

Cuestionario (Econometría, GADE)

2023-07-07

Clases prácticas

- El objetivo de las clases prácticas es **afianzar los contenidos vistos en la clase de Teoría**.
- Repasaremos la teoría, pero trabajaremos generalmente con casos y ejercicios con **datos**.
- Por ello, necesitamos usar un programa informático: [Gretl](#).
- Aunque no siempre es del todo posible, he agrupado los **ejercicios por temas** de la Guía Docente.
- Hay un tema adicional, el **Tema 0** cuyo objetivo es empezar a usar Gretl y recordar algunos conceptos e instrumentos básicos de Estadística.
- Muchos de los ejercicios utilizan datos. Puedes descargarte los datos [aquí](#)

Tema 0: Gretl y repaso de estadística

¿Qué es Gretl?

- En el curso usaremos [Gretl](#). Gretl es un paquete de software para **análisis econométrico**. Con Gretl se pueden hacer análisis estadísticos y estimar una amplia gama de **modelos econométricos**.
- Gretl es **software libre**. La página web de Gretl está [aquí](#).
- Lo usaremos a través de **menús**, pero incluye un potente lenguaje de programación: **Hansl**.

¿Cómo me instalo Gretl?

- Gretl es un programa libre y gratuito. Puedes descargarlo [aquí](#).

¿Cómo aprendo a usar Gretl?

- Usaremos Gretl a través de menús, por lo que en 2-3 clases os manejaréis bastante bien con Gretl.
- Si quieres aprender a usar Gretl en profundidad, entonces sí necesitas un manual.

Manuales de Gretl

- El propio Gretl tiene manuales de Gretl, concretamente en la pestaña **Ayuda** encontrarás varias guías. Los más interesantes son: **Guía del usuario**, **Guía de instrucciones** y **Hansl primer**.
- El manual más reciente que conozco está [aquí](#). Son 386 paginas.
- Puedes encontrar fácilmente manuales en castellano buscando en internet. Por ejemplo, [aquí](#), [aquí](#) o [aquí](#) tienes algunos.

Ejercicio 1

Objetivo

Empezar a usar Gretl mientras recordamos algunos instrumentos y conceptos estadísticos básicos.

Datos (los datos siempre están en aula virtual)

- Están en el fichero `t0_ej_01_interest-rate.gdt`
- Datos referentes a 2 tipos de interés (`i_long` e `i_short`) y un agregado monetario (`M2`)
- Son datos antiguos, pero ilustran muy bien las ideas y conceptos que quiero recordar

1) Abre en Gretl el fichero de datos `t0_ej_01_interest-rate.gdt`

Pista

Localiza el fichero `t0_ej_01_interest-rate.gdt` en tu ordenador. Pincha 2 veces en él con el ratón y se abrirá en Gretl.

2) Visualiza los datos de las variables `i_long` e `i_short`. ¿Qué valores toman las variables en 1990?

Pista

Una vez hayas cargado los datos en Gretl, puedes seguir la siguiente ruta de menús: **Datos > Mostrar Valores**.

3) Haz un gráfico temporal de esas 2 variables.

Pista

Puedes seguir la siguiente ruta de menús: **Ver > Gráficos > Gráfico de series temporales** y se abrirá una interfaz con la que podrás seleccionar las variables a graficar.

4) Con la variable `i_long`, haz un **histograma**, también llamado gráfico de barras o gráfico de distribución de frecuencias. Interpreta el gráfico, ¿qué información proporciona? ¿Se distribuye `i_long` aproximadamente como una normal?

Pista

En Gretl, selecciona la variable `i_long` con el ratón. Con `i_long` seleccionado, pincha con el botón derecho del ratón. Se abrirá un menú contextual. Selecciona la opción **Distribución de frecuencias**

5) Estadísticos descriptivos para `i_long`. Interpreta.

6) Gráfico de dispersión entre `i_long` e `i_short`. Interpreta. ¿Están relacionadas las 2 variables?

- 7) Gráfico de dispersión entre `i_long` y `M2`. Interpreta.
- 8) Matriz de correlaciones entre `i_long`, `i_short` y `M2`. Interpreta.
-

Ejercicio 2

Objetivo

Recordar, de forma intuitiva, qué es una **función de densidad**. Nos vendrá bien cuando hagamos contrastes de hipótesis.

1. Supongamos que la altura de la población española se **distribuye como una Normal** con media 180 centímetros y desviación típica 10 centímetros.
 - Dibuje la distribución de la variable.
 - ¿Qué probabilidad hay de que al seleccionar a un español al azar mida más de 2 metros?
 2. Vamos ahora a recordar algunas propiedades de la $N(0, 1)$
 - Dibuje a mano una $N(0, 1)$
 - Situe en el eje x los valores: $+1,96$, 0 y $-1,96$.
 - ¿Qué probabilidad hay de que aparezca un valor mayor o igual que $+1,96$? ¿Y mayor que 0 ? ¿Y de que sea mayor que $-1,96$?
 - Haga lo mismo con los valores $-1,64$ y $+1,64$
-

Ejercicio 3

Objetivo

- Recordar, de forma intuitiva, una propiedad de los **logaritmos** (naturales o neperianos). Concretamente que “los cambios en logaritmos **aproximan** tasas de variación en las variables originales”.
- Nos vendrá bien cuando usemos modelos con las variables en logaritmos. Será allá por el tema 5.

Abajo tienes una tabla con la evolución de los salarios de Juan y María. Rellena los huecos de la tabla para saber cuanto se han incrementado sus salarios y, sobre todo, ver cómo podríamos aproximar el crecimiento de una variable (en este caso el salario) si sólo tuviésemos información sobre el logaritmo de la variable.

Salario de Juan	variación absoluta	Tasa de variación	Variación en %	Ln (salario)	Incremento del logaritmo
100				4,605	
105	5	0,050	5,0%	4,654	0,049
107	2			4,673	
112	5	0,047			
115	3			4,745	4,745
Salario de María	variación absoluta	Tasa de variación	Variación en %	Ln (salario)	Incremento del logaritmo
100				4,605	
150	50			5,011	
400	250	1,667	166,7%		
1300	900			7,170	

Solución

Salario de Juan	variación absoluta	Tasa de variación	Variación en %	Ln (salario)	Incremento del logaritmo	exp (Δ ln)	1-exp (Δ ln)
100				4,605			
105	5	0,050	5,0%	4,654	0,049	1,050	0,050
107	2	0,019	1,9%	4,673	0,019	1,019	0,019
112	5	0,047	4,7%	4,718	0,046	1,047	0,047
115	3	0,027	2,7%	4,745	0,026	1,027	0,027
Salario de María	variación absoluta	Tasa de variación	Variación en %	Ln (salario)	Incremento del logaritmo	exp (Δ ln)	1-exp (Δ ln)
100				4,605			
150	50	0,500	50,0%	5,011	0,405	1,500	0,500
400	250	1,667	166,7%	5,991	0,981	2,667	1,667
1300	900	2,250	225,0%	7,170	1,179	3,250	2,250

Fíjate en 2 cosas:

- Los cambios en logaritmos aproximan las tasas de variación de la variable original **pero** solo si los cambios no son grandes. Fíjate que los logaritmos lo hacen relativamente bien con el salario de Juan que presenta crecimientos moderados, pero no para el salario de María que presenta crecimientos elevados.
- la tabla con las soluciones tiene 2 columnas adicionales que muestran como calcular tasas de variación exactas con datos en logaritmos. Generalmente nos bastará con aproximarlas con los incrementos de logaritmos.

Tema 1: Modelos Econométricos y datos económicos

Repaso de teoría (preguntas rápidas)

1. ¿Qué es la Econometría? ¿Qué hace la Econometría?
2. La ecuación $Y = AK^\alpha L^\beta$ ¿es lineal? ¿es un modelo econométrico?
3. ¿Qué etapas requiere un estudio econométrico?
4. ¿En qué se diferencian los datos de corte transversal de los de corte temporal?

Tema 2: Regresión lineal simple: geometría

Repaso de teoría (preguntas rápidas)

1. Escriba un MRLS y nombre sus componentes
2. ¿Cómo se estiman los parámetros del modelo anterior?
3. ¿Cómo se interpretan β_1 y β_2 ? ¿Y $\hat{\beta}_1$ y $\hat{\beta}_2$?
4. Enuncie las 4 propiedades descriptivas
5. ¿Qué es el R^2 ? Si $R^2 = 0.8$ ¿cómo se interpreta?

Ejercicio 1 (estimación a mano)

Objetivo

Recordar y practicar el proceso de estimación por MCO y los elementos del modelo teórico y del modelo estimado (o recta de regresión). Para ello usaremos un conjunto de datos muy simple, con sólo 5 observaciones.

Vamos a estimar “a mano”, bueno, en realidad usaremos Gretl para calcular lo que nos haga falta. También podemos mirar [aquí](#)

Datos

- En el fichero `t2_ej_01_estimar-a-mano.gdt`.
- Datos simples con solo 5 observaciones para x e y

Los datos pueden verse en la tabla adjunta:

x	y
5	6
7	3
4	9
6	3
3	9

1. Plantee un modelo de regresión lineal simple donde y sea la variable dependiente y x la variable independiente.
2. Escriba las expresiones de los estimadores MCO

3. Obtenga las estimaciones MCO y especifique el modelo estimado

Resultados

	X	Y	$(X - \bar{X})$	$(X - \bar{X})^2$	$(Y - \bar{Y})$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
Juan	5	6	0	0	0	0	0
Carlos	7	3	2	4	-3	9	-6
Susana	4	9	-1	1	3	9	-3
Veronica	6	3	1	1	-3	9	-3
Andrea	3	9	-2	4	3	9	-6
Suma	25	30	0	10	0	36	-18
Media	5	6		2		7,2	-3,6

$$\hat{y} = 15 - 1.8x$$

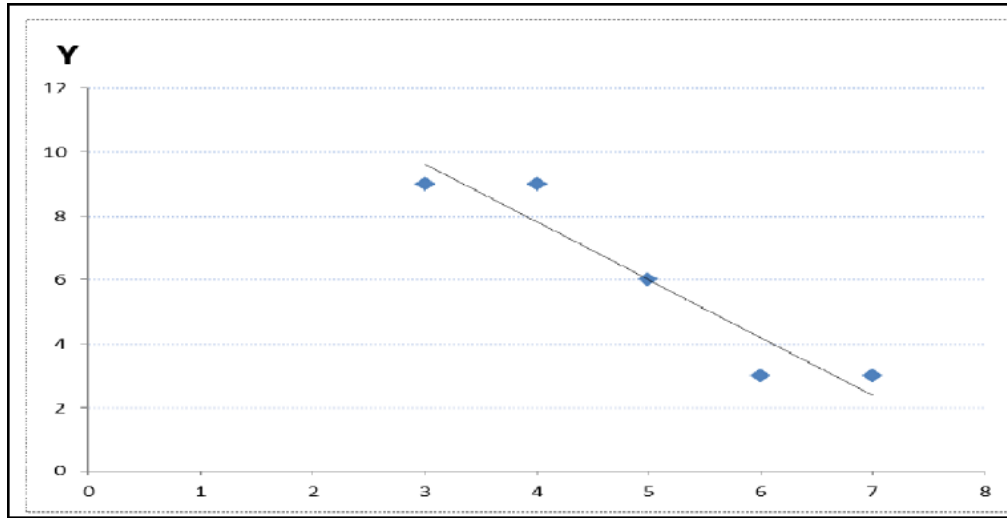
- Interprete el valor del parámetro asociado a la variable independiente o explicativa (β_2)
- Interprete el valor del **estimador** de β_2 (!!)
- Interprete el valor de la **estimación** de β_2
- Calcule los valores ajustados o estimados (\hat{y}) de la variable dependiente o regresando (y)

Resultados

	X	Y	\hat{Y}	\hat{u}
Juan	5	6	6	0
Carlos	7	3	2,4	0,6
Susana	4	9	7,8	1,2
Veronica	6	3	4,2	-1,2
Andrea	3	9	9,6	-0,6

- Calcule los residuos (\hat{u}) del modelo
- Represente gráficamente las observaciones (diagrama de dispersión o nube de puntos x - y), la recta de regresión ajustada, señalando los valores ajustados y los residuos.

Resultados



10. Verifique el cumplimiento de las 4 propiedades descriptivas
11. Descomponga la varianza total de la variable dependiente en varianza explicada por la regresión y varianza residual
12. Calcule el coeficiente de determinación (R^2) por los métodos que conozca (hay 2 métodos de cálculo)
13. Interprete el valor del coeficiente de determinación
14. Finalmente, estima el modelo otra vez, pero ahora con Gretl. Los datos están en el fichero `t2_ej_01_estimar-a-mano.gdt`

Resultados

Modelo 1: MCO, usando las observaciones 1-5

Variable dependiente: y

	coeficiente	Desv. típica	Estadístico t	valor p	
const	15,0000	1,80000	8,333	0,0036	***
x	-1,80000	0,346410	-5,196	0,0138	**
Media de la vble. dep.	6,000000	D.T. de la vble. dep.	3,000000		
Suma de cuad. residuos	3,600000	D.T. de la regresión	1,095445		
R-cuadrado	0,900000	R-cuadrado corregido	0,866667		
F(1, 3)	27,00000	Valor p (de F)	0,013847		
Log-verosimilitud	-6,273432	Criterio de Akaike	16,54686		
Criterio de Schwarz	15,76574	Crit. de Hannan-Quinn	14,45040		

Ejercicio 2 (zoólogo)

Objetivo

Seguir practicando, ahora de manera menos mecánica, más abierta, la especificación, estimación e interpretación de un MRLS

Enunciado y datos

- Un zoólogo piensa que existe una relación aproximadamente lineal entre los pesos y las longitudes de una especie de mamíferos. Para estudiar esta relación dispone de una muestra formada por veinte ejemplares de los que se dispone las siguientes variables: *peso* y *longitud*. la variable *peso* está expresada en kilogramos y la variable *longitud* en centímetros.
- Los datos están en el fichero `t2_ej_02_zoologo.gdt`.

En la siguiente tabla puedes ver los datos de las 5 primeras observaciones:

EJEMPLAR	peso	longitud
1	2.1	44

2	2.7	49
3	2.5	51
4	1.4	43
5	1.9	39

1. Plantee un MRLS que relacione el peso y la longitud de los animales
2. Estime con ordenador el modelo anterior

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-20				
Variable dependiente: peso				
	coeficiente	Desv. típica	Estadístico t	valor p
const	-0,736887	0,973953	-0,7566	0,4591
longitud	0,0582545	0,0217000	2,685	0,0151 **
Media de la vble. dep.	1,870000	D.T. de la vble. dep.	0,385391	
Suma de cuad. residuos	2,015176	D.T. de la regresión	0,334596	
R-cuadrado	0,285905	R-cuadrado corregido	0,246233	
F(1, 18)	7,206736	Valor p (de F)	0,015140	
Log-verosimilitud	-5,428511	Criterio de Akaike	14,85702	
Criterio de Schwarz	16,84849	Crit. de Hannan-Quinn	15,24578	

3. Interprete el **estimador** de la pendiente
4. Interprete la **estimación** de β_2
5. Interprete el valor del coeficiente de determinación (R^2)
6. ¿Qué peso tendrá un animal de 45 centímetros?
7. ¿Qué peso tendrá un animal de 300 centímetros?
8. Si un animal es tres centímetros **más** largo que otro, ¿qué diferencia de peso cabe esperar?
9. Si un animal es dos centímetros **menos** largo que otro, ¿qué diferencia de peso cabe esperar?
10. Interprete el término independiente (β_1) y su estimación

11. Si el peso se expresase en gramos y la longitud en centímetros, ¿qué estimación se obtendría?
 12. Si el peso se expresase en kilogramos pero la longitud en metros, ¿qué estimación se obtendría?
-

Ejercicio 3 (val_acc-val_con)

Objetivo

Seguir practicando la especificación, estimación e interpretación en un MRLS pero con un caso más realista, en el que aparecen tópicos, que no podemos analizar en profundidad a estas alturas del curso, pero que aparecen más adelante en el curso, como observaciones anómalas, heterocedasticidad y logaritmos. Seguramente pediré este ejercicio para casa.

Datos

- Los datos están en el fichero `t2_ej_03_valac-valcon.gdt`.
- Son datos antiguos pero dan mucho juego por la presencia de un valor anómalo.
- Este fichero recoge observaciones de 161 empresas españolas sobre las siguientes variables:
 - **valor_acc** = Valor total de las acciones de la empresa
 - **valor_contable** = Valor contable de la empresa

1. Análisis previo de los datos:
 - a. ¿qué tipo de datos son?
 - b. Obtén estadísticos descriptivos de las 2 variables.
 - c. Representa las 2 variables en un gráfico de dispersión. ¿hay relación entre ellas?
 - d. ¿Detectas algún valor anómalo?
2. Plantea y estima por MCO un MRLS que relacione las 2 variables:

Pista

Para decidir que variable es el regresando (y) piensa si un aumento del valor de las acciones tiene (inmediatamente) alguna repercusión en el valor contable de la empresa

- a. Escribe la ecuación de la recta o modelo estimado
 - b. Interpreta el coeficiente estimado para β_2
 - c. Interpreta el coeficiente estimado para β_1
 - d. Interpreta el valor de R^2
 - e. ¿Qué problema crees que puede originar en la estimación el valor anómalo?
3. Elimina el valor anómalo y vuelve a estimar el modelo.

Pista

Se trata de estimar el mismo modelo, **pero** eliminando una observación. La de la empresa Nicolás Corr. Para eliminar esa observación has de pinchar en gretl en una de los 2 variables y seleccionar **Editar valores**, para después localizar el dato de Nicolás Correa (es más o menos la observación nº 60) y borrar el valor de la variable. Antes de cerrar la ventana donde estás editando los datos has de darle al icono verde de OK para que se guarden los cambios. Sabrás que lo has hecho bien si cuando estimes el modelo por segunda vez, se utilizan 160 observaciones, una menos que antes.

- a. Observa cómo cambia el coeficiente estimado.
 - b. Representa la variable endógena estimada y los residuos.
 - c. ¿Observas alguna relación entre los residuos y la variable explicativa? Sí, se aprecia que posiblemente haya “heterocedasticidad” (!!!!)
4. Vamos a estimar otra vez el mismo modelo (otra vez sin Nicolás Correa) pero con las las variables del modelo en logaritmos (!!!)

Pista

Se trata de estimar el mismo modelo, sin Nicolas Correa; es decir, con 160 observaciones, **pero** con las **variables en logaritmos**. Para crear las variables en logaritmos has de pinchar en **Añadir > Logaritmos de las variables seleccionadas**.

- a. Estima el modelo en logaritmos (modelo doblemente logarítmico) sin Niscolas Correa, o sea, con 160 observaciones
 - b. Interpreta el coeficiente estimado para β_2
 - c. Visualiza la variable dependiente estimada y los residuos para sacar conclusiones. ¿Hay “heterocedasticidad”?
-

Tema 3: Regresión lineal simple: estadística y contraste de hipótesis

Repaso de teoría (preguntas rápidas)

1. ¿Cuales son las h.e.b? ¿Para qué se necesitan?
2. ¿Cómo son los estimadores MCO si se cumplen las h.e.b? ¿Qué significa que los estimadores sean ELIO? ¿Propiedades probabilísticas? ¿De que depende la varianza de los estimadores?
3. ¿Cómo se distribuyen los estimadores MCO? ¿Cómo se estima la desviación típica de los estimadores?
4. ¿Qué es el ratio t ? ¿Cómo se hace un contraste con el t -ratio? ¿Diferencias entre contrastes a una cola y a 2 colas? ¿Qué es el p-value?

Ejercicio 1 (contrastos con t -ratio)

Objetivo

Recordar y practicar los contrastes de hipótesis con el t -ratio en MRLS

Datos

- Los datos están en el fichero `t3_ej_01_contrastes.gdt`.
- El fichero tiene 3 variables: y , x y z

1. Estime el modelo $y = \beta_1 + \beta_2 x$, e interprete la estimación de β_2

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-44				
Variable dependiente: y				
	coeficiente	Desv. típica	Estadístico t	valor p
const	160,003	16,2649	9,837	1,83e-12 ***
x	-8,40982	1,64481	-5,113	7,37e-06 ***
Media de la vble. dep.	170,6647	D.T. de la vble. dep.	134,6957	
Suma de cuad. residuos	480849,2	D.T. de la regresión	106,9990	
R-cuadrado	0,383642	R-cuadrado corregido	0,368967	
F(1, 42)	26,14224	Valor p (de F)	7,37e-06	
Log-verosimilitud	-267,0139	Criterio de Akaike	538,0278	
Criterio de Schwarz	541,5962	Crit. de Hannan-Quinn	539,3512	

2. Obtenga, “a mano” $\hat{\sigma}_{\beta_2}^2$, Gretl la llama Desv. típica. Para ello antes ha de obtener $\hat{\sigma}^2$, Gretl la llama D.T. de la regresión
3. Contraste la significatividad individual de x
4. Interprete el resultado del contraste anterior ¿qué significa, que implica el resultado?
5. Contraste, otra vez, la significatividad individual de x , pero ahora usando el **p-value**
6. Construya un intervalo de confianza (IC) al 95% para β_2
7. Un IC al 90% ¿será más amplio, o más estrecho? Calcúlelo

Ejercicio 1b

Objetivo

Seguir practicando los contrastes de hipótesis con el t -ratio en MRLS. Esta vez la variable será **no significativa**

Datos

- Usaremos los mismos datos del ejercicio anterior. Están en el fichero `t3_ej_01_contrastes.gdt`.

1. Estime el modelo $y = \beta_1 + \beta_2 z$, e interprete la estimación de β_2

Resultados de estimación

Modelo 2: MCO, usando las observaciones 1-44				
Variable dependiente: y				
	coeficiente	Desv. típica	Estadístico t	valor p
-----	-----	-----	-----	-----
const	40,6873	83,9850	0,4845	0,6306
z	3,09897	1,94507	1,593	0,1186
Media de la vble. dep.	170,6647	D.T. de la vble. dep.	134,6957	
Suma de cuad. residuos	735682,6	D.T. de la regresión	132,3490	
R-cuadrado	0,056994	R-cuadrado corregido	0,034541	
F(1, 42)	2,538416	Valor p (de F)	0,118605	
Log-verosimilitud	-276,3693	Criterio de Akaike	556,7386	
Criterio de Schwarz	560,3070	Crit. de Hannan-Quinn	558,0620	

2. Obtenga, “a mano” $\hat{\sigma}_{\beta_2}^2$, Gretl la llama Desv. típica. Para ello antes ha de obtener $\hat{\sigma}^2$, Gretl la llama D.T. de la regresión
3. Contraste la significatividad individual de z
4. Interprete el resultado del contraste anterior ¿qué significa, que implica el resultado?
5. Contraste, otra vez, la significatividad individual de z , pero ahora usando el **p-value**
6. Construya un intervalo de confianza (IC) al 95% para β_2
7. Un IC al 90% ¿será más amplio, o más estrecho? Calcúlelo

Ejercicio 2 (más contrastes)

Objetivo

Seguir practicando los contrastes de hipótesis con el t -ratio en MRLS. Esta vez sabremos qué son las variables, por lo que podremos hablar de **significatividad práctica** o económica. También realizaremos contrastes para distintos niveles de significación (α).

Datos

- Usaremos los datos del fichero `t3_ej_02_salario.gdt`.
- El fichero tiene 3 variables: *salario*, *educacion*, *experiencia* y *antiguedad*

1. Estime el modelo $\text{salario} = \beta_1 + \beta_2 \text{educacion}$

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-222
Variable dependiente: salario

	coeficiente	Desv. típica	Estadístico t	valor p
const	-0,711695	1,31464	-0,5414	0,5888
educacion	0,555555	0,100524	5,527	9,19e-08 ***
Media de la vble. dep.	6,417928	D.T. de la vble. dep.	4,014527	
Suma de cuad. residuos	3127,524	D.T. de la regresión	3,770413	
R-cuadrado	0,121909	R-cuadrado corregido	0,117918	
F(1, 220)	30,54350	Valor p (de F)	9,19e-08	
Log-verosimilitud	-608,6348	Criterio de Akaike	1221,270	
Criterio de Schwarz	1228,075	Crit. de Hannan-Quinn	1224,017	

2. Contraste la significatividad individual del regresor *educacion* al 5%
3. Contraste la significatividad individual del regresor *educacion* al 10% y al 1%
4. Interprete el resultado de los contrastes anterior ¿qué significa, que implican los resultados?
5. La *educacion* es una variable estadísticamente significativa, para cualquier nivel de significación, pero ¿es significativa en la práctica?
6. Repita los contrastes, pero ahora usando el p -value

Ejercicio 2b

Objetivo

Seguir practicando los contrastes de hipótesis con el t -ratio en MRLS. Esta vez el regresor no será significativo para todos los niveles de significación.

Datos

- Usaremos los datos del fichero `t3_ej_02_salario.gdt`.
- El fichero tiene 3 variables: *salario*, *educacion* y *experiencia*

1. Estime el modelo $\text{salario} = \beta_1 + \beta_2 \text{experiencia}$

Resultados de estimación

Modelo 2: MCO, usando las observaciones 1-222
Variable dependiente: salario

	coeficiente	Desv. típica	Estadístico t	valor p	
const	5,66586	0,414938	13,65	3,74e-31	***
experiencia	0,0477161	0,0201693	2,366	0,0189	**
Media de la vble. dep.	6,417928	D.T. de la vble. dep.	4,014527		
Suma de cuad. residuos	3473,366	D.T. de la regresión	3,973415		
R-cuadrado	0,024809	R-cuadrado corregido	0,020377		
F(1, 220)	5,596936	Valor p (de F)	0,018860		
Log-verosimilitud	-620,2768	Criterio de Akaike	1244,554		
Criterio de Schwarz	1251,359	Crit. de Hannan-Quinn	1247,301		

2. Contraste la significatividad individual del regresor *experiencia* al 5%
3. Contraste la significatividad individual del regresor *experiencia* al 10% y al 1%
4. Interprete el resultado de los contrastes anterior ¿qué significa, que implican los resultados?
5. La *experiencia* es una variable estadísticamente significativa, al menos al 5%, pero ¿es significativa en la práctica?

6. Repita los contrastes, pero ahora usando el p-value
 7. Obtenga un IC para β_2 al 95%. Interprete
 8. Obtenga un IC para β_2 al 99%. Interprete
-

Ejercicio 2c

Objetivo

Seguir practicando los contrastes de hipótesis con el t -ratio en MRLS. Esta vez el regresor no será significativo al 5%.

Datos

- Usaremos los datos del fichero `t3_ej_02_salario.gdt`.
- El fichero tiene 3 variables: *salario*, *educacion* y *experiencia*

1. Estime el modelo $\text{salario} = \beta_1 + \beta_2 \text{antiguedad}$

Resultados de estimación

Modelo 3: MCO, usando las observaciones 1-222

Variable dependiente: salario

	coeficiente	Desv. típica	Estadístico t	valor p
const	6,21296	0,294855	21,07	1,11e-54 ***
antiguedad	0,0480999	0,0286779	1,677	0,0949 *
Media de la vble. dep.	6,417928	D.T. de la vble. dep.	4,014527	
Suma de cuad. residuos	3516,762	D.T. de la regresión	3,998160	
R-cuadrado	0,012626	R-cuadrado corregido	0,008138	
F(1, 220)	2,813155	Valor p (de F)	0,094915	
Log-verosimilitud	-621,6550	Criterio de Akaike	1247,310	
Criterio de Schwarz	1254,115	Crit. de Hannan-Quinn	1250,058	

2. Contraste la significatividad individual del regresor *antiguedad* al 5%
3. Contraste la significatividad individual del regresor *antiguedad* al 1% y al 10%

4. Interprete el resultado de los contrastes anterior ¿qué significa, que implican los resultados?
 5. La *antigüedad* es una variable estadísticamente significativa, aunque solo al 10%, pero ¿es significativa en la práctica?
 6. Repita los contrastes, pero ahora usando el **p-value**
 7. Obtenga un IC para β_2 al 95%. Interprete
 8. Obtenga un IC para β_2 al 90%. Interprete
-

Ejercicio 2d

Objetivo

Seguir practicando los contrastes de hipótesis con el t -ratio en MRLS. Esta vez los contrastes no serán contrastes de significatividad individual.

Datos

- Usaremos los datos del fichero `t3_ej_02_salario.gdt`.
- El fichero tiene 3 variables: *salario*, *educacion* y *experiencia*

1. Estime el modelo $salario = \beta_1 + \beta_2 educacion$

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-222
Variable dependiente: salario

	coeficiente	Desv. típica	Estadístico t	valor p
const	-0,711695	1,31464	-0,5414	0,5888
educacion	0,555555	0,100524	5,527	9,19e-08 ***
Media de la vble. dep.	6,417928	D.T. de la vble. dep.	4,014527	
Suma de cuad. residuos	3127,524	D.T. de la regresión	3,770413	
R-cuadrado	0,121909	R-cuadrado corregido	0,117918	
F(1, 220)	30,54350	Valor p (de F)	9,19e-08	
Log-verosimilitud	-608,6348	Criterio de Akaike	1221,270	
Criterio de Schwarz	1228,075	Crit. de Hannan-Quinn	1224,017	

2. Contraste si $\beta_2 > 0.7$ ¿Puede hacer este contraste con el p-value que ofrece Gretl?
3. Contraste si $\beta_2 > 1$ ¿Puede hacer este contraste con el p-value que ofrece Gretl?
4. Contraste si $\beta_2 < 0$ ¿Puede hacer este contraste con el p-value que ofrece Gretl?
5. Contraste si $\beta_2 < 0.5$ ¿Puede hacer este contraste con el p-value que ofrece Gretl?
6. Construya un IC para β_2 al 95%. Interprete y relacione con los resultados de los apartados anteriores
7. Construya un IC para β_2 al 99%. Interprete y relacione con los resultados de los apartados anteriores

Ejercicio 3 (Examen)

Objetivo

Ver alguna pregunta, en este caso más de tipo teórico, de examen. ¿Verdad que no es muy complicada?

(Examen de julio de 2007). Argumente la veracidad o falsedad de las siguientes afirmaciones:

- a) En un modelo de regresión, los residuos son iguales a las perturbaciones aleatorias

- b) En un modelo de regresión, el coeficiente de determinación (R^2) es igual al coeficiente de determinación corregido (\overline{R}^2)
-

Tema 4: Generalización: regresión lineal múltiple

Repaso de teoría (preguntas rápidas)

1. ¿En qué se diferencian el MRLS y el MRLM?
2. ¿Cómo se interpretan las estimaciones de un MRLM?
3. Diferencie entre R^2 y \bar{R}^2
4. ¿Qué es el estadístico F? ¿Para qué sirve?
5. ¿Cómo podemos efectuar predicciones a partir de un modelo estimado?

Ejercicio 1 (Y-X2-X3)

Objetivo

Ver que en un MRLM la mecánica de los contrastes con el t -ratio es igual que en el MRLS; eso sí, en el modelo múltiple podemos plantear **contrastos con H_0 compuestas**, como por ejemplo, el **contraste de significatividad global del modelo**.

Hay que diferenciar claramente los contrastes de significatividad individual del contraste de significatividad global.

Para hacer contrastes con H_0 compuestas de más de una restricción, usaremos el **estadístico F** o compararemos los **MG y MR**.

Con los resultados mostrados en el siguiente cuadro:

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX				
Modelo 1: MCO, usando las observaciones 1-44				
Variable dependiente: Y				
	coeficiente	Desv. típica	Estadístico t	valor p
const	72,7492	67,5993	1,076	0,2881
X2	2,08891	1,57170	1,329	0,1912
X3	-8,12574	1,64396	-4,943	1,35e-05 ***
Media de la vble. dep.	170,6647	D.T. de la vble. dep.	134,6957	
Suma de cuad. residuos	460987,9	D.T. de la regresión	106,0359	
R-cuadrado	0,409101	R-cuadrado corregido	0,380276	
F(2, 41)	14,19287	Valor p (de F)	0,000021	
Log-verosimilitud	-266,0859	Criterio de Akaike	538,1718	
Criterio de Schwarz	543,5244	Crit. de Hannan-Quinn	540,1568	

1. Contraste **detalladamente** la significatividad conjunta del modelo
 2. Contraste la significatividad de β_2
 3. Contraste la significatividad de β_3
 4. ¿Es β_2 igual a 4?
 5. ¿Es $\beta_3 = -2$?
 6. Estime la varianza de las perturbaciones
-

Ejercicio 2

Objetivo

Seguimos con contrastes, pero esta vez las H_0 no son tan mecánicas o directas.

Con 34 observaciones se ha estimado el siguiente modelo:

$$\widehat{y}_i = \underbrace{2,4}_{(1,1)} + \underbrace{6,2}_{(2,4)} x_{2,i} - \underbrace{5,8}_{(1,2)} x_{3,i} + \underbrace{12,5}_{(2,7)} x_{4,i}$$

donde los valores entre paréntesis son las desviaciones típicas estimadas de los estimadores correspondientes (errores estándar).

1. Contraste la significatividad de la variable x_4
 2. Contraste si x_4 es igual a 10 frente a la posibilidad de que sea mayor que 10
 3. Contraste si x_4 es igual a 2 frente a la posibilidad de que sea menor que 2
 4. Cuando x_3 aumenta una unidad, ¿la variable Y disminuye 4 unidades?
-

Ejercicio 3 (sqrt-bdrms)

Objetivo

Ver alguna pregunta de examen, en este caso más de tipo práctico. ¿Verdad que no es muy complicada? (!!!)

(Examen de enero de 2007) Con una muestra formada por 88 viviendas de una determinada zona se ha obtenido la siguiente estimación: $\hat{P} = 19,3 + 0,128\text{sqft} + 15,2\text{bdrms}$. Donde:

- **P**: precio de la vivienda en miles de dólares USA
 - **sqft**: superficie de la vivienda en pies cuadrados
 - **bdrms**: número de dormitorios de la vivienda
- a) Si se mantiene fija la superficie total de la vivienda, ¿cuál será el aumento estimado del precio de una vivienda si se le dota de un dormitorio adicional?
- b) ¿Cuál será el aumento estimado del precio de una vivienda si se construye un dormitorio adicional aumentando también la superficie de la vivienda en 100 pies cuadrados?
- c) Una vivienda de 2.500 pies cuadrados y 4 dormitorios se pone a la venta por 270.000 dólares. ¿Qué opina del precio?
- d) Exprese el modelo estimado utilizando como unidad de medida de superficie el metro cuadrado (sqmts) y expresando el precio en miles de euros (!!!)

NOTAS: 1 pie cuadrado equivale aproximadamente a 0.1 metros cuadrados. Utilice como tipo de cambio 1 euro = 1.28 dólares USA.

Ejercicio 4 (salario-educación)

Objetivo

Practicar, otra vez, interpretación de coeficientes y contrastes de hipótesis en el marco del MRLM. Seguramente este ejercicio lo pediré para casa.

Datos

- Usaremos los datos del fichero `t4_ej_04_educacion.gdt`.
- El fichero tiene 4 variables: *salario*, *educacion*, *experiencia* y *antiguedad*

1. Estime el modelo $\text{salario} = \beta_1 + \beta_2 \text{educacion} + u$ e interprete la estimación de β_2

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-526

Variable dependiente: salario

	coeficiente	Desv. típica	Estadístico t	valor p
const	-0,904852	0,684968	-1,321	0,1871
educacion	0,541359	0,0532480	10,17	2,78e-22 ***
Media de la vble. dep.	5,896103	D.T. de la vble. dep.	3,693086	
Suma de cuad. residuos	5980,682	D.T. de la regresión	3,378390	
R-cuadrado	0,164758	R-cuadrado corregido	0,163164	
F(1, 524)	103,3627	Valor p (de F)	2,78e-22	
Log-verosimilitud	-1385,712	Criterio de Akaike	2775,423	
Criterio de Schwarz	2783,954	Crit. de Hannan-Quinn	2778,764	

2. Ahora estime el modelo

$$\text{salario} = \beta_1 + \beta_2 \text{educacion} + \beta_3 \text{experiencia} + \beta_4 \text{antiguedad} + u$$

e interprete la estimación de β_2

Resultados de estimación

Modelo 3: MCO, usando las observaciones 1-526				
Variable dependiente: salario				
	coeficiente	Desv. típica	Estadístico t	valor p
const	-2,87273	0,728964	-3,941	9,22e-05 ***
educacion	0,598965	0,0512835	11,68	3,68e-28 ***
experiencia	0,0223395	0,0120568	1,853	0,0645 *
antigüedad	0,169269	0,0216446	7,820	2,93e-14 ***
Media de la vble. dep.	5,896103	D.T. de la vble. dep.	3,693086	
Suma de cuad. residuos	4966,303	D.T. de la regresión	3,084476	
R-cuadrado	0,306422	R-cuadrado corregido	0,302436	
F(3, 522)	76,87317	Valor p (de F)	3,41e-41	
Log-verosimilitud	-1336,831	Criterio de Akaike	2681,662	
Criterio de Schwarz	2698,723	Crit. de Hannan-Quinn	2688,342	

- Interprete la estimación de β_2 en el segundo modelo. ¿Por qué son diferentes las estimaciones? ¿Son diferentes los estimadores de β_2 en los 2 modelos? ¿Son diferentes los β_2 en los 2 modelos?
- Realice y explique **con palabras** de forma detallada el contraste la significatividad individual de la variable educación en el segundo modelo.
- La estimación puntual de β_2 es 0,598 ¿es la **estimación** óptima? Explique.
- ¿Cómo ha calculado Gretl la “Desv. Típica” de $\hat{\beta}_2$? Utiliza Gretl para calcularla tú mismo. (!!!!)
- Realice y explique con palabras de forma detallada el contraste de significatividad individual de la variable experiencia.
- Vuelve a efectuar el contraste de significatividad individual de la variable experiencia, pero ahora al 10%.
- Utilice el “valor p” (también llamado p-value o nivel de significación crítico) para contrastar la significatividad individual de la variable antigüedad.
- Contraste si el efecto de la educación es menor que 0,8
- Contraste si el efecto de la educación es menor que 0,6
- Construya un intervalo de confianza al 95% para el efecto de la educación
- Contraste si el efecto de la antigüedad es mayor que cero
- Contraste si el efecto de la antigüedad es mayor que 0,15

15. Construya un intervalo de confianza al 95% para el efecto de la experiencia. Interprete
16. Construya un intervalo de confianza al 90% para el efecto de la experiencia. Interprete
17. Contraste si el efecto de la educación es igual al efecto de la experiencia. (!!!!)

Pista

Este contraste no se puede hacer con el t -ratio tal y como lo hemos explicado en clase. Hay que usar el estadístico F (que tampoco lo hemos enseñado a hacer a mano) o comparando el MG y MR. Lo podéis hacer de dos maneras:

- comparando el MG y MR
- usando Gretl: en la ventana de estimación del segundo modelo, pincha en la pestaña `Contraste`

Ejercicio 5 (MG-MR)

Objetivo

Empezar a realizar contrastes mediante la comparación del MG y del MR

Se plantea el siguiente modelo:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$$

El modelo se estimó con 34 observaciones y se obtuvo una SCR de 57,29.

1. Se quiere contrastar la siguiente hipótesis: $\beta_2 = \beta_3 + \beta_4$. Obtenga el modelo restringido.
2. Tras estimar el modelo restringido anterior, su SCR fue de 69,11. Realice el contraste de la hipótesis al 5% y al 1%.
3. Escriba el modelo restringido necesario para poder estimar la siguiente hipótesis nula:

$$H_0 = \begin{cases} \beta_2 = \beta_3 + \beta_4 \\ \beta_3 = \beta_4 \end{cases}$$
4. La SCR del modelo restringido que incorpora las dos restricciones del aparatado anterior fue de 74,35. Realice el contraste para un nivel de significación del 5% y del 1%.

Ejercicio 6

Objetivo

Seguir con contrastes comparando MG y MR. Esta vez en lugar de tener la H_0 tenemos el MR y hemos de recuperar la H_0 . (!!!)

Considere el siguiente modelo de regresión lineal:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

A partir de una muestra de 26 observaciones se han obtenido por MCO los siguientes 2 modelos estimados:

$\hat{y}_t = 2.0 + 3.5x_{1,i} - 0.7x_{2,i} - 2.0x_{3,i}$ (1.8) (0.3) (1.3)	$R^2 = 0.982$
$\hat{y}_t = 1.5 + 3.0x_{1,i} + x_{2,i} - 0.6x_{3,i}$ (1.1) (0.3)	$R^2 = 0.876$

donde las cifras entre paréntesis representan las desviaciones típicas de los estimadores. Se pide:

1. ¿Qué restricciones lineales incorpora el segundo modelo respecto al primero?
2. Contraste la hipótesis nula especificada en el apartado previo (!!)

Ejercicio 7 (sleep)

Objetivo

Con este ejercicio podemos hacer un repaso rápido a las distintas opciones que tenemos para hacer contrastes de hipótesis, además de una forma menos mecánica que en otros ejercicios y que hace evidente que para interpretar los resultados nunca hay que olvidarse de la escala en la que están medidas las variables.

Datos

- Usaremos los datos del fichero `t4_ej_07_sleep75.gdt`.
- Los datos provienen del libro Introductory Econometrics de J. M. Wooldridge.
- Los datos incluyen, entre otras, las siguientes variables de una muestra de 706 personas adultas:
 - **SLEEP**: Tiempo de sueño cada semana (en minutos)
 - **EDUC**: Años de educación
 - **TOTWRK**: Tiempo de trabajo cada semana (en minutos)
 - **AGE**: Edad en años

Con estos datos estime el siguiente modelo: $SLEEP_i = \beta_1 + \beta_2 TOTWORK_i + \beta_3 EDUC_i + \beta_4 AGE_i + u_i$

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-706				
Variable dependiente: SLEEP				
	coeficiente	Desv. típica	Estadístico t	valor p
const	3638,25	112,275	32,40	1,47e-141 ***
TOTWRK	-0,148373	0,0166935	-8,888	5,19e-18 ***
EDUC	-11,1338	5,88457	-1,892	0,0589 *
AGE	2,19988	1,44572	1,522	0,1285
Media de la vble. dep.	3266,356	D.T. de la vble. dep.	444,4134	
Suma de cuad. residuos	1,23e+08	D.T. de la regresión	419,3589	
R-cuadrado	0,113364	R-cuadrado corregido	0,109575	
F(3, 702)	29,91889	Valor p (de F)	3,28e-18	
Log-verosimilitud	-5263,106	Criterio de Akaike	10534,21	
Criterio de Schwarz	10552,45	Crit. de Hannan-Quinn	10541,26	

1. ¿Cuánto dormirá diariamente una persona de 20 años de edad, con 14 de educación y que trabaja 40 horas semanales?
2. Analice la significatividad de cada una de las variables
3. Analice la significatividad conjunta del Modelo 1
4. Al trabajar más, ¿se duerme menos?
5. ¿Una hora adicional de trabajo implica quince minutos menos de sueño?

6. ¿Una hora adicional de trabajo implica diez minutos menos de sueño?
7. ¿Influyen la educación y la edad conjuntamente en el tiempo de sueño?

Pista

Para el apartado 7, seguramente te hará falta el siguiente cuadro:

Modelo 2: MCO, usando las observaciones 1-706				
Variable dependiente: SLEEP				
	coeficiente	Desv. típica	Estadístico t	valor p
const	3586,38	38,9124	92,17	0,0000 ***
TOTWRK	-0,150746	0,0167403	-9,005	1,99e-18 ***
Media de la vble. dep.	3266,356	D.T. de la vble. dep.	444,4134	
Suma de cuad. residuos	1,25e+08	D.T. de la regresión	421,1357	
R-cuadrado	0,103287	R-cuadrado corregido	0,102014	
F(1, 704)	81,08987	Valor p (de F)	1,99e-18	
Log-verosimilitud	-5267,096	Criterio de Akaike	10538,19	
Criterio de Schwarz	10547,31	Crit. de Hannan-Quinn	10541,71	

Tema 5: Modelos no lineales y transformación de variables

Repaso de teoría (preguntas rápidas)

1. ¿Qué ocurre si se cambia la escala de un regresor? ¿Y si se cambia la escala del regresando?
2. ¿Por qué algunos modelos se estiman en logaritmos? ¿Cómo se interpretan los β en esos parámetros?
3. ¿Qué es el AIC? ¿Para qué sirve? ¿Es similar a R^2 ? ¿a $\overline{R^2}$?

Ejercicio 1 (sleep)

Objetivo

Con este ejercicio, que utiliza datos ya conocidos, recordaremos cuales son los efectos de cambios de escala en regresores y regresando.

Datos

- Usaremos los datos del fichero `t5_ej_01_sleep75.gdt`.
- Los datos provienen del libro *Introductory Econometrics* de J. M. Wooldridge.
- Los datos incluyen, entre otras, las siguientes variables de una muestra de 706 personas adultas:
 - **SLEEP**: Tiempo de sueño cada semana (en minutos)
 - **EDUC**: Años de educación
 - **TOTWRK**: Tiempo de trabajo cada semana (en minutos)
 - **AGE**: Edad en años

Con estos datos estime el siguiente modelo: $SLEEP_i = \beta_1 + \beta_2 TOTWORK_i + \beta_3 EDUC_i + \beta_4 AGE_i + u_i$

1. Interprete la estimación de β_2

2. ¿Cómo cambiará la estimación de β_2 si el tiempo de trabajo pasase a medirse en horas por semana?
3. ¿Cómo cambiará la estimación de β_2 si el tiempo de trabajo pasase a medirse en días por semana?
4. ¿Cómo cambiará la estimación de β_2 si el tiempo de sueño pasase a medirse en horas por semana?
5. ¿Cómo cambiará la estimación de β_2 si las dos variables, tanto el tiempo de sueño como el tiempo de trabajo se midiesen en horas por semana?

Ejercicio 2 (logs)

Objetivo

Afianzar la interpretación de parámetros en modelos con variables en logaritmos

Datos

- Usaremos los datos del fichero `t5_ej_02_alim.gdt`.
- Los datos incluyen las siguientes variables (y sus logaritmos):
 - **ALIM**: Gasto familiar en alimentos (en euros anuales)
 - **RDISP**: Renta familiar disponible (en euros anuales)

El archivo `t5_ej_02_alim.gdt` contiene las siguientes variables:

Estima e interpreta la estimación de β_2 en los siguientes modelos:

1. $ALIM_i = \beta_1 + \beta_2 RDISP_i$

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-80				
Variable dependiente: ALIM				
	coeficiente	Desv. típica	Estadístico t	valor p
const	1766,59	433,700	4,073	0,0001 ***
RDISP	0,211427	0,0163746	12,91	4,67e-21 ***
Media de la vble. dep.	6975,438	D.T. de la vble. dep.	2506,518	
Suma de cuad. residuos	1,58e+08	D.T. de la regresión	1424,139	
R-cuadrado	0,681264	R-cuadrado corregido	0,677178	
F(1, 78)	166,7169	Valor p (de F)	4,67e-21	
Log-verosimilitud	-693,4082	Criterio de Akaike	1390,816	
Criterio de Schwarz	1395,580	Crit. de Hannan-Quinn	1392,726	

$$2. \ln(ALIM_i) = \beta_1 + \beta_2 RDISP_i$$

Resultados de estimación

Modelo 2: MCO, usando las observaciones 1-80				
Variable dependiente: l_ALIM				
	coeficiente	Desv. típica	Estadístico t	valor p
const	7,93772	0,0789756	100,5	2,79e-84 ***
RDISP	3,39104e-05	2,98177e-06	11,37	3,13e-18 ***
Media de la vble. dep.	8,773155	D.T. de la vble. dep.	0,420126	
Suma de cuad. residuos	5,245737	D.T. de la regresión	0,259332	
R-cuadrado	0,623798	R-cuadrado corregido	0,618975	
F(1, 78)	129,3356	Valor p (de F)	3,13e-18	
Log-verosimilitud	-4,530646	Criterio de Akaike	13,06129	
Criterio de Schwarz	17,82534	Crit. de Hannan-Quinn	14,97134	

$$3. \ln(ALIM_i) = \beta_1 + \beta_2 \ln(RDISP_i)$$

Resultados de estimación

```

Modelo 3: MCO, usando las observaciones 1-80
Variable dependiente: l_ALIM

      coeficiente  Desv. típica  Estadístico t  valor p
-----
const      1,32709      0,601309      2,207      0,0303  **
l_RDISP    0,743112      0,0599478     12,40     4,01e-20  ***

Media de la vble. dep.  8,773155  D.T. de la vble. dep.  0,420126
Suma de cuad. residuos  4,694919  D.T. de la regresión   0,245339
R-cuadrado              0,663301  R-cuadrado corregido   0,658984
F(1, 78)                153,6607  Valor p (de F)         4,01e-20
Log-verosimilitud       -0,093250  Criterio de Akaike     4,186500
Criterio de Schwarz      8,950554  Crit. de Hannan-Quinn  6,096546
  
```

$$4. ALIM_i = \beta_1 + \beta_2 \ln(RDISP_i)$$

Resultados de estimación

```

Modelo 4: MCO, usando las observaciones 1-80
Variable dependiente: ALIM

      coeficiente  Desv. típica  Estadístico t  valor p
-----
const     -36820,8      3685,32      -9,991      1,31e-15  ***
l_RDISP    4370,83       367,410      11,90      3,32e-19  ***

Media de la vble. dep.  6975,438  D.T. de la vble. dep.  2506,518
Suma de cuad. residuos  1,76e+08  D.T. de la regresión   1503,642
R-cuadrado              0,644684  R-cuadrado corregido   0,640129
F(1, 78)                141,5229  Valor p (de F)         3,32e-19
Log-verosimilitud       -697,7540  Criterio de Akaike     1399,508
Criterio de Schwarz      1404,272  Crit. de Hannan-Quinn  1401,418
  
```

Ejercicio 3 (ex)

Objetivo

Ejercicio de examen sobre interpretación de coeficientes en modelos con variables en logaritmos.

(Examen de julio de 2008) Razona la veracidad o falsedad de las siguientes afirmaciones, en el siguiente modelo:

$$\ln(y_i) = \beta_1 + \beta_2 x_{2i} + \beta_3 \ln(x_{3i}) + u_i$$

- a) Cuando la variable x_2 aumenta una unidad, la variable y experimenta una tasa de variación porcentual igual a β_2
 - b) Cuando la variable x_3 aumenta una unidad, la variable y aumenta β_3 unidades
-

Ejercicio 4 (ex)

Objetivo

Otro ejercicio de examen sobre interpretación de coeficientes en modelos con variables en logaritmos.

(Examen de junio de 2012). Argumente la veracidad o falsedad de las siguientes afirmaciones utilizando los siguientes modelos de regresión estimados:

$$\widehat{\ln(y_i)} = 3 + 0,17x_{2i} - 1,4\ln(x_{3i})$$

$$\widehat{y_i} = 2 - 9,3\ln(x_{4i}) + 3,5x_{5i}$$

- a) Cuando la variable x_2 aumenta en una unidad, la variable y experimenta una tasa de variación igual al 17%
 - b) Cuando x_4 disminuye en un 9,3%, la variable y disminuye en una unidad
-

Ejercicio 5 (Cobb)

Objetivo

Además de seguir trabajando en la interpretación de coeficientes en modelos con variables en logaritmos, en este ejercicio se aprecia muy bien la diferencia entre modelo no-lineal en parámetros y modelo lineal en parámetros (aunque con regresores en logaritmos).

Dada la función de producción Cobb-Douglas $Q = AL^\alpha K^\beta e^u$, en la que Q designa la producción, L la ocupación y K el capital:

1. Linealice el modelo anterior
 2. Interprete los parámetros α y β
-

Ejercicio 6 (edad-mates)

Objetivo

El segundo modelo de este ejercicio incluye un término cuadrático. Esto nos sirve para comprender mejor la diferencia entre regresor y variable explicativa; además, volvemos a trabajar la idea de que un modelo puede ser lineal en parámetros aunque no sea “lineal en variables”.

Datos

Se desea analizar la relación existente entre la edad de los maestros y la calificación obtenida en el examen de matemáticas por sus estudiantes.

Usaremos los datos del fichero `t5_ej_06_mates.gdt` que contiene una muestra formada por 1.000 observaciones de las siguientes variables:

- **edad** : Edad del maestro en años
- **nota**: Puntuación obtenida por el estudiante en una prueba de matemáticas (escala de 0 a 10)
- **sq_edad** : $nota^2$ (la puntuación al cuadrado, la nota al cuadrado)

1. Plantee y estime un modelo de regresión que relacione la edad del maestro con la puntuación en el examen de los estudiantes

Pista

Para el apartado 7, seguramente te hará falta el siguiente cuadro:

Modelo 1: MCO, usando las observaciones 1-1000				
Variable dependiente: nota				
	coeficiente	Desv. típica	Estadístico t	valor p
const	7,88207	0,349578	22,55	2,59e-91 ***
edad	-0,0441001	0,00761936	-5,788	9,54e-09 ***
Media de la vble. dep.	5,887600	D.T. de la vble. dep.	1,890185	
Suma de cuad. residuos	3453,309	D.T. de la regresión	1,860169	
R-cuadrado	0,032477	R-cuadrado corregido	0,031507	
F(1, 998)	33,49984	Valor p (de F)	9,54e-09	
Log-verosimilitud	-2038,605	Criterio de Akaike	4081,210	
Criterio de Schwarz	4091,026	Crit. de Hannan-Quinn	4084,941	

- Interprete el estimador de la pendiente
- Introduce en el modelo un nuevo regresor, concretamente la edad del maestro en forma cuadrática; es decir, la variable **sq_edad**. En este nuevo modelo interpreta la relación entre la edad de los maestros y la puntuación

Pista

Para el apartado 7, seguramente te hará falta el siguiente cuadro:

Modelo 2: MCO, usando las observaciones 1-1000				
Variable dependiente: nota				
	coeficiente	Desv. típica	Estadístico t	valor p
const	0,235757	1,47101	0,1603	0,8727
edad	0,304468	0,0656215	4,640	3,95e-06 ***
sq_edad	-0,00385655	0,000721257	-5,347	1,11e-07 ***
Media de la vble. dep.	5,887600	D.T. de la vble. dep.	1,890185	
Suma de cuad. residuos	3357,042	D.T. de la regresión	1,834978	
R-cuadrado	0,059448	R-cuadrado corregido	0,057562	
F(2, 997)	31,50812	Valor p (de F)	5,39e-14	
Log-verosimilitud	-2024,469	Criterio de Akaike	4054,937	
Criterio de Schwarz	4069,660	Crit. de Hannan-Quinn	4060,533	

- En el modelo del tercer apartado ¿Cómo se interpreta β_2 ?
- ¿Cuál sería la edad óptima de los maestros?
- ¿Qué notas cabe esperar en los alumnos de maestros de 30, 40, 50 y 60 años?

7. Comente los resultados del ejercicio (no se olvide de comparar los modelos 1 y 2 y decir que modelo le parece más razonable)

Tema 6: Análisis con información cualitativa

Repaso de teoría (preguntas rápidas)

1. ¿Qué son las variable ficticias? ¿Para qué sirven?
2. ¿Qué es la categoría de referencia? ¿Cómo se interpreta el β que acompaña a una dummy?
3. ¿La interpretación del coeficiente difiere si la dummy es aditiva o es multiplicativa?
4. Si tenemos una característica con q grupos ¿Cuántas dummies podemos definir? ¿Cuál será la categoría de referencia?
5. ¿Qué ocurre si queremos introducir en el modelo varias variables cualitativas? ¿Estas dummies pueden interactuar?

Ejercicio 1 (bebés)

Objetivo

Empezar a trabajar con modelos con dummies. Es un ejercicio muy pautado en el que se van introduciendo poco a poco más dummies.

Datos

- Usaremos los datos del fichero `t6_ej_01_bebes.gdt`. Son datos relacionados con nacimientos de bebés.
- Entre otras puedes encontrar las siguientes **variables cuantitativas**:
 - **peso**: peso del bebé en gramos
 - **semanas**: semanas de gestación del bebé
 - **edad**: edad de la madre en años
- Las variables cuantitativas **también se incluyen en logaritmos**
- Además contiene **variables ficticias** para poder introducir en el modelo distintos aspectos cualitativos:
 - el **genero del bebé** (con las variables dummy **bebida** y **bebido**)

- el **nivel de estudios de la madre** (con las dummies **primarios**, **secundarios** y **superiores**)
- si el parto se realizó con **cesarea** (con las dummies **cesarea** y **vaginal**)

Con estos datos se analizará la influencia de distintas variables sobre el peso con el que finalmente nacen los bebés. Para ello estime e interprete los siguientes modelos:

Aditivas

$$1. \text{ peso} = \beta_1 + \beta_2 \text{semanas} + \beta_3 \text{bebito} + u$$

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-1000				
Variable dependiente: peso				
	coeficiente	Desv. típica	Estadístico t	valor p
const	-2241,33	292,312	-7,668	4,15e-14 ***
semanas	140,081	7,45934	18,78	1,33e-67 ***
bebito	121,598	27,6313	4,401	1,19e-05 ***
Media de la vble. dep.	3273,443	D.T. de la vble. dep.	507,1946	
Suma de cuad. residuos	1,89e+08	D.T. de la regresión	435,1485	
R-cuadrado	0,265392	R-cuadrado corregido	0,263919	
F(2, 997)	180,0935	Valor p (de F)	1,69e-67	
Log-verosimilitud	-7493,124	Criterio de Akaike	14992,25	
Criterio de Schwarz	15006,97	Crit. de Hannan-Quinn	14997,84	

$$2. \log(\text{peso}) = \beta_1 + \beta_2 \text{semanas} + \beta_3 \text{bebito} + u$$

Resultados de estimación

Modelo 2: MCO, usando las observaciones 1-1000
Variable dependiente: l_peso

	coeficiente	Desv. típica	Estadístico t	valor p	
const	5,92999	0,100429	59,05	0,0000	***
semanas	0,0547018	0,00256279	21,34	1,49e-83	***
bebito	0,0388867	0,00949322	4,096	4,54e-05	***
Media de la vble. dep.	8,079256	D.T. de la vble. dep.		0,180579	
Suma de cuad. residuos	22,28408	D.T. de la regresión		0,149503	
R-cuadrado	0,315940	R-cuadrado corregido		0,314567	
F(2, 997)	230,2368	Valor p (de F)		6,23e-83	
Log-verosimilitud	483,0030	Criterio de Akaike		-960,0059	
Criterio de Schwarz	-945,2826	Crit. de Hannan-Quinn		-954,4100	

$$3. \text{ peso} = \beta_1 + \beta_2 \text{semanas} + \beta_3 \text{secundarios} + \beta_4 \text{superiores} + u$$

Resultados de estimación

Modelo 3: MCO, usando las observaciones 1-1000
Variable dependiente: peso

	coeficiente	Desv. típica	Estadístico t	valor p	
const	-2114,01	293,183	-7,211	1,10e-12	***
semanas	137,668	7,49701	18,36	4,32e-65	***
medios	27,1338	33,4155	0,8120	0,4170	
superiores	63,3781	33,6282	1,885	0,0598	*
Media de la vble. dep.	3273,443	D.T. de la vble. dep.		507,1946	
Suma de cuad. residuos	1,92e+08	D.T. de la regresión		438,7933	
R-cuadrado	0,253784	R-cuadrado corregido		0,251536	
F(3, 996)	112,9114	Valor p (de F)		6,19e-63	
Log-verosimilitud	-7500,963	Criterio de Akaike		15009,93	
Criterio de Schwarz	15029,56	Crit. de Hannan-Quinn		15017,39	

$$4. \log(\text{peso}) = \beta_1 + \beta_2 \text{semanas} + \beta_3 \text{secundarios} + \beta_4 \text{superiores} + u$$

Resultados de estimación

Modelo 4: MCO, usando las observaciones 1-1000

Variable dependiente: l_peso

	coeficiente	Desv. típica	Estadístico t	valor p	
const	5,96366	0,100465	59,36	0,0000	***
semanas	0,0539837	0,00256899	21,01	1,96e-81	***
medios	0,0173593	0,0114504	1,516	0,1298	
superiores	0,0281573	0,0115233	2,443	0,0147	**
Media de la vble. dep.	8,079256	D.T. de la vble. dep.	0,180579		
Suma de cuad. residuos	22,51792	D.T. de la regresión	0,150361		
R-cuadrado	0,308761	R-cuadrado corregido	0,306679		
F(3, 996)	148,2971	Valor p (de F)	1,91e-79		
Log-verosimilitud	477,7833	Criterio de Akaike	-947,5667		
Criterio de Schwarz	-927,9357	Crit. de Hannan-Quinn	-940,1055		

$$5. \text{ peso} = \beta_1 + \beta_2 \text{semanas} + \beta_3 \text{bebido} + \beta_4 \text{cesarea} + u$$

Resultados de estimación

Modelo 5: MCO, usando las observaciones 1-1000

Variable dependiente: peso

	coeficiente	Desv. típica	Estadístico t	valor p	
const	-2285,13	296,096	-7,718	2,87e-14	***
semanas	141,014	7,52703	18,73	2,54e-67	***
bebido	121,018	27,6402	4,378	1,32e-05	***
cesarea	29,3659	31,5577	0,9305	0,3523	
Media de la vble. dep.	3273,443	D.T. de la vble. dep.	507,1946		
Suma de cuad. residuos	1,89e+08	D.T. de la regresión	435,1778		
R-cuadrado	0,266030	R-cuadrado corregido	0,263820		
F(3, 996)	120,3348	Valor p (de F)	1,67e-66		
Log-verosimilitud	-7492,689	Criterio de Akaike	14993,38		
Criterio de Schwarz	15013,01	Crit. de Hannan-Quinn	15000,84		

Multiplicativas

$$6. \text{ peso} = \beta_1 + \beta_2 \text{semanas} + \beta_3(\text{semanas} * \text{bebito}) + u$$

Resultados de estimación

Modelo 6: MCO, usando las observaciones 1-1000				
Variable dependiente: peso				
	coeficiente	Desv. típica	Estadístico t	valor p
const	-2171,24	290,584	-7,472	1,72e-13 ***
semanas	138,289	7,43415	18,60	1,57e-66 ***
semanas_x_bebito	3,11959	0,708412	4,404	1,18e-05 ***
Media de la vble. dep.	3273,443	D.T. de la vble. dep.	507,1946	
Suma de cuad. residuos	1,89e+08	D.T. de la regresión	435,1431	
R-cuadrado	0,265411	R-cuadrado corregido	0,263937	
F(2, 997)	180,1105	Valor p (de F)	1,67e-67	
Log-verosimilitud	-7493,111	Criterio de Akaike	14992,22	
Criterio de Schwarz	15006,95	Crit. de Hannan-Quinn	14997,82	

$$7. \log(\text{peso}) = \beta_1 + \beta_2 \text{semanas} + \beta_3(\text{semanas} * \text{bebito}) + u$$

Resultados de estimación

Modelo 7: MCO, usando las observaciones 1-1000				
Variable dependiente: l_peso				
	coeficiente	Desv. típica	Estadístico t	valor p
const	5,95240	0,0998352	59,62	0,0000 ***
semanas	0,0541288	0,00255413	21,19	1,39e-82 ***
semanas_x_bebito	0,000997775	0,000243387	4,100	4,48e-05 ***
Media de la vble. dep.	8,079256	D.T. de la vble. dep.	0,180579	
Suma de cuad. residuos	22,28348	D.T. de la regresión	0,149501	
R-cuadrado	0,315958	R-cuadrado corregido	0,314586	
F(2, 997)	230,2561	Valor p (de F)	6,15e-83	
Log-verosimilitud	483,0162	Criterio de Akaike	-960,0324	
Criterio de Schwarz	-945,3091	Crit. de Hannan-Quinn	-954,4365	

Aditivas y multiplicativas

$$9. \text{ peso} = \beta_1 + \beta_2 \text{semanas} + \beta_3 \text{bebito} + \beta_4(\text{semanas} * \text{bebito}) + u$$

Resultados de estimación

Modelo 9: MCO, usando las observaciones 1-1000

Variable dependiente: peso

	coeficiente	Desv. típica	Estadístico t	valor p	
const	-2185,88	447,495	-4,885	1,21e-06	***
semanas	138,663	11,4338	12,13	1,14e-31	***
bebito	25,3373	588,650	0,04304	0,9657	
semanas_x_bebito	2,47070	15,0920	0,1637	0,8700	
Media de la vble. dep.	3273,443	D.T. de la vble. dep.	507,1946		
Suma de cuad. residuos	1,89e+08	D.T. de la regresión	435,3611		
R-cuadrado	0,265412	R-cuadrado corregido	0,263199		
F(3, 996)	119,9541	Valor p (de F)	2,54e-66		
Log-verosimilitud	-7493,110	Criterio de Akaike	14994,22		
Criterio de Schwarz	15013,85	Crit. de Hannan-Quinn	15001,68		

$$10. \log(\text{peso}) = \beta_1 + \beta_2 \text{semanas} + \beta_3 \text{bebito} + \beta_4 (\text{semanas} * \text{bebito}) + u$$

Resultados de estimación

Modelo 10: MCO, usando las observaciones 1-1000

Variable dependiente: l_peso

	coeficiente	Desv. típica	Estadístico t	valor p	
const	5,94917	0,153745	38,70	1,13e-200	***
semanas	0,0542113	0,00392827	13,80	9,01e-40	***
bebito	0,00559727	0,202241	0,02768	0,9779	
semanas_x_bebito	0,000854429	0,00518513	0,1648	0,8691	
Media de la vble. dep.	8,079256	D.T. de la vble. dep.	0,180579		
Suma de cuad. residuos	22,28347	D.T. de la regresión	0,149576		
R-cuadrado	0,315958	R-cuadrado corregido	0,313898		
F(3, 996)	153,3505	Valor p (de F)	1,06e-81		
Log-verosimilitud	483,0166	Criterio de Akaike	-958,0332		
Criterio de Schwarz	-938,4022	Crit. de Hannan-Quinn	-950,5720		

$$11. \log(\text{peso}) = \beta_1 + \beta_2 \log(\text{semanas}) + \beta_3 \text{bebito} + \beta_4 (\log(\text{semanas}) * \text{bebito}) + u$$

Resultados de estimación

Modelo 11: MCO, usando las observaciones 1-1000

Variable dependiente: l_peso

	coeficiente	Desv. típica	Estadístico t	valor p
const	0,709448	0,518739	1,368	0,1717
l_semanas	2,00797	0,141523	14,19	9,42e-42 ***
bebito	0,121575	0,678052	0,1793	0,8577
l_semanas_x_bebi~	-0,0226611	0,185170	-0,1224	0,9026
Media de la vble. dep.	8,079256	D.T. de la vble. dep.	0,180579	
Suma de cuad. residuos	21,94258	D.T. de la regresión	0,148427	
R-cuadrado	0,326423	R-cuadrado corregido	0,324394	
F(3, 996)	160,8906	Valor p (de F)	4,97e-85	
Log-verosimilitud	490,7246	Criterio de Akaike	-973,4492	
Criterio de Schwarz	-953,8182	Crit. de Hannan-Quinn	-965,9881	

Ejercicio 2 (trampa)

Objetivo

Trabajar en comprender la situación conocida como **trampa de las ficticias**.

Indique y explique detalladamente en cuales de los siguientes modelos se incurre en la trampa de las ficticias:

1. $\text{peso} = \beta_1 + \beta_2 \text{semanas} + \beta_3 \text{bebito} + \beta_4 \text{bebida} + u$
2. $\text{peso} = \beta_1 + \beta_2 \text{semanas} + \beta_3 (\text{semanas} * \text{bebito}) + \beta_4 \text{bebito} + u$
3. $\text{peso} = \beta_1 + \beta_2 \text{semanas} + \beta_3 (\text{semanas} * \text{bebito}) + \beta_4 \text{bebida} + u$
4. $\text{peso} = \beta_1 + \beta_2 (\text{semanas} * \text{primarios}) + \beta_3 \text{secundarios} + \beta_4 \text{superiores} + u$

Ejercicio 3 (salario)

Objetivo

Seguir trabajando con dummies, ahora con preguntas más abiertas.

Datos

Usaremos los datos del fichero `t6_ej_03_salario.gdt`. Tiene información de las siguientes variables:

- **SALARI**: Salario del trabajador en euros brutos anuales.
- **EXPLAB**: Experiencia laboral del trabajador en años.
- **GENERE**: Variable ficticia que toma el valor 1 para los hombres y 0 para las mujeres

1. Estime el siguiente modelo $SALARI_i = \beta_1 + \beta_2 EXPLAB_i + \beta_3 GENERE_i + u_i$

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-200

Variable dependiente: SALARI

	coeficiente	Desv. típica	Estadístico t	valor p
const	23712,7	1414,39	16,77	7,54e-40 ***
EXPLAB	434,197	68,2652	6,360	1,36e-09 ***
Media de la vble. dep.	31510,89	D.T. de la vble. dep.	10917,20	
Suma de cuad. residuos	1,97e+10	D.T. de la regresión	9973,194	
R-cuadrado	0,169655	R-cuadrado corregido	0,165462	
F(1, 198)	40,45517	Valor p (de F)	1,36e-09	
Log-verosimilitud	-2124,314	Criterio de Akaike	4252,628	
Criterio de Schwarz	4259,224	Crit. de Hannan-Quinn	4255,297	

2. ¿Cuál es el género de referencia?
3. Interprete los parámetros β_2 y β_3 , y sus estimaciones
4. ¿Influye la experiencia laboral en el salario?

5. ¿Cuánto puede esperarse que aumente el salario si aumenta su experiencia laboral en un año?
6. Considerando las estimaciones, ¿hay discriminación laboral por género?
7. Según el modelo estimado, ¿qué proporción de las variaciones salariales son explicadas por los factores género y experiencia?
8. Supón ahora que el valor 1 de la variable *GENERE* corresponde a las mujeres y el 0 a los hombres. ¿Cómo se interpretará ahora el parámetro que acompaña a esta variable?
9. Suponga que definimos las siguientes variables ficticias: *HOME* (con valor 1 para los hombres y 0 para las mujeres) y *DONA* (con valor 1 para las mujeres y 0 para los hombres). ¿Qué ocurrirá al estimar los siguientes 3 modelos?

$$SALARI_i = \beta_1 + \beta_2 EXPLAB_i + \beta_3 GENE_i + \beta_4 HOME_i + u_i$$

$$SALARI_i = \beta_1 + \beta_2 EXPLAB_i + \beta_3 GENE_i + \beta_4 DONA_i + u_i$$

$$SALARI_i = \beta_1 + \beta_2 EXPLAB_i + \beta_3 HOME_i + \beta_4 DONA_i + u_i$$

10. Plantee/especifique/escriba ahora un modelo que permita analizar si la experiencia laboral se retribuye de igual manera para hombres y mujeres.
11. Utilice el cuadro siguiente para contrastar si el género influye en la determinación de los salarios

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-200

Variable dependiente: SALARI

	coeficiente	Desv. típica	Estadístico t	valor p
const	23712,7	1414,39	16,77	7,54e-40 ***
EXPLAB	434,197	68,2652	6,360	1,36e-09 ***
Media de la vble. dep.	31510,89	D.T. de la vble. dep.	10917,20	
Suma de cuad. residuos	1,97e+10	D.T. de la regresión	9973,194	
R-cuadrado	0,169655	R-cuadrado corregido	0,165462	
F(1, 198)	40,45517	Valor p (de F)	1,36e-09	
Log-verosimilitud	-2124,314	Criterio de Akaike	4252,628	
Criterio de Schwarz	4259,224	Crit. de Hannan-Quinn	4255,297	

Ejercicio 4 (pescado)

Objetivo

Seguir trabajando con dummies, ahora con preguntas más abiertas.

Datos

Usaremos los datos del fichero `t6_ej_04_pescado.gdt`

Contiene datos referentes a gastos en diferentes categorías de alimento, entre otras las siguientes variables:

- **PESCADO**: Gasto anual en pescado en miles de pesetas per cápita
- **RDISP**: Renta anual disponible en miles de pesetas per cápita
- **ESTMED**: Variable ficticia que toma el valor 1 si el máximo nivel de estudios obtenido por el sustentador principal ha sido el de estudios medios y 0 en caso contrario
- **ESTSUP**: Variable ficticia que toma el valor 1 si el máximo nivel de estudios obtenido por el sustentador principal ha sido el de estudios superiores y 0 en caso contrario

contrario

Con estos datos se estimaron los siguientes modelos:

$$PESCADO_i = \beta_1 + \beta_2 RDISP_i + \beta_3 ESTMED_i + \beta_4 ESTSUP_i + u_i$$

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-40					
Variable dependiente: PESCAD0					
	coeficiente	Desv. típica	Estadístico t	valor p	
const	-3,18603	1,50993	-2,110	0,0419	**
RDISP	0,0196422	0,00119886	16,38	2,86e-18	***
ESTMED	0,918848	1,28809	0,7133	0,4802	
ESTSUP	4,93771	1,73458	2,847	0,0073	***
Media de la vble. dep.	25,85700	D.T. de la vble. dep.	12,51456		
Suma de cuad. residuos	383,5283	D.T. de la regresión	3,263980		
R-cuadrado	0,937208	R-cuadrado corregido	0,931976		
F(3, 36)	179,1083	Valor p (de F)	1,09e-21		
Log-verosimilitud	-101,9682	Criterio de Akaike	211,9364		
Criterio de Schwarz	218,6920	Crit. de Hannan-Quinn	214,3790		

1. ¿Cuál es el nivel de estudios de referencia? Interprete las estimaciones de β_3 y β_4
2. Contraste si el gasto en pescado es igual para las familias con estudios básicos y con estudios superiores
3. Contraste si el gasto en pescado es igual con estudios medios y con estudios superiores

Resultados

Restricción:

$$b[\text{ESTMED}] - b[\text{ESTSUP}] = 0$$

Estadístico de contraste: $F(1, 36) = 6,64545$, con valor $p = 0,0141804$

Estimaciones restringidas:

	coeficiente	Desv. típica	Estadístico t	valor p	
const	-4,33171	1,54923	-2,796	0,0082	***
RDISP	0,0206897	0,00121089	17,09	3,84e-19	***
ESTMED	1,81246	1,33185	1,361	0,1818	
ESTSUP	1,81246	1,33185	1,361	0,1818	

Desviación típica de la regresión = 3,50415

4. Indique las consecuencias de introducir una nueva variable ficticia que tome el valor 1 si el máximo nivel de estudios obtenido por el sustentador principal ha sido el de estudios básicos y 0 en caso contrario

Ejercicio 5 (ex)

Objetivo

Ver un ejemplo de pregunta de examen con dummies.

(Examen febrero 2006) Con una muestra formada por 25 trabajadores se han obtenido las siguientes estimaciones:

donde:

- **SALARI**: Salario bruto anual del trabajador en miles de euros
- **EXPELAB**: Experiencia laboral del trabajador en años

A

$$SALARI = 11,7 + 0,84 \cdot EXPELAB + 1,23 \cdot GÈNERE - 0,12 \cdot ALEMANY + 1,88 \cdot FRANCÈS$$

(0,69) (0,054) (0,58) (0,57) (0,52)

Suma de cuadrados de los residuos, SCR = 20,71; $R^2 = 0,967$

B

$$SALARI = 12,2 + 0,81 \cdot EXPELAB + 1,72 \cdot GÈNERE$$

(0,71) (0,066) (0,70)

Suma de cuadrados de los residuos, SCR = 36,92; $R^2 = 0,942$

- **GÈNERE**: Variable ficticia que toma el valor 1 si el trabajador es un hombre y 0 en caso contrario
- **ALEMANY**: Variable ficticia que toma el valor 1 si el trabajador sabe alemán y 0 en caso contrario
- **FRANCÈS**: Variable ficticia que toma el valor 1 si el trabajador sabe francés y 0 en caso contrario

y los valores entre paréntesis son los errores estándar.

- a) ¿Influye la experiencia laboral en el salario? (Tome como referencia el Modelo A)
- b) ¿Influye el conocimiento de los dos idiomas extranjeros en el salario de un trabajador?
- c) ¿Qué diferencia salarial cabe esperar entre un hombre sin conocimientos de francés y una mujer con conocimientos de francés?
- d) Plantee un modelo de regresión que permita analizar si la diferencia salarial entre hombres y mujeres aumenta con la experiencia laboral e indique cómo realizaría el contraste pertinente. (Tome como referencia el Modelo B)

Ejercicio 6 (estacionalidad)

Objetivo

Ejercicio que muestra que las dummies sirven para incorporar en nuestros modelos una amplia cantidad de fenómenos, como por ejemplo la **estacionalidad**. Además, como los resultados de estimación no se han efectuado con Gretl si no con otro programa, **Eviews**, sirve para ver que somos capaces de leer resultados de cualquier software econométrico.

Con datos trimestrales comprendidos entre el primer trimestre de 1981 y el cuarto trimestre de 2008 se ha obtenido la siguiente estimación:

=====				
Dependent Variable: CONSUM				
Method: Least Squares				
Sample: 1981Q1 2008Q4				
Included observations: 112				
=====				
	Coefficien	Std. Error	t-Statistic	Prob.
=====				
C	4.147752	1.206761	3.437096	0.0008
RENDA	0.111547	0.037210	2.997803	0.0034
T2	-0.502171	0.517375	-0.970613	0.3339
T3	-1.727730	0.518511	-3.332101	0.0012
T4	1.468067	0.517362	2.837600	0.0054
=====				
R-squared	0.318068	Mean dependent var	7.349508	
Adjusted R-squared	0.292575	S.D. dependent var	2.300232	
S.E. of regression	1.934692	Akaike info crit.	4.201390	
Sum squared resid	400.5047	Log likelihood	-230.2778	
F-statistic	12.47676	Durbin-Watson stat	1.918160	
Prob(F-statistic)	0.000000			
=====				

donde CONSUM es el consumo en miles de euros, RENDA es la renta en miles de euros, y T2, T3 y T4 son variables ficticias que toman el valor 1 en el segundo, tercer y cuarto trimestre, respectivamente, y 0 en los respectivos casos contrarios.

1. ¿Difiere el consumo en el primer y segundo trimestre?
2. ¿Es el consumo menor en el segundo trimestre que en el primero?
3. ¿Qué contraste plantearía para analizar si existe estacionalidad o diferencias entre los trimestres?
4. Se introduce una nueva ficticia, T1, que toma el valor 1 en el primer trimestre y 0 en los otros trimestres. ¿Qué ocurrirá?

Ejercicio 7

Objetivo

Ejercicio muy abierto con dummies (3 categorías)

Datos

Usaremos los datos del fichero `t6_ej_03_ocio.gdt`

Contiene información de las siguientes variables:

- **OCI:** Gasto en ocio en euros anuales
- **RENTA:** Renta disponible en euros anuales
- **PRIM:** Variable ficticia que toma el valor 1 si el máximo nivel de estudios conseguidos es el nivel primario y 0 en caso contrario
- **SEC:** Variable ficticia que toma el valor 1 si el máximo nivel de estudios conseguidos es el nivel secundario y 0 en caso contrario
- **UNI:** Variable ficticia que toma el valor 1 si el máximo nivel de estudios conseguidos son estudios universitarios y 0 en caso contrario

1. Plantee un modelo teórico sin interacciones entre variables explicativas que permita explicar el gasto en ocio a partir de la renta disponible y el nivel de estudios. ¿Cuál es el nivel de estudios de referencia?
2. Interprete los parámetros del modelo anterior
3. Estime el modelo planteado

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-1200

Variable dependiente: OCI

	coeficiente	Desv. típica	Estadístico t	valor p	
const	-367,649	117,669	-3,124	0,0018	***
RENDA	0,0727363	0,00553279	13,15	5,71e-37	***
SEC	375,885	68,6544	5,475	5,33e-08	***
UNI	921,918	105,315	8,754	6,90e-18	***
Media de la vble. dep.	1721,191	D.T. de la vble. dep.	1249,409		
Suma de cuad. residuos	1,03e+09	D.T. de la regresión	926,9725		
R-cuadrado	0,450919	R-cuadrado corregido	0,449541		
F(3, 1196)	327,3945	Valor p (de F)	3,7e-155		
Log-verosimilitud	-9899,032	Criterio de Akaike	19806,06		
Criterio de Schwarz	19826,42	Crit. de Hannan-Quinn	19813,73		

4. Considerando el modelo estimado en el apartado 3, ¿existen diferencias significativas en el gasto en ocio entre las personas con estudios secundarios y las personas con estudios primarios?
5. Considerando el modelo estimado en el apartado 3, ¿existen diferencias significativas en el gasto en ocio entre las personas con estudios universitarios y las personas con estudios primarios?
6. Considerando el modelo estimado en el apartado 3, ¿existen diferencias significativas en el gasto en ocio entre las personas con estudios secundarios y las personas con estudios universitarios?

Resultados

En el modelo del apartado 3 no lo podemos resolver directamente. tenemos 3 opciones:

- 1) Plantear el contraste directamente con el estadístico F

Restricción:

$$b[\text{SEC}] - b[\text{UNI}] = 0$$

Estadístico de contraste: $F(1, 1196) = 34,6983$, con valor $p = 4,99799\text{e-}09$

- 2) Reformular el modelo para que, por ejemplo, la categoría de referencia sean los estudios universitarios

Modelo 3: MCO, usando las observaciones 1-1200

Variable dependiente: OCI

	coeficiente	Desv. típica	Estadístico t	valor p	
const	554,269	199,463	2,779	0,0055	***
RENDA	0,0727363	0,00553279	13,15	5,71e-37	***
PRIM	-921,918	105,315	-8,754	6,90e-18	***
SEC	-546,033	92,6969	-5,891	5,00e-09	***
Media de la vble. dep.	1721,191	D.T. de la vble. dep.	1249,409		
Suma de cuad. residuos	1,03e+09	D.T. de la regresión	926,9725		
R-cuadrado	0,450919	R-cuadrado corregido	0,449541		
F(3, 1196)	327,3945	Valor p (de F)	3,7e-155		
Log-verosimilitud	-9899,032	Criterio de Akaike	19806,06		
Criterio de Schwarz	19826,42	Crit. de Hannan-Quinn	19813,73		

- 3) Con MG y MR

7. En el modelo estimado en el apartado 3 ¿Hay diferencias en el gasto en ocio asociadas al nivel educativo?

Pistas

Se puede responder de 2 formas:

- 1) Usando el estadístico F

Conjunto de restricciones

1: b[SEC] = 0

2: b[UNI] = 0

Estadístico de contraste: $F(2, 1196) = 39,1342$, con valor $p = 3,44508e-17$

Estimaciones restringidas:

	coeficiente	Desv. típica	Estadístico t	valor p	
const	-968,782	96,3141	-10,06	6,63e-23	***
RENTA	0,109412	0,00375320	29,15	1,22e-141	***
SEC	0,00000	0,00000	NA	NA	
UNI	0,00000	0,00000	NA	NA	

Desviación típica de la regresión = 956,024

2) Comparando el MG (apartado 3) con el siguiente modelo (MR)

Modelo 2: MCO, usando las observaciones 1-1200

Variable dependiente: OCI

	coeficiente	Desv. típica	Estadístico t	valor p	
const	-968,782	96,3141	-10,06	6,63e-23	***
RENTA	0,109412	0,00375320	29,15	1,22e-141	***
Media de la vble. dep.	1721,191	D.T. de la vble. dep.	1249,409		
Suma de cuad. residuos	1,09e+09	D.T. de la regresión	956,0243		
R-cuadrado	0,414986	R-cuadrado corregido	0,414497		
F(1, 1198)	849,8132	Valor p (de F)	1,2e-141		
Log-verosimilitud	-9937,065	Criterio de Akaike	19878,13		
Criterio de Schwarz	19888,31	Crit. de Hannan-Quinn	19881,97		

8. Especifique ahora un modelo que incluya interacciones entre la renta y el nivel de estudios

9. Interprete los parámetros del modelo anterior

10. Estime el modelo planteado

Resultados de estimación

Modelo 2: MCO, usando las observaciones 1-1200

Variable dependiente: OCI

	coeficiente	Desv. típica	Estadístico t	valor p
-----	-----	-----	-----	-----
const	-42,8484	199,264	-0,2150	0,8298
RENDA	0,0566196	0,00971030	5,831	7,09e-09 ***
SEC	-94,4295	319,670	-0,2954	0,7677
UNI	269,614	381,921	0,7059	0,4804
RENDA_x_SEC	0,0219018	0,0137495	1,593	0,1114
RENDA_x_UNI	0,0256156	0,0134440	1,905	0,0570 *
Media de la vble. dep.	1721,191	D.T. de la vble. dep.	1249,409	
Suma de cuad. residuos	1,02e+09	D.T. de la regresión	926,1398	
R-cuadrado	0,452821	R-cuadrado corregido	0,450530	
F(5, 1194)	197,6204	Valor p (de F)	1,5e-153	
Log-verosimilitud	-9896,949	Criterio de Akaike	19805,90	
Criterio de Schwarz	19836,44	Crit. de Hannan-Quinn	19817,40	

11. Considerando el modelo estimado en el apartado 10, ¿existen diferencias significativas en el gasto en ocio entre las personas con estudios secundarios y las personas con estudios universitarios?

Pistas

Se puede responder de 2 formas:

- 1) Usando el estadístico F

Conjunto de restricciones

- 1: $b[SEC] = 0$
- 2: $b[UNI] = 0$
- 3: $b[REDA_x_SEC] = 0$
- 4: $b[REDA_x_UNI] = 0$

Estadístico de contraste: $F(4, 1194) = 20,6402$, con valor $p = 1,83055e-16$

Estimaciones restringidas:

	coeficiente	Desv. típica	Estadístico t	valor p	
const	-968,782	96,3141	-10,06	6,63e-23	***
REDA	0,109412	0,00375320	29,15	1,22e-141	***
SEC	0,00000	0,00000	NA	NA	
UNI	0,00000	0,00000	NA	NA	
REDA_x_SEC	0,00000	0,00000	NA	NA	
REDA_x_UNI	0,00000	0,00000	NA	NA	

Desviación típica de la regresión = 956,024

- 2) Comparando el MG (apartado 3) con el siguiente modelo (MR)

Modelo 5: MCO, usando las observaciones 1-1200

Variable dependiente: OCI

	coeficiente	Desv. típica	Estadístico t	valor p	
const	-968,782	96,3141	-10,06	6,63e-23	***
REDA	0,109412	0,00375320	29,15	1,22e-141	***
Media de la vble. dep.	1721,191	D.T. de la vble. dep.	1249,409		
Suma de cuad. residuos	1,09e+09	D.T. de la regresión	956,0243		
R-cuadrado	0,414986	R-cuadrado corregido	0,414497		
F(1, 1198)	849,8132	Valor p (de F)	1,2e-141		
Log-verosimilitud	-9937,065	Criterio de Akaike	19878,13		
Criterio de Schwarz	19888,31	Crit. de Hannan-Quinn	19881,97		

12. Considerando el modelo estimado en el apartado 10, construya una tabla de doble entrada que muestre el gasto en ocio imputable a las personas con rentas iguales a 10.000, 20.000 y 30.000 euros y con diferentes niveles de estudios

Ejercicio 8

Objetivo

Mostrar otra pregunta de examen focalizada en las variables ficticias.

(Examen de enero de 2012). Para analizar las retribuciones salariales de determinados profesionales sanitarios, se plantea el siguiente modelo:

$$SAL_i = \beta_1 + \beta_2 EXPLAB_i + \beta_3 GENE_i + \beta_4 NACIO_i + \beta_5 (GENE_i \times NACIO_i) + u_i$$

- SAL : Salario en euros anuales
- $EXPLAB$: Experiencia laboral en años
- $GENE$: Variable ficticia que toma el valor 1 si el profesional es hombre y 0 en caso contrario
- $NACIO$: Variable ficticia que toma el valor 1 si el profesional tiene nacionalidad española y 0 en caso contrario

- a) ¿Qué salario cabe esperar para cada una de las posibles combinaciones de género y nacionalidad?
 - b) Interprete el significado de β_5
-

Ejercicio 9 (absentismo)

Objetivo

Un caso completamente abierto. Se trata de ver si sois capaces de utilizar lo visto en clase para sacar información sobre un fenómeno a partir de un conjunto de datos

Datos

- Usaremos los datos del fichero `t6_ej_09_absentismo.gdt`.
- La descripción de las variables puedes encontrarla en el mismo fichero de datos

¿Qué variables pueden explicar el comportamiento del absentismo laboral en esa empresa?

Tema 7: Incumplimiento de las hipótesis básicas

Repaso de teoría (preguntas rápidas)

1. ¿Qué significa que se incumpla alguna h.e.b? ¿Qué consecuencias tiene?
2. ¿Qué es multicolinealidad? ¿Qué consecuencias tiene? ¿Cómo se detecta? ¿Soluciones?
3. ¿Qué significa la h.e.b de Normalidad? ¿Qué consecuencias tiene? ¿Cómo se detecta? ¿Soluciones?
4. ¿Qué significa la existencia de heterocedasticidad? ¿En que fenómenos suele ocurrir? ¿Qué consecuencias tiene? ¿Cómo se detecta su presencia? ¿Soluciones?
5. ¿Qué significa la presencia de autocorrelación? ¿Cuando suele ocurrir? ¿Qué consecuencias tiene? ¿Cómo se detecta su presencia? ¿Soluciones?

Ejercicio 1 (bebes)

Objetivo

Recordar las principales ideas y contrastes del tema 7: normalidad, colinealidad, heterocedasticidad, ...

Datos

Usaremos los datos del fichero `t6_ej_01_bebes.gdt`.

Son datos que ya hemos usado, concretamente en el ejercicio 1 del tema 6. Son datos relacionados con nacimientos de bebés.

Estime el siguiente modelo:

$$peso = \beta_1 + \beta_2 semanas + \beta_3 edad + \beta_4 bebuto + u$$

Resultados de estimación

Modelo 1: MCO, usando las observaciones 1-1000

Variable dependiente: peso

	coeficiente	Desv. típica	Estadístico t	valor p
const	-2358,58	313,567	-7,522	1,20e-13 ***
semanas	140,507	7,47050	18,81	9,10e-68 ***
edad	2,98253	2,88725	1,033	0,3019
bebito	121,343	27,6315	4,391	1,25e-05 ***
Media de la vble. dep.	3273,443	D.T. de la vble. dep.	507,1946	
Suma de cuad. residuos	1,89e+08	D.T. de la regresión	435,1339	
R-cuadrado	0,266179	R-cuadrado corregido	0,263968	
F(3, 996)	120,4261	Valor p (de F)	1,51e-66	
Log-verosimilitud	-7492,588	Criterio de Akaike	14993,18	
Criterio de Schwarz	15012,81	Crit. de Hannan-Quinn	15000,64	

1. ¿cree que el modelo tiene problemas serios de colinealidad entre los regresores? Explique

Resultados

Coeficientes de correlación, usando las observaciones 1 - 1000
Valor crítico al 5% (a dos colas) = 0,0620 para n = 1000

peso	edad	semanas	bebito	
1,0000	0,0009	0,5011	0,0745	peso
	1,0000	-0,0563	0,0139	edad
		1,0000	-0,0888	semanas
			1,0000	bebito

Factores de inflación de varianza (VIF)

Mínimo valor posible = 1.0

Valores mayores que 10.0 pueden indicar un problema de colinealidad

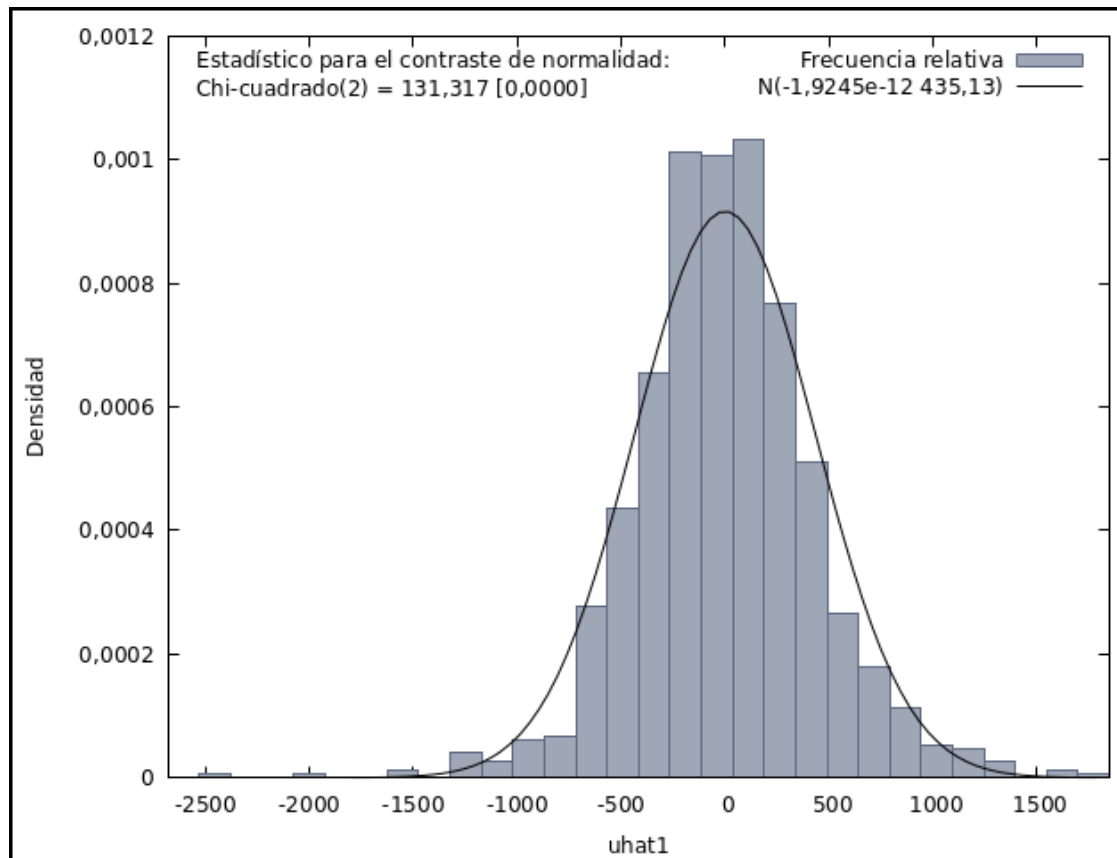
semanas	1,011
edad	1,003
bebito	1,008

$VIF(j) = 1/(1 - R(j)^2)$, donde $R(j)$ es el coeficiente de correlación múltiple entre la variable j y las demás variables independientes

2. Suponga que sí que hubiese un problema importante de colinealidad ¿cuales serían las consecuencias?

3. Contraste la normalidad de las perturbaciones. Explique el resultado y las consecuencias del contraste

Resultados



- d) Efectúe el test RESET de Ramsey y explique los resultados (!!!!)
- e) Efectúe el test de White para contrastar la homocedasticidad de las perturbaciones. Explique el contraste y las consecuencias del resultado obtenido

Resultados

Contraste de heterocedasticidad de White
MCO, usando las observaciones 1-1000
Variable dependiente: uhat^2

	coeficiente	Desv. típica	Estadístico t	valor p	
const	8,86759e+06	2,44735e+06	3,623	0,0003	***
semanas	-368379	98949,2	-3,723	0,0002	***
edad	-76875,5	54741,6	-1,404	0,1605	
bebido	-22777,7	565431	-0,04028	0,9679	
sq_semanas	3772,97	1143,36	3,300	0,0010	***
X2_X3	1824,75	1243,38	1,468	0,1425	
X2_X4	2614,86	13493,3	0,1938	0,8464	
sq_edad	103,719	380,010	0,2729	0,7850	
X3_X4	-1820,98	5169,40	-0,3523	0,7247	

R-cuadrado = 0,034154

Estadístico de contraste: $TR^2 = 34,154312$,
con valor p = $P(\text{Chi-cuadrado}(8) > 34,154312) = 0,000038$

Ejercicio 2 (ROA)

Objetivo

Otro ejercicio para recordar las principales ideas y contrastes del tema 7: normalidad, heterocedasticidad, ...

Datos

Usaremos los datos del fichero `t7_ej_02_ROA.gdt`.

Se quiere conocer los determinantes de la rentabilidad sobre activos (ROA) de una muestra de 5.375 entidades financieras europeas. Para ello se ha estimado el siguiente modelo:

$$ROA = \beta_1 + \beta_2 E_TA + \beta_3 LACTIVO + \beta_4 LACTIVO^2 + u$$

donde E_TA es una medida de la solvencia de la entidad (Recursos propios / Activo total), $LACTIVO$ mide el tamaño de la entidad (logaritmo del activo total), $LACTIVO^2$ es el cuadrado de esta variable.

Resultados de estimación

```

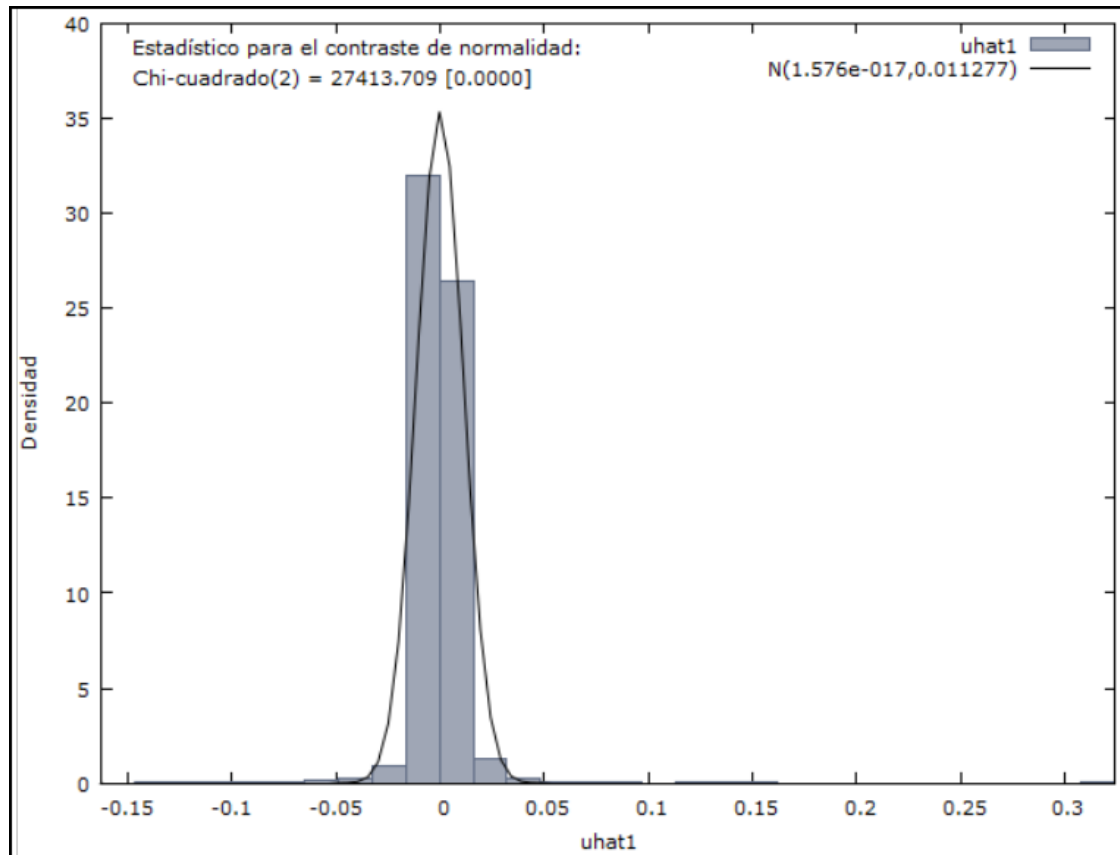
Modelo 1: MCO, usando las observaciones 1-5375
Variable dependiente: ROA

      Coeficiente   Desv. Típica  Estadístico t   Valor p
-----
const   -0.0305802    0.00668821     -4.572         4.93e-06 ***
E_TA     0.0942888      0.00246119     38.31         4.15e-284 ***
LACTIVO   0.00389075    0.000932189      4.174         3.04e-05 ***
LACTIVO2 -0.000118662    3.22575e-05     -3.679         0.0002 ***

Media de la vble. dep.  0.008128   D.T. de la vble. dep.  0.012722
Suma de cuad. residuos  0.682994   D.T. de la regresión   0.011277
R-cuadrado              0.214771   R-cuadrado corregido   0.214333
F(3, 5371)              489.6814   Valor p (de F)         2.8e-281
Log-verosimilitud       16482.19   Criterio de Akaike     -32956.37
Criterio de Schwarz     -32930.02   Crit. de Hannan-Quinn  -32947.17
  
```

1. Contraste la normalidad de las perturbaciones

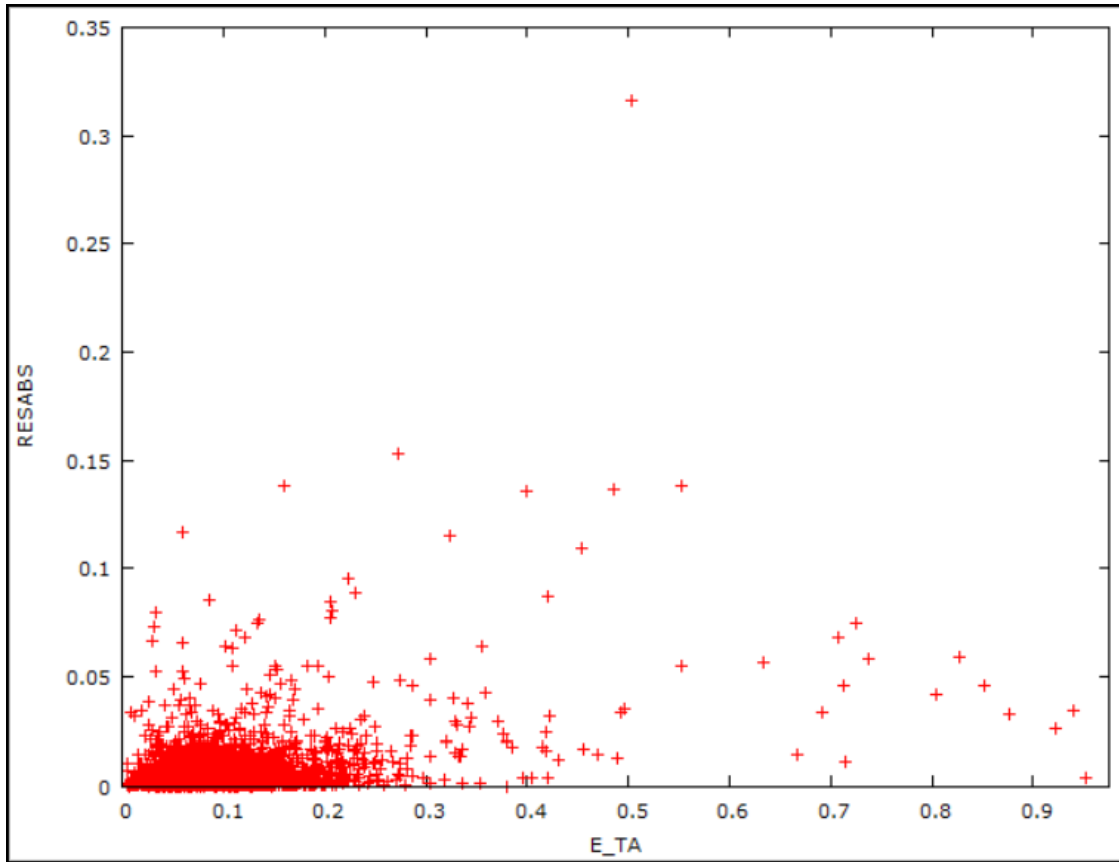
Resultados



2. ¿Presenta heterocedasticidad el modelo estimado?

Resultados

- La variable $RESABS = |\hat{u}|$; es decir es igual a los residuos del modelo estimado en valor absoluto.



```

Contraste de heterocedasticidad de White
MCO, usando las observaciones 1-5375
Variable dependiente: uhat^2
Omitidas debido a colinealidad exacta: sq_LACTIVO

-----
                Coeficiente    Desv. Típica    Estadístico t    Valor p
-----
const           0.0243410      0.0227751         1.069           0.2852
E_TA            -0.0317556      0.0128216        -2.477           0.0133  **
LACTIVO         -0.00637895      0.00627081        -1.017           0.3091
LACTIVO2        0.000604714      0.000641312         0.9429          0.3458
sq_E_TA         0.00297430      0.00109224         2.723           0.0065  ***
X2_X3           0.00570584      0.00181359         3.146           0.0017  ***
X2_X4           -0.000221037      6.34994e-05        -3.481           0.0005  ***
X3_X4           -2.50892e-05      2.88390e-05        -0.8700          0.3844
sq_LACTIVO2     3.90293e-07      4.80695e-07         0.8119          0.4169

ATENCIÓN: ¡matriz de datos casi singular!

R-cuadrado = 0.056136

Estadístico de contraste: TR^2 = 301.732384,
con valor p = P(Chi-cuadrado(8) > 301.732384) = 0.000000

```

3. Vuelva a estimar el modelo pero con “errores típicos robustos a la heterocedasticidad”

Resultados

```

Modelo 2: MCO, usando las observaciones 1-5375
Variable dependiente: ROA
Desviaciones típicas robustas ante heterocedasticidad, variante HC1

      Coeficiente   Desv. Típica  Estadístico t   Valor p
-----
const      -0.0305802    0.00823233      -3.715         0.0002    ***
E_TA       0.0942888      0.0104354       9.035         2.24e-019 ***
LACTIVO    0.00389075     0.00108477       3.587         0.0003    ***
LACTIVO2   -0.000118662    3.63469e-05     -3.265         0.0011    ***

Media de la vble. dep.  0.008128   D.T. de la vble. dep.  0.012722
Suma de cuad. residuos  0.682994   D.T. de la regresión   0.011277
R-cuadrado             0.214771   R-cuadrado corregido   0.214333
F(3, 5371)             28.15083   Valor p (de F)         4.69e-18
Log-verosimilitud      16482.19   Criterio de Akaike     -32956.37
Criterio de Schwarz    -32930.02   Crit. de Hannan-Quinn  -32947.17
  
```

4. Señale las diferencias que hay en las 2 estimaciones del modelo

Resultados (modelo 1)

```

Modelo 1: MCO, usando las observaciones 1-5375
Variable dependiente: ROA

      Coeficiente   Desv. Típica  Estadístico t   Valor p
-----
const      -0.0305802    0.00668821      -4.572         4.93e-06    ***
E_TA       0.0942888    0.00246119     38.31         4.15e-284 ***
LACTIVO    0.00389075     0.000932189      4.174         3.04e-05    ***
LACTIVO2   -0.000118662    3.22575e-05     -3.679         0.0002     ***

Media de la vble. dep.  0.008128   D.T. de la vble. dep.  0.012722
Suma de cuad. residuos  0.682994   D.T. de la regresión   0.011277
R-cuadrado             0.214771   R-cuadrado corregido   0.214333
F(3, 5371)             489.6814   Valor p (de F)         2.8e-281
Log-verosimilitud      16482.19   Criterio de Akaike     -32956.37
Criterio de Schwarz    -32930.02   Crit. de Hannan-Quinn  -32947.17
  
```
