

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

U.A.C.P. y P. del C.C.H.

I.I.M.A.S.

Análisis Estadístico de un Estudio Educativo

en la E.N.E.P. "Zaragoza"

TESINA

Que Para Obtener el Diploma de

ESPECIALIZACION EN ESTADISTICA APLICADA

PRESENTA

LUZ MARGARITA GUZMAN ARELLANO

BIBLIOTECA  
JUAN A. ESCALANTE H.  
UNIDAD ACADÉMICA DE  
LOS CICLOS PROFESIONAL  
Y DE POSGRADO / CCH  
UNAM

MAYO 1987



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.



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.

Mi agradecimiento a la Maestra  
en Ciencias Belem Trejo Valdivia  
por su apoyo y atinada supervi-  
sión en la realización de este  
trabajo

## INDICE

	Página
Introducción.....	4
Antecedentes.....	5
Metodología.....	11
Análisis Estadístico..	24
Conclusiones.....	49
Apéndice.....	52
Bibliografía.....	58

## I N T R O D U C C I O N

La Escuela Nacional de Estudios Profesionales "Zaragoza", dependiente de la UNAM, inició sus labores en enero de 1976. En el área químico-biológica de dicha Escuela se cursan las carreras de Biólogo, Ingeniero Químico y Químico Farmacéutico Biólogo. En esta área se observó desde sus inicios que una cantidad grande de alumnos, más del 60 %, quedaban en calidad de "irregulares" debido al alto índice de reprobación en los tres primeros semestres correspondientes al Tronco Común. Este hecho originó la realización de un estudio, mediante la aplicación de cuestionarios a alumnos de primer ingreso al ciclo profesional y el análisis de sus calificaciones.

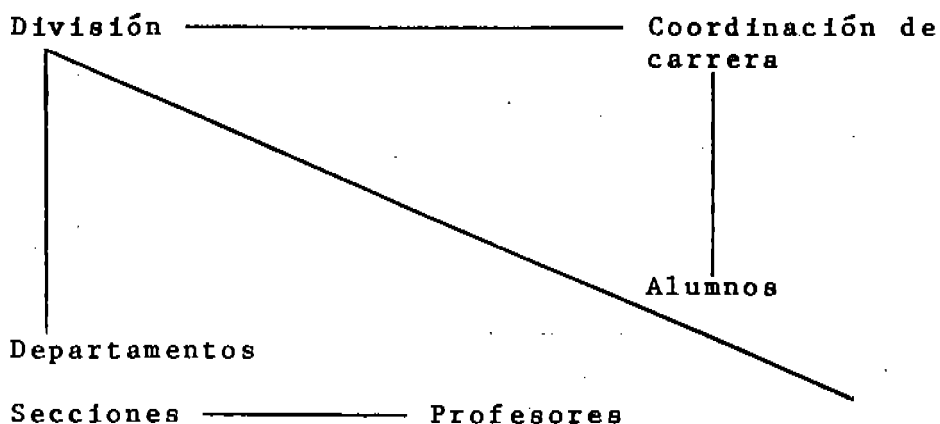
En el presente trabajo se hace un análisis estadístico sobre los datos obtenidos a través de un cuestionario aplicado a los alumnos de la generación que ingresó en el semestre 81-1. Se pretende estudiar la interrelación de algunas variables mediante tablas de contingencia y modelos logarítmicos lineales.

### ANTECEDENTES

La Escuela Nacional de Estudios Profesionales Zaragoza se crea en 1976 como parte del proceso de descentralización educativa iniciado por la UNAM en 1974. De 1954 a 1972 el alumnado de la UNAM aumentó de 33.600 a 107 900 estudiantes, la saturación de las instalaciones de la Ciudad Universitaria generó el proceso de descentralización y se propuso la creación de un sistema de nuevos centros educativos de estudios profesionales independientes académica y administrativamente de las Facultades y Escuelas de la Ciudad Universitaria.

Se propuso como organización académica de las nuevas escuelas una estructura matricial

Dirección

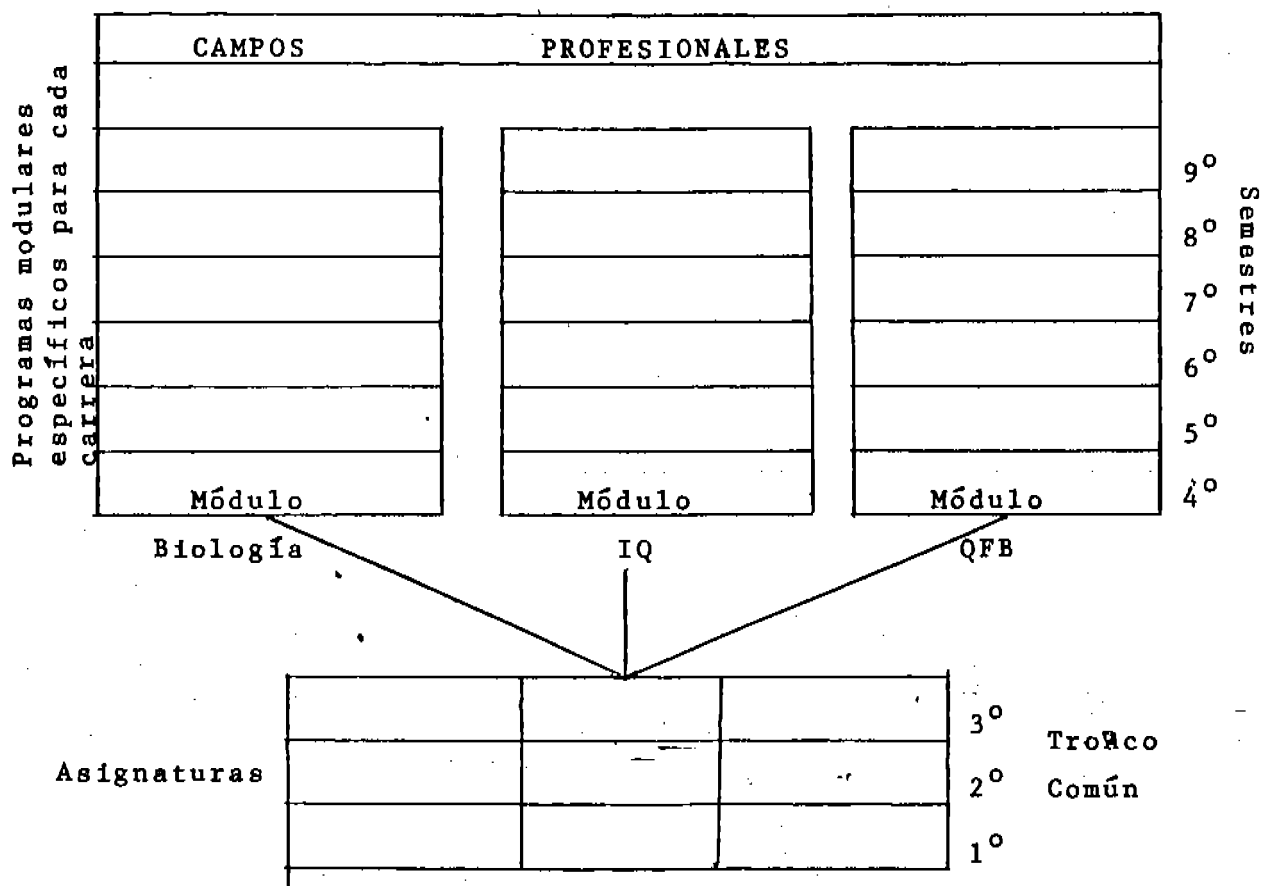


En la ENEP Zaragoza se implantó un sistema de enseñanza modular en el que los módulos se consideran como un "conjunto de actividades multidisciplinares y de aprendizaje que, en un lapso flexible permite alcanzar objetivos educativos de capacidades, destrezas y actitudes que le permitan al alumno desempeñar funciones profesionales..." (1)

(1) Lineamientos generales de la investigación 1983

En esta escuela se pretende realizar un trabajo multidisciplinario e interdisciplinario, ya que se imparten siete carreras: Medicina, Enfermería, Odontología y Psicología en el área de la salud. Y Biología, Ingeniería Química y Química farmacobiológica en el área químico biológica.

En esta última área el sistema modular propiamente dicho se inicia en el cuarto semestre de cada una de las carreras y previo a éste existe un "tronco común" con duración de tres semestres, donde se analizan problemas provenientes de los diferentes campos profesionales, con el objeto de que los alumnos adquieran una formación científica básica y un lenguaje común que facilitará la comunicación entre los diversos profesionistas y la formación de grupos de trabajo interdisciplinarios.



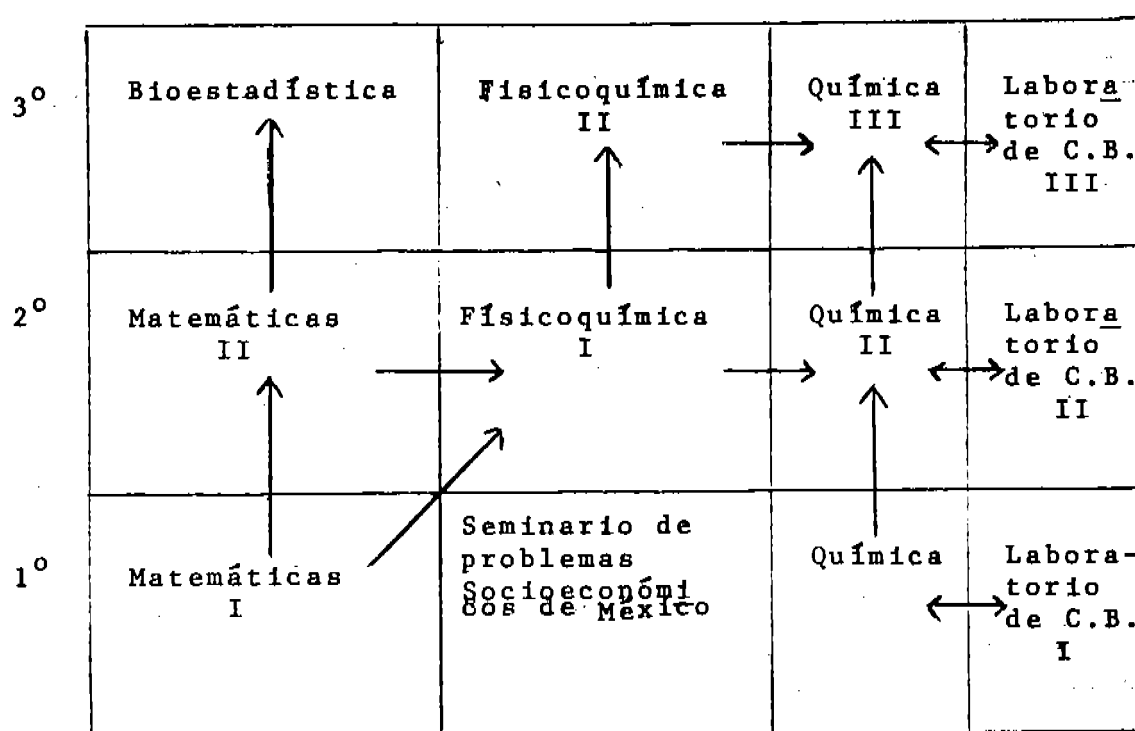
En los últimos semestres de las carreras se pretende una orientación al servicio de la comunidad.



En el Tronco Común se decidió reestructurar la organización tradicional de los conocimientos básicos en tres grandes áreas: Matemáticas, Fisicoquímica y Química General en Unidades didácticas diseñadas de acuerdo a objetivos de aprendizaje, agrupadas - en tres asignaturas teóricas y un laboratorio, con cierto carácter modular, por semestre

TRONCO COMUN  
Area Químico-Biológica

Semestres



El sistema de evaluación se basa en exámenes departamentales comunes y simultáneos a todos los alumnos, que establecen criterios mínimos de conocimientos para toda la población escolar.

El presente trabajo se llevó a cabo con alumnos del 1er. semestre del Tronco Común del Area Químico Biológica.

-8-

Previamente a este estudio se llevaron a cabo los siguientes trabajos:

-Investigación sobre la situación Académica de los estudiantes del segundo semestre y las Técnicas Didácticas utilizadas para su preparación. Margarita Eugenia Gutiérrez y Luz Margarita Guzmán Arellano. División de Estudios de posgrado. Especialización en Docencia. Facultad de Química. UNAM. abril 1974.

-Investigación acerca de la situación Académica de los alumnos del área Químico-Biológica de la ENEP "Zaragoza". Margarita Eugenia Gutiérrez, Georgina Rosales y Luz Margarita Guzmán. UNAM, (I), octubre 1979 y (II) marzo 1981.

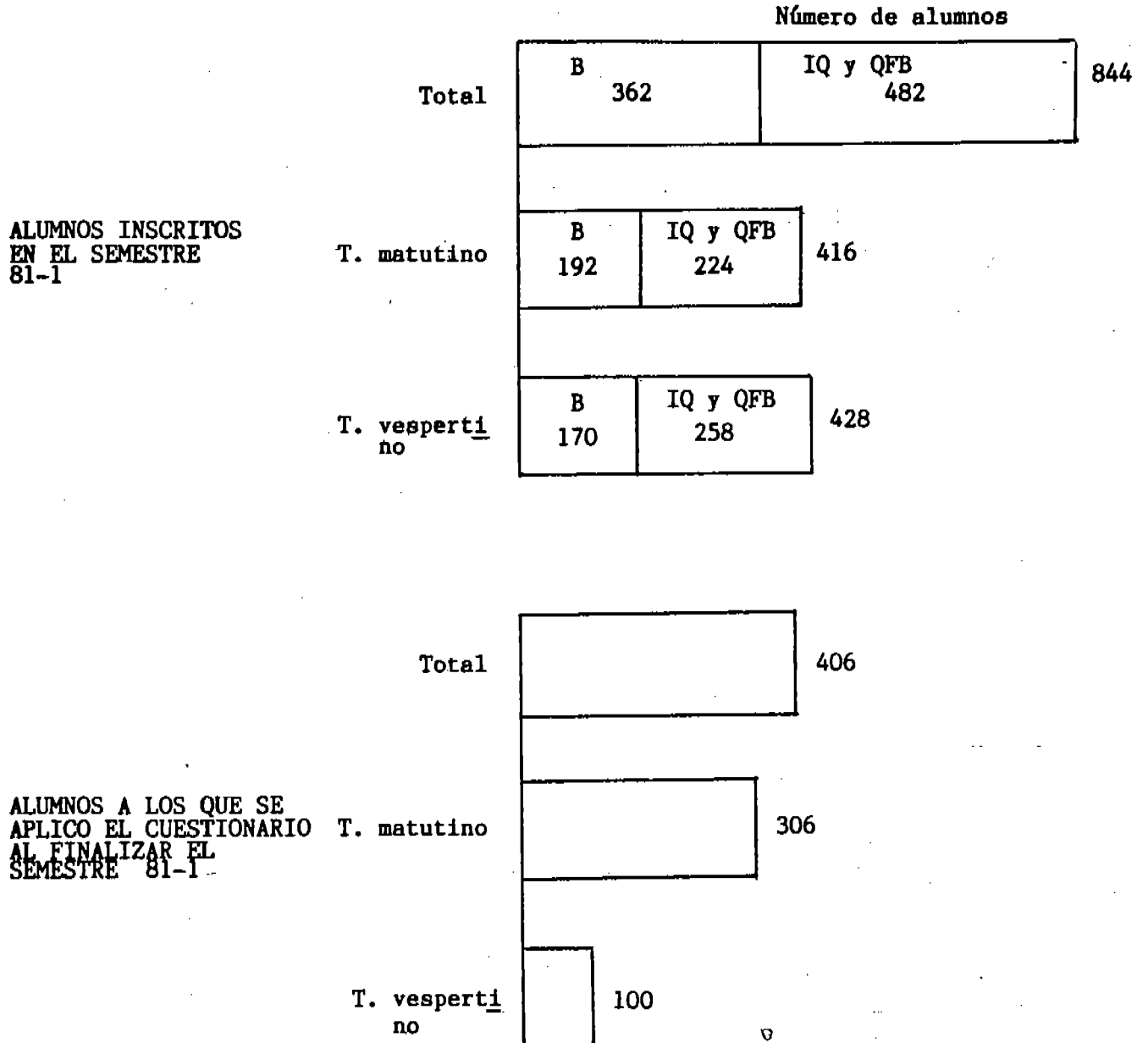
En los trabajos anteriores se llevó a cabo un cómputo simple de los datos obtenidos a través de la aplicación de cuestionarios llegando sólo a expresiones de porcentajes.

En el presente trabajo se aplicó el cuestionario (ver anexo) a los 406 alumnos que asistían regularmente a clases al finalizar el primer semestre de 1981 (gráfica 1). En dicho semestre aparecen en las actas inscritos 844 alumnos, 416 en el turno matutino y 428 en el turno vespertino; suprimiendo los que tenían como calificación NP (no presentado) quedaban 598 (70.8% del total), 309 en la mañana y 289 en la tarde. Al intentar la aplicación del cuestionario por censo, en la tarde sólo se captó al 34.6% del alumnado (100 cuestionarios) mientras en la mañana fue el 99% (306 cuestionarios), esto se explica por el ausentismo tan acentuado en el turno vespertino.

El cuestionario consta de 63 preguntas, de las cuales las siete primeras permiten conocer las características generales de la población en estudio como turno, edad, sexo, estado civil, carrera y escuela de procedencia. Los items fundamentales y las preguntas correspondientes, cuya medición se consideró de importancia para la investigación son:

-9-  
GRAFICA I

ENEP "Zaragoza"      Area Químico-Biológica



1.- Preparación adquirida por los alumnos en sus estudios de educación media superior

1.1 Cumplimiento de objetivos en programas educativos del bachillerato.

1.2 Hábitos de estudio

preguntas: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 28, 29, 40, 41, 42.

2.- Situación socioeconómica

2.1 Trabajo extraescolar remunerado

2.2 Características culturales del núcleo familiar

preguntas: 8, 9, 20, 21, 22, 23, 24, 25, 26, 27, 30, 31, 32, 33, 34, 35, 36.

3.- Vocación

3.1 Selección de la carrera

3.2 Adaptación a la carrera

preguntas: 37, 38, 39, 43, 44, 45, 46, 47, 48, 49, 50.

4.- Conceptos relativos a la ENEP "Zaragoza"

4.1 Técnicas didácticas modernas, su aplicación.

4.2 Concepción modular de las carreras

4.3 Empleo de cartas descriptivas

4.4 Aplicación de exámenes departamentales

preguntas: 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63

Lo observado en esos primeros trabajos sobre la situación académica de los estudiantes de la Facultad de Química y del área Químico-biológica de la ENEP "Zaragoza" fue que ellos en su mayoría:

- Dependen económicamente de su familia, generalmente numerosa
- Los padres tienen baja escolaridad
- Proviene de escuelas oficiales del Distrito Federal
- Consideran deficientes sus estudios preparatorios y eligen la carrera sin ningún fundamento
- No tienen el concepto de módulo
- Consideran favorable el uso de programas en forma de carta descriptiva
- Los exámenes departamentales, aunque marcan el nivel mínimo necesario del curso, por errores de elaboración y aplicación en ocasiones los desfavorecen.

-11-

## METODOLOGIA

Para estudiar la interrelación de algunas variables se emplearon tablas de contingencia y la prueba de  $\chi^2$  de independencia.

### Tablas de Contingencia Bidimensionales

Cuando la muestra de una población es clasificada según dos o más atributos, se construye una tabla de contingencia con la siguiente forma general:

Columna (variable 2 )

	1	2		C	Total
1	$n_{11}$	$n_{12}$		$n_{1c}$	$n_{1.}$
2	$n_{21}$				
...					
r	$n_{r1}$			$n_{rc}$	$n_{r.}$
Total	$n_{.1}$			$n_{.c}$	$n_{..} = N$

Renglones (variable 1)

La muestra con N observaciones está clasificada respecto a dos variables cualitativas (tabla bidimensional), una con r categorías y la otra con c categorías y se conoce como una tabla de contingencia r x c

Las categorías en que se clasifican deben ser:

- exhaustiva (debe haber suficientes categorías para colocar a todos los miembros de la población)
- mutuamente excluyente (se define que cada miembro de la población -- puede ser ubicado en una y sólo una categoría)

Los datos de la tabla están en forma de frecuencias (conteo del número de individuos en cada categoría). La frecuencia observada en la i-ésima categoría de la variable renglón y en la j-ésima de la variable columna es la frecuencia en la ij-ésima celda de la tabla, representada por  $n_{ij}$ .

Los totales marginales son

$n_{i.}$  (número total de observaciones en la  $i$ -ésima categoría de la variable renglón)

$$n_{i.} = n_{i1} + n_{i2} + \dots + n_{ic} = \sum_{j=1}^c n_{ij}$$

$n_{.j}$  (número total de observaciones en la  $j$ -ésima categoría de la variable columna)

$$n_{.j} = n_{1j} + n_{2j} + \dots + n_{rj} = \sum_{i=1}^r n_{ij}$$

El número total de observaciones  $n_{..} = N$  es:

$$n_{..} = \sum_{i=1}^r \sum_{j=1}^c n_{ij} = \sum_{i=1}^r n_{i.} = \sum_{j=1}^c n_{.j}$$

Es importante determinar si las variables de la tabla de contingencia son independientes o no.

Si en la población, la probabilidad de una observación de la  $i$ -ésima categoría de la variable renglón y de la  $j$ -ésima categoría de la variable columna es  $p_{ij}$ , la frecuencia esperada en la  $ij$ -ésima celda de la tabla de  $N$  individuos es

$$F_{ij} = N p_{ij}$$

Si  $p_{i.}$  es la probabilidad en la población de una observación de la  $i$ -ésima categoría de la variable renglón y  $p_{.j}$  es la probabilidad de la  $j$ -ésima categoría de la variable columna, se tendrá -- que, bajo la hipótesis de independencia entre dos variables:

$$F_{ij} = N p_{i.} p_{.j}$$

Los valores de probabilidad son desconocidos pero pueden estimarse mediante estimadores de máxima verosimilitud:

$$\hat{p}_{i.} = \frac{n_{i.}}{N} \qquad \hat{p}_{.j} = \frac{n_{.j}}{N}$$

Mediante ellos, se puede estimar la frecuencia esperada en la celda  $ij$  de la tabla, si las dos variables son independientes.

El estimador  $E_{ij}$  está dado por:

$$E_{ij} = N \hat{p}_{i.} \hat{p}_{.j} = \frac{n_{i.} n_{.j}}{N}$$

Quando las dos variables son independientes las frecuencias estimadas mediante esta fórmula y las frecuencias observadas difieren en cantidades atribuibles a factores aleatorios. Si las variables no son independientes habrá grandes diferencias entre lo observado y lo esperado.

### PRUEBA $\chi^2$ DE INDEPENDENCIA

La hipótesis nula  $H_0$  es  $p_{ij} = p_i \cdot p_j$ ,  $i=1, \dots, r$  y  $j=1, \dots, c$

La prueba que investiga si esta hipótesis se rechaza o no, es la de  $\chi^2$  y se basa en las diferencias entre las frecuencias estimadas y las frecuencias observadas:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

La magnitud de esta estadística depende de las diferencias de los valores  $(n_{ij} - E_{ij})$ . Si las dos variables son independientes, esas diferencias son pequeñas y por lo tanto  $H_0$  no se rechaza cuando  $\chi^2$  es pequeña.

Si se supone que las frecuencias observadas tienen una cierta distribución (multinomial) y que las frecuencias esperadas no son pequeñas, la estadística puede mostrar una distribución aproximadamente Ji-cuadrada, con  $(r-1)(c-1)$  grados de libertad.

### Tablas de Contingencia $r \times c$

El análisis de tablas de contingencia  $r \times c$ , cuando  $r$  o  $c$  o ambos son mayores de dos presenta problemas especiales, la prueba de  $\chi^2$  de independencia, en este caso, no tiene una interpretación clara y es necesario un análisis más detallado para decidir en que parte de la tabla hay independencia; además, la construcción de la  $\chi^2$  no considera ningún orden en las categorías por lo que, en caso de tener variables ordinales no sería posible, mediante esta prueba, detectar tendencias asociadas a dicho orden.

Se han sugerido varias técnicas que dan más información respecto a la no independencia y si ésta se presenta en toda la tabla o en una parte específica de ella.

### TABLAS MULTIDIMENSIONALES

Las técnicas para el análisis de tablas de contingencia de dos variables categóricas son bien conocidas pero para tres o más variables no lo son

### TABLAS TRIDIMENSIONALES

Nomenclatura.- Una tabla  $r \times c \times l$  tiene renglón, columna y estrato. La frecuencia observada en la  $ijk$ -ésima celda de la tabla - está representada por  $A_{ijk}$  para :

$$i = 1, 2, \dots, r, \quad j = 1, 2, \dots, c \quad \text{y} \quad k = 1, 2, \dots, l$$

Las frecuencias se determinan a través de un muestreo del siguiente tipo:

Se selecciona una muestra aleatoria de la población y a cada elemento se le clasifica de acuerdo a tres criterios los cuales determinan el renglón, la columna y el estrato, siendo todo variable; este tipo de selección produce el modelo multinomial. También se puede tener un modelo Poisson o producto multinomial, pero no se presentan en este trabajo, pues el estudio por analizar corresponde al modelo multinomial.

Se obtienen diversos totales marginales por las sumatorias de  $n_{ijk}$ . Por ejemplo, sumando sobre todos los valores, tanto de  $i$  como de  $j$ , se tiene el total para la  $k$ -ésima categoría del estrato. Similarmen- te pueden obtenerse los totales para la  $i$ -ésima categoría del renglón y la  $j$ -ésima categoría de la columna sumando  $n_{ijk}$  sobre  $j$  y  $k$  y sobre  $i$  y  $k$  respectivamente.

Esos totales se conocen como totales marginales simples:

$$\begin{array}{l}
n_{i..} = \sum_{j=1}^c \sum_{k=1}^l n_{ijk} \qquad n_{.j.} = \sum_{i=1}^r \sum_{k=1}^l n_{ijk} \qquad n_{...k} = \sum_{i=1}^r \sum_{j=1}^c n_{ijk}
\end{array}$$



Sumando  $n_{ijk}$  sobre un solo subíndice queda:

$$n_{ij.} = \sum_{k=1}^l n_{ijk} \quad n_{i.j} = \sum_{j=1}^c n_{ijk} \quad n_{.jk} = \sum_{i=1}^r n_{ijk}$$

El gran total de frecuencias llamado  $n_{...}$  está dado por:

$$n_{...} = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^l n_{ijk} \text{ y se denota como } N, \text{ que es el único total fijo.}$$

Prueba de independencia global de las variables en las tablas de tres entradas.- La hipótesis es la siguiente:

$$H_0: P_{ijk} = P_{i..} P_{.j.} P_{..k}$$

Donde  $p_{ijk}$  es la probabilidad de que al elegir aleatoriamente a un individuo de la población en estudio, éste sea clasificado en la  $i$ -ésima categoría de la variable 1, en la  $j$ -ésima categoría de la variable 2 y en la  $k$ -ésima categoría de la variable 3, - es decir se clasifica en la  $ijk$ -ésima celda de la tabla de contingencia y  $p_{i..}$ ,  $p_{.j.}$  y  $p_{..k}$  son las probabilidades marginales de las variables renglón, columna y estrato respectivamente; este es el equivalente tridimensional de la hipótesis de independencia en las tablas de dos entradas. Para probar esta hipótesis es necesario estimar las frecuencias esperadas cuando  $H_0$  es cierta, después comparar estos valores con las frecuencias observadas por medio de la estadística  $\chi^2$ , por último se compara la  $\chi^2$  calculada con la de tablas.

Los valores esperados se obtienen, de manera similar a las tablas de doble entrada:

$$E_{ijk} = N \hat{P}_{i..} \hat{P}_{.j.} \hat{P}_{..k}$$

donde  $\hat{P}_{i..}$ ,  $\hat{P}_{.j.}$  y  $\hat{P}_{..k}$  son estimaciones de las probabilidades  $P_{i..}$ ,  $P_{.j.}$  y  $P_{..k}$

Las mejores estimaciones se basan en los totales de las variables marginales simples

$$\hat{p}_{i..} = \frac{N_{i..}}{N} \quad \hat{p}_{.j.} = \frac{n_{.j.}}{N} \quad \hat{p}_{..k} = \frac{n_{..k}}{N}$$

estos son estimadores de máxima verosimilitud

Sustituyendo queda:

$$E_{ijk} = \frac{N n_{i..} n_{.j.} n_{..k}}{N^2}$$

$$= \frac{n_{i..} n_{.j.} n_{..k}}{N^2}$$

Una vez obtenidos los valores esperados se calcula

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^l \frac{(n_{ijk} - E_{ijk})^2}{E_{ijk}}$$

Para determinar los grados de libertad de la estadística  $\chi^2$ , es necesario calcular el número de parámetros estimados en el procedimiento y el número de celdas de la tabla, ya que:

grados de libertad = número de celdas - número de parámetros estimados.

En este caso el número de celdas es  $rcl$  y el número de parámetros estimados son:

$r - 1$  por las estimaciones de  $\hat{p}_{i..}$ ,  $-1$  es debido a la restricción

$$\sum_{i=1}^r \hat{p}_{i..} = 1$$

$c - 1$  por las estimaciones de  $\hat{p}_{.j.}$

$l - 1$  por las estimaciones de  $\hat{p}_{..k}$

$1$  por la estimación global

En total el número de parámetros estimados son  $r + c + l - 2$ , con lo cual los grados de libertad asociados con:  $rcl - r - c - l + 2$

Si la prueba de independencia global resulta no significativa, un análisis más profundo de la tabla es innecesario, pero cuando resulta significativa pueden existir importantes asociaciones entre al menos dos de las variables; por ejemplo, puede haber una asociación entre dos de ellas y la tercera ser completamente independiente, en este caso se tendrá interés en probar una hipótesis de independencia parcial; en otros casos hay dos variables con independencia condicional, dado el nivel de la tercera.

Las hipótesis de independencia parcial son:

$H_0^{(1)}: p_{ijk} = p_{i..} p_{.jk}$  (clasificación por renglón independiente de las clasificaciones por columna y por estrato)

$H_0^{(2)}: p_{ijk} = p_{.j.} p_{i.k}$  (clasificación por columna independiente de las de renglón y estrato)

$H_0^{(3)}: p_{ijk} = p_{..k} p_{ij.}$  (clasificación por estrato independiente de las de renglón y columna)

$H_0^{(1)}$  Implica que la probabilidad de una observación en la  $ijk$ -ésima celda ( $p_{ijk}$ ) está dada por el producto de la probabilidad de la  $i$ -ésima categoría de la variable renglón ( $p_{i..}$ ) con la probabilidad en la  $jk$ -ésima celda de la columna por el estrato ( $p_{.jk}$ ). Si la hipótesis es cierta, esto implica que la clasificación por renglón es independiente tanto de la clasificación por columna como por estrato y que la siguiente hipótesis compuesta es cierta:

$$p_{ij.} = p_{i..} p_{.j.} \quad \text{y} \quad p_{i.k} = p_{i..} p_{..k}$$

Para probar la hipótesis se procede de la misma manera que en la prueba de independencia global, con el cálculo de los valores esperados

$$E_{ijk} = N \hat{p}_{i..} \hat{p}_{.jk}$$

En que los estimadores  $\hat{p}_{i..}$  y  $\hat{p}_{.jk}$  de las probabilidades  $p_{i..}$  y  $p_{.jk}$  son obtenidos de los totales marginales:

$$\hat{p}_{i..} = \frac{n_{i..}}{N} \quad \hat{p}_{.jk} = \frac{n_{.jk}}{N}$$

usando esas estimaciones de probabilidad se calcula:

$$\begin{aligned} E_{ijk} &= \frac{N n_{i..}}{N} \frac{n_{.jk}}{N} \\ &= \frac{n_{i..} n_{.jk}}{N} \end{aligned}$$

Después se calcula  $\chi^2$  y los grados de libertad están dados por:

$$gl = clr - cl = r + 1$$

Las hipótesis de independencia condicional son:

$H_0^{(1)}$ :  $p_{ijk} = p_{i.k} p_{.jk}$  (la primera y segunda variables son independientes para cada nivel de la tercera variable)

$H_0^{(2)}$ :  $p_{ijk} = p_{ij.} p_{.jk}$  (la primera y tercera variables son independientes para cada nivel de la segunda variable)

$H_0^{(3)}$ :  $p_{ijk} = p_{i.k} p_{.ij}$  (la segunda y tercera variables son independientes para cada nivel de la primera variable)

Si  $H_0^{(1)}$  es cierta, quiere decir que la clasificación por renglón es independiente de la clasificación por columna para cada nivel de la clasificación por estrato.

Para probar esta hipótesis se calculan los valores esperados considerando el nivel k fijo, los cuales están dados por:

$$E_{ijk} = n_{..k} p_{ijk} = n_{..k} p_{i.k} p_{.jk}$$

los estimadores se obtienen de los totales marginales (para k fijo)

$$\hat{p}_{i.k} = \frac{n_{i.k}}{n_{..k}} \qquad \hat{p}_{.jk} = \frac{n_{.jk}}{n_{..k}}$$

Con éstos se calcula  $E_{ijk}$

$$E_{ijk} = \frac{n_{..k} n_{i.k} n_{.jk}}{n_{..k} n_{..k}}$$

$$= \frac{n_{i.k} n_{.jk}}{n_{..k}}$$

### Modelos logarítmicos lineales

En los modelos loglineales el logaritmo de los valores esperados de las observaciones son una combinación lineal de un número de parámetros, los cuales son estimados mediante técnicas de máximas verosimilitud o de mínimos cuadrados. La ventaja de estos modelos es que permiten el análisis de tablas multidimensionales complejas así como estimar la magnitud y la importancia de los efectos de interés. Los modelos a considerar son los jerárquicos en

el sentido que cuando una interacción esté presente, todas las interacciones de menor orden también están.

Recordando la hipótesis de independencia en tablas de dos dimensiones:  $P_{ij} = P_{i.} P_{.j}$  se encuentra que mediante los modelos logarítmicos lineales,  $p_{ij}$  o alguna función de ella puede expresarse con la suma de las probabilidades marginales o alguna función de ellas en forma similar al modelo de análisis de varianza; tomando logaritmos naturales:  $\ln p_{ij} = \ln p_{i.} + \ln p_{.j}$  o bien, en términos de frecuencias:  $\ln F_{ij} = \ln F_{i.} + \ln F_{.j} - \ln N$

Sumando sobre i se tiene:

$$\sum_{j=1}^r \ln F_{ij} = \sum_{i=1}^r \ln F_{i.} + r \ln F_{.j} - r \ln N$$

Sumando sobre j se tiene:

$$\sum_{j=1}^c \ln F_{ij} = c \ln F_{i.} + \sum_{j=1}^c \ln F_{.j} - c \ln N$$

Sumando sobre i y j:

$$\sum_{i=1}^r \sum_{j=1}^c \ln F_{ij} = c \sum_{i=1}^r \ln F_{i.} + r \sum_{j=1}^c \ln F_{.j} - rc \ln N$$

La ecuación de  $\ln F_{ij}$  puede reescribirse como:

$$\ln F_{ij} = u + u_1(i) + u_2(j) \quad \text{donde} \quad u = \frac{\sum_{i=1}^r \sum_{j=1}^c \ln F_{ij}}{rc}$$

$$u_1(i) = \frac{\sum_{j=1}^c \ln F_{ij}}{c} - \frac{\sum_{i=1}^r \sum_{j=1}^c \ln F_{ij}}{rc}$$

$$u_2(j) = \frac{\sum_{i=1}^r \ln F_{ij}}{r} - \frac{\sum_{i=1}^r \sum_{j=1}^c \ln F_{ij}}{rc}$$

La hipótesis nula plantea que este modelo se ajusta a los datos; en este tipo de modelos, el interés se centra en analizar el efecto de las interacciones entre las variables, quedando en segundo término el estudio de los efectos principales, a diferencia del análisis de varianza usual. Por lo anterior se introduce un término extra que representa la interacción entre dos variables y constituye, en este caso, el modelo saturado.

$$\ln F_{ij} = u + u_1(i) + u_2(j) + u_{12}(ij) \quad \begin{array}{l} \text{frecuencia} \\ \text{esperada} \end{array} = \begin{array}{l} \text{frecuencia} \\ \text{observada} \end{array}$$

El ajuste es perfecto  $g_1 = 0$

Los subíndices numéricos denotan las variables particulares involucradas y los subíndices alfabéticos, las categorías de esas variables en el mismo orden, así  $u_{12}(ij)$  representa el efecto de la interacción entre los niveles  $i$  y  $j$  de las variables 1 y 2 respectivamente.

La hipótesis de independencia será  $u_{12}(ij) = 0$  que es equivalente a afirmar que el primer modelo, sin interacción se ajusta a los datos.

La estimación de los efectos de interacción será útil en la identificación de las categorías responsables de la no independencia.

El modelo saturado para una tabla de tres dimensiones es:

$$\ln F_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk) + u_{123}(ijk)$$

Otra hipótesis en términos de parámetros es la de independencia global que es específica que no hay asociaciones entre las tres variables (no hay interacciones de primero ni de segundo orden).

$$H_0: u_{12}=0 \quad u_{13}=0 \quad u_{23}=0 \quad u_{123}=0$$

en este caso el modelo, que es el más simple, es:

$$\ln F_{ijk} = u + u_1(i) + u_2(j) + u_3(k)$$

Si se considera específicamente el modelo en que  $u_{12}=0$ , por ser modelos jerárquicos  $u_{123}=0$  y el modelo loglineal quedaría

$$\ln F_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{13}(ik) + u_{23}(jk)$$

esto indica que no hay interacción entre las variables 1 y 2 en cada nivel de la variable 3 o que las variables 1 y 2 son condicionalmente independientes dada la variable 3.

Similarmente, la hipótesis de independencia parcial puede considerarse en términos del modelo saturado en el que algunos parámetros son igual a cero y se especifica que  $u_{123}$  y un par de  $u_{12}$ ,  $u_{13}$  y  $u_{23}$  es cero, por ejemplo:

$$H_0^{(1)}: \ln F_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{23}(jk)$$

Si se continúan quitando términos del modelo saturado hasta tener pocos términos se tiene un modelo completamente independiente que no incluye todas las variables (conservando la característica de modelo jerárquico).

Los valores esperados correspondientes a algunos modelos no pueden obtenerse directamente de los totales marginales de valores observados, ya que las ecuaciones de máxima verosimilitud no tienen una solución explícita; una forma de obtener los valores esperados, bajo el modelo propuesto, es mediante un método iterativo propuesto por Deming y Stephan en 1940 y descrito más tarde en 1969 por Bishop.

Por ejemplo, para una tabla tridimensional con un modelo sin interacciones de segundo orden:

$$\ln F_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk)$$

El procedimiento iterativo asume un valor inicial  $E_{ijk}^{(0)}$  para cada  $E_{ijk}$  donde  $E_{ij.} = n_{ij.}$  (primera restricción marginal)

$$E_{ijk}^{(1)} = \frac{E_{ijk}^{(0)} \times n_{ij.}}{E_{ij.}^{(0)}}$$

Los valores esperados  $E_{ijk}^{(1)}$  son ahora ajustados para satisfacer la segunda restricción marginal  $E_{i.k} = n_{i.k}$

$$E_{ijk}^{(2)} = \frac{E_{ijk}^{(1)} \times n_{i.k}}{E_{i.k}^{(1)}}$$

El ciclo se completa cuando los valores dados por la expresión anterior son ajustados para satisfacer  $E_{.jr} = n_{.jr}$  usando

$$E_{ijk}^{(3)} = \frac{E_{ijk}^{(2)} \times n_{.jk}}{E_{.jk}^{(2)}}$$

Para la interpretación de las pruebas en que se emplea un modelo loglineal se procede de la siguiente manera:

Si se trata, por ejemplo, de una tabla tridimensional; se examinará primero un modelo que contenga sólo los efectos principales lo que implica la hipótesis de que las tres variables son globalmente independientes; se calculan los valores esperados mediante la fórmula:

$$E_{ijk} = \frac{n_{i..} \cdot n_{.j.} \cdot n_{..k}}{N^2}$$

y son comparados con las frecuencias observadas mediante la razón de máxima verosimilitud ( $X^2_L$  o deviance)

$$2 \sum O \ln \left[ \frac{E}{O} \right] \quad \text{En donde } O \text{ son los valores observados}$$

E son los valores esperados

y la suma se realiza sobre el número de celdas en la tabla

Grados de libertad = Número de celdas en la tabla - Número de parámetros en el modelo ajustado que requiere la estimación

$$g1 = r1 - r - c - 1 + 2$$

Si el valor  $X^2_L$  (para cierto número de grados de libertad) es mayor que la  $X^2$  de tablas, es significativo e indica que el modelo que incluye sólo efectos principales no proporciona un adecuado ajuste a los datos. Si los grados de libertad son mayores de 100 se utiliza una aproximación normal,  $Z = \sqrt{2 X^2} - \sqrt{2n-1}$  \*

Debe ahora considerarse un modelo con interacciones de primer orden:

$$\ln F_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk)$$

En este caso los valores esperados se obtienen por el procedimiento iterativo que ya se mencionó. Los grados de libertad son:

$$g1 = r1 - rc - r1 - c1 + r + c + 1 - 1$$

Si  $X^2_L$  es no significativa, no hay necesidad de postular interacciones de segundo orden.

Para llegar al modelo más apropiado que incluya las interacciones de primer orden más importantes se examinan los efectos estimados estandarizados, que se obtienen dividiendo el parámetro estimado entre su error estándar; estos efectos estandarizados tienen una distribución asintóticamente normal estándar bajo la hipótesis de que este efecto es cero, por lo cual se comparan con una tabla Z a un particular nivel de significancia, por ejemplo si  $\alpha = 0.05$  los efectos estandarizados superiores, en valor absoluto, a 1.96 son significativos. Esto permite seleccionar un modelo que incluya sólo las interacciones donde haya valores significativos, el cual puede ser un modelo de independencia condicional o parcial.

Si al calcular  $X^2_L$  ésta resulta menor que  $X^2$  de tablas ello demuestra que el modelo se ajusta adecuadamente a los datos.

Si se vuelven a calcular y examinar los valores estandarizados se podrá identificar en qué niveles de cada variable se dan las asociaciones.

La ausencia en el modelo de una interacción de segundo orden implica que las variables 1 y 3 interactúan de la misma manera con todos los niveles de la variable 2, que la interacción entre las variables 2 y 3 es la misma para todos los niveles de la variable 1 y que la interacción entre las variables 1 y 2 es la misma para todos los niveles de la variable 3

Si el número de dimensiones de la tabla aumenta, aumenta también el número de posibles modelos; para conocer cuál es el modelo que se ajusta más adecuadamente a los datos, es muy útil el procedimiento de examinar los efectos estandarizados ya que esto indica qué parámetros

\*Tomado de: Arkin, H. y Colton, R. Tables for Statisticians. Barnes & Noble Books Nueva York (1963)



pueden ser excluidos y qué modelos más simples que el saturado son dignos de considerarse. En todo caso un modelo que incluya parámetros extra, sólo se preferirá al modelo simple si produce un mejoramiento significativo al ajustarse a los datos.

Los métodos de cálculo se basan en que las frecuencias, en todas las celdas, son mayores o iguales a uno, por lo que puede presentar problemas de existencia de ceros. Los ceros pueden ser de dos tipos:

- Ceros muestrales, se presentan debido a la variabilidad de la muestra pero existen individuos en la población que caen dentro de la clasificación correspondiente a la celda. Una forma de corregir este problema es aumentar el tamaño de la muestra o adicionar una constante pequeña, como 1, a todas las -- frecuencias de las celdas.
- Ceros estructurales o ceros "a priori", se presentan cuando es imposible observar valores para ciertas combinaciones de variables, quiere decir que ningún individuo en la población cumple con esa combinación de características y por lo tanto esa celda tiene una probabilidad de cero.

Las tablas con ceros estructurales se conocen como tablas de contingencia incompletas y su análisis presenta problemas especiales. Los grados de libertad asociados a un modelo particular, en este caso son:

$gl = N_1 - N_2 - N_3$  donde  $N_1$  = número de celdas en la tabla

$N_2$  = número de parámetros en el modelo

$N_3$  = número de celdas con ceros "a priori"

El ajuste de modelos (y de aquí el cálculo de valores esperados) puede hacerse mediante el paquete GLIM (Generalized Linear Iterative Modeling) que es un programa Fortran para ajustar ciertos modelos lineales generalizados, diseñado por J.A Nelder de la Universidad de Oxford el cual está implementado en la Computadora Burroughs 7800 de la UNAM. Consta de una parte central relativa a la especificación y ajuste de los modelos y una parte auxiliar que efectúa lectura y transformación de datos, gráficas, etc.

El paquete GLIM es muy útil como herramienta para comparar el ajuste de varios modelos para un conjunto de datos, ya que puede usarse en forma interactiva.

-24-

## ANÁLISIS ESTADÍSTICO

Para iniciar el análisis estadístico se decidió formar dos grupos de variables, las que se tratarían de manera descriptiva y aquellas en las que se pretendía encontrar alguna interrelación; las primeras se refieren a las características generales de la población en estudio (turno, edad, sexo, estado civil, carrera, escuela de procedencia) y a las relacionadas con algún ítem como situación socioeconómica.

De las gráficas: 2, 3 y 4 se desprende que predomina en el turno vespertino:

- una edad mayor a 19 años
- más individuos del sexo masculino
- más individuos que desempeñan trabajo remunerado.

En la gráfica 5 se observa que en los dos turnos la escuela de procedencia es oficial.

En las gráficas 6 y 7 puede interpretarse que la escolaridad de los padres en general es baja (11% o menos, son profesionales), con los siguientes matices:

- más baja en el turno vespertino que en el matutino
- más baja en la madre que en el padre.

La gráfica 8 indica que el ingreso familiar es proporcionalmente más bajo en el turno vespertino que en el matutino.

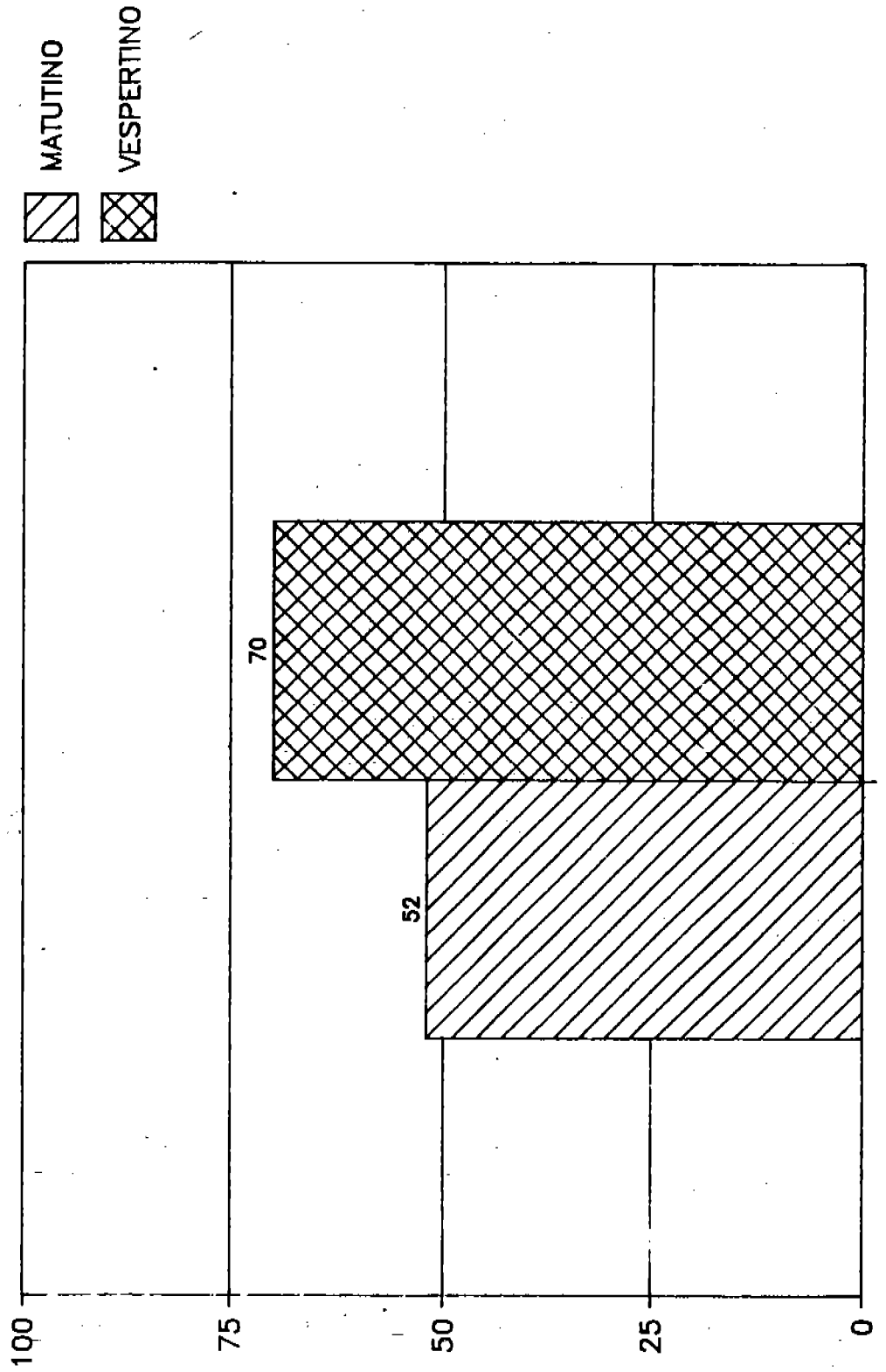
En el segundo grupo, se definieron parejas de variables en las que se harían pruebas de ji cuadrada de independencia, empleando el paquete SPSS (Statistical Package for the Social Sciences), en la computadora Burroughs 7800 de la UNAM.

Una vez obtenidas las tablas de contingencia se encontró que no había independencia, al 5%, en los siguientes pares de variables:

- VI (turno) con V16 (técnicas de estudio)
- VI (turno) con V21 (trabajo remunerado actual)
- VI (turno) con V33 (actividad de la madre)
- VI (turno) con V63 (opinión sobre el uso de técnicas didácticas modernas)
  
- V5 (carrera) con V16 (técnicas de estudio)
- V5 (carrera) con V39 (rendimiento en el primer semestre)
- V5 (carrera) con V46 (% de exámenes aprobados en química)
- V5 (carrera) con V48 (% de exámenes aprobados en matemáticas)
- V5 (carrera) con V49 (% de exámenes aprobados en laboratorio)
- V5 (carrera) con V63 (opinión sobre el uso de técnicas didácticas)

GRAFICA 2

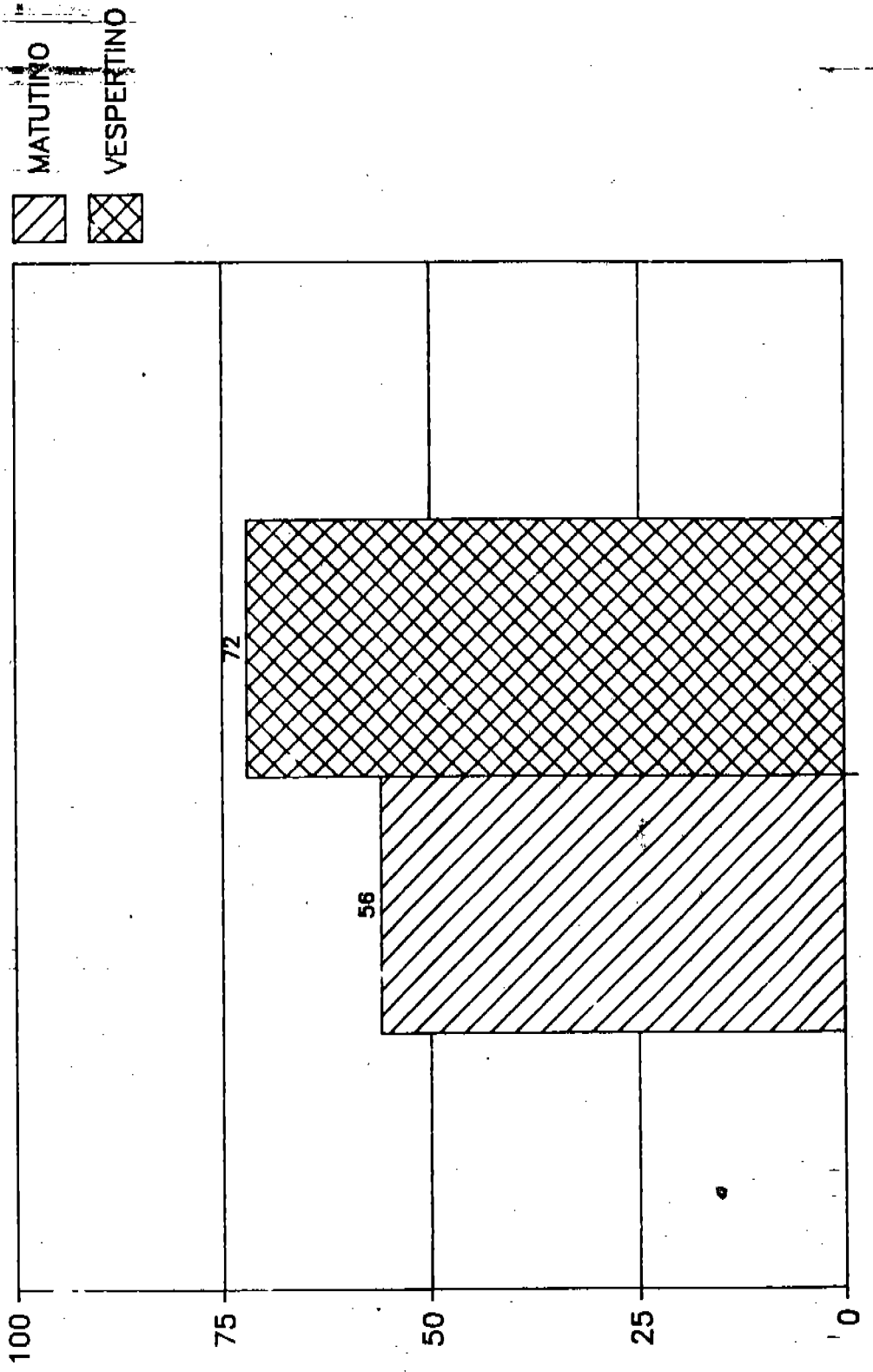
EDAD: MAYORES DE 19 AÑOS  
ENEP - Zaragoza



GRAFICA 3

# SEXO: MASCULINO

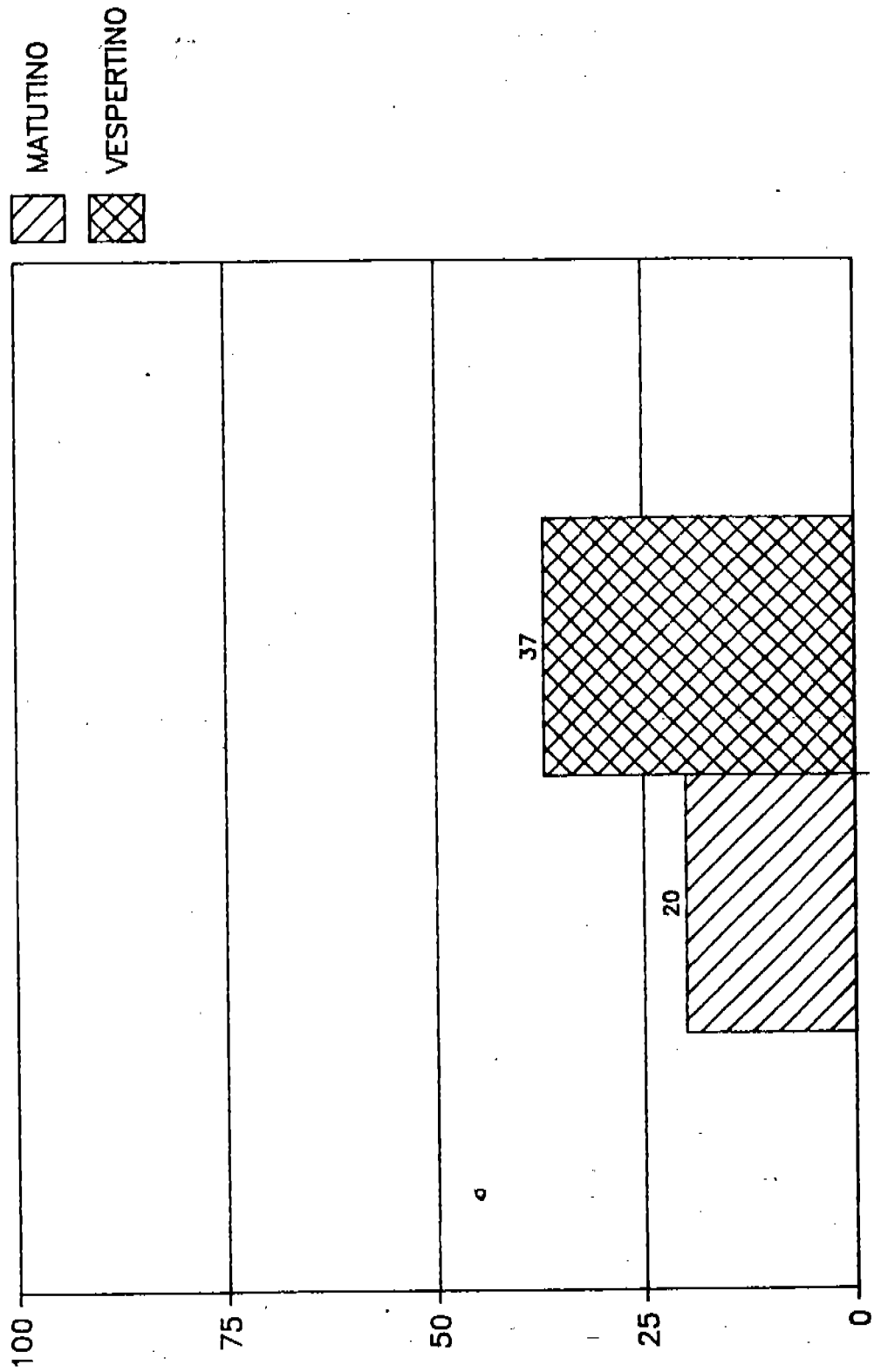
ENEP-Zaragoza



GRAFICA 4

# TRABAJO REMUNERADO

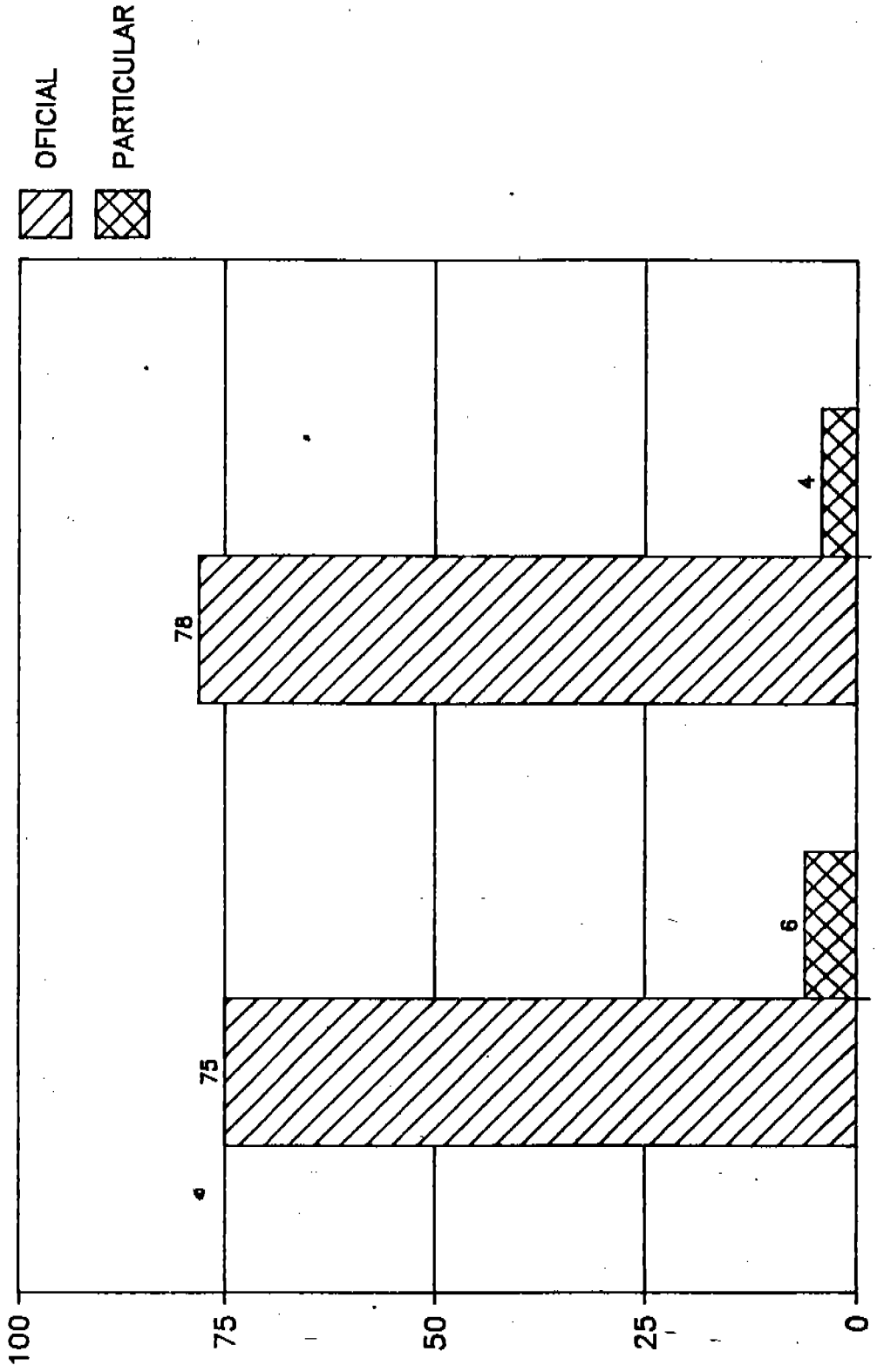
ENEP-Zaragoza



GRAFICA 5

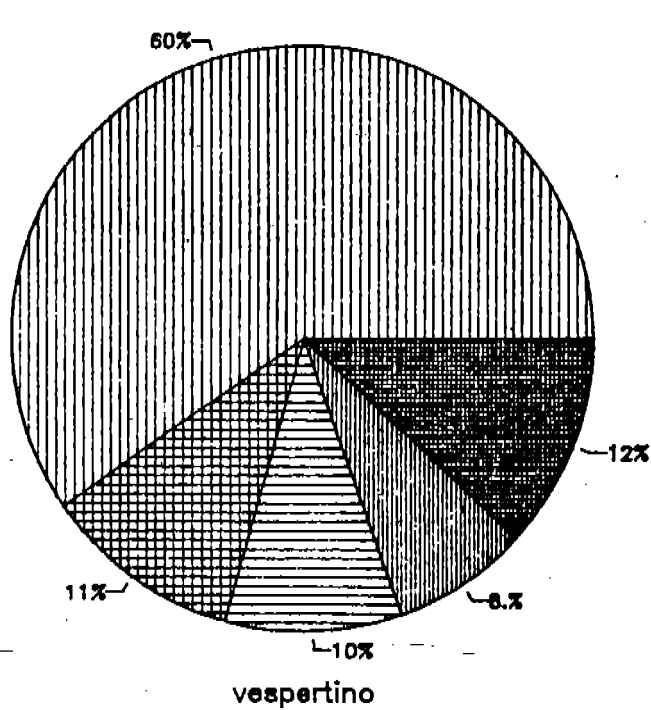
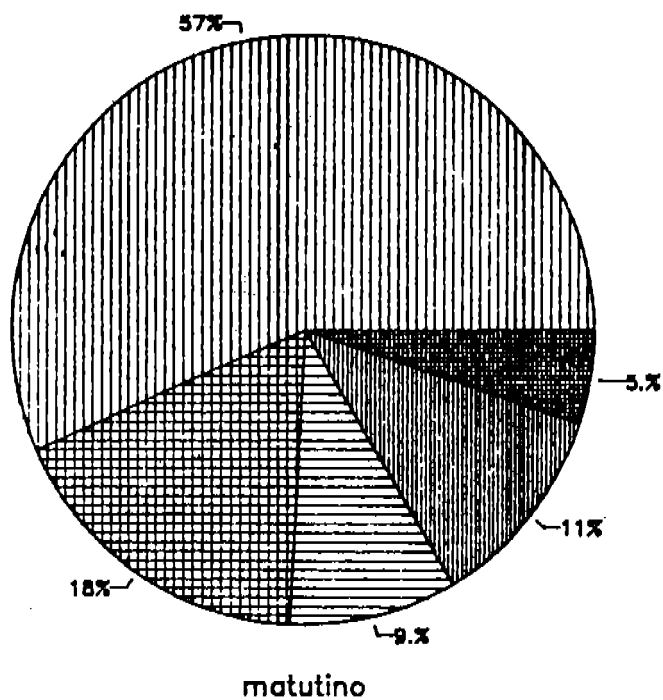
# ESCUELA DE PROCEDENCIA

ENEP-Zaragoza

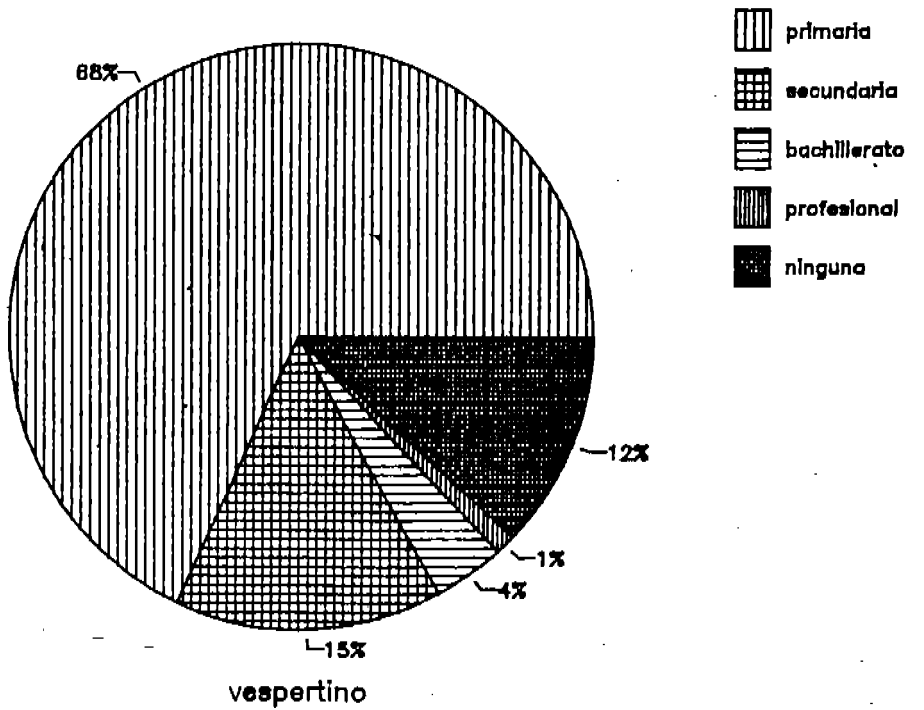
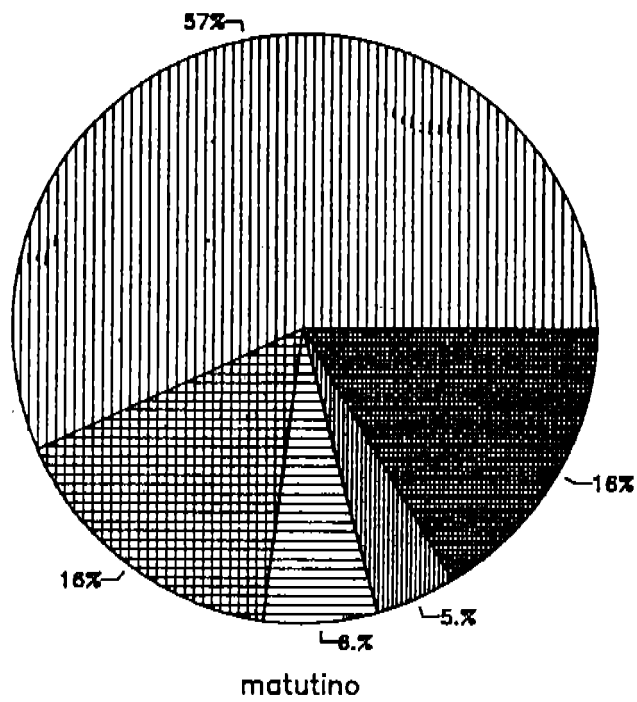


### ENEP-ZARAGOZA

escolaridad del padre



### ENEP-ZARAGOZA escolaridad de la madre





V6 (escuela de procedencia en el D.F.)	con V21 (trabajo remunerado)
V6 "	con V30 (escolaridad del padre)
V6 "	con V31 (actividad del padre)
V6 "	con V49 (% exámenes de exámenes aprobados en el laboratorio)
V6 "	con V52 (sistema en la ENEP Zaragoza)
V6 "	con V55 (opinión sobre cartas descriptivas)
V6 "	con V57 (opinión sobre exámenes departamentales)
V7 (escuela de procedencia de provincia)	con V16 (técnica de estudio)
V7 "	con V31 (actividad del padre)
V7 "	con V34 (ingreso familiar)
V7 "	con V39 (rendimiento en el primer semestre)
V9 (no vive con su familia)	con V31 (actividad del padre)
V9 "	con V39 (rendimiento en el primer semestre)
V9 "	con V57 (opinión sobre exámenes departamentales)
V21 (trabajo remunerado)	con V39 (rendimiento en el primer semestre)
V30 (escolaridad del padre)	con V39 (rendimiento en el primer semestre)
V30 "	con V48 (% de exámenes aprobados en matemáticas)
V30 "	con V49 (% de exámenes aprobados en el laboratorio)
V31 (actividad del padre)	con V46 (% de exámenes aprobados en química)
V31 "	con V49 (% de exámenes aprobados en el laboratorio)
V32 (escolaridad de la madre)	con V39 (rendimiento en el primer semestre)
V32 "	con V46 (% de exámenes aprobados en química)
V32 "	con V48 (% de exámenes aprobados en matemáticas)

V6 (escuela de procedencia en el D.F.)	con V21 (trabajo remunerado)
V6 "	con V30 (escolaridad del padre)
V6 "	con V31 (actividad del padre)
V6 "	con V49 (% exámenes de exámenes aprobados en el laboratorio)
V6 "	con V52 (sistema en la ENEP Zaragoza)
V6 "	con V55 (opinión sobre cartas descriptivas)
V6 "	con V57 (opinión sobre exámenes departamentales)
V7 (escuela de procedencia de provincia)	con V16 (técnica de estudio)
V7 "	con V31 (actividad del padre)
V7 "	con V34 (ingreso familiar)
V7 "	con V39 (rendimiento en el primer semestre)
V9 (no vive con su familia)	con V31 (actividad del padre)
V9 "	con V39 (rendimiento en el primer semestre)
V9 "	con V57 (opinión sobre exámenes departamentales)
V21 (trabajo remunerado)	con V39 (rendimiento en el primer semestre)
V30 (escolaridad del padre)	con V39 (rendimiento en el primer semestre)
V30 "	con V48 (% de exámenes aprobados en matemáticas)
V30 "	con V49 (% de exámenes aprobados en el laboratorio)
V31 (actividad del padre)	con V46 (% de exámenes aprobados en química)
V31 "	con V49 (% de exámenes aprobados en el laboratorio)
V32 (escolaridad de la madre)	con V39 (rendimiento en el primer semestre)
V32 "	con V46 (% de exámenes aprobados en química)
V32 "	con V48 (% de exámenes aprobados en matemáticas)

V33 (actividad del padre)		con V46 (% de exámenes aprobados en química)
V33	"	con V48 (% de exámenes aprobados en matemáticas)
V33	"	con V49 (% de exámenes aprobados en el laboratorio)
V34 (ingreso familiar)		con V39 (% rendimiento en el primer semestre)
V34	"	con V46 (% de exámenes aprobados en química)
V34	"	con V48 (% de exámenes aprobados en matemáticas)
V34	"	con V49 (% de exámenes aprobados en laboratorio)

Mediante modelos logarítmicos lineales y empleando el paquete GLIM se corrieron modelos de dos variables con su interacción, por ejemplo:

$$u + V1 + V16 + V1.V16, \quad u + V1 + V21 + V1.V21, \quad u + V1 + V33 + V1.V33$$

etcétera, de esta manera se corroboró la interrelación de esas variables y en algunos casos como en V1 y V33 resultaron independientes .

A partir de las asociaciones por pares de variables se formaron grupos que llevaron a analizar asociaciones más complejas que permiten satisfacer los objetivos planteados. Los grupos resultantes incluyen entre dos y cuatro variables y son los siguientes:

grupo 1: V6, V39, V55 y V57 (escuela de procedencia en el D.F., rendimiento en el primer semestre, opinión sobre cartas descriptivas y opinión sobre exámenes departamentales)

grupo 2: V1, V16, V21 y V63 (turno, técnica de estudio, trabajo remunerado, opinión sobre técnicas didácticas modernas)

grupo 3: V5, V16, V21 y V39 (carrera, técnica de estudio, trabajo remunerado y rendimiento en el primer semestre)

grupo 4: V5, V39, V46 y V49 (carrera, rendimiento en el primer semestre, % de exámenes aprobados en química y % de exámenes aprobados en laboratorio)

grupo 5: V6 y V31 (escuela de procedencia en el D.F. y actividad del padre)

Posteriormente se analizaron, con el paquete GLIM, las asociaciones de estos grupos, surgiendo la necesidad en algunas variables de reducir categorías para evitar la presencia de un gran número de ceros en las celdas; así, V6 (escuela de procedencia en el D.F.) disminuyó de 5 a 3 categorías; V39 (rendimiento en el primer semestre) disminuyó de 5 a 3 categorías; V35 (opinión sobre el uso de cartas descriptivas) disminuyó de 3 a 2 categorías; V57 (opinión sobre exámenes departamentales) disminuyó de 3 a 2 categorías; V46 (exámenes aprobados en química) disminuyó de 4 a 3 categorías. La disminución se hizo de manera que las variables conservaran la misma información, lo cual produjo que las categorías que se unieron fueran contiguas, debido a que casi todas las variables son de tipo ordinal.

El uso de modelos logarítmicos lineales con el grupo 1 se explicará detalladamente para describir todos los pasos seguidos en el tratamiento de cada grupo de variables.

#### Análisis del grupo 1

El grupo de variables V6, V39, V55 y V57 se obtuvo del resultado de tablas de ji cuadrada que determinaba posibles asociaciones entre pares de variables y de la necesidad de satisfacer los objetivos planteados.

Al correr el modelo:  $u + V6 + V39 + V55 + V57$  (con 180 datos; listado 1), que implica la hipótesis de que las cuatro variables son independientes se obtiene un valor para la deviance de 172.7 con 168 grados de libertad, que al calcular  $z = \sqrt{2(172.7)} - \sqrt{2(168) - 1} = 0.27$  resulta no significativa con respecto a 1.96, es decir, que el modelo que sólo incluye efectos principales parece dar un ajuste adecuado a los datos; sin embargo, esta decisión puede no ser conveniente, ya que, debido a la gran cantidad de ceros que presenta la tabla, se crean sesgos en la estimación de los efectos por lo cual se hace necesario llevar a cabo una reducción de categorías en las variables quedando de la siguiente manera:

V6 (bachillerato cursado en el D.F.) disminuye de 5 a 3 categorías:

- a) Preparatoria de la UNAM
- b) CCH de la UNAM
- c) otras

V39 (aprobación en el primer semestre de estudios profesionales) disminuye de 5 a 3 categorías:

- a) Rendimiento bueno (75% a 100% de exámenes aprobados)
- b) Rendimiento medio (50% a 74% de exámenes aprobados)
- c) Rendimiento bajo (menor a 49% de exámenes aprobados)

V55 (opinión sobre uso de cartas descriptivas) disminuye de 3 a 2 categorías:

- a) Favorecen el aprendizaje
- b) No favorecen el aprendizaje

V57 (opinión sobre exámenes departamentales) disminuye de 3 a 2 categorías:

- a) Favorecen el aprendizaje
- b) No favorecen el aprendizaje

Al correr el modelo de independencia global (listado 2), con las variables reducidas en sus categorías (36 datos) se obtiene una deviance de 69.51 con 29 grados de libertad, que comparada con el valor de tablas en que  $\alpha = 0.05$ ,

$\chi^2_{29} = 42.6$  resulta significativa, es decir, se rechaza la hipótesis de independencia.

Al incluir las interacciones de primer orden de estas cuatro variables en el modelo:  $u + V6 + V39 + V55 + V57 + V6.V39 + V6.V55 + V6.V57 + V39.V55 + V39.V57 + V55.V57$

se obtiene una deviance de 10.93 con 16 grados de libertad, que comparada con  $\chi^2_{16} = 26.3$  es no significativa, esto es, el modelo explica apropiadamente el comportamiento de la tabla y para verificar la necesidad de todas las interacciones se analiza la significancia de los efectos estandarizados. Al estandarizar, se observa que las interacciones significativas son: V6.V39 y V55.V57 .

Valor estimado	Error estándar	Parámetro	Efecto estandarizado
3.184	0.1996	ZGM	
0.7589	0.2398	V6(2)	
-0.9187	0.3748	V6(3)	
0.2639	0.2636	V39(2)	
-0.8940	0.3632	V39(3)	
-3.715	0.8091	V55(2)	
-3.178	0.6635	V57(2)	
-0.4249	0.3267	V6 (2).V39(2)	-1.3005
0.1301	0.4269	V6 (2).V39(3)	0.3047
-0.0947	0.4909	V6 (3).V39(2)	0.1929
1.374	0.5394	V6 (3).V39(3)	2.5472 ←
1.169	0.7886	V6 (2).V55(2)	1.4823
-7.739	29.73	V6 (3).V55(2)	-0.2603
0.1422	0.5707	V39(2).V55(2)	0.2491
-1.407	0.8519	V39(3).V55(2)	-1.6516
1.068	0.6542	V6 (2).V57(2)	1.6325
-0.0992	0.9638	V6 (3).V57(2)	-0.1029
-0.9213	0.5892	V39(2).V57(2)	-1.5636
0.8884	0.5030	V39(3).V57(2)	1.7662
2.344	0.5721	V55(2).V57(2)	4.0971 ←

Los efectos estimados estandarizados se obtienen al dividir el parámetro estimado entre su error estándar; los que son significativos son aquellos mayores, en valor absoluto, con respecto al valor de tablas de z al nivel de significancia deseado (1.96 para  $\alpha = 0.05$ ).

Al correr el modelo  $u + V6 + V39 + V55 + V57 + V6.V39 + V55.V57$  se obtiene una deviance de 38.99 ( $\chi^2_{24} = 36.4$ ). con los siguientes valores estandarizados:

Valor estimado	Error estándar	Parámetro	Efecto estandarizado
3.098	0.1978	ZGM	
0.9466	0.2311	V6 (2)	
-0.9555	0.3721	V6 (3)	
0.2384	0.2622	V39(2)	
-0.8602	0.3597	V39(3)	
-3.109	0.3232	V55(2)	
-2.416	0.2333	V57(2)	
-0.4919	0.3208	V6 (2).V39(2)	-1.5333
0.1520	0.4178	V6 (2).V39(3)	0.3638
0.09806	0.4901	V6 (3).V39(2)	0.1999
1.391	0.5368	V6 (3).V39(3)	2.5912 ←
2.311	0.5153	V55(2).V57(2)	4.4850 ←

El modelo  $u + V6 + V39 + V55 + V57 + V6.V39 + V55.V57$  se puede interpretar de la siguiente manera:

Hay asociación entre V6 y V39 y entre V55 y V57 sin que haya asociación entre esos dos grupos (son conjuntamente independientes) y se pueden interpretar por separado.

V6(3).V39(3).- Es mucho más frecuente que un rendimiento bajo en el primer semestre esté asociado con la escuela de procedencia diferente de CCH y preparatoria, es el efecto más marcado en el contexto de la tabla que relaciona V6 con V39, es el de mayor significancia.

V6(2).V39(2).- Es más probable que se encuentre un rendimiento medio en egresados de la preparatoria que en egresados del CCH.

V55(2).V57(2).- Entre alumnos de la carrera del área químico-biológica están fuertemente asociadas las opiniones de que no son útiles las cartas descriptivas ni los exámenes departamentales y esto es independiente tanto de la escue-

la de procedencia como del rendimiento.

### Análisis del grupo 2

Este grupo está formado por las variables: V1 (turno), V16 (técnica de estudio), V21 (trabajo remunerado) y V63 (opinión sobre técnicas didácticas modernas).

Se corrió el modelo  $u + V1 + V16 + V21 + V63$  y se obtuvo una deviance de 54, con 32 grados de libertad, que frente a la  $\chi^2_{12} = 46.2$  resulta significativa. Al introducir las interacciones y correr el modelo :  $u + V1 + V16 + V21 + V63 + V1.V16 + V1.V21 + V1.V63 + V16.V21 + V16.V63 + V21.V63$  se obtuvo una deviance de 18.57 con 17 grados de libertad, que comparada con  $\chi^2_{17} = 27.6$  es no significativa.

Los efectos estandarizados que resultaron significativos fueron: V1.V16, V1.V21, V1.V63 y al correrse el modelo con los efectos principales y estas tres interacciones, la deviance fue de 23.96 con 26 grados de libertad, que no es significativa ( $\chi^2_{26} = 38.9$ ).

Al observar los efectos estandarizados se infiere que una gran tendencia en los alumnos del turno vespertino a considerar que el uso del material didáctico no favoreció su aprendizaje porque las técnicas didácticas fueron aplicadas defectuosamente y de manera esporádica (V1(2).V63(5) y V1(2).V63(4)).

Una gran cantidad de alumnos del turno vespertino opina que su técnica de estudio hizo que aprobaran con dificultad los exámenes en el primer semestre del ciclo profesional (V1(2).V16(2)).

El efecto estandarizado correspondiente a V1(2).V21(2) es negativo y se interpreta como que hay menos estudiantes del turno vespertino que no trabajan y estudian que los del turno matutino.

La independencia de las variables 16, 21 y 63 está condicionada al turno, en general, no puede decirse que el trabajar y estudiar no determina la opinión acerca del material didáctico (independencia entre V21 y V63), sino que esto sólo es posible afirmarlo para alumnos del turno vespertino. Igual debe considerarse para los otros pares de variables, V16 y V21 así como V16 y V63.



### Análisis del grupo 3

Otro grupo de variables asociadas, obtenido a través de tablas de contingencia fué: V5 (carrera), V16 (técnica de estudio), V21 Trabajo remunerado) y V39 (rendimiento en el primer semestre).

Como V39 presentaba gran número de ceros, se redujo en sus categorías de tres a cinco como ya se indicó en los grupos anteriores.

Se corrió el modelo  $u + V5 + V16 + V21 + V39$  y se obtuvo una deviance de 80.75 con 29 grados de libertad, que frente al valor de tablas  $\chi^2_{29} = 42.5$  es significativa; se corrió entonces el modelo con interacciones:

$u + V5 + V16 + V21 + V39 + V5.V16 + V5.V21 + V5.V39 + V16.V21 + V16.V39 + V21.V39$  y se obtuvo una deviance de 24.43 con 16 grados de libertad que resultó no significativa ( $\chi^2_{16} = 26.29$ )

Al calcular los valores estandarizados, son significativas las siguientes interacciones: V5.V16 y V16.V39; V5.V39 resulta significativa sólo en dos combinaciones de categorías.

Al correr el modelo  $u + V5 + V16 + V21 + V39 + V5.V16 + V16.V39$  la deviance es de 42.48 con 25 grados de libertad, significativa ( $\chi^2_{25} = 37$ ).

Al correr el modelo  $u + V5 + V16 + V21 + V39 + V5.V16 + V5.V39 + V16.V39$  la deviance es de 33.95 con 21 grados de libertad, que frente a  $\chi^2_{21} = 32.67$ , es no significativa.

Al observar los valores estandarizados, resultan significativos en todos sus niveles V5.V16 y V16.V39, de manera práctica ésto se interpreta como que es muy frecuente que en la carrera de Biología, en especial y también en la de QFB los alumnos tengan una técnica de estudio defectuosa que ha hecho que aprueben con dificultad las asignaturas del ciclo profesional (V5.V16) y que una técnica de estudio defectuosa esté fuertemente asociada a un rendimiento medio y bajo en el primer semestre (V16.V39).

En la interacción V5.V39, es muy significativa la asociación entre la carrera de Biología y un rendimiento de los alumnos correspondientes, bajo en el primer semestre. En menor grado, esto mismo ocurre para alumnos de la carrera de QFB.

### Análisis del grupo 4

El grupo: V5 (carrera), V39 (rendimiento en el primer semestre), V46 (exámenes aprobados en Química I) y V49 (exámenes aprobados en Laboratorio de Ciencia Básica) se obtuvo del resultado de tablas de  $\chi^2$ . Es interesante hacer notar que la variable 48 (exámenes aprobados en Matemáticas I) no aparece en este grupo, esto probablemente se debe a que el resultado en matemáticas por experiencia se sabe que es muy deficiente pero muy homogéneo, es decir, no hay diferencia respecto a la carrera (V5) y opinión sobre el rendimiento en el primer semestre (V39); respecto a ésta última, parece estar influida especialmente por los resultados en química y laboratorio, pero no en matemáticas.

Al correr el modelo  $u + V5 + V39 + V46 + V49$ , con 240 datos se obtiene una deviance de 536.6 con 277 grados de libertad. Para analizar la significancia se calcula  $z = \sqrt{2(536.6)} - \sqrt{2(277)-1} = 11.4$  que resulta mucho mayor que 1.96. De nuevo se duda respecto a tomar una decisión basándose en esos datos, debido a la gran cantidad de ceros en la tabla; se decidió entonces disminuir categorías a las variables, quedando de la siguiente manera:

#### V5 Carrera

- a) IQ
- b) QFB
- c) Biología

#### V39 Aprobación en el primer semestre (disminuye de 5 a 3 categorías)

- a) Rendimiento bueno, 75 a 100% de exámenes aprobados
- b) Rendimiento medio, 50 a 74% de exámenes aprobados
- c) Rendimiento bajo, menos de 49% de exámenes aprobados

#### V46 Exámenes aprobados en Química I (disminuye de 4 a 3 categorías)

- a) Más del 80%
- b) 50 a 79%
- c) Menos del 50%

#### V49 Exámenes aprobados en Laboratorio de Ciencia Básica (disminuye de 4 a 3 categorías)

- a) Más del 50%
- b) del 50 al 79%
- c) Menos del 50%

Al correr el modelo  $u + V5 + V39 + V46 + V49$ , con la reducción de categorías, se obtiene una deviance de 315.5 con 72 grados de libertad, significativa;

al introducir las seis interacciones la deviance fue de 39.95 con 48 grados de libertad, no significativa comparada con  $\chi^2_{50} = 67.5$ .

Al examinar los efectos estandarizados de las interacciones, las que resultan significativas al 5% son: V5.V46, V39.V46, V39.V49 y V46.V49.

Se corrió entonces el modelo:  $u + V5 + V39 + V46 + V49 + V5.V46 + V39.V46 + V39.V49 + V46.V49$ , la deviance fue del 45.12 con 56 grados de libertad ( $\chi^2_{50} = 67.5$ ), no significativa, por tanto es el modelo que se ajusta mejor a los datos.

En este caso las cuatro variables tienen una gran dependencia entre sí. Probablemente la interacción de más peso es V39.V46 en la que en todos los niveles hay dependencia, especialmente en el tercer nivel, lo que indica que es muy frecuente un rendimiento bajo en el primer semestre con un porcentaje de exámenes aprobados de Química I menor de 50.

En la interacción V39.V49, es muy frecuente que un rendimiento bajo en el primer semestre esté asociado a un rendimiento medio en el laboratorio y un rendimiento bajo en el primer semestre tiene menos tendencia a asociarse con un rendimiento bajo en el laboratorio.

En la interacción V5.V46, el efecto de más peso es en el tercer nivel de las dos variables, hay más probabilidad de que los alumnos de Biología tengan menos del 50% de exámenes aprobados en Química I y, en menor grado, ésta tendencia se repite para QFB. Los alumnos de la carrera de Biología tienden a tener del 50 al 79% de los exámenes aprobados en laboratorio.

En el caso de V46.V49, todas las interacciones son negativas; al tratar de interpretar se llegaba a una contradicción en términos de la experiencia que se tiene; al analizar los residuales del ajuste se encontró que el dato número 62 tenía asociado un residual extraordinariamente grande (5.049), al identificar el caso correspondiente se encontró que es un estudiante de Biología con buen rendimiento en el primer semestre (V39) y sin embargo con un porcentaje de exámenes aprobados de Química I inferior a 50 y en el laboratorio aproximadamente 50%, lo cual es contradictorio; se decidió eliminar este dato, por lo que la frecuencia en esa celda es cero, y correr de nuevo el modelo.

Se encontró que V46 y V49 presentan independencia condicionada a la presencia de V39 (rendimiento en el primer semestre), así, un rendimiento bajo en el primer semestre implica bajo rendimiento en Química I, pero no puede afirmarse nada respecto al laboratorio.

En este caso no fue posible simplificar el análisis colapsando con alguna variable puesto que la independencia que se detectó fue de tipo condicional, por lo cual, la magnitud del efecto de las interacciones, que originalmente son significativas, cambiaría al colapsar con cualquiera de las variables.

#### Análisis del grupo 5

El grupo está formado por V6(escuela de procedencia en el D.F.) y V31 (actividad del padre).

Al correr el modelo  $u + V6 + V31$  se obtiene un valor de la deviance de 23.11 con 12 grados de libertad, que comparado con la  $\chi^2_{12} = 21.02$  es significativa, es decir, que el modelo que sólo incluye efectos principales no da un ajuste adecuado a los datos, por lo que es necesario introducir el efecto de interacción con lo cual se obtiene el modelo saturado:  $u + V6 + V31 + V6.V31$  para el cual se tiene un valor pequeño de la deviance, ese valor debía ser cero pero por la aproximación numérica se reporta  $0.4540 \times 10^{-4}$ , que comparado con  $\chi^2_1 = 3.84$  es no significativa, por lo tanto el modelo da un ajuste adecuado a los datos como se esperaba.

Al examinar los efectos estimados estandarizados y comparar con una tabla de Z a un nivel de significancia del 5% se observa que no hay valores significativos, sin embargo, a un nivel de 10% sí los hay para las interacciones:

V6(2).V31(3).- Esto indica que se observan menos estudiantes del CCH con padre industrial, comerciante o agricultor que los correspondientes que provienen de la preparatoria, ya que el efecto es negativo; es decir, es poco probable que un hijo de industrial, comerciante o agricultor provenga de un CCH cuando estudia una carrera del área químico-biológica en la ENEP "Zaragoza".

V6(2).V31(2).- Hay cierta tendencia a que haya menos hijos de obreros que provienen del CCH que quieren estudiar una carrera del área químico-biológica.

V6(3).V31(4).- El efecto es positivo y se interpreta como que los alumnos del Colegio de Bachilleres, que estudian una carrera del área químico-biológica tienen padres campesinos o jornaleros.

V6(3).V31(3).- Es el efecto más pequeño registrado y se infiere que es poco probable encontrar que un individuo que estudia en el área químico-biológica y provenga del Colegio de Bachilleres tenga un padre con actividad industrial, de comercio o sea agricultor.

A nivel práctico se interpreta que hay asociación entre la escuela de procedencia en el D.F. y la actividad del padre, pero que esta dependencia se da de manera semejante en todas las categorías de cada variable.

-44-

LISTADO 1

R\*SERVICIO/GLIM  
 #RUNNING 8021  
 #?

GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON

\$UNITS 180

\$DATA Y

\$READ

```

1 4 0 0 0 0 0 0 0 0 16 1 0 4 0 0 0 0
9 21 0 0 2 0 0 1 0 3 4 1 1 0 0 0 0 1
0 1 0 0 0 0 0 0 0 3 10 1 0 2 0 0 2 0
2 26 4 0 7 2 0 3 2 4 35 0 0 5 2 0 4 2
0 12 5 0 4 1 0 0 1 0 7 1 0 2 0 0 0 0
0 0 0 0 0 0 0 0 0 0 4 0 0 1 0 0 0 0
1 1 0 0 5 0 0 0 0 2 2 1 0 1 0 0 0 0
0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0
2 0 0 0 1 0 0 0 0 2 2 1 0 2 0 0 0 0
0 5 0 0 2 0 0 0 0 0 2 0 0 1 0 0 0 0
  
```

\$FACTOR V6 4: V39 5: V55 3: V57 3

\$CAL V6=%GL(4,45) : V39=%GL(5,9) : V55=%GL(3,3) : V57=%GL(3,1)

\$LINK L

\$ERROR F

\$YVAR Y

\$FIT + V6 + V39+ V55+ V57\$

SCALED		
CYCLE	DEVIANCE	DF
5	172.7	168

\$DISPLAY E\$

	ESTIMATE	S.E.	PARAMETER
1	-0.4607	0.2811	%GM
2	0.7555	0.1449	V6(2)
3	-1.304	0.2587	V6(3)
4	-1.157	0.2444	V6(4)
5	1.099	0.2309	V39(2)
6	1.376	0.2238	V39(3)
7	0.6098	0.2485	V39(4)
8	-0.5108	0.3266	V39(5)
9	-1.460	0.1637	V55(2)
10	-2.516	0.2598	V55(3)
11	1.843	0.1903	V57(2)
12	-0.2076	0.2640	V57(3)
SCALE PARAMETER TAKEN AS			1.000

SCALED  
CYCLE DEVIANCE DF  
10 72.88 124

\$DISPLAY E\*

	ESTIMATE	S.E.	PARAMETER
1	0.7068	0.5631	XGM
2	0.2847	0.6161	V6(2)
3	-9.811	34.14	V6(3)
4	-0.7164	0.8985	V6(4)
5	<del>-.4778E-01</del>	<del>0.7479</del>	V39(2)
6	1.347	0.6040	V39(3)
7	-.9976E-01	0.7383	V39(4)
8	-10.69	28.22	V39(5)
9	-3.714	1.015	V55(2)
10	-10.71	22.81	V55(3)
11	0.2033	0.5944	V57(2)
12	-4.040	1.347	V57(3)
13	-0.8665	0.6049	V6(2).V39(2)
14	-1.021	0.5794	V6(2).V39(3)
15	-0.8371	0.6691	V6(2).V39(4)
16	0.7132	1.185	V6(2).V39(5)
17	8.096	34.14	V6(3).V39(2)
18	8.053	34.14	V6(3).V39(3)
19	9.103	34.14	V6(3).V39(4)
20	9.608	34.17	V6(3).V39(5)
21	-0.9326	1.078	V6(4).V39(2)
22	-0.5760	0.9533	V6(4).V39(3)
23	0.6623	1.004	V6(4).V39(4)
24	2.259	1.477	V6(4).V39(5)
25	0.5632	0.4668	V6(2).V55(2)
26	1.036	0.7962	V6(2).V55(3)
27	1.646	0.6386	V6(3).V55(2)
28	-6.481	27.03	V6(3).V55(3)
29	1.621	0.6416	V6(4).V55(2)
30	-6.438	27.85	V6(4).V55(3)
31	0.3788	0.7071	V39(2).V55(2)
32	-0.5194	0.9369	V39(2).V55(3)
33	0.2213	0.6983	V39(3).V55(2)
34	-.3890E-01	0.8870	V39(3).V55(3)
35	0.1097	0.7629	V39(4).V55(2)
36	-1.204	1.142	V39(4).V55(3)
37	-.5776E-01	0.9271	V39(5).V55(2)
38	-8.136	26.81	V39(5).V55(3)
39	1.181	0.4815	V6(2).V57(2)
40	2.199	0.7929	V6(2).V57(3)
41	-0.1198	0.7426	V6(3).V57(2)
42	-0.2736	1.381	V6(3).V57(3)
43	-0.7597	0.6333	V6(4).V57(2)
44	-0.8677	1.331	V6(4).V57(3)
45	1.853	0.7411	V39(2).V57(2)
46	3.131	1.323	V39(2).V57(3)
47	0.7682	0.5974	V39(3).V57(2)
48	1.151	1.279	V39(3).V57(3)
49	1.002	0.7153	V39(4).V57(2)
50	3.358	1.313	V39(4).V57(3)
51	9.648	28.21	V39(5).V57(2)
52	10.44	28.25	V39(5).V57(3)
53	1.478	0.7857	V55(2).V57(2)
54	1.777	0.9513	V55(2).V57(3)
55	7.680	22.80	V55(3).V57(2)
56	9.602	22.80	V55(3).V57(3)

SCALE PARAMETER TAKEN AS 1.000

LISTADO 2

GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON

\$UNITS 36

\$DATA Y

\$READ

25	0	1	0	32	1	0	0	9	0	1	1
50	5	7	5	44	4	2	2	25	0	7	1
10	0	0	0	13	0	1	0	16	0	1	0

\*FACTOR U6 3 : U39 3 : U55 2 : U57 2

\*CAL U6=%GL(3,12) : U39=%GL(3,4) : U55=%GL(2,1) : U57=%GL(2,2)

\*LINK L

\*ERROR P

\*YVAR Y

\*FIT +U6 +U39 + U55 + U57\*

SCALED

CYCLE	DEVIANCE	DF
4	69.51	29

\*DISPLAY E\*

	ESTIMATE	S.E.	PARAMETER
1	3.119	0.1447	%GM
2	0.7754	0.1444	U6(2)
3	-0.5349	0.1966	U6(3)
4	-.3961E-01	0.1407	U39(2)
5	-0.5239	0.1615	U39(3)
6	-2.553	0.2373	U55(2)
7	-2.088	0.1965	U57(2)
SCALE PARAMETER TAKEN AS			1.000



\$FIT +U6.V39 + U6.V55 + U6.V57 + U39.V55 + U39.V57 + U55.V57\$

CYCLE	SCALED DEVIANCE	DF
9	10.93	16

\$DISPLAY E\$

	ESTIMATE	S.E.	PARAMETER
1	3.184	0.1996	ZGM
2	0.7589	0.2398	U6(2)
3	-0.9187	0.3748	U6(3)
4	0.2639	0.2636	U39(2)
5	-0.8940	0.3632	U39(3)
6	-3.715	0.8091	U55(2)
7	-3.178	0.6635	U57(2)
8	-0.4249	0.3267	U6(2).U39(2)
9	0.1301	0.4269	U6(2).U39(3)
10	.9471E-01	0.4909	U6(3).U39(2)
11	1.374	0.5394	U6(3).U39(3)
12	-1.169	0.7886	U6(2).U55(2)
13	-7.739	29.73	U6(3).U55(2)
14	0.1422	0.5707	U39(2).U55(2)
15	-1.407	0.8519	U39(3).U55(2)
16	1.068	0.6542	U6(2).U57(2)
17	-.9924E-01	0.9638	U6(3).U57(2)
18	-0.9213	0.5892	U39(2).U57(2)
19	0.8884	0.5030	U39(3).U57(2).PAGE.
20	2.344	0.5721	U55(2).U57(2)
SCALE PARAMETER TAKEN AS			1.000

\$FIT-U6.U55 - U6.U57 - U39.U55 -U39.U57\$

CYCLE	SCALED DEVIANCE	DF
4	38.99	24

\$DISPLAY E\$

	ESTIMATE	S.E.	PARAMETER
1	3.098	0.1978	%GM
2	0.9466	0.2311	U6(2)
3	-0.9555	0.3721	U6(3)
4	0.2384	0.2622	U39(2)
5	-0.8602	0.3597	U39(3)
6	-3.109	0.3232	U55(2)
7	-2.416	0.2333	U57(2)
8	-0.4919	0.3208	U6(2).U39(2)
9	0.1520	0.4178	U6(2).U39(3)
10	.9806E-01	0.4901	U6(3).U39(2)
11	1.391	0.5368	U6(3).U39(3)
12	2.311	0.5153	U55(2).U57(2)
SCALE PARAMETER TAKEN AS			1.000

\$FIT +U39.U57\$

CYCLE	SCALED DEVIANCE	DF
4	31.70	22

\$DISPLAY E\$

	ESTIMATE	S.E.	PARAMETER
1	3.080	0.2001	%GM
2	0.9466	0.2311	U6(2)
3	-0.9555	0.3721	U6(3)
4	0.3215	0.2659	U39(2)
5	-0.9241	0.3666	U39(3)
6	-3.109	0.3232	U55(2)
7	-2.263	0.3221	U57(2)
8	-0.4919	0.3208	U6(2).U39(2)
9	0.1520	0.4178	U6(2).U39(3)
10	.9806E-01	0.4901	U6(3).U39(2)
11	1.391	0.5368	U6(3).U39(3)
12	-0.9990	0.5464	U39(2).U57(2)
13	0.4207	0.4460	U39(3).U57(2)
14	2.311	0.5153	U55(2).U57(2)
SCALE PARAMETER TAKEN AS			1.000

\$END

## CONCLUSIONES

El uso de los modelos logarítmicos lineales da como resultado una mayor información que la obtenida con las tablas de contingencia; permite conocer en qué parte de la tabla se presenta la falta de independencia, así como su magnitud e importancia. Permite también estudiar la interrelación que existe entre grupos de tres o más variables, tales como:

- Los alumnos del área químico-biológica de la ENEP "Zaragoza" provenientes del Colegio de Ciencias y Humanidades (CCH) y de la Escuela Nacional Preparatoria, UNAM, tienen tendencia a un mejor rendimiento que los de otras escuelas como el Colegio de Bachilleres, Preparatoria Popular, de provincia, etcétera. Los del CCH presentan menos probabilidades de tener un rendimiento medio que los de Preparatoria.
- Hay gran tendencia en los alumnos del turno vespertino, que en su mayoría desempeñan trabajo remunerado, a considerar que el uso de material didáctico no favoreció su aprendizaje ya que las técnicas didácticas se aplicaron defectuosa y esporádicamente y que su técnica de estudio los hizo aprobar con dificultad los exámenes del primer semestre.
- Los alumnos de la carrera de Biología opinan que su técnica de estudio es defectuosa lo que dificulta la aprobación de las asignaturas y se asocia con un rendimiento bajo en el primer semestre; esto mismo ocurre, aunque en menor grado, con los alumnos de la carrera de QFB. Se encuentra también que una técnica de estudio defectuosa está en relación con un bajo rendimiento en el primer semestre.
- El rendimiento bajo en el primer semestre está asociado a un número que es menor del 50% de exámenes aprobados en Química y mayor del 50% en Laboratorio.
- Es poco probable encontrar que los padres de los alumnos del área químico-biológica, provenientes del CCH o del Colegio de Bachilleres, pertenezcan a estratos socioeconómicos de ingresos altos (dueños de fábricas, comercios o agroindustrias). Por otra parte hay pocos hijos de obreros que estudien en esta área y que provengan del CCH.

Con este análisis se corroboraron ideas concebidas con anterioridad acerca de los alumnos del área mencionada; tal es el caso de los estudiantes del turno vespertino y los de la carrera de Biología, quienes se encuentran en las condiciones más desfavorables respecto a su rendimiento escolar en el primer semestre del ciclo profesional.

Por otra parte, con este trabajo se llegó a datos más particulares que los conocidos con anterioridad:

- La opinión de los alumnos sobre el rendimiento en el primer semestre está influida sólo por los resultados en Química y Laboratorio y no en Matemáticas ni en el Seminario de Problemas Socioeconómicos de México, probablemente porque en estas últimas asignaturas, aunque en Matemáticas el índice de reprobación es sumamente alto y en el Seminario es muy bajo, el resultado es tan homogéneo y tan constante que no hay diferencia perceptible.
- Los padres de alumnos de carreras del área químico-biológica en la ENEP "Zaragoza", provenientes de un plantel del CCH, rara vez pertenecen a estratos socioeconómicos de ingreso alto y pocos son obreros; probablemente, en su mayoría, pertenecen a estratos medios.

Hay evidencias de que las características culturales y económicas del núcleo familiar son determinantes en el desempeño de los alumnos, ya que dependen de ellas aspectos como disciplina, hábitos de estudio, selección de la carrera, etcétera. Esto trae como consecuencia que el ámbito escolar les presente fuertes dificultades al ingresar al ciclo profesional.

En vista de los resultados obtenidos con este trabajo, parece claro que no todas las variables consideradas son relevantes para satisfacer los objetivos del estudio base, es decir, variables que no resultaron asociadas con alguna otra o variables cuyo comportamiento es evidente, proporcionan información poco útil para el estudio, provocan gasto de recursos y poca precisión en los análisis. Por lo tanto, si se desea aumentar la confiabilidad de las conclusiones de este trabajo, especialmente en términos de precisión, se desprende la

necesidad de realizar otro estudio en el cual se consideren solamente las variables que resultaron relevantes, aplicando un cuestionario más reducido y es pecífico y que se tome información de ellas en una muestra de alumnos, con lo cual se tendrá también una reducción de costos.

A P E N D I C E

CUESTIONARIO

Lea con cuidado las siguientes preguntas y conteste con la mayor veracidad, perforando en su tarjeta la opción que haya seleccionado.

1.- TURNO \_\_\_\_\_

- a) Matutino
- b) Vespertino

2.- EDAD \_\_\_\_\_

- a) de 17 a 19 años
- b) de 19 a 21
- c) más de 21

3.- SEXO \_\_\_\_\_

- a) Masculino
- b) Femenino

4.- ESTADO CIVIL \_\_\_\_\_

- a) soltero
- b) casado
- c) otro

5.- CARRERA \_\_\_\_\_

- a) Ingeniero Químico
- b) Químico Farmacéutico Biólogo
- c) Biólogo

6.- Escuela en la que realizó sus estudios de bachillerato: \_\_\_\_\_

- a) En el Distrito Federal
- b) Preparatoria UNAM
- c) C.C.N. UNAM
- d) Colegio de Bachilleres
- e) Preparatoria Particular
- f) C.C.N. Particular

7.- En Provincia \_\_\_\_\_

- a) Escuela Oficial, plan de 2 años
- b) Escuela particular, plan de 2 años
- c) Escuela Oficial, plan de 3 años
- d) Escuela particular, plan de 3 años

8.- Usted vive: \_\_\_\_\_

- a) Con su familia (padres, hermanos)
- b) Con familiares (tíos, primos)
- c) Con amigos de la familia o compañeros
- d) En casa de asistencia, pensión, hotel, etc.
- e) Con esposo(a)

9.- Si no vive con su familia, ¿año se debe a: \_\_\_\_\_

- a) Sus familiares no viven en la ciudad de México
- b) Problemas familiares
- c) Otras razones

10.- En sus estudios de bachillerato ¿cómo usted las \_\_\_\_\_

- a) En todas las materias
- b) En más de la mitad de las materias
- c) En menos de la mitad de las materias
- d) En ninguna materia

11.- En sus estudios de bachillerato ¿ha cumplido el pro- \_\_\_\_\_

- a) Programa académico establecido?
- b) En todas las materias
- c) En más de la mitad de las materias
- d) En menos de la mitad de las materias

12.- Considere usted que los conocimientos adquiridos en \_\_\_\_\_

- a) Preparatoria cubren los requisitos necesarios para sus estudios profesionales en su:
- b) 100%
- c) 75%
- d) 50%
- e) 25%

13.- El promedio obtenido por usted en el bachillerato fue: \_\_\_\_\_

- a) Porcentaje menor
- b) entre 7 y 8
- c) entre 8 y 9
- d) más de 9
- e) menor de 7

14.- Con qué frecuencia acostumbra usted estudiar: \_\_\_\_\_

- a) Diariamente
- b) Cada dos días
- c) Una vez a la semana
- d) Sólo cuando tiene exámenes
- e) Más de 12 horas
- f) Entre 6 y 12 horas
- g) Menos de 6 horas

15.- Su técnica de estudio le ha permitido aprobar las \_\_\_\_\_

- a) asignaturas en el ciclo profesional
- b) con dificultad
- c) con facilidad

- 17.- Asista a sus clases en bachillerato 17
  - a) Con regularidad
  - b) Con irregularidad
- 18.- Asiste a sus clases en profesional 18
  - a) Con regularidad
  - b) Con irregularidad
- 19.- Si existe o existe con irregularidad, es por razones. 19
  - a) De salud
  - b) De trabajo
  - c) Familiares
  - d) De compromisos deportivos, de actividades artísticas
  - e) Otras.
- 20.- Durante sus estudios de bachillerato ¿trabajaba usted 20 además de estudiar?
  - a) Si
  - b) No
- 21.- Actualmente ¿trabaja usted además de estudiar? 21
  - a) Si
  - b) No
- 22.- Si trabaja lo hace para: 22
  - a) Sostenerse usted solo
  - b) Ayudar a su familia
  - c) Sostenerse usted y su familia
- 23.- Si trabaja, qué tipo de trabajo desempeña: 23
  - a) Obrero
  - b) Esplendo
  - c) Artesano
  - d) Técnico o profesional
  - e) Otro
- 24.- Actualmente participa usted en actividades ajenas al 24 estudio
  - a) De carácter deportivo
  - b) De trabajo
  - c) De carácter político
  - d) De carácter cultural o social
  - e) Más de una de las actividades arriba mencionadas
- 25.- Si su respuesta anterior es afirmativa, cuánto tiempo 25 le dedica a la semana
  - a) Menos de 6 horas
  - b) De 7 a 12 horas
  - c) De 13 a 18 horas
  - d) De 19 a 24 horas.
- 26.- El tiempo diario que dedica a transportarse de su 26 domicilio a la ENEP o viceversa
  - a) Menos de una hora diaria
  - b) Entre una y dos horas diarias
  - c) Entre 2 y 3 horas diarias
  - d) Más de 3 horas diarias
- 27.- Indique cómo acostumbra afectuar sus comidas al mediodía 27
  - a) En su casa
  - b) En una cafetería o restaurante
  - c) Tortas o sandwiches, con el tiempo suficiente y en un lugar adecuado
  - d) Cualquier alimento entre horas
  - e) No come
- 28.- Usted estudia habitualmente en: 28
  - a) Apuntes tomados en la clase
  - b) Libros de la biblioteca de esta escuela
  - c) Libros propios
  - d) Libros de otras bibliotecas
- 29.- Un estudiante utiliza frecuentemente la biblioteca, si 29 usted no le hace señale la causa
  - a) Prefiere usar libros propios
  - b) Problemas de falta de tiempo por trabajar
  - c) Prefiere usar apuntes porque le es difícil estudiar en libros
  - d) No está familiarizado con el uso de la biblioteca
- 30.- Los estudios realizados por su padre fueron 30
  - a) Ninguno
  - b) Primaria
  - c) Secundaria
  - d) Bachillerato
- 31.- La actividad de su padre es: 31
  - a) Profesional o técnico
  - b) Obrero o empleado
  - c) Industrial, comerciante o agricultor (dueño del establecimiento)
  - d) Campesino o jornalero (ejidatario, peon, peonetero, etc).
  - e) Artesano



- 38.- Al elegir esta carrera
- a) Lo hizo como resultado de sugerencia de familiares - 38  
amigos o maestros
  - b) Lo hizo como resultado de un estudio vocacional
  - c) Porque es lo que más le gusta
  - d) No eligió esta carrera, pero no había cupo en la carrera de su preferencia
  - e) Lo hizo porque está bien remunerada
- 39.- En este primer semestre de estudios profesionales, ha aprobado: 39
- a) El 100% de los exámenes presentados
  - b) Del 75% al 99% de los exámenes presentados
  - c) Del 50% al 74% de los exámenes presentados
  - d) Del 25% al 49% de los exámenes presentados
  - e) Menos del 25% de los exámenes presentados
- 40.- Considere que la preparación en matemáticas obtenida por usted en el bachillerato es: 40
- a) Deficiente
  - b) Suficiente
  - c) Excelente
  - d) No cursó la materia
- 41.- Considere que la preparación en física obtenida por usted en el bachillerato es: 41
- a) Deficiente
  - b) Suficiente
  - c) Excelente
  - d) No cursó la materia
- 42.- Considere que la preparación en química obtenida por usted en el bachillerato es: 42
- a) Deficiente
  - b) Suficiente
  - c) Excelente
  - d) No cursó la materia
- 43.- Si tiene dificultades en comprender los conceptos que el maestro enseña, piense que se deba a: 43
- a) Deficiencias en sus conocimientos previos
  - b) Deficiencias en la exposición del maestro
  - c) Deficiencias en la relación maestro-alumno
  - d) Los motivos a y b
  - e) Todos los motivos anteriores

- 32.- Los estudios realizados por su madre fueron 32
- a) Ninguno
  - b) Primaria
  - c) Secundaria
  - d) Bachillerato
  - e) Profesional o posgrado
- 33.- La actividad de su madre es: 33
- a) Labores del hogar
  - b) Profesional o técnico (actualmente trabajando)
  - c) Industrial o comerciante (dueña del establecimiento)
  - d) Empleada
  - e) Obrera
- 34.- Los ingresos de su familia son: 34
- a) Menos de 3,000.00 Pesos
  - b) de \$3,000.00 a \$ 5,000.00
  - c) de \$5,000.00 a \$10,000.00
  - d) de \$10,000.00 a \$20,000.00
  - e) más de \$20,000.00
- 35.- Tiene usted hermanos? 35
- a) No
  - b) Uno
  - c) Dos
  - d) Tres
  - e) Más de tres
- 36.- Qué lugar ocupa usted entre su familia? 36
- a) Primero
  - b) Segundo
  - c) Tercero
  - d) Cuarto
  - e) Quinto u otro
- 37.- En el tiempo que lleva en esta escuela ¿piensa usted que la elección de carrera: 37
- a) Está de acuerdo con su vocación o aptitudes
  - b) No está de acuerdo con su vocación o aptitudes
  - c) No es la carrera que había elegido

- 51.- En las materias en las que hay profesor de taller y profesor de teoría 31
  - a) Considera que tuvo usted una mejor atención
  - b) El profesor de taller elaboró mucho para mejorar su rendimiento
  - c) Considera que es indiferente que haya profesor de taller
- 52.- Piensa usted que el sistema de enseñanza que se emplea en esta escuela es diferente al tradicional 32
  - a) En un 100%
  - b) En más del 50%
  - c) En menos del 50%
  - d) Solo es diferente en el laboratorio
  - e) No es diferente
- 53.- usted sabe en que consiste el sistema modular? 33
  - a) Si, se lo informaron al ingresar a la UNP-ZAHUACOSA
  - b) Si, usted lo averiguó fuera de UNP-ZAHUACOSA
  - c) No, aunque se lo informaron
  - d) No, nunca se lo informaron
- 54.- Usted ha utilizado en sus cursos programas en forma de cartas descriptivas? 34
  - a) Si
  - b) No
- 55.- Si la respuesta a la pregunta anterior es afirmativa - su opinión sobre el uso de cartas descriptivas es que: 35
  - a) Favorecen el aprendizaje del alumno
  - b) Favorecen al alumno pero el profesor no se apega a ellas
  - c) No favorecen al alumno
- 56.- Los exámenes que usted ha presentado han sido de tipo departamental? 36
  - a) Si
  - b) No
- 57.- Si la respuesta a la pregunta anterior es afirmativa, su opinión sobre este tipo de exámenes es que: 37
  - a) Favorecen siempre el aprendizaje del alumno
  - b) Favorecen al alumno solo si el profesor alcanza a revisar todos los objetivos y si está elaborado adecuadamente.
  - c) No favorecen al alumno en ningún caso

- 44.- Respecto a los profesores, usted considera: 44
  - a) La mayoría de los profesores conocen bien su materia
  - b) Algunos profesores conocen bien su materia
  - c) Casi ningún profesor conoce bien su materia
- 45.- Respecto a la asistencia y puntualidad de los profesores usted considera que: 45
  - a) La mayoría de los profesores asisten puntual y regularmente
  - b) Algunos profesores asisten puntual y regularmente
  - c) Casi ningún profesor asiste regularmente
- 46.- Del total de exámenes presentados en Química I qué porcentaje ha aprobado 46
  - a) 100%
  - b) 80%
  - c) 50%
  - d) -50%
- 47.- Del total de exámenes presentados en Seminarios de Problemas Socioeconómicos de México que porcentaje ha aprobado 47
  - a) 100%
  - b) 80%
  - c) 50%
  - d) Menos del 50%
- 48.- Del total de exámenes presentados en matemáticas I qué porcentaje ha aprobado 48
  - a) 100%
  - b) 80%
  - c) 50%
  - d) Menos del 50%
- 49.- Del total obligatorio de unidades en L.C.B. qué porcentaje ha aprobado 49
  - a) 100%
  - b) 80%
  - c) 50%
  - d) Menos del 50%
- 50.- Si ha tenido dificultades al cursar el primer semestre de profesional piensa que puede superarla con: 50
  - a) Mayor estudio y esfuerzo personal
  - b) Mayor comunicación con el profesorado
  - c) Cambio en sus actividades extracurriculares (trabajo, deportes, etc.)
  - d) Todos los anteriores
  - e) Otras soluciones

- 58.- En sus cursos teóricos sus maestros usan material didáctico (transparencias, películas, retroproyector, etc.) -- además del pizarrón? 58
- a) si
  - b) no
- 59.- Si su respuesta anterior es afirmativa indique con qué frecuencia lo usaban? 59
- a) En todas las clases
  - b) En más de la mitad de las clases
  - c) En menor de la mitad de las clases
  - d) En muy pocas clases
- 60.- En sus cursos teóricos sus maestros usan técnicas de enseñanza como seminario, mesa redonda y discusión abierta, además de la exposición oral? 60
- a) Si
  - b) No
- 61.- Qué porcentaje de maestros usan esas técnicas? 61
- a) El 100%
  - b) Más del 50%
  - c) Menos del 25%
- 62.- Los maestros que usan esas técnicas, con qué frecuencia lo hacen? 62
- a) En todas las clases
  - b) En más de la mitad de las clases
  - c) En menos de la mitad de las clases
  - d) En muy pocas clases
- 63.- En su opinión, el uso del material didáctico y de las técnicas educacionales mencionadas favorecieron su aprendizaje? 63
- a) Si, lo favorecieron notablemente
  - b) Lo favorecieron ligeramente
  - c) No, porque las técnicas se aplicaron defectuosamente
  - d) No, porque aunque se aplicaron de manera adecuada, esto se hizo esporádicamente
  - e) No, porque las técnicas se aplicaron defectuosamente y de manera esporádica

BIBLIOGRAFIA

1. Everitt B.S., The Analysis of Contingency Tables. Chapman and Hall, London 1977.
2. Fienberg Stephen E., The Analysis of Cross-Classified Categorical Data MIT Press. Massachusetts 1979.
3. Guzmán Arellano L.M., Gutiérrez M.E. "Investigación sobre la Situación Académica de los Estudiantes de Segundo Semestre y las Técnicas Didácticas Utilizadas Para su Preparación". Trabajo presentado para obtener el diploma de Especialización en Docencia. Facultad de Química -- UNAM. 1974
4. Guzmán L.M., Gutiérrez M.E., Rosales G. "Investigación Sobre la Situación Académica de los Alumnos del Area Químico Biológica de la ENEP -- "Zaragoza" UNAM. la parte presentada en el XIV Congreso de Química - Pura y Aplicada (Sociedad Química de México); octubre 1979, Monterrey Nuevo León. 2a parte presentada en el I Congreso de Educación en el - Area de la Química (Sociedad Química de México); marzo 1981. Cuernavaca, Morelos.
5. Guzmán L.M. Barba N., Gutiérrez M.E., Hernández G., Rosales G., "Estudio Comparativo Sobre el Aprovechamiento Escolar de los Alumnos de Primer Ingreso de la Facultad de Química y del Area Químico-Biológica de la ENEP "Zaragoza" UNAM. Presentado en el II Congreso de Educación en el Area de la Química (Sociedad Química de México); agosto 1982. Guanajuato, Guanajuato.
6. Lineamientos Generales de la Investigación. ENEP "Zaragoza" UNAM 1983
7. Ruiz Velasco Acosta, Silvia, Modelos Lineales Generalizados y su Relación con GLIM. Tesis Profesional. Facultad de Ciencias UNAM. México 1982.