



EL COLEGIO DE MÉXICO

CENTRO DE ESTUDIOS ECONÓMICOS

MAESTRÍA EN ECONOMÍA

TRABAJO DE INVESTIGACIÓN PARA OBTENER EL GRADO DE
MAESTRO EN ECONOMÍA

*Aprovechamiento y Salarios de los Recursos
Humanos en Ciencia y Tecnología en México*

Ricardo Leyva Bernal

Promoción 1995-1997

ASESOR:

DR. ADRIÁN JIMÉNEZ GÓMEZ

Noviembre 2006

Resumen

En este trabajo se hace un análisis sobre el aprovechamiento de los Recursos Humanos en Ciencia y Tecnología (RHCyT). Se plantean las consideraciones para medir este aprovechamiento de acuerdo con el *Manual de Canberra*¹, y se propone como medida del aprovechamiento de los recursos humanos a la probabilidad de que éstos estén ocupados en las actividades para las cuales se les educó. Se estiman modelos *logit* para calcular esta probabilidad.

Se concluye que a pesar de que las mediciones del *Manual de Canberra* sugieren un desaprovechamiento importante de todos los recursos humanos, con la medición propuesta en este trabajo se muestra que en general se tiene un mejor aprovechamiento de los mismos. Sin embargo, al desagregar por los diferentes campos de conocimiento el aprovechamiento de los recursos humanos dista de ser óptimo en ciencias naturales y exactas, y en ingeniería y tecnología. Para demostrar que esto no es explicado por los salarios se estima una función de ingresos con el método de dos etapas de Heckman. Los resultados de dicha estimación confirman que no existe una discriminación salarial negativa para las ingenierías y las ciencias exactas y naturales. Se prueba que el hecho de que los RHCyT no esten siendo aprovechados no se debe a que estos tengan estímulos salariales para salirse de las ocupaciones propias de su campo de conocimiento.

¹ Manual on the Measurement of Human Resources Devoted to S&T “Canberra Manual”. OECD, Paris, 1995.

Índice General

Resumen.....	2
Índice General.....	3
Introducción.....	4
Capítulo 1: Recursos Humanos en Ciencia y Tecnología en México.....	6
Capítulo 2: La Regresión Logística.....	13
Capítulo 3: Una regresión Logística para medir el Aprovechamiento de los Recursos Humanos en Ciencia y Tecnología.....	20
Capítulo 4: Salarios Promedio y Salarios Relativos entre las Diferentes Áreas del Conocimiento.....	41
Capítulo 5: Estimación de una Función de Ingresos.....	52
Conclusiones.....	67
Anexo 1: Fuentes de Información y Población Considerada.....	71
Anexo 2: Resultados de las Estimaciones de los Modelos Logit.....	73
Bibliografía.....	83
Índice de Cuadros.....	86

Introducción

En este trabajo se hace un análisis sobre el aprovechamiento de los Recursos Humanos en Ciencia y Tecnología (RHCyT), utilizando como fuente de información la base de datos de la Encuesta Nacional de Empleo de 1998². Se plantean las consideraciones para medir este aprovechamiento de acuerdo con el *Manual de Canberra*³, y se propone como medida del aprovechamiento de los recursos humanos a la probabilidad de que éstos estén ocupados en las actividades para las cuales se les educó.

Se concluye que, a pesar de que las mediciones del *Manual de Canberra* sugieren un desaprovechamiento importante de los recursos humanos, en términos generales, con la medición propuesta en este trabajo se muestra que se tiene un mejor aprovechamiento de los mismos. Sin embargo, al desagregar por los diferentes campos de conocimiento el aprovechamiento de los recursos humanos dista de ser óptimo en el caso de las ciencias naturales y exactas, y en ingeniería y tecnología, las que quizá sean las áreas de mayor importancia para el país.

Esto no es explicado por los salarios que obtienen los recursos humanos en las áreas de ciencia y tecnología y en las diferentes ingenierías, pues no resultan ser más bajos que en las otras áreas de conocimiento. Para demostrar esta afirmación, se estima una función de ingresos con el método de dos etapas de Heckman; los resultados de dicha estimación confirman que no existe una discriminación salarial negativa para las ingenierías y las ciencias exactas y naturales; al contrario, el efecto de estudiar en dichas áreas de conocimiento es positivo sobre los salarios y el coeficiente asociado a dichos efectos es más alto que en las demás áreas de conocimiento.

² Instituto Nacional de Estadística Geografía e Informática (INEGI) y Secretaría del Trabajo y Previsión Social (STPS), *Encuesta Nacional de Empleo 1998*, México, 1998.

³ Manual on the Measurement of Human Resources Devoted to S&T “Canberra Manual”. OECD, Paris, 1995.

Este trabajo está organizado en cinco capítulos. En el capítulo 1 se plantean las consideraciones del *Manual de Canberra* para definir el acervo de recursos humanos en ciencia y tecnología y el aprovechamiento de los mismos. En el capítulo 2 se expone en qué consiste la regresión logística y después en el capítulo 3 se estiman modelos *logit* para calcular una medida del aprovechamiento de los recursos humanos. Para cada nivel de instrucción, se presentan los principales resultados del modelo logístico, y se comparan con los obtenidos de la metodología del *Manual de Canberra*.

En el capítulo 4 se hace una descripción de los salarios promedios y los salarios relativos entre las diferentes disciplinas o áreas de conocimiento para posteriormente, en el capítulo 5, presentar lo referente a la estimación de una función de ingresos por el método de dos etapas de Heckman⁴, con la que se prueba que no existe una discriminación salarial negativa para las ingenierías y las ciencias exactas y naturales, y se presentan las conclusiones. Adicionalmente, en el anexo 1 se describen la fuente de información y la población considerada y en el anexo 2 se presentan las estimaciones de los modelos *logit* utilizados para medir el aprovechamiento de los recursos humanos. Finalmente se enumera la bibliografía considerada.

⁴ James Heckman, “*Sample selection bias as a specification error*”, *Econometrica* 47: 153-161, 1979.

Capítulo 1

Recursos Humanos en Ciencia y Tecnología en México

Es mucho lo que se ha dicho en la literatura económica sobre la importancia de los Recursos Humanos (RH) como una de las fuentes primordiales de crecimiento económico, en particular, la de los Recursos Humanos en Ciencia y Tecnología (RHCyT). Abramovitz (1956), Fabricant (1954), Kendrick (1956) y Solow (1957) encontraron que la mayor parte del crecimiento económico no podía ser explicada por las medidas convencionales de Capital y Trabajo y argumentaban que la calidad de la fuerza de trabajo era una posible explicación. Actualmente, nadie duda que el nivel educativo de la fuerza de trabajo es uno de los determinantes del crecimiento económico.

De acuerdo con Mincer (1984): *“El Capital Humano tiene un doble papel en el crecimiento económico: (1) como acervo de habilidades, es un factor de producción en combinación con el capital físico y la fuerza de trabajo, (2) como un acervo de conocimiento, que es la fuente de innovación, una causa básica del crecimiento económico”*. Por lo tanto, uno de los objetivos de cualquier nación es elevar la calidad de la fuerza de trabajo.

Hablar de la importancia de los RHCyT en el nuevo milenio y de cómo las economías se encaminan a un avance tecnológico cada vez más acelerado, en la que los RHCyT serán parte fundamental en la competitividad y el desarrollo de los países, tanto para la creación como para la adopción de nuevas y cada vez más avanzadas tecnologías, es caer en el lugar común. En consecuencia, todos los países tienen un reto, tanto en la formación como en el aprovechamiento de los RHCyT.

Por todo lo anterior, la importancia de contar con una estimación sobre el tamaño de los RHCyT cobra relevancia. En el *Manual de Canberra*⁵ se describe la manera en la que han de hacerse las mediciones con el objetivo de poder hacer comparaciones a nivel internacional. Estas medidas han sido utilizadas por países

⁵ Manual on the Measurement of Human Resources Devoted to S&T “Canberra Manual”. OECD, París, 1995.

Europeos como Francia o Suecia y también en México⁶, aunque algunos países no las han adoptado, como es el caso de los Estados Unidos de América.

En este manual se define el Acervo de Recursos Humanos en Ciencia y Tecnología (ARHCyT), como el subconjunto de la población que ha cubierto satisfactoriamente la educación de tercer nivel, de acuerdo con la ISCED⁷, en un campo de la Ciencia y la Tecnología; y/o está empleada en una ocupación de CyT que, generalmente, requiere estudios de tercer nivel.

El tercer nivel, de acuerdo con la ISCED, comprende los niveles educativos posteriores al bachillerato, estudios conducentes a grados universitarios o superiores (ISCED 6: licenciaturas; ISCED 7: especialidades, maestrías y doctorados) y estudios no equivalentes a los universitarios, pero que crean habilidades específicas (ISCED 5: carreras de técnico profesional).

Las ocupaciones consideradas como de CyT son un subconjunto de las ocupaciones consideradas en la ISCO⁸.

La clasificación de estudios y ocupaciones que se consideran de CyT puede estar sujeta a discusión. Por ejemplo, los estudios en áreas humanísticas pudieran entenderse como estudios que no son propios de la CyT. Sin embargo, esto no necesariamente es así. El desarrollo de la Ciencia ha sido resultado de avances e interacciones entre diferentes áreas de conocimiento, en las que se incluyen las humanidades.

La clasificación de los estudios y ocupaciones de la CyT no es una tarea fácil y, en la mayoría de los casos, se tienen clasificaciones arbitrarias. En este trabajo no se pretende contribuir a la discusión de qué debería considerarse como áreas de conocimiento y ocupaciones propias de la CyT.

⁶ CONACYT, Indicadores de Actividades Científicas y Tecnológicas, 1998.

⁷ UNESCO, International Standard Classification of Education, ISCED. 1997.

⁸ International Labor Office (ILO), International Standard Classification of Occupations, ISCO. 1988.

Se toman como referencias las recomendaciones del *Manual de Canberra*, las cuales, a pesar de ser de alguna forma arbitrarias, constituyen un consenso internacional sobre el tema.

En términos de escolaridad, el *Manual de Canberra* propone una clasificación muy amplia (ver cuadro 1). Se considera una lista con un núcleo que incluye disciplinas en el nivel de licenciatura o superior de los diferentes campos de la ciencia (ciencias naturales y exactas, ingeniería y tecnología, ciencias de la salud, ciencias sociales, y ciencias agropecuarias). Además del núcleo, se considera una extensión en la que se incluyen las disciplinas del área de humanidades y otras áreas de conocimiento, y el nivel educativo de técnico profesional en los diferentes campos de la Ciencia. Finalmente, para integrar una lista completa, se incluyen las disciplinas del nivel técnico profesional en todos los campos de conocimiento.

Cuadro 1

CAMPO DE CONOCIMIENTO Y NIVEL CONSIDERADOS EN EL *MANUAL DE CANBERRA*

Campo de conocimiento	Licenciatura y posgrado (ISCED 6/7)	Técnico profesional (ISCED 5)
Ciencias naturales y exactas	Núcleo	Extensión
Ingeniería y tecnología	Núcleo	Extensión
Ciencias de la salud	Núcleo	Extensión
Ciencias agropecuarias	Núcleo	Extensión
Ciencias sociales	Núcleo	Extensión
Humanidades	Extensión	Completa
Otros	Extensión	Completa

Fuente: OCDE, *Manual de Canberra*, 1995.

En términos de ocupación, el *Manual de Canberra* propone una clasificación en la que se considera, de acuerdo con la ISCO, a los grupos dos y tres y a los subgrupos 122, 123, 131 del grupo uno (ver cuadro dos). Al igual que en el caso de la escolaridad se propone una lista como núcleo, una extensión y una lista completa.

Cuadro 2

SUBGRUPOS DE OCUPACIÓN ISCO-88 CONSIDERADOS EN EL *MANUAL DE CANBERRA*

ISCO	Grupo de ocupación	
122	Administradores de los departamentos de producción y operación	Extensión
123	Administradores de otros departamentos	Extensión
131	Administradores generales	Extensión
21	Profesionales de las Ciencias físico-matemáticas e Ingenierías	Núcleo
22	Profesionales de las Ciencias de la salud y de la vida	Núcleo
23	Profesionales de la educación	Extensión
24	Otros profesionales	Extensión
31	Técnicos de las Ciencias físico-matemáticas e Ingenierías	Extensión
32	Técnicos de las Ciencias de la salud y de la vida	Extensión
33	Técnicos de la educación	Completa
34	Otros técnicos	Completa

Fuente: OCDE, *Manual de Canberra*, 1995.

Con estas clasificaciones, la recomendación principal del *Manual de Canberra* es contar con mediciones del Acervo de Recursos Humanos en Ciencia y Tecnología (ARHCyT) de acuerdo a criterios ocupacionales (RHCyTO: Recursos Humanos en CyT que estén ocupados en actividades clasificadas como de la CyT) y educacionales (RHCyTE: Recursos Humanos en CyT que tengan preparación de nivel técnico profesional o superior). El núcleo del Acervo lo constituyen las personas que cumplen con los dos criterios: educacional y ocupacional (RHCyTC).

Con la información utilizada en este estudio (ENE-98)⁹ y considerando las listas propuestas como núcleo, extensión y completa en el *Manual de Canberra*, se tienen las siguientes estimaciones del tamaño del Acervo de Recursos Humanos.

⁹ Instituto Nacional de Estadística Geografía e Informática (INEGI) (1998), *Encuesta Nacional de Empleo*.

Cuadro 3

TAMAÑO DEL ACERVO DE RECURSOS HUMANOS

Acervo	Lista núcleo	Lista extendida	Lista completa
ARHCyT	4,458,139	6,678,048	7,005,889
RHCyTE	4,452,056	5,281,060	5,290,515
RHCyTO	560,389	3,880,092	4,299,456
RHCyTC	554,306	2,483,104	2,584,082

El objetivo de la recomendación del *Manual de Canberra*, de clasificar a los Recursos Humanos de acuerdo con criterios, tanto educacionales como ocupacionales, es contar con una idea de los Recursos Humanos que están siendo aprovechados (RH con preparación en CyT y con ocupación en CyT) y desaprovechados (RH con preparación en actividades de CyT pero que no se ocupan en actividades clasificadas como de CyT).

De acuerdo con este criterio de aprovechamiento y las cifras anteriores, en el caso de la lista núcleo, de un total de 4,452,056 del acervo de RHCyTE sólo 554,306 se consideran también Recursos Humanos por Ocupación (RHCyTC). Es decir, se aprovechan alrededor del 12% de los RH.

En el caso de la lista extendida, de un total de 5,281,060 del acervo de RHCyTE sólo 2,483,104 se consideran también Recursos Humanos por Ocupación (RHCyTC), así que se aprovechan alrededor del 47% de los RH.

Para la lista completa, de un total de 5,290,515 del acervo de RHCyTE sólo 2,584,082 se consideran también Recursos Humanos por Ocupación (RHCyTC), de manera que se aprovechan alrededor del 49% de los RH. Si en este último caso se desagrega por nivel de instrucción, se tiene que en el nivel técnico de 722,721

personas en el acervo de RHCyTE, sólo 165,205 son RHCyTC, de manera que únicamente se aprovechan al 23% de los RHCyTE.

En el nivel de licenciatura, de 4,228,545 personas en el acervo de RHCyTE sólo 2,154,986 son RHCyTC, por lo que sólo se aprovechan alrededor del 50%; Finalmente, en el nivel de posgrado, de 339,249 personas en el acervo de RHCyTE sólo 263,891 son RHCyTC y, por lo tanto, sólo se aprovechan alrededor del 78%.

De acuerdo con esta información, se tiene un desaprovechamiento importante de los Recursos Humanos en el nivel técnico y licenciatura y en menor medida, en el nivel de posgrado.

Sin embargo, estas mediciones del Acervo sugieren, en términos muy generales, la medida en la que el Capital Humano no está aprovechado de manera óptima y, en conclusión, se requieren mediciones más precisas del aprovechamiento de los RH. Este trabajo es un esfuerzo en esa dirección.

Es importante considerar que la subutilización de los RH puede ser de diferente naturaleza. Por ejemplo, considere el caso de algún científico que trabaja como taxista y, por otro lado, alguno que trabaja en alguna institución del sistema financiero con un ingreso mayor del que podría obtener en un instituto de investigación. En ambos casos, se tienen RH que son subutilizados. Mientras en el primer caso el mercado de trabajo lo obliga a tener una ocupación que no es de CyT¹⁰; en el segundo caso, le da incentivos para tener una ocupación clasificada como de CyT, pero que no es propiamente de su área de especialización.

Por lo tanto, en el mercado de trabajo, los RHCyT enfrentan dos problemas de diferente naturaleza. El menos grave, cuando una persona está ocupada fuera de su área de especialización, pero en actividades clasificadas como de la CyT, y el grave en donde la persona no se ocupa de actividades clasificadas como de CyT.

¹⁰ Se hace el supuesto de que el taxista no trabaja en el taxi por vocación.

En este trabajo se hace un análisis sobre el empleo en la población con educación superior, a nivel de entidad federativa, con base en la Encuesta Nacional de Empleo (ENE-98).

Las preguntas a las que se quiere dar respuesta son las siguientes: ¿están los RHCyT siendo aprovechados de manera óptima? Es decir, la población con nivel de educación superior, ¿está ocupada en las áreas para las que se preparó? ¿Es variable este comportamiento entre las diferentes disciplinas y entidades federativas?

Es importante subrayar que todos los esfuerzos de formación de RH serán efectivos en la medida que estos recursos sean aprovechados de manera óptima; la formación de RH es tan importante como el aprovechamiento de los mismos.

Esto no quiere decir que algunos RH deberían dejar de formarse. Generalmente se afirma que las instituciones de educación superior deberían adecuar su oferta a la demanda de largo plazo de RH en el mercado de trabajo. Sin embargo, esta postura conllevaría algunos riesgos, pues además de tomar en cuenta la rentabilidad de las áreas de especialización, se deben considerar las demandas sociales, así como el proyecto de nación.

Capítulo 2

La Regresión Logística

La regresión lineal ha llegado a ser la herramienta más usada en cualquier análisis económico empírico concerniente a describir la relación que guarda alguna variable de interés (dependiente o respuesta) ante cambios en otras variables (explicativas o independientes). La regresión lineal es un caso particular del análisis de regresión, el cual consiste en modelar el valor medio de la variable respuesta como una función de las variables explicativas. Esto es:

$$E(Y|\mathbf{X}) = f(\mathbf{X})$$

en donde \mathbf{X} es el vector conformado por las variables explicativas y Y es la variable respuesta. La regresión lineal proviene del hecho de considerar que $f(\mathbf{X})$ es una función lineal, es decir,

$$E(Y|\mathbf{X}) = \beta_0 + \boldsymbol{\beta}'\mathbf{X}$$

en donde $\boldsymbol{\beta}$ es un vector de coeficientes asociados a cada una de las variables explicativas. Este modelo clásico de regresión lineal se basa en el siguiente conjunto de supuestos:

S1: linealidad, $y = \beta_0 + \boldsymbol{\beta}'\mathbf{X} + \varepsilon$

S2: rango completo, \mathbf{X} es $n \times K$ con rango K

S3: residual cero, $E[\varepsilon|\mathbf{X}] = 0$

S4: homoscedasticidad, $E[\varepsilon\varepsilon'|\mathbf{X}] = \sigma^2\mathbf{I}$

S5: regresores no estocásticos, \mathbf{X} es una matriz no estocástica.

S6: normalidad de los residuos, $\varepsilon|\mathbf{X} \sim N[\mathbf{0}, \sigma^2\mathbf{I}]$

Sin embargo, cuando la variable respuesta es discreta, por ejemplo, dicotómica¹¹, el análisis de regresión lineal no es apropiado. Para mostrarlo considere el siguiente modelo:

$$y = E[Y|\mathbf{X}] + (y - E[Y|\mathbf{X}]) = \beta_0 + \boldsymbol{\beta}'\mathbf{X} + \varepsilon$$

Si Y toma los valores cero o uno, el residuo ε toma los valores $-(\beta_0 + \boldsymbol{\beta}'\mathbf{X})$ o $1 - (\beta_0 + \boldsymbol{\beta}'\mathbf{X})$. De manera que $\text{var}[\varepsilon|\mathbf{X}] = (\beta_0 + \boldsymbol{\beta}'\mathbf{X})[1 - (\beta_0 + \boldsymbol{\beta}'\mathbf{X})]$ y, por lo tanto, se tendría un problema de heteroscedasticidad en los residuos, con lo que se incumple con el supuesto cuatro de los mencionados con anterioridad.

En el caso de que la variable respuesta sea discreta, el análisis de regresión que ha mostrado su utilidad en el razonamiento econométrico es la regresión logística. Este tipo de regresión proviene del caso general al considerar que $f(\mathbf{X})$ es la función de densidad de probabilidad logística:

$$f(\mathbf{X}) = \frac{e^{\beta_0 + \boldsymbol{\beta}'\mathbf{X}}}{1 + e^{\beta_0 + \boldsymbol{\beta}'\mathbf{X}}}$$

Esta forma funcional es extremadamente flexible, debido a las propiedades de la función exponencial, y provee una útil interpretación de los coeficientes como consecuencia de que la *transformación logit* es una función lineal.

$$g(x) = \log_e \left[\frac{f(\mathbf{X})}{1 - f(\mathbf{X})} \right] = \beta_0 + \boldsymbol{\beta}'\mathbf{X}$$

La regresión lineal y la regresión logística guardan algunas similitudes como compartir el espíritu, inherente al análisis de regresión general, de encontrar el modelo con mejor ajuste y parsimonia que describa la relación entre una variable respuesta y un grupo de variables explicativas. Sin embargo, ambas regresiones guardan algunas diferencias en los supuestos que se adoptan.

¹¹ Se considera variable dicotómica a aquella que ha sido codificada como cero o uno (o cualquier otro par de valores), representando la ausencia o presencia de alguna característica.

Como se mencionó con anterioridad, uno de los supuestos considerados en la regresión lineal es el llamado supuesto de normalidad. Se considera que la variable respuesta puede ser expresada como $Y = E(Y|\mathbf{X}) + \varepsilon$, en donde se supone que ε , el llamado término de error, admite una distribución normal con media cero y varianza constante a través de todos los valores de las variables independientes.

En la regresión logística, cuando la variable respuesta es dicotómica, se tiene que puede ser escrita como $y = \pi(\mathbf{X}) + \varepsilon$. En ese caso el término de error ε toma dos diferentes valores. Si $y = 1$ entonces $\varepsilon = 1 - \pi(\mathbf{X})$ con probabilidad $\pi(\mathbf{X})$, mientras que si $y = 0$, $\varepsilon = -\pi(\mathbf{X})$ con probabilidad $1 - \pi(\mathbf{X})$. De manera que ε sigue una distribución binomial con media cero y varianza $\pi(\mathbf{X})[1 - \pi(\mathbf{X})]$.

Estimación del modelo

En el caso de la regresión lineal para la estimación de los parámetros del modelo que represente de la mejor manera la información disponible, se utiliza el método de Gauss, conocido en la literatura como *Mínimos Cuadrados Ordinarios* (MCO), el cual consiste en ajustar la función lineal que minimiza el cuadrado de las distancias de las observaciones a la recta de ajuste. Bajo los supuestos mencionados arriba, los estimadores de los parámetros, con esta técnica de ajuste, tienen varias propiedades estadísticas deseables. Sin embargo, al tratar de utilizar el método de MCO en la regresión logística se observa que los estimadores no conservan estas propiedades.

Un método general de estimación es la llamada *estimación de máxima verosimilitud* que, en el caso particular de la regresión lineal, (cuando los residuos son normalmente distribuidos) conduce a la estimación por MCO. Este método provee los elementos para realizar la estimación en el caso de la regresión logística. En términos generales, este método estima los valores de los parámetros que maximizan la probabilidad de obtener el conjunto de valores observados.

Si la variable dependiente Y es codificada como cero o uno, se tiene que $P(Y = 1 | \mathbf{X}) = \pi(\mathbf{X}) = F(\mathbf{X}'\boldsymbol{\beta})$, y $P(Y = 0 | \mathbf{X}) = 1 - \pi(\mathbf{X})$, por lo que cualquier "par" (\mathbf{x}_i, y_i)

contribuye a la función de verosimilitud con el término $\zeta(\mathbf{x}_i) = \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i}$, de manera que la función de verosimilitud está dada por:

$$v(\boldsymbol{\beta}) = \prod_{i=1}^n \zeta(\mathbf{x}_i) = \prod_{i=1}^n F(\mathbf{X}'\boldsymbol{\beta})^{y_i} [1 - F(\mathbf{X}'\boldsymbol{\beta})]^{1-y_i}.$$

Los estimadores de máxima verosimilitud son elegidos como aquellos que maximizan esta función. Aunque se considera al logaritmo natural de esta función por ser más fácil de tratar.

$$L(\boldsymbol{\beta}) = \log_e(v(\boldsymbol{\beta})) = \sum_{i=1}^n \{y_i \log_e [F(\boldsymbol{\beta}\mathbf{x}_i)] + (1 - y_i) \log_e [1 - F(\boldsymbol{\beta}\mathbf{x}_i)]\}$$

Para encontrar el valor que maximiza $L(\boldsymbol{\beta})$ se calculan las derivadas con respecto a cada β_i y se iguala cada una de las expresiones a cero, de manera que al considerar la función de distribución logística F con función de densidad f se tiene que:

$$\frac{\partial \log_e L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \left[\frac{y_i f_i}{F_i} + (1 - y_i) \frac{-f_i}{(1 - F_i)} \right] \mathbf{x}_i = 0$$

tras algunas manipulaciones se llega a:

$$\frac{\partial \log_e L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n (y_i - F_i) \mathbf{x}_i = 0$$

En el análisis de regresión lineal, estas ecuaciones son lineales en los parámetros y, por tanto, fácilmente solubles. Sin embargo, en el caso de la regresión logística, estas últimas ecuaciones son no lineales en los parámetros y, por tanto, se requieren métodos especiales para su solución. Estos métodos son iterativos y se encuentran disponibles en algunos paquetes de *software* de análisis estadístico. Uno de los más utilizados es el método de Newton que, generalmente, converge al máximo con muy pocas iteraciones.

Al calcular la matriz de segundas derivadas, es decir, el Hessiano:

$$\frac{\partial^2 \log_e L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = - \sum_{i=1}^n F_i(1-F_i) \mathbf{x}_i \mathbf{x}_i'$$

se tiene que esta matriz es siempre definida negativa. Por lo que la función de verosimilitud es una función globalmente cóncava y la condición de derivada igual a cero asegura la detección de un máximo.

Efectos marginales

Tanto la regresión lineal como la regresión logística se obtienen a través del modelo de regresión general

$$E(Y|\mathbf{X}) = F(\boldsymbol{\beta}'\mathbf{X})$$

al considerar una función de distribución F con función de densidad f .

Por lo que, en general, los efectos marginales están dados por:

$$\frac{\partial}{\partial \mathbf{X}} E(Y|\mathbf{X}) = \frac{\partial F(\boldsymbol{\beta}'\mathbf{X})}{\partial (\boldsymbol{\beta}'\mathbf{X})} \boldsymbol{\beta} = f(\boldsymbol{\beta}'\mathbf{X}) \boldsymbol{\beta}$$

En el caso de la regresión lineal se tiene que $f(\boldsymbol{\beta}'\mathbf{X})=1$ por lo que $\frac{\partial}{\partial \mathbf{X}} E(Y|\mathbf{X}) = \boldsymbol{\beta}$, de manera que la $\boldsymbol{\beta}$ resulta ser el efecto marginal, el cual es independiente de \mathbf{X} . Este resultado ha hecho que el modelo de regresión sea ampliamente utilizado por su facilidad de interpretación.

Pero en esta propiedad no se cumple en el caso de la regresión logística, en ese caso se tiene que el efecto marginal está dado por:

$$\frac{\partial}{\partial \mathbf{X}} E(Y|\mathbf{X}) = \frac{\partial F(\boldsymbol{\beta}'\mathbf{X})}{\partial (\boldsymbol{\beta}'\mathbf{X})} \boldsymbol{\beta} = f(\boldsymbol{\beta}'\mathbf{X})[1-f(\boldsymbol{\beta}'\mathbf{X})]\boldsymbol{\beta}$$

Por lo que el efecto marginal depende de \mathbf{X} y, por lo tanto, la interpretación de la β es diferente al caso de la regresión lineal. Es frecuente que la derivada anterior sea calculada en algún valor de interés del vector \mathbf{X} como podría ser el vector conformado por los valores medios de las variables.

Cuando la variable independiente es dicotómica se tiene una transformación que es útil para facilitar la interpretación de la β .

Suponga que y es una variable dependiente dicotómica y x es una variable independiente también dicotómica.

	$x = 1$	$x = 0$
$y = 1$	$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
$y = 0$	$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$

De acuerdo con la tabla anterior, la razón de probabilidad entre éxito o fracaso en la variable dependiente, cuando la variable independiente toma el valor 1, es $\frac{\pi(1)}{1 - \pi(1)}$; de la misma manera cuando la variable independiente toma el valor cero la razón sería $\frac{\pi(0)}{1 - \pi(0)}$. Si definimos la razón de momios (odds-ratio) como

$$\psi = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}}$$

tenemos que el logaritmo es igual a la diferencia de la transformación *logit*, es decir:

$$\log(\psi) = g(1) - g(0)$$

pero

$$\psi = \frac{\left(\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}\right)\left(\frac{1}{1 + e^{\beta_0}}\right)}{\left(\frac{e^{\beta_0}}{1 + e^{\beta_0}}\right)\left(\frac{1}{1 + e^{\beta_0 + \beta_1}}\right)} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Este resultado ha hecho famosa a la regresión logística para las cuestiones de epidemiología. El ejemplo clásico para entender esta razón de momios (odds-ratio), ψ , es considerar a un individuo con variable dependiente: tener o no tener cáncer y como variable explicativa: fumar o no fumar. Afirar que fumar es uno de los condicionantes del cáncer es resultado de un modelo *logit* en donde se obtuvieron momios mayores que uno.

Capítulo 3

Una regresión logística para medir el aprovechamiento de los Recursos Humanos en Ciencia y Tecnología

El análisis presenta varias limitaciones. Una de ellas es el hecho de no tomar en cuenta las preferencias de los individuos. Por ejemplo, el caso del Actuario que obtiene una mayor satisfacción vendiendo seguros no es diferenciado. La fuente de información considerada no permite determinar esta cuestión, se tiene información de la escolaridad y la ocupación de la población, pero no de las motivaciones por las que se tiene dicha ocupación. Se necesitaría una variable que midiera la orientación vocacional para saber si la persona está fuera de su área de especialidad por su gusto y no por restricciones en el mercado laboral. Pero con los datos disponibles, dicha variable no se puede construir.

Otra limitación importante es el no disponer de información acerca del ingreso patrimonial o no laboral de las personas, como pueden ser herencias, pensiones, becas, subsidios familiares, etcétera. Así como esta información ha sido importante para determinar la oferta laboral (una persona puede trabajar menos si sabe que recibirá una cierta cantidad de dinero proveniente de sus parientes), seguramente influirá en la decisión de tomar un empleo que no es de su área de especialidad (si la persona sabe que recibirá una cierta cantidad de dinero de sus parientes puede esperar hasta ocuparse en su área de formación, en caso contrario podría no tener alternativa y ocuparse en áreas ajenas).

Hernández (1996)¹² sugiere aproximar esta información con la Encuesta Nacional de Ingreso Gasto de los Hogares (ENIG), ajustando un modelo *Tobit* del ingreso patrimonial y con éste clasificar a las familias como pobres y ricas y, con esta aproximación, estimar la situación de los entrevistados en la ENE.

¹²Hernández Gonzalo, *Efectos de la Pobreza Familiar Sobre la Participación en el Mercado de Trabajo, las Horas Trabajadas y el Desempleo en México*, Tesis Doctoral. Universidad de Oxford, 1996.

Sin embargo, el autor mismo señala las limitaciones del análisis debido a los supuestos que hay que hacer. Uno de ellos es suponer que el ingreso no laboral de la familia se distribuye equitativamente entre sus miembros, y aunque es un supuesto necesario es demasiado fuerte. Además de los supuestos, hay que hacer correcciones en las estimaciones de cualquier modelo, pues éstas se harán a su vez con estimadores.

La información que sería relevante es el nivel de ingreso no laboral de las personas. Con las recomendaciones anteriores, sólo se clasifica a la población en dos casos: pobres y ricos. La clasificación es tan amplia y subjetiva que no aportaría mucho en el análisis. Por lo que se decidió no hacerlo y enfrentar la limitación de tener esta variable omitida, en lugar de enfrentar el problema de hacer estimaciones con estimadores.

Para estudiar en qué medida la población con educación superior se ocupa en cuestiones propias de su área de especialidad, es necesario clasificar a los individuos entre los que están ocupados en cuestiones para las que se formaron y quienes no lo hacen.

Existen principalmente dos metodologías para hacer esta clasificación. Una forma es diseñar una encuesta donde se le pregunta al individuo si está ocupado en las cuestiones para las que se educó. Otra, es hacer una clasificación con base a un catálogo de requerimientos de educación para cada ocupación o un análisis estadístico de la concentración de requerimientos de educación para cada ocupación.

En este trabajo se utiliza la base de datos de la Encuesta Nacional de Empleo (ENE-98) como fuente de información. En ésta encuesta se tiene información sobre la escolaridad tanto en nivel de instrucción, como en área de conocimiento y años aprobados, así como la ocupación del entrevistado. Dicha información está clasificada de acuerdo al Catálogo Nacional de Ocupaciones y al Catálogo Nacional de Escolaridad.

Ya que en la encuesta no se le preguntó al entrevistado si considera que está ocupado en su área de especialización de acuerdo a su educación, se construyó una variable dicotómica: *oae* (ocupado en su área de especialidad), que toma los valores **1** si lo está y **0** si no es el caso. La clasificación es subjetiva y de acuerdo al criterio del autor. No se optó por una clasificación estadística por que se pueden tener imprecisiones de lo que se quiere medir. Por ejemplo, si un gran número de contadores se ocupan como taxistas, un criterio estadístico consideraría la ocupación de taxista como propia de los contadores, dado que sería la ocupación de un gran número de ellos. Por lo anterior, se optó por la clasificación subjetiva.

Sería ideal que se hiciera una discusión entre especialistas de diferentes formaciones y opiniones de las ocupaciones que son propias de todas y cada una de las áreas de educación para llegar a un consenso, y con esto obtener un manual o catálogo para clasificar a los individuos en ocupados en su área de especialidad o no.

En este estudio se analizan por separado tres diferentes niveles de educación, el nivel de técnico profesional, el de licenciatura y el de posgrado para todas y cada una de las disciplinas del Catálogo Nacional de Escolaridad. Se consideraron las ocupaciones del Catálogo Nacional de Ocupaciones para construir la variable *oae*. Esta variable se construyó con la población ocupada excluyendo iniciadores, y todos aquellos casos donde la información no estaba suficientemente especificada o donde la variable *oae* no se podía construir.

En el nivel técnico, esta variable toma los siguientes valores: $oae = 0$ en 292,729 casos; mientras que $oae = 1$ en 244,727 casos. Por lo tanto, de acuerdo con la clasificación utilizada, sólo 46% de la población en el nivel técnico se encuentra ocupado en las áreas para las que se preparó.

En el nivel de licenciatura la variable *oae* toma los siguientes valores: $oae = 0$ en 1,135,431 casos; mientras que $oae = 1$ en 2,241,869 casos. Por lo tanto, sólo el 66% de la población en el nivel de licenciatura se encuentra ocupada en las áreas para las que se preparó.

En el nivel de posgrado, la variable *OAE* toma los siguientes valores: $oae = 0$ en 35,378 casos; mientras que $oae = 1$ en 262,642 casos. Por lo tanto, el 88% de la población en el nivel de posgrado se encuentra ocupada en las áreas para las que se preparó.

3.1 Variables del modelo

Debido a la útil interpretación de la β de la regresión logística en términos de la razón de momios (odds-ratio) a la que se hace mención en la parte final del capítulo anterior, en este trabajo se consideran puras variables dicotómicas.

La variable dependiente del modelo de regresión logística es: *oae*, las variables independientes que resultaron significativas en el análisis se definen a continuación. Para todas las variables del modelo se construyeron variables dicotómicas de diseño que toman el valor 0 o 1. En algunos casos, estas variables no tienen el número suficiente de observaciones (30), no son de interés o son los casos donde la información no está suficientemente especificada. Por lo tanto, estas variables no son interpretadas. Sin embargo, son incorporadas en el modelo para aislar estos casos.

1. Edad

Algunas personas argumentan que el hecho de que una persona esté fuera de su área de especialidad puede ser un problema de inexperiencia y, por lo tanto, se remedia con la edad. Para considerar la edad, se construyó la variable *ge*: grupos de edad. No se consideró apropiado incorporar la edad directamente en el modelo, pues las diferencias por año no son tan importantes. Tener 26 o 27 años o tener 45 y 46 años no hacen gran diferencia, para lo que se quiere medir se requiere considerar diferencias más amplias.

Cuadro 4

DEFINICIÓN DE LA VARIABLE: GRUPOS DE EDAD

<i>ge</i>	Grupos de edad
1	Menos de 30 años
2	30 a 39
3	40 a 49

4	50 a 59
5	60 a 69
6	Más de 70

Para modelar la edad en el ámbito de la regresión logística, se consideraron seis variables dicotómicas de diseño: $ge_i = 1$ si la persona pertenece al grupo $i = 1, \dots, 6$ y $ge_i = 0$ en otro caso.

2. Sexo

Para analizar si el estar o no en su área de especialidad es un comportamiento diferente entre hombres y mujeres se toma en cuenta el sexo del entrevistado. Se consideraron dos variables dicotómicas de diseño: $sex_1 = 1$, si el entrevistado es hombre y, análogamente, para las mujeres se construyó la dicotómica sex_2 .

3. Hijos

En el caso de las mujeres se consideró una variable de interacción: el número de hijos.

Cuadro 5
DEFINICIÓN DE LA VARIABLE: NÚMERO DE HIJOS

h	Hijos
0	Ninguno
1	1 o 2
2	3 a 5
3	Más de 5
4	No especificado

Se atendieron cinco variables dicotómicas de diseño, h_i para considerar los cinco casos anteriores, pero en el modelo se trataron las variables de interacción: $hij_i = sex_2 * h_i$, $i = 0, \dots, 3$ pues la información del número de hijos sólo es contabilizada en el caso de las mujeres.

4. Estado civil

La variable estado civil toma los siguientes valores:

Cuadro 6
DEFINICIÓN DE LA VARIABLE: ESTADO CIVIL

<i>ec</i>	Estado civil
1	Soltero
2	Casado
3	Unión libre
4	Divorciado
5	Separado
6	Viudo
9	No especificado

Se construyeron cuatro variables de diseño: $ecv_1 = 1$ en el caso de soltero; $eciv_2 = 1$ para casado o en unión libre; $eciv_3 = 1$ para divorciado o separado y $eciv_4 = 1$ para viudo y no especificado.

5. Posición en el hogar

Se tomaron en cuenta los siguientes casos:

Cuadro 7
DEFINICIÓN DE LA VARIABLE: POSICIÓN EN EL HOGAR

<i>pos</i>	Posición en el hogar
1	Jefe de familia
2	Cónyuge
3	Hijo
4	Otro o no especificado

Se consideraron cuatro variables de diseño, $pos_i = 1$ para cada uno de los casos anteriores.

6. Estado

Para analizar si el comportamiento de la variable *oae* es diferente entre las entidades federativas se consideraron 32 variables de diseño, $edo_i = 1$, para cada uno de los 32 estados ordenados alfabéticamente.

Cuadro 8
DEFINICIÓN DE LA VARIABLE: ESTADO

<i>edo</i>	Estado
1	Aguascalientes
2	Baja California Norte
3	Baja California Sur
4	Campeche
5	Coahuila
6	Colima
7	Chiapas
8	Chihuahua
9	Distrito Federal
10	Durango
11	Guanajuato
12	Guerrero
13	Hidalgo
14	Jalisco
15	México
16	Michoacán
17	Morelos
18	Nayarit
19	Nuevo León
20	Oaxaca
21	Puebla
22	Querétaro
23	Quintana Roo
24	San Luis Potosí
25	Sinaloa
26	Sonora
27	Tabasco
28	Tamaulipas
29	Tlaxcala
30	Veracruz
31	Yucatán
32	Zacatecas

7. Campo de Conocimiento

Uno de los objetivos de este estudio es analizar la variable *oae* entre las diferentes áreas de conocimiento. Para cada nivel de instrucción se consideraron todas las

disciplinas del Catálogo de Escolaridad, sin considerar los casos en donde la información estaba insuficientemente especificada o se tenía un número reducido de observaciones.

En el nivel técnico se atendieron 24 variables de diseño, correspondientes a cada disciplina.

Cuadro 9

DEFINICIÓN DE LA VARIABLE: CAMPO DE CONOCIMIENTO EN EL NIVEL TÉCNICO

Esc	Campo de la ciencia y disciplina
	Ciencias Agropecuarias
2 021	TÉCNICAS AGROPECUARIAS, PESQUERAS, FORESTALES Y AMBIENTALES
	Ciencias de la Salud
3 031	TÉCNICAS DE ENFERMERIA AUXILIAR
4 032	TÉCNICAS DENTALES
5 033	OTRAS TÉCNICAS RELACIONADAS CON LA SALUD
6 034	TÉCNICAS ASISTENCIALES
	Ciencias Naturales y Exactas
7 041	TÉCNICAS APLICADAS A LAS CIENCIAS QUÍMICAS Y AFINES
11 054	TÉCNICAS MATEMÁTICAS Y ESTADÍSTICAS
	Ciencias Sociales
8 051	TÉCNICAS CONTABLES Y FINANCIERAS
9 052	TÉCNICAS ECONÓMICO-ADMINISTRATIVAS Y DE ARCHIVO
10 053	TÉCNICAS SECRETARIALES
12 061	TÉCNICAS DE LA PUBLICIDAD, COMUNICACIÓN, PERIODISMO, Y MERCADOTECNIA
13 062	TÉCNICAS EN TURISMO
16 081	TÉCNICAS EDUCATIVAS
23 111	TÉCNICAS EN INSTRUCCIÓN MILITAR
24 112	TÉCNICAS EN INSTRUCCIÓN POLICIAL
	Ingeniería y Tecnología
1 011	TÉCNICAS EN DIBUJO, DISEÑO Y DECORACIÓN
17 091	TÉCNICAS DE LA CONSTRUCCIÓN
18 092	TÉCNICAS DE LA EXPLORACIÓN Y EXPLOTACIÓN DE LOS RECURSOS NATURALES NO RENOVABLES
19 093	TÉCNICAS ELECTRÓNICAS Y DE COMPUTACION
20 094	TÉCNICAS EN PRODUCCIÓN INDUSTRIAL, ARTESANAL, DE REPARACIÓN Y MANTENIMIENTO
21 095	TÉCNICAS EN LA OPERACIÓN DE COMUNICACIONES Y TRANSPORTES
	Humanidades y Otros
14 063	IDIOMAS
15 071	TÉCNICAS ARTÍSTICAS
22 101	TÉCNICAS EN SERVICIOS PERSONALES

En el nivel licenciatura se consideraron 53 variables de diseño

Cuadro 10
DEFINICIÓN DE LA VARIABLE: CAMPO DE CONOCIMIENTO EN EL NIVEL
LICENCIATURA

Esc	Campo de la ciencia y disciplina
	Ciencias Agropecuarias
31 231	AGRONOMÍA
32 232	VETERINARIA Y ZOOTECNIA
33 233	INGENIERIA PESQUERA
34 234	FORESTALES
	Ciencias de la Salud
35 241	MEDICINA GENERAL Y AFINES
36 242	ODONTOLOGÍA Y ESPECIALIDADES DENTALES
37 243	EMFERMERÍA
38 244	NUTRICIÓN
39 245	BIOMÉDICAS
	Ciencias Naturales y Exactas
28 221	BIOLOGÍA, BOTÁNICA Y ZOOLOGÍA
29 222	ECOLOGÍA Y CIENCIAS ATMOSFERICAS
30 223	CIENCIAS DEL MAR
46 261	QUÍMICA, INGENIERIA QUÍMICA Y QUÍMICA INDUSTRIAL
47 262	QUÍMICA FARMACEUTICA BIOLÓGICA
48 263	BIOQUÍMICA
76 321	MATEMÁTICAS, ESTADÍSTICA Y ACTUARÍA
77 322	FÍSICA Y ASTRONOMÍA
	Ciencias Sociales
44 255	PSICOLOGÍA
49 271	SOCIOLOGÍA, TRABAJO SOCIAL Y AFINES
50 272	CIENCIAS POLÍTICAS Y ADMINISTRACIÓN PÚBLICA
51 273	CIENCIAS DE LA COMUNICACIÓN
52 274	DERECHO
53 275	GEOGRAFÍA
57 291	ECONOMÍA
58 292	ADMINISTRACIÓN
59 293	CONTADURÍA Y AFINES
60 294	TURISMO
61 295	ARCHIVONOMÍA Y BIBLIOTECONOMÍA
62 296	MERCADOTÉCNIA
63 301	FORMACIÓN DOCENTE EN EDUCACIÓN BASICA
64 302	FORMACIÓN DOCENTE EN EDUCACIÓN MEDIA Y NORMAL
65 303	PEDAGOGÍA Y CIENCIAS DE LA EDUCACION
66 304	FORMACION DOCENTE EN EDUCACION ESPECIAL
67 305	EDUCACION FISICA Y DEPORTES
68 306	EDUCACION MUSICAL, DANZA Y CANTO
	Ingeniería y Tecnología
25 211	ARQUITECTURA Y URBANISMO
26 212	DISEÑO INDUSTRIAL Y DE INTERIORES
27 213	DISEÑO GRAFICO
69 311	INGENIERIA CIVIL Y DE LA CONSTRUCCION
70 312	INGENIERIA EXTRACTIVA, METALURGICA Y ENERGETICA
71 313	INGENIERIA EN COMPUTACION E INFORMATICA
72 314	INGENIERIA ELECTRICA Y ELECTRONICA
73 315	INGENIERIA MECANICA E INDUSTRIAL
74 316	INGENIERIA DEL TRANSPORTE AERONAUTICA, NAVAL Y PILOTOS AVIADORES
75 317	INGENIERIA TOPOGRAFICA, HIDRAULICA, GEOLOGICA Y GEODESTA

Humanidades y Otros		
40	251	FILOSOFIA
41	252	HISTORIA
42	253	ANTROPOLOGIA Y ARQUEOLOGÍA
43	254	LETRAS LITERATURA E IDIOMAS
45	256	TEOLOGIA Y RELIGIÓN
54	281	ARTES PLASTICAS
55	282	MUSICA Y DANZA
56	283	TEATRO Y CINEMATOGRAFIA

Finalmente, en el nivel de posgrado se trataron 49 variables de diseño

Cuadro 11

DEFINICIÓN DE LA VARIABLE: CAMPO DE CONOCIMIENTO EN EL NIVEL POSGRADO

Esc	Campo de la ciencia y disciplina
Ciencias Agropecuarias	
84	431 AGRONOMIA
85	432 VETERINARIA Y ZOOTECNIA
86	433 INGENIERIA PESQUERA
Ciencias de la Salud	
87	441 MEDICINA GENERAL Y AFINES
88	442 ODONTOLOGÍA
89	443 ENFERMERIA
90	444 NUTRICION
91	445 BIOMÉDICAS
Ciencias Naturales y Exactas	
81	421 BIOLOGIA, BOTANICA Y ZOOLOGÍA
82	422 ECOLOGÍA Y CIENCIAS ATMOSFERICAS
83	423 CIENCIAS DEL MAR
98	461 QUIMICA, INGENIERIA QUIMICA Y QUIMICA INDUSTRIAL
99	462 QUIMICA FARMACEUTICA BIOLÓGICA
100	463 BIOQUÍMICA
125	521 MATEMATICAS, ESTADISTICA Y ACTUARIA
126	522 FISICA Y ASTRONOMIA
Ciencias Sociales	
96	455 PSICOLOGÍA
101	471 SOCIOLOGIA, TRABAJO SOCIAL Y AFINES
102	472 CIENCIAS POLITICAS Y ADMINISTRACION PUBLICA
103	473 CIENCIAS DE LA COMUNICACIÓN
104	474 DERECHO
105	475 GEOGRAFIA
108	491 ECONOMIA
109	492 ADMINISTRACIÓN
110	493 CONTADURIA Y AFINES
111	494 TURISMO
112	495 ARCHIVONOMIA Y BIBLIOTECONOMIA
113	496 MERCADOTECNIA
114	501 FORMACION DOCENTE EN EDUCACION BASICA
115	502 FORMACION DOCENTE EN EDUCACION MEDIA Y NORMAL
116	503 PEDAGOGIA Y CIENCIAS DE LA EDUCACIÓN
117	504 FORMACION DOCENTE EN EDUCACION ESPECIAL
Ingeniería y Tecnología	
78	411 ARQUITECTURA Y URBANISMO
79	412 DISEÑO INDUSTRIAL Y DE INTERIORES

80	413	DISEÑO GRAFICO
118	511	INGENIERIA CIVIL Y DE LA CONSTRUCCIÓN
119	512	INGENIERIA EXTRACTIVA, METALURGICA Y ENERGETICA
120	513	INGENIERIA EN COMPUTACION E INFORMATICA
121	514	INGENIERIA ELECTRICA Y ELECTRÓNICA
122	515	INGENIERIA MECANICA E INDUSTRIAL
123	516	INGENIERIA DEL TRANSPORTE AERONAUTICA, NAVAL Y PILOTOS AVIADORES
124	517	INGENIERIA TOPOGRAFICA, HIDRAULICA, GEOLOGICA Y GEODESTA
Humanidades y Otros		
92	451	FILOSOFIA
93	452	HISTORIA
94	453	ANTROPOLOGIA Y ARQUEOLOGIA
95	454	LETRAS, LITERATURA E IDIOMAS
97	456	TEOLOGIA Y RELIGIÓN
106	481	ARTES PLASTICAS
107	482	MUSICA Y DANZA

8. Rama de actividad económica

Para diferenciar el comportamiento de los entrevistados entre las distintas ramas de actividad económica se manejaron 76 variables de diseño, de acuerdo con la Clasificación de Actividad Económica considerada en las Encuestas de Empleo (CAE-ENEU) realizada por el INEGI.

Cuadro 12
DEFINICIÓN DE LA VARIABLE: RAMA DE ACTIVIDAD ECONÓMICA

Ram

Gran División 1: Agropecuario, Silvicultura y Pesca

- 01 AGRICULTURA
- 02 GANADERIA
- 03 SILVICULTURA
- 04 CAZA, PESCA Y ACUACULTURA

Gran División 2: Minería

- 05 EXTRACCION Y BENEFICIO DE CARBON DE GRAFITO Y SUS DERIVADOS
- 06 EXTRACCION DE PETROLERO CRUDO Y GAS NATURAL
- 07 EXTRACCION Y/O BENEFICIO DE MINERAL DE HIERRO
- 08 EXTRACCIÓN, BENEFICIO, FUNDICION Y REFINACION DE MINERALES METALICOS NO FERROSOS
- 09 EXPLOTACION DE CANTERAS Y EXTRACCION DE ARENA, GRAVA, ARCILLA Y PIEDRA
- 10 EXTRACCION Y BENEFICIO DE OTROS MINERALES NO METALICOS

Gran División 3: Industria Manufacturera

- 11 CARNES Y LACTEOS
- 12 PREPARACION DE FRUTAS Y LEGUMBRES
- 13 MOLIENDA DE TRIGO Y OTROS CEREALES
- 14 MOLIENDA DE NIXTAMAL
- 15 BENEFICIO Y MOLIENDA DE CAFÉ, INCLUYE EL ENVASADO DE TÉ
- 16 AZÚCAR
- 17 ACEITES Y GRASAS COMESTIBLES
- 18 ALIMENTOS PARA ANIMALES
- 19 OTROS PRODUCTOS
- 20 BEBIDAS ALCOHOLICAS

- 21 CERVEZA Y MALTA
- 22 REFRESCOS, AGUAS GASEOSAS Y PURIFICADAS
- 23 TABACO
- 24 HILADOS Y TEJIDOS DE FIBRAS BLANDAS
- 25 HILADOS Y TEJIDOS DE FIBRAS DURAS
- 26 OTRAS INDUSTRIAS TEXTILES
- 27 PRENDAS DE VESTIR Y OTROS ARTICULOS DE PUNTO
- 28 CUERO, CALZADO, Y OTROS ARTICULOS DE PIEL
- 29 ASERRADEROS, TRIPLAY Y TABLEROS
- 30 FABRICACION DE OTROS PRODUCTOS DE MADERA Y CORCHO
- 31 FABRICACION DE PAPEL Y CARTON
- 32 IMPRENTA Y EDITORIALES
- 33 INDUSTRIA DEL PETROLEO
- 34 PETROQUÍMICA BÁSICA
- 35 QUÍMICA BÁSICA
- 36 ABONOS, FERTILIZANTES Y SEMILLAS MEJORADAS
- 37 RESINAS SINTÉTICAS Y FIBRAS ARTIFICIALES
- 38 PRODUCTOS FARMACEUTICOS
- 39 JABONES, DETERGENTES Y COSMÉTICOS
- 40 OTROS PRODUCTOS QUÍMICOS
- 41 PRODUCTOS DE HULE
- 42 ARTÍCULOS DE PLÁSTICO
- 43 VIDRIO Y PRODUCTOS DE VIDRIO
- 44 FABRICACIÓN DE CEMENTO Y CONCRETO PREMEZCLADO
- 45 PRODUCTOS A BASE DE MINERALES NO METÁLICOS
- 46 INDUSTRIAS BÁSICAS DEL HIERRO Y EL ACERO
- 47 INDUSTRIAS BASICAS DE METALES NO FERROSOS
- 48 MUEBLES METÁLICOS
- 49 PRODUCTOS METÁLICOS ESTRUCTURALES
- 50 OTROS PRODUCTOS METÁLICOS, EXCEPTO MAQUINARIA
- 51 MAQUINARIA Y EQUIPO
- 52 FABRICACIÓN Y ENSAMBLE DE MAQUINARIA, EQUIPO Y APARATOS ELECTRICOS PARA LA INDUSTRIA
- 53 APARATOS ELECTRODOMESTICOS
- 54 EQUIPOS Y APARATOS ELECTRONICOS
- 55 EQUIPOS Y APARATOS ELECTRICOS DE USO GENERAL
- 56 FABRICACIÓN Y/O ENSAMBLE DE AUTOMÓVILES, INDUSTRIA AUTOMOTRIZ
- 57 CARROCERÍAS METÁLICAS, MOTORES, PARTES Y ACCESORIOS PARA AUTOMÓVILES
- 58 EQUIPO Y MATERIAL DE TRANSPORTE, EXCEPTO PARA AUTOMOTORES
- 59 OTRAS INDUSTRIAS MANUFACTURERAS
- Gran División 4: Construcción**
- 60 CONSTRUCCION
- Gran División 5: Electricidad Gas y Agua Potable**
- 61 ELECTRICIDAD, GAS Y AGUA POTABLE
- Gran División 6: Comercios Restaurantes y Hoteles**
- 62 COMERCIO
- 63 RESTAURANTES Y HOTELES
- Gran División 7: Transportes, Almacenamientos, y Comunicaciones**
- 64 TRANSPORTES
- 65 COMUNICACIONES
- Gran División 8: Servicios Financieros, Seguros y bienes inmuebles**
- 66 SEVICIOS FINANCIEROS
- 67 ALQUILER DE INMUEBLES
- Gran División 9: Servicios Comunales, Sociales y Personales**
- 68 SRVICIOS PROFESIONALES Y TECNICOS ESPECIALIZADOS
- 69 SERVICIOS DE EDUCACIÓN
- 70 SERVICIOS MEDICOS, DE SALUD Y ASISTENCIA
- 71 SERVICIOS DE ESPARCIMIENTO, RECREATIVOS Y DEPORTIVOS

9. Nivel de ingreso laboral

Es importante considerar el nivel de ingreso laboral para inspeccionar si el hecho de estar fuera de su área de especialidad se debe al estímulo monetario. Esto permitirá ubicar si se trata de un problema de subempleo. Es decir, si está fuera de su área de especialidad por obtener mayores ingresos o no. Se consideraron los siguientes casos.

Cuadro 13

DEFINICIÓN DE LA VARIABLE: NIVEL DE INGRESO LABORAL

<i>sal</i>	Nivel de ingreso laboral
1	Hasta 4 salarios mínimos
2	Más de 4 hasta 8
3	Más de 8 hasta 12
4	Más de 12 hasta 16
5	Más de 16 hasta 20
6	Más de 20 a 25
7	Más de 25 hasta 30
8	Más de 30
9	No especificado o no recibe ingreso

Al igual que en las demás variables se dispuso de nueve variables de diseño para contemplar estos casos.

3.2 El modelo

Para medir si la población considerada está en su área de especialidad o no, se propone un modelo de respuesta cualitativa *logit* para medir la probabilidad de tal evento. Esta probabilidad puede ser interpretada de dos maneras. Una es considerar que, para un individuo con vector de características X , representa la probabilidad esperada de estar en su área de especialidad. Otra interpretación, la cual se utiliza en este estudio, es considerar esta probabilidad como la situación que se tiene en los entrevistados. Por ejemplo, si la probabilidad es 0.3 puede interpretarse como que sólo el 30% de la población con características X se ocupa en su área de especialidad y el 70% no lo hace.

Para cada individuo sea $\pi(\mathbf{x}) = \Pr(oae = 1 | \mathbf{X})$, es decir, la probabilidad de estar ocupado en su área de especialidad. El vector \mathbf{x} está formado por las 10 variables consideradas en el modelo, cada una de las cuales está conformada por variables dicotómicas (0,1) de acuerdo con las definiciones dadas con anterioridad. Se considera que esta probabilidad admite una distribución logística.

El *logit* del modelo de regresión logística, en el caso general, está dado por:

$$g(\mathbf{x}) = \beta_0 + \sum_i \beta_i ge_i + \sum_j \beta_j sex_j + \sum_k \beta_k hij_k + \sum_l \beta_l eciv_l + \sum_m \beta_m pos_m + \sum_n \beta_n edo_n + \sum_o \beta_o esc_o + \sum_p \beta_p ram_p + \sum_q \beta_q sct_q + \sum_r \beta_r sal_r$$

y entonces la probabilidad es:

$$\pi(\mathbf{X}) = \frac{e^{g(\mathbf{X})}}{1 + e^{g(\mathbf{X})}}$$

Los sumandos del *logit* representan las sumas que consideran a las variables de diseño de cada una de las 10 variables del modelo. Por ejemplo, $\sum_i \beta_i ge_i = \beta_1 ge_1 + \dots + \beta_6 ge_6$ atiende las seis dicotómicas que representan a cada uno de los seis grupos de la variable edad definidos con anterioridad.

Es importante mencionar que para la estimación del modelo, algunas variables quedaron fuera del análisis para evitar problemas de colinealidad, además de las que se tuvieron que excluir para ser consideradas como casos de referencia.

Para cada uno de los niveles de instrucción se realizó la estimación por el método de máxima verosimilitud. Las estimaciones de los modelos se realizaron con el paquete estadístico y econométrico *STATA*¹³, los resultados de dichas estimaciones se muestran en el anexo 2, en todos los casos se presentan en términos de la razón de momios (odds-ratio), $\psi = e^\beta$, expuesta al final del tercer capítulo.

Para añadir o eliminar variables al modelo se realizaron pruebas de razón de verosimilitud (log-likelihood). Además, se probó la significancia de cada uno de los

¹³ StataCorp. 1999, *Stata Statistical Software: Release 6.0*, College Station, TX: Stata Corporation.

coeficientes de cada variable por medio de la prueba de Wald. También se realizaron pruebas de bondad de ajuste con la χ^2 de Pearson, la prueba de Hosmer-Lemeshow, tablas de predictibilidad, así como consideraciones con la pseudo- R^2 .

Como ya se mencionó, fueron dejadas fuera del análisis ciertas variables (con sus respectivas observaciones), pues se tenía un problema de colinealidad entre estas y el resto de las consideradas en el modelo.

Resultados

A continuación se presentan los resultados de las predicciones de probabilidad obtenidos en cada uno de los niveles de instrucción.

▪ Nivel técnico

Como se esperaba, la edad tiene un papel importante para explicar el hecho de que la población esté ocupada en su área de especialidad. En el caso de la población menor de 30 años, sólo el 0.39 se ocupaba en cuestiones para las que se preparó. En el caso de la población entre 30 y 39 años esta cifra aumenta hasta 0.50; para edades de entre 40 y 49 años es 0.62; mientras que para edades de 50 a 59 la cifra es 0.79. Por lo tanto, en el nivel técnico, el hecho de no estar ocupado en su área de especialidad puede ser explicado en parte por la falta de experiencia. Sin embargo, no desaparece con la edad.

Las mujeres resultaron estar ocupadas en su área de especialidad en mayor medida que los hombres. En promedio, el 0.43 de los hombres está ocupado en su área de especialidad, mientras que para las mujeres lo hace el 0.50.

La variable de interacción: número de hijos, indicó que para las mujeres que tienen uno o dos hijos, la probabilidad de estar en su área de especialidad disminuye en un 5%, mientras que si el número de hijos se eleva de tres a cinco la probabilidad disminuye hasta un 45%.

Para el estado civil se encontró que, para los solteros, la probabilidad de estar ocupado en su área de especialidad aumenta alrededor de un 30%, con respecto a la

situación de casado o en unión libre. Mientras que en el caso de los separados o divorciados la probabilidad disminuye un 60%.

La posición en el hogar influye en la probabilidad, en el caso de jefe de familia es 0.49, en el caso de cónyuge es 0.57, y en el caso de hijo es 0.40.

Por entidad federativa se encontraron algunas diferencias: en el Distrito Federal la probabilidad es de 0.48, al igual que el estado de Guanajuato.

En algunos estados la probabilidad es menor: Aguascalientes 0.44, Baja California Norte 0.43, Baja California Sur 0.41, Campeche 0.45, Colima 0.25, Durango 0.37, Hidalgo 0.43, Estado de México 0.44, Michoacán 0.47, Morelos 0.32, Nayarit 0.42, Puebla 0.45, Quintana Roo 0.38, San Luis Potosí 0.47, Sonora 0.42, Tabasco 0.41, Tamaulipas 0.30, Veracruz 0.40, Zacatecas 0.29.

En otros estados la probabilidad es mayor que en el DF: Coahuila 0.55, Chiapas 0.60, Chihuahua 0.54, Guerrero 0.62, Jalisco 0.50, Nuevo León 0.49, Oaxaca 0.58, Querétaro 0.45, Sinaloa 0.49, Tlaxcala 0.55 y Yucatán 0.58.

Si se consideran las diferentes ramas de actividad económicas se tienen las siguientes probabilidades: Agropecuario, Silvicultura y Pesca, 0.49; Minería, 0.7; Industria Manufacturera 0.48; Construcción, 0.44; Electricidad, Gas y Agua Potable 0.41; Comercios, Restaurantes y Hoteles 0.16; Transportes, Almacenamientos y Comunicaciones 0.27; Servicios financieros, Seguros y bienes inmuebles 0.55; Servicios Comunes, Sociales y Personales 0.63.

Si se toma en cuenta el nivel de ingreso, en el nivel 1 la probabilidad es 0.42, en el nivel 2 es 0.61, en el nivel 3 es 0.53, en el nivel 4 es 0.35, en el nivel 5 es 0.45, en el nivel 6 es 0.67, en el nivel 7 es 0.44 y finalmente en el nivel 8 es 0.78.

Se observa que aunque la probabilidad tiene un comportamiento de sube y baja, hay una tendencia a subir conforme se consideran mayores niveles de ingreso. Es decir, a mayor nivel de ingreso la probabilidad de estar ocupado en su área de especialidad es más alta. Esto refuerza la idea de que la población, en el nivel técnico, no está fuera de su área de especialidad por obtener mayores ingresos. Es más probable que se trate de sub-empleo (en donde la población está fuera de su área de

especialidad con menores ingresos) que otras cuestiones como podría ser la orientación vocacional.

En promedio, la probabilidad de estar en su área de especialidad es 0.43 para este nivel, lo que puede interpretarse como que el 43% de la población se ocupa de su área de especialidad, mientras que el 57% no lo hace. Esto indica un uso ineficiente de los Recursos Humanos a este nivel.

Si se atienden las diferentes áreas de conocimiento, se tiene que para las ciencias agropecuarias la probabilidad promedio es 0.33; ciencias de la salud, 0.56; ciencias naturales y exactas 0.32; ciencias sociales, 0.49; ingeniería y tecnología, 0.43; humanidades y otras áreas de conocimiento, 0.79.

▪ **Nivel licenciatura**

En el nivel de licenciatura la edad no resultó ser una variable que muestre grandes diferencias, como en el caso del nivel técnico. Para la población considerada con edad menor de 30 años, la probabilidad de estar en su área de especialidad es 0.62, para la población con edad entre 30 y 39 años es 0.67, para los que tienen entre 40 y 49 años es 0.69, para los que tienen de entre 50 y 59 años es 0.72 y finalmente, para la población con 60 y 69 años es 0.68. Con esta información, pensar que la Población está fuera de su área de especialidad por juventud e inexperiencia puede ser equivocado. En todos los grupos de edad la probabilidad es casi la misma.

Al igual que en el nivel técnico, las mujeres resultaron estar ocupadas en su área de especialidad en mayor medida que los hombres. En promedio, el 0.63 de los hombres está ocupado en su área de especialidad, mientras que para las mujeres lo hace el 0.70.

Al contrario que en el caso del nivel técnico, la variable de interacción: número de hijos no repercute en una disminución de la probabilidad de estar ocupado en su área de especialidad, para las mujeres que no tienen hijos la probabilidad es 0.68, si tienen uno o dos hijos, la probabilidad aumenta a 0.75, mientras que si el número de hijos se eleva de tres a cinco la probabilidad aumenta a 0.77.

En lo que respecta al estado civil no se encontraron diferencias importantes en las probabilidades. Para los solteros la probabilidad de estar ocupado en su área de especialidad es 0.63, la situación de casado o en unión libre tiene una probabilidad de 0.68. Mientras que en el caso de los separados o divorciados la probabilidad es 0.66.

La posición en el hogar influye en la probabilidad, en el caso de jefe de familia es 0.66, en el caso de cónyuge es 0.76 y en el caso de hijo es 0.61.

Al considerar las entidades federativas se tienen algunas diferencias: en el Distrito Federal la probabilidad es de 0.65, al igual que el estado de Quintana Roo, Sonora y Tamaulipas.

En pocos estados la probabilidad es menor: Aguascalientes 0.60, Baja California Norte 0.61, Coahuila 0.62, Jalisco 0.63, Nayarit 0.59, Nuevo León 0.60.

En la mayoría de los estados la probabilidad es mayor que en el Distrito Federal: Colima, Chihuahua, Morelos y Zacatecas 0.66; Baja California Sur, Campeche, Chiapas, Guanajuato, Puebla Tlaxcala y Yucatán tienen una probabilidad de 0.70; Durango y San Luis Potosí 0.67; Guerrero y Veracruz 0.69; Hidalgo y Oaxaca 0.74; Estado de México, Michoacán y Sinaloa 0.68; Querétaro 0.71 y Tabasco 0.72.

Si se consideran las diferentes ramas de actividad económicas se tienen las siguientes probabilidades: Agropecuario, Silvicultura y Pesca, 0.28; Minería, 0.72; Industria Manufacturera 0.55; Construcción, 0.82; Electricidad, Gas y Agua Potable 0.60; Comercios, Restaurantes y Hoteles 0.27; Transportes, Almacenamientos y Comunicaciones 0.38; Servicios financieros, Seguros y bienes inmuebles 0.42; Servicios Comunales, Sociales y Personales 0.81.

Si se toma en cuenta el nivel de ingreso, en el nivel 1 la probabilidad es 0.62, en el nivel 2 es 0.75, en el nivel 3 es 0.69, en el nivel 4 es 0.69, en el nivel 5 es 0.58, en el nivel 6 es 0.66, en el nivel 7 es 0.75 y finalmente en el nivel 8 es 0.65.

A diferencia del nivel técnico a mayor nivel de ingreso no corresponde una probabilidad de estar ocupado en su área de especialidad más alta. De manera que la población, en el nivel licenciatura, puede estar fuera de su área de especialidad por

obtener mayores ingresos. Por lo tanto, es poco factible que se trate de sub-empleo (en donde la población esta fuera de su área de especialidad con menores ingresos).

Para este nivel, la probabilidad (en promedio) de estar en su área de especialidad es 0.66. Esta cifra puede interpretarse como que el 66% de la población considerada se ocupa en su área de especialidad, mientras que el 34% no lo hace. Al igual que en el nivel técnico se tiene un uso ineficiente de los Recursos Humanos. Sin embargo, el desaprovechamiento es en menor medida.

En las diferentes áreas de conocimiento se tiene que para las ciencias agropecuarias la probabilidad promedio es 0.33; ciencias de la salud, 0.56; ciencias naturales y exactas, 0.32; ciencias sociales, 0.49; ingeniería y tecnología, 0.43; humanidades y otras áreas de conocimiento, 0.79.

▪ **Nivel de Posgrado**

Para la población considerada en el nivel posgrado con edad menor de 30 años, la probabilidad de estar en su área de especialidad es 0.82, para la población con edad entre 30 y 39 años es 0.92, para los que tienen entre 40 y 49 años es 0.86, al igual que los que tienen entre 50 y 59 años. Finalmente, para la población con 60 y 69 años es 0.93.

En este nivel las diferencias por género son prácticamente nulas. En promedio, el 0.88 de los hombres está ocupado en su área de especialidad, mientras que para las mujeres lo hace el 0.87.

Al contrario que en el caso del nivel técnico, la variable de interacción: número de hijos no repercute en una disminución de la probabilidad de estar ocupado en su área de especialidad. Para las mujeres que no tienen hijos la probabilidad es 0.87, si tienen uno o dos hijos, la probabilidad aumenta a 0.88, mientras que si el número de hijos se eleva de tres a cinco la probabilidad es de 0.86.

Para el estado civil se encontraron diferencias mínimas en las probabilidades de soltero y casado o unión libre. Para los solteros, la probabilidad de estar ocupado en su

área de especialidad es 0.86, la situación de casado o en unión libre tiene una probabilidad de 0.89. En el caso de los separados o divorciados la probabilidad es 0.76.

En el caso de la posición en el hogar se tienen diferencias mínimas en la probabilidad: en el caso de jefe de familia es 0.87, en el caso de cónyuge es 0.89.

Al considerar las entidades federativas se tiene que en el Distrito Federal la probabilidad es de 0.87, al igual que el estado de Colima, Durango y Morelos.

En algunos estados la probabilidad es menor: Baja California Norte 0.76, Campeche 0.73, Coahuila 0.84, Chihuahua 0.78, Nuevo León 0.76, Oaxaca 0.8, Querétaro 0.86, Quintana Roo 0.59, San Luis Potosí 0.63 y Tabasco 0.86.

En los demás estados la probabilidad es mayor que en el Distrito Federal: Aguascalientes 0.97, Baja California Sur 0.92, Chiapas 0.91, Guanajuato 0.93, Guerrero 0.94, Hidalgo 0.91, Jalisco 0.92, Estado de México 0.95, Michoacán 0.91, Nayarit 0.97, Puebla 0.95, Sinaloa 0.98, Sonora 0.9, Tlaxcala 0.94, Veracruz y Yucatán 0.93. Finalmente, Zacatecas 0.89.

Si se consideran las diferentes ramas de actividad económicas se tienen las siguientes probabilidades: Agropecuario, Silvicultura y Pesca, 0.86; Minería, 0.49; Industria Manufacturera 0.62; Construcción, 0.77; Electricidad, Gas y Agua Potable 0.46; Comercios, Restaurantes y Hoteles 0.57; Transportes, Almacenamientos y Comunicaciones 0.83; Servicios financieros, Seguros y bienes inmuebles 0.67; Servicios Comunes, Sociales y Personales 0.93.

Si se toma en cuenta el nivel de ingreso, en el nivel 1 la probabilidad es 0.81, en el nivel 2 es 0.88, en el nivel 3 es 0.86, en el nivel 4 es 0.89, en el nivel 5 es 0.95, en el nivel 6 es 0.99, en el nivel 7 es 0.72 y finalmente en el nivel 8 es 0.92.

A diferencia de la población en el nivel licenciatura, no hay evidencia para pensar que la población considerada pueda estar fuera de su área de especialidad por obtener mayores ingresos.

Para este nivel, la probabilidad (en promedio) de estar en su área de especialidad es 0.88. Al igual que los casos anteriores, esta cifra puede interpretarse como que el 88% de la población considerada se ocupa en su área de especialidad,

mientras que el 12% no lo hace. Al igual que en el nivel técnico y licenciatura se tiene un uso ineficiente de los Recursos Humanos. Sin embargo, el desaprovechamiento es mucho menor, que en los casos anteriores.

En las diferentes áreas de conocimiento se tiene que para las ciencias agropecuarias la probabilidad promedio es 0.8; ciencias de la salud, 0.95; ciencias naturales y exactas, 0.77; ciencias sociales, 0.87; ingeniería y tecnología, 0.72; humanidades y otras áreas de conocimiento, 0.96.

Capítulo 4

Salarios promedio y salarios relativos entre las diferentes áreas del conocimiento

Partiendo del hecho que la educación es una inversión redituable, medir los rendimientos de esta inversión es complejo y controversial debido a las derramas sociales que no pueden ser medidas. Los beneficios a la sociedad van mucho más allá de las retribuciones monetarias. La educación no sólo influye en el desarrollo económico sino también en el desarrollo cultural y social que puede ser difícilmente medido o valorado. A pesar de todo esto, es relevante conocer los rendimientos económicos de la educación en los diferentes campos del conocimiento.

Este trabajo se restringe al análisis de los rendimientos económicos de la educación en el nivel superior, en términos de los ingresos monetarios que reciben los individuos con educación en este nivel en el mercado de trabajo. Es decir, el enfoque de este trabajo supone que el ingreso salarial es una señal de mercado que es un indicador del valor económico de la educación. Por lo tanto, las diferencias salariales dan una idea de las diferentes valoraciones económicas de la educación en el nivel superior en los distintos campos de conocimiento.

Por lo anterior, uno de los objetivos de este trabajo es medir las diferencias salariales entre los diferentes campos de conocimiento para contribuir a tener un marco de información acerca de los rendimientos económicos que obtienen los recursos humanos con educación en el nivel superior. Para tal fin se utiliza como fuente de información la Encuesta Nacional de Empleo de 1998.

A pesar que la ENE no está diseñada para medir el nivel de ingreso total de la población, se dispone de información referente al ingreso laboral. El ingreso total de la población se compone de otros rubros (como pueden ser: herencias, intereses de inversiones, aportaciones económicas de familiares, becas, etcétera) cuya información no reporta la encuesta. La encuesta que está diseñada para obtener esta información

es: La Encuesta Nacional Ingreso Gasto de los Hogares (ENIGH). En este trabajo sólo se considera el ingreso laboral que reporta la ENE.

Es importante señalar que tanto la ENIGH como la ENE son encuestas a hogares, por lo que la información no es proporcionada directamente. La entrevista es hecha en las viviendas y generalmente el cuestionario es contestado por la esposa o algún otro familiar. Es una costumbre en México que la esposa (o los demás miembros de la familia) no conozca el nivel de ingreso que obtiene su cónyuge y, por lo tanto, las respuestas a las preguntas relacionadas con los ingresos pueden resultar sesgadas. En la mayoría de los casos son conjeturas y no se dispone de información sobre de la calidad de las mismas. Debido a esto, la información proporcionada por las encuestas debe ser tomada con las debidas reservas.

Los problemas asociados con la calidad de la información de los ingresos laborales no son particulares. En todos los países se tienen problemas con la medición de estas variables. Por ejemplo, una parte de los ingresos por el trabajo de las personas puede ser no monetario. Es muy frecuente que se pague una parte del ingreso laboral en forma de: acciones de la empresa, seguros de vida, de gastos médicos, fondos de pensión o créditos con tasas preferenciales, por considerar algunos ejemplos. En la mayoría de los casos, medir los ingresos tomando en cuenta todo lo anterior, no es posible ni en México ni en el resto del mundo.

En la ENE se obtiene información acerca del ingreso laboral que perciben las personas así como el número de horas que la población dedica al trabajo. Este trabajo se elaboró con dicha información a partir de la base de datos de la encuesta de 1998 por ser la más reciente al momento de la elaboración de este documento.

Se consideró a la población mayor de 18 años y menor de 70, con estudios de al menos cuatro años completos a nivel licenciatura que declaró percibir al menos 1.5 salarios mínimos. Esta población considerada fue clasificada en dos niveles: licenciatura y posgrado. En el caso de posgrado fueron clasificados todos aquellos que contaban con al menos un año completo de estudios en algún diplomado, especialización, maestría y doctorado. La base de datos no permite diferenciar en entre

todos estos casos, sólo se puede aislar el caso de doctorado, pero la representatividad en ese caso es dudosa, por lo tanto no se desagregó y se dejaron los dos grupos mencionados: con y sin posgrado.

En la base de datos se presenta información acerca de la zona salarial en donde habita el entrevistado (A, B o C). Para todas las personas consideradas que viven en zonas salariales B o C, el salario fue ponderado de acuerdo con la relación que guarda el salario mínimo en dichas zonas con respecto al de la zona A, lo anterior como un intento de considerar las diferencias entre las zonas geográficas en cuanto a salarios.

4.1 Salarios en el nivel de educación superior por campo de la ciencia

Con la información acerca de los ingresos laborales de las personas y el número de horas que dedican de su tiempo al trabajo, fue posible calcular el salario por hora de la población con educación en el nivel superior. El salario promedio de este subconjunto de la población fue de 33.29 pesos por hora, la información que se obtiene al desagregar por los diferentes campos de conocimiento se presenta en el cuadro 14. A manera de referencia, el salario mínimo en el periodo del levantamiento de la encuesta fue de 3.77 pesos por hora.

Cuadro 14

SALARIO PROMEDIO DE LA POBLACIÓN CON EDUCACIÓN EN EL NIVEL SUPERIOR (LICENCIATURA Y POSGRADO) POR CAMPO DE LA CIENCIA

	Número de observaciones	Salario por hora promedio	Desviación estándar	Mínimo	Máximo	Salario relativo al Total
Total	22,427	33.29	32.24	5.67	1,278.97	1.00
Ciencias Agropecuarias	1,048	30.33	46.17	5.68	1,278.97	0.91
Ciencias de la Salud	2,186	35.48	37.45	5.76	881.12	1.07
Ciencias Naturales y Exactas	1,254	35.75	37.52	5.70	744.23	1.07
Ciencias Sociales	12,447	32.38	28.87	5.67	1,115.38	0.97
Ingeniería y Tecnología	5,067	34.67	33.29	5.74	578.85	1.04
Humanidades y Otros	420	32.21	22.89	5.71	169.54	0.97

En el cuadro anterior puede apreciarse que el salario promedio por hora es más alto en las ciencias exactas y naturales, aunque la diferencia con el salario promedio de las ciencias de la salud es marginal. En estos casos el salario por hora promedio es 7% más alto que el salario promedio total. En lo que respecta a ingeniería y tecnología el salario promedio es 4% más alto que en el promedio total. En el caso de las ciencias sociales, en las que se agrupa la mayoría de los recursos humanos, el salario promedio esta por debajo del promedio total en un 3% al igual que el área de humanidades. El área donde se encuentran los salarios más bajos es la correspondiente a las ciencias agropecuarias con un 9% por debajo del salario promedio total.

4.2 Salarios en el nivel de licenciatura por campo de la ciencia

El salario promedio en el nivel de licenciatura fue de 32.02 pesos por hora, la información que se obtiene al desagregar por los diferentes campos de conocimiento se presenta en el cuadro 15.

Cuadro 15

SALARIO PROMEDIO DE LA POBLACIÓN CON EDUCACIÓN EN EL NIVEL LICENCIATURA POR CAMPO DE LA CIENCIA

	Número de observaciones	Salario por hora promedio	Desviación estándar	Mínimo	Máximo	Salario relativo al Total
Total	20,619	32.02	30.75	5.67	1,278.97	1.00
Ciencias Agropecuarias	1,003	29.58	46.68	5.68	1,278.97	0.92
Ciencias de la Salud	1,600	30.96	36.93	5.76	881.12	0.97
Ciencias Naturales y Exactas	1,143	34.25	35.91	5.70	744.23	1.07
Ciencias Sociales	11,629	31.31	26.60	5.67	1,115.38	0.98
Ingeniería y Tecnología	4,867	34.17	32.72	5.74	578.85	1.07
Humanidades y Otros	373	30.52	21.63	5.71	169.54	0.95

De acuerdo con el cuadro anterior, al considerar los recursos humanos en el nivel licenciatura, el salario promedio por hora más alto lo obtienen los recursos humanos en ingeniería y tecnología y le siguen las ciencias naturales y exactas, estando ambos alrededor de un 7% más alto que el salario por hora promedio total.

4.3 Salarios en el nivel de posgrado por campo de la ciencia

El salario promedio en el nivel de posgrado fue de 47.76 pesos por hora, 49 % más alto que en el nivel de licenciatura. La información que se obtiene al desagregar por los diferentes campos de conocimiento se presenta en el cuadro 16, esta información guarda algunas diferencias con respecto al nivel licenciatura, una de ellas es que la dispersión de los salarios relativos es menor.

Cuadro 16

**SALARIO PROMEDIO DE LA POBLACIÓN CON EDUCACIÓN
EN EL NIVEL POSGRADO POR CAMPO DE LA CIENCIA**

	Número de observaciones	Salario por hora promedio	Desviación estándar	Mínimo	Máximo	Salario relativo Al Total
Total	1,808	47.76	43.45	5.70	748.74	1.00
Ciencias Agropecuarias	45	47.08	27.96	5.95	152.97	0.99
Ciencias de la Salud	586	47.82	36.08	6.39	384.62	1.00
Ciencias Naturales y Exactas	111	51.28	48.82	9.27	342.38	1.07
Ciencias Sociales	818	47.59	48.77	5.70	748.74	1.00
Ingeniería y Tecnología	200	46.97	43.28	5.82	431.65	0.98
Humanidades y Otros	47	45.63	27.96	8.03	154.16	0.96

En estos casos el salario por hora promedio total es el mismo que el salario promedio en el caso de las ciencias sociales. Con relación a las ciencias exactas y naturales, el salario promedio por hora es 7% más alto que el caso de las ciencias sociales. A diferencia de la situación de licenciatura, en ingeniería y tecnología el salario promedio es 2% más bajo que en el promedio total; en el área de humanidades el salario promedio es 4% por debajo del salario promedio total. En contraste con el nivel de licenciatura, en el nivel de posgrado el salario promedio de las ciencias agropecuarias esta sólo 1% debajo del promedio total.

En la próxima sección, en los cuadros 17 y 18 se presentan los salarios promedios y relativos entre las diferentes disciplinas desagregando cada uno de los campos de conocimiento. Es importante destacar que para cada disciplina se incluye en la primera columna el tamaño de muestra, es decir, el número de observaciones que se tuvieron en la base de datos de cada disciplina para calcular los promedios. En los casos donde este número es pequeño, los salarios promedio deben tomarse con las debidas providencias, pues la disciplina en cuestión pudiera no estar bien representada.

4.4 Salarios en el nivel de licenciatura por disciplina

En el cuadro 17 se presentan los salarios promedio de la población con educación en el nivel de licenciatura para cada una de las disciplinas o grupos de disciplinas considerados en el catálogo de carreras de licenciatura¹⁴.

Cuadro 17
SALARIO PROMEDIO Y SALARIOS RELATIVOS DE LA POBLACIÓN CON EDUCACIÓN EN EL NIVEL LICENCIATURA POR DISCIPLINA

	Número de observaciones	salario por hora promedio	desviación estándar	mínimo	máximo	salario relativo al total	salario relativo al campo de la ciencia
Ciencias Agropecuarias							
AGRONOMÍA	688	29.7	53.1	5.9	1,279.0	0.93	1.00
VETERINARIA Y ZOOTECNIA	281	29.2	28.0	5.7	345.3	0.91	0.99
INGENIERÍA PESQUERA	19	34.0	35.3	6.4	165.4	1.06	1.15
FORESTALES	15	26.1	22.7	6.9	89.2	0.82	0.88
Ciencias de la Salud							
MEDICINA GENERAL	747	36.1	49.5	5.8	881.1	1.13	1.17
ODONTOLOGÍA	390	29.8	24.7	5.8	259.0	0.93	0.96
ENFERMERÍA	437	23.5	11.7	5.8	88.3	0.73	0.76
NUTRICIÓN	24	27.0	20.6	6.1	103.0	0.84	0.87
BIOMÉDICAS	1	15.4		15.4	15.4	0.48	0.50
Ciencias Naturales y Exactas							
BIOLOGÍA, BOTÁNICA Y ZOOLOGÍA	183	31.6	54.9	5.7	744.2	0.99	0.92
ECOLOGÍA Y CIENCIAS ATMOSFÉRICAS	9	21.5	12.4	5.8	46.2	0.67	0.63
CIENCIAS DEL MAR	20	27.0	18.4	10.3	78.5	0.84	0.79

¹⁴ Catálogo de carreras de nivel técnico profesional, licenciatura y posgrado, INEGI, 1994.

	Número de observaciones	salario por hora promedio	desviación estándar	mínimo	Máximo	salario relativo al total	salario relativo al campo de la ciencia
QUÍMICA, ING. QUÍMICA Y QUÍMICA IND.	524	38.5	37.1	5.8	535.4	1.20	1.12
QUÍMICA FARMACÉUTICA BIOLÓGICA	224	28.0	18.5	5.7	156.9	0.87	0.82
BIOQUÍMICA	68	31.5	24.7	6.3	143.9	0.98	0.92
MATEMÁTICAS, ESTADÍSTICA Y ACTUARÍA	77	35.3	16.9	8.0	89.2	1.10	1.03
FÍSICA Y ASTRONOMÍA	35	36.1	30.6	7.7	162.2	1.13	1.06
Ciencias Sociales							
PSICOLOGÍA	384	30.0	21.5	5.7	172.7	0.94	0.96
SOCIOLOGÍA, TRABAJO SOCIAL Y AFINES	224	25.5	15.7	5.7	95.2	0.80	0.81
CIENCIAS POLÍTICAS Y ADMÓN. PUBLICA	172	29.9	21.6	6.4	153.8	0.94	0.96
CIENCIAS DE LA COMUNICACIÓN	295	33.5	73.8	5.8	1,115.4	1.05	1.07
DERECHO	1,729	32.1	30.6	5.7	413.5	1.00	1.02
GEOGRAFÍA	17	28.1	14.6	7.4	64.2	0.88	0.90
ECONOMÍA	256	37.0	28.1	5.8	171.3	1.16	1.18
ADMINISTRACIÓN	1,713	32.4	28.5	5.7	311.2	1.01	1.03
CONTADURÍA Y AFINES	2,981	28.4	24.7	5.7	297.4	0.89	0.91
TURISMO	159	21.8	14.0	5.8	81.1	0.68	0.70
ARCHIVONOMÍA Y BIBLIOTECONOMÍA	7	24.1	11.9	9.5	41.3	0.75	0.77
MERCADOTECNIA	155	32.7	41.1	5.8	479.6	1.02	1.04
FORMACIÓN DOCENTE EDUCACIÓN BÁSICA	1,838	32.3	13.7	5.9	115.1	1.01	1.03
FORMACIÓN DOCENTE EDUCACIÓN MEDIA Y NORMAL	929	34.6	18.4	5.8	335.7	1.08	1.10
PEDAGOGÍA Y CIENCIAS DE LA EDUCACIÓN	456	34.4	19.5	5.8	198.5	1.07	1.10
FORMACIÓN DOCENTE EN EDUCACIÓN ESPECIAL	108	33.8	13.3	5.8	92.3	1.06	1.08
EDUCACIÓN FÍSICA Y DEPORTES	183	33.4	16.8	6.4	107.1	1.04	1.07
EDUCACIÓN MUSICAL, DANZA Y CANTO	11	39.1	24.5	6.7	99.2	1.22	1.25
Ingeniería y Tecnología							
ARQUITECTURA Y URBANISMO	612	33.3	38.6	5.8	578.8	1.04	0.97
DISEÑO INDUSTRIAL Y DE INTERIORES	66	33.1	37.6	6.4	256.4	1.03	0.97
DISEÑO GRÁFICO	120	27.1	21.8	7.1	182.5	0.85	0.79
INGENIERÍA CIVIL Y DE LA CONSTRUCCIÓN	926	33.8	33.7	5.7	576.9	1.05	0.99
INGENIERÍA EXTRACTIVA, METALÚRGICA Y ENERGÉTICA	59	46.3	50.6	8.9	361.0	1.45	1.36

	Número de observaciones	salario por hora promedio	desviación estándar	mínimo	Máximo	salario relativo al total	salario relativo al campo de la ciencia
INGENIERÍA EN COMPUTACIÓN E INFORMÁTICA	820	30.5	27.3	5.8	307.8	0.95	0.89
INGENIERÍA ELÉCTRICA Y ELECTRÓNICA	700	33.2	25.0	5.8	208.2	1.04	0.97
INGENIERÍA MECÁNICA E INDUSTRIAL	1,312	36.5	31.8	5.7	396.9	1.14	1.07
INGENIERÍA DEL TRANSPORTE AERONÁUTICA, NAVAL Y PILOTOS	105	47.2	62.7	6.0	575.5	1.48	1.38
INGENIERÍA TOPOGRÁFICA, HIDRÁULICA, GEOLÓGICA Y GEODESTA	124	37.8	28.6	5.8	197.8	1.18	1.11
Humanidades y Otros							
FILOSOFÍA	48	34.8	26.5	5.7	169.5	1.09	1.14
HISTORIA	43	30.4	20.5	8.9	113.4	0.95	1.00
ANTROPOLOGÍA Y ARQUEOLOGÍA	29	36.6	24.7	6.2	121.9	1.14	1.20
LETRAS LITERATURA E IDIOMAS	153	30.6	18.3	5.8	153.5	0.96	1.00
TEOLOGÍA Y RELIGIÓN	33	16.7	13.8	6.4	76.7	0.52	0.55
ARTES PLÁSTICAS	25	26.2	18.6	7.4	80.3	0.82	0.86
MÚSICA Y DANZA	31	38.0	28.0	5.7	143.9	1.19	1.24
TEATRO Y CINEMATOGRAFÍA	7	18.5	11.9	7.5	41.6	0.58	0.61

4.5 Salarios en el nivel de posgrado por disciplina

En el cuadro 18 se presentan los salarios promedio de la población con educación en el nivel de posgrado para cada una de las disciplinas o grupos de disciplinas considerados en el catálogo de carreras de posgrado¹⁵.

¹⁵ Op. Cit.

Cuadro 18
SALARIO PROMEDIO Y SALARIOS RELATIVOS DE LA POBLACIÓN CON EDUCACIÓN EN EL NIVEL DE POSGRADO POR DISCIPLINA

	Número de observaciones	Salario por hora promedio	Desviación estándar	Mínimo	máximo	salario relativo al total	salario relativo al campo de la ciencia
Ciencias Agropecuarias							
AGRONOMÍA	28	47.95	23.18	8.89	97.34	1.00	1.02
VETERINARIA Y ZOOTECNIA	14	49.83	37.61	5.95	152.97	1.04	1.06
INGENIERÍA PESQUERA	1	29.74		29.74	29.74	0.62	0.63
Ciencias de la Salud							
MEDICINA GENERAL	508	48.74	37.02	6.69	384.62	1.02	1.02
ODONTOLOGÍA	43	49.57	32.87	6.39	165.38	1.04	1.04
ENFERMERÍA	24	32.03	21.94	16.54	126.41	0.67	0.67
NUTRICIÓN	5	29.68	6.93	18.74	37.18	0.62	0.62
BIOMÉDICAS	6	36.33	16.22	19.83	66.92	0.76	0.76
Ciencias Naturales y Exactas							
BIOLOGÍA, BOTÁNICA Y ZOOLOGÍA	24	48.21	30.80	10.33	143.41	1.01	0.94
ECOLOGÍA Y CIENCIAS ATMOSFÉRICAS	10	42.68	26.98	14.39	100.38	0.89	0.83
CIENCIAS DEL MAR	5	38.70	7.25	28.83	47.47	0.81	0.75
QUÍMICA, ING. QUÍMICA Y QUÍMICA IND.	35	60.26	60.52	9.27	342.38	1.26	1.18
QUÍMICA FARMACÉUTICA BIOLÓGICA	5	33.46	10.09	20.08	46.15	0.70	0.65
BIOQUÍMICA	7	44.84	26.69	16.47	93.69	0.94	0.87
MATEMÁTICAS, ESTADÍSTICA Y ACTUARÍA	16	46.88	42.82	10.71	160.62	0.98	0.91
FÍSICA Y ASTRONOMÍA	8	67.88	98.18	10.71	285.54	1.42	1.32
Ciencias Sociales							
PSICOLOGÍA	52	36.76	16.63	12.72	89.23	0.77	0.77
SOCIOLOGÍA, TRABAJO SOCIAL Y AFINES	20	34.85	13.71	13.85	57.36	0.73	0.73
CIENCIAS POLÍTICAS Y ADMÓN. PUBLICA	21	52.76	33.52	8.08	133.85	1.10	1.11
CIENCIAS DE LA COMUNICACIÓN	12	41.49	18.72	12.75	69.72	0.87	0.87
DERECHO	82	47.15	42.56	6.39	287.77	0.99	0.99
GEOGRAFÍA	2	37.42	0.70	36.92	37.92	0.78	0.79

	Número de observaciones	salario por hora promedio	desviación estándar	mínimo	Máximo	salario relativo al total	salario relativo al campo de la ciencia
ECONOMÍA	35	91.78	161.79	19.93	748.74	1.92	1.93
ADMINISTRACIÓN	232	54.04	47.85	6.39	346.15	1.13	1.14
CONTADURÍA Y AFINES	56	44.40	35.58	5.70	199.20	0.93	0.93
TURISMO	1	7.44		7.44	7.44	0.16	0.16
ARCHIVONOMÍA Y BIBLIOTECONOMÍA	2	30.78	13.25	21.42	40.15	0.64	0.65
MERCADOTECNIA	16	62.05	53.89	6.17	178.46	1.30	1.30
FORMACIÓN DOCENTE EDUCACIÓN BÁSICA	41	39.18	20.66	13.38	110.77	0.82	0.82
FORMACIÓN DOCENTE EDUCACIÓN MEDIA Y NORMAL	51	37.58	15.80	14.28	75.70	0.79	0.79
PEDAGOGÍA Y CIENCIAS DE LA EDUCACIÓN	173	40.54	20.26	8.92	172.09	0.85	0.85
FORMACIÓN DOCENTE EN EDUCACIÓN ESPECIAL	18	39.42	17.63	19.72	90.72	0.83	0.83
Ingeniería y Tecnología							
ARQUITECTURA Y URBANISMO	27	37.99	28.22	5.82	153.92	0.80	0.81
DISEÑO INDUSTRIAL Y DE INTERIORES	5	38.32	19.90	15.62	70.27	0.80	0.82
DISEÑO GRAFICO	1	9.34		9.34	9.34	0.20	0.20
INGENIERÍA CIVIL Y DE LA CONSTRUCCIÓN	26	44.25	27.24	13.38	133.85	0.93	0.94
INGENIERÍA EXTRACTIVA, METALÚRGICA Y ENERGÉTICA	9	107.85	134.34	12.79	431.65	2.26	2.30
INGENIERÍA EN COMPUTACIÓN E INFORMÁTICA	45	42.59	30.34	10.31	168.35	0.89	0.91
INGENIERÍA ELÉCTRICA Y ELECTRÓNICA	25	47.92	42.09	11.15	180.42	1.00	1.02
INGENIERÍA MECÁNICA E INDUSTRIAL	46	49.12	30.68	9.37	138.92	1.03	1.05
INGENIERÍA DEL TRANSPORTE AERONÁUTICA, NAVAL Y PILOTOS	5	33.47	28.22	10.41	77.61	0.70	0.71
INGENIERÍA TOPOGRÁFICA, HIDRÁULICA, GEOLÓGICA Y GEODESTA	10	33.23	13.81	13.38	50.36	0.70	0.71
Humanidades y Otros							
FILOSOFÍA	8	45.10	16.47	29.74	80.31	0.94	0.99
HISTORIA	13	54.04	31.87	19.89	140.06	1.13	1.18
ANTROPOLOGÍA Y ARQUEOLOGÍA	4	44.47	14.42	24.34	56.60	0.93	0.97
LETRAS LITERATURA E IDIOMAS	14	44.39	35.63	15.00	154.16	0.93	0.97
TEOLOGÍA Y RELIGIÓN	1	18.25		18.25	18.25	0.38	0.40
ARTES PLÁSTICAS	3	42.83	30.87	8.03	66.92	0.90	0.94
MÚSICA Y DANZA	3	35.95	15.77	23.08	53.54	0.75	0.79

Capítulo 5

Estimación de una función de ingresos

Dentro de la literatura económica la importancia de la educación como determinante de los ingresos ha sido destacada desde el siglo XVIII por economistas clásicos como Adam Smith y John Stuart Mill, pero es hasta la década de los sesenta cuando el tema adquiere una mayor atención. Destacan principalmente dos teorías que tratan de explicar cuáles son los determinantes de los ingresos laborales de las personas.

La teoría que ha sido aceptada con mayor amplitud es la teoría de capital humano¹⁶, que ha explicado las desigualdades en los ingresos en términos de diferencias educativas. Al incrementar el nivel educativo se aumentan las habilidades productivas y, por tanto, a mayor nivel educativo mayor ingreso laboral. En conclusión, las diferencias salariales entre los individuos son resultado de diversos patrones de inversión en habilidades productivas, en donde los incrementos en los ingresos laborales obtenidos por educación adicional son vistos como rendimientos a los costos de invertir tiempo y esfuerzo en preparación.

Como alternativa a la teoría de capital humano, la teoría de señalización¹⁷ ha cuestionado el papel de la educación como proveedor de habilidades productivas. De acuerdo con esta teoría, la educación, más que incrementar la productividad, es una “señal” sobre habilidades innatas. Por tanto, el sistema educativo funciona principalmente como un mecanismo para identificar las habilidades individuales, en donde no necesariamente se imparten habilidades productivas. Desde el punto de vista de esta teoría, los salarios se interpretan como rentas a la información.

Aunque hasta este momento ninguna de las dos teorías anteriores resulta ser concluyente, en este trabajo se utiliza la teoría de capital humano.

¹⁶ Jacob Mincer (1958,1962,1974), Gary Becker (1962,1964,1975), Theodore Schultz(1960,1961)

¹⁷ Keneth Arrow (1973), Michael Spence (1973,1974)

5.1 Forma funcional de la función de ingresos

Varios investigadores han examinado los datos correspondientes a los ingresos, y han propuesto varios tipos de formas funcionales para modelarlos. Típicamente la mediana de los ingresos está por debajo de su media, por lo que se han propuesto distribuciones estadísticas como la log-normal.

En la literatura econométrica de la determinación de los ingresos se han planteado, por lo general, ecuaciones de regresión de la forma:

$$\ln y_i = f(s_i, x_i, z_i) + u_i$$

en donde el $\ln y_i$ es el logaritmo natural de los salarios para el individuo i , s_i es una medida de su escolaridad o nivel de educación, x_i es una variable que mide la experiencia laboral y z_i son otros factores que pueden afectar los ingresos como podrían ser sexo, raza, región geográfica, etcétera; u_i es una perturbación aleatoria que representa características no observadas y normalmente se supone que obedece a una distribución normal con media cero y varianza constante.

Se han hecho varios esfuerzos por derivar una ecuación como la anterior en un contexto de maximización de beneficios en donde los individuos eligen carreras y formación de capital humano de acuerdo con sus funciones de utilidad individual. Sin embargo, estos esfuerzos han sido poco exitosos y hasta el momento es imposible obtener soluciones al planteamiento de este tipo de problemas de elección intertemporal. De manera que la ecuación de ingresos debe ser vista como una aproximación aceptable de este problema de elección.

De acuerdo con la teoría de capital humano, y algunos supuestos adicionales se obtiene una guía acerca de la forma funcional de la función de ingresos. Por ejemplo, si se supone que el único costo de la educación es el de oportunidad por los ingresos no percibidos por el hecho de estar estudiando y si llamamos Y_i a los ingresos después de

i años de preparación, se tiene que la tasa de rendimiento del primer año de educación sería:

$$r_1 = \frac{(Y_1 - Y_0)}{Y_0}$$

la ecuación anterior puede describirse como:

$$Y_1 = Y_0(1 + r_1)$$

procediendo de la misma forma para los años posteriores de educación obtenemos que para i años de preparación:

$$Y_i = Y_0(1 + r_1)(1 + r_2)\dots(1 + r_i)$$

si además se supone que la tasa de rendimiento de la escolaridad es la misma para cualquier año (es decir, $r_1 = r_2 = \dots = r_i$) y que $(1 + r)^i$ puede ser aproximado por e^{ri} , se tiene que los ingresos en el año i están dados por:

$$Y_i = Y_0 e^{ri}$$

esta expresión puede describirse en forma logarítmica y añadiendo un término de error como:

$$\ln Y_i = \ln Y_0 + ri + u$$

Esta ecuación es la forma más básica de una función de ingresos en la que puede verse que, si esta ecuación es ajustada por una línea de regresión, el coeficiente correspondiente a los años de educación, r , corresponde a la tasa de rendimiento de los años de escolaridad y el intercepto, $\ln Y_0$, representa el nivel de ingresos en ausencia de escolaridad.

Esta especificación básica de la función de ingresos ha sido generalizada por Mincer (1974) para tomar en cuenta los efectos de la capacitación en el trabajo o el capital humano adquirido a través de la experiencia. Una posibilidad es añadir un término a la última ecuación:

$$\ln Y_i = \ln Y_0 + \beta_1 i + \beta_2 k_i X_i + u$$

en donde β_1 es la tasa de rendimiento de la escolaridad, β_2 es la tasa de rendimiento de los años de experiencia y k_i es la proporción del tiempo que el trabajador dedica a la capacitación y X_i son los años de experiencia laboral. Sin embargo, los datos acerca de k son difícilmente mensurables, y en la mayoría de los casos esa información no está disponible.

De acuerdo con la teoría de capital humano, los ingresos no son constantes a lo largo del tiempo sino que son una función de tipo parabólico con un pico en alguna parte de la vida productiva. Esto llevó a Mincer a proponer una forma funcional lineal en la escolaridad, pero cuadrática en la experiencia:

$$\ln Y_i = \ln Y_0 + \beta_1 i + \beta_2 X_i + \beta_3 X_i^2 + u_i$$

Otro resultado de la teoría de capital humano indica que si las habilidades de los individuos están correlacionadas con los años de escolaridad, los individuos con mayor escolaridad reciben más capacitación en el trabajo, por lo que los ingresos deben ser mayores para los individuos más educados. Para incorporar este elemento se incluyen términos de interacción entre la escolaridad y la experiencia:

$$\ln Y_i = \ln Y_0 + \beta_1 i + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 i X_i + u$$

En los estudios de los ingresos se incorporan variables dicotómicas para tener en cuenta diferencias asociadas a diferentes categorías como pueden ser sexo, raza, rama de actividad, etcétera. En este trabajo se incluirán variables dicotómicas que consideren el área de conocimiento en las que el individuo tiene su preparación.

5.2 Método de estimación en “dos etapas de Heckman”

El método de estimación conocido como “dos etapas de Heckman”, considera que el proceso de selección induce un sesgo como resultado de utilizar muestras no aleatorias para estimar funciones de comportamiento. Al haber un proceso de selección (inclusive autoselección) las observaciones no se obtienen al azar, por lo que el sesgo surge debido a que los datos referentes a alguna variable sólo se observan cuando el

individuo es elegido. Un ejemplo clásico del método de “dos etapas de Heckman” es la estimación de una función de ingresos.

El problema de selección surge al estimar una función de ingresos mediante una regresión lineal por el método de mínimos cuadrados. Como los ingresos de los individuos sólo son observados si ellos están ocupados en el mercado de trabajo se tiene que el supuesto de que los datos provienen de una muestra aleatoria se incumple y esto induce un sesgo en la estimación, conocido en la literatura como sesgo de selección.

James Heckman (1979) demostró que el sesgo en que se incurre al realizar estimaciones con muestras seleccionadas no aleatoriamente por un proceso de selección, puede ser visto como un error de especificación por omisión de variables y que este sesgo puede aislarse estimando un modelo *probit* para modelar el proceso de selección y con la estimación de este modelo *probit* construir un parámetro, conocido como “lambda de Heckman”, que cuando se añade como regresor en la función de comportamiento corrige el sesgo de selección.

Para presentar el método de Heckman considere que las variables y y z tienen una distribución con coeficiente de correlación ρ ; medias μ_y y μ_z y desviaciones estándar σ_y y σ_z . La densidad conjunta de y y z cuando z esta truncada esta dada por:

$$f(y, z | z > a) = \frac{f(y, z)}{P[z > a]}$$

los momentos de esta distribución truncada están dados por:

$$E[y | z > a] = \mu_y + \rho\sigma_y\lambda(\alpha_z)$$

$$\text{var}[y | z > a] = \sigma_y^2(1 - \rho^2\delta(\alpha_z))$$

en donde:

$$\alpha_z = \frac{a - \mu_z}{\sigma_z}$$

$$\lambda(\alpha_z) = \frac{\phi(\alpha_z)}{1 - \Phi(\alpha_z)}$$

$$\delta(\alpha_z) = \lambda(\alpha_z)(\lambda(\alpha_z) - \alpha_z)$$

Si el truncamiento ocurre cuando $z < a$ se considera el siguiente reemplazo:

$$\lambda(\alpha_z) = \frac{-\phi(\alpha_z)}{\Phi(\alpha_z)}$$

Como puede verse, la media truncada es empujada en la dirección de la correlación si el truncamiento es por debajo, y en la dirección opuesta si es por arriba.

Por otro lado, considere el modelo de regresión:

$$y = \beta x + u$$

y supóngase que y es observada sólo si $y > 0$

En ese caso el modelo sería:

$$\begin{aligned} y &= \beta x + u && \text{si } \beta x + u > 0 \text{ ó } u > -\beta x \\ y &= 0 && \text{en cualquier otro caso} \end{aligned}$$

De manera que, para estimar la ecuación de regresión por el método de mínimos cuadrados ordinarios, no se pueden utilizar sólo las observaciones para las cuales se cumple que $y > 0$, pues nunca se cumpliría el supuesto de que $E(u_i) = 0$.

Como ya se mencionó, Heckman estableció un mecanismo de estimación en el que puede utilizarse el método de mínimos cuadrados ordinarios eliminando al error en que se incurre al considerar muestras con sesgos de selección.

Para capturar este sesgo se estima primeramente un modelo *probit* que relacione la probabilidad de ser seleccionado con respecto a cierto número de características relevantes, sobre esta base se calcula la “ λ de Heckman” (también llamado el inverso de la razón de Mills) el cual es el mismo estadístico definido con anterioridad en los momentos de la distribución truncada; éste estadístico recoge y cuantifica la magnitud del sesgo de selección y se incluye como un regresor a la ecuación que ahora puede ser estimada por el método de mínimos cuadrados ordinarios. Este método de estimación en dos etapas de Heckman ha sido utilizado ampliamente para modelar funciones de ingreso laboral.

Para estimar una función de ingresos, el método de dos etapas puede utilizarse como sigue:

1ª etapa: probit para el mecanismo de selección

Los ingresos de los individuos sólo son observados si ellos están ocupados en el mercado de trabajo, por lo que la primera etapa consiste en estimar un modelo probit que indique la probabilidad de un individuo de estar en la fuerza de trabajo.

El modelo puede ser expresado como:

$$P[z = 1] = \Phi(\gamma'w)$$

$$P[z = 0] = 1 - \Phi(\gamma'w)$$

en donde Φ es la distribución acumulada de $\gamma'w$; con base en esta estimación se calcula el inverso de la razón de Mills, también llamado "lamda de Heckman", para cada observación:

$$\hat{\lambda}_i = \frac{\phi(\gamma'w_i)}{\Phi(\gamma'w_i)}$$

la cual será añadida como variable de regresión en la función de ingresos.

Es importante notar que este modelo *probit* puede representarse en términos de una variable latente Z^* de la siguiente manera:

$$Z^* = \gamma'w + u$$

$$Z = 1 \text{ si } Z^* > 0$$

$$Z = 0 \text{ si } Z^* \leq 0$$

2ª etapa: estimación de la ecuación de ingresos

Consiste en la estimación de la función de ingresos la cual tiene la siguiente forma:

$$w = \beta'x + \varepsilon, \text{ observada sólo si } Z = 1$$

donde (u, ε) se distribuyen como una función normal bivariada con parámetros $[0,0,1, \sigma_\varepsilon, \rho]$. Por lo que se tiene que la ecuación de regresión para los ingresos puede representarse por:

$$\begin{aligned} E[y | Z = 1] &= E[Z^* > 0] \\ &= E[y / u > -\gamma'w] \end{aligned}$$

$$\begin{aligned}
&= \beta'x + E[\varepsilon/u > -\gamma'w] \\
&= \beta'x + \rho\sigma_\varepsilon\lambda(\alpha)
\end{aligned}$$

en donde $\lambda(\alpha)$ es la inversa de la razón de Mills definida con anterioridad y $\alpha = \gamma'w$. Como se observa, en esta ecuación se requiere estimar β y $\beta_\lambda = \rho\sigma_\varepsilon$ por mínimos cuadrados ordinarios de w sobre x y $\hat{\lambda}$.

Por último, es importante señalar que los términos de error de la ecuación de regresión para estimar la función de ingresos resultan ser heterocedásticos, las estimaciones de los parámetros son ineficientes y los errores estándar asociados a la estimación son sesgados e inconsistentes. Por lo que el método de Heckman no es útil para hacer experimentos de cómo ajustar una función de comportamiento cuando se considera que existe un sesgo de selección. El método de Heckman parte del hecho de que se tiene una línea de regresión que ajusta de manera satisfactoria a los datos observados y únicamente se requiere aislar o eliminar el sesgo en la estimación como consecuencia del mecanismo de selección.

5.3 Variables utilizadas en las estimaciones

Para estimar el mecanismo de selección se atendieron las variables: edad; edad al cuadrado; años de educación formal; años de educación formal al cuadrado; sexo; estado civil; posición en el hogar; y finalmente en el caso de las mujeres: el número de hijos.

Para estimar la ecuación de ingresos se consideraron las siguientes variables: sexo, edad, edad al cuadrado, años de educación formal o escolaridad, años de educación formal al cuadrado, años de experiencia potencial¹⁸, años de experiencia potencial al cuadrado; una variable de interacción entre escolaridad y experiencia potencial, además de variables dicotómicas para diferenciar entre el nivel de licenciatura y posgrado así como cada uno de los campos de la ciencia y las ramas de actividad económica.

El mecanismo de selección consiste básicamente en participar o no en la fuerza de trabajo, pues los salarios son observados sólo en el caso de participar. Por lo que se utilizó un modelo probit para medir la probabilidad de participar o no en la fuerza de trabajo. Se tomaron en cuenta variables que han mostrado ser importantes para determinar dicha probabilidad; variables tales como: edad; edad al cuadrado; años de educación formal; años de educación formal al cuadrado; sexo; estado civil; posición en el hogar y en el caso de las mujeres: el número de hijos. Todas estas variables resultaron significativas en la estimación del modelo *probit*, lo que se ha llamado en el contexto del método de Heckman, el mecanismo de selección.

El modelo *probit* estimado es:

$$P(pfl = 1) = P_0 + \alpha_1 E + \alpha_2 E^2 + \alpha_3 S + \alpha_4 S^2 + \alpha_5 h_0 + \alpha_6 h_1 + \alpha_7 h_2 + \alpha_8 h_3 + \alpha_9 pos + \alpha_{10} ec_1 + \alpha_{11} ec_3 + \alpha_{12} ec_4 + \alpha_{13} ph_2 + \alpha_{14} ph_3 + \alpha_{15} ph_4 + u_1$$

en donde:

$$pfl = \text{dicotómica para participación en la fuerza de trabajo.}$$

¹⁸ Como la encuesta sólo le pregunta al entrevistado si tiene o no experiencia, se construyó la variable experiencia potencial, como es usual en la literatura sobre este tema: $X = \text{edad} - \text{años de educación formal} - 6$.

E = Edad
 E^2 = edad al cuadrado
 S_i = años de educación formal
 S_i^2 = años de educación formal al cuadrado
 h_0 = dicotómica para mujeres sin hijos
 h_1 = dicotómica para mujeres con 1 ó 2 hijos
 h_2 = dicotómica para mujeres con 3 a 5 hijos
 h_3 = dicotómica para mujeres con más de 5 hijos
 pos = dicotómica 0,1 para licenciatura ó posgrado
 ec_1 = dicotómica para estado civil: soltero
 ec_3 = dicotómica para divorciado o separado
 ec_4 = dicotómica para viudo
 ph_2 = dicotómica para posición en el hogar cónyuge
 ph_3 = dicotómica para hijo
 ph_4 = dicotómica para no especificado
 u_1 = término de error

Los resultados de la estimación del modelo probit se muestran más adelante junto con la estimación de la función de ingresos. El modelo probit estimado proporciona una buena estimación de la probabilidad y la mayoría de las variables resultaron estadísticamente significativas.

En el caso de la estimación de la función de ingresos, algunas de las variables mencionadas con anterioridad no fueron significativas en el análisis. Por ejemplo, la variable sexo no resultó ser estadísticamente significativa en la ecuación de ingresos. A partir de los datos se puede inferir que no hay diferencias que sean estadísticamente significativas en el promedio de los ingresos con relación al género. Esta variable ha mostrado ser importante en los trabajos de estimación de los ingresos para la población en general. Sin embargo, es importante destacar que este hecho está relacionado con el nivel de escolaridad. A mayor nivel de instrucción hay una menor discriminación por sexo en términos de los salarios.

Uno de los determinantes de los ingresos, quizá el más importante, es la experiencia laboral. Sin embargo, no es posible disponer de información en la mayoría

de las encuestas a este respecto. Ha sido típico en la literatura del tema aproximarla con una variable *proxi* llamada experiencia potencial definida como:

$$\text{Experiencia} = \text{edad} - \text{años de educación formal} - 6$$

Es decir, ésta variable supone que una vez terminada la educación formal todo el tiempo posterior es experiencia. Es claro que éste es un supuesto demasiado fuerte. En muchos casos, lo que se tiene en el empleo es antigüedad y no experiencia. Sin embargo, esta variable ha probado su utilidad en los estudios sobre los ingresos laborales de las personas.

Las variables edad y edad al cuadrado no resultaron significativas en la estimación de los ingresos, esto se desprende de la incorporación de la variable experiencia potencial, la cual está correlacionada con la edad.

De acuerdo con todo lo que se ha presentado en este capítulo, se estimó por el método de dos etapas de Heckman una ecuación de ingresos con la siguiente forma:

$$\ln w_i = \ln w_0 + \beta_1 S_i + \beta_2 S_i^2 + \beta_3 pos + \beta_4 X_i + \beta_5 X_i^2 + \beta_6 SX_i + \sum_{j=7}^{11} \beta_j cc_j + \sum_{k=12}^{19} \beta_k gd_k + \beta_\lambda \lambda_i + u_{2i}$$

en donde:

- $\ln w$ = logaritmo natural de los ingresos, medidos en pesos por hora.
- S = años de educación formal
- S^2 = años de educación formal al cuadrado
- pos = dicotómica 0,1 para licenciatura o posgrado
- X = años de experiencia potencial
- X^2 = años de experiencia potencial al cuadrado
- SX = interacción entre años de educación formal y experiencia potencial
- cc_7 = dicotómica para ciencias de la salud
- cc_8 = dicotómica para ciencias naturales y exactas
- cc_9 = dicotómica para ciencias sociales
- cc_{10} = dicotómica para ingeniería y tecnología
- cc_{11} = dicotómica para humanidades y otros
- gd_{12} = dicotómica para gran división: agropecuario, silvicultura y pesca
- gd_{13} = dicotómica para gran división: minería
- gd_{14} = dicotómica para gran división: industria manufacturera

- gd_{15} = dicotómica para gran división: construcción
 gd_{16} = dicotómica para gran división: electricidad, gas y agua potable
 gd_{17} = dicotómica para gran división: comercios, restaurantes y hoteles
 gd_{18} = dicotómica para gran división: transportes, almacenamientos y comunicaciones
 gd_{19} = dicotómica para gran división: servicios financieros, seguros y bienes inmuebles
 λ = lamda de Heckman o inversa de la razón de Mills
 u_2 = término de error

Es importante señalar que, para no inducir un problema de colinealidad entre las variables, no se incorporaron directamente variables dicotómicas para el campo de las *ciencias agropecuarias* ni para la gran división: *servicios comunales, sociales y personales*. Éstas quedaron como caso de referencia, por lo que en el caso de que todas las variables dicotómicas sean cero, el número:

$$\ln w_0 + \beta_1 S_i + \beta_2 S_i^2 + \beta_3 pos + \beta_4 X_i + \beta_5 X_i^2 + \beta_6 SX_i$$

representa el salario correspondiente a este caso de referencia, y todos los coeficientes de las variables dicotómicas deben de interpretarse como desviaciones con respecto a este caso, que es elegido de manera arbitraria, pues su elección no afecta el proceso de estimación. Se eligieron las *ciencias agropecuarias* por ser el caso en donde el salario promedio es más bajo y la gran división: *servicios comunales, sociales y personales* por ser la que agrupa a la mayoría de las observaciones.

Lo mismo ocurrió para el modelo *probit*, por lo que el caso de ser *hombre sin posgrado; casado y jefe de familia* quedó como referencia y los coeficientes de la estimación representan el efecto en la probabilidad de las desviaciones con respecto al caso de referencia.

5.4 Resultados de la estimación de la función de ingresos

Al igual que en los modelos *logit* del capítulo 3, la estimación de la función de ingresos se realizó con el paquete estadístico y econométrico *STATA*, los resultados se muestran a continuación:

```
Heckman selection model -- two-step estimates      Number of obs   =    31496
(regression model with sample selection)          Censored obs    =    9069
                                                  Uncensored obs  =    22427

                                                  Wald chi2(22)   =    2276.81
                                                  Prob > chi2     =    0.0000
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

lw						
s	.524437	.1847438	2.839	0.005	.1623458	.8865283
s2	-.0105354	.0049802	-2.115	0.034	-.0202964	-.0007745
pos	.0456698	.0424246	1.076	0.282	-.0374809	.1288205
x	.0367401	.0140727	2.611	0.009	.0091581	.064322
x2	-.0002762	.0000558	-4.947	0.000	-.0003856	-.0001668
xs	-.0007218	.0008148	-0.886	0.376	-.0023187	.0008752
cc7	.1009935	.025998	3.885	0.000	.0500383	.1519488
cc8	.1780499	.0280285	6.352	0.000	.123115	.2329847
cc9	.1736146	.0221858	7.825	0.000	.1301313	.217098
cc10	.1994529	.0235045	8.486	0.000	.1533849	.2455209
cc11	.1183815	.0386769	3.061	0.002	.0425762	.1941868
gd12	-.1097107	.0488958	-2.244	0.025	-.2055448	-.0138766
gd13	.3886734	.0389254	9.985	0.000	.312381	.4649658
gd14	.0468547	.0149359	3.137	0.002	.0175808	.0761287
gd15	-.0627672	.0242646	-2.587	0.010	-.1103249	-.0152095
gd16	-.1556956	.0404365	-3.850	0.000	-.2349497	-.0764415
gd17	-.2000765	.0146509	-13.656	0.000	-.2287917	-.1713613
gd18	-.0766264	.026559	-2.885	0.004	-.1286811	-.0245716
gd19	.1446289	.0254331	5.687	0.000	.0947809	.1944769
lw0	-2.880772	1.724865	-1.670	0.095	-6.261445	.499902

pfl						
edad	.1359262	.0055952	24.294	0.000	.1249599	.1468925
edad2	-.0018045	.0000664	-27.187	0.000	-.0019346	-.0016745
s	3.498474	.263094	13.297	0.000	2.982819	4.014129
s2	-.0863917	.007375	-11.714	0.000	-.1008466	-.0719369
h0	-.0330973	.0254282	-1.302	0.193	-.0829356	.016741
h1	-.2147899	.0322617	-6.658	0.000	-.2780217	-.1515581
h2	-.2758867	.0404259	-6.824	0.000	-.3551201	-.1966534
h3	-.4304635	.1400696	-3.073	0.002	-.7049949	-.155932
pos	-.5667948	.0650704	-8.710	0.000	-.6943305	-.4392591
ec1	-.1109961	.0299787	-3.702	0.000	-.1697533	-.052239
ec3	.1443861	.0500601	2.884	0.004	.04627	.2425022
ec4	-.0942969	.0845722	-1.115	0.265	-.2600553	.0714615
ph2	-.5880191	.0343673	-17.110	0.000	-.6553778	-.5206604
ph3	-.4728072	.0307292	-15.386	0.000	-.5330354	-.412579
ph4	-.3786396	.0383306	-9.878	0.000	-.4537662	-.3035131
_cons	-35.88291	2.346511	-15.292	0.000	-40.48199	-31.28383

mills						
lambda	-.3077853	.0319466	-9.634	0.000	-.3703995	-.2451711

rho	-0.48093					
sigma	.63998353					
lambda	-.30778529	.0319466				

De acuerdo con la estimación, la mayoría de los coeficientes del modelo *probit* resultan ser significativos, la edad y los años de educación tienen un efecto positivo en la probabilidad de participar o no en la fuerza de trabajo.

En el caso de las mujeres sin hijos, no se encontró evidencia de un efecto negativo sobre la probabilidad de participar o no en la fuerza de trabajo. Sin embargo, para las mujeres con hijos hay evidencia para afirmar que existe un efecto negativo que se acentúa con respecto al número de hijos.

Por otra parte, el estado civil de soltero tiene un efecto negativo sobre la probabilidad con respecto al caso de casado; esto es consistente con la idea de que el casado quizá tenga un mayor grado de responsabilidad que no le permite estar fuera de la fuerza de trabajo por mucho tiempo. De manera similar ocurre con la posición en el hogar; con respecto al caso de ser jefe de familia, los demás casos tienen un efecto negativo sobre la probabilidad.

La “ λ de Heckman” resulta ser estadísticamente significativa, lo que confirma la idea de que existe un sesgo de selección, que debe ser considerado en la estimación de la función de ingresos, la cual tiene un efecto positivo en la variable años de escolaridad sobre la función de ingresos.

El hecho de contar o no con estudios de posgrado tiene un efecto positivo, aunque de menor magnitud, que parece no ser significativo; esto se explica por el hecho de que tal efecto está descontado en la variable años de escolaridad. Es claro que al contar con estudios de posgrado el número de años de educación formal aumenta.

A pesar de que la variable experiencia potencial es sólo una *proxi* de la experiencia que contempla la teoría de capital humano (experiencia con aprendizaje, no antigüedad) ésta resulta ser significativa en la ecuación de ingresos. Aunque con un efecto de menor magnitud que los años de educación formal, esto se explica por que sólo se está considerando a la población con educación superior. En estudios donde se contempla al grueso de la población esta variable resulta con un efecto mayor.

Con respecto al campo de la ciencia, todos los campos considerados tienen un efecto positivo sobre los ingresos, esto es natural debido a que el caso de referencia: *ciencias agropecuarias* fue elegido así por ser el menor. De manera que lo que es de interés es la magnitud del coeficiente que mide el efecto de haber estudiado en dicha área del conocimiento. Para las ciencias de la salud el coeficiente es 0.10; para las ciencias naturales y exactas es 0.178; para las ciencias sociales es 0.173; para ingeniería y tecnología el coeficiente es 0.199; finalmente, para el caso de las humanidades el coeficiente es 0.11.

Estas cifras muestran que en las ciencias exactas y naturales y en las ingenierías no se tiene una discriminación negativa sobre los salarios.

En el tipo de actividad económica, se encuentra un efecto positivo en los ingresos para las divisiones: minería; industria manufacturera; y servicios financieros, seguros e inmuebles; con respecto a los servicios comunales, sociales y personales. En el resto de los casos el efecto fue negativo con respecto a este caso de referencia.

Conclusiones

En este trabajo se hizo un análisis sobre el aprovechamiento de los Recursos Humanos en Ciencia y Tecnología (RHCyT). Se plantearon las consideraciones para medir este aprovechamiento, de acuerdo con el Manual de Canberra y, se hizo una inspección de las condiciones del mercado de trabajo para la Población con educación en el nivel posterior al bachillerato (Población considerada).

Las mediciones del Manual de Canberra sugieren un desaprovechamiento importante de los RH, en términos generales. Sin embargo, la medición propuesta en este trabajo muestra que se tiene un mejor aprovechamiento de los mismos. Se sugiere la probabilidad de la población considerada de estar ocupada en cuestiones para las que se educó, como una medida del aprovechamiento de los Recursos Humanos y, se estimó un modelo de regresión logística para medir dicha probabilidad. Se llegaron a las siguientes conclusiones:

De acuerdo con el Manual de Canberra¹⁹ se tiene que de un total de 5,353,103 personas del acervo de Recursos Humanos por educación (RHCyTE), sólo 2,667,659 se consideran también Recursos Humanos por Ocupación (RHCyTC). Si se considera el cociente entre estos números como una medida del aprovechamiento de los RHCyT, se tendría que sólo se aprovechan alrededor del 50% de los mismos, lo cual sugiere un gran desaprovechamiento. Sin embargo, de acuerdo con el modelo logit propuesto en este trabajo, se tiene que la probabilidad promedio de que una persona este en actividades para las que se preparó es 0.65, lo cual muestra que los RH están mejor aprovechados de lo que sugieren las mediciones del Manual de Canberra (50% vs 65%).

Al considerar cada uno de los niveles de instrucción por separado las diferencias son más amplias. Se tiene que en el nivel técnico de 731,401 personas en el acervo de

¹⁹ En el caso de las listas completas de estudios y ocupaciones considerados de CyT.

RHCyTE sólo 170,404 son RHCyTC, de manera que sólo se aprovechan al 23%. Sin embargo, de acuerdo con el modelo presentado en este trabajo, la probabilidad promedio de estar en su área de especialidad es 0.43, por lo que el 43% de la Población se ocupa en su área de especialidad. Así que los RH están mejor aprovechados de lo que sugiere el Manual de Canberra (23% vs 43%).

En el nivel de licenciatura, de 4,277,248 personas en el acervo de RHCyTE sólo 2,229,420 son RHCyTC, por lo que sólo se aprovechan al 52%. De acuerdo con el modelo logit, para este nivel, la probabilidad promedio de estar en su área de especialidad es 0.66. Entonces, el 66% de la Población considerada se ocupa en su área de especialidad. Al igual que en el nivel técnico los RH están mejor aprovechados de lo que sugiere el Manual de Canberra (52% vs 66%).

En el nivel de posgrado, de 344,454 personas en el acervo de RHCyTE solo 267,835 son RHCyTC y, por lo tanto solo se aprovechan al 78%. Según el modelo logit, para este nivel la probabilidad promedio de estar en su área de especialidad es 0.88. Al igual que en el nivel técnico y licenciatura se tiene un mejor aprovechamiento de los RH de lo que sugiere el Manual de Canberra (78% vs 88%).

A pesar de que en términos generales parece haber un buen aprovechamiento de los RH, no es el caso cuando se consideran las diferentes Áreas de Conocimiento. Se obtuvo que, en el nivel técnico, la probabilidad promedio de estar ocupado en su área de especialidad es: 0.33 para las Ciencias Agropecuarias; 0.56 para las Ciencias de la Salud; 0.32 para las Ciencias Naturales y Exactas; 0.49 para las Ciencias Sociales; 0.43 para Ingeniería y Tecnología; y 0.79 para Humanidades y Otras Áreas de Conocimiento. La probabilidad promedio en este nivel es 0.43, por lo que los RH cuyo aprovechamiento está por debajo del promedio están en Ciencias de la Salud, y las Ciencias Naturales y Exactas.

En el nivel de licenciatura se tiene que para las Ciencias Agropecuarias la probabilidad promedio de estar ocupado en su área de especialización es 0.33; Ciencias de la Salud, 0.56; Ciencias Naturales y Exactas, 0.32; Ciencias Sociales, 0.49; Ingeniería y Tecnología, 0.43; Humanidades y otras áreas de conocimiento, 0.79. En este nivel la

probabilidad promedio es 0.66, de manera que los RH en los diferentes campos de las Ciencias, al nivel de licenciatura, no son aprovechados de manera óptima.

En el nivel de posgrado, para las Ciencias Agropecuarias la probabilidad promedio es 0.8; Ciencias de la Salud, 0.95; Ciencias Naturales y Exactas, 0.77; Ciencias Sociales, 0.87; Ingeniería y Tecnología, 0.72; Humanidades y otras áreas de conocimiento, 0.96. La probabilidad promedio en este nivel es 0.88, de manera que al igual que en el caso de licenciatura, los RH en los diferentes campos de la Ciencia no están siendo aprovechados de maneaa óptima.

En conclusión, las Ciencias Exactas y Naturales y las Ingenierías deben ser mejor aprovechadas, si es que se tienen aspiraciones de crecimiento económico. Todos los esfuerzos de formación de RHCyT serán efectivos en la medida en que estos sean aprovechados de manera óptima, lo cual al terminar el Siglo XX, no parece ser el caso.

Por otro lado se hizo un análisis de los rendimientos económicos de la educación en el nivel superior, en términos de los ingresos monetarios que reciben los individuos con educación en este nivel en el mercado de trabajo. Dado que el ingreso salarial es una señal de mercado acerca del valor económico de la educación se presentaron cifras, con base en la Encuesta Nacional de Empleo 1998, acerca de las diferencias salariales en los diferentes campos de conocimiento para la educación en el nivel superior.

A partir de los datos se pudo inferir que no hay diferencias estadísticamente significativas en el promedio de los ingresos con relación al genero. Esta variable ha resultado ser significativa en las estimaciones de los ingresos con la población general. Sin embargo es importante destacar que la discriminación salarial por sexo esta relacionada con el nivel de escolaridad, y en este trabajo al considerar sólo a la población con educación en el nivel superior, se tiene una menor discriminación por sexo en los salarios.

Variables tales como los años de educación y los años de experiencia potencial resultaron ser positivos, tal y como lo predice la teoría de capital humano.

De acuerdo con las cifras obtenidas en este trabajo se llega a la conclusión de que los salarios en las áreas de ciencia y tecnología y en las diferentes ingenierías no resultan ser más bajos que en las otras áreas de conocimiento, como pueden ser las humanidades o las ciencias sociales. Para demostrar esta afirmación, se estimó una función de ingresos, de acuerdo con la teoría de capital humano, con el método de dos etapas de Heckman; los resultados de dicha estimación confirman que no existe una discriminación salarial negativa para las ingenierías y las ciencias exactas y naturales; al contrario, el efecto de estudiar en dichas áreas de conocimiento es positivo sobre los salarios y el coeficiente asociado a dichos efectos es más alto que en las demás áreas de conocimiento.

La conclusión general de este trabajo es que los recursos humanos en las ciencias exactas y naturales y en las ingenierías son los recursos menos aprovechados y sin embargo no se tiene una discriminación salarial negativa en esos casos, por lo que se puede inferir que en el mercado de trabajo de la población con educación en el nivel superior se tienen incentivos monetarios para ocuparse en cuestiones diferentes para las que se prepararon.

ANEXO 1

Fuentes de Información y Población Considerada

El Instituto Nacional de Estadística, Geografía e Informática (INEGI) en coordinación con la Secretaría del Trabajo y Previsión Social (STPS), llevan a cabo el levantamiento de diversas encuestas con el fin de conocer el comportamiento del mercado laboral. Una de ellas es: La Encuesta Nacional de Empleo (ENE).

Mediante esta encuesta se obtiene información estadística sobre las características ocupacionales de la población a nivel nacional, así como la composición de la población total.

A la fecha, se han llevado a cabo siete levantamientos, los cuales se realizan durante el segundo trimestre del año, La información captada por la encuesta se refiere a la semana anterior al levantamiento.

Los primeros cuatro levantamientos (1988, 1991, 1993, 1995) proporcionan información a nivel nacional para las áreas más urbanizadas (localidades mayores a 100,000 habitantes y/o capitales de estado), así como para las menos urbanizadas (localidades con menos de 100,000 habitantes).

En 1996 los datos son representativos para cada una de las entidades federativas, y a partir de ese año se alternan los levantamientos, en los años nones tendrá representatividad nacional, y en los pares estatal. De manera que el levantamiento de 1997 tiene representatividad nacional mientras que el de 1998 nuevamente tiene representatividad estatal.

El objetivo de este trabajo es estudiar en que medida los RHCyT están siendo aprovechados de manera óptima a nivel de entidad federativa. La fuente de información considerada, la ENE-98, tiene tal representatividad.

La base de datos de la ENE-98 para la Población con edad mayor a 12 años considera 371,190 registros.

Un subconjunto de estos registros es utilizado como insumo en este trabajo: los que corresponden a entrevistados con estudios posteriores al bachillerato, que forman parte de la Población Ocupada.

El universo considerado originalmente para el análisis contenía a las carreras de técnico profesional a nivel universitario. Es decir, egresados de Universidades Tecnológicas, y no consideraba el técnico a nivel medio superior. Sin embargo, con la Información considerada en este trabajo no se puede diferenciar entre estos dos casos. De manera que se considera a toda la Población con nivel de educación a nivel técnico que haya terminado su carrera, ya sea en el nivel superior o medio superior.

El universo sujeto de análisis quedo comprendido de la siguiente forma: Todos aquellos con bachillerato terminado y carrera terminada en el nivel técnico, todos aquellos con al menos 4 años de estudios terminados en el nivel de licenciatura y aquellos que tengan al menos 1 año completo en el nivel de posgrado (diplomado, especialidad, maestría o doctorado).

Cabe señalar que en este trabajo se considera solo a la Población que participa en la fuerza de trabajo y que esta ocupada. De manera que no se considera la Población que esta desempleada ni a la PEI (como es el caso de los estudiantes, las amas de casa, los incapacitados, los pensionados y jubilados, etcétera). Lo anterior debido a que el objetivo del trabajo es analizar las ocupaciones de la Población con estudios en el nivel superior.

Anexo 2:

Resultados de las estimaciones de los modelos logit

Para el nivel técnico se calcularon las probabilidades con el siguiente modelo:

```
. logistic oae geda2-geda6 sex2 hijos0 hijos1-hijos3 eciv1 eciv3 eciv4 pos2-pos
> 9 edo1-edo8 edo10-edo32 escl-esc19 esc21-esc24 ram1-ram56 ram58-ram99 sal2-sa
> 19 [fw=fac]
```

```
Logit estimates                               Number of obs   =    534679
                                                LR chi2(136)    =  256091.50
                                                Prob > chi2     =    0.0000
Log likelihood = -240559.28                    Pseudo R2      =    0.3474
```

oae	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
geda2	1.342012	.0135564	29.121	0.000	1.315703	1.368847
geda3	2.353187	.0384414	52.386	0.000	2.279036	2.42975
geda4	1.908419	.0459274	26.855	0.000	1.820493	2.000592
geda5	.9962171	.0752243	-0.050	0.960	.8591714	1.155123
geda6	.4189866	.0683431	-5.333	0.000	.304338	.576825
sex2	1.269665	.0156499	19.370	0.000	1.23936	1.300712
hijos1	.9290556	.0165837	-4.122	0.000	.8971141	.9621344
hijos2	.5358149	.0147118	-22.725	0.000	.5077424	.5654395
eciv1	1.307068	.0169764	20.618	0.000	1.274215	1.340768
eciv3	.4174546	.0121595	-29.991	0.000	.39429	.4419802
eciv4	1.658738	.0962883	8.718	0.000	1.480356	1.858614
pos2	1.129102	.0222816	6.153	0.000	1.086265	1.173629
pos3	.7849974	.0109027	-17.429	0.000	.7639167	.8066599
pos9	.6563478	.0113136	-24.428	0.000	.634544	.6789008
edo1	.5703945	.0153451	-20.869	0.000	.5410977	.6012775
edo2	1.007864	.0239871	0.329	0.742	.9619298	1.055992
edo3	1.067821	.0324115	2.162	0.031	1.006149	1.133274
edo4	.9091115	.0347683	-2.492	0.013	.8434582	.9798751
edo5	1.481767	.0311835	18.686	0.000	1.421891	1.544163
edo6	.414608	.0130495	-27.973	0.000	.3898044	.4409899
edo7	2.947799	.0733061	43.472	0.000	2.807567	3.095036
edo8	1.330301	.0304788	12.457	0.000	1.271885	1.3914
edo10	.6950443	.0218483	-11.573	0.000	.6535148	.7392128
edo11	1.038604	.0205112	1.918	0.055	.999171	1.079593
edo12	5.294252	.1229413	71.770	0.000	5.058693	5.54078
edo13	.7926516	.0179192	-10.279	0.000	.7582973	.8285624
edo14	1.031025	.0214058	1.472	0.141	.9899123	1.073845
edo15	.7957454	.0118549	-15.336	0.000	.7728461	.8193231
edo16	2.210446	.0830874	21.102	0.000	2.053452	2.379443
edo17	.4784359	.0124386	-28.357	0.000	.4546674	.503447
edo18	.6990632	.0258489	-9.682	0.000	.6501924	.7516071
edo19	.9239782	.0228537	-3.197	0.001	.8802541	.9698742
edo20	1.470939	.0342929	16.553	0.000	1.405239	1.539711
edo21	1.252656	.0566488	4.981	0.000	1.146405	1.368755
edo22	1.061152	.0390391	1.613	0.107	.98733	1.140493
edo23	1.802923	.0577374	18.405	0.000	1.693238	1.919713
edo24	1.759375	.0575291	17.278	0.000	1.650157	1.875821
edo25	.5760889	.0231261	-13.738	0.000	.5324998	.6232461
edo26	.3058922	.0127587	-28.399	0.000	.2818804	.3319494
edo27	1.167668	.0331835	5.454	0.000	1.104408	1.234552
edo28	.6438543	.0122104	-23.216	0.000	.6203617	.6682365
edo29	.6574646	.0295385	-9.334	0.000	.602046	.7179845
edo30	.6443818	.0115335	-24.553	0.000	.6221684	.6673883

edo31		2.251537	.0937497	19.492	0.000	2.075088	2.442988
edo32		.3389563	.0209191	-17.530	0.000	.3003383	.3825398
esc1		.0232359	.0009362	-93.368	0.000	.0214715	.0251453
esc2		.1858665	.0032913	-95.027	0.000	.1795263	.1924305
esc3		.9787116	.0273661	-0.770	0.442	.9265182	1.033845
esc4		.3145141	.0215746	-16.863	0.000	.274948	.359774
esc5		.1556335	.0046339	-62.478	0.000	.1468111	.1649862
esc6		.0033342	.0001979	-96.074	0.000	.0029679	.0037456
esc7		.1557771	.0033843	-85.584	0.000	.1492832	.1625535
esc8		.2336883	.0031003	-109.580	0.000	.2276902	.2398444
esc9		.5686639	.0092853	-34.570	0.000	.5507531	.5871572
esc10		1.474171	.0261118	21.910	0.000	1.423871	1.526248
esc12		.0252017	.0020464	-45.330	0.000	.0214937	.0295494
esc13		.3549084	.0087287	-42.119	0.000	.3382062	.3724355
esc14		.6189559	.04182	-7.100	0.000	.5421856	.7065965
esc15		1.48427	.0982094	5.969	0.000	1.303742	1.689795
esc16		1.206963	.0322975	7.030	0.000	1.145292	1.271954
esc17		.1362657	.0045441	-59.769	0.000	.1276443	.1454695
esc18		.2057152	.012904	-25.208	0.000	.1819168	.232627
esc19		.2043368	.0026212	-123.790	0.000	.1992633	.2095395
esc21		.4165011	.0311327	-11.718	0.000	.3597413	.4822164
esc22		2.598388	.2551899	9.723	0.000	2.143417	3.149932
esc23		3.552575	.4448203	10.124	0.000	2.779481	4.5407
esc24		12.38255	1.542769	20.196	0.000	9.699672	15.8075
ram1		3.045592	.1047399	32.384	0.000	2.847071	3.257955
ram2		.6434334	.0332697	-8.528	0.000	.5814213	.7120594
ram4		.0485952	.0050699	-28.987	0.000	.0396084	.0596209
ram6		5.13771	.3603216	23.336	0.000	4.477881	5.894767
ram7		.429705	.0964577	-3.763	0.000	.2767563	.6671804
ram10		.418879	.1006428	-3.622	0.000	.2615609	.6708173
ram11		1.276849	.0580518	5.375	0.000	1.167991	1.395851
ram12		1.728779	.1280547	7.390	0.000	1.495164	1.998895
ram13		.1587957	.0093145	-31.371	0.000	.1415499	.1781426
ram14		.1550563	.0170732	-16.928	0.000	.124958	.1924044
ram18		.3561484	.0448556	-8.197	0.000	.2782435	.4558656
ram19		.8545389	.0466177	-2.881	0.004	.767885	.9509714
ram22		.1509043	.0065943	-43.276	0.000	.1385177	.1643985
ram23		1.042577	.1041358	0.417	0.676	.8572106	1.268028
ram24		.3825646	.0311047	-11.818	0.000	.32621	.4486548
ram26		.4917704	.0441578	-7.904	0.000	.4124108	.5864011
ram27		.2102098	.0073637	-44.523	0.000	.1962615	.2251493
ram28		9.808907	.9121002	24.555	0.000	8.174666	11.76986
ram29		.2166091	.0306737	-10.802	0.000	.1641117	.2859
ram30		.4982015	.0185894	-18.673	0.000	.4630673	.5360014
ram31		2.351524	.164456	12.226	0.000	2.050311	2.696987
ram32		1.701328	.06578	13.744	0.000	1.577165	1.835265
ram33		.1454812	.0181181	-15.479	0.000	.1139723	.1857011
ram34		4.020032	.7443181	7.514	0.000	2.796582	5.778716
ram35		1.2118	.2105394	1.106	0.269	.8620696	1.703412
ram37		1.31262	.0722514	4.942	0.000	1.178381	1.462151
ram38		2.130363	.0994577	16.200	0.000	1.944082	2.334494
ram39		82.09851	8.613636	42.013	0.000	66.8388	100.8421
ram40		1.732533	.2570806	3.704	0.000	1.295318	2.317322
ram41		8.990473	.7109027	27.774	0.000	7.699731	10.49759
ram42		1.899961	.0989699	12.322	0.000	1.715557	2.104186
ram43		1.969906	.1501354	8.896	0.000	1.696569	2.28728
ram45		.3833294	.022824	-16.104	0.000	.3411069	.4307784
ram46		1.142244	.0779849	1.948	0.051	.9991812	1.30579
ram47		.6010412	.0972693	-3.146	0.002	.4376735	.825388
ram48		12.61955	1.143252	27.985	0.000	10.56648	15.07153
ram49		.6015013	.0321456	-9.512	0.000	.5416846	.6679234
ram50		1.762867	.0945145	10.574	0.000	1.587022	1.958195
ram51		10.3614	.9606424	25.218	0.000	8.639734	12.42614

ram52		.6454728	.0569602	-4.961	0.000	.5429541	.7673489
ram53		1.246641	.1034028	2.658	0.008	1.059591	1.46671
ram54		2.968259	.1109396	29.109	0.000	2.758595	3.193859
ram55		1.999024	.1240831	11.159	0.000	1.770037	2.257635
ram56		2.331	.0963209	20.481	0.000	2.149657	2.527641
ram58		.2230872	.027029	-12.382	0.000	.1759316	.282882
ram59		6.417474	.4612969	25.862	0.000	5.574149	7.388387
ram60		1.363676	.0438237	9.652	0.000	1.280432	1.452331
ram61		.8800636	.0391097	-2.875	0.004	.8066533	.9601546
ram62		.1668663	.0041431	-72.116	0.000	.1589403	.1751875
ram63		.162275	.0047548	-62.061	0.000	.1532182	.1718671
ram64		.1943733	.0057452	-55.417	0.000	.1834329	.2059662
ram65		2.043382	.0802557	18.195	0.000	1.891986	2.206893
ram66		1.515969	.0446244	14.134	0.000	1.430982	1.606003
ram67		.2181929	.023978	-13.853	0.000	.1759134	.2706339
ram68		1.666551	.0459689	18.517	0.000	1.578846	1.759128
ram69		4.924036	.135592	57.891	0.000	4.665325	5.197094
ram70		2.004969	.0581705	23.976	0.000	1.894138	2.122285
ram71		.354787	.0126101	-29.155	0.000	.3309128	.3803837
ram72		1.543617	.0384942	17.409	0.000	1.469984	1.620938
ram73		.719624	.0185039	-12.796	0.000	.6842557	.7568204
ram88		.433808	.0268972	-13.470	0.000	.3841678	.4898624
sal2		1.524773	.0176748	36.392	0.000	1.490521	1.559811
sal3		1.15635	.0308892	5.438	0.000	1.097366	1.218505
sal4		.3854722	.0224032	-16.402	0.000	.3439714	.4319802
sal5		.9589974	.0612219	-0.656	0.512	.846208	1.08682
sal6		3.692382	.2994794	16.105	0.000	3.149691	4.328579
sal7		1.857715	.3939701	2.920	0.003	1.225922	2.815111
sal8		11.18091	2.251389	11.990	0.000	7.534932	16.5911
sal9		1.242386	.0146794	18.369	0.000	1.213945	1.271493

A pesar de que la pseudo-R2 es satisfactoria se considera el número de éxitos que predice el modelo como una medida complementaria de bondad de ajuste.

Logistic model for oae

Classified	True		Total
	D	~D	
+	180157	49106	229263
-	64038	241378	305416
Total	244195	290484	534679

Classified + if predicted Pr(D) >= .5

True D defined as oae ~= 0

Sensitivity	Pr(+ D)	73.78%
Specificity	Pr(- ~D)	83.10%
Positive predictive value	Pr(D +)	78.58%
Negative predictive value	Pr(~D -)	79.03%
False + rate for true ~D	Pr(+ ~D)	16.90%
False - rate for true D	Pr(- D)	26.22%
False + rate for classified +	Pr(~D +)	21.42%
False - rate for classified -	Pr(D -)	20.97%
Correctly classified		78.84%

En el nivel licenciatura se estimó el siguiente modelo para calcular las probabilidades

```
. logistic oae geda1 geda3-geda6 sex2 hijos0-hijos9 eciv1 eciv3 eciv4 pos2-pos9
> edo1-edo8 edo10-edo32 esc26-esc77 ram1-ram59 ram61-ram99 sal2-sal9 [fw=fac]
```

```
Logit estimates                                     Number of obs   =   3375019
                                                    LR chi2(177)    =  1319703.25
                                                    Prob > chi2     =    0.0000
Log likelihood = -1495549.3                       Pseudo R2      =    0.3061
```

oae	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
geda1	1.001265	.0040976	0.309	0.757	.9932655	1.009328
geda3	.9289027	.0037771	-18.138	0.000	.9215292	.9363352
geda4	1.172927	.0070291	26.616	0.000	1.159231	1.186785
geda5	.8007556	.0102036	-17.438	0.000	.7810045	.8210062
geda6	.3838678	.0079691	-46.120	0.000	.3685622	.3998089
sex2	.6969765	.0064075	-39.268	0.000	.6845305	.7096487
hijos0	1.616564	.0148769	52.191	0.000	1.587668	1.645987
hijos1	1.08967	.0089575	10.447	0.000	1.072254	1.107369
hijos9	.4144491	.0380114	-9.604	0.000	.3462604	.4960662
eciv1	.8963527	.0050087	-19.582	0.000	.8865895	.9062235
eciv3	.8205762	.0070325	-23.074	0.000	.8069078	.8344762
eciv4	.7822089	.0142659	-13.468	0.000	.7547422	.8106752
pos2	1.216298	.0083533	28.511	0.000	1.200035	1.23278
pos3	.9729896	.0054615	-4.878	0.000	.962344	.983753
pos9	.8797026	.0061253	-18.408	0.000	.8677788	.8917903
edo1	.6756743	.0097022	-27.302	0.000	.6569233	.6949604
edo2	.700694	.006827	-36.506	0.000	.6874403	.7142032
edo3	1.060957	.0229985	2.730	0.006	1.016825	1.107005
edo4	.8016041	.0146384	-12.110	0.000	.7734208	.8308144
edo5	.7191905	.0063959	-37.065	0.000	.7067634	.7318361
edo6	1.057982	.0193511	3.082	0.002	1.020726	1.096597
edo7	1.090963	.012511	7.592	0.000	1.066716	1.115762
edo8	.7985575	.0073893	-24.310	0.000	.7842052	.8131725
edo10	.8719546	.0121637	-9.822	0.000	.8484372	.8961239
edo11	1.063352	.010138	6.443	0.000	1.043667	1.083409
edo12	.976106	.0101727	-2.321	0.020	.9563701	.9962491
edo13	.93477	.0125708	-5.016	0.000	.9104536	.9597359
edo14	.8548271	.0056877	-23.575	0.000	.8437519	.8660478
edo15	1.100161	.0059045	17.786	0.000	1.088649	1.111795
edo16	.7707985	.0076622	-26.188	0.000	.7559262	.7859635
edo17	.9510765	.0105787	-4.510	0.000	.9305671	.9720379
edo18	1.082303	.0176265	4.856	0.000	1.048301	1.117408
edo19	.809057	.0055405	-30.941	0.000	.7982703	.8199894
edo20	1.15731	.0129236	13.083	0.000	1.132255	1.182919
edo21	.9679302	.0076077	-4.147	0.000	.9531336	.9829564
edo22	1.130291	.0136998	10.105	0.000	1.103756	1.157464
edo23	.9593979	.0162216	-2.451	0.014	.9281251	.9917243
edo24	.9628545	.0105212	-3.464	0.001	.9424526	.983698
edo25	1.080313	.0098912	8.437	0.000	1.0611	1.099875
edo26	.8284208	.0080714	-19.320	0.000	.8127513	.8443924
edo27	1.102889	.0151321	7.138	0.000	1.073626	1.13295
edo28	.8184246	.00703	-23.327	0.000	.8047613	.8323198
edo29	.8961236	.0138511	-7.096	0.000	.869383	.9236867
edo30	.8522684	.006593	-20.664	0.000	.8394439	.8652888
edo31	1.270027	.0159098	19.082	0.000	1.239224	1.301596
edo32	.6717067	.0100228	-26.669	0.000	.6523468	.6916411
esc26	1.224451	.0248221	9.989	0.000	1.176755	1.274081
esc27	1.810764	.0290425	37.020	0.000	1.754727	1.86859

esc28		.3973971	.0064882	-56.522	0.000	.3848817	.4103194
esc29		.3053181	.021806	-16.611	0.000	.2654356	.351193
esc30		.0895421	.0064759	-33.365	0.000	.0777081	.1031783
esc31		.3451867	.0037406	-98.156	0.000	.3379326	.3525966
esc32		.5671502	.0073986	-43.474	0.000	.552833	.5818382
esc33		.2480506	.0191273	-18.080	0.000	.2132573	.2885206
esc34		.1302162	.0114508	-23.182	0.000	.1096007	.1547096
esc35		1.862491	.0216747	53.441	0.000	1.82049	1.905461
esc36		1.265646	.0167355	17.816	0.000	1.233267	1.298876
esc37		.7474998	.0103069	-21.106	0.000	.7275692	.7679765
esc38		1.36574	.0637919	6.673	0.000	1.246262	1.496672
esc39		5.636532	.7529596	12.945	0.000	4.338141	7.323526
esc40		.8345915	.0268609	-5.618	0.000	.7835712	.8889339
esc41		.3300691	.0121833	-30.030	0.000	.3070336	.3548329
esc42		.3315117	.0139861	-26.170	0.000	.3052021	.3600892
esc43		.3999168	.0083284	-44.009	0.000	.3839221	.4165779
esc45		9.622662	.4289308	50.793	0.000	8.81765	10.50117
esc46		1.088243	.0121097	7.599	0.000	1.064766	1.112239
esc47		.8856324	.0137981	-7.796	0.000	.8589975	.9130933
esc48		.6058685	.0188417	-16.113	0.000	.5700423	.6439463
esc49		.8979107	.0146451	-6.602	0.000	.8696607	.9270783
esc50		.4091543	.0066801	-54.737	0.000	.3962688	.4224587
esc51		.6707737	.0075863	-35.308	0.000	.6560684	.6858086
esc52		1.428497	.0110558	46.079	0.000	1.406992	1.450331
esc53		.3181157	.0129717	-28.088	0.000	.293681	.3445833
esc54		1.564114	.0610321	11.464	0.000	1.448953	1.688428
esc55		1.772716	.072952	13.912	0.000	1.635348	1.921624
esc56		7.804049	.4721726	33.959	0.000	6.931373	8.786597
esc57		1.214489	.0143936	16.396	0.000	1.186603	1.243303
esc58		1.629732	.0125508	63.421	0.000	1.605318	1.654518
esc59		3.546532	.0256962	174.726	0.000	3.496524	3.597255
esc60		.2431165	.0049672	-69.217	0.000	.2335732	.2530497
esc61		1.331181	.06084	6.259	0.000	1.217121	1.455929
esc62		.2178904	.0037871	-87.669	0.000	.2105928	.2254409
esc63		1.854206	.0199632	57.350	0.000	1.815489	1.893749
esc64		.9826311	.012848	-1.340	0.180	.9577693	1.008138
esc65		.9626425	.0138804	-2.640	0.008	.9358183	.9902357
esc66		4.099199	.1620228	35.693	0.000	3.793629	4.429382
esc68		1.176583	.1053368	1.816	0.069	.9872261	1.402261
esc69		.839166	.0076475	-19.241	0.000	.8243103	.8542893
esc70		.3833537	.0119775	-30.687	0.000	.3605827	.4075629
esc71		1.023993	.0093111	2.607	0.009	1.005905	1.042406
esc72		.6598432	.0062955	-43.576	0.000	.6476188	.6722983
esc73		.9649887	.0079714	-4.314	0.000	.9494908	.9807396
esc74		.6113867	.0116137	-25.902	0.000	.5890427	.6345783
esc75		.3926863	.0081134	-45.241	0.000	.3771019	.4089147
esc76		1.278128	.032201	9.740	0.000	1.216548	1.342825
esc77		.3916383	.0144583	-25.392	0.000	.3643015	.4210264
ram1		.0654301	.0010938	-163.119	0.000	.0633211	.0676093
ram2		.1716855	.0034064	-88.811	0.000	.1651372	.1784934
ram3		1.031119	.104019	0.304	0.761	.8461357	1.256545
ram4		.3002565	.0105883	-34.117	0.000	.2802047	.3217432
ram5		10.84691	1.526386	16.940	0.000	8.232356	14.29183
ram6		.34941	.0072404	-50.744	0.000	.3355034	.363893
ram7		.9977628	.0430949	-0.052	0.959	.9167748	1.085905
ram8		1.078547	.0708502	1.151	0.250	.9482507	1.226746
ram9		.1702279	.0116089	-25.964	0.000	.14893	.1945716
ram10		2.526509	.2384663	9.820	0.000	2.099808	3.03992
ram11		.1535354	.0033648	-85.502	0.000	.1470801	.160274
ram12		.4535454	.0148003	-24.229	0.000	.4254454	.4835013
ram13		.1408365	.0031879	-86.598	0.000	.134725	.1472252
ram14		.031357	.0011851	-91.613	0.000	.0291183	.0337679
ram15		.2760516	.018488	-19.219	0.000	.2420934	.3147731

ram16		.35115	.014169	-25.936	0.000	.3244489	.3800485
ram17		.130507	.006039	-44.007	0.000	.1191917	.1428966
ram18		.2783173	.0200152	-17.785	0.000	.2417275	.3204456
ram19		.2095303	.0043432	-75.398	0.000	.2011883	.2182182
ram21		.5139019	.0206149	-16.596	0.000	.475045	.5559371
ram22		.1837382	.0032435	-95.976	0.000	.1774898	.1902066
ram23		.1227391	.0082962	-31.035	0.000	.1075099	.1401255
ram24		.4490997	.014101	-25.495	0.000	.4222955	.4776053
ram25		.0893275	.0111494	-19.352	0.000	.0699427	.1140847
ram26		.1338451	.0035532	-75.756	0.000	.1270591	.1409936
ram27		.199985	.0034746	-92.637	0.000	.1932895	.2069124
ram28		.3597256	.0077823	-47.260	0.000	.3447914	.3753066
ram29		.0781947	.0037729	-52.820	0.000	.071139	.0859504
ram30		.0986689	.0018591	-122.916	0.000	.0950915	.1023808
ram31		.1953263	.0049068	-65.009	0.000	.1859421	.2051841
ram32		.2023713	.0027198	-118.875	0.000	.1971101	.2077728
ram33		.2368104	.004867	-70.090	0.000	.2274609	.2465442
ram34		.3366846	.0118453	-30.942	0.000	.3142507	.3607202
ram35		.1148449	.0033695	-73.763	0.000	.1084271	.1216427
ram36		.6759486	.0363659	-7.280	0.000	.6083019	.7511181
ram37		.2692912	.0087315	-40.463	0.000	.2527103	.28696
ram38		.1598832	.0034285	-85.494	0.000	.1533027	.1667462
ram39		.0347756	.0013717	-85.153	0.000	.0321884	.0375707
ram40		.4585095	.0137066	-26.085	0.000	.4324169	.4861766
ram41		.3421084	.0132048	-27.789	0.000	.3171822	.3689936
ram42		.1979668	.0039959	-80.241	0.000	.1902879	.2059557
ram43		.2244663	.0092135	-36.399	0.000	.2071155	.2432707
ram44		.2603255	.0082345	-42.547	0.000	.2446763	.2769755
ram45		.3590913	.0079288	-46.385	0.000	.3438826	.3749726
ram46		.4185486	.0089206	-40.865	0.000	.4014246	.4364031
ram47		.2112704	.0100244	-32.764	0.000	.1925088	.2318605
ram48		.2414524	.0109275	-31.400	0.000	.2209574	.2638485
ram49		.2704783	.0063098	-56.050	0.000	.2583898	.2831324
ram50		.1962332	.0043109	-74.127	0.000	.1879633	.204867
ram51		.2806905	.0049804	-71.604	0.000	.2710969	.2906236
ram52		.5086464	.0166627	-20.636	0.000	.4770144	.5423759
ram53		.2952939	.0092871	-38.784	0.000	.2776412	.3140689
ram54		.3280328	.0054621	-66.941	0.000	.3175001	.3389149
ram55		.1214285	.0028736	-89.095	0.000	.1159249	.1271932
ram56		.2142121	.0042459	-77.735	0.000	.2060498	.2226978
ram57		.8261281	.0131753	-11.977	0.000	.8007044	.8523591
ram58		.24093	.0147714	-23.214	0.000	.2136504	.2716928
ram59		.0311455	.0008359	-129.257	0.000	.0295495	.0328277
ram61		.3438956	.0050564	-72.598	0.000	.3341267	.35395
ram62		.0544695	.0004766	-332.588	0.000	.0535433	.0554117
ram63		.0887489	.0010656	-201.717	0.000	.0866848	.0908621
ram64		.1051012	.0011649	-203.259	0.000	.1028427	.1074094
ram65		.1320635	.0020613	-129.704	0.000	.1280845	.136166
ram66		.1121583	.0011625	-211.083	0.000	.1099028	.1144601
ram67		.0313067	.0007386	-146.829	0.000	.0298921	.0327883
ram68		.7306092	.0064995	-35.283	0.000	.7179808	.7434598
ram69		4.941591	.0497536	158.684	0.000	4.845031	5.040075
ram70		1.476205	.0165514	34.737	0.000	1.444118	1.509004
ram71		.2458067	.0033597	-102.663	0.000	.2393092	.2524807
ram72		.1313853	.0013711	-194.489	0.000	.1287253	.1341002
ram73		.3285438	.0028397	-128.782	0.000	.3230251	.3341569
ram88		.1494605	.0040696	-69.806	0.000	.1416933	.1576534
ram99		2.641689	.3332321	7.701	0.000	2.063042	3.382637
sal2		2.426256	.0092877	231.544	0.000	2.40812	2.444528
sal3		2.45663	.0136472	161.790	0.000	2.430027	2.483524
sal4		3.395256	.0295143	140.619	0.000	3.337899	3.453598
sal5		2.332191	.0250747	78.761	0.000	2.28356	2.381858
sal6		3.159431	.0394207	92.200	0.000	3.083105	3.237647

sal7		6.245205	.1572721	72.741	0.000	5.94444	6.561186
sal8		2.567212	.0467331	51.792	0.000	2.477231	2.660461
sal9		1.726143	.0077157	122.125	0.000	1.711086	1.741332

Al igual que el caso anterior se considera el número de éxitos que predice el modelo como un medida complementaria de bondad de ajuste.

Logistic model for oae

Classified	True		Total
	D	~D	
+	1952070	418098	2370168
-	287518	717333	1004851
Total	2239588	1135431	3375019

Classified + if predicted Pr(D) >= .5

True D defined as oae ~= 0

Sensitivity	Pr(+ D)	87.16%
Specificity	Pr(- ~D)	63.18%
Positive predictive value	Pr(D +)	82.36%
Negative predictive value	Pr(~D -)	71.39%
False + rate for true ~D	Pr(+ ~D)	36.82%
False - rate for true D	Pr(- D)	12.84%
False + rate for classified +	Pr(~D +)	17.64%
False - rate for classified -	Pr(D -)	28.61%
Correctly classified		79.09%

Finalmente para el nivel posgrado se calcularon las probabilidades con el siguiente modelo:

```
. logistic oae geda1 geda2 geda4-geda6 sex2 hijos0 hijos1 eciv1 eciv3 eciv4 pos
> 2-pos9 edo1-edo8 edo10-edo32 esc78-esc86 esc88-esc126 ram1-ram69 ram71-ram99
> sal1 sal3-sal9 [fw=fac]
```

```
Logit estimates                               Number of obs   =    280224
                                                LR chi2(121)   =   83747.58
                                                Prob > chi2    =    0.0000
Log likelihood = -60991.465                    Pseudo R2      =    0.4071
```

oae	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
geda1	2.906742	.097958	31.662	0.000	2.720951	3.105219
geda2	3.734858	.0849861	57.909	0.000	3.571948	3.905198
geda4	1.137447	.0300077	4.882	0.000	1.080128	1.197809
geda5	3.188036	.2201012	16.793	0.000	2.784559	3.649976
geda6	7.578882	.9770761	15.710	0.000	5.88664	9.757595
sex2	.1522034	.0069588	-41.175	0.000	.1391576	.1664722
hijos0	3.364481	.1516743	26.913	0.000	3.079959	3.675286
hijos1	2.375121	.0877182	23.423	0.000	2.209271	2.553421
eciv1	.7424353	.0274761	-8.047	0.000	.6904899	.7982885
eciv3	.5078711	.0190775	-18.037	0.000	.4718232	.5466731
eciv4	.7727578	.0821672	-2.424	0.015	.6273868	.9518127
pos2	2.45137	.1014483	21.666	0.000	2.260385	2.658492
pos3	1.830766	.0739711	14.967	0.000	1.691377	1.981642
pos9	3.849144	.270396	19.187	0.000	3.354044	4.417329
edo1	29.62711	5.849704	17.163	0.000	20.11983	43.62688
edo2	.8563314	.0441827	-3.006	0.003	.7739696	.9474578
edo3	1.145846	.1693095	0.921	0.357	.8577351	1.530732
edo4	.3559328	.0342236	-10.744	0.000	.2947972	.4297467
edo5	.8849608	.0378963	-2.854	0.004	.813717	.9624422
edo6	.6492635	.0536558	-5.226	0.000	.5521751	.7634228
edo7	.7894081	.0646	-2.890	0.004	.6724265	.926741
edo8	.7592086	.0397979	-5.255	0.000	.6850794	.8413589
edo10	1.161301	.0936242	1.855	0.064	.9915645	1.360094
edo11	1.641954	.1116157	7.295	0.000	1.437138	1.875959
edo12	5.722071	.5196394	19.208	0.000	4.789089	6.83681
edo13	1.558742	.1002356	6.903	0.000	1.37416	1.768117
edo14	.9057235	.0316495	-2.834	0.005	.8457682	.9699289
edo15	2.747977	.0990951	28.032	0.000	2.560459	2.949228
edo16	3.392232	.2616635	15.835	0.000	2.916266	3.94588
edo17	.6880781	.0667967	-3.851	0.000	.5688601	.832281
edo18	4.129348	.4272115	13.707	0.000	3.371463	5.057601
edo19	.6748122	.0206277	-12.867	0.000	.6355698	.7164775
edo20	.8124847	.0477364	-3.534	0.000	.7241093	.9116462
edo21	6.481803	.5512944	21.975	0.000	5.486544	7.657601
edo22	1.85125	.1349847	8.446	0.000	1.60472	2.135653
edo23	.1260047	.0177744	-14.685	0.000	.0955685	.166134
edo24	.4846628	.032073	-10.945	0.000	.4257068	.5517836
edo25	4.882187	.4333374	17.864	0.000	4.102633	5.809866
edo26	2.667332	.1898217	13.786	0.000	2.320069	3.066572
edo27	1.482419	.1163075	5.018	0.000	1.271123	1.728839
edo28	.7451702	.0327045	-6.702	0.000	.6837501	.8121076
edo29	4.314583	.6809022	9.264	0.000	3.166702	5.878553
edo30	2.897111	.1822206	16.912	0.000	2.561102	3.277204
edo31	4.291447	.3848316	16.244	0.000	3.599756	5.116047
edo32	.9643223	.0790419	-0.443	0.658	.8212066	1.132379
esc78	.106765	.0067144	-35.572	0.000	.0943837	.1207704

esc81		.1260392	.0060088	-43.444	0.000	.1147957	.138384
esc82		.101531	.0126874	-18.305	0.000	.0794753	.1297075
esc84		.1207563	.0078717	-32.430	0.000	.106273	.1372134
esc85		.1055318	.0089629	-26.477	0.000	.0893491	.1246453
esc88		.2058843	.0085988	-37.841	0.000	.1897022	.2234468
esc89		.4602646	.0353112	-10.114	0.000	.396008	.5349475
esc90		.2036609	.0549454	-5.898	0.000	.1200227	.3455828
esc95		.5580329	.0511213	-6.368	0.000	.466317	.6677875
esc98		.0266507	.001673	-57.746	0.000	.0235654	.0301399
esc100		.4190131	.0800619	-4.552	0.000	.2881287	.6093528
esc101		.0364932	.0021138	-57.156	0.000	.0325768	.0408804
esc102		.1328548	.0108086	-24.811	0.000	.1132731	.1558216
esc103		.4872478	.0599334	-5.845	0.000	.3828677	.6200848
esc104		.5255489	.0273699	-12.353	0.000	.4745518	.5820264
esc105		.2375977	.0564309	-6.051	0.000	.1491684	.3784493
esc106		.096472	.0182221	-12.381	0.000	.0666228	.1396948
esc108		.4423631	.0338576	-10.656	0.000	.3807409	.5139587
esc109		.2134064	.008076	-40.814	0.000	.1981505	.229837
esc110		1.8596	.0979001	11.784	0.000	1.677287	2.06173
esc112		.1889924	.0452661	-6.956	0.000	.1181872	.3022165
esc113		.1020814	.0077111	-30.210	0.000	.0880335	.1183709
esc114		.1757042	.0132476	-23.064	0.000	.1515667	.2036856
esc115		.2833689	.0177977	-20.077	0.000	.2505477	.3204897
esc116		.1958768	.010233	-31.206	0.000	.1768132	.2169958
esc117		.2836297	.0415111	-8.610	0.000	.2128985	.37786
esc118		.1140297	.0079275	-31.232	0.000	.0995042	.1306757
esc119		1.181016	.125267	1.569	0.117	.9593374	1.453919
esc120		.16979	.0085596	-35.174	0.000	.1538158	.1874233
esc121		.1340522	.0098295	-27.405	0.000	.116107	.1547709
esc122		.056567	.0031982	-50.803	0.000	.0506334	.0631958
esc123		.0215606	.0036077	-22.930	0.000	.0155321	.029929
esc124		.0228984	.0027541	-31.401	0.000	.0180896	.0289857
esc125		.2294259	.0396967	-8.508	0.000	.1634415	.3220495
ram1		.336331	.0402432	-9.107	0.000	.2660219	.4252225
ram2		1.644755	.2318851	3.529	0.000	1.247656	2.168241
ram6		.0449752	.0031085	-44.876	0.000	.0392773	.0514997
ram11		.1764835	.0252656	-12.116	0.000	.1333046	.2336487
ram14		1.492517	.2793497	2.140	0.032	1.034195	2.153952
ram18		.0367731	.0085711	-14.171	0.000	.0232879	.0580668
ram32		.0562098	.0067475	-23.981	0.000	.0444255	.0711199
ram33		.6554413	.0848405	-3.264	0.001	.5085739	.8447214
ram35		.0901727	.0146461	-14.813	0.000	.0655874	.1239737
ram36		.0193623	.0062872	-12.147	0.000	.0102462	.0365892
ram39		.4461731	.034026	-10.583	0.000	.3842281	.5181049
ram40		5.597996	1.147763	8.401	0.000	3.7455	8.366722
ram42		.0057811	.0025	-11.917	0.000	.0024769	.0134927
ram45		.0483344	.008143	-17.983	0.000	.0347418	.0672451
ram46		.0737133	.0058101	-33.083	0.000	.0631617	.0860276
ram49		.0125691	.0027785	-19.798	0.000	.0081497	.0193853
ram51		.4608627	.0458756	-7.782	0.000	.3791757	.5601477
ram52		.0230843	.0056137	-15.497	0.000	.0143324	.0371803
ram53		.0358875	.0070714	-16.886	0.000	.0243904	.0528041
ram54		.056677	.0049515	-32.856	0.000	.0477576	.0672622
ram57		.0057283	.000415	-71.263	0.000	.0049701	.0066022
ram60		.3263436	.0188413	-19.396	0.000	.291428	.3654424
ram61		.0733432	.0066322	-28.892	0.000	.0614312	.0875651
ram62		.0336084	.0015672	-72.764	0.000	.030673	.0368247
ram63		.0953395	.0069348	-32.312	0.000	.0826719	.109948
ram64		.0501733	.0051734	-29.020	0.000	.0409926	.0614102
ram65		.1430465	.011806	-23.561	0.000	.1216817	.1681625
ram66		.0300061	.0014461	-72.757	0.000	.0273016	.0329785
ram67		.0143	.0018062	-33.628	0.000	.0111641	.0183167
ram68		.3420881	.0152557	-24.054	0.000	.313457	.3733344

ram69		6.136794	.2422761	45.956	0.000	5.679849	6.630502
ram71		.1150034	.0140473	-17.707	0.000	.0905189	.1461108
ram72		.0524338	.0036101	-42.821	0.000	.0458148	.060009
ram73		.7694202	.0298525	-6.756	0.000	.7130797	.8302121
sal1		.7032472	.0169257	-14.627	0.000	.6708437	.7372158
sal3		1.208122	.0306353	7.456	0.000	1.149546	1.269683
sal4		2.482069	.0963679	23.415	0.000	2.300199	2.678319
sal5		16.56293	.9679707	48.033	0.000	14.77037	18.57304
sal6		34.96336	4.186225	29.686	0.000	27.65013	44.21086
sal7		11.5363	.6835946	41.270	0.000	10.27136	12.95703
sal8		5.626994	.4217541	23.049	0.000	4.858221	6.517419
sal9		4.417557	.1334429	49.180	0.000	4.163605	4.686997

Como en los casos anteriores se considera el número de éxitos que predice el modelo como una medida complementaria de bondad de ajuste.

Logistic model for oae

Classified	----- True -----		Total
	D	~D	
+	240112	17809	257921
-	6463	15840	22303
Total	246575	33649	280224

Classified + if predicted Pr(D) >= .5
True D defined as oae ~= 0

Sensitivity	Pr(+ D)	97.38%
Specificity	Pr(- ~D)	47.07%
Positive predictive value	Pr(D +)	93.10%
Negative predictive value	Pr(~D -)	71.02%

False + rate for true ~D	Pr(+ ~D)	52.93%
False - rate for true D	Pr(- D)	2.62%
False + rate for classified +	Pr(~D +)	6.90%
False - rate for classified -	Pr(D -)	28.98%

Correctly classified		91.34%

Bibliografía

Abramovitz, M. *Resource and Output Trends in the U.S. since 1870*, American Economic Review 46, 1956.

Arrow, K. *Higher Education as a Filter*, Journal of Public Economics, 1973.

Becker, G. *Investment in Human Capital: A Theoretical Analysis*, Journal of Political Economy 70, 1962.

_____. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, New York, National Bureau of Economic Research (NBER), 1964.

_____. *Human Capital*, 2nd ed., Chicago University Press, 1975.

Berndt, E. *The Practice of Econometrics Classic and Contemporary*, Addison Wesley, USA, 1991.

Bracho, T. y A. Zamudio, *Los Rendimientos Económicos de la Escolaridad en México, 1989*, Economía Mexicana, vol. 3, núm. 2, México, Centro de Investigación y Docencia Económica (CIDE), 1994.

Consejo Nacional de Ciencia y Tecnología (CONACYT), *Indicadores de Actividades Científicas y Tecnológicas*, 1998.

Fabricant, S. *Economic Progress and Economic Change*, Cambridge MA, National Bureau of Economic Research (NBER), 1954.

Greene, W. *Econometric Analysis*, 2nd ed., New York, Macmillan, 1993.

Heckman, J. *Sample selection bias as a specification error*, Econometrica 47: 153-161, vol.1, 1979.

Hernández, G. *Efectos de la Pobreza Familiar Sobre la Participación en el Mercado de Trabajo, las Horas Trabajadas y el Desempleo en México*, Tesis Doctoral. Universidad de Oxford, 1996.

Hosmer D.W. & S. Lemeshow, *Applied Logistic Regression*, New York, Wiley, 1989.

Instituto Nacional de Estadística Geografía e Informática (INEGI), *Catálogo de Carreras de Nivel Técnico Profesional, Licenciatura y Posgrado*, México, 1994.

_____. *Clasificación de Actividades Económicas de la Encuesta Nacional de Empleo Urbano (CAE-ENEU 94)*, México, 1994.

_____. *Clasificación Mexicana de Ocupaciones (CMO 94)*, México, 1994.

_____. *Encuesta Nacional de Empleo Urbano. Instructivo de codificación de escolaridad*, México, 1994.

_____. *Definición Operativa de las Principales Variables Captadas en la Encuesta Nacional de Empleo*, México, 1996. (Mimeo)

_____. *Los profesionistas en México. XI Censo General de Población y Vivienda 1990*, México, 1993.

Instituto Nacional de Estadística Geografía e Informática (INEGI) y Secretaría del Trabajo y Previsión Social (STPS), *Encuesta Nacional de Empleo 1998*, México, 1998.

_____. *Base de Datos de la Encuesta Nacional de Empleo 1998*, México.

International Labor Office (ILO), *International Standard Classification of Occupations (ISCO)*, 1988.

Jusidman, C. *El sector informal en México*, Secretaría del Trabajo y Previsión Social (STPS), 1993.

Kendrick, J.W. *Productivity Trends: Capital and Labor*, New York, National Bureau of Economic Research (NBER), 1956.

Mincer, J. *Investment on Human Capital and Personal Income Distribution*, *Journal of Political Economy* 66, 1958.

_____. *On-The-Job Training: Costs, Returns, and Some Implications*, *Journal of Political Economy*, 1962.

_____. *Schooling, Experience, and Earnings*, New York, Columbia University Press, 1974.

_____. *Human Capital and Economic Growth*, Economics of Education Review, Vol. 3, #3, 1984.

OECD, *Manual on the Measurement of Human Resources Devoted to S&T "Canberra Manual"*, París, 1995.

Revenga, A. & M. Riboud, *Unemployment in Mexico: an analysis of its characteristics and determinants*, Policy Research Working Paper 1230, The World Bank, 1993.

Schultz, T. *Capital Formation by Education*, Journal of Political Economy 68, 1960.

_____. *Investment in Human Capital*, American Economic Review 51, 1961.

Solow, R.M. *Technical Change and the Aggregate Production Functions*, Review of Economics and Statistics 39, 1957.

Spence, M. *Job Market Signalling*, Quarterly Journal of Economics 87, 1973.

_____. *Market Signalling: Informational Transfer in Hiring and Related Screening Processes*, Cambridge MA, Harvard University Press, 1974.

UNESCO, *International Standard Classification of Education (ISCED)*, 1997.

Zamudio, A. *¿Sobre-educación en México?*, Documento de trabajo #126 del Centro de Investigación y Docencia Económica (CIDE), 1998.

Índice de Cuadros

Cuadro 1	Campo de Conocimiento y Nivel Considerados en El Manual de Canberra.....	8
Cuadro 2	Subgrupos de Ocupación ISCO-88 Considerados en El Manual de Canberra.....	9
Cuadro 3	Tamaño del Acervo de Recursos Humanos.....	10
Cuadro 4	Definición de la variable: Grupos de edad.....	23
Cuadro 5	Definición de la variable: Número de hijos.....	24
Cuadro 6	Definición de la variable: Estado civil.....	25
Cuadro 7	Definición de la variable: Posición en el hogar.....	25
Cuadro 8	Definición de la variable: Estado.....	26
Cuadro 9	Definición de la variable: Campo de conocimiento en el nivel técnico.....	27
Cuadro 10	Definición de la variable: Campo de conocimiento en el nivel licenciatura.....	28
Cuadro 11	Definición de la variable: Campo de conocimiento en el nivel posgrado.....	29
Cuadro 12	Definición de la variable: Rama de actividad económica.....	30
Cuadro 13	Definición de la variable: Nivel de ingreso laboral.....	32
Cuadro 14	Salario promedio de la población con educación en el nivel superior (licenciatura y posgrado) por campo de la ciencia	44
Cuadro 15	Salario promedio de la población con educación en el nivel licenciatura por campo de la ciencia	45
Cuadro 16	Salario promedio de la población con educación en el nivel posgrado por campo de la ciencia.....	46
Cuadro 17	Salario promedio y salarios relativos de la población con educación en el nivel licenciatura por disciplina.....	47
Cuadro 18	Salario promedio y salarios relativos de la población con educación en el nivel de posgrado por disciplina	50