



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

**FACULTAD DE CIENCIAS**

**SALIDA DEL SISTEMA EDUCATIVO: NIVEL  
ALCANZADO Y CAUSAS DE ABANDONO ESCOLAR  
ENTRE LOS JÓVENES DE DISTINTOS SECTORES  
SOCIOECONÓMICOS DE MÉXICO**

**REPORTE DEL SEMINARIO DE  
TITULACIÓN**

**QUE PARA OBTENER EL TÍTULO DE:**

**ACTUARIA**

**P R E S E N T A:**

**ROCÍO ZARAHÍ MARTÍNEZ BARREDA**



**TUTORA:  
DRA. MARTA MIER Y TERÁN ROCHA**

**2009**



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.

Martínez Barreda Rocío Zarahi  
56 87 26 22  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Actuaría  
30172109-7  
[ch\\_io@hotmail.com](mailto:ch_io@hotmail.com)

Dra. Marta Mier y Terán Rocha

Dra. María Edith Pacheco Gómez Muñoz  
Mtro. Ricardo Aparicio Jiménez  
Mtra. Rosa María Camarena Córdova  
M. en D. Alejandro Mina Valdés

Salida del sistema educativo: nivel alcanzado  
y causas de abandono escolar entre los  
jóvenes de distintos sectores socioeconómicos de México  
82 p.  
2009

## AGRADECIMIENTOS

Quiero agradecer y dedicar el presente trabajo primeramente a Dios por siempre haber estado conmigo y darme la paciencia, la fuerza y la sabiduría para la realización de todos los proyectos en mi vida.

A toda la gente que no creyó en mí diciendo cosas que me desanimaron haciéndome dudar de mi misma. Gracias especialmente a ellos porque sin darse cuenta al lastimarme en donde ni ellos se imaginan me dieron la fuerza para seguir adelante y demostrarle al mundo y a mi misma lo que soy y puedo lograr; y lo que no debo ser.

A mi familia. A *Rocío Barreda Parra*, mi madre, porque siempre ha estado conmigo. Por ser mi mamá antes que todo, mi amiga y un ejemplo a seguir de fortaleza y tenacidad. Porque tú sabes mejor que nadie lo que este trabajo significa. Gracias por ser una gran mujer y enseñarme a serlo. Gracias a ti enfrente la vida todos los días con fuerza aunque pueda estar temblando por dentro. A mi hermano, *David Martínez Barreda* por ser mi apoyo incondicional, por emocionarse conmigo de todos los logros obtenidos, por querer ser un ejemplo a seguir para él y enseñarle a no rendirse jamás cuando de perseguir los sueños se trata. A mi padre, *David Martínez Eguiluz*, por todos sus consejos y apoyo durante mi vida como estudiante y por los que me ha dado ahora en mi vida laboral. Por ser el hombre de mi vida y el que me ha hecho la mujer que soy ahora. Y a toda mi familia materna, especialmente a Carlos "Botito" Barreda por ser un ejemplo a seguir.

A Rosalía Suárez Ávalos porque no sólo ha sido mi terapeuta si no una mujer ejemplar, un apoyo incondicional y me ha ayudado a descubrirme a mí misma y lo que puedo llegar a ser si me lo propongo.

A todos mis amigos porque cada uno de ellos me enseñó algo y me escuchó cuando lo necesité. ¡Gracias por estar conmigo siempre! Son como unos hermanos para mí y han sido un impulso valioso en mi vida.

Al Act. Erick Mier Moreno y su familia por ser siempre mi segunda familia y por sus palabras de aliento. Por todo el conocimiento que Erick me transmitió tanto académico como a nivel personal, por siempre animarme a ser mejor, porque siempre que sentía que ya no podía seguir con la carrera él me daba ánimos. Gracias a ti soy la acturia que soy y una mejor persona. Me enseñaste más de lo que tú crees.

A los profesores de la Facultad de Ciencias por todos sus valiosos conocimientos, por inspirarme amor a la carrera, por impulsarme a seguir y en muchas otras ocasiones por obligarme a demostrar que sí podía.

A la Dra. Marta Mier por ser la guía de este trabajo y a mis compañeras de seminario porque las cuatro hicimos un buen equipo, siempre dándonos mucho ánimo. A los sinodales la Dra. Edith Pacheco, el Mtro. Ricardo Aparicio, la Mtra. Rosa Ma. Camarena y al Mtro. Alejandro Mina por sus valiosos comentarios para mejorar este trabajo.

A mi jefe en Banco de México, Luis García Ibarra que me apoyó todo el tiempo durante la realización de este trabajo y creyó en mí desde el principio. Gracias, además, a todo el equipo de trabajo que me apoyó de la Subgerencia de Información Financiera.

Y finalmente, pero muy importante, a ti... ... por estar conmigo a donde quiera que voy y por hacerme sentir la vida en mi piel. Gracias, además, porque me ayudaste en la realización de este proyecto. Es por ti que he dado los pasos que he dado en este último año.

No lo olvides, ¡tú corres por mis venas!

## CONTENIDO

INTRODUCCIÓN	1
CAPÍTULO I.- ELABORACIÓN DE LOS SECTORES SOCIO ECONÓMICOS	5
I.I COEFICIENTE DE CORRELACIÓN DE PEARSON	7
I.II EL ANÁLISIS FACTORIAL APLICADO A DATOS DEL CENSO DEL 2000	9
I.II.I MATRIZ DE CORRELACIONES ENTRE LAS VARIABLES	11
I.II.II EXTRACCIÓN DE LOS FACTORES	16
I.II.III ROTACIÓN DE FACTORES	23
CAPÍTULO II.- ANÁLISIS DESCRIPTIVO DE LA CONDICIÓN EDUCATIVA DE LOS JÓVENES DE 12 A 24 AÑOS	29
II.I METODOLOGÍA	29
II.II ASISTENCIA	31
II.III SITUACIÓN EDUCATIVA	33
II.IV SITUACIÓN DE ACTIVIDAD	41
CAPÍTULO III.- MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL PARA EXPLICAR LA CONDICIÓN EDUCATIVA DE LOS JÓVENES DE 12 A 24 AÑOS	46
III.I ECUACIÓN BÁSICA DEL MODELO	46
III.II PRUEBAS DE HIPÓTESIS Y DE AJUSTE DEL MODELO	49
III.III ANÁLISIS DE LOS RESULTADOS OBTENIDOS	53
III.III.I ECUACIÓN DEL MODELO	55
III.III.II RESULTADOS DEL MODELO APLICADO	60
CONCLUSIONES	68
ANEXO I	70
ANEXO II	76
BIBLIOGRAFÍA	81

## INTRODUCCIÓN

El sistema educativo mexicano ha tenido grandes cambios y transiciones a lo largo del tiempo. El desarrollo industrial, el movimiento del 68, la migración rural-urbana, la urbanización, el crecimiento demográfico, los diferentes regímenes políticos del poder ejecutivo y legislativo, crisis de todos tipos, y más recientemente la introducción de las computadoras e Internet en la vida y el trabajo diario; así como el incremento de las tecnologías en general son algunos de los acontecimientos que, como consecuencia, han traído cambios al sistema educativo, ya sea para mejorarlo o para indicar que todavía queda un largo camino por recorrer.

Si bien es cierto que la cobertura de la educación básica se ha ampliado, ésta ha sido un proceso desigual y lento a lo largo del territorio nacional al existir actualmente sectores de la población que todavía no tienen acceso a una educación de calidad. La mayoría de los estudiosos de la educación coinciden en que la insatisfactoria calidad del sistema educativo; así como las oportunidades desiguales de ingreso, permanencia y aprendizaje son los problemas más importantes que el Sistema Educativo debe enfrentar (Schmelkes, 1998).

Hoy en día existen varios programas educativos que buscan atender a aquellos sectores de la población más desfavorecidos. Por ejemplo, la población indígena, y mantener a su población infantil y joven en la escuela (Cervini, 2004). En respuesta a la creciente oferta educativa; así como a los mayores requisitos educativos en el mercado de trabajo, las familias han ido valorando cada vez más la educación formal de los hijos y han propiciado que ellos permanezcan en la escuela hasta terminar sus estudios (SITEAL, 2009).

La calidad deficiente y desigual de la formación son dos conceptos que están íntimamente relacionados con el problema de deserción educativa. Lo anterior debido a que la calidad está íntimamente relacionada con la inequidad y la desigualdad que se vive dentro de la sociedad mexicana. La desigualdad se refiere a las diferencias en la noción social, económica o de cualquier otra índole que hacen a una persona diferente de otra (Ornelas, 1998). La inequidad se manifiesta en diferencias en las oportunidades de ingreso, permanencia, egreso y aprendizaje dentro del sistema educativo (Schmelkes, 1998). Es decir, la inequidad es una desigualdad que no debería ser. Esto se refleja fuertemente en las escuelas al estar altamente estratificadas según el nivel socioeconómico de los alumnos, por tanto la

calidad educativa que ofrecen depende del nivel socioeconómico de los alumnos; y como consecuencia, los resultados académicos de los alumnos son un reflejo del colegio o institución educativa en donde estudian y no tanto de su capacidad intelectual. Todo lo anterior nos lleva a afirmar que la enorme desigualdad social se reproduce en la escuela. (Reimers, 2000)

Las condiciones de vida de los adolescentes afectan directamente la deserción del sistema educativo de la población joven y es un concepto que está ligado fuertemente al concepto de inequidad educativa; ya que, como se dijo anteriormente, el acceso a la educación de calidad está acotado por la capacidad económica de las familias. Por lo tanto, la liquidez y solvencia económicas de las personas encargadas de la educación del estudiante son fundamentales para su adecuado desarrollo educativo. Los hogares con mayores recursos podrán solventar los gastos educativos de los hijos, pero los hogares pobres con frecuencia no pueden afrontar los gastos por lo que requieren del ingreso del trabajo de los jóvenes para subsistir. Esto sucede por procesos externos, independientes del sistema educativo, relativos al funcionamiento de los mercados de trabajo (Calero, 2004).

No se debe perder de vista que dentro de la sociedad son aceptables ciertas diferencias, siempre y cuando se garantice la igualdad en ciertas condiciones básicas que den oportunidades a todos los ciudadanos; es decir, se busca que la desigualdad que exista sea producto del esfuerzo de las personas y no de la falta de oportunidades que se les brinden (Reimers, 2000).

Por lo anterior, para hacer frente al problema de la inequidad educativa es necesario tomar en cuenta las características geográficas, culturales, las necesidades y las condiciones socioeconómicas de los grupos que conforman la población para que todos los mexicanos tengan oportunidades iguales para ingresar y permanecer en el sistema educativo. El gobierno mexicano ha destinado recursos adicionales para la educación de los hijos de los pobres. El objetivo de las políticas educativas ha sido mejorar la cobertura del sistema educativo, pero no de igualar las oportunidades ni de mejorar la calidad de la educación de los sectores socioeconómicos desfavorecidos del país. Por tanto, la igualdad de oportunidades educativas queda aún pendiente y debe ser objetivo prioritario de las políticas educativas (Reimers, 2000).

Las escuelas pueden combatir las inequidades formando comunidades en donde se aprenda a valorar a quienes son diferentes y se desarrollen experiencias positivas de cooperación entre grupos diversos de su población estudiantil. Dichas acciones deben

estar contempladas dentro del objetivo de las políticas antes mencionadas y destinar recursos a que se cumplan (Reimers, 2000).

En términos generales, la diversidad de la población y del territorio nacional es un detalle importante al momento de querer resolver el problema de la inequidad; ya que, tomando en cuenta las desigualdades inherentes a las personas que habitan en este país se puede resolver de mejor manera las desigualdades educativas que no deberían de existir para la población de México (Reimers, 2000).

Uno de los fines de la educación es el instrumental, que posibilita a los individuos que la reciben el tener un marco más amplio de oportunidades para desarrollarse socialmente, puesto que las personas más educadas tienen mayores oportunidades de participar en el sector laboral y así obtener mejores empleos e ingresos económicos más altos. Por lo tanto, lograr la igualdad de oportunidades educativas erradicando las inequidades mencionadas anteriormente; así como, mejorar las oportunidades de ingreso al mercado de trabajo deben constituir un eje fundamental de la política pública de México (Márquez, 2004).

Otro de los problemas que afectan la deserción del nivel educativo es el que tiene que ver con la naturaleza misma del ser adolescente; ya que, la adolescencia es una etapa de la vida caracterizada por la búsqueda de identidad, esto aunado a problemas personales, de familia o por desorientación por parte de los padres de familia podría desencadenar que los jóvenes no vayan a la escuela (SITEAL, 2009).

Conocer los motivos de abandono entre los niños y los jóvenes del país, así como el nivel que alcanzaron en el sistema educativo permite orientar las acciones de la política pública con el objeto de hacerlas más eficientes. Además, ante las grandes desigualdades sociales y educativas que privan en el país, es de sumo interés ubicar a los jóvenes en los sectores socioeconómicos a los que pertenecen sus hogares.

El objetivo de este trabajo es definir estratos socioeconómicos de los hogares de los jóvenes de 12 a 24 años de edad e investigar la relación entre el sector al que pertenecen y el abandono del sistema educativo; así como el nivel que alcanzaron dentro de él. La fuente de datos es una submuestra del Censo Nacional de Población y Vivienda de 2000.<sup>1</sup>

---

<sup>1</sup> La definición de los datos utilizados se explicara en los capítulos posteriores.



La hipótesis de trabajo que guía este estudio es que, de acuerdo al sector socio económico al que pertenecen los jóvenes, el nivel de estudios alcanzado y las causas de abandono de la escuela cambian, ya que estos ambientes propician o no la permanencia en el sistema educativo y el logro de niveles educativos más altos.

## CAPITULO I

### ELABORACIÓN DE LOS SECTORES SOCIOECONÓMICOS – ANÁLISIS FACTORIAL<sup>2</sup>

La elaboración de los estratos socio económicos es parte fundamental de este proyecto para poder exponer de manera adecuada el análisis descriptivo de los datos de los hogares donde se desenvuelven los niños y jóvenes de 12 a 24 años respecto al objetivo del trabajo.

Para poder elaborarlos se hizo uso de una herramienta estadística de análisis multivariado llamado Análisis Factorial.

El *Análisis Factorial* es una técnica estadística de reducción de datos usada para explicar la varianza entre las variables observadas en términos de un número menor de variables no observadas llamadas *factores*.

Es una técnica que busca encontrar grupos homogéneos de variables a partir de un conjunto numeroso de variables. Esos grupos homogéneos se forman con las variables que correlacionan mucho entre si y procurando, inicialmente, que unos grupos sean independientes de otros.

En el modelo matemático del análisis factorial cada característica individual o variables observadas se expresa como una combinación lineal de factores F:

$$X_{ij} = F_{1i}a_{i1} + F_{2i}a_{i1} + \dots + F_{ki}a_{ik} + V_i,$$

donde:

$X_{ij}$  = la puntuación del individuo i en la variable j.

$F_{ki}$  = los coeficientes factoriales.

$a_{ik}$  = las puntuaciones factoriales.

$V_i$  = el factor único de cada variable<sup>3</sup>.

Su propósito último consiste en buscar el número de dimensiones capaces de explicar el máximo de información contenida en los datos. Para lograrlo y que el análisis

---

<sup>2</sup> Este apartado se basa fundamentalmente en: [SPSS 10 Guía para el análisis de datos](#) y del documento de Cuesta, Marcelino, [Introducción al Análisis Factorial](#).

<sup>3</sup> Los factores únicos explican, por sí mismos, una parte de la variabilidad de cada variable, sin relación con los demás ni con los factores comunes.

factorial tenga sentido, deben cumplirse dos condiciones básicas: parsimonia e interpretabilidad. Esto quiere decir que se busca explicar la mayor parte de la varianza con la menor cantidad de factores posibles que sean interpretables.

Los pasos que se suelen seguir en el análisis factorial son:

1. *Calcular la matriz de correlaciones entre todas las variables y examinarla para decidir qué tan propicias son las variables a un Análisis de Factores.* Es decir, hacer una serie de pruebas a esta matriz para saber si las variables que se van a involucrar en el análisis están suficientemente correlacionadas. Las pruebas que se aplican a la matriz de correlaciones en este trabajo se explican más adelante.
2. *Extracción de los factores necesarios que representen las características de los datos incluidos en el análisis factorial.* Es durante este proceso en donde se cumple la condición básica de parsimonia del Análisis Factorial mencionada anteriormente.
3. *Rotación de los factores con objeto de facilitar su interpretación.* Es posible que cuando se extraen los factores estos logren explicar las características de los datos que se estudian; sin embargo, algunas veces los factores aun son difíciles de interpretar por lo que se recurre a la rotación de los ejes coordenados que representan a los factores para intentar que se ajusten mejor a las variables de los datos y de esta manera se puedan interpretar y nombrar fácilmente. En este caso, fue necesaria una rotación la cual se explica más adelante.

Antes de exponer lo que se elaboró en este proyecto, es de interés definir lo que es el coeficiente de correlación ya que es a partir de este coeficiente que se elabora el análisis de factores al iniciar éste con un examen de la matriz de correlaciones entre las variables tomadas en cuenta para elaborar los sectores socio económicos.

## I.1 COEFICIENTE DE CORRELACIÓN DE PEARSON

El *coeficiente de correlación de Pearson* (Mood, 1974) es un índice estadístico que mide la relación lineal entre dos variables cuantitativas.

Cuando se habla de correlación implícitamente se habla de una relación o dependencia lineal. Es decir, que la dependencia es de la siguiente forma:

$X_i$  está correlacionada con  $X_j$  si  $X_i = \alpha_0 + \alpha_j X_j$  con  $j=1, \dots, n$ ,  $i \neq j$ , para alguna  $\alpha_j \neq 0$  donde  $X_i$  es la variable dependiente de  $X_1, X_2, \dots, X_n$ .

Si  $X_i$  está correlacionada con  $X_j$  entonces existe una combinación lineal finita no trivial del vector cero; es decir, alguna  $\alpha_j$  con  $j=1, \dots, n$  e  $i \neq j$  va a ser diferentes de cero. Sin embargo, si  $X_i$  no está correlacionada con  $X_j$  no garantiza que  $X_i$  no dependa de  $X_j$ , lo único que implica es que  $X_i$  no depende linealmente de  $X_j$ , pudiendo existir relaciones no lineales entre las variables.

La medida de dependencia lineal se hace considerando el *coeficiente de correlación* denominado  $\rho$ . Este número es una medida de la dependencia lineal de una variable con respecto a otra.

El cálculo del coeficiente de correlación lineal se realiza dividiendo la covarianza entre el producto de las desviaciones estándar de ambas variables:

$$\rho_{ij} = \frac{\sigma_{X_i X_j}}{\sigma_{X_i} \sigma_{X_j}},$$

donde:

$\sigma_{X_i X_j}$  = la covarianza entre la variable  $X_i$  y la variable  $X_j$

$\sigma_{X_i}$  = la desviación estándar de la variable  $X_i$

El valor del índice de correlación (Salas, 2007) varía en el intervalo  $[-1, +1]$ :

- Si  $\rho_{ij} = 0$ , no existe relación lineal. Pero esto no necesariamente implica una independencia total entre las dos variables como se mencionó anteriormente.
- Si  $\rho_{ij} = 1$ , existe una correlación positiva perfecta<sup>4</sup>. El índice indica una dependencia total entre las dos variables denominada relación directa: cuando una de ellas aumenta, la otra también lo hace en la misma proporción.
- Si  $0 < \rho_{ij} < 1$ , existe una correlación positiva.
- Si  $\rho_{ij} = -1$ , existe una correlación negativa perfecta. El índice indica una dependencia total entre las dos variables llamada relación inversa: cuando una de ellas aumenta, la otra disminuye en la misma proporción.
- Si  $-1 < \rho_{ij} < 0$ , existe una correlación negativa.

---

<sup>4</sup> En la realidad es difícil encontrar una correlación perfecta, ya sea positiva o negativa, entre dos variables distintas. Este apartado es sólo para guiar al lector acerca de los posibles valores que pueda tener el coeficiente de correlación.

## I.II EL ANÁLISIS FACTORIAL APLICADO A DATOS DEL CENSO DEL 2000

Las gráficas y tablas de este trabajo fueron el resultado del análisis de los datos obtenidos del Censo Nacional de Población y Vivienda del 2000 del Instituto Nacional de Estadística y Geografía (INEGI) accesibles en la página: [www.inegi.org.mx](http://www.inegi.org.mx)

Estos datos son una submuestra del 10% del total de encuestas aplicadas a la población a la cual se le aplicó el cuestionario ampliado. Éste último incluye preguntas acerca de la vivienda las cuales son de gran utilidad para la elaboración de los factores al mostrar las condiciones en las que viven las personas. Es gracias a este cuestionario ampliado que se puede medir la solvencia económica que tienen las familias mexicanas y con esto tratar de saber que tanta capacidad monetaria se puede invertir en la educación de los jóvenes.

Es importante mencionar que se utilizó una extracción de esta submuestra del 10% que toma en cuenta solo las edades de 12 a 24 años y solo los descendientes y parientes con edades menores a las de jefe<sup>5</sup>. Lo anterior porque estas edades son las que comprenden generalmente a los estudiantes y, por lo tanto, muestran tener mayor dependencia económica del hogar.

Además, para la elaboración de las tablas y resultados se escogió a un joven por cada hogar para poder describir las características de la vivienda, así como para la elaboración de los factores. De esta manera, se elimina la posibilidad de elegir a dos o más jóvenes dentro de la muestra de edad de 12 a 24 años que cohabiten en el mismo hogar.

Para obtener los resultados de todos los capítulos se utilizó la muestra ponderada. Para ello primeramente se obtuvo  $N = 10,020,181$  que es el número de casos totales expandidos de la muestra de los jóvenes descendientes y dependientes (del jefe de familia) de 12 a 24 años. Posteriormente se obtuvo  $n = 1,021,616$  que es el número de casos totales sin expandir de la muestra de los jóvenes (descendientes y dependientes) del jefe de familia de 12 a 24 años. Ahora bien el ponderador se calculó de la siguiente manera:

$$Pondera = F_e(n/N),$$

---

<sup>5</sup> La sintaxis de la variable utilizada para la elaboración de la muestra que solo considera a los descendientes y parientes menores que el jefe se encuentra en el Anexo II.

donde:

$F_e$  = el factor de expansión

El primer paso que se siguió para elaborar los estratos fue decidir qué variables involucrar dentro del Análisis Factorial que estuvieran correlacionadas y explicaran las condiciones de vida en el hogar en dónde viven los jóvenes.

Para ello, como se expuso anteriormente, se hizo uso del cuestionario ampliado que aplica el INEGI a una parte de la población; en el cual se indaga no solamente acerca de las características de las personas si no que también de las características de la vivienda en la que habitan las personas encuestadas. Cabe mencionar que no todas las variables que maneja el cuestionario ampliado fueron integradas al Análisis Factorial. Se decidió excluir a aquellas variables (como material de construcción de la vivienda, antigüedad de la vivienda, material del techo, de las paredes, etc.) que presentaban ambigüedad y poca claridad para reflejar las condiciones de altos o bajos recursos económicos que tengan los habitantes de una vivienda.

Las variables que intervienen en el análisis y que por inferencia reflejan la condición económica de los habitantes son<sup>6</sup>:

1. **Cocina.** Esta variable indica si el cuarto donde cocinan es independiente al cuarto donde duermen. Esta variable adopta el valor de 3 para los que no duermen en la cocina y 4 para los que sí duermen.
2. **Hacinamiento.** Esta variable se obtiene al dividir el número de personas que habitan en la vivienda entre el número total de cuartos de la vivienda incluida la cocina.
3. **Disponibilidad de agua.** Indica la condición en la que el agua llega a la vivienda. Tiene el valor de 1 si la vivienda cuenta con agua entubada dentro de la vivienda; 2 si cuenta con agua entubada fuera de la vivienda, pero dentro del terreno; 3 si cuenta con agua entubada de llave pública; 4 si cuenta con agua entubada pero que los habitantes acarrearán de otra vivienda; 5 si obtiene el agua de una pipa; y finalmente 6 si obtiene el agua de algún pozo, río, lago, arroyo u otra.

---

<sup>6</sup> Las frecuencias de las variables utilizadas se encuentran en el anexo I

4. **Electricidad.** Muestra si la vivienda cuenta con energía eléctrica. Esta variable adopta el valor 1 si la vivienda cuenta con electricidad y 2 si no cuenta con ella.
5. **Combustible.** Esta variable indica qué combustible se utiliza para cocinar. Esta variable adopta el valor de 1 si es con gas, 2 si es con electricidad, 3 si es con petróleo, 4 si es con carbón o 5 si es con leña.
6. **Eliminación de basura.** Muestra cómo se elimina la basura de la vivienda. Esta variable tiene el valor de 1 si la basura la recoge un camión de basura; 2 si la depositan en un contenedor o depósito; 3 si la queman o entierran; y finalmente, 4 si la tiran a la intemperie.
7. **Drenaje.** La variable explica de qué manera la vivienda se deshace del agua. Adopta el valor de 1 si el drenaje o desagüe de aguas sucias de la vivienda va a dar a la red pública; 2 si va a dar a una fosa séptica; 3 si va hacia una tubería que va a dar a la intemperie; o 4 si la vivienda cuenta con drenaje.
8. **Bienes en la vivienda**<sup>7</sup>. Esta parte está integrada por 11 variables que explican la existencia de los siguientes bienes en la vivienda. Para todas estas variables las respuestas son únicamente sí y no. Los bienes tomados en cuenta son: Radio o radiograbadora, televisión, videocasetera, licuadora, refrigerador, lavadora, teléfono, calentador de agua (boiler), automóvil propio y computadora.

### I.II.I MATRIZ DE CORRELACIONES ENTRE LAS VARIABLES

Ya definido lo que es el coeficiente de correlación Pearson y establecidas las variables que se van a analizar dentro de este proceso, el primer paso en el Análisis Factorial es calcular la *matriz de correlaciones* entre todas las variables que entran en el análisis.

Esta matriz de correlaciones es del tipo:

---

<sup>7</sup> La variable "radio" no se incluye en el Análisis Factorial; ya que, los factores explicaban débilmente su varianza y por lo tanto no resultaba relevante incluirlo en el análisis.



	$x_1$	$x_2$	...	$x_j$
$x_1$	1	$\rho_{1,2}$	...	$\rho_{1,j}$
$x_2$	$\rho_{2,1}$	1	...	$\rho_{2,j}$
.	.	.	...	.
.	.	.	...	.
.	.	.	...	.
$x_j$	$\rho_{j,1}$	$\rho_{j,2}$	...	1

donde:

$\rho_{kj}$  = coeficiente de correlación de Pearson entre la variable k y la variable j con  $k \neq j$

Una vez que se dispone de esta matriz se examina para comprobar si sus características son adecuadas para realizar un Análisis Factorial. Uno de los requisitos que deben cumplirse para que el análisis factorial tenga sentido es que las variables estén correlacionadas.

En este proyecto, las variables tienen la siguiente matriz de correlaciones:

**Cuadro.I.1 Matriz de correlaciones**

Matriz de correlaciones																
	Cocina	Hacinamiento	Disponibilidad de agua	Electricidad	Combustible	Eliminación de basura	Drenaje	Televisión	Videocasetera	Licuada	Refrigerador	Lavadora	Teléfono	Boiler	Auto propio	Computadora
Cocina	1.000	.474	.139	.090	.091	.091	.145	.129	.160	.153	.198	.168	.180	.202	.151	.096
Hacinamiento		1.000	.304	.194	.333	.286	.345	.291	.319	.334	.408	.342	.367	.389	.290	.229
Disponibilidad de agua			1.000	.352	.492	.506	.544	.375	.313	.410	.424	.389	.355	.395	.239	.185
Electricidad				1.000	.342	.288	.311	.429	.164	.408	.311	.227	.161	.171	.112	.072
Combustible					1.000	.618	.611	.532	.352	.574	.557	.467	.366	.381	.265	.166
Eliminación de basura						1.000	.672	.410	.341	.457	.467	.425	.410	.404	.224	.195
Drenaje							1.000	.426	.369	.486	.500	.444	.431	.461	.255	.213
Televisión								1.000	.291	.533	.467	.364	.270	.284	.218	.124
Videocasetera									1.000	.341	.431	.450	.473	.467	.393	.344
Licuada										1.000	.529	.422	.331	.352	.249	.156
Refrigerador											1.000	.552	.448	.437	.350	.216
Lavadora												1.000	.462	.459	.385	.268
Teléfono													1.000	.543	.428	.384
Boiler														1.000	.424	.355
Auto propio															1.000	.371
Computadora																1.000

Fuente: Elaboración propia en SPSS a partir de los datos del censo de población y vivienda del 2000 del INEGI

Esta matriz de correlaciones (ver Cuadro.I.1) muestra que todas las variables están suficientemente correlacionadas lo cual indica que el análisis factorial puede ser una herramienta adecuada y eficiente para lograr elaborar un índice de estratos socioeconómicos y así, conocer las condiciones en las que viven los jóvenes que pueden o no propiciar su educación.

El hacinamiento y la cocina son variables que tienen correlación alta entre ellas, esto podría indicar que uno de los factores puede explicar eficientemente el comportamiento de éstas.

Otra característica de esta matriz, es que el acceso a los servicios públicos como el agua, el drenaje o la eliminación de basura e incluso el combustible y los algunos enseres, licuadora y refrigerador, están correlacionados y presentan las correlaciones más grandes. Lo anterior puede indicar que el análisis factorial puede generar otro factor que explique el comportamiento de estas variables.

Por lo tanto, al observar la matriz de correlaciones (Cuadro.I.1) Se puede inferir que cuando se apliquen las pruebas correspondientes a esta matriz de correlaciones para evaluar qué tan pertinente es utilizar un análisis factorial a estas variables, éstas van a dar resultados positivos.

Pueden utilizarse diferentes métodos para comprobar el grado de asociación entre las variables. Algunos de ellos son<sup>8</sup>:

- **El determinante de la matriz de correlaciones**

Un determinante muy bajo indicará altas intercorrelaciones entre las variables, pero no debe de ser cero, pues esto indicaría que algunas de las variables son linealmente dependientes y no se podrán realizar ciertos cálculos necesarios en el Análisis Factorial.

La matriz de correlaciones mostrada anteriormente arrojó un determinante igual a 0.002 el cual es suficientemente bajo para poder aplicar el Análisis Factorial.

---

<sup>8</sup> Este apartado se basa fundamentalmente en: SPSS 10 Guía para el análisis de datos y del documento de Cuesta, Marcelino, Introducción al Análisis Factorial.

- **Test de esfericidad de Barlett**

Comprueba que la matriz de correlaciones se ajuste a la matriz identidad (I), es decir, ausencia de correlación significativa entre las variables. Esto significa que la nube de puntos se ajustará a una esfera perfecta, expresando así la hipótesis nula por:

$$H_0: R = I$$

Es decir, que el determinante de la matriz de correlaciones es 1.

$$H_0: |R| = 1$$

La fórmula correspondiente se muestra en la siguiente expresión:

$$\chi^2 = - \left[ n - 1 - \frac{1}{6} * (2 * v + 5) \right] * \ln |R|,$$

donde:

n = tamaño muestral.

v = número de variables.

R = matriz de correlaciones de las variables.

Si se acepta la hipótesis nula ( $p > 0.05$ ) significa que las variables no están intercorrelacionadas y por tanto no tiene mucho sentido llevar a cabo un Análisis Factorial.

- **Índice KMO de Kaiser-Meyer-Olkin**

La fórmula del índice KMO es la siguiente:

$$KMO = \frac{\sum_{i \neq j} \sum_{j \neq i} r_{ij}^2}{\sum_{i \neq j} \sum_{j \neq i} r_{ij}^2 + \sum_{i \neq j} \sum_{j \neq i} a_{ij}^2},$$

donde:

$r_{ij}$  = correlación simple.

$a_{ij}$  = correlación parcial.

Valores bajos del índice KMO indican que no es recomendable la utilización del Análisis Factorial. La siguiente es la escala que se utiliza para saber que tan aptas son las variables para un análisis factorial:

1  $\geq$  KMO  $\geq$  0.9 muy bueno  
0.9  $\geq$  KMO  $\geq$  0.8 meritorio  
0.8  $\geq$  KMO  $\geq$  0.7 mediano  
0.7  $\geq$  KMO  $\geq$  0.6 mediocre  
0.6  $\geq$  KMO  $\geq$  0.5 bajo  
KMO  $\leq$  0.5 inaceptable

En estas últimas dos pruebas la matriz de correlaciones dio los siguientes resultados:

**Cuadro.I.2 Resultados de las pruebas de KMO y Bartlett**

KMO y Prueba de Bartlett		
Medida de adecuación de la muestra Índice de Kaiser-Meyer-Olkin		.932
Prueba de esfericidad de Bartlett	Aprox. Chi-Square	5,693,032.146
	Grados de libertad	120
	Sig.	.000
Fuente: Elaboración propia en SPSS a partir de los datos del censo de población y vivienda del 2000 del INEGI		

La tabla (ver Cuadro.I.2) muestra lo altamente recomendable que es la aplicación del análisis factorial al alcanzar un índice KMO con valor de 0.932; así mismo, la prueba de esfericidad de Bartlett arroja un p-value  $< 0.05$  lo que señala que las variables están altamente correlacionadas. Estos resultados, en conjunto con la prueba del determinante de la matriz de correlaciones, respaldan la adecuada decisión de llevar a cabo el Análisis Factorial de las variables involucradas como se había pensado previamente.

## I.II.II EXTRACCIÓN DE LOS FACTORES<sup>9</sup>

La *extracción de factores* es la segunda etapa dentro del Análisis Factorial. Este paso permite decidir el modelo factorial a utilizar para estimar las saturaciones de las variables en los factores resultantes, en qué matriz de datos basar el análisis y los factores que desean extraerse.

El modelo factorial utilizado permite transformar un conjunto de variables, intercorrelacionadas, en otros conjuntos de variables no correlacionadas denominados factores. Como se dijo anteriormente, el objetivo es explicar la mayor cantidad de varianza de las variables originales a través del menor número de factores o componentes.

En este proyecto, la base del análisis factorial es la matriz de correlaciones a partir de la cual se obtendrán las ecuaciones lineales que representan a las variables originales en relación con los factores resultantes.

Inicialmente, se muestra una matriz factorial llamada *matriz de componentes* que muestra la solución factorial propiamente dicha. Esta matriz contiene los pesos factoriales que existen entre las variables y los factores obtenidos. Los factores principales que muestre la matriz de componentes son tales que la variabilidad explicada por cada uno de ellos sea la máxima que puedan representar.

### PESOS FACTORIALES

Los *pesos factoriales* son los índices de correlación entre las variables y los factores. Aunque estrictamente, sólo son correlaciones cuando los factores no están relacionados entre sí, es decir son ortogonales. Sin embargo, a menudo sucede que los factores sí están correlacionados, por lo tanto, se busca que esto suceda en la menor medida posible.

Como su nombre lo indica, los pesos factoriales exponen “el peso de cada variable en cada factor”. Lo ideal es que cada variable obtenga alto en un factor y bajo en los demás. Las cargas factoriales pueden tener como valor máximo 1, es decir que ese

---

<sup>9</sup> Este apartado se basa fundamentalmente de: SPSS 10 Guía para el análisis de datos, del documento de Cuesta, Marcelino, Introducción al análisis Factorial, y en el documento Modelos Sociodemográficos: Atlas Social de la ciudad de Alicante.

factor explique por completo el comportamiento de una variable. Los *factores bipolares* son aquellos factores en los que unas variables cargan positivamente y otras tienen carga negativa.

La matriz de componentes obtenida en este proyecto es la siguiente:

**Cuadro.I.3 Matriz de Componentes o matriz factorial**

<b>Matriz factorial<sup>a</sup></b>			
	Componentes		
	1	2	3
Cocina	.294	.277	.801
Hacinamiento	.571	.235	.554
Disponibilidad de agua	.659	-.206	-.043
Electricidad	.453	-.434	.153
Combustible	.749	-.328	-.084
Eliminación de basura	.711	-.254	-.175
Drenaje	.751	-.210	-.102
Televisión	.627	-.366	.102
Videocasetera	.623	.316	-.141
Licuada	.693	-.303	.079
Refrigerador	.752	-.057	.044
Lavadora	.699	.115	-.086
Teléfono	.667	.359	-.148
Boiler	.679	.324	-.095
Auto propio	.520	.459	-.137
Computadora	.407	.511	-.241
Método de extracción: Componentes principales.			
a. 3 componentes extraídos.			
Fuente: Elaboración propia en SPSS a partir de los datos del censo de población y vivienda del 2000 del INEGI			

Observando los pesos factoriales de las variables (ver Cuadro.I.3) se puede observar que se extraen tres factores. Para que la interpretación factorial sea más fácil se debe intentar dar nombre a los componentes. Sin embargo, en este caso, la mayoría de las variables caen en el primer factor, por lo que no hay una distinción contundente que facilite el nombramiento del primer y segundo componente al tener el primero variables de todos los tipos, tanto bienes como servicios. Mientras tanto, se puede observar que las características propias de la vivienda (hacinamiento y cocina) tienen pesos altos en el tercer factor, aunque hacinamiento tenga mayor peso en el primer factor probablemente con la rotación de factores el peso más alto pase al tercer factor. Por lo que el tercer factor se podría nombrar con respecto a las características de la vivienda.

De esta tabla podemos concluir que la interpretabilidad no es fácil aún. Por lo que se tendrá que hacer uso de la rotación factorial para mejorar este punto. Más adelante se expone en qué consiste este proceso.

Junto con el análisis de la matriz factorial anterior se debe de hacer el análisis de la matriz de comunalidades y de la tabla de porcentaje de varianza explicada (tabla de valores propios) con el fin de decidir qué factores conservar y conocer la proporción de la *varianza única* (parte de la variación de la variable que es propia de ésta) y la *varianza común* (parte de la variación de una variable que es compartida con las demás) que están dentro del modelo para poder conocer qué tan representativos son los factores.

## COMUNALIDADES

Se denomina *comunalidad* a la proporción de la varianza explicada por los factores comunes en una variable; es decir, muestran la varianza única de cada variable. Estudiando las comunalidades obtenidas se puede valorar cuáles de las variables son mejor y peor explicadas por el modelo.

El Análisis Factorial comienza sus cálculos a partir de lo que se conoce como *matriz reducida* compuesta por los coeficientes de correlación entre las variables y las comunalidades en la diagonal.

La comunalidad es la suma de los pesos factoriales al cuadrado en cada una de las filas de matriz factorial. Lo anterior se resume en la siguiente expresión:

$$h^2 = P_{k1}^2 + P_{k2}^2 + \dots + P_{kj}^2,$$

donde:

$P_{kj}^2$  = peso factorial al cuadrado entre la variable  $X_k$  y el factor  $F_j$ ,  $j = 1, 2, \dots, m$   
donde  $m$  es el número de factores obtenidos.

Uno de los problemas del Análisis Factorial es que para conocer la matriz factorial se necesita de las comunalidades y para conocer estas últimas se necesita de la primera. Por lo tanto, es necesario estimar las comunalidades, estas estimaciones son las que aparecen en la tabla con el nombre de *comunalidades iniciales*.

En el análisis de Componentes Principales como no suponemos la existencia de ningún factor común la comunalidad toma como valor inicial 1. Es decir, dicho método de extracción asume que es posible explicar el 100% de la varianza observada, y por ello, todas las comunalidades iniciales son iguales a la unidad.

La tabla de comunalidades obtenida en este proyecto es:

**Cuadro.I.4 Comunalidades**

Comunalidades		
	Inicial	Extracción
Cocina	1.000	.804
Hacinamiento	1.000	.688
Disponibilidad de agua	1.000	.479
Electricidad	1.000	.417
Combustible	1.000	.675
Eliminación de basura	1.000	.601
Drenaje	1.000	.619
Televisión	1.000	.538
Videocasetera	1.000	.508
Licuada	1.000	.578
Refrigerador	1.000	.570
Lavadora	1.000	.509
Teléfono	1.000	.595
Boiler	1.000	.576
Auto propio	1.000	.500
Computadora	1.000	.485
Método de extracción: Componentes principales.		
Fuente: Elaboración propia en SPSS a partir de los datos del censo de población y vivienda del 2000 del INEGI		

En general las variables están suficientemente explicadas por los factores. Sin embargo, la variable "Cocina" es la variable mejor explicada por los factores y la variable "Electricidad" es la peor explicada por el modelo.

Ahora bien, para decidir la cantidad de factores que se van a extraer se utilizan los valores propios. Los cuales se explican en la siguiente sección.

## VALORES PROPIOS

El cuadrado de un carga factorial indica la proporción de la varianza explicada por un factor en una variable en particular. La suma de los cuadrados de los pesos de cualquier columna es lo que denominamos *valores propios* ( $\lambda$ ) y expresan la cantidad total de varianza que explica ese factor para las variables consideradas como grupo.



Si dividimos el valor propio entre el número de variables nos indica la proporción de la varianza de las variables que explica el factor. Lo anterior se resume en las siguientes expresiones:

<b>Factor</b>	<b>Valores propios ( <math>\lambda_m</math> )</b>	<b>Proporción de varianza explicada por el factor m</b>
1	$\lambda_1 = P_{11}^2 + P_{21}^2 + P_{31}^2 + \dots + P_{n1}^2$	$\frac{\lambda_1}{n}$
2	$\lambda_2 = P_{12}^2 + P_{22}^2 + P_{32}^2 + \dots + P_{n2}^2$	$\frac{\lambda_2}{n}$
.	.	.
.	.	.
.	.	.
m	$\lambda_m = P_{1m}^2 + P_{2m}^2 + P_{3m}^2 + \dots + P_{nm}^2$	$\frac{\lambda_m}{n}$

Donde m es le número de factores obtenidos y n es el número de variables incluidas en el análisis. En este caso m = 3 y n = 16

Como se mencionó anteriormente, las cargas factoriales pueden tener como valor máximo 1 por esto los valores propios pueden tomar un valor máximo igual al número total de variables involucradas en el análisis. Es decir, que ese factor explique por completo el comportamiento de todas las variables involucradas.

En la tabla de *porcentajes de varianza explicada* (ver Cuadro.1.5) se ofrece un listado de los valores propios y del porcentaje de varianza que representa cada uno de ellos. La tabla de porcentajes de varianza explicada resultante en el proyecto es la siguiente:

**Cuadro.1.5 Varianza total explicada**

Total de varianza explicada						
Componente	Valores propios Iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de Varianza	% Acumulado	Total	% de Varianza	% Acumulado
1	6.342	39.636	39.636	6.342	39.636	39.636
2	1.627	10.166	49.802	1.627	10.166	49.802
3	1.174	7.338	57.140	1.174	7.338	57.140
4	.875	5.468	62.609			
5	.754	4.713	67.322			
6	.627	3.919	71.241			
7	.597	3.728	74.969			
8	.542	3.386	78.355			
9	.529	3.309	81.664			
10	.488	3.051	84.716			
11	.473	2.958	87.673			
12	.460	2.872	90.546			
13	.447	2.793	93.339			
14	.406	2.540	95.879			
15	.343	2.143	98.022			
16	.316	1.978	100.000			
Método de extracción: Componentes principales.						
Fuente: Elaboración propia en SPSS a partir de los datos del censo de población y vivienda del 2000 del INEGI						

Por defecto, se extraen tantos factores como autovalores mayores que 1 tiene la matriz analizada. Esto porque, por ejemplo, el primer valor propio  $\lambda_1 = 6.342$  indica que ese factor representa tanta varianza como 6.342 variables; por lo tanto, no tiene mucho sentido incluir a los autovalores menores a uno ya que no alcanzan a representar a una variable completa.

En este caso se extrajeron 3 factores que representan el 57.140% de la varianza común de las variables. La información de estos factores está en las columnas incluidas dentro de "Extraction Sums of Squared Loadings".

Se observa que el primer factor representa el 39.636% de la varianza común que es igual a la varianza de 6.342 variables. El segundo factor representa el 10.166% de la varianza común que es igual a 1.627 variables. Y por último, el tercer factor representa el 7.338% de la varianza común que es igual a la varianza de 1.174 variables.

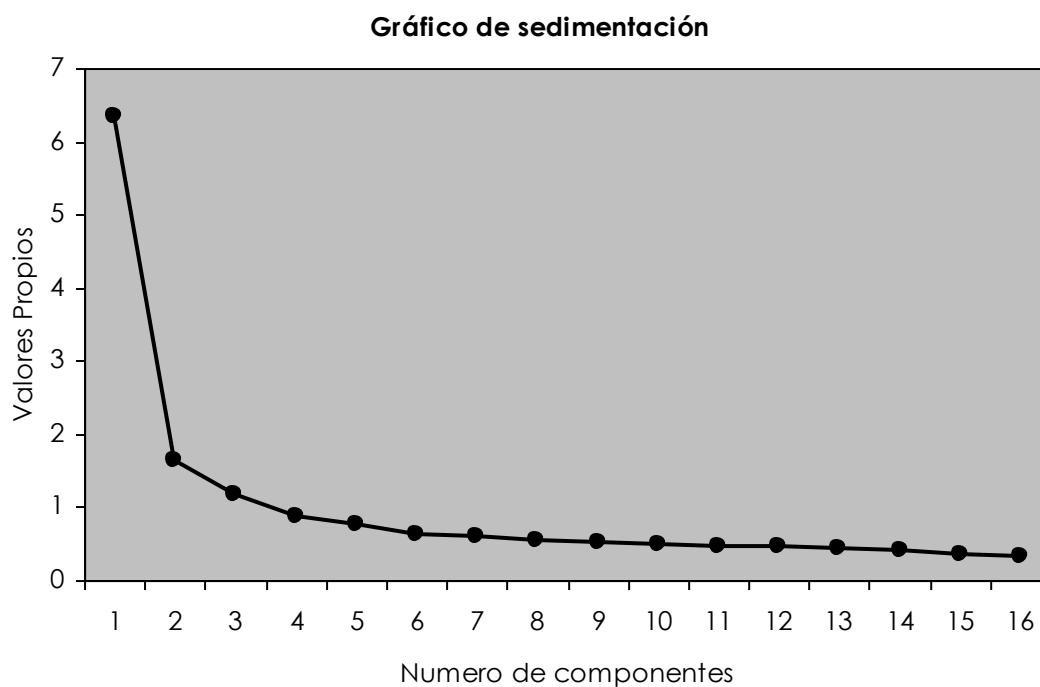
## GRÁFICO DE SEDIMENTACIÓN

El *gráfico de sedimentación* sirve para determinar el número óptimo de factores. Originalmente propuesto por Catell (1966) consiste simplemente en una representación gráfica del tamaño de los autovalores. Tanto la tabla de porcentajes de varianza explicada como el gráfico de sedimentación ofrecen los autovalores ordenados de mayor a menor. Si un valor propio se aproxima a cero, esto significa que el factor correspondiente a ese valor propio es incapaz de explicar una cantidad relevante de la varianza total y por tanto se considera un factor residual y carente de sentido en el análisis.

Es conveniente inspeccionar el gráfico de sedimentación de izquierda a derecha, buscando el punto de inflexión en el que los eigen valores dejan de formar una pendiente significativa y comienzan a describir una caída de poca inclinación. Es importante resaltar que el gráfico de sedimentación no varía con el número de factores seleccionado.

En este caso el gráfico de sedimentación es el siguiente:

**Gráfico.I.1 Gráfico de sedimentación**



Fuente: Elaboración propia en SPSS a partir de los datos del censo de población y vivienda del 2000 del INEGI

El gráfico de sedimentación (ver Gráfico.1.1) concuerda con la tabla de porcentajes de varianza explicada (Cuadro.1.5); ya que ambos muestran que los primeros tres factores son los más representativos. En el gráfico, es a partir del 3er punto que ya no presenta una inclinación relevante, por lo que esta es otra prueba de que el número de factores a extraer son 3.

### **I.II.III ROTACIÓN DE FACTORES**

Las *rotaciones factoriales* se utilizan para facilitar la interpretación de los factores al seleccionar una solución más sencilla e interpretable.

Existen varios métodos de rotación dependiendo de la correlación que exista entre los factores resultantes. Así, se habla de *rotación ortogonal* cuando la correlación entre factores es nula; es decir, los ejes factoriales tienen un ángulo de 90° entre sí. En tanto, se dice que la rotación es *oblicua* cuando la correlación entre los ejes factoriales no es nula o cuando es relativamente alta; es decir, si el ángulo es distinto de 90°

Lo más recomendable es la rotación ortogonal, aunque en el caso de que existan razones para pensar que los factores están correlacionados entonces se utiliza la rotación oblicua.

Para fines de este proyecto, se utilizó la rotación ortogonal llamada *varimax*.<sup>10</sup> Este método de rotación minimiza el número de variables que tienen valores altos en cada factor. De esta manera simplifica la interpretación de los factores optimizando la solución por columna. En síntesis se hacen girar los ejes de coordenadas, que representan los factores, hasta conseguir que se aproxime al máximo a las variables en que están saturados.

La rotación de factores transforma la matriz inicial en otra denominada *matriz factorial rotada* de más fácil interpretación. Esta última es una combinación lineal de la matriz factorial inicial y la *matriz de transformación de los factores o componentes* (Cuadro.1.6).

La matriz de transformación de los componentes es la matriz utilizada para rotar la solución factorial inicial. Esto se resume con la siguiente expresión:

---

<sup>10</sup> Durante el proceso se intentó hacer una rotación oblicua Oblimin con  $\delta = 0$ , sin embargo, los resultados de los pesos factoriales rotados no presentaban variación significativa.

$$\Lambda^* = \Lambda T,$$

donde:

$\Lambda$  = la matriz factorial inicial.

T = la matriz de transformación de los componentes.

$\Lambda^*$  = la matriz factorial después de la rotación.

En este caso la matriz de transformación utilizada para la rotación de los ejes fue:

**Cuadro.1.6 Matriz de transformación de componentes**

Matriz de transformación			
Componente	1	2	3
1	.761	.602	.241
2	-.648	.712	.269
3	-.010	-.361	.932
Método de extracción: Componentes principales.			
Método de rotación: Varimax con Normalización de Kaiser.			
Fuente: Elaboración propia en SPSS a partir de los datos del censo de población y vivienda del 2000 del INEGI			

Es importante mencionar que, con la rotación factorial, las communalidades no se alteran. Lo que cambia es la matriz factorial (Cuadro.1.8) y la tabla de porcentajes de varianza explicada (Cuadro.1.7). En esta última lo que se altera es la varianza explicada por cada factor más no el total de varianza explicada por los factores.

En este trabajo la tabla resultante de porcentajes de varianza al hacer la rotación es:

**Cuadro.1.7 Varianza total explicada con factores rotados**

Total de varianza explicada									
Componente	Eigenvalues Iniciales			Sumas de las saturaciones al cuadrado de la extracción			Sumas de las saturaciones al cuadrado de la rotación		
	Total	% de Varianza	% Acumulado	Total	% de Varianza	% Acumulado	Total	% de Varianza	% Acumulado
1	6.342	39.636	39.636	6.342	39.636	39.636	4.359	27.246	27.246
2	1.627	10.166	49.802	1.627	10.166	49.802	3.276	20.474	47.719
3	1.174	7.338	57.140	1.174	7.338	57.140	1.507	9.421	57.140
4	.875	5.468	62.609						
5	.754	4.713	67.322						
6	.627	3.919	71.241						
7	.597	3.728	74.969						
8	.542	3.386	78.355						
9	.529	3.309	81.664						
10	.488	3.051	84.716						
11	.473	2.958	87.673						
12	.460	2.872	90.546						
13	.447	2.793	93.339						
14	.406	2.540	95.879						
15	.343	2.143	98.022						
16	.316	1.978	100.000						
Método de extracción: Componentes principales.									
Fuente: Elaboración propia en SPSS a partir de los datos del censo de población y vivienda del 2000 del INEGI									

Las primeras dos partes de la tabla (Cuadro.1.7) se mantienen igual que cuando se obtiene la tabla antes de la rotación, mientras que en las columnas incluidas dentro de "Sumas de las saturaciones al cuadrado de la rotación" los datos han cambiado. Estos nuevos datos nos dicen que el primer componente es ahora capaz de representar el 27.246% de la varianza común que es igual a la varianza de 4.359 variables. El segundo factor representa ahora el 20.474% de la varianza común que es igual a 3.276 variables. Y por último, el tercer factor ahora representa el 9.421% de la varianza común que es igual a la varianza de 1.507 variables. Esto quiere decir que se mejoraron las representaciones de los factores para las variables y el porcentaje de varianza explicada por el primer factor en la primera tabla de "Total de varianza explicada" se repartió entre los otros dos factores.

Ahora se muestra la matriz factorial (Cuadro.1.8) que es la solución propiamente dicha del Análisis Factorial aplicado en este trabajo para las variables seleccionadas.

**Cuadro.I.8 Matriz factorial**

<b>Matriz factorial rotada</b>			
	Componente		
	1	2	3
Cocina	.037	.085	.892
Hacinamiento	.277	.311	.717
Disponibilidad de agua	.636	.265	.063
Electricidad	.625	-.091	.135
Combustible	.784	.247	.013
Eliminación de basura	.708	.311	-.060
Drenaje	.709	.340	.030
Televisión	.714	.080	.147
Videocasetera	.271	.651	.104
Licuada	.723	.173	.159
Refrigerador	.609	.396	.206
Lavadora	.459	.534	.119
Teléfono	.276	.710	.119
Boiler	.308	.674	.162
Auto propio	.099	.690	.122
Computadora	-.019	.696	.011
Método de extracción: Componentes principales.			
Método de rotación: Varimax con Normalización de Kaiser.			
a. Rotación lograda con 5 iteraciones.			
Fuente: Elaboración propia en SPSS a partir de los datos del censo de población y vivienda del 2000 del INEGI			

<b>Factor 1</b>	Servicios y enseres básicos
<b>Factor 2</b>	Enseres medios
<b>Factor 3</b>	Condiciones de la vivienda

A diferencia de la matriz factorial inicial (Cuadro.I.3) ahora es más claro cuales son las variables que definen a cada factor. Esto facilita enormemente la interpretación de cada uno de ellos; y por lo tanto, es más fácil nombrarlos.

De esta manera, el factor 1 representa los servicios y enseres básicos o comunes que presentan las viviendas de los niños y jóvenes de entre 12 y 24 años. El factor dos representa los bienes que no son básicos en las viviendas y que, dependiendo de los recursos económicos de cada hogar, las familias e individuos deciden adquirir. Y por último, el tercero se refiere a las características propias de la vivienda, este último factor no cambió con respecto a la matriz factorial inicial.

Una vez realizados los factores, se guardaron como variables en SPSS para obtener el valor de sus cuartiles y de esta manera poder utilizarlos para hacer la clasificación de los sectores. Los valores de los cuartiles para cada factor rotado son los siguientes:

<b>Cuartiles</b>	<b>Factor rotado 1</b>	<b>Factor rotado 2</b>	<b>Factor rotado 3</b>
<b>25</b>	- 0.6504683	- 0.5936	- 0.457496
<b>50</b>	- 0.3872622	0.1677272	- 0.2580187
<b>75</b>	0.1653029	0.8296232	- 0.0010153

Es importante recalcar que las preguntas del cuestionario ampliado clasifican las condiciones más desfavorables con los números más altos y las mejores con los números más chicos. Por lo tanto, los valores pequeños se refieren a los estratos altos y los grandes a los bajos. Así se hicieron 4 grupos distintos: Sector bajo, Sector medio bajo, Sector medio alto y Sector alto para cada factor rotado.

La siguiente sintaxis muestra el criterio utilizado para la clasificación de los sectores en los tres diferentes factores.

Para el factor rotado 1: "Servicios y enseres básicos"

Nombre de la variable: *cuartilesF1*

```

DATASET ACTIVATE DataSet1.
IF (FAC1_1 < -.6504683) cuartilesF1=4.
IF (FAC1_2 > -.6504683 & FAC1_2 < -.3872622) cuartilesF1=3.
IF (FAC1_2 > -.3872622 & FAC1_2 < .1653029) cuartilesF1=2.
IF (FAC1_2 > .1653029) cuartilesF1=1.
EXECUTE.
VARIABLE LABELS cuartilesF1 "Cuartiles Factor 1".
VALUE LABELS cuartilesF1 1 "Sector bajo" 2 "Sector medio bajo" 3
"Sector medio alto" 4 "Sector alto".

```

Para el factor rotado 2: "Enseres medios"

Nombre de la variable: *cuartilesF2*

```

DATASET ACTIVATE DataSet1.
IF (FAC2_1 < -.5936) cuartilesF2=4.
IF (FAC2_1 > -.5936 & FAC2_1 < .1677272) cuartilesF2=3.
IF (FAC2_1 > .1677272 & FAC2_1 < .8296232) cuartilesF2=2.
IF (FAC2_1 > .8296232) cuartilesF2=1.
EXECUTE.
VARIABLE LABELS cuartilesF2 "Cuartiles Factor 3".
VALUE LABELS cuartilesF2 1 "Sector bajo" 2 "Sector medio bajo" 3
"Sector medio alto" 4 "Sector alto".

```

Para el factor rotado 3: "Condiciones de la vivienda"

Nombre de la variable: *cuartilesF3*

```

DATASET ACTIVATE DataSet1.
IF (FAC3_1 < -.457496) cuartilesF3=4.
IF (FAC3_1 > -.457496 & FAC3_1 < -.2580187) cuartilesF3=3.
IF (FAC3_1 > -.2580187 & FAC3_1 < -.0010153) cuartilesF3=2.
IF (FAC3_1 > -.0010153) cuartilesF3=1.

```



```
EXECUTE.  
VARIABLE LABELS cuartilesF3 "Cuartiles Factor 3".  
VALUE LABELS cuartilesF3 1 "Sector bajo" 2 "Sector medio bajo" 3  
"Sector medio alto" 4 "Sector alto".
```

## **CAPITULO II**

### **ANÁLISIS DESCRIPTIVO DE LA CONDICIÓN EDUCATIVA DE LOS JÓVENES DE 12 A 24 AÑOS**

Describir las condiciones educativas de la población de jóvenes de 12 a 24 años es una parte importante de este trabajo; ya que ayuda a descubrir los posibles problemas que enfrentan los jóvenes dentro en su educación y respaldar, así, las hipótesis que se plantean en torno a ésta.

Se sabe que uno de los mayores problemas que enfrentan las familias mexicanas es la falta de recursos para apoyar la educación formal de los menores ya que muchos hogares tienen que hacer un gran esfuerzo para cubrir primeramente las necesidades básicas. Por eso, en ocasiones, las personas en edad de asistir a la escuela se ven con frecuencia obligadas a apoyar económicamente a la familia, consiguiendo un trabajo, que si dejarán de desempeñar la familia tendría que afrontar un costo de oportunidad muy alto; es decir, dejarían de percibir recursos económicos para atender las necesidades familiares por ocuparse de las tareas escolares.

El problema anterior es el que, principalmente, se ve reflejado dentro de la información expuesta en este capítulo. Se puede notar cómo las diferencias para apoyar económicamente a los estudiantes depende mucho de los sectores socioeconómicos en donde el hogar se encuentre.

#### **II.1 METODOLOGIA**

Para la elaboración de las tablas y las gráficas de este capítulo, de la muestra de 12 a 24 se hicieron dos grupos, un grupo que abarca de los 12 a 17 años de edad y otro de los 18 a 24 años. Esta agrupación se hizo porque el primer grupo es el que generalmente comprende las edades en las que niños y jóvenes asisten a la educación básica o media superior, mientras que el grupo de 18 a 24 años comprende las edades de la educación media superior y profesional. Cada grupo de edad presenta características educativas y laborales muy diferentes por lo que amerita tratarlos por separado, tanto en el análisis descriptivo como en el modelo de regresión logística.

Para este segundo capítulo se utilizó el primer factor sin rotar; ya que este expone claramente la situación educativa de la población de estudio al concentrar y describir en un solo factor el comportamiento de las variables incluidas en el Análisis Factorial (ver Cuadrol.3). Los valores de los cuartiles para este factor son:

<b>Cuartiles</b>	<b>Factor 1 sin rotar</b>
<b>25</b>	- 0.7517350
<b>50</b>	- 0.3110871
<b>75</b>	0.4598459

Al igual que con los factores rotados, los valores menores a - 0.7517350 son lo que se refieren al sector alto; los que están dentro de -0.751735 y -0.3110871 son los de sector medio alto; los que están dentro de -0.3110871 y 0.4598459 son los de sector medio bajo; y, finalmente, los que son mayores a 0.4598459 son lo que pertenecen al sector bajo.

Es importante recalcar que para cumplir el objetivo de este trabajo no solo se utilizaron las variables originales del censo; si no que además, se tuvieron que construir variables nuevas a partir de las variables originales. Para las nuevas, la construcción se irá explicando a la largo de esta sección.<sup>11</sup>

Primeramente se construyó la variable dependiente o resultante; es decir, la variable que queremos describir (capítulo 2), explicar y modelar (capítulo 3). Para efectos de este trabajo, esta variable debía mostrar la situación educativa de los jóvenes de entre 12 a 24 años de edad. Por lo tanto se construyó la variable “Situación educativa” la cual incluye las diferentes causas de abandono en relación a la asistencia de los jóvenes de los distintos grupos de edad. Es decir, si el joven Asiste actualmente a la escuela la variable “Situación educativa” toma el valor de 1; si el joven no asiste a la escuela y la causa de abandono es “No quiso o no le gustó estudiar” entonces la variable toma el valor de 2; si la causa de abandono es “Falta de dinero o tenía que trabajar” entonces toma el valor de 3; si la causa es “Se casó o unió” entonces toma el valor de 4; si la causa es “No lo dejaron o por encargarse del hogar” entonces toma el valor de 5; si la causa es “Terminó sus estudios” entonces toma el valor de 6; y finalmente, si fue por “Otro motivo” entonces toma el valor de 7.

Para fines prácticos, en este trabajo, nos referiremos a la variable de “Situación Educativa” como “causas de abandono”.

---

<sup>11</sup> Las sintaxis de la programación elaboradas y utilizadas en SPSS para la construcción de las diferentes variables utilizadas en este capítulo se encuentran en el Anexo del capítulo dos al final del trabajo.

## II.II ASISTENCIA<sup>12</sup>

Tomando en cuenta que en México la educación básica generalmente abarca las edades entre 6 y 15 años, las Gráficas.II.1 muestran la asistencia de los jóvenes de 12 a 17 años de los distintos sectores. Se puede ver que para cada uno de los sectores más del 60% asiste a una institución educativa sin importar el sexo; sin embargo, sí existen diferencias de asistencia claras entre los sectores. Se ve claramente como al aumentar el sector aumenta la asistencia, por lo que, la gran diferencia entre el sector alto y el sector bajo se vuelve muy evidente. Para el sector alto el 93.06% de los jóvenes varones asiste a la escuela, mientras que en el sector bajo asisten el 64.12%. Para las mujeres ocurre algo similar, para el sector alto el 94.42% asiste a la escuela, mientras que en el sector bajo asisten el 62.21%.

En comparación con las gráficas mencionadas anteriormente (Gráficas.II.1), las Gráficas.II.2, que abarcan las edades de 18 a 24 años, dejan ver la transición entre los sectores; es decir, es evidente el aumento de los porcentajes referentes a la no asistencia para cada uno de los sectores. Sin embargo, se sigue evidenciando la diferencia entre los sectores.

Ahora en estas edades podemos notar como la mayor parte de la población para los sectores bajo, medio bajo y alto no asiste a la escuela. Nuevamente se aprecia, además, la gran diferencia entre el sector bajo y el sector alto. Para el sector alto el 57.87% de los jóvenes varones asiste a la escuela, mientras que en el sector bajo asisten solo el 10.71%. Para las mujeres ocurre algo similar, para el sector alto el 57.542% asiste a la escuela, mientras que en el sector bajo asisten solo el 11.59%.

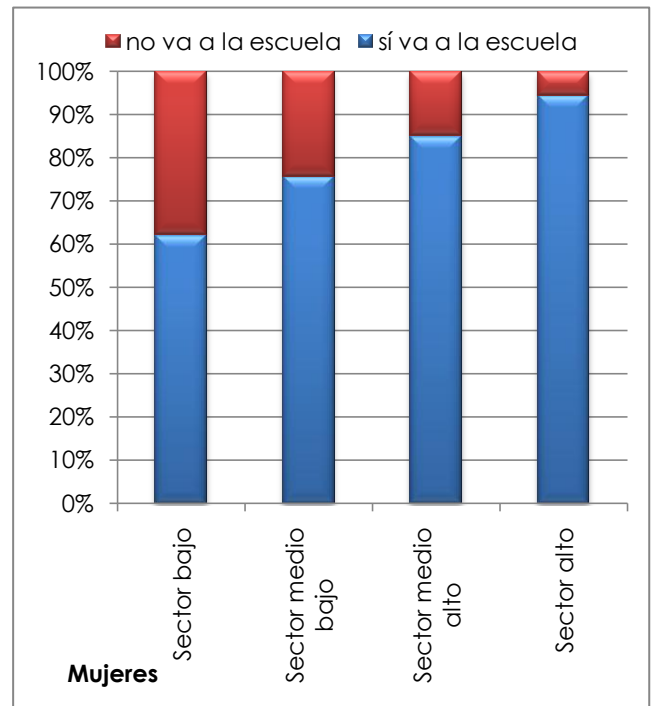
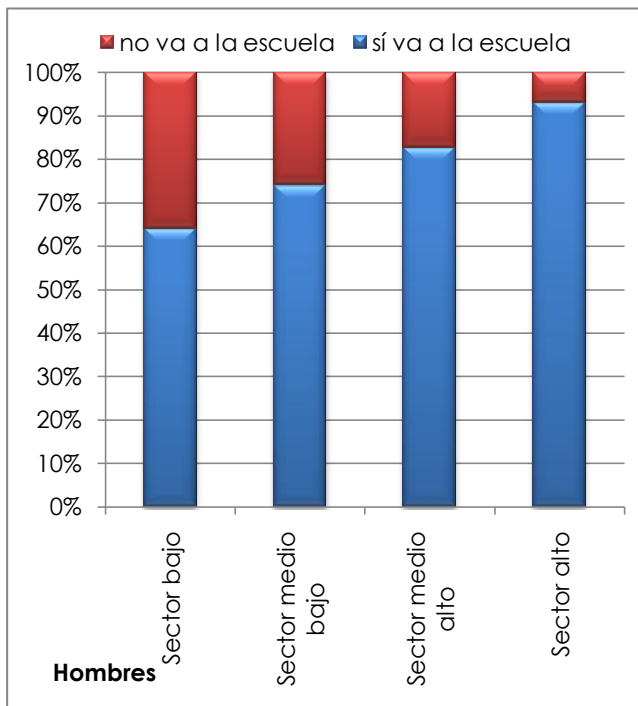
Se observa en las cuatro gráficas que el comportamiento de los datos de asistencia y no asistencia no se altera con respecto al sexo para cada uno de los grupos de edad.

---

<sup>12</sup> Las tablas correspondientes a estas gráficas se encuentran en el Anexo al final del trabajo.

**Gráficas.II.1**

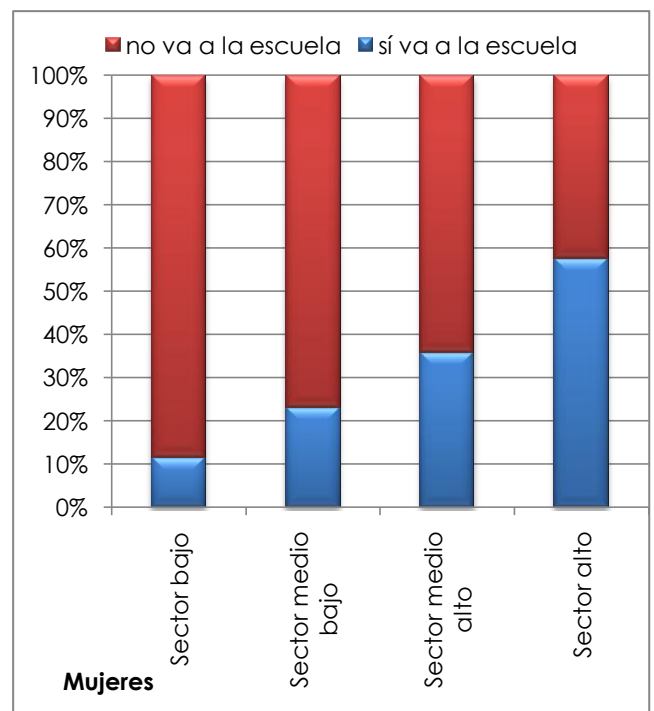
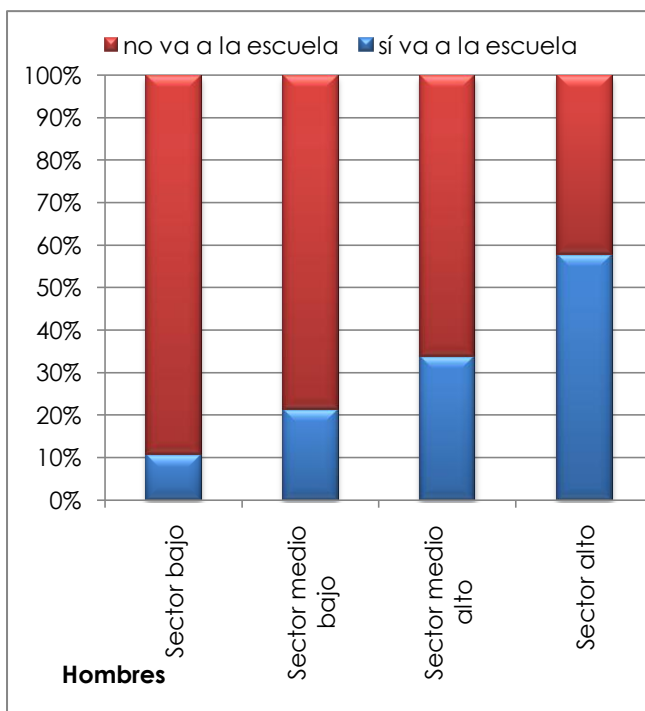
**Asistencia por sexo (edades de 12 a 17 años)**



Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI

**Gráficas.II.2**

**Asistencia por sexo (edades de 18 a 24 años)**



Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI

## **II.III SITUACIÓN EDUCATIVA**

### **CAUSAS DE ABANDONO**

Anteriormente se expuso que la mayoría de las personas dentro del grupo de edad de 12 a 17 años asisten a la escuela y que las de 18 a 24 asisten en menor medida; sin embargo, es ahora de interés conocer el motivo de la deserción y el nivel de estudios alcanzado.

En las edades de 12 a 17 años (Cuadro.II.1), los motivos de la deserción para los sectores bajos se centran principalmente en la falta de interés del estudiante hacia la escuela; y en segundo lugar, la falta de recursos económicos. Para los demás sectores el primero es también el más representativo, lo cual indica que dentro del país aún falta incentivar el ingreso y la permanencia a la educación por parte de las escuelas y los padres de familia generando y sembrando el gusto por el conocimiento en la conciencia de los jóvenes. Es importante, también, destacar que por otro lado esta causa de abandono ("No quiso no le gustó") podría ser el reflejo de las pocas oportunidades de trabajo que hay en el país. Probablemente para los jóvenes terminar la educación básica no sea de mucha ayuda al momento de querer participar en la vida laboral; ya que, hay una desvalorización de las cualificaciones educativas inferiores por la división del trabajo (Calero, 2004). Es decir, los jóvenes pueden pensar que tener o no educación básica no va a hacer una diferencia en la percepción de su ingreso en el sector laboral y esto podría traer como consecuencia falta de interés de asistir a la escuela.

De la causa "Falta de dinero o tenía que trabajar" se puede efectivamente afirmar que las familias pertenecientes a los sectores bajos tienen dificultades en asignar recursos económicos a la educación de los menores y que esta causa va disminuyendo en porcentaje conforme se avanza de sector, esto sucede tanto entre los jóvenes de 12 a 17 años como entre los de 18 a 24 años y en ambos sexos.

Analizando el sexo, se puede destacar que las mujeres de 12 a 17 años son las que desertan menos y que, posiblemente con el paso del tiempo y de las generaciones, la decisión de casarse, de encargarse del hogar o de la obstaculización por parte de la familia ya no figura dentro de las razones principales de las mujeres para no asistir a la alguna institución de educación básica. Apenas las mujeres del sector más bajo que abandonan la escuela por dedicarse a los quehaceres del hogar representan un 1.87%

Sin embargo, no se debe dejar de resaltar que en comparación con los hombres estos motivos siguen siendo más representativos del sexo femenino.

**Cuadro.II.1 Situación educativa 12 a 17 años**

Situación educativa por sexo a partir del Factor 1 (edades de 12 a 17)									
Sector		Asisten	Causa de abandono de los que no asisten:						Total
		Sí	No quiso o no le gustó	Falta de dinero o por trabajo	Se casó o unió	No lo dejaron o por encargarse del hogar	Terminó sus estudios	Otro motivo	
Hombres	Sector bajo	67.96%	15.35%	13.84%	0.06%	0.74%	1.25%	0.80%	75,125
	Sector medio bajo	76.52%	13.94%	7.27%	0.05%	0.32%	0.64%	1.26%	69,235
	Sector medio alto	84.55%	9.80%	3.45%	0.07%	0.17%	0.65%	1.30%	64,832
	Sector alto	94.10%	3.78%	0.87%	0.03%	0.05%	0.34%	0.82%	60,889
Total		80.03%	11.05%	6.74%	0.06%	0.34%	0.74%	1.04%	270,081
Mujeres	Sector bajo	66.83%	13.76%	14.96%	0.38%	1.87%	1.37%	0.84%	68,851
	Sector medio bajo	78.09%	10.81%	7.60%	0.47%	0.98%	0.88%	1.18%	67,652
	Sector medio alto	86.98%	6.88%	3.38%	0.37%	0.56%	0.73%	1.09%	62,748
	Sector alto	95.54%	2.26%	0.76%	0.15%	0.13%	0.38%	0.78%	58,422
Total		81.20%	8.70%	6.99%	0.35%	0.93%	0.86%	0.97%	257,673

Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.

Para las edades de 18 a 24 años se observa un cambio significativo en las causas de abandono. Ahora "Falta de dinero o por trabajo" es la causa principal dentro de los sectores bajos mientras que, como es de esperarse, para los sectores altos lo es en menor medida. Sin embargo, es importante destacar que la causa de "No quiso o no le gustó" es ahora más fuerte en comparación con el Cuadro.II.1 para todos los sectores y en ambos sexos, pero sobre todo para los hombres. Por ejemplo, para el sector alto, los hombres de 18 a 24 años presentan un 14.59% mientras que los de 12 a 17 años muestran un 3.78%. En comparación con el sector, los hombres de 18 a 24 años presentan el 14.59% en sector alto y en contraste el sector bajo presenta el 32.42%.

Comparando por sexo en la causa "No quiso o no le gustó" del Cuadro.II.2, el porcentaje de los hombres de 18 a 24 para sector alto (14.59%) está muy por encima del de las mujeres que es de 8.38%. Nótese la repetición de este comportamiento en los demás sectores para ambos sexos.

Ya en este grupo de edad las mujeres, como en el Cuadro.II.1, son las que desertan más que los hombres por casarse o unirse, porque no la dejaron seguir estudiando o por dedicarse a los quehaceres del hogar; y en comparación con el cuadro de las mujeres de 12 a 17 años ahora estas causas tienen más peso. Por ejemplo para las mujeres de 18 a 24 años de sector bajo "Se casó o unió" y "No lo dejaron o por

dedicarse al hogar" presentan un 3.08% y un 3.91% respectivamente; mientras que, para las de 12 a 17 años representa un 0.38% y un 1.87% respectivamente.

Ya en los jóvenes de 18 a 24 años se observan cifras mayores al 2% en la causa de abandono "Terminó sus estudios". Y nuevamente observamos que las cifras van aumentando conforme al sector y al género. En el sector alto existe la mayor cantidad de personas que terminan sus estudios, pero además, son las mujeres las que tienen el porcentaje más alto (18.02%).

Otra característica importante de este grupo de edad es que además existe la causa "Otro motivo" por la cual los jóvenes deciden desertar y ésta empieza a ser ligeramente significativa.

**Cuadro.II.2 Situación educativa 18 a 24 años**

Situación educativa por sexo a partir del Factor 1 (edades de 18 a 24)									
Sector		Asisten	Causas de abandono de los que no asisten						
		Sí	No quiso o no le gustó	Falta de dinero o por trabajo	Se casó o unió	No lo dejaron o por encargarse del hogar	Terminó sus estudios	Otro motivo	Total
Hombres	Sector bajo	12.38%	32.42%	46.20%	1.05%	1.67%	4.86%	1.43%	34,240
	Sector medio bajo	23.51%	32.83%	33.25%	1.16%	0.73%	6.40%	2.11%	39,689
	Sector medio alto	36.57%	27.42%	23.06%	1.25%	0.58%	8.59%	2.54%	45,004
	Sector alto	61.03%	14.59%	10.34%	0.79%	0.22%	10.72%	2.30%	50,745
Total		35.95%	25.86%	26.31%	1.05%	0.73%	7.96%	2.14%	169,678
Mujeres	Sector bajo	13.66%	27.39%	44.11%	3.08%	3.91%	6.04%	1.82%	26,142
	Sector medio bajo	25.56%	23.35%	30.82%	4.85%	2.03%	10.90%	2.49%	35,924
	Sector medio alto	38.84%	17.72%	19.24%	4.16%	1.47%	16.04%	2.54%	41,860
	Sector alto	61.13%	8.38%	7.61%	2.21%	0.52%	18.02%	2.13%	47,360
Total		38.31%	17.80%	22.65%	3.53%	1.73%	13.71%	2.28%	151,286

Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.



## NIVEL DE ESTUDIOS ALCANZADO

En este apartado se tuvo que construir una variable que englobara el nivel de estudios alcanzado por los jóvenes al momento de ser entrevistados. La variable construida lleva el nombre de "Nivel de estudios" (*nivest*) y toma los valores de 0 si la persona nunca ha ido a la escuela; 1 si la persona cursó o cursa preescolar; 2 "Primaria" si la persona cursó o cursa primaria; 3 "Secundaria" si el entrevistado cursó o cursa secundaria o alguna carrera técnica o comercial, y además, dice tener como antecedente escolar la primaria terminada; 4 "Preparatoria" si el joven cursó o cursa el bachillerato o preparatoria, normal o alguna carrera técnica o comercial, y además, dice tener como antecedente la secundaria terminada; 5 "Profesional" si la persona cursó o cursa una licenciatura, alguna carrera técnica o comercial, normal, y además, dice tener como antecedente la preparatoria terminada; 6 "Postgrado" si el joven cursó o cursa alguna maestría o doctorado; y finalmente, 9 si responde No Especificado.

Con la variable anterior se pudo crear una nueva variable que agrupa solamente la educación primaria, secundaria, media superior, y profesional. Las clasificaciones de "Ninguno", "Kinder" y "No especificado" de la variable *nivest* se sacaron de la muestra al representar un porcentaje bajo de la muestra de la población de estudio. Además, la educación "Profesional" y de "Postgrado" se juntaron en una sola llamada "Profesional". El nombre de la nueva variable de nivel educativo es *nivestn* que representan "El nivel de estudios" y toma el valor de 1 si el joven cursó o cursa la primaria, 2 si cursó o cursa la secundaria, 3 si cursó o cursa la Preparatoria o similar, y 4 si cursó o cursa alguna licenciatura o similar o postgrado.

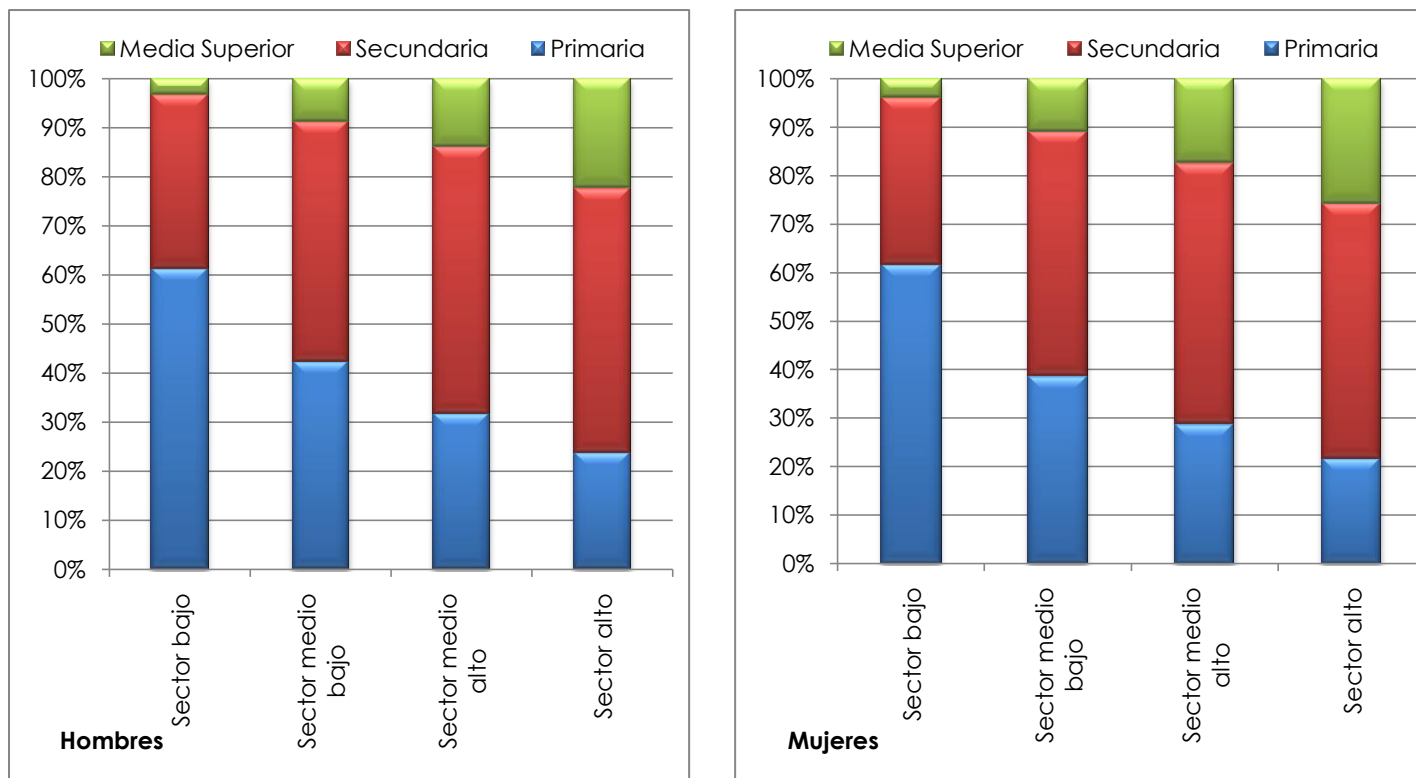
En general, para los hombres y mujeres en las edades de 12 a 17 años (Gráficas.II.3), se observa que la mayoría cuentan con educación secundaria principalmente, seguido del nivel de primaria y, finalmente, de media superior. Es importante destacar cómo los sectores bajos son los que presentan mayor rezago, al presentar las mayores proporciones en primaria y proporciones bajas en secundaria y media superior. Los sectores restantes son los que en su mayoría se encuentran dentro de la educación secundaria, lo que sería de esperarse por pertenecer a este grupo de edad en el que muchos cursan este nivel.

Nuevamente, se observan grandes diferencias entre sectores. Los individuos que pertenecen a los sectores altos son los que tienen los mayores porcentajes de

educación media superior; mientras que, el sector bajo apenas y alcanza a tener algunos individuos que alcanzan este nivel de educación.

**Gráficas.II.3**

**Nivel de estudios (edades de 12 a 17 años)**



Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI

Ya para los hombres y mujeres en las edades de 18 a 24 años (Gráficas.II.4), las gráficas muestran cómo el nivel de estudios depende enormemente de los sectores a los que pertenecen los estudiantes. Sobre todo porque esta edad es en la que los jóvenes deberían de estar terminando la educación media superior o cursando los estudios profesionales o incluso haber concluido esta última, lo cual se comprueba que no sucede.

Efectivamente, en este grupo de edad ya algunos jóvenes tienen educación profesional; sin embargo, ésta es más significativa en el sector alto y es mínima en el sector bajo. Como se esperaría, pasa lo contrario con la educación primaria y secundaria, los sectores altos presentan un porcentaje mínimo de personas que cuentan sólo con este tipo de educación (2.99% para hombres y 2.44% para mujeres) y el sector bajo presenta más de la mitad de jóvenes únicamente con ésta (51.37% para hombres y 52.76% para mujeres).

La otra diferencia marcada entre sectores es en la educación profesional. Por una parte, los hombres de sector alto que tienen educación profesional son el 37.43%; mientras que, los del bajo son apenas el 2.44%. Para las mujeres sucede algo muy similar, las de sector alto que cuentan con esta educación son el 44.19% y las de sector bajo son el 2.88%.

En conclusión, en este grupo de edad (18 a 24 años) se presentan diferencias en cuanto a primaria y secundaria, por una parte, y media superior y profesional por otra. La educación básica es más representativa en los sectores bajos para este grupo de edad; mientras que en contraste, la educación media superior y profesional es más representativa de los sectores altos.

En cuanto a sexo, la diferencia entre hombres y mujeres es importante. Retomando las cifras anteriores, se compara un 37.43% de los hombres de sector alto con educación profesional contra un 44.19% de las mujeres de este mismo sector. Este fenómeno se manifiesta en los diferentes sectores, lo que quiere decir que la pertenencia a un sector en específico para las mujeres es importante y decisiva para seguir o no con su educación.

Analizando lo anterior, se puede ver claramente como los estudiantes pertenecientes a sectores altos son los que reciben más apoyo económico y cultural por parte de los jefes de familia y de la familia como tal y tienen mayores oportunidades para permanecer en el sistema educativo. Esto debido probablemente a que al pertenecer a los sectores altos los recursos económicos facilitan el enriquecimiento cultural de una persona lo cual ayuda al crecimiento y enriquecimiento integral de la educación de los jóvenes.

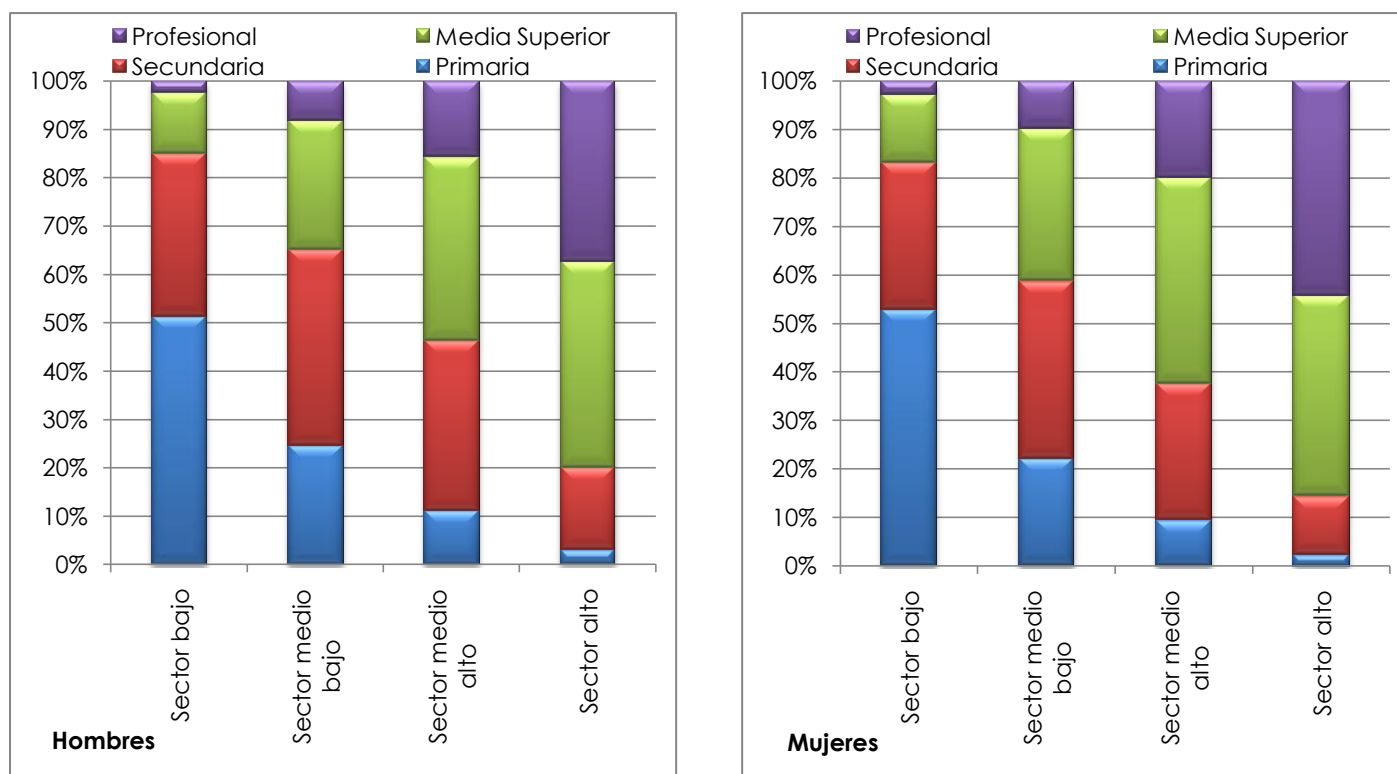
Es posible, además, que los resultados obtenidos sean un reflejo de las muchas o pocas aspiraciones inculcadas a los jóvenes en la vida. Por lo tanto, se puede concluir que los estímulos realizados en las familias, junto con el apoyo económico que se tenga pueden generar efectos positivos en la educación de los jóvenes de la población.

Por otro lado, los sectores bajos tienen los menores niveles de estudios, lo que indica que no contaron con apoyo para continuar sus estudios, que tuvieron que empezar a trabajar o que simplemente no les interesó continuar. Esto último puede ser debido a lo que se mencionó anteriormente acerca de las pocas oportunidades laborales que hay en el mercado laboral. Mercado en donde la educación básica ya no es suficientemente restrictiva y, por lo tanto, los jóvenes pueden llegar a pensar que entrar

y concluir la educación básica o no hacer ninguna de las dos crea poca diferencia en el sector laboral. Por lo tanto, se muestra un atraso importante de educación en comparación con los sectores altos.

**Gráficas.II.4**

**Nivel de estudios (edades de 18 a 24 años)**



Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.

**NÚMERO MEDIO DE AÑOS APROBADOS**

El número medio de años aprobados es un índice ilustrativo e interesante de los logros educativos de los jóvenes.

En los cuadros (Cuadro.II.3 y Cuadro.II.4) se puede una vez más apreciar que los sectores bajos efectivamente son los que muestran menores años de educación y en promedio apenas alcanzan a cubrir la educación primaria, esto sucede tanto en hombres como en mujeres.

En los sectores altos para las edades de 12 a 17 (Cuadro.II.3) la media de años aprobados es cercana a los 9 años, tanto para mujeres como hombres.

Para este mismo sector pero dentro de las edades de 18 a 24 (Cuadro.II.4) se expone claramente como los de sectores bajos son los que se quedan rezagados; mientras que los estudiantes de sector alto han seguido avanzando y por tanto su media de años aprobados es de 13 para los hombres y 14 para las mujeres.

**Cuadro.II.3 Promedio de años aprobados acumulados edades de 12 a 17 años**

Medias de años aprobados acumulados (edades de 12 a 17)		
Sector	Media de años aprobados acumulados	
Hombres	Sector bajo	6.33
	Sector medio bajo	7.29
	Sector medio alto	8.00
	Sector alto	8.52
Mujeres	Sector bajo	6.35
	Sector medio bajo	7.66
	Sector medio alto	8.16
	Sector alto	8.65

Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.

**Cuadro.II.4 Promedio de años aprobados acumulados edades de 18 a 24 años**

Medias de años aprobados acumulados (edades de 18 a 24)		
Sector	Media de años aprobados acumulados	
Hombres	Sector bajo	7.49
	Sector medio bajo	9.66
	Sector medio alto	11.12
	Sector alto	13.09
Mujeres	Sector bajo	7.56
	Sector medio bajo	10.24
	Sector medio alto	12.03
	Sector alto	13.95

Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.

En las edades de 18 a 24 años se pueden notar las diferencias abruptas entre los sectores para cada uno de los sexos. Mientras que para los hombres de sector bajo la media de años aprobados es de 7.49, para los de sector alto es de 13.09, Es decir, los sectores bajos tienen, en general, únicamente la educación primaria acreditada; mientras que los de sector alto tienen, en general, la educación básica y media superior. Incluso se pueden encontrar cursando los estudios profesionales.

Para las mujeres sucede algo similar en las edades de 18 a 24 años. Para los sectores bajos, las mujeres tienen tan solo 7.56 como media de años aprobados. Es decir, generalmente, solo cuentan con la educación primaria y parte de la secundaria.

Mientras que en los sectores altos, la media de años aprobados es 13.95 que además resulta ser mayor que la de los hombres. Es decir, las mujeres dentro de este grupo de edad y sector terminan la educación básica, media superior e incluso pueden estar cursando los estudios profesionales.

En conclusión, para la parte de la situación educativa, los jóvenes dentro de las edades de 12 a 24 años en su mayoría cuentan con educación básica y pocos son los que logran llegar a la educación superior; ya que está íntimamente relacionado con el sector al que pertenecen. Es importante destacar que la causa de "No quiso o no le gustó" muestra cifras altas dentro de los jóvenes lo cual nos indica que a pesar de que las familias mexicanas han incentivado más la inserción a la escuela (SITEAL, 2009) aún falta hacer de la escuela una institución más atractiva e interesante para los jóvenes, así como brindar más apoyos económicos a los estudiantes de bajos recursos para que continúen en el sistema educativo.

## **II.IV SITUACIÓN DE ACTIVIDAD**

### **ACTIVIDAD PRINCIPAL**

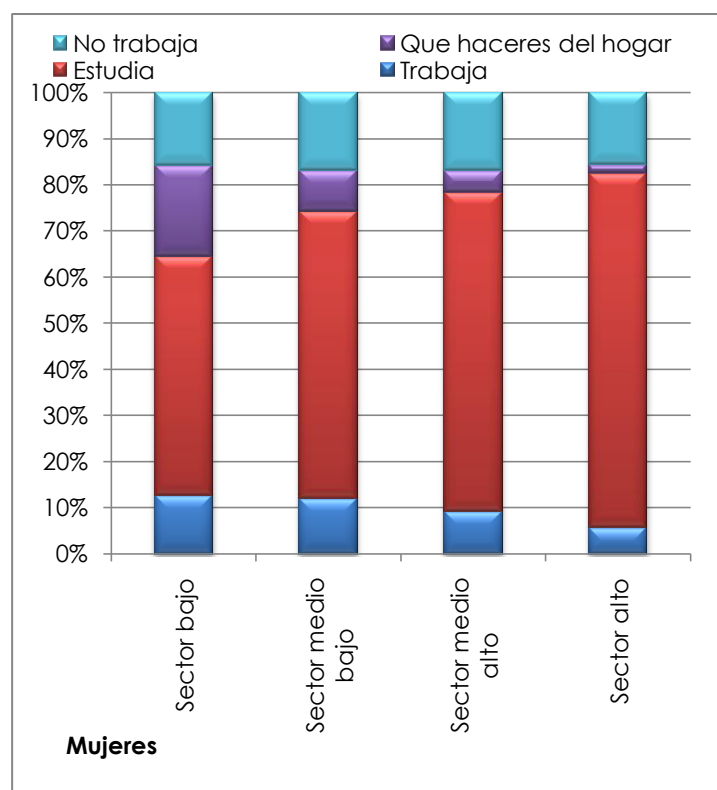
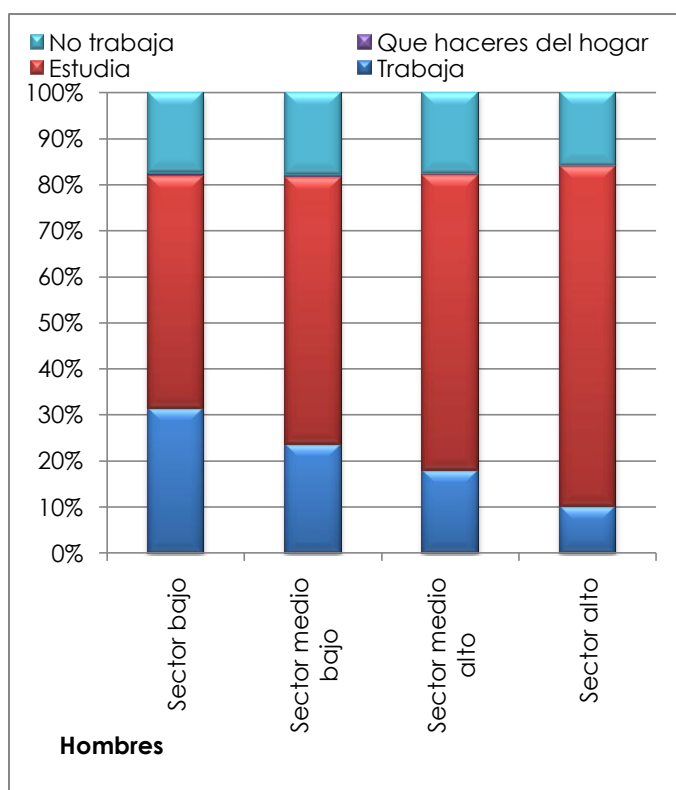
Para describir la actividad principal fue necesaria la construcción de una variable que englobara las diferentes categorías de actividad principal que proporciona el cuestionario del censo para poder exponerla y analizarla con mayor facilidad. El nombre de esta variable es "Actividad principal" (*actividadn*) y se basa en la variable de condición de actividad, que es la variable original del cuestionario y en la que se pregunta a la persona cual fue su actividad principal la semana anterior a la aplicación del cuestionario. La variable nueva de actividad principal toma el valor de 1 "Trabaja" si la persona declara que: trabaja, busca trabajo y en la verificación se rescata que trabaja, es estudiante y en la verificación trabaja, se dedica a los quehaceres del hogar y en la verificación trabaja, es jubilado y en la verificación trabaja, no trabaja y en la verificación trabaja, no tiene información de condición de actividad y en la verificación trabaja, tiene trabajo pero no trabajo; 2 "Busca trabajo" si declara haber buscado actividad laboral; 3 "Estudia" si declara ser estudiante; 4 "Quehaceres del hogar" si declara haberse dedicado al hogar; 5 "No trabaja" si no realiza alguna actividad laboral o está jubilado o pensionado; 6 "Incapacitado" si se encuentra incapacitado permanentemente para trabajar; y finalmente, 9 que es "No especificado".

Las clasificaciones 6 y 9 se sacaron del análisis debido a que no eran significantes en la muestra de la población analizada.

De las gráficas obtenidas se puede observar que los niños de edades entre 12 y 17 años afirman que su actividad principal es el estudio (Gráficas.II.5), lo cual concuerda con el análisis realizado en las gráficas y cuadros de las secciones anteriores. Se ha mostrado que la mayoría de la población de este grupo de edad estudia mostrando aumentos graduales con respecto al sector (de bajo hacia alto). Sin embargo, en esta gráfica es más clara la diferencia entre hombres y mujeres. Hay un porcentaje importante de mujeres del sector bajo (19.70% para ser exactos) que sostienen que su actividad principal es "Quehaceres del hogar". La gráfica de mujeres muestra que esta actividad va disminuyendo dependiendo del sector. El alto presenta una proporción muy baja de mujeres que se dedican a los quehaceres del hogar (1.89%) y en su mayoría declaran que se dedican principalmente al estudio (76.74%).

También, las gráficas (Gráficas.II.5) muestran que hay una proporción importante de esta población en todos los sectores que afirma que su actividad principal es "No trabaja" lo cual puede interpretarse como jóvenes que dedican su tiempo al ocio, lo cual no es una decisión positiva para su desarrollo personal e intelectual.

Gráficas.II.5

Actividad principal (edades de 12 a 17 años)<sup>13</sup>

Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.

Por otro lado, en la población de 18 a 24 años (Gráfica.II.6) la diferencia de actividades según el sexo es aún más marcada; ya que, "Quehaceres del hogar" muestra una proporción aún más grande en las mujeres que en la gráfica anterior sobre todo en el sector bajo en donde el porcentaje de mujeres que se dedican a esta actividad es del 40.56%.

Comparando a los hombres y las mujeres que pertenecen al sector bajo se puede apreciar una marcada diferencia de actividades; ya que los hombres en su mayoría trabajan (75.33%) y las mujeres se reparten entre el trabajo (36.88%) y las que se dedican al hogar (40.56%). Sin embargo, en ambos sexos se aprecia que son pocos los que siguen estudiando (6.51% de los hombres estudian y 7.54% para las mujeres).

En contraste, para los sectores altos, las actividades principales de las mujeres son el trabajo (40.34%) y el estudio (38.31%) y sólo el 6.51% se dedica a los quehaceres del hogar. Igualmente para los hombres, las actividades ahora se reparten principalmente entre trabajo (50.47%) y estudio (35.33%) y no únicamente en el trabajo como sucede en el sector bajo.

<sup>13</sup> La clasificación "Busca trabajo" se sacó de la gráfica debido al porcentaje bajo que representaba. La tabla con los datos, incluyendo esta clasificación, se encuentra en el anexo de este capítulo.

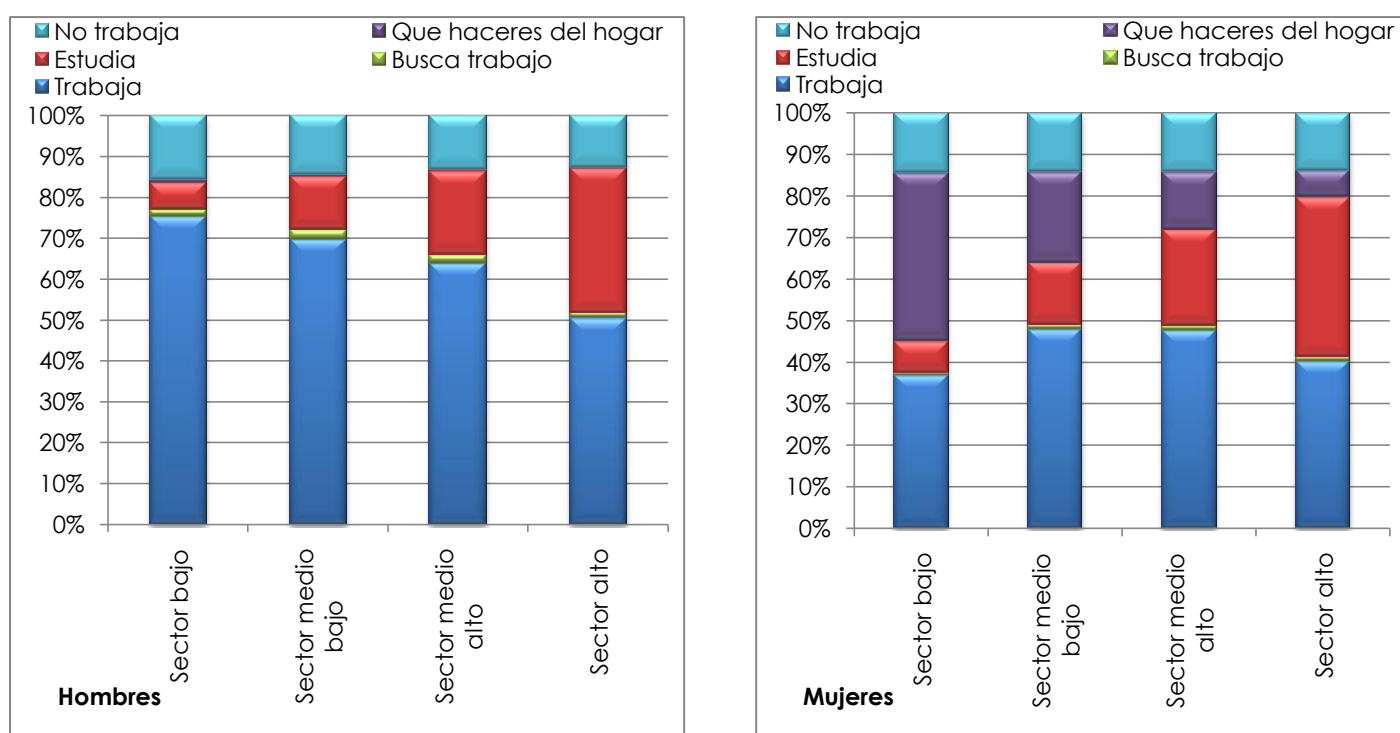


Además, estas gráficas (Gráficas.II.6) también muestran una concordancia con el análisis hecho en las secciones anteriores al mostrar un considerable aumento de personas que declaran que su principal actividad es trabajar. En los sectores altos nuevamente podemos observar que las mujeres se dedican más a estudiar que los hombres. Por lo que se confirma nuevamente que para las mujeres el seguir con sus estudios, para edades relativamente altas, tiene una estrecha relación con el sector al que pertenecen. Igualmente para los hombres, pero en menor medida.

La actividad “Busca trabajo” es una actividad que figura poco dentro de esta población ya que en su mayoría se dedica a estudiar, trabajar o al hogar (que es el caso de la mujeres como se dijo anteriormente).

**Gráficas.II.6**

**Actividad principal (edades de 18 a 24 años)**



Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.

Es importante resaltar que las gráficas anteriores (Gráficas.II.5 y Gráficas.II.6) muestran la gran diferencia que existe entre hombres y mujeres con respecto al trabajo. En general se expone que las mujeres se dedican más a estudiar que a trabajar o a los que haces del hogar para los sectores bajos y edades de 18 a 24 años; mientras que los hombres muestran el comportamiento contrario.

En los sectores bajos es en donde esta diferenciación se hace más evidente. Primeramente, en las edades de 12 a 17 años (Gráficas.II.5) los hombres de sector bajo muestran mayor proporción en la actividad “Trabaja” con un 31.22% y las mujeres con

el 12.58%. Ya para las edades de 18 a 24 años esta diferencia es más clara al presentar que el 75.33% de los varones trabajan y el 36.88% de las mujeres lo hacen.

Se conoce que la educación se ha ido universalizando y en la actualidad prácticamente toda la población joven del país asiste o asistió en algún momento de su vida a la escuela. Sin embargo, poco se sabe sobre las condiciones de las escuelas tanto en instalaciones, personal didáctico, como en contenido de las materias impartidas y libros de texto (Schmelkes, 1998). Los datos aquí presentados no muestran estas condiciones, simplemente se puede tratar de inferir lo que puede estar pasando dentro de las escuelas y los hogares a partir de estos datos numéricos.

Las diferencias entre sectores son muy marcadas, los sectores bajos tienen una escolaridad muy limitada son los que están más atrasados y los sectores altos son los que cuentan con mayor escolaridad les ayudará a tener un mejor puesto y una mejora en su bienestar y calidad de vida. A lo largo de los años ha habido cambios positivos que ha hecho que sean cada vez menos las personas que nunca han ido a la escuela; sin embargo, como se dijo anteriormente, aún queda por mejorar la calidad de la escuela como institución, de manera que se incentive a los jóvenes a seguir estudiando y se den mayores oportunidades a los estudiantes de menores recursos.

## CAPÍTULO III

### MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL PARA EXPLICAR LA CONDICIÓN EDUCATIVA DE LOS JÓVENES DE 12 A 24 AÑOS<sup>14</sup>

La regresión logística es una técnica de análisis multivariante en la que existe una variable de respuesta que depende de otras variables que pueden ser de carácter cualitativo o cuantitativo.

Este modelo se clasifica en dos tipos fundamentalmente: regresión logística simple, si en el modelo hay una sola variable dependiente; y regresión logística múltiple, si en el modelo hay varias variables independientes.

Según el objetivo de este proyecto de investigación, planteado anteriormente, se desea conocer la relación entre el sector socioeconómico y las causas de abandono (variable de respuesta) entre los jóvenes de 12 a 24 años de edad; por lo tanto, se utiliza el modelo de regresión múltiple al existir diversas variables independientes o explicativas que podrían influir en la variable de respuesta.

#### III.1 ECUACIÓN BÁSICA DEL MODELO

El modelo de regresión logística y el modelo de probabilidad lineal tienen como variables de respuesta probabilidades. Sin embargo, el modelo lineal presenta ciertos problemas:

1. La probabilidad estimada  $P$  toma valores imposibles; es decir, negativos o mayores a uno.
2. Se violan seriamente la linealidad y la homocedasticidad de los datos al tomar  $P$  valores fuera de cero o uno. Y por lo tanto, los procedimientos usuales para establecer las pruebas de hipótesis son inválidas.

Estos problemas dificultan el análisis de la relación entre las variables de respuesta y las variables explicativas del fenómeno a investigar. Es por esto que el modelo de regresión logística es utilizado frecuentemente; ya que, éste ajusta una curva en forma de "S" alargada cuyos límites están por arriba del cero y por debajo de 1; resolviendo de esta manera los problemas que presenta el modelo de probabilidad lineal.

---

<sup>14</sup> Todo este capítulo está basado en los textos de: Álvarez, Rafael (1994), "Estadística multivariada y no paramétrica con SPSS"; Retherford, R.D. y M.K. Choe (1993), "Statistical Models for Causal Analysis"; M. Mood, Alexander, Graybill, Franklin A., Boes, Duane C. (1974), "Introduction to the theory of Statistics"; Rodríguez, Jaime (2001), "Modelos sociodemográficos: Atlas social de la ciudad de Alicante"

En regresión logística, la variable dependiente es dicotómica (sí – no, estudia – no estudia, etc.) Para poder construir un modelo matemático se necesitan de valores numéricos los cuales se pueden obtener si se considera la probabilidad de que se dé un determinado valor de la variable dependiente. En este caso, se quiere plantear un modelo que permita calcular la probabilidad de que los jóvenes de entre 12 y 24 años de edad, que presenten ciertas características, decidan o no abandonar la escuela. Si se piensa en una ecuación que modele estas probabilidades entonces se tiene que el miembro de la izquierda de la expresión sería la probabilidad de que el suceso ocurra (en este caso la situación educativa representada por  $P$ ) y del lado derecho la dependencia de dicha probabilidad en función de las variables dependientes según un modelo logístico.

El modelo de interés pretende se den a conocer los factores que aumentan o disminuyen la probabilidad de que ocurra un suceso determinado. Se sabe que la probabilidad es un número que puede oscilar entre 0 y 1; por lo tanto, la relación funcional de la expresión del modelo sólo podrá tomar valores entre 0 y 1 para que sea congruente. Entre los muchos modelos que cumplen las condiciones anteriores esta el modelo logístico cuya expresión es la siguiente:

$$P = \frac{1}{1+e^{-Z}}, \dots\dots (3.1)$$

donde:

$Z$  es la variable dependiente de nuestro modelo que puede estar modelada por las variables explicativas de nuestro interés.

La función logística básica (3.1) modela el tipo de curvas en forma de “S” alargadas mencionadas anteriormente; además, es la más utilizada por su facilidad para trabajar con ella.

Si al numerador y denominador de la ecuación (3.1) los multiplicamos por  $e^Z$  entonces la función logística toma la siguiente forma:

$$P = \frac{e^Z}{1+e^Z} = \frac{\exp(Z)}{1+\exp(Z)}, \dots\dots (3.2)$$

Como se puede notar las expresiones anteriores son bivariadas para hacerlas multivariadas basta con crear un modelo en donde nuestra variable dependiente  $Z$  este explicada por otras variables; es decir,

$$Z = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k, \dots\dots (3.3)$$

donde:

$X_1, X_2, \dots, X_k$  son las  $k$  variables explicativas.

La expresión anterior (3.3) se sustituye en la ecuación (3.1) y de esta manera se obtiene la forma de la ecuación logística multinomial, que es:

$$P = \frac{1}{1 + e^{-(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k)}} \dots\dots(3.4)$$

Todas las propiedades de la función logística básica se preservan para la ecuación (3.4). La función se encuentra entre 0 y 1; y, su máximo valor de cambio ( $P=0.5$ ) para cualquier  $X_i$ .

Recordando los valores de los factores obtenidos en el capítulo I, se tiene:<sup>15</sup>

<b>Cuartiles</b>	<b>Factor rotado 1</b>	<b>Factor rotado 2</b>	<b>Factor rotado 3</b>
<b>25</b>	- 0.6504683	- 0.5936	- 0.457496
<b>50</b>	- 0.3872622	0.1677272	- 0.2580187
<b>75</b>	0.1653029	0.8296232	- 0.0010153

En este trabajo las variables dicotómicas explicativas utilizadas para explicar las causas de abandono (variable dependiente) son:<sup>16</sup>

- *DFR1c*: Variable dicotómica que devuelve el valor de 1 si el factor rotado 1 "Servicios y enseres básicos" es  $<0.1653029$ , que corresponde al sector alto, y 0 si es  $\geq 0.1653029$ , que corresponde al sector bajo, para la persona a analizar.
- *DFR2c*: Variable dicotómica que devuelve el valor de 1 si el factor rotado 2 "Enseres medios" es  $<0.8296232$ , que corresponde al sector alto, y 0 si es  $\geq 0.8296232$ , que corresponde al sector bajo, para la persona a analizar.
- *DFR3c*: Variable dicotómica que devuelve el valor de 1 si el factor rotado 3 "Condiciones de la vivienda" es  $<-0.0010153$ , que corresponde al sector alto, y 0 si es  $\geq -0.0010153$ , que corresponde al sector bajo, para la persona a analizar.
- *Sec*: Variable dicotómica que toma el valor de 1 si la persona asiste a secundaria y 0 en cualquier otro caso.

<sup>15</sup> Se partieron los factores en dos sectores únicamente: alto y bajo a partir del cuartil 75 debido a que de esta manera se ajusta mejor el modelo y las probabilidades resultantes son más fáciles de interpretar.

<sup>16</sup> La sintaxis de la programación utilizada para crear estas variables se encuentra en el Anexo II.

- *MedSup*: Variable dicotómica que toma el valor de 1 si la persona asiste a la escuela medio superior y 0 en cualquier otro caso.
- *Prof*:<sup>17</sup> Variable dicotómica que toma el valor de 1 si la persona asiste a la escuela medio superior y 0 en cualquier otro caso.
- *M*: Variable dicotómica que toma el valor de 1 si la persona es mujer y 0 si es hombre
- *E* (13 a 17 y de 19 a 24 según grupo de edad): Variable dicotómica que toma el valor de 1 si la persona tiene 13 y 0 en cualquier otro caso para E13, 1 si la persona tiene 14 y 0 en cualquier otro caso para E14, y así sucesivamente.

### III.II PRUEBAS DE HIPÓTESIS Y DE AJUSTE DEL MODELO

En la expresión (3.3) se pueden observar los coeficientes del modelo representados por  $b_0, b_1, b_2, \dots, b_k$ . El primero de ellos ( $b_0$ ) es el término independiente y puede ser igual a cero; mientras que, los demás coeficientes ( $b_1, b_2, \dots, b_k$ ) son los coeficientes de regresión logística y por lo menos uno de ellos debe ser significativamente distinto de cero. Si lo anterior no sucede no se tiene evidencia de que las variables explicativas estén relacionadas con la variable de respuesta; y por lo tanto, el modelo no tendría sentido.

El signo de los coeficientes es importante ya que si alguno es positivo eso significa que la variable explicativa, por ejemplo edad, aumenta la probabilidad del suceso (en este caso situación educativa). Por el contrario si es negativo, el factor cuyo coeficiente es negativo disminuye la probabilidad del suceso.

Para conocer si tiene sentido aplicar el modelo de regresión logística, primeramente se calculan los estimadores de estos coeficientes mediante el método de Máxima Verosimilitud. Posteriormente se prosigue a evaluar estos estimadores por medio de pruebas de hipótesis.

La hipótesis que se quiere contrastar es si las variables independientes influyen significativamente en la probabilidad del suceso de estudio. Las hipótesis conceptuales serían:

---

<sup>17</sup> Esta variable solo se utilizó como variable explicativa de la regresión logística multinomial para las edades de 18 a 24 años; ya que para las edades de 12 a 17 años se presentan 0 casos de jóvenes que asisten a educación profesional.

$H_0$  Las variables explicativas no influyen significativamente sobre la variable de respuesta.

$H_1$  Las variables explicativas influyen significativamente sobre la variable de respuesta.

Matemáticamente las hipótesis anteriores se plantean de la siguiente manera:

$$\begin{array}{ll} H_0 & b_1 = b_2 = \dots = b_k = 0 \\ H_1 & b_i \neq 0 \text{ para alguna } i \in k \end{array}$$

Es decir, al menos un coeficiente de regresión logística  $b_i$  debe ser estadísticamente distinto de cero para que exista influencia significativa en la variable dependiente.

Para este proyecto, se utilizaron la Evaluación del estadístico  $-2\log L_0$  y la prueba de la Pseudo- $R^2$ .

### **EVALUACIÓN DEL ESTADÍSTICO $-2\log L_0$**

Esta prueba comienza conociendo el valor de este estadístico si el modelo considera sólo el término constante  $b_0(\delta_0)$ . Este estadístico  $\delta_0$  se distribuye como una Chi-Cuadrada con  $n-1$  grados de libertad donde  $n$  es el tamaño de la muestra. A continuación se calcula el valor del estadístico nuevamente, pero con todas las variables explicativas incluidas en el modelo ( $\delta_1$ ). En este caso los grados de libertad son  $n-K-1$  donde  $K$  es el número de variables explicativas incluidas. Se calcula la diferencia entre los dos valores del estadístico y esta diferencia indica que tanto se está ajustando el modelo a nuestras variables. Entre más grande sea la diferencia y el valor del estadístico con todas las variables explicativas incluidas sea más chico, mejor se ajustará el modelo.

Para cada uno de los grupos de edad se obtuvieron los siguientes resultados de la evaluación del estadístico  $-2\log L_0$ :

### Para el grupo de edad de 12 a 17 años

En este caso  $\delta_0 = 202,895.386$  y la diferencia es  $\Delta = 190,883.19$ ; por lo que  $\delta_1 = 12,012.196$  la cual se distribuye como una Chi-Cuadrada con 66 grados de libertad.

Se puede notar gracias a la diferencia obtenida y al valor reducido de  $\delta_1$  comparado con  $\delta_0$  que el modelo se redujo queriendo decir que el modelo se ajusta. Entonces se rechaza la hipótesis nula y concluimos que las variables explicativas influyen significativamente en las variables resultantes.

**Tabla.III.I Información de ajuste del modelo edades de 12 a 17 años**

Información de ajuste del modelo				
Modelo	Criterio de ajuste del modelo	Prueba de la razón de Verosimilitud		
	-2 Log Verosimilitud	Chi-Cuadrada	gl	Sig.
Solo con $\delta_0$	202,895.386			
Final	12,012.196	190,883.19	66	.000
Elaboración propia con SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI				

### Para el grupo de edad de 18 a 24 años

En este caso  $\delta_0 = 252,670.705$  y la diferencia es  $\Delta = 233,666.536$ ; por lo que

$\delta_1 = 19,004.169$  la cual se distribuye como una Chi-Cuadrada con 78 grados de libertad.

Se puede notar gracias a la diferencia obtenida y al valor reducido de  $\delta_1$  comparado con  $\delta_0$  que el modelo se redujo queriendo decir que el modelo se ajusta. Entonces se rechaza la hipótesis nula y concluimos que las variables explicativas influyen significativamente en las variables resultantes.

**Tabla.III.II Información de ajuste del modelo edades de 18 a 24 años**

Información de ajuste del modelo				
Modelo	Criterio de ajuste del modelo	Prueba de la razón de Verosimilitud		
	-2 Log Verosimilitud	Chi-Cuadrada	gl	Sig.
Solo con $\delta_0$	252,670.705			
Final	19,004.169	233,666.536	78	.000
Elaboración propia con SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI				



## PRUEBA DE LA $R^2$

En regresión múltiple, el indicador más frecuente para medir la bondad de ajuste del modelo es la  $R^2$  la cual mide la proporción de la varianza de la variable de respuesta (situación educativa) explicada por las variables independientes.

Las variables de respuesta en el caso de la regresión logística varían entre 0 y 1 por ser probabilidades; por lo que es difícil ajustar a la perfección un modelo y el máximo valor que puede alcanzar  $R^2$  en el modelo depende de la media y la varianza de las probabilidades obtenidas en el modelo.

Una alternativa para calcular  $R^2$  es utilizando una ecuación que se deriva por medio del método de máxima verosimilitud. Algunas de las expresiones generalmente más utilizadas por los programas estadísticos son:

$$\text{pseudo} - R^2 = 1 - (\log L_0 / \log L_1)^{2/n},$$

$$\text{pseudo} - R^2 = \frac{2\log L_1 - 2\log L_0 - 2k}{-2\log L_0},$$

donde:

$n$  es el tamaño de la muestra.

$k$  es el número de coeficientes estimados.

$-2\log L_0 - 2\log L_1$  se distribuye como una Chi-Cuadrada.

En este caso para la  $R^2$  se obtuvieron los siguientes valores:

**Tabla.III.IV Información de la Prueba de la  $R^2$**

Pseudo R-Cuadrada		
	Gpo de edad 12 a 17 años	Gpo de edad 18 a 24 años
Cox and Snell	.306	.523
Nagelkerke	.405	.550
McFadden	.260	.246
Elaboración propia con SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI		

En general, en ambos casos se aprecia que la prueba indica que las variables resultantes (causas de abandono) se ajustan en buena medida al modelo de regresión logística obtenido. Esto es de interés debido a que se requiere saber que tan apegado esta el modelo a las variables, qué tanto explica su comportamiento; así como la relación que guardan las variables independientes con la variables explicativas que puedan ocasionar que las probabilidades de ocurrencia de las variables de respuesta se incrementen o disminuyan.

### III.III ANÁLISIS DE LOS RESULTADOS OBTENIDOS

En la sección anterior se ha visto que analizar la significancia de los coeficientes es importante ya que con ello podemos conocer el ajuste de las variables al modelo que se obtuvo. Sin embargo, además de la significancia estadística hecha con pruebas de hipótesis, en una segunda etapa se debe estudiar la fuerza de asociación estadística. Esto se mide con el ODDS ratio o razón de predominio y con el Log de P.

#### ODDS Y LOGIT DE P

El *ODDS* y el *logit de P* (Choe, 1993) se obtienen a partir de la ecuación básica del modelo de regresión logística visto anteriormente:

$$P = \frac{1}{1 + e^{-Z}} \dots\dots (3.1 \text{ repetida})$$

De aquí se sigue que:

$$1 - P = 1 - \frac{1}{1 + e^{-Z}} = \frac{e^{-Z}}{1 + e^{-Z}} \dots\dots (3.5)$$

Dividiendo (3.1) entre (3.5) se obtiene:

$$\frac{P}{1 - P} = e^Z \dots\dots (3.6)$$

Sacando el logaritmo natural de ambos lados de la ecuación (3.6) se obtiene,

$$\log \frac{P}{1 - P} = Z \dots\dots (3.7)$$

A la cantidad  $P/(1-P)$  se le llama *ODDS* y se le denota con una  $\Omega$ ; mientras que, al logaritmo natural se le denomina *LOGIT DE P*; es decir,

$$\text{Odds} = \frac{P}{1-P} = \Omega \dots\dots (3.8)$$

Y

$$\text{logit}P = \log \frac{P}{1-P} = \log \Omega \dots\dots (3.9)$$

Con estas definiciones y junto con la expresión (3.3), la expresión (3.7) puede escribirse como:

$$\text{logit}P = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k \dots\dots (3.10)$$

$$\log \frac{P}{1-P} = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k \dots\dots (3.11)$$

$$\log \Omega = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k \dots\dots (3.12)$$

Las ecuaciones (3.10)-(3.12) ahora se encuentran en la forma más familiar y fácil de interpretar al tener la misma forma que una ecuación de regresión múltiple. Además, éstas son las que se van a utilizar para obtener el ODDS ratio el cual es explicado a continuación.

## ODDS RATIO

En regresión logística, la medida de asociación más utilizada es el *ODDS Ratio* por su sencillez de cálculo. Para obtenerlo se realiza lo siguiente:

Se tiene la expresión (3.12) a la cual se le aplica la función e de ambos lados para obtener:

$$\Omega = e^{b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k} \dots\dots (3.13)$$

Suponiendo que se quisiera incrementar en una unidad alguna de las variables explicativas  $X_1, X_2, \dots, X_k$  dejando las demás constantes, por ejemplo  $X_1$  entonces de la ecuación (3.13) se tiene:

$$\Omega^* = e^{b_0 + b_1(X_1+1) + b_2X_2 + \dots + b_kX_k}$$

$$\Leftrightarrow \Omega^* = e^{b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + b_1}$$

$$\Leftrightarrow \Omega^* = e^{b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k} e^{b_1}$$

$$\Leftrightarrow \Omega^* = \Omega e^{b_1} \dots\dots (3.14)$$

$$\Leftrightarrow \frac{\Omega^*}{\Omega} = e^{b_1} \dots\dots(3.15)$$

De la expresión (3.14) se aprecia que es evidente que un incremento de la variable explicativa  $X_1$ , dejando las demás variables constantes, provoca que aumente o disminuya la probabilidad (ODDS) de ocurrencia de la variable resultante a razón del factor  $e^{b_1}$ . Esta cantidad ( $e^{b_1}$ ) es lo que se conoce como *ODDS Ratio*. Por ejemplo, para la deducción de la expresión (3.15) se utilizó la variable  $X_1$ ; por lo tanto, el factor  $e^{b_1}$  representa el efecto multiplicativo sobre la variable dependiente (o resultante) al incrementar en una unidad la variable explicativa  $X_1$ .

El efecto de aumentar en una unidad alguna de las variables independientes o explicativas también se puede analizar como un efecto aditivo. A partir de la ecuación (3.12), la cual tiene la forma de una ecuación de regresión lineal, el efecto de incrementar en una unidad la variable  $X_1$  (con todas las demás variables explicativas constantes) se expresa de la siguiente manera:

$$\log \Omega^* = \log \Omega + b_1 \dots\dots(3.16)$$

Si aplicamos la función  $e$  en ambos lados de la ecuación (3.16) tenemos nuevamente:

$$\Omega^* = \Omega e^{b_1} \dots\dots(3.17)$$

Por las deducciones de las expresiones (3.15) y (3.17) se puede ver claramente las diferencias entre los efectos multiplicativos y aditivos, los cuales llevan al mismo resultado. Sin embargo, el *ODDS ratio* es el más utilizado debido a que se obtiene de manera relativamente más sencilla y sobre todo por ser fácil de interpretar intuitivamente; es decir, obedece de mejor manera al sentido común.

### III.III.I ECUACIÓN DEL MODELO

Todos los resultados anteriores obtenidos se basan en la ecuación básica del modelo de regresión; sin embargo, la naturaleza de la variable de respuesta en este estudio no sólo se basa en una categoría de respuesta si no en varias. Es decir, la variable de respuesta "Situación educativa" engloba varias categorías: Asistencia, no quiso o no le gustó estudiar, falta de dinero o tenía que trabajar, se casó o unió, no lo dejaron o

por dedicarse a las tareas del hogar, terminó sus estudios u otro motivo. Esto quiere decir que no es suficiente la utilización de una sola ecuación para expresar la probabilidad de ocurrencia de la variable de respuesta, si no que se necesitan de 7 ecuaciones que expliquen las diferentes salidas de la variable de respuesta expuestas anteriormente.

Por lo tanto, para el caso de la variable de respuesta "Situación educativa" se considera el modelo general de *regresión logística multinomial* (Choe, 1993) el cual es utilizado para las variables de respuesta que contengan diversas categorías mutuamente excluyentes.

Las expresiones generales de este modelo son:

$$\log \frac{P_j}{P_J} = \sum_k b_{jk} X_k \text{ con } j=1,2,\dots,J-1, \dots (3.18)$$

$$\sum_j P_j = 1, \dots (3.19)$$

donde:

$$\sum_k \text{ con } k=0 \text{ hasta, } k=K.$$

$$\sum_j \text{ con } j=0 \text{ hasta } j=J.$$

Y la probabilidad  $P_J$  (categoría de referencia) está dada por la expresión siguiente:

$$P_J = \frac{1}{1 + \sum_j e^{\sum_k b_{jk} X_k}} \dots (3.20)$$

Por conveniencia se define  $X_0 = 1$  y así se incluya la intersección  $b_{j0} = b_{j0} X_0$  de la ecuación de regresión.

Se puede notar que estas expresiones son las mismas que las expuestas anteriormente, con la diferencia de que efectivamente éstas incluyen las diferentes categorías que tiene una variable de respuesta.

Con las expresiones anteriores y haciendo uso de SPSS se obtienen las siguientes expresiones que modelan con regresión logística multinomial la variable "Situación educativa" para los dos grupos de edad:

## Para el grupo de edad de 12 a 17 años

Ecuación del modelo de regresión logística multinomial (3.21)

$$\log \frac{P_i}{P_7} = \begin{pmatrix} -3.129 & -0.177 & -0.348 & -0.241 & -2.465 & -5.450 & -0.082 & 1.413 & 2.633 & 3.820 & 4.409 & 5.774 \\ -3.145 & -0.832 & -0.652 & -0.510 & -2.290 & -5.350 & 0.211 & 1.449 & 2.553 & 3.955 & 5.156 & 6.131 \\ -10.378 & 0.066 & -0.364 & -0.462 & -1.649 & -4.269 & 2.028 & .762 & 3.549 & 4.796 & 6.531 & 7.766 \\ -6.155 & -0.840 & -0.138 & -0.123 & -2.733 & -6.018 & 1.202 & 1.328 & 2.402 & 3.576 & 4.576 & 5.416 \\ -5.907 & -0.902 & -0.050 & -0.143 & -1.620 & -3.520 & 0.300 & 1.166 & 2.014 & 3.500 & 4.707 & 5.685 \\ -6.110 & 0.528 & -0.098 & -0.103 & -1.742 & -3.449 & 0.028 & 1.039 & 1.879 & 3.334 & 4.203 & 4.754 \end{pmatrix} \begin{pmatrix} 1 \\ DFR1c \\ DFR2c \\ DFR3c \\ Sec \\ MedSup \\ M \\ E13 \\ E14 \\ E15 \\ E16 \\ E17 \end{pmatrix}$$

con  $i=1, \dots, 6$

$$P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 = 1$$

donde:

$P_1 = P(Y = \text{No quiso o no le gustó estudiar})$

$P_2 = P(Y = \text{Faltade dinero o teníaque trabajar})$

$P_3 = P(Y = \text{Se casó o unió})$

$P_4 = P(Y = \text{No lo dejaron por encargarse del hogar})$

$P_5 = P(Y = \text{Terminó sus estudios})$

$P_6 = P(Y = \text{Otro motivo})$

$P_7 = P(Y = \text{Asisten a la escuela})$  la cual es categoría de referencia.

Es de interés saber, además, los valores del ODDS ratio para cada una de las variables explicativas en las distintas categorías de respuesta, recordando que estos valores representan el efecto multiplicativo sobre la propensión de ocurrencia de la categoría que se desee analizar de la variable "Situación educativa". En este caso como todos los números son mayores el efecto es que todas las probabilidades aumentan.

Tabla.III.V Valores de ODDS ratio  $\exp(b_{jk})$

$\exp(b_{jk})$ edades de 12 a 17 años						
Variables explicativas	Causa de abandono					
	No quiso o no le gustó estudiar	Falta de dinero o tenía que trabajar	Se casó o unió	No lo dejaron o por encargarse del hogar	Terminó sus estudios	Otro motivo
DFR1c	0.838	0.435	1.069	0.431	0.406	1.695
DFR2c	0.706	0.521	0.695	0.871	0.951	0.907
DFR3c	0.786	0.600	0.630	0.884	0.867	0.903
Sec	0.085	0.101	0.192	0.065	0.198	0.175
MedSup	0.004	0.005	0.014	0.002	0.030	0.032
M	0.922	1.236	7.600	3.327	1.350	1.029
E13	4.110	4.259	2.142	3.773	3.209	2.825
E14	13.919	12.844	34.780	11.047	7.497	6.548
E15	45.609	52.206	120.973	35.716	33.128	28.053
E16	134.844	173.458	685.847	97.099	110.668	66.862
E17	321.738	460.122	2360.083	224.889	294.516	116.103
a. La categoría de referencia es: Asisten.						
Elaboración propia con SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI						

### Para el grupo de edad de 18 a 24 años

Ecuación del modelo de regresión logística mutlinomial (3.22)

$$\log \frac{P_i}{P_7} = \begin{pmatrix} 1.958 & 0.160 & -0.281 & 0.022 & -1.519 & -4.236 & -7.866 & -0.303 & 0.594 & 0.999 & 1.311 & 1.636 & 2.038 & 2.425 \\ 2.271 & -0.360 & -0.641 & -0.245 & -1.156 & -3.480 & -6.882 & -0.084 & 0.689 & 1.200 & 1.575 & 1.975 & 2.401 & 2.815 \\ -2.373 & 0.228 & -0.389 & -0.410 & -0.507 & -2.580 & -5.700 & 1.267 & 0.879 & 1.457 & 1.977 & 2.427 & 2.864 & 3.395 \\ -1.254 & -0.579 & -0.286 & 0.198 & -1.840 & -4.277 & -7.344 & 0.939 & 0.625 & 1.128 & 1.466 & 1.772 & 2.283 & 2.684 \\ -1.614 & -0.156 & -0.162 & 0.016 & -0.434 & -0.869 & -1.987 & 0.560 & 0.663 & 1.114 & 1.535 & 2.121 & 2.712 & 3.228 \\ -1.346 & 0.519 & -0.101 & 0.066 & -1.694 & -2.609 & -4.842 & 0.098 & 0.511 & .714 & 0.967 & 1.210 & 1.490 & 1.828 \end{pmatrix} \begin{pmatrix} 1 \\ DFR1c \\ DFR2c \\ DFR3c \\ Sec \\ MedSup \\ Prof \\ M \\ E19 \\ E20 \\ E21 \\ E22 \\ E23 \\ E24 \end{pmatrix}$$

con  $i=1, \dots, 6$

$$P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 = 1$$

donde:

$$P_1 = P(Y = \text{No quiso o no le gustó estudiar})$$

$$P_2 = P(Y = \text{Falta de dinero o tenía que trabajar})$$

$$P_3 = P(Y = \text{Se casó o unió})$$

$$P_4 = P(Y = \text{No lo dejaron o por encargarse del hogar})$$

$P_5 = P(Y = \text{Terminó sus estudios})$

$P_6 = P(Y = \text{Otro motivo})$

$P_7 = P(Y = \text{Asisten a la escuela})$  la cual es la categoría de referencia.

Es de interés saber, al igual que en las edades de 12 a 17 años, los valores del ODDS ratio para cada una de las variables explicativas en las distintas categorías de respuesta, recordando que estos valores representan el efecto multiplicativo sobre la probabilidad de ocurrencia de la categoría que se desee analizar de la variable "Situación educativa". En este caso como todos los números son mayores el efecto es que todas las probabilidades aumentan.

**Tabla.III.VI Valores de ODDS ratio**

<b>exp(b<sub>jk</sub>) edad de 18 a 24 años</b>						
<b>Variables explicativas</b>	<b>Causa de abandono</b>					
	<b>No quiso o no le gustó estudiar</b>	<b>Falta de dinero o tenía que trabajar</b>	<b>Se casó o unió</b>	<b>No lo dejaron o por encargarse del hogar</b>	<b>Terminó sus estudios</b>	<b>Otro motivo</b>
DFR1c	1.173	0.698	1.256	0.561	0.856	1.681
DFR2c	0.755	0.527	0.678	0.752	0.851	0.904
DFR3c	1.023	0.783	0.664	1.219	1.016	1.068
Sec	0.219	0.315	0.602	0.159	0.648	0.184
MedSup	0.014	0.031	0.076	0.014	0.419	0.074
Prof	0.000	0.001	0.003	0.001	0.137	0.008
M	0.738	0.919	3.551	2.556	1.751	1.103
E19	1.812	1.991	2.409	1.869	1.941	1.667
E20	2.716	3.321	4.294	3.088	3.045	2.043
E21	3.711	4.829	7.220	4.333	4.641	2.631
E22	5.137	7.206	11.328	5.880	8.341	3.352
E23	7.678	11.030	17.531	9.810	15.065	4.438
E24	11.305	16.701	29.806	14.640	25.227	6.221
a. The reference category is: Asisten.						
Elaboración propia con SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI						



### III.III.II RESULTADOS DEL MODELO APLICADO

Para presentar los resultados del análisis logístico multivariado se hace uso de una tabla que muestre las probabilidades  $P_1, P_2, \dots, P_7$ , este cuadro lleva el nombre de *Formato de Análisis de Clasificación Múltiple* (MCA por sus siglas en inglés)

Para la elaboración de esta tabla (Choe, 1993) se hizo lo siguiente<sup>18</sup>:

Se saben las ecuaciones del modelo logístico multinomial para las edades de 12 a 17; la matriz de  $6 \times 12$  son los coeficientes del vector de las variables explicativas (VE):

$$VE^T = (1 \text{ DFR1c DFR2c DFR3c Sec MedSup M E13 E14 E15 E16 E17})$$

Nuevamente se exponen las ecuaciones del modelo de regresión (3.21):

Ecuación (3.21.a)

$$\log \frac{P_i}{P_7} = \begin{pmatrix} -3.129 & -0.177 & -0.348 & -0.241 & -2.465 & -5.450 & -0.082 & 1.413 & 2.633 & 3.820 & 4.409 & 5.774 \\ -3.145 & -0.832 & -0.652 & -0.510 & -2.290 & -5.350 & 0.211 & 1.449 & 2.553 & 3.955 & 5.156 & 6.131 \\ -10.378 & 0.066 & -0.364 & -0.462 & -1.649 & -4.269 & 2.028 & .762 & 3.549 & 4.796 & 6.531 & 7.766 \\ -6.155 & -0.840 & -0.138 & -0.123 & -2.733 & -6.018 & 1.202 & 1.328 & 2.402 & 3.576 & 4.576 & 5.416 \\ -5.907 & -0.902 & -0.050 & -0.143 & -1.620 & -3.520 & 0.300 & 1.166 & 2.014 & 3.500 & 4.707 & 5.685 \\ -6.110 & 0.528 & -0.098 & -0.103 & -1.742 & -3.449 & 0.028 & 1.039 & 1.879 & 3.334 & 4.203 & 4.754 \end{pmatrix} \begin{pmatrix} 1 \\ \text{DFR1c} \\ \text{DFR2c} \\ \text{DFR3c} \\ \text{Sec} \\ \text{MedSup} \\ \text{M} \\ \text{E13} \\ \text{E14} \\ \text{E15} \\ \text{E16} \\ \text{E17} \end{pmatrix}$$

con  $i=1, \dots, 6$

$$P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 = 1, \dots, \dots (3.21.b)$$

El lado derecho de la ecuación (3.21.a) se denominará con el nombre de  $\Delta_i$  donde  $i=1, \dots, 6$  para referirse al número de renglones. Así las ecuaciones del modelo (3.21) ahora tienen las siguientes expresiones:

$$\log \frac{P_i}{P_7} = \Delta_i \text{ para cada } i=1, \dots, 6 \dots (3.21.c)$$

$$P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 = 1, \dots, \dots (3.21.d)$$

Aplicando la función  $e$  en ambos lados de la ecuación (3.21.c) se tiene que:

$$P_i = P_7 e^{\Delta_i} \dots (3.22) \text{ para cada } i=1, \dots, 6$$

<sup>18</sup> Esta sección se mostrará con las edades de 12 a 17, para el caso de 18 a 24 se aplicó el mismo procedimiento.

Utilizando la expresión (3.21.d) y la (3.22) se obtiene:

$$1 = \sum_{i=1}^7 P_i \quad \text{y} \quad \sum_{i=1}^7 P_i = P_7 \sum_{i=1}^6 e^{\Delta_i} + P_7$$

$$\Rightarrow 1 = P_7 \sum_{i=1}^6 e^{\Delta_i} + P_7$$

$$\Rightarrow P_7 = \frac{1}{1 + \sum_{i=1}^6 e^{\Delta_i}}$$

∴ Al sustituir el nuevo valor de  $P_7$  en la expresión (3.22) obtenemos los valores de la probabilidad de ocurrencia para cada categoría de la variable explicativa: causa de abandono.

Expresiones de las probabilidades para las diferentes causas de abandono (3.23)

$$P_i = \frac{e^{\Delta_i}}{1 + \sum_{i=1}^6 e^{\Delta_i}} \quad \text{para cada } i=1, \dots, 6$$

Ahora bien, la tabla MCA se construye sustituyendo combinaciones de ceros, unos y valores medios adecuadamente. Por ejemplo, si queremos la probabilidad  $P_1$  (es decir, la probabilidad de abandonar la escuela porque no quiso o no le gustó estudiar) para todos aquellos jóvenes de 12 a 17 que asisten a Secundaria entonces en la ecuación (3.23) en el vector columna VE de  $\Delta_1$  se sustituye Sec=1, MedSup=0,  $FR1 = \overline{FR1}$ ,  $FR2 = \overline{FR2}$ ,  $FR3 = \overline{FR3}$ ,  $M = \overline{M}$ ,  $E13 = \overline{E13}$ ,  $E14 = \overline{E14}$ ,  $E15 = \overline{E15}$ ,  $E16 = \overline{E16}$ ,  $E17 = \overline{E17}$

## TABLAS DE ANÁLISIS DE CLASIFICACIÓN MÚLTIPLE (Tablas de probabilidades)

Tabla.III.VII Para el grupo de edad de 12 a 17 años

Variables explicativas <sup>19</sup>	n	P1	P2	P3	P4	P5	P6	P7	Total
		No quiso o no le gustó estudiar	Falta de dinero o tenía que trabajar	Se casó o unió	No lo dejaron o por encargarse del hogar	Terminó sus estudios	Otro motivo	Asisten a la escuela	
Factor 1 SB	150,722	6.83%	6.50%	0.04%	0.52%	0.96%	0.49%	84.66%	100.00%
Factor 1 SA	396,810	6.04%	2.99%	0.04%	0.24%	0.41%	0.89%	89.39%	100.00%
Factor 2 SB	147,265	7.65%	5.83%	0.05%	0.34%	0.56%	0.74%	84.84%	100.00%
Factor 2 SA	400,267	5.69%	3.21%	0.04%	0.31%	0.56%	0.71%	89.49%	100.00%
Factor 3 SB	146,885	7.21%	5.38%	0.06%	0.34%	0.60%	0.75%	85.67%	100.00%
Factor 3 SA	400,648	5.90%	3.36%	0.04%	0.31%	0.54%	0.70%	89.15%	100.00%
Primaria	250,939	24.45%	14.01%	0.10%	1.48%	1.17%	1.57%	57.23%	100.00%
Secundaria	287,358	3.39%	2.31%	0.03%	0.16%	0.38%	0.45%	93.29%	100.00%
Medio Superior	74,413	0.18%	0.12%	0.00%	0.01%	0.06%	0.09%	99.55%	100.00%
Hombre	320,880	6.63%	3.61%	0.02%	0.18%	0.49%	0.71%	88.36%	100.00%
Mujer	305,382	6.05%	4.42%	0.12%	0.59%	0.65%	0.72%	87.45%	100.00%
Edad 12	126,816	0.43%	0.25%	0.00%	0.03%	0.05%	0.08%	99.16%	100.00%
Edad 13	117,959	1.73%	1.02%	0.00%	0.10%	0.15%	0.23%	96.77%	100.00%
Edad 14	107,464	5.49%	2.88%	0.05%	0.26%	0.33%	0.49%	90.49%	100.00%
Edad 15	101,122	14.41%	9.39%	0.13%	0.68%	1.18%	1.68%	72.52%	100.00%
Edad 16	88,075	27.16%	19.89%	0.46%	1.19%	2.52%	2.55%	46.23%	100.00%
Edad 17	84,826	36.15%	29.43%	0.88%	1.53%	3.74%	2.47%	25.79%	100.00%
Total	626,262	77,650	56,603	1,340	3,470	7,185	6,942	473,072	626,262

Elaboración propia con SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI

Gracias a este cuadro (Tabla.III.VII) se pueden apreciar las diferentes probabilidades de las causas de abandono para los jóvenes de 12 a 17 años para cada una de las categorías de las variables explicativas.

Se puede ver nuevamente como la probabilidad de asistir es alta para los jóvenes de 12 a 17 años debido a que son las edades en las que normalmente se asiste a alguna institución educativa. Sin embargo para los factores se puede ver que a pesar de que en ambos sectores la mayoría de la población tiene probabilidades altas de asistir a la escuela existen diferencias entre los sectores altos y bajos. Las variables: Factor 1 SB, Factor 2 SB y Factor 3 SB, que se refieren a los sectores bajos, son los que presentan las probabilidades bajas de asistencia con respecto a los sectores altos (alrededor del 85% contra un 90% aproximadamente para los sectores altos). Con lo anterior se comprueba nuevamente que los sectores bajos tienen menos posibilidades de asistir a la escuela.

<sup>19</sup> Las iniciales SB y SA de los factores se refieren al sector bajo y al sector alto, respectivamente.

Para la parte de las causas de abandono, se ven las diferencias entre los sectores sobre todo para "Falta de dinero o tenía que trabajar" en la cual los sectores bajos son los que presentan nuevamente probabilidades más altas que el sector alto (aproximadamente 6% para los sectores bajos contra un aproximado de 3% para los sectores altos). Se puede apreciar que para este grupo de edad los tres factores se comportan de manera muy semejante y dan información congruente con el análisis descriptivo realizado anteriormente; así como con las características que se esperan de cada sector.

Para la parte de nivel de estudios alcanzados, los tres niveles analizados tienen probabilidades significativas de asistencia; es decir, hay probabilidades relativamente altas de asistencia para este grupo edad. Sin embargo, la diferencia entre primaria y los otros dos niveles es significativa siendo la educación medio superior la que presenta la probabilidad más alta.

Respecto a las causas de abandono en especial las de "No quiso o no le gustó estudiar" y "Falta de dinero o tenía que trabajar" se aprecia como las probabilidades más altas de abandono corresponden a las personas que cuentan solo con educación primaria (24.45% y 14.01% respectivamente) y se nota el considerable decremento de las probabilidades para la educación secundaria (3.39% para "No quiso o no le gustó estudiar" y 2.31% para "Falta de dinero o tenía que trabajar" ) y para la educación medio superior (0.18% para "No quiso o no le gustó estudiar" y 0.12% para "Falta de dinero o tenía que trabajar" ) con respecto a la educación primaria.

Con respecto al sexo, se aprecia como las mujeres tienen probabilidades ligeramente más altas para las causas de "Se casó o unió" y la de "No lo dejaron o por encargarse del hogar" que los hombres. Sin embargo, para este grupo de edad, estas causas no alcanzan ni siquiera el 1% de probabilidad lo que quiere decir que estas ya no son causas tan representativas para este grupo de edad y han dejado de provocar que los jóvenes, principalmente las mujeres, dejen la escuela.

Las probabilidades de las diferentes edades también presentan diferencias. Para la categoría de referencia "Asisten a la escuela" las probabilidades más altas se presentan para la edad 12 (99.16%) y van decreciendo según vaya aumentando la edad (25.79% para la edad 17). Por otro lado pasa lo contrario para las causas de abandono, todas las causas de abandono van incrementando sus probabilidades conforme aumenta la edad. El salto se aprecia sobre todo a partir de la edad 15 en

adelante; por lo que podemos decir que conforme se aumenta la edad las probabilidades de asistencia a alguna institución educativa van bajando debido a que los jóvenes no tienen el apoyo económico para seguir con estudios que necesitan de mayor inversión de recursos; o porque presentan poco interés en el estudio debido a que probablemente no ven a las instituciones educativas como buenas oportunidades de desarrollo, como se mencionó anteriormente en el capítulo II; o porque se casan o se dedican a los quehaceres del hogar.

Para la parte de "Terminó sus estudios" se aprecia que en general las probabilidades son bajas, es decir, existe poca oportunidad de que los jóvenes terminen su ciclo educativo, mínimo de educación básica. Por la tabla se puede ver que la probabilidad más alta de terminar los estudios corresponde a los jóvenes de 17 años (3.74%).

**Tabla.III.VIII Para el grupo de edad de 18 a 24 años<sup>20</sup>**

Variables explicativas	n	P1	P2	P3	P4	P5	P6	P7	Total
		No quiso o no le gustó estudiar	Falta de dinero o tenía que trabajar	Se casó o unió	No lo dejaron o por encargarse del hogar	Terminó sus estudios	Otro motivo	Asisten a la escuela	
Factor 1 SB	74,367	18.42%	36.19%	1.83%	1.73%	12.61%	1.97%	27.25%	100.00%
Factor 1 SA	278,458	23.50%	27.81%	2.55%	1.04%	11.77%	3.60%	29.74%	100.00%
Factor 2 SB	77,825	21.62%	37.74%	2.43%	1.22%	10.89%	2.62%	23.48%	100.00%
Factor 2 SA	275,000	21.94%	27.17%	2.25%	1.21%	12.52%	3.19%	31.72%	100.00%
Primaria	77,817	47.85%	40.53%	1.62%	2.62%	2.34%	2.39%	2.66%	100.00%
Secundaria	109,336	36.30%	43.05%	3.23%	1.46%	5.24%	1.53%	9.19%	100.00%
Medio Sup	122,468	11.85%	20.41%	1.95%	0.64%	16.75%	3.03%	45.37%	100.00%
Profesional	72,471	0.60%	1.30%	0.16%	0.06%	10.48%	0.62%	86.78%	100.00%
Hombre	209,373	25.32%	31.38%	1.28%	0.78%	9.26%	2.89%	29.09%	100.00%
Mujer	185,988	18.21%	28.15%	4.43%	1.95%	15.81%	3.10%	28.35%	100.00%
Edad 18	79,060	14.30%	16.07%	0.92%	0.70%	5.89%	2.63%	59.49%	100.00%
Edad 19	62,242	18.98%	23.34%	1.61%	0.96%	8.37%	3.21%	43.54%	100.00%
Edad 20	61,024	21.52%	29.33%	2.16%	1.20%	9.92%	2.98%	32.89%	100.00%
Edad 21	48,758	22.81%	32.94%	2.80%	1.31%	11.70%	2.97%	25.46%	100.00%
Edad 22	53,904	23.01%	35.82%	3.20%	1.30%	15.34%	2.76%	18.57%	100.00%
Edad 23	46,735	23.53%	37.46%	3.39%	1.48%	18.94%	2.50%	12.70%	100.00%
Edad 24	43,639	23.57%	38.46%	3.90%	1.51%	21.55%	2.38%	8.63%	100.00%
Total	395,361	81,067	114,794	9,424	4,568	48,148	11,040	126,321	395,361

Elaboración propia con SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI

Gracias a este cuadro de probabilidades (Tabla.III.VIII) se pueden vislumbrar las características y comportamientos del grupo de edad de 18 a 24 años debido a que las probabilidades de asistencia y de las diferentes causas de abandono cambian en comparación con la tabla correspondiente a las edades de 12 a 17 años.

En general se puede notar que para este grupo de edad las otras causas de abandono aparte de "No quiso o no le gustó estudiar" y "Falta de dinero o tenía que trabajar" ya tienen más relevancia y probabilidades más altas que para las edades de 12 a 17 años en donde estas dos causas son las más relevantes. Incluso para las edades de 18 a 24 años se ve como "Otro motivo" presenta probabilidades más altas que para el grupo de los jóvenes de la tabla anterior.

Para ambos factores se destaca que la "Falta de dinero o tenía que trabajar" es más significativa para los sectores bajos, lo cual es congruente con lo esperado y con el análisis descriptivo realizado. Igualmente pasa lo esperado para la asistencia, donde se aprecia que los sectores altos son los que presentan las probabilidades más altas en comparación con los sectores bajos. Sin embargo, para las demás causas las

<sup>20</sup> Para este grupo de edad se sacó del análisis al factor 3 "Condiciones de la vivienda" debido a que los coeficientes de este factor para las categorías más comunes (asistencia, no quiso o no le gustó estudiar y falta de dinero) de la variable dependiente no son estadísticamente significativas lo que causaba problemas en las interpretaciones de los resultados.

diferencias no son tan notorias e incluso en algunos casos se presenta lo contrario a lo esperado, lo que podría decirnos que probablemente para este grupo de edad las demás probabilidades de las causas de abandono no dependen de los sectores en los que se encuentren y únicamente son representativos de los recursos económicos con los que cuentan los jóvenes para asistir a alguna institución educativa posterior a la educación básica. Es decir, para este grupo de edad la probabilidad de dejar la escuela debido a la falta de recursos económicos es la más importante y la que obviamente marca la diferencia entre el sector bajo y alto.

Respecto al nivel de estudios alcanzado, se ve que la probabilidad de asistencia aumenta conforme se va aumentando de nivel educativo. Mientras que las probabilidades de "No quiso o no le gustó estudiar" y "No lo dejaron o por encargarse de las tareas del hogar" van disminuyendo conforme se aumente de nivel educativo alcanzado (47.85% y 2.62% respectivamente para los jóvenes que están en primaria contra un 0.60% y 0.06% respectivamente para los de educación profesional). Para las causas de "Falta de dinero o tenía que trabajar" y "Se casó o unió" las probabilidades más altas se presentan para los jóvenes que tienen educación Secundaria. Sin embargo para la falta de recursos económicos las probabilidades entre educación Profesional y los demás niveles presentan un salto fuerte (1.30% para los de educación profesional); lo que significa que la falta de dinero es una causa de poco peso y tiene probabilidades bajas de ocurrencia para los jóvenes que tienen la oportunidad de continuar con sus estudios profesionales. Lo mismo sucede con "No quiso o no le gustó estudiar" (0.60% para los de educación profesional) lo que quiere decir que las probabilidades de abandonar la escuela a nivel licenciatura o similar por esta razón son muy pocas. Lo anterior obedece al sentido común al esperarse que los jóvenes que atienden una institución profesional lo hagan por gusto. Con respecto a "Terminó sus estudios" podemos ver que hay más probabilidad de terminar para los niveles de educación medio superior y profesional.

Para el sexo, se destaca que las causas de "Se casó o unió" y "No lo dejaron o por encargarse del hogar" tienen probabilidades más altas si se trata de las mujeres. Estas razones ya son ligeramente más significativas para este grupo de edad que para las mujeres de 12 a 17 años. Sin embargo, es relevante destacar que para las mujeres la probabilidad de terminar sus estudios es significativamente más alta que para los hombres (9.26% para los hombres y 15.81% para las mujeres); es decir, parece ser que las mujeres presentan mayores oportunidades de terminar con sus estudios. Además presentan menor probabilidad de dejar la escuela debido a que no le haya gustado o no haya querido estudiar con respecto a los hombres (25.32% para los hombres y

18.21% para las mujeres) lo cual también puede influir en que las mujeres tengan la probabilidad más alta de concluir sus estudios.

Ya para las edades, se alcanza a resaltar que en general las probabilidades según la edad de las diferentes causas de abandono van aumentando conforme aumenta ésta. Por el contrario, las probabilidades de asistencia para las diferentes edades van disminuyendo conforme aumenten estas últimas.



## CONCLUSIONES

Los factores obtenidos en el presente trabajo ayudaron a corroborar la fuerte relación que existe entre tener una alta calidad educativa y mayores oportunidades de desarrollo con las condiciones socioeconómicas en las que viven los jóvenes de 12 a 24 años.

Al describir la asistencia, la situación educativa y la situación de actividad en el capítulo II se pudo observar como los sectores más bajos son los más desfavorecidos en las situaciones descritas. Además se pudo apreciar cómo es que un importante porcentaje se dedica a trabajar en lugar de estudiar, asunto que no debería de ser tan frecuente debido a los rangos de edad. Por lo tanto, las condiciones en las que se desarrollen los jóvenes determinan las posibilidades de participar en el ambiente educativo o bien en el ambiente laboral.

El capítulo III le da más fuerza a estas afirmaciones debido a que están respaldadas por un modelo matemático que nos indica las probabilidades de dejar la escuela según el sector, el sexo, el nivel de estudios alcanzado y la edad. Se puede notar nuevamente como es que las probabilidades de deserción son más altas en los sectores bajos que en los altos.

Como bien se ha afirmado en ese trabajo, a lo largo de la historia del país se han hecho políticas para incrementar las posibilidades de al menos cursar la educación básica; sin embargo, poco se ha hecho por la calidad de esta educación. Y es en este punto donde recae, sobre todo, las diferencias y desigualdades socioeconómicas que vive actualmente la población; es decir las inequidades (Ornelas, 1998).

Algo que se debe tomar en cuenta es que los resultados obtenidos en el presente trabajo sólo son cifras que describen las situaciones en las que viven los jóvenes y las probabilidades que tienen de desertar según sus características. Sin embargo, estas cifras no reflejan como se vive el ambiente escolar o familiar que propicia que los jóvenes deserten, ya sea porque no les guste ni les atraiga el ambiente educativo o porque se vean obligados a dejarlo por cuestiones que van más allá de su control; por ejemplo, que tengan que trabajar para ayudar a que la familia salga adelante (Ornelas, 1998).

Por lo tanto atender con prontitud y calidad a la población más desfavorecida para ofrecer mejores servicios educativos en cuanto a instalaciones y docentes; así como

mejorar la experiencia educativa en los jóvenes para evitar las deserciones son cuestiones prioritarias en las que la población debe trabajar, incluidos gobierno y sociedad. Por lo tanto, la participación y el apoyo de los padres de familia son fundamentales en la educación de los hijos ya que son ellos los que están más cercanos a la población joven y pueden lograr que haya menos deserciones al mejorar las experiencias educativas y de vida y así incrementar los estímulos y las aspiraciones de los jóvenes sin importar el sector socioeconómico en donde se encuentren (Schmelkes, 1998).

## ANEXO I

Variables para los cuartiles del Factor 1

Cocina			
Sector	En el cuarto donde cocinan ¿también duermen?		
	No	Sí	Total
Sector bajo	77.60%	22.40%	225,094
Sector medio bajo	88.75%	11.25%	225,082
Sector medio alto	97.07%	2.93%	224,981
Sector alto	98.97%	1.03%	225,200
<b>Total</b>	<b>90.60%</b>	<b>9.40%</b>	<b>900,357</b>

Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.

Disponibilidad de agua							
Sector	Agua entubada						Total
	Dentro de la vivienda	Fuera de la vivienda	De la llave pública	Se acarrea de otra vivienda	De pipa	De un pozo, río, lago arroyo u otra	
Sector bajo	13.98%	45.67%	4.48%	3.33%	3.26%	29.28%	225,094
Sector medio bajo	52.88%	38.49%	2.27%	0.87%	1.92%	3.57%	225,083
Sector medio alto	83.77%	14.35%	0.84%	0.09%	0.43%	0.52%	224,981
Sector alto	97.83%	1.81%	0.28%	0.01%	0.04%	0.04%	225,200
<b>Total</b>	<b>62.12%</b>	<b>25.08%</b>	<b>1.97%</b>	<b>1.07%</b>	<b>1.41%</b>	<b>8.35%</b>	<b>900,358</b>

Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.

Electricidad			
Sector	Sí	No	Total
Sector bajo	85.22%	14.78%	225,094
Sector medio bajo	99.75%	0.25%	225,083
Sector medio alto	99.96%	0.04%	224,981
Sector alto	99.99%	0.01%	225,200
<b>Total</b>	<b>96.23%</b>	<b>3.77%</b>	<b>900,358</b>

Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.

Combustible con el que cocinan						
Sector	Gas	Electricidad	Petróleo	Carbón	Leña	Total
Sector bajo	35.00%	0.09%	0.04%	0.50%	64.37%	225,094
Sector medio bajo	96.36%	0.12%	0.01%	0.10%	3.41%	225,083
Sector medio alto	99.68%	0.12%	0.00%	0.02%	0.18%	224,981
Sector alto	99.76%	0.24%	0.00%	0.00%	0.00%	225,200
<b>Total</b>	<b>82.70%</b>	<b>0.14%</b>	<b>0.01%</b>	<b>0.15%</b>	<b>16.99%</b>	<b>900,358</b>

Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.

Eliminación de basura					
Sector	Recoge camión	Depositan en contenedor	Queman o entierran	La tiran	Total
Sector bajo	19.57%	1.35%	67.10%	11.98%	225,094
Sector medio bajo	74.49%	3.59%	18.81%	3.11%	225,083
Sector medio alto	90.70%	4.94%	3.54%	0.82%	224,981
Sector alto	95.99%	3.81%	0.15%	0.05%	225,200
Total	70.19%	3.42%	22.40%	3.99%	900,358

Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.

Drenaje					
Sector	El drenaje de esta vivienda va a dar a:				
	Red pública	Fosa séptica	Intemperie	No tiene	Total
Sector bajo	13.89%	14.03%	5.69%	66.38%	225,093
Sector medio bajo	64.01%	18.50%	5.13%	12.35%	225,083
Sector medio alto	86.99%	9.82%	1.91%	1.28%	224,981
Sector alto	96.25%	3.31%	0.37%	0.07%	225,200
Total	65.29%	11.42%	3.27%	20.02%	900,357

Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.

Televisión			
Sector	Sí	No	Total
Sector bajo	61.06%	38.94%	225,094
Sector medio bajo	97.75%	2.25%	225,083
Sector medio alto	99.78%	0.22%	224,981
Sector alto	100.00%	0.00%	225,199
Total	89.65%	10.35%	900,357

Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.

Videocasetera			
Sector	Sí	No	Total
Sector bajo	4.11%	95.89%	225,093
Sector medio bajo	21.41%	78.59%	225,082
Sector medio alto	56.08%	43.92%	224,981
Sector alto	92.29%	7.71%	225,200
Total	43.48%	56.52%	900,356

Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.

Licuadora			
Sector	Sí	No	Total
Sector bajo	43.96%	56.04%	225,094
Sector medio bajo	93.13%	6.87%	225,082
Sector medio alto	98.95%	1.05%	224,981
Sector alto	99.96%	0.04%	225,200
Total	84.00%	16.00%	900,357

Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.

Refrigerador			
Sector	Sí	No	Total
Sector bajo	19.39%	80.61%	225,094
Sector medio bajo	76.33%	23.67%	225,083
Sector medio alto	97.91%	2.09%	224,981
Sector alto	99.97%	0.03%	225,200
<b>Total</b>	<b>73.40%</b>	<b>26.60%</b>	<b>900,358</b>
Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.			

Lavadora			
Sector:	Sí	No	Total
Sector bajo	8.37%	91.63%	225,094
Sector medio bajo	47.26%	52.74%	225,083
Sector medio alto	82.97%	17.03%	224,980
Sector alto	98.26%	1.74%	225,199
<b>Total</b>	<b>59.22%</b>	<b>40.78%</b>	<b>900,356</b>
Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.			

Teléfono			
Sector	Sí	No	Total
Sector bajo	1.03%	98.97%	225,094
Sector medio bajo	12.94%	87.06%	225,083
Sector medio alto	55.52%	44.48%	224,981
Sector alto	95.88%	4.12%	225,200
<b>Total</b>	<b>41.35%</b>	<b>58.65%</b>	<b>900,358</b>
Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.			

Boiler			
Sector	Sí	No	Total
Sector bajo	1.94%	98.06%	225,094
Sector medio bajo	18.07%	81.93%	225,083
Sector medio alto	62.10%	37.90%	224,981
Sector alto	97.82%	2.18%	225,199
<b>Total</b>	<b>44.99%</b>	<b>55.01%</b>	<b>900,357</b>
Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.			

Auto propio			
Sector	Sí	No	Total
Sector bajo	6.60%	93.40%	225,094
Sector medio bajo	17.28%	82.72%	225,083
Sector medio alto	39.27%	60.73%	224,981
Sector alto	80.58%	19.42%	225,199
<b>Total</b>	<b>35.94%</b>	<b>64.06%</b>	<b>900,357</b>
Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.			

Computadora			
Sector	Sí	No	Total
Sector bajo	0.04%	99.96%	225,094
Sector medio bajo	0.45%	99.55%	225,083
Sector medio alto	3.75%	96.25%	224,981
Sector alto	43.73%	56.27%	225,199
<b>Total</b>	<b>12.00%</b>	<b>88.00%</b>	<b>900,357</b>
Fuente: Elaboración propia en SPSS de los censos de población y vivienda del 2000 del INEGI.			

Tablas de las gráficas elaboradas en el capítulo II

## ASISTENCIA

Asistencia por sexo a partir del Factor 1 ( edades 12 a 17 )				
Sector		Sí va a la escuela	No va a la escuela	Total
Hombres	Sector bajo	64.12%	35.88%	79,629
	Sector medio bajo	74.16%	25.84%	71,435
	Sector medio alto	82.75%	17.25%	66,244
	Sector alto	93.06%	6.94%	61,570
<b>Total</b>		<b>77.51%</b>	<b>22.49%</b>	<b>278,878</b>
Mujeres	Sector bajo	62.21%	37.79%	73,968
	Sector medio bajo	75.69%	24.31%	69,791
	Sector medio alto	85.13%	14.87%	64,113
	Sector alto	94.42%	5.58%	59,115
<b>Total</b>		<b>78.37%</b>	<b>21.63%</b>	<b>266,987</b>
Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.				

Asistencia por sexo a partir del Factor 1 ( edades de 18 a 24 )				
Sector		Sí va a la escuela	No va a la escuela	Total
Hombres	Sector bajo	10.71%	89.29%	39,562
	Sector medio bajo	21.44%	78.56%	43,521
	Sector medio alto	33.94%	66.06%	48,485
	Sector alto	57.87%	42.13%	53,517
<b>Total</b>		<b>32.95%</b>	<b>67.05%</b>	<b>185,085</b>
Mujeres	Sector bajo	11.59%	88.41%	30,792
	Sector medio bajo	23.25%	76.75%	39,497
	Sector medio alto	35.83%	64.17%	45,368
	Sector alto	57.54%	42.46%	50,316
<b>Total</b>		<b>34.92%</b>	<b>65.08%</b>	<b>165,973</b>
Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.				

## NIVEL DE ESTUDIOS

Nivel de estudios por sexo a partir del Factor 1 ( edades 12 a 17 )					
Sector		Primaria	Secundaria	Media Superior	Total
Hombres	Sector bajo	61.24%	35.44%	3.32%	76,644
	Sector medio bajo	42.31%	48.91%	8.78%	70,611
	Sector medio alto	31.74%	54.37%	13.88%	65,741
	Sector alto	23.72%	54.04%	22.24%	61,311
Total		40.91%	47.60%	11.49%	274,307
Mujeres	Sector bajo	61.58%	34.49%	3.93%	71,119
	Sector medio bajo	38.60%	50.65%	10.76%	69,037
	Sector medio alto	28.68%	54.00%	17.33%	63,643
	Sector alto	21.66%	52.64%	25.71%	58,874
Total		38.62%	47.53%	13.85%	262,673

Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.

Nivel de estudios a partir del Factor 1 (edades de 18 a 24)						
Sector		Primaria	Secundaria	Media Superior	Profesional	Total
Hombres	Sector bajo	51.37%	33.68%	12.52%	2.44%	36,672
	Sector medio bajo	24.78%	40.50%	26.49%	8.23%	42,492
	Sector medio alto	11.15%	35.22%	37.87%	15.76%	47,811
	Sector alto	2.99%	17.17%	42.41%	37.43%	53,232
Total		20.14%	30.82%	31.37%	17.68%	180,207
Mujeres	Sector bajo	52.76%	30.43%	13.94%	2.88%	28,496
	Sector medio bajo	22.08%	36.69%	31.41%	9.81%	38,676
	Sector medio alto	9.49%	28.16%	42.59%	19.76%	44,748
	Sector alto	2.44%	11.96%	41.41%	44.19%	49,893
Total		17.95%	25.60%	34.51%	21.94%	161,813
Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.						

Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.

## ACTIVIDAD PRINCIPAL

Actividad principal por sexo a partir del Factor 1 ( edades 12 a 17 )							
Sector		Trabaja	Busca trabajo	Estudia	Que hacedes del hogar	No trabaja	Total
Hombres	Sector bajo	31.22%	0.60%	50.29%	0.78%	17.10%	79,231
	Sector medio bajo	23.44%	0.66%	57.73%	0.48%	17.69%	70,945
	Sector medio alto	17.80%	0.44%	64.06%	0.34%	17.36%	65,742
	Sector alto	10.07%	0.24%	73.69%	0.19%	15.81%	61,040
Total		21.38%	0.50%	60.62%	0.47%	17.03%	276,958
Mujeres	Sector bajo	12.58%	0.22%	51.72%	19.70%	15.78%	73,874
	Sector medio bajo	11.92%	0.26%	62.11%	9.05%	16.66%	69,537
	Sector medio alto	9.32%	0.24%	68.96%	4.65%	16.83%	63,843
	Sector alto	5.66%	0.12%	76.74%	1.89%	15.59%	58,839
Total		10.09%	0.21%	64.10%	9.37%	16.22%	266,093
Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.							

Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.

Actividad principal a partir del Factor 1 ( edades 18 a 24 )							
Sector		Trabaja	Busca trabajo	Estudia	Que hacedes del hogar	No trabaja	Total
Hombres	Sector bajo	75.33%	1.76%	6.51%	0.80%	15.60%	39,512
	Sector medio bajo	69.72%	2.42%	13.01%	0.38%	14.48%	43,461
	Sector medio alto	64.00%	2.12%	20.42%	0.31%	13.16%	48,458
	Sector alto	50.47%	1.47%	35.33%	0.25%	12.49%	53,543
Total		63.85%	1.92%	20.02%	0.41%	13.80%	184,974
Mujeres	Sector bajo	36.88%	0.59%	7.54%	40.56%	14.43%	30,824
	Sector medio bajo	47.83%	1.11%	15.08%	21.83%	14.15%	39,502
	Sector medio alto	47.66%	1.22%	23.10%	13.81%	14.21%	45,381
	Sector alto	40.34%	1.10%	38.31%	6.51%	13.73%	50,355
Total		43.48%	1.04%	22.91%	18.47%	14.09%	166,062

Fuente: Elaboración propia en SPSS a partir de los datos del Censo de Población y Vivienda del 2000 del INEGI.



## ANEXO II

Sintaxis de los programas utilizados para la construcción de las variables.

- Variable para establecer el parentesco con el jefe de familia

La variable *paredesc* no indica cual es el parentesco de las persona con el jefe de familia En este caso nos interesan los hijos(as), nietos(as), bisnietos(as), tataranietos(as), sobrinos(as), tutelados(as) y ahijados(as) porque son de generaciones abajo del jefe que pueden llegar a ser influenciadas por las decisiones que tome este último, por el nivel educativo que tenga el jefe y porque generalmente son los dependientes económicos en el hogar. Esta variable es importante porque con ella fue con la que se hicieron las submuestras de la población total.

El programa para crearla fue:

```
DATASET ACTIVATE DataSet1.
IF (otropare = 300) paredesc=300.
IF (otropare = 608) paredesc=608.
IF (otropare = 609) paredesc=609.
IF (otropare = 610) paredesc=610.
IF (otropare = 612) paredesc=612.
IF (otropare = 622) paredesc=622.
EXECUTE.
VARIABLE LABELS paredesc "Parentesco descendientes".
VALUE LABELS paredesc 300 "Hijo(a)" 608 "Nieto(a)" 609
"Bisnieto(a)" 610 "Tataranieto(a)" 612 "Sobrino(a)" 620
"Tutelado" 622 "Ahijado(a)".
```

- Variable para establecer la situación educativa de los jóvenes

Las causas de abandono son una parte esencial del trabajo de investigación, por lo que para cada grupo de edad se establecieron las causas de abandono más significativas. El cuestionario del censo tiene las siguientes causas de abandono: nunca ha ido a la escuela, no quiso o no le gustó estudiar, falta de dinero o tenía que trabajar, la escuela estaba muy lejos o no había, Se casó o unió, su familia ya no lo dejo o por encargarse de las tareas del hogar, terminó sus estudios, otro motivo y no especificado. Sin embargo, las causas de abandono: nunca ha ido a la escuela, la escuela estaba muy lejos y el no especificado se sacaron del análisis al representar a una proporción baja de la muestra en estudio.

El programa para crearla fue:

```
DATASET ACTIVATE DataSet1.
RECODE abanESCO (1=1) (2=2) (3=3) (5=4) (4=SYSMIS) (9=SYSMIS)
(7=5) (6=SYSMIS) (0=SYSMIS) INTO abandono12.
VARIABLE LABELS abandono 'Causa de abandono para los de edades
de 12 a 17'.
VALUE LABELS abandono 1 "No quiso o no le gustó" 2 "Falta de
dinero o tenía que trabajar" 3 "Se casó o unió" 4 "Ya no lo
dejaron o por encargarse del hogar" 5 "Otro motivo".
EXECUTE.
```

Para hacer la variable "Situación educativa" se hizo una nueva variable donde se tomaran en cuenta a los que asisten y no asisten a la escuela y que expusiera la causa para los que no asisten.

Esta variable de situación educativa es muy importante porque además fue la que se utilizó para realizar el análisis de regresión logística múltiple.

El programa para crearla fue:

```
DATASET ACTIVATE DataSet1.
compute situedu=9.
IF (asisten = 1) situedu=1.
IF (asisten = 2 & abandono=1) situedu=2.
IF (asisten = 2 & abandono=2) situedu=3.
IF (asisten = 2 & abandono=3) situedu=4.
IF (asisten = 2 & abandono=4) situedu=5.
IF (asisten = 2 & abandono=5) situedu=6.
IF (asisten = 2 & abandono=6) situedu=7.
VARIABLE LABELS situedu "Situación educativa".
VALUE LABELS situedu 1 "Asisten" 2 "No quiso o no le gustó
estudiar" 3 "Falta de dinero o tenía que trabajar" 4 "Se casó o
unió" 5 "No lo dejaron o por encargarse del hogar" 6 "Terminó
sus estudios" 7 "Otro motivo".
```

- Variable para definir el nivel de estudios alcanzado

Al combinar las variables *nivacad* y *antesc* que miden el nivel académico y el antecedente escolar respectivamente dan como resultado la variable *nivest* que mide el nivel de estudios desde kínder a posgrado de la población a analizar.

El programa para crearla fue:

```
DATASET ACTIVATE DataSet1.
compute nivest=9.
IF (abanESCO = 0) nivest=0.
IF (nivacad = 0) nivest=0.
IF (nivacad = 1) nivest=1.
IF (nivacad = 2) nivest=2.
IF (nivacad = 3 | (nivacad=6 & antesc = 1)) nivest=3.
IF (nivacad = 4 | ((nivacad=5 | nivacad=6) & (antesc = 2)))
nivest=4.
IF ((nivacad > 4 & nivacad < 6) & antesc = 3) nivest=5.
IF (nivacad = 6) nivest=6.
EXECUTE.
var label nivest "Nivel de estudios".
```

```
val label nivest 0 "Ninguno" 1 "Kinder o preescolar" 2
"Primaria" 3 "Secundaria" 4 "Media Superior" 5 "Profesional" 6
"Posgrado" 9 "No Especificado".
```

Con la variable anterior se pudo crear una nueva variable que agrupa nada más la educación primaria, secundaria, media superior, y profesional. Las clasificaciones de "Ninguno", "Kinder" y "No especificado" de la variable *nivest* se sacaron de la muestra al no ser estas variables que representarían a la mayoría de la población de estudio. Y la educación "Profesional" y de "Posgrado" se juntaron en una sola llamada "Profesional". El nombre la nueva variable de nivel educativo es *nivestn*.

El programa para crearla fue:

```
DATASET ACTIVATE DataSet1.
RECODE nivest (2=1) (3=2) (4=3) (5 THRU 6=4) (9=9) INTO nivestn.
VARIABLE LABELS nivestn 'Nivel de estudios nuevo'.
VALUE LABELS nivestn 1 "Primaria" 2 "Secundaria" 3 "Media
Superior" 4 "Profesional".
EXECUTE.
```

- Variable para definir la actividad principal de los jóvenes

A continuación se presenta la sintaxis utilizada para crear la variable "Actividad principal" expuesta y explicada en el capítulo II

El programa para crearla fue:

```
DATASET ACTIVATE DataSet1.
RECODE conact (30=2) (40=3) (50=4) (60=5) (80=5) (70=6) (9=9)
(10 thru 20=1) INTO actividadn.
VARIABLE LABELS actividadn 'Actividad principal'.
VALUE LABELS actividadn 1 "Trabaja" 2 "Busca trabajo" 3
"Estudia" 4 "Que haces del hogar" 5 "No trabaja" 6
"Incapacitado" 9 "NE".
EXECUTE.
```

- Variables utilizadas en el Análisis de Regresión Logística multinomial convertidas en variables dicotómicas:

Factor 1 "Servicios y enseres básicos", categoría base: Sector bajo

```
DATASET ACTIVATE DataSet1.
IF (FAC1_1 <= .1653029) cuartilesF1R_2=2.
IF (FAC1_1 > .1653029) cuartilesF1R_2=1.
IF (cuartilesF1R_2 = 2) DFR1c=1.
IF (cuartilesF1R_2 = 1) DFR1c=0.
VARIABLE LABELS cuartilesF1R_2 "Cuartiles factor 1 rotado dos
grupo".
value labels cuartilesF1R_2 1 "Bajo", 2 "Alto".
variable labels DFR1c "Dummy para el factor rotado 1 dos
grupos".
EXECUTE.
```

Factor 2 "Enseres medios", categoría base: Sector bajo

```
DATASET ACTIVATE DataSet2.
IF (FAC2_1 <= .8296232) cuartilesF2R_2=2.
IF (FAC2_1 > .8296232) cuartilesF2R_2=1.
IF (cuartilesF2R_2 = 2) DFR2c=1.
IF (cuartilesF2R_2 = 1) DFR2c=0.
VARIABLE LABELS cuartilesF2R_2 "Cuartiles factor 2 rotado dos
grupo".
value labels cuartilesF2R_2 1 "Bajo", 2 "Alto".
variable labels DFR2c "Dummy para el factor rotado 2 dos
grupos".
EXECUTE.
```

Factor 3 "Condiciones de la vivienda", categoría base: Sector bajo

```
DATASET ACTIVATE DataSet2.
IF (FAC3_1 <= -.0010153) cuartilesF3R_2=2.
IF (FAC3_1 > -.0010153) cuartilesF3R_2=1.
IF (cuartilesF3R_2 = 2) DFR3c=1.
IF (cuartilesF3R_2 = 1) DFR3c=0.
VARIABLE LABELS cuartilesF3R_2 "Cuartiles factor 3 rotado dos
grupo".
value labels cuartilesF3R_2 1 "Bajo", 2 "Alto".
variable labels DFR3c "Dummy para el factor rotado 3 dos
grupos".
EXECUTE.
```

Edades de 12 a 17 años, categoría base: edad 12

```
DATASET ACTIVATE DataSet1.
IF (edad = 14) dedad14=1.
IF (edad=13 | edad=12 | edad=15 | edad=16 | edad=17) dedad14=0.
IF (edad = 13) dedad13=1.
IF (edad=12 | edad=14 | edad=15 | edad=16 | edad=17) dedad13=0.
IF (edad = 15) dedad15=1.
IF (edad=12 | edad=13 | edad=14 | edad=16 | edad=17) dedad15=0.
IF (edad = 16) dedad16=1.
IF (edad=12 | edad=13 | edad=14 | edad=15 | edad=17) dedad16=0.
IF (edad = 17) dedad17=1.
IF (edad=12 | edad=13 | edad=14 | edad=15 | edad=16) dedad17=0.
variable labels dedad13 "Dummy edad 13".
variable labels dedad15 "Dummy edad 15".
variable labels dedad14 "Dummy edad 14".
variable labels dedad16 "Dummy edad 16".
variable labels dedad17 "Dummy edad 17".
EXECUTE.
```

Edades de 18 a 24 años, categoría base: edad 18

```
DATASET ACTIVATE DataSet3.
IF (edad = 19) dedad19=1.
IF (edad=18 | edad=20 | edad=21 | edad=22 | edad=23 |
edad=24)dedad19=0.
IF (edad = 20) dedad20=1.
IF (edad=18 | edad=19 | edad=21 | edad=22 | edad=23 |
edad=24)dedad20=0.
IF (edad = 21) dedad21=1.
IF (edad=18 | edad=19 | edad=20 | edad=22 | edad=23 |
edad=24)dedad21=0.
IF (edad = 22) dedad22=1.
IF (edad=18 | edad=19 | edad=20 | edad=21 | edad=23 |
edad=24)dedad22=0.
```

```

IF (edad = 23) dedad23=1.
IF (edad=18 | edad=19 | edad=20 | edad=21 | edad=22 |
edad=24)dedad23=0.
IF (edad = 24) dedad24=1.
IF (edad=18 | edad=19 | edad=20 | edad=21 | edad=22 |
edad=23)dedad24=0.
variable labels dedad19 "Dummy edad 19".
variable labels dedad20 "Dummy edad 20".
variable labels dedad21 "Dummy edad 21".
variable labels dedad22 "Dummy edad 22".
variable labels dedad23 "Dummy edad 23".
variable labels dedad24 "Dummy edad 24".
EXECUTE.

```

Nivel de estudios alcanzado, categoría base: primaria

```

DATASET ACTIVATE DataSet3.
IF (nivestn = 2) dsec=1.
IF (nivestn =1 | nivestn=3 | nivestn=4) dsec=0.
IF (nivestn = 3) dmedsup=1.
IF (nivestn =1 | nivestn=2 | nivestn=4) dmedsup=0.
IF (nivestn = 4) dprof=1.
IF (nivestn =1 | nivestn=2 | nivestn=3) dprof=0.
variable labels dsec "Dummy para Secundaria".
variable labels dmedsup "Dummy para Medio superior".
variable labels dprof "Dummy para Profesional".
EXECUTE.

```

Sexo, categoría base: Hombre

```

DATASET ACTIVATE DataSet1.
IF (sexo = 2) dm=1.
IF (sexo = 1) dm=0.
variable labels dm "Dummy para mujeres".
EXECUTE.

```

## BIBLIOGRAFÍA

Babbie, Earl R. (1997) The Practice of Social Research, 8ª edición, E.U.A., Wadsworth Pub Co.

Latapí, Pablo (1998) "Un siglo de educación nacional: una sistematización" en Pablo Latapí (coord.) Un siglo de educación en México, vol. I, México, Consejo Nacional para la Cultura y las Artes y Fondo de Cultura Económica, págs. 21 a 42.

Ornelas, Carlos (1998) "La cobertura de la educación básica" en Pablo Latapí (coord.) Un siglo de educación en México, vol. II, México, Consejo Nacional para la Cultura y las Artes y Fondo de Cultura Económica, págs. 111 a 140.

Schmelkes, Silvia (1998) "La educación básica" en Pablo Latapí (coord.) Un siglo de educación en México, vol. II, México, Consejo Nacional para la Cultura y las Artes y Fondo de Cultura Económica, págs. 173 a 194.

Márquez Jiménez, Alejandro (2009), "Las diferentes posibilidades de los jóvenes mexicanos de participar en la educación y el trabajo. Un análisis a partir de la encuesta nacional de la juventud 2005",

<http://www.comie.org.mx/congreso/memoria/v9/ponencias/at10/PRE1178887300.pdf>

Cervini, Rubén, Tenti, Emilio (2004) "Notas sobre la masificación de la escolarización en seis países de América Latina", Debate 1: Equidad en el acceso y la permanencia en el sistema educativo, Sistema de Información de Tendencias Educativas en América Latina (SITEAL).

[http://www.siteal.iipe-](http://www.siteal.iipe-oei.org/modulos/DebatesV1/upload//deb_23/art_15/ART_Cervini-Tenti.pdf)

[oei.org/modulos/DebatesV1/upload//deb\\_23/art\\_15/ART\\_Cervini-Tenti.pdf](http://www.siteal.iipe-oei.org/modulos/DebatesV1/upload//deb_23/art_15/ART_Cervini-Tenti.pdf)

Calero, Jorge (2004) "La incidencia de la educación sobre los ingresos y sobre el riesgo de pobreza", Debate 3: La incidencia de la educación sobre el bienestar de los hogares, Sistema de Información de Tendencias Educativas en América Latina (SITEAL).

[http://www.siteal.iipe-](http://www.siteal.iipe-oei.org/modulos/DebatesV1/upload//deb_38/art_70/SITEAL_Debate_3_Calero_Articulo.pdf)

[oei.org/modulos/DebatesV1/upload//deb\\_38/art\\_70/SITEAL\\_Debate\\_3\\_Calero\\_Articulo.pdf](http://www.siteal.iipe-oei.org/modulos/DebatesV1/upload//deb_38/art_70/SITEAL_Debate_3_Calero_Articulo.pdf)

Sistema de Información de Tendencias Educativas en América Latina (SITEAL) (2009), "La escuela y los adolescentes: Datos destacados", Informe sobre tendencias sociales y educativas en América Latina 2008.

<http://www.siteal.iipe-oei.org/datosdestacados/detalle.asp?DatoDestacadoID=35>

"Capítulo 20: Análisis Factorial", SPSS 10 Guía para el análisis de datos.

<http://www2.uca.es/serv/ai/formacion/spss/Pantalla/verguia.pdf>

Cuesta, Marcelino, Herrero Francisco J. (2009), "Introducción al Análisis Factorial", Tutorial de Análisis Factorial, España, Universidad de Oviedo.

[http://www.psico.uniovi.es/Dpto\\_Psicologia/metodos/tutor.1/indice.html](http://www.psico.uniovi.es/Dpto_Psicologia/metodos/tutor.1/indice.html)

Reimers, Fernando (2000), "Educación, desigualdad y opciones de política en América Latina en el siglo XXI", Revista Iberoamericana, Núm. 23, pág. 21-47

Salas, Roberto (2007), "Nota técnica: Correlación e Interdependencia", México, Metrología Integral y Desarrollo S.A. de C.V., pág. 3-4

[http://www.midelab.com.mx/notas\\_tec/correlacion\\_e\\_independencia.pdf](http://www.midelab.com.mx/notas_tec/correlacion_e_independencia.pdf)

Álvarez, Rafael (1994), Estadística multivariada y no paramétrica con SPSS. Aplicación a las ciencias de la salud, Díaz de Santo, España, págs. 155-202.

Retherford, R.D. y M.K. Choe (1993), Statistical Models for Causal Analysis, Nueva York, John Wiley y Sons, págs. 119-130.

M. Mood, Alexander, Graybill, Franklin A., Boes, Duane C. (1974), "Introduction to the theory of Statistics", Singapur, McGraw Hill, pág. 155, 271 y 401

Rodriguez, Jaume (2001), "Modelos sociodemográficos: Atlas social de la ciudad de Alicante", Capítulo 5 El análisis factorial en la construcción de indicadores e índices sociales, Alicante, pág. 193-202.