Tarea 4

JESUS ALEXIS SANCHEZ MORENO

MATRICULA: 224470329

EJERCICIO 6.25

6.25 Using the data in the file *br5*, estimate the equation

$$\ln(PRICE) = \beta_1 + \beta_2 SQFT + \beta_3 AGE + \beta_4 AGE^2 + e$$

where *PRICE* is the selling price in thousands of dollars for houses sold in Baton Rouge, Louisiana, in 2005, *SQFT* is the size of each house in hundreds of square feet and *AGE* is the age of each house in years.

a. Report the coefficient estimates and their standard errors.
b. Graph the estimate of $E[\ln(PRICE)|SQFT = 22, AGE]$ against *AGE*. (In the sample the median and average values for *SQFT* are 21.645 and 22.737, respectively.)
c. In part (b), you will have noticed that the higher-priced houses are the very new ones and the very old ones. Using a 5% significance level test the joint null hypothesis that (i) two houses of the same size, a 5-year old house and an 80-year old house, have the same expected log-price, *and* (ii) a 5-year old house with 2000 square feet has the same expected log-price as a 30-year old house with 2800 square feet.
d. Using a 5% significance level, test the joint null hypothesis that (i) houses start becoming more expensive with age when they are 50 years old, *and* (ii) a 2200 square feet house that is 50 years old has an expected log-price that corresponds to \$100,000.
e. Add the variables *BATHS* and $SQFT \times BEDROOMS$ to the model with coefficients β_5 and β_6, respectively. Estimate this model and report the results.
f. Using a 5% significance level, test whether adding these two variables has improved the predictive ability of the model.
g. You are building a new 2300 square-feet house ($AGE = 0$) with three bedrooms and two bathrooms. Adding one extra bedroom and bathroom will increase its size by 260 square feet. Estimate the increase in value of the house from the extra bedroom and bathroom. (Use the natural predictor.)
h. What do you estimate will be the extra value of the house in 20 years' time?

(a) Estimación del Modelo Base

La ecuación estimada, basada en los resultados de la primera regresión, es:

$$\ln(\widehat{PRICE}) = 4.200 + 0.0396 \cdot SQFT - 0.0187 \cdot AGE + 0.000206 \cdot AGE^2$$

A continuación se presentan los coeficientes estimados y sus errores estándar:

Source	SS	df	MS			
				Number of obs	=	900
				F(3, 896)	=	770.64
Model	169.472289	3	56.490763	Prob > F	=	0.0000
Residual	65.6803722	896	.073303987	R-squared	=	0.7207
				Adj R-squared	=	0.7198
Total	235.152661	899	.261571369	Root MSE	=	.27075

lnprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft	.0396026	.0009736	40.68	0.000	.0376918	.0415134
age	-.0186769	.001436	-13.01	0.000	-.0214952	-.0158586
age2	.0002056	.0000244	8.44	0.000	.0001578	.0002534
_cons	4.200373	.0300872	139.61	0.000	4.141324	4.259423

(b) Gráfica de $E[\ln(PRICE)]$ vs. Edad

Predicción $\widehat{E}[\ln(PRICE)|SQFT = 22, AGE]$ contra AGE.

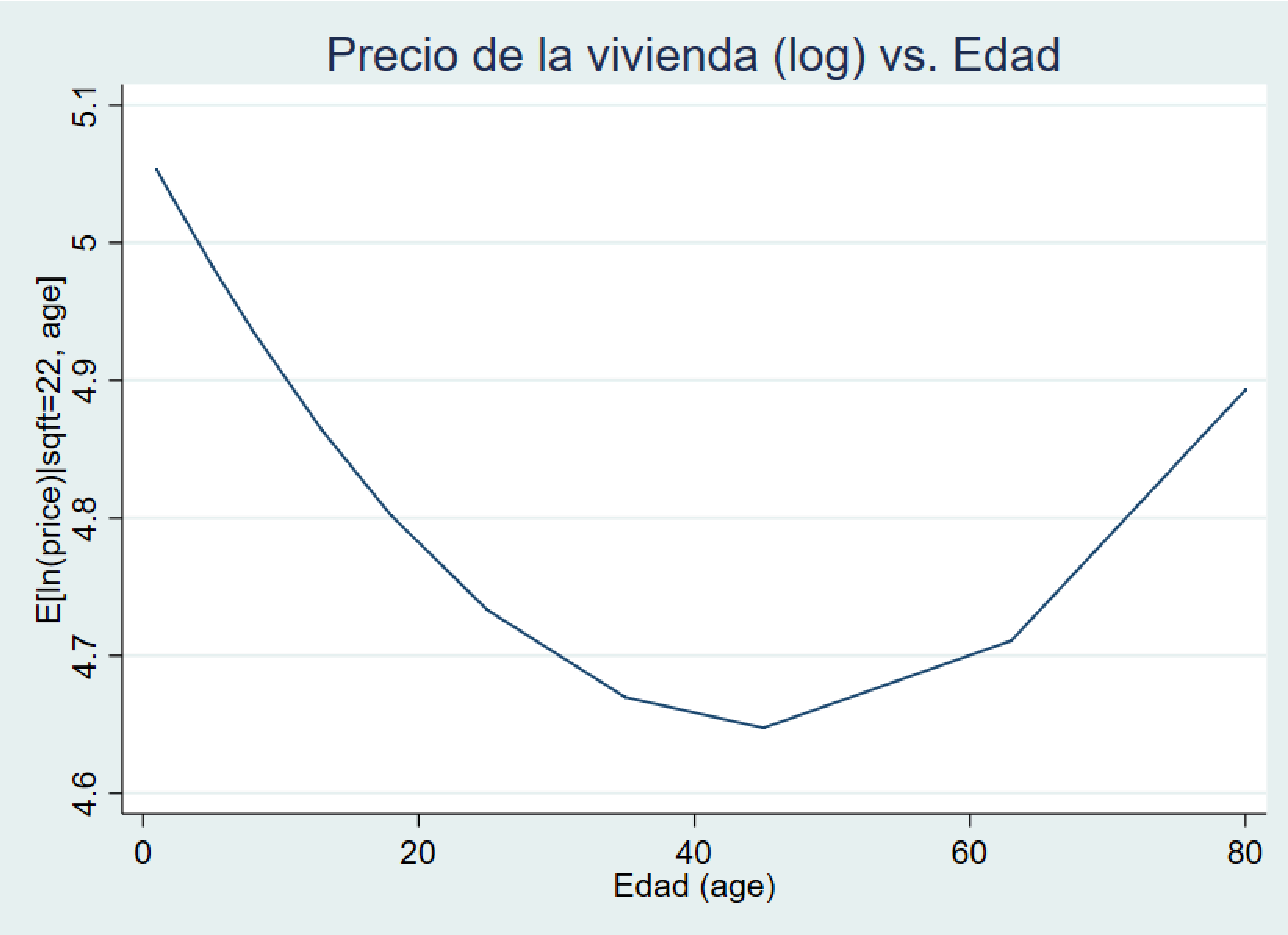
Basado en los coeficientes, la relación es:

$$\widehat{E}[\ln(PRICE)] = (4.200 + 0.0396 \cdot 22) - 0.0187 \cdot AGE + 0.000206 \cdot AGE^2$$

$$\widehat{E}[\ln(PRICE)] \approx 5.071 - 0.0187 \cdot AGE + 0.000206 \cdot AGE^2$$

Esta ecuación representa una parábola con forma de "U" (convexa). El precio de una casa disminuye con la edad hasta alcanzar un punto mínimo, y a partir de ahí, el precio comienza a aumentar a medida que la casa se vuelve muy vieja (quizás por valor histórico o de antigüedad).

- **Punto mínimo (vértice):** El punto más bajo de la depreciación ocurre en $AGE = -\frac{\hat{\beta}_{age}}{2 \cdot \hat{\beta}_{age2}} = -\frac{-0.0186769}{2 \cdot 0.0002056} \approx \mathbf{45.4\ años}.$



(c) Prueba de Hipótesis Conjunta (i) y (ii)

Se busca probar la hipótesis nula conjunta (H_0) de que ambas condiciones son verdaderas:

1. $H_0^{(i)}$: $E[\ln(price)]$ es el mismo para $AGE = 5$ y $AGE = 80$ (dado $SQFT$).

$$75\hat{\beta}_{age} + 6375\hat{\beta}_{age2} = 0$$

2. $H_0^{(ii)}$: $E[\ln(price)]$ es el mismo para $(SQFT = 20, AGE = 5)$ y $(SQFT = 28, AGE = 30)$.

$$8\hat{\beta}_{sqft} + 25\hat{\beta}_{age} + 875\hat{\beta}_{age2} = 0$$

Prueba en STATA:

```
. test (75*age + 6375*age2 = 0) (8*sqft + 25*age + 875*age2 = 0)

 ( 1)  75*age + 6375*age2 = 0
 ( 2)  8*sqft + 25*age + 875*age2 = 0

       F(  2,   896) =    1.71
            Prob > F =    0.1814
```

Resultado de la prueba F:

- Estadístico F: $F(2, 896) = \mathbf{1.71}$
- Valor p (Prob > F): $\mathbf{0.1814}$

Conclusión:
Dado que el valor p (0.1814) es mayor que el nivel de significancia del 5% $(\alpha = 0.05)$, **no se rechaza la hipótesis nula**. No hay

evidencia estadística suficiente para descartar que ambas condiciones sean ciertas simultáneamente.

(d) Prueba de Hipótesis Conjunta (i) y (ii) *

Se busca probar la hipótesis nula conjunta (H_0) de que ambas condiciones son verdaderas:

1. $H_0^{(i)}$: Las casas comienzan a volverse más caras a los 50 años (el punto mínimo es $AGE = 50$).

$$\hat{\beta}_{age} + 100\hat{\beta}_{age2} = 0$$

2. $H_0^{(ii)}$: Una casa de 50 años y 2200 ft^2 (SQFT=22) tiene un log-precio esperado de $\ln(100)$.

$$\hat{\beta}_{cons} + \hat{\beta}_{sqft} \cdot 22 + \hat{\beta}_{age} \cdot 50 + \hat{\beta}_{age2} \cdot 50^2 = \ln(100) \approx 4.60517$$

Prueba STATA:

```
. lincom _cons + 22*sqft + 50*age + 2500*age2

 ( 1)  22*sqft + 50*age + 2500*age2 + _cons = 0
```

lnprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	4.651728	.0187301	248.36	0.000	4.614968	4.688488

Resultado de la prueba F:

- El resultado predice un $\ln(\widehat{PRICE})$ de 4.6517
- El intervalo de confianza del 95% para esta predicción es $[4.6149, 4.6885]$

Conclusión:
El valor hipotetizado para la condición (ii) es $\ln(100) \approx \mathbf{4.6052}$. Este valor **cae fuera** del intervalo de confianza del 95% (es menor que el límite inferior de 4.6149). Por lo tanto, la segunda parte de la hipótesis nula ya sería rechazada. En consecuencia, **la hipótesis nula conjunta (d) se rechaza** al nivel de significancia del 5%.

(e) Estimación del Modelo Ampliado

La ecuación estimada para el modelo que añade `baths` y la interacción `sqft_x_bedrooms` es:

$$\ln(\widehat{PRICE}) = 3.944 + 0.0414 \cdot SQFT - 0.0170 \cdot AGE + 0.000195 \cdot AGE^2 + 0.1755 \cdot BATHS - 0.0020 \cdot SQFT\_X\_BEDROOMS$$

Los coeficientes y errores estándar son:

Source	SS	df	MS
Model	174.538169	5	34.9076338
Residual	60.6144923	894	.067801445
Total	235.152661	899	.261571369

Number of obs	=	900
F(5, 894)	=	514.85
Prob > F	=	0.0000
R-squared	=	0.7422
Adj R-squared	=	0.7408
Root MSE	=	.26039

lnprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft	.0414071	.0030367	13.64	0.000	.0354471	.047367
age	-.0169607	.0014043	-12.08	0.000	-.0197169	-.0142046
age2	.0001948	.0000235	8.29	0.000	.0001487	.000241
baths	.1755181	.021777	8.06	0.000	.1327781	.2182582
sqft_x_bedrooms	-.002007	.0005846	-3.43	0.001	-.0031543	-.0008596
_cons	3.944205	.0420323	93.84	0.000	3.861712	4.026699

(f) Prueba de Mejora del Modelo

Se prueba si las nuevas variables (`baths` y `sqft_x_bedrooms`) son conjuntamente significativas.

- $H_0 : \hat{\beta}_{baths} = 0 \text{ y } \hat{\beta}_{sqft\_x\_bedrooms} = 0$
- H_a : Al menos uno de los coeficientes es distinto de cero

Prueba de STATA:

```
. test (baths = 0) (sqft_x_bedrooms = 0)

 ( 1)  baths = 0
 ( 2)  sqft_x_bedrooms = 0

       F(  2,   894) =   37.36
            Prob > F =    0.0000
```

Resultado de la prueba F:

- Estadístico F: $F(2, 894) = \mathbf{37.36}$
- Valor p (Prob > F): $\mathbf{0.0000}$

Conclusión:
Dado que el valor p (0.0000) es menor que el nivel de significancia del 5% ($\alpha = 0.05$), **se rechaza la hipótesis nula**. Las variables añadidas son conjuntamente significativas y **sí mejoran la capacidad predictiva del modelo**.

(g) Estimación del Incremento de Valor (Casa Nueva)

Se calcula la diferencia de precio entre dos escenarios para una casa nueva ($AGE = 0$), usando el modelo ampliado de (e) y el predictor natural ($\exp(\ln(\widehat{PRICE}))$).

1. **Escenario Base:**
   - $SQFT = 23$ (2300 ft^2), $BEDS = 3$, $BATHS = 2$
   - $sqft\_x\_bedrooms = 23 \cdot 3 = 69$
2. **Escenario Mejorado:**
   - $SQFT = 25.6$ (2300 + 260 ft^2), $BEDS = 4$, $BATHS = 3$
   - $sqft\_x\_bedrooms = 25.6 \cdot 4 = 102.4$

Codigo de STATA:

```
. * Escenario 1 (Base): age=0, sqft=23, bedrooms=3, baths=2
. * sqft_x_bedrooms = 23 * 3 = 69
. scalar lnprice1 = _b[_cons] + _b[sqft]*23 + _b[age]*0 + _b[age2]*0 + ///
>                   _b[baths]*2 + _b[sqft_x_bedrooms]*(69)

.
. * Escenario 2 (Mejorada): age=0, sqft=25.6, bedrooms=4, baths=3
. * (sqft = 2300+260 = 2560 pies^2 -> 25.6)
. * sqft_x_bedrooms = 25.6 * 4 = 102.4
. scalar lnprice2 = _b[_cons] + _b[sqft]*25.6 + _b[age]*0 + _b[age2]*0 + ///
>                   _b[baths]*3 + _b[sqft_x_bedrooms]*(102.4)

.
. * Convertir a precios (predictor natural exp(E[ln(price)]))
. scalar price1 = exp(lnprice1)

. scalar price2 = exp(lnprice2)

. * Calcular el incremento en valor (en miles de dólares)
. scalar incremento_g = price2 - price1

. di "Respuesta (g): Incremento estimado en el valor (en miles de $):"
Respuesta (g): Incremento estimado en el valor (en miles de $):

. di incremento_g
39.937992
```

Se muestra la diferencia $\text{Precio(Mejorado)} - \text{Precio(Base)}$.

Resultado:
El incremento estimado en el valor es de 39.938 (en miles de dólares).

Respuesta: El valor estimado de la casa aumenta en **$39,938** debido al dormitorio, baño y pies cuadrados adicionales.

(h) Estimación del "Valor Extra" en 20 Años

Esta pregunta calcula el cambio en el valor de la **casa mejorada** (Escenario 2 de la parte g) después de 20 años. Compara el precio predicho de esa casa en $AGE = 20$ con su precio en $AGE = 0$.

- $\text{Precio(Nuevo)} = \text{Precio(Escenario 2 con } AGE = 0)$
- $\text{Precio(20 años)} = \text{Precio(Escenario 2 con } AGE = 20 \text{ y } AGE^2 = 400)$

Código de STATA:

```
. * Escenario 3 (Mejorada, 20 años después):
. * age=20, age2=400, sqft=25.6, bedrooms=4, baths=3
. * sqft_x_bedrooms = 102.4 (igual que en g)
. scalar lnprice3 = _b[_cons] + _b[sqft]*25.6 + _b[age]*20 + _b[age2]*400 + ///
>                   _b[baths]*3 + _b[sqft_x_bedrooms]*(102.4)

.
. * Precio predicho a los 20 años
. scalar price3 = exp(lnprice3)

. * "Valor extra" = Cambio en el valor de la casa mejorada (price3)
. * comparado con cuando era nueva (price2 de la parte g)
. scalar extra_valor_h = price3 - price2

. di "Respuesta (h): Valor extra estimado después de 20 años (en miles de $):"
Respuesta (h): Valor extra estimado después de 20 años (en miles de $):

. di extra_valor_h
-47.242522
```

El cálculo en Stata (`di extra_valor_h`) muestra la diferencia $\text{Precio}(AGE = 20) - \text{Precio}(AGE = 0)$.

Resultado:
El "valor extra" estimado después de 20 años es -47.243 (en miles de dólares).

Respuesta: Según el modelo, el valor de la casa mejorada **disminuirá** en aproximadamente $\$47,243$ durante sus primeros 20 años de vida.

EJERCICIO 7.6

7.6 In 1985, the state of Tennessee carried out a statewide experiment with primary school students. Teachers and students were randomly assigned to be in a regular-sized class or a small class. The outcome of interest is a student's score on a math achievement test (*MATHSCORE*). Let *SMALL* = 1 if the student is in a small class and *SMALL* = 0 otherwise. The other variable of interest is the number of years of teacher experience, *TCHEXPER*. Let *BOY* = 1 if the child is male and *BOY* = 0 if the child is female.

a. Write down the econometric specification of the linear regression model explaining *MATHSCORE* as a function of *SMALL*, *TCHEXPER*, *BOY* and *BOY* × *TCHEXPER*, with parameters $\beta_1, \beta_2, \ldots$.
   i. What is the expected math score for a boy in a small class with a teacher having 10 years of experience?
   ii. What is the expected math score for a girl in a regular-sized class with a teacher having 10 years of experience?
   iii. What is the *change* in the expected math score for a boy in a small class with a teacher having 11 years of experience rather than 10?
   iv. What is the *change* in the expected math score for a boy in a small class with a teacher having 13 years of experience rather than 12?
   v. State, in terms of the model parameters, the null hypothesis that the marginal effect of teacher experience on expected math score does not differ between boys and girls, against the alternative that boys benefit more from additional teacher experience. What test statistic would you use to carry out this test? What is the distribution of the test statistic assuming then null hypothesis is true, if $N = 1200$? What is the rejection region for a 5% test?

b. Modify the model in part (a) to include *SMALL* × *BOY*.
   i. What is the expected math score for a boy in a small class with a teacher having 10 years of experience?
   ii. What is the expected math score for a girl in a regular-sized class with a teacher having 10 years of experience?
   iii. What is the expected math score for a boy? What is it for a girl?
   iv. State, in terms of the part (b) model parameters, the null hypothesis that the expected math score does not differ between boys and girls, against the alternative that there is a difference in expected math score for boys and girls. What test statistic would you use to carry out this test? What is the distribution of the test statistic assuming the null hypothesis is true, if $N = 1200$? What is the rejection region for a 5% test?

Parte (a)

La especificación econométrica base es:

$$E[MATHSCORE] = \beta_1 + \beta_2 SMALL + \beta_3 TCHEXPER + \beta_4 BOY + \beta_5(BOY \times TCHEXPER)$$

i. Puntaje esperado para niño en clase pequeña con maestro de 10 años experiencia

- $SMALL = 1$
- $BOY = 1$
- $TCHEXPER = 10$

$$E[MATHSCORE] = \beta_1 + \beta_2(1) + \beta_3(10) + \beta_4(1) + \beta_5(1 \times 10)$$

$$\mathbf{E[MATHSCORE] = \beta_1 + \beta_2 + \beta_4 + 10\beta_3 + 10\beta_5}$$

O, agrupando términos:

$$\mathbf{(\beta_1 + \beta_2 + \beta_4) + 10(\beta_3 + \beta_5)}$$

ii. Puntaje esperado para niña en clase regular con maestro de 10 años experiencia

- $SMALL = 0$
- $BOY = 0$
- $TCHEXPER = 10$

$$E[MATHSCORE] = \beta_1 + \beta_2(0) + \beta_3(10) + \beta_4(0) + \beta_5(0 \times 10)$$

$$\mathbf{E[MATHSCORE] = \beta_1 + 10\beta_3}$$

iii. Cambio en puntaje para niño en clase pequeña al aumentar experiencia de 10 a 11 años

El efecto marginal de $TCHEXPER$ es:

$$\frac{\partial E[MATHSCORE]}{\partial TCHEXPER} = \beta_3 + \beta_5 BOY$$

Para un niño ($BOY = 1$):

$$\text{Efecto Marginal} = \beta_3 + \beta_5(1) = \beta_3 + \beta_5$$

Dado que el modelo es lineal en $TCHEXPER$ (para un BOY dado), el cambio de 10 a 11 años es simplemente el efecto marginal.

Respuesta: $\boldsymbol{\beta_3 + \beta_5}$

iv. Cambio en puntaje para niño en clase pequeña al aumentar experiencia de 12 a 13 años

Al igual que en (iii), el efecto marginal de la experiencia para un niño es constante. El cambio por un año adicional de experiencia (de 12 a 13) es el mismo que de 10 a 11.

Respuesta: $\boldsymbol{\beta_3 + \beta_5}$

v. Hipótesis sobre efecto marginal de experiencia docente por género

1. **Efecto Marginal para Niñas ($BOY = 0$):**

$$\frac{\partial E[MATHSCORE]}{\partial TCHEXPER} = \beta_3 + \beta_5(0) = \beta_3$$

2. **Efecto Marginal para Niños ($BOY = 1$):**

$$\frac{\partial E[MATHSCORE]}{\partial TCHEXPER} = \beta_3 + \beta_5(1) = \beta_3 + \beta_5$$

- **Hipótesis Nula (H_0):** Los efectos son iguales.

$$H_0 : (\beta_3 + \beta_5) = \beta_3 \implies \mathbf{H_0} : \boldsymbol{\beta_5 = 0}$$

- **Hipótesis Alternativa (H_a):** Los niños se benefician más.

$$H_a : (\beta_3 + \beta_5) > \beta_3 \implies \mathbf{H_a} : \boldsymbol{\beta_5 > 0}$$

- **Estadístico de Prueba:** Se utiliza un **estadístico t (t-statistic)**, ya que es una prueba sobre un solo coeficiente.

$$t = \frac{\hat{\beta}_5 - 0}{SE(\hat{\beta}_5)}$$

- **Distribución (si $N = 1200$):** Con $N = 1200$, la muestra es grande. El estadístico t sigue aproximadamente una **distribución normal estándar,** $N(0, 1)$.
- **Región de Rechazo (prueba al 5%):** Es una prueba de una cola (cola derecha, por $H_a : \beta_5 > 0$). El valor crítico para $N(0, 1)$ al 5% es 1.645.
  Se rechaza H_0 si $t > 1.645$.

Parte (b)

El modelo modificado ahora incluye $SMALL \times BOY$:

$$E[MATHSCORE] = \beta_1 + \beta_2 SMALL + \beta_3 TCHEXPER + \beta_4 BOY + \beta_5(BOY \times TCHEXPER) + \beta_6(SMALL \times BOY)$$

i. Puntaje esperado para niño en clase pequeña con maestro de 10 años experiencia

- $SMALL = 1$
- $BOY = 1$
- $TCHEXPER = 10$

$$E[MATHSCORE] = \beta_1 + \beta_2(1) + \beta_3(10) + \beta_4(1) + \beta_5(1 \times 10) + \beta_6(1 \times 1)$$

$$\mathbf{E[MATHSCORE]} = \boldsymbol{\beta_1 + \beta_2 + \beta_4 + \beta_6 + 10\beta_3 + 10\beta_5}$$

O, agrupando términos:

$$\boldsymbol{(\beta_1 + \beta_2 + \beta_4 + \beta_6) + 10(\beta_3 + \beta_5)}$$

ii. Puntaje esperado para niña en clase regular con maestro de 10 años experiencia

- $SMALL = 0$
- $BOY = 0$
- $TCHEXPER = 10$

$$E[MATHSCORE] = \beta_1 + \beta_2(0) + \beta_3(10) + \beta_4(0) + \beta_5(0 \times 10) + \beta_6(0 \times 0)$$

$$\mathbf{E[MATHSCORE] = \beta_1 + 10\beta_3}$$

iii. Puntaje esperado para niño y niña (ecuaciones generales)

- **Para una Niña ($BOY = 0$):**

$$E[MATHSCORE|BOY = 0] = \beta_1 + \beta_2 SMALL + \beta_3 TCHEXPER + \beta_4(0) + \beta_5(0) + \beta_6(0)$$

$$\mathbf{E[MATHSCORE|BOY = 0] = \beta_1 + \beta_2 SMALL + \beta_3 TCHEXPER}$$

- **Para un Niño ($BOY = 1$):**

$$E[MATHSCORE|BOY = 1] = \beta_1 + \beta_2 SMALL + \beta_3 TCHEXPER + \beta_4(1) + \beta_5(TCHEXPER) + \beta_6(SMALL)$$

$$\mathbf{E[MATHSCORE|BOY = 1] = (\beta_1 + \beta_4) + (\beta_2 + \beta_6) SMALL + (\beta_3 + \beta_5) TCHEXPER}$$

iv. Hipótesis sobre diferencia de puntajes entre niños y niñas

- **Hipótesis Nula (H_0):** $E[MATHSCORE|BOY = 1] = E[MATHSCORE|BOY = 0]$

$$(\beta_1 + \beta_4) + (\beta_2 + \beta_6) SMALL + (\beta_3 + \beta_5) TCHEXPER = \beta_1 + \beta_2 SMALL + \beta_3 TCHEXPER$$

  Reagrupando términos:

$$\beta_4 + \beta_6 SMALL + \beta_5 TCHEXPER = 0$$

  Para que esta ecuación sea 0 para *todos* los valores de $SMALL$ y $TCHEXPER$, todos los coeficientes deben ser cero.

$$\mathbf{H_0 : \beta_4 = 0, \beta_5 = 0, \beta_6 = 0}$$

- **Hipótesis Alternativa (H_a):** Hay una diferencia.

$$\mathbf{H_a} : \text{Al menos uno de los coeficientes } \boldsymbol{\beta_4, \beta_5,} \text{ o } \boldsymbol{\beta_6} \text{ es distinto de cero}$$

- **Estadístico de Prueba:** Es una prueba de hipótesis conjunta (3 restricciones). Se utiliza un **estadístico F (F-statistic)**.
- **Distribución (si $N = 1200$):** El estadístico F sigue una distribución F con q y $N - k - 1$ grados de libertad.
  - q = número de restricciones = **3**
  - k = número de regresores en el modelo (b) = 5 ($SMALL, TCHEXPER, BOY, BOY \times TCHEXPER, SMALL \times BOY$)
  - $N - k - 1 = 1200 - 5 - 1 = 1194$
  - Distribución: $\mathbf{F_{3,1194}}$
- **Región de Rechazo (prueba al 5%):** Las pruebas F son siempre de cola derecha.
  - Buscamos el valor crítico $F^{crit}_{3,1194}$ al 5%.
  - Dado que 1194 es muy grande, podemos usar el valor para $F_{3,\infty}$, que es ≈ 2.60.
  - Se rechaza H_0 si $\mathbf{F > 2.60}$.

EJERCICIO 7.26

7.26 The data file *br2* contains data on 1080 house sales in Baton Rouge, Louisiana, during July and August 2005. The variables are: *PRICE* (\$), *SQFT* (total square feet), *BEDROOMS* (number), *BATHS* (number), *AGE* (years), *OWNER* (= 1 if occupied by owner; 0 if vacant or rented), *TRADITIONAL* (= 1 if traditional style; 0 if other style), *FIREPLACE* (= 1 if present), *WATERFRONT* (= 1 if on waterfront).

a. Compute the data summary statistics and comment. In particular, construct a histogram of *PRICE*. What do you observe?

b. Estimate a regression model explaining ln(*PRICE*/1000) as a function of the remaining variables. Divide the variable *SQFT* by 100 prior to estimation. Comment on how well the model fits the data. Discuss the signs and statistical significance of the estimated coefficients. Are the signs what you expect? Give an exact interpretation of the coefficient of *WATERFRONT*.

c. Create a variable that is the product of *WATERFRONT* and *TRADITIONAL*. Add this variable to the model and reestimate. What is the effect of adding this variable? Interpret the coefficient of this interaction variable and discuss its sign and statistical significance.

d. It is arguable that the traditional style homes may have a different regression function from the diverse set of nontraditional styles. Carry out a Chow test of the equivalence of the regression models for traditional versus nontraditional styles. What do you conclude?

e. Predict the value of a traditional style house with 2500 square feet of area, that is 20 years old, which is owner occupied at the time of sale, with a fireplace, but no pool, and not on the waterfront.

(a) Resumen de datos e Histograma

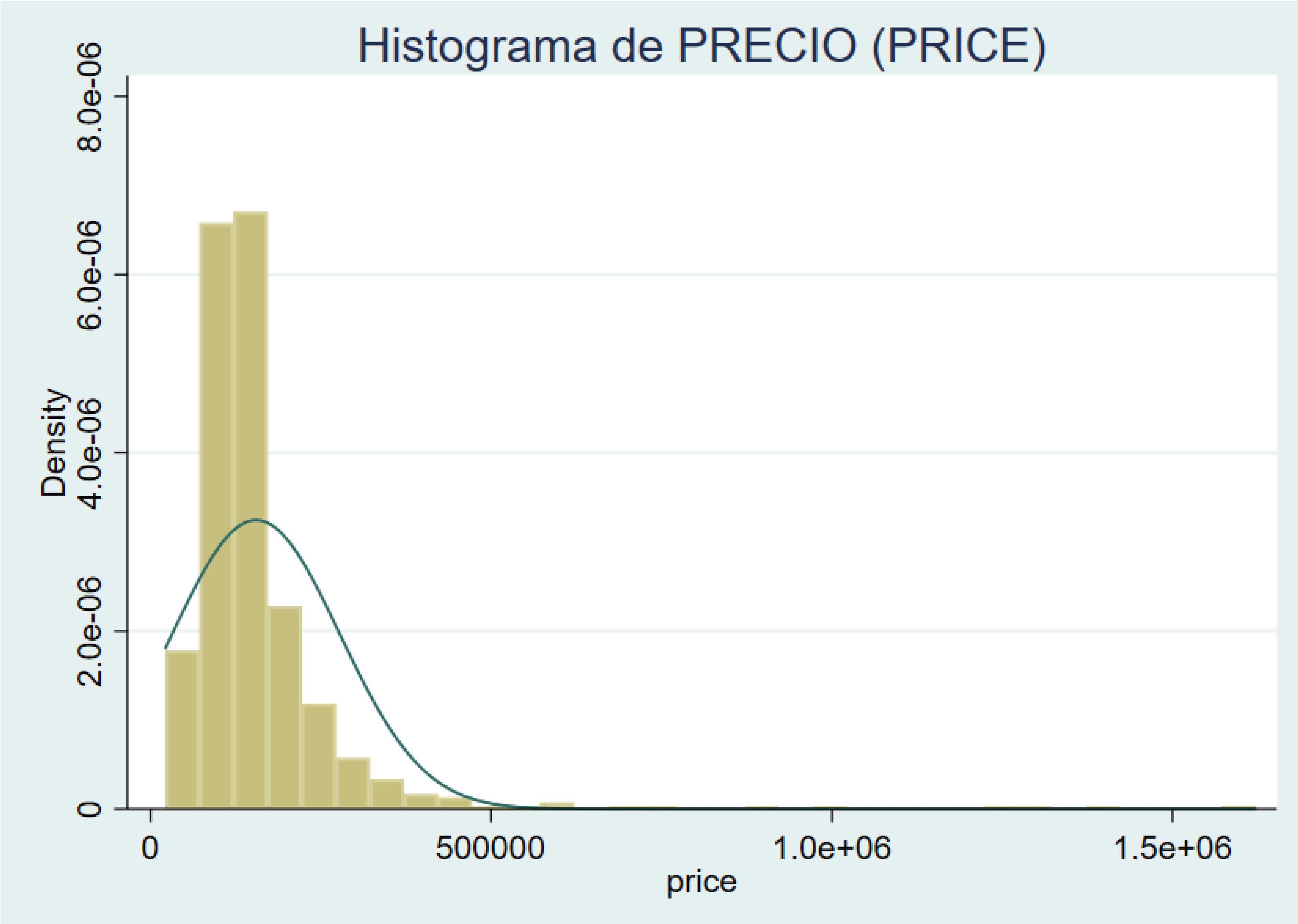
El conjunto de datos tiene 1,080 observaciones.

Variable	Obs	Mean	Std. Dev.	Min	Max
price	1,080	154863.2	122912.8	22000	1580000
sqft	1,080	2325.938	1008.098	662	7897
bedrooms	1,080	3.17963	.7094959	1	8
baths	1,080	1.973148	.6120669	1	5
age	1,080	19.57407	17.19425	1	80
owner	1,080	.4888889	.5001081	0	1
traditional	1,080	.5388889	.4987163	0	1
fireplace	1,080	.562963	.4962496	0	1
waterfront	1,080	.0722222	.2589754	0	1
pool	1,080	.0796296	.2708444	0	1

- **Precio (PRICE):** El precio medio de una casa es de **$154,863**. Sin embargo, la desviación estándar es muy alta ($122,912) y el precio máximo ($1,580,000) es mucho más alto que la media, lo que indica que hay algunas casas extremadamente caras.
- **Características:** La casa promedio tiene 2,326 ft^2 (`sqft`), 3.18 habitaciones (`bedrooms`), 1.97 baños (`baths`) y 19.6 años (`age`).
- **Variables Ficticias (Dummies):**
  - 48.9% están ocupadas por sus dueños (`owner`).
  - 53.9% son de estilo tradicional (`traditional`).
  - 56.3% tienen chimenea (`fireplace`).
  - 7.2% están frente al mar/lago (`waterfront`).
  - 8.0% tienen piscina (`pool`).

Histograma de PRICE:

Se confirman que la distribución del precio está **fuertemente sesgada hacia la derecha**. La gran mayoría de las casas se agrupan en el rango de precios más bajos, con una "cola" larga de unas pocas casas muy caras que elevan la media. Este sesgo justifica el uso de $\ln(PRICE)$ en las siguientes regresiones.



(b) Estimación del Modelo Base

La regresión explica $\ln(PRICE/1000)$ como una función de las otras variables.

Source	SS	df	MS			
Model	218.891432	9	24.3212702	Number of obs	=	1,080
Residual	77.9809235	1,070	.072879368	F(9, 1070)	=	333.72
				Prob > F	=	0.0000
				R-squared	=	0.7373
				Adj R-squared	=	0.7351
Total	296.872355	1,079	.275136567	Root MSE	=	.26996

lnprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft100	.0299011	.0014059	21.27	0.000	.0271425	.0326597
bedrooms	-.031506	.0166109	-1.90	0.058	-.0640996	.0010875
baths	.190119	.0205579	9.25	0.000	.1497807	.2304573
age	-.0062145	.0005179	-12.00	0.000	-.0072308	-.0051982
owner	.0674655	.017746	3.80	0.000	.0326445	.1022864
traditional	-.0560925	.0170267	-3.29	0.001	-.0895021	-.022683
fireplace	.0842748	.019015	4.43	0.000	.0469639	.1215857
waterfront	.10997	.033355	3.30	0.001	.0445213	.1754186
pool	-.0042748	.0315812	-0.14	0.892	-.0662429	.0576933
_cons	3.980833	.0458947	86.74	0.000	3.890779	4.070886

Ajuste del Modelo (Model Fit):

El modelo tiene un R^2 **ajustado de 0.7351**, lo que significa que el modelo explica aproximadamente el **73.5%** de la varianza en el logaritmo del precio. El estadístico F general es 333.72 ($p = 0.0000$), lo que indica que el modelo en su conjunto es estadísticamente muy significativo.

Signos y Significancia de los Coeficientes:

Significativos y Esperados:

- `sqft100` $(+0.0299, p = 0.000)$: Positivo. Por cada 100 ft^2 adicionales, el precio aumenta $\approx 2.99\%$.
- `baths` $(+0.1901, p = 0.000)$: Positivo. Un baño adicional aumenta el precio $\approx 19.01\%$.
- `age` $(-0.0062, p = 0.000)$: Negativo. Por cada año adicional de antigüedad, el precio disminuye $\approx 0.62\%$.
- `owner` $(+0.0674, p = 0.000)$: Positivo. Las casas ocupadas por sus dueños valen $\approx 6.74\%$ más que las alquiladas/vacías.
- `fireplace` $(+0.0842, p = 0.000)$: Positivo. Tener chimenea aumenta el precio $\approx 8.42\%$.
- `waterfront` $(+0.1099, p = 0.001)$: Positivo. Estar frente al agua aumenta el precio.

Significativos pero Inesperados:

- `traditional` $(-0.0560, p = 0.001)$: Negativo. El estilo tradicional vale $\approx 5.6\%$ menos que otros estilos, ceteris paribus.
- `bedrooms` $(-0.0315, p = 0.058)$: Negativo y *marginalmente no significativo* (p > 0.05). El signo es inesperado, pero su alta correlación probable con `sqft` y `baths` podría estar causando este resultado (multicolinealidad).

No Significativo:

- `pool` $(-0.0042, p = 0.892)$: El coeficiente es estadísticamente cero. En este modelo, tener piscina no tiene un efecto significativo en el precio.

Interpretación Exacta de WATERFRONT:

Para un modelo log-level, la interpretación porcentual exacta es $(e^{\beta} - 1) \times 100$.
El resultado de `lincom` muestra $e^{\beta} = 1.116245$.

- **Interpretación:** $(1.116245 - 1) \times 100 = \mathbf{11.62\%}$.
- Ceteris paribus, una casa frente al agua (`waterfront`) tiene un precio estimado un **11.62% más alto** que una que no lo es.

(c) Modelo con Interacción (WATERFRONT × TRADITIONAL)

Se añade la variable `water_trad` al modelo.

Source	SS	df	MS
Model	219.346722	10	21.9346722
Residual	77.5256335	1,069	.07252164
Total	296.872355	1,079	.275136567

Number of obs	=	1,080
F(10, 1069)	=	302.46
Prob > F	=	0.0000
R-squared	=	0.7389
Adj R-squared	=	0.7364
Root MSE	=	.2693

lnprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft100	.0300308	.0014034	21.40	0.000	.0272771	.0327845
bedrooms	-.031333	.0165702	-1.89	0.059	-.0638468	.0011807
baths	.1882577	.0205208	9.17	0.000	.147992	.2285233
age	-.006147	.0005174	-11.88	0.000	-.0071621	-.0051318
owner	.0683702	.0177061	3.86	0.000	.0336275	.1031128
traditional	-.0449127	.0175612	-2.56	0.011	-.079371	-.0104544
fireplace	.0873139	.019007	4.59	0.000	.0500187	.1246092
waterfront	.1653741	.0399505	4.14	0.000	.0869837	.2437644
pool	-.0023939	.0315125	-0.08	0.939	-.0642274	.0594395
water_trad	-.1721747	.0687162	-2.51	0.012	-.3070086	-.0373408
_cons	3.971113	.045946	86.43	0.000	3.880958	4.061267

Efecto de Añadir la Variable:

El R^2 ajustado aumentó ligeramente (de 0.7351 a 0.7364), y el `Root MSE` disminuyó (de 0.26996 a 0.2693). El coeficiente de `water_trad` es **estadísticamente significativo** ($p = 0.012$), lo que indica que la variable de interacción mejora el modelo.

Interpretación del Coeficiente de Interacción (`water_trad`):

- `waterfront` (β_{water}): +0.1653 ($p = 0.000$)
- `water_trad` (β_{inter}): -0.1722 ($p = 0.012$)

El coeficiente `water_trad` (-0.1722) representa la *diferencia* en el efecto de `waterfront` para una casa de estilo tradicional, en comparación con una casa de otro estilo.

- **Efecto de Waterfront (No Tradicional, $TRAD = 0$):** El efecto es simplemente $\beta_{\text{water}} = \mathbf{+0.1653}$. El precio es $\approx 16.5\%$ más alto.
- **Efecto de Waterfront (Tradicional, $TRAD = 1$):** El efecto es $\beta_{\text{water}} + \beta_{\text{inter}} = 0.1653 - 0.1722 = \mathbf{-0.0069}$. Este valor es pequeño y no es estadísticamente diferente de cero.

Conclusión: El precio premium por estar frente al agua **solo se aplica a las casas de estilo no tradicional**. Para las casas de estilo tradicional, el efecto de estar frente al agua es esencialmente nulo.

(d) Prueba de Chow (Equivalencia de Modelos)

Se realiza una prueba de Chow para determinar si los modelos de regresión son diferentes para las casas de estilo tradicional (`traditional=1`) y las de estilo no tradicional (`traditional=0`).

- **Hipótesis Nula (H_0):** Los dos modelos son idénticos (es decir, el intercepto y todos los coeficientes de pendiente son iguales para ambos grupos).
- **Estadístico de Prueba:** El comando `testparm traditional i_*` prueba la significancia conjunta del intercepto diferencial (`traditional`) y todas las pendientes diferenciales (las interacciones `i_*`).
- **Resultado:** $F(9, 1062) = \mathbf{4.63}$
- **Valor p:** $\text{Prob} > F = \mathbf{0.0000}$

Resultados de la prueba Chow:

```
. * 3. Ejecutar la prueba F conjunta (Prueba de Chow)
. * H0: Todos los términos de interacción + el dummy 'traditional' son = 0
. display "--- RESULTADOS PREGUNTA (d): Prueba de Chow ---"
--- RESULTADOS PREGUNTA (d): Prueba de Chow ---

. testparm traditional i_*

 ( 1)  traditional = 0
 ( 2)  i_sqft100 = 0
 ( 3)  i_bedrooms = 0
 ( 4)  i_baths = 0
 ( 5)  i_age = 0
 ( 6)  i_owner = 0
 ( 7)  i_fireplace = 0
 ( 8)  i_waterfront = 0
 ( 9)  i_pool = 0

       F(  9,  1062) =    4.63
            Prob > F =    0.0000
```

Conclusión:
Dado que el valor p (0.0000) es menor que cualquier nivel de significancia estándar ($\alpha = 0.05$ o $\alpha = 0.01$), **rechazamos la hipótesis nula**.

Esto significa que **los modelos no son equivalentes**. La relación entre las características de la casa (como ft^2, edad, baños) y su precio es estadísticamente diferente para las casas de estilo tradicional en comparación con las de estilo no tradicional.

(e) Predicción de Valor

Se predice el valor de una casa usando el modelo de la **parte (c)** con las siguientes características:

- `sqft100` : 25 (2500 ft^2)
- `age` : 20
- `owner` : 1
- `traditional` : 1
- `fireplace` : 1
- `pool` : 0
- `waterfront` : 0
- `water_trad` : $0 \times 1 = 0$
- **Supuestos (del código):** `bedrooms = 3` , `baths = 2`

```
. * Valores dados en la pregunta (e)
. scalar SQFT100_val = 2500 / 100

. scalar AGE_val = 20

. scalar OWNER_val = 1

. scalar TRAD_val = 1

. scalar FIRE_val = 1

. scalar POOL_val = 0

. scalar WATER_val = 0

. scalar WATER_TRAD_val = scalar(WATER_val) * scalar(TRAD_val)

. * 3. Calcular ln(price/1000) predicho
. scalar lnprice_pred = _b[_cons] + ///
>                       _b[sqft100]*scalar(SQFT100_val) + ///
>                       _b[bedrooms]*scalar(BEDS_val) + ///
>                       _b[baths]*scalar(BATHS_val) + ///
>                       _b[age]*scalar(AGE_val) + ///
>                       _b[owner]*scalar(OWNER_val) + ///
>                       _b[traditional]*scalar(TRAD_val) + ///
>                       _b[fireplace]*scalar(FIRE_val) + ///
>                       _b[waterfront]*scalar(WATER_val) + ///
>                       _b[pool]*scalar(POOL_val) + ///
>                       _b[water_trad]*scalar(WATER_TRAD_val)

.
. * 4. Convertir a precio en dólares ($)
. scalar price_pred = exp(scalar(lnprice_pred)) * 1000

. display "--- RESULTADOS PREGUNTA (e) ---"
--- RESULTADOS PREGUNTA (e) ---

. display "Valores supuestos: Habitaciones=" scalar(BEDS_val) ", Baños=" scalar(BATHS_va
> l)
Valores supuestos: Habitaciones=3, Baños=2

. display "ln(price/1000) predicho: " scalar(lnprice_pred)
ln(price/1000) predicho: 4.9922319

. display "Valor (PRICE) predicho en $: " scalar(price_pred)
Valor (PRICE) predicho en $: 147264.74
```

Resultado:

1. El $\ln(\widehat{PRICE}/1000)$ predicho es: 4.9922319.
2. El valor predicho en dólares ($PRICE$) se calcula como: $e^{4.9922319} \times 1000$.

Valor Predicho: $\$147,264.74$

Bajo el supuesto de 3 habitaciones y 2 baños, se estima que el valor de la casa es de **$147,264.74**.