

Buenos Aires | 13-16 de agosto de 2003

6^o

**Congreso
Nacional
de Estudios
del Trabajo**

**Los trabajadores
y el trabajo en la crisis**

LA DESOCUPACIÓN EN JUJUY. VARIABLES A TENER EN CUENTA PARA DISEÑAR POLÍTICAS TENDIENTES A DISMINUIRLA.

María Elena Marcoleri

marcoleri@imagine.com.ar

Facultad de Ciencias Económicas. UNJu. Alvear 843

1.- INTRODUCCIÓN

Las políticas de ajuste implementadas para afrontar los efectos de la desaceleración del crecimiento económico mundial, y de la crisis que afecta a varios países de América Latina, particularmente a Argentina, da como resultados que la mayor parte de las economías registrara en 2002 un crecimiento inferior al esperado inicialmente, y que algunos de los países hayan entrado en una fase de abierta recesión. Este comportamiento de la economía repercute directamente sobre el mercado laboral de la región, provocando que el empleo se expanda a una tasa inferior que la del aumento de la fuerza de trabajo, lo que explica el aumento del desempleo.

Según la Organización Internacional del trabajo (OIT) (2002), los principales indicadores que dan cuenta del progreso laboral se agrupan en tres categorías: indicadores de empleo, como la tasa de desempleo abierto urbano y la proporción de empleo informal; indicadores de ingreso, como los salarios reales de la industria y los salarios mínimos reales; y finalmente la productividad es un indicador intermedio entre las dos categorías anteriores. En general, se registró un desempeño laboral negativo en 2002, comparado con el año 2001. En Argentina, el bajo desempeño laboral se debe al deterioro generalizado de los indicadores de calidad del mercado laboral. La reducción de las tasas de empleo obedece a las crisis de los países, las que se han transformado en recesiones de índole local, como en Venezuela, y subregional, en los países del MERCOSUR, como resultado de la contracción de la economía argentina.

La política de liberalización económica ha trastocado las relaciones entre el Estado, el mundo del trabajo y el ámbito empresarial. En los logros económicos influyen hoy más las fuerzas del mercado que la mediación por conducto de actores sociales, normas legales o intervenciones del Estado. Los mercados internacionales de capital se han desconectado de los mercados de trabajo nacionales, acarreado beneficios y riesgos asimétricos para el capital y para el trabajo. Se tiene la impresión de que la economía «real» ha perdido contacto con los sistemas financieros, y viceversa.

La evolución de los mercados de trabajo, de las relaciones laborales y de las pautas de empleo, han impactado profundamente en los sindicatos y en las organizaciones de empleadores. La mundialización ha traído consigo prosperidad y desigualdades, que están convocando imperativamente a una responsabilidad social colectiva.

Para la OIT, cuyo campo de actuación se sitúa en la intersección de la sociedad, la economía y las vidas de los seres humanos, tales cambios han revestido proporciones de cataclismo, pero están sentando también las bases para su misión futura. Las mismas fuerzas que transformaron el antiguo orden están engendrando nuevas exigencias y nuevas oportunidades de acción social.

La evolución de los sistemas tecnológicos y de producción ha transformado la conciencia social y suscitado un nuevo modo de entender la identidad personal y los derechos humanos. Debido a las mayores posibilidades de elección de los consumidores y de acceso al saber, y a nuevos medios de comunicación, los individuos y las instituciones sociales no son ya meramente sujetos sino también actores en potencia de la mundialización. Las preferencias sociales influyen en el funcionamiento del mercado y repercuten en el prestigio de las empresas.

El cambio no es únicamente económico y social. En el orden político, muchos países constatan hoy que están sometidos al ojo crítico de los mercados y de la opinión pública, sin el beneficio de las subvenciones financieras.

Por otra parte, los problemas de inseguridad de los seres humanos y de desempleo han vuelto a ser uno de los elementos capitales del quehacer político en la mayoría de los países. La dimensión social de la mundialización y los problemas y exigencias que impone al mundo del trabajo tienen hoy una proyección pública. Se percibe con creciente claridad que los mercados no operan independientemente de su entorno social y político. Se estima cada vez más que la protección social y el diálogo social, por ejemplo, son elementos insoslayables del propio ajuste. La experiencia de las economías en transición, la creciente polarización social y la crisis reciente de los mercados incipientes, han puesto de manifiesto la necesidad de contar con un sólido marco social para apuntalar la nueva arquitectura financiera.

1.1. Situación Regional del NOA

Según los criterios más generalizados para encarar la división regional de la Argentina, se han considerado las cinco regiones establecidas a partir de espacios geográficos, económicos e históricos diferenciados: el Noroeste (NOA), el Nordeste (NEA), Cuyo, la Patagonia y el Área Pampeana. Las provincias que constituyen la región NOA son Catamarca, Jujuy, Santiago del Estero, Salta y Tucumán.

Las cifras del Censo de 2001 indican que la región NOA reunía una población cercana a los 3 millones y medio de habitantes, con un PBG regional estimado en 17.397 millones de pesos de 1998 en ese mismo año, y una superficie de 470.184 km². La provincia con mayor densidad de población es Tucumán, con 50.7 habitantes por km²; de entre las otras provincias de la región, la más densamente poblada es Jujuy, con 9.6 hab/km², y la menos densamente poblada es Catamarca, con 2.6 hab/km².

La población de la región representa un 11 % del total de población del país, y la distribución porcentual por provincia, dentro de la región, se muestra en el Cuadro 1, donde puede observarse que la participación de Jujuy en el total regional es de aproximadamente el 15 %.

Cuadro 1 : Población del NOA por provincia, según Censo Nacional de Población de 2001, en cifras absolutas y relativas.

PROVINCIA	Población	% Sobre Total
Catamarca	264.234	7.6
Jujuy	512.329	14.8
Salta	866.1253	25.1
Santiago del Estero	671.988	19.4
Tucumán	1.142.105	33.0
Total región NOA	3.456.809	100.0

Fuente: elaboración propia con datos del Censo Nacional 2001.

El NOA es la región que reúne mayor población, con excepción de la región pampeana, y la que, junto con el NEA, tiene mayor cantidad de población con NBI. Dado que esta situación empeoró en todo el país para 2001 y el NOA mantuvo su alto crecimiento poblacional y de desempleo, es probable que siga siendo la región del país en peores condiciones económicas y sociales. Estas cinco provincias tienen una característica social común, de mayor pobreza relativa junto con las del NEA. Sin embargo, su potencial productivo más o menos inmediato es diferente. Salta presenta alternativas mucho más variadas que las restantes del NOA, Catamarca está en el principio de una versión moderna de su desarrollo y Santiago del Estero, si bien tiene población con NBI menor que

las del resto del NOA, presuntamente debido a políticas sociales más eficaces y una distribución del ingreso algo menos regresiva, sufre un subdesarrollo productivo mayor.

Desde el punto de vista productivo, la provincia más diversificada del NOA es Salta, seguida por Tucumán (en un territorio mucho más reducido), con mayor desarrollo industrial pero con problemas de reconversión, desempleo y crisis social sumamente graves y con la más alta densidad de población de las provincias argentinas. Siguen luego, en cuanto a diversificación de la producción, Catamarca y Jujuy. Esta última, de una superficie menor que la de Salta, tiene características similares al norte de esta provincia (Puna y selvas subtropicales con pequeños valles intermedios). En último lugar se ubica Santiago del Estero.

1.2. Indicadores de la situación laboral

Desocupación

En la Argentina de la época del cambio de milenio, la desocupación abierta¹ medida en áreas urbanas, registró un sostenido aumento, pasando de tener un valor de 13.7 % en octubre de 1997, a 17.8 % en octubre de 2002. Esto implica un incremento de 30 % en el transcurso de los seis años analizados. Las cifras mencionadas corresponden al total de aglomerados urbanos relevados en el país, y se pueden observar en el Cuadro 2.

Cuadro 2: Evolución de la tasa de desocupación de las provincias del NOA y del total del país, durante el período 1997-2002 (Onda Octubre)

PROVINCIAS	1997	1998	1999	2000	2001	2002
Catamarca	13.2	12.5	16.0	16.8	19.0	20.5
Jujuy	15.5	15.6	16.1	19.1	19.5	20.5
Salta	15.3	12.0	14.1	14.9	18.3	17.7
Santiago del Estero	9.4	4.8	8.7	10.5	12.3	10.1
Tucumán	15.5	14.9	15.9	18.4	17.9	17.8
Total País	13.7	12.4	13.8	14.7	18.3	17.8

Fuente: Información de Prensa. INDEC. Mayo y octubre de 2002.

Jujuy no se aparta de esa tendencia nacional de crecimiento del desempleo, pasando de tener una tasa de desocupación de 15.5 % en octubre de 1997, a 20.5% en octubre de 2002, siendo el au-

¹ Se refiere estrictamente a personas que, no teniendo ocupación, están buscando activamente trabajo. No incluye otras formas de precariedad laboral (también relevadas por la EPH) tales como las referidas a personas que realizan trabajos transitorios mientras buscan activamente una ocupación, a aquellas que trabajan jornadas involuntariamente por debajo de lo normal, a los desocupados que han suspendido la búsqueda por falta de oportunidades visibles de empleo, a los ocupados en puestos con remuneración por debajo de la mínima o en puestos por debajo de su calificación, etc. (INDEC).

mento de 32 %. Esto muestra que el desempleo se incrementó a un ritmo más acelerado en Jujuy, que en el total de aglomerados urbanos.

Considerando la variación de este indicador entre octubre de 2001 y octubre de 2002, se observa que Salta y Santiago del Estero siguieron la tendencia nacional a disminuir la desocupación, mientras que Catamarca y Jujuy aumentaron, permaneciendo Tucumán prácticamente igual en ambos años.

Empleo

Recordando que al hablar de empleo se habla de individuos ocupados, es útil estudiar las variables relacionadas con este fenómeno, tales como la intensidad de la ocupación, la cantidad de horas semanales que se trabaja, la categoría ocupacional, entre otras. Estas variables permiten estudiar la problemática del empleo precario y del deterioro de los ingresos. La tasa de empleo refleja la proporción de personas ocupadas sobre la población total, y en la capital de la provincia dicha tasa alcanzó en octubre de 2002 el 30 %, mostrando un leve aumento a lo largo del período considerado. Para el total de aglomerados urbanos relevados, en octubre de 2002 la tasa de actividad era de 35.3 %, inferior en 1.2 puntos porcentuales a la de octubre de 1997.

A, continuación, se presenta el Cuadro 3, que muestra los datos comentados.

Cuadro 3: Evolución de la tasa de empleo de las provincias del NOA y del total del país, durante el período 1997-2002 (Onda Octubre)

PROVINCIAS	1997	1998	1999	2000	2001	2002
Catamarca	32.2	30.6	32.6	30.9	31.4	32.0
Jujuy	29.6	27.7	27.9	28.9	28.2	30.3
Salta	32.5	34.2	34.7	33.1	32.2	30.9
Santiago del Estero	29.4	29.5	29.4	30.9	30.7	31.0
Tucumán	31.9	32.4	32.7	32.6	31.3	29.8
Total País	36.5	36.9	36.8	36.5	34.5	35.3

Fuente: Información de Prensa. INDEC. Mayo y octubre de 2002.

Como puede observarse en el Cuadro anterior, todas las provincias de la región presentan una tasa de empleo inferior a la del total de aglomerados urbanos que abarca la Encuesta Permanente de Hogares, siendo Tucumán y Jujuy las jurisdicciones que presentan menores tasas de empleo.

Actividad

Para medir los niveles de actividad de la población se utiliza la tasa de actividad². Durante la década de los noventa la tasa de actividad mostró niveles más altos que en décadas anteriores, tal vez respondiendo al crecimiento de hogares cuyos jefes de hogar son mujeres que han tenido que incorporarse al mercado laboral, como también a la inserción más temprana de los jóvenes, en busca de complementar con sus ingresos los del jefe del hogar o de lograr independencia económica, entre otras causas.

A partir de 1997, la tasa de actividad del total de aglomerados urbanos del país, se ubicó entre 42 y 43 %, siendo superior a la de todas las provincias del NOA. Mientras Catamarca y Jujuy siguen la tendencia nacional al aumento de la tasa de actividad, en Salta, Santiago del Estero y Tucumán disminuyeron los indicadores de actividad. Particularmente en Jujuy, el incremento fue de casi el 9 % durante el período de análisis. En el Cuadro 4 se presentan las tasas de actividad que dan lugar a este comentado.

Cuadro 4: Evolución de la tasa de actividad de las provincias del NOA y del total del país, durante el período 1997-2002 (Onda Octubre)

PROVINCIAS	1997	1998	1999	2000	2001	2002
Catamarca	37.2	34.9	38.8	37.1	38.7	40.3
Jujuy	35.0	32.8	33.3	35.7	35.0	38.1
Salta	38.0	38.9	40.4	38.9	39.4	37.6
Santiago del Estero	32.5	31.0	32.2	34.5	35.0	34.5
Tucumán	37.7	38.0	38.9	40.0	38.2	36.3
Total País	42.3	42.1	42.7	42.7	42.2	42.9

Fuente: Información de Prensa. INDEC. Mayo y octubre de 2002.

Subocupación

La subocupación visible, según definiciones del INDEC, se refiere a las personas ocupadas que trabajan menos de 35 horas semanales, desean trabajar más horas y se encuentran buscando activamente otra ocupación³. Dentro de la población subocupada se consideran dos categorías: demandante y no demandante. Son demandantes quienes son subocupados por causas involuntarias y dispuestos a trabajar más horas, que además buscan activamente otra ocupación; mientras que los

² Representa la proporción de población económicamente activa sobre el total de la población, o lo que es lo mismo, el peso que tienen en la población total los individuos que trabajan y los que buscan activamente trabajo.

³ Los motivos para esta actitud pueden ser diversos, pero en la actualidad es apropiado pensar que la grave crisis económica y social que atraviesa el país en su totalidad, y la provincia de Jujuy en particular, la incipiente inflación con el consiguiente deterioro de los salarios, lleva a las personas a necesitar trabajar mayor cantidad de horas para sostener un nivel de vida apenas digno.

subocupados no demandantes son aquellas personas que estarían dispuestas a trabajar más horas, pero que no buscan otra ocupación activamente.

Comentando solamente la evolución de la tasa de subocupación de los demandantes, debido a que presenta mayor magnitud, puede observarse que en todas las provincias de la región, como en el total de aglomerados urbanos del país, dicha tasa aumentó sostenidamente, siendo Salta la que presenta un comportamiento similar al total del país. Jujuy, se ubica en segundo lugar, después de Catamarca; con una tasa de 15 %, cifra esta que representa un incremento porcentual de 55.6 entre 1997 y 2002. En ese mismo período la tasa del total del país pasa de 8.1 % a 13.8 %, lo cual indica una variación porcentual de 70 %. Santiago del Estero resulta ser la provincia con menor tasa de subocupación demandante, no alcanzando el 10 %. En el Cuadro 5 se presentan los datos que corroboran estos comentarios.

A modo de conclusión, puede decirse que la situación laboral de los habitantes de Jujuy es tanto o más grave que la de otras provincias de la región y de la Argentina, porque a la recesión económica que lleva a muchas empresas a cerrar sus puertas creando más desempleo, se suma el aumento del fenómeno de la sobreocupación de los ocupados, ya sea para obtener mayores ingresos o para mantener los actuales, en el caso de los trabajadores por cuenta propia, como por la presión de los empleadores en el caso de los asalariados.

Cuadro 5: Evolución de la tasa de subocupación de las provincias del NOA y del total del país, durante el período 1997-2002 (Onda Octubre)

PROVINCIAS	1997		1998		1999		2000		2001		2002	
	D	No D										
Catamarca	9.7	4.4	8.7	4.4	12.0	2.9	11.6	3.1	14.1	3.9	17.5	7.0
Jujuy	9.7	6.3	10.7	5.0	9.5	6.5	11.5	4.2	12.1	4.5	15.1	4.5
Salta	8.8	6.1	9.3	4.8	10.1	5.6	13.6	3.4	14.2	6.3	13.8	7.7
Santiago del Estero	6.0	5.3	8.6	4.9	6.6	4.5	4.1	4.2	7.6	4.5	9.6	3.4
Tucumán	11.2	4.7	12.4	4.4	14.0	4.1	12.0	4.5	11.6	6.1	16.1	6.0
Total País	8.1	5.0	8.4	5.2	9.1	5.2	9.3	5.3	10.7	5.6	13.8	6.1

Fuente: Información de Prensa. INDEC. Mayo y octubre de 2002, octubre 2000.

1.3. Objetivos

El objetivo principal de este trabajo, sin pretender realizar predicciones con respecto al desempleo en Jujuy, es investigar si las variables propuestas como independientes o explicativas en el modelo de regresión logística, contribuyen a aumentar o disminuir la probabilidad de estar desocupado que tiene un individuo, y que, si resultaran estadísticamente significativas, debieran tenerse en

cuenta al planificar políticas de acción social tendientes a disminuir el desempleo a límites normales e inevitables.

Para alcanzar ese objetivo general se proponen los siguientes objetivos específicos:

- a) Probar si las variables demográficas, edad y sexo, influyen en alguna medida en la probabilidad que tiene un individuo de estar desocupado.
- b) Analizar si las variables de capacitación, el nivel educacional alcanzado y la calificación de la ocupación, contribuyen a determinar la probabilidad de que un individuo sea desocupado.
- c) Investigar si la rama de la actividad económica y la categoría ocupacional son variables explicativas de la probabilidad que tiene un individuo de estar desocupado.

1.4. Antecedentes

Según Monza (1995), el fenómeno de crecimiento del desempleo es de carácter mundial, y alcanza a las sociedades más ricas del planeta; pero en el caso argentino, hasta principios de los 90, no puede asociarse al desarrollo de procesos de reestructuración productiva, ni a una incorporación importante y difundida de innovaciones tecnológicas.

En Argentina, el nivel de desempleo aumentó dramáticamente como resultado de una caída en la demanda de mano de obra en las industrias, la construcción y los servicios públicos, entre otras actividades, en un período en que se expandió la oferta de fuerza de trabajo (Marshall, 1996).

Actualmente, las consecuencias de la reestructuración económica se manifiesta especialmente en el deterioro del nivel y las condiciones de empleo. En este marco, aparecen manifestaciones xenófobas que adjudican a los migrantes limítrofes la responsabilidad del aumento de la desocupación, de la pobreza y de otros problemas sociales. Se percibe al extranjero como una “amenaza” de las fuentes de trabajo. Sin embargo, los resultados de la EPH muestran que la migración originaria de países vecinos no explica ni contribuye al aumento de los problemas de empleo. En cambio, indican que los migrantes limítrofes revelan una menor proporción de desocupación que los migrantes internos y los no migrantes. Este hecho podría deberse a que aquel grupo es menos exigente a la hora de conseguir empleo, aceptando condiciones más precarias de inserción laboral (Maguid, 1995).

Por ser el desempleo, como los demás indicadores laborales de actividad, empleo y subocupación, dependiente de diferentes factores relacionándose en forma compleja en el ámbito económico, demográfico y social, requiere de métodos estadísticos especiales para describir el comportamiento conjunto de las variables que intervienen.

Sana y Pantelides (1999), utilizaron modelos de regresión logística para realizar un análisis multivariado de la pobreza entre los ancianos, en el cual la condición de pobre determinada por el método de la Línea de Pobreza (LP) es considerada como variable dependiente.

Cortés (1996), intenta dar una visión panorámica del uso de la regresión logística en las ciencias sociales, en general, y en la Ciencia Política, en particular. Muestra dos ejemplos en que se aplica regresión logística. El primero de ellos trata del análisis de votos tomando como unidad de análisis las divisiones político administrativas de México y tiene un carácter netamente teórico. El segundo ejemplo, muestra los resultados a los que se llegó, al estudiar las determinantes de la pobreza, también en México. En este caso la intención no es entrar en los vericuetos estadísticos sino limitarse al tipo de conocimiento que se genera al aplicar esta técnica de análisis de datos. La variable dependiente es el tipo de hogar, con dos categorías: *pobre* y *no pobre*. Considera como variables explicativas el nivel de instrucción del jefe de hogar, la tasa de dependencia (el tamaño del hogar dividido por el número de preceptores), el nivel de ingreso monetario medio por preceptor de los hogares expresado en moneda constante, la zona de localización del hogar (urbana o rural), el sexo del jefe de hogar, y la composición por sexo de los preceptores de salarios. En las conclusiones del documento destaca las ventajas de aplicar la técnica de regresión logística en el análisis de datos resultantes de la investigación en ciencias sociales.

Salvia (1993) aplicó el procedimiento de regresión logística para explicar el comportamiento estratégico de retiro voluntario de un conjunto de trabajadores ocupados en una empresa minera del sector público argentino.

Una descomposición de la varianza del desempleo, considerando la tasa de desempleo en términos de la población total, en lugar de tomarla sobre la fuerza laboral, consiste en un análisis de regresión por mínimos cuadrados ordinarios de la proporción de empleo (relativo de la población) contra la tasa de desempleo (relativo de la población). Esto equivale a una descomposición de varianza, pero evita preocuparse por los posibles problemas de covarianza entre los movimientos de la oferta y la demanda de trabajo.

Mondino et al (1998) muestran los hechos empíricos básicos referentes a los mercados de trabajo en las provincias en las últimas décadas y, a partir de modelos autorregresivos univariados, calculan funciones de impulso-respuesta para las variables crecimiento del empleo relativo, desempleo relativo y participación relativa.

Jiang (1998), en un artículo sobre estimadores consistentes en Modelos Lineales Generalizados Mixtos, muestra mediante un ejemplo con variable de respuesta binaria y transformación logit, utilizando método de simulación Monte Carlo, que los estimadores obtenidos por el método de

los momentos son consistentes, pero que para muestras pequeñas estos estimadores presentan una varianza sustancialmente mayor que los estimadores de quasi-verosimilitud penalizados. Esto puede deberse al hecho de que los estimadores del método de los momentos suelen ser ineficientes.

Para analizar el cambio en la probabilidad de la participación femenina en los mercados de trabajo en el NOA, en los 80, como función de un conjunto de variables independientes, Paz (1994) utiliza el modelo logit, considerando como covariables la edad, el estado civil y el nivel de educación, con sus interacciones.

Siqueira y Taylor (1999) consideran la aplicación de un modelo logístico para una variable de respuesta binaria que indica tratamiento o grupo, utilizando la transformación de Box-Cox. Dicen que este modelo es aplicable en estudios observacionales donde se investiga un efecto tratamiento después de controlar una variable confundida X , continua, a la cual se le aplica la citada transformación. Concentran su atención en la inferencia del parámetro que representa el efecto de tratamiento. Prueban dos métodos y llegan a la conclusión de que bajo la hipótesis nula de que ese parámetro es cero, ambos métodos son asintóticamente equivalentes si X tiene la misma distribución en los dos grupos. Los dos procedimientos difieren asintóticamente si el parámetro en cuestión es distinto de cero.

Nebot et al. (2002) analizaron la relación entre el apoyo social y la mortalidad de una cohorte de población de 60 años o más, residente en Barcelona, España, no institucionalizada. Realizaron un análisis bivariado para identificar las variables asociadas con la mortalidad, y un análisis multivariado mediante regresión logística para determinar la influencia independiente de esas variables.

Con el objetivo de revisar las diferentes soluciones que distintos autores han dado al problema del cálculo del tamaño muestral, cuando se aplica la regresión logística no condicionada en la modelación estadística para predicción del riesgo en estudios epidemiológicos, Ortega Calvo y Cayuela Domínguez (2002), realizaron un exhaustivo análisis bibliográfico sobre el tema.

2. METODOLOGÍA

2.1. Las muestras

Las principales fuentes de datos utilizadas para realizar este trabajo son la Encuesta Permanente de Hogares (EPH), de la cual se tomó la Onda de octubre de los años 1997 a 2002, el Censo Nacional de Población de 2001 y las publicaciones de octubre de 2000, 2001 y 2002 de Información de Prensa sobre el mercado de trabajo, del INDEC.

Del total de individuos que constituyen la BUA⁴ se seleccionaron aquellos que componen la población económicamente activa (PEA), es decir, ocupados más desocupados, quienes suman entre 936 y 1072 personas, cifra que representa aproximadamente el 35% del total de las muestras.

En los archivos resultantes se seleccionaron aleatoriamente la mitad de los individuos, de manera tal que quedaran dos submuestras de igual tamaño, una utilizada para la construcción del modelo, llamada *desarrollo del modelo*, y la otra para validación del modelo, llamada *submuestra de validación*⁵.

El tamaño de la submuestra de validación, es un tema que presenta distintos criterios. Al respecto, Edward W. Frees (1996), dice que varios investigadores han recomendado diferentes proporciones para la asignación del tamaño de dicha submuestra. Entre ellos menciona a Snee (1977), quien considera que el tamaño de la muestra de validación debe ser mayor que $2(k+1) + 20$, donde k es el número de variables explicativas que intervienen en el modelo. También se refiere a Picard y Berk (1990), quienes opinan que la cantidad de parámetros que se deben estimar debe ser una guía para determinar la proporción de observaciones necesarias para la submuestra de desarrollo del modelo; y como regla práctica señalan que para muestras de menos de 100 elementos se deben usar entre el 25-35% de las observaciones para la submuestra de validación, y para muestras de más de 500 observaciones, es conveniente utilizar el 50% de las observaciones para la validación del modelo. Siguiendo estas indicaciones es que se decidió la división de la muestra disponible para el análisis en dos submuestras de igual tamaño.

2.2. Las variables a analizar

Las variables seleccionadas con sus respectivas modalidades, como se presentan en la base de datos original, se describen a continuación:

Variable dependiente:

Y : Estado (Condición de actividad)

Modalidades: Desocupado (0)
Ocupado (1)

⁴ Base Usuaría Ampliada: consiste en una base de datos que contiene información de la EPH con mayor nivel de desagregación conjuntamente con variables construidas.

⁵ La validación es un proceso para confirmar que el modelo propuesto es apropiado, especialmente si el propósito de la investigación es realizar predicciones. Ante la imposibilidad de contar con dos muestras diferentes, a fin de utilizar una de ellas para la validación externa del modelo, se recurrió a esta técnica de división de la muestra total.

Variables independientes (explicativas):

X₁ : Sexo	Modalidades: Masculino	(1)
	Femenino	(2)
X₂ : Posición en el hogar	Modalidades: Jefe del hogar	(1)
	No Jefe	(2)
X₃ : Grupos de Edad	Modalidades: Hasta 13 años	(0)
	De 14 a 19 años	(1)
	De 20 a 24 años	(2)
	De 24 a 49 años	(3)
	De 50 a 59 años	(4)
	60 años y más	(5)
X₄ : Nivel educativo alcanzado	Modalidades: Bajo (Hasta Primario Incompleto)	(1)
	Medio (Hasta Secundario Incompleto)	(2)
	Alto (Superior)	(3)
X₅ : Categoría Ocupacional	Modalidades: Patrón o empleador	(1)
	Trabajador por su cuenta (Cta. Propia)	(2)
	Obrero o empleado (Asalariado)	(3)
	Trabajador sin salario	(4)
X₆ : Rama de Actividad	Modalidades: Industria	(1)
	Construcción	(2)
	Comercio, Restaurantes, Hoteles	(3)
	Servicios Comunes, Sociales	(4)
	Actividades Primarias	(5)
X₇ : Calificación de la ocupación	Modalidades: Calificación Profesional	(1)
	Calificado/Semicalificado	(2)
	No Calificado	(3)

Recodificación de variables cualitativas

Puesto que la metodología empleada para la estimación del modelo logístico se basa en la utilización de variables cuantitativas, igual que cualquier otro procedimiento de regresión, es incorrecto que en él intervengan variables cualitativas, ya sean nominales u ordinales.

En todo análisis de regresión, el tema más importante es la interpretación de los coeficientes del modelo estimados, y desde que esta interpretación depende fundamentalmente de cómo las va-

variables explicativas hayan sido codificadas, o en términos técnicos, cómo el modelo ha sido parametrizado, este punto requiere especial atención y tratamiento.

La codificación de las variables asignando un número a cada categoría es un procedimiento apropiado cuando las variables cualitativas son dicotómicas, donde se hace corresponder el código 1 a la presencia del suceso de interés y 0 a la ausencia del mismo. Sin embargo, no resuelve este inconveniente cuando las variables tienen más de dos categorías de respuestas, ya que conduce a interpretaciones sin sentido. Por ejemplo, la variable nivel de instrucción con sus tres modalidades, bajo, medio y alto, con códigos 1, 2 y 3 respectivamente, estos números significan a efectos del modelo, que al haber alcanzado nivel educativo alto, el riesgo de estar desocupado es tres veces mayor que si se tiene nivel bajo, lo cual carece de sentido. Esta interpretación es más absurda aún cuando se trata de variables cualitativas con escala nominal, sin ninguna relación de orden entre las categorías, como podría ser el sexo de los individuos, su lugar de nacimiento, o su posición en el hogar.

Este problema se soluciona creando, para cada una de las variables explicativas, tantas variables dicotómicas como cantidad de respuestas – 1. Estas nuevas variables, creadas artificialmente, reciben el nombre de **variables internas, indicadoras, o variables diseño**⁶.

De acuerdo a las consideraciones anteriores, se procedió a recodificar la variable dependiente, estado, y las independientes sexo y posición en el hogar, todas dicotómicas, y las indicadoras (dummy) correspondientes a variables originales con más de dos categorías se construyeron tomando la primera de las posibles respuestas como categoría o nivel de referencia. El programa de computación SPSS (Statistical Program for Social Sciences) utilizado para el procesamiento de los datos, recodifica automáticamente las variables mediante esta técnica, y presenta la opción de considerar la primera o la última respuesta como categoría de referencia.

El Cuadro 5 es la salida de computadora que muestra la recodificación de las variables.

⁶ En la literatura anglosajona son llamadas variables “dummy”.

Cuadro 5: Variables categorizadas recodificadas

Categorical Variables Codings

		Frequency	Parameter coding				
			(1)	(2)	(3)	(4)	(5)
Rama de actividad	Industria	37	,000	,000	,000	,000	,000
	Construcción	41	1,000	,000	,000	,000	,000
	Comercio, Restaurantes Hoteles	82	,000	1,000	,000	,000	,000
	Servicios Comunales, Sociales	324	,000	,000	1,000	,000	,000
	Actividades Primarias	11	,000	,000	,000	1,000	,000
	Nuevos trabajadores	32	,000	,000	,000	,000	1,000
	Categoría ocupacional	No responde	32	,000	,000	,000	,000
CALIFICA	Patrón o empleador	14	1,000	,000	,000	,000	
	Trabajador por su cuenta	110	,000	1,000	,000	,000	
	Obrero o empleado	370	,000	,000	1,000	,000	
	Trabajador sin salario	1	,000	,000	,000	1,000	
Grupos de edad	Calificado	92	,000	,000	,000	,000	
	Semicalificado	224	1,000	,000	,000	,000	
	No calificado	149	,000	1,000	,000	,000	
	Calificación profesional	29	,000	,000	1,000	,000	
	0	33	,000	,000	,000	1,000	
Nivel de instrucción	De 14 a 19 años	22	,000	,000	,000	,000	
	De 20 a 24 años	81	1,000	,000	,000	,000	
	De 25 a 49 años	348	,000	1,000	,000	,000	
	De 50 a 59 años	54	,000	,000	1,000	,000	
	De 60 años y más	22	,000	,000	,000	1,000	
Sexo	Bajo	142	,000	,000			
	Medio	257	1,000	,000			
	Alto	128	,000	1,000			
Posición en el hogar	Mujer	229	,000				
	Varón	298	1,000				
Posición en el hogar	No Jefe	291	,000				
	Jefe de hogar	236	1,000				

En esta forma de codificación el coeficiente de la ecuación de regresión para cada variable indicadora (transformado con la función exponencial), corresponde al *odds ratio* de esa categoría con respecto al nivel de referencia (la primera respuesta), por ejemplo, cuantifica cómo cambia el riesgo de estar desocupado al haber alcanzado el nivel medio de instrucción con respecto a tener nivel bajo; o también, cuantifica cómo cambia el riesgo de estar desocupado si se es asalariado con respecto a si se es patrón o empleador.

2.3. Fundamentos del uso del análisis de regresión logística

En este trabajo se busca estimar la probabilidad de la presencia o ausencia de individuo desocupado, teniendo en cuenta los valores del vector de covariables.

La variable dependiente "Situación Laboral" requiere de una respuesta "sí" o "no" en las encuestas, es decir, es dicotómica por naturaleza.

Debido a que la técnica de regresión logística tiene como objetivo principal modelar cómo influye en la probabilidad de aparición de un suceso, habitualmente dicotómico, la presencia o ausencia de diversos factores considerados relevantes, y el valor o nivel de los mismos, el modelo de regresión logística múltiple con las variables seleccionadas resulta adecuado para modelar el fenómeno del desempleo.

La función logística encuentra la probabilidad π de que cada individuo presente el efecto de interés (en este caso desocupado) según los valores de una serie de covariables (X_i). Por lo tanto, la regresión logística consiste en obtener una función logística de las variables independientes que permite clasificar a los individuos en una de las dos subpoblaciones o grupos establecidos por las dos modalidades de la variable dependiente.

También es conveniente destacar que, además de predecir el riesgo del individuo de estar desocupado, la regresión logística puede ser útil para estimar la fuerza de la asociación de cada factor de riesgo de forma independiente, es decir, eliminando la posibilidad de que un factor confunda el efecto de otro.

Una de las características que hacen tan interesante la regresión logística es la relación que los coeficientes de regresión guardan con un parámetro de cuantificación de riesgo, conocido en la literatura como "**odds ratio**" (aunque puede tener traducción al castellano, renunciamos a ello para evitar confusión ya que siempre se utiliza la terminología inglesa). El odds asociado a un suceso es el cociente entre la probabilidad de que ocurra frente a la probabilidad de que no ocurra (Molineiro, 2001).

En regresión logística no se asume el supuesto de linealidad de la relación entre las variables independientes y la dependiente, no se requiere que las variables y el término error sean normalmente distribuidas, no se asume homoscedasticidad y en general tiene requerimientos menos condicionantes, comparada con la regresión lineal. Sin embargo, la solución sería más estable si las variables explicativas tuvieran distribución normal multivariada; la multicolinealidad entre dichas variables puede conducir a estimaciones sesgadas.

Como regla general, la regresión logística ajusta con mayor precisión que la regresión por mínimos cuadrados, para propósitos de clasificación de individuos. Esta es otra razón para preferir la regresión logística sobre los mínimos cuadrados para datos que involucran variables dependientes binarias.

2.4. Método de selección de variables

Los modelos se estimaron utilizando el método de entrada en bloque de las variables por pasos hacia adelante⁷, basado en la razón de máxima verosimilitud para la entrada de las variables al modelo y para la eliminación de ellas.

El empleo de este procedimiento de selección de las variables paso a paso, puede proveer un medio rápido y efectivo de elegir un gran número de variables, y ajustar, simultáneamente, varias ecuaciones de regresión logística.

El proceso de selección o eliminación de las variables del modelo está basado en un algoritmo estadístico que prueba la importancia de las mismas, y según una regla de decisión fijada, las incluye o excluye. La importancia de una variable se define en términos de una medida de la significación estadística de su coeficiente de regresión. La estadística que se aplica depende de los supuestos del modelo en cuanto a distribución de los residuos. En regresión logística se supone que los errores tienen distribución binomial, y esa significación se evalúa por un test chi-cuadrado de razón de verosimilitud⁸. Entonces, en cada paso del proceso, la variable más importante será aquella que produce un cambio mayor en el logaritmo de la razón de verosimilitud, comparada con el del modelo que no contiene esa variable.

El procedimiento de selección stepwise identifica variables como candidatas para el modelo solamente desde el punto de vista estadístico. Entonces, luego de la selección de efectos principales todas las variables deben ser cuidadosamente analizadas por su credibilidad aparente en el tema de que se trata. En general, las interacciones entre las variables deben alcanzar al menos un nivel de significación estadística moderado, para modificar los estimadores puntuales y por intervalos de confianza de los efectos principales del modelo. Por lo tanto, el procedimiento de selección stepwise de interacciones puede proveer una contribución valiosa para la identificación del modelo, especialmente cuando existe un gran número de interacciones posibles generadas desde los efectos principales.

4. RESULTADOS

⁷ Stepwise forward selection.

⁸ En regresión lineal se usa un test F porque se supone que los errores están normalmente distribuidos.

Se analizan los resultados del último block de las salidas de computadora, de los cuales se adjunta, a modo de ejemplo, un Anexo con el block 7 de los datos correspondientes al año 2001. A los fines de comparación con el año anterior, en el análisis de los resultados se consignan los cuadros correspondientes a 2002.

4.1. Test ómnibus sobre los coeficientes del modelo

El primer cuadro que aparece, luego de la historia de la iteración del procedimiento, es el resultado de un test ómnibus sobre los coeficientes del modelo, presentando tres valores de la estadística chi cuadrado: para el paso, para el block y para el modelo. A continuación se presenta dicho cuadro para el año 2002:

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	10,912	3	,012
Block	10,912	3	,012
Model	185,783	18	,000

a) *Chi cuadrado del paso (step chi-square)* representa el cambio en chi cuadrado del modelo debido a la regresión logística paso a paso. Los datos analizados muestran que este chi cuadrado resulta estadísticamente significativo ($p < 0.05$); y esto es así para todos los años, excepto para 1998.

b) *Chi cuadrado del block* es un test de razón de verosimilitud que representa el cambio en chi cuadrado del modelo debido a la entrada de las variables en bloques. Con las variables categorizadas, chi cuadrado del bloque se usa para testear el efecto de la entrada de una variable categórica. En tal caso, todas las variables dummy asociadas con la variable categórica son entradas como un bloque. Para los años analizados en este trabajo, este chi cuadrado resulta estadísticamente significativo ($p < 0.05$); excepto para 1998.

Los resultados obtenidos en los puntos a) y b) explican que las variables introducidas en cada paso y en cada block tienen alguna influencia en la probabilidad de estar desocupados de los individuos.

c) El estadístico *Chi cuadrado del modelo*, es también conocido como G_M , G de Hosmer y Lemeshow, o bondad de ajuste. La función chi cuadrado del modelo es un test de significación para

el modelo logístico⁹, prueba la hipótesis nula de que todos los coeficientes de regresión logística poblacionales son ceros. Es un test sobre todo el modelo, el cual no asegura que cada variable independiente sea significativa. En todos los años analizados se rechaza la hipótesis nula mencionada ($p < 0.05$), por lo tanto, los coeficientes del vector de covariables del modelo resultante son adecuados para cuantificar la relación entre las variables explicativas y la probabilidad de ser desocupado..

4.2. Test de Hosmer y Lemeshow

El cuadro siguiente muestra el resultado del test de Hosmer y Lemeshow, que se basa en una estadística de bondad de ajuste que se obtiene calculando la estadística χ^2 de una tabla de contingencia de $2 \times g$ de frecuencias observadas y esperadas, siendo g el número de grupos de individuos que se obtiene por distintas formas de colapsar la tabla de datos. Prueba la hipótesis nula de que los datos son generados por el modelo ajustado por el investigador, es decir, de que no hay diferencias entre los valores observados de la variable dependiente y los que se pueden predecir por el modelo.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8,845	8	,356

En todos los años analizados, se observa que no existen evidencias suficientes para rechazar dicha hipótesis nula, por lo tanto, los estimadores del modelo ajustan a los datos a un nivel de significación aceptable, el cual, en el cuadro anterior, es $0.356 > 0.05$, como es deseable.

El test divide a los individuos en deciles basados en las probabilidades predichas, luego computa el χ^2 a partir de las frecuencias observadas y esperadas, dando lugar a la tabla de contingencia que se muestra a continuación:

⁹ Como lo es el test F para el modelo de regresión por mínimos cuadrados ordinarios, o el test de razón de verosimilitud (G2) en análisis loglineal.

Contingency Table for Hosmer and Lemeshow Test

		Situación laboral = Ocupado		Situación laboral = Desocupado		Total
		Observed	Expected	Observed	Expected	
Step	1	53	54,129	2	,871	55
1	2	50	49,336	1	1,664	51
	3	53	51,359	1	2,641	54
	4	51	50,315	3	3,685	54
	5	59	56,176	4	6,824	63
	6	43	44,335	8	6,665	51
	7	33	37,745	12	7,255	45
	8	38	38,143	11	10,857	49
	9	35	35,206	20	19,794	55
	10	8	6,259	42	43,741	50

donde puede observarse que tanto para los ocupados como para los desocupados, las frecuencias observadas se aproximan bastante a las esperadas, estimadas por el modelo.

4.3. Tabla de clasificación

A continuación, en la salida de computadora, se registra una tabla de clasificación de los individuos en dos grupos que responden a las dos categorías de la variable dependiente: Ocupado y Desocupado. Es una tabla 2x2 que cuenta los casos correcta e incorrectamente clasificados. Las columnas contienen los dos valores predichos de la variable de respuesta, mientras que las filas muestran los dos valores observados de esa variable. En un modelo de ajuste perfecto, todos los casos deberían estar sobre la diagonal y el porcentaje correcto sobre el total debe ser 100 %. Si el modelo logístico tiene homoscedasticidad¹⁰, el porcentaje correctamente clasificado debe ser aproximadamente el mismo para ambas filas. Como esta tabla toma la forma de un cuadro de contingencia, se pueden calcular medidas de asociación como una vía para resumir la de la tabla. La versión 10.0 de SPSS agrega a esta tabla una subdivisión de los valores predichos en dos columnas, una para los casos seleccionados como submuestra para ajustar el modelo¹¹, y la otra para la submuestra de validación. Entonces, la comparación también se realiza en el sentido horizontal de la tabla, de modo que se comparan los resultados obtenidos para ambas submuestras.

A continuación se transcribe la tabla de clasificación resultante en el análisis de los datos de octubre de 2002:

¹⁰ Aunque no es una suposición de regresión logística.

¹¹ Desarrollo del modelo.

Classification Table

Observed	Predicted						
	Selected Cases ^a			Unselected Cases ^{b,c}			
	Situación laboral		Percentage Correct	Situación laboral		Percentage Correct	
	Ocupado	Desocupado		Ocupado	Desocupado		
Step 1 Situación laboral Ocupado	339	84	80,1	304	111	73,3	
Desocupado	31	73	70,2	36	77	68,1	
Overall Percentage			78,2			72,2	

a. Selected cases 536 from the first 1072 cases (SAMPLE) EQ 0

b. Unselected cases 536 from the first 1072 cases (SAMPLE) NE 0

c. Some of the unselected cases are not classified due to either missing values in the independent variables or category values out of the range of the selected cases.

d. The cut value is ,190

En la submuestra de los casos seleccionados para desarrollo del modelo, el porcentaje de individuos correctamente clasificados varía entre el 61 % y el 80 % para los ocupados, y oscila entre 65 % y el 89 % para los desocupados, para todos los años analizados.

En la submuestra de validación del modelo, el porcentaje de individuos correctamente clasificados varía entre el 57 % y el 74 % para los ocupados, y oscila entre 56 % y el 70 % para los desocupados, considerando todos los años analizados.

La tasa total de individuos mal clasificados, es el número de observaciones mal clasificadas dividida por el número total de observaciones. En este análisis particular, la tasa total de individuos desocupados mal clasificados varía entre 1.5 % y 7.3 %, encontrándose en general, los valores más bajos en la muestra para desarrollo del modelo, mientras que la tasa total de individuos ocupados mal clasificados, oscila entre el 15.7 % y el 35.8 %.

La tabla siguiente muestra las tasas de individuos mal clasificados por año, según situación laboral y la submuestra a la que pertenecen.

Cuadro 6: Tasas de individuos mal clasificados por año, según Situación Laboral y la submuestra a la que pertenecen

Años	Situación Laboral	Submuestra Modelo	Submuestra Validación
1997	Desocupado	4 %	3 %
	Ocupado	32.5 %	35.8 %
1998	Desocupado	1.5 %	5.9 %
	Ocupado	29.6 %	30.9 %
1999	Desocupado	4.8 %	5.4 %
	Ocupado	26.2 %	27 %
2000	Desocupado	4.2 %	7.3 %
	Ocupado	19.4 %	20.6 %
2001	Desocupado	5.2 %	7 %
	Ocupado	23.8 %	30.5 %
2002	Desocupado	5.8 %	6.7 %
	Ocupado	15.7 %	20.7 %

Fuente: Elaboración propia con los datos de las tablas de clasificación.

4.4. Variables en la ecuación

En este cuadro aparecen las variables que han sido incluidas en el modelo, las estimaciones de los coeficientes con sus respectivos errores estándares, la transformación exponencial de los coeficientes, que constituyen las razones de odds y los intervalos del 95 % de confianza de estas últimas. En la página siguiente se presenta la tabla de variables en la ecuación de Octubre de 2002.

Los coeficientes logit¹², también llamados coeficientes de regresión logística no estandarizados o coeficientes efectos, corresponden a los coeficientes b en regresión por mínimos cuadrados ordinarios y se usan en la ecuación de regresión logística para estimar los *odds* de la variable dependiente igual a 1. Si el logit para una variable independiente dada es b_1 , entonces, una unidad de incremento en la variable independiente está asociada con b_1 unidades de incremento en el log *odds* de la variable dependiente¹³. Es notable que mientras la regresión por mínimos cuadrados tiene una función de enlace identidad, la regresión logística tiene una función link logit, esto es, la regresión logística calcula cambios en el log *odds* de la variable dependiente, no cambios en la variable dependiente como la regresión por mínimos cuadrados ordinarios.

La probabilidad de que la variable dependiente asuma el valor 1 (es decir, $y = 1$, individuo desocupado) es una función de los coeficientes logit. Por ejemplo, sea $y=0$ o $y=1$, y sean x_1 , x_2 y x_3 variables continuas independientes para el modelo logístico $y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$. La esti-

12 Llamados B en las salidas de SPSS.

13 Es el logaritmo natural de la probabilidad de que la variable dependiente sea igual a 1, dividida por la probabilidad de que la variable dependiente sea igual a 0.

La estimación de la probabilidad $p(y=1)$ es el logaritmo natural e elevado a la potencia de un término que es la ecuación de regresión logística.¹⁴

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	POSICIÓN(1)	-1,595	,365	19,040	1	,000	,203
	SEXO(1)	,703	,343	4,217	1	,040	2,021
	EDAD			5,549	4	,235	
	EDAD(1)	-,995	,730	1,861	1	,173	,370
	EDAD(2)	-1,410	,700	4,052	1	,044	,244
	EDAD(3)	-1,039	,854	1,479	1	,224	,354
	EDAD(4)	-2,057	1,253	2,695	1	,101	,128
	CALIFICA			1,319	4	,858	
	CALIFICA(1)	-,147	,444	,110	1	,740	,863
	CALIFICA(2)	,164	,482	,115	1	,734	1,178
	CALIFICA(3)	-,792	1,101	,518	1	,472	,453
	CALIFICA(4)	-8,011	99,631	,006	1	,936	,000
	RAMACTIV			14,022	5	,015	
	RAMACTIV(1)	,965	,597	2,616	1	,106	2,625
	RAMACTIV(2)	-,305	,571	,285	1	,593	,737
	RAMACTIV(3)	-,825	,494	2,792	1	,095	,438
	RAMACTIV(4)	-1,543	1,210	1,624	1	,202	,214
	RAMACTIV(5)	7,440	141,872	,003	1	,958	1701,941
	CATEGORI			5,065	3	,167	
	CATEGORI(1)	-17,361	102,932	,028	1	,866	,000
CATEGORI(2)	-11,184	99,633	,013	1	,911	,000	
CATEGORI(3)	-10,311	99,632	,011	1	,918	,000	
Constant	10,651	99,636	,011	1	,915	42219,513	

a. Variable(s) entered on step 1: CATEGORI.

El odds asociado a un suceso es el cociente entre la probabilidad de que ocurra frente a la probabilidad de que no ocurra:

$$odds = \frac{p}{1-p}$$

siendo p la probabilidad del suceso. Así, por ejemplo, podemos calcular el odds de presencia de individuo desocupado cuando el nivel de instrucción es igual o superior a un cierto grado, que en realidad determina cuántas veces es más probable que haya desocupación a que no la haya en esa situación. Igualmente se podría calcular el odds de presencia de desempleo cuando el nivel de instrucción es inferior a ese grado. Si se divide el primer odds entre el segundo, se calcula un cociente de odds, esto es un odds ratio, que de alguna manera cuantifica cuánto más probable es la aparición de desocupación cuando se tiene un nivel de instrucción bajo (primer odds) respecto a cuando se tiene un nivel más alto. La noción que se está midiendo es parecida a la que encontramos en lo que

¹⁴ SPSS usa todos los coeficientes logit en el cálculo de p , sean significativos o no. Para evitar esto, vuelve a correr el modelo logístico bajando las variables independientes no significativas.

se denomina **riesgo relativo** que corresponde al cociente de la probabilidad de que aparezca un suceso (desocupación) cuando está presente el factor (bajo nivel de instrucción) respecto a cuando no lo está. De hecho cuando la prevalencia del suceso es baja (< 20 %) el valor del odds ratio y el riesgo relativo es muy parecido.

Si en la ecuación de regresión tenemos un factor dicotómico, como puede ser por ejemplo si el sujeto es no jefe de hogar, el coeficiente ***b*** de la ecuación para ese factor está directamente relacionado con el odds ratio ***OR*** de ser jefe de hogar respecto a no serlo

$$OR = \exp(b)$$

es decir que $\exp(b)$ es una medida que cuantifica el riesgo que representa poseer el factor correspondiente respecto a no poseerlo, suponiendo que el resto de variables del modelo permanecen constantes.

El intervalo de 95 % de confianza alrededor del coeficiente de regresión logística, se calcula así: $B \pm 1.96 \times \text{error estándar asintótico del coeficiente } b \text{ logístico.}$

Asintótico significa el menor valor posible para el error estándar cuando los datos se ajustan al modelo. Es también la precisión más alta posible.

5. CONCLUSIONES

Los modelos de regresión logística resultantes, en todos los años analizados, son estadísticamente significativos, y a través del test de bondad de ajuste de Hosmer y Lemeshow se prueba que, con estos modelos, el ajuste entre los datos reales y los estimados es adecuado.

Con respecto a las variables explicativas que intervienen en los modelos, puede decirse que todas están presentes en las ecuaciones de por lo menos tres años de los seis analizados, por lo tanto, se considera explican de alguna manera y de alguna manera el fenómeno del desempleo, siendo esta una razón suficiente para tenerlas en cuenta al diseñar políticas tendientes a mejorar la situación laboral de la población.

En el Cuadro 7 (una síntesis gráfica) se indica con una X la presencia de las variables en los modelos de todos los años, con las modalidades respectivas.

Cuadro 7: Variables que intervienen en los modelos de regresión logística según los años analizados

VARIABLES	A Ñ O S					
	1997	1998	1999	2000	2001	2002
SEXO (1) = Varón	X	X	X	X		X
POSICIÓN (1) = Jefe de hogar		X	X	X	X	X
EDAD (1) = De 14 a 19 años	X	X	X	X	X	X
EDAD (2) = De 20 a 24 años	X	X	X	X	X	X
EDAD (3) = De 25 a 49 años	X	X	X	X	X	X
EDAD (4) = De 50 a 59 años	X	X	X	X	X	X
EDAD (5) = De 60 años y más	X		X			
Nivel de Instrucción (1) = Bajo	X	X		X	X	
Nivel de Instrucción (2) = Medio	X	X		X	X	
Nivel de Instrucción (3) = Alto						
Calificación (1) = Profesional	X	X	X		X	X
Calificación (2) = Calificado/Semi	X	X	X		X	X
Calificación (3) = No calificado		X	X		X	X
Rama (1) = Industria		X	X	X	X	X
Rama (2) = Construcción		X	X	X	X	X
Rama (3) = Comercio, Rest., Hotel.		X	X	X	X	X
Rama (4) = Serv. Com., Soc., Pers.		X	X	X	X	X
Rama (5) = Actividades Primarias		X	X	X		X
Categoría (1) = Patrón/empleador				X	X	X
Categoría (2) = Cta. propia				X	X	X
Categoría (3) = Asalariado				X	X	X
Categoría (4) = Trabaj. sin salario				X	X	

Fuente: Elaboración propia con datos de los cuadros de Variables en la ecuación.

Observando los coeficientes de regresión en la tabla de las variables en la ecuación, puede concluirse que:

- Ser Varón disminuye la probabilidad de ser desocupado, lo mismo que ser Jefe de hogar.
- La Edad aparece alternativamente con signo positivo y negativo en las ecuaciones, pero en la mayoría de ellas, el signo es negativo, lo cual indica que tener una edad determinada disminuye la probabilidad de ser desocupado, con respecto a tener menos de 14 años (categoría de referencia).
- El Nivel de instrucción alcanzado por los individuos en su modalidad “alto”, no interviene en ninguno de los modelos, mientras que tener nivel medio disminuye la probabilidad de ser desocupado, con respecto a tener nivel bajo.
- La Calificación de la ocupación interviene en algunos de los modelos indicando que disminuye la probabilidad de ser desempleado y en otros modelos que la aumenta. Esto representa la dualidad que se presenta en los últimos tiempos, cuando, por ejemplo, una persona con título profesional de posgrado se postula para ingresar a una empresa, y re-

cibe la respuesta de que es una calificación muy alta, que la empresa no podría afrontar el costo de sus servicios.

- e) La Rama de actividad interviene en las ecuaciones de regresión de todos los años a partir de 1998, siendo la modalidad de Comercio, Restaurantes y Hoteles, la que presenta un comportamiento estable disminuyendo la probabilidad de los individuos de ser desocupado, mientras que las otras ramas aparecen en algunos años como que la aumentan y en otros la disminuyen.
- f) La Categoría ocupacional cobra importancia como variable explicativa del desempleo a partir del año 2000. Pero aparece con la característica de que, en general, ser trabajador por cuenta propia o asalariado también disminuye la probabilidad de estar desempleado, con respecto a ser patrón o empleador.

Como conclusión global puede decirse que no es posible formular un modelo de regresión logística único que permita estimar los parámetros para todos los años y realizar predicciones, esto ya se descartó a la hora de plantear el objetivo general de este trabajo, pero creo conveniente recomendar que se deberían tener en cuenta todas las variables independientes analizadas, para diseñar políticas económicas, sociales, demográficas, educativas y empresariales, tendientes a disminuir los niveles de desempleo de la población de la provincia de Jujuy.

BIBLIOGRAFÍA

Agresti, A. (1990), "Categorical Data Analysis". John Wiley & Sons.

Carpio, Jorge, Orsatti, Alvaro, Sobrón, Claudia y López, Néstor, "Precariedad laboral en el conurbano bonaerense. Resultados de un estudio sobre pobreza familiar". En Galín, Pedro y Novick, Marta (comps.), "La precarización del empleo en la Argentina", CEAL / CIAT / CLACSO, Buenos Aires, 1990.

Cortés F. (1996). "La regresión logística en la investigación social: potencialidades y limitaciones". www.rau.edu.uy/

García Borges Demétrio, C., (1993), "Modelos Lineares Generalizados na Experimentação Agronômica", Universidade Federal do Rio Grande do Sul, Brasil.

Hosmer, D.W. Jr., Lemeshow, S. (1990). "Applied Logistic Regression". John Wiley & Sons.

INDEC. Información de prensa. Octubre 2000, mayo y octubre de 2002. ISSN 0327-7968. Bs. As.

Jiang, J., (1998), "Consistent Estimators in Generalized Linear Mixed Models", Journal of the American Statistical Association, 93, 720-729. USA.

Johnson, R. A., Wichern, D. W., (1998), "Applied Multivariate Statistical Analysis", (Fourth Edition), Prentice Hall.

Langford, I. H., Lewis, T., (1998), "Outliers in multilevel data" (with comments), Journal of the Royal Statistical Society A, 161, Part 2, 121-160. U.K.

Lebart L., Morineau A., Piron M., (1995), "Statistique exploratoire multidimensionnelle", DUNOD. París. Francia.

Lindenboim, J., (1996), "Las condiciones del mercado de trabajo en los 90. Desocupación y precariedad", Trabajo presentado en el 3er. Congreso Nacional de Estudios del Trabajo. Bs. As. Setiembre de 1996.

Maguid, A. (1995). "Migrantes Limítrofes en la Argentina: su inserción e impacto en el mercado de trabajo". Estudios del Trabajo (aset). N° 10: 47-76. ISSN 0327-5744.

Marcoleri, M:E. (2002) "La Regresión Logística aplicada para modelar el Desempleo en Jujuy". Trabajo de Tesis aprobado para acceder al título de Posgrado de Magíster en Estadística Aplicada. Univ. Nac. de Córdoba. Argentina.

Marcoleri. M.E. et al. (2001) "Transformaciones Socio-Laborales en Tiempos de Convertibilidad. Empleo, Desempleo, Pobreza y Migraciones en Jujuy". ISBN: 950-721-162-4. EdiUnju. Jujuy. Argentina.

Marshall, A. (1996). "Protección del empleo en América Latina: las reformas de los años 1990 y sus efectos en el mercado de trabajo. Estudios del Trabajo (aset). N° 1: 3-29. Bs. As. Argentina.

McCullagh, P., Nelder, J.A. (1989). "Generalized Linear Model" (Second Edition). Chapman & Hall.

Molinero, L.M. (2001). "La regresión logística binaria". www.seh-lelha.org/

Mondino, G. et al., (1998), "Los mercados de trabajo regionales en la Argentina", Desarrollo Económico. Revista de Ciencias Sociales, Número especial, Vol 38, 247-265.

Monza, A. et al., (1995), "El libro blanco del empleo en la Argentina", Ministerio de Trabajo y Seguridad Social de la Nación, Bs. As.

Nebot M. et al. (2002). "Efecto Protector del Apoyo Social en la Mortalidad en Población Anciana: un estudio longitudinal". Revista Especial Salud Pública; 76: 673-682; N° 6 – Nov.-Dic. 2002. España.

Paz, J. (1994). "Mercados Urbanos de Trabajo en el Norte Argentino. La participación femenina en los 80". Ponencia presentada en el 2° Congreso Nacional de Estudios del Trabajo. Aset. Bs. As. Argentina.

OIT (2002) "Panorama Laboral 2002". www.oit.org.pe/

Ortega Calvo L. y Cayuela Domínguez A. (2002). Regresión Logística no condicionada y tamaño de muestra: Una revisión bibliográfica. Revista Especial Salud Pública; 76: 85-93; N° 2 – Marzo-abril 2002. España.

Pok, Cynthia, (1992). "Precariedad laboral : Personificaciones sociales en la frontera de la estructura del empleo", Documento de Trabajo N° 29, CEIL.

Salvia, A., (1993), "La adopción del retiro voluntario por trabajadores de una empresa pública minera (Una decisión ajustada a condiciones sociales de existencia)", Estudios del Trabajo, N° 6, 81-112.

Sana, M., Pantelides, E.A., (1999), "La pobreza entre los ancianos. Lo que dicen los datos a la luz de las limitaciones de la medición". Desarrollo Económico. Revista de Ciencias Sociales, N° 152, 38, 1005-1026.

Síntesis Regional (2002). Sitio <http://fapep.netsys3.com/>

Siqueira, A.L., Taylor, M. G., (1998), Journal of the American Statistical Association, 94, 240-246. USA.

ANEXO

Análisis de Regresión Logística aplicada a la base de datos de la PEA. Octubre de 2001

Block 7: Method = Forward Stepwise (Likelihood Ratio)

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1	18,787	4	,001
Block	18,787	4	,001
Model	134,723	18	,000

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	4,652	8	,794

Contingency Table for Hosmer and Lemeshow Test

		Situación laboral = Ocupado		Situación laboral = Desocupado		Total
		Observed	Expected	Observed	Expected	
Step 1	1	51	50,766	1	1,234	52
	2	49	48,686	3	3,314	52
	3	46	44,876	3	4,124	49
	4	43	44,896	7	5,104	50
	5	29	31,623	7	4,377	36
	6	48	45,794	5	7,206	53
	7	47	45,804	8	9,196	55
	8	41	42,108	12	10,892	53
	9	38	35,940	14	16,060	52
	10	17	18,505	46	44,495	63

Classification Tablé

Observed	Predicted						
	Selected Cases ^a				Unselected Cases ^b		
	Situación laboral		Percentage Correct	Situación laboral		Percentage Correct	
	Ocupado	Desocupado		Ocupado	Desocupado		
Step 1 Situación laboral Ocupado	286	123	69,9	263	157	62,6	
Desocupado	27	79	74,5	36	59	62,1	
Overall Percentage			70,9			62,5	

a. Selected cases 515 from the first 1030 cases (SAMPLE) EQ 0

b. Unselected cases 515 from the first 1030 cases (SAMPLE) NE 0

c. The cut value is ,160

Variables in the Equation

Step	Variable	B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I. for EXP(B)	
								Lower	Upper
1	EDAD5			10,714	4	,030			
	EDAD5(1)	-1,105	,728	2,302	1	,129	,331	,079	1,380
	EDAD5(2)	-1,789	,704	6,459	1	,011	,167	,042	,664
	EDAD5(3)	-2,012	,783	6,597	1	,010	,134	,029	,621
	EDAD5(4)	-1,013	,935	1,175	1	,278	,363	,058	2,268
	NIVEL3			3,353	2	,187			
	NIVEL3(1)	-,658	,430	2,344	1	,126	,518	,223	1,203
	NIVEL3(2)	-,255	,478	,285	1	,594	,775	,303	1,979
	POSICIÓN(1)	-,482	,287	2,807	1	,094	,618	,352	1,085
	CATEGORI			,251	4	,993			
	CATEGORI(1)	5,206	14,665	,126	1	,723	182,362	,000	5,5E+14
	CATEGORI(2)	5,263	14,665	,129	1	,720	193,047	,000	5,9E+14
	CATEGORI(3)	-1,234	37,423	,001	1	,974	,291	,000	2,1E+31
	CATEGORI(4)	20,831	63,647	,107	1	,743	1,1E+09	,000	1,67E+63
	CALIF			8,528	3	,036			
	CALIF(1)	1,146	1,063	1,161	1	,281	3,145	,391	25,279
	CALIF(2)	1,938	1,093	3,146	1	,076	6,946	,816	59,135
	CALIF(3)	-4,555	60,442	,006	1	,940	,011	,000	2,95E+49
	RAMA5			18,279	4	,001			
	RAMA5(1)	1,491	,600	6,168	1	,013	4,442	1,369	14,412
	RAMA5(2)	,017	,614	,001	1	,977	1,017	,305	3,390
	RAMA5(3)	,170	,543	,098	1	,754	1,186	,409	3,435
	RAMA5(4)	1,726	,849	4,134	1	,042	5,618	1,064	29,659
	Constant	-6,402	14,716	,189	1	,664	,002		

a. Variable(s) entered on step 1: RAMA5.