

## *EDUCACIÓN E INGRESO COMO PREDICTORES DE LA ESPERANZA DE VIDA: EVIDENCIAS DE UN ANÁLISIS DE REGRESIÓN APLICADO A INDICADORES DE DESARROLLO HUMANO*

Eiliana Montero Rojas

### RESUMEN

Usando datos del Informe sobre Desarrollo Humano de 1998 se ajustó un modelo de regresión lineal múltiple para estimar la *esperanza de vida al nacer* como función de un indicador de poder adquisitivo, el *PIB per cápita en dólares*, y un indicador de educación, la *tasa de matriculación combinada*. Los resultados muestran que la importancia relativa del indicador de educación en la predicción de la esperanza de vida es prácticamente tres veces la importancia relativa del indicador del poder adquisitivo. El estudio también ilustra la utilidad del enfoque de regresión múltiple como herramienta para la investigación en ciencias sociales.

### ABSTRACT

A linear regression model was adjusted with data from the Human Development Report 1998. The purpose was to predict Life Expectancy at Birth as a function of an indicator of education, the Combined First, Second and Third Level Gross Enrolment Ratio, and an indicator of monetary or buying power, the Real GDP per capita in dollars. Results show that the relative importance of the education indicator is practically three times the importance of the buying power indicator. The study also shows the usefulness of the multiple regression approach as a tool for research in Social Sciences.

### INTRODUCCIÓN

Los métodos cuantitativos constituyen una útil herramienta para la investigación y la toma de decisiones. El presente trabajo pretende mostrar el uso de una técnica relativamente sencilla de análisis de datos, la regresión múltiple, para explorar el grado en que un indicador de longevidad, la *esperanza de vida*, puede ser predicho o estimado usando la *tasa de matriculación combinada*, como indicador de educación y el Producto

Interno Bruto (PIB) *per cápita* en dólares, como indicador de poder adquisitivo o ingreso. Se trabajó con una muestra aleatoria sistemática de 35 países y los datos utilizados son los que aparecen en el Informe sobre desarrollo humano 1998 (Programa de las Naciones Unidas para el Desarrollo, 1998).

Desde el punto de vista sustantivo resulta interesante explorar estas relaciones y establecer cuál de los dos predictores (el indicador de educación o el de ingreso) arroja una mayor asociación con el indicador de

esperanza de vida. Los resultados y su interpretación pueden ser útiles para informar la generación de políticas públicas, por ejemplo, en términos de decisiones de inversión para diferentes programas de desarrollo. Atendiendo a las evidencias empíricas sobre la importancia relativa de cada uno de los predictores, un agente tomador de decisiones podría inclinarse a favorecer el fortalecimiento de programas específicos en una de las dos dimensiones, educativa o económica, suponiendo que el objetivo final sea el aumentar los niveles de longevidad de la población.

## OBJETIVOS

Por medio de la aplicación de un modelo de regresión múltiple con datos extraídos del Informe sobre desarrollo humano 1998, se pretende:

1. Establecer el grado en que la *esperanza de vida*, como medida de longevidad, puede ser predicha o estimada conociendo el *PIB per cápita* y la *tasa de matriculación combinada* de un país.
2. Determinar la importancia relativa de cada una de las dos variables independientes, *tasa de matriculación combinada* y *PIB per cápita* en la predicción de la esperanza de vida.
3. Ilustrar, por medio de la aplicación de una técnica sencilla de análisis estadístico de datos, cómo se pueden establecer evidencias empíricas que alimenten la toma de decisiones y la investigación.

## METODOLOGÍA

### DESCRIPCIÓN DE LOS INDICADORES

A continuación se describen los indicadores empleados en el análisis, según se definen en el Informe sobre desarrollo humano 1998.

*Esperanza de vida al nacer*: Es el número de años, en promedio, que vivirá un recién nacido si las pautas de mortalidad imperantes en el momento de su nacimiento siguieran siendo las mismas a lo largo de toda su vida. Este es un indicador clásico, ampliamente conocido y utilizado para juzgar aspectos de desarrollo humano, dada la dependencia que existe entre las pautas de mortalidad de un país y sus condiciones de salud y calidad de vida.

*PIB per cápita en dólares*: El PIB es la producción total para uso final de bienes y servicios de una economía. El *PIB per cápita* es el cociente que resulta de dividir el PIB entre la población total del país. De esta forma se obtiene un promedio por habitante. Se convierte luego a dólares de Estados Unidos sobre la base de la paridad de poder adquisitivo de la moneda del país (tipo de cambio). El *PIB per cápita* es también una medida ampliamente conocida y utilizada tradicionalmente en los ámbitos del desarrollo humano. Es un indicador de poder adquisitivo o ingreso, aunque debe hacerse notar que al ser una medida promedio, no hace referencia a la variabilidad, dispersión, o heterogeneidad en la distribución del ingreso.

*Tasa de matriculación combinada*: La tasa de matriculación para un nivel dado de educación (primaria, secundaria, superior) es la cantidad de alumnos matriculados en ese nivel de educación expresada como porcentaje del total de la población del grupo de edades correspondiente a ese nivel. Así, la tasa de matriculación combinada es el promedio ponderado de las tasas de matriculación en los tres niveles (primaria, secundaria, superior). Las ponderaciones están dadas por el número de personas en cada uno de los grupos de edad.

Comparada con la esperanza de vida y el PIB per cápita, la tasa de matriculación combinada es una medida menos conocida. Hasta hace relativamente poco tiempo el indicador más común de educación en los ámbitos de desarrollo humano era la *tasa de alfabetización* (porcentaje de personas mayores de 12 años que saben leer y escribir). Sin embargo, es claro a partir de su defini-

ción, que la tasa de matriculación combinada es una medida más sensible y representativa como indicador de educación, dado que involucra la cobertura educativa en los tres niveles: primario, secundario y superior. Se puede ilustrar el cálculo de este indicador por medio del siguiente ejemplo: Supongamos que la población de un país es de un millón de habitantes, un 30% (300 000 personas) están en el rango de edad que contempla usualmente la educación primaria, un 20% (200 000) están en el rango de edad de la educación secundaria y un 10% (100 000) en el rango de edad más común para la educación superior. Supongamos también que la cobertura en primaria es de 80%, en secundaria de 50% y en educación universitaria de 20%. Así, la tasa de matriculación combinada para este país sería:

$$(300\ 000 \times 0,8 + 200\ 000 \times 0,5 + 100\ 000 \times 0,2) / (300\ 000 + 200\ 000 + 100\ 000) = 60\%$$

Del ejemplo anterior se deduce claramente que la tasa de matriculación combinada es en realidad una proporción o porcentaje.

Finalmente, debe considerarse que cada uno de los indicadores incluidos en este estudio puede ser evaluado desde el punto de vista de su validez y confiabilidad. Por ejemplo, puede argumentarse que el *PIB per cápita*, al no medir el grado de concentración del ingreso en un país es una medida imperfecta de poder adquisitivo. Aunque existen otros indicadores que podrían considerarse más eficientes para representar estas dimensiones, hay una importante restricción para su empleo, dado que la información necesaria para su cálculo no se recoge regularmente en muchos países (especialmente en los menos desarrollados). De ahí que se empleen las medidas definidas anteriormente, en vista de que están disponibles para la gran mayoría de las naciones del mundo.

#### DESCRIPCIÓN DE LA MUESTRA UTILIZADA

Utilizando la publicación del Informe sobre desarrollo humano 1998 (PNUD) y

tomando el listado de los países con sus respectivos indicadores de ingreso, educación y longevidad se escogió una muestra aleatoria sistemática de uno de cada cinco países. En una muestra aleatoria sistemática la “semilla” o punto de arranque es un número escogido al azar entre uno y el grado de espaciamiento de la muestra, en este caso un número entre uno y cinco. A partir de ahí se cuentan cinco unidades en la lista para el segundo elemento de la muestra y así sucesivamente. De esta forma se escogieron 35 de un total de 174 países en el listado.

#### PAÍSES SELECCIONADOS EN LA MUESTRA (EN ORDEN ALFABÉTICO)

Angola	Guatemala	Omán
Argentina	Indonesia	Paraguay
Bolivia	Italia	Perú
Canadá	Líbano	Rep. Árabe Siria
Chile	Luxemburgo	Rep. Dem. Pop. Lao
China	Malawi	Seychelles
Comoras	Malí	Santo Tomé y Príncipe
Croacia	Mauricio	Suiza
Dominica	Mongolia	Venezuela
España	Mozambique	Yemen
Finlandia	Myanmar	Zambia
Granada	Nicaragua	

Fuente: La autora, basada en PNUD (1998).

La decisión de utilizar este tamaño de muestra se fundamentó en la necesidad de que las pruebas de inferencia estadística pudieran detectar como significativas asociaciones de importancia práctica. Así, para inferencias acerca del coeficiente de correlación de Pearson y trabajando con un nivel de significancia del 10% en la prueba de hipótesis, con una muestra de alrededor de 35 elementos se detectarían como significativas asociaciones observadas mayores o iguales a 0,30 en valor absoluto (Brewer, 1991).

#### LA TÉCNICA DE ANÁLISIS: REGRESIÓN MÚLTIPLE

En un modelo de regresión múltiple se construye una ecuación en donde la variable dependiente o criterio (Y), aquella que interesa predecir o estimar, se concibe como una función lineal de un conjunto de variables in-

dependientes o predictoras ( $X_1, X_2, X_3, \dots, X_n$ ). Esta función es llamada ecuación de regresión y matemáticamente se puede expresar así:

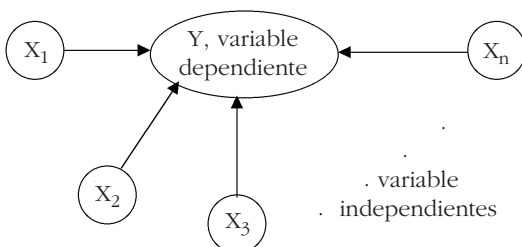
$$Y = f(X_1, X_2, X_3, \dots, X_n) = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_nX_n$$

Se parte del supuesto de que el nivel o valor que observamos en la variable dependiente es el resultado de la influencia simultánea de varios factores. Las variables independientes son estos factores relevantes. Así, se busca que el conjunto de variables independientes permita predecir, con un grado razonable de precisión, el comportamiento de la variable dependiente. Esta técnica de la estadística multivariable permite también analizar la fuerza o intensidad de la asociación entre las variables. Determina el grado de relación entre cada una de las variables independientes y la dependiente, “controlando”, “aislando” o “manteniendo constante” el efecto de las otras variables independientes en el modelo. Una vez que el modelo es ajustado, es decir, una vez que se ha estimado la ecuación de regresión que mejor reproduce el comportamiento de los datos empíricos, se puede juzgar la importancia relativa de cada variable independiente en la predicción de la dependiente. Esto se hace examinando la magnitud de los llamados “coeficientes de regresión”.

En la figura 1 se ilustra esquemáticamente un modelo de regresión múltiple con “n” variables independientes.

FIGURA 1

MODELO DE REGRESIÓN MÚLTIPLE: ILUSTRACIÓN



$$Y = f(X_1, X_2, X_3, \dots, X_n) = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_nX_n$$

Fuente: La autora

El coeficiente  $a_0$  es la constante de la ecuación de regresión y representa el valor que toma  $Y$ , la variable dependiente, cuando todas las variables independientes son iguales a cero. En general, este resultado no suele tener una interpretación sustantiva de importancia. Los valores  $a_1, a_2, a_3, \dots, a_n$  son los coeficientes de regresión y representan los pesos o ponderaciones de las correspondientes variables independientes  $X_1, X_2, X_3, \dots, X_n$ . El criterio fundamental que guía el procedimiento de estimación de estos valores en una ecuación de regresión específica es la minimización de los residuos o errores de estimación, definiendo estos como la diferencia entre los valores observados de la variable dependiente y los valores estimados con base en el modelo de regresión. Cuando  $a_i$  es positivo existe una relación directa entre  $X_i$  y  $Y$  (cuando aumenta una, aumenta la otra), manteniendo constantes o controlando las otras variables independientes en el modelo. Por el contrario, si el coeficiente de regresión  $a_i$  es negativo la relación entre  $X_i$  y  $Y$  es inversa, es decir al aumentar una disminuye la otra. Estos coeficientes también se interpretan en el mismo sentido que una pendiente en una función lineal, cada uno de ellos indica en cuántas unidades cambia  $Y$  por unidad de cambio en  $X_i$ .

Aunque en un análisis de regresión las evidencias empíricas pueden interpretarse solamente en términos correlacionales y no causales, es importante mencionar que si el referente teórico está dado en términos de un modelo causal, este tipo de análisis ayuda a establecer si hay evidencias empíricas para apoyar la hipótesis de causalidad en una de sus condiciones necesarias, la existencia de asociación.

Para un tratamiento más profundo de esta técnica y descripción de sus posibles usos en contextos de ciencias sociales, pueden consultarse, entre otros a Pagano (1999), Hopkins, Hopkins y Glass (1997) y Daniel (1988).

En nuestro caso particular se eligió como variable dependiente o criterio la esperanza de vida ( $Y$ ), y como variables independientes o predictoras el producto interno bruto per cápita ( $X_1$ ) como indicador del poder

adquisitivo y la tasa de matriculación combinada ( $X_2$ ) como indicador de educación. Es decir,

$$Y = f(X_1, X_2) = a_0 + a_1X_1 + a_2X_2$$

Los análisis fueron realizados con el paquete estadístico SPSS para Windows, versión 8.

## RESULTADOS

### ANÁLISIS DESCRIPTIVO

El cuadro 1 muestra las estadísticas descriptivas para cada una de las tres variables de interés en la muestra analizada.

CUADRO 1

ESTADÍSTICAS DESCRIPTIVAS PARA LAS VARIABLES DEL ANÁLISIS  
n = 35

Medidas	PIB per cápita (en \$ anuales) ( $X_1$ )	Tasa de matriculación combinada (en porcentaje) ( $X_2$ )	Esperanza de vida al nacer (en años) (Y)
Promedio	7331,31	62,80	65,69
Mediana	3971,00	63,00	69,10
Desviación estándar	7963,17	18,19	10,76
Mínimo	565,00	18,00	41,00
Máximo	34 004,00	100,00	79,10

Fuente: La autora, basada en PNUD (1998).

Es interesante observar aquí la considerable variabilidad que presentan los datos, especialmente en el caso del *PIB per cápita*. Como se recordará, la desviación estándar nos dice cuántas unidades en promedio, se desvía cada dato del valor promedio del conjunto. Así, en el caso del *PIB per cápita*, la desviación estándar es incluso mayor que el valor del promedio. Esta gran dispersión también se evidencia al analizar los valores mínimo y máximo de esta variable, ya que el valor más bajo (565 dólares anuales) representa solamente un 1,66% del *PIB per cápita* más alto (34 004 dólares anuales). Tendencias similares, aunque quizá menos dramáticas, se

observan para la tasa de matriculación combinada y la esperanza de vida.

El cuadro 2, por su parte, contiene la matriz de correlaciones simples (de Pearson) para las tres variables. Debe recordarse que la magnitud del coeficiente de correlación indica la fuerza o intensidad de la asociación entre las dos variables. El valor más alto que puede tomar esta medida es la unidad. En general, cuando se trata de variables sociales o educativas, un valor de 0,30 o más para el coeficiente de correlación se considera ya como indicativo de una asociación lineal de importancia entre las dos variables. Además el signo del coeficiente indica la dirección de la asociación. Es positivo si la relación es directa, y negativo si la relación es inversa.

CUADRO 2

MATRIZ DE CORRELACIONES PARA LAS VARIABLES DEL ANÁLISIS<sup>1</sup>  
n = 35

Variable	Esperanza de vida (Y)	PIB per cápita ( $X_1$ )
PIB per cápita ( $X_1$ )	.665**	
Tasa de matriculación combinada ( $X_2$ )	.692**	.495**

1. Se trata de coeficientes de correlación de Pearson.  
\*\* Significativos al 1%

Fuente: La autora, basada en PNUD (1998).

Entonces a partir de lo que muestra el cuadro 2, se concluye que existen niveles relativamente altos de asociación entre las variables, siendo el valor mayor el que corresponde a la correlación entre esperanza de vida y tasa de matriculación combinada, con un valor de 0,692 y el más bajo el que representa la correlación entre tasa de matriculación combinada y PIB real per cápita, con una magnitud de 0,495.

### DIAGRAMAS DE DISPERSIÓN

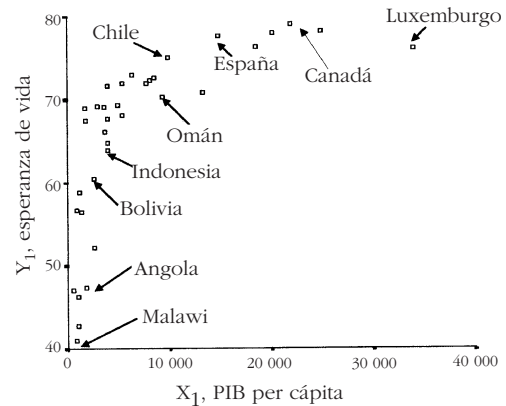
Uno de los supuestos fundamentales para la aplicación de un modelo de regre-

sión es el que se refiere a la existencia de una relación lineal entre las variables independientes y la variable dependiente. Para comprobar si este supuesto se cumple de manera aceptable se procedió a generar dos diagramas de dispersión, uno con la esperanza de vida (Y) como variable dependiente y el PIB per cápita ( $X_1$ ) como variable independiente, y otro con la misma variable dependiente y la tasa de matriculación combinada ( $X_2$ ) como variable independiente. Cada punto en el diagrama de dispersión representa el valor de un par ordenado (X,Y) en una unidad de estudio específica. Así, en nuestro caso particular, las unidades de estudio son los países y los pares (X,Y) son, para el primer diagrama de dispersión, los valores del PIB per cápita y la esperanza de vida para cada uno de los treinta y cinco países y, en el segundo diagrama, los valores correspondientes de la tasa de matriculación combinada y la esperanza de vida. Inspeccionando visualmente estos gráficos se puede establecer si es razonable suponer una relación lineal, al comprobar si “la nube” de puntos (X,Y) sigue la tendencia de una línea recta.

En el gráfico 1 se presenta el diagrama de dispersión que corresponde al PIB per cápita y la esperanza de vida. Del examen de este gráfico parece claro que existe un par ordenado que claramente se aleja de la tendencia general de los datos. Este corresponde al país de Luxemburgo, dado el valor extraordinariamente alto que presenta su PIB per cápita. Por su parte el gráfico 2 da cuenta de la relación entre la tasa de matriculación combinada y la esperanza de vida. Su inspección revela la existencia de dos valores extremos, pero esta vez en el extremo opuesto de la escala. Para Malawi y Zambia se observan valores excepcionalmente bajos de la esperanza de vida que corresponden con valores no tan bajos de la tasa de matriculación combinada.

GRÁFICO 1

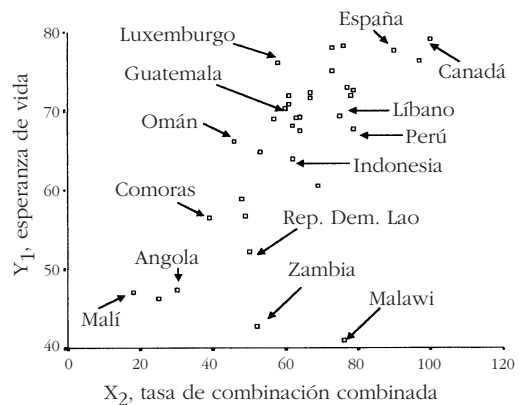
DIAGRAMA DE DISPERSIÓN PIB PER CÁPITA EN DÓLARES) Y ESPERANZA DE VIDA AL NACER



Fuente: La autora, basada en PNUD (1998).

GRÁFICO 2

DIAGRAMA DE DISPERSIÓN TASA DE MATRICULACIÓN COMBINADA Y ESPERANZA DE VIDA AL NACER



Fuente: La autora, basada en PNUD (1998).



Estas observaciones que se desvían de la tendencia general de los datos se llaman “outliers” o valores extremos en el contexto de una regresión múltiple. Los “outliers” pueden tener un efecto considerable sobre los resultados e interpretación del análisis, al disminuir considerablemente el poder predictivo del modelo de regresión. Los textos antiguos de regresión recomendaban la exclusión de estos casos del análisis, debido a sus posibles efectos adversos para la bondad del ajuste del modelo. Autores más recientes como Pedhazur (1982), Stevens (1986) y Kleinbaum *et al.* (1998) aconsejan ser más cuidadoso y reflexivo en cuanto a la posible eliminación de estos valores del análisis de regresión y tratar, más bien, de establecer posibles explicaciones para su comportamiento.

En nuestro contexto particular, puede haber varias razones que explican la existencia de valores extremos en estos tres países. En el caso de Luxemburgo, se debe recordar que es un país diminuto, con un área de 2586 Km<sup>2</sup> y una población menor a los 450000 habitantes. Goza de un estándar alto de vida y aproximadamente 30% de su población es extranjera. Algunas de estas características tan distintivas podrían estar asociadas al valor excesivamente alto para el PIB *per cápita* de Luxemburgo.

En el caso de Malawi y Zambia una razón que puede explicar el comportamiento diferencial de sus observaciones es que estos países han sido dos de los más azotados por la epidemia del SIDA en el área del Africa sub-Sahariana. Esta situación obviamente resulta en una esperanza de vida más baja de lo que se hubiera esperado en otras condiciones, si la epidemia no hubiera surgido. El otro hecho interesante es que Zambia y Malawi son países limítrofes, ambos en el sur del continente africano y ninguno de los dos tiene salida al mar (Instituto del Tercer Mundo de Uruguay, 1999). Finalmente, según el mismo Informe sobre desarrollo humano 1998 (PNUD, 1998) los datos de la tasa de matriculación combinada para estos dos países provienen de esti-

maciones de la UNESCO y no de estadísticas oficiales. Podría entonces argumentarse que quizá estos valores se encuentran sobrestimados, contribuyendo así a que las observaciones se alejen del comportamiento general del grupo de países.

Como se dijo anteriormente, la decisión de excluir o no del análisis de regresión a estos tres países (Luxemburgo, Malawi y Zambia) no debe ser en modo alguno automática e irreflexiva. En realidad lo más aconsejable es comparar el modelo de regresión ajustado con y sin estas observaciones, para así establecer su grado de influencia sobre el poder explicativo de la ecuación de regresión. Este estudio procedió según esta recomendación.

#### LAS ECUACIONES DE REGRESIÓN Y SU INTERPRETACIÓN

El cuadro 3 muestra los resultados más importantes del ajuste de las ecuaciones de regresión bajo dos contextos: incluyendo y excluyendo los valores extremos u “outliers”.

Examinando los datos del cuadro 3 los resultados que llaman la atención de manera más evidente son los que se refieren a los valores de R<sup>2</sup> para ambos modelos. Esta medida, R<sup>2</sup>, se denomina coeficiente de determinación y se emplea para evaluar sustantivamente la capacidad predictiva del modelo de regresión. El valor de R<sup>2</sup> indica el porcentaje de la variabilidad en la variable dependiente que puede ser explicada a partir de su relación lineal con las variables independientes. En este caso se concluye que, para el primer modelo (incluyendo todos los 35 países de la muestra), el porcentaje de la variabilidad en la esperanza de vida que puede explicarse a partir de su relación lineal con el PIB *per cápita* y la matriculación combinada es aproximadamente 62%. Para el segundo modelo (aquel en donde se excluyen los tres países con valores extremos) se obtiene un valor para R<sup>2</sup> sustancialmente más alto, igual a 0,80. Esto significa que, en el segundo modelo, el 80% de la variabilidad en los valores observados de la esperanza de vida

CUADRO 3  
RESULTADOS DEL ANÁLISIS DE REGRESIÓN

Variable dependiente: esperanza de vida, Y	R <sup>2</sup>	Coeficientes de regresión		Coeficientes estandarizados		Significancia de los coeficientes	
		PIB per cápita X <sub>1</sub>	Tasa de matricu- lación X <sub>2</sub>	PIB per cápita X <sub>1</sub>	Tasa de matricu- lación X <sub>2</sub>	PIB per cápita X <sub>1</sub>	Tasa de matricu- lación X <sub>2</sub>
Primer modelo (incluye los 35 países)	0.616	5,77x10 <sup>-4</sup>	0.284	0.427	0.480	0.002	0.001
Segundo modelo (excluye a Luxemburgo, Malawi y Zambia)	0.804	3,35x10 <sup>-4</sup>	0.353	0.240	0.718	0.038	0.000

FUENTE: La autora, basada en PNUD (1998).

$$\text{MODELO GENERAL } Y = a_0 + a_1X_1 + a_2X_2$$

ECUACIÓN PARA EL PRIMER MODELO (incluye los 35 países):

$$\text{Esperanza de vida} = 43,628 + 5,77 \times 10^{-4} \times \text{PIB per cápita} + 0,284 \times \text{Tasa de matriculación}$$

ECUACIÓN PARA EL SEGUNDO MODELO (excluye a Luxemburgo, Malawi y Zambia):

$$\text{Esperanza de vida} = 42,358 + 3,35 \times 10^{-4} \times \text{PIB per cápita} + 0,353 \times \text{Tasa de matriculación}$$

*Nota:* La constante en las ecuaciones de regresión, 43,63 y 42,36 para el primer y segundo modelo respectivamente, se interpreta como el valor que tomaría la variable dependiente cuando ambas variables independientes son iguales a cero. En nuestro caso particular, estos valores no poseen interés sustantivo.

puede explicarse a partir de su relación lineal con el PIB *per cápita* y la tasa de matriculación. Es claro a partir de estos datos que, efectivamente, al excluir los tres casos que presentan comportamientos atípicos, el ajuste del modelo y su capacidad predictiva se hacen mayores.

En los contextos sociales y educativos valores de R<sup>2</sup> como los obtenidos se consideran bastante elevados. Entonces, se logra un grado relativamente alto de exactitud en la predicción si se usan el indicador de educación y el indicador de ingreso para estimar la esperanza de vida. Ambos modelos de regresión resultaron además altamente significativos a nivel estadístico ( $p < 0,01$ ), o sea que es muy poco probable que estos resultados se hayan presentado por azar.

Ahora bien, si analizamos los resultados individualmente para cada una de las dos variables independientes también se notan diferencias dramáticas entre los dos modelos. En primer lugar se pueden comparar los coeficientes de regresión, que son los coeficientes que se usan para formar la ecuación de regresión. En el caso de la variable PIB *per cápita* se obtiene un valor de  $5,77 \times 10^{-4}$  para el primer modelo y  $3,35 \times 10^{-4}$  para el segundo modelo. Debe recordarse que estos valores se interpretan en el mismo sentido que se interpreta la pendiente en una función lineal, es decir, nos indican en cuántas unidades cambia la variable dependiente cuando la variable independiente cambia en una unidad. Por tanto, en el primer modelo, la esperanza de vida aumenta en  $5,77 \times 10^{-4}$



= 0,000577 años cuando el PIB *per cápita* aumenta en un dólar anual. Para hacer este valor más fácil de interpretar se puede usar un factor de ampliación de 1000 para estas cifras y decir que la esperanza de vida aumenta, en promedio, 0,577 años cuando el PIB *per cápita* aumenta en 1000 dólares anuales. En el caso del segundo modelo, la esperanza de vida aumenta, en promedio, 0,335 años cuando el PIB *per cápita* aumenta en 1000 dólares anuales.

En cuanto a la tasa de matriculación combinada, se tiene en el primer modelo que la esperanza de vida se incrementa en 0,284 años cuando la tasa de matriculación combinada aumenta en una unidad (recuérdese que esta última variable es un porcentaje). Mientras, en el segundo modelo, la esperanza de vida aumenta, en promedio, 0,353 años cuando la tasa de matriculación se eleva en una unidad. Si usamos un factor de ampliación de 10 para estos valores podemos decir, alternativamente, que en el caso de la matriculación combinada un aumento de 10 puntos porcentuales estará asociado a un incremento promedio en la esperanza de vida de 2,84 y 3,53 años para el primer y segundo modelo respectivamente.

A partir de las interpretaciones anteriores podría argumentarse entonces, desde la perspectiva de políticas públicas de inversión, que es relativamente más realista para un país lograr un aumento de 10 puntos en su cobertura educativa que elevar su nivel de PIB *per cápita* en 1000 dólares anuales. Además estos aumentos en la cobertura educativa estarán asociados a mayores impactos sobre los niveles de la esperanza de vida. La evidencia empírica sugiere que es potencialmente más beneficioso invertir en programas que refuercen la cobertura educativa, si el objetivo es aumentar la longevidad de la población considerada esta como indicador de calidad de vida.

Otra manera de ilustrar este punto es discutiendo la importancia relativa de las dos variables independientes. Para esto se deben analizar los valores de los coeficientes de regresión estandarizados (o coeficientes *beta*) para cada una de las dos variables independientes. Estos coeficientes pueden compararse

directamente de manera válida, ya que la estandarización ha eliminado el efecto de las diferentes unidades de medida de las variables. El coeficiente *beta* nos dice en cuántas desviaciones estándar cambia la esperanza de vida cuando la variable independiente en cuestión cambia en una desviación estándar. A partir del ajuste del primer modelo y comparando los coeficientes estandarizados para ambas variables concluiríamos que ambas tienen un grado similar de importancia para la predicción de la esperanza de vida. Sin embargo, si analizamos los resultados del segundo modelo la interpretación cambia radicalmente, pues aquí se tiene que el coeficiente estandarizado para la tasa de matriculación (0,718) es prácticamente 3 veces el coeficiente para el PIB *per cápita* (0,240). Esto significa que la importancia relativa del indicador de educación en la predicción de la esperanza de vida es prácticamente tres veces la importancia relativa del indicador de poder adquisitivo.

## CONCLUSIONES

En este punto, quizá la pregunta más obvia que se genera en la mente del lector (o lectora) se refiere a cuál de los dos modelos de regresión es el más adecuado y el que mejor refleja la relación empírica entre las variables. Todo depende de cuál sea el contexto en que se están analizando e interpretando los datos. Si el interés es producir, por ejemplo, conclusiones generales que se refieran al comportamiento más común de las variables, que se manifiesta en la mayoría de las unidades de estudio, entonces parece evidente que la recomendación sería utilizar los resultados arrojados por el segundo análisis de regresión, pues a partir de la medida de bondad de ajuste, este modelo tiene claramente un mayor poder predictivo. Este enfoque podría tomarse en el caso de que interese alimentar la toma de decisiones en cuanto a inversión pública en programas de desarrollo. La evidencia arrojada por este segundo modelo de regresión establece de manera muy sólida que el indicador de educación

está más fuertemente asociado al indicador de longevidad que el indicador de poder adquisitivo. El mayor grado de asociación entre esperanza de vida y matriculación combinada sugiere que se pueden lograr cambios más significativos en la calidad de vida invirtiendo comparativamente más en programas que eleven la cobertura educativa.

Pero al hacer este tipo de interpretaciones no debe olvidarse que pueden existir en realidad algunas observaciones que se desvían del patrón general de comportamiento y que merecen ser analizadas separadamente. El estudio a profundidad de estos valores extremos u "outliers" podría incluso ayudar a enriquecer la teoría o ayudar a generar nuevas hipótesis acerca del fenómeno de interés.

#### BIBLIOGRAFÍA

- Brewer, James K. *Introductory Statistics for Researchers*. Edina, Minnesota, Burgess International Group Inc., Bellwether Press Division, 1991.
- Daniel, Wayne W. *Estadística con aplicaciones a las ciencias sociales y a la educación*. México D.F., Mc Graw-Hill, 1988.
- Hopkins, Kenneth D., B.R. Hopkins y Gene V. Glass. *Estadística básica para las ciencias sociales y del comportamiento*. Naucalpan de Juárez, Edo. de México, Prentice-Hall Hispanoamericana, 1997.
- Instituto del Tercer Mundo de Uruguay. *The World Guide 1999: An Alternative Reference to the Countries of our Planet*. Publicación en CD-ROM. Adelaide, Australia, New Internationalist Publications, 1999.
- Pagano, Robert R. *Estadística para las ciencias del comportamiento*. México D.F., Editorial Thomson, 1999.
- Programa de las Naciones Unidas para el Desarrollo (PNUD). *Informe sobre desarrollo humano 1998*. Madrid, Ediciones Mundi-Prensa, 1998.
- Stevens, James. *Applied Multivariate Statistics for the Social Sciences*. Hillsdale, New Jersey, Lawrence Erlbaum Associates, 1986.
- Pedhazur, Elazar J. *Multiple Regression in Behavioral Research: Explanation and Prediction*. Fort Worth Texas, Harcourt Brace College Publishers, 1982.

Eiliana Montero Rojas  
emontero@cariari.ucr.ac.cr