_{n-3} \right) \right) =$$

$$\dots = \frac{1}{n} \left( \frac{1}{n-1} \left( \frac{1}{n-2} \dots \left( \frac{1}{n-(n-1)} D_{n-n} \right) \right) \dots \right)$$

$$= \frac{1}{n!}$$

Por lo tanto el contenido del simplejo será  $\frac{1}{n!}$  veces el paralelepípedo correspondiente.

Si el paralelogramo está definido por  $(n+1)$  puntos, excluyendo al origen, el contenido será:

$$\frac{1}{n!} C \quad \text{ver [5]}$$

donde

$$C = \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1n} & x_{1,n+1} \\ x_{21} & x_{22} & & x_{2n} & x_{2,n+1} \\ \vdots & & & & \\ x_{n1} & x_{n2} & & x_{nn} & x_{n,n+1} \\ 1 & & & 1 & 1 \end{vmatrix}$$

El contenido se puede expresar también en términos de las longitudes de sus lados.

Sea  $D$  el contenido del simplejo, i.e.  $D = \frac{1}{n!} C$ , entonces:

$$n! D = \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1,n+1} \\ x_{21} & x_{22} & & x_{2,n+1} \\ \vdots & & & \\ x_{n1} & x_{n2} & & x_{n,n+1} \\ 1 & 1 & & 1 \end{vmatrix}$$

el cual se puede escribir como:

$$(-1)^n 2^n n! D = \begin{vmatrix} 1 & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i2}^2 & \dots & \sum_{i=1}^n x_{i,n+1}^2 \\ 0 & -2x_{11} & -2x_{12} & & -2x_{1,n+1} \\ 0 & -2x_{21} & -2x_{22} & & -2x_{2,n+1} \\ \vdots & & & & \\ 0 & -2x_{n1} & -2x_{n2} & & -2x_{n,n+1} \\ 0 & 1 & 1 & & 1 \end{vmatrix} \quad \text{--- (37)}$$

o de una forma equivalente

$$(-1)^{2n+1} n! D = \begin{vmatrix} 0 & 0 & 0 & \dots & 0 & 1 \\ 1 & x_{11} & x_{21} & & x_{n1} & \sum x_{i1}^2 \\ 1 & x_{12} & x_{22} & & x_{n2} & \sum x_{i2}^2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & x_{1,n+1} & x_{2,n+1} & & x_{n,n+1} & \sum x_{i,n+1}^2 \end{vmatrix} \quad \dots (38)$$

Multipliando (37) por (38) obtendremos la siguiente expresión:

$$(-1)^{n+1} 2^n (n!)^2 D^2 = \begin{vmatrix} 0 & 1 & 1 & \dots & 1 \\ 1 & 0 & a_{12}^2 & & a_{1,n+1}^2 \\ 1 & a_{21}^2 & 0 & & a_{2,n+1}^2 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & a_{n+1,1}^2 & a_{n+1,2}^2 & & 0 \end{vmatrix}$$

ya que  $\sum x_{ij}^2 - 2 \sum x_{ij} + \sum x_{ik}^2 = \sum (x_{ij} - x_{ik})^2 = a_{jk}^2$   $J \neq K$   
 $= 0$   $J = K$

donde  $a_{jk}$  es la longitud del lado determinado por  $x_j$  y  $x_k$ .

Para un simplejo regular, todas las  $a_{ij}$  son iguales a  $a$ , entonces:

$$(-1)^{n+1} 2^n (n!)^2 D^2 = \begin{vmatrix} 0 & 1 & 1 & \dots & 1 \\ 1 & 0 & a^2 & & a^2 \\ 1 & a^2 & 0 & & a^2 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & a^2 & a^2 & & 0 \end{vmatrix}$$

rotando la última columna a todas las demás exceptuando a la primera, tenemos:



$$(-1)^{n+1} 2^n (n!) D^2 = \begin{vmatrix} 0 & 0 & 0 & \dots & 1 \\ 1 & -a^2 & 0 & & a^2 \\ 1 & 0 & -a^2 & & a^2 \\ \vdots & & & & \\ \vdots & & & & \\ 1 & a^2 & a^2 & & 0 \end{vmatrix}$$

$$= (-1)^{n+1} a^{2n} = \begin{vmatrix} -1 & 0 & \dots & 1 \\ 0 & -1 & & 1 \\ \vdots & & & \\ \vdots & & & \\ 0 & 0 & & 1 \\ 1 & 1 & & 1 \end{vmatrix}$$

$$= (-1)^{n+1} (n+1) a^{2n}$$

Como  $(-1)^{n+1} 2^n (n!) D^2 = (-1)^{n+1} (n+1) a^{2n}$  entonces

$$D = \frac{a^n}{n!} \left( \frac{n+1}{2^n} \right)^{1/2}$$

$$\text{Si } n=3 \Rightarrow D = \frac{a^3}{6} \sqrt{2}$$

= volumen de un tetraedro regular

1.12

## Radio de Curvatura

Sea  $\alpha: I = (a, b) \rightarrow \mathbb{R}^3$  una curva parametrizada por la longitud de arco, como el vector tangente  $\alpha'(s)$  tiene longitud unitaria, la norma  $\|\alpha''(s)\|$  de la segunda derivada mide la razón de cambio del ángulo que las tangentes vecinas hacen con la tangente  $s$ . La norma de la segunda derivada da por lo tanto una medida de qué tan rápido la curva se aleja de la tangente de una vecindad de  $s$ .

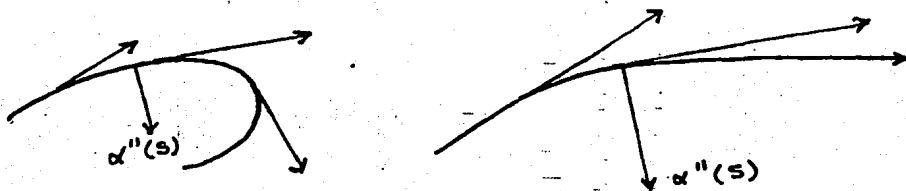


fig. 1.26

Esto sugiere la siguiente definición:

Definición:

Sea  $\alpha: I \rightarrow \mathbb{R}^3$  una curva parametrizada por la longitud de arco  $s \in I$ . El número  $\|\alpha''(s)\| = K(s)$  se llamará curvatura de  $\alpha(s)$ .

Si  $\alpha$  es una recta,  $\alpha(s) = us + v$  donde  $u$  y  $v$  son vectores constantes ( $\|u\| = 1$ ) entonces  $K \equiv 0$ , recíprocamente, si integramos  $K = \|\alpha''(s)\| \equiv 0$  obtendremos  $\alpha(s) = us + v$  lo que representa a una línea recta.

El radio de curvatura en  $s$  será igual a

$$R = 1/K$$

es decir, será el inverso de la curvatura.

Así pues un círculo de radio  $r$ , tiene radio de curvatura igual a  $r$ .

## CAPITULO 2

Después de los planteamientos geométricos, vistos en el capítulo anterior, empezaremos a analizar en forma gradual, las interpretaciones geométricas de la Estadística.

## 2.1

## INTERPRETACION GEOMETRICA DEL MODELO DE REGRESION LINEAL BIVARIADO.

Consideremos el modelo

$$Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon \quad \text{ver [8]}$$

donde  $Y$  es el vector columna de variables dependientes,  $X_1$  y  $X_2$  son vectores columna de variables independientes y  $\epsilon$  denota un vector aleatorio de error.

La regresión de este modelo se determina encontrando un vector fijo  $\hat{Y}$ , el cual es una combinación lineal de  $X_1$  y  $X_2$ , y de acuerdo al principio de mínimos cuadrados  $\hat{Y}$  se localiza a una distancia mínima de  $Y$ .

Si al espacio de estimación denotado por  $L$ , se le considera como el conjunto de todas las combinaciones lineales de  $X_1$  y  $X_2$ , entonces  $\hat{Y} \in L$ .

Geométricamente  $\hat{Y}$  representa la proyección perpendicular de  $Y$  sobre  $L$ .

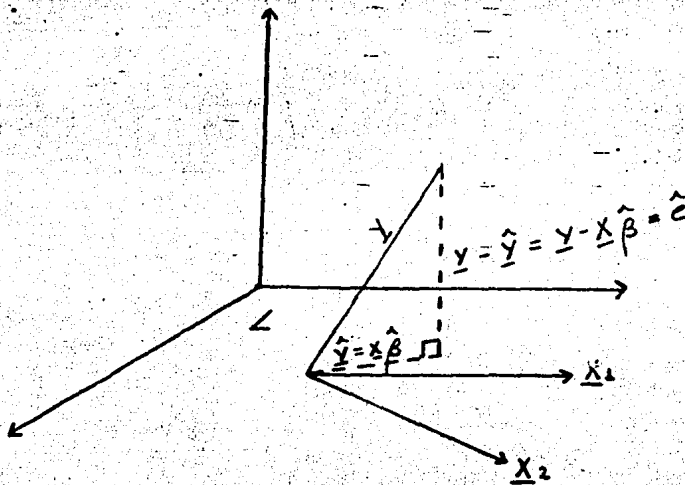


fig. 2.1

Por lo tanto  $\underline{y}$  se puede expresar de la siguiente forma

$$\underline{y} = \hat{\underline{y}} + \hat{\underline{e}}$$

donde  $\hat{\underline{e}} = \underline{y} - \hat{\underline{y}} = \underline{y} - X \hat{\underline{\beta}}$ . ----- (2.1)

De la ecuación normal de mínimos cuadrados ordinarios

$$X^T \underline{y} = X^T X \hat{\underline{\beta}} \quad \text{--- (2.2) ver [8]}$$

se determina el vector de coeficientes  $\hat{\underline{\beta}}$ , donde  $X$  es la matriz de vectores independientes. De esta ecuación se obtiene la siguiente expresión

$$\begin{aligned} X^T \underline{y} - X^T X \hat{\underline{\beta}} &= 0 \\ \Rightarrow X^T (\underline{y} - X \hat{\underline{\beta}}) &= X^T \hat{\underline{e}} = 0 \end{aligned}$$

lo que significa que el vector de error estimado es perpendicular a cada una de las columnas de  $X$ .

La proyección perpendicular de un vector de variables dependientes  $\underline{y}$  que se encuentra en un espacio muestral, sobre el espacio de estimación  $L$ , transforma a  $\underline{y}$  linealmente.

Sea  $\hat{\underline{y}} = P \underline{y}$  donde  $P$  es el operador de proyección perpendicular de  $\underline{y}$  sobre  $L$ . El operador  $P$  elimina al vector de error  $\hat{\underline{e}}$  componente de  $\underline{y}$  y determina únicamente al vector  $\hat{\underline{y}} \in L$ . El sub-espacio de error es ortogonal al espacio de estimación  $L$ , y se representa por  $L^T$ .

Dado que  $\underline{y}$  se puede expresar como:

$$\begin{aligned} \underline{y} &= \hat{\underline{y}} + \hat{\underline{e}} \\ \Rightarrow \underline{y} &= P \underline{y} + \hat{\underline{e}} \end{aligned}$$

donde  $P \underline{y} = \hat{\underline{y}} \in L$

entonces  $\hat{\underline{e}} = \underline{y} - P \underline{y} = (I - P) \underline{y} \in L^T$

Por otra parte tenemos de (2.1) que

$$\hat{\underline{y}} = X \hat{\underline{\beta}}$$

y de (2.2) que  $X \hat{\underline{\beta}} = X (X^T X)^{-1} X^T \underline{y} = P \underline{y}$

$$\Rightarrow P = X (X^T X)^{-1} X^T$$

Se denotará por  $P_x$  a la proyección perpendicular de cierto vector sobre el espacio  $L$ , generado por las columnas de la matriz  $X$ . Así pues, como se vio en el Capítulo 1 sección 1.14,  $P_x \underline{y}$  representará la proyección del vector  $\underline{y}$  sobre el

sub-espacio de estimación  $L$ .

Utilizando esta notación  $\underline{y}$  quedará expresado de la siguiente forma:

$$\underline{y} = \underline{\hat{y}} + \underline{\hat{e}} = P_x \underline{y} + (I - P_x) \underline{y}$$

la cual es una suma directa de dos proyecciones mutuamente perpendiculares  $P_x \underline{y}$  y  $(I - P_x) \underline{y}$ .

## 2.2

## INTERPRETACION GEOMETRICA DEL MODELO DE REGRESION LINEAL SIMPLE

Ahora analicemos el modelo

$$\underline{y} = \alpha \underline{1} + \beta \underline{x} + \underline{e}$$

Los vectores  $\underline{1}$  y  $\underline{x}$  generan el subespacio de estimación, cuyo operador de proyección perpendicular se denotará por  $P_{[\underline{1}, \underline{x}]}$ .  $[\underline{1}, \underline{x}]$  se refiere a la matriz cuyos vectores columna  $\underline{1}$  y  $\underline{x}$  generan el plano de estimación  $L$ .

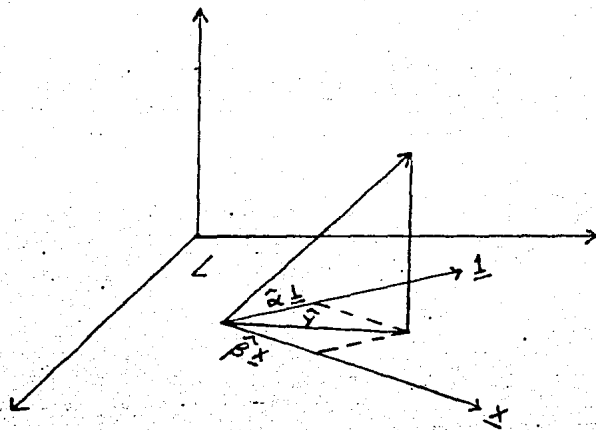


fig. 2.2

Analicemos que sucede cuando se aplican los operadores  $P_1$  y  $P_{[\underline{1}, \underline{x}]}$  a  $\underline{y} = \alpha \underline{1} + \beta \underline{x} + \underline{e}$ .

### Caso 1

Si aplicamos  $P_1$  a  $\underline{y}$  tendremos

$$\begin{aligned}
 P_1 \underline{y} &= P_1 (\alpha \underline{1} + \beta \underline{x} + \underline{e}) \\
 &= P_1 \alpha \underline{1} + P_1 \beta \underline{x} + P_1 \underline{e} \\
 &= \underline{1} (\underline{1}^T \underline{1})^{-1} \underline{1}^T \alpha \underline{1} + \underline{1} (\underline{1}^T \underline{1})^{-1} \underline{1}^T \beta \underline{x} + 0 \\
 &= \underline{1} \underline{1}^{-1} (\underline{1}^T)^{-1} \underline{1}^T \alpha \underline{1} + \underline{1} n^{-1} \underline{1}^T \beta \underline{x} \\
 &= \alpha \underline{1} + \underline{1} n^{-1} \beta \sum x_i \\
 &= \alpha \underline{1} + \beta \frac{\sum x_i}{n} \underline{1} \\
 &= (\alpha + \beta \bar{x}) \underline{1} \\
 &= \bar{y} \underline{1}
 \end{aligned}$$

Caso 2

Aplicuemos el operador  $P_{[1X]}$  al vector  $\underline{y}$ , siguiendo un procedimiento análogo al caso 1 obtendremos

$$P_{[1X]} \underline{y} = \hat{\alpha} \underline{1} + \hat{\beta} \underline{x} = \hat{\underline{y}}$$

el vector  $\hat{\underline{y}}$  representa un vector fijo en el plano de estimación  $L$ .

Vamos ahora qué sucede con el vector de error, para el cual tendremos que:  $(I - P_1) \underline{y} = \underline{y} - \bar{y} \underline{1}$ , esto representa la desviación muestral del vector de medias. La varianza de  $\underline{y}$  será igual al promedio de las diferencias al cuadrado.

Si aplicamos  $(I - P_{[1X]})$  a  $\underline{y}$  tendremos:

$$(I - P_{[1X]}) \underline{y} = \underline{y} - \hat{\underline{y}} = \underline{\underline{e}}$$

donde la varianza del residual será igual al promedio de la distancia al cuadrado de los residuales estimados.

Para obtener los estimadores de  $\alpha$  y  $\beta$ , supondremos que  $\hat{\beta}$  es conocido de tal forma que el modelo  $\underline{y} = \alpha \underline{1} + \hat{\beta} \underline{x} + \underline{e}$  quedará expresado como un modelo univariado, es decir:

$$\underline{y} - \hat{\beta} \underline{x} = \alpha \underline{1} + \underline{e}$$

Si el operador proyección  $P_1$ , se aplica a la expresión anterior, tendremos:

$$\begin{aligned} \hat{\alpha} \underline{1} &= P_1 (\underline{y} - \hat{\beta} \underline{x}) = (\bar{y} - \hat{\beta} \bar{x}) \underline{1} \\ \Rightarrow \hat{\alpha} \underline{1} &= \bar{y} \underline{1} - \hat{\beta} \bar{x} \underline{1} \end{aligned} \quad (2.3)$$

así pues tendremos que  $\hat{\alpha}$  es una combinación lineal de medias y por lo tanto  $\hat{\alpha}$  es una media.

Con seguida se obtendrá  $\hat{\beta}$ , es decir el estimador de  $\beta$ . Habíamos visto anteriormente que  $(I - P_1) \underline{y} = \underline{y} - \bar{y} \underline{1}$ , representa la desviación muestral del vector de medias. Si el operador  $(I - P_1)$  se aplica a la expresión del modelo lineal bivariado tendremos:

$$\begin{aligned} (I - P_1) \underline{y} &= (I - P_1) (\alpha \underline{1} + \beta \underline{x} + \underline{e}) \\ &= (I - P_1) \alpha \underline{1} + \beta (I - P_1) \underline{x} + I \underline{e} - P_1 \underline{e} \\ &= I \alpha \underline{1} - \alpha \underline{1} + \beta (\underline{x} - \bar{x} \underline{1}) + \underline{e} \\ &= \beta (\underline{x} - \bar{x} \underline{1}) + \underline{e} \end{aligned}$$

Si se aplica el operador perpendicular  $P_{[x-\bar{x} \perp]}$  a  $(y-\bar{y} \perp)$  por definición de proyección ortogonal tendremos:

$$P_{[x-\bar{x} \perp]} (y-\bar{y} \perp) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} (x - \bar{x} \perp)$$

$\therefore$  el estimador de  $\beta$  es:

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad \text{--- (2.4)}$$



## 2.3

## REPRESENTACION GEOMETRICA DE LA MEDIA

Sea

$$\underline{y} = \beta \underline{1} + \underline{\varepsilon}$$

la expresión de un modelo de regresión de una muestra de  $n$  observaciones sobre  $y$ , donde  $\beta$  representa un parámetro desconocido,  $\underline{1}$  un vector unitario  $n \times 1$  y  $\underline{\varepsilon}$  un vector aleatorio de error.

Si al modelo  $\underline{y} = \beta \underline{1} + \underline{\varepsilon}$ , aplicamos el operador de proyección  $P_1$ , obtendremos:

$$\begin{aligned} P_1 \underline{y} &= \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{n \times 1} \left[ (\underline{1}, \dots, \underline{1})_{1 \times n} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{n \times 1} \right]^{-1} (\underline{1}, \dots, \underline{1})_{1 \times n} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}_{n \times 1} \\ &= \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} n^{-1} \sum_{i=1}^n y_i = \frac{\sum_{i=1}^n y_i}{n} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= \bar{y} \underline{1} \end{aligned}$$

$$\therefore \hat{\beta} = \bar{y}$$

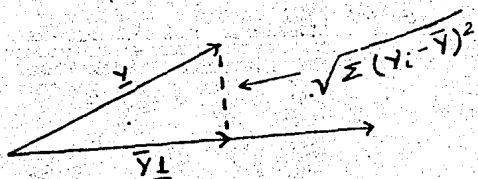


fig. 2.3

La  $\bar{y}$  es la proyección perpendicular de  $\underline{y}$  sobre la línea  $\beta \underline{1}$  y el sub-espacio de estimación generado por el vector unitario es una línea.

2.4

## REPRESENTACION GEOMETRICA DE LA VARIANZA

Si se aplica el operador proyección  $(I - P_1)$  al modelo

$$\underline{y} = \beta \underline{1} + \underline{\varepsilon}$$

$$\begin{aligned} \text{obtenemos: } (I - P_1) \underline{y} &= \underline{y} - P_1 \underline{y} \\ &= \underline{y} - \underline{1} \bar{y} \\ &= \underline{y} - \bar{y} \underline{1} \\ &= \underline{y} - \hat{y} \\ &= \hat{\underline{\varepsilon}} \end{aligned}$$

Por mínimos cuadrados

$$\begin{aligned} \hat{\underline{\varepsilon}}^T \hat{\underline{\varepsilon}} &= \underline{y}^T (I - P_1)^T (I - P_1) \underline{y} \\ &= (\underline{y} - \bar{y} \underline{1})^T (\underline{y} - \bar{y} \underline{1}) \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 \end{aligned}$$

Desde el punto de vista geométrico la suma de los errores al cuadrado es la distancia de  $y_i$  a  $\bar{y}$ , al cuadrado, sobre la línea definida por  $\beta \underline{1}$ .

Así pues la varianza de  $\underline{y}$  en términos geométricos es el promedio de la distancia al cuadrado del vector estimado de error  $\hat{\underline{\varepsilon}}$ , i.e.

$$\text{Var}_y = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$$

## 2.5

## COLINEALIDAD EN EL MODELO DE REGRESION LINEAL SIMPLE

Se dice que hay colinealidad perfecta entre dos vectores cuando el ángulo entre ellos es cero.

En el modelo bivariado se supone que el vector unitario y el vector  $\underline{x}$  son casi colineales.

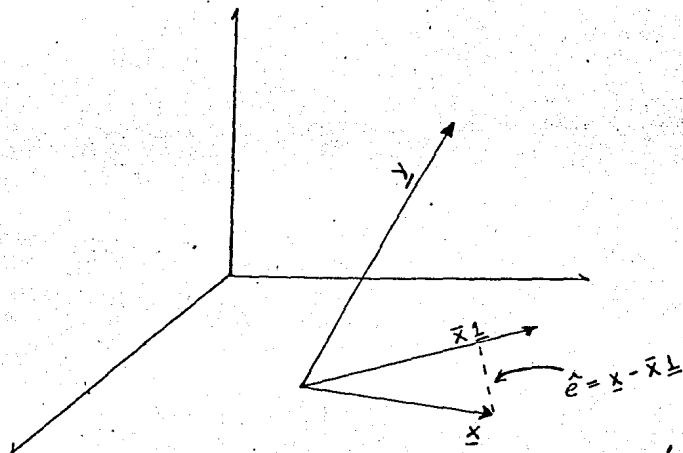


fig.2.4

Para medir la proximidad entre el vector  $\underline{1}$  y el vector  $\underline{x}$ , se utiliza el promedio de la distancia al cuadrado entre ellos.

Sin embargo también se puede utilizar el análisis geométrico de medias y varianzas, antes visto.

Como podemos ver en la figura anterior,  $\underline{x}$  se puede proyectar perpendicularmente sobre  $\underline{1}$  para así obtener  $\bar{x}_1$ , el vector de medias de  $\underline{x}$ .

El vector de error estimado para esta proyección es:

$$\hat{e} = (I - P_1) \underline{x} = \underline{x} - \bar{x}_1$$

donde el promedio de la distancia al cuadrado de  $\hat{e}$ , es la varianza de  $\underline{x}$ , i.e.  $Var_x = \frac{\sum (x_i - \bar{x})^2}{n}$ .

Un ángulo pequeño entre  $\underline{1}$  y  $\underline{x}$  corresponde a la varianza mínima de  $\underline{x}$ .

Como se había mencionado anteriormente, los vectores  $\underline{1}$  y  $\underline{X}$  generan al plano de estimación  $L$ , si dichos vectores son colineales el plano  $L$  estará mal definido. Así pues no se podrá llevar a cabo una proyección perpendicular del vector  $\underline{y}$  sobre  $L$ , en el caso de severa colinealidad entre los vectores regresores  $\underline{1}$  y  $\underline{X}$ .

Ejemplo:

Consideremos un modelo de regresión lineal bivariado

$$\underline{y} = \beta_1 \underline{X}_1 + \beta_2 \underline{X}_2 + \underline{e}$$

Supongamos que  $X_1 = 2X_2$ , i.e.  $X_1$  está en función de  $X_2$ , en este caso ocurre una colinealidad y el subespacio de estimación  $L$ , es una línea en lugar de un plano.

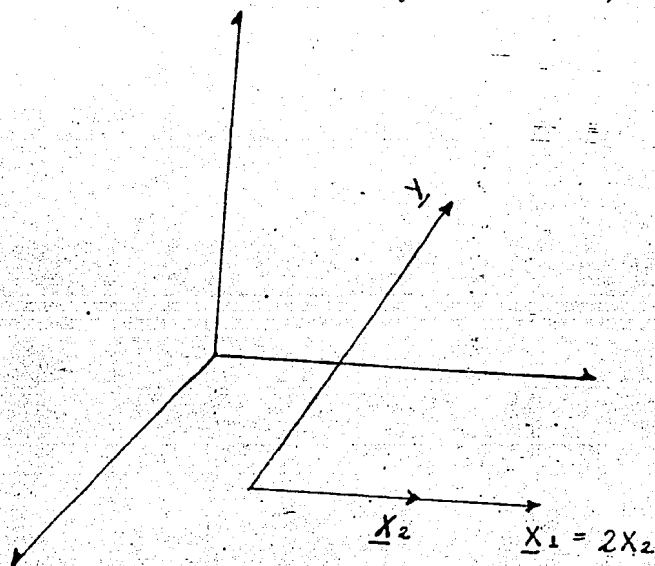


fig.2.5

## 2.6

## PRUEBAS DE HIPOTESIS

Consideremos el modelo de regresión lineal simple, con media  $E(\underline{Y}) = \alpha \underline{1} + \beta \underline{X}$  y varianza  $\sigma^2$ .

La hipótesis a probar es que  $E(\underline{Y})$  se encuentra en un sub-espacio unidimensional  $M$  generado por un vector unitario, de esta forma tendremos que la hipótesis nula a probar es  $H_0: \beta = 0$  contra la hipótesis alternativa  $H_a: \beta \neq 0$ .

Geométricamente la realización de esta prueba involucra dos proyecciones perpendiculares de  $\underline{Y}$ .

Bajo la hipótesis alternativa el vector  $\hat{\underline{Y}}$  se encontrará en el sub-espacio  $L$  generado por  $K=2$  vectores independientes,  $\underline{1}$  y  $\underline{X}$ .

Bajo la hipótesis nula el vector  $\hat{\underline{Y}}$  se encontrará en el sub-espacio  $M$  generado por el vector unitario que representa las  $K-1=1$  variables independientes.

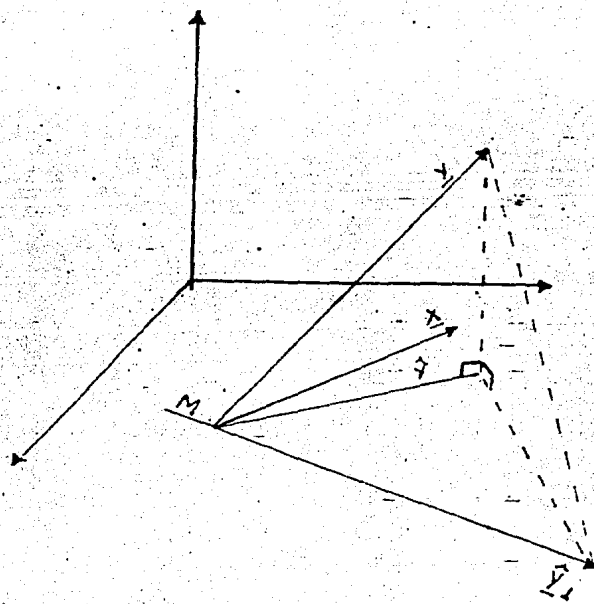


fig. 2.6

El vector  $\hat{\underline{Y}}$  es la proyección perpendicular de  $\underline{Y}$  sobre  $L$  y podemos denotar a  $\hat{\underline{Y}}_1$  como su proyección sobre  $M$ .

Dado que la distancia de  $\hat{Y}$  a  $Y$  y de  $\hat{Y}_1$  a  $Y$  es fija, entonces la cercanía de  $\hat{Y}_1$  a  $\hat{Y}$  define la veracidad o falsedad de la hipótesis nula, entonces mientras más cercano es  $\hat{Y}_1$  de  $\hat{Y}$  es más probable que sea cierta la hipótesis nula.

## 2.6.1

## ANÁLISIS DE LOS RESIDUALES BAJO LA HIPÓTESIS ALTERNATIVA

La ecuación normal sin restricciones se expresa de la siguiente forma:

$$[\mathbf{1} \ \mathbf{X}]^T [\mathbf{1} \ \mathbf{X}] \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = [\mathbf{1} \ \mathbf{X}]^T \underline{Y}$$

Por otra parte tenemos que:

$$\begin{aligned} \underline{Y}^T P_{[\mathbf{1} \ \mathbf{X}]} \underline{Y} &= \underline{Y}^T [\mathbf{1} \ \mathbf{X}] ([\mathbf{1} \ \mathbf{X}]^T [\mathbf{1} \ \mathbf{X}])^{-1} [\mathbf{1} \ \mathbf{X}]^T \underline{Y} \\ &= \underline{Y}^T [\mathbf{1} \ \mathbf{X}] \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \\ &= \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}^T [\mathbf{1} \ \mathbf{X}]^T \underline{Y} \end{aligned}$$

La suma de cuadrados de los residuales basada sobre la hipótesis de que ambos coeficientes no son cero, es:

$$\begin{aligned} \sum e_i^2 &= \underline{e}^T \underline{e} \\ &= (\underline{Y} - \hat{Y})^T (\underline{Y} - \hat{Y}) \\ &= (\underline{Y} - \mathbf{X} \hat{\beta})^T (\underline{Y} - \mathbf{X} \hat{\beta}) \end{aligned}$$

desarrollando, tendremos:

$$\begin{aligned} \sum e_i^2 &= \left\{ \underline{Y} - \mathbf{X} ([\mathbf{1} \ \mathbf{X}]^T [\mathbf{1} \ \mathbf{X}])^{-1} [\mathbf{1} \ \mathbf{X}]^T \underline{Y} \right\}^T \\ &\quad \cdot \left\{ \underline{Y} - \mathbf{X} ([\mathbf{1} \ \mathbf{X}]^T [\mathbf{1} \ \mathbf{X}])^{-1} [\mathbf{1} \ \mathbf{X}]^T \underline{Y} \right\} \\ &= \underline{Y}^T \left[ \mathbf{I} - [\mathbf{1} \ \mathbf{X}] ([\mathbf{1} \ \mathbf{X}]^T [\mathbf{1} \ \mathbf{X}])^{-1} [\mathbf{1} \ \mathbf{X}]^T \right] \\ &\quad \cdot \left[ \mathbf{I} - [\mathbf{1} \ \mathbf{X}] ([\mathbf{1} \ \mathbf{X}]^T [\mathbf{1} \ \mathbf{X}])^{-1} [\mathbf{1} \ \mathbf{X}]^T \right] \underline{Y} \end{aligned}$$

Dado que la matriz  $(I - P_{[1X]}) = \{I - [1X]([1X]^T[1X])^{-1}[1X]^T\}$  es una matriz idempotente (i.e.  $[I - P_{[1X]}]^2 = [I - P_{[1X]}]$ ), entonces:

$$\begin{aligned}\sum e_i^2 &= \underline{Y}^T [I - [1X]([1X]^T[1X])^{-1}[1X]^T] \underline{Y} \\ &= \underline{Y}^T (I - P_{[1X]}) \underline{Y}\end{aligned}$$

representa la suma de cuadrados de los residuales.

Se dice que la suma tiene  $n-2$  grados de libertad ya que  $(\frac{\sum e_i^2}{\sigma^2} \sim \chi^2_{(n-2)})$  [6] y está representada geométricamente como la distancia de  $\hat{\underline{Y}}$  a  $\underline{Y}$ .

### 2.6.2

#### ANÁLISIS DE LOS RESIDUALES BAJO LA HIPÓTESIS NULA

El modelo a considerar bajo la hipótesis nula es:

$$\underline{Y} = \alpha \underline{1} + \underline{e}$$

y su ecuación normal estará dada por:

$$\underline{1}^T \underline{1} \hat{\alpha} = \underline{1}^T \underline{Y} = \underline{Y}^T \underline{1}$$

de aquí tendremos que:

$$\hat{\alpha} = \underline{Y}^T \underline{1} [\underline{1}^T \underline{1}]^{-1} \quad \text{--- (2.5)}$$

como  $P_1 \underline{Y} = \underline{1} (\underline{1}^T \underline{1})^{-1} \underline{1}^T \underline{Y}$

entonces  $\underline{Y}^T P_1 \underline{Y} = \underline{Y}^T [\underline{1} (\underline{1}^T \underline{1})^{-1} \underline{1}^T] \underline{Y}$

de (2.5)

$$\underline{Y}^T P_1 \underline{Y} = \hat{\alpha} \underline{1}^T \underline{Y}$$

La suma de cuadrados de residuales basada en la hipótesis nula, i.e.  $\beta = 0$ , se expresa de la siguiente forma:

$$\begin{aligned}\sum e_i^2 &= \underline{e}^T \underline{e} \\ &= (\underline{Y} - \hat{\underline{Y}}_1)^T (\underline{Y} - \hat{\underline{Y}}_1) \\ &= \{ \underline{Y} - \underline{1} (\underline{1}^T \underline{1})^{-1} \underline{1}^T \underline{Y} \}^T \{ \underline{Y} - \underline{1} (\underline{1}^T \underline{1})^{-1} \underline{1}^T \underline{Y} \} \\ &= \underline{Y}^T [I - \underline{1} (\underline{1}^T \underline{1})^{-1} \underline{1}^T] [I - \underline{1} (\underline{1}^T \underline{1})^{-1} \underline{1}^T] \underline{Y} \\ &= \underline{Y}^T [I - \underline{1} (\underline{1}^T \underline{1})^{-1} \underline{1}^T] \underline{Y} \\ &= \underline{Y}^T [I - P_1] \underline{Y}\end{aligned}$$

2.7

## DISTRIBUCION NORMAL

La función de densidad de una distribución normal con media  $\mu$  y varianza  $\sigma^2$  se expresa de la siguiente forma:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

de aquí que si tenemos  $X_1, X_2, \dots, X_n$  variables aleatorias independientes e idénticamente distribuidas, i.e.:  $X_i \sim N(\mu, \sigma^2)$  entonces, su función de densidad conjunta  $f(x_1, \dots, x_n)$  se obtendrá de la siguiente manera:

$$\begin{aligned} f(x_1, \dots, x_n) &= f(x_1) f(x_2) \cdots f(x_n) \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_2-\mu)^2}{2\sigma^2}} \cdots \\ &\quad \cdots \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \end{aligned}$$

Debe hacer notar la importancia de esta función ya que es la base para poder resolver problemas de estimación, predicción, para obtener estadísticas suficientes, etc.

Esta función de densidad se puede expresar también como:

$$dF(x) = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\} dx_1 \cdots dx_n \quad (2.6)$$

Si consideramos la estadística  $T = T(x_1, \dots, x_n)$ , su función de distribución en  $T_0$  se obtendrá, por lo tanto, integrando (2.6) sobre la región  $T \leq T_0$ , siendo  $T$  un valor dado.

La frontera de esta región será la variedad  $V_{n-1}$  dada por  $T = T_0$  en  $\mathbb{R}^n$ .

Si realizamos la transformación  $x_i - \mu = y_i$ , la función de densidad  $f(y_1, \dots, y_n)$ , quedará expresada de la siguiente forma:

$$f(y_1, \dots, y_n) = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n y_i^2} |J|, \quad |J| = 1$$



si  $\sum_{i=1}^n y_i^2 = r^2$   
 entonces  $f(r) = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} e^{-\frac{r^2}{2\sigma^2}}$

Sea  $P$  el punto muestral con coordenadas  $(y_1, \dots, y_n)$  y  $O$  el origen, entonces  $\sum y_i^2$  representa el cuadrado de la distancia del punto  $P$  al origen, esta distancia no sufre ningún cambio con la rotación de los ejes coordenados.

Procedamos a obtener la función de distribución de la estadística  $r^2 = \sum y_i^2$ .

La densidad será constante sobre la superficie  $r^2 = \text{cte.}$ , i.e. sobre la superficie de una hiper-esfera de  $n$  dimensiones, ya que una hiper-esfera de radio  $a$  está determinada por  $\sum x_i^2 = a^2$ .

La función de distribución de  $z = r^2$  es la integral de esta densidad constante entre las hiper-esferas  $z$  y  $z + dz$ , considerando estas hiper-esferas concéntricas, por lo tanto la función de distribución es proporcional a  $\exp(-\frac{1}{2} r^2)$  veces el volumen de la hiper-esfera, la cual es proporcional a la  $n$ -ésima potencia del radio  $r$ .

Tendremos entonces que:

$$dF(z) \propto \exp\left(-\frac{1}{2} r^2\right) \frac{d}{dz} r^n dz$$

$$= K e^{-\frac{1}{2} z} z^{\frac{1}{2}(n-2)} dz$$

evaluando la constante  $K$  tendremos:

$$dF(z) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{1}{2}n\right)} e^{-\frac{1}{2} z} z^{\frac{1}{2}(n-2)} dz \quad \text{--- (2.7)}$$

Supongamos que  $x_1, \dots, x_n$  es una muestra aleatoria que se distribuye en forma normal con varianzas unitarias y está sujeta a  $P$  restricciones lineales independientes de la forma:

$$a_1 x_{11} + a_2 x_{12} + \dots + a_n x_{1n} = 0$$

$$a_1 x_{21} + a_2 x_{22} + \dots + a_n x_{2n} = 0$$

$$a_1 x_{p1} + a_2 x_{p2} + \dots + a_n x_{pn} = 0$$

Las variables  $x_i$  en el espacio de  $n$ -dimensiones se encuentran por lo tanto en  $p$  hiperplanos. El primer hiperplano corta a la hiper-esfera de densidad constante, obteniéndose de este corte una hiper-esfera de dimensión menor, cuya densidad también es constante, el corte del segundo hiperplano será una hiper-esfera cuya dimensión es menor a la obtenida por el primer corte, y así sucesivamente.

Por lo tanto bajo estas restricciones lineales la función de densidad de  $\bar{x}$  se expresará de la misma forma que en (2.7) pero con  $n-p$  en lugar de  $n$ , i.e.:

$$dF(\bar{x}) = \frac{1}{2^{\frac{1}{2}(n-p)} \Gamma(\frac{1}{2}(n-p))} e^{-\frac{1}{2}\bar{x}^2} \bar{x}^{\frac{1}{2}(n-p)} d\bar{x} \quad \text{--- (2.8)}$$

Para una muestra que se distribuya  $N(0, 1)$ , la densidad en el punto con coordenadas  $x_1, \dots, x_n$  es proporcional a:

$$\exp\left(-\frac{1}{2} \sum x_i^2\right) = \exp\left\{-\frac{1}{2} (n\bar{s}^2 + n\bar{x}^2)\right\}$$

ya que la varianza muestral  $s^2$  es igual a:

$$\frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{n} \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

Sea  $O$  el origen,  $Q$  el punto cuyas coordenadas son  $(1, \dots, 1)$ ,  $P$  el punto cuyas coordenadas son  $x_1, \dots, x_n$  y  $\overrightarrow{PH}$  un vector perpendicular al vector  $\overrightarrow{OQ}$ . La longitud del vector  $OH$  será:

$$\frac{x_1}{\sqrt{n}} + \frac{x_2}{\sqrt{n}} + \dots + \frac{x_n}{\sqrt{n}} = \bar{x} \sqrt{n}$$

ya que la proyección de P a Q representada por M está dada por:

$$M = \frac{(x_1, \dots, x_n) \cdot (1, \dots, 1)}{\|(1, \dots, 1)\|^2} (1, \dots, 1)$$

$$= \frac{x_1 + \dots + x_n}{n} (1, \dots, 1)$$

$$= \left( \frac{x_1 + \dots + x_n}{n}, \frac{x_1 + \dots + x_n}{n}, \dots, \frac{x_1 + \dots + x_n}{n} \right)$$

por lo tanto la longitud del origen O a M es:

$$\|\vec{OM}\| = \sqrt{\frac{(x_1 + \dots + x_n)^2}{n^2} + \dots + \frac{(x_1 + \dots + x_n)^2}{n^2}}$$

$$= \sqrt{n \left( \frac{x_1 + \dots + x_n}{n} \right)^2}$$

$$= \sqrt{\frac{1}{n} (x_1 + \dots + x_n)^2}$$

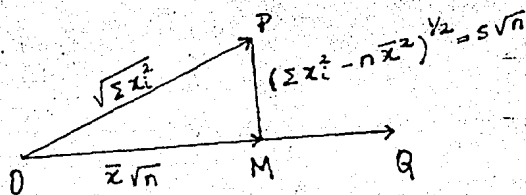
$$= \frac{x_1 + \dots + x_n}{\sqrt{n}}$$

la longitud de OP será  $\sqrt{\sum x_i^2}$  y la longitud de PM estará dada por:

$$\|\vec{PM}\| = \sqrt{\left( x_1 - \frac{x_1 + \dots + x_n}{n} \right)^2 + \dots + \left( x_n - \frac{x_1 + \dots + x_n}{n} \right)^2}$$

$$= \sqrt{\sum x_i^2 - 2 \sum x_i (\bar{x}) + n \bar{x}^2}$$

$$= (\sum x_i^2 - n \bar{x}^2)^{1/2} = s \sqrt{n}$$



Si el vector  $\vec{OP}$  lo hacemos girar alrededor del vector  $\vec{OM}$  se generará un cono y si se varía el ángulo entre  $\vec{OP}$  y  $\vec{OQ}$ , se generarán varios conos con vértice común en el origen  $O$ . Al realizar un corte a dichos conos con el hiperplano que contiene al vector  $PM$ , obtendremos hiper-esferas concéntricas de radio  $s\sqrt{n}$ , centradas en  $M$ .

Por lo tanto el volumen de los conos será igual al producto del volumen de las hiper-esferas por un incremento en  $\vec{OM}$ , igual a  $d\bar{x}\sqrt{n}$ , i.e.  $K d\bar{x}s^{n-2}ds$  multiplicado por otras variables independientes de  $\bar{x}$  y  $s$ .

Así pues la función de densidad de la media y la desviación estandar es proporcional a:

$$dF \propto \exp\left\{-\frac{1}{2}(ns^2 + n\bar{x}^2)\right\} s^{n-2} d\bar{x} ds$$

esta expresión se puede dividir en dos factores:

$$dF_1 \propto \exp\left(-\frac{1}{2}n\bar{x}^2\right) d\bar{x} \quad \text{--- (2.9)}$$

$$dF_2 \propto \exp\left(-\frac{1}{2}ns^2\right) s^{n-2} ds \quad \text{--- (2.10)}$$

## 2.8

## DISTRIBUCION T DE STUDENT

Basándonos en la sección anterior tendremos que:

$$\frac{\bar{x}\sqrt{n}}{s\sqrt{n}} = \frac{DM}{PM} = \cot \phi$$

donde  $\phi$  representa el ángulo formado por PDM.

Definamos la estadística  $\bar{z} = \bar{x}/s$ ,  $\bar{z}$  será constante sobre el cono generado al rotar DP alrededor del vector unitario DQ, siendo el ángulo  $\phi$  constante.

La distribución de  $\bar{z}$  se obtendrá calculando el volumen entre los conos definidos por  $\phi$  y  $d\phi$ .

Consideremos la intersección de estos conos con la hiper-esfera de radio OP.

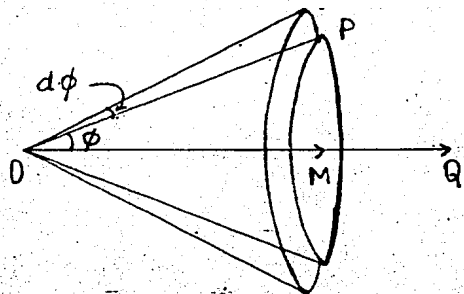


fig. 2.7

Cuando se intersecan las bases de los conos con la hiper-esfera se generará un anillo sobre la hiper-esfera cuyo contenido o volumen será proporcional a:

$$OP d\phi PM^{n-2} = OP^{n-1} \text{sen}^{n-2} \phi d\phi$$

La función de densidad es constante y proporcional a  $\exp\{-\frac{1}{2} OP^2\}$  sobre la hiper-esfera, y por lo tanto la distribución entre los conos será proporcional a:

$$e^{-\frac{1}{2} OP^2} OP^{n-1} \text{sen}^{n-2} \phi d\phi d(OP) \\ \propto \text{sen}^{n-2} \phi d\phi, \quad 0 \leq \phi \leq \pi$$

La distribución de  $z = \cot \phi$  es proporcional a

$$dF \propto \frac{k}{(1+z^2)^{\frac{1}{2}n}} dz$$

evaluando la constante  $k$  tendremos

$$dF = \frac{1}{B\{\frac{1}{2}(n-1), \frac{1}{2}\}} \frac{dz}{(1+z^2)^{\frac{1}{2}n}} \quad \text{--- (2.11)}$$

donde  $z$  es la razón de dos funciones de dimensión uno en las variables, esta distribución es válida para muestras de una distribución normal independientemente de la escala.

Si consideramos

$$t = \frac{\bar{x}\sqrt{n}}{\left\{\frac{1}{n-1} \sum (x_i - \bar{x})^2\right\}^{1/2}} = \sqrt{(n-1)} z$$

la distribución en (2.11) se puede expresar en términos de  $t$  de la siguiente forma, utilizando el Teorema de Cambio de Variable.

$$dF = \frac{1}{\sqrt{n-1} B\{\frac{1}{2}(n-1), \frac{1}{2}\}} \frac{dt}{\left(1 + \frac{t^2}{n-1}\right)^{\frac{1}{2}n}}$$

$$dF = \frac{\Gamma\{\frac{1}{2}(v+1)\}}{\sqrt{v\pi} \Gamma\{\frac{1}{2}v\}} \frac{dt}{\left(1 + t^2/v\right)^{\frac{1}{2}(v+1)}}$$

donde  $v = n-1$

## 2.9

## DISTRIBUCION DE LA MEDIA MUESTRAL DE UNA POBLACION UNIFORME.

Consideremos una variable uniformemente distribuida sobre cierto intervalo unitario, la distribución de dicha variable se pueda expresar como:

$$dF(x) = dx \quad 0 \leq x \leq 1$$

$$= 0 \quad \text{en otro lado}$$

Si tenemos una muestra de  $n$  valores, la densidad valdrá 1 en el interior del hiper-cubo  $0 \leq x_i \leq 1$ , y 0 fuera de él.

$$f(x) = \begin{cases} 1 & 0 \leq x_i \leq 1 \\ 0 & \text{en otro lado} \end{cases}$$

Supongamos que un vector unitario se encuentra sobre la diagonal de este hiper-cubo, (por ejemplo en  $\mathbb{R}^3$  la diagonal se encuentra en la cara superior del cubo, ver fig. 8). Si  $P$  es el punto muestral  $(x_1, \dots, x_n)$  y  $PM$  la perpendicular a esta diagonal, entonces, como se vio anteriormente en la sección 1.7, la distancia del origen  $O$  al punto  $M$ , está dada por  $\bar{x}\sqrt{n}$ . Para obtener la distribución de  $\bar{x}$ , es necesario encontrar el volumen entre los hiperplanos  $\bar{x}$  y  $\bar{x} + d\bar{x}$ . El hiperplano general  $\bar{x} = cte.$ , representa la frontera del hiper-cubo en una región acotada por hiperplanos, pero cambia su contorno conforme  $\bar{x}$  se incrementa.

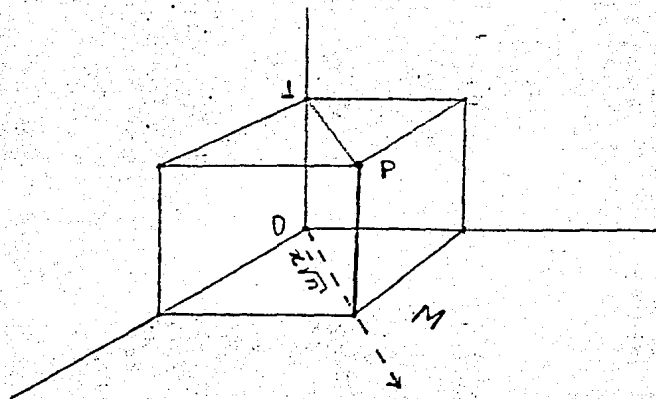


fig. 2.8

Las esquinas del hiper-cubo pueden constar de 0 ó 1 ó 2 ... ó  $n$ , coordenadas iguales a la unidad y el resto tener valor de 0.

Si los hiperplanos que acotan al hiper-cubo, se extienden indefinidamente sobre valores positivos de las  $x$ 's, definen octantes (análogos a los cuadrantes en 2 dimensiones).

Consideremos los octantes definidos por  $x_j \geq r_j$  donde  $r_j$  puede ser 0 ó 1. Estos en total contienen todas las esquinas del cubo, y cada esquina corresponde a la intersección de los hiperplanos que forman un octante.

El número de esquinas se agrupará en  $n+1$  conjuntos de acuerdo al tipo de esquina (las esquinas pueden tener 0, ó 1, ó 2, ..., ó  $n$  coordenadas iguales a la unidad), así pues tendremos que:

$$\sum_{j=1}^n r_j = r = 0, 1, \dots, n$$

Definamos a  $Q_i$  como el  $i$ -ésimo conjunto que cumple con estas características.

Tomando como ejemplo un cubo en  $\mathbb{R}^3$  encontraremos que: si  $Q_0 = \{(x, y, z) \mid x=0, y=0, z=0\}$ , el número de elementos de  $Q_0$  será 1 porque solo existe una esquina cuyas coordenadas son  $(0, 0, 0)$ .

$Q_1$  será el conjunto cuyos elementos tienen una sola coordenada igual a la unidad. El número de elementos de  $Q_1$  será 3, ya que  $Q_1 = \{(0, 0, 1), (1, 0, 0), (0, 1, 0)\}$ .

$Q_2$  será el conjunto cuyos elementos tienen 2 coordenadas iguales a la unidad. El número de elementos de  $Q_2$  será 3, ya que  $Q_2 = \{(1, 0, 1), (0, 1, 1), (1, 1, 0)\}$ .

$Q_3 = \{(x, y, z) \mid x=1, y=1, z=1\}$ , el número de elementos de  $Q_3$  es 1, porque solo existe una esquina cuyas coordenadas son  $(1, 1, 1)$ .



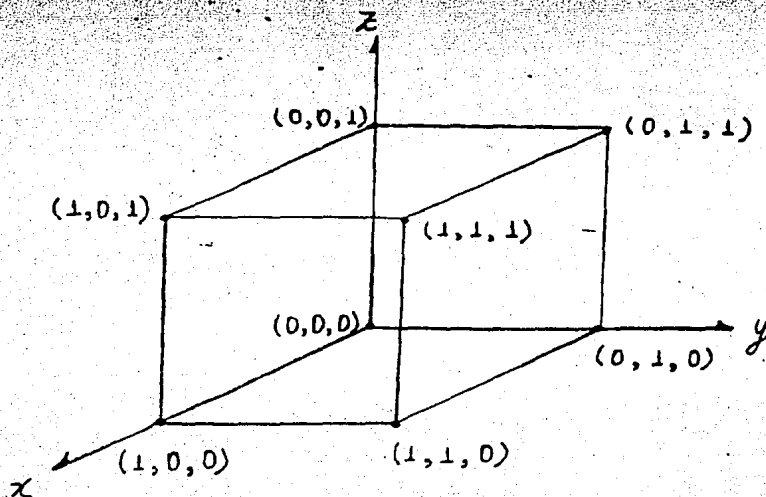


fig. 2.9

En general el número de elementos de  $Q_p$  es el coeficiente de  $t^p$  en la expansión binomial  $(1+t)^n$ . Hay  $2^n$  esquinas correspondientes a  $2^n$  octantes, en  $\mathbb{R}^3$  habrá  $2^3 = 8$  octantes y  $n+1 = 4$  conjuntos  $\{Q_0, Q_1, Q_2, Q_3\}$  que definen a cada esquina.

A cada octante se le asignará una densidad, sobre todo su dominio, desde 0 o 1 hasta infinito. Así pues de acuerdo al conjunto  $Q_p$  al que pertenezca, al octante se le asignará la densidad  $(-1)^p$ .

Sea  $P$  un punto con  $x$ 's no negativas, tal que  $s$  de sus coordenadas son más grandes que la unidad.

Por consiguiente la densidad total será cero en cualquier lugar donde  $s \geq 1$ .

$$0 = \sum_{p=0}^s (-1)^p \binom{s}{p}$$

esto nos indica que el punto  $P$  se encuentra fuera del cubo. Si el punto  $P$  se encontrará sobre o en el interior del cubo, la densidad sería unitaria. Consideremos el segmento del hiperplano  $\sum x = z$  que se encuentra en el octante referido al conjunto  $Q_0$  y que tiene como contenido a  $C_n(z)$ .

Si este hiperplano está situado en algún otro cuadrante referido al conjunto  $Q_r$ , con  $r$  coordenadas unitarias, su

contenido será  $C_n(z-r)$ . Si  $z \leq r$  el contenido será igual a cero. Considerando el desarrollo anterior, la densidad total requerida será:

$$\sum_{r=0}^K (-1)^r \binom{n}{r} C_n(z-r)$$

donde  $K = [z]$  es el más grande entero menor que  $z$ . Solo falta encontrar  $C_n(z)$ .

El contenido  $C_n(z)$  se encuentra en el hiperplano  $\sum x = z$ , el cual intersecta a los ejes coordenados en el punto con coordenada  $z$ .

El contenido de la pirámide determinado por los planos coordenados y el hiperplano  $\sum x = z$ , será  $z^n/n!$ .

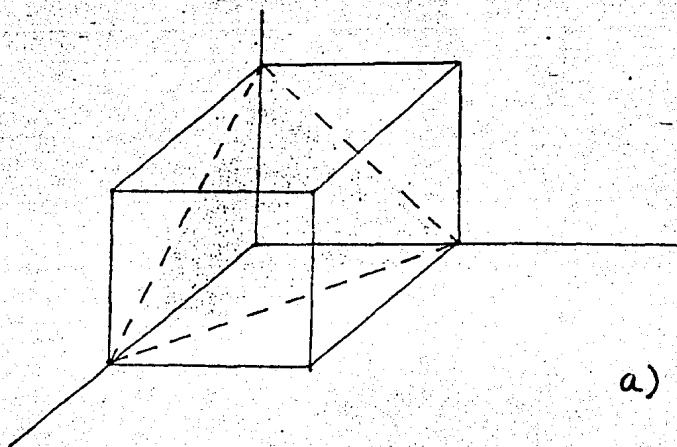
El contenido  $C_n(z)$  de la cara de la pirámide que no se encuentra en los planos coordenados se obtiene por diferenciación con respecto a  $z$ , i.e., es  $z^{n-1}/(n-1)!$ .

Por lo tanto tendremos que la función frecuencia en los octantes referidos a  $Q_r$ , se expresará como:

$$\frac{1}{(n-1)!} \sum_{r=0}^K (-1)^r \binom{n}{r} (z-r)^{n-1}, \quad K \leq z \leq K+1$$

La función frecuencia para  $\bar{x} = z/n$  se obtendrá a través de una substitución simple:

$$dF = \frac{n^n}{(n-1)!} \sum_{r=0}^K (-1)^r \binom{n}{r} \left(\bar{x} - \frac{r}{n}\right)^{n-1} d\bar{x}, \quad \frac{K}{n} \leq \bar{x} \leq \frac{K+1}{n}$$



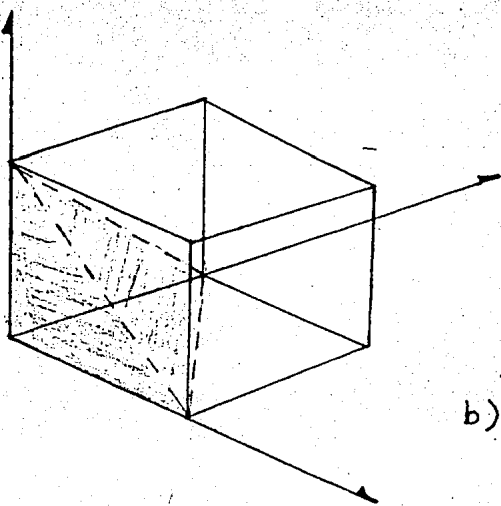


fig. 2.10

## 2.10

COEFICIENTE DE CORRELACION EN LA DISTRIBUCION  
NORMAL BIVARIADA

En Análisis de Regresión es importante saber si el ajustar los datos a una línea recta, es el mejor ajuste que pudimos obtener, ya que la recta puede distar mucho de los valores reales de nuestros datos.

De aquí la necesidad de encontrar una medida de la relación lineal entre  $X$  y  $Y$ , esta medida se llama coeficiente de correlación.

Sea

$$\phi(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right\}$$

la expresión de la distribución normal bivariada, donde  $\mu_1$ ,  $\mu_2$  son las medias de  $x$ ,  $y$  respectivamente;  $\sigma_1^2$ ,  $\sigma_2^2$  son las varianzas de  $x$ ,  $y$ , y por último  $\rho$  representa el coeficiente de correlación.

Consideremos el caso en que las medias valen cero y las varianzas son unitarias, entonces:

$$\phi(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2 - 2xy + y^2)\right\} \quad \text{--- (2.12)}$$

De acuerdo a la definición de Coeficiente de Correlación,  $\rho$  es independiente del origen y las escalas de las variables [7]; por lo tanto sin pérdida de generalidad se puede utilizar 2.12.

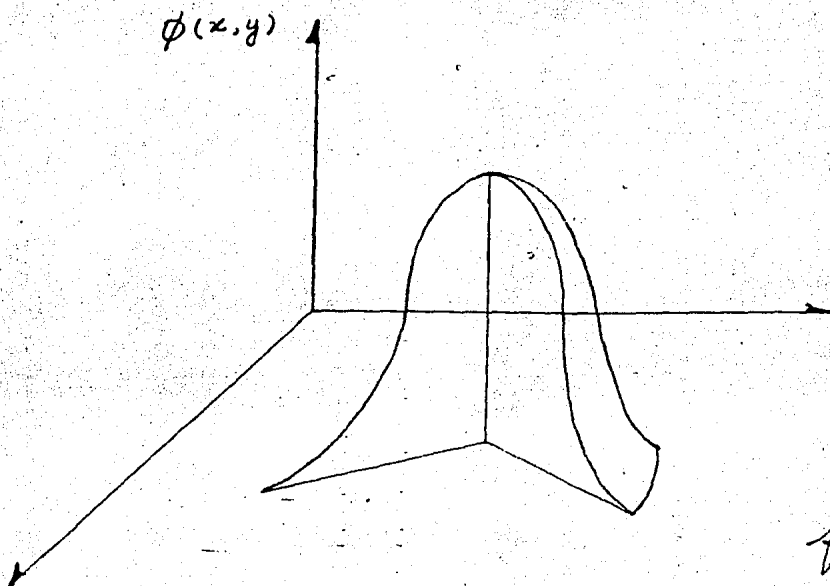


fig 2.11

Se requiere encontrar la función de distribución del coeficiente de correlación, en una muestra de  $n$  observaciones. La expresión para el estimador del coeficiente  $\rho$  es:

$$r = \hat{\rho} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left\{ \sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2 \right\}^{1/2}}$$

Aplicando la transformación de Helmert, vista en la sección 1.7.1, la función de densidad conjunta de una muestra de tamaño  $n$ , se expresará como:

$$f(u_1, \dots, u_n, v_1, \dots, v_n) = K \exp \left\{ -\frac{1}{2(1-\rho^2)} \left( \sum_{i=1}^{n-1} u_i^2 - 2\rho \sum_{i=1}^{n-1} u_i v_i + \sum_{i=1}^{n-1} v_i^2 \right) \right\} \\ du_1, \dots, du_{n-1} \times dv_1, \dots, dv_{n-1} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left( u_n^2 - 2\rho u_n v_n + v_n^2 \right) \right\} du_n dv_n \quad (2.13)$$

donde  $u_n = \bar{x}$ ,  $v_n = \bar{y}$ , y las demás  $u$ 's son independientes de  $u_n$  y  $v_n$ .

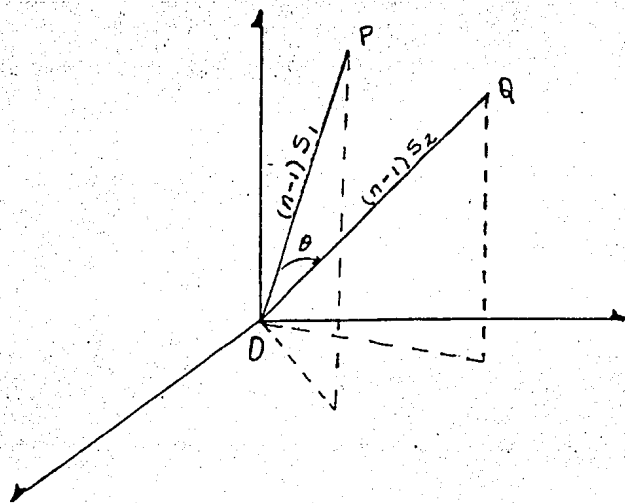


fig. 2.12

Considérese el vector  $OP$  fijo, y el vector  $OQ$  que forma un ángulo  $\theta$  con  $OP$ , que varía en el espacio de las  $v_i$  sobre un anillo de radio  $s_2 \sqrt{(n-1) \operatorname{sen} \theta}$ , la superficie obtenida será proporcional a:

$$\operatorname{sen}^{n-3} \theta d\theta = \operatorname{sen}^{n-4} \theta d \cos \theta = (1-r^2)^{\frac{1}{2}(n-4)} dr$$

el elemento diferencial será proporcional a:

$$s_1^{n-2} s_2^{n-2} (1-r^2)^{\frac{1}{2}(n-4)}$$

multiplicando este elemento por:

$$K \exp \left\{ -\frac{n-1}{2(1-p^2)} (s_1^2 - 2pr s_1 s_2 + s_2^2) \right\}$$

se obtendrá la función de distribución conjunta de  $s_1$ ,  $s_2$  y  $r$ . Para encontrar solo la distribución de  $r$ , se tendrá que realizar la integral parcial con respecto a  $s_1$  y  $s_2$ .

En este sistema coordenado, la distribución de

$$r = \frac{\sum_{i=1}^{n-1} u_i v_i}{\left\{ \sum_{i=1}^{n-1} u_i^2 \sum_{i=1}^{n-1} v_i^2 \right\}^{1/2}}$$

se obtendrá considerando el factor de la izquierda en (2.13), por lo tanto si se disminuye la dimensión, se elimina la variación de las medias  $\bar{x}$  y  $\bar{y}$ .

Para encontrar la distribución de  $r$ , consideraremos:

$$S_1 = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} u_i^2} \quad \text{y} \quad S_2 = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} v_i^2}$$

de tal forma que la función de frecuencias quedará expresada en términos de  $S_1$  y  $S_2$  como sigue:

$$K \exp \left\{ -\frac{n-1}{2(1-\rho^2)} (S_1^2 - 2\rho r S_1 S_2 + S_2^2) \right\}$$

Solo resta encontrar la forma de expresar  $du_1, \dots, du_{n-1}$ ,  $dv_1, \dots, dv_{n-1}$  en términos de  $S_1$ ,  $S_2$  y  $r$ .

Consideremos dos espacios de dimensión  $n-1$ , uno para  $u$  y otro para  $v$ , superpuestos uno sobre otro.

La distancia de  $P$ , punto con coordenadas  $u_i$ , a  $O$  (el origen) estará dado por  $(n-1)S_1$ , y la distancia de  $Q$ , punto con coordenadas  $v_i$ , a  $O$ , estará dada por  $(n-1)S_2$ , (ver fig. 12), por consiguiente  $r = \frac{\sum_{i=1}^{n-1} u_i v_i}{\sqrt{\sum_{i=1}^{n-1} u_i^2 \sum_{i=1}^{n-1} v_i^2}}$  será igual al coseno del

ángulo entre  $OP$  y  $OQ$ , puesto que:

$$\cos \theta = \frac{u \cdot v}{\|u\| \|v\|} = \frac{\sum_{i=1}^{n-1} u_i v_i}{\|u\| \|v\|}$$

$$= \frac{\sum_{i=1}^{n-1} u_i v_i}{\sqrt{\sum_{i=1}^{n-1} u_i^2} \sqrt{\sum_{i=1}^{n-1} v_i^2}} = \frac{\sum_{i=1}^{n-1} u_i v_i}{\sqrt{\sum_{i=1}^{n-1} u_i^2 \sum_{i=1}^{n-1} v_i^2}}$$

2.11

## DISTRIBUCION WISHART

El procedimiento de cálculo de distribuciones conjuntas, visto anteriormente, se puede extender para  $P$  formas de variación.

Considérese el arreglo de una muestra de tamaño  $n$ , de un conjunto, de  $P$  variables.

$$\begin{array}{cccc} x_{11} & \dots & \dots & x_{1p} \\ x_{21} & \dots & \dots & x_{2p} \\ \vdots & & & \\ x_{n1} & \dots & \dots & x_{np} \end{array}$$

La distribución normal multivariada de  $P$  variables  $x_1 \dots x_p$ , con media cero y varianza unitaria, está dada por:

$$f(x) = \frac{1}{2\pi^{P/2} |\Sigma|^{1/2}} e^{-\frac{1}{2} x' \Sigma^{-1} x}$$

Sea  $A$  la matriz inversa de la matriz de correlación  $(\Sigma^{ij})$ ,  $i, j = 1, \dots, p$ , entonces:

$$f(x) = \frac{|A|^{1/2}}{2\pi^{P/2}} e^{-\frac{1}{2} x' A x}$$

Para una muestra  $n$ , tendremos la siguiente función de frecuencias:

$$\begin{aligned} f(x)_n &= \frac{|A|^{n/2}}{2\pi^{\frac{nP}{2}}} e^{-\frac{1}{2} x_n' A x_n} \\ &= \frac{|A|^{n/2}}{2\pi^{\frac{nP}{2}}} e \left\{ -\frac{1}{2} \sum_{k=1}^n \sum_{i,j=1}^p A^{ij} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \right. \\ &\quad \left. - \frac{n}{2} \sum_{k=1}^n \sum_{i,j=1}^p A^{ij} x_i x_j \right\} \end{aligned}$$



Haciendo uso de la Transformación de Helmer, se eliminará  $\bar{x}$  y se obtendrá la siguiente expresión.

$$f(\underline{x})_n = \frac{|A|^{\frac{1}{2}(n-1)}}{n^{P/2} (2\pi)^{\frac{(n-1)P}{2}}} e^{-\frac{1}{2} U'AU} \quad \dots (2.14)$$

Donde  $U'U = \sum_{k=1}^{n-1} U_{ki} U_{kj} = \xi_{ij} = n a_{ij}$ , donde  $a_{ij}$  representa las varianzas o covarianzas.

Si se divide  $\frac{|A|^{n/2}}{(2\pi)^{\frac{nP}{2}}}$  entre la constante  $\frac{n^{P/2} |A|^{1/2}}{(2\pi)^{P/2}}$  se hará que la distribución de medias, tenga frecuencia unitaria, i.e. obtendremos la constante  $\frac{|A|^{n/2}}{n^{P/2} (2\pi)^{\frac{(n-1)P}{2}}}$  que es la que se encuentra en la ecuación (2.14). La ecuación (2.14) se puede expresar de la siguiente forma:

$$f(\underline{x})_n = \frac{|A|^{\frac{1}{2}(n-1)}}{n^{P/2} (2\pi)^{\frac{(n-1)P}{2}}} e^{-\frac{1}{2} \sum A^{ij} a_{ij}}$$

Consideremos  $P$  hiper-espacios de  $n-1$  dimensiones uno para cada  $u$ , y sean  $P_1, \dots, P_P$  puntos muestrales.

Se considerará la variación de estos puntos en forma recursiva, i.e. se toma en cuenta la variación de  $P_1$ , en seguida la de  $P_2$  dada por  $P_1$ , la de  $P_3$  dada por  $P_1$  y  $P_2$ , y así sucesivamente.

Multiplicando todas estas variaciones, se obtendrá la variación conjunta de  $P_1, \dots, P_P$ .

Consideremos fijo el vector  $OP_m$ , así como los ángulos que se forman entre  $P_m OP_1, P_m OP_2, \dots, P_m OP_{m-1}$ , donde  $O$  es el origen y  $P_m$  está dado por  $P_1, \dots, P_{m-1}$ ;  $P_m$  se encontrará sobre una hiper-esfera en  $n-m$  dimensiones, ( $m \leq n$ ).

Consideremos ahora que los espacios están superpuestos, sea  $l_m$  la longitud de la perpendicular de  $P_m$  al hiper-plano de dimensión  $m-1$  determinado por  $O, P_1, \dots, P_{m-1}$ .

Como  $P_m$  se encuentra en una hiper-esfera, el contenido o volumen se obtendrá con la siguiente ecuación:

$$\frac{2\pi^{\frac{1}{2}(n-m)} t_m^{n-m-1}}{\Gamma\left\{\frac{1}{2}(n-m)\right\}} \quad \text{--- (2.15)}$$

Ya que el contenido de la superficie de una hiper-esfera está dado por  $\frac{2\pi^{\frac{1}{2}n} a^{n-1}}{\Gamma\left(\frac{1}{2}n\right)}$ , donde  $a$  representa el radio, (ver sección 1.13.1).

2.12

## CORRELACIONES

El conjunto de  $n$  valores  $x_1, \dots, x_n$  se representó como un punto  $P_i$  en  $n$  dimensiones, o en forma equivalente, como un vector  $OP_i$ , donde  $O$  es el origen.

Así pues un conjunto de  $n$  observaciones de  $P$  variables especificado por:

$$(x_{ij}), \quad i=1, 2, \dots, n \quad ; \quad j=1, 2, \dots, P$$

se puede representar por  $P$  puntos  $P_1, \dots, P_P$  o como  $P$  vectores,  $\vec{OP}_1, \dots, \vec{OP}_P$ .

En el caso uni-dimensional, la función de densidad está representada por un conjunto de puntos, en el caso  $p$ -dimensional se representará por conjuntos de  $P$  puntos o vectores cada uno.

El cuadrado de la longitud del vector  $OP_j$  estará dada por:

$$(OP_j)^2 = \sum_{i=1}^n x_{ij}^2$$

Consideremos el origen  $O$  como la media de las  $n$  observaciones, i.e., el punto cuya  $j$ -ésima coordenada es  $\sum_{i=1}^n x_{ij} / n = \bar{x}_j$ , entonces:

$$(\vec{OP}_j)^2 = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 = n \widehat{\text{Var}} x_j$$

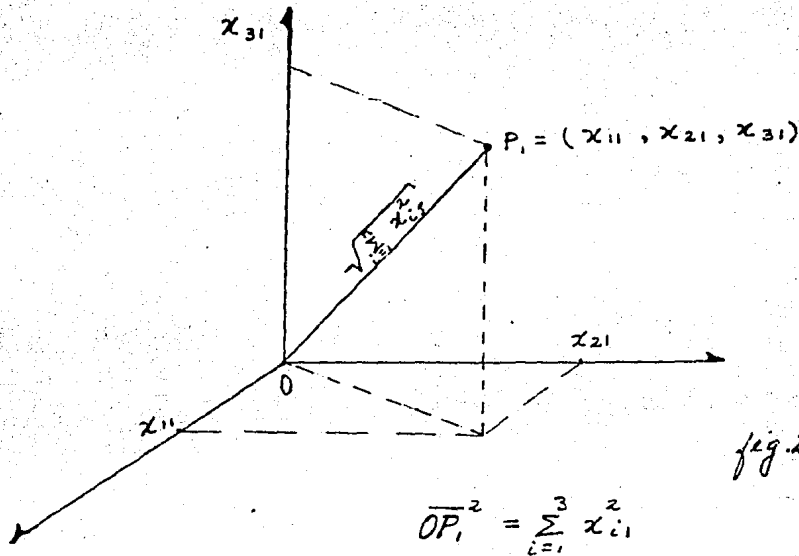
donde  $\widehat{\text{Var}} x_j$  es el estimador máximo verosímil de la varianza de  $x_j$ ,  $j=1, \dots, P$ .

Sea  $\theta$  el ángulo que se forma entre los puntos  $P_j$  y  $P_k$ , entonces:

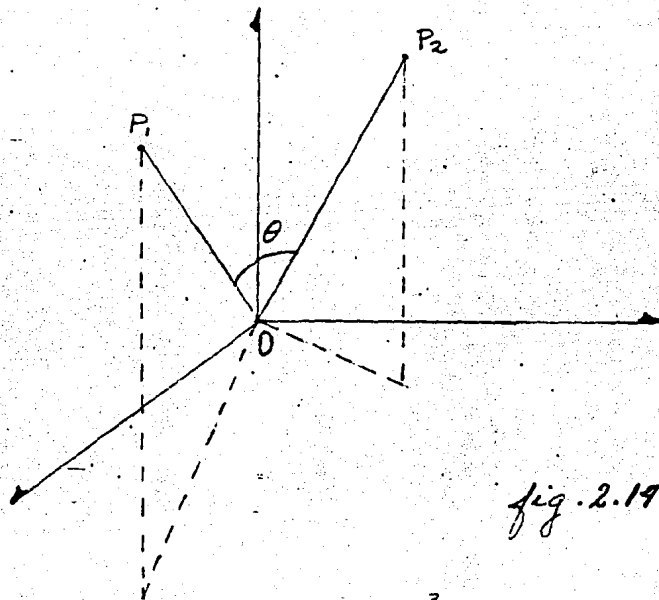
$$\cos \theta = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\left\{ \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2 \right\}^{1/2}}$$

$$= \text{correlación entre } x_j \text{ y } x_k = r_{jk}$$

Para el caso  $n=3$  y  $J=1$ , tendremos:



Supongamos ahora que  $J=1, 2$ , i.e. contaremos con 2 vectores  $\overrightarrow{OP_1}$  y  $\overrightarrow{OP_2}$



$$\begin{aligned} \cos \theta &= \frac{(x_{11} - \bar{x}_1) \cdot (x_{12} - \bar{x}_2)}{\| (x_{11} - \bar{x}_1) \| \| (x_{12} - \bar{x}_2) \|} = \frac{\sum_{i=1}^3 (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\left\{ \sum_{i=1}^3 (x_{i1} - \bar{x}_1)^2 \sum_{i=1}^3 (x_{i2} - \bar{x}_2)^2 \right\}^{1/2}} \\ &= \text{correlación entre } x_1 \text{ y } x_2 = r_{12} \end{aligned}$$

Podemos ver entonces cómo la variancia y la correlación de  $P$  variables (cada una con  $n$  observaciones), se pueden representar por longitudes y ángulos.

Si  $r_{ij} = 1$ , el ángulo entre los vectores es cero y la variancia es una función lineal de la otra. Si  $r_{ij} = 0$ , las variables no están correlacionadas.

Consideremos tres variables correspondientes a  $OP_1$ ,  $OP_2$ ,  $OP_3$  y proyectemos  $OP_1$  y  $OP_2$  sobre el hiperplano de dimensión  $(n-1)$ , perpendicular a  $OP_3$ . Si los puntos proyectados son  $Q_1$  y  $Q_2$ , la correlación entre las variables  $x_1$  y  $x_2$  en este hiperplano está medida por el coseno del ángulo formado por los puntos  $Q_1, O, Q_2$ , esta medida de relación entre  $x_1$  y  $x_2$ , es independiente de  $x_3$  y se expresa como:

$$r_{12 \cdot 3} = \frac{r_{12} - r_{13}r_{23}}{\{(1-r_{13}^2)(1-r_{23}^2)\}^{1/2}}$$

La cual recibe el nombre de correlación parcial.

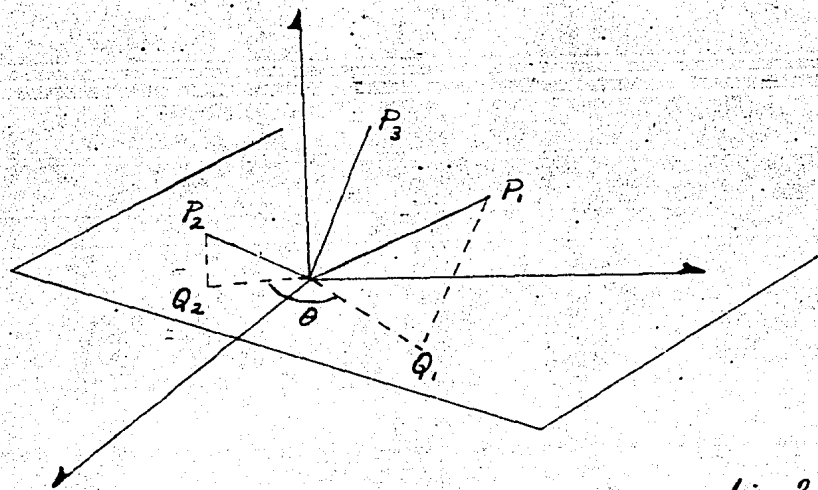


fig. 2.15

De igual forma, si se realiza la proyección a un hiperplano ortogonal a un vector adicional  $x_4$ , obtendremos una correlación parcial de segundo orden,  $r_{12.34}$ , la cual expresa la relación entre las variables  $x_1$  y  $x_2$  cuando la dependencia de  $x_3$  y  $x_4$  ha sido eliminada, y así sucesivamente.

Para una distribución normal, las superficies de densidad constante de la variable  $x_i$ , son hiper-esferas centradas en 0.

Resumiendo lo expuesto anteriormente, las correlaciones entre las variables, estarán representadas por los cosenos de los ángulos entre los vectores, definidos por el origen y los puntos que representan a las variables.

Veamos ahora cómo la representación de todas las correlaciones es invariante bajo una rotación de los ejes.

Consideremos a  $OP_3$  coincidente con alguno de los ejes coordenados, si se proyectan los vectores  $OP_1$  y  $OP_2$  al espacio  $R^{n-1}$  ortogonal a  $OP_3$ , vemos que las hiper-esferas de densidad constante son las mismas y que las correlaciones totales son reemplazadas por correlaciones parciales. Por lo tanto la distribución muestral de  $r_{12.3}$  será exactamente la misma que la de  $r_{12}$  a excepción de que nos encontramos en una dimensión menor.

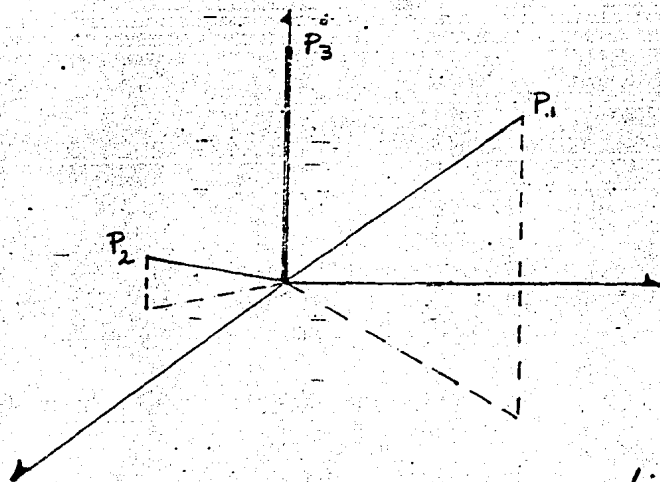


fig 2.16

2.13

## CORRELACION MULTIPLE

En la sección 2.10 tratamos ya el concepto de Coeficiente de Correlación para el caso de una distribución Normal Bivariada. En esta sección veremos qué sucede con  $p$  variables.

Los puntos  $P_1$  a  $P_p$ , junto con el origen definirán un espacio de  $p$  dimensiones, este espacio formará un ángulo con el vector  $OY$ . El coseno de este ángulo es llamado coeficiente de correlación múltiple de  $Y$  sobre  $x_1, x_2, \dots, x_p$ . Este coeficiente se denota usualmente por  $R$ .

Si  $Y$  es una función lineal de las  $x$ 's

$$Y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (2.16)$$

entonces el vector  $OY$  se encuentra enteramente en el espacio de las  $x$ 's y  $R$  es unitario. Si  $Y$  no depende de las  $x$ 's  $OY$  es ortogonal y  $R=0$ .  $R$  puede asumir cualquier valor en el rango  $-1 \leq R \leq 1$ . Así pues la correlación múltiple puede ser vista como el coseno del ángulo entre  $OY$  y la proyección ortogonal de  $OY$  sobre el espacio de las  $x$ 's. Esto como se mencionó en la sección 2.10, mide la exactitud con la cual es realizada la representación lineal de  $Y$  en términos de  $x$ 's.

Además  $R$  es el coseno del mínimo ángulo de la familia de ángulos entre  $OY$  y cualquier vector  $OX$ , i.e.  $R$  es un máximo, por lo tanto los coeficientes  $\beta$ 's de la ecuación 2.16 se encontrarán al maximizar  $R$ , i.e. al maximizar el coeficiente de correlación entre  $Y$  y la función lineal  $\sum_{j=1}^p \beta_j x_j$ . Esto es equivalente a minimizar el cuadrado de la distancia de  $Y$  a su proyección sobre el espacio de las  $x$ 's, i.e. minimiza la longitud del vector  $Y - \sum \beta_j x_j$  representada de la siguiente forma:

$$\sum_{\text{muestra}} \left( Y - \sum_{j=1}^p \beta_j x_j \right)^2$$

la cual nos lleva a las ecuaciones de mínimos cuadrados de la Teoría de Regresión Lineal.

Consideremos el contenido del paralelepípedo determinado por  $Y, P_1, \dots, P_p$ , y un sistema coordenado en el espacio de  $P+1$  dimensiones; denotemos las coordenadas de  $P_j$  como  $\xi_{ji}$ ,  $j=1, \dots, p$  y las de  $Y$  por  $\eta_{ii}$ , donde  $i=1, \dots, P+1$ . El contenido del paralelepípedo será entonces:

$$C = \begin{vmatrix} \eta_{11} & \xi_{11} & \dots & \dots & \xi_{p1} \\ \eta_{12} & \xi_{12} & \dots & \dots & \xi_{p2} \\ \vdots & \vdots & \dots & \dots & \vdots \\ \eta_{1P} & \xi_{1P} & \dots & \dots & \xi_{pP} \\ \eta_{1P+1} & 0 & \dots & \dots & 0 \end{vmatrix}$$

Multipliquemos esta matriz por su transpuesta y denotando por  $r_{ij}$  al coeficiente de correlación entre  $\xi_i$  y  $\xi_j$ , y por  $r_{yi}$  al coeficiente de correlación entre  $Y$  y  $\xi_i$ , tendremos:

$$C^2 = \begin{vmatrix} 1 & r_{y1} & r_{y2} & \dots & \dots & r_{yp} \\ r_{y1} & 1 & r_{12} & \dots & \dots & r_{1p} \\ r_{y2} & r_{12} & 1 & \dots & \dots & r_{2p} \\ \vdots & \vdots & \vdots & \dots & \dots & \vdots \\ r_{yp} & r_{p1} & \dots & \dots & \dots & 1 \end{vmatrix} \quad \text{--- (2.17)}$$

Ahora si  $\theta$  es el ángulo entre  $OY$  y el espacio de las  $x$ 's, el contenido  $C$  es  $\sin \theta$  veces el contenido del paralelepípedo determinado por  $O, P_1, \dots, P_p$ , (ver seccs. 1.13.3 y 1.13.4), el cual es el menor del primer elemento a mano izquierda en la expresión (2.17).



Así que:

$$\frac{W}{W_{11}} = \sin^2 \theta = 1 - R^2$$

donde  $W$  es la matriz cuyo determinante está dado en (2.17) y  $W_{11}$  es el menor del elemento en la esquina superior izquierda.

Pero el contenido se puede obtener también de la siguiente forma:

Consideremos el área del plano generado por los vectores  $\vec{OY}$  y  $\vec{OP}_1$ , la longitud de la perpendicular del vector  $\vec{OP}_2$  a este plano se multiplicará por la longitud de la perpendicular del vector  $\vec{OP}_3$  al plano generado por los vectores  $\vec{OY}$ ,  $\vec{OP}_1$ ,  $\vec{OP}_2$ ; y así sucesivamente. Estas perpendiculares son senos de ángulos, cada uno de los cuales es una correlación parcial.

Por lo tanto encontraremos que:

$$1 - R^2 = (1 - r_{y1}^2) (1 - r_{y2 \cdot 1}^2) (1 - r_{y3 \cdot 12}^2) \cdots (1 - r_{yp \cdot 12 \cdots (p-1)}^2)$$

la cual es una descomposición de  $1 - R^2$ , que se puede expresar en diferentes formas, de acuerdo al orden en que se seleccionen los vectores.

Consideremos ahora la distribución de  $R$ , para  $n$  muestras, como una distribución normal multivariada en la cual la variación representada por  $\vec{OY}$ , es independiente de otras variaciones.

Esto significa que muestRANDO repetidamente, el vector  $\vec{OY}$  será aleatoriamente orientado con respecto al espacio  $S_p$  de las  $x$ 's. No se perderá generalidad si se supone las  $x$ 's fijas, especialmente el espacio  $S_p$  se considerará fijo, se tomará en cuenta también la distribución del ángulo formado por el espacio  $S_p$  y el vector  $\vec{OY}$  orientado aleatoriamente, o en forma equivalente el ángulo determinado por un punto  $Y$  que se mueve sobre una hiper-esfera de radio unitario.

Consideremos un vector  $\vec{OZ}$  en el plano  $S_p$  y el ángulo  $\theta$  como uno de los ángulos entre  $\vec{OY}$  y  $\vec{OZ}$ , supondremos que este ángulo es el mínimo posible para distintas variaciones de  $Z$ .

Si  $\vec{OZ}$  y  $\theta$  son fijos,  $Y$  varía sobre la superficie de una hiper-esfera en  $n-p-1$  dimensiones con contenido proporcional a la integral de  $(\sin \theta)^{n-p-2}$ . El punto  $O$  puede variar independientemente en un plano  $S_p$  sobre la superficie de una hiper-esfera con contenido igual a la integral de  $(\cos \theta)^{p-1}$ , (ver sec 1.13.1). Para el ángulo  $\theta$ , tendremos que:

$$dF \propto (\sin \theta)^{n-p-2} (\cos \theta)^{p-1} d\theta$$

si hacemos  $R = \cos \theta$ , entonces:

$$dF \propto R^{p-1} (1-R^2)^{\frac{1}{2}(n-p-3)} dR$$

ahora, si expresamos esta función en términos de  $R^2$  y evaluamos la constante, tendremos:

$$dF = \frac{1}{B\left\{\frac{1}{2}p, \frac{1}{2}(n-p+1)\right\}} (R^2)^{\frac{1}{2}(p-1)} (1-R^2)^{\frac{1}{2}(n-p-3)}$$

2.14

## CORRELACIONES CANONICAS

La generalización de la regresión de una variable  $Y$  sobre un conjunto de variables  $x_1, \dots, x_p$ , se da de la siguiente forma: consideremos la relación entre un conjunto de  $q$  variables  $Y_1, Y_2, \dots, Y_q$  sobre  $x_1, x_2, \dots, x_p$ .

Se puede considerar la correspondencia de las  $Y$ 's con  $q$  vectores en un espacio  $S_q$  y la correspondencia de las  $x$ 's con  $P$  vectores en un espacio  $S_p$ . La correlación entre los espacios depende de los ángulos entre los vectores.

Es posible encontrar transformaciones lineales de las  $x$ 's a  $\xi$ 's y de las  $Y$ 's a  $\eta$ 's, tales que:

- 1.- Todas las  $\xi$ 's son independientes (i.e. los ejes en el espacio de orden  $p$  son ortogonales).
- 2.- Todas las  $\eta$ 's son independientes, (i.e. todos los ejes en el espacio de orden  $q$  son ortogonales).
- 3.- Todas las  $\xi$ 's no están correlacionadas con todas las  $\eta$ 's excepto para  $p < q$  correlaciones entre:  $(\xi_1, \eta_1), (\xi_2, \eta_2), \dots, (\xi_p, \eta_p)$ , estos son valores fijos de las posibles correlaciones entre un vector en un espacio y un vector en el otro; y corresponden además a los ángulos canónicos entre los espacios. Ellos expresan la relación entre los dos conjuntos de variables, y son conocidos como Correlaciones Canónicas.

2.15

MEDIDA DE INFORMACION

Sea  $f(x, \theta_1, \theta_2, \dots, \theta_p)$  la función de densidad de una distribución de  $P$  parámetros. Si  $x_1, x_2, \dots, x_n$  es una muestra de  $n$  observaciones independientes de la distribución, la función de verosimilitud de la muestra se expresará de la siguiente forma:

$$L = \log \prod_{i=1}^n f(x_i, \theta_j)$$

Es necesario hacer notar que la condición necesaria y suficiente para que una distribución que depende de  $P$  parámetros admita un conjunto de  $P$  estadísticas suficientes, es que la función de densidad de la distribución deberá ser de la forma:  
[ver [9]]

$$f(x, \theta_j) = \exp \left\{ \sum_{k=1}^P \mu_k(\theta_j) v_k(x) + A(x) + B(\theta_j) \right\} \quad \text{--- (2.18)}$$

donde las  $\mu_k$ 's y  $B$  son funciones de  $\theta_j$ 's y las  $v_k$ 's y  $A$  son funciones de  $x$ .

Expresando  $L$  de la forma (2.18), tendremos

$$\begin{aligned} L &= \log \prod_{i=1}^n f(x_i, \theta_j) \\ &= \sum_{i=1}^n \log f(x_i, \theta_j) \\ &= \sum_{k=1}^P \mu_k(\theta_j) \sum_{i=1}^n v_k(x_i) + \sum_{i=1}^n A(x_i) + nB(\theta_j) \\ &= \sum_{k=1}^P \mu_k(\theta_j) T_k + \sum_{i=1}^n A(x_i) + nB(\theta_j) \end{aligned}$$

donde  $T_k = \sum_{i=1}^n v_k(x_i)$

$$y \quad \frac{\partial L}{\partial \theta_r} = \sum_{k=1}^P \frac{\partial \mu_k}{\partial \theta_r} T_k + n \frac{\partial B}{\partial \theta_r} \quad (r=1, 2, \dots, P) \quad \text{--- (2.19)}$$

Sea  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_P$  una solución del sistema de ecuaciones de verosimilitud:

$$\frac{\partial L}{\partial \theta_r} = 0 \quad (r=1, 2, \dots, P)$$

Las  $\hat{\theta}_j$ 's son estimadores máximo verosímiles de los  $\theta_j$ 's.

Entonces

$$\frac{\partial L}{\partial \theta_r} \Big|_{\hat{\theta}_j} = \sum_{k=1}^P \left( \frac{\partial \mu_k}{\partial \theta_r} \right)_{\hat{\theta}_j} T_k + n \left( \frac{\partial B}{\partial \theta_r} \right)_{\hat{\theta}_j} = 0 \quad \text{--- (2.20)}$$

La notación  $(\partial \mu_k / \partial \theta_r)_{\hat{\theta}_j}$  indica que la derivada está valuada en  $\hat{\theta}_j$ .

Dado que  $E(\partial L / \partial \theta_r) \equiv 0$ , tendremos

$$\sum_{k=1}^P \frac{\partial \mu_k}{\partial \theta_r} E(T_k) + n \frac{\partial B}{\partial \theta_r} = 0 \quad \text{--- (2.21)} \\ (r=1, 2, \dots, P)$$

$$y \quad \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} = \sum_{k=1}^P \frac{\partial^2 \mu_k}{\partial \theta_r \partial \theta_s} T_k + n \frac{\partial^2 B}{\partial \theta_r \partial \theta_s} \quad \text{--- (2.22)} \\ (r=1, 2, \dots, P)$$

De aquí que:

$$E \left( \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right) = \sum_{k=1}^P \frac{\partial^2 \mu_k}{\partial \theta_r \partial \theta_s} E(T_k) + n \frac{\partial^2 B}{\partial \theta_r \partial \theta_s} \quad \text{--- (2.23)}$$

$$y \quad \left( \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right)_{\hat{\theta}_j} = \sum_{k=1}^P \left( \frac{\partial^2 \mu_k}{\partial \theta_r \partial \theta_s} \right)_{\hat{\theta}_j} T_k + n \left( \frac{\partial^2 B}{\partial \theta_r \partial \theta_s} \right)_{\hat{\theta}_j} \quad \text{--- (2.24)}$$

Al resolver las P ecuaciones lineales simultáneas, expresadas en (2.21), obtendremos el valor de  $E(T_k)$ , si este valor obtenido se sustituye en (2.23), obtendremos una expresión en términos del parámetro  $\theta$ :

$$E \left( \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right) \Big|_{\hat{\theta}_j} = \phi(\theta_j) \Big|_{\hat{\theta}_j} \quad \text{--- (2.25)}$$

Las  $p$  ecuaciones lineales simultáneas dadas en (2.20) nos permiten expresar a las  $T_k$ 's en términos de las  $\hat{\theta}_j$ 's, por tanto se pueden sustituir las  $T_k$ 's en la expresión (2.24).

En forma análoga a (2.25) el resultado de sustituir las  $T_k$ 's en (2.24) será:

$$\left( \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right)_{\hat{\theta}_j} = \phi(\hat{\theta}_j) \quad \text{--- (2.26)}$$

De (2.25) y (2.26) tenemos la siguiente propiedad

$$\left( \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right)_{\hat{\theta}_j} = \left\{ E \left( \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right) \right\}_{\hat{\theta}_j} \quad \text{--- (2.27)}$$

Si  $r, s = 1, 2, \dots, p$ ; (2.27) puede ser escrito en forma matricial

$$\left[ \left( \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right)_{\hat{\theta}_j} \right] = \left[ \left\{ E \left( \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right) \right\}_{\hat{\theta}_j} \right] \quad \text{--- (2.28)}$$

Si multiplicamos por  $-1$  a ambos lados de la igualdad, obtendremos la siguiente expresión:

$$\left[ \left( - \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right)_{\hat{\theta}_j} \right] = \left[ \left\{ E \left( - \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right) \right\}_{\hat{\theta}_j} \right] \quad \text{--- (2.29)}$$

donde  $\left[ E \left( - \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right) \right]$  es la matriz información de Fisher.

En seguida analizaremos cuál es el significado geométrico de la expresión (2.29), considerando el caso de un solo parámetro, i.e.

$$\left( - \frac{\partial^2 L}{\partial \theta^2} \right)_{\hat{\theta}} = \left\{ E \left( - \frac{\partial^2 L}{\partial \theta^2} \right) \right\}_{\hat{\theta}} = I(\hat{\theta}) \quad \text{--- (2.30)}$$

donde  $I(\theta)$  es la medida de información de una muestra de  $n$  observaciones.

La curva representada por la función de verosimilitud de una muestra, es llamada curva de verosimilitud.

Si  $\rho$  es el radio de curvatura de la curva de verosimilitud en el punto  $\theta = \hat{\theta}$ , (ver secc. 1.15),

$$\frac{1}{\rho} = \frac{\left(-\frac{\partial^2 L}{\partial \theta^2}\right)_{\hat{\theta}}}{\left\{1 + \left(\frac{\partial L}{\partial \theta}\right)_{\hat{\theta}}^2\right\}^{1/2}} = \left(-\frac{\partial^2 L}{\partial \theta^2}\right)_{\hat{\theta}} \quad \text{donde } \left(\frac{\partial L}{\partial \theta}\right)_{\hat{\theta}} = 0$$

entonces  $\frac{1}{\rho} = I(\hat{\theta})$ .

Así pues la función de información, mide la curvatura de la curva de verosimilitud en el punto representado por el estimador máximo verosímil.

En muestras grandes la varianza de  $\hat{\theta}$  es  $\frac{1}{I(\theta)}$  y el estimador de la varianza de  $\hat{\theta}$  es  $\frac{1}{I(\hat{\theta})}$ .

De aquí que el radio de curvatura de la curva de verosimilitud evaluada en el estimador máximo verosímil da un valor estimado de la varianza de dicho estimador.

## CONCLUSIONES

Normalmente en los cursos de Estadística y en documentos sobre la materia, el concepto que más comúnmente es tratado en forma geométrica es el del Modelo de Regresión Lineal, a pesar de la utilidad, en cuanto a facilidad de asimilación del concepto se refiere, que una representación geométrica nos brinda.

Si bien es cierto que resulta imposible representar geoméricamente un concepto en su forma generalizada, el hacerlo para una, dos o tres dimensiones nos da la pauta para poderlo concebir en  $n$  dimensiones.

De aquí que este trabajo haya sido enfocado al Análisis Estadístico Univariado, sabiendo que a partir de los conceptos aquí tratados, se pueden desarrollar estos para un enfoque en Análisis Estadístico Multivariado.

Así pues el objetivo primordial del presente trabajo, como se planteó en la introducción, espero que pueda servir de apoyo a aquellas personas interesadas en el estudio de la Estadística, de tal forma que puedan asimilar con mayor facilidad la generalización de los conceptos estadísticos.

Entre el conjunto de interpretaciones que se plantean, la que me pareció más interesante fue la del concepto de Distribución  $t$  de Student, primero en cuanto a su importancia en Estadística y segundo porque en su desarrollo se manejaron varios conceptos como lo son el de proyección, la media, la desviación estandar, y la generación y cálculo del volumen de un sólido.

Por último, es de vital importancia mencionar que en este trabajo, no se pretenden cubrir todos los conceptos estadísticos y los que se mencionan fueron seleccionados bajo el criterio de considerarlos básicos para poder desarrollar otros conceptos.



## REFERENCIAS

- Aitken, A. C. (1949). ON THE WISHART DISTRIBUTION IN STATISTICS. *Biometrika*, Vol. 36, p.p. 59-62.
- Apostol, Tom M. (1978). CALCULUS (segunda edición) Vol. I, p.p. 335-341 [2]
- Apostol, Tom M. (1978). CALCULUS (segunda edición) Vol. II, p.p. 501-504 [4]  
p.p. 96-110 [5]
- Graybill, Franklin A. (1961). AN INTRODUCTION TO LINEAR STATISTICAL MODELS. Mc Graw-Hill, Vol 1, Teo. 4.6 [6]
- Haaser, Norman (1970). ANALISIS MATEMATICO. (primera edición), Vol 2, p. 57 [1]
- Hogg, R. V. and Craig A. T. (1970). INTRODUCTION TO MATHEMATICAL STATISTICS (Third edition), Collier Mc. Millan Int. E. [9]
- Huzar bazar, V. S. (1949). ON A PROPERTY OF DISTRIBUTION ADMITING SUFFICIENT STATISTICS. *Biometrika*, Vol 36, p.p. 71-74.
- Johnston, J. (1975). ECONOMETRIC METHODS (segunda edición), Mc. Graw-Hill Kogakusha, p.p. 33-35 [7]  
Capt. 5 [8]

Kaplan, Wilfred (1985). *CALCULO AVANZADO* (segunda edición), p.p. 125 - 133 y 220 - 224.

Kendall, M. G. (1961). *A COURSE IN THE GEOMETRY OF N DIMENSIONS* (primera edición) [3]

Kendall and Stuart (1957). *THE ADVANCED THEORY OF STATISTICS*. (fourth edition), Vol 1, p.p. 270 - 276 y 411 - 421.

Margolis, Marvin S. (1979) *PERPENDICULAR PROJECTIONS AND ELEMENTARY STATISTICS*. *The American Statistician*, Vol. 33, N° 3.