METODO DE MODELOS DE RESPUESTA DISCRETA

Juan Carlos Flores

1. Introducción

Dentro de un análisis estadístico, en muchos estudios existe información cualitativa, para ello es necesario desarrollar modelos que respondan a esta información, como alternativa, en esta oportunidad desarrollamos los modelos categóricos denominados también modelos de respuesta discreta, partir de las variables cualitativas que son discretizadas en muchos estudios y este proceso nos da la posibilidad de poder analizar la información a partir de las probabilidades que proporcionan los modelos de respuesta categórica.

2. Marco conceptual

En el caso de un modelo de elección dicotómica. Así, sea p la probabilidad de que ocurra el suceso E. Evidentemente.

$$q = 1 - p$$

Es la probabilidad de que no ocurra ($\bar{E}$)(el complementario de E). En particular, E puede ser el suceso consistente en que un joven de bachillerato este de acuerdo con el gobierno de Evo Morales, o que una persona sepa leer; en estos casos, $\bar{E}$ sería el suceso consistente en que un joven de bachillerato superior no está de acuerdo con el gobierno de Evo Morales, o que el individuo no sepa leer.

Parece natural considerar p como un ordenada de una función de distribución (FD) y, por tanto, escribir

$$p = F(t) = \int_{-\infty}^{t} f(\zeta)d\zeta \text{ ,} \qquad (1)$$

Donde F(˙) es una función de distribución. Si f(˙) es la función de densidad asociada a la misma, de manera que tendríamos que calcular tantas probabilidades como individuos haya. Y realmente, es cierto que tendremos que hacerlo, pero para ello necesitamos calcular solamente una función de probabilidad. Explicitaremos más esta idea volviendo a la notación relativa al suceso intermedio, E. Así, definamos las variables.

$$y_i \begin{cases} = 1 \; si \; el \; i - esimo \; individuo \; escoge \; la \; alternativa \; xorrespondiente \; al \; suce \; E \\ = 0 \; en \; otro \; caso \end{cases}$$

Si disponemos de una muestra de n individuos, los aspectos probabilísticas de la muestra quedaran realmente caracterizados por

$$P_r[y_i = 1] = F(x_i \cdot \beta),\ \ P_r[y_i = 0] = 1 - F(x_i \cdot \beta),$$

Evidentemente hay muchas formas de escoger las FD, F(.). Obsérvese también que los primeros miembros de (2) no son observables. No obstante, se pueden estimar si al individuo j-ésimo le hemos presentado repetidamente el mismo problema de elección y conocemos la alternativa que ha escogido a cada caso. Entonces podemos definir

$$y_{it} = \begin{cases} 1 & Si\ en\ el\ instante\ t\ el\ i-esimo\ individuo\ escoge\ la\ alternativa\ correspondiente\ al\ suceso\ E \\ 0 & otro\ caso \end{cases}$$

Donde $i=1,2.......n;\ t= 1,2.....,T.$

Evidentemente, como el vector X_i. Corresponde a los atributos de las alternativas y/o del j-ésimo individuo, sus elementos no dependen de t, o sea, del lugar, dentro de la sucesión de T elecciones, que ocupa la j-ésimo elección.

Por lo tanto, podemos calcular

$$\hat{p}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}. \qquad (3)$$

La razón para hacer resto es que podemos considerar la T elecciones del j-ésimo individuo como un sucesión de ensayos Bernoulli, para que la probabilidad de que ocurra el sucesos E es

$$F(x_i \cdot \beta).$$

Y la probabilidad de que no ocurra $\bar{E}$ es

$$1 - F(x_i \cdot \beta).$$

La implicación sustantiva de esta afirmación es que las T elecciones realizadas por el j–ésimo individuo son independientes. Por tanto, podemos considerar este proceso como un proceso binomial y, en consecuencia, estimar la probabilidad de éxitos como (3), ya que T es fijo y la aleatoriedad reside en los valores que toman las variables y_{it}, $t=1,2,....,T.$

En caso en que T>1 que acabamos de considerar recibe normalmente el apelativo de varias observaciones por elemento e lustra el significado de los términos análisis "probit" y "logit" en el sentido siguiente. Intuitivamente, resulta claro que $\hat{p}_i$ es una estimación de la ordenada

$$F(x_i \cdot \beta).$$

Si suponemos que F(.) es la función de distribución normal estándar, podemos definir el probit de $\hat{p}_i$ encontrando la abscisa a la que corresponde; o sea, definiremos.

$$Probit(\hat{p}_i) = \hat{t}_i + 5 \qquad (4)$$

donde $\hat{t}_i$ es un valor tal que $F(\hat{t}_i)=\hat{p}_i$; se le añade 5 para la posibilidad de que el segundo miembro de (4) sea negativo, a todos los efectos prácticos. En consecuencia, el probit de $\hat{p}_i$ no es sino el valor de la función inversa de $F(x_i \cdot \beta)$ centrado adecuadamente.

Analógicamente, si F (.) es la función de distribución logística estándar, podemos definir el logit de $\hat{p}_i$ como

$$Logit(\hat{p}_i) = \ln\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = \hat{t}_i. \qquad (5)$$

Este valor se define así porque la distribución (FD) logística estándar viene dada por

$$F(t) = \frac{1}{1+e^{-t}} \qquad (6)$$

Por tanto

$$\ln\left(\frac{F(t)}{1-F(t)}\right) = \ln(e^t) = t.$$

Obsérvese que tanto si estamos trabajando con el logit como si lo hacemos con el probita, siempre esta en juego la "inversion" de la FD, F(•). Es decir, dada una ordenada, $F(t_i)$, tenemos que contrar la abscisa, d_i, correspondiente. para la FD normal, este trabajo se puede realizar consultando cualquier tabla de la distribución normal. Para la distribución logística, se dispone de manera explicita de la función inversa de F(•), que viene dada sencillamente por

$$t_i = \ln\left[\frac{F(t_i)}{1-F(t_i)}\right].$$

En un caso muy particular la regresión logística se tendrá

$$\ln\left(\frac{F(t)}{1-F(t)}\right) = \ln\left(\frac{p}{1-p}\right) = \alpha_0 + \alpha_1 x.$$

$$\frac{p}{1-p} = \alpha_0 + \alpha_1 x.$$

Asumiendo que ε tiene distribución logística tenemos

$$p = F(\alpha + \beta X_i) = \frac{e^{\alpha_0 + \alpha_1 X_i}}{1 - e^{\alpha_0 + \alpha_1 X_i.}} = \frac{e^{\alpha_0 + \beta X_i}}{1 - e^{\alpha_0 + \beta X_i}}$$

El efecto no lienal de X_i puede ser también entendido calculando la derivada de la probabilidad acumulada con respecto a X_i :

$$\frac{\partial F(\alpha + \beta X_i)}{\partial X_i} = F(\alpha + \beta X_i)[1 - F(\alpha + \beta X_i)]\,\beta$$

Esto muestra que el efecto de un cambio en X_i depende no solo del valor de β sino tambien de valor tomado por la función logística.

Estimación Para realizar la estimación del modelo debemos recurrir al método de máxima verosimilitud.

$$L(\alpha, \beta; X_i) = \prod_{i=1}^{n} \left[\frac{e^{\alpha + \beta X_i}}{1 + e^{\alpha + \beta X_i}}\right]^{Y_i} \left[1 - \frac{e^{\alpha + \beta X_i}}{1 + e^{\alpha + \beta X_i}}\right]^{1-Y_i}$$

$$\ell = \sum_{i=1}^{n} \left\{Y_i\, Ln\left[\frac{e^{\alpha + \beta X_i}}{1 + e^{\alpha + \beta X_i}}\right] + (1 - Y_i) Ln\left[1 - \frac{e^{\alpha + \beta X_i}}{1 + e^{\alpha + \beta X_i}}\right]\right\}$$

Las condiciones de primer orden para la maximización de la función de verosimilitud son:

$$\frac{\partial \ell}{\partial \hat{\alpha}} = \sum_{i=1}^{n} \left(Y_i - \frac{e^{\hat{\alpha} + \hat{\beta} X_i}}{1 + e^{\hat{\alpha} + \hat{\beta} X_i}}\right) = 0$$

$$\frac{\partial \ell}{\partial \hat{\beta}} = \sum_{i=1}^{n} \left(Y_i - \frac{e^{\hat{\alpha} + \hat{\beta} X_i}}{1 + e^{\hat{\alpha} + \hat{\beta} X_i}}\right) X_i = 0$$

Como se puede observar en las condiciones de primer grado las incógnitas de ambas ecuaciones ($\hat{\alpha}$ y $\hat{\beta}$) entran en forma no lineal y por lo tanto no pueden resolverse por métodos lineales. Ameniya(1985) demostró que la función de verosimilitud del modelo Logit es globalmente cóncava por lo que las condiciones de segundo orden para un máximo se cumplen.

Las condiciones de segundo orden vienen dadas por las siguientes expresiones:

$$\frac{\partial^2 \ell}{\partial \hat{\alpha}^2} = \sum_{i=1}^{n} \left(\frac{e^{\hat{\alpha} + \hat{\beta} X_i}}{1 + e^{\hat{\alpha} + \hat{\beta} X_i}} \right) \left(1 - \frac{e^{\hat{\alpha} + \hat{\beta} X_i}}{1 + e^{\hat{\alpha} + \hat{\beta} X_i}} \right)$$

$$\frac{\partial \ell}{\partial \hat{\alpha} \partial \hat{\beta}} = - \sum_{i=1}^{n} \left(\frac{e^{\hat{\alpha} + \hat{\beta} X_i}}{1 + e^{\hat{\alpha} + \hat{\beta} X_i}} \right) \left(1 - \frac{e^{\hat{\alpha} + \hat{\beta} X_i}}{1 + e^{\hat{\alpha} + \hat{\beta} X_i}} \right) X_i$$

$$\frac{\partial^2 \ell}{\partial \hat{\beta}^2} = - \sum_{i=1}^{n} \left(\frac{e^{\hat{\alpha} + \hat{\beta} X_i}}{1 + e^{\hat{\alpha} + \hat{\beta} X_i}} \right) \left(1 - \frac{e^{\hat{\alpha} + \hat{\beta} X_i}}{1 + e^{\hat{\alpha} + \hat{\beta} X_i}} \right) X_i^2$$

^

Llamando Pi a $e^{\hat{\alpha}+\hat{\beta}X_i}/(1+e^{\hat{\alpha}+\hat{\beta}X_i})$ para abreviar notación, estas condiciones de segundo orden pueden agruparse en la matris Hesiana(la matriz de las segundas derivadas).

Generalizando los modelos de regresión logística se tiene

$$H(\hat{\alpha}, \hat{\beta}) = - \begin{bmatrix} \sum_{i=1}^{n} P_i (1 - P_i) & \sum_{i=1}^{n} P_i (1 - P_i) X_i \\ \sum_{i=1}^{n} P_i (1 - P_i) X_i & \sum_{i=1}^{n} P_i (1 - P_i) X_i^2 \end{bmatrix}$$

$$\Pr(Y_i = j | X, W, Z) = \frac{e^{\beta_{0,j} + \beta_{1,j} X_i + \gamma_1 W_{i,j} + \delta Z_j}}{\sum_{l=1}^{J} e^{\beta_{0,j} + \beta_{1,j} X_i + \gamma_1 W_{i,j} + \delta Z_j}}$$
$$j = 1,2, \dots, J$$

Donde $\beta_{0,J} = \beta_{1,J} = 0$ por motivos de identificación.

3. Estimación e Inferencia en Modelos de Elección Múltiple

La estimación de los parámetros de cualquiera de los modelos de elección mútiple descritos anteriomente, se realiza mediante el método de maximización de la función de verosimilitud. La función de verosimilitud por:

$$L(\theta) = \prod_{i=1}^{N} \prod_{j=1}^{J} \Pr[Y_i = j | \bullet]^{I[Y_i = j]}$$

Note que para cada i,

$$\sum_{j=1}^{J} I[Y_i = j] = 1$$

Donde I[•] es la función indicador que asume el valor 1 cuando el argumento de la función es verdadero y asume el valor 0 cuando el argumento de la función es falso.

Obteniendo el logaritmo de la función de verosimilitud, y derivando respecto de los parámetros que serían las condiciones de primer orden, es inmediatamente obvio que no se pueden resolver las condiciones de primer orden despejando las incógnitas por tratarse de ecuaciones no lineales en las mismas. Para obtener los estimadores de máxima verosimilitud del modelo hay que recurrir a algoritmos no lineales. El procedimiento de maximización requiere que se cumplan las condiciones de segundo orden representadas por la siguiente matriz Hesiana:

$$H(\hat{\theta}) = \left\{ \begin{array}{cccccccc}
\frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.1}^2} & \cdots & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.1} \partial \hat{\beta}_{0.J-1}} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.1} \partial \hat{\beta}_{1.1}} & \cdots & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.1} \partial \hat{\beta}_{1.J-1}} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.1} \partial \hat{\gamma}_1} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.1} \partial \hat{\delta}} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.J-1} \partial \hat{\beta}_{0.1}} & \cdots & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.J-1}^2} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.J-1} \partial \hat{\beta}_{1.1}} & \cdots & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.J-1} \partial \hat{\beta}_{1.J-1}} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.J-1} \partial \hat{\gamma}_1} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.J-1} \partial \hat{\delta}} \\
\frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{1.1} \partial \hat{\beta}_{0.1}} & \cdots & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{1.1} \partial \hat{\beta}_{0.J-1}} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{1.1}^2} & \cdots & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{1.1} \partial \hat{\beta}_{1.J-1}} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{1.1} \partial \hat{\gamma}_1} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{1.1} \partial \hat{\delta}} \\
\vdots & \cdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.J-1} \partial \hat{\beta}_{0.1}} & \cdots & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{1.J-1} \partial \hat{\beta}_{0.J-1}} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{0.J-1} \partial \hat{\beta}_{1.1}} & \cdots & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{1.J-1}^2} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{1.J-1} \partial \hat{\gamma}_1} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\beta}_{1.J-1} \partial \hat{\delta}} \\
\frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\gamma}_1 \partial \hat{\beta}_{0.1}} & \cdots & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\gamma}_1 \partial \hat{\beta}_{0.J-1}} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\gamma}_1 \partial \hat{\beta}_{1.1}} & \cdots & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\gamma}_1 \partial \hat{\beta}_{1.J-1}} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\gamma}_1^2} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\gamma}_1 \partial \hat{\delta}} \\
\frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\delta} \partial \hat{\beta}_{0.1}} & \cdots & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\delta} \partial \hat{\beta}_{0.J-1}} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\delta} \partial \hat{\beta}_{1.1}} & \cdots & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\delta} \partial \hat{\beta}_{1.J-1}} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\delta} \partial \hat{\gamma}_1} & \frac{\partial^2 \ell(\hat{\theta})}{\partial \hat{\delta}^2}
\end{array} \right\}$$

Para obtener esta matriz necesitamos las segundas derivadas de la función de verosimilitud con respecto al vector θ.

Puede demostrarse que la función de verosimilitud es globalmente cóncava y por lo tanto las condiciones de segundo orden se satisfacen.

El estimador de máxima verosimilitud del vector θ es insesgado, consistente y eficiente con la siguiente distribución:

$$\theta \overset{A}{\sim} N\left(\theta, \left(-H(\hat{\theta})\right)^{-1}\right)$$

Estos mismo modelos presentados para la función logística, pueden desarrollarse para la función de distribución normal dando origen a los denominados modelo multinomial probit y modelo probit condicional.

4. Medidas de Bondad del Ajuste

Como pasaba en los modelos de variable dependiente binaria, aquí se pueden utilizar como medidas de bondad del ajuste los denominados Pseudo-R^2. Uno de esos estadísticos es el R^2 de McFadden:

$$R^2_{MF} = 1 - \frac{\ell(\hat{\theta})}{\ell(\hat{\theta}_0)}$$

Donde $\ell(\hat{\theta})$ es el valor del logaritmo de la función de verosimilitud evaluada en Los estimadores de MV y $\ell(\hat{\theta}_0)$ es el valor del logaritmo de la función de Verosimilitud de un modelo que tiene solo una constante.

El R^2_{MF} tiene la misma interpretación que el R^2 común. Es decir, nos dice que porcentaje de la variabilidad de la variable dependiente está explicado por la regresión.

Aplicaciones de los Modelos Logit

Se presenta la aplicación de la teoría de los modelos de respuesta discreta o modelos categóricos. Una de las variables que se toma en cuenta es si sabe leer y escribir y la variable edad de la encuesta de hogares MECOVI realizada en la gestión 2002 a Personas de 7 años y más. Los tradicionales contenidos de la Encuesta se mantienen: Información general de los miembros del hogar, migración, salud, educación, empleo, ingresos no laborales, gastos en consumo, vivienda, contingencias y préstamos del hogar e ingreso del productor agropecuario independiente.

Considerando las variables anteriormente mencionadas y procediendo a la utilización del modelo, las salidas fueron las siguientes:

Variables en la ecuación

Tabla 4.4

		B	E.T.	Wald	gl	Exp(B)
Paso 1(a)	EDAD	-,058	,001	2343,479	1	,944
	Constante	4,002	,056	5154,283	1	54,722

a Variable(s) introducida(s) en el paso 1 : EDAD

Esta tabla 4.4 nos proporciona los coeficientes del modelo de la forma

$$p_i = P(Y_i = 1) = F(x_i'\beta) = \frac{e^{x_i'\beta}}{1 + e^{x_i'\beta}}$$

En este caso

$$p_i = P(Y_i = 1) = F(x_i'\beta) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}$$

$$p_i = P(Y_i = 1) = F(x_i'\beta) = \frac{e^{4.002\pm0.058x}}{1 + e^{4.002\pm0.058x}}$$

Ejemplo si X=20, entonces se tendrá

$p_i = P(Y_i = 1) = F(x_i'\beta) = \dfrac{e^{4.002-0.058(20)}}{1+e^{4.002-0.058(20)}}$ es la probabilidad de que un ciudadano de 20 años sepa leer, esta interpretación es válida para todo análisis a nivel nacional.

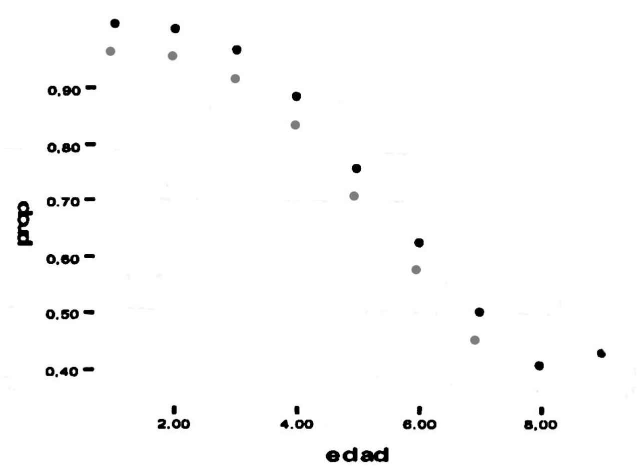
Proseguimos con la prueba de **Hosmer y Lemeshow** en la tabla 4.5

Tabla 4.5
Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	46.142	4	.061

Mide la bondad de ajuste, observando su significación es 0.061 la hipótesis nula es el modelo de regresión logística es adecuado al nivel del 5%. Es decir que existe evidencia de que el modelo no sea malo.

Cuya gráfica es del modelo es



La figura nos da la gráfica que nos proporciona el modelo de regresión logística. El cual como interpertación nos da lo siguiente:

De acuerdo al gráfico se puede interpretar que las edades que se encuentran en el rango 1 y 2 es decir el 1 es rango de edad 7 y 17 años; 2 es rango de edad 18 y 27; 3 es rango de edad 28 y 37, la probabilidad de que sepa leer y escribir es bastante alta aproximadamente de 0.90 a 0.98. Las edades que se encuentran en 4 es decir el rango de edad 38 y 47, la probabilidad de que sepan leer y escribir la probabilidad esta entre 0.8 a 0.7. Si se sigue observando el siguiente rango de edad 5 el rango de edad 48 y 57; la probabilidad de que sepa leer y escribir esta entre 0.7 y 0.59 aproximadamente. Siguiendo con este proceso de interpretación la probabilidad de que las personas sepan leer y escribir el siguiente rango es decir 6 el rango de edad 58 y 67 es 0.65 a 0.48 aproximadamente, y por último los rangos de edad de 7 es rango de edad 68 y 77; 8 es rango de edad 78 y 87 por último 9 es rango de edad es el rango 88y 98, la probabilidad de que sepan leer y escribir son menor a 0.4.

Esta información obtenida por el modelo y graficado tiene sentido en la vida real de nuestro país y se debe a distintos factores puede ser explicado por la información que se tiene de acuerdo a los resultados que se tiene del censo de población y vivienda del 2001, las encuestas de hogares efectuadas por el INE permiten identificar diversas causas de la inasistencia a la escuela, los motivos que explica esta causa es: 1.- por problemas económicos el 45% de las personas no van a la escuela por esta causa, prosiguiendo la inasistencia se debe también a problemas ámbito personal y familiar con el 41.7%, problemas educativos el 5%, otros 4.2% y por enfermedad y discapacidad el 3.1%. Un importante punto que llama la atención de que la mayoría de los encuestados declaró no estar asistiendo a la escuela debido a problemas económicos o aquellos relacionados con el ámbito personal y familiar. Lo que sugiere que las intervenciones dedicadas a mejorar la oferta educativa no estarían afectando a la inasistencia a la escuela. Sin embargo, algunas causas de inasistencia, incluidos dentro de las categorías problemas económicos y ámbito familiar y personal, pueden ser el resultado de las limitaciones de oferta educativa para responder las necesidades de la población. También la declaración de no asistir a los centros de estudios escuela, colegios, etc. Por trabajo o por falta de interés podrían estar mostrando la falta de respuesta a la expectativa que tiene los alumnos y las familias respecto a la educación.

Índice