



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

"ANÁLISIS DE LA MORTALIDAD INFANTIL DE LOS AÑOS  
1999 Y 2000, A TRAVÉS DE COMPONENTES PRINCIPALES  
Y DISCRIMINANTE"

**T E S I S**

QUE PARA OBTENER EL TÍTULO DE:

**A C T U A R I A**

P R E S E N T A:

**MARISOL LUNA CONTRERAS**



DIRECTORA DE TESIS: M. en A. P. MA. DEL PILAR ALONSO REYES



2004

FACULTAD DE CIENCIAS  
SECCION ESCOLAR



Universidad Nacional  
Autónoma de México



**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 esta 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.

ESTA TESIS NO SALE  
DE LA BIBLIOTECA





UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

NOMBRE: Marisol Luna Contreras  
FECHA: 30-04-04  
FIRMA: [Signature]

**ACT. MAURICIO AGUILAR GONZÁLEZ**  
**Jefe de la División de Estudios Profesionales de la**  
**Facultad de Ciencias**  
**Presente**

Comunicamos a usted que hemos revisado el trabajo escrito:

"Análisis de la Mortalidad Infantil de los años 1999 y 2000, a través de componentes principales y discriminante".

realizado por

Luna Contreras Marisol

con número de cuenta 09423540-5 , quien cubrió los créditos de la carrera de: Actuaría.

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

- |                               |   |                    |
|-------------------------------|---|--------------------|
| Director de Tesis Propietario | M. en A.P. María del Pilar Alonso Reyes | <u>[Signature]</u> |
| Propietario                   | M. en C. José Antonio Flores Díaz       | <u>[Signature]</u> |
| Propietario                   | Act. Jaime Vázquez Alamilla             | <u>[Signature]</u> |
| Suplente                      | Act. Lucio Gerardo Chávez Heredia       | <u>[Signature]</u> |
| Suplente                      | Act. MaryPaola Janett Maya López        | <u>[Signature]</u> |

Consejo Departamental de Matemáticas

[Signature]  
M. en C. José Antonio Flores Díaz

# AGRADECIMIENTOS

A lo largo de mi vida muchas personas han estado inmersas en mi desarrollo, no sólo como ser humano, sino también, académico. Por lo cual, dedico este trabajo a todos ellos, ya que fue fruto del esfuerzo conjunto de todos ellos y mío. Tal vez, no se hayan dado cuenta de esto, pero con el simple hecho de darme una palabra de apoyo para que continuara fue más que suficiente para que yo viera que todo es posible en esta vida (bueno, casi todo).

Sin temor a que mis padres se sientan un poco desplazados, quiero empezar agradeciendo a dos personas maravillosas, las cuales desgraciadamente ya no están conmigo, mi abuelita Consuelo y mi tía Luisa, pues, sin sus cuidados y hasta sin sus regañíos, tal vez ahora no sería la persona que soy. Creo que no sólo yo les estoy agradecida sino también mis papás. A mi abuelita le agradezco que me haya enseñado el valor de ser mujer. A mi tía Luisa le estoy en deuda, por todos esos consejos que me dio cuando más los necesitaba, y por todas las veces que lloró conmigo y me consoló. En fin les agradezco a ambas por haber sido mis "grandes mamás".

A mis papás les agradezco todo su apoyo, tanto económico como moral, espero algún día recompensarlos con todo lo que se merecen. Y recuerden que puede que no siempre este de acuerdo con ustedes, pero la manera en como pienso y como soy es fruto de sus enseñanzas, consejos y hasta de las cosas en las que no concordamos.

A mis tres hermanos, Rosalía, José Ramón y Monserrat, les agradezco el estar conmigo y saber que puedo contar con ustedes siempre que los necesite, y que me hayan aguantado mis cambios de humor cuando me sentía más presionada.

A todos mis amigos: Ana Lilia, Esperanza, Yesenia, Mirra, Anabel, Lupita, Nora, Raquel, Omar Riverroll, Omar, Rodrigo, Edgar, Carlos (El pareja), Carlos (Charlie Boy), Samuel, Nadia, Maribel, Sandra, Lola. Gracias por compartir conmigo más que un lugar en el salón de clases, así como por el apoyo que me han brindado, y sobre todo por su gran amistad.

A mis amigos de la secundaria y preparatoria: Nohemí, Carlos Nieves y Jerónimo. En fin a todos aquellos que han dejado huella en mi vida.

Gracias a mis sinodales por tomarse la molestia de leer mi trabajo y ayudarme a mejorarlo: M. en P. María del Pilar Alonso, M. en C. José Antonio Flores, Act. Jaime Vazquez, Act. Gerardo Chávez, Act. Maripaola Maya.

También le agradezco a la Dra. Catalina Stern por darme la facilidad de trabajar en el Laboratorio de Acústica. También les agradezco a ustedes, Manuel y César, por permitir compartir un espacio en el laboratorio, espero no haberles causado muchas molestias.

Señora Leticia Mayén, le agradezco el apoyo que me ha brindado en estos años, sus consejos me han ayudado mucho. Señor Benjamín García, gracias por haber leído mi trabajo sus consejos fueron muy valiosos para mí.

Gracias a la Universidad Nacional Autónoma de México por haberme permitido ser parte de ella.

Finalmente, Héctor te agradezco que seas un gran motor para mi desarrollo personal y académico, y por apoyarme en todos los proyectos que tengo, así como por el tiempo que invertiste en ayudarme a realizar este trabajo. Gracia por tu cariño, y por ser un gran novio y amigo al mismo tiempo.

# ÍNDICE

## INTRODUCCIÓN

### CAPÍTULO I: Base teórica sobre mortalidad y fuentes de información

<b>1.1</b>	<b>INTRODUCCIÓN</b>	<b>1</b>
<b>1.2</b>	<b>MORTALIDAD</b>	
1.2.1	Definición de mortalidad	10
1.2.2	Características fundamentales	10
1.2.3	Causas de la mortalidad	11
1.2.4	Mortalidad infantil	12
<b>1.3</b>	<b>FUENTES DE INFORMACIÓN</b>	
1.3.1	Estadísticas vitales	13
1.3.2	Encuesta Nacional de Empleo	15
1.3.3	Boletín de Información Estadística	16

### CAPÍTULO II: Metodología

<b>2.1</b>	<b>ANÁLISIS DE COMPONENTES PRINCIPALES</b>	
2.1.1	Introducción	18
2.1.2	Planteamiento general del problema	19

2.1.3	Interpretación geométrica de los componentes principales	23
2.1.4	Definiciones y propiedades	24
2.1.5	Componentes principales en una población	24
2.1.6	Estandarización de los componentes Principales	31
2.1.7	Componentes principales generados a partir de una muestra	32
2.1.8	Interpretación geométrica de los componentes principales bajo normalidad de observaciones	36
2.1.9	Estimación máxima verosímil para datos normales	37
2.1.10	Intervalo de confianza para una valor propio	38
2.1.11	Pruebas de hipótesis sobre los componentes principales	40
2.1.12	Reglas de corte	44
2.1.13	Algunas herramientas adicionales para los componentes principales	45

## **2.2 ANÁLISIS DISCRIMINANTE**

2.2.1	Introducción	47
2.2.2	Formulación del problema	48
2.2.3	Clasificación en dos poblaciones normales multivariadas	52
2.2.4	Matrices de covarianza diferentes	58
2.2.5	Regla discriminante de la razón de verosimilitudes	67



## **CAPÍTULO III: Análisis Estadístico**

<b>3.1</b>	<b>INTRODUCCIÓN</b>	<b>73</b>
<b>3.2</b>	<b>ANÁLISIS ESTADÍSTICO DEL AÑO 1999</b>	
3.2.1	Análisis de Componentes Principales	75
3.2.2	Análisis Discriminante	84
<b>3.3</b>	<b>ANÁLISIS ESTADÍSTICO DEL AÑO 2000</b>	
3.3.1	Análisis de Componentes Principales	88
3.3.2	Análisis Discriminante	94
	<b>CONCLUSIONES</b>	<b>98</b>
	<b>APÉNDICE</b>	<b>101</b>
	<b>GLOSARIO</b>	<b>106</b>
	<b>BIBLIOGRAFÍA</b>	<b>108</b>

# INTRODUCCIÓN

Uno de los campos de estudio en la investigación de la dinámica demográfica es la mortalidad, cuyo comportamiento se encuentra asociado a las condiciones de vida de la población y a su estado de salud; éstos a su vez se ven influenciados por los modos de organización social, que determinan los mecanismos de socialización de la producción, circulación, distribución y consumo de los bienes y servicios requeridos para la reproducción social y la generación de la riqueza individual y colectiva. Las consecuencias de la distribución de esta riqueza se manifiesta de distinta manera, una de las cuales es la conformación de condiciones de vida diferentes, donde la mortalidad se expresa en forma diferencial por grupos de edad, sexo, ocupacionales y en general estratos socioeconómicos, así como por regiones.

Como reflejo de las condiciones ambientales de carácter económico, social y cultural, por mencionar algunos, la mortalidad infantil es particularmente significativa, al referirse a un grupo sumamente vulnerable –ya que esta se refiere a las defunciones de niños menores de un año-. Como consecuencias de ello, el nivel que presenta la mortalidad infantil permite conocer el grado en que se satisfacen algunas necesidades sociales básicas, entre las que destacan las educativas, culturales, médicas e ideológicas; todo esto se relaciona con el papel que juega la formación e información de los padres para la toma de decisiones en la atención materno-infantil con respecto al tipo de alimentación, protección sanitaria y el espaciamiento de los hijos, así como de lo económico –donde el tipo de ocupación y el monto de los ingresos de la población determinan las alternativas para satisfacer las distintas necesidades materiales-.

En la historia de México la mortalidad infantil ha sido de grandes dimensiones, disminuyendo paulatinamente a lo largo de los años, esto a través de los avances en materia educativa y de salud, así como otros factores socioeconómicos.

El estudio de la mortalidad infantil se ha realizado desde diversos puntos de vista. El que se utilizó en este trabajo fue agrupar a las entidades federativas de la República Mexicana según su mortalidad infantil, pero considerando otras variables; esto para poder darnos

una idea de las posibles diferencias en su comportamiento. Se optó por el análisis a nivel de las entidades federativas, pues, cuando se analiza por total nacional se pierden las diferencias que pudieran existir a nivel estatal; por lo que el uso de técnicas como componentes principales y discriminante permite hacer un análisis de diversas variables junto con la del índice de mortalidad infantil, dando una dimensión multivariada del fenómeno.

En este sentido, en el capítulo I del trabajo se definen algunos conceptos relacionados con la mortalidad como son algunas definiciones y características más importantes de ella, incluyendo un breve análisis de la evolución de la mortalidad dentro del República Mexicana, todo esto haciendo énfasis en la mortalidad infantil. También se da un marco general de las fuentes de información que se utilizaron para la realización del trabajo, las cuales son: las estadísticas vitales, la Encuesta Nacional de Empleo (ENE) y el Boletín de Información Estadística.

El capítulo II se dividió en dos partes, en la primera se describe detalladamente la técnica del análisis multivariado denominado componentes principales, mostrando algunos de los teoremas y resultados más importantes. En la segunda parte, se expone la técnica del análisis discriminante, sus características matemáticas y sus cualidades.

El capítulo III muestra el análisis de la mortalidad infantil en las distintas entidades federativas de los años 1999 y 2000, esto a través de componentes principales y discriminante, donde además se consideraron otras variables como: población económicamente activa e inactiva (masculina y femenina), número de camas censables, número de hospitalización general, número de unidades de consulta externa, número de quirófanos y número de laboratorios; todo esto con el objetivo de obtener la agrupación del fenómeno multivariado.

Por último se muestran las conclusiones a las que se llegaron, así como un breve glosario de términos y un apéndice de conceptos matemáticos que se utilizaron a lo largo del trabajo.

# CAPÍTULO I: Base teórica sobre mortalidad y fuentes de información

## 1.1 INTRODUCCIÓN

Entre los componentes del comportamiento de la población, la mortalidad ha desempeñado históricamente un papel importante en la determinación del crecimiento poblacional.

Según Sergio Camposoterga Cruz de la Universidad Autónoma de Hidalgo (1997), la evolución de la mortalidad en México durante el siglo XX puede dividirse en cuatro etapas:

- La primera, que comprende hasta 1920, se caracteriza por la permanencia de niveles de mortalidad altos y fluctuantes; donde la esperanza de vida al nacer<sup>1</sup> pasó de 25.4 años en 1900 a 28.9 años en 1920, con un decremento en 1915 en donde se redujo a 24.2 años (esto último a causa de la guerra de Revolución)
- Ya en los años treinta la esperanza de vida al nacer alcanza los 34 años y en 1940 los 40.5 años. La disminución de la mortalidad en este periodo se relaciona con las modificaciones en el nivel de vida de la población, ya que durante esos años se empiezan a notar los primeros efectos de la Revolución, como son los primeros programas de salud pública y modificaciones en favor de las condiciones sociales.
- La disminución de la mortalidad presenta una aceleración muy marcada de 1940 a 1960, donde la esperanza de vida al nacer aumenta de 40.4 a 57.8 años. Esta aceleración es una de las más rápidas observadas a nivel mundial; los países europeos nunca lograron incrementos superiores a medio año y muy pocos países en vías de desarrollo han obtenido incrementos similares en periodos tan cortos. La disminución de la mortalidad en este periodo radica en la introducción de las nuevas

---

<sup>1</sup> Esperanza de vida al nacer es el número promedio de años que vivirían los integrantes de una población que permaneciese sujetas a la mortalidad imperante desde su nacimiento hasta su extinción.

técnicas médicas de prevención y salud, así como los programas de eliminación de agentes patógenos. En el progreso económico y social del país, en dichos años, también son considerados como factores determinantes.

- Entre 1960 y 1995, la velocidad en la disminución de la mortalidad fue más lenta, en comparación con los periodos antes mencionados. El incremento de la esperanza de vida al nacer pasó de 0.9 años durante los años cuarenta y cincuenta a 0.4 años desde los años sesenta. Siendo en 1995 la esperanza de vida al nacer de 72.3 años. Esta desaceleración se encuentra asociada a las causas de la defunción, tales como: muertes asociados a los accidentes y violencia, enfermedades circulatorias y cáncer.

De acuerdo a las estimaciones de la tabla 1.1 (pág. 3), en 1999 las estadísticas reflejan que la mortalidad general en México se ubica en 4.8 defunciones por cada mil habitantes. Haciendo un análisis por entidad federativa y a nivel nacional las cifras muestran lo siguiente:

- En el grupo de menores de 15 años de edad, los valores fluctúan entre 4.6% en Durango y 19.6% en Puebla. La proporción de muertes para el país es de 12.6%. En el rango de 6.0% y 9.9% se encuentran los siguientes estados, Coahuila, Distrito Federal, Morelos, Nayarit, Nuevo León, Sinaloa, Sonora y Tamaulipas. Los valores más altos del promedio nacional los tienen las entidades con un proceso más lento en lo social y económico.
- En el grupo de 15-64 años, el promedio nacional se sitúa en 37.8%. El valor más alto lo tiene Quintana Roo con el 48.1%. En quince de las 32 entidades, el porcentaje de defunciones sobrepasan la cifra nacional, y sólo Sinaloa y Morelos registran valores que igualan al del país; en las quince entidades restantes, los valores varían entre el 29.8% de Yucatán y el 37.6% de Oaxaca.
- Con respecto al grupo de entidades de 65 años y más, tienen como principal característica que la mayoría de las entidades cuenta con una alta concentración de muertes. A nivel nacional las defunciones en estas edades representan casi el 50% del total de defunciones, cifra que supera en 12 puntos porcentuales a las defunciones correspondientes al grupo de edades 15-64 años y sólo cuatro entidades, Baja California, Chiapas, Estado de México y Quintana Roo, los porcentajes de este grupo de edades son

inferiores a los obtenidos en dicho grupo. En los 28 estados restantes los datos varían en 42.6% de Tabasco y el 57.7% en Nayarit.

**Tabla 1.1 México, proporción de muertes por grandes grupos de edad según entidad federativa, 1999**

ENTIDADES	GRUPOS DE EDAD		
	MENORES DE 15	15-64	65 Y MÁS
Aguascalientes	14.2	34.2	51.6
Baja California	13.7	46.4	39.9
Baja California Sur	11.7	40.9	47.4
Campeche	11.8	40.1	48.2
Coahuila	7.4	39.7	52.8
Colima	9.5	35.6	54.8
Chiapas	16.9	44.7	38.5
Chihuahua	11.3	41.3	47.5
Distrito Federal	8.7	35.9	55.4
Durango	4.6	38.2	57.2
Guanajuato	16.0	32.3	51.7
Guerrero	12.1	43.3	44.6
Hidalgo	12.5	38.1	49.4
Jalisco	10.9	33.5	55.6
Edo. de Méx.	17.9	41.3	40.9
Michoacán	10.0	36.0	54.0
Morelos	9.9	37.8	52.3
Nayarit	7.9	34.3	57.7
Nuevo León	8.5	36.3	55.2
Oaxaca	13.5	37.6	48.9
Puebla	19.6	34.5	45.9
Querétaro	16.8	37.2	46.0
Quintana Roo	19.0	48.1	32.9
San Luis Potosí	12.6	32.4	55.0
Sinaloa	6.4	37.8	55.8
Sonora	9.9	39.7	51.3
Tabasco	15.6	41.8	42.6
Tamaulipas	8.5	39.8	51.7
Tlaxcala	19.3	32.1	48.6
Veracruz	11.3	39.5	49.2
Yucatán	10.2	29.8	59.9
Zacatecas	11.2	30.3	58.4
<b>TOTAL</b>	<b>12.6</b>	<b>37.8</b>	<b>49.6</b>

Fuente: Morelos, José B. (2000), "La mortalidad general a finales del segundo milenio", en DEMOS: Carta Demográfica sobre México, 2000, México: IISUANM, pp. 14-16.

Los índices, antes mencionados, han sido influenciados por los cambios en las estructuras por edades de las poblaciones de las distintas entidades; dichas variaciones resultan del efecto combinado de la caída de la mortalidad y de la baja de la natalidad. Cuando se presenta una disminución en la tasa de natalidad<sup>2</sup> y la población en su mayoría es adulta, da como resultado una reducción en las muertes de los menores de 15 años y aumenta la mortalidad en los otros dos grupos. Este fenómeno también ocurre a la inversa.

Numerosos indicadores provenientes del análisis de la mortalidad son de gran utilidad para el diagnóstico y análisis de las condiciones de vida de la población. La mortalidad infantil ha sido reconocida como un indicador "clásico" de las condiciones de vida de la población. También se le suele denominar como un indicador internacional que mide el estado de salud de una población.

En México, desde el final de la Revolución, la mortalidad infantil ha mostrado un descenso continuo. De acuerdo con las estadísticas vitales, la Tasa de Mortalidad Infantil (TMI)<sup>3</sup> en 1930 era de 130 por mil; en la mitad de siglo XX fue de 100 por mil y para 1995 fue de 17 defunciones por cada mil nacidos vivos. Por otro lado, la tendencia e disminución no ha sido uniforme:

- Entre 1950 y 1960, la reducción fue superior al 30%, pasando de 132 a 90 muertes de menores de un año por cada mil nacidos vivos.
- De 1960 a 1970 decreció un 15%.
- Entre 1970 y 1980 se presentó una disminución más significativa, la cual fue superior al 30%.
- Según datos del Consejo Nacional de Población (CONAPO), entre 1980 y 1987 la disminución de TMI fue sólo del 14%, al pasar de 53 a 47 defunciones por mil nacidos vivos.

---

<sup>2</sup> La tasa de natalidad indica el número de nacidos vivos ocurridos por cada 1000 habitantes en un año determinado.

<sup>3</sup> La TMI es el cálculo de la división de las defunciones de menores de un año ocurridas en un año calendario entre el número de nacidos vivos el mismo año.

## BASE TEÓRICA SOBRE MORTALIDAD

Los factores que explican estas diferencias en la disminución de TMI están determinados por las condiciones de vida de la población, pues, la mortalidad infantil es sensible a la presencia o ausencia de recursos orientados a su atención, sobre todo si se consideran que gran parte de estas muertes son evitables. La gran disminución que se presentó entre 1970 y 1980, coincide con la puesta en práctica de una política de protección social y de apoyos orientados a zonas rurales y a grupos marginales urbanos. En cambio, entre 1980 y 1990, estas acciones han tenido un escaso respaldo presupuestal, resultado de una reducción considerable del gasto público en aspectos de bienestar social. Pero a pesar de las crisis económicas por las que ha atravesado nuestro país en los últimos años, la TMI no ha detenido su descenso, dicha baja es el resultado de varios factores como son: el uso de terapia de hidratación oral que ha combatido la mortalidad por infecciones intestinales. Por otro lado, a causa de la práctica de la planificación familiar cada vez nacen menos niños con alto riesgo de morir, como los hijos de mujeres en los extremos del periodo reproductivo, los niños de orden alto (del cuarto en adelante), así como aquellos entre los que hay un intervalo intergenésicos más cortos (menos de dos años), así como la intensificación y continuidad de las campañas de vacunación han sido parte de los elementos responsables de los mejoramientos de los indicadores de salud infantil en el país.<sup>4</sup>

Alejandro Aguirre, del Colegio de México, en 1997 hizo un estudio en donde analiza diferentes causas de muerte infantil y las agrupa según lo fácil o difícil que es evitarlas, de la siguiente manera:

- Enfermedades Inmunoprevenibles (EIP).
- Enfermedad Diarreica Aguda (EDA).
- Infección Respiratoria Aguda (IRA).
- Afecciones Perinatales (AP).
- Anomalías Congénitas (AC).

---

<sup>4</sup> Cárdenas, Rosario, "La disminución de la mortalidad y las causas de muerte", DEMOS: Carta Demográfica sobre México, IISUNAM, 1998, p. 11.



## CAPÍTULO I

El grupo más fácil de controlar es el de las EIP, ya que con la aplicación de una o varias dosis de vacunas, en la mayoría de los casos es posible evitar estas enfermedades. La mortalidad por EDA puede prevenirse de manera relativamente fácil con la terapia de hidratación oral. Combatir las IRA es un poco más difícil, pero con una serie de medidas de atención primaria oportuna de salud se pueden evitar. En cuanto a las defunciones por AP es necesario emplear medidas más costosas, como son una adecuada atención prenatal y durante el parto, pero para ello es necesario de infraestructura hospitalaria, así como contar con recursos humanos capacitados, elementos que están asociados al desarrollo económico. Por otro lado, las AC no sólo se consideran difíciles de evitar sino también de detectar, por lo que con los conocimientos científicos y tecnología actuales, disponibles en el país, es poco lo que se puede hacer en este campo para disminuir la mortalidad infantil.

**Tabla 1.2 TASAS DE MORTALIDAD INFANTIL DERIVADAS DE ESTADÍSTICAS VITALES Y ESTIMADAS Y PRINCIPALES CAUSAS DE MUERTES 1922-1995**

	TMI			CAUSAS			
	Porcentaje de certificación	EV	Estimada	1	2	3	4
1922		222		EDA	IRA	AC	TF
1930		130	189	EDA	IRA	AC	TF
1940	52	126	163	EDA	IRA	AC	Violencia
1950	53	96	129	IRA	AP	EDA	TF
1960	63	74	90	AP	EDA	IRA	AC
1970	74	68	76	ARA	EDA	AP	AC
1980	86	40	57	AP	EDA	IRA	AC
1990	94	24	41	AP	EDA	IRA	AC
1995	96	17	30	AP	AC	IRA	EDA

TF=tosferina, EIP=enfermedades inmunoprevenibles, EDA=enfermedad diarreica aguda, IRA= infección respiratoria aguda, AP=afecciones perinatales, AC=acciones congénitas.

Nota: la clasificación de las causas de muerte ha cambiado de la 3ª revisión de la CIE en 1992 a la 9ª revisión en 1995.

Fuentes: SSA, Dirección General de Estadística, Informática y Evaluación, *Compendio Histórico Estadísticas Vitales 1893-1993*, Estados Unidos Mexicanos, SSA, Dirección General de Estadística e Informática, *Mortalidad 1995*. Campsoriga, S. "Cambios en la mortalidad en México", DEMOS, 1997.

La tabla 1.2 (pág. 6) muestra algunos datos acerca de las principales causas de mortalidad infantil que se han presentado desde 1922 hasta 1995. Se puede observar que, hasta 1930 la causa dominante de mortalidad infantil fueron los padecimientos infecciosos, como son: las

## BASE TEÓRICA SOBRE MORTALIDAD

diarreas, infecciones respiratorias y los padecimientos inmunoprevenibles. En 1922 y en 1930 las EDA y las IRA ocupan los dos primeros lugares, y una EIP, la TP se ubican en el cuarto. Al ir avanzando los años, los padecimientos inmunoprevenibles pierden importancia, y es común que las enfermedades diarreicas y las infecciones respiratorias ocupen los dos primeros lugares entre las causas de mortalidad infantil, y para 1970 las afecciones perinatales ocupan el tercer lugar. Entre 1980 y 1990 las EDA y las IRA dejan de ser las dos causas más importantes de mortalidad infantil, descendiendo al segundo y tercer sitios, y las AP pasaron al primer lugar; por otro lado las AC alcanzaron el cuarto lugar. Finalmente, en 1995 la principal causa de mortalidad infantil fueron las AP, AC, IRA y EDA.

Antonio Escobedo Aguirre, del Instituto Nacional de Estadística, Geografía e Informática (INEGI), en 2000, obtiene tasa refinadas de mortalidad infantil<sup>5</sup>, las cuales describen algunas relaciones cuantitativas entre ellas y tres rasgos socioeconómicos de la población femenina en edades reproductivas, los cuales son:

- Escolaridad.
- Derechohabencia a servicios médicos.
- Participación en la actividad económica.

Su objetivo era mostrar los niveles de mortalidad infantil y sus diferenciales según algunas características socioeconómicas de las mujeres en edades entre 12 y 49 años, que declararon haber tenido hijos nacidos vivos durante 1999.

Obtuvo como resultados que los niveles del indicador de la mortalidad infantil (Tabla 1.3) son mayores mientras menor sea el tamaño de la localidad de residencia, pasando de 13.80 para las localidades con mas de 100 mil habitantes a 23.08 en las localidades rurales.

En cuanto a la instrucción de la madre, se puede ver que los valores más altos a nivel nacional, corresponden a las mujeres sin instrucción 28.1, a quienes participan en actividades económicas reenumeradas 21.2 y a las no derechohabientes de servicios médicos 20.2; en

<sup>5</sup> Referirse a CELADÉ, *Mortalidad*, Apuntes de clase, 1990.

## CAPÍTULO I

tanto que los valores más bajos son: 13.4 para las derechohabientes de servicios médicos, 16.5 a las económicamente inactivas y 16.6 para quienes tienen instrucción escolar, como se muestra en la Tabla 1.3.

Los resultados de Escobedo Aguirre, muestran que las mujeres sin instrucción tienen vinculados los niveles de mortalidad infantil más altos, siguen en orden de magnitud las que participan en actividades económicas reenumeradas, las no derechohabientes, las que tienen instrucción escolar, posteriormente las mujeres económicamente inactivas y en la mejor situación las derechohabientes a servicios de salud.

**Tabla 1.3 Tasas refinadas de mortalidad infantil por tamaño de localidad, según instrucción, derechohabiencia a servicios médicos y condición de actividad 1999**

Tamaño de localidad y grupos de edad	Mujeres de 12 a 49 años	Sin instrucción	Con instrucción	Sin derechohabiencia	Con derechohabiencia	PEI	PEA
<b>Total</b>	17.51	28.13	16.61	20.15	13.35	16.49	21.20
<b>Menos de 2500 habitantes</b>	23.08	28.51	21.85	24.12	17.93	21.82	31.71
<b>2500 a 14999 habitantes</b>	20.54	33.65	19.03	21.38	18.68	18.76	28.37
<b>15000 a 99999 habitantes</b>	14.46	21.86	14.13	15.87	12.68	13.82	16.65
<b>100 mil y más habitantes</b>	13.80	23.67	13.63	16.27	11.81	12.22	17.67

Fuente: XII Censo General de Población y Vivienda 2000. Indicadores derivados del cuestionario ampliado (información estimada).

En cuanto al plano internacional, México en el año 2001 ocupó el 11° lugar, con respecto de otros países de América Latina con una TMI del 27%, es decir, murieron 27 niños menores de un año por cada mil nacidos vivos; Cuba y Canadá presentaron una TMI del 6%, siendo esta la más baja de toda la región llamada de las Américas, y la TMI más alta la tuvo Guatemala con 45 niños menores de un año muertos por cada mil que nacieron vivos. Cuba alcanzó niveles tan bajo de mortalidad infantil debido a la atención que el gobierno ha puesto en cuanto a cuidados en el periodo de gestación, lo cual se muestra en la Tabla 1.4.

Tabla 1.4 Tasas de Mortalidad infantil en las Américas 2001

Países	Tasa de Mortalidad
Cuba y Canadá	6
E.U.A.	7
Chile	11
Costa Rica	13
Uruguay	15
Argentina	19
Venezuela	20
Panamá	21
Colombia	26
Ecuador	27
México	27
Paraguay	27
Brasil	34
El Salvador	35
Nicaragua	28
Perú	42
República Dominicana	42
Guatemala	45

Fuente: UNICEF, Estado Mundial de la Infancia, 2001

La actual situación de la mortalidad infantil en México puede explicarse por las grandes desigualdades socioeconómicas y regionales que se viven en el país y sólo se pueden combatir con un mejoramiento en las condiciones de vida de la población más necesitada y aquí es donde entran las soluciones, en donde el sector gubernamental debe asignar mayores recursos al bienestar social. A lo largo de las últimas décadas, se ha demostrado que la mortalidad infantil resulta un fenómeno sensible a la existencia o ausencia de medios elementales para la supervivencia, como son: alimentación, educación, trabajo y salud. Es importante señalar que la mejora en los servicios de salud no sirven de nada si la población no tiene acceso a ellos.

## 1.2 MORTALIDAD

### 1.2.1 Definición de mortalidad

La mortalidad es empleada para expresar la acción de la muerte sobre los integrantes de una población. Para muchos sociólogos, demógrafos y demás profesionales del área, el concepto de mortalidad señala la salida de un individuo de la población debido a su fallecimiento.<sup>6</sup>

El Diccionario de Sociología de Giner define mortalidad como el fenómeno demográfico que se refiere a la acción de la muerte sobre una población. Los sucesos individuales se denominan muertes o defunciones, y a la suma de ellos en una población y en un momento dado se le llama mortalidad. Y dice que la demografía estudia las condiciones subyacentes que influyen en la existencia de mortalidad, así como el impacto que está tienen sobre el tamaño y distribución de la población.

### 1.2.2 Características fundamentales

Al realizar una análisis de la mortalidad, es importante mencionar algunas de sus características fundamentales:

- La muerte es un riesgo al que se está expuesto a lo largo de toda la vida y es un hecho que sólo le ocurre una vez a cada persona.
- Los determinantes de la mortalidad, están íntimamente ligados a factores biológicos, así como a la estructura por edad de la población.<sup>7</sup>
- Existe un criterio consensual, tanto a nivel científico, político y ético, respecto a la necesidad de reducir la mortalidad, la cual se debe de posponer lo más posible. Todo esto facilita la definición de programas y políticas para enfrentarla.

---

<sup>6</sup> Ortega, Osana José Antonio, "Revisión de conceptos demográficos", <http://www.eumed.net>, 2001.

### 1.2.3 Causas de la mortalidad

Toda defunción tiene una causa, entendiéndose por tal la enfermedad, traumatismo o lesión que provocan la muerte.

Existen diversas formas de clasificar las defunciones según su causa. La más general es la que la clasifica en dos grupos, según su naturaleza:

- a) *Las causas endógenas:* provienen de la constitución genética del individuo, de las malformaciones congénitas, del traumatismo provocado por el nacimiento o de la degeneración producida por el envejecimiento del organismo.
- b) *Las causas exógenas:* corresponden a circunstancias o factores externos asociados a los individuos, como son las enfermedades infecciosas y parasitarias y los traumatismos accidentales.

Otra clasificación que es utilizada por las Naciones Unidas agrupa las enfermedades según su comportamiento frente a la acción sanitaria, es decir, su mayor o menor resistencia a los progresos médicos y a los programas de la salud. Consta de los siguientes cinco grupos:

*GRUPO I:* enfermedades infecciosas y parasitarias, enfermedades del aparato respiratorio, gripe, neumonía y bronquitis antes de los cinco años.

*GRUPO II:* cáncer.

*GRUPO III:* enfermedades cardiovasculares y bronquitis después de los cinco años.

*GRUPO IV:* violencia.

*GRUPO V:* restantes causas de muerte y causas mal definidas y desconocidas.

---

<sup>7</sup> La estructura de una población es una reclasificación en subgrupos de la población ya sea por edad y/o sexo.

Se puede reclasificar estos cinco grupos en causas endógenas y exógenas; los grupos I, IV y V corresponden a causas exógenas y los grupos II, III a causas endógenas.

Es importante decir que la distribución por causas depende de la estructura por edad, ya que una población con una estructura envejecida tenderá a registrar una mayor proporción de muertes debidas a enfermedades degenerativas, como son el cáncer y las enfermedades cardiovasculares y las pertenecen a las causas endógenas. En cambio, una población joven tendrá una mayor proporción de muertes debido a enfermedades de tipo exógeno, como los accidentes y las enfermedades infecciosas. También a medida que los progresos médicos se imponen, la mortalidad debida causas endógenas se ha tornado más importante.

#### 1.2.4 Mortalidad infantil

Se denomina mortalidad infantil aquella que ocurre antes de cumplir un año. Este tipo de mortalidad presenta intensidades muy altas, las cuales se ven reflejadas en las tasas de mortalidad que son significativamente mayores que las registradas en las edades siguientes.

Dentro del primer mes de vida la mortalidad no es uniforme. Esto se debe, a que en dicho lapso el niño está expuesto a una elevada probabilidad de morir causada ya sea por factores genéticos, por problemas congénitos o daños que se producen durante la gestación o el nacimiento. Después del primer mes de vida, empiezan a cobrar importancia los factores relacionados con las condiciones del medio en que se desarrolla el niño como son nutrición, exposición a microorganismos, condiciones sanitarias, etc. Por lo anterior y para fines de análisis, la mortalidad infantil se divide en:

- *Mortalidad neonatal*: comprende la mortalidad que ocurre dentro del primer mes de vida, es decir, desde el nacimiento y hasta antes de cumplir 28 días.
- *Mortalidad postneonatal*: ésta se refiere a la mortalidad que ocurre entre el momento de cumplir 28 días de nacido y hasta antes del primer cumpleaños.

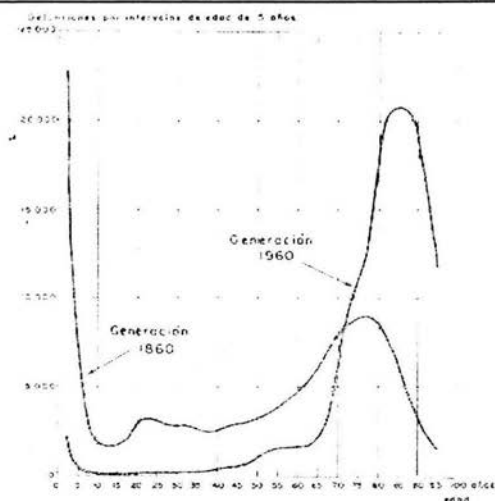


Figura 1.10. Distribución de las defunciones, según la edad, en las generaciones femininas

La mortalidad infantil es igual a la suma de las dos anteriores. La mortalidad postneonatal por estar vinculada a causas exógenas, es más fácil de evitar, ya sea con programas masivos de vacunación, mejoramiento en la higiene ambiental, así como con la introducción de tratamientos de bajo costo, por ejemplo, terapias de rehidratación oral para contrarrestar la diarrea. En cambio, la mortalidad neonatal se relaciona más con enfermedades o malformaciones, las cuales son más difíciles de evitar y también son más costosas de atender. El INEGI da otra subdivisión de la mortalidad, aparte de las dos mencionadas anteriormente, la denominada mortalidad hebdomadal, la cual se refiere a la que ocurre durante los primeros 7 días de vida.

La mortalidad infantil es un foco de atención para todas las políticas de salud de cualquier país, ya que se ha comprobado que una adecuada política de salud puede bajar la mortalidad infantil sin los costos que significa reducir la mortalidad en otras edades. Programas de bajo costo y de una ejecución relativamente baja, como vacunaciones masivas, controles preventivos, entrega de alimentos, etc., pueden ayudar a evitar un importante número de muertes.<sup>8</sup> Como se menciona en la sección 1.1 la mortalidad infantil es considerada como un

<sup>8</sup>Walti, Chanes Carlos, "México: Programa Latinoamericano de Actividades de Población", Instituto de Investigaciones Sociales UNAM, 1997, México. p.81



muy buen indicador de las condiciones de salud y mortalidad de una población, siendo éste un indicador para la medición del desarrollo social.

## 1.3 FUENTES DE INFORMACIÓN

### 1.3.1 Estadísticas vitales

La fuente básica para el estudio de la mortalidad es el sistema de estadísticas vitales. Las estadísticas vitales son el resultado del recuento de los hechos más importantes de la vida de población de un país, como son los nacimientos, defunciones, defunciones fetales, matrimonios y divorcios<sup>9</sup>, por lo que se generan estadísticas de cada uno de estos sucesos. En este caso en particular sólo interesará para este trabajo las estadísticas de defunciones, pues, es la que está ligada directamente al fenómeno de la mortalidad.

Con el término de estadística de defunciones se identifica lo que es el recuento de las muertes ocurridas en el país refiriéndose a un período determinado, el cual es generalmente de un año. Aquí también se investigan las características principales que identifican el hecho: edad, sexo, causa de muerte, lugar de residencia habitual del fallecido, etc. Con lo anterior es posible colocar en su adecuada perspectiva el nivel de mortalidad que se está observando en el país, así como determinar las causas que están originando las muertes e identificar las diferencias que existen a nivel de entidades, municipios y localidades.

La importancia de las estadísticas vitales radica en que la generación de información demográfica en cualquier país se puede considerar como una base fundamental para analizar la composición y estructura de su población, esto permite que se propongan nuevas políticas para planear el desarrollo económico y social. La utilidad de las estadísticas vitales en cuanto al fenómeno de la mortalidad es permitir ahondar sobre la tendencia del crecimiento basándose en las tasas de mortalidad.

---

<sup>9</sup> INEGI, "El ABC de las Estadísticas Vitales", 1997, México.

## FUENTES DE INFORMACIÓN

A pesar de la gran utilidad de las estadísticas vitales existen serias deficiencias, éstas son en cuanto a su recaudación. Referente a las estadísticas de defunciones se pueden mencionar las siguientes:

- La omisión de los fallecimientos, en donde se puede encontrar ciertas diferencias según el estrato socioeconómico y la localización espacial, provocando esto la no óptima realización de un análisis comparativo, sin embargo, a pesar de esto puede ser utilizado para ciertos grupos sociales y áreas geográficas.
- Los datos registrados frecuentemente son poco precisos y carecen de confiabilidad, ejemplo de ello, es que por lo regular la causa de muerte no es señalada con claridad y en ocasiones no se consigna. Por otro lado la edad y el lugar de residencia del fallecido no se anotan en el acta de defunción.
- La información social, como nivel educacional, ocupación, estado civil, etc., es recogida escasamente, y carece de confiabilidad; esto impide que se realice un trabajo de calidad en torno a probables determinantes sociales de la mortalidad.

### 1.3.2 Encuesta Nacional de Empleo

La Encuesta Nacional de Empleo (ENE), constituye un esfuerzo conjunto realizado por el Instituto Nacional de Estadística, Geografía e Informática (INEGI) y la Secretaría de Trabajo y Previsión Social (STPS), para obtener información a nivel nacional y para las áreas más y menos urbanizadas del país.

Los temas que trata la ENE son: la composición por edad y sexo de la población, la distribución de la Población Económicamente Activa (PEA) e Inactiva (PEI)<sup>10</sup> por edad y sexo, la población ocupada por posición en el trabajo, ocupación principal, rama de actividad, horas semanales dedicadas al trabajo, causas para trabajar jornada menor a 35 horas por semana, ingresos, así como forma de pago y prestaciones. Cabe señalar que para la realización de esta tesis, sólo se considera la información de PEA y PEI por sexo.

---

<sup>10</sup> Ver glosario

El objetivo general de la ENE es la de contar con una base de información estadística sobre las características ocupacionales de la población a nivel nacional, así como sobre otras variables demográficas y económicas que permitan profundizar en el análisis de los aspectos laborales.

### 1.3.3 Boletín de Información Estadística

El Boletín de Información Estadística es presentado por la Secretaría de Salud, la cual es responsable del Sistema Nacional de Información Estadística en Salud.

El Boletín es una publicación anual y por lo general se presenta en tres volúmenes, los cuales contienen la siguiente información:

- Respecto a coberturas institucionales.
- De recursos físicos, materiales y humanos desagregados por Institución y entidad federativa.
- Información sobre tipo de unidad para ciertas variables, tales como: camas censables, camas no censables, consultorios, médicos y enfermeras.
- Información sobre las principales variables de recursos físicos, materiales, humanos y los principales servicios otorgados por municipio.
- Prestación de servicios médicos otorgados por Institución y entidad federativa, considerando aspectos como las actividades de consulta externa (general, especialización, de urgencias y odontológica).
- Servicios de hospitalización, servicios auxiliares de diagnóstico y tratamiento.
- Datos relacionados con la Asistencia Social.

## *FUENTES DE INFORMACIÓN*

---

- Información sobre los recursos otorgados por las unidades médicas privadas.
- Datos sobre lo relativo a los daños a la salud de la población como: morbilidad y mortalidad hospitalaria por entidad federativa.
- Lo referente a salud reproductiva y de la mujer, atención de salud del adulto y del anciano y sobre la salud del niño.

Para la realización de esta tesis sólo se utilizaron algunas de las variables mencionadas en los puntos anteriores.

# CAPÍTULO II: Metodología

## 2.1 ANÁLISIS DE COMPONENTES PRINCIPALES

### 2.1.1 Introducción

El Análisis de Componentes Principales (A.C.P.) es una de las técnicas del llamado Análisis Multivariado; dicho análisis se puede definir como un conjunto de técnicas que describen o modelan el comportamiento de dos o más características hechas sobre cada individuo proveniente de una o varias poblaciones.

Dicho análisis se considera como una técnica descriptiva para reducir la dimensión del espacio en el que originalmente se encuentra la muestra y en la que ningún supuesto distribucional se hace inicialmente, es decir, no existe un objeto inferencial, se trata de una técnica exploratoria que consiste en obtener el máximo de información con el mínimo posible de hipótesis. Pero, si se supone que la población tiene una distribución normal o Gaussiana, entonces la muestra observada puede ser utilizada para hacer inferencia estadísticas a partir de pruebas de hipótesis que ayuden a conocer la estructura de la población origen.

El objetivo primario es el de construir combinaciones lineales con base en las originales, tal que acumulen la mayor variabilidad de la muestra inicial. Esto se hace porque en algunas ocasiones se observa un conjunto de  $p$  variables que están correlacionadas entre sí, y es necesario transformarlas en un conjunto que contenga un menor número de variables que no estén tan correlacionadas, a las cuales se les llama *componentes principales*, las nuevas variables que subsisten al análisis guardan la información más relevante de la muestra.

El A.C.P. también es considerado un método geométrico de carácter descriptivo, visto desde esta perspectiva tiene como objetivo descubrir la estructura subyacente en un conjunto de  $n$  observaciones estudiadas bajo una serie de  $p$  variables cuantitativas.

## ANÁLISIS DE COMPONENTES PRINCIPALES

---

Los primeros trabajos que se hicieron con el A.C.P. se le atribuyen a Karl Pearson (1901), él publicó un trabajo sobre el ajuste de puntos en un espacio multidimensional a través de una línea o de un plano; su trabajo se centraba en las combinaciones lineales de variables originales para los cuales la varianza no era mínima. Estas combinaciones formaban un plano, el cual estaba en función de las variables originales y aquí el ajuste del sistema de puntos es el "mejor" por ser mínima la suma de las distancias de cada punto al plano de ajuste. Después este enfoque fue retomado por Hotelling (1933), quien fue el primero en introducir el A.C.P. tal como se conoce en la actualidad, su trabajo se centraba en el análisis de las componentes que sintetizaban la máxima variabilidad del sistema de observaciones, muy probablemente a esto se le deba el calificativo de "Principal"; por medio de la inspección de estos componentes, los cuales resumen la mayor proporción posible de la variabilidad total entre el conjunto de puntos, puede encontrarse un medio para clasificar o detectar relaciones entre los puntos.

Actualmente, de las técnicas del Análisis Multivariado, el A.C.P. se considera como una de las más conocidas y es utilizada como una técnica exploratoria. Los objetivos más importantes del A.C.P. son:

- Generar variables transformadas que puedan expresar la información del conjunto original de datos.
- Ayudar a detectar algunas de las variables originales que aportan poca información.

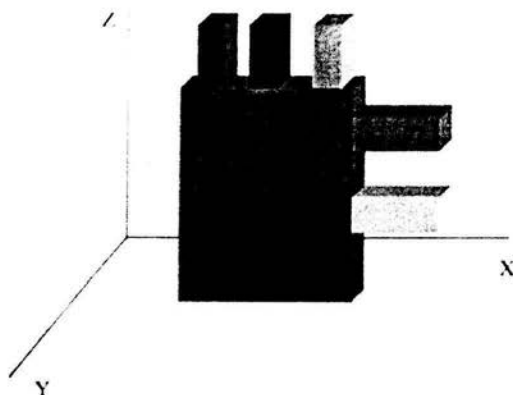
### 2.1.2 Planteamiento general del problema

Supóngase que se tienen  $n$  observaciones caracterizadas mediante una sola variable; es fácil describir a dichas observaciones representándolas sobre una recta ( $y_i = mx_i + b$ ). Si fuesen dos variables se representarían a las observaciones sobre un plano ( $y_i = mx_i + mx_2 + b$ ), de hecho si fuesen tres variables se puede recurrir a una la representación gráfica llamada "nube de puntos". Pero si el número de variables es igual o superior a cuatro sería imposible su representación gráfica.

## CAPÍTULO II

Ahora supóngase que pesar de que el conjunto de variables es superior a dos se pretende representar a las observaciones, entonces esto se hara proyectando a la nube de puntos sobre el plano. Entonces la nueva proyección sobre el plano resultaría una nube de puntos, pero está resultara deformada, pues, las distancias entre los puntos ya no van a ser las originales; por lo anterior, se intentara buscar el plano en el cual la deformación sea la mínima, es decir, aquel plano en donde la nube se extienda al máximo.

Lo anterior se puede comprender más fácilmente con los siguientes dibujos:



Las proyecciones sobre los planos serían:



## ANÁLISIS DE COMPONENTES PRINCIPALES

---

Y se puede observar que la mejor proyección se consigue sobre el plano XZ.

Como se menciona anteriormente, el A.C.P. es un método que permite transformar un conjunto de variables originales, en otro conjunto de variables llamado conjunto de componentes principales. Estos componentes principales son combinación lineal de las variables originales y se caracterizan por no estar correlacionadas entre sí. En principio se generarán  $p$  componentes principales, es decir, tantas como variables originales.

Si se toman las variables originales y se calcula la matriz de correlación se podrá observar que, por lo regular, existe un alto grado de correlación entre algunas de las variables; lo que lleva a pensar en trabajar sobre un conjunto de variables no correlacionadas de menor dimensión. Por lo cual, si las variables originales están muy correlacionadas entre sí, se esperaría que la información se pueda expresar a través de unos pocos componentes principales. Por el contrario, si las variables originales están muy poco correlacionadas entre sí, se esperará que el número de componentes principales sea muy similar al de las variables originales.

Sea  $Y$  la matriz de  $n \times p$  de variables originales y  $\Sigma$  su matriz de covarianzas, las cuales son de la siguiente manera, los pasos para obtener los componentes principales son como se muestran a continuación.

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \quad y \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{pn} \end{bmatrix}$$

### *PASO 1: buscar combinaciones lineales de las variables originales.*

Toda combinación lineal  $c$  de las variables originales puede expresarse de la siguiente forma:

$$c = Yv$$

donde  $v$  es el vector que permite obtener la combinación lineal.



**PASO 2: obtener del primer componente principal.**

El primer componente principal se define como la combinación lineal de variables originales de varianza máxima. Se busca  $v_1$  de norma uno tal que la varianza del primer componente principal  $c_1$  sea máxima.

La varianza de las componentes principales  $c$  se escribe como:

$$c = Yc$$

Donde  $V_y$  es la matriz de vectores propios asociados a la matriz de variables  $Y$ . Entonces el planteamiento del problema se puede reducir de la siguiente manera: buscar  $v$ , tal que  $v^t V_y v$  sea máxima, sujeto a  $\|v\| = 1$ .

La resolución de ese problema de maximización lleva a la ecuación:

$$V_y v = \lambda v$$

De donde se concluye que  $v$  es el vector propio de la matriz de covarianzas de los datos originales. Como se ha puesto de condición que la varianza sea máxima, se escogerá el vector propio con mayor valor propio asociado.

Es decir, el primer componente principal  $c_1$  se obtiene haciendo  $c_1 = Yv_1$ , donde  $v_1$  es el vector propio de la matriz de varianzas con mayor valor propio asociado.

**PASO 3: obtener el resto de los componentes principales.**

El segundo componente principal será aquella combinación lineal de variables originales de varianza máxima y ortogonal a  $c_1$ . Para obtener éste y los siguientes componentes principales no es necesario retomar el problema de maximización del paso anterior, se va a proceder de la siguiente manera: como toda matriz de covarianzas es simétrica y semi-definida positiva, por tanto tiene  $p$  vectores propios ortogonales dos a dos y sus valores

propios asociados son todos positivos o nulos. Los vectores propios de la matriz  $V_y$  están asociados a los valores propios escritos en forma decreciente son por tanto los vectores buscados. Estos vectores permiten calcular los componentes principales a través de la siguiente ecuación:

$$c = Yv$$

donde la varianza de cada una de los componentes principales viene dada por los valores propios.

Finalmente, el número de valores propios no nulos da la dimensión del espacio de los componentes principales.

### 2.1.3 Interpretación geométrica de los componentes principales

En términos geométricos es fácil mostrar que el primer componente principal define la mejor línea en las observaciones en un espacio de dimensión  $p$ . Por lo tanto, las observaciones pueden ser representadas en una dimensión, tomando su proyección sobre dicha línea, esto es, encontrar su primer componente resultante. Si en las observaciones se da el caso de que existiese colinealidad, entonces su representación deberá explicarse completamente por la variación de los datos y la matriz de covarianzas de la muestra deberá tener solo un valor propio o eigenvector. En la práctica, por supuesto, este tipo de colinealidad no se da y una mejor representación se deberá tener proyectando las observaciones en el espacio  $p$ -dimensional que mejor se ajuste, esto será definido por los primeros dos componentes principales. Entonces, el primer  $p^*$  componente principal tendrá el mejor ajuste en la  $p^*$  dimensiones. Si las observaciones se ajustan exactamente en el espacio de dimensión  $p^*$ , se verá esto indicado por que habrá  $p - p^*$  eigenvectores. Esto deberá implicar la presencia de  $p - p^*$  relaciones lineales entre las variables.

### 2.1.4 Definiciones y propiedades

Al considerar una muestra de  $n$  observaciones, los cuales se denotan con  $x_i$ ; cada uno de estos individuos tiene asociadas  $p$  características y sobre cada una de ellas se va aplicar el A.C.P.

La muestra puede agruparse en una matriz de  $n \times p$  llamada  $\Gamma$  la cual tiene la siguiente forma:

$$\Gamma = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (2.1)$$

También se puede suponer que cada individuo  $x_i$  tiene un vector de medias y una matriz de covarianzas ya definidos y dados de la siguiente manera:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \quad \text{y} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{pn} \end{bmatrix} \quad (2.2)$$

### 2.1.5 Componentes principales en una población

Si la matriz de covarianzas asociadas a un vector aleatorio  $X$  es conocida, entonces la transformación que lleva a estas variables a su correspondiente vector de componentes principales  $Y$  está dada por la siguiente definición:

**Definición 2.1.1.** Sea  $X$  es un vector aleatorio con media  $\mu$  y matriz de covarianzas  $\Sigma$ , se define la transformación de los componentes como:

$$X \rightarrow Y = \Gamma'(X - \mu)$$

donde  $\Sigma$  es una matriz ortogonal.

Por otro lado  $\Gamma'\Sigma\Gamma = \Lambda$  es una matriz diagonal con los valores propios de la matriz  $\Sigma$ , los cuales cumplen con la propiedad de que  $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n \geq 0$ , estos valores serán estrictamente positivos si la matriz  $\Sigma > 0$ .

El  $j$ -ésimo componente principal de  $X$  se define como el  $j$ -ésimo elemento del vector  $Y$ , cuya expresión es la siguiente:

$$Y_j = \Gamma'_{(j)}(X - \mu)$$

donde  $\Gamma'_{(j)}$  es la  $j$ -ésima columna de  $\Gamma$  y se le llama el  $j$ -ésimo vector correspondiente a el  $j$ -ésimo componente principal.

## TEOREMAS Y DEFINICIONES

**Teorema 2.1.1** Si  $X$  es un vector de parámetros  $\mu$  y  $\Sigma$  donde la matriz  $\Sigma$  se puede descomponer en  $\Sigma = \Gamma\Lambda\Gamma'$  con  $\Gamma'\Gamma = I_p$  y  $Y = \Gamma'(x - \mu)$ , entonces se satisfacen las siguientes propiedades:

1.  $E(Y_j) = 0 \quad \forall j$

Demostración:

$$\begin{aligned} E(Y_j) &= E[\Gamma'_{(j)}(X - \mu)] \\ &= E(\Gamma'_{(j)}X) - E(\Gamma'_{(j)}\mu) \\ &= \Gamma' E(X) - \Gamma'_{(j)}\mu \\ &= \Gamma'_{(j)}\mu - \Gamma'_{(j)}\mu \end{aligned}$$

$$= 0$$

$$\therefore E(Y_j) = 0 \quad \forall j$$

2.  $V(Y) = \Lambda$

Demostración:

$$\begin{aligned} V(Y) &= V(\Gamma'(X - \mu)) \\ &= V(\Gamma'X) \\ &= \Gamma'V(X)\Gamma \\ &= \Gamma'\Gamma\Lambda\Gamma'\Gamma \\ &= \Lambda \end{aligned}$$

$$\therefore V(Y) = \Lambda$$

3.  $V(Y_1) \geq V(Y_2) \geq \dots \geq V(Y_p) \geq 0$

Demostración:

$V(Y_1) \geq V(Y_2) \geq \dots \geq V(Y_p) \geq 0$  se cumple implícitamente de que  $\Sigma > 0$ .

4.  $\sum_{j=1}^p V(Y_j) = tr \Sigma$

Demostración: (porque se cumple la segunda igualdad)

$$\begin{aligned} tr(\Sigma) &= tr(\Gamma\Lambda\Gamma') \\ &= tr(\Gamma\Gamma'\Lambda) \\ &= tr(\Lambda) \\ &= \sum_{j=1}^p V(Y_j) \end{aligned}$$

$$\therefore \sum_{j=1}^p V(Y_j) = tr(\Sigma)$$

$$5. \prod_{j=1}^p V(Y_j) = |\Sigma|$$

Demostración:

$$\begin{aligned} |\Sigma| &= |\Gamma \Lambda \Gamma'| \\ &= |\Gamma \Lambda \Gamma'| \\ &= |\Lambda| \\ &= \prod_{j=1}^p \lambda_j \\ &= \prod_{j=1}^p V(Y_j) \end{aligned}$$

$$\therefore \prod_{j=1}^p V(Y_j) = |\Sigma|$$

**Definición 2.1.2** Una combinación lineal estandarizada de  $X \in \mathfrak{R}^p$  se define como:

$$Y = a' X$$

donde  $a' a = I$

**Teorema 2.1.2** Sea  $Y = a' X$  una combinación lineal estandarizada, donde  $E(X) = \mu$  y  $V(X) = \Sigma \geq 0$ , entonces  $V(Y) \leq \lambda_1$ , donde la descomposición espectral de la matriz  $\Sigma$  se define por  $\Sigma = \Gamma \Lambda \Gamma'$ ,  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$  y  $\Gamma' \Gamma = I$ . También se satisface que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$  y  $V(\Gamma'_{(1)} X) = \lambda_1$ .

Demostración:

Tenemos que todas las columnas de  $\Gamma$  son linealmente independientes, por lo cual se puede decir que:

$$a = \Gamma b$$

con  $b$  definida como  $b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix}$ , por lo que  $b' = (b_1 \ b_2 \ \dots \ b_p)$ .

De lo anterior se tiene que:

$$\begin{aligned} a &= (\Gamma_{(1)} \ \dots \ \Gamma_{(p)}) \\ &= \sum_{j=1}^p b_j \Gamma_{(j)} \end{aligned}$$

Por la def. 2.1.2, se tiene:

$$\begin{aligned} I &= a' a \\ &= (\Gamma b)' (\Gamma b) \\ &= b' b \end{aligned}$$

Entonces:

$$\begin{aligned} V(Y) &= V(a' X) \\ &= a' V(X) a \\ &= b' \Lambda b \\ &= \sum_{j=1}^p b_j^2 \lambda_j \end{aligned}$$

Por lo que  $V(Y)$  se maximiza si:

$$b = e_{(j)} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

y con esto se tiene que:

$$a = \Gamma b = \Gamma_{(1)}$$

Y finalmente  $V(Y) = \Gamma'_{(1)} \Lambda \Gamma_{(1)} = \lambda_1$

**Teorema 2.1.3** Si  $Y = a'X$  es una combinación lineal estandarizada, la cual no está correlacionada con los primeros  $k$  componentes principales de  $X$ , definidos por:

$$Y_j = \Gamma'_{(j)}(X - \mu) \quad j=1,2,\dots,k$$

Entonces  $V(Y)$  se maximiza en  $a = \Gamma_{(k+1)}$ .

Demostración:

Como las columnas de  $\Gamma$  son linealmente independientes se puede escribir:

$$a = \Gamma b$$

pero también se sabe que  $bb' = I$  y utilizando este resultado se obtiene que:

$$\begin{aligned} 0 &= C(Y, \Gamma'_{(j)}(X - \mu)) \\ &= C(a'X, \Gamma'_{(j)}X) \\ &= a' C(X, X) \Gamma_{(j)} \\ &= a' \Gamma \Lambda \Gamma' \Gamma_{(j)} \\ &= b' \Lambda \Gamma' \Gamma_{(j)} \end{aligned} \tag{2.3}$$

Y como



$$e_{(j)} = \begin{pmatrix} \Gamma_{(1)} \\ \Gamma_{(2)} \\ \vdots \\ \Gamma_{(p)} \end{pmatrix} \Gamma_{(j)}$$

se puede escribir la ecuación (2.3) como:

$$b' \Lambda e_{(j)} = 0$$

$$b_j \lambda_j = 0$$

Por otro lado, puede suponerse que  $\lambda_j \geq 0 \quad \forall j = 1, 2, \dots, k$  debido a que si alguna  $\lambda_j = 0$ , entonces las  $\lambda_r = 0 \quad \forall r \neq j, r = 1, 2, \dots, p$ .

Por lo tanto

$$b_j = 0, \quad \forall j = 1, 2, \dots, k,$$

así se cumple que

$$\begin{aligned} V(Y) &= \sum_{j=1}^p b_j^2 \lambda_j \\ &= \sum_{j=k+1}^p b_j^2 \lambda_j \end{aligned}$$

Por lo que si  $b_j = e_{(k+1)}$   $V(Y)$  se maximiza, entonces

$$\begin{aligned} a &= \Gamma b \\ &= \Gamma e_{(k+1)} \\ &= \Gamma_{(k+1)} \end{aligned}$$

quedando así que:

$$\begin{aligned}V(Y) &= V(\Gamma'_{(k+1)}X) \\ &= \Gamma'_{(k+1)}\Lambda\Gamma_{(k+1)} \\ &= e'_{(k+1)}\Lambda e_{(k+1)} \\ &= \lambda_{(k+1)}\end{aligned}$$

### 2.1.6 Estandarización de los componentes principales

Cuando las variables que se estudian están medidas en unidades diferentes, se debe de aplicar una estandarización, esto se hace para que los efectos de la escala no influyan en la determinación de los componentes principales.

La estandarización se realiza de la siguiente manera:

- Primero se debe suponer que la matriz  $X_{n \times p}$  tiene asociado un vector de medias y un vector de varianzas.
- Con dichos datos la estandarización para la  $j$ -ésima variable del elemento  $X_j$ , se hace de la siguiente manera:

$$Z_j = \frac{X_j - \mu_j}{\sqrt{\sigma_j^2}}, \quad \forall j$$

Si  $\Delta = \text{diag}\Sigma$  se tiene la ecuación anterior se puede escribir de la siguiente manera:

$$Z = \left( \frac{1}{\Delta^2} \right)^{-1} (X - \mu).$$

---

**ALGUNAS PROPIEDADES**

$$\begin{aligned}
 1. \quad E(\mathbf{Z}) &= 0 \\
 2. \quad V(\mathbf{Z}) &= V\left[\left(\Delta^2\right)^{-1} (X - \mu)\right] \\
 &= \left(\Delta^2\right)^{-1} \Sigma \left(\Delta^2\right)^{-1} \\
 &= \rho
 \end{aligned}$$

En donde  $\rho$  denota la matriz de correlación poblacional, la cual puede descomponerse en  $\rho = \Gamma' \mathbf{Z}$  y las columnas de la matriz  $\Gamma$  son los vectores propios de  $\rho$ , los cuales son ortogonales. Entonces, la transformación que va a definir a los componentes principales será  $Y = \Gamma' \mathbf{Z}$ .

$$\begin{aligned}
 3. \quad V(Y) &= V(\Gamma' \mathbf{Z}) \\
 &= \Gamma' \rho \Gamma \\
 &= \tilde{\Lambda}
 \end{aligned}$$

### 2.1.7 Componentes principales generados a partir de una muestra

Cuando se desconoce la matriz de covarianzas se va a estimar mediante la muestra observada y se supondrá que los valores de las  $p$  variables observadas  $X_1, X_2, \dots, X_p$  obtenidos de la muestra observada se concentrarán en una matriz de  $n \times p$ .

Cuando el A.C.P. se basa en una matriz de covarianzas desconocida se procederá a utilizar una matriz de covarianzas muestral  $S_x$ , donde:

$$X_j = \frac{1}{n} \sum_{i=1}^n X_{ij} \quad \forall j = 1, \dots, p$$

es la media de los valores observados para la  $j$ -ésima variable sobre los  $n$  individuos y

$$S_{jk} = \frac{1}{n} \sum_{i=1}^n (X_{ij} - X_j)(X_{ik} - X_k)$$

es la covarianza muestral entre las variables  $X_j$  y  $X_k$ . Entonces  $S_x = S_{jk}$  es la matriz de covarianzas de las  $p$  variables.

**Definición 2.1.3** Sea  $U$  una matriz ortogonal cuyos elementos en la diagonal son positivos tales que cumplan que:

$$U' S_x U = \tilde{\Lambda}$$

y  $U'U = I_p$  donde  $\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_p$  son los valores propios ordenados y que están asociados a la matriz  $S_x$ . La transformación de los componentes principales de un vector  $X \in \mathbb{R}^p$  está definida como:

$$Y_{px1} = U'(X - \bar{X})$$

en donde la  $i$ -ésima nueva observación está dada por:

$$Y_i = U'(X - \bar{X}) \quad \forall i = 1, \dots, n$$

Y la media muestral de las nuevas observaciones será:

$$\begin{aligned} \bar{Y} &= \frac{1}{n} \sum_{i=1}^n Y_i \\ &= \frac{1}{n} \sum_{i=1}^n U'(X_i - \bar{X}) \\ &= \frac{1}{n} U' \sum_{i=1}^n (X_i - \bar{X}) \\ &= 0 \end{aligned}$$

La matriz de covarianzas muestral  $S_y$  se podrá definir como:

$$\begin{aligned} S_y &= \frac{1}{n} \sum_{i=1}^n U'(Y_i - \bar{Y})(Y_i - \bar{Y})' \\ &= \frac{1}{n} \sum_{i=1}^n U'(X_i - \bar{X})(X_i - \bar{X})' U \\ &= U' S_x U \end{aligned}$$

La forma matricial de las nuevas observaciones se puede escribir como:

$$Y_{n \times p} = \begin{pmatrix} Y_1' \\ Y_2' \\ \vdots \\ Y_n' \end{pmatrix} = \begin{pmatrix} (X_1 - \bar{X})' \\ (X_2 - \bar{X})' \\ \vdots \\ (X_n - \bar{X})' \end{pmatrix} U$$

En donde el  $j$ -ésimo elemento de  $Y_i$ , está dado por  $Y_{ij}$  y representa el puntaje del  $j$ -ésimo componente sobre la  $i$ -ésima observación. En términos de dicha observación se puede escribir la transformación del componente principal como:

$$Y_{ij} = U'_{(j)}(X_i - \bar{X}) \quad \forall i=1, \dots, n \text{ y } j=1, \dots, p.$$

Si se ve la correlación que existe entre el vector de componentes principales  $Y$  y el punto  $X$  definido como:

$$X \rightarrow Y = \Gamma'(x - \mu)$$

se tendrá que la covarianza se calculará como:

$$\begin{aligned} C(X, Y) &= C(X, \Gamma'(X - \mu)) \\ &= C(X, \Gamma' X) \end{aligned}$$

$$\begin{aligned}
 &= C(X, X\Gamma) \\
 &= V(X) \\
 &= \Gamma\Lambda\Gamma'\Gamma \\
 &= \Gamma\Lambda
 \end{aligned}$$

La covarianza entre  $X_k$  y  $Y_j$  se definirá como:

$$C(X_k, Y_j) = \Gamma_{kj}\lambda_j$$

Si se define la varianza de  $X$  de la siguiente manera:

$$V(X) = \Sigma = \{\sigma_{kj}\} \text{ con } j=1,2,\dots,p$$

y a  $\Lambda$  como la matriz diagonal de las varianzas de  $Y$ , la correlación entre las variables  $X_{kj}$  y  $Y_j$  se va a obtener de la siguiente manera:

$$\begin{aligned}
 \tau_{kj} &= \frac{\Gamma_{kj}\lambda_j}{\sqrt{\sigma_{kk}\lambda_j}} \\
 &= \frac{\Gamma_{kj}\sqrt{\lambda_j}}{\sqrt{\sigma_{kk}}}
 \end{aligned}$$

Entonces la proporción de variabilidad explicada de  $X_k$  por la componente  $Y_j$  es  $\tau_{kj}^2$ , donde:

$$\tau_{kj}^2 = \frac{\Gamma_{kj}^2\lambda_j}{\sigma_{kk}}$$

Como los elementos de  $Y$  no están correlacionados, entonces cualquier subconjunto de  $I$  de componentes principales va a explicar una proporción de la varianza de  $X_k$ , con:

$$\tau_{kj} = \sum_{j \in I} \tau_{kj}^2$$

$$= \frac{1}{\sigma_{kk}} \sum_{j \in I} \lambda_j \Gamma_{kj}^2$$

El numerador va a proporcionar la variación acumulada por el conjunto  $I$  y el denominador la variación de  $X_k$  que va a ser explicada. Cuando  $I$  incluye todos los componentes principales la proporción acumulada será de uno.

### 2.1.8 Interpretación geométrica de los componentes principales bajo normalidad de observaciones

La función de densidad de un vector  $X \in \mathfrak{R}^p$ , si  $X$  es una  $N_p(\mu, \Sigma)$ ,  $\Sigma > 0$  se puede escribir como:

$$f_x(X) = |2\pi \Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(X - \mu)' \Sigma^{-1}(X - \mu)\right\}$$

Como  $|2\pi \Sigma|^{-\frac{1}{2}}$  es una constante, la cual no depende del vector  $X$  y tiene la forma cuadrática  $(X - \mu)' \Sigma^{-1}(X - \mu) = c$  (2.7), entonces se define una elipse en un espacio de dimensión  $p$ . Y cuando varía la constante  $c$  se genera una familia de elipses.

Por otro lado, se sabe que la matriz  $\Sigma$  es definida positiva y por definición  $\Sigma^{-1}$  también lo va a ser, por lo cual podemos escribir que:

$$\Sigma^{-1} = U \Lambda^{-1} U'$$

Con lo anterior se puede escribir a (2.7) de la siguiente manera:

$$(X - \mu)' U \Lambda^{-1} U' (X - \mu) = c \quad (2.4)$$

Entonces los principales ejes de la elipse van a ser los vectores propios de la matriz  $\Sigma$ .

Ahora, si tomamos en cuenta que  $Y = U'(X - \mu)$ , (2.3) se transforma de la siguiente manera:

$$Y' \Lambda^{-1} Y = c \Leftrightarrow \sum_{j=1}^p \frac{Y_j^2}{\lambda_j} = c$$

Y la magnitud del  $j$ -ésimo eje principal está dado por:

$$y_j = \pm \sqrt{\lambda_j c}.$$

## 2.1.9 Estimación máxima verosímil para datos normales

Cuando se tiene cuenta con una muestra muy pequeña de valores y vectores propios de una matriz de covarianzas  $S$  su distribución es muy complicada, aún si no existe correlación; esto se debe a que los valores propios son funciones no racionales de los elementos de  $S$ .

Pero cuando se tiene una muestra grande y los resultados son conocidos, algunas de las propiedades de la muestra de componentes principales para datos normales están contenidos en los resultados de máxima verosimilitud.

A continuación se mencionan algunos teoremas importantes.

**Teorema 2.1.4** *Para datos normales cuando los valores propios de  $\Sigma$  son distintos, los componentes principales y los valores propios muestrales, son los estimadores máximo verosímiles de los parámetros poblacionales correspondientes.*

**Teorema 2.1.5** *Sea  $\Sigma$  una matriz definida positiva con valores propios distintos y sean  $M \approx W(\Sigma, m)$  y  $W = m^{-1}M$ . Ahora, considérese la descomposición espectral  $\Sigma = \Gamma \Lambda \Gamma'$  y  $W = GLG$ , y sean  $\Psi = \text{diag}(\Lambda)$  y  $\Phi = \text{diag}(L)$ . Entonces las siguientes distribuciones asintóticas se satisfacen siempre que  $m \rightarrow \infty$ :*



- $\Phi \approx N_p\left(\Psi, \frac{2\Psi^2}{m}\right)$ , es decir, que los valores propios de  $W$  son asintóticamente normales, insesgados e independientes.
- $g(i) \approx N_p\left(\Gamma_{(i)}, \frac{V_i}{m}\right)$ , donde  $V_i = \lambda_i \sum \frac{\lambda_j}{(\lambda_j)^2} \Gamma_{(i)} \Gamma_{(j)}'$ , con  $\Gamma_{(i)}$  el vector correspondiente a la  $i$ -ésima columna de la matriz  $\Gamma$ . Los vectores propios de  $W$  son asintóticamente normales e insesgados, con matriz de covarianza asintótica  $\frac{V_i}{m}$ .
- La covarianza entre el  $r$ -ésimo elemento del  $g(i)$  y el  $t$ -ésimo elemento de  $g(j)$  es  $-\frac{\lambda_i \lambda_j \tau_{rj} \tau_{ti}}{m(\lambda_i - \lambda_j)^2}$ .
- Los elementos de  $L$  son asintóticamente independientes de los elementos de  $G$ .

**Teorema 2.1.6** Sea  $\hat{\Sigma}$  el estimador máximo verosímil para  $\Sigma$ , basado en una muestra de tamaño  $n$ , de una población  $N(\mu, \Sigma)$ . Sean  $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p)$  y  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$  en donde  $\hat{\lambda}_i$  y  $\lambda_i$  son los valores propios de  $\hat{\Sigma}$  y  $\Sigma$  respectivamente. Sean  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ , entonces si  $\Sigma > 0$  y todos los valores propios son diferentes, i.e.,  $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$ ,  $\sqrt{n-1}(\hat{\lambda} - \lambda)$  se distribuye asintóticamente como una  $N(\mu, 2\Lambda^2)$ .

### 2.1.10 Intervalo de confianza para un valor propio

Por lo anterior se puede construir un intervalo de confianza para algún  $\lambda_j$ , dicho intervalo dará la dispersión de este valor.

El intervalo de confianza se construye de la siguiente forma:

## ANÁLISIS DE COMPONENTES PRINCIPALES

---

Como  $\sqrt{n-1}(\hat{\lambda}_j - \lambda_j)$  se distribuye como  $N(\mu, 2\lambda_j^2) \forall j=1,2,\dots,p$ , si se estandariza se obtiene:

$$\frac{\sqrt{n-1}(\hat{\lambda}_j - \lambda_j)}{\sqrt{2\lambda_j}} \approx N(0,1),$$

dicha expresión es una cantidad pivotal con la que se puede obtener el intervalo mediante la siguiente probabilidad:

$$p \left[ z_1 < \frac{\sqrt{n-1}(\hat{\lambda}_j - \lambda_j)}{\sqrt{2\lambda_j}} < z_2 \right] = 1 - \alpha,$$

⇔

$$p \left[ z_1 < \frac{\frac{\sqrt{n-1}(\hat{\lambda}_j - \lambda_j)}{\lambda_j}}{\frac{\sqrt{2\lambda_j}}{\lambda_j}} < z_2 \right] = 1 - \alpha,$$

⇔

$$p \left[ z_1 \sqrt{2} < \sqrt{n-1} \frac{\hat{\lambda}_j}{\lambda_j} - \sqrt{n-1} < z_2 \sqrt{2} \right] = 1 - \alpha,$$

⇔

$$p \left[ z_1 \sqrt{2} + \sqrt{n-1} < \sqrt{n-1} \frac{\hat{\lambda}_j}{\lambda_j} < z_2 \sqrt{2} + \sqrt{n-1} \right] = 1 - \alpha,$$

⇔

$$p \left[ \frac{z_1 \sqrt{2}}{\sqrt{n-1}} + 1 < \frac{\hat{\lambda}_j}{\lambda_j} < \frac{z_2 \sqrt{2}}{\sqrt{n-1}} + 1 \right] = 1 - \alpha,$$

⇔

$$p \left[ \frac{1}{1 + \frac{z_1 \sqrt{2}}{\sqrt{n-1}}} < \frac{\lambda_j}{\hat{\lambda}_j} < \frac{1}{1 + \frac{z_2 \sqrt{2}}{\sqrt{n-1}}} \right] = 1 - \alpha,$$

⇔

$$p \left[ \frac{\hat{\lambda}_j}{1 + \frac{z_1 \sqrt{2}}{\sqrt{n-1}}} < \lambda_j < \frac{\hat{\lambda}_j}{1 + \frac{z_2 \sqrt{2}}{\sqrt{n-1}}} \right] = 1 - \alpha.$$

∴ El intervalo de confianza para  $\lambda_j$  al  $(1 - \alpha) \times 100$  % está definido por:

$$\left( \frac{\hat{\lambda}_j}{1 + \frac{z_2 \sqrt{2}}{\sqrt{n-1}}}, \frac{\hat{\lambda}_j}{1 - \frac{z_2 \sqrt{2}}{\sqrt{n-1}}} \right)$$

### 2.1.11 Pruebas de hipótesis sobre los componentes principales

Se debe contar con un procedimiento para decidir cuantos  $k$  componentes principales se deben incluir en la variación que se considera importante en la matriz de observaciones  $X$ . Se espera ignorar a  $(p - k)$  si sus correspondientes valores propios son iguales a cero, pero esto sólo ocurre cuando la matriz  $\Sigma$  asociada a la muestra es de rango  $(p - k)$ , i.e.,  $(p - k)$  valores propios son iguales a cero, pero este caso no ocurre con frecuencia en la práctica.

Otra alternativa es hacer una prueba de hipótesis, en la que la proporción de variabilidad explicada por las  $k$  componentes principales sea menor que un cierto valor de  $\pi$ . Otra hipótesis que puede probarse es que cuando los últimos  $(p - k)$  valores propios son iguales, esto implica que la variación es igual en todas las direcciones del espacio generado por los últimos  $(p - k)$  vectores propios, esta situación es denominada variación isotrópica e implica que si alguna componente es eliminada, entonces deben ser eliminadas todas las demás.

Este tipo de pruebas pueden realizarse con la suposición de que la muestra original es normal.

## TIPOS DE PRUEBAS PARA COMPONENTES PRINCIPALES

### 1. Prueba de la proporción de variabilidad explicada por los primeros $k$ componentes principales

Sean  $\lambda_1, \lambda_2, \dots, \lambda_p$  los valores propios de  $\Sigma$  y  $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p$  los valores propios de  $S$ . La hipótesis a probar es:

$$H_0: \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^p \lambda_j} = \pi \text{ con } k < p \text{ vs } H_a: \text{no } H_0.$$

Sea  $\hat{\pi}$  el estimador muestral de  $\pi$ , entonces por el teorema 2.1.7 se sabe que los elementos de  $\hat{\lambda}_j$  tienen una distribución normal asintótica, y  $\hat{\pi}$  tiene una distribución normal con media  $\pi$  y varianza de la forma:

$$V(\hat{\pi}) = \frac{2tr(\Sigma^2)}{(n-1)(tr(\Sigma))^2} (\pi^2 - 2c\pi + c),$$

donde  $c$  es un número definido como:

$$c = \frac{\sum_{j=1}^k \lambda_j^2}{\sum_{j=1}^p \lambda_j^2}$$

## CAPÍTULO II

---

La estimación de la  $V(\pi)$  se hace utilizando la matriz  $S_x$  y los valores propios de ella, i.e.,

$$\hat{\Sigma} = S, \text{tr}(\hat{\Sigma}) = \sum_{j=1}^c \hat{\lambda}_j.$$

Entonces resulta que:

$$V(\hat{\pi}) = \frac{2\text{tr}(S^2)}{(n-1)(\text{tr}(S))^2} (\pi^2 - 2c\pi + c).$$

Y queda que:

$$\hat{\pi} \approx N(\pi, V(\hat{\pi}))$$

Con la expresión anterior se puede hacer un intervalo de confianza estandarizado, el cual tiene un nivel de confianza de  $(1 - \alpha) \times 100\%$ , el cual se obtiene de la siguiente manera:

$$\frac{(\hat{\pi} - \pi)}{\sqrt{V(\hat{\pi})}} \approx N(0,1)$$

y calculando se obtiene que:

$$P\left[z_1 < \frac{(\hat{\pi} - \pi)}{\sqrt{V(\hat{\pi})}} < z_2\right] = 1 - \alpha,$$

⇔

$$P[z_1 \sqrt{V(\hat{\pi})} < (\hat{\pi} - \pi) < z_2 \sqrt{V(\hat{\pi})}] = 1 - \alpha,$$

⇔

$$P[z_1 \sqrt{V(\hat{\pi})} - \hat{\pi} < -\pi < z_2 \sqrt{V(\hat{\pi})} - \hat{\pi}] = 1 - \alpha,$$

⇔

$$P[\hat{\pi} - z_2 \sqrt{V(\hat{\pi})} < \pi < \hat{\pi} - z_1 \sqrt{V(\hat{\pi})}] = 1 - \alpha.$$

Se define  $z_1 = -z_2$  para minimizar la longitud del intervalo, esta definición va a llevar a la expresión final dada como:

$$(\hat{\pi} - z_2 \sqrt{V(\hat{\pi})}, \hat{\pi} + z_2 \sqrt{V(\hat{\pi})})$$

En donde  $Z_2$  es el cuantil de  $\left(1 - \frac{\alpha}{2}\right)$  de una normal estándar.

## 2. Prueba de Esfericidad

Esta prueba es usada para determinar el número de componentes principales que serán utilizados para describir el comportamiento de los datos. Aquí se desea probar que los últimos  $(p-k)$  valores propios son iguales, es decir, las últimas  $(p-k)$  componentes principales tienen la misma varianza; esto significa que si se incluye una de ellas deben de incluirse todas las demás.

La hipótesis a probar es:

$$H_0 : \lambda_{k+1} = \dots = \lambda_p \quad \text{vs} \quad H_a : \text{no } H_0$$

La estadística de prueba se obtiene con el método de la razón de verosimilitud, el cual tiene asociada la siguiente expresión:

$$-2 \ln \Lambda = np[a - \ln(g) - 1],$$

donde la constante  $a$  está definida en (2.2). Como  $a$  y  $g$  corresponden a la media aritmética y geométrica de los valores propios de  $\hat{\Sigma}^{-1} S$ , donde  $\hat{\Sigma}$  es el estimador máximo verosímil de  $\Sigma$  bajo la hipótesis nula. Sea:

$$a_0 = \frac{\sum_{j=k+1}^p \hat{\lambda}_j}{(p-k)} \quad (2.9)$$

que denota la media aritmética de los valores propios  $\hat{\lambda}_{k+1}, \dots, \hat{\lambda}_{k+p}$  asociados a la  $\hat{\Sigma}$  y

$$g_0 = \left( \prod_{j=k+1}^p \hat{\lambda}_j \right)^{\frac{1}{(p-k)}}$$

la media geométrica. Por lo tanto, la estadística de prueba es:

$$-2 \ln \Lambda = np[a_0 - \ln(g_0) - 1]$$

en donde  $n$  es el tamaño de la muestra y  $p$  es la dimensión; así  $-2 \ln \Lambda$  en la ecuación (2.9) se distribuye como una  $\chi_r^2$ , donde  $r$  es el número de grados de libertad.

### 2.1.12 Reglas de corte

Las reglas de corte son utilizadas como herramientas para determinar el número de componentes principales que deben de ser retenidos.

Dichas reglas son:

1. Cattell en 1966, desarrolla una forma práctica de observar empíricamente la contribución de varios componentes principales, la cual consiste en observar la gráfica conocida como "Screeplot" y graficar el valor propio de  $\lambda_j$  y  $j$ . Dicha gráfica indica donde terminan los valores propios grandes y en que punto comienzan los pequeños.
2. Incluir los componentes principales que en conjunto acumulen un 90% de la variación total.

3. Kaiser propone excluir los componentes cuyos valores propios sean menores que la media, es decir, aquellos menores a la unidad si es que se ha utilizado una matriz de correlación.

### 2.1.13 Algunas herramientas adicionales para los componentes principales

Las siguientes propiedades sirven como herramienta para la interpretación de los resultados que se obtienen con este método estadístico.

#### 1. *La media de las nuevas observaciones*

Cuando se define la media muestral de las observaciones originales  $\bar{X}$  y se usa como matriz muestral de covarianzas a  $S_x$ , es entonces cuando la transformación general es de la forma:

$$Y = U'(X - \bar{X})$$

Dicha transformación consta de una translación seguida de una rotación, razón por la cual la media de los componentes principales es igual a cero.

Si  $U$  se denota como la matriz de los vectores propios asociada a la matriz de correlación  $R$ , entonces dicha transformación puede ser usada después de estandarizar las variables  $(X - \bar{X})$ , con lo que se obtiene que cada variable va a tener una varianza unitaria.

#### 2. *Valores propios iguales a cero*

Cuando algunas variables originales son linealmente dependientes, se obtiene que algunos valores propios de  $\Sigma$  son iguales a cero.



**3. Componentes principales bajo cambios de escala de las variables**

Los componentes principales de un vector aleatorio no son invariantes con respecto a la escala.

**4. Valores propios repetidos**

En algunas ocasiones los valores propios de  $\Sigma$  son iguales si y sólo si  $\lambda_{q+1} = \dots = \lambda_{q+k} = \lambda$ , con lo cual puede decirse que la raíz de  $\lambda$  es de multiplicidad  $k$ . Los vectores propios que corresponden a las raíces múltiples no son únicos y sus correspondientes componentes principales tendrán la misma varianza.

**5. Proporción de variabilidad explicada**

El cociente  $\frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^p \lambda_j}$  representa la variabilidad explicada de los primeros  $k$  componentes principales.

**6. El rango de la matriz  $X$** 

Si la matriz de covarianza de la muestra  $X_{n \times p}$  es de rango  $k < p$ , entonces la variabilidad de  $X$  puede ser explicada totalmente por los primeros  $k$  componentes principales.

## 2.2 ANÁLISIS DISCRIMINANTE

### 2.2.1 Introducción

La idea básica del Análisis Discriminante consiste en asignar a un individuo o grupo de individuos a una o varias poblaciones conocidas o desconocidas. Esta técnica tiene como objetivo primordial la clasificación entre una variable categórica y un grupo de variables interrelacionadas, es decir, clasifica individuos en uno y sólo uno de los grupos o poblaciones que se tienen como alternativas.

Las áreas donde esta técnica puede ser interesante son numerosas, algunas de ellas son:

- Diagnóstico médico: aquí las variables que describen a cada individuo pueden ser los resultados de varias pruebas clínicas y los grupos pueden ser colecciones de pacientes, los cuales tienen enfermedades distintas.
- Arqueología: las variables pueden ser mediciones particulares de objetos de interés y los grupos, características taxonómicas de dichos objetos.
- Reconocimiento de diálogo: aquí los objetos que van a ser clasificados, usualmente tienen forma de onda y las variables de los parámetros acústicos son extraídos de la pronunciación de una palabra específica por un individuo.

El Análisis Discriminante, puede ser basado en supuestos distribucionales que se hacen sobre la muestra que va a ser observada, éstos son:

- Discriminante Logístico: supone que la forma específica de las densidades es desconocida, pero el logaritmo del cociente de las densidades es lineal en los parámetros asociados.
- Discriminante no Paramétrico: supone que la distribución de la muestra no es conocida y deber ser estimada.

- Discriminante Logístico: supone que las poblaciones tienen una distribución normal multivariada.

En la literatura científica, el Análisis Discriminante tiene varios sinónimos, tales como clasificación, reconocimiento de patrones, recolección de características, identificación, predicción y selección. El origen de esta técnica es muy antigua y se desarrolló gracias a Pearson, Neyman y Waldian.

## 2.2.2 Formulación del problema de análisis discriminante

Supóngase que se tienen  $k$  poblaciones distintas  $\pi_1, \pi_2, \dots, \pi_k$ . Dichas poblaciones se desean clasificar individualmente con observación  $x = (x_1, \dots, x_p)$  o en un grupo de  $N$  individuos con observaciones  $x^\alpha = (x_{\alpha 1}, \dots, x_{\alpha p})$  y  $\alpha = 1, \dots, N$ , en  $p$  diferentes características, el individuo o grupo se caracteriza, dentro de  $\pi_1, \pi_2, \dots, \pi_k$ . Cuando se caracterice al grupo de individuos se hará bajo la suposición de que todo el grupo corresponde a una sola población entre las  $k$  dadas. Además, se deseará asumir que cada una de las  $\pi_i$  pueden ser especificadas a través de su media o de su función de distribución  $F_i$  (o por su función de densidad de probabilidad  $f_i$  con respecto a la medida de Lebesgue) para un vector aleatorio  $X = (X_1, \dots, X_p)'$ ,

Dado lo anterior se tienen las siguientes definiciones:

**Definición 2.2.1** Una regla discriminante corresponde a una división del espacio  $E^p$  en regiones disjuntas (o mutuamente excluyentes),  $R_1, \dots, R_k$  con  $\bigcup_{i=1}^k R_i \equiv E^p$  donde la regla está dada como:

asignar  $x$  a  $\pi_i$  si  $X$  con  $i = 1, \dots, k$ .

## ANÁLISIS DISCRIMINANTE

La discriminación será más exacta en la medida en que  $\pi_i$  tenga la mayor parte de su probabilidad centrada en  $R_i$  para cada  $i$ .

Y, la probabilidad de una clasificación errónea de un individuo con la observación  $x$  de  $\pi_i$  en  $\pi_j$ , con la regla discriminante es:

$$P(j|i, R) = \int_{R_j} f_i(x) dx$$

donde  $dx = \prod_{i=1}^p dx_i$ .

La esperanza del costo del error de clasificación en una observación  $\pi_i$ , usando la regla discriminante estará dada por:

$$r_i(R) = \sum C(j|i) P(j|i, R), \text{ con } i = 1, \dots, k.$$

Para encontrar la regla discriminante óptima se necesitará hacer una comparación del vector de costos  $r_i(\bar{R}) = (r_1(R), \dots, r_k(R))$ , para diferentes  $R$ .

**Definición 2.2.2** Dadas dos reglas discriminante  $R$  y  $R^*$  se dirá que  $R$  es mejor que  $R^*$  si  $r_i(R) \leq r_i(R^*)$  para todo  $i$ , y  $R$  es mejor que  $R^*$  si la desigualdad es estricta.

**Definición 2.2.3 Regla discriminante admisible.** Se dice que una regla discriminante  $R$  es admisible si no existe una regla discriminante  $R^*$  que sea mejor que  $R$ .

**Definición 2.2.4 Regla discriminante Minmax.** Se dice que una regla discriminante  $R^*$  es minmax de regla discriminante  $R$  si:

$$\max_i r_i(R^*) = \min_R \max_i r_i(R)$$

**Definición 2.2.5 Regla Discriminante de Bayes.** Dada  $p$ , una regla discriminación  $R$  que minimiza el costo promedio de una mala clasificación es llamada una Regla Discriminante de Bayes, con respecto a  $p$ .

Se deben evaluar las formas explícitas de esta regla, en casos en los que cada  $\pi_i$  admite una función de densidad de probabilidad  $f_i, i = 1, \dots, k$ . También se supondrá que todos los procedimientos de clasificación considerados son los mismos si éstos sólo difieren en un conjunto de probabilidad de medida cero.

**Teorema 2.2.1** Si las probabilidades a priori  $p_i, i = 1, \dots, k$ , son conocidas y si  $\pi_i$  admite una función de densidad de probabilidad  $f_i$  con respecto a la medida de Lebesgue, entonces la Regla Discriminante de Bayes  $R^* = (R_1^*, \dots, R_k^*)$  que minimiza el costo promedio es definida asignando a  $x$  a la región  $R_i^*$  si:

$$\sum_{i=1, i \neq l}^k p_i f_i(x) c(l|i) < \sum_{i=1, i \neq j}^k p_i f_i(x) c(j|i), \quad j = 1, \dots, k, \quad j \neq l \quad (2.5)$$

Demostración:

Sea

$$h_i(x_i) = \sum_{i=1, i \neq j}^k p_i f_i c(j|i)$$

entonces el costo promedio esperado de la regla discriminante  $R = (R_1, \dots, R_k)$  con respecto de las probabilidades a priori  $p_i, i = 1, \dots, k$ , está dado por:

$$\sum_{j=1}^k \int_{R_j} h_j(x) dx = \int h(x) dx$$

## ANÁLISIS DISCRIMINANTE

---

donde  $h(x) = h_j(x)$  si  $x \in R_j$ .

Para la Regla Discriminante de Bayes  $R^*$ ,  $h(x)$  es igual a:

$$h^*(x) = \min_j h_j(x).$$

En otras palabras,  $h^*(x) = h_j(x) = \min_j h_j(x)$  para  $x \in R_j^*$ . La diferencia de costo promedio esperado para alguna Regla Discriminante de Bayes  $R$  y  $R^*$  es:

$$\int [h(x) - h^*(x)] dx = \sum_j \int_{R_j} [h(x) - \min_i h_i(x)] dx \geq 0$$

y la igualdad se mantiene si  $h(x) = \min_i h_i(x)$  para  $x \in R_j, \forall j$ .

Observaciones:

- Si (2.7) se mantiene para toda  $j \neq l$ , excepto para  $h$  índices, para los cuales la desigualdad es reemplazada por la igualdad, entonces a  $x$  se le puede asignar a cada uno de esos  $(h+1)\pi_i$ , términos.
- Si  $c(i|j) = c$  ( $c \neq 0$ ),  $\forall (i, j), i \neq j$ , entonces en  $R_l^*$

$$\sum_{i=1, i \neq l}^k p_i f_i(x) < \sum_{i=1, i \neq j}^k p_i f_i(x), j = 1, \dots, k, j \neq l,$$

implica que en  $R_l^*$

$$p_l f_l(x) < p_l f_l(x), j = 1, \dots, k, j \neq l.$$

En otras palabras, el punto  $x \in R_l^*$  si  $l$  es el índice para el cual  $p_l f_l(x)$  es un máximo.

La razón de probabilidad de una regla de discriminante  $R = (R_1, \dots, R_k)$  está definida como:

$$R_j : C_j f_j(x) > \max_{i, i \neq j} C_i f_i(x),$$

para constantes positivas  $C_1, \dots, C_k$ . En particular, si todas las  $C_i$  son iguales, la regla discriminante es llamada una *Regla Discriminante de Máxima Probabilidad*.

**Teorema 2.2.2** *Toda regla discriminante admisible es una regla discriminante de Bayes con respecto a ciertas probabilidades a priori en  $\pi_1, \dots, \pi_k$ .*

**Teorema 2.2.3** *La clase de toda regla discriminante admisible es completa.*

**Teorema 2.2.4** *Para todo conjunto de probabilidades a priori  $p = (p_1, \dots, p_k)$  en  $(\pi_1, \dots, \pi_k)$ , existe una regla discriminante de Bayes admisible.*

**Teorema 2.2.5** *Para  $k=2$ , existe una única regla discriminante mínima  $R$  en donde  $r_1(R) = r_2(R)$ .*

**Teorema 2.2.6** *Supóngase que  $C(j|i) = C > 0, \forall i \neq j$  y que las funciones de distribución  $F_1, \dots, F_k$  caracterizan a las poblaciones  $\pi_1, \dots, \pi_k$  son absolutamente continuas. Entonces existe una única regla discriminante  $R$  para cualquier  $r_1(R) = \dots = r_k(R)$ .*

### 2.2.3 Clasificación en dos poblaciones normales multivariadas

Considérese el problema de clasificar un individuo, con observación  $x$  de una de dos poblaciones normales  $p$ -variadas conocidas con medias  $\mu_1$  y  $\mu_2$ , respectivamente, e iguales matrices de covarianzas definidas positivas  $\Sigma$ . Entonces la función de probabilidad es:

$$f_i(x) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu_i)' \Sigma^{-1}(x-\mu_i)\right\}, \quad i = 1, 2.$$

Y el cociente de densidad es:

$$\frac{f_1(x)}{f_2(x)} = \exp\left\{-\frac{1}{2}(x-\mu_1)' \Sigma^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)' \Sigma^{-1}(x-\mu_2)\right\}$$

La regla discriminante de Bayes  $R = (R_1, R_2)$  contra las probabilidades *a priori*  $(p_1, p_2)$  está dada por:

$$R_1 : \left(x - \frac{1}{2}(\mu_1 + \mu_2)\right)' \Sigma^{-1}(\mu_1 - \mu_2) \geq k, \tag{2.8}$$

$$R_2 : \left(x - \frac{1}{2}(\mu_1 + \mu_2)\right)' \Sigma^{-1}(\mu_1 - \mu_2) < k$$

donde  $k = \frac{\log(p_2 C(1|2))}{p_1 C(2|1)}$ .

Por simplicidad la región de bondad es asignada a  $R_1$ . La función lineal

$\left(x - \frac{1}{2}(\mu_1 + \mu_2)\right)' \Sigma^{-1}(\mu_1 - \mu_2)$  de los componentes del vector de observaciones  $x$  se le

llama "*función discriminante*" y los componentes de  $\Sigma^{-1}(\mu_1 - \mu_2)$  son llamados "*coeficientes discriminantes*".

Es importante señalar que si  $p_1 = p_2 = \frac{1}{2}$  y  $C(1|2) = C(2|1)$  entonces  $k = 0$ .

Por otro lado, supóngase no se tienen las probabilidades *a priori* para  $\pi_i$ . En este caso no se podrá usar la técnica de Bayes para obtener la regla discriminante dada en (2.8). Sin embargo, se podrá utilizar la regla discriminante minmax, encontrando  $k$  tal que la regla discriminante de Bayes en (2.8) con  $k$  desconocida satisfaga:



$$C(2|1)P(2|1, R) = C(1|2)P(1|2, R) \quad (2.9)$$

Sea  $X$  el vector aleatorio correspondiente a la observación  $X$  con:

$$U = \left( X - \frac{1}{2}(\mu_1 + \mu_2) \right)' \Sigma^{-1}(\mu_1 - \mu_2) \quad (2.10)$$

Suponiendo que  $X$  está distribuido con respecto a  $\pi_1$ .  $U$  será entonces una normal con media y varianza:

$$\begin{aligned} E_1 &= \frac{1}{2}(\mu_1 - \mu_2)' \Sigma^{-1}(\mu_1 - \mu_2) = \frac{1}{2} \alpha \\ \text{var}(U) &= E \left\{ (\mu_1 - \mu_2)' \Sigma^{-1} (X - \mu_1) (X - \mu_1)' \Sigma^{-1} (\mu_1 - \mu_2) \right\} \\ &= (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) \\ &= \alpha \end{aligned} \quad (2.11)$$

Si  $X$  se distribuye como  $\pi_2$ , entonces  $U$  tendrá una distribución normal con media y varianza:

$$E_2 = -\frac{1}{2} \alpha, \quad (2.12)$$

$$\text{var}(U) = \alpha.$$

La cantidad  $\alpha$  es llamada la distancia de Mahalanobis entre dos poblaciones normales con la misma matriz de covarianza. Entonces la regla discriminante minmax  $R$  es, escribiendo  $u = U(x)$ ,

$$R_1 : u \geq k, \quad R_2 : u < k \quad (2.13)$$

donde la constante  $k$  estará dada por:

$$\begin{aligned} C(2|1) & \int_{-\infty}^k \frac{1}{(2\pi\alpha)^{1/2}} \exp\left\{-\frac{1}{2\alpha}(u-\alpha/2)^2\right\} du \\ & = C(1|2) \int_k^{\infty} \frac{1}{(2\pi\alpha)^{1/2}} \exp\left\{-\frac{1}{2\alpha}(u+\alpha/2)^2\right\} du \end{aligned}$$

o equivalente a:

$$C(2|1)\phi\left(\frac{k-\alpha/2}{\sqrt{\alpha}}\right) = C(1|2)\left(1-\phi\left(\frac{k+\alpha/2}{\sqrt{\alpha}}\right)\right) \quad (2.14)$$

donde  $\phi(z) = \int_{-\infty}^z (2\pi)^{-1} \exp\left\{-\frac{1}{2}t^2\right\} dt$ .

Ahora supóngase que se tienen un grupo de  $N$  individuos, con observaciones  $x^\alpha$ ,  $\alpha=1, \dots, N$ , que será completamente clasificado a uno de los  $\pi_i$ ,  $i=1, 2$ . Entonces,

escribiendo  $\bar{x} = \left(\frac{1}{N}\right) \sum_1^N x^\alpha$ ,

$$\prod_{\alpha=1}^N \frac{f_1(x^\alpha)}{f_2(x^\alpha)} = \exp\left\{N\left(\bar{x} - \frac{1}{2}(\mu_1 + \mu_2)\right)' \Sigma^{-1}(\mu_1 - \mu_2)\right\}$$

donde  $N\left(\bar{x} - \frac{1}{2}(\mu_1 + \mu_2)\right)' \Sigma^{-1}(\mu_1 - \mu_2)$  es una distribución normal con medias  $N\alpha/2$ ,  $-N\alpha/2$  y la misma varianza  $N\alpha$  bajo  $\pi_1$  y  $\pi_2$  respectivamente, la regla discriminante de Bayes  $R = (R_1, R_2)$  contra las probabilidades *a priori*  $(p_1, p_2)$  estará dada por:

$$R_1 : N \left( \bar{x} - \frac{1}{2}(\mu_1 + \mu_2) \right)' \Sigma^{-1} (\mu_1 - \mu_2) \geq k \quad (2.16)$$

$$R_2 : N \left( \bar{x} - \frac{1}{2}(\mu_1 + \mu_2) \right)' \Sigma^{-1} (\mu_1 - \mu_2) < k$$

La regla discriminante Minmax  $R = (R_1, R_2)$  está dada por (2.16) donde  $k$  es determinada por:

$$C(2|1) \phi \left( \frac{k - N\alpha/2}{\sqrt{N\alpha}} \right) = C(1|2) \left( 1 - \phi \left( \frac{k + N\alpha/2}{\sqrt{N\alpha}} \right) \right). \quad (2.17)$$

Si los parámetros son desconocidos, la estimación de estos parámetros se obtendrán de las muestras aleatorias Independientes de medida  $N_1$  y  $N_2$  de  $\pi_1$  y  $\pi_2$  respectivamente.

Sean  $x_\alpha^{(1)} = (x_{\alpha 1}^1, \dots, x_{\alpha p}^1)'$ ,  $\alpha = 1, \dots, N_1$ ,  $x_\alpha^{(2)} = (x_{\alpha 1}^2, \dots, x_{\alpha p}^2)'$ ,  $\alpha = 1, \dots, N_2$ , muestras de observaciones (independientes) de  $\pi_1$  y  $\pi_2$ , y, respectivamente, y sea

$$x^{(i)} = \frac{1}{N_i} \sum_{\alpha=1}^{N_i} x_\alpha^{(i)}, \quad i = 1, 2,$$

$$(N_1 + N_2 - 2)s = \sum_{i=1}^2 \sum_{\alpha=1}^{N_i} (x_\alpha^{(i)} - x^{(i)}) (x_\alpha^{(i)} - x^{(i)})'$$

Se sustituyen estos estimadores por los parámetros desconocidos en la expresión de  $U$  para obtener la muestra de la función discriminante  $[v(x)]$  :

$$v = \left( x - \frac{1}{2}(x^{(1)} + x^{(2)}) \right)' s^{-1} (x^{(1)} - x^{(2)}), \quad (2.18)$$

la cual es usada en el mismo modo que  $U$  en el caso de que los parámetros conocidos para definir la regla de discriminante  $R$ . Cuando se clasifica un grupo de  $N$  individuos en vez de uno solo, se puede obtener el estimador de  $\Sigma$  tomando su estimador como  $s$ , definido por:

$$(N_1 + N_2 + N - 3)s = \sum_{i=1}^2 \sum_{\alpha=1}^{N_i} (x_{\alpha}^{(i)} - \bar{x}^{(i)}) (x_{\alpha}^{(i)} - \bar{x}^{(i)})' + \sum_{\alpha=1}^N (x^{\alpha} - \bar{x}) (x^{\alpha} - \bar{x})'$$

La muestra de la función discriminante en este caso es:

$$v = N \left( \bar{x} - \frac{1}{2} (\bar{x}^{(1)} + \bar{x}^{(2)}) \right)' s^{-1} (\bar{x}^{(1)} + \bar{x}^{(2)}) \quad (2.19)$$

La regla de discriminante basada en  $v$  es una regla plug-in. Para encontrar los puntos cutoff  $k$  es necesario encontrar una distribución de  $V$ .

En 1963 Okamoto dio una expresión asintótica de la distribución de  $V$ , de la siguiente manera:

$$Z = \bar{X} - \frac{1}{2} (\bar{X}^{(1)} + \bar{X}^{(2)}), \quad Y = \bar{X}^{(1)} - \bar{X}^{(2)} \quad (2.20)$$

$$(N_1 + N_2 - 2)s = \sum_{i=1}^2 \sum_{\alpha=1}^{N_i} (X_{\alpha}^{(i)} - \bar{X}_{\alpha}^{(i)}) (X_{\alpha}^{(i)} - \bar{X}_{\alpha}^{(i)})'$$

$Z$  y  $Y$  se distribuyen normal  $p$ -variadas con:

$$E(Y) = \mu_1 - \mu_2, \quad \text{cov}(Y) = \left( \frac{1}{N_1} + \frac{1}{N_2} \right) \Sigma,$$

$$E_1(Z) = \frac{1}{2} (\mu_1 - \mu_2), \quad E_2(Z) = \frac{1}{2} (\mu_1 - \mu_2), \quad (2.21)$$

$$\text{cov}(Z) = \left(1 + \frac{1}{4N_1} + \frac{1}{4N_2}\right) \Sigma, \quad \text{cov}(Y, Z) = \left(\frac{1}{2N_1} - \frac{1}{2N_2}\right) \Sigma,$$

y  $(N_1 - N_2 - 2)$  es una distribución independiente de  $Z$ ,  $Y$  se distribuye como una función Wishart  $W_p(N_1 - N_2 - 2, \Sigma)$  donde  $N_i > p$ ,  $i = 1, 2$ . Si  $N_1 = N_2$ ,  $Y$  y  $Z$  son independientes y la distribución de  $X$  que viene de  $\pi_1$  es la misma que da  $-V$  si  $X$  viene de  $\pi_2$ . Un resultado similar considera la dependencia de  $V$  en  $X$ . De manera que si  $v \geq 0$  es la región  $R_1$  y  $v < 0$  es la región de  $R_2$ , entonces la probabilidad de tener un mala clasificación en  $x$  cuando en realidad viene de  $\pi_1$  es igual a la probabilidad de mala clasificación cuando viene de  $\pi_2$ . Lo que es más, dados  $X^{(i)} = x^{(i)}$ ,  $i = 1, 2$ ,  $S = s$ , la distribución condicional de  $V$  es normal con medias y varianza:

$$E_1(V) = \left( \mu_1 - \frac{1}{2}(\bar{x}^{(1)} + \bar{x}^{(2)}) \right)'_2 s^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)}),$$

$$E_2(V) = \left( \mu_2 - \frac{1}{2}(\bar{x}^{(1)} + \bar{x}^{(2)}) \right)'_2 s^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)}), \quad (2.22)$$

$$\text{var}(V) = (\bar{x}^{(1)} + \bar{x}^{(2)})' s^{-1}(\bar{x}^{(1)} + \bar{x}^{(2)}).$$

## 2.2.4 Matrices de covarianza diferentes

La suposición de igual covarianza se satisface rara vez, sin embargo, en algunos casos las dos matrices de covarianza son tan iguales que el resultado tiene muy poca o ninguna diferencia al suponer que son iguales. Cuando ellas son totalmente diferentes se tendrá:

$$\frac{f_1(x)}{f_2(x)} = \left( \frac{|\Sigma_2|}{|\Sigma_1|} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} x'(\Sigma_1^{-1} - \Sigma_2^{-1})x + x'(\Sigma_1^{-1} - \Sigma_2^{-1})\mu - \frac{1}{2} (\mu_1' \Sigma_1^{-1} \mu_1 - \mu_2' \Sigma_2^{-1} \mu_2) \right\}$$

Entonces la regla discriminante de Bayes  $R = (R_1, R_2)$  contra las probabilidades *a priori*  $(p_1, p_2)$  está dada por:

$$R_1 : \frac{1}{2} \log \left( \frac{|\Sigma_2|}{|\Sigma_1|} \right) - \frac{1}{2} \mu_1' \Sigma_1^{-1} \mu_1 + \frac{1}{2} \mu_2' \Sigma_2^{-1} \mu_2 - \frac{1}{2} (x'(\Sigma_1^{-1} - \Sigma_2^{-1})x - 2x'(\Sigma_1^{-1} \mu_1 - \Sigma_2^{-1} \mu_2)) \geq k$$

donde  $k = \log \left( \frac{p_2 C(1|2)}{p_1 C(2|1)} \right)$ .

La cantidad:

$$(x'(\Sigma_1^{-1} - \Sigma_2^{-1})x - 2x'(\Sigma_1^{-1} \mu_1 - \Sigma_2^{-1} \mu_2))$$

es llamada la "*Función Cuadrática la discriminante*" y en el caso de diferentes matrices de varianza una tendrá que utilizar dicha función cuadrática, ya que  $\Sigma_1^{-1} - \Sigma_2^{-1}$  no se elimina.

Para la regla discriminante  $R$  Minmax primero se tiene que encontrar  $k$  tal que (2.9) se satisfaga. Típicamente esto conlleva a encontrar la función cuadrática de discriminante cuando  $x$  viene de  $\pi_i$ ,  $i = 1, 2$ .

La distribución de esta función cuadrática es muy complicada, fue estudiada por Cavalli (1945) para el caso particular  $p = 1$ ; por Smith (1947), Cooper (1963, 1965) y Bunke (1964); Okamoto (1963) la estudió tomando el caso particular  $\mu_1 = \mu_2$ ; Bartlett y Please (1963) para el caso  $\mu_1 = \mu_2 = 0$  y

$$\Sigma_i = \begin{pmatrix} 1 & \rho_i & \dots & \rho_i \\ \rho_i & 1 & \dots & \rho_i \\ \vdots & \vdots & 1 & \vdots \\ \rho_i & \rho_i & \dots & 1 \end{pmatrix} \quad (2.31)$$

y Han (1968-1970) para diferentes formas especiales de  $\Sigma_i$ . Okamoto (1963) derivó la regla discriminante Minmax y la forma de regla discriminante de Bayes para el caso de parámetros conocidos. Él también estudió algunas propiedades de la función de riesgo de la regla discriminante de Bayes y sugirió un método para escoger las componentes. También trató el caso cuando  $\Sigma_i$  es desconocida y el valor común de  $\mu_i$  puede ser conocido o no. Bunke (1964) mostró que la regla discriminante Minmax es consistente. Siguiendo el método de Kiefer y Schwartz (1965), Nishida (1971) obtuvo una regla admisible de la regla discriminante de Bayes para cuando los parámetros son desconocidos.

A continuación se discutirá una solución al problema planteado por Bahadur y Anderson (1962), basado solamente en una función discriminante lineal.

Sean  $b \neq 0$  un vector  $p$ -columna y  $c$  un escalar. Una observación en  $x$  un individuo es clasificado como si viniera de  $\pi_1$  si  $b'x \leq c$  y como si viniera de  $\pi_2$  si  $b'x > c$ . Las probabilidades de mala clasificación con esta regla discriminante pueden ser fácilmente evaluada a partir del hecho de que  $b'x$  se distribuye normal con media  $b'\mu_1$  y varianza  $b'\Sigma_1 b$  si  $x$  viene de  $\pi_1$ , y con media  $b'\mu_2$  y varianza  $b'\Sigma_2 b$  si  $x$  viene de  $\pi_2$  y está dado por:

$$P_1 = P(2|1, R) = 1 - \phi(z_1), \quad P_2 = P(1|2, R) = 1 - \phi(z_2), \quad (2.32)$$

donde

$$z_1 = \frac{c - b'\mu_1}{(b'\Sigma_1 b)^{1/2}}, \quad z_2 = \frac{b'\mu_2 - c}{(b'\Sigma_2 b)^{1/2}} \quad (2.33)$$

Se asumirá que  $C(2|1) = C(1|2)$ . De aquí que cada procedimiento (obtenido al variar  $b$ ) pueden ser evaluados en términos de las dos probabilidades de mala clasificación  $P_1, P_2$ . Ya que la transformación por la distribución normal acumulativa  $\phi(z_2)$  es estrictamente monótona, comparaciones de los diferentes procedimientos lineales sólo pueden ser hechos en términos de los argumentos  $z_1, z_2$  y dados en (2.33). Dado  $z_2$ , eliminando  $c$ , se obtendrá de también de (2.33):

$$z_1 = \frac{b' \delta - z_2 (b' \Sigma_2 b)^{1/2}}{(b' \Sigma_1 b)^{1/2}},$$

donde  $\delta = (\mu_2 - \mu_1)$ . Ya que  $z_1$  es homogénea en  $b$  de grado 0, se puede restringir  $b$  a estar en una elipse, dicha  $b' \Sigma_1 b = cte$ , y en este dominio cerrado  $z_1$  es continua y por ello tiene un máximo. A través de los procedimientos lineales con una coordenada  $z_2$  específica (equivalentemente con una  $P_2$  específica) hay al menos un procedimiento que máxima la coordenada  $z_1$  (equivalentemente  $P_1$  se minimiza).

**Lema 2.2.1** *La coordenada máxima  $z_1$  es una función decreciente de  $z_2$ .*

Demostración:

Sea  $z_1^* > z_2^*$  y sea  $b^*$  un vector que maximice a  $z_1^*$  dado  $z_2^*$ . Entonces

$$\max_b z_1 = \max_b \frac{b' \delta - z_2 (b' \Sigma_2 b)^{1/2}}{(b' \Sigma_1 b)^{1/2}} \geq \frac{b^* \delta - z_2 (b^* \Sigma_2 b^*)^{1/2}}{(b^* \Sigma_1 b^*)^{1/2}} > \frac{b^* \delta - z_2 (b^* \Sigma_2 b^*)^{1/2}}{(b^* \Sigma_1 b^*)^{1/2}} = \max z_1^*$$

El conjunto  $z_2$  con su correspondiente  $z_1$  máximo es entonces una curva en el plano  $z_1, z_2$  que va hacia abajo a la derecha. Ya que  $\delta \neq 0$ , la curva está arriba y a la derecha del origen.



**Teorema 2.2.7** Una regla de discriminante lineal  $R$  con  $P_1 = 1 - \phi(z_1)$ ,  $P_2 = 1 - \phi(z_2)$ , donde  $z_1$  es maximizado con respecto de  $b$  para  $z_2$  dada es admisible.

Demostración:

Supóngase que  $R$  no es admisible. Entonces ahí hay una regla discriminante lineal  $R^* = (R_1^*, R_2^*)$  con argumentos  $(z_1^*, z_2^*)$  tal que  $z_1^* \geq z_1$ ,  $z_2^* \geq z_2$  con al menos una desigualdad estricta. Si  $z_2^* = z_2$ , entonces  $z_1^* > z_1$ , lo que contradice el hecho de que  $z_1$  es un máximo. Si  $z_2^* > z_2$  la coordenada máxima correspondiente a  $z_2^*$  debe ser menor que  $z_1$ , lo que contradice que  $z_1^* \geq z_1$ .

A continuación se caracterizará analíticamente la regla discriminante lineal admisible.

**Lema 2.2.2** Si un punto  $(\alpha_1, \alpha_2)$  con  $\alpha_i > 0$ ,  $i = 1, 2$  es admisible, entonces existe  $t_i > 0$ ,  $i = 1, 2$ , tal que la regla discriminante lineal definida por:

$$b = (t_1 \Sigma_1 + t_2 \Sigma_2)^{-1} \delta \quad (2.34)$$

$$c = b' \mu_1 + t_1 b' \Sigma_1 b = b' \mu_2 - t_2 b' \Sigma_2 b \quad (2.35)$$

Demostración:

Sea la regla discriminante lineal admisible definida por el vector  $\beta$  y el escalar  $\gamma$ . La recta

$$z_1 = \frac{s - \beta^{-1} \mu_1}{(\beta' \Sigma_1 \beta)^{1/2}}, \quad z_2 = \frac{\beta^{-1} \mu_2 - s}{(\beta' \Sigma_2 \beta)^{1/2}} \quad (2.36)$$

como  $s$  es parámetro, tiene pendiente negativa en el punto  $(\alpha_1, \alpha_2)$ . Por lo tanto ahí existen dos números positivos  $t_1, t_2$  tal que la recta (2.36) es tangente a la elipse:

$$\frac{z_1^2}{t_1} + \frac{z_2^2}{t_2} = k \quad (2.37)$$

en el punto  $(\alpha_1, \alpha_2)$ . Considérese que la recta definida por un vector arbitrario  $b$  y todos los escalares  $c$ . Esta recta es tangente a una elipse similar o concéntrica a (2.37) si  $c$  en (2.33) es elegido de manera que  $-z_1 t_2 / z_2 t_1$  es igual a la pendiente de dicha recta. Para  $b$  dado, los valores de  $c$  y  $z_1, z_2$  resultantes son:

$$c = \frac{t_1 b' \sum_1 b b' \mu_2 + t_2 b' \sum_2 b b' \mu_1}{t_1 b' \sum_1 b + t_2 b' \sum_2 b}$$

$$z_1 = \frac{t_1 (b' \sum_1 b)^{1/2} b' \delta}{t_1 b' \sum_1 b + t_2 b' \sum_2 b}, \quad z_2 = \frac{t_2 (b' \sum_2 b)^{1/2} b' \delta}{t_1 b' \sum_1 b + t_2 b' \sum_2 b} \quad (2.38)$$

Dicho punto  $(z_1, z_2)$  esta en la elipse, entonces:

$$\frac{z_1^2}{t_1} + \frac{z_2^2}{t_2} = \frac{(b' \delta)^2}{b' (t_1 \sum_1 + t_2 \sum_2) b} \quad (2.39)$$

El máximo del lado derecho de (2.39) con respecto de  $b$  ocurre cuando  $b$  está dada como en (2.34). Sin embargo, el máximo debe corresponder a el procedimiento admisible, si hubiera una  $b$  tal que la constante (2.39) fuera mayor que  $k$ , el punto  $(\alpha_1, \alpha_2)$  estaría dentro de la elipse con dicha constante y estaría más cerca del origen que la recta tangente en  $(z_1, z_2)$ . Entonces algunos puntos en esa recta (correspondientes a los procedimientos con  $b$  vector y  $c$  escalar) podrían ser mejores. La expresión para el valor de  $c$  en (2.38) y (2.35) son los mismos si usa el valor de  $b$  como el dado en (2.34).

Observación: si  $\sum_1, \sum_2$  son definidas positivas y  $t_i > 0, i = 1, 2, t_1 \sum_1 + t_2 \sum_2$  también es definida positiva, cualquier múltiplo de (2.34), (2.35) son soluciones equivalentes. Cuando  $b$  en (2.34) es normalizado se tiene que:

$$b' \delta = b'(t_1 \Sigma_1 + t_2 \Sigma_2) b = \delta''(t_1 \Sigma_1 + t_2 \Sigma_2)' \delta, \quad (2.40)$$

entonces de (2.38) se obtiene:

$$z_1 = t_1 (b' \Sigma_1 b)^{1/2}, \quad z_2 = t_2 (b' \Sigma_2 b)^{1/2}.$$

Entonces son homogéneos de grado 0 en  $t_1$  y  $t_2$  para dado  $b$  por (2.34) se encuentra conveniente tomar  $t_1 + t_2 = 1$  cuando  $t_i > 0, i = 1, 2, t_1 - t_2 = 1$  y cuando  $t_1 > 0, t_2 < 0$  y  $t_1 - t_2 = 1, t_2 > 0, t_1 < 0$ .

**Teorema 2.2.8** Una regla de discriminante lineal con las condiciones pedidas en (2.34) y (2.35) y para cualquier  $t_1, t_2$  tal que  $t_1 \Sigma_1 + t_2 \Sigma_2$  es definida positiva es admisible.

Demostración:

Si  $t_i > 0, i = 1, 2$ , las correspondientes  $z_1, z_2$  son también positivas. Si esta regla discriminante lineal no es admisible, habrá una regla discriminante lineal que puede ser mejor (ya que el conjunto de todas las reglas discriminante es completo) y ambos argumentos para esta regla pueden ser algunas veces positivos. Por el Lema 2.3.2 a regla discriminante definida por:

$$\beta = (\tau_1 \Sigma_1 + \tau_2 \Sigma_2)^{-1} \delta, \quad (2.40)$$

para  $\tau_i > 0, i = 1, 2$ , tal que  $\tau_1 + \tau_2 = 1$ . Por otro lado, por ser monótonas  $z_1, z_2$  como funciones de  $t_1$ , una de las coordenadas correspondientes a  $\tau_1$ , tendrá que ser menor que una de las coordenadas correspondientes a  $t_1$ . Esto

demuestra que la regla discriminante lineal correspondiente a  $\beta$  no es mejor que la definida por  $b$ . Por consiguiente el teorema está demostrado para  $t_i > 0, i = 1, 2$ . Si  $t_1 = 0$ , entonces  $z_1 = 0, b = \Sigma_2^{-1} \delta, z_2 = (\delta' \Sigma_2^{-1} \delta)^{1/2}$ . Por otro lado, para alguna  $b$  si  $z_1 = 0$ , entonces

$z_2 = b' \delta (\delta' \Sigma_2^{-1} \delta)^{1/2}$  y  $z_2$  es máximo si  $b = \Sigma_2^{-1} \delta$ . Similarmente si  $t_2 = 0$  la solución asumida en el teorema es óptima.

Ahora considérese  $t_1 > 0$ ,  $t_2 < 0$  y  $t_1 - t_2 = 1$ . Alguna elipse

$$\frac{z_1^2}{t_1} + \frac{z_2^2}{t_1} = k \quad (2.41)$$

para  $k > 0$  corta en el eje  $z_1$  en  $\pm (t_1 k)^{1/2}$ . La regla supuesta en el teorema tiene  $z_1 > 0$  y  $z_2 < 0$ . De (2.33) se tienen que:

$$\frac{(c - b' \mu_1)^2}{t_1 b' \Sigma_1 b} + \frac{(b' \mu_2 - c)^2}{t_2 b' \Sigma_2 b} = k \quad (2.42)$$

El máximo de dicha expresión con respecto de  $c$  dado  $b$  es obtenido por  $c$  como en (2.38). Entonces  $z_1$ ,  $z_2$  son de la forma (2.38) y (2.42) se reduce a (2.40). El máximo de (2.42) está entonces dado por  $b = (t_1 \Sigma_1 + t_2 \Sigma_2)^{-1} \delta$ . Este punto estará sobre una elipse con  $k$  mayor.

El caso en donde  $t_1 < 0$ ,  $t_2 > 0$  es demostrado similarmente.

Dados  $t_1$ ,  $t_2$  tal que  $t_1 \Sigma_1 + t_2 \Sigma_2$  es definida positiva, la  $b$  se calcularía óptima con:

$$(t_1 \Sigma_1 + t_2 \Sigma_2) b = \delta, \quad (2.43)$$

y calcular  $c$  como en (2.35). Usualmente  $t_1$ ,  $t_2$  no se dan. Una solución deseada puede ser obtenida de la siguiente manera. Para otro tipo de soluciones consultar Bahadur y Anderson (1962).

**"Minimización de una probabilidad de mala clasificación dada la otra"**

Supóngase que  $z_2$  está dada y además es positiva. Entonces si la máxima  $z_1 > 0$ , se puede encontrar con  $t_2 = 1 - t_1$  tal que  $z_2 = t_2(b' \Sigma_2 b)^{1/2}$  con  $b$  dada como en (2.43). La solución puede ser aproximada por ensayo y error. Para  $t_2 = 0$ ,  $z_2 = 0$  y  $t_2 = 1$ ,  $z_2 = t_2(b' \Sigma_2 b)^{1/2} = (b' \delta)^{1/2} = \delta' \Sigma_2^{-1} \delta$  donde  $\Sigma_2 b = \delta$ , se podría tratar con otros valores de  $t_2$  sucesivamente para resolver (2.43) e insertar la solución en  $b' \Sigma_2 b$  hasta que  $t_2(b' \Sigma_2^{-1} b)^{1/2}$  se acerca mucho a la  $z_2$  deseada.

Para  $t_1 > 0$ ,  $t_2 < 0$  y  $t_1 - t_2 = 1$ ,  $z_2$  es una función decreciente de  $t_2$  ( $t_2 \leq 1$ ) y en  $t_2 = 1$ ,  $z_2 = (\delta' \Sigma_2^{-1} \delta)^{1/2}$ . Si la  $z_2$  dada es mayor que  $(\delta' \Sigma_2^{-1} \delta)^{1/2}$ , entonces  $z_1 < 0$  y se busca un valor de  $t_2$  tal que  $z_2 = (\delta' \Sigma_2^{-1} \delta)^{1/2}$ . Se requiere que  $t_2$  sea bastante grande de manera que  $t_1 \Sigma_1 + t_2 \Sigma_2 = (t_2 - 1) \Sigma_1 + t_2 \Sigma_2$  sea definida positiva.

**"La Regla Discriminante Minmax"**

La regla discriminante lineal Minmax es una regla admisible con  $z_1 = z_2$ . Obviamente en el caso de  $z_1 = z_2 > 0$  y  $t_i > 0$ ,  $i = 1, 2$ . De aquí que se quiere encontrar  $t_2 = 1 - t_1$  de tal manera que:

$$0 = z_1^2 - z_2^2 = b'(t_1^2 \Sigma_1 - (1 - t_1)^2 \Sigma_2) b \quad (2.44)$$

Los valores de  $b$  y  $t_1$  satisfacen a (2.42) y (2.43), los cuales son obtenidos por el método de ensayo y error.

Ya que  $\Sigma_1, \Sigma_2$  son definidas positivas ahí existe una matriz no singular  $C$  tal que  $\Sigma_1 = C'\Delta C, \Sigma_2 = C'C$  donde  $\Delta$  es una matriz diagonal con  $\lambda_i, K, \lambda_p$  como elementos en la diagonal y las raíces del  $|(\Sigma_1 - \lambda \Sigma_2)| = 0$ .

Sea  $b^* = (b_1^*, K, b_p^*) = Cb$ . Entonces (2.44) puede ser escrita como:

$$\sum_{i=1}^p (\lambda_i - \theta) b_i^{*2} = 0 \quad (2.45)$$

donde  $\theta = \frac{(1-t_1^2)}{t_1^2}$ . Si  $\lambda_i - \theta$  son todos positivos o todos negativos, (2.45) puede no tener una solución para  $b^*$ .

## 2.2.5 Regla discriminante de la razón de verosimilitudes

Considerando el caso en que si  $X$  es un individuo que pertenece a  $P_r$ , entonces  $X$  se distribuye como  $N_p(\mu, \Sigma)$ . El criterio consiste en calcular las verosimilitudes de las hipótesis:

$H_r : X$  y los renglones de  $X_r$  pertenecen a  $P_r$  y las filas de  $X_r$  pertenecen al grupo de  $P_k, k \neq r$ .

La regla consiste en asignar  $X$  a la población cuya hipótesis  $H_r$  tiene la mayor verosimilitud; esta verosimilitud bajo  $H_r$  está dada por:

$$L_r(\mu_1, K, \mu_2, \Sigma) = \prod_{i=1}^R \prod_{j=1}^R |2\pi \Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(X_{ij} - \mu_i)' \Sigma^{-1}(X_{ij} - \mu_i)\right\} |2\pi \Sigma|^{-1/2}$$

$$\exp\left\{-\frac{1}{2}(X_{ij} - \mu_r)' \Sigma^{-1}(X_{ij} - \mu_r)\right\} \quad (2.46)$$

Ahora, si los parámetros son conocidos entonces la región de clasificación en  $P_r$  es:

$$R_r = \{X \in \mathfrak{R}^p | L_r(x) \geq L_k(x), \forall r \neq k\}, \quad r = 1, 2, k, p.$$

donde  $L_r(x)$  es la función de verosimilitud de la  $r$ -ésima población definida por (2.46).

Si los parámetros son desconocidos, entonces éstos deben reemplazarse por sus correspondientes estimadores máximo verosimiles. Si se definen como  $\mu_k^{(r)}$  y  $\Sigma^{(r)}$  a los estimadores para la media y para la matriz de varianzas asociados al  $k$ -ésimo grupo para la hipótesis  $H_r$ , toman entonces las siguientes expresiones:

$$\mu_k^{(r)} = \begin{cases} \bar{X}_k = \frac{n_r \bar{X}_r + X}{n_r + 1}, & k = r \\ \bar{X}_k, & k \neq r \end{cases}$$

y

$$\Sigma^{(r)} = W^{(r)} = \frac{1}{n+1} \left[ \sum_{k=1}^p (n_k + I_{(r)}(k)) S_k^{(r)} \right],$$

donde

$$S_k^{(r)} = \begin{cases} S_k = \frac{1}{n_k} \sum_{j=1}^{n_k} (X_{kj} - X_k)(X_{kj} - X_k)', & k \neq r \\ S_r^{(r)} = \frac{1}{n_r + 1} \left[ \sum_{j=1}^{n_r} (X_{rj} - X_r^{(r)})(X_{rj} - X_r^{(r)})' + (X - X_r^{(r)})(X - X_r^{(r)})' \right], & k = r \end{cases}$$

Por lo que las regiones de clasificación están dadas por:

$$R_r = \{X \in \mathbb{R}^p \mid L_r(x) \geq L_k(x), \forall r \neq k\}, \quad r = 1, 2, k, p.$$

y donde:

$$L_r(X) = \prod_{i=1}^p \prod_{j=1}^{n_j} 2\pi \Sigma^{-1/2} \exp\left\{-\frac{1}{2}(X_{ij} - X^{(r)})' W^{(r)-1} (X_{ij} - X^{(r)})\right\} 2\pi \Sigma_i^{-1/2} \exp\left\{-\frac{1}{2}(X - X^{(r)})' W^{(r)-1} (X - X^{(r)})\right\}, \quad r = 1, k, p. \quad (2.47)$$

De la ecuación anterior se puede observar que las regiones de clasificación están en términos de la matriz  $W^{(r)}$  la cual se puede simplificar cuando se desarrolla de la siguiente igualdad:

$$\begin{aligned} (n_r + 1)S_r^{(r)} &= \sum_{j=1}^{n_r} (X_{rj} - X_r^{(r)})(X_{rj} - X_r^{(r)})' + (X - X_r^{(r)})(X - X_r^{(r)})' \\ &= \sum_{j=1}^{n_r} (X_{rj} - X_r + X_r - X_r^{(r)})(X_{rj} - X_r + X_r - X_r^{(r)})' + (X - X_r^{(r)})(X - X_r^{(r)})' \\ &= \sum_{j=1}^{n_r} (X_{rj} - X_r)(X_{rj} - X_r)' + \sum_{j=1}^{n_r} (X_r - X_r^{(r)})(X_r - X_r^{(r)})' + (X - X_r^{(r)})(X - X_r^{(r)})' \\ &= n_r S_r + n_r (X_r - X_r^{(r)})(X_r - X_r^{(r)})' + (X - X_r^{(r)})(X - X_r^{(r)})' \\ &= n_r S_r + \frac{n_r}{(n_r + 1)^2} (X_r - X_r)(X_r - X_r)' + \frac{n_r^2}{(n_r + 1)^2} (X_r - X_r)(X_{rj} - X_r)' \\ &= n_r S_r + \frac{n_r}{(n_r + 1)} (X_r - X_r)(X_r - X_r)' \end{aligned}$$



Quedando:

$$\begin{aligned}
 W^{(r)} &= \frac{1}{n+1} \left[ \sum_{k+1}^p (n_k + I_{(r)}(k)) S_k^{(r)} \right] \\
 &= \frac{1}{n+1} \left[ \sum_{k+1}^p n_k S_k + S_r^{(r)} \right] \\
 &= \frac{n}{n+1} W + \frac{1}{(n+1)(n_r+1)} (X - X_r)(X - X_r)'.
 \end{aligned}$$

Con lo que la función de verosimilitud de la muestra  $X_1, K, X_p$  bajo la hipótesis  $H_r$  dada (2.46) se puede simplificar de la siguiente manera:

$$\hat{L}_r(X) = |2\pi W^{(r)}|^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{2} \text{tr}(W^{(r-1)})(n+1)W^{(r)}\right\} |2\pi W^{(r)}|^{-\frac{n-1}{2}} \exp\left\{-\frac{n+1}{2} p\right\}$$

Por lo que ahora se debe de encontrar:

$$\begin{aligned}
 \min_r |W^{(r)}| &= \min \left| \frac{n}{n+1} W^{(r)} + \frac{n_r}{(n+1)(n_r+1)} (X - X_r^{(r)})(X - X_r^{(r)})' \right| \\
 &= \min |W^{(r)}| \left[ 1 + \frac{n_r}{(n_r+1)} (X - X_r^{(r)})(nW^{(r)})^{-1} (X - X_r^{(r)})' \right] \\
 &= \min_r \left[ \frac{n_r}{(n_r+1)} (X - X_r^{(r)})(nW^{(r)})^{-1} (X - X_r^{(r)})' \right]
 \end{aligned}$$

y desarrollando la forma cuadrática en la ecuación anterior se obtiene que la regla que asigna  $X$  a  $P_r$  en:

$$\max_r \left[ \frac{n_r}{(n_r + 1)} a_r' \left( X - \frac{1}{2} X \right) \right]$$

donde se define:

$$a_r = W^{-1} X_r.$$

Ahora, si en particular se considera el caso de dos poblaciones normales con matriz de covarianzas igual, para las cuales un nuevo individuo  $X$  debe discriminarse, el juego de hipótesis es de la forma:

$H_1$  :  $X$  y los renglones de  $X_1$  pertenecen a  $P_1$  y las filas de  $X_2$  provienen de  $P_2$ .

$H_2$  : Los renglones de  $X_1$  provienen de  $P_1$  y las filas de  $X_2$  pertenecen de  $P_2$ .

Si los parámetros de  $X_1$  y  $X_2$  son desconocidos se deben reemplazarse por sus estimadores máximo verosímiles. Los estimadores para  $\mu_1$ ,  $\mu_2$  y  $\Sigma$  bajo la hipótesis  $H_1$  están dadas por:

$$\mu_i = \begin{cases} \bar{X}_i^{(1)} = \frac{n_1 \bar{X}_1 + X}{n_1 + 1}, & i = 1 \\ \bar{X}_2, & i = 2 \end{cases}$$

y

$$\hat{\Sigma}^{(1)} = \frac{1}{n_1 + n_2 + 1} \left\{ W + \frac{n_1}{n_1 + 1} (X - \bar{X}_1)(X - \bar{X}_1)' \right\}.$$

Donde  $nW = n_1 S_1 + n_2 S_2$ .

Bajo la hipótesis  $H_2$  los estimadores máximo verosímiles para  $\mu_1$ ,  $\mu_2$  y  $\Sigma$  están dados por:

$$\mu_i = \begin{cases} X_1, & i = 1 \\ X_2^{(2)} = \frac{n_2 X_2 + X}{n_2 + 1}, & i = 2 \end{cases}$$

y

$$\Sigma^{(2)} = \frac{1}{n_1 + n_2 + 1} \left\{ W + \frac{n_2}{n_2 + 1} (X - X_2)(X - X_2)' \right\}.$$

Entonces el cociente de verosimilitud es la proporción:

$$\left| \frac{\Sigma^{(2)}}{\Sigma^{(1)}} \right| = \frac{1 + \frac{n_2}{n_2 + 1} (X - X_2)' W^{-1} (X - X_2)}{1 + \frac{n_1}{n_1 + 1} (X - X_1)' W^{-1} (X - X_1)}$$

La prueba acepta  $H_0$ , es decir, se asigna  $X$  a  $P_i$  si y solo si:

$$1 + \frac{n_2}{n_2 + 1} (X - X_2)' W^{-1} (X - X_2) > 1 + \frac{n_1}{n_1 + 1} (X - X_1)' W^{-1} (X - X_1)$$

y si los tamaños de muestra son iguales, es decir, si  $N_1 = N_2$  es equivalente a la regla de máxima verosimilitud. Pero si los tamaños de muestra son diferentes entonces este método tiende a clasificar a  $X$  a la población que tiene el tamaño de muestra más grande.

# CAPÍTULO III: Análisis Estadístico

## 3.1 INTRODUCCIÓN

En este capítulo se realizó un análisis estadístico bajo dos vertientes, la del análisis con componentes principales y la del análisis discriminante.

Se analizaron las once variables contenidas en dos muestras que corresponden a los años 1999 y 2000, compuestas por las 32 entidades federativas, esto dio como resultado dos cuadros, los cuales se muestran en las tablas 3.1 y 3.2.

Las variables son<sup>1</sup>:

- HOM: mortalidad infantil de hombres.
- MUJ: mortalidad infantil de mujeres.
- PEAH: población económicamente activa de hombres.
- PEAM: población económicamente activa de mujeres.
- PEIH: población económicamente inactiva de hombres.
- PEIM: población económicamente inactiva de mujeres.
- CONEXT: número de unidades de consulta externa.
- HOS: número unidades de hospitalización general.

---

<sup>1</sup> Ver glosario

- CAM: número de camas censables.
- QUIROF: número de quirófanos.
- LAB: número de laboratorios

Tabla 3.1 Datos de las 32 entidades federativas correspondientes al año 1999

	HOM	MUJ	PEAH	PEAM	PEIH	PEIM	CONEXT	HOS	CAM	QUIR	LAB
Ags.	243	169	347079.0	303313.0	154925.5	175871.0	106	10	799	30	13
B.C.	653	516	945703.0	786446.0	442557.00	444302.50	206	19	1549	59	28
B.C.S.	83	66	557503.5	327184.5	173202.50	434171.50	146	19	611	32	20
Cam.	125	85	264559.0	209419.5	123947.00	122539.00	259	19	671	29	21
Coah.	297	224	1523812.0	1228911.0	622547.50	686167.00	350	41	2445	93	59
Col.	89	69	1152268.5	958000.0	527917.50	581588.50	150	10	457	15	14
Chis.	871	668	880012.5	743267.5	398929.50	405563.50	993	40	1752	80	72
Chih.	749	535	202395.5	175759.0	94693.00	97121.00	534	36	2319	76	61
D.F.	1841	1482	3233410.5	2969201.5	1543840.50	1811756.00	655	109	16581	559	307
Dgo.	81	68	551631.0	457519.5	242818.50	269667.00	453	18	1234	41	39
Gto.	1494	1083	2815501.5	2154697.0	1075347.50	1809301.00	586	30	2500	80	51
Gro.	492	315	1308696.5	1076613.5	563110.50	804169.50	961	28	1574	70	70
Hgo.	485	414	948254.0	791913.0	342674.50	462295.00	684	26	1361	39	41
Jal.	1354	1086	1821590.0	1798975.5	936885.50	755030.50	967	50	5946	218	91
Edo.Méx.	4318	3213	4107962.0	3869553.0	1967918.00	1604828.50	1191	64	7140	180	135
Mich.	670	563	1433598.0	1305237.5	644970.50	771091.00	850	45	2068	83	59
Mor.	260	217	582658.5	522137.5	260293.50	303881.50	265	12	834	32	20
Nay.	96	91	356494.0	307349.5	157090.50	166951.50	335	13	723	31	22
N.L.	567	457	1110756.0	536604.0	322878.50	914499.50	531	28	3791	117	74
Oax.	844	609	525987.0	229959.5	149738.50	472661.00	1138	37	1724	61	53
Pue.	2172	1721	1279095.0	625628.5	389739.00	1216943.00	923	51	3204	108	72
Qro.	409	339	343768.0	183857.5	112563.50	326142.00	253	8	768	27	18
Q.R.	197	136	244008.0	99590.0	58831.50	195182.00	186	13	517	26	31
S.L.P.	502	404	366081.0	417730.5	637438.00	910251.50	524	21	1402	44	33
Sin.	189	139	460301.5	499287.0	726055.50	921097.00	444	25	1950	82	39
Son.	433	310	425116.5	446033.0	634358.50	794749.50	358	42	2385	89	59
Tabas.	435	338	344715.5	350836.5	523037.50	721300.50	579	25	1359	61	56
Tamps.	377	313	469239.5	511737.5	788148.00	997521.50	460	37	2701	96	59
Tlaxc.	399	306	176550.5	184604.5	275094.00	370103.00	195	8	458	15	12
Ver.	1434	1096	1251450.0	1325730.5	1871928.50	2640035.50	1355	76	4320	173	100
Yuc.	340	249	292037.0	316637.0	471814.50	615873.00	304	18	1463	44	32
Zac.	280	199	339620.0	163598.5	108561.00	362861.00	377	12	708	33	32

Fuentes: Secretaría de Salud, Anuario de Mortalidad, 1999; DGEI, Boletín de Información Estadística, No. 19, Vol. 1, 1999; INEGI, Encuesta Nacional de Empleo, 1999.

**INTRODUCCIÓN**

Para la tabla anterior, las variables PEAH, PEAM, PEIH y PEIM se obtuvieron sacando la media muestral de los años 1998 y 2000, esto es debido a que los datos del año 1999 el INEGI no los publicados por entidad federativa.

**Tabla 3.2 Datos de las 32 entidades federativas correspondientes al año 2000**

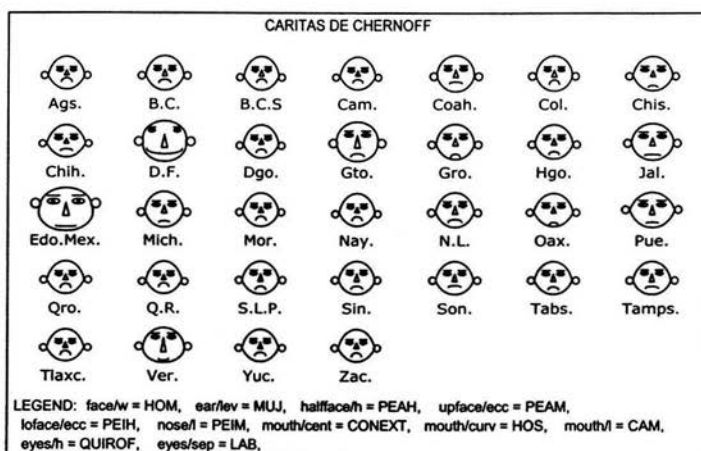
	HOM	MUJ	PEAH	PEAM	PEIH	PEIM	CONEXT	HOS	CAM	QUIR	LAB
Ags.	222	160	458505	487269	231847	124193	109	10	799	29	13
B.C.	600	502	1248684	1246961	694363	383420	210	19	1535	62	29
B.C.S.	87	70	212558	208544	120619	65582	137	18	575	28	21
Cam.	104	81	342673	346877	191823	74010	256	20	665	27	22
Coah.	296	230	1954864	2002322	1101793	556819	344	39	2434	89	61
Col.	97	70	1497882	1552837	812166	423734	150	11	536	19	15
Chis.	916	721	1138865	1156995	630194	302587	1471	43	1830	86	78
Chih.	713	505	265331	274142	150252	83884	526	34	2351	80	63
D.F.	1771	1357	4071271	4484313	2314187	1512422	684	111	16640	578	308
Dgo.	85	52	703458	737755	378863	183106	466	18	1231	40	46
Gto.	1424	1039	2172163	2491904	1113375	612642	582	30	2486	86	52
Gro.	373	269	1473523	1597454	772778	437689	982	27	1607	72	72
Hgo.	433	357	1078283	1154660	511659	255239	718	28	1261	38	43
Jal.	1310	1016	3039332	3271963	1730301	1023352	991	51	6103	225	97
Edo.Méx.	4079	3180	6424544	6708566	3548177	1843237	1185	63	6962	180	137
Mich.	684	473	1874885	2089047	1038308	527332	861	39	1977	78	63
Mor.	272	201	739708	811212	402204	223426	267	12	845	34	19
Nay.	98	74	441771	469545	249934	138115	329	12	693	29	22
N.L.	615	448	1117092	527983	327567	938072	525	29	3667	119	73
Oax.	806	597	931418	402092	264043	853020	1167	37	1688	60	53
Pue.	2042	1671	1314930	650007	406203	1229016	928	57	3371	114	64
Qro.	373	303	353246	202070	113470	318223	291	8	765	27	20
Q.R.	177	160	253203	105830	60591	207937	192	14	526	25	30
S.L.P.	480	381	179506	583944	1090033	1205205	527	20	1371	42	32
Sin.	145	144	214522	630509	1251773	1276321	434	29	1899	83	41
Son.	487	378	218521	550741	1099965	1099546	329	42	2354	89	57
Tabs.	492	320	150065	507580	925480	966761	582	25	1397	62	57
Tamps.	397	324	207964	637815	1372578	1378980	456	37	2650	91	55
Tlaxc.	337	253	78295	239007	468269	493965	189	8	496	18	13
Ver.	1312	1011	523720	1751505	3317515	3574231	1519	76	4222	172	99
Yuc.	309	244	128406	389916	810147	846871	306	18	1496	44	31
Zac.	272	191	333504	152263	112516	375823	397	12	712	34	32

Fuentes: Secretaría de Salud, Anuario de Mortalidad, 2000; DGEI, Boletín de Información Estadística, No. 19, Vol. 1, 2000; INEGI, Encuesta Nacional de Empleo, 2000.

## 3.2 ANÁLISIS ESTADÍSTICO DEL AÑO 1999

## 3.2.1 Análisis de Componentes Principales

La primer opción de análisis fue la de observar las características de cada entidad federativa de acuerdo con las variables mencionadas, para con ello tratar de identificar si existen semejanzas entre ellas y ver la posible formación de grupos; esto se hizo mediante las caritas de Chernoff.



Se puede observar que existen algunos grupos, por ejemplo: la mayoría de los estados presentan mucha similitud en cuanto a la variable HOM, que en las caritas es denominada *face/w*, es decir, el tamaño de la cara, las entidades que saldrían de dicho grupo son Estado de México, D.F., Veracruz, Jalisco y Puebla. Observando la variable MUJ, denominada por *ear/lev* o altura de las orejas, se puede ver que todos los estados se comportan de manera similar. Chiapas, Guerrero, Jalisco, Michoacán, Puebla, Sonora y Veracruz presentan distinta curvatura en la boca (*mouth/cur*), esta se refiere a la variable HOS. Se pueden seguir dando más ejemplos de formación de grupos, pero existe el problema de que resulta difícil determinarlos a simple vista. Por lo anterior, se procedió a realizar un análisis utilizando componentes principales.

El primer paso fue obtener las estadísticas descriptivas de cada una de las variables, las cuales arrojaron los siguientes resultados:

Tabla 3.3 Estadísticas Descriptivas

Variable	Medias	Desv. Sta.
HOM	711.844	842.263
MUJ	546.250	638.223
PEAH	958182.953	924266.885
PEAM	808666.641	849740.355
PEIH	541995.484	486296.051
PEIM	723922.359	571390.405
CONEXT	541.188	340.605
HOS	30.938	21.887
CAM	2416.063	3011.487
QUIROF	85.094	99.116
LAB	56.031	53.698

Fuente: cálculos propios

Para el cálculo de los componentes principales fue necesaria la obtención de la matriz de correlación de las variables, para verificar que existiese una correlación fuerte entre las variables. Esto debido a que como explico anteriormente el A.C.P. transforma un conjunto de variables correlacionadas en un nuevo conjunto de variables no correlacionadas, pues si las variables casi no estaban correlacionadas no hubiese tenido sentido hacer el análisis mediante este método. La matriz de correlación para los datos del año 1999 es:

Tabla 3.4 Matriz de correlaciones

	HOM	MUJ	PEAH	PEAM	PEIH	PEIM	CONEX T	HOS	CAM	QUIROF	LAB
HOM	1.000										
MUJ	0.999	1.000									
PEAH	0.827	0.829	1.000								
PEAM	0.820	0.821	0.981	1.000							
PEIH	0.733	0.735	0.789	0.853	1.000						
PEIM	0.634	0.639	0.673	0.677	0.897	1.000					
CONEXT	0.660	0.661	0.508	0.513	0.591	0.601	1.000				
HOS	0.673	0.689	0.692	0.717	0.778	0.747	0.673	1.000			
CAM	0.616	0.636	0.729	0.751	0.694	0.613	0.405	0.883	1.000		
QUIROF	0.541	0.563	0.674	0.699	0.672	0.612	0.407	0.893	0.990	1.000	
LAB	0.606	0.623	0.708	0.726	0.682	0.621	0.492	0.915	0.977	0.975	1.000

Fuente: cálculos propios



### CAPÍTULO III

En la matriz de correlación se observa que muchas de las variables están muy correlacionadas como son (HOM,MUJ), (HOM,PEAH), (HOM,PEAM), (MUJ,PEAH), (MUJ,PEAM), (PEAH,PEAM) (PEAM,PEAIH), (PEIH,PEIM), (HOS,CAM), (HOS,QUIROF), (CAM,QUIROF), (CAM,LAB), (QUIROF,LAB) los cuales presentan un coeficiente de correlación mayor a 0.800, esto indica que a través de esas relaciones tal vez se están midiendo variables que se pueden obtener a través de otras por medio de una relación específica. Por lo anterior, se verifica que sí existe una correlación fuerte entre algunas de las variables, entonces es viable la aplicación del A.C.P.

El siguiente paso fue la obtención de los valores propios asociados o eigenvalores, los cuales se muestran en la tabla 3.5, así como ver la asimilación de varianza para cada uno de éstos y la varianza acumulada que tiene cada uno de los componentes principales.

En la tabla 3.5 se puede apreciar el hecho de que es suficiente considerar hasta el segundo componente principal, ya que con ello se tiene poco más de un 85% de la explicación de la variación total.

**Tabla 3.5 Valores propios**

Componente	Eigenvalores	% Varianza Total	Acumul. Eigenval	% Acumul.
1	8.153	74.119	8.153	74.119
2	1.209	10.991	9.362	85.110
3	0.697	6.340	10.060	91.451
4	0.544	4.946	10.604	96.397
5	0.227	2.065	10.831	98.462
6	0.089	0.811	10.920	99.273
7	0.054	0.493	10.974	99.765
8	0.021	0.191	10.995	99.956
9	0.003	0.027	10.998	99.983
10	0.001	0.010	10.999	99.993
11	0.001	0.007	11.000	100.000

Fuente: cálculos propios

Ahora los coeficientes o eigenvectores de los primeros cinco componentes para las 12 variables son:

Tabla 3.6 Coeficientes de los Componentes Principales o eigenvectores

Variable	Comp. Princ. 1	Comp. Princ. 2	Comp. Princ. 3	Comp. Princ. 4	Comp. Princ. 5
HOM	0.856	-0.405	-0.157	-0.205	0.189
MUJ	0.865	-0.380	-0.151	-0.212	0.196
PEAH	0.894	-0.163	-0.331	0.075	-0.204
PEAM	0.910	-0.137	-0.307	0.113	-0.209
PEIH	0.891	-0.137	0.085	0.370	0.010
PEIM	0.813	-0.123	0.298	0.435	0.137
CONEXT	0.675	-0.364	0.530	-0.283	-0.215
HOS	0.916	0.228	0.246	-0.082	0.019
CAM	0.884	0.442	-0.088	-0.076	0.049
QUIROF	0.855	0.507	-0.004	-0.055	0.017
LAB	0.885	0.429	0.021	-0.127	-0.012

Fuente: cálculos propios

La calificación del  $j$ -ésimo componente principal para la  $r$ -ésima observación se calcula por medio de la ecuación (3.1.1). Para ilustrar como se calcula se hará para el caso 1, es decir, para el estado de Aguascalientes, cuya información de las 11 variables, se encuentran en las tabla 3.1 y en la 3.3.

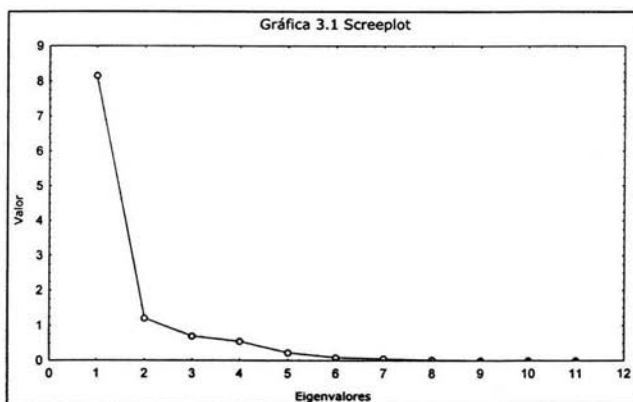
Entonces

$$\begin{aligned}
 y_{11} &= 0.8722(243-711.844) + 0.8812(169-546.250) + 0.8994(347079-958182.95) \\
 &\quad + 0.9136(303313-808666.641) + 0.8835(154925.5-541995.484) \\
 &\quad + 0.7996(175871-723922.359) + 0.6595(106-541.188) + 0.9039(10-30.938) \\
 &\quad + 0.8825(799-2416.063) + 0.8466(30-85.094) + 0.8801(13-56.031) \\
 &= -1799043.171
 \end{aligned}$$

En la tabla 3.7 se muestran los puntajes para los 32 casos y para los primeros cinco componentes principales, calculados como en el ejemplo anterior.

### CAPÍTULO III

Hasta ahora el análisis se ha hecho con cinco componentes principales, por lo cual es siguiente paso que interesa es determinar con cuantos componentes hay que quedarse, una opción es la de analizar la gráfica 3.1 llamada "Screeplot", en ella se puede apreciar que la pendiente que definen los segmentos de recta se estabilizan a partir del segundo valor propio, lo que llevo a decidir quedarse con los dos primeros componentes principales.



Por lo tanto la matriz de coeficientes principales o eigenvectores se muestra en la tabla 3.8:

**Tabla 3.8 Matriz de eigenvectores**

Variable	Comp. Princ. 1	Comp. Princ. 2
HOM	0.856	-0.405
YMUJ	0.865	-0.380
PEAH	0.894	-0.163
PEAM	0.910	-0.137
PEIH	0.891	-0.137
PEIM	0.813	-0.123
CONEXT	0.675	-0.364
HOS	0.916	0.228
CAM	0.884	0.442
QUIROF	0.855	0.507
LAB	0.885	0.429

Fuente: cálculos propios

Tabla 3.7 Valores de los primeros cinco componentes principales aplicados en la muestra

	Caso	Comp. Princ. 1	Comp. Princ. 2	Comp. Princ. 3	Comp. Princ. 4	Comp. Princ. 5
1	Ags.	-1799043.171	288637.778	161078.748	-484065.668	150908.944
2	B.C.	-348556.770	52722.768	-80926.704	-161797.028	-32004.646
3	B.C.S.	-1363276.645	216990.864	162669.053	-346383.629	138496.098
4	Cam.	-2029344.400	325683.157	198798.971	-535493.210	179616.551
5	Coah.	928035.927	-155643.439	-320545.133	103206.889	-207237.968
6	Col.	177880.900	-33012.784	-153541.355	-35346.676	-90462.241
7	Chis.	-515818.589	79750.090	-60849.904	-204910.558	-15442.976
8	Chih.	-2159598.847	347785.780	219475.365	-566262.538	195754.860
9	D.F.	5791092.955	-931123.805	-1007967.974	1256353.223	-754060.991
10	Dgo.	-1320937.660	210869.260	81684.605	-378113.576	90696.736
11	Gto.	4243366.258	-693239.575	-659190.398	960135.078	-504960.120
12	Gro.	640161.626	-106812.213	-172179.565	99176.694	-116279.841
13	Hgo.	-415770.804	62908.166	-86311.896	-190198.744	-32368.800
14	Jal.	2054148.393	-333108.244	-546928.679	335203.005	-373788.822
15	Edo.Méx.	7596305.467	-1235895.753	-1598676.863	1489811.063	-1143667.615
16	Mich.	1006531.711	-165534.494	-286694.472	150085.081	-192956.816
17	Mor.	-1191258.549	190128.158	63154.654	-347128.405	75836.094
18	Nay.	-1792378.467	287326.053	154415.538	-486002.870	146857.785
19	N.L.	-150398.345	19891.858	70905.958	-17477.853	49553.141
20	Oax.	-1466639.054	233590.619	212623.520	-352271.972	170305.124
21	Pue.	388812.286	-67452.640	83482.534	160887.433	39141.364
22	Qro.	-1825833.889	292744.313	239992.123	-448149.648	196632.842
23	Q.R.	-2146643.841	344019.001	255326.000	-541882.727	216010.253
24	S.L.P.	-649312.048	113627.460	379756.313	28319.554	228356.359
25	Sin.	-403156.103	74157.778	334344.096	82175.977	194468.198
26	Son.	-666812.006	115315.282	316717.897	-15506.523	194680.569
27	Tabs.	-984982.714	165147.143	341350.886	-105417.029	219662.492
28	Tamps.	-265312.383	53316.474	355525.194	140363.432	201216.053
29	Tlaxc.	-1794308.594	292051.383	322110.918	-381197.909	238134.566
30	Ver.	3479759.421	-535675.170	428736.068	1406059.794	107672.685
31	Yuc.	-1194862.506	198545.780	333374.923	-177954.782	222633.014
32	Zac.	-1821849.557	292288.953	258289.579	-436215.877	206597.108
	Suma	0.000	0.000	0.000	0.000	0.000

Fuente: cálculos propios

Y las dos ecuaciones de los componentes principales son:

$$C_1 = 0.856HOM + 0.865MUJ + 0.894PEAH + 0.910PEAM + 0.891PEIH + 0.813PEIM \\ + 0.675CONEXT + 0.916HOS + 0.887CAM + 0.855QUIROFA + 0.885LAB$$

$$C_2 = -0.405HOM - 0.380MUJ - 0.163PEAH - 0.137PEAM - 0.137PEIH - 0.123PEIM \\ - 0.364CONEXT + 0.228HOS + 0.442CAM + 0.507QUIROFA + 0.429LAB$$

Por último se gravitaron las dos ecuaciones de los componentes principales, tomando a  $C_1$  en el eje X y a  $C_2$  en el eje Y y con ello se determinó de que manera se agrupan las 32 entidades federativas, para ello fue necesario multiplicar la matriz de datos que se encuentra en la tabla 3.1 por la matriz de la tabla 3.8, los resultados se muestran en la tabla 3.9. Para tener una mejor apreciación de la gráfica se reetiquetaron los casos, en donde se les asoció el número 1 a Aguascalientes, el 2 a Baja California y así sucesivamente. (Gráfica 3.2)

En la gráfica 3.2 se ve la formación de 2 grupos, en el primero se agrupan la mayoría de las entidades federativas y en el otro grupo contiene a las que están dispersas.

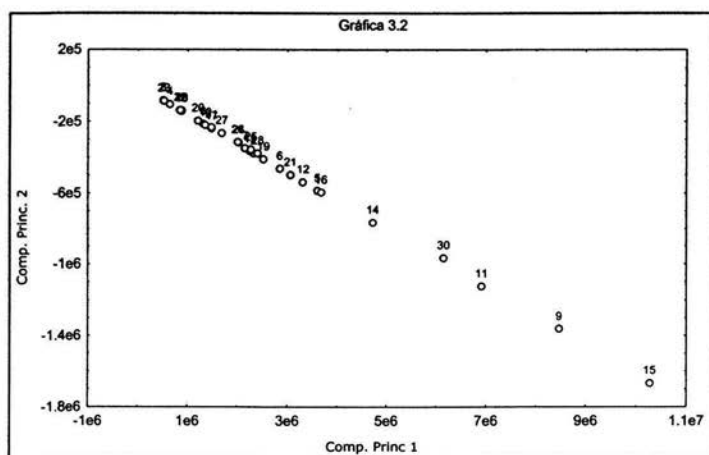


Tabla 3.9

	Caso	Comp. Princ. 1 (X)	Comp. Princ.. 2 (Y)
1	Ags.	868338.560	-140650.547
2	B.C.	2318824.961	-376565.558
3	B.C.S.	1304105.086	-212297.461
4	Cam.	638037.331	-103605.168
5	Coah.	3595417.658	-584931.764
6	Col.	2845262.631	-462301.110
7	Chis.	2151563.142	-349538.236
8	Chih.	507782.884	-81502.545
9	D.F.	8458474.686	-1360412.131
10	Dgo.	1346444.071	-218419.065
11	Gto.	6910747.989	-1122527.901
12	Gro.	3307543.357	-536100.538
13	Hgo.	2251610.927	-366380.159
14	Jal.	4721530.124	-762396.569
15	Edo.Méx.	10263687.198	-1665184.079
16	Mich.	3673913.442	-594822.820
17	Mor.	1476123.182	-239160.167
18	Nay.	875003.264	-141962.272
19	N.L.	2516983.386	-409396.467
20	Oax.	1200742.677	-195697.706
21	Pue.	3056194.017	-496740.966
22	Qro.	841547.842	-136544.012
23	Q.R.	520737.890	-85269.325
24	S.L.P.	2018069.683	-315660.866
25	Sin.	2264225.628	-355130.548
26	Son.	2000569.725	-313973.043
27	Tab.	1682399.016	-264141.183
28	Tamps.	2402069.348	-375971.851
29	Tlaxc.	873073.137	-137236.942
30	Ver.	6147141.152	-964963.496
31	Yuc.	1472519.225	-230742.546
32	Zac.	845532.173	-136999.372
Fuente: cálculos propios			

Finalmente los grupos que arrojo el análisis por componentes principales se muestran en la tabla 3.10.

**Tabla 3.10 Clasificación por grupos**

Grupo	Entidad Federativa
1	Ags., B.C., B.C.S., Camp., Coah., Col., Chis., Chih., Dgo., Gro., Hgo., Mich., Mor., Nay., N.L.Oax., Pue., Qro., Q.R., S.L.P., Sin., Son., Tabs., Tamps., Tlax., Yuc., Zac.
2	D.F. Edo. de Méx., Gto., Jal., Ver.

### 3.2.2 Análisis Discriminante

El análisis de discriminante se utilizo para comprobar que la formación de los grupos hecha mediante los componentes principales es correcta, por ello para la realización de este análisis se utilizo la tabla 3.10. Antes que nada es importante hacer el supuesto de normalidad de la muestra, así como considerar que las matrices de covarianza para las 12 variables son iguales y realizar la clasificación por individuo con base en las funciones de discriminante lineal.

En la tabla 3.11 se observan las medias de las 11 variables por grupo, en donde se ve una diferencia significativa en cada una de ellas.

**Tabla 3.11 Medias por grupo**

Grupo	HOM	MUJ	PEAH	PEAM	PEIH	PEIM	CONEXT	HOS	CAM	QUIROF	LAB
1	456.96	352.59	645627.44	509599.06	368442.06	538687.56	465.33	24.48	1512.11	56.04	41.07
2	2088.20	1592.00	2645982.75	2423631.50	1479184.00	1724190.25	950.80	65.80	7297.40	242.00	136.80

Fuente: cálculos propios

La distancia cuadrada de Mahalanobis por grupo y los niveles de significancia para determinar si vale la pena realizar el análisis discriminante se muestran en la tablas 3.12 y 3.13 respectivamente.

Grupo	1	2
1	0.000	73.250
2	73.250	0.000

Fuente: cálculos propios

Grupo	1	2
1		1.942E-07
2	1.942E-07	

Fuente: cálculos propios

Los niveles de significancia para los dos grupos, mostrados en la tabla 3.13, son muy pequeños, por lo cual es conveniente la realización del análisis discriminante. Cuando los costos de los errores de clasificación de los grupos son iguales y cuando las matrices de covarianza también lo son, se puede calcular las funciones de clasificación para cada grupo, las cuales se muestran en la tabla 3.14, en dicha tabla se aprecia que aproximadamente el 84% está clasificado en el Grupo 1 y el 16% restante en el Grupo 2.

Tabla 3.14 Funciones de Clasificación

Variable	Grupo	
	1 p=.84375	2 p=.15625
HOM	0.003211	0.170596
MUJ	-0.005831	-0.213257
PEAH	0.000012	-0.000062
PEAM	-0.000013	0.000099
PEIH	0.000012	-0.000077
PEIM	-0.000005	0.000056
CONEXT	0.004155	0.018761
HOS	0.033155	-0.262603
CAM	-0.001256	-0.004245
QUIROF	0.04932	0.685787
LAB	-0.021007	-0.890569

Fuente: cálculos propios

Tabla 3.15 Matriz de clasificación

Grupo	% Correcto	1 p=.84375	2 p=.15625
1	100	27	0
2	100	0	5
Total	100	27	5

Fuente: cálculos propios

El porcentaje de clasificación y el número de casos clasificados correctamente se muestran en las tablas 3.15 y 3.16. En la tabla 3.15 se puede apreciar que la clasificación hecha por el análisis de componentes principales es muy buena, pues, en ambos grupos el error de clasificación es cero, por la cual el porcentaje de clasificación correcta es del 100%, lo anterior se puede corroborar en la tabla 3.16. La matriz de distancias cuadradas de Mahalanobis para cada caso se muestra en tabla 3.17.



Tabla 3.16 Clasificación de los grupos

Clasificaciones incorrectas marcadas con *			
	Observac	G_1:1	G_2:2
	Clasif.	p=.84375	p=.15625
Ags.	G_1:1	G_1:1	G_2:2
B.C.	G_1:1	G_1:1	G_2:2
B.C.S	G_1:1	G_1:1	G_2:2
Cam.	G_1:1	G_1:1	G_2:2
Coah.	G_1:1	G_1:1	G_2:2
Col.	G_1:1	G_1:1	G_2:2
Chis.	G_1:1	G_1:1	G_2:2
Chih.	G_1:1	G_1:1	G_2:2
D.F.	G_2:2	G_2:2	G_1:1
Dgo.	G_1:1	G_1:1	G_2:2
Gto.	G_2:2	G_2:2	G_1:1
Gro.	G_1:1	G_1:1	G_2:2
Hgo.	G_1:1	G_1:1	G_2:2
Jal.	G_2:2	G_2:2	G_1:1
Edo.Mex.	G_2:2	G_2:2	G_1:1
Mich.	G_1:1	G_1:1	G_2:2
Mor.	G_1:1	G_1:1	G_2:2
Nay.	G_1:1	G_1:1	G_2:2
N.L.	G_1:1	G_1:1	G_2:2
Oax.	G_1:1	G_1:1	G_2:2
Pue.	G_1:1	G_1:1	G_2:2
Qro.	G_1:1	G_1:1	G_2:2
Q.R.	G_1:1	G_1:1	G_2:2
S.L.P.	G_1:1	G_1:1	G_2:2
Sin.	G_1:1	G_1:1	G_2:2
Son.	G_1:1	G_1:1	G_2:2
Tabas.	G_1:1	G_1:1	G_2:2
Tamps.	G_1:1	G_1:1	G_2:2
Tlaxc.	G_1:1	G_1:1	G_2:2
Ver.	G_2:2	G_2:2	G_1:1
Yuc.	G_1:1	G_1:1	G_2:2
Zac.	G_1:1	G_1:1	G_2:2
Grupo 1=G_1:1			
Grupo 2=G_2:2			
Fuente: cálculos propios			

Tabla 3.17 Distancias cuadradas de Mahalanobis

	Observac	G_1:1	G_2:2
	Clasif	p=.84375	p=.15625
Ags.	G_1:1	5.579	59.858
B.C.	G_1:1	6.638	62.278
B.C.S	G_1:1	3.599	78.92
Cam.	G_1:1	3.231	87.275
Coah.	G_1:1	13.252	80.42
Col.	G_1:1	6.272	77.27
Chis.	G_1:1	11.913	94.985
Chih.	G_1:1	11.41	91.11
D.F.	G_2:2	107.09	22.758
Dgo.	G_1:1	4.966	99.843
Gto.	G_2:2	77.622	17.234
Gro.	G_1:1	17.919	52.805
Hgo.	G_1:1	11.118	96.932
Jal.	G_2:2	79.02	18.559
Edo.Mex.	G_2:2	105.65	22.478
Mich.	G_1:1	15.995	60.518
Mor.	G_1:1	2.489	66.278
Nay.	G_1:1	1.8	83.889
N.L.	G_1:1	17.726	94.964
Oax.	G_1:1	14.849	77.768
Pue.	G_1:1	23.305	74.328
Qro.	G_1:1	3.678	80.262
Q.R.	G_1:1	7.127	108.25
S.L.P.	G_1:1	7.187	61.473
Sin.	G_1:1	10.642	39.911
Son.	G_1:1	7.309	72.564
Tabas.	G_1:1	8.094	90.6
Tamps.	G_1:1	5.039	71.992
Tlaxc.	G_1:1	3.344	73.226
Ver.	G_2:2	72.591	17.585
Yuc.	G_1:1	3.452	68.798
Zac.	G_1:1	3.451	79.002
Grupo 1=G_1:1			
Grupo 2=G_2:2			
Fuente: cálculos propios			

Para ejemplificar el análisis de la tabla 3.18 se tomó el caso 1, es decir Aguascalientes, entonces la probabilidad de que pertenezca al grupo 1 es de 1 y de que pertenezca al 2 es de  $8.83928E-14$  y como  $1 > 8.83928E-14$ , entonces con certeza debe de pertenecer al grupo 1. Para los demás casos el análisis es similar.

Tabla 3.18. Probabilidades posteriores

	Observac Clasif.	G_1:1 p=.84375	G_2:2 p=.15625
Ags.	G_1:1	1	3.027E-13
B.C.	G_1:1	1	1.533E-13
B.C.S	G_1:1	1	8.166E-18
Cam.	G_1:1	1	1.041E-19
Coah.	G_1:1	1	4.812E-16
Col.	G_1:1	1	7.092E-17
Chis.	G_1:1	1	1.693E-19
Chih.	G_1:1	1	9.137E-19
D.F.	G_2:2	2.632E-18	1
Dgo.	G_1:1	1	4.628E-22
Gto.	G_2:2	4.162E-13	1
Gro.	G_1:1	1	4.922E-09
Hgo.	G_1:1	1	4.298E-20
Jal.	G_2:2	4.013E-13	1
Edo.Mex.	G_2:2	4.693E-18	1
Mich.	G_1:1	1	3.979E-11
Mor.	G_1:1	1	2.606E-15
Nay.	G_1:1	1	2.769E-19
N.L.	G_1:1	1	3.131E-18
Oax.	G_1:1	1	4.027E-15
Pue.	G_1:1	1	1.542E-12
Qro.	G_1:1	1	4.34E-18
Q.R.	G_1:1	1	2.037E-23
S.L.P.	G_1:1	1	3.017E-13
Sin.	G_1:1	0.9999999	8.166E-08
Son.	G_1:1	1	1.253E-15
Tab.	G_1:1	1	2.247E-19
Tamps.	G_1:1	1	5.358E-16
Tlaxc.	G_1:1	1	1.239E-16
Ver.	G_2:2	6.138E-12	1
Yuc.	G_1:1	1	1.197E-15
Zac.	G_1:1	1	7.277E-18

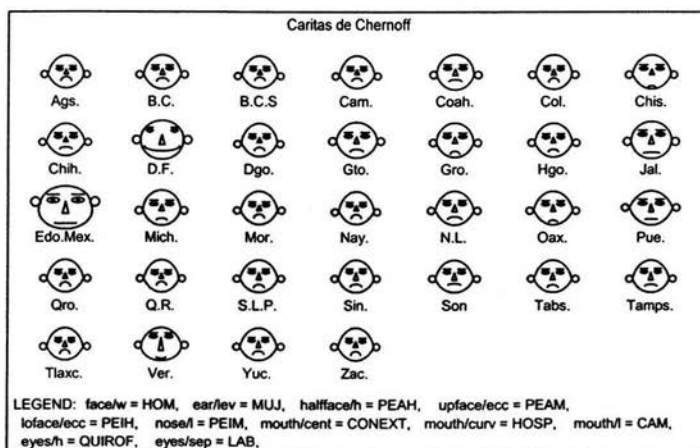
Grupo 1=G\_1:1 Grupo 2=G\_2:2  
Fuente: cálculos propios

Por todo lo anterior la clasificación hecha por el análisis discriminante es la misma que por componentes principales, entonces la forma en se agrupan las entidades federativas es la que se muestra en la tabla 3.10.

### 3.3 ANÁLISIS ESTADÍSTICO DEL AÑO 2000

#### 3.3.1 Análisis de Componentes Principales

Primero se vio como era cada entidad federativa de acuerdo con las variables de la tabla 3.2, y con ello observar si algunas de ellas se parecen con la finalidad de poder obtener grupos con características similares; para este propósito se utilizaron las caritas de Chernoff



Se puede observar que es muy factible la formación de grupos, por ejemplo la variable PEIM, que en las caritas es denominada *nose/l*, es decir, el largo de la nariz, se aprecia que es muy similar en casi todas las entidades federativas; también Chiapas, Guerrero y Oaxaca se parecen mucho en la variable HOSP, que es *mouth/curv* o lo curvo de la boca y en LAB que es *eyes/sep* o la separación de los ojos. En cuanto a MUJ, que es *ear/lev* o altura de las orejas, se ve como que todas las entidades federativas se comportan de manera similar. Se podría dar muchísimos más ejemplos, pero lo que queda claro es que a simple vista no se

puede realizar una clasificación, por ello se procederá a realizar un análisis utilizando componentes principales, similar al de la sección 3.1.1, buscando estimar la ecuación (3.1.1).

Las estadísticas descriptivas de cada una de las variables para este año se pueden observar en la tabla 3.19. Al hacer una comparación de dicha tabla y la del año 1999 (tabla 3.2), se observa que las medias son muy parecidas, así como sus desviaciones estándares.

**Tabla 3.19 Estadísticas Descriptivas**

	<b>Medias</b>	<b>Desv. Esta.</b>
<b>HOM</b>	681.500	796.657
<b>MUJ</b>	524.438	622.778
<b>PEAH</b>	1098209.125	1335568.252
<b>PEAM</b>	1200738.375	1405937.818
<b>PEIH</b>	862906.031	851843.638
<b>PEIM</b>	735461.250	706267.842
<b>CONEXT</b>	565.938	385.997
<b>HOS</b>	31.156	22.167
<b>CAM</b>	2410.750	3014.996
<b>QUIROF</b>	86.250	102.367
<b>LAB</b>	56.813	53.904

Fuente: cálculos propios

Como se menciona antes, para el análisis de componentes principales es necesario ver si las variables están muy correlacionadas o no y con ello decidir si es factible la realización de dicho análisis. La matriz de correlaciones para este año se muestra en la tabla 3.20. En la matriz se ve que (MUJ,HOM) tienen una correlación muy fuerte igual a 1, esto quiere decir que las dos variables se explican una a la otra, también se aprecia que algunas de las otras variables están muy correlacionadas, con un nivel de correlación mayor a 0.80, como son, (PEAH,HOM), (PEAH,MUJ), (PEAM,MUJ), (PEAM,PEAH), (PEIH,PEAM), (PEIM,PEIH), (CAMAS,HOSP), (QUIROF,HOSP), (QUIROF,CAMAS), (LAB,HOSP), (LAB,CAMAS), (LAB,QUIROF), a parte de las correlaciones anteriores que son lo cual significa que es factible la aplicación de componentes principales.

Tabla 3.20 Matriz de correlaciones

	HOM	MUJ	PEAH	PEAM	PEIH	PEIM	CONEXT	HOS	CAM	QUIROF	LAB
HOM	1.000										
MUJ	0.999	1.000									
PEAH	0.835	0.828	1.000								
PEAM	0.820	0.811	0.970	1.000							
PEIH	0.709	0.704	0.661	0.800	1.000						
PEIM	0.538	0.540	0.289	0.420	0.841	1.000					
CONEXT	0.605	0.599	0.432	0.460	0.551	0.571	1.000				
HOS	0.683	0.680	0.618	0.670	0.738	0.688	0.644	1.000			
CAM	0.620	0.614	0.697	0.723	0.641	0.494	0.364	0.885	1.000		
QUIROF	0.548	0.542	0.631	0.667	0.616	0.496	0.374	0.892	0.991	1.000	
LAB	0.608	0.599	0.680	0.711	0.636	0.489	0.473	0.910	0.974	0.973	1.000

Fuente: cálculos propios

Tabla 3.21 Valores propios

Componente	Eigenval	% Varianza total	Acumul. Eigenval	% Acumul.
1	7.716	70.143	7.716	70.143
2	1.243	11.303	8.959	81.446
3	1.084	9.859	10.044	91.305
4	0.548	4.981	10.592	96.286
5	0.299	2.722	10.891	99.009
6	0.053	0.479	10.944	99.487
7	0.034	0.309	10.978	99.796
8	0.018	0.163	10.996	99.959
9	0.003	0.023	10.998	99.982
10	0.001	0.013	10.999	99.995
11	0.001	0.005	11.000	100.000

Fuente: cálculos propios

Después se procedió a la obtención de los valores propios asociados o eigenvalores, los cuales se muestran en la tabla anterior, en dicha tabla se puede observar que tomando hasta el segundo componente principal es suficiente, pues con ello se tendría poco más del 81% de explicación de la variación total.

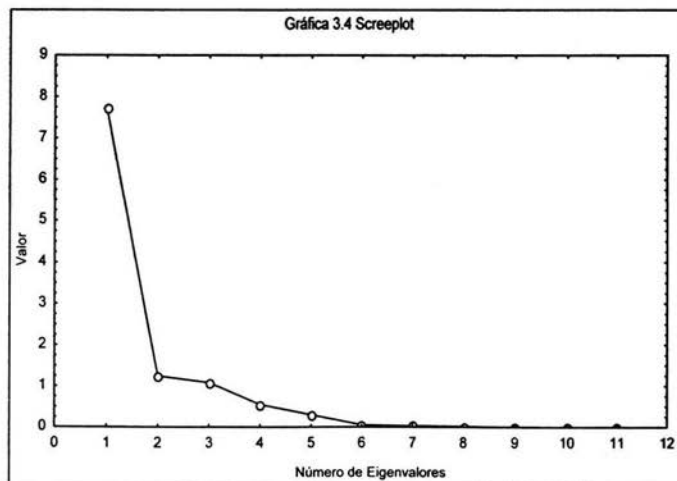
En la tabla 3.22 se muestran los coeficientes o eigenvectores de los primeros cinco componentes para todas las variables.

Tabla 3.22 Coeficientes de los Componentes Principales o eigenvectores

Variable	Comp. Princ. 1	Comp. Princ. 2	Comp. Princ. 3	Comp. Princ. 4	Comp. Princ. 5
HOM	0.867	-0.382	0.185	0.060	0.249
MUJ	0.862	-0.387	0.182	0.056	0.265
PEAH	0.842	-0.147	0.485	-0.027	-0.146
PEAM	0.884	-0.140	0.349	-0.153	-0.228
PEIH	0.855	-0.193	-0.232	-0.372	-0.176
PEIM	0.678	-0.169	-0.641	-0.284	0.092
CONEXT	0.643	-0.363	-0.371	0.517	-0.216
HOS	0.916	0.222	-0.231	0.135	0.045
CAM	0.882	0.455	0.061	-0.008	0.069
QUIROF	0.851	0.515	-0.012	0.023	0.028
LAB	0.885	0.433	0.006	0.113	-0.020

Fuente: cálculos propios

Como ya se menciono antes la calificación de la  $j$ -ésima componente principal para el  $r$ -ésimo individuo se calcula por medio de la ecuación (3.1.1) y el procedimiento de cómo se saca se ilustra en la sección 3.1.1. En la tabla 3.22 se muestran los puntajes.



La gráfica 3.4 "Screeplot" muestra que a partir de la cuarta componente se estabiliza la pendiente, pero si se regresa a la tabla 3.20 se puede apreciar que con toman los cuatros componentes se acumula poco más del 96%, por lo cual no existe mucha diferencia si se toman sólo los dos primeros, además de que es más fácil realizar el análisis.

Tabla 3.23 Valores de los primeros cinco componentes principales aplicados en la muestra

Caso	Comp. Princ. 1	Comp. Princ. 2	Comp. Princ. 3	Comp. Princ. 4	Comp. Princ. 5	
1	Ags.	-2125873.169	418882.178	-20816.708	534145.030	310915.977
2	B.C.	-216335.532	63308.655	354069.428	151214.999	-34989.111
3	B.C.S.	-2714601.007	525395.934	-174031.193	641382.820	424437.321
4	Cam.	-2415969.677	471725.106	-84613.149	587922.876	362154.737
5	Coah.	1512449.396	-253584.255	754300.277	-183970.919	-366119.728
6	Col.	390209.253	-45781.913	528272.125	42421.740	-158377.723
7	Chis.	-496385.946	117613.144	335793.524	215432.622	5322.036
8	Chih.	-2571590.179	499950.505	-144004.170	613977.600	398561.725
9	D.F.	7189129.680	-1301980.742	1753846.455	-1342175.028	-1364808.062
10	Dgo.	-1532089.662	309650.820	113343.558	418031.881	197492.993
11	Gto.	2178020.024	-366464.022	992327.477	-284522.080	-505770.753
12	Gro.	386973.861	-43125.207	532058.759	47419.760	-156736.630
13	Hgo.	-684651.808	158073.033	363562.817	274504.644	31177.575
14	Jal.	4407177.103	-790083.429	1278570.553	-772966.951	-881002.199
15	Edo.Méx.	12411136.250	-2259833.388	3172836.895	-2297933.757	-2401726.477
16	Mich.	1448052.947	-237339.276	779324.071	-162734.313	-365675.443
17	Mor.	-1389526.518	282471.600	125396.327	385628.867	175156.464
18	Nay.	-2130864.942	417922.642	-48370.313	526736.336	315367.678
19	N.L.	-898214.824	161276.132	-231355.120	243515.444	263474.986
20	Oax.	-1278927.068	231525.043	-296337.051	315925.354	322457.789
21	Pue.	-356947.420	49441.832	-297379.655	108030.659	220167.750
22	Qro.	-2435977.018	464126.245	-268408.760	569560.795	429891.440
23	Q.R.	-2725847.319	521220.971	-267545.367	637875.791	465548.407
24	S.L.P.	-807633.041	97476.503	-1014871.936	-98543.000	277259.181
25	Sin.	-550573.935	43027.195	-1064817.325	-186989.176	239435.297
26	Son.	-866493.348	112890.984	-941944.729	-68348.577	267811.396
27	Tabs.	-1202205.225	184665.988	-864712.797	42788.139	306116.508
28	Tamps.	-375712.231	2419.363	-1159204.887	-261932.250	226972.548
29	Tlaxc.	-2212807.302	400935.539	-583871.231	389648.998	415074.004
30	Ver.	4029350.867	-947381.541	-2476679.793	-1785861.400	-214498.604
31	Yuc.	-1504723.222	247109.074	-812564.983	138065.574	345436.090
32	Zac.	-2458548.990	464465.288	-332173.096	561747.521	449472.827
	Suma	0.000	0.000	0.000	0.000	0.000

Por lo tanto las dos ecuaciones de los componentes principales son:

$$C_1 = 0.867HOM + 0.862MUJ + 0.842PEAH + 0.884PEAM + 0.855PEIH + 0.678PEIM \\ + 0.643CONEXT + 0.916HOS + 0.882CAM + 0.851QUIROFA + 0.885LAB$$

$$C_2 = -0.382HOM - 0.387MUJ - 0.147PEAH - 0.140PEAM - 0.193PEIH - 0.169PEIM \\ - 0.363CONEXT + 0.222HOS + 0.455CAM + 0.515QUIROFA + 0.433LAB$$

Y la matriz de coeficientes principales o eigenvectores queda de la siguiente manera:

Tabla 3.24 Matriz de eigenvectores

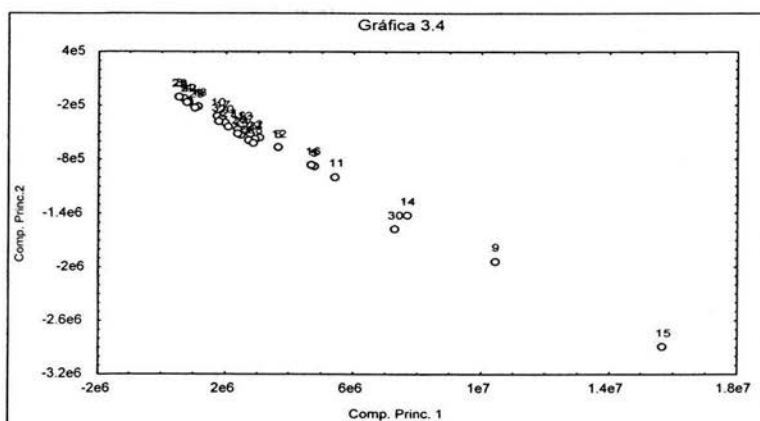
Variable	Comp. Princ. 1	Comp. Princ. 2
HOM	0.867	-0.382
MUJ	0.862	-0.387
PEAH	0.842	-0.147
PEAM	0.884	-0.140
PEIH	0.855	-0.193
PEIM	0.678	-0.169
CONEXT	0.643	-0.363
HOS	0.916	0.222
CAM	0.882	0.455
QUIROF	0.851	0.515
LAB	0.885	0.433

Fuente: cálculos propios

Como último paso se graficaron las dos ecuaciones de los componentes principales siendo esta la gráfica 3.4, en ella se tomo a  $C_1$  en el eje X y a  $C_2$  en el eje Y. Con lo anterior se determino de que manera agrupa el análisis de componentes principales a las entidades federativas. La manera de realizar los cálculos es la misma que en la sección 3.1.1, así como la reetiquetación de los casos.

Ahora analizando las gráfica 3.4 se aprecia la formación de dos grupos, en el primer grupo están contenidas la mayoría de las entidades federativas y en el segundo están contenidas los estados más dispersos, que son solo 4 de los 32 estados.





Por lo tanto la formación de grupos que arroja el análisis por componentes principales es:

**Tabla 3.25 Clasificación por grupos**

Grupo	Entidad Federativa
1	Ags., B.C., B.C.S., Camp., Coah., Col., Chis., Chih., Dgo., Gro., Gto., Hgo., Mich., Mor., Nay., N.L.Oax., Pue., Qro., Q.R., S.L.P., Sin., Son., Tabs., Tamps., Tlax., Yuc., Zac.
2	D.F. Edo. de Méx. , Jal., Ver.

### 3.3.2 Análisis Discriminante

Al igual que en la sección 3.1.2 también se va a hacer el supuesto de normalidad de la muestra y se considerara que las matrices de covarianza para las 12 variables son iguales.

Las medias por grupo para las 12 variables son, se observa que existe una diferencia grande entre las medias de los dos grupos:

Tabla 3.26 Medias por grupo

Grupo	HOM	MUJ	PEAH	PEAM	PEIH	PEIM	CONEXT	HOS	CAM	QUIROF	LAB
1	476.29	364.93	752993.75	793117.19	596529.06	556482.81	490.39	24.86	1543.46	57.32	42.04
2	2118.00	1641.00	3514716.75	4054086.75	2727545.00	1988310.50	1094.75	75.25	8481.75	288.75	160.25

Fuente: cálculos propios

La distancia cuadrada de Mahalanobis para los grupos y los niveles de significancia se muestran a continuación:

Tabla 3.27 Distancia cuadrada de Mahalanobis

Grupo	1	2
1	0	57.84
2	57.84	0

Fuente: cálculos propios

Tabla 3.28 p-levels

Grupo	1	2
1		1.02E-05
2	1.02E-05	

Fuente: cálculos propios

Tabla 3.29  
Funciones de Clasificación

Variable	Grupo	
	1 p=.87500	2 p=.12500
HOM	0.00473	0.012864
MUJ	-0.007654	-0.013696
PEAH	0.000005	0.000011
PEAM	-0.000004	-0.000017
PEIH	0.000004	0.000034
PEIM	0	-0.000012
CONEXT	0.001462	0.021968
HOSP	0.084522	-0.359758
CAM	-0.001632	0.002243
QUIROF	0.040635	0.293752
LAB	0.003354	-0.423094

Fuente: cálculos propios

Si se compara la tabla 3.27 con la tabla 3.11 de la sección 3.1.2, se ve que la distancia entre los grupos para el año 2000 es un poco más pequeña que la del año 1999. En la 3.28 se aprecia que el *p-level* es muy pequeño, por lo cual se puede continuar con el análisis

### CAPÍTULO III

---

En la sección 3.1.2 se mencionó que si los errores de clasificación de los grupos son iguales, así como las matrices de covarianza, entonces se pueden calcular las funciones de clasificación por grupo y éstas se ven en la tabla 3.29. Dicha tabla dice que aproximadamente el 87.5% de la muestra está clasificado en el grupo 1 y el resto que es el 12.5% en el 2.

El porcentaje de clasificación y el número de casos clasificados correctamente están en la tablas 3.30 y 3.31 respectivamente. En la primer tabla se ve que la clasificación hecha por el análisis de componentes principales es muy buena, pues en los dos grupos el error de clasificación es cero, por lo cual el porcentaje de clasificación correcta es del 100%, lo anterior se corrobora en la tabla 3.30, donde no existen datos mal clasificados.

**Tabla 3.30 Matriz de clasificación**

<b>Grupo</b>	<b>% Correcto</b>	<b>1 p=.84375</b>	<b>2 p=.15625</b>
<b>1</b>	100	28	0
<b>2</b>	100	0	4
<b>Total</b>	100	28	4

Fuente: cálculos propios

La matriz de distancias cuadradas de Mahalanobis para cada caso se muestra en la tabla 3.32.

Las probabilidades posteriores para cada caso se muestran en la tabla 3.33, su análisis se hace de la misma manera que se explico en la sección 3.1.2.

Por todo lo anterior la clasificación por análisis discriminante para el año 2000 es la misma que el hecho por componentes principales, por lo tanto la clasificación final para este año es la dada en la tabla 3.25.

Tabla 3.31 Clasificación de los grupos

Clasificaciones incorrectas marcadas con *			
	Observac	G_1:1	G_2:2
	Clasif.	p=.87500	p=.12500
Ags.	G_1:1	G_1:1	G_2:2
B.C.	G_1:1	G_1:1	G_2:2
B.C.S	G_1:1	G_1:1	G_2:2
Cam.	G_1:1	G_1:1	G_2:2
Coah.	G_1:1	G_1:1	G_2:2
Col.	G_1:1	G_1:1	G_2:2
Chis.	G_1:1	G_1:1	G_2:2
Chih.	G_1:1	G_1:1	G_2:2
D.F.	G_2:2	G_2:2	G_1:1
Dgo.	G_1:1	G_1:1	G_2:2
Gto.	G_1:1	G_1:1	G_2:2
Gro.	G_1:1	G_1:1	G_2:2
Hgo.	G_1:1	G_1:1	G_2:2
Jal.	G_2:2	G_2:2	G_1:1
Edo.Mex.	G_2:2	G_2:2	G_1:1
Mich.	G_1:1	G_1:1	G_2:2
Mor.	G_1:1	G_1:1	G_2:2
Nay.	G_1:1	G_1:1	G_2:2
N.L.	G_1:1	G_1:1	G_2:2
Oax.	G_1:1	G_1:1	G_2:2
Pue.	G_1:1	G_1:1	G_2:2
Qro.	G_1:1	G_1:1	G_2:2
Q.R.	G_1:1	G_1:1	G_2:2
S.L.P.	G_1:1	G_1:1	G_2:2
Sin.	G_1:1	G_1:1	G_2:2
Son.	G_1:1	G_1:1	G_2:2
Tabs.	G_1:1	G_1:1	G_2:2
Tamps.	G_1:1	G_1:1	G_2:2
Tlaxc.	G_1:1	G_1:1	G_2:2
Ver.	G_2:2	G_2:2	G_1:1
Yuc.	G_1:1	G_1:1	G_2:2
Zac.	G_1:1	G_1:1	G_2:2

Fuente: cálculos propios

Tabla 3.32 Distancias cuadradas de Mahalanobis

	Observac	G_1:1	G_2:2
	Clasif	p=.87500	p=.12500
Ags.	G_1:1	2.928	63.926
B.C.	G_1:1	8.036	49.328
B.C.S	G_1:1	4.652	83.297
Cam.	G_1:1	3.965	77.699
Coah.	G_1:1	12.698	63.991
Col.	G_1:1	6.682	62.141
Chis.	G_1:1	19.99	56.805
Chih.	G_1:1	13.381	80.713
D.F.	G_2:2	92.038	21.07
Dgo.	G_1:1	4.481	73.767
Gto.	G_1:1	19.888	48.34
Gro.	G_1:1	11.141	62.24
Hgo.	G_1:1	12.687	76.027
Jal.	G_2:2	61.985	15.117
Edo.Mex.	G_2:2	82.086	21.634
Mich.	G_1:1	8.584	52.438
Mor.	G_1:1	1.761	56.598
Nay.	G_1:1	1.485	63.763
N.L.	G_1:1	15.525	63.328
Oax.	G_1:1	12.407	68.87
Pue.	G_1:1	23.564	71.33
Qro.	G_1:1	4.023	68.415
Q.R.	G_1:1	9.27	95.267
S.L.P.	G_1:1	8.527	46.829
Sin.	G_1:1	8.131	36.481
Son.	G_1:1	7.292	58.447
Tabs.	G_1:1	10.083	63.01
Tamps.	G_1:1	8.64	38.962
Tlaxc.	G_1:1	3.36	60.484
Ver.	G_2:2	57.247	18.635
Yuc.	G_1:1	5.012	52.501
Zac.	G_1:1	5.353	76.843

Grupo 1=G\_1:1  
Grupo 2=G\_2:2  
Fuente: cálculos propios

Tabla 3.33 Probabilidades posteriores

	Observac Clasif.	G_1:1 p=.87500	G_2:2 p=.12500
Ags.	G_1:1	1	8.115E-15
B.C.	G_1:1	1	1.543E-10
B.C.S	G_1:1	1	1.195E-18
Cam.	G_1:1	1	1.393E-17
Coah.	G_1:1	1	1.039E-12
Col.	G_1:1	1	1.294E-13
Chis.	G_1:1	1	1.447E-09
Chih.	G_1:1	1	3.418E-16
D.F.	G_2:2	2.72E-15	1
Dgo.	G_1:1	1	1.287E-16
Gto.	G_1:1	0.9999999	9.478E-08
Gro.	G_1:1	1	1.145E-12
Hgo.	G_1:1	1	2.516E-15
Jal.	G_2:2	4.654E-10	1
Edo.Mex.	G_2:2	5.226E-13	1
Mich.	G_1:1	1	4.287E-11
Mor.	G_1:1	1	1.766E-13
Nay.	G_1:1	1	4.279E-15
N.L.	G_1:1	1	5.951E-12
Oax.	G_1:1	1	7.839E-14
Pue.	G_1:1	1	6.061E-12
Qro.	G_1:1	1	1.487E-15
Q.R.	G_1:1	1	3.027E-20
S.L.P.	G_1:1	1	6.881E-10
Sin.	G_1:1	0.9999999	9.973E-08
Son.	G_1:1	1	1.114E-12
Taba.	G_1:1	1	4.59E-13
Tamps.	G_1:1	0.9999999	3.72E-08
Tlaxc.	G_1:1	1	5.633E-14
Ver.	G_2:2	2.889E-08	1
Yuc.	G_1:1	1	6.961E-12
Zac.	G_1:1	1	4.277E-17
Grupo 1=G_1:1			
Grupo 2=G_2:2			
Fuente: cálculos propios			

# CONCLUSIONES

Tanto para el año 1999 como para el 2000, las entidades federativas se distribuyeron en dos grupos, los cuales no difieren mucho en cuanto a su composición.

En ambos años 1999 y 2000, el primer grupo contiene a la mayoría de las entidades federativas, esto quiere decir que dichos estados se comportan de manera similar en cuanto al fenómeno de la mortalidad infantil asociado con las distintas variables de salud y económicas consideradas en este trabajo. Lo anterior no significa que todas las entidades tengan la misma cantidad defunciones de niños o niñas menores de un año, o que el número de camas censables sea el mismo o el mismo número de quirófanos, entre otros factores, sino significa que las proporciones entre todas las variables consideradas es muy similar para todos los estados que componen el grupo.

El segundo grupo en ambos años es el que tiene menos entidades federativas, dicho grupo está definido de esa manera no por que su mortalidad infantil se comporte de manera similar, sino que en estas entidades federativas dicho fenómeno es totalmente diferente entre ellas, dichas diferencias se pueden deber a que probablemente el peso que tienen las variables, con respecto a la de la mortalidad infantil, es muy distinto o tal vez, a una variable o varias que tienen una mayor influencia.

En el análisis para ambos años se pudo apreciar que los casos, es decir, las entidades federativas, en su mayoría se mantienen en el mismo grupo, excepto el estado de Guanajuato, ya que este estado en el año 1999 está en el grupo dos, ello significa que su mortalidad infantil no tiene un comportamiento similar al de la mayoría de las entidades federativas que conforman la República Mexicana, y en el año 2000 pertenece al primer grupo, es decir, su mortalidad infantil tiene un comportamiento diferente al del año anterior, esto no significa que aumente o disminuya el número de niños(as) muertos(as) o algunas de las otras variables, sino simplemente cambiaron las proporciones entre todas ellas. Claro que se esperaría que hubiese un avance en cuanto a su mortalidad infantil, pero esto no puede precisarse con el análisis hecho anteriormente, ya que como se mencionó en el capítulo I, el fenómeno de la mortalidad infantil no es aislado, ni tampoco está relacionado

## CONCLUSIONES

---

con factores económicos y de salud únicamente, sino también influyen en él otros como: escolaridad de la madre y edad así como su residencia geográfica, por citar algunos. Otra explicación podría deberse a que hay una deficiencia en la toma de datos, es decir, que estos realmente no reflejan la magnitud de ellos.

Sin hacer un estudio de las tasas de mortalidad infantil, se puede concluir que para los años 1999 y 2000 el fenómeno es similar en la mayoría de las entidades federativas que conforman la República Mexicana, considerando para su análisis algunos factores económicos y de salud, y que hay cuatro estados que son totalmente diferentes con respecto a los demás, siendo estos: D. F., Estado de México, Jalisco y Veracruz.

Las técnicas aplicadas en este trabajo nos reflejan que existen condiciones de desigualdad entre las entidades federativa, esto se puede deber a varios factores, los cuales van desde que todavía existen entidades en las cuales la mortalidad infantil es muy alta -sin olvidar que tienen un comportamiento que tiene a disminuir-, hasta la gran desigualdad en cuanto a servicios de salud y a la oportunidades laborales.

Valdría la pena hacer un estudio que viera cuales son las diferencias existentes entre los grupos por separado, y si la mayoría está mejor o peor que el grupo al que pertenece, con el fin de disminuir las condiciones (consideradas en este trabajo) que puedan ser la causa de la mortalidad infantil.

# APÉNDICE

## Definiciones y propiedades de las matrices

Sea  $A$  una matriz cuadrada, definida de la siguiente forma:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

- La **traza de una matriz** cuadrada  $p \times p$  se define como la suma de sus elementos en la diagonal, y se escribe:

$$\text{tr}(A) = \sum_{i=1}^p a_{ii}$$

- Se dice que una matriz cuadrada es una **matriz diagonal** si todos sus elementos que no están en la diagonal son ceros.
- El **determinante de una matriz cuadrada**  $A$ , se define como:

$$|A| = \sum_{i=1}^p a_{ii} A_{ij}$$

Donde  $A_{ij} = (-1)^{i+j}$  multiplicado por el determinante de la matriz obtenida a partir de  $A$  al eliminar el  $i$ -ésimo renglón y su  $j$ -ésima columna.

- Si  $|A| \neq 0$ , entonces se dice que  $A$  es una **matriz no singular**. En este caso, existe la inversa de  $A$  y se denota por  $A^{-1}$ . La inversa satisface  $AA^{-1} = A^{-1}A = I$ .
- Se dice que un conjunto de vectores  $x_1, x_2, \dots, x_p$  es **linealmente dependiente** si existen las constantes  $c_1, c_2, \dots, c_p$ , las cuales no son todas iguales a cero, tales que



$\sum_{i=1}^p c_i x_i$ . De lo contrario, se dice que los vectores son linealmente independientes. Un conjunto de vectores es linealmente dependiente si y sólo si por lo menos uno de los vectores se puede escribir como una combinación lineal de los restantes.

- El **rango de una matriz** se define como el número máximo de renglones (columnas) en  $A$  que son linealmente independientes. De modo equivalente, el rango de  $A$  es la dimensión del subespacio vectorial generado por los renglones (columnas) de la matriz.
- Dos vectores,  $x$  y  $y$ , son **ortogonales** si  $x'y = 0$ . Además, se dice que dos vectores ortogonales son ortonormales si  $x'x = 1$  y  $y'y = 1$
- Se dice que una matriz cuadrada  $A$  es **ortogonal** si  $AA' = I$  o bien  $A'A = I$ . En este caso,  $A = A^{-1}$
- Si  $A$  es una matriz ortogonal  $|A|$ , entonces es igual a 1 o a  $-1$ .

## Definiciones y propiedades de los Eigenvalores y Eigenvectores

Los eigenvalores (valores propios) y eigenvectores (vectores propios) de una matriz son funciones especiales de los elementos de esta última que desempeñan un papel en extremo importante en muchas de las técnicas del análisis multivariado.

Sea una matriz simétrica de  $p \times p$ , definida de la siguiente forma:

- Los **eigenvalores** (también llamados raíces características) de son las raíces de la ecuación polinomial dada por:

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

A cada eigenvalor de  $\Sigma$  le corresponde un vector diferente de cero llamado **eigenvector** (también conocido como vector característico) que satisface  $\Sigma c_i = \lambda_i c_i$  para  $i = 1, 2, \dots, p$ .

Debido a que tiene  $p$  eigenvalores, tendrá  $p$  *eigenvectors*.

- Si  $\Sigma$  es una matriz simétrica de número reales, entonces sus eigenvalores y eigenvectores también consistirán en números reales.
- Los **eigenvalores** de  $\Sigma$  se denotan por  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ .
- Los eigenvalores no son únicos, de modo que a menudo se normalizan de tal forma que  $c_i' c_i = 1$ .
- Si una matriz es simétrica y si todos sus eigenvalores son positivos, se dice que es una **matriz positiva definida**.
- La traza de una matriz simétrica es igual a la suma de sus eigenvalores, es decir,
 
$$tr(\Sigma) = \sum_{i=1}^p \lambda_i .$$
- El determinante de una matriz simétrica siempre es igual al producto de sus eigenvalores, es decir,  $|\Sigma| = \prod_{i=1}^p \lambda_i$ .
- Una matriz simétrica es positiva definida si y sólo si, para cada  $\lambda_i > 0$  para cada  $i$ , y por lo menos un eigenvalor es igual a cero.

## Estadística

- **Covarianza:** medida paramétrica que expresa el grado de variación conjunta de dos variables. Operativamente la covarianza es la medida de la suma de los productos entre las desviaciones respecto de su media, de los valores de dos variables. La covarianza se basa en las unidades de medida originales de las dos variables. Por es su valor varía entre en -1 y 1.

$$Cov = \frac{(x_i - \bar{x}_i)(x_j - \bar{x}_j)}{N}$$

- **Matriz de varianza-covarianza:** matriz cuadrada simétrica que resulta de dividir cada una de las entradas de una matriz de cuadrados de sumas y productos por  $n$ . el número de valores de la distribución, o bien por  $n-1$ , según se trabaje con la población total o con una muestra de ella, respectivamente. Se llama de varianza-covarianza porque los valores restantes de la diagonal principal son las covarianzas de las dos variables que se cruzan en cada casilla, y los elementos de que forma la diagonal son las varianzas de cada variable.
- **Matriz de correlaciones:** es una matriz cuadrada y simétrica, donde sus elementos son los coeficientes de correlación entre una serie de variables. Los elementos de la diagonal principal, expresan la correlación de cada variable consigo misma, por lo cual su valor es igual a 1. Los elementos que se encuentran a la izquierda de la diagonal principal, expresan las correlaciones entre las mismas variables en orden invertido, que las que están a la derecha de dicha diagonal, es decir,  $r_{ij} = r_{ji}$ , donde

$$r_{ij} = \frac{Cov_{x_i x_j}}{\sqrt{\sigma_i^2} \sqrt{\sigma_j^2}}$$

La matriz de correlaciones se puede obtener dividiendo las entradas de la matriz de varianza-covarianza por el producto de la desviación estándar de las los variables que intervienen en la covarianza.

- **Distancia de Mahalanobis:** se define como  $d_{ij} = \left[ (x_i - x_j)' \Sigma^{-1} (x_i - x_j) \right]^{1/2}$ . Donde  $\Sigma$  es la matriz de varianza-covarianza. Se utiliza cuando las variables están correlacionadas fuertemente entre sí.

# GLOSARIO

**Cama censable:** es la cama de servicio instalada en el área de hospitalización para uso regular de pacientes internos, debe contar con los recursos indispensables de espacio y personal para la atención médica, es controlada por el servicio de admisión de la unidad y se asigna al paciente en el momento de su ingreso hospitalario para ser sometido a observación, diagnóstico, cuidado o tratamiento. Incluye las de medicina interna, cirugía, gineco-obstetricia, pediatría y especialidades derivadas de ellas.

**Consulta externa:** es la atención médica que se imparte tanto en el consultorio como en el domicilio del paciente, sin utilización de la cama hospitalaria.

**Laboratorio de análisis clínicos:** área o local dotado de personas, mobiliario, equipo e instalaciones para realizar estudios de hematología, microbiología, clínicos e inmunológicos de muestras orgánicas provenientes de pacientes ambulatorios y hospitalizados.

**Población económicamente activa:** comprende a todas las personas de 12 años y más que realizaron algún tipo de actividad económica o que buscaron activamente hacerlo en los dos meses previos a la semana de referencia.

**Población económicamente activa femenina:** proporción de las personas de las personas de 12 años y más del sexo femenino que realizaron algún tipo de actividad económica o que buscaron activamente hacerlo en los dos meses previos a la semana de referencia.

**Población económicamente activa masculina:** proporción de las personas de 12 años y más del sexo masculino que realizaron algún tipo de actividad económica o que buscaron activamente hacerlo en los dos meses previos a la semana de referencia.

**Población económicamente inactiva:** comprende a todas las personas de 12 años y más que no realizaron actividades económicas en la semana de referencia, ni buscaron hacerlo en los dos meses previos a la semana de referencia.

**Quirófano:** área o local de un hospital dotada de personal, mobiliario, equipo e instalaciones para la ejecución de procedimientos quirúrgicos.

**Unidad de consulta externa:** establecimiento con personal médico y paramédico organizado, donde se proporciona atención médica ambulatoria; incluye a las unidades móviles. Incluye el número de unidades del primer nivel con que cuenta la institución y cuya función principal es dar consulta externa.

**Unidad de hospitalización general:** es el establecimiento de los sectores público, social y privado cualquiera que sea su denominación en el que se presta atención médica integral a la población, que cuenta con recursos materiales, humanos, tecnológicos y económicos para la atención de pacientes en las ramas troncales de la medicina: medicina interna, cirugía, pediatría, gineco-obstetricia y especialidades de mayor demanda, brindando en general servicio las 24 horas al día, 7 días a la semana y 365 días al año. Incluye el número de hospitales generales, centros de salud con hospital entre otros, que cuenten con camas censables para la hospitalización de pacientes.

# BIBLIOGRAFÍA

## Libros

- Baez, Revueltas Berenice; *Análisis de los Componentes Principales en la salud femina*, Tesis de Licenciatura, UNAM, México, 2001.
- Dallas, E. Johson; *Métodos multivariados aplicados al análisis de datos*, International Thomson Editores, México, 2000.
- DGEI; *Boletín de Información Estadística*, Núm. 19, Vol. 1, México, 1999.
- DGEI; *Boletín de Información Estadística*, Núm. 19, Vol. 1, México, 2000.
- Everitt, B. Sidney; *Applied Multivariate Data Analysis*, Londres, Inglaterra, 1991.
- Giri, Nayaran C.; *Multivariate statistical inference*, Academic Press, New York, 1977.
- Jiménez, Ornelas René A.; *La desigualdad de la mortalidad en México: tablas de mortalidad para la República Mexicana y sus entidades federativas 1990*, CRIM-UNAM, Cuernavaca, Mor., 1995.
- Jonson, Dallas E.; *Métodos multivaridos aplicados al análisis de datos*, Internacional Thomson Editorories, México, 2000.

## BIBLIOGRAFÍA

---

- Harris, Richard A.; *A primer of multivariate statistics*, Academic Press, New York, 1975.
- Hernández, Bringas Héctor H.; *Aportes de Investigación Núm. 48: Un panorama de la mortalidad infantil en México*, CRIM-UNAM, Cuernavaca, Mor., 1991.
- INEGI; *El ABC de las Estadísticas Vitales*, México, 1997.
- Mina, Valdés Alejandro; *Curso Básico de Demografía*, Serie: Notas de clase, Vínculos Matemáticos núm. 118, Taller de Publicaciones de Matemáticas de la Facultad de Ciencias, UNAM, 7ª ed., México, 2001.
- Welti, Chanes Carlos; *México: programa Latinoamericano de Actividades de Población*, Instituto de Investigaciones Sociales, UNAM, México, 1997.

## Revistas

- Aguirre, Alejandro; *Cambios en la mortalidad infantil*, DEMOS, México, 1997.
- Camposortega, Cruz Sergio; *Cambios en la mortalidad: cien años de la mortalidad infantil*, DEMOS, México, 1995.
- Hernández, Bringas Héctor H; *Algunos determinantes de la mortalidad infantil*, DEMOS, México, 1998.



- Morelos, José B.; *Mortalidad: la mortalidad general a finales del segundo milenio*, DEMOS, México, 2000.

## Páginas Web

- [www.inegi.gob.mx](http://www.inegi.gob.mx)
- [www.ssa.gob.mx](http://www.ssa.gob.mx)
- [www.eumed.net/cursecon/colaboraciones/index.htm](http://www.eumed.net/cursecon/colaboraciones/index.htm)