

Un estudio de la migración utilizando análisis multivariado, en áreas urbanas del Noroeste Argentino (NOA). Período 2000- 2002.

Gladys del Valle Rosales.

Cita:

Gladys del Valle Rosales (2007). *Un estudio de la migración utilizando análisis multivariado, en áreas urbanas del Noroeste Argentino (NOA). Período 2000- 2002. IX Jornadas Argentinas de Estudios de Población. Asociación de Estudios de Población de la Argentina, Huerta Grande, Córdoba.*

Dirección estable: <https://www.aacademica.org/000-028/69>

ARK: <https://n2t.net/ark:/13683/eQa4/ctY>

**Un estudio de la migración utilizando análisis multivariado, en áreas urbanas del
Noroeste Argentino (NOA). Período 2000- 2002**

Dra. Rosales, Gladys del Valle

Instituto de Investigaciones Estadísticas y Demográficas
Facultad de Ciencias Económicas y de Administración. Universidad Nacional de Catamarca
(UNCa.)- gladysrosales@arnet.com.ar

Resumen

En el período 2000- 2002, nuestro país atravesó una de sus mayores crisis política y socio- económica; lo que repercutió negativamente en el mercado laboral.

Las áreas urbanas del NOA fueron unas de las más afectadas: con elevados niveles de desocupación, subocupación y precariedad laboral; y con importantes caídas del salario real y del PBI.

Las migraciones están muy ligadas a la estructura de la participación laboral y los recursos disponibles.

En esta investigación se modeliza aplicando la técnica estadística multivariada Modelo de Regresión Logística. Se considera como variable dependiente la condición de migración (migrante- no migrante) y como variables explicativas las más relevantes (según estudio previo) referidas al mercado de trabajo.

Se trabaja con las Base Usuario de la Encuesta Permanente de Hogares (EPH) y con el software estadístico SPSS.

Introducción

De estudios realizados previamente surge que existen diferencias significativas entre migrantes y no migrantes ($p_{60r}=1$, $p_{60r}=2$) de las áreas urbanas del NOA, según las variables: grupo de edad (h12), condición de actividad (estador), categoría ocupacional (p17r), nivel de instrucción (ninst), ingreso total (p47t), descuento jubilatorio (jubi) ; mientras que no existen diferencias significativas según rama de actividad (ramar1) y sexo (h13) de la población ocupada (salvo algunas excepciones).

Por ello se modeliza utilizando regresión logística. Se considera como variable dependiente la condición de migración, que es una variable dicotómica (1=Migrante, 2=No migrante) y como variables independientes a las mencionadas en el párrafo anterior.

En realidad se hacen dos análisis de regresión logística: en uno de ellos se considera la población total y en el otro la población ocupada. Finalmente se elige el segundo por dar mayor porcentaje de observaciones clasificadas correctamente por el modelo y porque permite incluir variables que solo corresponden a la población ocupada (categoría ocupacional, rama de actividad, ingreso individual total). La variable descuento jubilatorio no se incluye porque se refiere solo a asalariados.

Cabe recordar que la EPH define como migrante a toda persona que residió fuera del área por más de 6 meses. Vale decir que incluye a migrantes de retorno.

Metodología

La regresión logística es una técnica de análisis multivariante en la que la variable dependiente es dicotómica y la variable o variables independientes pueden ser cuantitativas o cualitativas. Si la variable independiente es cualitativa con k categorías se generan $k-1$ variables dummy, para que todas las posibilidades de la variable queden representadas en el modelo

Con modelos de regresión logística se puede investigar factores causales de una característica de la población y también estudiar qué factores modifican la probabilidad de un suceso

Para construir un modelo matemático se necesitan valores numéricos, los cuales se pueden obtener considerando la probabilidad de que se de una característica o valor determinado de la variable dependiente.

Se trata de conocer los factores que aumentan o disminuyen la probabilidad de que ocurra un suceso determinado. En nuestro caso, ser migrante .

El modelo de regresión logística es el siguiente:

$$\text{Prob.}(\text{migrante})=1/(1+e^{-z})$$

$$\text{Donde } Z= b_0 +b_1 x +b_2 x^2 +\dots +b_n x^n$$

e es la base de los logaritmos neperianos (e= 2,71828)

x_i son las variables independientes

Los b_i son los coeficientes del modelo; b_0 es el término independiente y puede ser igual a cero. Los demás coeficiente son los coeficientes de regresión logística y al menos uno de ellos debe ser significativamente distinto de cero. Si el coeficiente de una variable independiente es significativamente distinto de cero esto indica que existe relación según el modelo logístico entre esa variable independiente y la variable dependiente.

La estimación de los coeficientes logísticos se hizo utilizando el principio de máxima verosimilitud, o sea eligiendo como estimadores de los coeficientes b_i los valores que maximizan la función de verosimilitud. Esta es una técnica de cálculo diferencial en la que se emplean cálculos iterativos hasta que la diferencia con el valor de la función es menor que un valor predeterminado, generalmente 0,01. El número de iteraciones es fijo y también ajustable por el experimentador. Si la función no converge en el número de iteraciones predeterminado se dice que no tiene solución. Esto se hace con ordenadores y en este caso se utilizó el software estadístico SPSS.

En el caso de las variables independientes cualitativas con más de dos categorías se generan tantas variables dummy como el número de categorías menos 1. Esto permite comparar la importancia relativa – para discriminar las categorías de la variable dependiente- de las demás categorías de la variable con una de ellas elegida como de referencia

El signo de los coeficientes tiene un significado importante. Si el coeficiente de cada variable es positivo, significa que la variable aumenta la probabilidad del suceso. Si el coeficiente es negativo, el factor cuyo coeficiente es negativo disminuye la probabilidad del suceso (ser no migrante, de acuerdo a la codificación considerada, ya que en lugar de 0:No Migrante, 1: Migrante se tomó: 1: Migrante y 2: No Migrante).

La hipótesis nula del modelo establece que todos los coeficientes de regresión logística son nulos y la hipótesis alternativa, que es la negación de la nula, que al menos uno de ellos es distinto de cero. Se utilizan diversos estadísticos para contrastar estas hipótesis.

En este trabajo se usó el método backward stepwise (WALD), el cual selecciona variables con pruebas de entrada y salida basadas en la significación del estadístico Wald, para probar la hipótesis nula de que el coeficiente para cada variable es nula

Resultados obtenidos

Las variables utilizadas, categorías y categorías de referencia de cada variable independiente cualitativa son:

Variable dependiente: **Condición de migración (P60R)**: 1: Migrante- 2: No Migrante

Variables independientes:

Ingreso Total (P47T), es continua

Categoría Ocupacional (P17R): P17R (1): Patrón o Socio- P17R (2): Cuenta Propia- P17R (3): Asalariado- P17R(4) Trabajador sin Salario Fijo (Categoría de Referencia)

Nivel de Instrucción (NINST): NINST (1): Primario- NINST (2): Secundario- NINST (3): Terciario- NINST (4): Universitario (Categoría de Referencia)

Edad (H12), es continua

Sexo (H13): H13 (1): Varón- H13 (2): Mujer (Categoría de Referencia)

Rama de Actividad (RAMAR1): RAMAR1 (1): Industria - RAMAR1 (2): Construcción- RAMAR1 (3): Comercio- RAMAR1 (4): Servicios- RAMAR1 (5): Otra (Categoría de Referencia)

A continuación se presentan los resultados obtenidos para el total área urbana NOA y para cada una de las áreas de la región

Regresión logística. Región Urbana NOA

Dependent Variable..	P60R	P60R	
	Chi-Square	df	Significance
Model	977,357	13	,0000
Block	977,357	13	,0000
Step	977,357	13	,0000

Classification Table for P60R
The Cut Value is ,50

Observed		Predicted		Percent Correct
		1,00	2,00	
1,00	1	I 4062 I	3983 I	50,49%
2,00	2	I 2776 I	6077 I	68,64%
Overall				60,00%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
P47T	-,0003	4,414E-05	40,0344	1	,0000	-,0403	,9997
P17R			19,6762	3	,0002	,0242	
P17R(1)	,0370	,1665	,0494	1	,8241	,0000	1,0377
P17R(2)	-,0268	,1412	,0360	1	,8495	,0000	,9736
P17R(3)	,1518	,1386	1,1984	1	,2736	,0000	1,1639
NINST			142,6992	3	,0000	,0765	
NINST(1)	,3120	,0541	33,3025	1	,0000	,0366	1,3662
NINST(2)	,5605	,0493	129,1417	1	,0000	,0737	1,7515
NINST(3)	,3229	,0659	24,0075	1	,0000	,0307	1,3812
H12	-,0295	,0014	435,7423	1	,0000	-,1362	,9709
H13(1)	,0545	,0349	2,4436	1	,1180	,0044	1,0560
RAMAR1			24,1059	4	,0001	,0262	
RAMAR1(1)	,0163	,1400	,0136	1	,9073	,0000	1,0164
RAMAR1(2)	-,0756	,1411	,2866	1	,5924	,0000	,9272
RAMAR1(3)	,1961	,1361	2,0779	1	,1494	,0018	1,2167
RAMAR1(4)	,1706	,1324	1,6610	1	,1975	,0000	1,1860
Constant	,6781	,1962	11,9505	1	,0005		

Regresión logística. Santiago del Estero- La Banda

Dependent Variable..	P60R	P60R	
	Chi-Square	df	Significance
Model	111,922	13	,0000
Block	111,922	13	,0000
Step	111,922	13	,0000

Classification Table for P60R
The Cut Value is ,50

Observed		Predicted		Percent Correct
		1,00	2,00	
		1	2	
1,00	1	61	675	8,29%
2,00	2	34	1726	98,07%
Overall				71,59%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
P47T	-,0003	,0001	4,1402	1	,0419	-,0266	,9997
P17R			6,7699	3	,0796	,0159	
P17R(1)	-,0753	,5187	,0210	1	,8847	,0000	,9275
P17R(2)	,0134	,4623	,0008	1	,9768	,0000	1,0135
P17R(3)	,2768	,4580	,3652	1	,5456	,0000	1,3189
NINST			35,6946	3	,0000	,0990	
NINST(1)	,5400	,1528	12,4883	1	,0004	,0589	1,7160
NINST(2)	,8409	,1436	34,2911	1	,0000	,1033	2,3185
NINST(3)	,5247	,1889	7,7183	1	,0055	,0435	1,6899
H12	-,0210	,0038	30,2284	1	,0000	-,0966	,9793
H13(1)	-,0667	,0983	,4601	1	,4976	,0000	,9355
RAMAR1			1,2052	4	,8772	,0000	
RAMAR1(1)	-,0051	,4616	,0001	1	,9911	,0000	,9949
RAMAR1(2)	,0763	,4636	,0271	1	,8693	,0000	1,0793
RAMAR1(3)	,1577	,4446	,1258	1	,7229	,0000	1,1708
RAMAR1(4)	,0365	,4382	,0069	1	,9337	,0000	1,0371
Constant	,9923	,6507	2,3255	1	,1273		

Regresión logística. Jujuy- Palpalá

Dependent Variable..	P60R	P60R
	Chi-Square	df Significance
Model	233,532	13 ,0000
Block	233,532	13 ,0000
Step	233,532	13 ,0000

Classification Table for P60R
The Cut Value is ,50

Observed		Predicted		Percent Correct
		1,00	2,00	
		1	2	
1,00	1	I 974	I 392	I 71,30%
2,00	2	I 526	I 603	I 53,41%
Overall				63,21%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
P47T	-,0005	,0001	13,3361	1	,0003	-,0574	,9995
P17R			3,3996	3	,3340	,0000	
P17R(1)	,1831	,5388	,1155	1	,7340	,0000	1,2009
P17R(2)	-,2818	,4655	,3665	1	,5449	,0000	,7544
P17R(3)	-,1426	,4616	,0954	1	,7574	,0000	,8671
NINST			47,8265	3	,0000	,1103	
NINST(1)	,4190	,1714	5,9737	1	,0145	,0340	1,5204
NINST(2)	,9029	,1533	34,6974	1	,0000	,0975	2,4667
NINST(3)	,6394	,1776	12,9610	1	,0003	,0565	1,8954
H12	-,0347	,0041	69,8581	1	,0000	-,1405	,9659
H13(1)	,0242	,0936	,0671	1	,7956	,0000	1,0245
RAMAR1			10,7670	4	,0293	,0284	
RAMAR1(1)	,3539	,3707	,9117	1	,3397	,0000	1,4247
RAMAR1(2)	-,1138	,3645	,0976	1	,7548	,0000	,8924
RAMAR1(3)	,2103	,3594	,3424	1	,5585	,0000	1,2340
RAMAR1(4)	,3960	,3429	1,3337	1	,2481	,0000	1,4858
Constant	,5364	,5863	,8368	1	,3603		

Regresión logística. Gran Catamarca

Dependent Variable..	P60R	P60R	
	Chi-Square	df	Significance
Model	217,099	13	,0000
Block	217,099	13	,0000
Step	217,099	13	,0000

Classification Table for P60R
The Cut Value is ,50

Observed		Predicted		Percent Correct
		1,00	2,00	
		1	2	
1,00	1	I 697 I	650 I	51,74%
2,00	2	I 430 I	1072 I	71,37%
Overall				62,09%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
P47T	-,0005	,0001	18,7153	1	,0000	-,0651	,9995
P17R			1,8301	3	,6084	,0000	
P17R(1)	,1491	,3965	,1414	1	,7069	,0000	1,1608
P17R(2)	-,0183	,3112	,0034	1	,9532	,0000	,9819
P17R(3)	,1225	,3019	,1647	1	,6848	,0000	1,1304
NINST			33,9456	3	,0000	,0842	
NINST(1)	,4875	,1355	12,9428	1	,0003	,0527	1,6282
NINST(2)	,6552	,1180	30,8118	1	,0000	,0855	1,9256
NINST(3)	,2064	,1537	1,8021	1	,1795	,0000	1,2292
H12	-,0313	,0038	69,1707	1	,0000	-,1306	,9692
H13(1)	-,1224	,0868	1,9899	1	,1583	,0000	,8848
RAMAR1			7,8633	4	,0967	,0000	
RAMAR1(1)	-,2361	,2789	,7168	1	,3972	,0000	,7897
RAMAR1(2)	-,3686	,2774	1,7665	1	,1838	,0000	,6917
RAMAR1(3)	,0473	,2689	,0310	1	,8602	,0000	1,0485
RAMAR1(4)	-,0745	,2541	,0860	1	,7694	,0000	,9282
Constant	1,0968	,4128	7,0594	1	,0079		

Regresión logística. Salta

Dependent Variable.. P60R P60R

	Chi-Square	df	Significance
Model	291,475	13	,0000
Block	291,475	13	,0000
Step	291,475	13	,0000

Classification Table for P60R
The Cut Value is ,50

Observed		Predicted		Percent Correct
		1,00	2,00	
		1	2	
1,00	1	I 1428 I	403 I	77,99%
2,00	2	I 720 I	565 I	43,97%
Overall				63,96%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
P47T	-,0002	,0001	4,8703	1	,0273	-,0261	,9998
P17R			22,9435	3	,0000	,0633	
P17R(1)	,2047	,3792	,2915	1	,5893	,0000	1,2272
P17R(2)	,2774	,3145	,7781	1	,3777	,0000	1,3197
P17R(3)	,6693	,3091	4,6879	1	,0304	,0252	1,9528
NINST			58,5315	3	,0000	,1115	
NINST(1)	-,2737	,1307	4,3861	1	,0362	-,0238	,7606
NINST(2)	,4140	,1103	14,0926	1	,0002	,0535	1,5128
NINST(3)	,4646	,1516	9,3969	1	,0022	,0418	1,5914
H12	-,0325	,0033	94,4037	1	,0000	-,1479	,9680
H13(1)	,0923	,0840	1,2082	1	,2717	,0000	1,0967
RAMAR1			2,6512	4	,6178	,0000	
RAMAR1(1)	-,4000	,4394	,8288	1	,3626	,0000	,6703
RAMAR1(2)	-,5164	,4406	1,3737	1	,2412	,0000	,5967
RAMAR1(3)	-,4437	,4325	1,0525	1	,3049	,0000	,6417
RAMAR1(4)	-,5252	,4271	1,5124	1	,2188	,0000	,5914
Constant	,7166	,5384	1,7714	1	,1832		

Regresión logística. La Rioja

Dependent Variable..	P60R	P60R	
	Chi-Square	df	Significance
Model	206,713	13	,0000
Block	206,713	13	,0000
Step	206,713	13	,0000

Classification Table for P60R
The Cut Value is ,50

Observed		Predicted		Percent Correct
		1,00	2,00	
		1	2	
1,00	1	I 1441 I	I 273 I	84,07%
2,00	2	I 748 I	I 371 I	33,15%
Overall				63,96%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
P47T	-,0004	,0001	10,4745	1	,0012	-,0472	,9996
P17R			5,4621	3	,1409	,0000	
P17R(1)	-,4932	,4427	1,2409	1	,2653	,0000	,6107
P17R(2)	-,7472	,3849	3,7680	1	,0522	-,0216	,4737
P17R(3)	-,5454	,3742	2,1246	1	,1450	-,0057	,5796
NINST			42,6516	3	,0000	,0982	
NINST(1)	,6102	,1462	17,4276	1	,0000	,0637	1,8408
NINST(2)	,8247	,1354	37,0934	1	,0000	,0961	2,2813
NINST(3)	,2582	,1898	1,8507	1	,1737	,0000	1,2946
H12	-,0333	,0038	75,4287	1	,0000	-,1390	,9672
H13(1)	,0236	,0884	,0710	1	,7899	,0000	1,0238
RAMAR1			18,5389	4	,0010	,0527	
RAMAR1(1)	,1320	,3968	,1107	1	,7394	,0000	1,1411
RAMAR1(2)	,4153	,4042	1,0558	1	,3042	,0000	1,5148
RAMAR1(3)	-,0019	,3961	,0000	1	,9961	,0000	,9981
RAMAR1(4)	,4574	,3862	1,4029	1	,2362	,0000	1,5800
Constant	,5299	,5537	,9158	1	,3386		

Regresión logística. Gran Tucumán- Tafí Viejo

Dependent Variable..	P60R	P60R	
	Chi-Square	df	Significance
Model	169,190	13	,0000
Block	169,190	13	,0000
Step	169,190	13	,0000

Classification Table for P60R
The Cut Value is ,50

Observed		Predicted		Percent Correct
		1,00	2,00	
1,00	1	I 147 I	904 I	13,99%
2,00	2	I 147 I	1911 I	92,86%
Overall				66,19%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
P47T	-7,4E-05	9,101E-05	,6539	1	,4187	,0000	,9999
P17R			1,7137	3	,6339	,0000	
P17R(1)	,1748	,3822	,2092	1	,6474	,0000	1,1911
P17R(2)	,1751	,3275	,2858	1	,5929	,0000	1,1913
P17R(3)	,2732	,3202	,7280	1	,3935	,0000	1,3142
NINST			14,4411	3	,0024	,0461	
NINST(1)	,1309	,1178	1,2353	1	,2664	,0000	1,1399
NINST(2)	,3872	,1154	11,2655	1	,0008	,0483	1,4729
NINST(3)	,3348	,1756	3,6338	1	,0566	,0203	1,3977
H12	-,0339	,0032	109,4761	1	,0000	-,1644	,9667
H13(1)	,1547	,0869	3,1659	1	,0752	,0171	1,1673
RAMAR1			4,8641	4	,3015	,0000	
RAMAR1(1)	,5836	,3160	3,4112	1	,0648	,0188	1,7925
RAMAR1(2)	,4637	,3255	2,0288	1	,1543	,0027	1,5899
RAMAR1(3)	,6180	,3009	4,2196	1	,0400	,0236	1,8553
RAMAR1(4)	,6001	,2935	4,1792	1	,0409	,0234	1,8222
Constant	,8743	,4341	4,0566	1	,0440		

Las tablas de predicciones del modelo logístico aplicado al total urbano NOA y a cada una de las áreas de la región muestran que en todos los casos el porcentaje de observaciones correctamente clasificadas por el modelo es superior al 60%

Además en las tablas de “variables en la ecuación” se indican: nombres de las variables, valor de los coeficientes (B) para las variables, los errores estándares de los coeficientes (S. E.), el estadístico de Wald, los grados de libertad, la significación de los coeficientes.

En el modelo múltiple los estadísticos del modelo pueden ser significativos ($\text{sig.} \leq 0,05$) y algún coeficiente puede no serlo.

También se muestran los valores del coeficiente de correlación parcial y el valor de Exp. (B) que es el OR, “riesgo/ efecto/ peso de cada categoría de la variable con respecto a otra (categoría de referencia)” en relación a ser no migrante; vale decir indica que la categoría correspondiente ofrece un efecto tantas veces mayor que la otra, a ser no migrante, como el valor de OR.

Conclusiones:

En los estudios previos realizados, utilizando comparaciones y análisis de independencia, test de χ^2 , se vio que existen diferencias significativas entre migrantes y no migrantes ocupados de las áreas urbanas del NOA con respecto a las variables: edad, categoría ocupacional, nivel de instrucción, ingresos, descuento jubilatorio; no así con respecto a rama de actividad y sexo.

Así resultó que en los migrantes son mayores- que en los no migrantes- las proporciones de población de 25 años y más (en edades de mayor actividad económica), de cuenta propia y patrón, de ocupados con instrucción primaria o universitaria, con ingresos más altos. Esto se verifica en el modelo aplicado.

También son mayores las proporciones de asalariados con descuentos jubilatorios y los niveles de actividad; y menores los niveles de desocupación.

El estudio de regresión logística- por las mayores exigencias del estadístico de Wald- muestra que además de las variables rama de actividad y sexo, la categoría ocupacional tampoco establece diferencias significativas entre migrantes y no migrantes.

El hecho de que los coeficientes (B) correspondientes a ingreso total y edad sean negativos se debe a una relación inversa entre cada una de estas variables y la condición de migración (acorde a los valores que asume). Vale decir que los migrantes (1) poseen mayor proporción de población con mayores ingresos y mayores edades que los no migrantes (2).

Bibliografía

Álvarez Cáceres, Rafael (1995). Estadística Multivariante y No Paramétrica con SPSS. Ediciones Diaz de Santos S.A. Madrid. España.

INDEC. Encuesta Permanente de Hogares. Base Usuaría. Ondas de Octubre de 2000, 2001 y 2002. Argentina

Rosales, Gladys (2002). Gran Catamarca, Mercado Laboral y Migración. Período 1995- 1999. Secretaría de Ciencia y Tecnología. Universidad Nacional de Catamarca. Arg.

Rosales, Gladys (2006). Mercado de Trabajo y Migraciones. Gran Catamarca. Período 1995- 2002. Capítulo de Libro en Libro: “Aplicaciones de las Estadísticas Oficiales en la Investigación Científica- Serie Investigación Científica”. Dirección Provincial de Estadísticas y Censos de Catamarca. ISSN 1850- 6526. Diciembre de 2006. Argentina

Rosales, Gladys (2007). Mercado de Trabajo de Áreas Urbanas del Noroeste Argentino (NOA). Período 2000- 2002. Revista de Investigación en Ciencias Económicas y Administración (RICEA). FACEYA. UNCa., con referato. Con evaluación externa aprobada y aceptado para publicación. Diciembre de 2006. Argentina

Rosales, Gladys (2007). Migración y Mercado de Trabajo de Áreas Urbanas del Noroeste Argentino (NOA). Período 2000- 2002. Revista de Investigación en Ciencias Económicas y Administración (RICEA). FACEYA. UNCa., con referato. Con evaluación externa aprobada y aceptado para publicación. Diciembre de 2006. Argentina

SPSS for Windows, Advanced Statistics, Release 14. Inc Chicago. USA

