CAPÍTULO 4-2
EL MODELO LINEAL GENERAL Y LA REGRESIÓN MÚLTIPLE APLICADOS AL ANÁLISIS DE VARIANZA[192]

Vimos que el análisis de regresión es un método estadístico, el cual se aplica cuando un modelo teórico propone una relación entre una variable dependiente continua y una o más variables independientes continuas o discretas. En cuanto al análisis de varianza (analisis of variance, ANOVA), se aplica cuando las variables independientes son todas discretas. Sus características sobresalientes son por lo tanto:

- Un modelo del tipo estímulo-reacción.
- Una reacción que se mide por medio de variables continuas.
- Unos estímulos que se miden por medio de variables discretas.[193]

(En caso de medir los estímulos por medio de variables, siendo algunas discretas y otras continuas, se trata de aná-

[192] Wonnacott y Wonnacott (1992, pp. 503-507); Iman y Conover (1989, caps. 16-17).
[193] Dependiendo del contexto y de la disciplina, se usa diferentes términos para designar las variables independientes, como factores, efectos, categorías, variables cualitativas, variable de clasificación (clasification variables), etc.

lisis de covarianza y es necesario, entonces, usar el modelo lineal general.)

La pregunta de investigación que se propone usualmente es: ¿es la reacción al estímulo significativamente diferente entre categorías?

En el caso del análisis de varianza, existen procedimientos específicos y formatos estándares de presentación de los resultados. Sin embargo, es posible efectuar de manera equivalente un análisis de varianza por medio de la regresión lineal. Esto trae consigo algunas ventajas. Primero, y esto no es el caso del análisis de varianza, el análisis de regresión no impone restricciones en cuanto al plan de muestreo (número de observaciones por categorías), Segundo, es posible combinar el análisis de la varianza con variables independientes continuas (como ya mencionamos, este tipo de modelo es, a veces, conocido como un modelo de análisis de covarianza). Finalmente, el análisis de regresión ofrece una más amplia flexibilidad en cuanto a las hipótesis que se pueden someter a tests estadísticos.

4-2.1 Un ejemplo

En el marco de la construcción de una matriz de contabilidad social para Quebec, Robichaud *et al*. (1998) estudiaron el ahorro de los hogares de Quebec a partir de los datos contenidos en el archivo de micro-datos de gran difusión de la encuesta de *Statistique Canada* sobre los gastos de las familias en 1992. El archivo contiene 1900 observaciones para Quebec.

Se obtuvieron las seis variables, que presentamos a continuación, del archivo de microdatos de gran difusión de la encuesta de *Statistique Canada* sobre los gastos de las familias en 1992:

1. Composición del hogar
   - Personas solas.

- Parejas[194] sin hijos.
- Parejas con hijos.[195]
- Familias monoparentales.
- Otros hogares.[196]

2. Número de hijos menores de 16 años.
3. Edad de la persona de referencia.[197]
4. Ingreso del hogar con impuestos deducidos.
5. Variación neta del activo y del pasivo.
6. Seguro.[198]

La formulación del modelo lineal cuyos parámetros se estimarán, se basa en el modelo conceptual siguiente. El monto del ahorro de un hogar (variable dependiente) aumenta con el ingreso con impuestos deducidos, pero depende de la edad del hogar (hipótesis del ciclo de vida) y de la presencia de hijos (gastos más elevados); es posible también que el ahorro se vea influenciado por el hecho de que la responsabilidad del hogar esté en manos de una sola persona (lo que implica, por lo general, que hay un solo ingreso y que, en la mayoría de los casos, no existen más adultos que mantener); además, queremos verificar que la categoría heteróclita de los “Otros hogares” es diferente.

---

[194] Casados o juntados.
[195] Con relación a la composición del hogar, nos referimos a hijos de cualquier edad, nunca casados y que viven bajo el mismo techo que sus padres.
[196] Esta categoría contiene las parejas sin hijos que viven con un familiar que no es su hijo, así como los hogares donde vive por lo menos una persona que no es familiar de la “persona de referencia” (vea la nota siguiente). En particular, encontramos en esta categoría heteróclita, los hogares sin hijos con inquilinos y los grupos de estudiantes que comparten un departamento.
[197] En la encuesta sobre los gastos de las familias, la “persona de referencia” es el miembro del hogar que el contestador designa como el principal sostén financiero, lo que corresponde, normalmente, a la persona con el ingreso más elevado.
[198] Primas de seguros de vida, etcétera.

Es importante aclarar que la selección y la definición de las variables independientes fueron dictadas con el objetivo de construir una matriz de contabilidad social, y que el modelo que presentamos aquí, no debe considerarse como un ejemplo de un modelo de comportamiento de ahorro de los hogares. Lo único importante que se debe inferir de este ejemplo, es la manera de tratar las variables independientes categóricas en la regresión lineal. Añadamos que la presencia del ingreso entre las variables independientes vuelve imposible aplicar a este modelo un análisis de la varianza clásico (al menos que cambiemos el ingreso por una variable categórica), lo que, nuevamente, demuestra la más grande polivalencia del análisis de regresión.

Definimos entonces la variable dependiente:

AHORRO = Variación neta del activo y del pasivo más seguros.

Las variables independientes son:

REVAPIMP = Ingreso del hogar con impuestos deducidos.

Edad.

Composición del hogar.

Veremos luego cómo fueron especificadas las variables de edad y de composición del hogar. Combinando las variables categóricas, obtenemos una repartición de las 1900 observaciones según la composición del hogar y la edad de la persona de referencia. Se presenta esta repartición en la tabla que sigue, en la cual podemos observar que la repartición no se conforma a un plan de muestreo “equilibrado” (con el mismo número de observaciones en cada celda) ni tampoco a un plan de muestreo biproporcional. En estas condiciones, efectuar un análisis de varianza clásico se revelaría muy difícil; sin embargo, el uso de la regresión múltiple no encierra restricciones semejantes en cuanto a la estructura de los datos. No obstante, es importante notar que algunas celdas no tienen más que un número pequeño de observaciones, lo que nos pide actuar con mucha prudencia al momento de interpretar los resultados.

Composición del hogar	Menos de 35	35-45	45-65	65 y más	Total
Personas solas	101	77	132	142	452
Parejas sin hijos	95	49	177	133	454
Parejas con hijos	163	267	239	26	695
Familias mono.	30	71	48	10	159
Otros sin hijos	24	17	41	21	103
Otros con hijos	8	16	12	1	37
Total	421	497	649	333	1900

4-2.1.1 Variables independientes de edad

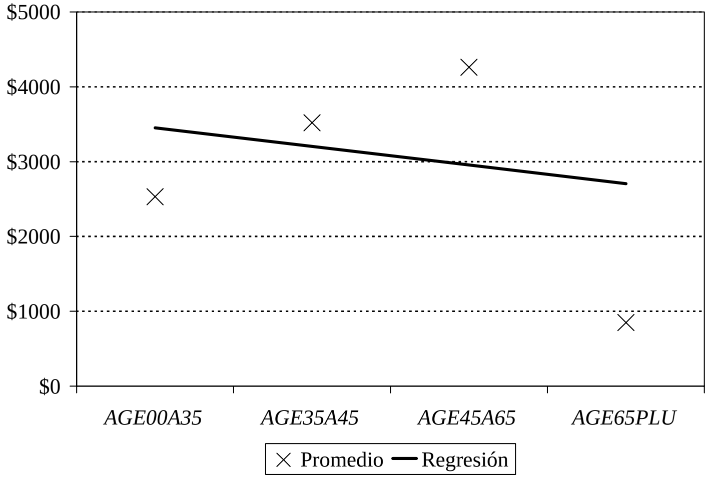
En primer lugar, definimos la variable independiente *GROUPAGE*:

Edad de la persona de referencia	Valor de la variable *GROUPAGE*
Menos de 35 años	1
35 años o más y menos de 45 años	2
45 años o más y menos de 65 años	3
65 años y más	4

La variable *GROUPAGE* es una variable ordinal de orden incompleto (ver cap. 1-1).

Pero esta variable no puede entrar tal cual en la regresión. ¿Por qué? Porque impondría artificialmente una relación lineal entre el grupo de edad y el ahorro, lo cual es contrario a los hechos, como se puede ver en el siguiente gráfico.

Valor promedio del ahorro por grupo de edad
y regresión lineal sobre *GROUPAGE*



Por tanto, no podemos especificar el modelo antes de haber reemplazado la variable GROUPAGE por una serie de variables dicotómicas, puesto que al emplearla tal cual como variable independiente, estaríamos diciendo que el ahorro aumenta (o disminuye) de manera lineal con la categoría edad. Por esta razón, se crean las cuatro variables dicotómicas siguientes:

7. *AGE0A35* = 1 si *GROUPAGE* = 1 (edad < 35); = 0 de otra manera.
8. *AGE35A45* = 1 si *GROUPAGE* = 2 (edad ≥ 35 y < 45); = 0 de otra manera.
9. *AGE045A65* = 1 si *GROUPAGE* = 3 (edad ≥ 45 y < 65); = 0 de otra manera.
10. *AGE65PLU* = 1 si *GROUPAGE* = 4 (edad ≥ 65); = 0 de otra manera.

4-2.1.2 Variables independientes de composición del hogar

La composición del hogar es una variable categórica politómica. Para poder distingüir entre hogares "otros" con y sin hijos, usaremos la información complementaria dada por el número de hijos menores de 16 años. Obtendremos así 6 tipos de hogares:

- Personas solas.
- Parejas sin hijos.
- Parejas con hijos.
- Familias monoparentales.
- Otros hogares sin hijos.
- Otros hogares con hijos.

Sin embargo, hay que reemplazar la variable de composición del hogar, tal como la de edad, por una serie de variables dicotómicas, y por las mismas razones. Es aún más necesario, dado que la composición del hogar no es una variable ordinal (cuando sí lo es *GROUPAGE*).

Presentamos a continuación dos formas de modelización, una que usa 5 variables y otra que usa 3. Ambas formas son representadas por los árboles de clasificación correspondientes.

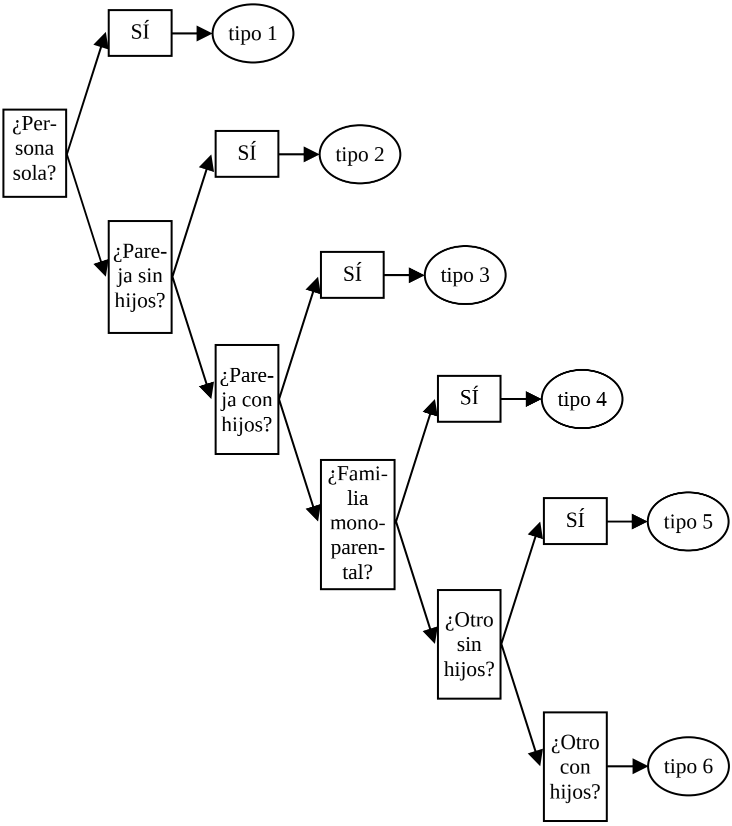
La clasificación con 5 variables consiste sencillamente en definir tantas variables dicotómicas como hay tipos de hogares, sin quitar una porque es redundante (vamos a profundizar eso luego). En cuando al método con 3 variables, se constituye por las siguientes variables dicotómicas:

11. *SEULMONO* = 1 para una persona sola o una familia monoparental; *SEULMONO* = 0 de otra manera
12. *AUTRE* = 1 si el hogar pertenece a la categoría "Otro"; *AUTRE* = 0 de otra manera
13. *ENFANTS* = 1 si el hogar cuenta, al menos, con un hijo menor de 16 años; *ENFANTS* = 0 de otra manera

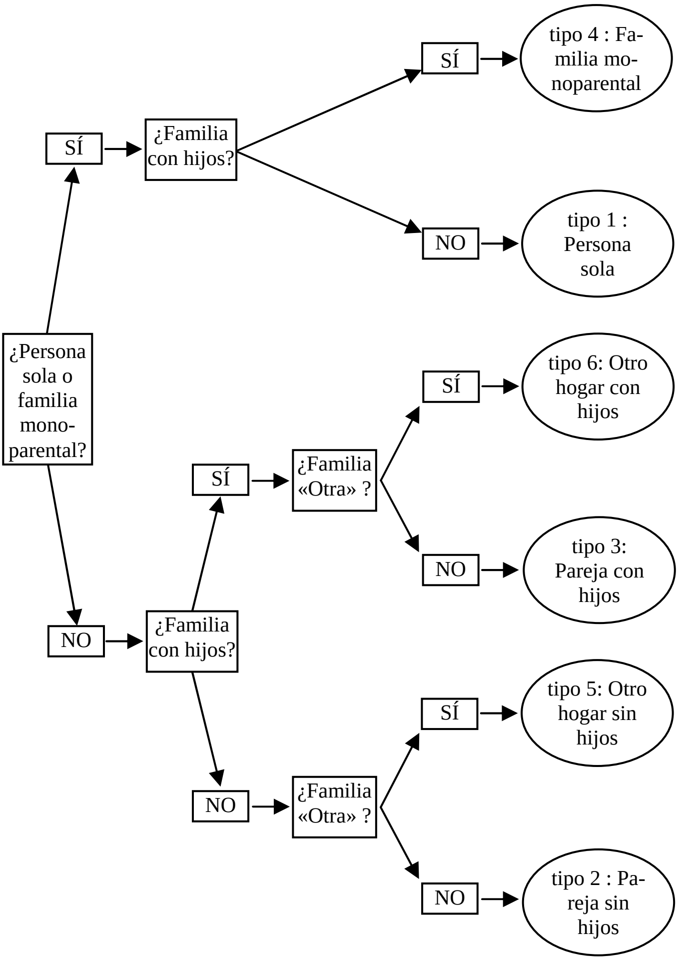
Por hijos entendemos hijos de cualquier edad, nunca casados y que viven con sus padres, salvo para los hogares

“*AUTRE*”, para los cuales *ENFANTS* significa solamente hijos menores de 16 años.

Dos árboles de clasificación con variables dicotómicas
I – Clasificación con 5 variables



Dos árboles de clasificación con variables dicotómicas
II – Clasificación con 3 variables



La que usamos fue la de 3 variables. Combinando esas tres variables se obtiene la clasificación que sigue:

Composición de los hogares	Número de hijos	Valor de la variable SEULMONO	Valor de la variable AUTRE	Valor de la variable ENFANTS
Personas solas	0	1	0	0
Parejas sin hijos	0	0	0	0
Parejas con hijos	> 0	0	0	1
Familias mono.	> 0	1	0	1
Otros hogares	0	0	1	0
	> 0	0	1	1

Se ve en este cuadro que cada tipo de hogar corresponde a una combinación única de las variables dicotómicas.

¿Cuál es la diferencia entre los dos esquemas de clasificación? De alguna manera, el esquema que escogimos impone una cierta coherencia en el modelo. Por ejemplo, con la tripleta *SEULMONO*, *ENFANTS* y *AUTRE*, el efecto de tener hijos debe ser el mismo independientemente de las demás características del hogar. Esto implica, por lo tanto, restricciones para el modelo. Sin embargo, veremos cómo se pueden evitar estas restricciones con la introducción de variables de interacción (vea 4-2.4).

4-2.2 Eliminación de la redundancia entre las variables independientes

No se deben incluir las cuatro variables dicotómicas de edad juntas entre las variables independientes, porque una de estas variables es redundante. En efecto, si *AGE0A35* = 0 y *AGE45A65* = 0 y *AGE65PLU* = 0, entonces forzosamente *AGE35A45* = 1, es decir, generalizando, que si para una observación dada, tres de las cuatro variables toman el valor cero, la cuarta toma necesariamente el valor 1. Por consiguiente, es necesario descartar una de las variables del modelo; en estas condiciones, el caso que corresponde a la variable descartada llega a ser el caso de referencia. En nuestro ejemplo, escogemos el grupo de edad de 35 a 45 como caso de referencia.

> De manera formal, al incluir las cuatro variables, estaríamos violando la condición H4 del modelo clásico de la regresión lineal puesto que su suma es siempre igual a 1, es decir igual a la constante del modelo:
>
> *AGE0A35* + *AGE35A45* + *AGE45A65* + *AGE65PLU* = 1 = *CONSTANTE*

Es importante observar que, al momento de definir las variables dicotómicas *ENFANTS*, *SEULMONO* y *AUTRE*, eliminamos, de manera implícita, las variables redundantes. En efecto, evitamos definir dos variables correspondientes a una por categoría. Por ejemplo, hubiéramos podido definir

- *AVECENFANTS* = 1 si el hogar cuenta con, por lo menos, un hijo;

  AVECENFANTS = 0 de otra manera;
- *SANSENFANTS* = 0 si el hogar cuenta con, por lo menos, un hijo;

  SANSENFANTS = 1 de otra manera.

No hicimos tal cosa puesto que una de estas dos variables hubiera sido redundante.

Vimos en el apartado 4-2.1 cómo se reemplazaba la variable de composición por la tripleta *SEULMONO*, *ENFANTS* y *AUTRE*. En caso de que hubiéramos querido utilizar el otro esquema de clasificación, tendríamos, por las mismas razones por las cuales se efectuó en el caso de la variable *GROUPAGE*, que haber eliminado la redundancia. Es por eso que el primer esquema no cuenta con 6 variables dicotómicas, sino con 5.

4-2.3 ESPECIFICACIÓN DE UN MODELO SIN INTERACCIÓN

Estamos listos ahora para enunciar una primera especificación del modelo:

$$\begin{aligned} EPARGNE = {} & \beta_1 + \beta_2\, REVAPIMP + \beta_3\, SEULMONO \\ & + \beta_4\, AUTRE + \beta_5\, ENFANTS + \beta_6\, AGE00A35 \\ & + \beta_7\, AGE45A65 + \beta_8\, AGE65PLU \end{aligned}$$

Donde es posible notar la ausencia de la variable *AGE35A45* que sería redundante.

Veamos, ahora, lo que este modelo significa para cada una de las 24 posibilidades que alojaron nuestros datos. Los 24 casos se presentan en la siguiente tabla.

Podemos notar en esta tabla que a cada uno de los 24 casos posibles corresponde una combinación única de valores de las variables dicotómicas; esto muestra que no faltan variables puesto que cada caso tiene una representación distinta. Observamos también que el caso de referencia cuando todas las variables dicotómicas son nulas, corresponde a una pareja sin hijos cuya persona de referencia tiene entre 35 y 45 años. Se deduce que los coeficientes de las variables dicotómicas representan las diferencias con relación a este caso de referencia; por ejemplo, el modelo predice que entre el hogar de referencia y una familia monoparental cuya persona de referencia tiene menos de 35 años, teniendo los dos hogares el mismo ingreso, la diferencia será igual a $\beta_3 + \beta_5 + \beta_6$. Podríamos pensar que cada uno de estos tres coeficientes sea negativo; sin embargo, sólo la estimación del modelo podrá aclarar este hecho.

Interpretación del modelo sin variables de interacción

Grupo de edad de la persona de referencia	SEULMONO	AUTRE	ENFANTS	AGE00A35	AGE45A65	AGE65PLU	AHORRO predicho por el modelo
Personas solas (Número de hijos = 0)							
<35	1	0	0	1	0	0	$\beta_1 + \beta_2\, REVAPIMP + \beta_3 + 0 + 0 + \beta_6 + 0 + 0$
≥35 y <45	1	0	0	0	0	0	$\beta_1 + \beta_2\, REVAPIMP + \beta_3 + 0 + 0 + 0 + 0 + 0$
≥45 y <65	1	0	0	0	1	0	$\beta_1 + \beta_2\, REVAPIMP + \beta_3 + 0 + 0 + 0 + \beta_7 + 0$
≥65	1	0	0	0	0	1	$\beta_1 + \beta_2\, REVAPIMP + \beta_3 + 0 + 0 + 0 + 0 + \beta_8$
Parejas sin hijos (Número de hijos = 0)							
<35	0	0	0	1	0	0	$\beta_1 + \beta_2\, REVAPIMP + 0 + 0 + 0 + \beta_6 + 0 + 0$
≥35 y <45	0	0	0	0	0	0	$\beta_1 + \beta_2\, REVAPIMP + 0 + 0 + 0 + 0 + 0 + 0$
≥45 y <65	0	0	0	0	1	0	$\beta_1 + \beta_2\, REVAPIMP + 0 + 0 + 0 + 0 + \beta_7 + 0$
≥65	0	0	0	0	0	1	$\beta_1 + \beta_2\, REVAPIMP + 0 + 0 + 0 + 0 + 0 + \beta_8$
Parejas con hijos (Número de hijos > 0)							
<35	0	0	1	1	0	0	$\beta_1 + \beta_2\, REVAPIMP + 0 + 0 + \beta_5 + \beta_6 + 0 + 0$
≥35 y <45	0	0	1	0	0	0	$\beta_1 + \beta_2\, REVAPIMP + 0 + 0 + \beta_5 + 0 + 0 + 0$
≥45 y <65	0	0	1	0	1	0	$\beta_1 + \beta_2\, REVAPIMP + 0 + 0 + \beta_5 + 0 + \beta_7 + 0$
≥65	0	0	1	0	0	1	$\beta_1 + \beta_2\, REVAPIMP + 0 + 0 + \beta_5 + 0 + 0 + \beta_8$

Continua...

Interpretación del modelo sin variables de interacción (continuación)

Grupo de edad de la persona de referencia	SEULMONO	AUTRE	ENFANTS	AGE00A35	AGE45A65	AGE65PLU	AHORRO predicho por el modelo
Familias monoparentales (Número de hijos > 0)							
<35	1	0	1	1	0	0	$\beta_1 + \beta_2$ *REVAPIMP* $+ \beta_3 + 0 + \beta_5 + \beta_6 + 0 + 0$
≥35 y <45	1	0	1	0	0	0	$\beta_1 + \beta_2$ *REVAPIMP* $+ \beta_3 + 0 + \beta_5 + 0 + 0 + 0$
≥45 y <65	1	0	1	0	1	0	$\beta_1 + \beta_2$ *REVAPIMP* $+ \beta_3 + 0 + \beta_5 + 0 + \beta_7 + 0$
≥65	1	0	1	0	0	1	$\beta_1 + \beta_2$ *REVAPIMP* $+ \beta_3 + 0 + \beta_5 + 0 + 0 + \beta_8$
Otros hogares sin hijos (Número de hijos = 0)							
<35	0	1	0	1	0	0	$\beta_1 + \beta_2$ *REVAPIMP* $+ 0 + \beta_4 + 0 + \beta_6 + 0 + 0$
≥35 y <45	0	1	0	0	0	0	$\beta_1 + \beta_2$ *REVAPIMP* $+ 0 + \beta_4 + 0 + 0 + 0 + 0$
≥45 y <65	0	1	0	0	1	0	$\beta_1 + \beta_2$ *REVAPIMP* $+ 0 + \beta_4 + 0 + 0 + \beta_7 + 0$
≥65	0	1	0	0	0	1	$\beta_1 + \beta_2$ *REVAPIMP* $+ 0 + \beta_4 + 0 + 0 + 0 + \beta_8$
Otros hogares con hijos (Número de hijos > 0)							
<35	0	1	1	1	0	0	$\beta_1 + \beta_2$ *REVAPIMP* $+ 0 + \beta_4 + \beta_5 + \beta_6 + 0 + 0$
≥35 y <45	0	1	1	0	0	0	$\beta_1 + \beta_2$ *REVAPIMP* $+ 0 + \beta_4 + \beta_5 + 0 + 0 + 0$
≥45 y <65	0	1	1	0	1	0	$\beta_1 + \beta_2$ *REVAPIMP* $+ 0 + \beta_4 + \beta_5 + 0 + \beta_7 + 0$
≥65	0	1	1	0	0	1	$\beta_1 + \beta_2$ *REVAPIMP* $+ 0 + \beta_4 + \beta_5 + 0 + 0 + \beta_8$

Por otro lado, al observar la tabla, entendemos mejor por qué hubiera sido absurdo incluir en el modelo la variable politómica *GROUPAGE*.

Se presentan los resultados de la estimación en la tabla siguiente.

Variable	Descripción	Símbolo	Coeficiente estimado	*t* de Student	Probabildad crítica
CONSTANTE		β_1	–7727	–11.062	0.0001
REVAPIMP	Ingreso después de impuestos	β_2	0.340	28.468	0.0001
ENFANTS	Presencia de hjos	β_5	–2260	–4.937	0.0001
SEULMONO	Persona sola o monoparentale	β_3	1903	3.834	0.0001
AUTRE	Hogar « Otro »	β_4	–2578	–3.309	0.0010
AGE00A35	Edad 00-35	β_6	258	0.444	0.6574
AGE45A65	Edad 45-65	β_7	419	0.796	0.4263
AGE65PLU	Edad 65+	β_8	875	1.322	0.1862

$n = 1900$
$R^2 = 0.33$

En particular, constatamos que los coeficientes de las variables correspondientes a la edad no son, de manera significativa, diferentes de cero. ¿Debemos entender, por lo tanto, que la edad no tiene efecto sobre el comportamiento de ahorro?

4-2.4 Introducción de los efectos de interacción

El modelo que se presentó en la tabla anterior no toma en cuenta la posibilidad de efectos de interacción. Existe un gran número de interacciones posibles. Por tanto, no se encuentra a menudo que un modelo las contenga todas.

Por ejemplo, el modelo predice que el efecto sobre el ahorro de la presencia de hijos es igual a β_5, independientemente de la edad de la persona de referencia y de la composición del hogar. ¿Así sucede en la realidad? En otras palabras, ¿no habrá alguna interación entre la variable *ENFANTS* y las variables *SEULMONO*, *AUTRE*, *AGE0A35*, *AGE45A65* y *AGE65PLU*? Es importante entender que cada uno de los efectos de interacción que evocamos en la frase anterior es simétrico; por ejemplo, en lugar de preguntarse si el efecto de la presencia de hijos (*ENFANTS*) cambia con pertenecer al grupo de los menores de 35 años (*AGE0A35*), es posible preguntarse de manera equivalente si el efecto de pertenecer al grupo de los menores de 35 años cambia con la presencia de hijos.

Para poder incluir la posibilidad de interacción en el modelo, es necesario agregar variables a las ocho que ya tiene el modelo, Así, se define:

9. *MONOMONO* = 1
   si *ENFANTS* = 1 y *SEULMONO* = 1
   MONOMONO = 0 de otra manera

Para ser más conciso, se define matemáticamente[199]

$$MONOMONO = ENFANTS \times SEULMONO$$

De la misma manera, tenemos

10. AUTRENFA = ENFANTS × AUTRE
11. *ENFA0035 = ENFANTS* × *AGEOA35*
12. *ENFA4565 = ENFANTS* × *AGE45A65*
13. *ENFA65PL = ENFANTS* × *AGE65PLU.*
14. *AUTAOO35 = AUTRE* × *AGEOOA35.*
15. *AUTA4565 = AUTRE* × *AGE45A65.*

[199] Las variables dicotómicas son variables lógicas o variables de Boole. En álgebra booleania, la conjunción “y” se representa con la multiplicación.

16.*AUTA65PL* = *AUTRE* × *AGE65PLU*.

17.*SOLA0035* = *SEULMONO* × *AGE00A35*.

18.*SOLA4565* = *SEULMONO* × *AGE45A65*.

19.*SOLA65PL* = *SEULMONO* × *AGE65PLU*.

En esta lista podemos notar que no se incluyeron todas las variables posibles (por ejemplo, no hay ninguna variable de interacción con *AGE35A45*) porque, al igual que para otros grupos de variables categóricas, en el caso de las variables de interacción el hecho de incluir en el modelo todas las variables posibles implica redundancia.

Se interpretan los coeficientes de las variables de interacción como unas diferencias. Por ejemplo, vimos en la tabla del apartado 4-2.3 como β_5, el coeficiente de la variable *ENFANTS*, representaba la diferencia, en cuanto al monto del ahorro, entre dos hogares idénticos en todo menos en la presencia de hijos; igualmente, β_7, el coeficiente de la variable *AGE45A65* representa la diferencia, en cuanto al ahorro, entre dos hogares idénticos en todo menos en la edad, puesto que uno pertenece al grupo de edad de referencia (35-45 años) y el otro al grupo de los 45-65 años. En ausencia de variables de interacción, estas diferencias se suman; por ejemplo, el modelo descrito en el apartado 4-2.3 predice que entre un hogar sin hijos del grupo 35-45 años y un hogar con hijos del grupo 45-65 años, la diferencia será igual a $\beta_5 + \beta_7$. Si agregamos a este modelo la variable de interacción *ENFA4565*, esta diferencia será entonces igual a $\beta_5 + \beta_7$, más el coeficiente de la variable de interacción *ENFA4565*.[200]

[200] Es posible efectuar una analogía con la farmacología: el efecto de una combinación de dos medicamentos puede implicar efectos muy diferentes que los efectos de cada uno de estos medicamentos empleados solos. Los medicamentos juntados pueden reforzarse mutuamente o, por lo contrario, anularse el uno al otro.

Además, puede suceder que haya interacción entre una variable categórica y una variable continua. Así, el modelo predice que, independientemente de las características del hogar, un alza de un dólar del ingreso con impuestos retenidos repercutirá en un alza del ahorro de β_2 dólares. ¿Podría este efecto ser diferente para los hogares con hijos? Con el fin de examinar este problema, es necesario incluir, en el modelo, unas variables de interacción. Consideremos, por lo tanto, las tres variables suplementarias siguientes:

20. $REVENFAN = REVAPIMP \times ENFANTS$

21. $REVSELMO = REVAPIMP \times SEULMONO$

22. $REVAUTRE = REVAPIMP \times AUTRE$

Los coeficientes de estas variables se pueden interpretar también como diferencias. Por ejemplo, si comparamos dos hogares idénticos menos en la presencia de hijos, el coeficiente *REVENFAN* representa la diferencia entre los dos hogares en cuanto a su propensión marginal para ahorrar.

Después de incluir unas variables de interacción, el modelo completo se enuncia de la manera siguiente:

$$
\begin{aligned}
EPARGNE = {} & \beta_1 + \beta_2\ REVAPIMP \\
& + \beta_3\ SEULMONO + \beta_4\ AUTRE + \beta_5\ ENFANTS \\
& + \beta_6\ AGE00A35 + \beta_7\ AGE45A65 + \beta_8\ AGE65PLU \\
& + \gamma_1\ MONOMONO + \gamma_2\ AUTRENFA \\
& + \gamma_3\ ENFA0035 + \gamma_4\ ENFA4565 + \gamma_5\ ENFA65PL \\
& + \gamma_6\ AUTA0035 + \gamma_7\ AUTA4565 + \gamma_8\ AUTA65PL \\
& + \gamma_9\ SOLA0035 + \gamma_{10}\ SOLA4565 + \gamma_{11}\ SOLA65PL \\
& + \alpha_1\ REVENFAN + \alpha_2\ REVSELMO + \alpha_3\ REVAUTRE
\end{aligned}
$$

Después de ejecutar el procedimiento backward para eliminar las variables cuyos coeficientes no son significativos, obtenemos los resultados que se presentan en la tabla que sigue.

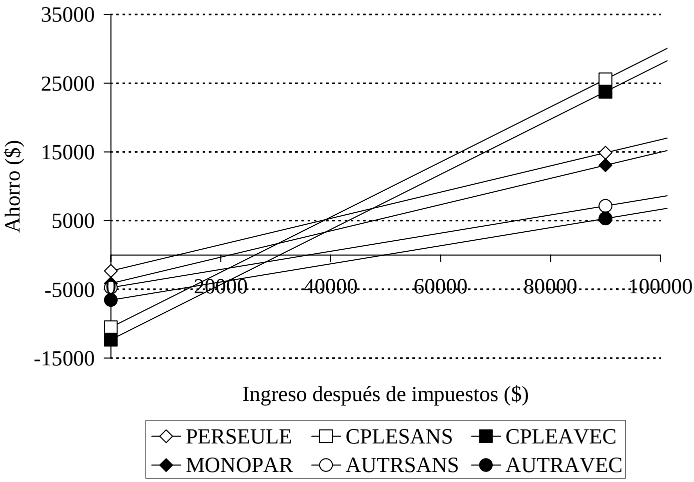
Variable	Descripción	Símbolo	Coeficiente estimado	Error Estandar	Probabilidad crítica
CONSTANTE		β_1	–10487	729	0.0001
REVAPIMP	Ingresos después de impuestos	β_2	0.400	0.013	0.0001
ENFANTS	Presencia de hijos	β_5	–1927	561	0.0006
SEULMONO	Pers. sola o monoparental	β_3	8233	1000	0.0001
AUTRE	Hogar "Otro"	β_4	7969	1771	0.0001
AGE45A65	Edad 45-65	β_7	1767	779	0.0234
AGE65PLU	Edad 65+	β_8	1513	770	0.0497
Variables de interacción					
ENFA4565	*ENFANTS* × *AGE45A65*	γ_4	–1506	897	0.0932
SOLA4565	*SEULMONO* × *AGE45A65*	γ_{10}	–1983	1008	0.0494
SOLA65PL	*SEULMONO* × *AGE65PLU*	γ_{11}	–1996	1138	0.0796
REVSELMO	*REVAPIMP* × *SEULMONO*	α_2	–0.211	0.030	0.0001
REVAUTRE	*REVAPIMP* × *AUTRE*	α_3	–0.297	0.046	0.0001

El número de observaciones es de 1900 y el coeficiente de determinación múltiple R^2 es de 0.36.

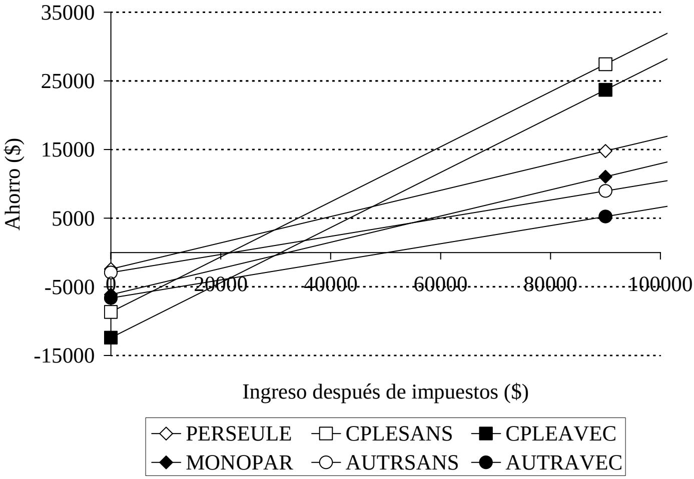
No es fácil concluir algo claro con todos estos coeficiente y hemos de preguntarnos qué significan realmente. Las figuras que presentamos a continuación ilustran las predicciones del modelo.

Leyenda	
Personas solas	PERSEULE
Parejas sin hijos	CPLESANS
Parejas con hijos	CPLEAVEC
Familias monoparentales	MONOPAR
Otros hogares sin hijos	AUTRSANS
Otros hogares con hijos	AUTRAVEC

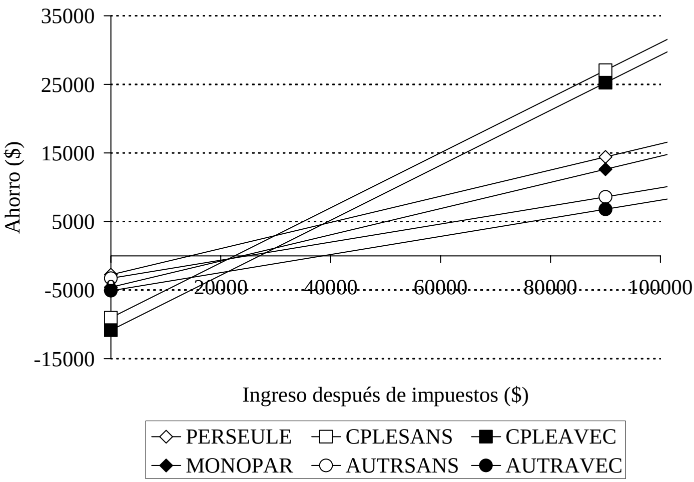
Ahorro de los menos de 45 años



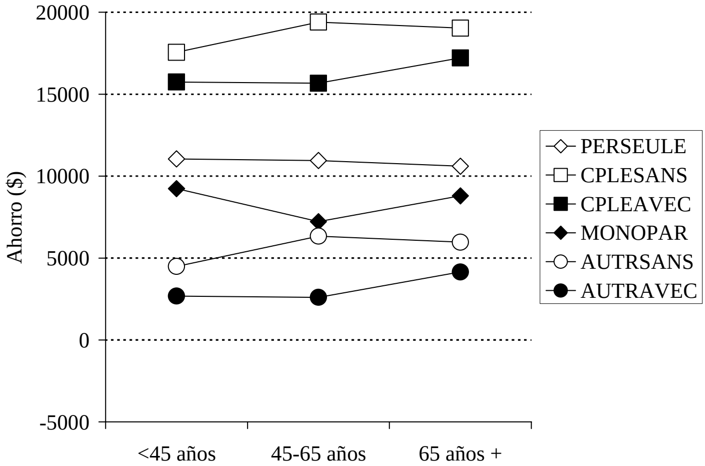
Ahorro de los 45-65 años



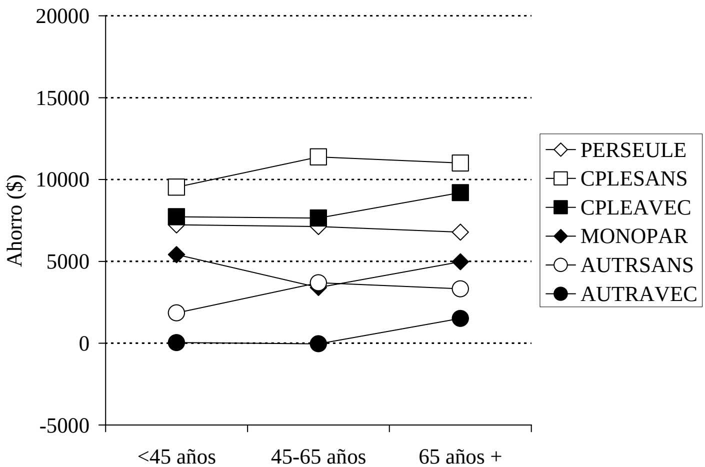
Ahorro de los 65 años y más



Ahorro según el grupo de edad, con un ingreso de $70,000



Ahorro según el grupo de edad, con un ingreso de $50,000



Ahorro según el grupo de edad, con un ingreso de $25,000