



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

DETERMINACIÓN DE GRUPOS DE ESTADOS DE LA REPÚBLICA MEXICANA SOBRE EL TRABAJO MASCULINO; UNA APLICACIÓN DE COMPONENTES PRINCIPALES Y DISCRIMINANTE

298392

T E S I S
QUE PARA OBTENER EL TÍTULO DE:
A C T U A R I O
P R E S E N T A :
LUIS FELIPE GONZÁLEZ CONTRERAS

DIRECTOR DE TESIS: M. EN A.P. MARIA DEL PILAR ALONSO REYES



MÉXICO, D.F.

FACULTAD DE CIENCIAS
UNAM



2001



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.



LIBERTAD NACIONAL
JUSTITIA
MEXICO

M. EN C. ELENA DE OTEYZA DE OTEYZA
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis:

Determinación de grupos de estados de la República Mexicana sobre el trabajo masculino; una aplicación de componentes principales y discriminante

realizado por Luis Felipe González Contreras

con número de cuenta 09003868-2, pasante 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

Propietario

M. en C. José Antonio Flores Díaz

Propietario

Act. Jaime Vázquez Alamilla

Suplente

Act. María Guadalupe Tzintzun Cervantes

Suplente

Lic. Reyna Pineda González

Consejo Departamental de



M. en C. José Antonio Flores Díaz

FACULTAD DE CIENCIAS
CONSEJO DEPARTAMENTAL
DE
MATEMÁTICAS

A DIOS por la vida
y por su compañía

*Por lo cual estoy cierto que ni la
muerte, ni la vida, ni ángeles, ni
principados, ni potestades, ni lo
presente, ni lo porvenir, ni lo alto,
ni lo bajo, ni ninguna criatura nos
podrá apartar del amor de Dios,
que es en Cristo Jesús Señor
Nuestro.*

Romanos 8: 38,39

A mis padres
Felipe (Conchito)
Alicia (Chonita)
Por su amor, su apoyo, su paciencia, su motivación

A mis hermanas
María Elena, María Dolores y Guillermina
Por su amor, comprensión y complicidad

A mis maestros; en especial
A Pilar por el apoyo brindado
A José Antonio por su tiempo
A Guadalupe, Reyna y Jaime por sus comentarios

A mis amigos de INEGI
María Teresa, Rita, Rene, Paco y Juan
A todos gracias por su orientación y consejos

A mis otros amigos
Alejandro, Francisco, Fernando, Ivonne y Adriana
Por su respeto y amistad

CONTENIDO

| | |
|---|----|
| INTRODUCCIÓN..... | 4 |
| 1. El empleo en México (1998)..... | 5 |
| 1.1. El empleo en general..... | 5 |
| 1.1.1. Conceptualización del trabajo y definiciones..... | 6 |
| 1.1.2. Definiciones operacionales de empleo y desempleo..... | 9 |
| 1.1.3. Definición de la población económicamente activa y sus subgrupos..... | 11 |
| 1.1.4. Condicionantes sociales de la inserción masculina en la fuerza de trabajo..... | 12 |
| 1.2. Empleo y género..... | 15 |
| 1.2.1. Participación económica de la población (1970 - 1998)..... | 15 |
| 1.2.2. Tendencias generales y diferencias por sexo..... | 15 |
| 1.2.3. Participación por edad y sexo..... | 17 |
| 1.2.4. Participación por nivel de escolaridad y sexo..... | 18 |
| 1.2.5. Tendencias del empleo por sector de actividad y sexo..... | 20 |
| 1.2.6. Estado civil fecundidad y empleo..... | 24 |
| 1.2.7. Subestimación de la participación económica de la mujer..... | 26 |
| 1.2.8. Segregación y discriminación de la mujer en el mercado de trabajo..... | 29 |

| | |
|--|----|
| 2. Análisis de componentes principales..... | 37 |
| 2.1. Procedimiento de cálculo de los componentes principales..... | 38 |
| 2.2. Sentido geométrico de los componentes principales..... | 48 |
| 2.3. Determinación del número de componentes principales a usar..... | 55 |
| 2.4. Representación de los datos..... | 59 |
| 3. Análisis discriminante..... | 61 |
| 3.1. Planteamiento del objetivo principal del análisis discriminante..... | 62 |
| 3.2. Discriminación cuando las poblaciones son conocidas (distribuciones y parámetros)..... | 63 |
| 3.2.1. La regla discriminante por máxima verosimilitud..... | 63 |
| 3.2.2. La regla discriminante de Bayes..... | 67 |
| 3.3. Discriminación cuando se conoce la distribución de las poblaciones pero los parámetros son estimados..... | 70 |
| 3.3.1. Clasificación por mínima distancia..... | 72 |
| 3.4. Función lineal discriminante de Fisher..... | 73 |

| | | |
|--------|--|-----|
| 3.5. | Probabilidades de mala clasificación..... | 79 |
| 3.5.1. | Probabilidades cuando los parámetros son estimados..... | 79 |
| 3.5.2. | Método de resubstitución..... | 81 |
| 3.5.3. | El método U de Jack-Kniff..... | 81 |
| 4. | Análisis estadístico..... | 83 |
| 4.1. | Aplicación de la técnica de componentes principales..... | 84 |
| 4.1.1. | Análisis de resultados de la población ocupada..... | 84 |
| 4.1.2. | Análisis de resultados de la población desocupada abierta..... | 91 |
| 4.2. | Aplicación de la técnica de análisis discriminante..... | 97 |
| 4.2.1. | Análisis de resultados de la población ocupada..... | 97 |
| 4.2.2. | Análisis de resultados de la población desocupada abierta..... | 102 |
| 5. | Conclusiones..... | 112 |

ANEXOS

| | | |
|-----------|--|-------|
| Anexo I | Información sobre la Encuesta Nacional de Empleo..... | A-I |
| Anexo II | Resultados de álgebra lineal..... | A-II |
| Anexo III | Definición de variables; población ocupada..... | A-III |
| Anexo IV | Definiciones de variables; población desocupada abierta..... | A-IV |

Bibliografía

INTRODUCCIÓN

El presente trabajo tiene como objetivo principal el determinar grupos de entidades federativas de la República Mexicana, seguido de la validación de la agrupación propuesta que de acuerdo a dos conjuntos de variables al interior de ellos la situación del empleo masculino sea similar.

En el primer capítulo se realiza un análisis del término trabajo, desde su forma más simple hasta el punto de vista económico. Se definen conceptos importantes vinculados al empleo. Y en una segunda parte se realiza un análisis descriptivo-comparativo de la población económicamente activa por sexo; finalmente se revisan algunas problemáticas, que son más comunes a la población femenina.

Por otra parte, debido al número de variables involucradas en el estudio se hace necesario para la obtención de la agrupación de las entidades federativas, el uso de técnicas multivariadas. Por lo que en los siguientes dos capítulos se hace una descripción, y se presentan algunos resultados importantes del análisis de componentes principales (segundo capítulo) y discriminante (tercer capítulo).

En el cuarto capítulo, se realiza un análisis estadístico, que involucra según se ha mencionado a la población masculina de las 32 entidades federativas del país exclusivamente. Primeramente se realiza un examen exploratorio, seguido de la aplicación de la técnica de componentes principales (que permite la representación gráfica de las 32 entidades federativas) para concluir con la validación de la agrupación propuesta vía la técnica del discriminante de Fisher.

Por último se presentan las conclusiones en el capítulo cinco, seguido de los anexos, tres de los cuales corresponden a información referente a la Encuesta Nacional de Empleo (ENE-1998), realizada por el Instituto Nacional de Estadística Geografía e Informática y otro de ellos presenta resultados matemáticos.

1. El empleo en México (1998).

1.1. El empleo en general.

A través de la revisión, de los significados de los términos empleo y trabajo es posible hacer notar la relación que existe entre ellos; se encontró que:

Empleo. Es la acción de emplear; o bien hace referencia a ocupación u oficio de una persona o individuo.

Emplear. Es ocupar a alguien encargándole un negocio, comisión o puesto.

Trabajo. Es toda acción que requiere un esfuerzo físico o mental para alcanzar un fin determinado.

De tal forma que un empleo supone la realización de un trabajo; pero la ejecución de un cierto trabajo no implica necesariamente la existencia de un empleo, es decir, el individuo por su propia cuenta puede llevar a cabo una actividad que nada tenga que ver con la encomienda de otro.

Cuando a los términos empleo y trabajo son tratados desde un punto de vista económico, no es difícil ver porque para el Instituto Nacional de Estadística Geografía e Informática (INEGI), trabajo, es "toda actividad económica que en general tiene como propósito la obtención de un ingreso monetario o en especie o la contribución para generarlo. Para los trabajadores sin pago su actividad tiende a contribuir a la generación de un ingreso familiar o es una forma de conseguir capacitación o empleo".

De acuerdo a lo anterior se tiene la siguiente relación: Todo empleo es un trabajo económico; pero no todo trabajo económico es un empleo, ya que este último supone un pago monetario o en especie.

En los apartados siguientes se hace un análisis más detallado de los términos trabajo, empleo, y un tratamiento de temas relacionados.

1.1.1. Conceptualización del trabajo y definiciones.

Debido a la formulación de planteamientos filosóficos y religiosos en los últimos siglos, así como a los adelantos científicos y tecnológicos de cada época, el concepto del "trabajo" ha cambiado y es difícil tratarlo.

Tales doctrinas sobre el trabajo han tenido repercusiones distintas en la historia de las relaciones humanas.

Enseguida se citan algunas de las perspectivas ideológicas que han tenido influencia histórica sobre la concepción y organización del trabajo en la sociedad occidental.

1. Concepto judeocristiano.

Debido a la influencia del cristianismo sobre la valoración social del trabajo en la sociedad actual se aborda esta conceptualización.

Se parte de que el trabajo es un castigo impuesto por Dios, y se asume con todas sus consecuencias la frase "*ganarás el pan con el sudor de tu frente*". De manera que se difunde la interpretación de que el trabajo aparece como consecuencia del pecado, siendo un sistema de redención del mismo.

Es una interpretación negativa y un concepto pobre de lo que es el trabajo. Sin embargo es difícil ponderar su vigencia cuando se contraponen creencias aprendidas en el proceso de socialización de los individuos con el hecho de vivir en una sociedad en continuo cambio y transformación.

2. *Concepto calvinista.*

La ética protestante representada por la moral calvinista, que según Max Weber, tuvo una gran influencia en el desarrollo del espíritu capitalista, produjo un efecto favorable sobre la aparición y desarrollo de la sociedad industrial.

Se plantea una interpretación del trabajo que ayuda a sobrellevar la indeterminación de no saber si el individuo al final de la vida iba a ser salvo o al contrario a condenarse, y se afirma que el hombre es el encargado de realizar la obra de Dios a través de su trabajo; el que más trabajaba mejor realizaba esa obra, y el que triunfa es porque estaba predestinado para la salvación.

A pesar de que tampoco es un concepto muy elevado de lo que es el trabajo sirvió como un paso importante en la revolución inicial de la producción, ya que desde esta perspectiva de análisis es relevante el hecho de que introduce una valoración positiva del trabajo, en un medio en el que prevalecía la consideración punitiva del cristianismo. Sin embargo se sigue sin reconocer valores, más que intrínsecos al trabajo por sí mismo, sino que se le trata como un medio necesario para el logro de unas metas extraterrenales deseables.

3. *Concepto económico.*

El trabajo organizado surge con la aparición y desarrollo de la sociedad industrial (Revolución Industrial), se caracteriza por una relación asalariada, de compraventa, donde

existe una transacción entre un hombre que realiza un trabajo a otra persona y ésta le paga por su trabajo un salario que posibilita la satisfacción de sus necesidades.

Este es un concepto materialista del trabajo, es decir, se trabaja por necesidad. Refuerza la consideración del trabajo como un mal necesario, lo cual unido a las inhumanas condiciones en que se desarrollaba en los inicios de la Revolución Industrial, favorece el que fuera visto con unas connotaciones destructivas considerables.

4. *Concepto del trabajo como creación.*

Parte de la idea de que el hombre es un ser creador y mediante su trabajo busca la autorrealización. Así se ha de considerar al trabajo como una actividad a través de la cual el hombre se enfrenta a su medio ambiente natural y social, transformándolo en función de sus intereses. El hombre es pues un creador esencialmente un transformador que se exterioriza a través de las modificaciones que impone a su medio, las cuales son el camino para expresar su creatividad.

Para los psicólogos el trabajo es una conducta: así, Ombredame y Faverge (1955) lo definen como un *"complemento adquirido por aprendizaje y que debe adaptarse a las exigencias de una tarea"*. Estas varían notablemente según el nivel técnico-socioeconómico de la organización en la cual se inserta la tarea. Para Tredman (1976), el único factor común a las actividades denominadas "trabajo" es un elemento de compulsión (apremio, obligación, disciplina), que puede ser de origen interno o externo; en cuanto a conducta, puede caracterizarse también por su complejidad y por el hecho de que, en general, por lo menos en la sociedad actual, sólo se considera con relación a la gente mayor a cierta edad.

Desde un punto de vista social, Marx escribe: *"Para producir los hombres contraen determinados vínculos y relaciones y a través de estos vínculos y relaciones sociales, y sólo a través de ellos es como se relacionan con la naturaleza y como se efectúa la producción"*. El trabajo social es pues el trabajo enlazado de los hombres en un proceso de producción de determinados productos sociales, esto es, útiles para la sociedad.

Los hombres contrajeron vínculos de trabajo desde los inicios de la sociedad, los cuales crecieron y se hicieron cada vez más diversos a medida que se desarrollaron las fuerzas productivas y, con esto, la división y cooperación del mismo.

Desde un punto de vista social y económico. Se puede decir que el trabajo es un esfuerzo humano, físico e intelectual, aplicado a la transformación de la naturaleza, para la satisfacción de las necesidades (fisiológicas y sociales) del hombre.

Finalmente el trabajo puede definirse como la actividad humana racional encaminada a la producción de valores de uso dirigidos a satisfacer las unidades. Esta es un proceso que comprende tres factores: una acción adecuada a un propósito o fin; el objeto, es decir, aquello sobre lo que se realiza la actividad; y, el medio, constituido por el conjunto de objetos que el trabajador usa en la ejecución de su tarea.

Así pues el trabajo responde a tres necesidades primarias y fundamentales de la naturaleza humana las cuales son:

1. La necesidad de subsistir (función económica).
2. La necesidad de crear (función psicológica).
3. La necesidad de colaborar (función social).

1.1.2. Definiciones operacionales de empleo y desempleo.

Las personas comprendidas en el empleo son todas aquellas que tengan más de cierta edad especificada y que estén dentro de las categorías siguientes:

1. Que estén trabajando; es decir las personas que realizan algún trabajo remunerado en especie o monetariamente, durante un breve período especificado, ya sea durante una semana o un día.

2. Que tengan un empleo, pero que no estén trabajando, esto es, las personas que hayan trabajado ya en su empleo actual, pero que se hallen temporalmente ausentes del trabajo en el curso de período especificado debido a enfermedad o accidente, conflicto de trabajo, vacaciones u otra clase de permiso, ausencia sin permiso, interrupción del trabajo a causa de determinados eventos como por ejemplo el mal tiempo.
3. Se considera que los trabajadores familiares no remunerados que ordinariamente exploten o ayuden a explotar un negocio cualquiera o una explotación agrícola tienen un empleo si han trabajado por lo menos un tercio del tiempo normal de trabajo durante el período especificado.

No se considerarán empleadas a las personas comprendidas en las categorías siguientes:

- a) Los trabajadores que durante el período especificado hayan sido suspendidos temporal o indefinidamente sin goce de remuneración
- b) Las personas que no tengan ningún empleo o que no exploten un negocio cualquiera o una explotación agrícola, pero que hayan obtenido un nuevo empleo, negocio o explotación a comenzar en una fecha subsiguiente al período de referencia.
- c) Los trabajadores familiares no remunerados que trabajen menos de un tercio del tiempo normal de trabajo durante el período especificado en un negocio o explotación agrícola familiar.

Por otra parte las personas comprendidas en el desempleo son todas aquellas que tengan más de cierta edad especificada y que, en un día concreto, se hallen en las siguientes categorías:

1. Los trabajadores disponibles para el empleo cuyo contrato de trabajo haya expirado o esté suspendido temporalmente, que estén sin empleo y busquen trabajo remunerado durante un breve período especificado con referencia a una semana.¹

¹ En la Encuesta Nacional de Empleo (ENE), el período de referencia es de dos meses, mientras que en los censos es de una semana.

2. Las personas que no hayan estado empleadas nunca y aquellas cuya categoría de ocupación más reciente sea distinta de la de asalariada en unión de las que estén jubiladas, cuyas personas se hallan disponibles para trabajar en el curso del período especificado y estén buscando trabajo remunerado.
3. Las personas sin empleo que en el momento de que se trate se hallen disponibles para trabajar y hayan logrado un nuevo empleo que deba empezar en una fecha subsiguiente al período especificado.
4. Las personas que hallan sido suspendidas temporal o indefinidamente, sin goce de remuneración.

No se consideran desempleadas a las personas comprendidas en las siguientes categorías.

- a) Las personas que tengan el propósito de establecer por su cuenta un negocio cualquiera o explotación agrícola; pero que no hayan tomado medidas en esa dirección y que no estén buscando trabajo remunerado.
- b) Los antiguos trabajadores familiares no remunerados que no estén trabajando ni buscando trabajo remunerado.

1.1.3. Definición de la población económicamente activa y sus subgrupos.

Población Económicamente Activa. Son todas las personas de 12 años y más ² que en la semana de referencia, realizaron algún tipo de actividad económica, o formaban parte de la población desocupada abierta.

² En 1930 se captó a las personas de 6 años y más que realizaban algún tipo de actividad económica. A partir del censo de 1940 se considera económicamente activa exclusivamente a las personas de 12 años y más. Aunque en 1960 se registraron a las de 8 años y más.

Población ocupada. Son todas las personas de 12 y más años que en la semana de referencia:

- a) Participaron en actividades económicas al menos una hora o un día a la semana a cambio de un ingreso monetario o en especie o que lo hicieron sin recibir pago.
- b) No trabajaron pero cuentan con un empleo.
- c) Iniciarán alguna ocupación en el término de un mes o menos.

Población ocupada excluyendo a los iniciadores de un próximo trabajo. Son todas las personas de 12 años y más que en la semana de referencia estuvieron en la condición descrita en los incisos a) o b) de la definición anterior.

Población desocupada abierta. Son las personas de 12 años y más, que sin estar ocupados en la semana de referencia, buscaron incorporarse a alguna actividad económica en el mes previo a la semana de levantamiento o entre uno y dos meses aun cuando no lo haya buscado en el último mes por causas ligadas al mercado de trabajo, pero que estén dispuestos a incorporarse de inmediato.

1.1.4. Condicionantes sociales de la inserción masculina en la fuerza de trabajo.

Es posible que siempre haya existido una cierta división sexual del trabajo. Sin embargo la concepción del trabajo como una "obligación" primaria de los hombres y una actividad secundaria de las mujeres es relativamente reciente. En las sociedades recolectoras y cazadoras, las mujeres trabajaban por lo menos lo mismo que los hombres, sino es que más. En estas sociedades primitivas los hombres eran los encargados de la caza, de la fabricación de todos los utensilios para ella como: herramientas, armas, trampas, equipos para pesca, botes, etcétera. Mientras que las mujeres se dedicaban a preparar lo que se comía, a fabricar lo que se vestía y en ocasiones contribuir a la construcción y mantenimiento de la "vivienda" en la mayoría de las sociedades tribales que dependían de una agricultura sedentaria, eran principalmente las mujeres quienes realizaban las tareas

en los sembradíos, mientras que la alimentación y la crianza del ganado estaba destinada a los hombres. Inclusive en las sociedades pastoriles más desarrolladas, las mujeres pasaban una gran parte de su tiempo en los campos, y toda la familia trabajaba como un equipo de producción. Cuando la caza o la agricultura constituyen el tipo principal de trabajo disponible, la división sexual del trabajo no es entre trabajo y no-trabajo, sino simplemente entre diferentes tipos de tareas. Los hombres se encargan de las tareas que requieren mayor movilidad, fuerza; aquellas que implican mayores riesgos físicos como la defensa y ataque. Y por otra parte a las mujeres les corresponden tareas que implican poca fortaleza física y que permitan la crianza de los hijos.

La idea de que el trabajo es principalmente una responsabilidad, una obligación de los hombres surgió con mucha certeza junto con el auge de las ciudades y el desarrollo de una tecnología industrial. Una vez que empezó a notarse la distinción entre trabajo pagado y no-pagado, se tendió a pensar que sólo el primero era trabajo autentico. Esta distinción se agudizo, por supuesto al surgir un gran número de fábricas y establecerse firmemente una economía monetaria. El trabajo se convirtió entonces cada vez más en una actividad que se realizaba fuera del hogar y por la que se recibía un salario. Entre la crianza de los niños y el trabajo doméstico, la esposa del obrero podía terminar desgastada, pero esto no se consideraba ya trabajo porque no se recibía un pago de manera directa.

En la sociedad moderna existe una relación muy estrecha entre el trabajo y la compensación monetaria, que sólo se habla de trabajador cuando el individuo está realizando una actividad por la que se le paga un sueldo o salario.

La mayor cantidad de mujeres que trabajan lo hacen antes del matrimonio y después de que sus hijos crecen lo suficiente como para no requerir un cuidado continuo. Por supuesto, esta correlación con el estado civil, o con la edad o presencia de hijos no se encuentra en el caso del trabajador masculino.

Más allá de las diferencias biológicas, aunque guardando una relación indirecta con ellos, es obvio que las "chicas" y los "muchachos" enfrentan presiones y exigencias, sociales y culturales muy diferentes. En las sociedades como la de México, se hace saber a los "muchachos" por diversas vías que se espera de ellos que "se ganen la vida" cuando

crezcan, que sean independientes para que puedan proporcionar una seguridad económica a su pareja y a sus hijos. Desde una época relativamente temprana, los "muchachos" se enfrentan a poderosas presiones culturales, cuyo objetivo es producir un adulto responsable en todos los aspectos de su vida, sobre la base de algún tipo de trabajo remunerado. Si bien las "chicas" no están exentas de algunas de estas presiones, los imperativos culturales para las jóvenes son los de casarse y procrear hijos. Existe una creencia amplia y profundamente arraigada de que el trabajo femenino es temporal e incidental.

En países donde se ha avanzado hacia la igualdad de los sexos en lo legal y social del trabajo, la presión ejercida sobre las mujeres para que se incorporen al mercado de trabajo sigue siendo menor que para los hombres. La sociedad muestra una mayor tolerancia hacia la mujer que no trabaja que hacia el hombre desocupado; pero por otra parte la mujer que trabaja enfrenta problemas de adaptación mucho más difíciles que el hombre. La trabajadora casada no sólo tiene que aportar gran parte del día al trabajo remunerado, sino que también ocuparse de las tareas domésticas.

Evidentemente la actual división del trabajo está condicionada por las tradiciones culturales, por lo menos en la misma medida que por las necesidades técnicas de éste. Aún los problemas referentes al cuidado de los hijos podrían ser substancialmente reducidos, aunque no eliminados por completo, si las tareas del hogar empezaran a ser compartidas entre hombres y mujeres.

A pesar de los considerables avances de los sexos en lo que respecta al trabajo, la sociedad todavía distingue en forma bastante aguda entre las tareas de desarrollo planteadas a sus niños. A los "muchachos" se les cría para que se conviertan primero en trabajadores y luego en esposos y padres; a las "chicas", para que sean esposas y madres primero, y trabajadoras después. Cada vez se hace más difícil encontrar tareas puramente técnicas que justifiquen estas diferencias. Hay una amplia evidencia de que las mujeres pueden desempeñar tan bien como los hombres las mismas ocupaciones. Sin embargo, las tradiciones culturales son todavía muy poderosas. Puede esperarse, por lo menos que los componentes de la personalidad laboral se inserten más firmemente en los "muchachos" que en las chicas por el momento. Sin embargo, están ocurriendo cambios

culturales masivos, y no es en absoluto seguro que esta situación se mantenga indefinidamente.

1.2. Empleo y género.

1.2.1. Participación económica de la población (1970 - 1998).

Para exponer las tendencias globales de la participación económica de la población a lo largo del periodo comprendido entre 1970 y 1998 es necesario hacer algunas consideraciones; la comparación entre distintos censos, y entre éstos y las encuestas de empleo, involucra una serie de limitaciones que provienen de los cambios en las definiciones, los períodos de referencia y las formas de captación de la información, que dificultan la apreciación de las tendencias; sin embargo el cálculo de los indicadores con los datos de cada estudio, sirven para hacer una comparación.

1.2.2. Tendencias generales y diferencias por sexo.

Entre 1970 y 1998, la población mantiene un promedio de participación neta de alrededor del 50%. Esto es que de la población mayor a cierta edad especificada más o menos la mitad se encontraba ubicada en la Población Económicamente Activa (PEA), sin embargo esta tendencia es hacia la baja en las dos siguientes décadas a 1950, existiendo una sensible recuperación para las tres últimas décadas del siglo. Una parte de este descenso es debido a la combinación de un conjunto de factores económicos, demográficos y culturales vinculados al proceso de desarrollo económico, con un impacto diferencial sobre la fuerza de trabajo masculino y femenino.³

Cuadro 1.1 Participación económica de la población, 1970-1998

| Año | 1970 | 1979 | 1991 | 1996 | 1998 |
|------------|------|------|------|------|------|
| Porcentaje | 43.6 | 43.7 | 53.5 | 53.4 | 55.7 |

³ Oliveira, Orlandina de, Marina Ariza y Marcela Etemod, 1998.

La participación económica masculina excede a la femenina a lo largo del período manejado. En 1970 cerca del 81% del total de la población económicamente activa estaba constituida por hombres y el restante 19% por mujeres; para 1998 la participación masculina ha disminuido hasta representar un poco más del 66%, mientras que la femenina aumentó a cerca del 34%.

Cuadro 1.2 Distribución de la PEA por sexo, 1970-1998

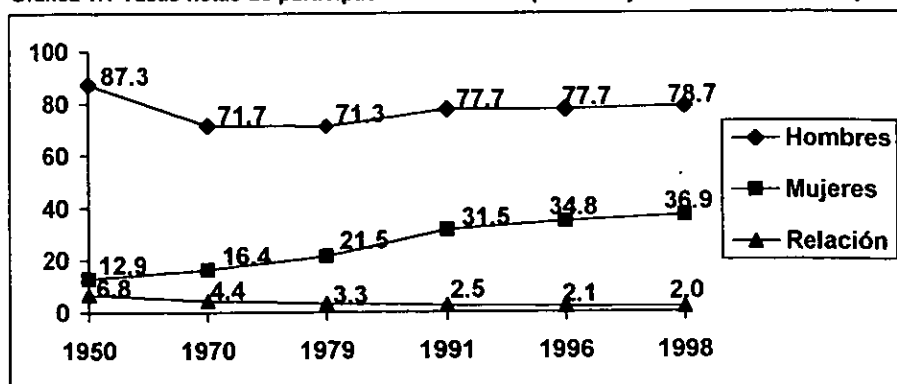
| Año | 1970 | 1979 | 1991 | 1996 | 1998 |
|---------|------------|------------|------------|------------|------------|
| Hombres | 10,488,800 | 14,483,000 | 21,630,013 | 24,627,936 | 26,176,026 |
| Mujeres | 2,466,257 | 4,594,000 | 9,599,035 | 11,952,810 | 13,331,037 |
| Total | 12,957,027 | 19,078,979 | 31,231,039 | 36,582,742 | 39,509,061 |

Cuadro 1.3 Distribución porcentual de la PEA por sexo, 1970-1998

| Año | 1970 | 1979 | 1991 | 1996 | 1998 |
|---------|--------|--------|--------|--------|--------|
| Hombres | 80.96 | 75.92 | 69.26 | 67.32 | 66.25 |
| Mujeres | 19.04 | 24.08 | 30.74 | 32.68 | 33.75 |
| Total | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

La disimilitud entre la tasa de participación económica masculina y la femenina ha experimentado una disminución progresiva (ensanchamiento de las distancias). En 1950 por cada mujer económicamente activa existían casi 7 hombres, a partir de entonces las diferencias no han dejado de acotarse. En efecto la tasa neta de participación femenina ha venido experimentado aumentos significativos, mientras que la masculina después de una caída de un poco menos del 16% (1950-1970) ha experimentado ascensos muy discretos hasta llegar a situarse en 1998 en 78.7%.

Gráfica 1.1 Tasas netas de participación económica por sexo y relación hombres-mujeres



Para 1998 la tasa de participación masculina es casi poco más de dos veces la femenina; o lo que es lo mismo, la participación económica de las mujeres se aproxima a la mitad de la participación masculina.

1.2.3. Participación por edad y sexo.

Como lo han señalado algunos autores "en la mayoría de las sociedades industrializadas, la institución del matrimonio y la estructura familiar misma, contribuyen a desarrollar una firme inclinación laboral en los hombres en la medida en que ejercen presión sobre ellos para lograr la satisfacción de una serie de necesidades del consumo" ⁴. En las mujeres por el contrario la segregación entre la producción y la reproducción condiciona el modo de participación económica, dando lugar a pautas muy diversas.

En el período 1970-1998 se experimenta un aumento en las tasas de participación económica para todos los grupos de edad, a excepción de los últimos dos grupos más viejos masculinos que sufren un descenso superior a los 5 y 18 puntos porcentuales respectivamente.

En el caso de la población masculina los grupos de edad 25-34,35-44 y 45-54 mantienen sus tasas de participación por arriba de los 90 puntos, así mismo el incremento más representativo ocurre en el grupo de edad 12-19 al pasar de 33.8 en 1970 a 47.8 en 1991.

Para la población femenina los grupos de edad que sufren el mayor incremento en sus tasas de participación son los de 25-34 y 35-44 al pasar de 16.6 y 12.9 en 1970 a 46.6 y 48.8 en 1998 respectivamente. Por otra parte aunque de manera discreta también se observa un aumento en las tasas de participación de los grupos 20-24, que pasa de 24.1 en 1970 a 44.2 en 1998 y 45-54 que va de 16.2 en 1970 a 39.9 en 1998.

⁴ Standing, 1978.

Cuadro 1.4 Tasas específicas de participación por grupos de edad y sexo, 1970-1998

| Grupos de Edad | 1970 | | 1979 | | 1991 | | 1996 | | 1998 | |
|----------------|------|------|------|------|------|------|------|------|------|------|
| | H | M | H | M | H | M | H | M | H | M |
| 12 - 19 | 33.8 | 14.9 | 35.4 | 15.4 | 47.8 | 22.6 | 46.1 | 21.1 | 46.7 | 23.2 |
| 20 - 24 | 79.6 | 24.1 | 82.5 | 33.4 | 84.1 | 40.4 | 85.6 | 42.8 | 85.9 | 44.2 |
| 25 - 34 | 91.8 | 16.6 | 95.9 | 27.3 | 96.6 | 39.2 | 96.7 | 44.9 | 96.4 | 46.6 |
| 35 - 44 | 94.1 | 12.9 | 96.7 | 24.5 | 98.2 | 40.6 | 97.8 | 44.8 | 97.8 | 48.8 |
| 45 - 54 | 93.2 | 16.2 | 93.4 | 21.1 | 95.3 | 32.8 | 94.0 | 38.5 | 95.5 | 39.9 |
| 55 - 64 | 88.5 | 14.7 | 85.1 | 16.3 | 85.9 | 24.4 | 80.2 | 27.8 | 83.3 | 28.3 |
| 65 y más | 70.4 | 10.9 | 53.2 | 9.3 | 55.2 | 12.4 | 52.0 | 14.1 | 53.9 | 15.7 |

Es de llamar la atención que en las década de los 80 y los 90 la mujer no abandona el mercado de trabajo en la edad reproductiva como sucedía en la década de los 70 para dedicarse a la vida familiar.⁵

En consecuencia, la brecha entre los niveles de participación ha empezado a acotarse entre 1970 y 1998 se han registrado descensos de alrededor de 32 puntos porcentuales en las diferencias en las tasas de actividad masculina y femenina entre los 35 y 44 años, y de 25 puntos entre los 25 y 34 .

En la actualidad México presenta una disimilitud de alrededor de 50 puntos porcentuales entre las tasas de participación entre hombres y mujeres en el intervalo reproductivo de edad. Esta diferencia se asemeja a países como España, Italia e Irlanda que se caracterizan por una disparidad de entre 40 y 60 puntos porcentuales; pero contrasta con países europeos de mayor equidad como Suecia y Finlandia en los que la disparidad de los niveles de participación de hombres y mujeres de 24 a 54 años no excede los 7 puntos porcentuales.⁶

1.2.4. Participación por nivel de escolaridad y sexo.

La escolaridad es un aspecto fundamental en la explicación de la participación económica de hombres y mujeres en el mercado de trabajo. Se sostiene la teoría de que una mayor participación es consecuencia de un mayor nivel de escolaridad, debido a los requisitos

⁵ Oliveira, Orlandina de, Marina Ariza y Marcela Etemod , Op. cit. p. 5.

⁶ Organization for Economic Cooperation and Development, 1994.

cada vez más específicos que se imponen en el mercado de trabajo para el desempeño de distintas ocupaciones, esto supone que se hace necesario prolongar el periodo de formación escolar, retrasando con ello la edad de ingreso al mercado de trabajo; por otro lado, el establecimiento de una edad de retiro reduce significativamente la participación de los grupos más viejos. Sin embargo es importante aclarar que son las mujeres jóvenes (20 - 34 años) con mediana y alta escolaridad las que han enfrentado mayores dificultades para incrementar su presencia en el mercado de trabajo, en un contexto de crisis económica y contracción del empleo no manual.

La escolaridad de la población en México ha experimentado avances importantes en el último cuarto de siglo. Mientras que en 1976, 50% de la población femenina económicamente activa no tenía instrucción o sólo contaba con primaria incompleta, en 1998 esta cifra llegó a ser del 25%, y para los hombres este porcentaje se situó en un 28%.

La estructura por nivel de instrucción de la población económicamente activa ha tenido cambios muy marcados. A pesar de que es difícil una comparación (ya que no existe un estudio que de cifras acerca de los niveles de instrucción de la población masculina económicamente activa para 1976), se observa que para 1988 alrededor del 40% de la PEA total no contaba con instrucción o sólo tenía algunos años de primaria; esta tendencia se ha revertido ya que a finales de los años noventa alrededor del 50% de la PEA cuenta con por lo menos un año de secundaria.

Cuadro 1.5 Distribución de la PEA por nivel de instrucción y sexo, 1976-1998

| Nivel de instrucción | 1976 | | 1988 | | 1991 | | 1996 | | 1998 | |
|---------------------------|------|------|------|------|------|------|------|------|------|------|
| | H | M | H | M | H | M | H | M | H | M |
| Sin escolaridad | n.d. | 50.8 | 41.2 | 38.8 | 37.3 | 30.5 | 30.0 | 26.2 | 28.2 | 25.1 |
| Primaria completa | n.d. | 22.1 | 20.7 | 20.3 | 20.8 | 22.0 | 21.8 | 21.2 | 21.3 | 20.4 |
| Secundaria | n.d. | 15.7 | 20.9 | 27.9 | 24.0 | 31.6 | 26.1 | 31.5 | 27.6 | 32.4 |
| Medio superior y superior | n.d. | 11.4 | 17.2 | 13.0 | 17.9 | 15.9 | 22.1 | 21.1 | 23.0 | 22.1 |

n.d. Información no disponible.

Es de importancia el aumento inesperado de las tasas de participación económica de la población femenina sin escolaridad o con primaria incompleta, de diferentes edades, en la década de los 80. Este aumento llegó hasta valores muy cercanos a las tasas de

participación correspondientes al nivel de instrucción "primaria completa". Esta tendencia muy probablemente tuvo su origen en el deterioro de las condiciones de vida en estos años, y esto a su vez hizo necesaria la búsqueda de estrategias para sobrevivir a la población de bajos ingresos.

En 1988 las tasas de participación económica más altas para los hombres fueron las correspondientes a la población sin escolaridad⁷ (78.3), y primaria completa (80.0). Mientras que en las mujeres las tasas más representativas correspondieron a la población femenina con secundaria (35.9), y nivel medio superior y superior (45.7). Grupos que se mantienen como los más importantes hasta 1998 al llegar a 40.2% y 50.7% respectivamente. Por otra parte para los hombres cambia la situación ya que los niveles de instrucción que más participan son la población con primaria completa (81.3) y medio superior y superior (80.0).

Por último cabe destacar el descenso y estancamiento a lo largo del período en los niveles de participación de la población femenina con mayor escolaridad (medio superior y superior) sin embargo es necesario aclarar que estos sectores ya han alcanzado niveles bastante elevados de participación económica (66.2% en 1976 y 50.7% en 1998).

Cuadro 1.6 Tasas específicas de participación por nivel de instrucción y sexo, 1976-1998

| Nivel de Instrucción | 1976 | | 1988 | | 1991 | | 1996 | | 1996 | |
|---------------------------|------|------|------|------|------|------|------|------|------|------|
| | H | M | H | M | H | M | H | M | H | M |
| Sin escolaridad | n.d. | 21.5 | 78.3 | 29.1 | 79.9 | 24.2 | 76.9 | 27.3 | 77.4 | 29.3 |
| Primaria completa | n.d. | 27.8 | 80.0 | 29.2 | 80.5 | 30.1 | 80.6 | 31.9 | 81.3 | 33.3 |
| Secundaria | n.d. | 40.2 | 66.8 | 35.9 | 72.7 | 37.1 | 74.5 | 38.0 | 77.1 | 40.2 |
| Medio superior y superior | n.d. | 66.2 | 75.0 | 45.7 | 77.6 | 47.9 | 80.2 | 50.4 | 80.0 | 50.7 |

n.d. Información no disponible.

1.2.5. Tendencias del empleo por sector de actividad y sexo.

Los cambios de sector de la población económicamente activa a lo largo del presente siglo han sido verdaderamente importantes. La mano de obra en las actividades del sector primario ha disminuido 2.6 veces su importancia relativa, mientras la del terciario ha

⁷ Bajo el rubro de sin escolaridad se agrupa a la población sin instrucción y a aquellos con primaria incompleta.

crecido casi cuatro veces. La creciente incorporación de la población a las actividades terciarias ha sido de tal magnitud que México se ha transformado a finales de los 90 en una economía de servicios, ya que en él se encuentra poco más del 52% de la PEA. En este siglo la mano de obra en las actividades del sector secundario ha reducido ligeramente su importancia relativa.

En los años setenta se hablaba en América Latina de una incorporación excesiva de la mano de obra en actividades terciarias de muy baja calificación y escasa remuneración, tales como los servicios personales y el comercio ambulante. Estas actividades eran vistas como refugio de mano de obra que expulsada del sector agropecuario no encontraba cabida en el sector industrial de la economía.

A partir de los ochenta en un marco de crisis de las economías latinoamericanas, de una sustantiva reducción del estado benefactor y del aumento de los niveles de pobreza han cobrado vigencia las explicaciones que asocian el crecimiento del sector terciario con la expansión de las actividades no asalariadas de baja calificación, propias del sector informal urbano. Se enfatiza, en el mismo sentido, la pérdida de capacidad del sector industrial para absorber mano de obra en países que, como México habían sobresalido por un fuerte dinamismo del mismo.⁸

Durante la transición al nuevo modelo de desarrollo (1970-1979) a pesar de la pérdida en el dinamismo de la inversión pública y privada, la reducción del ritmo de crecimiento del producto industrial, el aumento de la inflación, la devaluación del peso y la fuga de capitales, la importancia relativa de la PEA industrial total y asalariada en la minería, energía y manufactura se mantuvo gracias al breve repunte de la economía en los últimos años de la década.⁹

En los años de ajuste, crisis económica y apertura comercial (1979-1995), se revierte por primera vez la tendencia de rápida expansión del empleo industrial iniciada en los años cincuenta. La industria nacional a excepción de las maquiladoras ha sido afectada por la crisis y las políticas económicas implantadas en esos años, esto ha provocado la

⁸ Rendón y Salas, 1992, 1993; García y Oliveira 1994, García 1996.

⁹ Rendón y Salas, 1987; García, 1988; Negrete, 1988; García y Oliveira, 1994.

contracción sistemática de la PEA industrial. Este fenómeno es observable en la disminución de los trabajadores asalariados por el efecto combinado de los avances tecnológicos, del cierre de empresas y el aumento de las importaciones.¹⁰ A pesar de ello, el proceso de salarización de la mano de obra industrial recupera su tendencia ascendente en la primera mitad de los noventa (1991-1995).

Cuadro 1.7 Distribución de la PEA por sector de actividad, 1970-1998

| Sector | 1970 | 1979 | 1991 | 1995 | 1998 |
|------------|------|------|------|------|------|
| Primario | 39.4 | 29.1 | 27.0 | 24.9 | 24.3 |
| Secundario | 25.3 | 27.6 | 23.2 | 21.4 | 23.3 |
| Terciario | 35.3 | 43.3 | 49.8 | 53.7 | 52.4 |

En México, ha prevalecido una clara división sexual entre trabajo agrícola y el no agrícola a lo largo del siglo, la población masculina se ha dedicado en su mayoría (esto en forma decreciente) a labores agrícolas; mientras que las mujeres a las no agropecuarias.

En el período 1970-1998 se nota una fuerte baja de la actividad agropecuaria masculina, de ser la más importante a principios de la década de los setenta, concentrando al 46% de la PEA masculina, pasa a ocupar un segundo lugar con un peso relativo del 29%. Este 17% de diferencia es reflejado en el aumento de las actividades terciarias que pasan de aglutinar de cerca del 28% de la PEA masculina a un poco más del 45%.

En cuanto al sector secundario se observa que de concentrar al 25.3% de la PEA, reúne para 1998 al 23.3%. Experimentándose el aumento más significativo (2.3%) en el periodo 1970-1979 y la disminución más grande de 1979-1991 (4.4%).

A diferencia de los hombres, las mujeres han mantenido a cerca del 70% de su PEA en el sector terciario a lo largo de estos últimos 28 años. Y han tenido como segundo sector más importante al secundario donde se observa una gradual disminución, al pasar de un 22% en 1970 a un 18.9% en 1998; esto debido a la fuerte contracción de la actividad industrial en los ochenta y principios de los noventa, cabe aclarar que esta misma situación repercute en las actividades del sector secundario de la población masculina.

¹⁰ García y Oliveira, 1994; García, 1997.

Cuadro 1.8 Distribución de la PEA por sector de actividad y sexo 1970-1998

| Sector | 1970 | | 1979 | | 1991 | | 1995 | | 1998 | |
|------------|------|------|------|------|------|------|------|------|------|------|
| | H | M | H | M | H | M | H | M | H | M |
| Primario | 46.1 | 10.8 | 36.5 | 5.7 | 34.1 | 10.9 | 31.3 | 11.1 | 29.2 | 14.9 |
| Secundario | 26.1 | 22.0 | 29.3 | 22.4 | 24.9 | 19.2 | 24.4 | 15.1 | 25.7 | 18.9 |
| Terciario | 27.8 | 67.2 | 34.2 | 71.9 | 41.0 | 69.9 | 44.3 | 73.8 | 45.1 | 66.2 |

Para analizarse de manera conjunta la participación económica de hombres y mujeres las variaciones en los índices de feminización de la fuerza de trabajo por sectores son de gran utilidad. Con ellos es posible ver la segregación por sexo entre las actividades agrícolas y las no agrícolas. Por ejemplo la elevada masculinización de las actividades del sector primario en 1970 solamente 5.5 mujeres por cada 100 hombres realizaban labores en el sector agrícola; 28 años después la cifra casi se quintuplica, sin embargo sigue siendo baja.

Las actividades del sector secundario, que a principios de siglo se caracterizaban por una elevada feminización de la mano de obra han experimentado un importante proceso de masculinización, como resultado de la salida masiva de las mujeres en la primera década del siglo.

Será únicamente durante los años ochenta cuando en un contexto de fuerte contracción de la actividad industrial en general, la expansión de las industrias contribuya a incrementar el grado de feminización de las actividades secundarias hasta llegar a 34 mujeres por cada 100 hombres en 1991, sin embargo este proceso se revierte en la primera mitad de los noventa, cuando vuelve a disminuir la presencia femenina respecto de la masculina.

En contraste con el sector primario y secundario, el terciario ha presentado a lo largo del siglo, un elevado grado de feminización. En 1970 por cada cien hombres existían 56.8 mujeres en actividades de servicios o comercio y para 1995 son ya cerca de 79 mujeres por cada cien hombres.

Cuadro 1.9 Índices de feminización de la PEA por sector de actividad y sexo 1970-1998

| Sector | 1970 | 1979 | 1991 | 1995 | 1998 |
|------------|------|------|------|------|------|
| Total | 23.5 | 31.7 | 43.8 | 47.3 | 52.0 |
| Primario | 5.5 | 5.0 | 14.0 | 16.8 | 26.6 |
| Secundario | 19.8 | 24.2 | 33.8 | 29.3 | 38.3 |
| Terciario | 56.8 | 66.7 | 74.7 | 78.9 | 76.3 |

1.2.6. Estado civil fecundidad y empleo.

En la sociedad mexicana, el hecho de que una mujer se case o se una a un hombre sugiere un cambio fundamental en su curso de vida; esto constituye el período de formación de la familia, de procreación y conlleva casi siempre mayores responsabilidades. En México la distribución porcentual de la población femenina no ha variado con el paso de las dos últimas décadas: las mujeres casadas y en unión libre constituyen 54.1, 53.8, 51.84% en 1970, 1980 y 1995 respectivamente, según los censos de población y el Censo de 1995, siguen en orden de importancia las solteras con 36.8, 37.4 y 37% en los mismos años y por último, las viudas, separadas y divorciadas que representan cerca del 9% para las dos primeras fechas y un 10.6% para la última.

A menudo la mayor carga de trabajo doméstica que involucra el vivir en pareja, junto con los obstáculos existentes en el mercado de trabajo (segregación y discriminación) contribuyen para que las mujeres tengan menor oportunidad de situarse en actividades extradomésticas, en comparación con las solteras, viudas, divorciadas y separadas. Sin embargo en épocas recientes la participación de la mujer casada o en unión libre se ha incrementado en forma marcada. Las mujeres unidas entre 20 y 49 años presentan 62% de aumento en su incursión en el mercado de trabajo entre 1976 y 1987; esto es la tasa específica de participación de las unidas paso de 17.4% a 28.2% y para 1995 esta llegó al 32.9%.

Muchos de los análisis recientes apuntan que es la fecundidad la que determina la participación económica de la mujer y no al revés. Se destaca además el efecto limitante del número de hijos sobre el tipo de trabajo que las mujeres desempeñan.¹¹

¹¹ Naciones Unidas, 1987; García y Oliveira, 1989.

En el período 1970 - 1996 se observa que las tasas de participación económica son mayores para las mujeres unidas sin hijos que para las unidas con hijos; además las primeras incrementan de manera importante su presencia en el mercado de trabajo para el período mencionado al pasar de un 28 en 1970 a un 41.9% en 1987, a pesar de una disminución de tres puntos porcentuales en la tasa de participación, que llegó a situarla en 38.9% para 1995. Por otra parte, entre las mujeres con hijos se observa aumentos significativos en la participación económica, aunque sus niveles permanezcan por debajo de las que no tienen. En un primer período (1976-1982), se presenta un ligero incremento en las tasas de participación de las mujeres con hijos al pasar esta de un 16% a un 18.3%. En un segundo período (1982-1987), esta pasó a un 26.8% y en fechas más recientes (1995) se encuentra en 29.2%. En resumen las mujeres con hijos han incrementado su presencia en el mercado de trabajo de 1976 a 1995 en un 82.5%.

Una de las causas más importantes de que las mujeres con niños pequeños se incorporen cada vez más al mercado de trabajo, ha sido el descenso de la fecundidad, en el período señalado, el cual pudo haber liberado tiempo del que usualmente era dedicado al cuidado de los hijos.¹²

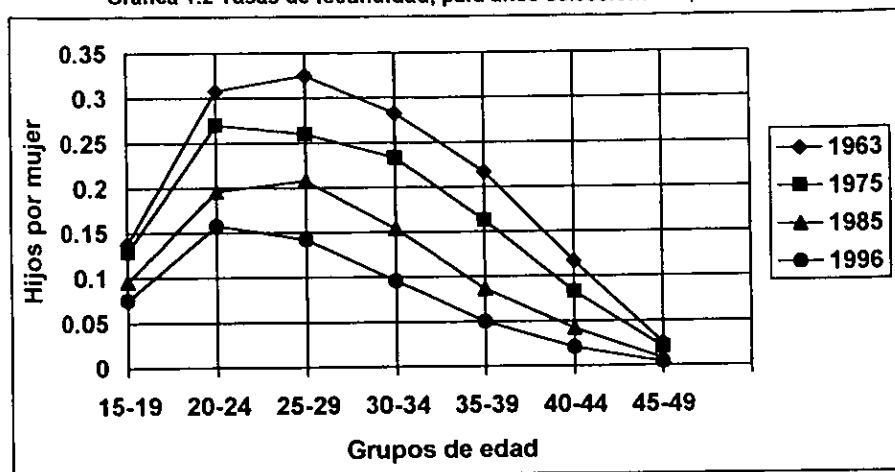
El porcentaje de mujeres unidas que utilizaba algún método de control natal paso de 30.2 a 47.7% entre 1976 y 1982, y se incrementó a 52.7 en 1987¹³, y para 1995 este porcentaje es del 66.5%. Junto con la disminución del nivel reproductivo se constata también algunos cambios en la estructura de la fecundidad. En el período 1963-1975 se advierte un mayor descenso de la fecundidad en el grupo de edad 25-29 años, de modo que al final del período la fecundidad del grupo 20-24 era la más elevada. Es claro el paulatino rejuvenecimiento del perfil por edades de la fecundidad a partir de 1975, el cual se vincula a disminución de nivel reproductivo. Se observa una mayor reducción después de los 25 años que antes de esa edad: por cada 100 mujeres del grupo 20-24 años que procrean hijos, 70 del intervalo 35-39 y 8 del grupo 45-49 años dieron a luz en 1963; en cambio en 1996 sólo 34 y 2 de cada cien respectivamente dieron a luz. A mediados de los ochenta se advierte una notoria reversión de la tendencia al rejuvenecimiento, la cual se encuentra ligada a un freno en el descenso de la fecundidad entre los 25 y 34 años de edad entre

¹² Palma, 1988.

¹³ Bronfman, López y Tuiran, 1986; Cervantes y Sandoval 1988.

1982 y 1986; sin embargo se espera que en el futuro continúe la tendencia de rejuvenecimiento de largo plazo.

Gráfica 1.2 Tasas de fecundidad, para años seleccionados, 1963-1996



Otra de las posibles causas de que las mujeres hallan incrementado su participación económica, ha sido los períodos de crisis que el país ha atravesado, principalmente la recesión de la década de los 80 y principios del presente sexenio. Indiscutiblemente, la fuerte contracción del salario ha hecho indispensable la obtención de varios ingresos para mantener a una familia, elemento que posiblemente ha movilizadado una oferta potencial de mano de obra constituida principalmente por mujeres unidas, las cuales salen al mercado en búsqueda de un trabajo o crean su propio empleo.

1.2.7. Subestimación de la participación económica de la mujer.

Se sabe que las fuentes de información (Censos, conteos, encuestas, etcétera) tienden a subestimar la participación económica de la mujer. Con mucha frecuencia el problema inicia desde la forma en que se plantean las preguntas sobre el tema del trabajo y en quién se selecciona para que las responda (en los censos generalmente responde el jefe de familia, quien comúnmente es hombre).

Existe una serie de ideas que inducen a la subestimación de la condición de actividad de la población; por ejemplo, que trabajar es una actividad sólo cuando se hace fuera del hogar y por una jornada completa; que siempre implica el recibir un salario en moneda, etcétera.

En el subsector de agricultura la subestimación es muy clara, en tanto que para la mujer a su manera de percibir sus obligaciones no existe diferencia entre su contribución al predio familiar (cría de animales, siembra y cosecha, etcétera) y su labor de ama de casa propiamente dicha.

Al definir un rubro que aglomera a los trabajadores familiares no remunerados, se disminuye una parte de la subestimación, si existe una restricción en cuanto a las horas a trabajar para ser considerado como persona económicamente activa, por lo menos deben ser 1/3 de la jornada normal, sin embargo esto se presta a la polémica de fijar el número de horas que constituyen a esta.

Para superarse de esta última causa de subestimación, sería necesario considerar el volumen total de horas trabajadas, para lo cual se tendrían que sumar las correspondientes a cada trabajador, aunque no lleguen a las fijadas por un límite arbitrario. En las tabulaciones se pueden tomar opciones para clasificar a la PEA según número de horas laboradas de acuerdo a investigaciones específicas sobre trabajadores de jornada completa, jornada doble, medio tiempo, etcétera.

Otra práctica que contribuye a la subestimación de la participación económica de las mujeres es el establecer un límite para la edad mínima con el fin de considerar a la persona como activa, valor que muchas veces no corresponde a la realidad. En el caso de América Latina la participación de menores de 12 años es muy alta sobre todo en actividades del campo. En México no es raro encontrar trabajadoras domésticas que no superan la edad de los 12 años.

Otro aspecto que ha contribuido a la subestimación ha sido el volumen de actividades insuficientemente especificadas, que si bien no se trata de un problema de subestimación numérica, si lo es cualitativamente, ya que todo análisis de la estructura de la población

económicamente activa se ve limitado si este grupo es muy grande. Por citar un ejemplo en 1970 el 9.7% de la PEA femenina fue clasificada en el rubro de "actividad insuficientemente especificada".

Una probable explicación a este último punto es que en el trabajo de campo el problema radica en que con frecuencia se registran en actividades insuficientemente especificadas a las personas que tienen más de una ocupación y les es imposible elegir una como principal. Algunos censos captan ocupación secundaria práctica que ha sido muy útil para disminuir la subestimación.

En el caso de las mujeres no es raro que se combinen, es más, en muchos casos tienen carácter de complementarias, aunque desde el punto de vista de clasificaciones pertenezcan a diferentes ocupaciones y ramas de actividad.

El que se den más comúnmente actividades combinadas entre las mujeres que entre los hombres puede obedecer a su mayor movilidad, o la contratación temporal únicamente en estaciones que requieren de mucha mano de obra. Esto por considerar a la fuerza de trabajo femenina como secundaria, por los mercados de trabajo restringidos para la mujer y por la necesidad de combinar sus actividades económicas con las del hogar.

Por otra parte, el registrar las características de la mujer trabajadora sólo con respecto a su actividad principal puede interactuar con el límite de horas, para subestimar el nivel de participación. Puede ser que una trabajadora no reuniera las horas marcadas como mínimas para considerarla económicamente activa en una sola actividad; pero la suma de horas de todas sus actividades sea muy superior a dicho límite.

También existen dificultades con la clasificación por posición en la ocupación, lo cual altera el análisis de la estructura de la población económicamente activa femenina.

1.2.8. Segregación y discriminación de la mujer en el mercado de trabajo.

Es importante antes que nada citar qué se debe entender por segregación y discriminación, en términos sencillos se describe que:

Segregación. Es la acción de separar de un todo, de poner aparte o aislar. En el sentido que interesa. Tendencia de ciertos sectores de opinión, que intentan la separación o aislamiento de los individuos de ciertos grupos (mujeres, ancianos, minorías étnicas, discapacitados, homosexuales, etcétera), haciéndolos objeto de discriminación en los planos social y jurídico. Estas medidas se traducen, principalmente, en prohibir su acceso a lugares de expansión, centros de enseñanza, medios de locomoción, etcétera, declarándolos asimismo ineptos para el ejercicio de determinadas funciones.

Discriminación. Es la acción de distinguir o diferenciar una cosa de otra. Se refiere a la distinción, exclusión o restricción basada en el hecho de "pertener" o no a un grupo determinado de personas, que tenga por resultado el menoscabar o anular el reconocimiento, goce o ejercicio por parte del individuo, independientemente de su estado civil, sobre la base de igualdad del hombre y la mujer de los derechos humanos y las libertades fundamentales en las esferas políticas, económica, social, cultural y civil.

En cuanto a las anteriores definiciones resulta interesante mencionar que dicen las leyes del país.

La Constitución Política de los Estados Unidos Mexicanos señala en su artículo 4º párrafo segundo que:

"El varón y la mujer son iguales ante la Ley. Esta protegerá la organización y el desarrollo de la familia".

El artículo 3º de la Ley Federal del Trabajo en sus párrafos dos y tres establece que:

“El trabajo es un derecho y un deber sociales. No es artículo de comercio, exige respeto para las libertades y dignidad de quien lo presta y debe efectuarse en condiciones que aseguren la vida, la salud y un nivel económico decoroso para el trabajador y su familia”.

“No podrán establecerse distinciones entre los trabajadores por motivo de raza, sexo, edad, credo religioso, doctrina política o condición social”.

Por otra parte el artículo 133 señala que:

“Queda prohibido a los patrones: Negarse a aceptar trabajadores por razón de edad o de sexo”.

Una muestra de segregación por parte de Las leyes hacia las mujeres, lo constituye: El Reglamento de Labores Peligrosas e insalubres, promulgado en 1934. En su conjunto los artículos 3, 13, 17 y 21 prohíben a las mujeres:

- 1. Trabajar en lugares donde se encuentren maquinarias, cuyas partes peligrosas no sean cubiertas debidamente con tapas protectoras.*
- 2. Realizar trabajos subterráneos o submarinos.*
- 3. Sólo durante los tres meses anteriores al parto, desempeñar trabajos que exijan esfuerzos físicos.*
- 4. Trabajar en determinadas plantas, talleres y fabricas donde existan emanaciones nocivas, peligro de envenenamiento, vapores, deletéreos o polvos nocivos o peligrosos. En dichos talleres o lugares de trabajo, las mujeres no pueden trabajar ni entrar a los mismos salvo en el caso de las que por su profesión de químicas o farmacéuticas estuvieran autorizadas para ejercer esas labores.*

La mujer trabajadora desarrolla sus actividades en una supuesta igualdad con el hombre. Igualdad expresamente consagrada tanto en la Constitución, como en la Ley Federal del Trabajo.

No obstante lo anterior, en su vida diaria se enfrenta cada vez en más ocasiones, por las necesidades económicas actuales, a una serie de problemas en el mercado de trabajo.

En la práctica, es común la violación de las leyes laborales cuando se establecen perfiles de trabajo de manera discriminatoria.

El hecho de que la mujer sufra prácticas de segregación y discriminación es debido a razones fundamentalmente sociales y culturales.

La existencia de un modelo patriarcal, recrea valores que excluyen a la mujer de derechos y decisiones y busca someterla a condiciones en ocasiones injustas. Las cualidades económicas existentes en México si bien no minan por si mismas los fundamentos del modelo patriarcal, crean circunstancias favorables para la cada vez más creciente inserción de la mujer en el trabajo.

Una primera forma de segregación se da dentro de la familia ya que a las mujeres, les son asignadas casi exclusivamente las tareas domésticas y la crianza de los hijos. Por consecuencia a la gran concentración de las mujeres en éste tipo de trabajo le corresponde una desigual e inequitativa ubicación en el trabajo extradoméstico.

En cifras se puede citar que a mediados de la década de los 90 sólo un 37.8% de los hombres mayores de 12 años realizaban alguna actividad doméstica, en contraste con el 92.3% de las mujeres.

Resultado de esta desigual distribución de las tareas domésticas, y la no-valoración de la dimensión social de éstas, es que las mujeres sufran de sobrecarga de trabajo en el número total de horas laboradas. Considerando ambos tipos de trabajo ellas exceden en al menos 9.3 horas semanales a los hombres; mientras que entre los casados con bajos niveles de instrucción esta diferencia llega a ser de 14.3 horas.

Las repercusiones de este modo básico de segregación son claras y es importante citarlas.

1. Se impide la igualdad de condiciones en el acceso al trabajo remunerado y refuerza la situación de desventaja social de las mujeres en ambos tipos de trabajo.
2. Que una parte importante de las actividades laborales de las mujeres constituya una extensión de las tareas tradicionalmente consideradas como femeninas, ya sea empleadas domésticas o de comercio, educadoras, enfermeras, etcétera, lo que reduce sus opciones ocupacionales y las mantiene en niveles jerárquicos menores.

Una vez que la mujer tiene la oportunidad de realizar algún tipo de trabajo remunerado, se enfrenta a otro tipo de segregación, la del mercado laboral, ésta consiste en que ahora se le sitúa en su mayoría en actividades bajo el rubro de "trabajadoras por su cuenta" y/o "trabajadoras a tiempo parcial".

Los procesos de crisis y reestructuración económica por los que ha atravesado México en las últimas décadas han ocasionado un aumento significativo de las actividades no asalariadas (trabajo por cuenta propia y ayuda no remunerada), con un impacto relativo mayor sobre las mujeres de la fuerza de trabajo. Las trabajadoras por cuenta propia y no remuneradas pasaron en su conjunto de representar el 26.6% de la PEA femenina en 1970 a un 40.5% en 1996.

Por su propia naturaleza el trabajo por cuenta propia es considerado como "precario" ya que se carece de un contrato laboral, prestaciones y de un sueldo fijo.

Cuando las mujeres trabajan por cuenta propia lo hacen con más frecuencia que los hombres, de manera unipersonal, en labores de subsistencia (la familia consume lo que produce), o que se gana menos de dos salarios mínimos mensuales; y ocupan posiciones extremas en cuanto a las horas que laboran (menos de 15 y más de 48 de horas promedio a la semana).

La creciente participación de las mujeres en el trabajo no asalariado es resultado de múltiples aspectos, entre otros: la mayor participación económica de las casadas con hijos pequeños, y sin escolaridad; la necesidad de ingreso adicional de las familias en el

prolongado período de crisis económica, y las estrategias empresariales de expansión del trabajo a domicilio y de la subcontratación.

La duración de la jornada laboral es el segundo rasgo de diferenciación del trabajo extradoméstico: el tiempo parcial o completo.

En México la proporción de mujeres en el trabajo de tiempo parcial se movió de un 50.1% en 1979, a un 48.6% de la población ocupada femenina en 1996. Resulta de interés evaluar el trabajo femenino a tiempo parcial con relación a el masculino. Para lo cual una revisión de los índices de feminización será útil.

Los datos del país señalan que a mediados de los noventa el trabajo a tiempo parcial es una actividad mucho más feminizada que el de tiempo completo, tanto en los asalariados como en aquellos que trabajan por cuenta propia. En efecto, 107 y 85 mujeres por cada 100 hombres desempeñan actividades de tiempo parcial asalariadas y por cuenta propia, respectivamente. Mientras que en el trabajo de tiempo completo las cifras correspondientes son de 44 y 23 mujeres por cada 100 hombres. En cierta manera, la creciente participación de las mujeres asalariadas de tiempo parcial obedece a la necesidad "ineludible" de hacer compatibles sus roles de madres y trabajadoras.

Un elevado grado, de segregación indica que las mujeres y los hombres se concentran en ocupaciones integradas en su mayoría por miembros de su propio sexo; o lo que es lo mismo, que el género sigue siendo un criterio para crear espacios laborales socialmente diferenciados y jerárquicos.¹⁴

El carácter excluyente de la segregación ocupacional por género resulta patente en varios aspectos: restringe el rango de posibilidades disponibles para las mujeres y las repliega a las ocupaciones de menor prestigio social tipificadas como "femeninas", ofrece pocas perspectivas de movilidad laboral a la vez que alta inestabilidad; genera y refuerza la disparidad salarial; y, por último, limita de manera importante las condiciones de autonomía que el trabajo extradoméstico es capaz de brindar.¹⁵

¹⁴ Reskin, 1984; Reskin y Hartmann, 1986; Blau y Ferber, 1986.

¹⁵ Anker y Hein, 1986; Benería y Roldán, 1987; García, 1997.

Las cifras que se presentaron muestran una alta segregación ocupacional de la fuerza de trabajo en el país a mediados de la presente década, y la necesidad de que alrededor de un 60.6% y un 41.7% respectivamente de los trabajadores por cuenta propia y asalariados de tiempo completo, y un 50% de los de tiempo parcial en ambas categorías, tendría que ser redistribuidos entre las diferentes ocupaciones para lograr una estructura ocupacional igualitaria entre los hombres y mujeres.¹⁶

Hasta aquí se han citado algunas de las formas de segregación a que están expuestas las mujeres en el mundo del trabajo; el acceso limitado y condicionado al mercado laboral, la separación de ellas en actividades por cuenta propia y de tiempo parcial, y la participación mayoritaria en ocupaciones catalogadas como femeninas. Lo que resta de esta sección se ocupará en determinar cuales son las prácticas más comunes de discriminación hacia las mujeres y sus consecuencias.

Las declaraciones de igualdad entre hombres y mujeres existentes en las Leyes del país son insuficientes, pues a pesar de ellas en la sociedad mexicana se siguen practicando discriminaciones laborales frente a las cuales la mujer no tiene ningún medio de defensa efectivo.

Prácticas como la de exigir la presentación de certificados de ingravidez para obtener un empleo, negar la admisión por la situación de gravidez, estado civil o por tener a su cargo el cuidado de los hijos menores, despedir a una trabajadora o coaccionarla directa o indirectamente para que renuncie por estar embarazada, por cambiar de estado civil, retribuir el trabajo con el salario menor al que se pague a otra por una tarea igual, y establecer cualquier tipo de distinción o exclusión basada en el sexo que tenga por objeto o por resultado menoscabar, anular o alterar la igualdad de oportunidades o de trato en el empleo con excepción de las sustentadas en las exigencias particulares de una actividad determinada.

¹⁶ Oliveira, Ariza y Eternod, 1996.

El hecho de que se niegue a la mujer la oportunidad de integrarse al trabajo extradoméstico con base al estado o riesgo de ingravidez, es debido tal vez a las obligaciones que los patrones tienen según la Ley.

El artículo 170 de la ley federal del trabajo cita:

Las madres trabajadora tendrán los siguientes derechos:

- I. Durante el período del embarazo, no realizarán trabajos que exijan esfuerzos considerables y signifiquen un peligro para su salud con relación a la gestación, tales como levantar, tirar o empujar grandes pesos que produzcan trepidación, estar de pie durante largo tiempo o que actúen o puedan alterar su estado psíquico y nervioso;*
- II. Disfrutaran de un descanso de seis semanas anteriores y seis posteriores al parto;*
- III. Los períodos de descanso a que se refiere la fracción anterior se prorrogarán por el tiempo necesario en el caso en que se encuentren imposibilitadas para trabajar a causa del embarazo o del parto*
- IV. En el período de lactancia tendrán dos reposos extraordinarios por día, de media hora cada uno, para alimentar a sus hijos, en el lugar adecuado e higiénico que designe la empresa;*
- V. Durante los períodos de descanso a que se refiere la fracción II, percibirán su salario integro, en los casos de prórroga mencionados en la fracción III, tendrán derecho al cincuenta por ciento de su salario por un período no mayor de sesenta días;*
- VI. A regresar al puesto que desempeñaba, siempre que no haya transcurrido más de un año de la fecha de parto; y*
- VII. A que se computen en su antigüedad los períodos pre y pos natales.*

Estas obligaciones de los patrones se traducen en gastos, motivo por el cual si se tiene la oportunidad de contratar a un hombre en el puesto vacante, se hace, siendo ésta una clara muestra de discriminación.

Por otra parte el hecho de que se condicione el otorgamiento o permanencia de un empleo a una mujer por su estado civil (principalmente a las casadas con hijos) se respalda en ideas tales como:

1. Que la mujer en cualquier momento regresara al "lugar donde le corresponde" o sea su hogar.
2. Que la mujer trabajadora al no poder desligarse de sus tareas domésticas, se traducirá en inseguridad, inadecuación e insuficiente dedicación al trabajo.

Esto tiene como consecuencia que las empresas o patrones, prefieran invertir en los hombres que en las mujeres para los cursos de capacitación y adiestramiento con lo que la mujer se ve limitada en sus posibilidades de ascenso.

Por último, la discriminación salarial es una práctica que consiste en otorgar diferente pago, a trabajos iguales realizados por personas con similar capacitación; lo común es que el hombre reciba más que la mujer. Bajo la creencia de que la mujer sólo trabaja para "darse sus gastos" o si acaso para ayudar en el ingreso familiar.

Las estadísticas muestran que en México existen elevados índices de discriminación salarial en perjuicio de las mujeres principalmente en las actividades asalariadas de tiempo completo. En estas actividades recibían en 1995, 37% menos de salario que los hombres, teniendo iguales niveles de instrucción. En contraste, en las actividades asalariadas de tiempo parcial la diferencia en los salarios es del 13%. Estas cifras ponen de manifiesto que, por lo regular, la discriminación salarial es mayor en las ocupaciones dominadas por los hombres que las dominadas por las mujeres.

2. Análisis de componentes principales.

Con frecuencia el estudio y análisis de problemas, naturales y sociales, se haya vinculado a fenómenos complejos, que requieren para su tratamiento de un considerable número de variables.

Por lo regular los valores medidos a cada individuo de la población o muestra son presentados en tablas bidimensionales que cruzan individuos y variables. Las cuales a simple vista no dan la mayor información.

Tradicionalmente se recurre en primer lugar, a realizar un análisis descriptivo; cuyo objetivo es la comprensión de cada una de las variables por separado. Esto sugiere la obtención de medidas de tendencia central, medidas de dispersión, asimetría, curtosis, representaciones gráficas, comprobación de supuestos paramétricos, detección de casos atípicos, depuración de la tabla de datos.

En segundo lugar, a la búsqueda de relación entre pares de variables, mediante coeficientes de correlación. Esto supone la obtención de las matrices de covarianzas y correlaciones.

En un tercer lugar se podría recurrir a graficar los puntos correspondientes a cada individuo. Esto supondría que el número de variables medidas fue como máximo tres.

Como en la mayoría de los casos no es posible realizar ésto último. El problema que se plantea es encontrar una técnica que permita el tratamiento de todas las variables a la vez.

El análisis de componentes principales permite dar representaciones geométricas de los datos que transforman en distancias euclidianas las proximidades estadísticas entre elementos.

Esto permite utilizar las facultades de percepción cotidianas, sobre los gráficos del análisis de componentes principales, se ve, en el sentido literal de la palabra, agrupaciones, oposiciones, tendencias, imposibles de notar sobre una gran tabla de números, incluso después de un examen prolongado.

Resumiendo. El objetivo del análisis de componentes principales (ACP), aplicado a una tabla de datos, es un estudio exploratorio. Las dos vías principales de esa exploración son:

1. Un balance de las semejanzas entre individuos, se trata de averiguar: ¿Qué elementos se parecen más?, ¿Qué individuos son diferentes?, ¿Existen grupos homogéneos en la población?, ¿Se puede poner en evidencia una tipología de los individuos?.
2. Un balance de las relaciones entre variables, se busca contestar preguntas tales como: ¿Qué variables son correlacionadas entre sí?, ¿De qué forma?, ¿Existen grupos de variables correlacionadas entre sí?, ¿Se puede poner en evidencia una tipología de las variables?.

2.1. Procedimiento del cálculo de los componentes principales.

Sea $X \in \mathfrak{R}^p$ vector aleatorio con alguna función de distribución multivariada, con vector de medias $\mu \in \mathfrak{R}^p$, matriz de varianzas y covarianzas $\Sigma \in \mathfrak{R}^{p \times p}$. Además, supongase que el

tamaño de la población es n .

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

El primer componente de las observaciones es la combinación lineal:

$$Y_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p = a_1'X'$$

donde $a_1' = [a_{11}, a_{21}, \dots, a_{p1}]$ es de dimensión p y de norma 1 es decir $a_1'a_1 = \sum_{h=1}^p a_{h1}^2 = 1$. Y_1 con varianza:

$$\begin{aligned} \text{Var}(Y_1) &= E[(Y_1 - E(Y_1))(Y_1 - E(Y_1))'] \\ &= E[(a_1'X - E(a_1'X))(a_1'X - E(a_1'X))'] \\ &= E[a_1'(X - E(X))(X - E(X))'a_1] \\ &= a_1'E[(X - E(X))'(X - E(X))]a_1 \\ &= a_1'\Sigma a_1 \end{aligned}$$

Se pretende que la varianza de Y_1 sea tan grande como sea posible, sujeta a que el vector de coeficientes a_1 sea de norma 1.

Un método para resolver este problema es mediante el uso de multiplicadores de Lagrange.

La función a maximizar en este caso es: $Var(Y_1) = a_1' \Sigma a_1$ y la restricción:

$$\sum_{h=1}^p a_{h1}^2 = 1$$

Así la ecuación combinada con la restricción es: $G_1(a_1, \lambda_1) = a_1' \Sigma a_1 + \lambda_1(1 - a_1' a_1)$, donde λ_1 es el multiplicador de Lagrange. Derivando $G_1(a_1, \lambda_1)$ con respecto al vector a_1 se obtiene:

$$\begin{aligned} \frac{\partial G_1(a_1, \lambda_1)}{\partial a_1} &= \frac{\partial}{\partial a_1} [a_1' \Sigma a_1 + \lambda_1(1 - a_1' a_1)] \\ &= 2\Sigma a_1 - 2\lambda_1 I_p a_1 \\ &= 2(\Sigma - \lambda_1 I_p) a_1 \end{aligned}$$

Por otra parte derivando con respecto al multiplicador λ_1 se obtiene:

$$\begin{aligned} \frac{\partial G_1(a_1, \lambda_1)}{\partial \lambda_1} &= \frac{\partial}{\partial \lambda_1} [a_1' \Sigma a_1 + \lambda_1(1 - a_1' a_1)] \\ &= (1 - a_1' a_1) \end{aligned}$$

Los coeficientes deben satisfacer el sistema lineal de p ecuaciones:

$$(\Sigma - \lambda_1 I_p) a_1 = 0_p \quad (2.1)$$

Donde I_p es la identidad de $p \times p$ y 0_p es el vector nulo de dimensión p .

Como a_1 debe ser distinto del vector nulo de \mathcal{R}^p ya que $\sum_{h=1}^p a_{h1}^2 = 1$ esto implica que λ_1 se debe de escoger de manera que:

$$|(\Sigma - \lambda_1 I_p)| = 0 \quad (2.2)$$

Esto es que el sistema tenga solución diferente a la trivial. (Aun más, una infinidad de soluciones).

De la expresión (2.1) se observa que λ_1 es un valor característico de la matriz Σ y a_1 es el vector propio asociado a λ_1 . Para determinar cual de las p raíces características debe usarse hay que premultiplicar la expresión (2.1) por el vector a_1' y replantear. Esto es:

$$a_1'(\Sigma - \lambda_1 I_p)a_1 = a_1'0_p$$

$$a_1'\Sigma a_1 - \lambda_1 a_1' a_1 = 0$$

$$\Rightarrow a_1'\Sigma a_1 - \lambda_1 = 0 \text{ ya que } a_1' a_1 = 1$$

$$\Rightarrow \lambda_1 = a_1'\Sigma a_1 = \text{Var}(Y_1)$$

Pero el vector de coeficientes fue tomado para maximizar la varianza de Y_1 , esto implica que λ_1 debe ser el valor característico más grande de la matriz de covarianzas Σ . Se puede resumir este resultado de la siguiente forma.

Definición 2.1 El primer componente principal Y_1 es una combinación lineal de las variables aleatorias X_1, X_2, \dots, X_p de la forma $Y_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p$ cuyos coeficientes a_{hi} son los elementos del vector característico asociado con el más grande valor propio λ_1 de la matriz de covarianzas Σ . Además el vector a_1 es de norma 1 y el valor propio λ_1 es interpretado como la varianza de Y_1 .

El segundo componente principal es la combinación lineal de las variables aleatorias X_1, X_2, \dots, X_p de la forma $Y_2 = a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p$ cuyos coeficientes tienen que ser elegidos, sujetos a las restricciones:

$$a_2^t a_2 = 1 \text{ y } a_1^t a_2 = 0 \quad (2.3)$$

De manera que la varianza de Y_2 sea máxima. La primera restricción garantiza la unicidad de los coeficientes y la segunda la no correlación entre los componentes.

Los coeficientes del segundo componente principal son encontrados, maximizando la varianza de Y_2 sujeta a las restricciones dadas en (2.3), de manera que se tiene la ecuación $G_2(a_2, \lambda_2, \mu) = a_2^t \Sigma a_2 + \lambda_2(1 - a_2^t a_2) + \mu a_1^t a_2$, donde λ_2 y μ son los multiplicadores de Lagrange.

Derivando $G_2(a_2, \lambda_2, \mu)$ respecto a a_2 se obtiene:

$$\begin{aligned} \frac{\partial G_2(a_2, \lambda_2, \mu)}{\partial a_2} &= \frac{\partial}{\partial a_2} [a_2^t \Sigma a_2 + \lambda_2(1 - a_2^t a_2) + \mu a_1^t a_2] \\ &= 2\Sigma a_2 - 2\lambda_2 I_p a_2 + \mu a_1 \\ &= 2(\Sigma - \lambda_2 I_p) a_2 + \mu a_1 \end{aligned}$$

Igualando la última expresión al vector nulo de \mathfrak{R}^p y premultiplicando por a_1^t , se sigue de las cualidades de ortogonalidad y normalización de a_1 que:

$$\begin{aligned} 2a_1^t (\Sigma - \lambda_2 I_p) a_2 + a_1^t \mu a_1 &= a_1^t 0_p \\ 2a_1^t \Sigma a_2 - 2\lambda_2 a_1^t a_2 + \mu a_1^t a_1 &= 0 \\ 2a_1^t \Sigma a_2 + \mu &= 0 \quad (2.4) \end{aligned}$$

De manera similar premultiplicando la ecuación (2.1) por a_2^t se sigue que:

$$a_2'(\Sigma - \lambda_1 I)a_1 = a_2'0_p$$

$$a_2'\Sigma a_1 - \lambda_1 a_2' a_1 = 0$$

$$a_2'\Sigma a_1 = 0$$

$$a_1'\Sigma a_2 = 0 \quad (2.5)$$

De la ecuación (2.4) y (2.5) se concluye que: $\mu = 0$. Y por otro lado que el segundo vector debe satisfacer:

$$(\Sigma - \lambda_2 I_p)a_2 = 0_p$$

Análogamente al primer componente principal los coeficientes del segundo vector son los elementos del vector propio asociado al segundo valor característico más grande.

Resumiendo y generalizando.

Definición 2.2 El j -ésimo componente principal, es una combinación lineal de las variables aleatorias X_1, X_2, \dots, X_p de la forma: $Y_j = a_{j1}X_1 + a_{j2}X_2 + \dots + a_{jp}X_p$ cuyos coeficientes a_{ij} son los elementos del vector propio de la matriz de covarianzas Σ correspondiente al j -ésimo valor propio más grande λ_j . Además si $\lambda_i \neq \lambda_j$, los coeficientes del i -ésimo y del j -ésimo componente son necesariamente ortogonales; además si $\lambda_i = \lambda_j$ los elementos del vector propio pueden ser tomados ortogonales, aunque existe una infinidad de vectores ortogonales. La varianza de la "muestra" acumulado por el j -ésimo componente es λ_j y el total de la varianza de la muestra está dado por $\lambda_1 + \lambda_2 + \dots + \lambda_p$.

Por lo que el porcentaje de varianza acumulada por el j -ésimo componente se calcula mediante la expresión.

$$\frac{\lambda_j}{\sum_{i=1}^p \lambda_i}$$

A continuación se enuncia un resultado que será usado para justificar algunas propiedades de los componentes principales; se trata del "Teorema de la descomposición espectral" o de "Diagonalización". Que dice: si A una matriz de p renglones por p columnas ($p \times p$), simétrica entonces existe una matriz ortogonal P tal que A pueda descomponerse como: $A = P\Lambda P'$ donde Λ es una matriz diagonal de los valores propios de A , y P cumple que $PP' = I_p$ (Ver anexo II).

Con lo expuesto anteriormente se puede dar una definición de los componentes principales en forma matricial.

Definición 2.3 Sea $X \in \mathfrak{R}^p$ un vector aleatorio con vector de medias $\mu \in \mathfrak{R}^p$ y matriz de varianzas y covarianzas Σ , los componentes principales, son la transformación lineal de las variables observadas $\{X_i\}$ de la forma $Y = A'X$.

Donde la j -ésima columna de la matriz A corresponde al vector propio con norma 1 asociado al j -ésimo valor característico de la matriz Σ . Esto es: $A = [a_1 | a_2 | \dots | a_p]$

Además $A'A = I_p$ y $\Sigma = A'\Lambda A$ donde Λ es una matriz diagonal de los valores característicos de Σ tal que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

Hasta aquí se ha visto qué son y cómo se obtienen los componentes principales, ahora será interesante conocer algunas propiedades referentes a ellos. Para lo cual se cita el siguiente resultado.

Teorema 2.1 Sea $X \in \mathfrak{R}^p$ un vector aleatorio con vector de medias $\mu \in \mathfrak{R}^p$ y matriz de varianzas y covarianzas Σ , definida positiva, entonces los componentes principales $Y = A'X$ tienen las siguientes propiedades:

i) $E(Y) = A'\mu$

ii) $Var(Y) = \Lambda$

$$\text{iii) } \text{Var}(Y_1) \geq \text{Var}(Y_2) \geq \dots \geq \text{Var}(Y_p) \geq 0$$

$$\text{iv) } \sum_{j=1}^p \text{Var}(Y_j) = \text{tr}\Sigma$$

$$\text{v) } \prod_{j=1}^p \text{Var}(Y_j) = |\Sigma|$$

Demostración.

$$\text{i) } E(Y_j) = E(a_j'X)$$

$$= a_j'E(X)$$

$$= a_j'\mu \quad J = 1, 2, \dots, p$$

Esto es que $E(Y) = A'\mu$

$$\text{ii) } \text{Var}(Y) = \text{Var}(A'X)$$

$$= \text{Cov}(A'X, A'X)$$

$$= A' \text{Cov}(X, X) A$$

$$= A'\Sigma A \quad \text{pero } \Sigma = A\Lambda A'$$

$$\Rightarrow A'\Sigma A = A'(A\Lambda A')A = (A'A)\Lambda(A'A)$$

$$= I_p \Lambda I_p = \Lambda$$

iii) Como $\text{Var}(Y_1) = \lambda_1, \text{Var}(Y_2) = \lambda_2, \dots, \text{Var}(Y_p) = \lambda_p$ y además $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Es inmediato el resultado

$$\text{iv) } \text{tr}(\Sigma) = \text{tr}(A\Lambda A') = \text{tr}(A'A\Lambda) = \text{tr}(\Lambda) = \sum_{j=1}^p \lambda_j = \sum_{j=1}^p \text{Var}(Y_j)$$

$$\text{v) } |\Sigma| = |A\Lambda A'|$$

$$= |A\Lambda| |A'|$$

$$= |A'| |A\Lambda|$$

$$= |A' A \Lambda|$$

$$= |I_p \Lambda|$$

$$= |\Lambda|$$

$$= \prod_{j=1}^p \lambda_j$$

$$= \prod_{j=1}^p \text{Var}(Y_j); \text{ ya que } \Lambda \text{ es diagonal.}$$

Para seguir con las propiedades de los componentes principales, se citan a continuación, resultados que involucran la correlación entre las variables observadas $\{X_i\}$ y los componentes principales $\{Y_j\}$.

Sean $X \in \mathbb{R}^p$ con vector de medias $\mu \in \mathbb{R}^p$ y matriz de varianzas y covarianzas Σ , $Y = A'X$ la transformación de los componentes principales. La matriz de varianzas y covarianzas entre el vector aleatorio X y el vector aleatorio de los componentes principales Y está dada por:

$$\begin{aligned}
\text{Cov}(X, Y) &= \text{Cov}(X, A'X) \\
&= E[(X - E(X))(A'X - E(A'X))'] \\
&= E[(X - E(X))(A'(X - E(X)))'] \\
&= E[(X - E(X))(X - E(X))'(A)'] \\
&= E[(X - E(X))(X - E(X))'A] \\
&= \text{Cov}(X, X)A \\
&= \Sigma A \\
&= A\Lambda A' \\
&= A\Lambda
\end{aligned}$$

De esta última expresión se concluye que la covarianza entre la variable aleatoria X_i y el j -ésimo componente principal está dado por:

$$\text{Cov}(X_i, Y_j) = a_{ij}\lambda_j.$$

Sean $\{\sigma_{ij}\}$ $i, j = 1, 2, \dots, p$ los elementos de la matriz de varianzas y covarianzas Σ , la correlación entre la variable X_i y el j -ésimo componente principal Y_j está dado por:

$$r_{ij} = \frac{a_{ij}\lambda_j}{\sqrt{\text{Var}(X_i)} \sqrt{\text{Var}(Y_j)}}$$

$$= \frac{a_{ij} \sqrt{\lambda_j}}{\sqrt{\sigma_{ii}}}$$

En términos sencillos se interpreta a la correlación r_{ij}^2 como la proporción de variabilidad de la variable aleatoria X_i explicada por el j -ésimo componente principal Y_j . Esto es:

$$r_{ij}^2 = \frac{a_{ij}^2 \lambda_j}{\sigma_{ii}}$$

Si se toma un subconjunto de componentes principales, éste acumula una proporción de variabilidad explicada de X_i . Esta proporción se puede calcular mediante:

$$\sum_{j=1}^r r_{ij}^2 = \frac{\sum_{j=1}^r a_{ij}^2 \lambda_j}{\sigma_{ii}} \quad \text{con } r \leq p \quad (2.6)$$

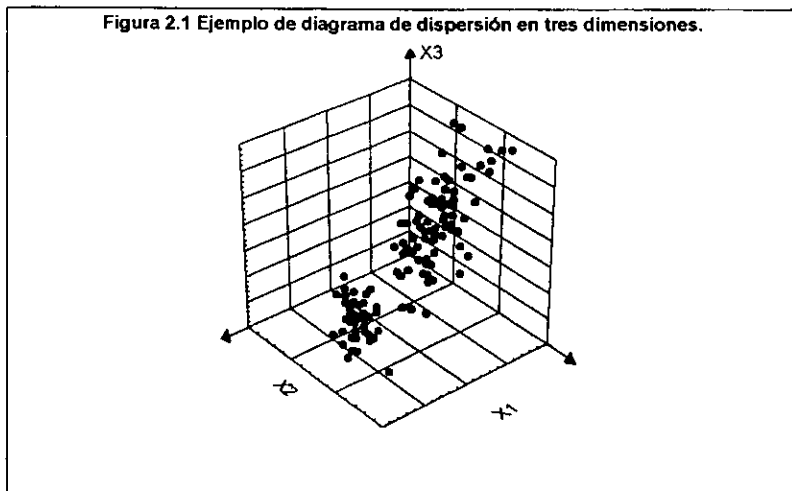
Este resultado se justifica debido a que de *ii*) del teorema 2.1; los componentes principales son variables aleatorias no correlacionadas al tener como matriz de varianzas y covarianzas a Λ que es una matriz diagonal.

El denominador de la expresión (2.6) representa la variabilidad de X_i que va a ser explicada, y el numerador indica la variabilidad acumulada por los primeros r componentes principales.

2.2. Sentido geométrico de los componentes principales.

Hasta ahora se ha dado solamente un tratamiento analítico a los componentes principales como combinaciones lineales de las variables aleatorias $\{X_i\}$, los cuales explican progresivamente pequeños porcentajes del total de la variabilidad de la muestra. Ahora se dará una interpretación geométrica de los componentes, como las variables correspondientes a los ejes principales de la dispersión de las observaciones en el espacio.

Supóngase que se toma una muestra de n individuos a los cuales se les midieron 3 características. El origen de los ejes X_1, X_2, X_3 son las medias muestrales. La nube de puntos generalmente adopta una forma elipsoidal, con eje mayor Y_1 .



Sea α_1 el ángulo que forma el eje Y_1 con el eje X_1 , α_2 el que forma Y_1 con X_2 y α_3 el que forma Y_1 con X_3 .

Si Y_1 pasa por el punto de las medias muestrales, esta orientación es completamente determinada por los cosenos directores, por lo tanto:

$$\cos \alpha_1 = a_{11} \quad \cos \alpha_2 = a_{21} \quad \cos \alpha_3 = a_{31}$$

Donde $a_{11}^2 + a_{21}^2 + a_{31}^2 = 1$. Es conocido de la geometría analítica que el valor de las observaciones $\{(x_{i1}, x_{i2}, x_{i3})\}$ en las nuevas coordenadas con eje principal Y_1 estarán dadas por:

$$y_{i1} = a_{11}(x_{i1} - \bar{x}_1) + a_{21}(x_{i2} - \bar{x}_2) + a_{31}(x_{i3} - \bar{x}_3)$$

Donde $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ $j = 1, 2, 3$. (media muestral de la j -ésima variable).

Debe notarse que la esperanza de la variable Y_1 está dada por:

$$\begin{aligned} E(Y_1) &= \frac{1}{n} \sum_{i=1}^n y_{i1} \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^3 a_{j1} (x_{ij} - \bar{x}_j) \\ &= \frac{1}{n} \sum_{i=1}^n [a_{11}(x_{i1} - \bar{x}_1) + a_{12}(x_{i2} - \bar{x}_2) + a_{13}(x_{i3} - \bar{x}_3)] \\ &= \frac{1}{n} \left[\sum_{i=1}^n a_{11}(x_{i1} - \bar{x}_1) + \sum_{i=1}^n a_{12}(x_{i2} - \bar{x}_2) + \sum_{i=1}^n a_{13}(x_{i3} - \bar{x}_3) \right] = 0 \end{aligned}$$

Por definición se sabe que $Var(X) = E((X - E(X))^2)$ donde X es una variable aleatoria cualquiera. En el caso de Y_1 se tiene que $Var(Y_1) = E(Y_1^2)$ ya que $E(Y_1) = 0$.

El problema pues que queda planteado es encontrar: $a'_1 = [a_{11}, a_{21}, a_{31}]$ que determine el eje mayor que pase por la dirección de máxima varianza de la nube de puntos.

En el presente caso de tres variables la varianza está dada por:

$$\begin{aligned} Var(Y_1) &= \frac{1}{n} \sum_{i=1}^n y_{i1}^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left[\sum_{j=1}^3 a_{j1} (x_{ij} - \bar{x}_j) \right]^2 \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})' \\ &= \frac{1}{n} \sum_{i=1}^n a'_1 (X_i - \bar{X}) (X_i - \bar{X})' a_1 \end{aligned}$$

$$\begin{aligned}
&= a_1' \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' \right] a_1 \\
&= a_1' \Sigma a_1 \qquad (2.7)
\end{aligned}$$

Donde X_i es el vector columna correspondiente a las p variables medidas al individuo i .

Restringiendo $Var(Y_1)$ a que: $a_{11}^2 + a_{21}^2 + a_{31}^2 = 1$

Por lo anterior, los ángulos de Y_1 deben encontrarse derivando la expresión (2.7) respecto a a_1 sujeta a la restricción $\sum_{j=1}^3 a_{j1}^2 = 1$. La solución es entonces el vector propio asociado al valor característico más grande de la matriz de varianzas y covarianzas y Y_1 corresponde pues al primer componente principal.

Este último resultado es válido para el caso general de p variables.

Sean los cosenos directores del primer eje principal Y_1 las coordenadas del vector:

$$a_1' = [a_{11}, a_{21}, \dots, a_{p1}] \text{ con la restricción } a_1' a_1 = 1.$$

La varianza de las proyecciones está dada por:

$$\begin{aligned}
Var(Y_1) &= \frac{1}{n} \sum_{i=1}^n y_{i1}^2 \\
&= \frac{1}{n} \sum_{i=1}^n \left[\sum_{j=1}^p a_{j1} (x_{ij} - \bar{x}_j) \right]^2 \\
&= \frac{1}{n} \sum_{i=1}^n \left[(X_i - \bar{X})' a_1 \right]^2 \\
&= \frac{1}{n} \sum_{i=1}^n a_1' (X_i - \bar{X})(X_i - \bar{X})' a_1
\end{aligned}$$

$$\begin{aligned}
&= a_1' \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' \right] a_1 \\
&= a_1' \Sigma a_1
\end{aligned}$$

Introduciendo la restricción. La ecuación de Lagrange que resulta es: $G_1(a_1, \lambda_1) = a_1' \Sigma a_1 + \lambda_1(1 - a_1' a_1)$ donde λ_1 es el multiplicador de Lagrange.

Que es justamente la expresión que se utilizó para encontrar los coeficientes del primer componente principal (ver pagina 40).

Los restantes valores propios y vectores característicos de Σ determinan el tamaño y orientación de el segundo más grande eje. Si dos valores propios sucesivos λ_i y λ_{i+1} son iguales, la configuración de la dispersión no tiene un único eje mayor siguiente a λ_{i-1} en este plano generado por los vectores asociados a λ_i y λ_{i+1} la nube de puntos es más circular que elíptica.

Con lo anterior se puede dar la siguiente definición geométrica de los componentes principales.

Definición 2.4 Los componentes principales de una muestra de n observaciones, p -dimensionales son las nuevas variables especificadas por los ejes de una rotación rígida del sistema de coordenadas de las variables originales en una orientación correspondiente a la dirección de máxima varianza en la dispersión de la muestra. Los cosenos directores de los nuevos ejes son los vectores característicos de norma 1 correspondientes a los sucesivamente más pequeños valores propios de la matriz de varianzas y covarianzas Σ . Además si dos o más valores propios son iguales las direcciones de los ejes asociados no son únicas y pueden ser tomadas en una infinidad de posiciones ortogonales; si los componentes son calculados mediante la matriz de correlaciones, la misma interpretación geométrica permanece, aunque las coordenadas del sistema de las variables originales es expresada en unidades estándares de medias cero y varianzas unitarias.

Otra propiedad de los "ejes componentes", es el hecho de que la elección de los nuevos ejes coordenados es tal que la suma de los cuadrados de las distancias de cada punto a su proyección sobre los ejes sucesivos será mínima.

En el caso general el cuadrado de la i -ésima distancia está dado por:

$$\begin{aligned}
 d^2(P'_i, P_i) &= d^2(0, P_i) - d^2(0, P'_i) \\
 &= \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2 - [y_{i1} - E(Y_1)]^2 \\
 &= \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2 - \left[\sum_{j=1}^p a_{j1} (x_{ij} - \bar{x}_j) \right]^2 \quad (2.8)
 \end{aligned}$$

Donde P'_i es la proyección de P_i sobre el eje Y_1 .

Minimizar la suma $\sum_{i=1}^n d^2(P'_i, P_i)$ de todos los cuadrados de las distancias a los nuevos ejes es equivalente a maximizar el segundo término de la expresión (2.8), que es proporcional a la varianza del componente i -ésimo.

Esto es maximizar la suma de cuadrados:

$$\begin{aligned}
 \sum_{i=1}^n \left[\sum_{j=1}^p a_{j1} (x_{ij} - \bar{x}_j) \right]^2 &= \sum_{i=1}^n \left[(X_i - \bar{X})' a_1 \right]^2 \\
 &= \sum_{i=1}^n a_1' (X_i - \bar{X}) (X_i - \bar{X})' a_1 \\
 &= a_1' \left[\sum_{i=1}^n (X_i - \bar{X}) (X_i - \bar{X})' \right] a_1 \\
 &= n a_1' \Sigma a_1 \quad \text{Sujeta a la restricción: } a_1' a_1 = 1
 \end{aligned}$$

La solución a este problema se ha visto que es elegir a a_1 como el vector propio asociado al valor característico más grande de Σ .

Por otra parte la distancia del punto P a la proyección P' sobre el segundo eje componente

está dada por:

$$\sum_{j=1}^p (x_{ij} - \bar{x}_j)^2 - \sum_{i=1}^n \left[\sum_{j=1}^p a_{j2} (x_{ij} - \bar{x}_j) \right]^2$$

minimizar esta distancia es equivalente a maximizar la suma de cuadrados:

$$\sum_{i=1}^n \left[\sum_{j=1}^p a_{j2} (x_{ij} - \bar{x}_j) \right]^2 = na_2^t \Sigma a_2 \quad \text{Sujeta a las restricciones: } a_2^t a_2 = 1 \quad \text{y} \quad a_1^t a_2 = 0$$

Cuya solución según se ha visto es: a_2 el vector característico asociado al segundo valor propio de Σ más grande.

Por lo tanto. La suma de los cuadrados de las distancias de cada punto a su proyección sobre los ejes componentes es mínima.

Si el primer valor propio, es notoriamente más grande que los $p - 1$ restantes, una excelente aproximación a los coeficientes de el primer componente pueden ser obtenidos de la recta que une a los dos puntos extremos de la nube de puntos. Sean $x_{(1)}$ y $x_{(n)}$ las observaciones separadas por la distancia más grande en la nube de puntos dada por:

$$d^2 = \sum_{j=1}^p (x_{(n)j} - x_{(1)j})^2$$

Los cosenos directores de la recta que determinan están dados por:

$$a_j = \frac{x_{(N)j} - x_{(1)j}}{d}$$

donde d es la raíz cuadrada de la distancia de $x_{(1)}$ a $x_{(N)}$.

2.3. Determinación del número de componentes principales a usar.

El ACP frecuentemente es usado con el propósito de generar un conjunto reducido de variables que expliquen la mayor variabilidad posible de los datos originales, y que puedan ser usados en subsecuentes análisis

Se debe por lo tanto justificar cuantos componentes retener. Desafortunadamente no existen métodos que sean aceptados universalmente para hacer esto. La decisión en muchos de ellos es en gran parte de ellos subjetiva.

Un número considerable de procedimientos para determinar cuantos componentes retener han sido sugeridos, a continuación se citan algunos.

1) El cálculo de la variabilidad explicada de los datos originales, por los primeros m componentes principales, es un primer criterio, y tal vez el más simple. Ya que quien realiza el análisis, elige el número de componentes que va a retener en función de la variabilidad explicada mediante:

$$\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j}$$

Si los componentes principales fueron extraídos a partir de la matriz de correlaciones se sugieren los dos siguientes métodos.

2) Se elige a los componentes principales que tengan asociados a los valores propios más grandes o iguales a 1. La razón para este criterio es que el componente retenido debe explicar mayor variabilidad que cualquier variable estandarizada.

3) Los valores propios de cada componente son gráficos en orden sucesiva de su extracción, se seleccionan sólo los componentes que quedan antes del punto de inflexión. Esta prueba es llamada "scree test".¹

4) Si los valores propios, posteriores al m -ésimo son iguales, no hay ejes principales a partir del eje $m + 1$, en el sentido de que no hay direcciones de máxima variabilidad. La variabilidad en las $(n - m)$ dimensiones es esférica.² Para decidir lo anterior se debe probar la hipótesis:

$$H_0 : \lambda_{m+1} = \lambda_{m+2} = \dots = \lambda_p$$

Si esta hipótesis es cierta entonces el estadístico es:

$$\chi^2 = -(n-1) \sum_{j=m+1}^p \log(l_j) + (p-m)(n-1) \log \left\{ \frac{1}{p-m} \sum_{j=m+1}^p l_j \right\}$$

Se distribuye asintóticamente χ^2 con $r = \frac{1}{2}(n-p)(n-p+1) - 1$ grados de libertad, si el número de individuos N es "grande".

Donde l_j representa el valor propio calculado sobre la matriz de varianzas y covarianzas muestral.

La regla de decisión está dada por:

i) Si $\chi^2 > \chi^2_{(r,\alpha)}$ se rechaza la hipótesis nula con un nivel de significancia α .

ii) Si $\chi^2 \leq \chi^2_{(r,\alpha)}$ no se rechaza la hipótesis nula con un nivel de significancia α .

Realizando esta prueba, el criterio de decisión sería, quedarse con los primeros m componentes principales si no se rechazó H_0 .

¹ Propuesto por Catell (1966).

² Propuesto por Anderson (1963).

Las tres siguientes pruebas obedecen a una concepción más empírica que racional del problema. La formulación matemática de lo que pretenden demostrar está probablemente justificada en términos de inferencia estadística.

5) La prueba del bastón roto.³ La variabilidad total a explicar está dada por:

$$\sum_{j=1}^p \lambda_j = V$$

Supóngase que se tiene un segmento de recta de longitud V , y que se divide en p trozos, por $p - 1$ puntos trazados al azar.

Sean: $L_1 \geq L_2 \geq \dots \geq L_p$ los trozos ordenados de acuerdo a su longitud. Se demuestra que los valores medios son:

$$E(L_p) = \frac{1}{p^2} \quad E(L_{p-1}) = \frac{1}{p} \left(\frac{1}{p} + \frac{1}{p-1} \right)$$

Suponiendo $V = 1$ para estandarizar el problema y en general: $E(L_j) = \frac{1}{p} \sum_{i=0}^{p-j} \frac{1}{j+i}$

Si se expresan estos valores medios, cuya suma es 1, en porcentajes de la longitud total, se obtiene el modelo teórico de la descomposición de la varianza en p componentes obtenidos al azar.

Los m primeros componentes se retienen si explican claramente mayor varianza que los m primeros valores medios del modelo teórico. Se considera que los demás componentes descomponen la varianza residual al azar.

6) La idea en que se sustenta esta prueba es que a partir de cierta dimensión la restante variabilidad de los datos originales es debida a causas aleatorias. En esencia, debidas a fluctuaciones del muestreo. Asimilando el "ruido" a variables independientes, la significancia

³ También se le conoce como criterio de Frontier (1974).

de la dimensión m queda resuelta cuando la varianza explicada supera claramente a la varianza explicada por el ruido.⁴

Se sabe que la varianza explicada por los primeros m componentes principales es:

$$V_m = \lambda_1 + \lambda_2 + \dots + \lambda_m$$

La prueba consiste en realizar k análisis sobre p variables independientes para un tamaño muestral N . Ordenando las varianzas explicadas, se tiene.

$$V_m^{(i_1)} < V_m^{(i_2)} < \dots < V_m^{(i_k)}$$

La probabilidad de que se de una cierta ordenación es: $\frac{1}{k!}$

Considerando el evento: "La varianza explicada por el k -ésimo análisis supera la varianza de los demás", es decir:

$$V_m^{(i_1)} < V_m^{(i_2)} < \dots < V_m^{(i_{k-1})} < V_m^{(i_k)}$$

Ya que se pueden formar $(k-1)!$ permutaciones de $\{1, \dots, k-1\}$ la probabilidad de este evento es: $\frac{1}{k}$

Sea $\varepsilon = 0.05$ el nivel de significancia y V_m la varianza explicada por el análisis real. Sea $k-1 = 19$ el número de análisis generados al azar. Si V_m proviene de variables aleatorias independientes, la probabilidad de que supere a las varianzas explicadas por los análisis simulados es $\frac{1}{20} = 0.05$. De este modo se tiene una prueba no paramétrica. Si V_m supera a la varianza explicada por los 19 análisis simulados, se puede afirmar, con probabilidad de error 0.05, que la dimensión m es significativa en el sentido dado anteriormente.

⁴ Prueba de Lebart y Fenelon (1973)

En general para decidir si se retienen m componentes principales con un nivel de significancia ε es necesario realizar $k \approx \frac{1}{\varepsilon}$ análisis y que V_m cumpla con:

$$V_m^{(1)} < V_m^{(2)} < \dots < V_m^{(m-1)} < V_m$$

7) Esta prueba consiste en añadir una variable E formada por datos generados al azar a las p variables observadas en el problema.⁵ Se repite entonces el ACP si después de la componente $m+1$ E queda resaltada en la estructura factorial (la saturación de E en la componente $m+1$ es alta), el número significativo de componentes no puede ser superior a m , pues los demás componentes explicarían una variabilidad inferior a la que es debida a la variable arbitraria E .

2.4. Representación de los datos.

Para satisfacer el objetivo de describir, a los individuos más semejantes, o bien la formación de grupos, el graficar los primeros m componentes principales ($m = 1, 2$ o 3) ayuda a visualizarlo de manera más sencilla.

Dados los n individuos a los cuales se les han medido p características (variables $\{X_i\}$), se propone realizar una representación canónica sobre la tabla de datos.

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

Los individuos originalmente se sitúan en una configuración de n puntos en \mathfrak{R}^p . Para representarlos en un espacio de dimensión $m \leq p$, con dispersión máxima se recurre a aplicar la transformación de los componentes principales a la tabla de datos (D) esto es:

⁵ Prueba E de Ibáñez (1973).

$$Y = A'X$$

Donde Y adopta la forma:

$$Y = \begin{bmatrix} y_{11} & y_{21} & \cdots & y_{n1} \\ \vdots & \vdots & \vdots & \vdots \\ y_{1m} & y_{2m} & \cdots & y_{nm} \\ \vdots & \vdots & \vdots & \vdots \\ y_{1p} & y_{2p} & \cdots & y_{np} \end{bmatrix}$$

En seguida se decide cuantos componentes han de retenerse (según los criterios ya descritos). Supóngase que se eligen m .

Dado lo anterior se tiene a los n individuos representados en un espacio m -dimensional, con dispersión máxima.

La dispersión de los datos de cada renglón va disminuyendo de arriba hacia abajo. Además si $r = \min(p, n - 1)$ es inferior a p , a partir del renglón $m + 1$, los elementos de los renglones son exactamente iguales.

3. Análisis discriminante.

El problema de clasificar surge cuando se miden ciertas características (variables) a un individuo y se decide clasificarlo a uno y sólo uno de los grupos que se tienen con base a los valores de sus características. Por lo general no se puede identificar a simple vista a que grupo pertenece, pero se puede hacer uso de sus medidas. En muchos casos puede ser asumido que hay un número finito de grupos o categorías de los cuales puede provenir el individuo y cada población es caracterizada por una distribución de probabilidad de las medias. Así, un individuo es considerado como una observación aleatoria de su población. El problema es pues: Dado un individuo con ciertas medidas, descubrir de que población proviene.

El llevar a cabo una clasificación puede ser considerado un problema de "funciones estadísticas de decisión". Se tiene un número de hipótesis: cada una de ellas es que la distribución de la observación está dada por una de las funciones y rechazar las otras.

En algunas ocasiones, las poblaciones son especificadas de antemano en el sentido de que la distribución de probabilidad de las medidas es completamente conocida. En otros casos la

forma de cada una de ellas pueda ser conocida, pero los parámetros tengan que ser estimados a partir de una muestra de la población.

Ejemplos donde se hace necesaria una clasificación.

1. Ciertas larvas de crustáceos tienen diferentes estadios de crecimiento. El estadio 1 (Nauplius) y el estadio 2 (Metanauplius) son difícilmente identificables. Se trata de asignar una larva a partir de la longitud y anchura de su cuerpo a uno de los estadios.
2. Para otorgarse un crédito bancario, los solicitantes son clasificados de acuerdo a su tipo de riesgo en: alto, medio o bajo. Un cierto cliente se asignará a uno de los anteriores grupos de acuerdo a: su ingreso, egresos, dependencia económica, propiedades que posee.

3.1. Planteamiento del objetivo principal del análisis discriminante.

Considérense g poblaciones o grupos denotados por $\Pi_1, \Pi_2, \dots, \Pi_g$ $g \geq 2$. Supóngase que con cada población Π_j , está asociada una función de densidad de probabilidad $f_j(X)$ en \mathbb{R}^p , tal que si un individuo proviene de la población Π_j , el tiene función de densidad de probabilidad (f.d.p) a $f_j(X)$, entonces el objetivo principal del análisis discriminante es asignar un individuo a uno de esos g grupos con base a sus medidas. Además es deseable cometer tan pocos errores como sea posible al hacer las asignaciones.

Supóngase que al nuevo individuo a asignarse se le midieron p características (variables) y éstas son presentadas en el vector: $X = (X_1, X_2, \dots, X_p)$ donde X_1, X_2, \dots, X_p son variables aleatorias.

La clasificación puede desarrollarse bajo diferentes perspectivas, que en algunos casos dan lugar a la misma regla de decisión. En general la clasificación se lleva a cabo construyendo ciertas funciones de las variables.

$$F = g(X_1, X_2, \dots, X_p)$$

Llamadas funciones discriminantes, la decisión se toma a partir del valor:

$$f = g(x_1, x_2, \dots, x_p)$$

En algunos casos, tales funciones definen regiones disjuntas R_1, R_2, \dots, R_g del espacio euclídeano \mathfrak{R}^p . Donde $\bigcup_{i=1}^g R_i = \mathfrak{R}^p$, de modo que el criterio de decisión es:

$$X \in \Pi_j \text{ si } X \in R_j$$

Para $j = 1, \dots, g$ la discriminación sera más exacta si Π_j tiene mayor probabilidad que las otras concentrada en R_j , para cada j .

3.2. Discriminación cuando las poblaciones son conocidas (distribuciones y parámetros).

En esta sección se presenta el caso más difícil de encontrar en la practica; el conocer la distribución y los parametros de las poblaciones involucradas en el estudio.

3.2.1 La regla discriminante por máxima verosimilitud

Supóngase que las variables X_1, X_2, \dots, X_p tienen una distribución de probabilidad conocida $f_1(X)$ en $\Pi_1, f_2(X)$ en $\Pi_2, \dots, f_g(X)$ en Π_g . La probabilidad o verosimilitud de la observación (x_1, x_2, \dots, x_p) en Π_j es:

$$f_j(x_1, x_2, \dots, x_p)$$

Entonces una regla razonable de clasificación es asignar X a la población Π_j para la cual la verosimilitud de la observación es mayor.

Definición 3.1 La regla discriminante por máxima verosimilitud para asignar una observación X a una de $\Pi_1, \Pi_2, \dots, \Pi_g$ poblaciones, es asignar X a la población con la cual la verosimilitud de la observación es mayor. Esto es:

$$X \in \Pi_j \text{ si } L_j(X) = \max_i L_i(X) \quad i = 1, \dots, g$$

Teorema 3.1 a) Si Π_i es una población con distribución normal p -variada con vector de medias μ_i y matriz de varianzas y covarianzas Σ , esto es $N_p(\mu_i, \Sigma)$, $i = 1, \dots, g$ y Σ es definida positiva, entonces la regla discriminante de máxima verosimilitud asigna X a Π_j , donde $j \in \{1, \dots, g\}$ es aquel valor que minimiza el cuadrado de la distancia dada por ¹: $(X - \mu_i)' \Sigma^{-1} (X - \mu_i)$ b) Cuando $g = 2$, la regla asigna X a Π_1 si: $\alpha'(X - \mu) \geq 0$ y a Π_2 en otro caso. Donde $\alpha = \Sigma^{-1}(\mu_1 - \mu_2)$ y $\mu = \frac{1}{2}(\mu_1 + \mu_2)$

Demostración.

a) La verosimilitud de la i -ésima población es:

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

como $|2\pi\Sigma|^{-\frac{1}{2}}$ es constante. Basta con maximizar: $\exp\left\{-\frac{1}{2}(X - \mu_i)' \Sigma^{-1} (X - \mu_i)\right\}$ (3.1)

Ahora como Σ es definida positiva entonces Σ^{-1} también es definida positiva. Por lo tanto

$$(X - \mu_i)' \Sigma^{-1} (X - \mu_i) \geq 0 \quad (3.2)$$

Esto es maximizar (3.1) equivale a minimizar (3.2) que es la distancia de Mahalanobis al

¹ Esta distancia estadística fue introducida por Mahalanobis (1936). Para más información revisar el anexo A-II.

cuadrado. De ahí que debe hacerse $i = j$ donde j es el valor que minimiza la distancia.

b) Nótese que:

$$L_1(X) \geq L_2(X) \text{ si y sólo si } (X - \mu_1)' \Sigma^{-1} (X - \mu_1) \leq (X - \mu_2)' \Sigma^{-1} (X - \mu_2)$$

$$(X' - \mu_1') \Sigma^{-1} (X - \mu_1) \leq (X' - \mu_2') \Sigma^{-1} (X - \mu_2)$$

$$X' \Sigma^{-1} X - X' \Sigma^{-1} \mu_1 - \mu_1' \Sigma^{-1} X + \mu_1' \Sigma^{-1} \mu_1 \leq X' \Sigma^{-1} X - X' \Sigma^{-1} \mu_2 - \mu_2' \Sigma^{-1} X + \mu_2' \Sigma^{-1} \mu_2$$

$$-2X' \Sigma^{-1} \mu_1 + \mu_1' \Sigma^{-1} \mu_1 \leq -2X' \Sigma^{-1} \mu_2 + \mu_2' \Sigma^{-1} \mu_2$$

$$(2\mu_1' \Sigma^{-1} X - \mu_1' \Sigma^{-1} \mu_1) + (-2\mu_2' \Sigma^{-1} X + \mu_2' \Sigma^{-1} \mu_2) + (\mu_2' \Sigma^{-1} \mu_1 - \mu_1' \Sigma^{-1} \mu_1) \geq 0$$

$$2\left(\mu_1' \Sigma^{-1} X - \frac{1}{2} \mu_1' \Sigma^{-1} \mu_1 + -\mu_2' \Sigma^{-1} X + \frac{1}{2} \mu_2' \Sigma^{-1} \mu_2 + \frac{1}{2} \mu_2' \Sigma^{-1} \mu_1 - \frac{1}{2} \mu_1' \Sigma^{-1} \mu_1\right) \geq 0$$

$$\left(\mu_1' \Sigma^{-1} X - \frac{1}{2} \mu_1' \Sigma^{-1} \mu_1 - \frac{1}{2} \mu_1' \Sigma^{-1} \mu_2\right) - \left(\mu_2' \Sigma^{-1} X - \frac{1}{2} \mu_2' \Sigma^{-1} \mu_2 - \frac{1}{2} \mu_2' \Sigma^{-1} \mu_1\right) \geq 0$$

$$\mu_1' \left(\Sigma^{-1} X - \frac{1}{2} \Sigma^{-1} \mu_1 - \frac{1}{2} \Sigma^{-1} \mu_2\right) - \mu_2' \left(\Sigma^{-1} X - \frac{1}{2} \Sigma^{-1} \mu_2 - \frac{1}{2} \Sigma^{-1} \mu_1\right) \geq 0$$

$$(\mu_1' - \mu_2') \Sigma^{-1} \left(X - \frac{1}{2} \mu_1 - \frac{1}{2} \mu_2\right) \geq 0$$

$$(\mu_1 - \mu_2)' \Sigma^{-1} (X - \mu) \geq 0$$

$$\alpha' \Sigma^{-1} (X - \mu) \geq 0$$

Por lo tanto:

$$X \in \Pi_1 \text{ si } a' \Sigma^{-1} (X - \mu) \geq 0$$

$X \in \Pi_2$ en otro caso.

Teorema 3.2 a) Si Π_i es una población con distribución Normal p -variada con vector de medias μ_i y matriz de varianzas y covarianzas Σ_i , esto es $N_p(\mu_i, \Sigma_i)$, $i = 1, \dots, g$ y Σ_i es definida positiva, entonces la regla discriminante de máxima verosimilitud asigna X a Π_j , donde $j \in \{1, \dots, g\}$ si: $\ln|2\pi\Sigma_j| + (X - \mu_j)' \Sigma_j^{-1} (X - \mu_j) \leq \ln|2\pi\Sigma_k| + (X - \mu_k)' \Sigma_k^{-1} (X - \mu_k) \quad \forall k \neq j$

Demostración.

Supóngase que: $L_j(X) \geq L_k(X) \quad \forall k \neq j$

$$|2\pi\Sigma_j|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(X - \mu_j)' \Sigma_j^{-1} (X - \mu_j)\right\} \geq |2\pi\Sigma_k|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(X - \mu_k)' \Sigma_k^{-1} (X - \mu_k)\right\}$$

Tomando logaritmos de ambos lados de la desigualdad.

$$-\frac{1}{2} \ln|2\pi\Sigma_j| - \frac{1}{2}(X - \mu_j)' \Sigma_j^{-1} (X - \mu_j) \geq -\frac{1}{2} \ln|2\pi\Sigma_k| - \frac{1}{2}(X - \mu_k)' \Sigma_k^{-1} (X - \mu_k) \quad (3.3)$$

$$\ln|2\pi\Sigma_j| + (X - \mu_j)' \Sigma_j^{-1} (X - \mu_j) \leq \ln|2\pi\Sigma_k| + (X - \mu_k)' \Sigma_k^{-1} (X - \mu_k) \quad (3.4)$$

Por lo tanto:

X se asigna a Π_j si cumple con (3.4).

En el caso de dos poblaciones partiendo de (3.3)

$$-\frac{1}{2} \ln|2\pi\Sigma_1| - \frac{1}{2}(X - \mu_1)' \Sigma_1^{-1} (X - \mu_1) \geq -\frac{1}{2} \ln|2\pi\Sigma_2| - \frac{1}{2}(X - \mu_2)' \Sigma_2^{-1} (X - \mu_2)$$

$$-\frac{1}{2}(X - \mu_1)' \Sigma_1^{-1} (X - \mu_1) + \frac{1}{2}(X - \mu_2)' \Sigma_2^{-1} (X - \mu_2) - \frac{1}{2} \ln|2\pi\Sigma_1| + \frac{1}{2} \ln|2\pi\Sigma_2| \geq 0$$

$$-\frac{1}{2}(X - \mu_1)' \Sigma_1^{-1} (X - \mu_1) + \frac{1}{2}(X - \mu_2)' \Sigma_2^{-1} (X - \mu_2) - \frac{1}{2} \ln|\Sigma_1| + \frac{1}{2} \ln|\Sigma_2|$$

$$= \frac{1}{2} X' (\Sigma_2^{-1} - \Sigma_1^{-1}) X + X' (\Sigma_1^{-1} \mu_1 - \Sigma_2^{-1} \mu_2) + \frac{1}{2} \mu_2' \Sigma_2^{-1} \mu_2 - \frac{1}{2} \mu_1' \Sigma_1^{-1} \mu_1 + \frac{1}{2} \ln|\Sigma_2| - \frac{1}{2} \ln|\Sigma_1| \geq 0$$

Y X se asigna a Π_2 en otro caso.

3.2.2. La regla del discriminante de Bayes.

En ciertas situaciones tiene sentido suponer que las distintas poblaciones tienen asociadas probabilidades a priori.

Supóngase que un individuo X tiene una probabilidad a priori π_j de pertenecer a la población Π_j . Esta información puede ser incorporada al análisis usando la regla de discriminación de Bayes.

Definición 3.2 Si las poblaciones $\Pi_1, \Pi_2, \dots, \Pi_g$ tienen probabilidades a priori $\pi_1, \pi_2, \dots, \pi_g$, entonces la regla de discriminación de Bayes es:

$$X \in \Pi_j \quad \text{si} \quad \pi_j L_j(X) = \max_i \pi_i L_i(X) \quad i = 1, \dots, g$$

Debe notarse que la regla de discriminación por máxima verosimilitud es un caso particular de la regla de Bayes donde las probabilidades a priori son todas iguales, es decir:

$$\pi_1 = \pi_2 = \dots = \pi_g = \frac{1}{g}$$

En el caso de dos poblaciones la regla de discriminación de Bayes se puede escribir de la siguiente manera:

$$X \in \Pi_1 \quad \text{si} \quad \frac{\pi_1 L_1(X)}{\pi_2 L_2(X)} \geq 1$$

Tomando logaritmos de ambos lados de la desigualdad anterior

$$\ln(L_1(X)) - \ln(L_2(X)) + \ln\left(\frac{\pi_1}{\pi_2}\right) \geq 0$$

$$\text{Sea } B(X) = \ln(L_1(X)) - \ln(L_2(X)) + \ln\left(\frac{\pi_1}{\pi_2}\right)$$

$$X \in \Pi_1 \quad \text{si} \quad B(X) \geq 0$$

$$X \in \Pi_1 \quad \text{en caso contrario.}$$

Teorema 3.3 Si las poblaciones $\Pi_1, \Pi_2, \dots, \Pi_g$ tienen probabilidades a priori $\pi_1, \pi_2, \dots, \pi_g$ y además Π_j es una población con distribución normal p -variada con vector de medias μ_j y matriz de varianzas y covarianzas Σ , esto es $N_p(\mu_j, \Sigma)$, $i = 1, \dots, g$, la regla discriminante de Bayes asigna X a Π_j si: $(\mu_j - \mu_k)' \Sigma^{-1} \left(X - \frac{1}{2}(\mu_j + \mu_k) \right) \geq \ln\left(\frac{\pi_k}{\pi_j}\right) \quad \forall k \neq j$

Demostración.

Por definición X se asigna a Π_j si:

$$\pi_j L_j(X) \geq \pi_k L_k(X) \quad \forall k \neq j$$

$$\pi_j |2\pi\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(X - \mu_j)' \Sigma^{-1} (X - \mu_j)\right\} \geq \pi_k |2\pi\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(X - \mu_k)' \Sigma^{-1} (X - \mu_k)\right\}$$

$$\ln \pi_j - \frac{1}{2} \ln|2\pi\Sigma| - \frac{1}{2}(X - \mu_j)' \Sigma^{-1} (X - \mu_j) \geq \ln \pi_k - \frac{1}{2} \ln|2\pi\Sigma| - \frac{1}{2}(X - \mu_k)' \Sigma^{-1} (X - \mu_k)$$

$$\ln \pi_j - \frac{1}{2}(X - \mu_j)' \Sigma^{-1} (X - \mu_j) \geq \ln \pi_k - \frac{1}{2}(X - \mu_k)' \Sigma^{-1} (X - \mu_k)$$

$$(X - \mu_j)' \Sigma^{-1} (X - \mu_j) \leq (X - \mu_k)' \Sigma^{-1} (X - \mu_k) + 2 \ln \pi_j - 2 \ln \pi_k$$

$$X' \Sigma^{-1} X - X' \Sigma^{-1} \mu_j - \mu_j' \Sigma^{-1} X + \mu_j' \Sigma^{-1} \mu_j \leq X' \Sigma^{-1} X - X' \Sigma^{-1} \mu_k - \mu_k' \Sigma^{-1} X + \mu_k' \Sigma^{-1} \mu_k + 2 \ln \frac{\pi_j}{\pi_k}$$

$$-2 \mu_j' \Sigma^{-1} X + \mu_j' \Sigma^{-1} \mu_j \leq -2 \mu_k' \Sigma^{-1} X + \mu_k' \Sigma^{-1} \mu_k + 2 \ln \pi_j - 2 \ln \pi_k$$

$$(2 \mu_j' \Sigma^{-1} X - \mu_j' \Sigma^{-1} \mu_j) + (-2 \mu_k' \Sigma^{-1} X + \mu_k' \Sigma^{-1} \mu_k) + (\mu_k' \Sigma^{-1} \mu_j - \mu_k' \Sigma^{-1} \mu_j) \geq 2 \ln \frac{\pi_j}{\pi_k}$$

$$2 \left(\mu_j' \Sigma^{-1} X - \frac{1}{2} \mu_j' \Sigma^{-1} \mu_j + -\mu_k' \Sigma^{-1} X + \frac{1}{2} \mu_k' \Sigma^{-1} \mu_k + \frac{1}{2} \mu_k' \Sigma^{-1} \mu_j - \frac{1}{2} \mu_k' \Sigma^{-1} \mu_j \right) \geq 2 \ln \frac{\pi_j}{\pi_k}$$

$$\left(\mu_j' \Sigma^{-1} X - \frac{1}{2} \mu_j' \Sigma^{-1} \mu_j - \frac{1}{2} \mu_j' \Sigma^{-1} \mu_k \right) + \left(-\mu_k' \Sigma^{-1} X + \frac{1}{2} \mu_k' \Sigma^{-1} \mu_k + \frac{1}{2} \mu_k' \Sigma^{-1} \mu_j \right) \geq \ln \frac{\pi_j}{\pi_k}$$

$$\mu_j' \left(\Sigma^{-1} X - \frac{1}{2} \Sigma^{-1} \mu_j - \frac{1}{2} \Sigma^{-1} \mu_k \right) - \mu_k' \left(\Sigma^{-1} X - \frac{1}{2} \Sigma^{-1} \mu_k - \frac{1}{2} \Sigma^{-1} \mu_j \right) \geq \ln \frac{\pi_j}{\pi_k}$$

$$\mu_j' \Sigma^{-1} \left(X - \frac{1}{2} \mu_j - \frac{1}{2} \mu_k \right) - \mu_k' \Sigma^{-1} \left(X - \frac{1}{2} \mu_k - \frac{1}{2} \mu_j \right) \geq \ln \frac{\pi_j}{\pi_k}$$

$$(\mu_j' - \mu_k') \Sigma^{-1} \left(X - \frac{1}{2} (\mu_j + \mu_k) \right) \geq \ln \frac{\pi_k}{\pi_j}$$

$$(\mu_j - \mu_k)' \Sigma^{-1} \left(X - \frac{1}{2} (\mu_j + \mu_k) \right) \geq \ln \frac{\pi_k}{\pi_j} \quad \forall k \neq j$$

En el caso de que la matriz de varianzas y covarianzas sea distinta para cada población se puede citar el siguiente teorema.

Teorema 3.4 Si las poblaciones $\Pi_1, \Pi_2, \dots, \Pi_g$ tienen probabilidades a priori $\pi_1, \pi_2, \dots, \pi_g$ y además Π_i es una población con distribución Normal p -variada con vector de medias μ_i y matriz de varianzas y covarianzas Σ_i , esto es $N_p(\mu_i, \Sigma_i)$, $i = 1, \dots, g$, la regla discriminante de Bayes asigna X a Π_j si:

$$X'(\Sigma_k^{-1} - \Sigma_j^{-1})X + X'(\Sigma_j^{-1}\mu_j + \Sigma_k^{-1}\mu'_k) + \mu'_k\Sigma_k^{-1}\mu_k - \mu'_j\Sigma_j^{-1}\mu_j \geq 2\ln\left(\frac{\pi_k}{\pi_j}\right) + \ln\left(\frac{|2\pi\Sigma_j^{-1}|}{|2\pi\Sigma_k^{-1}|}\right) \quad \forall k \neq j$$

Demostración.

Por definición X se asigna a Π_j si:

$$\pi_j L_j(X) \geq \pi_k L_k(X) \quad \forall k \neq j$$

$$\pi_j |2\pi\Sigma_j|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(X - \mu_j)' \Sigma_j^{-1} (X - \mu_j)\right\} \geq \pi_k |2\pi\Sigma_k|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(X - \mu_k)' \Sigma_k^{-1} (X - \mu_k)\right\}$$

$$\ln \pi_j - \frac{1}{2} \ln |2\pi\Sigma_j| - \frac{1}{2}(X - \mu_j)' \Sigma_j^{-1} (X - \mu_j) \geq \ln \pi_k - \frac{1}{2} \ln |2\pi\Sigma_k| - \frac{1}{2}(X - \mu_k)' \Sigma_k^{-1} (X - \mu_k)$$

$$(X - \mu_j)' \Sigma_j^{-1} (X - \mu_j) \leq (X - \mu_k)' \Sigma_k^{-1} (X - \mu_k) + 2 \ln \pi_j - 2 \ln \pi_k + \ln |2\pi\Sigma_k| - \ln |2\pi\Sigma_j|$$

Desarrollando y simplificando se tiene que:

$$X'(\Sigma_j^{-1} - \Sigma_k^{-1})X - X'(\Sigma_j^{-1}\mu_j + \Sigma_k^{-1}\mu'_k) - \mu'_k\Sigma_k^{-1}\mu_k + \mu'_j\Sigma_j^{-1}\mu_j \leq 2\ln\left(\frac{\pi_k}{\pi_j}\right) + \ln\left(\frac{|2\pi\Sigma_j^{-1}|}{|2\pi\Sigma_k^{-1}|}\right)$$

$$X'(\Sigma_k^{-1} - \Sigma_j^{-1})X + X'(\Sigma_j^{-1}\mu_j + \Sigma_k^{-1}\mu'_k) + \mu'_k\Sigma_k^{-1}\mu_k - \mu'_j\Sigma_j^{-1}\mu_j \geq 2\ln\left(\frac{\pi_k}{\pi_j}\right) + \ln\left(\frac{|2\pi\Sigma_j^{-1}|}{|2\pi\Sigma_k^{-1}|}\right)$$

3.3. Discriminación cuando se conoce la distribución de las poblaciones pero los parámetros son estimados.

La situación más común al asignar un nuevo individuo a uno de las g poblaciones con distribución conocida, es que los parámetros de ésta son estimados.

En el caso en que se sepa que cada población Π_j tiene distribución multivariada con μ_j vector de medias y matriz de varianzas y covarianzas común Σ . Los vectores de medias y la matriz de varianzas y covarianzas pueden ser reemplazados por las estimaciones:

$$\hat{\mu}_j = \bar{X}_j, \quad \hat{\Sigma} = S = \frac{1}{n-g} \sum_{j=1}^g n_j S_j = \frac{1}{n-g} \sum_{j=1}^g \sum_{i=1}^{n_j} (x_{ij} - \bar{X}_j)(x_{ij} - \bar{X}_j)'$$

Definidas en términos de medias muestrales \bar{X}_j y de las matrices de covarianzas muestrales S_j para el j -ésimo grupo.

Si X es la nueva observación que se desea asignar a alguna de las poblaciones $\Pi_1, \Pi_2, \dots, \Pi_g$. Se asigna X a Π_i si $W_{ij} \geq 0 \quad \forall j \neq i$

$$W_{ij} = (\bar{X}_i - \bar{X}_j)' S^{-1} \left(X - \frac{1}{2} (\bar{X}_i + \bar{X}_j) \right) \geq 0 \quad (3.5)$$

Obsérvese que es exactamente el mismo resultado obtenido en el teorema (3.1 a), sólo que en lugar de los valores de los parámetros aparecen sus estimaciones.

Nótese inmediatamente que $W_{ij} = -W_{ji}$ y que cualquier $(k-1)$ W_{ij} linealmente independientes forman una base para el conjunto completo de las estadísticas si $(k-1) \leq p$. Donde k es el número total de W_{ij} 's.

Si $p \leq k-1$ el espacio de los W_{ij} tendrá rango p , y la regla de clasificación puede ser definida en términos de p puntajes.

Ejemplo. Considerese que $g = 3$ poblaciones y sea $p \geq 2$. Los distintos discriminantes son:

$$W_{12} = (\bar{X}_1 - \bar{X}_2)' S^{-1} \left(X - \frac{1}{2} (\bar{X}_1 + \bar{X}_2) \right)$$

$$W_{13} = (\bar{X}_1 - \bar{X}_3)' S^{-1} \left(X - \frac{1}{2} (\bar{X}_1 + \bar{X}_3) \right)$$

$$W_{23} = (\bar{X}_2 - \bar{X}_3)' S^{-1} \left(X - \frac{1}{2} (\bar{X}_2 + \bar{X}_3) \right)$$

Ya que $W_{23} = W_{13} - W_{12}$, sólo es necesario usar los estadísticos W_{12} y W_{13} , la regla de clasificación es definida de la siguiente forma:

$$\Pi_1 \quad \text{si} \quad W_{12} \geq 0 \text{ y } W_{13} \geq 0$$

$$\Pi_2 \quad \text{si} \quad W_{12} \leq 0 \text{ y } W_{13} \geq W_{12}$$

$$\Pi_3 \quad \text{si} \quad W_{13} \leq 0 \text{ y } W_{12} \geq W_{13}$$

3.3.1 Clasificación por mínima distancia.

La clasificación de la regla (3.5) puede ser expresada mediante la distancia de Mahalanobis.

$$D_i^2 = (X - \bar{X}_i)' S^{-1} (X - \bar{X}_i)$$

de la nueva observación a la media muestral del i -ésimo grupo.

Asignando X a la población i si $D_i^2 = \min\{D_1^2, D_2^2, \dots, D_k^2\}$. La equivalencia de las reglas se sigue inmediatamente de la relación $W_{ij} = -\frac{1}{2}D_i^2 + \frac{1}{2}D_j^2$. Así es que (3.5) se satisface si

$$D_i^2 \leq D_j^2 \quad \forall j \neq i$$

3.4. Función Lineal Discriminante de Fisher.

Otra aproximación para el problema de discriminación basado en la matriz de datos $X_{(n \times p)}$ puede ser hecho sin asumir alguna forma paramétrica de la distribución de las poblaciones

$\Pi_1, \Pi_2, \dots, \Pi_g$, pero lo suficientemente sensible para discriminar entre ellas. Fisher sugirió observar la función lineal $a'X$ que maximice el cociente de la suma de cuadrados entre grupos y la suma de cuadrados dentro de los grupos; es decir:

$$Y = \begin{bmatrix} X_1 a \\ \vdots \\ X_g a \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_g \end{bmatrix}$$

donde la matriz X_j $j = 1, \dots, g$ de tamaño $n_j \times p$ representa la muestra de n_j individuos de la población Π_j , es decir, los renglones son los individuos y las columnas son las variables.

Esto es Y es una combinación lineal de las columnas de X .

Entonces Y tiene la suma de cuadrados total.

$$\begin{aligned} \sum_{i=1}^g \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 &= \sum_{i=1}^g \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})(Y_{ij} - \bar{Y})' \\ &= \sum_{i=1}^g \sum_{j=1}^{n_i} (a'(X_{ij} - \bar{X})) (a'(X_{ij} - \bar{X}))' \\ &= \sum_{i=1}^g \sum_{j=1}^{n_i} a'(X_{ij} - \bar{X})(X_{ij} - \bar{X})' a \\ &= a' \left[\sum_{i=1}^g \sum_{j=1}^{n_i} (X_{ij} - \bar{X})(X_{ij} - \bar{X})' \right] a \end{aligned}$$

$$= a' T a \quad (3.6)$$

Donde T se conoce como la matriz total y:

$$\bar{Y} = \frac{1}{n} \sum_{j=1}^g \sum_{i=1}^{n_i} Y_{ij} \in \mathfrak{R}$$

Se puede demostrar que:

$$\begin{aligned} T &= \sum_{j=1}^g \sum_{i=1}^{n_i} (X_{ij} - \bar{X}_i) (X_{ij} - \bar{X}_i)' + \sum_{i=1}^g n_i (\bar{X}_i - \bar{X}) (\bar{X}_i - \bar{X})' \\ &= W + B \end{aligned}$$

Donde

$$W = \sum_{j=1}^g \sum_{i=1}^{n_i} (X_{ij} - \bar{X}_i) (X_{ij} - \bar{X}_i)'$$

Se conoce como la matriz dentro de los grupos y

$$B = \sum_{i=1}^g n_i (\bar{X}_i - \bar{X}) (\bar{X}_i - \bar{X})'$$

se conoce como la matriz entre grupos.

Por lo tanto se tiene que (3.6) se puede escribir como:

$$\begin{aligned} a' T a &= a' (W + B) a \\ &= a' W a + a' B a \end{aligned}$$

El criterio de Fisher es intuitivamente atractivo porque es más fácil identificar los grupos separadamente si la suma de cuadrados entre grupos para Y es relativamente grande con respecto a la suma de cuadrados dentro de los grupos. El cociente está dado por:

$$\frac{X'BX}{X'WX} \quad (3.7)$$

Sea a el vector que maximiza (3.7) se suele llamar a la función $a'X$ función lineal discriminante de Fisher o la "Primera variable Canónica".

$$\frac{a'Ba}{a'Wa} \quad (3.8)$$

Nótese que cualquier múltiplo escalar de a no afecta el valor de (3.8).

Hasta aquí se han mencionado características del vector a ; pero no se ha expuesto como se calcula, para lo cual se cita el siguiente teorema.

Teorema 3.5 El vector a en la función discriminante de Fisher es el vector propio asociado al valor característico más grande de la matriz $W^{-1}B$.

Demostración

$$\max \frac{X'BX}{X'WX} \quad \text{equivale a} \quad \max X'BX \quad \text{sujeto a que} \quad X'WX = 1 \quad (3.9)$$

Sea $W^{\frac{1}{2}}$ que denota la raíz cuadrada de W , es decir, $W^{\frac{1}{2}}W^{\frac{1}{2}} = W$ y sea $Y = W^{\frac{1}{2}}X$

Entonces el máximo de $X'BX$ sujeto que $X'WX = 1$ puede ser escrito como:

$$\max Y'W^{-\frac{1}{2}}B W^{-\frac{1}{2}}Y \quad \text{sujeto a que} \quad Y'Y = 1$$

Sea $W^{-\frac{1}{2}}BW^{-\frac{1}{2}} = \Gamma\Lambda\Gamma'$ la descomposición espectral de la matriz dada y sea $Z = \Gamma'Y$ entonces $Z'Z = Y'\Gamma\Gamma'Y = Y'Y$ así que (3.9) puede escribirse como:

$$\max Z'\Lambda Z = \max \sum_{i=1}^p \lambda_i Z_i^2 \quad \text{sueto a que } Z'Z = 1$$

Si $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$

$$\max \sum_{i=1}^p \lambda_i Z_i^2 \leq \lambda_1 \sum_{i=1}^p Z_i^2 = \lambda_1$$

Esto obliga a que $Z' = (1, 0, \dots, 0)$

Ahora bien $Z = \Gamma'Y$ premultiplicando por Γ se obtiene

$$\Gamma Z = \Gamma\Gamma'Y \Rightarrow Y = r_{(1)} \Rightarrow X = W^{-\frac{1}{2}}r_{(1)}$$

Por el teorema anterior $W^{-1}B$ y $W^{-\frac{1}{2}}BW^{-\frac{1}{2}}$ tienen los mismos valores propios ya que:

$W^{-\frac{1}{2}}W^{-\frac{1}{2}}B = W^{-1}B$ y $X = W^{-\frac{1}{2}}r_{(1)}$ es el eigenvector de $W^{-1}B$ correspondiente a λ_1 que maximiza el cociente $\frac{X'BX}{X'WX}$

Ahora lo más importante de saber como encontrar el valor de α ; es la descripción del proceso que se ha de seguir para asignar un nuevo individuo X a una de las g poblaciones.

Una vez que la función discriminante es encontrada. Una observación X puede ser asignada a una de las g poblaciones con base a su puntaje discriminante $\alpha'X$. Las medias muestrales \bar{X}_i tienen puntajes $\alpha'\bar{X}_i = \bar{y}_i$. Entonces X es asignado a aquella población cuyo puntaje de medias es más cercano a $\alpha'X$; esto es:

$$X \quad \text{se asigna a} \quad \Pi_j \quad \text{si} \quad |\alpha'X - \alpha'\bar{X}_j| \leq |\alpha'X - \alpha'\bar{X}_i| \quad \forall i \neq j$$

La función discriminante de Fisher en el caso especial de $g = 2$ poblaciones.

Si $g = 2$ entonces B tiene rango uno y puede calcularse mediante la expresión:

$$\begin{aligned}
 B &= \frac{1}{n} [n_1(\bar{X}_1 - \bar{X})(\bar{X}_1 - \bar{X})' + n_2(\bar{X}_2 - \bar{X})(\bar{X}_2 - \bar{X})'] \\
 &= \frac{1}{n} [n_1\bar{X}_1\bar{X}_1' - n_1\bar{X}_1\bar{X}' - n_1\bar{X}\bar{X}_1' + n_1\bar{X}\bar{X}' + n_2\bar{X}_2\bar{X}_2' - n_2\bar{X}_2\bar{X}' - n_2\bar{X}\bar{X}_2' + n_2\bar{X}\bar{X}'] \\
 &= \frac{1}{n} [n_1\bar{X}_1\bar{X}_1' - \frac{n_1}{n}\bar{X}_1(n_1\bar{X}_1' + n_2\bar{X}_2')\bar{X}' - \frac{n_1}{n}(n_1\bar{X}_1 + n_2\bar{X}_2)\bar{X}_1' \\
 &\quad + \frac{1}{n} [n\bar{X}\bar{X}' + n_2\bar{X}_2\bar{X}_2' - \frac{n_2}{n}\bar{X}_2(n_1\bar{X}_1' + n_2\bar{X}_2') - \frac{n_2}{n}(n_1\bar{X}_1 + n_2\bar{X}_2)\bar{X}_2'] \\
 &= \frac{1}{n} [n_1\bar{X}_1\bar{X}_1' - \frac{n_1^2}{n}\bar{X}_1\bar{X}_1' - \frac{n_1n_2}{n}\bar{X}_1\bar{X}_2' - \frac{n_2^2}{n}\bar{X}_1\bar{X}_1' - \frac{n_1n_2}{n}\bar{X}_2\bar{X}_1' + \frac{n_1^2}{n}\bar{X}_1\bar{X}_1' \\
 &\quad + \frac{1}{n} [\frac{n_1n_2}{n}\bar{X}_1\bar{X}_2' + \frac{n_1n_2}{n}\bar{X}_1\bar{X}_2' + \frac{n_2^2}{n}\bar{X}_2\bar{X}_2' + \frac{n_1n_2}{n}\bar{X}_2\bar{X}_1' + \frac{n_2^2}{n}\bar{X}_2\bar{X}_2'] \\
 &\quad + \frac{1}{n} [n_2\bar{X}_2\bar{X}_2' - \frac{n_1n_2}{n}\bar{X}_2\bar{X}_1' - \frac{n_2^2}{n}\bar{X}_2\bar{X}_2' - \frac{n_1n_2}{n}\bar{X}_1\bar{X}_2' - \frac{n_2^2}{n}\bar{X}_2\bar{X}_2'] \\
 &= \frac{1}{n^2} [nn_1\bar{X}_1\bar{X}_1' - n_1^2\bar{X}_1\bar{X}_1' + nn_2\bar{X}_2\bar{X}_2' - n_1n_2\bar{X}_2\bar{X}_1' - n_1n_2\bar{X}_1\bar{X}_2' - n_2^2\bar{X}_2\bar{X}_2'] \\
 &= \frac{1}{n^2} [n_1n_2\bar{X}_1\bar{X}_1' + n_1n_2\bar{X}_2\bar{X}_2' - n_1n_2\bar{X}_2\bar{X}_1' - n_1n_2\bar{X}_1\bar{X}_2'] \\
 &= \frac{1}{n^2} n_1n_2(\bar{X}_1 - \bar{X}_2)(\bar{X}_1' - \bar{X}_2') \\
 &= \frac{1}{n^2} n_1n_2(\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X}_2)'
 \end{aligned}$$

De lo anterior $W^{-1}B$ tiene un sólo valor propio, diferente de cero que puede encontrarse directamente mediante la expresión:

$$\begin{aligned}\lambda_1 &= tr(W^{-1}B) \\ &= \frac{n_1 n_2}{n} (\bar{X}_1 - \bar{X}_2)' W^{-1} (\bar{X}_1 - \bar{X}_2)\end{aligned}$$

Y el correspondiente valor propio:

$$a = W^{-1} (\bar{X}_1 - \bar{X}_2)$$

Entonces la regla discriminante está dada por:

$$X \in \Pi_1 \quad \text{si} \quad |a'X - a'\bar{X}_1| \leq |a'X - a'\bar{X}_2|$$

$$(a'X - a'\bar{X}_2) + (a'X - a'\bar{X}_1) \geq 0$$

$$2a'X - a'\bar{X}_2 - a'\bar{X}_1 \geq 0$$

$$a'X - \frac{1}{2}a'\bar{X}_2 - \frac{1}{2}a'\bar{X}_1 \geq 0$$

$$a' \left(X - \frac{1}{2}\bar{X}_2 - \frac{1}{2}\bar{X}_1 \right) \geq 0$$

$$a' \left(X - \frac{1}{2}(\bar{X}_2 + \bar{X}_1) \right) \geq 0$$

$X \in \Pi_2$ en otro caso.

Obsérvese que la regla de asignación anterior, es exactamente la misma que la regla discriminante por máxima verosimilitud muestral para dos grupos de una distribución normal con la misma matriz de covarianza. De cualquier modo las justificaciones para estas reglas son totalmente diferentes en los dos casos, en la regla por máxima verosimilitud muestral, se hace una suposición explícita de normalidad, mientras que con la regla discriminante de Fisher, simplemente se busca una regla "sensible" basada en una función lineal de X . Así de esta manera, se puede esperar que esta regla será más apropiada para poblaciones donde la hipótesis de normalidad no es exactamente satisfecha.

Para $g \geq 3$ poblaciones la regla de asignación basada en la función lineal discriminante de Fisher y la regla por máxima verosimilitud para poblaciones normales con la misma matriz de covarianzas no deben ser las mismas a menos que las medias muestrales sean colineales aunque las dos reglas serán similares si las medias son casi colineales.

En general $W^{-1}B$ tiene $\min(p, g-1)$ de valores propios diferentes de cero. Y los correspondientes valores propios definen la segunda, tercera, y subsecuentes variables canónicas.

3.5. Probabilidades de mala clasificación.

3.5.1 Probabilidades cuando los parámetros son estimados.

Formalmente, las probabilidades de mala clasificación P_{ij} están dadas por:

$$P_{ij} = \int_{R_i} \pi_j L_j(X) dx$$

donde P_{ij} es la probabilidad de asignar X a la población Π_i , cuando en realidad proviene de la población Π_j . R_i es la región de \mathcal{R}^p asociada por la regla de discriminación usada, a la población Π_i .

Como en la mayoría de los casos en que se conoce la distribución de las poblaciones; no se conocen sus parámetros, éstos tienen que ser estimados, hecho por el cual las probabilidades de mala clasificación también tienen que estimarse.

Si se tiene el caso de dos poblaciones normales $N_p(\mu_1, \Sigma)$ y $N_p(\mu_2, \Sigma)$, se sabe que la regla de discriminación está dada por:

Se asigna X a Π_1 si $W(X) = (X_1 - \bar{X}_2)' S^{-1} \left(X - \frac{1}{2}(X_1 + X_2) \right) \geq 0$, además

$$E(W(X)) = \hat{a}' \left(\mu_1 - \frac{1}{2}(\bar{X}_1 + \bar{X}_2) \right) = \hat{a}' \left(\mu_1 - \frac{1}{2}\bar{X} \right)$$

$$Var(W(X)) = Var \left[\hat{a}' \left(\mu_1 - \frac{1}{2}(\bar{X}_1 + \bar{X}_2) \right) \right] = \hat{a}' V(X) \hat{a} = \hat{a}' \Sigma \hat{a}$$

Donde $\hat{a} = S^{-1}(\bar{X}_1 - \bar{X}_2)$

$$\bar{X} = \frac{1}{2}(\bar{X}_1 + \bar{X}_2)$$

Por lo tanto $W(X) \sim N_1 \left(\hat{a}' \left(\mu_1 - \frac{1}{2}\bar{X} \right), \hat{a}' \Sigma \hat{a} \right)$

$$\hat{P}_{21} = P[W(X) \leq 0 \mid X \in \Pi_1]$$

$$= P \left[\frac{W(X) - \hat{a}' \left(\mu_1 - \frac{1}{2}\bar{X} \right)}{\sqrt{\hat{a}' \Sigma \hat{a}}} \leq \frac{-\hat{a}' \left(\mu_1 - \frac{1}{2}\bar{X} \right)}{\sqrt{\hat{a}' \Sigma \hat{a}}} \mid X \in \Pi_1 \right]$$

$$= P \left[\frac{\hat{a}' \bar{X} - \hat{a}' \mu_1}{\sqrt{\hat{a}' \Sigma \hat{a}}} \leq \frac{-\frac{1}{2} \hat{a}' \bar{X} - \hat{a}' \mu_1}{\sqrt{\hat{a}' \Sigma \hat{a}}} \mid X \in \Pi_1 \right]$$

$$= \Phi \left[\frac{\frac{1}{2} \hat{a}' \bar{X} - \hat{a}' \mu_1}{\sqrt{\hat{a}' \Sigma \hat{a}}} = z \mid X \in \Pi_1 \right]$$

donde Φ es la función de distribución acumulada hasta z de una $N(0, 1)$.

Similamente la opuesta probabilidad de mala clasificación está dada por:

$$\hat{P}_{12} = \Phi \left[\frac{\hat{a}' \mu_2 - \frac{1}{2} \hat{a}' \bar{X}}{\sqrt{\hat{a}' \Sigma \hat{a}}} = z \mid X \in \Pi_2 \right]$$

Desafortunadamente esta aproximación tiende a ser optimista; ya que tiende a subestimar las verdaderas probabilidades de mala clasificación cuando n es muy pequeño.

3.5.2. Método de resubstitución.

Supóngase que la discriminación es basada en la matriz de datos $X_{(n \times p)}$ de la cual n_j individuos provienen de la población j -ésima. Si la regla de discriminación es definida por la asignación de X a regiones R_i , sean n_{ij} el número de individuos de Π_j que están situados en r_i . Así $\sum_i n_{ij} = n_j$.

Entonces $\hat{P}_{ij} = \frac{n_{ij}}{n_j}$ es una estimación de P_{ij} , desafortunadamente este método también tiende a ser optimista con respecto a las probabilidades de mala clasificación.

3.5.3. El método U de Jack-Kniff.

Un problema adicional al método anterior es que las mismas observaciones son usadas para definir la regla de discriminación y para determinar si ésta es correcta. Una forma para evitar este problema es usar el método U de jack-knifing, definido enseguida.

Sea X_r , $r = 1, \dots, n$, los primeros n renglones de la matriz X representando los individuos de la población Π_1 . Para cada r , sean $R_1^{(r)}, \dots, R_g^{(r)}$ que denotan las regiones de asignación basada en la matriz de $(n-1) \times p$ obtenida al eliminar el r -ésimo renglón de la matriz X . Entonces X_r puede ser evaluado (si está bien asignado) con base a esta regla, que no es derivada usando X_r .

Si n_{1i}^* denota el número de los primeros n_1 individuos para los cuales $X_r \in R_i^{(r)}$, entonces $P_{1i}^* = \frac{n_{1i}^*}{n}$ es una estimación de las probabilidades de mala clasificación. Repitiendo este proceso para cada una de las demás poblaciones $j = 2, \dots, g$, se obtienen las estimaciones P_{ij}^* .

Para dos poblaciones multinormales con la misma matriz de covarianza, esta aproximación conduce a una estimación más confiable de las probabilidades de mala clasificación que los dos métodos anteriores.

4. Análisis estadístico.

Debido principalmente a la marginación, segregación y discriminación que las mujeres han sufrido, son muchos los estudios realizados con un enfoque de género. Temas como: la fecundidad, reproducción, y más recientemente el empleo, han sido abordados por demógrafos, sociólogos y otros profesionistas, considerando a las mujeres como objeto exclusivo de dichos estudios.

El objetivo principal del presente trabajo es la clasificación de las 32 entidades federativas de la república mexicana (1998) según dos grupos de variables económicas, y la posterior validación de dichas clasificaciones, considerando únicamente a la población masculina.

Así pues las 32 entidades constituyen la población objetivo del presente estudio, involucrando en el análisis 26 variables relacionadas con el tema del empleo. De las 26, 15 variables corresponden a la población ocupada masculina, clasificada tres veces en diversas categorías, mientras que las 11 restantes pertenecen a la población desocupada abierta masculina, clasificada en cuatro ocasiones, según igual número de criterios; para cada estado de la república.

Estas variables, si bien cubren algunos aspectos del tema del empleo, no son las únicas relacionadas con éste, ni necesariamente las más adecuadas para su medición. Sin embargo, cabe aclarar que se buscó involucrar variables que estuvieran estrechamente relacionadas entre si y con el tema.

Los datos fueron obtenidos de la Encuesta Nacional de Empleo (ENE) de 1998. Realizada año con año por el Instituto Nacional de Estadística e Informática (INEGI). Para más detalles revisar el Anexo I.

Se escogió el año de 1998, por ser la información más accesible. Editada en octubre de 1999, no es sino hasta principios del año 2000 que el INEGI la pone disponible para consulta.

4.1. Aplicación de la técnica de componentes principales.

En esta sección se disminuirá la dimensionalidad de las 32 observaciones con el objeto de obtener una representación grafica de las 32 entidades federativas que permita determinar grupos de estados en condiciones similares de acuerdo a cada conjunto de variables.

4.1.1. Análisis de resultados de la población ocupada masculina.

Primer grupo de variables¹. Este grupo corresponde al número de hombres ubicados en la población ocupada, por estado, según: posición en el trabajo (empleadores, trabajadores por su cuenta, asalariados, a destajo, sin pago), forma de pago (salario fijo, ingreso variable, ganancias, la familia consume lo que produce), nivel de ingresos (menos de un salario mínimo, de 1 a 2, más de 2 a 3, más de 3 a 5, más de 5 a 10, más de 10).

Antes de continuar, es importante realizar un primer acercamiento a la población sujeta al estudio y a las variables mediante la obtención de estadísticas básicas, seguido del calculo de correlaciones.

En el cuadro 4.1 se muestran medidas descriptivas correspondientes al primer grupo de variables. Según se puede observar el grueso de la población masculina es asalariada, percibiendo dos salarios mínimos o menos.

¹ Para consultar los datos por entidad federativa, revisar el anexo A-III.

La existencia de correlaciones altas entre las 15 variables originales, permite esperar buenos resultados, con el uso de la técnica, ya que se necesitará un número reducido de componentes para representar los datos.

En los cuadros 4.3 y 4.4 se han recogido las salidas básicas del paquete STATISTICA, para el cálculo de valores propios y los correspondientes vectores propios, de la matriz de correlaciones del primer grupo de variables.

Se obtuvieron 15 valores propios. En el cuadro 4.3 se presentan las varianzas de los 15 componentes principales, así como el porcentaje de varianza total explicada por cada uno de ellos. De acuerdo al mismo cuadro, un sólo componente principal explica el 77.8% de la varianza total del primer grupo de variables, mientras que dos componentes explican alrededor del 94.2%.

Cuadro 4.3 Obtención de valores propios y varianza explicada

| | Acumulados | | | |
|----|--------------|----------|--------------|----------|
| | Valor propio | Varianza | Valor propio | Varianza |
| 1 | 11.67217 | 77.81448 | 11.67217 | 77.81448 |
| 2 | 2.45375 | 16.35834 | 14.12592 | 94.17283 |
| 3 | 0.28569 | 1.90457 | 14.41161 | 96.07740 |
| 4 | 0.20952 | 1.39681 | 14.62113 | 97.47421 |
| 5 | 0.14912 | 0.99412 | 14.77025 | 98.46834 |
| 6 | 0.13606 | 0.90709 | 14.90631 | 99.37542 |
| 7 | 0.03909 | 0.26060 | 14.94540 | 99.63603 |
| 8 | 0.01771 | 0.11807 | 14.96312 | 99.75410 |
| 9 | 0.01513 | 0.10087 | 14.97825 | 99.85497 |
| 10 | 0.00853 | 0.05688 | 14.98678 | 99.91185 |
| 11 | 0.00719 | 0.04792 | 14.99397 | 99.95977 |
| 12 | 0.00364 | 0.02429 | 14.99761 | 99.98406 |
| 13 | 0.00185 | 0.01232 | 14.99946 | 99.99639 |
| 14 | 0.00054 | 0.00358 | 14.99999 | 99.99996 |
| 15 | 0.00001 | 0.00004 | 15 | 100 |

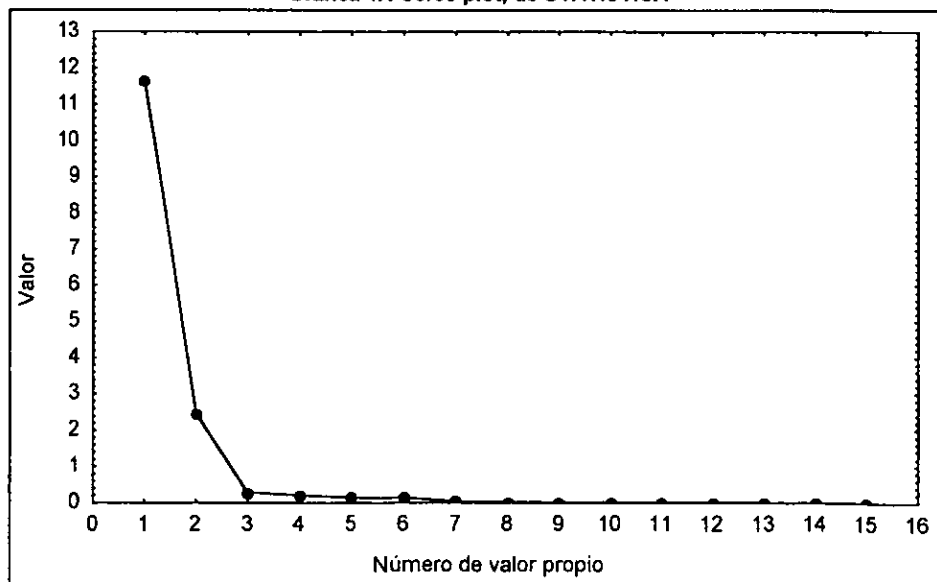
Enseguida en el cuadro 4.4 se presentan los ponderadores o puntajes (factor scores) que permiten calcular el valor de los primeros cinco componentes principales para cada una de las 32 entidades federativas de la república mexicana. Cabe mencionar que la norma de cada uno de los citados vectores es diferente de 1, sin embargo esto no representa problema, excepto el de escala.

Cuadro 4.4 Coeficientes de los primeros cinco componentes principales

| Variables | Componentes | | | | |
|------------|-------------|----------|----------|----------|----------|
| | 1 | 2 | 3 | 4 | 5 |
| E | 0.08136 | 0.07024 | -0.05109 | 0.88587 | 0.50612 |
| TSC | 0.07925 | -0.14946 | -0.14534 | -0.02279 | -0.13962 |
| TA | 0.08370 | 0.05629 | 0.15819 | -0.65539 | -0.32171 |
| TD | 0.08153 | 0.01851 | 0.84781 | 0.61230 | 0.60253 |
| TSP | 0.05342 | -0.30083 | -0.30340 | 0.64588 | -0.79886 |
| SF | 0.08303 | 0.07059 | 0.14602 | -0.72818 | -0.33676 |
| IV | 0.08218 | -0.02382 | 0.52428 | 0.63272 | 0.88051 |
| G | 0.08301 | -0.07285 | -0.04748 | 0.37004 | -0.63604 |
| LFCLP | 0.03432 | -0.34742 | -0.46981 | -0.83431 | 1.68950 |
| <1 SM | 0.06450 | -0.24133 | -0.13706 | -0.00772 | -0.97781 |
| >1 A 2 SM | 0.08445 | 0.02145 | 0.43303 | -0.29079 | -0.00443 |
| >2 A 3 SM | 0.08050 | 0.09528 | 0.41493 | -0.84103 | -0.16362 |
| >3 A 5 SM | 0.08079 | 0.12072 | -0.29287 | -0.26041 | -0.06021 |
| >5 A 10 SM | 0.07713 | 0.15507 | -0.65148 | -0.05394 | -0.07779 |
| >10 SM | 0.06560 | 0.21977 | -1.11037 | 0.37366 | 0.43066 |

Según los criterios 2 y 3 de la sección 2.3, del capítulo 2, se sugiere la retención de los dos primeros componentes principales. Ya que son éstos los únicos mayores a 1. El primero con un valor de 11.5 y el segundo muy próximo a 3, que en conjunto según se ha mencionado, explican casi la totalidad de la varianza de las 32 observaciones.

Gráfica 4.1 scree plot, de STATISTICA



Lo anterior sugiere la obtención de los dos primeros componentes principales para cada una de las 32 entidades federativas de la república mexicana. En el cuadro 4.5 se muestran los primeros cinco; el paquete los calcula como combinaciones lineales de las variables estandarizadas, es decir, aquellas con media cero y desviación estándar uno, asociadas a las originales.

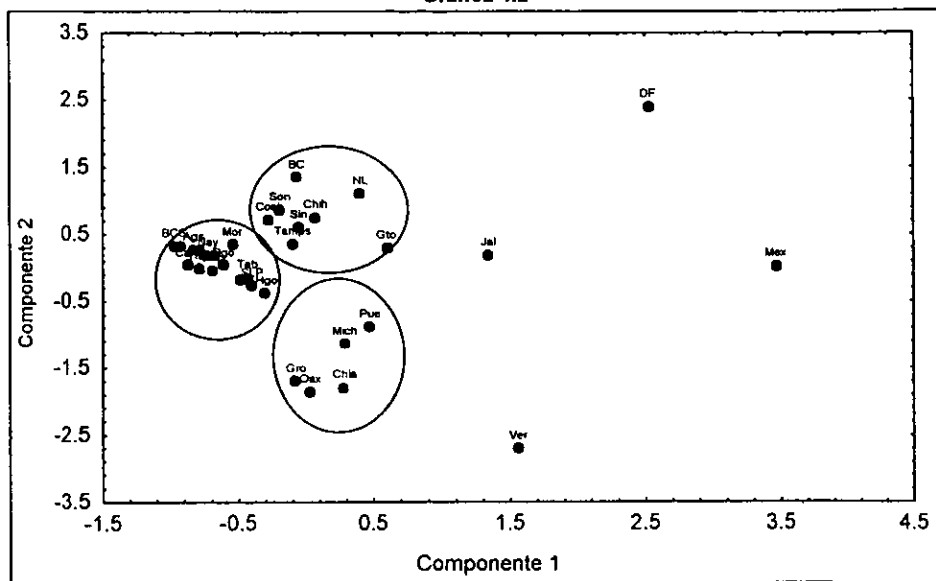
Cuadro 4.5 Componentes principales para las 32 entidades de la república mexicana (1998), según primer grupo de variables

| | Componentes | | | | |
|-------|-------------|----------|----------|----------|----------|
| | 1 | 2 | 3 | 4 | 5 |
| Ags | -0.83686 | 0.28669 | 0.55298 | -0.14665 | 0.15766 |
| BC | -0.07223 | 1.36099 | -1.61840 | -0.56306 | 0.22040 |
| BCS | -0.98122 | 0.34308 | 0.60457 | 0.08554 | -0.15495 |
| Cam | -0.88268 | 0.04948 | 0.17350 | -0.06808 | 0.38643 |
| Coah | -0.28109 | 0.73444 | -0.69915 | -0.73151 | -1.10137 |
| Col | -0.93889 | 0.33055 | 0.78277 | 0.26533 | 0.03799 |
| Chia | 0.27358 | -1.77760 | -1.45628 | 1.79051 | -3.33678 |
| Chih | 0.05658 | 0.76485 | -0.72090 | -0.72580 | -0.16933 |
| DF | 2.51779 | 2.41369 | -2.14686 | 1.92238 | 1.31487 |
| Dgo | -0.62170 | 0.04899 | 0.15524 | -0.37575 | 0.14234 |
| Gto | 0.59600 | 0.30922 | 1.82787 | 1.55741 | 0.43947 |
| Gro | -0.08294 | -1.68913 | -0.60579 | -0.39450 | 1.79887 |
| Hgo | -0.30511 | -0.36266 | 0.38565 | 0.64781 | -0.10531 |
| Jal | 1.33346 | 0.20524 | 1.76081 | 1.60339 | 0.28652 |
| Mex | 3.45983 | 0.03603 | 2.14599 | -2.74555 | -0.70621 |
| Mich | 0.28011 | -1.11973 | -0.29758 | 0.57118 | 2.27732 |
| Mor | -0.54884 | 0.35782 | 0.72315 | 0.46665 | -0.42892 |
| Nay | -0.73405 | 0.20909 | 0.60960 | 0.76538 | -0.41726 |
| NL | 0.39808 | 1.12436 | -1.63344 | -1.26700 | -1.29619 |
| Oax | 0.02289 | -1.84517 | -1.63193 | -1.53799 | 1.14309 |
| Pue | 0.46393 | -0.86905 | 0.40285 | -1.19410 | -0.66841 |
| Qro | -0.67815 | 0.19704 | -0.04689 | -0.40190 | 0.58193 |
| QR | -0.79125 | 0.27870 | 0.29378 | -0.29453 | 0.66815 |
| SLP | -0.40313 | -0.24916 | -0.26194 | -0.79692 | 0.22093 |
| Sin | -0.06236 | 0.61955 | 0.25247 | 0.61694 | -0.23547 |
| Son | -0.20081 | 0.87956 | 0.03160 | -0.07923 | 0.03823 |
| Tab | -0.43574 | -0.12517 | -0.09158 | 0.02799 | -1.22241 |
| Tamps | -0.10839 | 0.37197 | -0.06589 | -0.11576 | -0.41047 |
| Tlax | -0.80048 | 0.00382 | 0.68922 | -0.07683 | -0.22750 |
| Ver | 1.55677 | -2.68868 | -0.54955 | 1.21502 | -0.05448 |
| Yuc | -0.49322 | -0.17455 | -0.06509 | -0.19947 | 0.99475 |
| Zac | -0.69988 | -0.02428 | 0.49920 | 0.17909 | -0.17387 |

Ahora se procede a graficar a las 32 entidades federativas de la república considerando únicamente los primeros dos componentes principales.

Observando las gráficas 4.2 y 4.3, se sugiere la existencia de cuatro grupos, en donde se supone que la situación de los hombres ocupados es muy similar dentro de cada uno de ellos. A excepción del grupo IV que son casos aislados. Sin embargo hay que hacer notar que al interior de cada entidad federativa, es posible encontrar marcados contrastes en cuanto a la situación laboral de los hombres.

Gráfica 4.2



Así pues la agrupación que enseguida se propone, se basa en la percepción visual, apoyada en la proximidad entre las entidades en las gráficas.

GRUPOS

- I. Aguascalientes, Baja California Sur, Campeche, Colima, Durango, Hidalgo, Morelos, Nayarit, Querétaro, Quintana Roo, San Luis Potosí, Tabasco, Tlaxcala, Yucatán, Zacatecas.

II. Baja California, Coahuila, Chihuahua, Guanajuato, Nuevo León, Sinaloa, Sonora, Tamaulipas.

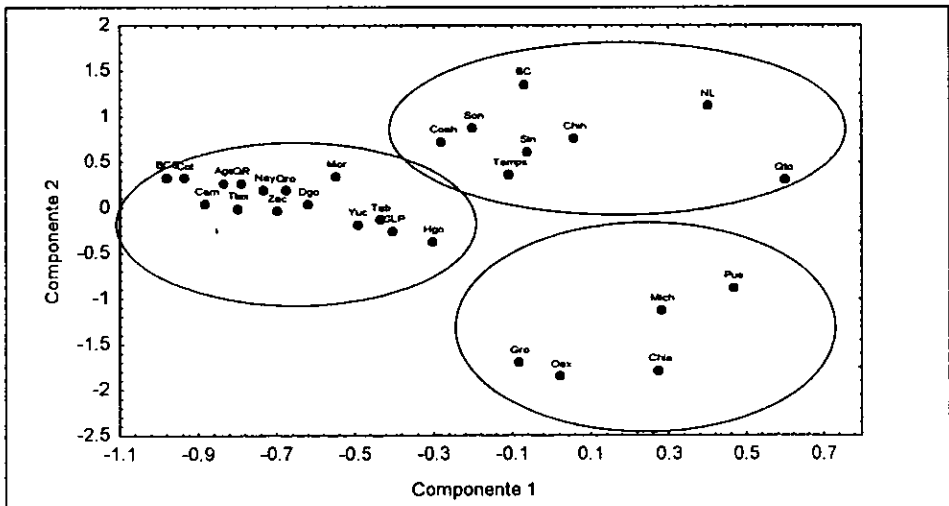
III. Chiapas, Guerrero, Michoacán, Oaxaca, Puebla.

IV. Distrito Federal, Jalisco, Veracruz, México.

Nótese que las entidades federativas tradicionalmente conocidos como las "más pobres", quedan ubicados en el grupo III. Por otra parte es de llamar la atención que el estado de Veracruz no esté incluido en este mismo grupo. Una posible explicación a esto es el hecho de que esta entidad es la número uno en cuanto a familias que consumen lo que producen y a trabajadores que no reciben ingreso; que en suma representan al 20.6% de la PEA de Veracruz.

También es curioso la similitud entre el estado de Guanajuato y el estado de Nuevo León. De hecho el primero también es muy próximo al estado de Jalisco, entidades que se ubican como de mejores condiciones.

Gráfica 4.3



4.1.2. Análisis de resultados de la población desocupada abierta.

Segundo grupo de variables². Este grupo corresponde al número de hombres ubicados en la población desocupada abierta, por estado, según: nivel de instrucción (sin instrucción, primaria, secundaria, nivel medio superior y superior), tipo de empleo buscado (asalariado, trabajo por su cuenta), experiencia laboral (con experiencia, sin experiencia), motivo del desempleo (por cese, trabajo temporal terminado, insatisfacción en el trabajo).

En el cuadro 4.6 se muestran las medidas descriptivas para el segundo grupo de variables. Obsérvese que de acuerdo al nivel de instrucción, en México el mayor número de desempleados abiertos cuenta con secundaria como grado máximo de estudios; que la mayor parte de la PDA masculina tiene experiencia laboral y busca trabajo asalariado; en cuanto a motivos de desempleo se ve que la causa principal es el cese.

Cuadro 4.6 Estadísticas básicas de variables de la Población Desocupada Abierta Masculina

| Variable | Etiqueta | Media | Mínimo | Máximo | Desv. Est. |
|--|----------|-----------|--------|---------|------------|
| PDA sin instrucción | PDASI | 520.69 | 0 | 3,804 | 910.84 |
| PDA con primaria | PDAP | 4,682.53 | 285 | 30,467 | 6,227.95 |
| PDA con secundaria | PDAS | 5,811.50 | 249 | 45,215 | 9,025.83 |
| PDA con educación media superior y superior | PDAMSYS | 5,007.75 | 359 | 37,965 | 8,248.39 |
| Busca trabajo asalariado | BTA | 15,707.13 | 1,349 | 107,962 | 22,937.02 |
| Busca trabajo por cuenta propia | BTCP | 276.59 | 0 | 3,168 | 625.91 |
| Con experiencia laboral | CE | 14,556.88 | 1,246 | 100,573 | 21,340.68 |
| Sin experiencia laboral | SE | 1,471.09 | 64 | 10,557 | 2,219.33 |
| Desempleado por cese | CESE | 4,815.75 | 195 | 33,765 | 7,945.10 |
| Desempleado por trabajo temporal terminado | TTT | 3,761.16 | 119 | 19,874 | 4,939.58 |
| Desempleado por insatisfacción en el trabajo | ICT | 4,124.47 | 178 | 34,774 | 6,729.65 |

En el cuadro siguiente se tienen los coeficientes de correlación para el segundo grupo de variables. Presentando al igual que el primer grupo, valores muy altos para la mayoría de los casos (35 de 55 mayores a 0.90). En este grupo la correlación mayor se da entre los desempleados con experiencia y los buscadores de trabajo asalariado; por otra parte la menor ocurre entre los desocupados con nivel de instrucción medio superior y superior, y los que no tienen instrucción.

² Para consultar los datos por entidad federativa, revisar el anexo A- IV.

Cuadro 4.7 Matriz de correlaciones del segundo grupo de variables

| | | | | | | | | | | | | |
|---------|-------|------|------|---------|------|------|------|------|------|------|------|--|
| | PDASI | | | | | | | | | | | |
| PDASI | 1.00 | PDAP | | | | | | | | | | |
| PDAP | 0.59 | 1.00 | PDAS | | | | | | | | | |
| PDAS | 0.53 | 0.98 | 1.00 | PDAMSYS | | | | | | | | |
| PDAMSYS | 0.38 | 0.91 | 0.93 | 1.00 | BTA | | | | | | | |
| BTA | 0.53 | 0.98 | 0.99 | 0.97 | 1.00 | BTCP | | | | | | |
| BTCP | 0.44 | 0.87 | 0.91 | 0.86 | 0.89 | 1.00 | CE | | | | | |
| CE | 0.54 | 0.98 | 0.99 | 0.96 | 1.00 | 0.90 | 1.00 | SE | | | | |
| SE | 0.47 | 0.95 | 0.97 | 0.95 | 0.97 | 0.90 | 0.97 | 1.00 | CESE | | | |
| CESE | 0.39 | 0.94 | 0.95 | 0.97 | 0.97 | 0.85 | 0.97 | 0.94 | 1.00 | TTT | | |
| TTT | 0.73 | 0.92 | 0.88 | 0.84 | 0.91 | 0.75 | 0.91 | 0.85 | 0.83 | 1.00 | ICT | |
| ICT | 0.49 | 0.96 | 0.98 | 0.92 | 0.97 | 0.92 | 0.97 | 0.97 | 0.92 | 0.83 | 1.00 | |

En los cuadros 4.8 y 4.9 se muestran nuevamente las salidas básicas del paquete, en este caso para las variables relacionadas con la población desocupada abierta.

Se obtuvieron los 11 valores propios y los correspondientes vectores propios, de la matriz de correlaciones mostrada en el cuadro 4.7.

Cuadro 4.8 Obtención de valores propios y varianza explicada

| | Valor propio | Varianza | Acumulados | |
|----|--------------|----------|--------------|----------|
| | | | Valor propio | Varianza |
| 1 | 9.65453 | 87.76848 | 9.65453 | 87.76848 |
| 2 | 0.86433 | 7.85752 | 10.51886 | 95.62600 |
| 3 | 0.23573 | 2.14299 | 10.75459 | 97.76899 |
| 4 | 0.08278 | 0.75250 | 10.83736 | 98.52149 |
| 5 | 0.06573 | 0.59758 | 10.90310 | 99.11907 |
| 6 | 0.05820 | 0.52909 | 10.96130 | 99.64816 |
| 7 | 0.02057 | 0.18697 | 10.98186 | 99.83513 |
| 8 | 0.01739 | 0.15807 | 10.99925 | 99.99320 |
| 9 | 0.00074 | 0.00669 | 10.99999 | 99.99989 |
| 10 | 0.00001 | 0.00010 | 11.00000 | 99.99999 |
| 11 | 0.00000 | 0.00001 | 11 | 100 |

Según se aprecia en el cuadro 4.8 un sólo componente principal explica alrededor del 88% de la varianza total de la población, mientras que, el restante 11% es explicado por los otros 10 componentes.

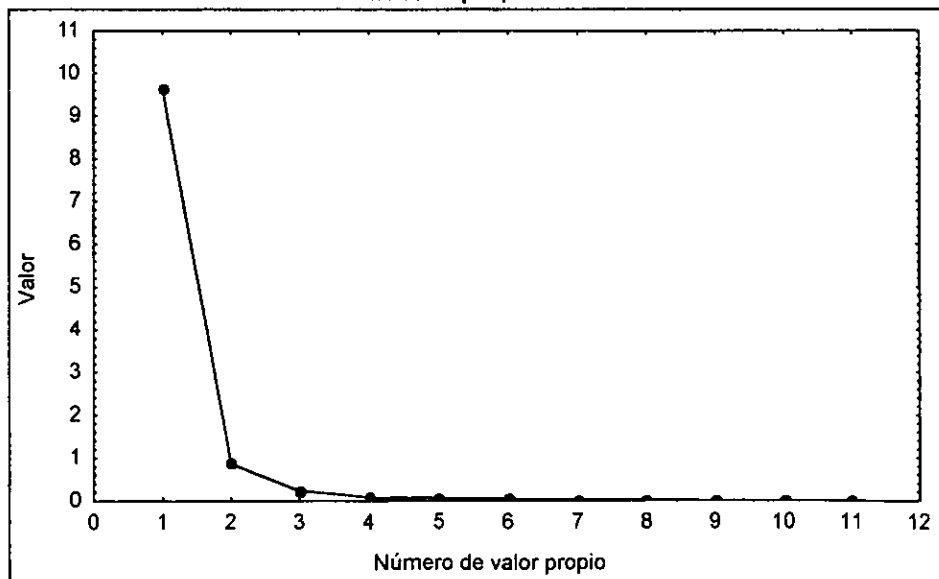
De manera análoga a la sección anterior en el cuadro 4.9 se presentan los ponderadores de cada una de las 11 variables, para cada uno de los primeros cinco componentes principales.

Cuadro 4.9 Coeficientes de los primeros cinco componentes principales

| Variables | Componentes | | | | |
|-----------|-------------|----------|----------|----------|----------|
| | 1 | 2 | 3 | 4 | 5 |
| PDASI | 0.05865 | -0.94106 | 0.34267 | -0.11487 | -1.60678 |
| PDAP | 0.10208 | -0.05555 | -0.13816 | 1.00561 | 1.09675 |
| PDAS | 0.10252 | 0.04612 | 0.17162 | 0.94975 | 0.31584 |
| PDAMSYS | 0.09882 | 0.22548 | -0.59078 | -1.45576 | -1.42289 |
| BTA | 0.10335 | 0.04292 | -0.20601 | 0.16137 | -0.19485 |
| BTCP | 0.09436 | 0.14333 | 1.50965 | -1.90527 | 0.75713 |
| CE | 0.10335 | 0.03702 | -0.19126 | 0.06190 | -0.11828 |
| SE | 0.10096 | 0.12687 | 0.09489 | 0.64444 | -0.46845 |
| CESE | 0.09929 | 0.21541 | -0.56236 | -0.40432 | -1.40796 |
| TTT | 0.09474 | -0.34394 | -0.83729 | -0.94355 | 2.54485 |
| ICT | 0.10075 | 0.10376 | 0.58548 | 1.70476 | -0.02565 |

A pesar de que el segundo eigenvalor es menor a 1, por motivos de graficación se optó por la retención de los primeros dos componentes principales; que en su conjunto explican cerca del 96% de la variabilidad de la población.

Gráfica 4.4 scree plot, de STATISTICA



En la grafica anterior, se aprecia que la pendiente que definen los segmentos de recta se empieza a estabilizar a partir del segundo valor propio, esto sugiere que la explicación de

varianza por los siguientes valores propios es mínima, por lo cual se reitera que sólo debieran de retenerse los dos primeros componentes principales.

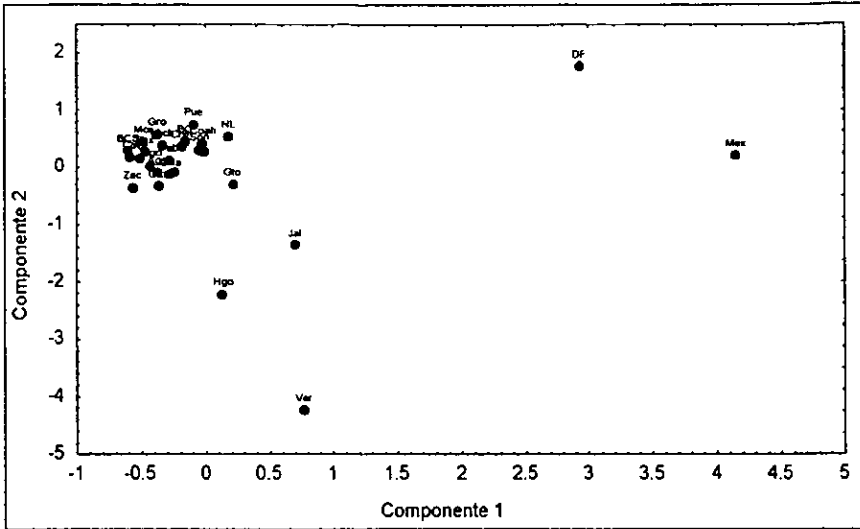
Cuadro 4.10 Componentes principales para las 32 entidades de la república mexicana (1998), según segundo grupo de variables

| | Componentes | | | | |
|-------|-------------|----------|----------|----------|----------|
| | 1 | 2 | 3 | 4 | 5 |
| Ags | -0.38324 | -0.07808 | 0.26012 | -0.02671 | -0.25901 |
| BC | -0.16660 | 0.45499 | 1.67831 | -1.79375 | 1.53465 |
| BCS | -0.61021 | 0.29972 | 0.36278 | -0.53294 | 0.15439 |
| Cam | -0.57687 | 0.29104 | 0.15443 | -0.14396 | -0.25816 |
| Coah | -0.03227 | 0.43938 | 0.09153 | 1.12128 | -0.30055 |
| Col | -0.59284 | 0.18978 | 0.45257 | -0.37787 | -0.11893 |
| Chia | -0.28231 | -0.11821 | -0.35462 | -0.12182 | -1.13176 |
| Chih | -0.19623 | 0.37142 | -0.39554 | -0.64700 | 0.61516 |
| DF | 2.92742 | 1.76833 | -2.72981 | -2.58010 | -1.33822 |
| Dgo | -0.42952 | 0.05316 | -0.05378 | -0.15141 | 0.73936 |
| Gto | 0.20721 | -0.28926 | -1.13673 | 1.55655 | 0.82585 |
| Gro | -0.37985 | 0.58690 | -0.20911 | -0.05984 | -1.25251 |
| Hgo | 0.12323 | -2.20062 | 2.02074 | -1.12102 | -2.12462 |
| Jal | 0.69210 | -1.35252 | -0.56067 | 1.90726 | -1.84305 |
| Mex | 4.14364 | 0.21317 | 2.64266 | 1.10287 | 0.89549 |
| Mich | -0.34600 | 0.39310 | -0.38003 | 0.12413 | 0.21436 |
| Mor | -0.49143 | 0.45759 | -0.01670 | 0.04270 | -0.76048 |
| Nay | -0.60967 | 0.23074 | 0.20046 | -0.12800 | -0.11139 |
| NL | 0.16923 | 0.54365 | 0.05742 | 1.66458 | 0.76254 |
| Oax | -0.36536 | -0.32738 | -0.24095 | -0.45834 | 1.49292 |
| Pue | -0.09915 | 0.76729 | -0.37219 | 1.59860 | -1.45067 |
| Qro | -0.35680 | -0.00892 | 0.31847 | 0.03008 | -0.59663 |
| QR | -0.58814 | 0.31489 | 0.32421 | -0.29793 | 0.08200 |
| SLP | -0.35447 | 0.46155 | 1.24896 | -1.51580 | 0.52004 |
| Sin | -0.24337 | -0.08370 | 0.11745 | 0.23790 | -0.37095 |
| Son | -0.05664 | 0.32340 | -1.24195 | 0.27181 | 1.61964 |
| Tab | -0.28738 | 0.13000 | -0.67819 | 0.06607 | 0.69541 |
| Tamps | -0.01424 | 0.28700 | -0.33747 | 1.19365 | 1.47811 |
| Tlax | -0.48024 | 0.27655 | 0.01184 | 0.19879 | -0.04144 |
| Ver | 0.76074 | -4.21956 | -1.69336 | -0.85808 | 1.17769 |
| Yuc | -0.51086 | 0.16492 | 0.09086 | 0.05921 | -0.25223 |
| Zac | -0.56987 | -0.34033 | 0.36830 | -0.36090 | -0.59699 |

En este caso el detectar una agrupación inmediata partiendo de la primera gráfica que ofrece el paquete, es difícil, debido a la proximidad entre la mayoría de las entidades, por

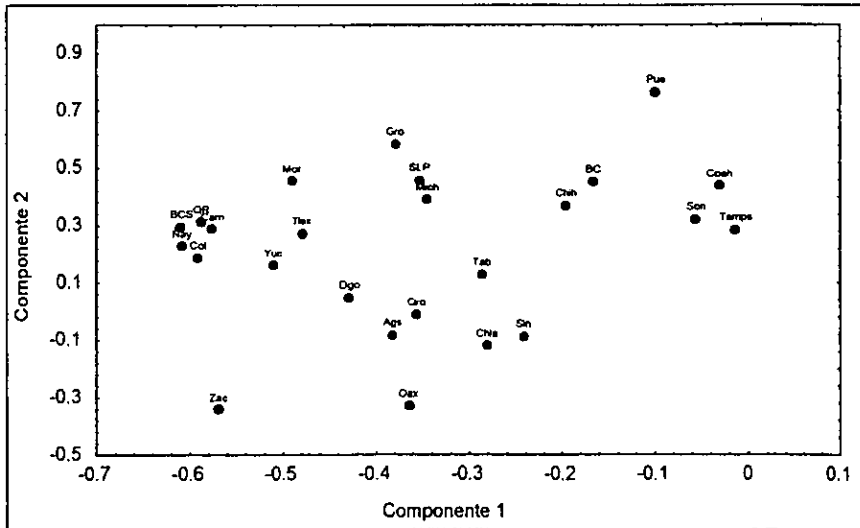
lo cual se hace necesario obtener gráficas adicionales con un acercamiento a la nube de puntos.

Gráfica 4.5

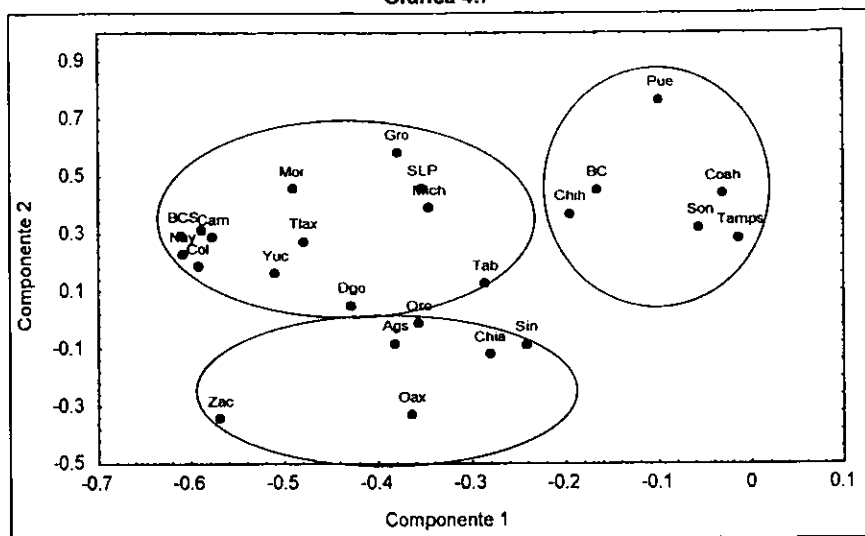


Son muchas las agrupaciones que se pueden proponer, debido a que estas proposiciones se basan en hechos visuales. Sin embargo se piensa que debiera darse una agrupación de entidades similar a la dada en la sección anterior.

Gráfica 4.6



Gráfica 4.7



Por lo tanto basándose nuevamente en la percepción visual; de las gráficas 4.5 y 4.6, se propone la siguiente agrupación de las 32 entidades de la república, según el segundo grupo de variables.

GRUPOS

- I. Baja California Sur, Campeche, Colima, Durango, Guerrero, Michoacán, Morelos, Nayarit, Quintana Roo, San Luis Potosí, Tabasco, Tlaxcala, Yucatán.
- II. Baja California, Coahuila, Chihuahua, Nuevo León, Puebla, Sonora, Tamaulipas,.
- III. Aguascalientes, Chiapas, Oaxaca, Querétaro, Sinaloa, Zacatecas.
- IV. Distrito Federal, Guanajuato, Hidalgo, Jalisco, México, Veracruz.

Nótese que las entidades federativas de la república que se mantienen en el mismo grupo, según la clasificación propuesta en la sección anterior son.

GRUPOS

- I. Baja California Sur, Campeche, Colima, Durango, Morelos, Nayarit, Quintana Roo, San Luis Potosí, Tabasco, Tlaxcala, Yucatán.
- II. Baja California, Coahuila, Chihuahua, Nuevo León, Sonora, Tamaulipas.
- III. Chiapas, Oaxaca.
- IV. Distrito Federal, Jalisco, México, Veracruz.

Es decir, 23 de las 32 entidades federativas (71.2%) quedan asignadas al mismo grupo según los dos conjuntos de variables.

4.2. Aplicación de la técnica de análisis discriminante.

En esta sección se pretende revisar que tan correctas son las clasificaciones dadas en la sección anterior; partiendo del supuesto que cada entidad federativa tiene la misma probabilidad de pertenecer a cualquiera de los cuatro grupos. Y en caso de ser necesario, la reasignación de individuos.

4.2.1. Análisis de resultados de la población ocupada.

Enseguida, de manera particular se muestra en que grupo se localizó inicialmente a cada una de las 32 entidades federativas de la república mexicana, y la probabilidad (a posteriori) con que cada observación corresponde a un grupo dado, de acuerdo al primer grupo de variables.

A menos que se indique otra cosa el paquete utiliza para los cálculos, a las variables originales, W (matriz de covarianzas dentro de los grupos) y a B (matriz de covarianzas entre los grupos).

La tercera columna del cuadro 4.11 sugiere a cual grupo deberían pertenecer y como siguientes posibilidades se ofrecen las restantes tres (en orden de importancia). Nótese que en este caso el paquete no marca "clasificaciones incorrectas".

Cuadro 4.11

| Las clasificaciones incorrectas son marcadas con * | | | | | |
|--|-------------|----------|----------|----------|----------|
| | Grupo | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Estado | Clasificado | p=.25000 | p=.25000 | p=.25000 | p=.25000 |
| Ags | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| BC | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| BCS | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Cam | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Coah | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| Col | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Chia | G 3:3 | G 3:3 | G 1:1 | G 2:2 | G 4:4 |
| Chih | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| DF | G 4:4 | G 4:4 | G 1:1 | G 2:2 | G 3:3 |
| Dgo | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Gto | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| Gro | G 3:3 | G 3:3 | G 1:1 | G 2:2 | G 4:4 |
| Hgo | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Jal | G 4:4 | G 4:4 | G 1:1 | G 2:2 | G 3:3 |
| Mex | G 4:4 | G 4:4 | G 1:1 | G 2:2 | G 3:3 |
| Mich | G 3:3 | G 3:3 | G 1:1 | G 2:2 | G 4:4 |
| Mor | G 1:1 | G 1:1 | G 3:3 | G 2:2 | G 4:4 |
| Nay | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| NL | G 2:2 | G 2:2 | G 3:3 | G 1:1 | G 4:4 |
| Oax | G 3:3 | G 3:3 | G 1:1 | G 2:2 | G 4:4 |
| Pue | G 3:3 | G 3:3 | G 1:1 | G 2:2 | G 4:4 |
| Qro | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| QR | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| SLP | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Sin | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| Son | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| Tab | G 1:1 | G 1:1 | G 3:3 | G 2:2 | G 4:4 |
| Tamps | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| Tlax | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Ver | G 4:4 | G 4:4 | G 1:1 | G 2:2 | G 3:3 |
| Yuc | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Zac | G 1:1 | G 1:1 | G 3:3 | G 2:2 | G 4:4 |

De gran ayuda para asegurarse de lo correcto de dicha asignación, se pueden revisar las probabilidades a posteriori de que cada uno de las entidades pertenezca al grupo señalado. Redondeadas con cuatro decimales se ve (cuadro 4.12) que prácticamente todas valen 1. Lo cual habla de una separación de grupos total y por consecuencia de una clasificación correcta.

Cuadro 4.12 Probabilidades a posteriori

| Las clasificaciones incorrectas son marcadas con * | | | | | |
|--|-------------|----------|----------|----------|----------|
| | Grupo | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Estado | Clasificado | p=.25000 | p=.25000 | p=.25000 | p=.25000 |
| Ags | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| BC | G 2:2 | 0.00000 | 1.00000 | 0.00000 | 0.00000 |
| BCS | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| Cam | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| Coah | G 2:2 | 0.00000 | 1.00000 | 0.00000 | 0.00000 |
| Col | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| Chia | G 3:3 | 0.00000 | 0.00000 | 1.00000 | 0.00000 |
| Chih | G 2:2 | 0.00000 | 1.00000 | 0.00000 | 0.00000 |
| DF | G 4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Dgo | G 1:1 | 0.99996 | 0.00004 | 0.00000 | 0.00000 |
| Gto | G 2:2 | 0.00000 | 1.00000 | 0.00000 | 0.00000 |
| Gro | G 3:3 | 0.00000 | 0.00000 | 1.00000 | 0.00000 |
| Hgo | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| Jal | G 4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Mex | G 4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Mich | G 3:3 | 0.00000 | 0.00000 | 1.00000 | 0.00000 |
| Mor | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| Nay | G 1:1 | 0.99998 | 0.00002 | 0.00000 | 0.00000 |
| NL | G 2:2 | 0.00000 | 1.00000 | 0.00000 | 0.00000 |
| Oax | G 3:3 | 0.00000 | 0.00000 | 1.00000 | 0.00000 |
| Pue | G 3:3 | 0.00000 | 0.00000 | 1.00000 | 0.00000 |
| Qro | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| QR | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| SLP | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| Sin | G 2:2 | 0.00000 | 1.00000 | 0.00000 | 0.00000 |
| Son | G 2:2 | 0.00021 | 0.99979 | 0.00000 | 0.00000 |
| Tab | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| Tamps | G 2:2 | 0.00000 | 1.00000 | 0.00000 | 0.00000 |
| Tlax | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| Ver | G 4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Yuc | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| Zac | G 1:1 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |

A manera de resumen en el cuadro 4.13 se muestra el porcentaje de clasificaciones correctas e incorrectas en cada grupo. Según se observa, se ha asignado a los 32 individuos de manera acertada. Perteneciendo 15 al grupo 1, 8 al 2, 5 al 3, y los restantes al 4.

Cuadro 4.13

| Rengiones: clasificaciones observadas | | | | | |
|--|------------|----------|----------|----------|----------|
| Columnas: clasificaciones predictorias | | | | | |
| | Porcentaje | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| | Correcto | p=.25000 | p=.25000 | p=.25000 | p=.25000 |
| G 1:1 | 100 | 15 | 0 | 0 | 0 |
| G 2:2 | 100 | 0 | 8 | 0 | 0 |
| G 3:3 | 100 | 0 | 0 | 5 | 0 |
| G 4:4 | 100 | 0 | 0 | 0 | 4 |
| Total | 100 | 15 | 8 | 5 | 4 |

A continuación se ofrecen los coeficientes de la función lineal discriminante de Fisher –o también llamada canónica-, estandarizados, toman este nombre debido a que se calculan sobre las variables estandarizadas. La magnitud de los coeficientes son un indicador de la importancia que tiene cada variable en el calculo de la función discriminante, también aparecen los tres valores propios de la matriz $W^{-1}B$ en el renglón "eigenval".

Cuadro 4.14 Coeficientes canónicos estandarizados

| | Raíz 1 | Raíz 2 | Raíz 3 |
|----------|-----------|------------|-----------|
| E | 2.30393 | -6.65560 | -1.25546 |
| TSC | 7.87061 | -25.29275 | -9.92810 |
| TA | 38.79320 | -104.58869 | -16.68199 |
| TD | 5.78173 | -12.32867 | -1.23173 |
| TSP | 1.27067 | 1.39230 | -0.30712 |
| SF | -27.70897 | 115.59202 | 22.92237 |
| IV | -4.26675 | 19.12210 | 3.80292 |
| G | -5.41402 | 23.70617 | 7.45642 |
| LFCLP | -1.14068 | 8.14851 | 2.93104 |
| MEN_DE_1 | -5.10400 | -1.49645 | -0.27657 |
| 1_A_2 | -6.13943 | -3.13737 | 0.06173 |
| 2_A_3 | -6.96314 | -3.97279 | -3.98790 |
| 3_A_5 | -2.56048 | 1.47584 | -3.34326 |
| 5_A_10 | 3.15619 | -5.61526 | 2.68109 |
| MAS_DE10 | -2.24815 | 0.58910 | -2.04831 |
| Eigenval | 34.48313 | 9.59535 | 4.26070 |
| Cum.Prop | 0.71336 | 0.91186 | 1 |

Asociados con cada valor propio es posible obtener los coeficientes de la función discriminante canónica no estandarizados, que son estrictamente proporcionales a los coeficientes de la función discriminante de Fisher.

Aunque el paquete no muestra los coeficientes canónicos no estandarizados; los utiliza para obtener los puntajes del mismo nombre para cada una de las 32 entidades federativas, mostrados en el cuadro 4.15.

Cuadro 4.15 Puntajes canónicos no estandarizados

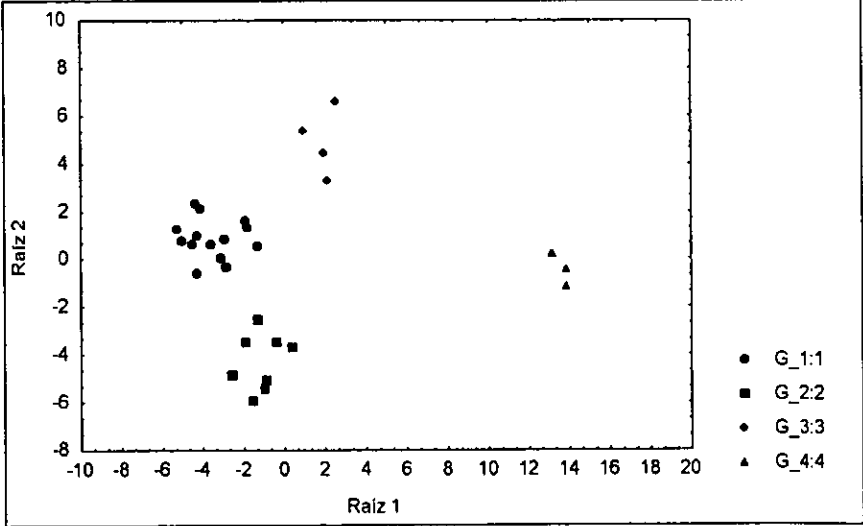
| | Grupo | Raíz 1 | Raíz 2 | Raíz 3 |
|-------|-------|----------|----------|----------|
| Ags | G 1:1 | -3.62293 | 0.66340 | 2.30596 |
| BC | G 2:2 | -1.57941 | -5.90887 | -3.42405 |
| BCS | G 1:1 | -5.31872 | 1.31496 | 1.97717 |
| Cam | G 1:1 | -4.32888 | 1.03748 | 1.36808 |
| Coah | G 2:2 | -2.58636 | -4.80482 | -0.54187 |
| Col | G 1:1 | -5.10190 | 0.76477 | 1.87584 |
| Chia | G 3:3 | 1.95594 | 4.43849 | -3.59470 |
| Chih | G 2:2 | -1.37546 | -2.52096 | -2.80541 |
| DF | G 4:4 | 13.63304 | -1.20540 | 2.60240 |
| Dgo | G 1:1 | -2.89146 | -0.26987 | 0.45263 |
| Gto | G 2:2 | -0.40525 | -3.48353 | -1.57951 |
| Gro | G 3:3 | 2.50315 | 6.63861 | -3.26775 |
| Hgo | G 1:1 | -1.37512 | 0.57998 | 2.08008 |
| Jal | G 4:4 | 13.77882 | -0.44951 | 2.17556 |
| Mex | G 4:4 | 13.84556 | -1.12177 | 1.39611 |
| Mich | G 3:3 | 1.95576 | 3.43704 | -4.06508 |
| Mor | G 1:1 | -4.18333 | 2.14275 | 1.09473 |
| Nay | G 1:1 | -4.30706 | -0.61136 | 0.25754 |
| NL | G 2:2 | 0.30280 | -3.66357 | -3.32983 |
| Oax | G 3:3 | 0.90813 | 5.36288 | -3.44230 |
| Pue | G 3:3 | 2.12557 | 3.31674 | -1.54854 |
| Qro | G 1:1 | -3.12196 | 0.09262 | 1.52649 |
| QR | G 1:1 | -2.93067 | 0.84481 | 1.05502 |
| SLP | G 1:1 | -4.33478 | 1.07170 | 1.35531 |
| Sin | G 2:2 | -1.01615 | -5.41266 | -1.45988 |
| Son | G 2:2 | -1.95447 | -3.44974 | 0.80720 |
| Tab | G 1:1 | -1.96101 | 1.67186 | 2.14363 |
| Tamps | G 2:2 | -0.96203 | -5.03744 | -1.18042 |
| Tlax | G 1:1 | -4.53393 | 0.60763 | 1.88520 |
| Ver | G 4:4 | 13.14269 | 0.17771 | 0.81822 |
| Yuc | G 1:1 | -1.82608 | 1.37628 | 2.67673 |
| Zac | G 1:1 | -4.43447 | 2.39986 | 0.38545 |

La utilidad de obtener el cuadro 4.15 recae en el hecho de que se permite realizar una representación gráfica de cada una de las 32 observaciones.

Según se observa en la figura siguiente los grupos están claramente definidos. Lo cual permite aseverar con certeza que la agrupación propuesta es correcta con base en el primer grupo de variables.

Sin embargo es importante señalar que la agrupación dada no es la única. Definiéndose de antemano cuantos grupos se pretenden tener, es posible dar una clasificación de las entidades federativas un tanto diferente a la anterior.

Gráfica 4.8 Raíz 1 Vs. Raíz 2



4.2.2. Análisis de resultados de la población desocupada abierta.

Procediendo de manera similar a la sección anterior, se obtienen los mismos seis cuadros para el caso del segundo grupo de variables relacionadas con la población desocupada abierta.

En esta ocasión los resultados no fueron los esperados, ya que existen cuatro individuos (Aguascalientes, Michoacán, Oaxaca y Tlaxcala) que fueron marcadas como mal clasificadas. Según se ve en el cuadro 4.16, Aguascalientes debió de ser asignado al grupo 2; Michoacán al 3; Oaxaca al 1; y Tlaxcala al 2. Ya que las probabilidades a posteriori, de pertenencia a dichos grupos, son de alrededor de 0.47, 0.58, 0.50, y 0.52 respectivamente.

Cuadro 4.16

| Las clasificaciones incorrectas son marcadas con * | | | | | |
|--|-------------|----------|----------|----------|----------|
| | Grupo | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Estado | Clasificado | p=.25000 | p=.25000 | p=.25000 | p=.25000 |
| *Ags | G 3:3 | G 2:2 | G 3:3 | G 1:1 | G 4:4 |
| BC | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| BCS | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Cam | G 1:1 | G 1:1 | G 3:3 | G 2:2 | G 4:4 |
| Coah | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| Col | G 1:1 | G 1:1 | G 3:3 | G 2:2 | G 4:4 |
| Chia | G 3:3 | G 3:3 | G 1:1 | G 2:2 | G 4:4 |
| Chih | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| DF | G 4:4 | G 4:4 | G 3:3 | G 1:1 | G 2:2 |
| Dgo | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Gto | G 4:4 | G 4:4 | G 3:3 | G 1:1 | G 2:2 |
| Gro | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Hgo | G 4:4 | G 4:4 | G 3:3 | G 1:1 | G 2:2 |
| Jal | G 4:4 | G 4:4 | G 3:3 | G 1:1 | G 2:2 |
| Mex | G 4:4 | G 4:4 | G 1:1 | G 2:2 | G 3:3 |
| *Mich | G 1:1 | G 3:3 | G 2:2 | G 1:1 | G 4:4 |
| Mor | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Nay | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| NL | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| *Oax | G 3:3 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Pue | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| Qro | G 3:3 | G 3:3 | G 2:2 | G 1:1 | G 4:4 |
| QR | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| SLP | G 1:1 | G 1:1 | G 3:3 | G 2:2 | G 4:4 |
| Sin | G 3:3 | G 3:3 | G 2:2 | G 1:1 | G 4:4 |
| Son | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| Tab | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Tamps | G 2:2 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| *Tlax | G 1:1 | G 2:2 | G 1:1 | G 3:3 | G 4:4 |
| Ver | G 4:4 | G 4:4 | G 3:3 | G 1:1 | G 2:2 |
| Yuc | G 1:1 | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Zac | G 3:3 | G 3:3 | G 1:1 | G 2:2 | G 4:4 |

Por otra parte 22 de las 32 entidades federativas presentan probabilidades superiores al 0.90, lo que permite reconocer que la clasificación dada no es del todo errónea.

Así pues, en el cuadro 4.17 se pueden revisar las probabilidades a posteriori de que cada entidad federativa pertenezca a cada uno de los cuatro grupos.

Cuadro 4.17 Probabilidades a posteriori

| Las clasificaciones incorrectas son marcadas con * | | | | | |
|--|-------------|----------|----------|----------|----------|
| | Grupo | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| Estado | Clasificado | p=.25000 | p=.25000 | p=.25000 | p=.25000 |
| *Ags | G 3:3 | 0.16608 | 0.47441 | 0.35952 | 0.00000 |
| BC | G 2:2 | 0.37026 | 0.61624 | 0.01350 | 0.00000 |
| BCS | G 1:1 | 0.94653 | 0.04873 | 0.00475 | 0.00000 |
| Cam | G 1:1 | 0.92422 | 0.02049 | 0.05529 | 0.00000 |
| Coah | G 2:2 | 0.00291 | 0.99679 | 0.00030 | 0.00000 |
| Col | G 1:1 | 0.87912 | 0.01212 | 0.10876 | 0.00000 |
| Chia | G 3:3 | 0.00003 | 0.00000 | 0.99997 | 0.00000 |
| Chih | G 2:2 | 0.00074 | 0.99878 | 0.00047 | 0.00000 |
| DF | G 4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Dgo | G 1:1 | 0.60681 | 0.39082 | 0.00237 | 0.00000 |
| Gto | G 4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Gro | G 1:1 | 0.99336 | 0.00369 | 0.00294 | 0.00000 |
| Hgo | G 4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Jal | G 4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Mex | G 4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| *Mich | G 1:1 | 0.04400 | 0.36606 | 0.58994 | 0.00000 |
| Mor | G 1:1 | 0.91561 | 0.04405 | 0.04034 | 0.00000 |
| Nay | G 1:1 | 0.94233 | 0.04779 | 0.00988 | 0.00000 |
| NL | G 2:2 | 0.00523 | 0.99465 | 0.00012 | 0.00000 |
| *Oax | G 3:3 | 0.50898 | 0.40595 | 0.08507 | 0.00000 |
| Pue | G 2:2 | 0.04082 | 0.92900 | 0.03018 | 0.00000 |
| Qro | G 3:3 | 0.00011 | 0.00053 | 0.99936 | 0.00000 |
| QR | G 1:1 | 0.93038 | 0.06943 | 0.00019 | 0.00000 |
| SLP | G 1:1 | 0.99148 | 0.00335 | 0.00517 | 0.00000 |
| Sin | G 3:3 | 0.00003 | 0.00006 | 0.99991 | 0.00000 |
| Son | G 2:2 | 0.14689 | 0.82165 | 0.03146 | 0.00000 |
| Tab | G 1:1 | 0.96399 | 0.02530 | 0.01072 | 0.00000 |
| Tamps | G 2:2 | 0.48024 | 0.51511 | 0.00465 | 0.00000 |
| *Tlax | G 1:1 | 0.46793 | 0.52827 | 0.00381 | 0.00000 |
| Ver | G 4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Yuc | G 1:1 | 0.78242 | 0.20957 | 0.00801 | 0.00000 |
| Zac | G 3:3 | 0.02798 | 0.00042 | 0.97160 | 0.00000 |

A manera de resumen se tiene que el 87.5% de las entidades federativas fueron asignadas correctamente. Que el grupo que presentó mayores problemas fue el tres, ya que de los seis casos propuestos, existió evidencia para pensar que dos de ellos (33.33%) pertenecen a otro grupo.

Cuadro 4.18

| Renglones: clasificaciones observadas | | | | | |
|--|------------|----------|----------|----------|----------|
| Columnas: clasificaciones predictorias | | | | | |
| | Porcentaje | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| | Correcto | p=.25000 | p=.25000 | p=.25000 | p=.25000 |
| G 1:1 | 84.62 | 11 | 1 | 1 | 0 |
| G 2:2 | 100.00 | 0 | 7 | 0 | 0 |
| G 3:3 | 66.67 | 1 | 1 | 4 | 0 |
| G 4:4 | 100.00 | 0 | 0 | 0 | 6 |
| Total | 87.50 | 12 | 9 | 5 | 6 |

En la búsqueda de una segunda representación gráfica de las 32 entidades federativas se llevo a cabo el cálculo de los tres valores propios diferentes de cero de la matriz W^1B (sobre las variables estandarizadas), que según se ha visto permiten encontrar los coeficientes canónicos estandarizados.

Cuadro 4.19 coeficientes canónicos estandarizados

| | Raíz 1 | Raíz 2 | Raíz 3 |
|----------|------------|------------|------------|
| PDASI | 8.27582 | -6.15982 | -10.50078 |
| PDAP | 39.19228 | -41.09064 | -88.49212 |
| PDAS | 57.66061 | -63.81499 | -143.24367 |
| PDAMSYS | 67.93473 | -69.74622 | -141.47157 |
| BTA | -177.58183 | -119.76001 | -91.83952 |
| BTCP | -6.15079 | -5.42282 | -2.87617 |
| CE | 0.72082 | 289.22623 | 397.65176 |
| SE | 0.79464 | 33.74170 | 48.28972 |
| CESE | 11.00953 | -13.57219 | 3.56425 |
| TTT | 2.82378 | -4.83828 | 2.51248 |
| ICT | 11.13945 | -12.53519 | 4.83465 |
| Eigenval | 29.67973 | 0.96180 | 0.42236 |
| Cum.Prop | 0.95544 | 0.98640 | 1 |

De acuerdo al cuadro anterior con las dos primeras raíces se cuenta con el 98.7% de la representación de la variabilidad de las observaciones.

Nuevamente con base en los coeficientes canónicos no estandarizados el paquete calcula las tres raíces para cada una de las 32 entidades federativas. Utilizando posteriormente las dos primeras para obtener una gráfica de la situación descrita en los primeros párrafos de la presente sección.

Para obtener las tres raíces para cada entidad federativa el paquete realiza la multiplicación del vector de atributos de cada una de ellas por los correspondientes coeficientes canónicos no estandarizados.

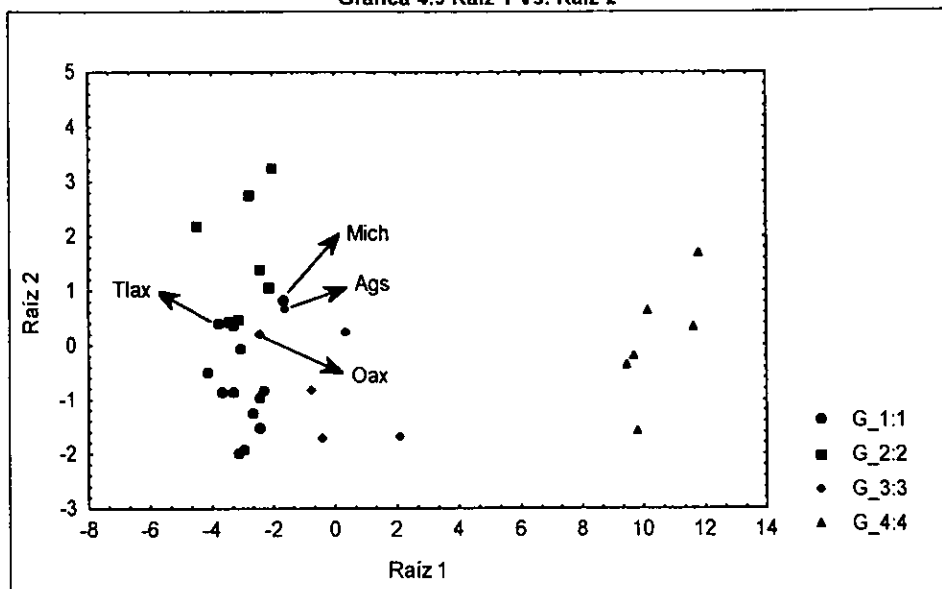
Cuadro 4.20 Puntajes canónicos no estandarizados

| | Grupo | Raíz 1 | Raíz 2 | Raíz 3 |
|-------|-------|----------|----------|----------|
| Ags | G 3:3 | -1.64377 | 0.67599 | 0.35329 |
| BC | G 2:2 | -3.17595 | 0.47770 | -0.00121 |
| BCS | G 1:1 | -3.66336 | -0.84537 | -0.02539 |
| Cam | G 1:1 | -2.66913 | -1.26027 | 0.06600 |
| Coah | G 2:2 | -2.81022 | 2.75140 | -0.36037 |
| Col | G 1:1 | -2.47622 | -1.51721 | 0.22978 |
| Chia | G 3:3 | 2.11543 | -1.66495 | 0.31154 |
| Chih | G 2:2 | -2.02923 | 3.23653 | -0.77871 |
| DF | G 4:4 | 9.69801 | -0.18202 | -0.36312 |
| Dgo | G 1:1 | -3.30947 | 0.36969 | -0.94065 |
| Gto | G 4:4 | 10.14285 | 0.63454 | 0.17897 |
| Gro | G 1:1 | -2.98560 | -1.89342 | -1.61946 |
| Hgo | G 4:4 | 9.80327 | -1.58554 | 0.99502 |
| Jal | G 4:4 | 9.44913 | -0.36253 | -1.20036 |
| Mex | G 4:4 | 11.79685 | 1.68451 | -1.36886 |
| Mich | G 1:1 | -1.68294 | 0.83827 | 1.28433 |
| Mor | G 1:1 | -2.33803 | -0.80252 | -0.70825 |
| Nay | G 1:1 | -3.34533 | -0.85469 | -0.03153 |
| NL | G 2:2 | -4.50442 | 2.19539 | 1.22642 |
| Oax | G 3:3 | -2.47303 | 0.20217 | 0.06903 |
| Pue | G 2:2 | -2.45392 | 1.38895 | 0.44720 |
| Qro | G 3:3 | 0.34573 | 0.24414 | 2.03176 |
| QR | G 1:1 | -4.13177 | -0.47933 | -1.53099 |
| SLP | G 1:1 | -3.18451 | -1.98205 | -0.73040 |
| Sin | G 3:3 | -0.74138 | -0.80414 | 4.09735 |
| Son | G 2:2 | -2.17733 | 1.07902 | -0.42523 |
| Tab | G 1:1 | -2.44030 | -0.96191 | -1.48082 |
| Tamps | G 2:2 | -3.45643 | 0.42441 | -0.22910 |
| Tlax | G 1:1 | -3.78678 | 0.41462 | 0.26131 |
| Ver | G 4:4 | 11.63278 | 0.33278 | 0.37824 |
| Yuc | G 1:1 | -3.07184 | -0.05227 | -0.58452 |
| Zac | G 3:3 | -0.43319 | -1.70181 | 0.44848 |

De la graficación de las dos primeras raíces se observa, como era de esperarse, que los grupos no están bien separados, es decir, existen individuos que habiendo sido asignados a un grupo en realidad pudieran pertenecer a otro.

Observando detenidamente la siguiente gráfica se ve que Michoacán y Aguascalientes se ubican en una zona que no corresponde al grupo en que fueron asignadas inicialmente. De manera similar, aunque es menos evidente ocurre, con Tlaxcala y Oaxaca.

Gráfica 4.9 Raíz 1 Vs. Raíz 2



Enseguida se procede a proponer la nueva agrupación y por consiguiente a la revisión de ésta mediante la obtención de cuatro de los seis cuadros que hasta ahora se habían presentado.

Llevando a cabo la reasignación de individuos sugerida, se encontró que Michoacán no debería de reasignarse al grupo 3, sino permanecer en el 2 y que las otras tres entidades (Aguascalientes, Oaxaca y Tlaxcala) deberían de cambiar de grupo.

De acuerdo a lo anterior, en esta ocasión el paquete no reporta ninguna observación como mal clasificada.

Al revisar las probabilidades a posteriori se observa que de los 32 casos, 28 de ellas son mayores a 0.9; por otra parte Durango y Tamaulipas la aumentan al pasar de 0.61 y 0.52 a 0.70 y 0.82 respectivamente; finalmente las entidades con probabilidades menores son Durango con 0.65 y Michoacán con 0.77.

Cuadro 4.21 Probabilidades a posteriori

| Las clasificaciones incorrectas son marcadas con * | | | | | |
|--|-------------|----------|----------|----------|----------|
| | Grupo | G_1:1 | G_2:2 | G_3:3 | G_4:4 |
| Estado | Clasificado | p=.25000 | p=.25000 | p=.25000 | p=.25000 |
| Ags | G_2:2 | 0.00486 | 0.99391 | 0.00124 | 0.00000 |
| BC | G_2:2 | 0.08450 | 0.91550 | 0.00000 | 0.00000 |
| BCS | G_1:1 | 0.94623 | 0.05377 | 0.00000 | 0.00000 |
| Cam | G_1:1 | 0.96024 | 0.03957 | 0.00019 | 0.00000 |
| Coah | G_2:2 | 0.00068 | 0.99932 | 0.00001 | 0.00000 |
| Col | G_1:1 | 0.97502 | 0.02352 | 0.00145 | 0.00000 |
| Chia | G_3:3 | 0.00000 | 0.00000 | 1.00000 | 0.00000 |
| Chih | G_2:2 | 0.00136 | 0.99864 | 0.00000 | 0.00000 |
| DF | G_4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Dgo | G_1:1 | 0.65779 | 0.34221 | 0.00000 | 0.00000 |
| Gto | G_4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Gro | G_1:1 | 0.99991 | 0.00009 | 0.00000 | 0.00000 |
| Hgo | G_4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Jal | G_4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Mex | G_4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Mich | G_3:3 | 0.00002 | 0.22958 | 0.77040 | 0.00000 |
| Mor | G_1:1 | 0.98067 | 0.01921 | 0.00012 | 0.00000 |
| Nay | G_1:1 | 0.95594 | 0.04406 | 0.00000 | 0.00000 |
| NL | G_2:2 | 0.00240 | 0.99760 | 0.00000 | 0.00000 |
| Oax | G_2:2 | 0.02607 | 0.97392 | 0.00001 | 0.00000 |
| Pue | G_2:2 | 0.00029 | 0.99970 | 0.00001 | 0.00000 |
| Qro | G_3:3 | 0.00000 | 0.00000 | 1.00000 | 0.00000 |
| QR | G_1:1 | 0.99821 | 0.00179 | 0.00000 | 0.00000 |
| SLP | G_1:1 | 0.99964 | 0.00036 | 0.00000 | 0.00000 |
| Sin | G_3:3 | 0.00000 | 0.00000 | 1.00000 | 0.00000 |
| Son | G_2:2 | 0.05686 | 0.94100 | 0.00214 | 0.00000 |
| Tab | G_1:1 | 0.98986 | 0.01014 | 0.00000 | 0.00000 |
| Tamps | G_2:2 | 0.29654 | 0.70346 | 0.00000 | 0.00000 |
| Tlax | G_2:2 | 0.17591 | 0.82409 | 0.00000 | 0.00000 |
| Ver | G_4:4 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |
| Yuc | G_1:1 | 0.90325 | 0.09675 | 0.00000 | 0.00000 |
| Zac | G_3:3 | 0.00138 | 0.00006 | 0.99856 | 0.00000 |

Cuadro 4.22

| Renglonas: clasificaciones observadas | | Columnas: clasificaciones predictorias | | | |
|---------------------------------------|------------|--|----------|----------|----------|
| | Porcentaje | G 1:1 | G 2:2 | G 3:3 | G 4:4 |
| | Correcto | p=.25000 | p=.25000 | p=.25000 | p=.25000 |
| G 1:1 | 100 | 11 | 0 | 0 | 0 |
| G 2:2 | 100 | 0 | 10 | 0 | 0 |
| G 3:3 | 100 | 0 | 0 | 5 | 0 |
| G 4:4 | 100 | 0 | 0 | 0 | 6 |
| Total | 100 | 11 | 10 | 5 | 6 |

Así pues el cuadro 4.22 muestra que el 100% de las entidades federativas están bien asignadas a los correspondientes grupos. En resumen los únicos cambios efectuados fueron los siguientes.

1. Aguascalientes que inicialmente se localizó en el grupo 3, se reasignó al grupo 2.
2. Oaxaca pasó del 3 al grupo 1.
3. Tlaxcala del 1 al 2.

Cuadro 4.23 Coeficientes canónicos estandarizados

| | Raíz 1 | Raíz 2 | Raíz 3 |
|----------|------------|------------|------------|
| PDASI | 8.68001 | -12.96774 | 5.20906 |
| PDAP | 31.36564 | -94.71075 | 50.14714 |
| PDAS | 44.52666 | -154.58562 | 81.04042 |
| PDAMSYS | 59.91432 | -157.99803 | 74.01227 |
| BTA | -291.88272 | -126.54491 | 86.18885 |
| BTCP | -10.23505 | -5.24456 | 2.58667 |
| CE | 118.78784 | 490.88760 | -234.44855 |
| SE | 15.43030 | 59.25284 | -28.60792 |
| CESE | 15.83177 | -11.01895 | -8.91798 |
| TTT | 5.05311 | -3.48804 | -4.74512 |
| ICT | 18.84135 | -11.26281 | -12.60358 |
| Eigenval | 45.32247 | 1.53342 | 0.98855 |
| Cum.Prop | 0.94729 | 0.97934 | 1 |

De manera similar a lo encontrado en el cuadro 4.19 se tiene que con las dos primeras raíces se representa alrededor del 98% de la variabilidad de las 32 observaciones.

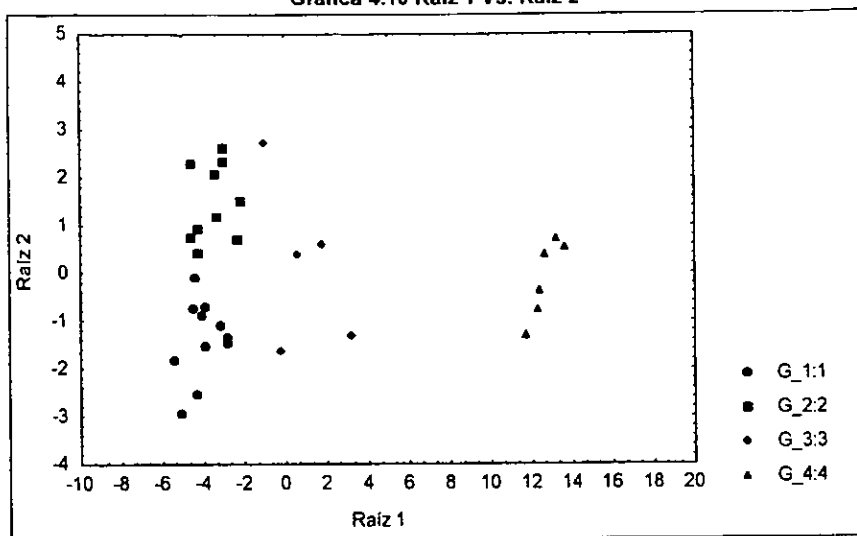
En seguida se presentan los puntajes no estandarizados para cada una de las 32 entidades federativas. Seguidos de la gráfica de las dos primeras raíces. En este caso se espera una separación casi total entre los grupos, dados los resultados de los cuadros anteriores.

Cuadro 4.24 Puntajes canónicos no estandarizados

| | Grupo | Raíz 1 | Raíz 2 | Raíz 3 |
|-------|-------|----------|----------|----------|
| Ags | G 2:2 | -2.15057 | 1.52697 | 0.29994 |
| BC | G 2:2 | -4.27145 | 0.92950 | 0.60083 |
| BCS | G 1:1 | -4.53587 | -0.73768 | -0.30267 |
| Cam | G 1:1 | -3.19770 | -1.09645 | -0.60863 |
| Coah | G 2:2 | -3.04722 | 2.33959 | 0.67554 |
| Col | G 1:1 | -2.85466 | -1.33154 | -0.91217 |
| Chia | G 3:3 | 3.22656 | -1.28356 | -1.36668 |
| Chih | G 2:2 | -3.42910 | 2.07875 | 2.42021 |
| DF | G 4:4 | 12.37491 | -0.36256 | -0.18723 |
| Dgo | G 1:1 | -4.46943 | -0.06818 | 1.01702 |
| Gto | G 4:4 | 12.59534 | 0.38577 | 0.11261 |
| Gro | G 1:1 | -5.12976 | -2.92855 | 1.63488 |
| Hgo | G 4:4 | 12.29013 | -0.74325 | -1.13382 |
| Jal | G 4:4 | 11.67040 | -1.29555 | 0.55159 |
| Mex | G 4:4 | 13.62902 | 0.53898 | 2.40027 |
| Mich | G 3:3 | -0.97084 | 2.74434 | -1.33670 |
| Mor | G 1:1 | -2.89748 | -1.44180 | 0.01367 |
| Nay | G 1:1 | -4.13659 | -0.88370 | -0.33208 |
| NL | G 2:2 | -4.61875 | 2.31029 | -1.00623 |
| Oax | G 2:2 | -3.36206 | 1.18236 | 0.52147 |
| Pue | G 2:2 | -3.05400 | 2.62635 | 0.46463 |
| Qro | G 3:3 | 1.78911 | 0.62450 | -2.57292 |
| QR | G 1:1 | -5.46208 | -1.79798 | 0.87414 |
| SLP | G 1:1 | -4.36961 | -2.51264 | 0.06135 |
| Sin | G 3:3 | 0.57992 | 0.38819 | -4.26283 |
| Son | G 2:2 | -2.35714 | 0.71717 | 0.21657 |
| Tab | G 1:1 | -3.98933 | -1.52501 | 1.48859 |
| Tamps | G 2:2 | -4.32232 | 0.44631 | 0.32342 |
| Tlax | G 2:2 | -4.60260 | 0.76007 | -0.06791 |
| Ver | G 4:4 | 13.22934 | 0.72370 | 1.30416 |
| Yuc | G 1:1 | -3.92522 | -0.67443 | 0.33579 |
| Zac | G 3:3 | -0.23096 | -1.63972 | -1.22670 |

En la siguiente gráfica se observa una separación casi total en los grupos. En esta ocasión no se observa traslape de ellos.

Gráfica 4.10 Raíz 1 Vs. Raíz 2



Finalmente se presenta la agrupación obtenida (validada) de acuerdo al segundo conjunto de variables.

GRUPOS

- I. Baja California Sur, Campeche, Colima, Durango, Guerrero, Michoacán, Morelos, Nayarit, Oaxaca, Quintana Roo, San Luis Potosí, Tabasco, Yucatán.
- II. Aguascalientes, Baja California, Coahuila, Chihuahua, Nuevo León, Puebla, Sonora, Tamaulipas, Tlaxcala.
- III. Chiapas, Querétaro, Sinaloa, Zacatecas.
- IV. Distrito Federal, Guanajuato, Hidalgo, Jalisco, México, Veracruz.

De acuerdo a la nueva agrupación 21 de las 32 (65.6%) entidades federativas permanecen en el mismo grupo de acuerdo a los dos conjuntos de variables.

5. Conclusiones.

En la actualidad suele pensarse que la persona que trabaja, recibe forzosamente a cambio un pago, cosa que no siempre es cierta. Por ello es importante conceptualizar el término para evitar ambigüedad; pero en general, el trabajo cualquiera que sea su contexto responde primordialmente a tres necesidades: subsistir, crear y colaborar.

De acuerdo a lo expuesto en el capítulo uno, desde un punto de vista económico un empleo supone la realización de un trabajo, por lo que en muchas ocasiones estos términos se usan como sinónimos.

De gran importancia resulta el estudio de temas económicos como el empleo, ya que las condiciones en que se encuentre este, se reflejarán en los niveles de vida de la población. Por lo anterior es importante contar con bases de información confiables, apegadas a la realidad.

Por lo regular, falta de empleos y salarios bajos son resultado de políticas económicas adoptadas por el estado; mientras que en otros casos son debidos a factores de la economía mundial.

En relación con lo anterior es de llamar la atención, algunas de las definiciones operacionales del empleo manejadas por INEGI. Por ejemplo, que se considere:

1. Empleados a los trabajadores familiares no remunerados, que han trabajado un tercio del tiempo normal de trabajo.
2. Ocupadas a las personas que trabajaron una hora o un día en la semana, o las que declararon iniciaran alguna ocupación en el término de un mes o menos.
3. Económicamente activos, a quienes buscaron incorporarse a alguna actividad económica en el mes previo a la semana de levantamiento.

Un ejercicio interesante, sería determinar que efectos causan las anteriores consideraciones en el calculo de indicadores económicos.

Por otra parte, a pesar de que cada vez es mas frecuente la incursión de la mujer en el mercado de trabajo; las actividades extradomésticas continúan siendo desempeñadas primordialmente por los hombres, para 1998 en México por cada mujer económicamente activa existían dos hombres en la misma situación. Esto debido en gran medida a una sociedad cuya cultura, tradiciones y costumbres segrega a la mujer destinándole un papel en primer lugar reproductor.

Es importante reconocer que a través del tiempo, en especial en los últimos 20 años, se ha dado un cambio en la percepción sobre el papel que juegan las mujeres en la sociedad. A causa de las condiciones generadas en periodos de crisis, que han hecho necesaria la incursión de mujeres y de los grupos de edad más jóvenes para asegurar la subsistencia de la población. Sin embargo se dista mucho de una igualdad laboral entre hombres y mujeres.

Otro punto importante de resaltar, es que México es un país que concentra a la mayor parte de su población económicamente activa en el sector terciario, que una parte importante de los empleos son generados por los mismos individuos, que al no

encontrarlos disponibles en el mercado de trabajo, son orillados a incorporarse por cuenta propia, a este sector principalmente.

Por su complejidad el tema del trabajo requiere para su tratamiento de un número importante de variables e indicadores.

Instrumentos imprescindibles para cualquier análisis estadístico en el que intervengan dos o más variables, los métodos estadísticos multivariados permiten el estudio de fenómenos que por su naturaleza requieren de dos o más de ellas. Con ellos se busca dar respaldo empírico a planteamientos teóricos y, por otro lado, analizar explicaciones teóricas sugeridas por los datos mismos.

Cabe mencionar que la determinación de los grupos de estados pudo llevarse a cabo mediante otras técnicas; por ejemplo mediante el análisis de clusters. Conformado por dos tipos de métodos: jerárquicos y no jerárquicos.

La determinación de grupos de estados de la república con base en variables económicas puede encontrar utilidad en los siguientes casos.

- 1) En primer lugar se obtiene un panorama general de la situación laboral de la población mexicana, siendo posible realizar comparaciones entre estados de la República.
- 2) Asignación de recursos (mediante la obtención de índices de marginación y bienestar social).
- 3) Implementación de programas y medidas económicas que en algunas entidades hallan dado buenos resultados. Permitiendo así mejorar las condiciones del empleo de los individuos y por consiguiente los niveles de bienestar.
- 4) Se permitiría una posible contrastación de las agrupaciones propuestas, con otras definidas después de transcurrido un cierto periodo.

- 5) El mismo procedimiento pudiera ser aplicado a rubros como: salud, educación, seguridad, etcétera.

Como era de esperarse al obtener las agrupaciones, el Distrito Federal, Jalisco y México son casos extremos aislados, en relación a los dos conjuntos de características consideradas de la población masculina de cada entidad federativa.

A excepción de Veracruz y las tres entidades mencionadas en el párrafo anterior se tienen grupos muy homogéneos respecto a las variables referentes a la PEA; mientras que con relación a la PDA se obtiene a un inicio un traslape de 25 entidades, lo que habla de que el desempleo abierto es muy similar en toda la república mexicana. En ambos casos México y Baja California Sur son los estados más diferentes.

Otras situaciones que llaman la atención son:

1. Que Nuevo León no sea un caso del tipo del Distrito Federal.
2. Que Guanajuato este cerca de los niveles de Jalisco.
3. Que Veracruz no quede contenido en el grupo de los estados considerados pobres.

Las representaciones simplificadas de grandes tablas de datos, que los dos métodos empleados permiten obtener, se han manifestado como un instrumento de síntesis notable.

En este trabajo en especial las representaciones gráficas del análisis de componentes principales (usado en un sentido descriptivo) y discriminante (en un sentido confirmatorio) son un medio de comunicación valioso ya que no es necesario ser un especialista en estadística para comprender que la proximidad entre dos puntos traduce la semejanza entre las entidades que representa sin que sea necesario entender la formalización matemática de esta semejanza.

ANEXO I

La encuesta Nacional de Empleo 1998, presenta información estadística sobre empleo y desempleo con diferentes niveles de cobertura geográfica: nacional, áreas más urbanizadas (localidades con más de 100 mil habitantes), áreas menos urbanizadas (localidades con menos de 100 mil habitantes) y por entidad federativa.

Mediante esta edición, por segunda vez se presenta la información a nivel entidad federativa, gracias a las ventajas que representa utilizar el método de encuestas para captar información sobre actividades económicas de la población. Esto fue posible debido a un incremento importante en el tamaño de muestra utilizada ¹ y al diseño de la misma, de tal manera que se publicaron 32 volúmenes correspondientes a cada una de las entidades federativas del país.

El objetivo general de la ENE es contar con una base de información estadística sobre las principales características demográficas y económicas de la población, profundizando sobre aspectos laborales que sean de utilidad para los sectores públicos, privados y social.

El diseño muestral de la ENE se caracteriza por ser probabilístico, por lo cual los resultados obtenidos de la encuesta se generalizan a toda la población. A su vez es polietápico, estratificado y por conglomerados, donde la unidad última de selección es la vivienda y la unidad de observación es la vivienda.

- a) Probabilístico. Esto significa que las unidades de selección tienen una probabilidad conocida y distinta de cero de entrar en la muestra para cada miembro de la población, lo cual, permite conocer la precisión de los resultados muestrales.
- b) Estratificado. Porque las unidades de selección son agrupadas por características similares de tipo geográficas y socioeconómicas.
- c) Polietápico. Por que la unidad última de selección (vivienda) es seleccionada después de varias etapas.
- d) Por conglomerados. Porque previamente se conforman conjuntos de unidades muestrales de los cuales se obtiene una muestra.

¹ Para 1998 el tamaño de muestra fue de 140, 824 viviendas a nivel nacional.

El marco de muestreo de la ENE-98 es el marco muestral de propósitos múltiples del INEGI, construido a partir de la información demográfica y cartográfica obtenida durante el Censo de Población y Vivienda de 1995.²

Cuadro 1. Indicadores por entidad federativa; población masculina 1998.

| Entidad | P>12 | PEA | PO | POEPI | PDA |
|---------------------|------------|------------|------------|------------|---------|
| Aguascalientes | 313,657 | 235,653 | 229,284 | 228,548 | 6,369 |
| Baja California | 833,473 | 642,722 | 632,674 | 628,228 | 10,048 |
| Baja California Sur | 155,990 | 120,556 | 119,118 | 118,465 | 1,438 |
| Campeche | 242,516 | 186,445 | 183,943 | 183,602 | 2,502 |
| Coahuila | 788,825 | 621,160 | 604,897 | 598,684 | 16,263 |
| Colima | 178,594 | 139,460 | 137,792 | 137,067 | 1,668 |
| Chiapas | 1,236,061 | 1,092,759 | 1,083,146 | 1,080,061 | 9,613 |
| Chihuahua | 1,050,324 | 806,655 | 794,034 | 790,366 | 12,621 |
| Distrito Federal | 3,169,046 | 2,395,552 | 2,307,155 | 2,298,900 | 88,397 |
| Durango | 501,578 | 394,804 | 388,649 | 386,644 | 6,155 |
| Guanajuato | 1,497,316 | 1,143,873 | 1,122,200 | 1,116,019 | 21,673 |
| Guerrero | 991,915 | 818,225 | 809,682 | 809,178 | 8,543 |
| Hidalgo | 747,318 | 603,848 | 588,884 | 587,626 | 14,964 |
| Jalisco | 2,179,040 | 1,791,381 | 1,758,528 | 1,747,785 | 32,853 |
| México | 4,496,165 | 3,458,841 | 3,347,711 | 3,305,277 | 111,130 |
| Michoacán | 1,243,944 | 992,311 | 984,711 | 978,958 | 7,600 |
| Morelos | 543,992 | 425,609 | 420,139 | 419,912 | 5,470 |
| Nayarit | 335,464 | 271,217 | 269,477 | 267,286 | 1,740 |
| Nuevo León | 1,422,614 | 1,104,424 | 1,083,970 | 1,076,466 | 20,454 |
| Oaxaca | 1,128,235 | 902,449 | 896,035 | 894,618 | 6,414 |
| Puebla | 1,616,538 | 1,243,263 | 1,228,586 | 1,225,073 | 14,677 |
| Querétaro | 445,947 | 334,290 | 327,021 | 324,828 | 7,269 |
| Quintana Roo | 291,885 | 234,813 | 232,280 | 231,761 | 2,533 |
| San Luis Potosí | 737,499 | 552,656 | 545,883 | 544,738 | 6,773 |
| Sinaloa | 906,419 | 706,081 | 694,874 | 687,235 | 11,207 |
| Sonora | 800,464 | 631,712 | 615,704 | 606,885 | 16,008 |
| Tabasco | 659,961 | 539,366 | 529,267 | 526,035 | 10,099 |
| Tamaulipas | 934,233 | 730,515 | 714,503 | 707,478 | 16,012 |
| Tlaxcala | 356,725 | 274,806 | 269,928 | 268,223 | 4,878 |
| Veracruz | 2,405,518 | 1,979,176 | 1,948,171 | 1,940,366 | 31,005 |
| Yucatán | 589,150 | 455,668 | 451,100 | 449,519 | 4,568 |
| Zacatecas | 450,342 | 345,736 | 343,727 | 340,496 | 2,009 |
| Nacional | 33,250,748 | 26,176,026 | 25,663,073 | 25,506,327 | 512,953 |

P>12 . Población de 12 años y más.

PEA. Población económicamente activa.

PO. Población ocupada.

POEPI. Población ocupada excluyendo a los iniciadores de un próximo trabajo.

PDA. Población desocupada abierta.

² Toda la información mostrada en el presente anexo fue tomada de la Encuesta Nacional de Empleo 1998.

ANEXO II

Teorema AII-1 Sea A una matriz de p renglones por p columnas ($p \times p$). Si A es simétrica entonces existe una matriz ortogonal P tal que A pueda descomponerse como: $A = P\Lambda P^t$ donde Λ es una matriz diagonal de los valores propios de A y P cumple que $PP^t = I_p$.

Demostración:

Supongase que se pueden encontrar los vectores ortonormales $r_{(1)}, \dots, r_{(p)}$ (vectores columna) tal que $Ar_{(i)} = \lambda_i r_{(i)}, \quad \forall i = 1, \dots, p$

$$r_{(i)}^t A r_{(i)} = r_{(i)}^t \lambda_i r_{(i)} = \lambda_i r_{(i)}^t r_{(i)} = \begin{cases} \lambda_i & i = j \\ 0 & i \neq j \end{cases}$$

En forma matricial se tiene que:

$$P^t A P = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix} = \Lambda \dots (1)$$

Donde $P = [r_{(1)} | r_{(2)} | \dots | r_{(p)}]$

Recuérdese que si C es una matriz de $p \times p$ no singular

$$|A - \lambda I| = |C||A - \lambda C^{-1}C||C^{-1}| = |CAC^{-1} - \lambda I|$$

Es decir A y CAC^{-1} tienen los mismos valores propios.

Por construcción $P^t P = I = P P^t$ de aquí que $P^{-1} = P^t$ Por lo que A y Λ tienen los

mismos valores propios.

Multiplicando por la izquierda por P y por P' la derecha de cada miembro de (1) se

tiene que:

$$P(P'AP)P' = P\Lambda P'$$

$$I_p A I_p = P\Lambda P'$$

$$A = P\Lambda P'$$

Corolario Si A es una matriz no singular, simétrica, entonces para cualquier entero n se tiene que: $\Lambda^n = \text{diag}(\lambda_i^n)$ y $A^n = \Gamma\Lambda^n\Gamma'$. Además si todos los eigenvalores de A son positivos entonces se puede definir para potencias racionales $A^{rs} = \Gamma\Lambda^{rs}\Gamma'$ donde $\Lambda^{rs} = \text{diag}(\lambda_i^{rs})$ para enteros $s > 0$ y r . Si algunos de los eigenvalores de A son cero, entonces los resultados son ciertos si los exponentes son restringidos para ser no negativos.

Demostración:

Ya que:

$$A^2 = (\Gamma\Lambda\Gamma')^2 = \Gamma\Lambda\Gamma'\Gamma\Lambda\Gamma' = \Gamma\Lambda^2\Gamma'$$

$$A^{-1} = \Gamma\Lambda^{-1}\Gamma', \Lambda^{-1} = \text{diag}(\lambda_i^{-1})$$

Se puede ver que la primera parte puede ser probada fácilmente por inducción.

$$(A^{r/s})^s = (\Gamma\Lambda^{r/s}\Gamma^t)(\Gamma\Lambda^{r/s}\Gamma^t)\dots(\Gamma\Lambda^{r/s}\Gamma^t) = \Gamma\Lambda^r\Gamma^t = A^r$$

Teorema AII-2 Para $A_{(n \times p)}$ y $B_{(p \times n)}$, los valores propios no nulos de AB y BA son los mismos y tienen la misma multiplicidad. Además si X es un eigenvector no trivial de AB para un eigenvalor $\lambda \neq 0$, entonces $Y = BX$ es un eigenvector no trivial de BA .

Demostración:

Supongase que $n \geq p$ entonces se tiene que:

$$\begin{vmatrix} -\lambda I_n & -A \\ B & I_p \end{vmatrix} = (-\lambda)^{n-p} |BA - \lambda I_p| = |AB - \lambda I_n|$$

Por lo tanto de los n eigenvalores de AB p son iguales a los eigenvalores de BA ,

y los restantes $n - p$ eigenvalores de AB son cero.

Para la segunda parte sustituyendo $Y = BX$ en la ecuación $B(ABX) = \lambda BX$ se tiene

que $BAY = \lambda Y$, por lo tanto, X es un eigenvector no trivial de BA . Ya que

$AY = ABX = \lambda X \neq 0$ se sigue que $y \neq 0$ también.

Definición A1 Sean X_1, X_2, \dots, X_p variables aleatorias con matriz de covarianzas Σ , ω_i , ω_j individuos con vectores de atributos: $X_i^t = (x_{i1}, \dots, x_{im})$, $X_j^t = (x_{j1}, \dots, x_{jm})$, supongase Σ no singular, se define la distancia (al cuadrado) de Mahalanobis entre ω_i y ω_j como:

$$D^2(\omega_i, \omega_j) = (X_i - X_j)^t \Sigma^{-1} (X_i - X_j)$$

$$D^2(\omega_i, \omega_j) = (X_i - X_j)' \Sigma^{-1} (X_i - X_j)$$

Propiedades

1) $D^2(\omega_i, \omega_j) = D^2(\omega_j, \omega_i)$

2) $D^2(\omega_i, \omega_j) \geq 0$, $D^2(\omega_i, \omega_j) = 0 \Leftrightarrow \omega_i = \omega_j$

3) $D^2(\omega_i, \omega_j) \leq D^2(\omega_i, \omega_k) + D^2(\omega_k, \omega_j)$

4) Es invariante bajo transformaciones lineales no singulares de las variables. En particular es invariante por cambios de escala.

5) esta expresada en unidades de desviación típica y e tiene en cuenta las correlaciones (redundancia) entre las variables.

6) Si se indica D_p^2 la distancia al cuadrado en función de p variables, entonces: $D_p^2 \leq D_{p+m}^2$

7) Si las variables X_1, X_2, \dots, X_p son estocasticamente independientes de las variables X_{p+1}, \dots, X_{p+m} entonces: $D_{p+m}^2 = D_p^2 + D_m^2$

ANEXO III

Posición en el trabajo. Clasificación que se hace de la población ocupada de acuerdo a la relación que guardan con los medios de producción y la propiedad de los bienes y servicios generados en el desempeño de su trabajo.

Empleador. Es aquella persona que trabaja en su propio negocio, empresa o establecimiento o ejerce un oficio o profesión, empleando a una o más personas a cambio de una remuneración monetaria o en especie.

Trabajadores por su cuenta. Personas que trabajan solas o apoyándose exclusivamente con fuerza de trabajo no pagada.

Trabajadores asalariados. Personas que trabajan para un patrón, empresa, negocio, institución o dependencia, regidos por un contrato escrito o de palabra a cambio de un jornal, sueldo o salario.

Trabajadores a destajo. Son aquellas personas que trabajan por una remuneración que está determinada exclusivamente por la cantidad de trabajo. Es decir, por el número de servicios, piezas, trabajadas u obra terminada.

Trabajadores sin pago. Son las personas que participan en actividades que sirven de apoyo a su sustento o al de su familias, sin percibir a cambio alguna remuneración monetaria o en bienes susceptibles de intercambiar en el mercado.

Forma de pago. Hace referencia al carácter de los ingresos que recibieron los ocupados (excepto iniciadores de un próximo trabajo) en la semana de referencia, a cambio de su trabajo principal.

Salario fijo. Son aquellas personas que reciben una remuneración fija, establecida a través de un convenio hablado o de un contrato firmado.

Ingreso variable. Son aquellas personas que reciben una remuneración por su trabajo; dependiendo de la cantidad y/o tiempo de trabajo (por horas o días), a destajo comisión, porcentaje y propinas o pagos en especie.

Ganancias. Esta compuesto por cooperativistas, empleadores o dueños de negocios que invierten en su empresa o negocio y por ello obtienen ciertas ganancias; así como también a quienes no obtuvieron su cuota de ganancias porque trabajaron con números rojos o tuvieron pérdidas.

La familia consume lo que produce. Son aquellas personas que destinan una parte de lo que producen (maíz, frijol, gallinas, etcétera) o venden para el consumo de ellas o de sus familias y el resto lo venden; o bien cuando un patrón prestador de servicios utiliza parte de esos para uso de su familia sin pago alguno.

No recibe ingresos. Son aquellas personas que no reciben ningún tipo de remuneración por la realización de su trabajo.

Nivel de ingresos. Mide la percepción monetaria mensual de los ocupados (excepto iniciadores de un próximo trabajo) en términos del salario mínimo vigente en el trimestre de levantamiento de la encuesta. Se incluye a los trabajadores sin pago en el rubro "no recibe ingresos" para mantener el mismo universo de los ocupados.

Rangos de salario mínimo. Agrupación del ingreso de los trabajadores con relación a los salarios mínimos urbanos vigentes según la región del país de la que se trate. Durante el segundo trimestre de 1998, dichos salarios mínimos diarios eran \$30.20, \$28.00 y \$26.05 en las zonas salariales A, B y C respectivamente.

| | TOTAL | | SEGUN POSICION EN EL TRABAJO | | | | | | TSP | Otros |
|---------------------|-----------|-----------|------------------------------|---------|-----------|---------|---------|-------|-----|-------|
| | PEA | PO | E | TSC | TA | TD | TSP | | | |
| Aguascalientes | 235,653 | 229,284 | 12,150 | 43,104 | 142,381 | 20,113 | 11,536 | 0 | | |
| Baja California | 642,722 | 632,674 | 49,680 | 97,837 | 446,990 | 26,399 | 8,589 | 3,179 | | |
| Baja California Sur | 120,556 | 119,118 | 7,303 | 21,479 | 69,780 | 12,709 | 4,851 | 2,996 | | |
| Campeche | 186,445 | 183,943 | 9,636 | 61,849 | 80,842 | 14,441 | 17,124 | 51 | | |
| Coahuila | 621,160 | 604,897 | 30,384 | 119,564 | 401,852 | 19,867 | 33,193 | 37 | | |
| Colima | 139,460 | 137,792 | 11,548 | 22,327 | 77,052 | 21,125 | 5,630 | 110 | | |
| Chiapas | 1,092,759 | 1,083,146 | 52,126 | 391,466 | 341,196 | 43,359 | 254,999 | 0 | | |
| Chihuahua | 806,655 | 794,034 | 49,708 | 191,220 | 473,472 | 38,634 | 41,000 | 0 | | |
| Distrito Federal | 2,395,552 | 2,307,155 | 166,362 | 452,972 | 1,405,284 | 219,964 | 62,573 | 0 | | |
| Durango | 394,804 | 388,649 | 19,511 | 104,227 | 203,394 | 20,154 | 41,192 | 171 | | |
| Guanajuato | 1,143,873 | 1,122,200 | 73,960 | 282,448 | 541,422 | 146,552 | 77,718 | 100 | | |
| Guerrero | 618,225 | 609,682 | 31,787 | 283,809 | 270,928 | 47,962 | 175,196 | 0 | | |
| Hidalgo | 603,848 | 588,884 | 30,314 | 164,431 | 270,347 | 54,612 | 69,180 | 0 | | |
| Jalisco | 1,791,381 | 1,758,528 | 127,736 | 390,308 | 922,977 | 169,819 | 147,688 | 0 | | |
| México | 3,458,841 | 3,347,711 | 144,721 | 726,285 | 1,994,270 | 283,931 | 195,893 | 2,611 | | |
| Michoacan | 992,311 | 984,711 | 69,186 | 320,536 | 367,417 | 70,954 | 153,840 | 2,778 | | |
| Morelos | 425,609 | 420,139 | 33,646 | 97,343 | 230,414 | 30,217 | 28,519 | 0 | | |
| Nayarit | 271,217 | 269,477 | 23,272 | 80,300 | 102,978 | 31,235 | 31,692 | 0 | | |
| Nuevo León | 1,104,424 | 1,083,970 | 59,747 | 215,942 | 722,549 | 38,271 | 47,461 | 0 | | |
| Oaxaca | 902,449 | 896,035 | 28,085 | 364,868 | 329,397 | 28,344 | 144,534 | 807 | | |
| Puebla | 1,243,263 | 1,228,586 | 46,773 | 323,360 | 663,849 | 74,399 | 130,205 | 0 | | |
| Queretaro | 334,290 | 327,021 | 18,782 | 76,129 | 187,498 | 21,010 | 23,602 | 0 | | |
| Quintana Roo | 234,813 | 232,280 | 11,380 | 48,954 | 131,611 | 26,309 | 13,747 | 279 | | |
| San Luis Potosi | 552,656 | 545,883 | 34,349 | 160,237 | 295,356 | 18,882 | 37,059 | 0 | | |
| Sinaloa | 706,081 | 694,874 | 39,973 | 173,221 | 373,944 | 65,609 | 41,304 | 823 | | |
| Sonora | 631,712 | 615,704 | 44,220 | 92,562 | 410,751 | 50,014 | 18,045 | 112 | | |
| Tabasco | 539,366 | 529,267 | 25,794 | 133,418 | 282,916 | 26,595 | 60,303 | 241 | | |
| Tamaulipas | 730,515 | 714,503 | 37,548 | 181,109 | 413,312 | 37,503 | 44,224 | 807 | | |
| Tlaxcala | 274,806 | 269,928 | 10,623 | 73,591 | 138,349 | 21,046 | 26,145 | 174 | | |
| Veracruz | 1,979,176 | 1,948,171 | 79,367 | 574,239 | 852,267 | 152,970 | 288,531 | 797 | | |
| Yucatan | 455,668 | 451,100 | 29,593 | 137,638 | 219,457 | 31,060 | 33,352 | 0 | | |
| Zacatecas | 345,736 | 343,727 | 19,846 | 96,342 | 161,516 | 24,167 | 40,792 | 1,064 | | |

A-III

| | POEPI | SEGUN FORMA DE PAGO | | | LFCLP | NRI | No Esp. |
|---------------------|-----------|---------------------|---------|---------|---------|---------|---------|
| | | SF | IV | G | | | |
| Aguascalientes | 228,548 | 135,073 | 28,361 | 45,388 | 7,987 | 11,536 | 203 |
| Baja California | 628,228 | 429,534 | 73,499 | 116,346 | 117 | 8,527 | 205 |
| Baja California Sur | 118,465 | 67,694 | 15,575 | 30,255 | 0 | 4,851 | 90 |
| Campeche | 183,602 | 78,873 | 23,067 | 47,103 | 17,314 | 17,124 | 121 |
| Coahuila | 598,684 | 388,703 | 33,228 | 142,637 | 1,265 | 32,678 | 173 |
| Colima | 137,067 | 74,094 | 26,755 | 29,774 | 728 | 5,630 | 86 |
| Chiapas | 1,080,061 | 322,998 | 74,206 | 376,450 | 51,456 | 254,951 | 0 |
| Chihuahua | 790,366 | 468,532 | 76,536 | 191,117 | 13,042 | 41,000 | 139 |
| Distrito Federal | 2,298,900 | 1,367,813 | 339,415 | 528,350 | 1,749 | 62,573 | 0 |
| Durango | 386,644 | 193,385 | 36,023 | 94,213 | 21,068 | 41,150 | 805 |
| Guanajuato | 1,116,019 | 503,703 | 210,704 | 310,598 | 13,267 | 77,718 | 29 |
| Guerrero | 809,178 | 239,430 | 105,343 | 192,600 | 96,609 | 175,196 | 0 |
| Hidalgo | 587,626 | 237,593 | 97,867 | 155,703 | 27,283 | 69,180 | 0 |
| Jalisco | 1,747,785 | 863,328 | 241,326 | 466,725 | 28,518 | 147,688 | 200 |
| México | 3,305,277 | 1,926,114 | 422,912 | 691,856 | 73,921 | 187,704 | 2,770 |
| Michoacan | 978,958 | 334,610 | 157,288 | 250,173 | 82,075 | 153,090 | 1,722 |
| Morelos | 419,912 | 221,594 | 58,988 | 110,457 | 192 | 28,519 | 162 |
| Nayarit | 267,286 | 99,221 | 33,997 | 100,602 | 1,604 | 31,692 | 170 |
| Nuevo León | 1,076,466 | 697,593 | 58,993 | 265,986 | 5,910 | 47,461 | 523 |
| Oaxaca | 894,618 | 312,463 | 65,206 | 263,274 | 109,141 | 144,534 | 0 |
| Puebla | 1,225,073 | 595,131 | 132,965 | 306,814 | 59,713 | 130,205 | 245 |
| Queretaro | 324,828 | 184,097 | 32,225 | 64,281 | 20,899 | 23,217 | 109 |
| Quintana Roo | 231,761 | 130,194 | 34,725 | 37,348 | 15,362 | 13,747 | 385 |
| San Luis Potosi | 544,738 | 283,247 | 36,347 | 148,949 | 38832 | 36,759 | 604 |
| Sinaloa | 687,235 | 349,438 | 116,276 | 177,944 | 986 | 41,304 | 1,287 |
| Sonora | 606,885 | 399,827 | 80,619 | 107,733 | 388 | 5,554 | 12,764 |
| Tabasco | 526,035 | 271,847 | 45,361 | 133,198 | 15,606 | 60,023 | 0 |
| Tamaulipas | 707,478 | 372,705 | 104,750 | 174,196 | 11,603 | 43,838 | 386 |
| Tlaxcala | 268,223 | 127,518 | 32,818 | 69,631 | 12,111 | 26,145 | 0 |
| Veracruz | 1,940,366 | 718,949 | 314,633 | 497,842 | 120,278 | 288,531 | 133 |
| Yucatan | 449,519 | 213,328 | 61,973 | 104,417 | 36,449 | 33,352 | 0 |
| Zacatecas | 340,496 | 156,466 | 36,923 | 90,722 | 14,176 | 40,029 | 2,180 |

| | SEGUN NIVEL DE INGRESOS | | | | | | | más de 10 NRI | No Esp. |
|---------------------|-------------------------|------------|------------|------------|-------------|--------------|--------------|---------------|---------|
| | Menos de 1 | 1 a 2 S.M. | 2 a 3 S.M. | 3 a 5 S.M. | 5 a 10 S.M. | 10 a 15 S.M. | 15 a 20 S.M. | | |
| Agascalientes | 20,546 | 80,158 | 40,632 | 28,808 | 16,552 | 7,103 | 18,553 | 16,196 | |
| Baja California | 16,644 | 136,361 | 181,260 | 143,268 | 95,058 | 46,162 | 8,644 | 831 | |
| Baja California Sur | 9,235 | 34,234 | 24,807 | 21,641 | 12,316 | 4,044 | 4,851 | 7,337 | |
| Campeche | 40,942 | 48,012 | 20,403 | 20,368 | 14,498 | 6,763 | 29,429 | 3,187 | |
| Coahuila | 51,400 | 193,028 | 134,646 | 90,322 | 61,726 | 28,655 | 33,916 | 4,991 | |
| Colima | 16,715 | 45,348 | 26,389 | 24,314 | 11,115 | 4,045 | 5,683 | 3,458 | |
| Chiapas | 324,235 | 232,209 | 82,099 | 66,753 | 45,659 | 11,406 | 298,547 | 19,153 | |
| Chihuahua | 37,575 | 266,567 | 172,245 | 136,607 | 82,619 | 29,739 | 51,145 | 13,869 | |
| Distrito Federal | 210,328 | 820,805 | 393,048 | 317,929 | 211,343 | 130,386 | 62,663 | 152,398 | |
| Durango | 39,024 | 131,378 | 67,554 | 43,834 | 26,042 | 8,743 | 60,978 | 9,091 | |
| Guanajuato | 142,981 | 398,556 | 225,223 | 131,073 | 68,631 | 25,776 | 90,224 | 33,555 | |
| Guerrero | 169,671 | 192,678 | 88,297 | 54,895 | 21,325 | 4,358 | 264,154 | 13,800 | |
| Hidalgo | 168,224 | 171,827 | 61,013 | 47,101 | 29,094 | 12,445 | 93,829 | 4,093 | |
| Jalisco | 139,211 | 590,478 | 315,790 | 172,793 | 87,869 | 32,895 | 172,453 | 236,296 | |
| México | 489,693 | 1,201,035 | 637,331 | 326,552 | 182,488 | 67,032 | 238,735 | 162,311 | |
| Michoacan | 147,260 | 309,716 | 130,386 | 89,010 | 42,420 | 14,844 | 215,193 | 30,129 | |
| Morelos | 46,313 | 154,731 | 71,435 | 51,900 | 25,894 | 9,762 | 28,857 | 31,020 | |
| Nayarit | 36,883 | 84,609 | 40,207 | 40,923 | 19,059 | 5,749 | 33,044 | 6,812 | |
| Nuevo León | 74,023 | 339,789 | 230,779 | 163,682 | 92,051 | 53,663 | 51,791 | 70,688 | |
| Oaxaca | 255,188 | 195,789 | 86,978 | 58,886 | 34,192 | 8,699 | 242,486 | 12,400 | |
| Puebla | 249,724 | 363,854 | 186,804 | 127,572 | 47,188 | 10,412 | 176,778 | 62,741 | |
| Queretaro | 34,526 | 105,451 | 59,802 | 38,062 | 27,611 | 13,826 | 42,817 | 2,733 | |
| Quintana Roo | 14,902 | 55,432 | 49,145 | 45,443 | 25,231 | 8,402 | 28,727 | 4,479 | |
| San Luis Potosi | 130,228 | 181,175 | 74,605 | 48,956 | 26,813 | 12,046 | 65,162 | 5,753 | |
| Sinaloa | 70,893 | 248,120 | 127,207 | 103,119 | 65,494 | 26,560 | 42,202 | 3,640 | |
| Sonora | 54,000 | 208,560 | 137,342 | 94,498 | 59,464 | 29,959 | 18,433 | 4,629 | |
| Tabasco | 151,810 | 135,211 | 56,708 | 57,175 | 34,678 | 10,440 | 74,967 | 5,046 | |
| Tamaulipas | 97,881 | 239,599 | 126,608 | 90,162 | 59,090 | 23,154 | 54,692 | 16,292 | |
| Tlaxcala | 60,673 | 94,532 | 36,092 | 22,889 | 11,932 | 2,696 | 33,065 | 6,344 | |
| Veracruz | 582,973 | 543,858 | 164,119 | 144,817 | 79,316 | 24,762 | 390,523 | 9,998 | |
| Yucatan | 99,318 | 150,354 | 48,346 | 37,440 | 30,993 | 12,133 | 67,288 | 3,647 | |
| Zacatecas | 60,031 | 111,053 | 47,111 | 28,479 | 10,877 | 7,319 | 51,489 | 24,137 | |

ANEXO IV

Nivel de instrucción o de escolaridad. Registra el último curso educativo formal aprobado por las personas de seis años y más.

Sin instrucción. Comprende a las personas de 6 años y más que nunca asistieron a algún centro de enseñanza ni aprobaron ningún grado escolar.

Con primaria. Comprende a las personas que aprobaron algún o algunos grados de instrucción primaria, entre primero y sexto año.

Con secundaria. Incluye a las personas que aprobaron algún grado de enseñanza secundario o bien el tercer grado.

Con educación media superior y superior. Incluye a las personas que aprobaron 1, 2 o 3 años de preparatoria, vocacional normal o equivalente y a las personas que aprobaron algún o algunos grados o concluyeron su instrucción profesional, independientemente si obtuvieron o no el título profesional.

Tipo de empleo buscado. Clasificación de población no ocupada que busca trabajo de acuerdo a si desea un trabajo asalariado, o bien por cuenta propia.

Experiencia laboral. Hace referencia a la experiencia laboral que las personas tienen en la semana de referencia.

Trabajadores con experiencia. Se refiere a las personas desocupadas que cuentan con experiencia laboral.

Trabajadores sin experiencia. Son personas que nunca han trabajado o que nunca han sido ayudantes sin pago.

Motivos de desempleo. Las personas que en situación de desocupación abierta declararon haber trabajado anteriormente se clasifican conforme a los motivos que tuvieron para dejar el empleo y que pueden ser voluntarios e involuntarios.

Desempleado por cese. Hace referencia a las personas que por cierre o quiebra de su fuente de trabajo, recorte de personal, suspensión por tiempo indefinido o despido sin causa justificada se encuentran sin empleo.

Desempleado por trabajo temporal terminado. Son las personas que son contratadas temporalmente para desempeñar un trabajo claramente determinado y que una vez que lo efectúan, quedan sin empleo.

Desempleado por insatisfacción con el trabajo. Son las personas que renuncian a su empleo por no estar satisfechas con el, ya sea por insatisfacción con sus ingresos, horarios, jefes, funciones, compañeros, etcétera.

| | POBLACION DESOCUPADA ABIERTA | | | | | | | PDAMSyS | No Esp. |
|---------------------|------------------------------|----------------------------|--------|--------|---------|---------|--|---------|---------|
| | TOTAL | SEGUN NIVEL DE INSTRUCCION | | | | | | | |
| | PDA | PDASI | PDAP | PDAS | PDAMSyS | No Esp. | | | |
| Agascalientes | 6,369 | 426 | 2,277 | 2,136 | 1,530 | 0 | | | |
| Baja California | 10,048 | 93 | 3,693 | 3,284 | 2,978 | 0 | | | |
| Baja California Sur | 1,438 | 0 | 492 | 249 | 697 | 0 | | | |
| Campeche | 2,502 | 45 | 759 | 729 | 969 | 0 | | | |
| Coahuila | 16,263 | 200 | 4,504 | 5,807 | 5,752 | 0 | | | |
| Colima | 1,668 | 122 | 535 | 437 | 574 | 0 | | | |
| Chiapas | 9,613 | 547 | 2,758 | 1,822 | 4,486 | 0 | | | |
| Chihuahua | 12,621 | 66 | 3,729 | 5,516 | 3,310 | 0 | | | |
| Distrito Federal | 88,397 | 669 | 20,550 | 29,213 | 37,965 | 0 | | | |
| Durango | 6,155 | 186 | 2,564 | 1,489 | 1,916 | 0 | | | |
| Guanajuato | 21,673 | 684 | 10,070 | 7,927 | 2,992 | 0 | | | |
| Guerrero | 8,543 | 0 | 1,694 | 4,572 | 2,277 | 0 | | | |
| Hidalgo | 14,964 | 2,580 | 4,691 | 4,777 | 2,916 | 0 | | | |
| Jalisco | 32,853 | 2,001 | 10,342 | 11,192 | 9,318 | 0 | | | |
| México | 111,130 | 2,599 | 30,467 | 45,215 | 32,849 | 0 | | | |
| Michoacan | 7,600 | 0 | 2,952 | 376 | 4,272 | 0 | | | |
| Morelos | 5,470 | 0 | 1,340 | 1,485 | 2,645 | 0 | | | |
| Nayarit | 1,740 | 61 | 477 | 761 | 441 | 0 | | | |
| Nuevo León | 20,454 | 111 | 4,838 | 10,086 | 5,419 | 0 | | | |
| Oaxaca | 6,414 | 418 | 3,273 | 1,296 | 1,427 | 0 | | | |
| Puebla | 14,677 | 0 | 3,935 | 6,180 | 4,562 | 0 | | | |
| Queretaro | 7,269 | 419 | 1,107 | 3,556 | 2,187 | 0 | | | |
| Quintana Roo | 2,533 | 0 | 285 | 1,451 | 797 | 0 | | | |
| San Luis Potosí | 6,773 | 53 | 1,893 | 2,321 | 2,506 | 0 | | | |
| Sinaloa | 11,207 | 497 | 2,984 | 3,717 | 3,775 | 234 | | | |
| Sonora | 16,008 | 0 | 5,367 | 4,510 | 6,131 | 0 | | | |
| Tabasco | 10,099 | 162 | 3,844 | 3,372 | 2,721 | 0 | | | |
| Tamaulipas | 16,012 | 121 | 5,469 | 6,806 | 3,616 | 0 | | | |
| Tlaxcala | 4,878 | 69 | 1,291 | 2,066 | 1,452 | 0 | | | |
| Veracruz | 31,005 | 3,804 | 10,052 | 11,256 | 5,893 | 0 | | | |
| Yucatan | 4,568 | 164 | 1,215 | 1,673 | 1,516 | 0 | | | |
| Zacatecas | 2,009 | 565 | 394 | 691 | 359 | 0 | | | |

| | SEGUN TIPO DE TRABAJO BUSCADO | | | SEGUN EXPERIENCIA | No Esp. | |
|---------------------|-------------------------------|-------|-------|-------------------|---------|----|
| | BTA | BTCP | Otros | | | |
| | SEGUN EXPERIENCIA | | | CE | SE | |
| Aguascalientes | 6,229 | 107 | 33 | 5,280 | 1,089 | 0 |
| Baja California | 9,106 | 825 | 117 | 8,906 | 1,142 | 0 |
| Baja California Sur | 1,349 | 89 | 0 | 1,246 | 192 | 0 |
| Campeche | 2,490 | 12 | 0 | 2,188 | 314 | 0 |
| Coahuila | 16,094 | 169 | 0 | 14,335 | 1,928 | 0 |
| Colima | 1,581 | 87 | 0 | 1,446 | 222 | 0 |
| Chiapas | 9,613 | 0 | 0 | 8,613 | 1,000 | 0 |
| Chihuahua | 12,078 | 214 | 329 | 12,147 | 474 | 0 |
| Distrito Federal | 86,717 | 1,680 | 0 | 80,348 | 8,049 | 0 |
| Durango | 6,098 | 57 | 0 | 5,757 | 398 | 0 |
| Guanajuato | 20,933 | 59 | 681 | 20,519 | 1,154 | 0 |
| Guerrero | 8,485 | 58 | 0 | 8,345 | 198 | 0 |
| Hidalgo | 14,259 | 675 | 30 | 13,393 | 1,571 | 0 |
| Jalisco | 32,660 | 193 | 0 | 30,170 | 2,683 | 0 |
| México | 107,962 | 3,168 | 0 | 100,573 | 10,557 | 0 |
| Michoacan | 7,600 | 0 | 0 | 5,764 | 1,836 | 0 |
| Morelos | 5,470 | 0 | 0 | 5,150 | 320 | 0 |
| Nayarit | 1,740 | 0 | 0 | 1,565 | 175 | 0 |
| Nuevo León | 20,198 | 256 | 0 | 17,917 | 2,537 | 0 |
| Oaxaca | 6,336 | 78 | 0 | 5,521 | 893 | 0 |
| Puebla | 14,677 | 0 | 0 | 12,240 | 2,437 | 0 |
| Queretaro | 7,002 | 110 | 157 | 6,351 | 918 | 0 |
| Quintana Roo | 2,404 | 64 | 65 | 2,397 | 78 | 58 |
| San Luis Potosi | 6,211 | 562 | 0 | 6,479 | 294 | 0 |
| Sinaloa | 11,094 | 113 | 0 | 11,143 | 64 | 0 |
| Sonora | 15,946 | 0 | 62 | 14,711 | 1,297 | 0 |
| Tabasco | 10,099 | 0 | 0 | 9,620 | 479 | 0 |
| Tamaulipas | 15,876 | 136 | 0 | 14,357 | 1,655 | 0 |
| Tlaxcala | 4,878 | 0 | 0 | 4,096 | 782 | 0 |
| Veracruz | 30,887 | 118 | 0 | 29,020 | 1,985 | 0 |
| Yucatan | 4,568 | 0 | 0 | 4,323 | 245 | 0 |
| Zacatecas | 1,988 | 21 | 0 | 1,900 | 109 | 0 |

| | SEGUN MOTIVOS DEL DESEMPLEO | | | | | No Esp. |
|---------------------|-----------------------------|--------|--------|--------|--------|---------|
| | Cese | TTT | IICT | Otros | | |
| Aguascalientes | 1,605 | 1,701 | 1,180 | 794 | 1,089 | |
| Baja California | 1,915 | 2,506 | 2,783 | 1,702 | 1,142 | |
| Baja California Sur | 228 | 574 | 178 | 266 | 192 | |
| Campeche | 1,089 | 453 | 530 | 116 | 314 | |
| Coahuila | 3,239 | 2,304 | 5,785 | 3,007 | 1,928 | |
| Colima | 367 | 466 | 539 | 74 | 222 | |
| Chiapas | 3,224 | 2,138 | 2,087 | 1,164 | 1,000 | |
| Chihuahua | 5,774 | 3,136 | 1,132 | 2,105 | 474 | |
| Distrito Federal | 33,765 | 17,044 | 19,570 | 9,969 | 8,049 | |
| Durango | 607 | 2,388 | 1,430 | 1,332 | 398 | |
| Guanajuato | 9,876 | 5,538 | 4,525 | 580 | 1,154 | |
| Guerrero | 6,557 | 119 | 854 | 815 | 198 | |
| Hidalgo | 2,080 | 4,184 | 4,187 | 2,942 | 1,571 | |
| Jalisco | 8,884 | 6,977 | 9,750 | 4,559 | 2,683 | |
| México | 33,362 | 19,874 | 34,774 | 12,563 | 10,557 | |
| Michoacan | 917 | 2,322 | 1,919 | 606 | 1,836 | |
| Morelos | 2,293 | 168 | 1,487 | 1,202 | 320 | |
| Nayarit | 560 | 479 | 356 | 170 | 175 | |
| Nuevo León | 4,646 | 4,195 | 7,564 | 1,512 | 2,537 | |
| Oaxaca | 243 | 4,276 | 653 | 349 | 893 | |
| Puebla | 6,480 | 818 | 3,928 | 1,014 | 2,437 | |
| Queretaro | 2,043 | 1,651 | 2,263 | 394 | 918 | |
| Quintana Roo | 428 | 570 | 633 | 766 | 136 | |
| San Luis Potosí | 2,032 | 1,153 | 2,002 | 1,292 | 294 | |
| Sinaloa | 2,786 | 2,269 | 4,427 | 1,661 | 64 | |
| Sonora | 3,662 | 5,537 | 3,800 | 1,712 | 1,297 | |
| Tabasco | 3,777 | 3,291 | 1,624 | 928 | 479 | |
| Tamaulipas | 3,748 | 4,506 | 5,198 | 905 | 1,655 | |
| Tlaxcala | 1,271 | 1,192 | 1,162 | 471 | 782 | |
| Veracruz | 5,396 | 16,568 | 3,996 | 3,060 | 1,985 | |
| Yucatan | 1,055 | 886 | 1,262 | 786 | 579 | |
| Zacatecas | 195 | 1,074 | 405 | 226 | 109 | |

BIBLIOGRAFÍA

- ANDERSON, T.W. (1984). *An introduction to multivariate statistical analysis*. New York: John Wiley and sons.
- ANKER, Richard, Catherine Hein (1987). *Empleo de la mujer fuera de la agricultura en países del Tercer Mundo: panorama general de las estadísticas ocupacionales*, en Desigualdades entre hombres y mujeres en los mercados de trabajo urbano del Tercer Mundo. Ginebra, Oficina Internacional de Trabajo.
- BENERIA, Lourdes, Marta Roldán (1987). *The Crossroads of Class and Gender. Industrial Homework, Subcontracting and Household Dynamics in México City*. Chicago: The University of Chicago Press. Traducción al español: Lourdes Benería y Marta Roldán, Las encrucijadas de clase y género Trabajo a domicilio, subcontratación y dinámica de la unidad doméstica en la ciudad de México. México: El colegio de México y Fondo de Cultura Económica (Economía Latinoamericana).
- BLAU, Francine D., Marianne A. Ferber (1986). *The economics of women, men and work*. Englewood cliffs, New Jersey: prentice Hall.
- ESCOFIER, Brigitte, Jérôme Pagès (1992). *Analices factorielles simples et multiples*. Traducción de Elena Abascal Fernández y Karnele Fernández Aguirre. Bilbao: Servicio editorial de la Universidad del País Vasco.
- BRONFMAN, Mario, Elsa López ,Rodolfo Tuirán (1986). *Practica anticonceptiva y clases sociales en México: la experiencia reciente*, en Estudios Demográficos y Urbanos, Vol. 1, núm. 2, mayo-agosto, pp. 165-203. México: El Colegio de México.
- CERVANTES, Carson Alejandro, Germán Sandoval (1988). "Regulación de la fecundidad", en Encuesta nacional sobre fecundidad y salud: 1987, Memoria de la reunión celebrada el 30 de septiembre de 1988. México: Secretaria de Salud.
- CUADRAS, Carlos M. (1996). *Métodos de análisis multivariante*. Barcelona: UUB.
- Dillon, William R., Mattew Goldstein (1984). *Multivariate Analysis: Methods and applications*. New York: John Wiley and sons.
- GARCÍA, Brígida (1988). *Desarrollo económico y absorción de la fuerza de trabajo en México: 1950-1980*. México: El Colegio de México.
- GARCÍA, Brígida (1997). *Economic restructuring ,women survival and transformation in México*. Ponencia presentada en el Seminario Female Empowerment and Demographic Proceses. Suecia 21-24 de abril, (mimeo).
- GARCÍA, Brígida (1996). *Fuerza de trabajo en 1995. Las implicaciones del nuevo modelo de desarrollo en Demos: carta demográfica sobre México*. México: IIS-UNAM.

GARCÍA, Brígida y Orlandina de Oliveira (1994). *La medición de la Población económicamente activa en México al inicio de los años noventa*. En Estudios Demográficos y Urbanos, Vol. 9, núm. 3 septiembre-diciembre. México: El Colegio de México.

GARCIA, Brígida ,Orlandina de Oliveira (1994). *Trabajo femenino y vida familiar en México*. México: El colegio de México.

JOEKES, Susan P. (1987). *La mujer y la economía mundial*. México: ED. Siglo XXI.

JUSIDMAN, Clara (1971). *Conceptos y definiciones en relación con el empleo, el desempleo y el subempleo*. En: Demografía y Economía Vol. 5(3) No. 15. México: El Colegio de México.

JUSIDMAN, Clara ,Marcela Eternod (1992). *La participación de la población en la actividad económica en México*. México: INEGI, IIS-UNAM.

MANLY, Bryan F. J. (1986). *Multivariate statistical methods*. London: Chapman and hall.

MANZANAREZ, Martínez Domingo Antonio (1992). *Temas de historia social del trabajo*. Barcelona: Promociones y Publicaciones Universitarias.

MARDIA, Kantilal Varichand (1979). *Multivariate Analysis*. London; New York: Academic.

MORRISON, Donald F. (1990). *Multivariate statistical methods*. New York: McGraw-Hill.

NACIONES UNIDAS (1987). *Women's Employment and Fertility*, en Fertility Behavior in the Context of Development. Evidence from the World. Fertility Survey, Nueva York: ONU.

NEGRETE Salas, María Eugenia (1988). *Cambios en la estructura y distribución de la fuerza de trabajo en México: la dimensión regional, 1950-1980*, en Memorias de la III Reunión Nacional sobre Investigación Demográfica, tomo I. México: Sociedad Mexicana de Demografía.

OLIVEIRA, Orlandina de (1989). *Empleo femenino en México en tiempos de recesión económica: tendencias recientes en Jennifer Cooper, Teresita de Barbieri, et al (comp.)*. *Fuerza de trabajo femenina urbana en México. Características y tendencias*. México, UNAM-Porrúa.

OLIVEIRA, Orlandina de, Marina Ariza ,Marcela Eternod (1996). *Trabajo e inequidad de género*, en Orlandina de Oliveira (coord.). *La condición femenina: una propuesta de indicadores*. Informe final. México: SOMEDE-CONAPO.

OLIVEIRA, Orlandina de, Marina Ariza ,Marcela Eternod (1998). *La fuerza de trabajo en México: Un siglo de Cambios* . México: mimeo.

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (1994). *Women and Structural change. New Perspective*. Paris: OECD.

- PALMA, Yolanda (1988). "Niveles de fecundidad", en Encuesta nacional de fecundidad y salud: 1987, Memoria de la reunión celebrada el 30 de septiembre de 1988, Dirección General de Planificación Familiar de la Secretaría de Salud.
- PEDRERO, Mercedes (1976). *La participación femenina en la actividad económica y su presupuesto de tiempo (Notas sobre problemas relativos a conceptos y captación)*. México: STPS, Centro Nacional de Información y Estadística del Trabajo.
- PEDUENO, M., Teresa Rendón, Antonieta Barrón (1997). *Segregación Ocupacional por genero*. México: UNAM-CRIM.
- RENDON, Teresa, Carlos Salas (1993). *El Empleo en México en los ochentas. Tendencias y cambios*, en Comercio Exterior, Vol. 43, núm. 8, agosto. México.
- RENDON, Teresa, Carlos Salas (1987). *Evolución del empleo en México: 1890-1980*, en Estudios Demográficos y Urbanos, Vol. 2, núm. 2, mayo-agosto. México.
- RENDON, Teresa, Carlos Salas (1992). *El mercado de trabajo no agrícola en México. Tendencias y cambios recientes en Ajuste estructural, mercados laborales y Tratado de Libre Comercio*. México: Centro de Estudios Sociológicos de El Colegio de México, Fundación Friedrich Ebert y El Colegio de la Frontera Norte.
- RESKIN, Barbara F. (1984). *Sex segregation in the workplace. Trends, explanations, remedies*. Washington: National Academy Press.
- RESKIN, Barbara F., Heidi I. Hartmann (1986). *Women's work, men's work. Sex segregation on the job*. Washington D.C.: National Academy Press.
- ROBINSON, Joan (1963). *Introduction to the theory of employment*. Traducción a español por Rafael Muñoz Rojas de Alarcón, Introducción a la teoría del empleo. Madrid: Sociedad de estudios y publicaciones.
- SECRETARIA DE TRABAJO Y PREVISIÓN SOCIAL (1986). *La mujer y el trabajo en México (Antología)*. México: STPS.
- STANDING, Guy (1978). *Labour Force participation and development*. Ginebra: OIT.
- URIEL, Jimenez Ezequiel (1995). *Análisis de datos: series de tiempo y análisis multivariante*. Madrid, España: Ed. AC.