MODELO DESARROLLADO

A causa de que los datos fueron agrupados por clases, por ejemplo la edad, y puesto que se deseó que esas clases se subclasificaran de acuerdo al sexo -hombres y mujeres- las variables explicatorias no pueden ser designadas en una forma númerica ordinaria. Una forma modificada de la técnica de regresión múltiple, fue usada para facilitar el análisis de la clasificación de las variables explicatorias. El procedimiento incorpora el uso de variables independientes que toman el valor de uno y cero. 16/

Como ya hemos mencionado anteriormente, el ingreso medio es considerado la variable dependiente en el análisis. Un ingreso medio diferente, es asociado con una combinación de los sub-grupos de las variables independientes. Básicamente son cuatro clases de variables independientes: sexo, edad, educación y ocupación. Dentro de esas cuatro clases, existen 18 sub-clases definidas: dos para el sexo; seis para los grupos de edad, cinco para los niveles educativos, y siete para los grupos de ocupación. Para facilitar el análisis entre sexos, los datos de edad; educación y ocupación, fueron subdivididos en hombres y mujeres. Por lo tanto, son 12 las variables independientes para la edad (seis para cada sexo), 10 variables independientes para la educación (cinco para cada sexo) y 14 variables independientes para la ocupación (7 para cada sexo). Una parte representativa de la forma matricial del modelo general, sería como la que se ilustra a continuación.

16/ Guajarati, D. Basic Econometric, Mc Graw Hill Ed., 1978 pp. 287-321.

		EDAD												EDUCACION					
		HOMBRES						MUJERES						HOMBRES					MUJERES
INGRESO MEDIO	SEXO	12-24	25-34	35-44	45-54	55-64	65 Y MAS	12-24	25-34	35-44	45-54	55-65	65 Y MAS	01-06	07-09	10-11	12-15	16 Y MAS	
Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	
Y_1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	
Y_2	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	

Para identificar los sexos, mujeres= 0 y hombres = 1.

La primera hilera indica que un ingreso medio de Y_1 está asociado con las mujeres, en el grupo de edad de 25 a 34 años más algún nivel de educación y grupo de ocupación no mostrados. Del mismo modo la hilera dos muestra un ingreso Y_2 asociado con los hombres, en el grupo de edad de 45 a 54 años, con 10 a 11 años de educación y algún tipo de ocupación no mostrado.

Cuatro ecuaciones fueron examinadas usando este procedimiento estadístico. Primero, los datos fueron clasificados por sexo (X_1). Segundo, los datos fueron corridos en forma separada en las ecuaciones (1) y (2), solamente para hombres y solamente para mujeres. Tercero, todos los datos fueron corridos en una sola ecuación (3). Esto se hizo para permitir las comparaciones entre ambos sexos. Cuarto, se desarrolló una ecuación que combina a hombres y mujeres en datos específicos de las variables. Por ejemplo, los datos tanto de hombres como mujeres fueron integrados en un grupo de edad de 25 a 34 años, lo mismo se hizo con los restantes grupos - de edad, educación y ocupación, el resultado de este procedimiento fue cortar a la mitad el número de variables independientes. La ecuación (4) en conjunto con la (3) facilitaron probar una excesiva diferencia en los niveles de ingreso medio de los dos sexos, vía una prueba "F" para diferencias en la variación explicada de las dos ecuaciones.

Para probar los efectos de la edad, educación y ocupación como grupos independientes, las ecuaciones (1) y (2) fueron modificadas en tres etapas, dejando fuera la edad, educación y ocupación sucesivamente. Como resultado, se obtuvieron seis ecuaciones modificadas, tres para hombres y tres para mujeres, las cuales fueron probadas con respecto a la ecuación original usando de nuevo una prueba "F" para la diferencia en la variación explicada entre las ecuaciones.

Los modelos de las cuatro ecuaciones generales son:

$$(1) \quad Y_n = a_1 + \sum_{j=2}^{6} b_j X_j + \sum_{k=1}^{3,5} b_k X_k + \sum_{j=1}^{5,7} b_i X_i + U$$

$$(2) \quad Y_m = a_1 + \sum_{j'=2}^{6} b_{j'} X_{j'} + \sum_{k'=1}^{3,5} b_{k'} X_{k'} + \sum_{i'=1}^{5,7} b_{i'} X_{i'} + u$$

$$(3) \quad Y_{nm} = a_1'' + a_2'' X_n + \sum_{j=2}^{6} b_j X_j + \sum_{j'=2}^{6} b_j' X_j' + \sum_{k=1}^{3,5} b_k X_k + \sum_{k'=1}^{3,5} b_{k'} X_{k'} + \sum_{i=1}^{6} b_i X_i + \sum_{i'=1}^{6} b_{i'} X_{i'} + u$$

$$(4) \quad Y'_{nm} = a_1''' + \sum_{q=2}^{6} b_q X_q \sum_{m=1}^{3,5} b_m X_m + \sum_{p=1}^{6} b_p X_p + u$$

Donde:

Y_n = Ingreso medio de los hombres

Y_m = Ingreso medio de las mujeres

Y_{nm} = Ingreso medio de hombres y mujeres (será igual a Y_n o Y_m)

Y'_{nm} = Ingreso medio de hombres y mujeres, asociado a las variables transformadas.

X_n = Toma el valor de 1 si la hilera de la matriz se refiere a hombres y 0 si se refiere a mujeres.

j = Grupo de edad (hombres)

k = Nivel educativo (hombres)

i = Ocupación (hombres)

j' = Grupo de edad (mujeres)

k' = Nivel educativo (mujeres)

i' = Ocupación (mujeres)

q = Variable transformada para los grupos de edad (j + j')

m = Variable transformada para los grupos educativos (k + k')

p = Variable transformada para los grupos ocupación (i + i')

u = Error aleatorio.

En las ecuaciones generales se han omitido los valores siguientes: de j y j' = 1, de k y k' = 4 y de i y i' = 6 excepto en las ecuaciones (3) y (4) donde; i y i' = 7 es la variable omitida.

En el ejemplo de la matriz formada por el modelo, las variables omitidas corresponden a X_2, X_8 y X_{17}. Su eliminación es requerida para prevenir que la matriz sea singular.[17/] El término constante en la regre

17/ Cuando se usan variables cualitativas en modelos econométricos, la regla general es: "Si una variable tiene "M" categorías se introducen solamente "m-1" variables dummy". Si esta regla no es seguida, se cae en lo que se conoce como: "dummy-variable trap", esto es, una situación de perfecta multicolinealidad.

sión (a_1) es llamado el "coeficiente de intercepción diferencial" porque explica la diferencia en el ingreso proveniente de la supresión de las variables antes mencionadas. Del mismo modo, las variables dentro de un grupo en particular, como por ejemplo de edad, tienen su base de comparación del nivel de ingreso medio con respecto al grupo de edad específico que es incorporado en el término constante (a_1). Específicamente, los coeficientes de regresión de las variables independientes suprimidas, son iguales a cero. Los coeficientes de regresión de las restantes variables independientes son leídos en términos de pesos semanales y son aditivos. Como un ejemplo, consideremos lo siguiente (los datos son tomados del Cuadro 2.5 para mujeres).

Término constante	a =	\$ 913.04	
j' = edad (25-34)	$b_{j'}$ =	270.21	
j' = educación (07-09)	$b_{k'}$ =		-380.45
k' = ocupación (oficinistas)	$b_{i'}$ =	346.16	
		\$ 1,529.41	-380.45
Ingreso esperado (Y_m)... =			\$ 1,148.96

El término constante "a" es, en efecto, el ingreso medio que resulta de la omisión de alguno de los sexos, grupos de edad, educación y ocupación. Los valores de "b" indican la contribución al ingreso de alguna variable relativa a la base sobre la cual es calculada. En nuestro ejemplo, un grado de educación comercial o de secundaria, podríamos esperar que tuviera un ingreso medio por semana, de 380 pesos menos que el que pudiera atribuirse al grupo de educación base, el cual en el modelo empleado es de 12 a 15 años, correspondiente a un nivel de normal o universitario.

Los grupos base incorporados en el término constante son: grupo de edad 12-24 años, nivel educativo 12-15 años. Las ecuaciones que tratan en forma separada a hombres y mujeres, usan la categoría ocupacional "operarios y artesanos en la industria" como base, mientras que las ecuaciones que integran los datos de hombres y mujeres, usan la categoría "otros trabajadores y servicios" como grupo base. El cambio del grupo base facilita la comparación entre sexos, de la categoría "operarios y artesanos en la industria".

Un supuesto primordial de este enfoque estadístico es que todas las categorías tienen la misma ponderación, es decir, las categorías ocupacionales, estructura de los grupos de edad y la estructura de la educación se supone que son idénticos entre y dentro de los sexos (igual número de individuos en cada categoría). Es obvio que estos supuestos no son reales y por lo tanto los resultados del análisis deben ser interpretados con cautela. Sin embargo, la característica de igual ponderación de este proceso estadístico tiene la ventaja de indicar el efecto "puro" de la edad, educación y ocupación sobre el ingreso dentro de cada sexo, en el sentido de que un coeficiente de regresión para un nivel de educación de 6 a 7 años, por ejemplo, indicará la contribución de ese grado de educación (puede ser de secundaria o comercial) al ingreso, sin tomar en cuenta el número de individuos que hayan alcanzado o no ese nivel educativo.