

MODELO DE REGRESION LOGISTICA
PARA ESTIMAR LA AUSENCIA O
PRESENCIA DE EMPLEO EN JOVENES
DEL DISTRITO DE SAN MIGUELITO
ENTRE LOS 18 - 29 AÑOS CON
RELACION A ALGUNAS VARIABLES
SOCIODEMOGRAFICAS:

JULIO DE 2003.

REGRESION LOGISTICA

El análisis de regresión logística es un modelo matemático formulado con el propósito de predecir el comportamiento de una variable dependiente en función de una o mas variables independientes .

La regresión logística es una de las herramientas estadísticas con mejor capacidad para el análisis de datos en investigación clínica y epidemiología, de ahí su amplia utilización.

REGRESION LOGISTICA

El objetivo fundamental que resuelve esta técnica es el de predecir cómo influye en la probabilidad de aparición de un suceso (dicotómico), la presencia o no de diversos factores y el valor o nivel de los mismos.

La predicción se realiza en función de un conjunto de variables independientes con capacidad explicativa respecto a la variable dependiente.

REGRESION LOGISTICA

El modelo logístico puede ser también utilizado para estimar la probabilidad de aparición de cada una de las posibilidades de un suceso con más de dos categorías (Politómicas)

Generalmente, este tipo de situaciones se aborda mediante técnicas de regresión lineal. Sin embargo, la metodología de la regresión lineal es aplicable, cuando los datos no se ajustan a una línea recta.

REGRESION LOGISTICA

La regresión logística es una variación de la regresión lineal para situaciones en que la estimación de la variable dependiente se realiza en términos de probabilidad.

En general, la variable dependiente Y puede ser dicotómica o Politémicas. Las variables explicativas X_i pueden ser dicotómicas, nominal, ordinal, intervalo o razón.

MODELO DE REGRESION LOGISTICA MULTIVARIABLE

La regresión logística múltiple es una extensión del modelo simple con el fin de aumentar la capacidad explicativa del modelo, ya que incluye más de dos variables predictoras o explicativas.

MODELO DE REGRESION LOGISTICA MULTIVARIABLE

La formula del modelo múltiple se deduce del modelo simple. La ecuación matemática del modelo es:

$$P(Y = 1 / X_K) = \frac{e^{(a+b_1X_1+\dots+b_kX_k)}}{1+e^{(a+b_1X_1+\dots+b_kX_k)}}$$

MODELO DE REGRESION LOGISTICA MULTIVARIABLE

Otras variantes matemáticas del modelo múltiple son:

$$\frac{P}{1-P} = e^{(a+b_i X_1 + \dots + b_k X_k)}$$

$$LN\left(\frac{P}{1-P}\right) = a + b_i X_1 + \dots + b_k X_k$$

ANTECEDENTES

Se deseaba llevar a cabo un estudio entre jóvenes mayores de 18 años del distrito de San Miguelito.

De manera que sustentara la necesidad de llevar a cabo proyectos de capacitación y aumentar el número de escuelas públicas y privadas que incorporaran el Sistema Educativo Laboral.

ANTECEDENTES

Para ello se selecciono una muestra de 1500 jóvenes entre los 18 y 34 años de ambos sexos, tomando como punto de partida la encuesta de hogares 2002.

RESULTADOS BASICOS

Estadísticas

	Actividad Económica	Sexo del joven	Grupo de edad del joven	Primaria y Sec. incomp. del joven	Secundaria y superior del joven
N	Valid	1500	1500	1061	1346
	Missing	0	0	439	154

Actividad Económica

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Desocupados	959	63.9	63.9	63.9
	Ocupados	541	36.1	36.1	100.0
	Total	1500	100.0	100.0	

Sexo del joven

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Mujer	758	50.5	50.5	50.5
	Hombre	742	49.5	49.5	100.0
	Total	1500	100.0	100.0	

Grupo de edad del joven

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	De 18-24	644	42.9	60.7	60.7
	De 25-29	417	27.8	39.3	100.0
	Total	1061	70.7	100.0	
Missing	System	439	29.3		
	Total	1500	100.0		

Nivel de educación del joven

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Primaria incompleta	16	1.1	1.2	1.2
	Primaria y secundaria incompleta	503	33.5	37.4	38.6
	Secundaria completa y superior	827	55.1	61.4	100.0
	Total	1346	89.7	100.0	
Missing	System	154	10.3		
	Total	1500	100.0		

Regresión Logística

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	945	63.0
	Missing Cases	555	37.0
	Total	1500	100.0
Unselected Cases		0	.0
Total		1500	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Desocupados	0
Ocupados	1

Block 1: Method = Enter

Variables en la Ecuacion

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	SEXO	1.000	.148	45.967	1	.000	2.719
	EGRUPO	.771	.146	27.821	1	.000	2.161
	PYSICJOV	-.465	.614	.572	1	.449	.628
	SYSUPJOV	-.251	.608	.170	1	.680	.778
	Constant	-1.277	.610	4.375	1	.036	.279

a. Variable(s) entered on step 1: SEXO, EGRUPO, PYSICJOV, SYSUPJOV.

RESULTADOS BASICOS

Nivel de educación del padre

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Ningún grado	194	12.9	16.2	16.2
	Primaria incompleta	112	7.5	9.3	25.5
	Primaria y secundaria incompleta	441	29.4	36.8	62.4
	Secundaria completa y superior	451	30.1	37.6	100.0
	Total	1198	79.9	100.0	
Missing	System	302	20.1		
	Total	1500	100.0		

Nivel de educación de la madre

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Ningún grado	124	8.3	9.8	9.8
	Primaria incompleta	112	7.5	8.9	18.7
	Primaria y secundaria incompleta	579	38.6	45.8	64.5
	Secundaria completa y superior	448	29.9	35.5	100.0
	Total	1263	84.2	100.0	
Missing	System	237	15.8		
	Total	1500	100.0		

Regresión Logística

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	714	47.6
	Missing Cases	786	52.4
	Total	1500	100.0
Unselected Cases		0	.0
Total		1500	100.0

a. If weight is in effect, see classification table for the total number of cases.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	SEXO	.824	.167	24.260	1	.000	2.280
	EGRUPO	.680	.168	16.451	1	.000	1.974
	PYSICJOV	-.213	.688	.096	1	.757	.808
	SYSUPJOV	-.169	.686	.060	1	.806	.845
	PINCMAD	.624	.428	2.130	1	.144	1.867
	PYSICMAD	.531	.340	2.447	1	.118	1.701
	SYSUPMAD	.538	.353	2.317	1	.128	1.712
	Constant	-1.755	.739	5.640	1	.018	.173

a. Variable(s) entered on step 1: PINCMAD, PYSICMAD, SYSUPMAD.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	SEXO	.860	.169	25.827	1	.000	2.362
	EGRUPO	.704	.169	17.394	1	.000	2.023
	PYSICJOV	-.238	.695	.117	1	.732	.789
	SYSUPJOV	-.305	.696	.192	1	.661	.737
	PINCMAD	.388	.453	.732	1	.392	1.474
	PYSICMAD	.329	.378	.758	1	.384	1.389
	SYSUPMAD	.184	.403	.208	1	.648	1.202
	PINCPAD	.698	.386	3.263	1	.071	2.009
	PYSINPAD	.358	.308	1.357	1	.244	1.431
	SYSUPPAD	.700	.328	4.549	1	.033	2.014
	Constant	-1.918	.754	6.470	1	.011	.147

a. Variable(s) entered on step 1: PINCPAD, PYSINPAD, SYSUPPAD.

Regresión Logística

Block 1: Method = Backward Stepwise (Conditional)

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	SEXO	1.000	.148	45.967	1	.000	2.719
	EGRUPO	.771	.146	27.821	1	.000	2.161
	PYSICJOV	-.465	.614	.572	1	.449	.628
	SYSUPJOV	-.251	.608	.170	1	.680	.778
	Constant	-1.277	.610	4.375	1	.036	.279
Step 2	SEXO	1.000	.148	45.975	1	.000	2.719
	EGRUPO	.770	.146	27.796	1	.000	2.160
	PYSICJOV	-.219	.153	2.044	1	.153	.803
	Constant	-1.522	.141	116.135	1	.000	.218
Step 3	SEXO	.983	.147	44.871	1	.000	2.672
	EGRUPO	.776	.146	28.289	1	.000	2.172
	Constant	-1.591	.133	142.418	1	.000	.204

a. Variable(s) entered on step 1: SEXO, EGRUPO, PYSICJOV, SYSUPJOV.

GRACIAS POR SU ATENCION