

# Factores que influyen en el empleo en Cochabamba - Año 2002

Jesús Rojas Vargas

Centro de Estadística Aplicada, Universidad Mayor de San Simón (UMSS)  
Departamento de Ciencias Exactas e Ingenierías, Universidad Católica Boliviana  
e-mail: rojasje@ucbcba.edu.bo

## Resumen

El año 2002, el Instituto Nacional de Estadística realizó la cuarta versión de la Encuesta de Hogares – MECOVI, la encuesta trata de presentar un panorama completo sobre las condiciones de vida de la población. En este sentido el presente trabajo constituye un diagnóstico de la situación del empleo y los factores que los explican en Cochabamba. Se analiza las características propias del empleo tratando de determinar la relación entre el hecho de trabajar o no y las características sociales del individuo, para lo que utilizamos un Modelo Probit y el método de selección de variables *Stepwise*. Finalmente, interpretamos los resultados, tanto desde un nivel netamente estadístico, como desde un nivel económico - social.

Palabras clave: MECOVI, Cochabamba, Empleo, Probit, *Stepwise*.

## 1 Antecedentes

Desde el año 1999, por iniciativa del Banco Mundial, se estableció el Programa de mejoramiento de las encuestas “Medición de las Condiciones de Vida (MECOVI)” con el objetivo de compilar la información sobre las condiciones de vida de la población boliviana para la generación de indicadores de pobreza y la consecuente formulación de políticas y programas que contribuyan a mejorar las condiciones de bienestar de los hogares. Como parte de este programa, se efectúa la Encuesta Continua de Hogares – MECOVI a partir del llenado de un cuestionario multitemático.

El año 2002, el Instituto Nacional de Estadística - INE realizó la cuarta versión de la Encuesta de Hogares – Programa MECOVI durante los meses de noviembre y diciembre. Los contenidos tradicionales de la encuesta han sido mantenidos: Información general de los miembros del hogar, la migración, la salud, la educación, el empleo, los ingresos, los gastos de consumos, la vivienda, etc. La encuesta trata de presentar un panorama completo sobre las condiciones de vida de la población, como un aporte para el estudio de la realidad nacional y una guía para las instancias de toma de

decisiones en materia de política y de reducción de la pobreza, para los organismos internacionales, para la comunidad académica, así como para los investigadores privados y los estudiantes.

El presente trabajo constituye un diagnóstico de la situación del empleo y los factores que los explican en el caso del departamento de Cochabamba, sobre la base de la Encuesta de Hogares – MECOVI 2002. Se analiza las características propias del empleo tratando de determinar la relación entre el hecho de trabajar o no y las características sociales del individuo.

La importancia de analizar la problemática del empleo en un país en vías de desarrollo como Bolivia, justifica todos los esfuerzos que buscan explicar nuestra realidad en general y la situación del empleo para Cochabamba en particular.

## 2 Objetivo

Intentamos responder a la pregunta: ¿Qué factores determinan el hecho de trabajar o no?

Por lo tanto nuestros objetivos serán:

- Identificar los factores estadísticamente significativos, que influyen en el empleo en el departamento de Cochabamba, para el año 2002
- Interpretar los resultados obtenidos, desde un punto de vista económico y social

## 3 Métodos y materiales

Se utilizó un **Modelo Probit** [2], que nos permitió estudiar la relación entre una variable respuesta binaria y un conjunto de variables explicativas, en nuestro caso la variable a explicar es “Trabajo”. En general, el resultado de una observación binaria es llamado “éxito” o “fracaso”; se representa matemáticamente por una variable aleatoria  $Y$  tal que  $Y = 1$  si tenemos un éxito y  $Y = 0$  si tenemos un fracaso. El resultado  $Y$  puede depender de los valores asumidos por  $p$  variables explicativas, al momento de la observación y nosotros deseamos estudiar esta relación.

El modelo:

$$\Phi^{-1}(\pi) = \beta_1 + \beta_2 x$$

$$\pi = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-\frac{1}{2}\left(\frac{s-\mu}{\sigma}\right)^2\right] ds = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Donde:

- $\pi = \Phi\left(\frac{x-\mu}{\sigma}\right)$  es la distribución normal acumulada en el punto  $x$  (centrado y reducido).
- $\beta_i$  son los parámetros de nuestro modelo.
- $\Phi^{-1}(\pi)$  es la función inversa de la distribución normal acumulada.

La regresión *Probit* emplea la distribución normal acumulada, esta la razón principal que nos llevó a escogerlo, puesto que nos permitió realizar comparaciones, en trabajos posteriores, con modelos de regresión múltiple, los que se construyen bajo el supuesto de la normalidad de los errores.

Por otro lado, utilizamos el método de selección de variables *Stepwise* [4] o *Forward* y *Backward*. Este método es una modificación del método de selección *Forward*, en el que todas las variables no necesariamente están en el modelo, es decir que se añaden y quitan variables de manera de obtener la mejor combinación.

### 3.1 Descripción de la Base de Datos

Trabajamos con la sección “Individuos” de la Base de Datos “MECOVI - 2002” [5], contamos con 15 variables:

Las variables son:

- **sexo:** variable binaria que otorga el valor 1 para los hombres y 2 para las mujeres
- **edad:** variable numérica para la edad
- **ecivil:** variable categórica que representa el estado civil del individuo
- **etnia:** variable categórica que indica el grupo étnico del individuo (Aymara, Quechua, Guarani, Blanco, etc.)
- **menor:** variable binaria que da el valor 1 para los individuos menores de 7 años y 2 a los mayores
- **trabajo:** variable binaria que señala el valor 1 a los individuos que trabajan y 0 a los individuos que no trabajan (independientemente de si perciben o no un salario)
- **desdepto:** variable cualitativa que representa los departamentos del país
- **depto:** variable categórica que representa los departamentos (códigos)
- **urbrur:** variable binaria que indica el valor 1 a los individuos que habitan los sectores urbanos y 2 a los que habitan sectores rurales
- **a\_oesc:** variable cuantitativa que representa los años de escolaridad del individuo

- **estud:** variable binaria que otorga el valor 1 si el individuo estudia y 2 si no lo hace
- **nocup:** variable cuantitativa que indica la media del número de personas ocupadas en la familia del individuo
- **z:** variable cuantitativa que representa la línea de pobreza del individuo
- **ocupa:** variable categórica que muestra la ocupación laboral del individuo

**Tabla 1.** Descripción de la variable ocupa

Valor	Categoría
2	Empleado(a) <sup>1</sup>
3	Trabajador por cuenta propia
4	Patrón, socio o empleador que recibe remuneración
5	Patrón, socio o empleador que no recibe remuneración
7	Trabajador, pariente o aprendiz sin remuneración
8	Empleado(a) de la casa (doméstico)
10	Trabajador asalariado

- **relación:** variable categórica que señala la ubicación del individuo en relación a su lugar en la familia.

**Tabla 2.** Descripción de la variable relacion

Valor	Categoría
1	Jefe del hogar
2	Cónyuge
3	Hijo (a)
4	Yerno (a)
5	Nieto(a)
6	Hermano(a) – cuñado(a)
7	Padres o suegros
8	Otro pariente
9	empleado(a)
10	Pariente del empleado(a)
11	Otro no pariente

<sup>1</sup> La categoría **Empleado(a)** hace referencia a cualquier trabajador dependiente que percibe un salario.

### 3.2 Limpieza de la base de datos

Nos interesamos en los individuos del departamento de Cochabamba, entonces se ha seleccionado todas las observaciones que le pertenecen. En consecuencia la base de datos cuenta con 13 variables, todas las anteriores menos “desdepto” y “depto” y 3918 observaciones.

Se recodificó las variables para transformar los valores (1 y 2) en 1 y 0 y las variables “ecivil” y “etnia”, que son categóricas, se las transformó en binarias.

- De “ecivil” tenemos “**casado**” que toma el valor 1 si el individuo está casado y 0 si no lo está.
- De “etnia” tenemos “**indígena**” que toma el valor 1 si el individuo pertenece a los pueblos originarios y 0 si no lo hace.

Así podemos incluirlas en el modelo y también para facilitar su interpretación.

De la misma forma, se produjo variables binarias para representar cada una de las categorías de las variables “ocupa” y “relación”, esto para poder utilizarlas en nuestra regresión Probit.

- ocupa: ocupa\_2, ocupa\_3, ocupa\_4,.....
- relacion: relacion\_1, relacion\_2, relacion\_3,.....

## 4 El modelo

Para responder a nuestro objetivo, utilizaremos el *Procedimiento Logistic* [1] del programa informático SAS v. 9.1, con las consideraciones siguientes:

- **Modelo:** Comenzamos y fijamos las variables (*sexo, edad, casado, indígena, urbrur, a\_oesc, relacion\_1, estud*) que consideramos, explican el hecho de trabajar o no y aumentamos las otras variables, de manera de encontrar el mejor modelo.
- **Selección de variables:** Hemos escogido el método de selección de variables *stepwise* con un error de primera especie de 0.05 para la entrada y la salida, con la consideración de fijar las 8 variables que deseamos conservar.
- **Calidad del modelo:** Para medir la calidad de nuestro modelo en cada etapa del proceso *stepwise*, utilizaremos la prueba de bondad de ajuste de *Hosmer and Lemeshow* [2].

## 5 Análisis de resultados

### 5.1 Nivel estadístico

Con las 8 variables que fijamos, la etapa 0 del proceso *stepwise* cumple el criterio de convergencia, además del criterio AIC (Criterio de información de Akaike), es decir, tenemos un mejor modelo añadiendo las variables iniciales.

**Tabla 3.** Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
<b>AIC</b>	1132.847	<b>1066.002</b>

El modelo inicial es globalmente válido.

**Tabla 4.** Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
<b>Likelihood Ratio</b>	82.8455	8	<.0001
<b>Store</b>	79.2576	8	<.0001
<b>Wald</b>	73.9580	8	<.0001

Revisando, la última etapa del proceso *stepwise*, el criterio AIC ha mejorado mucho más, mostrando la participación de las nuevas variables:

**Tabla 5.** Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
<b>AIC</b>	1132.847	<b>1003.130</b>

Una vez más nuestro modelo es válido:

**Tabla 6.** Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
<b>Likelihood Ratio</b>	153.7175	12	<.0001
<b>Store</b>	158.3016	12	<.0001
<b>Wald</b>	129.5665	12	<.0001

En el proceso, hemos añadido 4 variables:

- *ocupa\_7*: que corresponde a la categoría *trabajador pariente o aprendiz sin remuneración*.
- *ocupa\_3*: *trabajador por cuenta propia*.
- *relacion\_7*: que corresponde a los *padres o suegros* que no son jefes de familia.
- *ocupa\_5*: *Patrón, socio o empleador* que no recibe remuneración.

**Tabla 7.** Summary of Stepwise Selection

Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
	Entered	Removed						
1	<i>ocupa_7</i>		1	9	47.7990		<.0001	
2	<i>ocupa_3</i>		1	10	14.6534		0.0001	
3	<i>relacion_7</i>		1	11	6.6189		0.0101	
4	<i>ocupa_5</i>		1	12	4.7279		0.0297	

La calidad de ajuste de nuestro modelo se muestra con la prueba de *Hosmer and Lemeshow* (un p-valor pequeño, menor al nivel de significancia 0.05, sugeriría que el modelo no es adecuado), en nuestro caso tenemos todo lo contrario:

**Tabla 8.** Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
7.2268	8	0.5124

Además de mostrar un nivel de concordancia, de aproximadamente el 80%.

### 5.2 Nivel económico

Remarquemos que las variables que fueron añadidas muestran la existencia de trabajo, sin que exista remuneración (salario), este aspecto es importante por que es un indicador de la calidad del empleo.

Pasemos a la interpretación de cada uno de los factores:

**Tabla 9.** Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard	Wald	Pr > ChiSq
			Error	Chi-Square	
intercept	1	2.1206	0.2325	83.2190	<.0001
sexo	1	0.2085	0.1043	3.9993	0.0455
edad	1	-0.00149	0.00390	0.1461	0.7023
casado	1	0.0998	0.1070	0.8700	0.3510
indígena	1	0.3167	0.1130	7.8551	0.0051
urbrur	1	-0.4749	0.1156	16.8770	<.0001
a_oesc	1	-0.0191	0.0113	2.8508	0.0913
relacion_1	1	0.1484	0.1396	1.1309	0.2876
estud	1	-0.2219	0.1211	3.3551	0.0670
relacion_7	1	-0.8529	0.3530	5.8367	0.0157
ocupa_3	1	-0.6007	0.1443	17.3231	<.0001
ocupa_5	1	-0.4616	0.2262	4.1635	0.0413
ocupa_7	1	-1.1443	0.1510	57.3938	<.0001

- **sexo:** Diríamos que los hombres tienen una probabilidad de trabajar más alta que las mujeres. Pero este resultado es débilmente significativo ( $\alpha=0,05$ ), con lo que podemos decir que en el departamento de Cochabamba trabajan hombres y mujeres indistintamente.
- **edad:** La probabilidad de trabajar disminuiría cuando la edad aumenta, pero este resultado no es significativo, entonces la edad no tiene relación con el hecho de trabajar o no. Lo que podemos constatar fácilmente, puesto que tenemos un gran porcentaje de nuestra niñez trabajando y no existe la jubilación para la mayoría de nuestros ancianos.
- **casado:** La estimación indicaría que es más probable que una persona casada trabaje, pero no es un resultado significativo, por lo que podemos decir que en Cochabamba el matrimonio no es relevante en el hecho de trabajar o no.
- **indígena:** La probabilidad de trabajar de una persona indígena es más alta que la de una persona que no es indígena, lo que se puede explicar por el hecho que normalmente los indígenas tienen los trabajos menos remunerados.
- **urbrur:** En relación con la interpretación precedente, podemos ver que es menos probable que una persona que habita en un sector urbano tenga trabajo, lo que no sucede en las áreas rurales, en el campo todos trabajan. Una vez más debemos



aclarar que nos referimos al hecho de trabajar o no, independientemente de si se percibe salario o no.

- **a\_oesc:** Se podría decir que es menos probable trabajar cuando se tiene más años de escolaridad, pero este es un resultado no significativo. Es muy importante tomar en cuenta este resultado porque muestra una de las problemáticas del empleo en el departamento de Cochabamba.
- **relacion\_1:** El hecho de ser jefe de familia aumentaría la probabilidad de trabajar, pero es un resultado no significativo. Este resultado muestra, también, la difícil situación económica de nuestro país, no es suficiente que el jefe de familia trabaje para poder sostener a la familia, es una tarea que corresponde a todos.
- **estud:** La probabilidad de trabajar cuando un individuo está estudiando disminuiría pero es un resultado no significativo, una vez más podemos ver el reflejo de nuestra situación económica, muchos estudiantes tienen que trabajar.
- **relacion\_7:** La probabilidad de trabajar de los padres o suegros disminuye, en relación a los otros miembros de la familia.
- **ocupa\_3, ocupa\_5 e ocupa\_7:** Son ocupaciones (Trabajador por cuenta propia, Patron – socio o empleador sin remuneración y Trabajador pariente o aprendiz sin remuneración), que no perciben salario, por lo que podemos decir que la probabilidad de buscar un salario disminuye la probabilidad de trabajar, una vez más reflejamos la situación económica de nuestro país.

Tratemos de cuantificar los resultados a través de los *efectos marginales* [8] en un individuo particular (concreto).

Consideremos una persona con las siguientes características: (sexo=1) un hombre, (edad=36) con 36 años, (casado=1) que es casado, (indigena=1) indígena, (urbrur=1) que vive en la ciudad, (A\_oesc=17) que tiene 17 años de escolaridad, (relacion\_1=1) jefe de familia (relacion\_7=0), (etude=0) que no estudia, (ocupa\_3=0) no es Trabajador por cuenta propia, ni Patron socio o empleador sin remuneración (ocupa\_5=0), ni Trabajador pariente o aprendiz sin remuneración (ocupa\_7=0).

Recordemos que las variables Edad, Casado, A\_oesc, Relacion\_1 y Estud no son significativas, por lo tanto no consideraremos los efectos marginales en estas variables.

- *Meffsexo:* Si el individuo del ejemplo pasa de ser hombre a ser mujer su probabilidad de trabajar disminuye en 0,0094
- *Meffindigena:* Si el individuo del ejemplo pasa de ser indígena a no serlo su probabilidad de trabajar disminuye en 0,011568
- *Meffurbrur:* Si el sujeto del ejemplo pasa de vivir en la ciudad a vivir en el campo su probabilidad de trabajar aumenta en 0,02061
- *Meffrelacion\_7:* Si la persona del ejemplo pasa de ser jefe de familia a ser padre o suegro en el hogar, su probabilidad de trabajar disminuye en 0,03472

- *Meffocupa\_3*: Si el individuo del ejemplo pasa a tener la ocupación de Trabajador por cuenta propia su probabilidad de trabajar aumenta en 0,0279
- *Meffocupa\_5*: Si el sujeto del ejemplo pasa a tener la ocupación de Patrón socio o empleador sin remuneración, su probabilidad de trabajar aumenta en 0,01884
- *Meffocupa\_7*: Si el individuo del ejemplo pasa a tener la ocupación de Trabajador pariente o aprendiz sin remuneración, su probabilidad de trabajar aumenta en 0,0503

## 6 Conclusiones

Hemos determinado un modelo Probit, para el departamento de Cochabamba, que explica el hecho de trabajar o dejar de hacerlo. Pero lo más importante, es la manera como reflejamos la realidad económica – social, a través del mismo modelo.

VARIABLES como la edad, los años de escolaridad, el estado civil y el ser jefe de familia que son centrales en un modelo de estas características, en otras regiones del mundo, en nuestro caso no son factores significativos. Por el contrario encontramos que variables como el ser indígena, vivir en el área rural o tener una ocupación no remunerada juegan un rol central en el modelo.

En otras palabras reflejamos la crisis en la que nos encontramos. Problemáticas como la informalidad, el trabajo infantil, la discriminación, el subempleo, y la falta de oportunidades para las personas que invierten más años de estudio en su formación, aparecen con mucha claridad y están planteadas para poder seguir con el análisis, tanto desde el punto de vista estadístico, económico y social.

Para terminar queremos destacar la consistencia y precisión de los resultados obtenidos, este es un ejemplo de que cuando los métodos estadísticos son empleados correctamente, obtenemos un fiel reflejo del fenómeno que buscamos estudiar.

## 7 Anexo

Líneas de código del programa Mecovi.sas, que generamos para obtener los principales resultados.

```
ods rtf
file='c:\bruselas\SASods\Probit.rtf';

ods graphics on;

title 'Factores que influyen en el
empleo MECOVI-2002 (CBBA-BOLIVIA)';

footnote 'JRV';

estud relacion_2 relacion_3 relacion_4
relacion_5 relacion_6 relacion_7
relacion_8 relacion_9 relacion_10
ocupa_8 ocupa_2 ocupa_3 ocupa_4 ocupa_5
ocupa_7 /selection=stepwise include=8
lackfit rsq l=probit;

run;

/*Selección de variables*/

proc logistic data=Mecovi.cbba7;

model trabajo(event='1') = sexo edad
casado indigena urbrur a_oesc relacion_1
```

```

/*Estimación de los parámetros*/
proc logistic data=Mecovi.cbba7;
  model trabajo(event='1') = sexo edad
  casado indigena urbrur a_oesc relacion_1
  estud relacion_7 ocupa_3 ocupa_5 ocupa_7
  / lackfit rsq l=probit;
run;
ods graphics off;
ods rtf close;

/*Computo de los Efectos Marginales*/
proc logistic data=Mecovi.cbba7
  outmodel=om;
  model trabajo(event='1') = sexo edad
  casado indigena urbrur a_oesc relacion_1
  estud relacion_7 ocupa_3 ocupa_5
  ocupa_7;
  ods output parameterestimates=logparms;
  output out=outlog p=p;
run;

proc transpose data=logparms out=tlog
  (rename=(ocupa_7=tocupa_7 urbrur=turbrur
  relacion_7=trelacion_7
  ocupa_3=tocupa_3 indigena=tindigena
  relacion_1=trelacion_1 sexo=tsexo
  edad=tedad casado=tcasado a_oesc=ta_oesc
  ocupa_5=tocupa_5 etude=tetude));
  var estimate;
  id variable;
run;

data outlog;
  if _n_=1 then set tlog;
  set outlog;
  Meffsexo=p*(1-p)*tsexo;
  Meffedad=p*(1-p)*tedad;
  Meffcasado=p*(1-p)*tcasado;
  Meffindigena=p*(1-p)*tindigena;
  Meffurbrur = p*(1-p)*turbrur;
  Meffa_oesc=p*(1-p)*ta_oesc;
  Meffrelacion_1=p*(1-p)*trelacion_1;
  Meffestud=p*(1-p)*testud;
  Meffrelacion_7=p*(1-p)*trelacion_7;
  Meffocupa_3=p*(1-p)*tocupa_3;
  Meffocupa_5=p*(1-p)*tocupa_5;
  Meffocupa_7 = p*(1-p)*tocupa_7;
run;

proc print data=outlog (obs=15);
  var sexo edad casado indigena urbrur
  a_oesc relacion_1 estud relacion_7
  ocupa_3 ocupa_5
  ocupa_7 Meffsexo Meffedad Meffcasado
  Meffindigena Meffurbrur Meffa_oesc
  Meffrelacion_1 Meffestud Meffrelacion_7
  Meffocupa_3 Meffocupa_5 Meffocupa_7;
run;

```

## Referencias

- [1] Customer Support Center. *http://www.support.sas.com* (Verificado Abril–Mayo, 2005)
- [2] Dobson J. Annette. *An Introduction to Generalized Linear Models*. Chapman and Hall. 2001
- [3] Everitt Brain and DER Geoff. *A Handbook of Statistical Analysis using SAS*. Chapman and Hall, 1996
- [4] Greene W. *Econometric Analysis*. Prentice Hall. 2003.
- [5] Instituto Nacional de Estadística – Bolivia. *http://www.ine.gov.bo* (verificado Abril–Mayo, 2005)
- [6] Ludovic L. *Statistique Exploratoire Multidimensionnelle*. Dunod. 2000
- [7] Unidad de Análisis de Políticas Sociales y Económicas. *http://www.udape.gov.bo* (verificado Abril–mayo, 2005)
- [8] Verardi V. *Econometrics Course*. ECARES. 2004