

## **PUBLICACIONES DE 4º CURSO**

**Curso: 4º**

**Grado: Economía**

**Asignatura: ECONOMETRÍA III**

### **SOLUCIÓN EJERCICIOS EMPÍRICOS PARTE 2**

**Profesores: Antonio Aznar y Mª Isabel Ayuda**

**Departamento de ANÁLISIS ECONÓMICO**

**Curso Académico 2015/16**

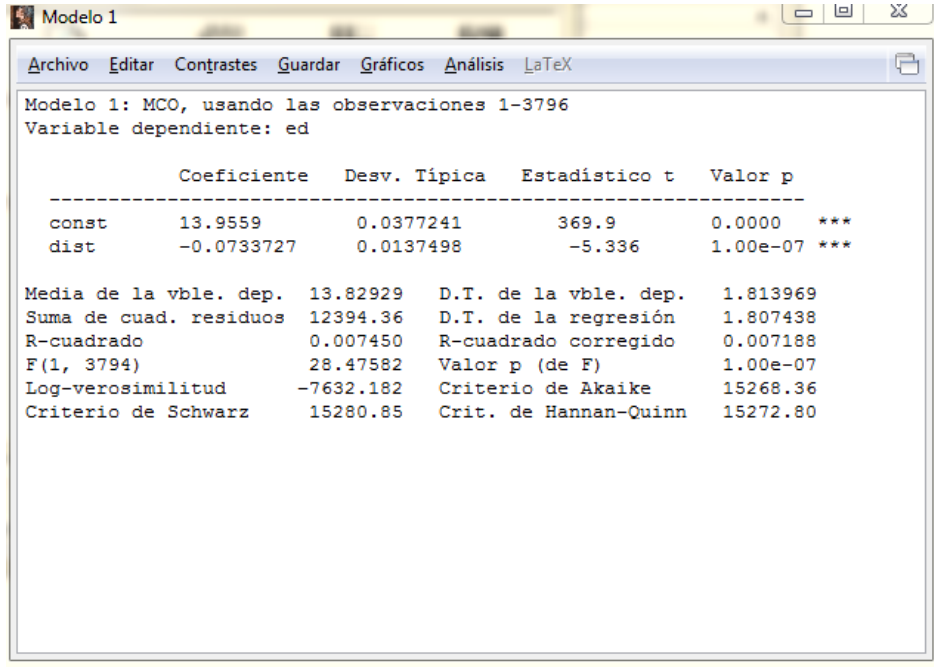


**Facultad de  
Economía y Empresa  
Universidad Zaragoza**

## SOLUCIÓN EJERCICIOS EMPÍRICOS PARTE 2

### Ejercicio empírico 2.1

#### 1. MODELO-MÍNIMOS CUADRADOS ORDINARIOS.



Modelo 1: MCO, usando las observaciones 1-3796  
Variable dependiente: ed

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	13.9559	0.0377241	369.9	0.0000 ***
dist	-0.0733727	0.0137498	-5.336	1.00e-07 ***

Media de la vble. dep.	13.82929	D.T. de la vble. dep.	1.813969
Suma de cuad. residuos	12394.36	D.T. de la regresión	1.807438
R-cuadrado	0.007450	R-cuadrado corregido	0.007188
F(1, 3794)	28.47582	Valor p (de F)	1.00e-07
Log-verosimilitud	-7632.182	Criterio de Akaike	15268.36
Criterio de Schwarz	15280.85	Crit. de Hannan-Quinn	15272.80

$$\widehat{ed} = 13.96 - 0.073 \times dist.$$

El modelo predice que por cada 10 millas que incremente la distancia de la universidad del estudiante con respecto al instituto en el que estudio, los años medios de estudios disminuyen en 0.073 años

2. El  $R^2 = 0.007$ , por lo que la distancia explica una proporción muy pequeña de la varianza de los años de educación.

3. ESR = 1807 años.

4. Como el  $p$ -valor de la variable *dist* es menor que 0.01 se rechaza la  $H_0$  para cualquiera de los dos niveles de significación por lo que la pendiente es significativa.

5. El procedimiento a seguir es dentro del modelo estimado: **ANÁLISIS-INTERVALOS DE CONFIANZA PARA LOS COEFICIENTES**

gretl: intervalos de confianza para los coeficientes

t(3794, 0.025) = 1.961

VARIABLE	COEFICIENTE	INTERVALO DE CONFIANZA 95%	
const	13.9559	13.8819	14.0298
dist	-0.0733727	-0.100330	-0.0464150

Se podrían calcular como se vio en teoría:  $\{\hat{\beta}_1 \pm 1.96 \cdot ES(\hat{\beta}_1)\}$

6. La predicción de los años de educación completados por Bob si su universidad está a 20 millas será:  $13.96 - 0.073 \times 2 = 13.81$ .

La predicción de los años de educación completados por Bob si su universidad está a 10 millas será:  $13.96 - 0.073 \times 1 = 13.89$ .

7. El **contraste de White** se realiza dentro del modelo estimado: **CONTRASTES-HETEROCEDASTICIDAD-CONTRASTE DE WHITE**

contraste LM(heterocedasticidad)

Contraste de heterocedasticidad de White  
MCO, usando las observaciones 1-3796  
Variable dependiente: uhat^2

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	3.34953	0.0688810	48.63	0.0000 ***
dist	-0.0601525	0.0444027	-1.355	0.1756
sq_dist	0.00256867	0.00404364	0.6352	0.5253

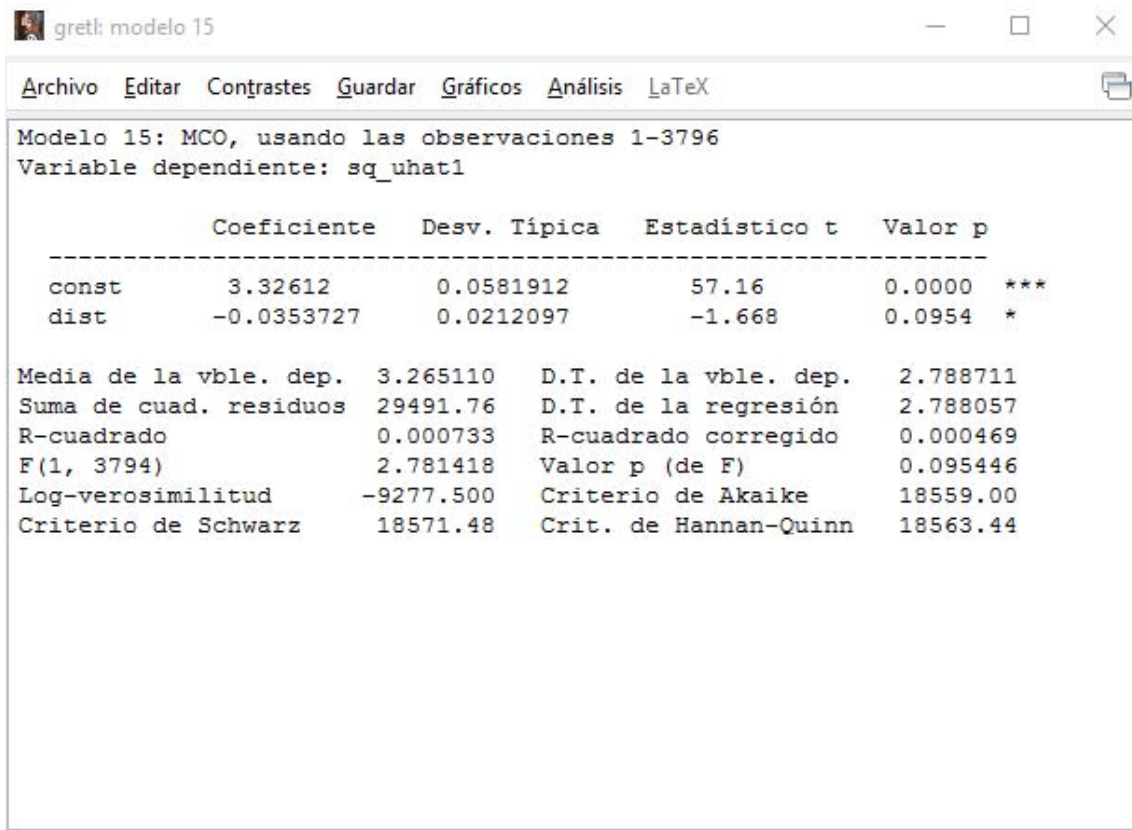
R-cuadrado = 0.000839

Estadístico de contraste:  $TR^2 = 3.184351$ ,  
con valor p =  $P(\text{Chi-cuadrado}(2) > 3.184351) = 0.203482$

Según el contraste de White se mantiene la hipótesis nula de homocedasticidad.

El **contraste de Breusch-Pagan** que hace Gretl es asintóticamente igual que el propuesto en clase y el procedimiento a seguir es: **CONTRASTES-HETEROCEDASTICIDAD-BREUSCH-PAGAN**

Se guardan los residuos del modelo y se hace la regresión de los residuos al cuadrado como endógena y como exógenas las variables que creamos que puedan causar la posible heterocedasticidad, en este caso, dist:



	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	3.32612	0.0581912	57.16	0.0000 ***
dist	-0.0353727	0.0212097	-1.668	0.0954 *

Media de la vble. dep.	3.265110	D.T. de la vble. dep.	2.788711
Suma de cuad. residuos	29491.76	D.T. de la regresión	2.788057
R-cuadrado	0.000733	R-cuadrado corregido	0.000469
F(1, 3794)	2.781418	Valor p (de F)	0.095446
Log-verosimilitud	-9277.500	Criterio de Akaike	18559.00
Criterio de Schwarz	18571.48	Crit. de Hannan-Quinn	18563.44

El estadístico  $BP = TR^2 = 3796 * 0.000733 = 2.78 < \chi^2(1) = 3.84$  por lo que no rechazamos la hipótesis de homocedasticidad.

El **contraste de Glejser** no lo tiene implementado Gretl por lo que se hará manualmente. Para ello se guardan los residuos de la regresión y se calcula el valor absoluto que guardaremos como auhat1:

**auhat1= abs(uhat1)**

Posteriormente haremos la regresión auxiliar de Glejser:

gretl: modelo 11

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

Modelo 11: MCO, usando las observaciones 1-3796  
Variable dependiente: auhat1

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	1.63972	0.0163878	100.1	0.0000	***
dist	-0.00709259	0.00597308	-1.187	0.2351	

Media de la vble. dep.	1.627486	D.T. de la vble. dep.	0.785215
Suma de cuad. residuos	2338.987	D.T. de la regresión	0.785173
R-cuadrado	0.000371	R-cuadrado corregido	0.000108
F(1, 3794)	1.409980	Valor p (de F)	0.235134
Log-verosimilitud	-4467.222	Criterio de Akaike	8938.445
Criterio de Schwarz	8950.928	Crit. de Hannan-Quinn	8942.881

Y calculamos el  $LM = T R^2 = 3796 \cdot 0.00037 = 1.4045 < \chi^2(1) = 3.84$  por lo que no rechazamos la hipótesis de homocedasticidad.

### Contraste de Goldfeld-Quandt

Como necesitamos reordenar los datos, vamos a añadir una variable índice para poder volver, cuando acabemos el contraste, a los datos originales: **AÑADIR-VARIABLE INDICE**

En primer lugar, ordenamos las variables en función de la variables que creemos que puede causar la posible heterocedasticidad: **DATOS-ORDENAR LOS DATOS (ascendente)** y le indicamos la variable *dist*. Posteriormente elegimos  $c = 796$  que serán las observaciones centrales que vamos a eliminar y trabajaremos con una muestra 1 con las primeras 1500 observaciones y otra muestra 2 con las últimas 1500 observaciones.

Volvemos a estimar el modelo con la primera muestra 1:

Recordar poner la muestra adecuada: **MUESTRA-ESTABLECER RANGO**

gretl: modelo 12

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

Modelo 12: MCO, usando las observaciones 1-1500  
Variable dependiente: ed

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	13.8182	0.0647027	213.6	0.0000	***
dist	-0.0734927	0.0274660	-2.676	0.0075	***

Media de la vble. dep.	13.69733	D.T. de la vble. dep.	1.798485
Suma de cuad. residuos	4825.526	D.T. de la regresión	1.794801
R-cuadrado	0.004757	R-cuadrado corregido	0.004092
F(1, 1498)	7.159749	Valor p (de F)	0.007537
Log-verosimilitud	-3004.749	Criterio de Akaike	6013.497
Criterio de Schwarz	6024.124	Crit. de Hannan-Quinn	6017.456

La SR1 = 4825.5

Volvemos a estimar el modelo con la segunda muestra que ira desde la observación 2297 hasta la 3796. **MUESTRA-ESTABLECER RANGO**

gretl: modelo 13

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

Modelo 13: MCO, usando las observaciones 2297-3796 (n = 1500)  
Variable dependiente: ed

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	14.1588	0.0561476	252.2	0.0000	***
dist	-0.0600017	0.0188089	-3.190	0.0015	***

Media de la vble. dep.	14.06200	D.T. de la vble. dep.	1.834957
Suma de cuad. residuos	5013.177	D.T. de la regresión	1.829366
R-cuadrado	0.006748	R-cuadrado corregido	0.006085
F(1, 1498)	10.17657	Valor p (de F)	0.001452
Log-verosimilitud	-3033.361	Criterio de Akaike	6070.723
Criterio de Schwarz	6081.349	Crit. de Hannan-Quinn	6074.682

SR2 = 5013.17

Por lo que el contraste  $GQ = SR2/SR1 = 5013.17 / 4825.5 = 1.04$

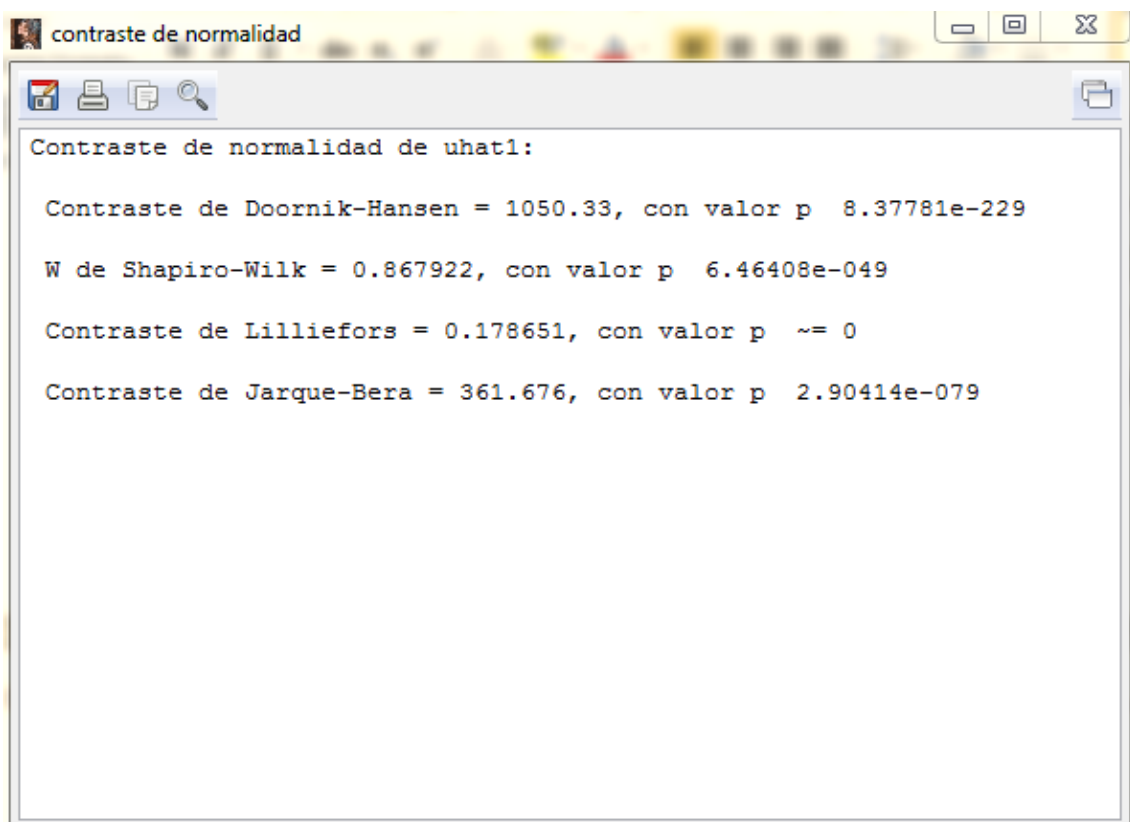
$1.04 < F_{0.05}(1498, 1498) = 1.09$  Por lo que mantenemos la hipótesis de homocedasticidad a un nivel de significación del 5%

RECORDAR VOLVER A LA MUESTRA COMPLETA : **MUESTRA-RECUPERAR EL RANGO COMPLETO**

RECORDAR ORDENAR LA MUESTRA SEGÚN INDEX

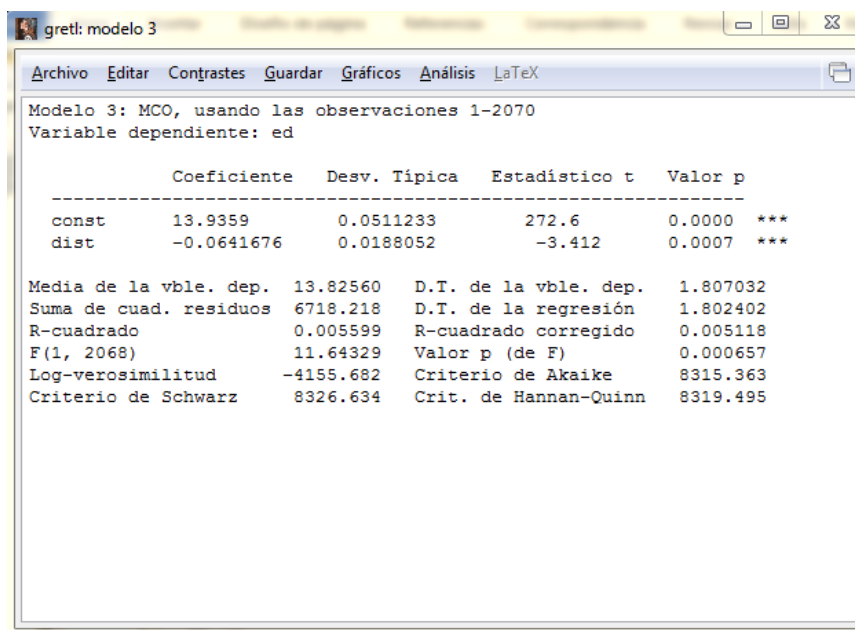
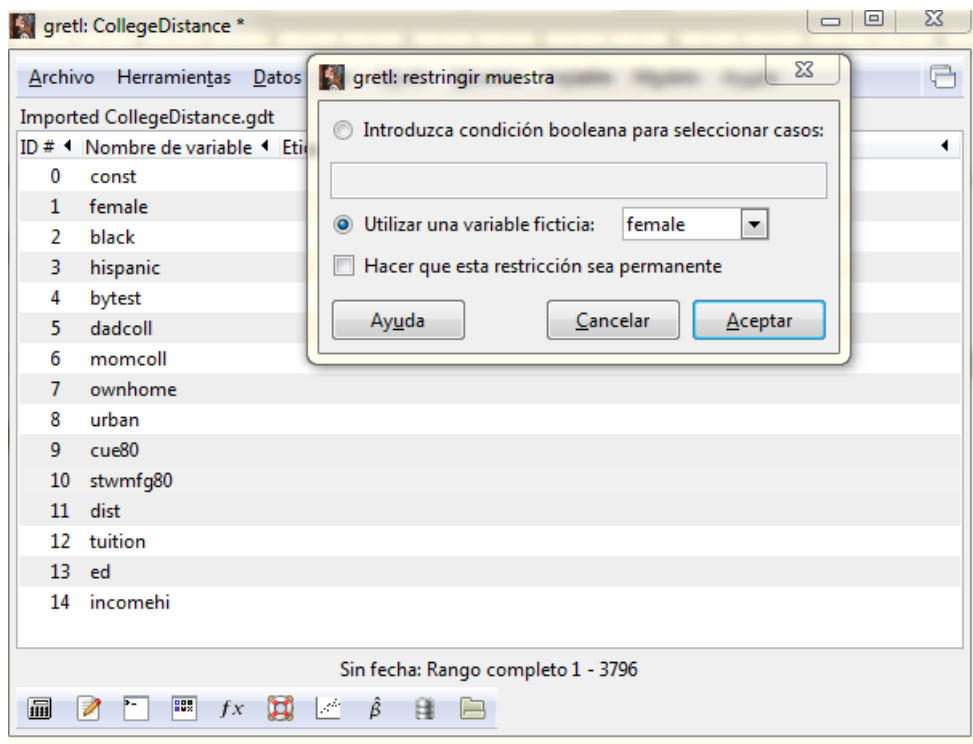
### Contraste de Jarque-Bera

Para ello guardar los residuos de la regresión original **GUARDAR – RESIDUOS** por ejemplo con el nombre *uhat1* e ir a **VARIABLE-CONTRASTE DE NORMALIDAD**



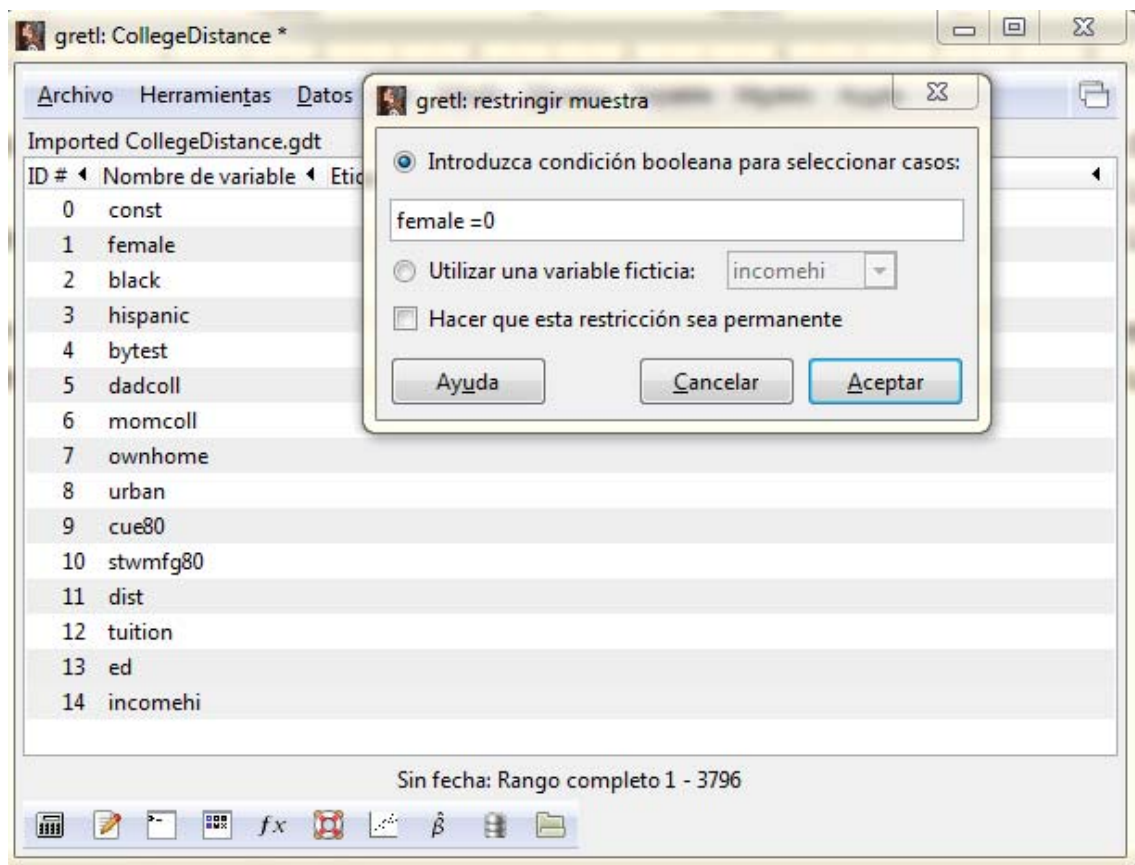
Según el contraste de Jarque-Bera se rechaza la hipótesis de Normalidad. El problema puede estar en la omisión de variables relevantes.

**8.** Para estimar la regresión sólo para las mujeres hay que ir a **MUESTRA-RESTRINGIR, APARTIR DE CRITERIO** y utilizar la variable ficticia *female*

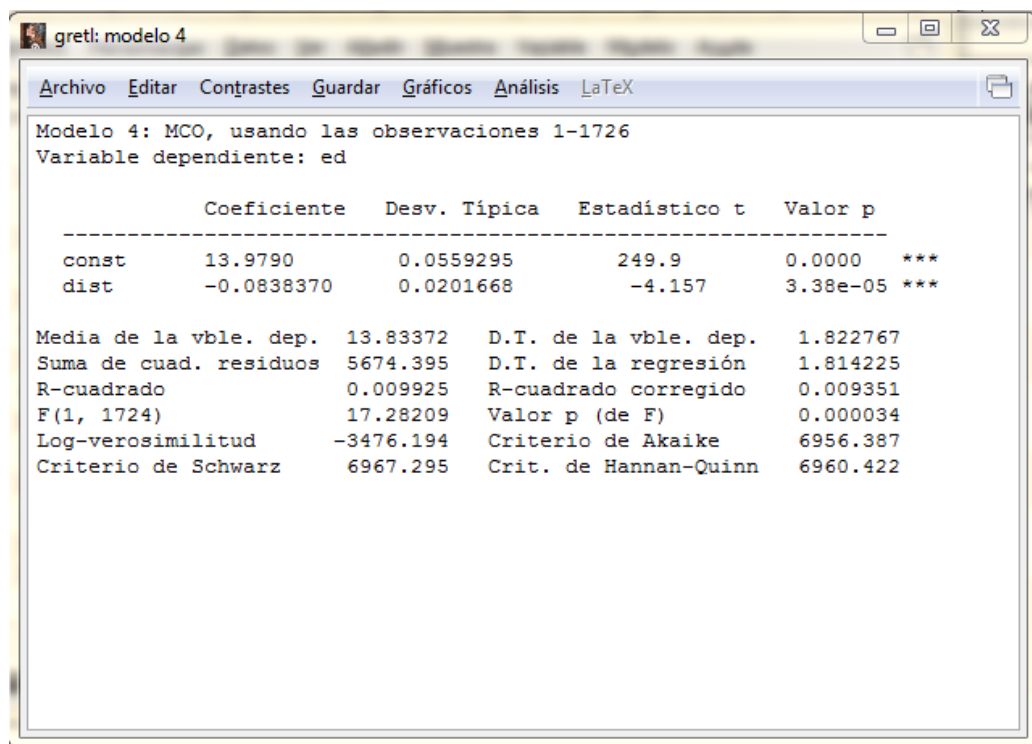


9.- Para hacerlo para los hombres hay que ir a MUESTRA- RECUPERAR RANGO COMPLETO Y PONER LA RESTRICCIÓN DE *female* = 0





El modelo para hombres sería:



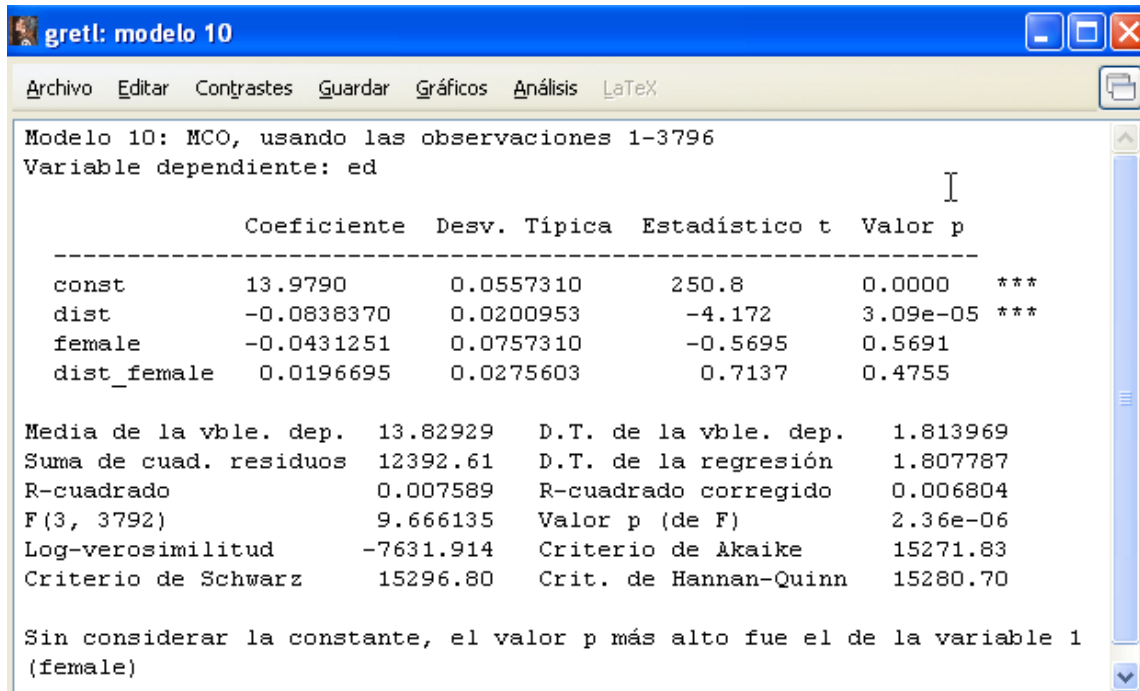
**10.-** El efecto de la distancia sobre los años contemplados de educación en las mujeres es -0,064 y en los hombres es -0,084. Por lo que sí que es distinto.

También se podría meter la ficticia multiplicativa:  $\text{dist\_female} = \text{dist} * \text{female}$

Para ello hay que añadir una nueva variable: **AÑADIR-DEFINIR NUEVA VARIABLE**

$\text{Dist\_female} = \text{dist} * \text{female}$

Y estimar el modelo añadiendo la ficticia *female* aditivamente y multiplicativamente



Modelo 10: MCO, usando las observaciones 1-3796  
Variable dependiente: ed

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	13.9790	0.0557310	250.8	0.0000	***
dist	-0.0838370	0.0200953	-4.172	3.09e-05	***
female	-0.0431251	0.0757310	-0.5695	0.5691	
dist_female	0.0196695	0.0275603	0.7137	0.4755	

Media de la vble. dep.	13.82929	D.T. de la vble. dep.	1.813969
Suma de cuad. residuos	12392.61	D.T. de la regresión	1.807787
R-cuadrado	0.007589	R-cuadrado corregido	0.006804
F(3, 3792)	9.666135	Valor p (de F)	2.36e-06
Log-verosimilitud	-7631.914	Criterio de Akaike	15271.83
Criterio de Schwarz	15296.80	Crit. de Hannan-Quinn	15280.70

Sin considerar la constante, el valor p más alto fue el de la variable 1 (female)

Se comprueba que aunque la diferencia del efecto de la distancia en los años de educación, entre hombres y mujeres, es 0.02, no es significativa.

## Resultado ejercicio empírico 2.2

1. Estimamos el modelo con todas las variables indicadas.

Modelo 5

Modelo 5: MCO, usando las observaciones 1-3796  
Variable dependiente: ed

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	8.86137	0.249705	35.49	2.25e-238	***
dist	-0.0308039	0.0123377	-2.497	0.0126	**
bytest	0.0924474	0.00316741	29.19	4.37e-169	***
female	0.143378	0.0504535	2.842	0.0045	***
black	0.353808	0.0712345	4.967	7.11e-07	***
hispanic	0.402351	0.0742642	5.418	6.41e-08	***
incomehi	0.366595	0.0606792	6.042	1.67e-09	***
ownhome	0.145642	0.0666409	2.185	0.0289	**
dadcoll	0.569915	0.0737182	7.731	1.36e-014	***
cue80	0.0244180	0.00960948	2.541	0.0111	**
stwmfg80	-0.0502044	0.0198013	-2.535	0.0113	**
momcoll	0.379184	0.0815498	4.650	3.44e-06	***
Media de la vble. dep.	13.82929	D.T. de la vble. dep.	1.813969		
Suma de cuad. residuos	8954.270	D.T. de la regresión	1.538295		
R-cuadrado	0.282935	R-cuadrado corregido	0.280850		
F(11, 3784)	135.7331	Valor p (de F)	1.9e-263		
Log-verosimilitud	-7015.121	Criterio de Akaike	14054.24		
Criterio de Schwarz	14129.14	Crit. de Hannan-Quinn	14080.86		

Ahora el efecto de la Dis es -0,031 y en el modelo sólo con la variable *dist* era -0.073

2. Si que presenta sesgo de estimación porque se habían omitido variables relevantes.
3. El  $R^2$  ha pasado de 0,007 en el modelo con una variable a ser 0,28 en el modelo con las variables de control.
4. Los estudiantes cuyos padres fueron al mismo instituto acaban 0,570 años más de educación, en media, que los estudiantes cuyos padres no fueron al mismo instituto.

Las dos variables recogen costes de oportunidad de estudiar. Cuando la variable *cue80* (tasa de paro) aumenta, es más difícil encontrar trabajo, lo cual indica menos costes de oportunidad de acudir a clase, por lo que los años de asistencia a clase aumentan. Por ello el signo del coeficiente es positivo. Cuando la variable *stwmfg80* (salario por hora) incrementa, los costes de oportunidad son mayores y por lo tanto los años de estudio disminuyen. Por lo que el signo negativo del coeficiente de esta variable es consistente. Las dos variables son significativa para un nivel de significación del 5%.

5. Lo años de educación predichos para Bob son:  $8.827 - 0.0315 \times 2 + 0.093 \times 58 + 0.145 \times 0 + 0.367 \times 1 + 0.398 \times 0 + 0.395 \times 1 + 0.152 \times 1 + 0.696 \times 0 + 0.023 \times 7.5 - 0.051 \times 9.75 = 14.79$

6. Los años de educación predichos para Jim son:  $2 \times 0.0315 = 0.0630$  menos que los de Bob. Es decir,  $14.75 - 0.063 = 14.69$ .

7. Hay que contrastar la restricción lineal  $b_2 = -0.075$  y para un nivel de significación del 5% como p valor = 0.00034 se rechaza la hipótesis nula de que los años universitarios completados por una persona se incrementarían en aproximadamente 0.15 años si la distancia a la universidad más cercana se redujera en 20 millas.

Para ello se contrasta la siguiente restricción:

Restricción:				
b[dist] = -0.075				
Estadístico de contraste: F(1, 3784) = 12.8321, con valor p = 0.00034502				
Estimaciones restringidas:				
	Coefficiente	Desv. Típica	Estadístico t	Valor p
const	8.98802	0.247576	36.30	6.90e-248 ***
dist	-0.0750000	0.00000	NA	NA
bytest	0.0916481	0.00316447	28.96	9.26e-167 ***
female	0.140020	0.0505236	2.771	0.0056 ***
black	0.320440	0.0707331	4.530	6.07e-06 ***
hispanic	0.388682	0.0742820	5.233	1.76e-07 ***
incomehi	0.356924	0.0607138	5.879	4.49e-09 ***
ownhome	0.158506	0.0666480	2.378	0.0174 **
dadcoll	0.553666	0.0736934	7.513	7.16e-014 ***
cue80	0.0331805	0.00930744	3.565	0.0004 ***
stwmfg80	-0.0575694	0.0197250	-2.919	0.0035 ***
momcoll	0.375442	0.0816705	4.597	4.43e-06 ***
Desviación típica de la regresión = 1.5407				

## 8. Contraste de White

Contraste de heterocedasticidad de White -

Hipótesis nula: No hay heterocedasticidad

Estadístico de contraste: LM = 285.162

con valor p =  $P(\text{Chi-cuadrado}(69) > 285.162) = 4.47892e-028$

Contraste de Breusch-Pagan

gretl: contraste LM(heterocedasticidad)

Contraste de heterocedasticidad de Breusch-Pagan  
MCO, usando las observaciones 1-3796  
Variable dependiente: uhat^2 escalado

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	-0.0602083	0.199985	-0.3011	0.7634	
dist	-0.0143181	0.00988110	-1.449	0.1474	
bytest	0.0221477	0.00253672	8.731	3.73e-018	***
female	0.0686375	0.0404074	1.699	0.0895	*
black	-0.0691234	0.0570505	-1.212	0.2257	
hispanic	0.0375360	0.0594770	0.6311	0.5280	
incomehi	0.00525923	0.0485970	0.1082	0.9138	
ownhome	0.0643858	0.0533716	1.206	0.2278	
dadcoll	-0.113179	0.0590397	-1.917	0.0553	*
cue80	0.00732376	0.00769607	0.9516	0.3413	
stwmfg80	-0.0160549	0.0158585	-1.012	0.3114	
momcoll	-0.0605374	0.0653119	-0.9269	0.3540	

Suma de cuadrados explicada = 156.791

Estadístico de contraste: LM = 78.395663,  
con valor p = P(Chi-cuadrado(11) > 78.395663) = 0.000000

Con ambos contrastes se rechaza la hipótesis de homocedasticidad.

Contraste de Jarque –Bera.

gretl: contraste de normalidad

Contraste de normalidad de uhat5:

Contraste de Doornik-Hansen = 171.722, con valor p 5.14134e-038

W de Shapiro-Wilk = 0.985548, con valor p 2.84111e-019

Contraste de Lilliefors = 0.0605603, con valor p ~ = 0

Contraste de Jarque-Bera = 98.4994, con valor p 4.08432e-022

Por lo que también se rechaza la hipótesis de normalidad.

Estos problemas podrían ser debidos a una forma funcional no correcta u omisión de variables relevantes.

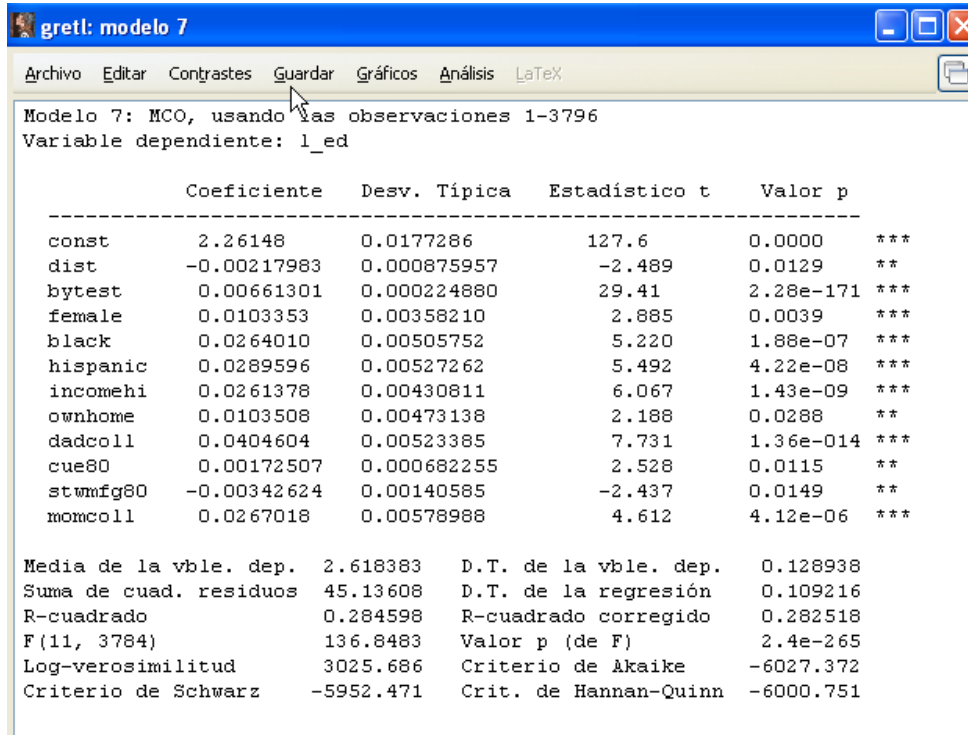
9. Habría que mejorar el modelo pero si no lo conseguimos para que los contrastes sean válidos habría que utilizar la estimación robusta a heterocedasticidad, siempre que no hubiera problemas de no normalidad.

### Resultado ejercicio empírico 2.3

1.- En primer lugar hay que añadir el  $\ln(ed)$  y después estimar por MCO. Para ello marcamos la variable *ed* y:

#### AÑADIR-LOGARITMOS DE LA VARIABLE SELECCIONADA

Estimamos el modelo:



	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	2.26148	0.0177286	127.6	0.0000	***
dist	-0.00217983	0.000875957	-2.489	0.0129	**
bytest	0.00661301	0.000224880	29.41	2.28e-171	***
female	0.0103353	0.00358210	2.885	0.0039	***
black	0.0264010	0.00505752	5.220	1.88e-07	***
hispanic	0.0289596	0.00527262	5.492	4.22e-08	***
incomehi	0.0261378	0.00430811	6.067	1.43e-09	***
ownhome	0.0103508	0.00473138	2.188	0.0288	**
dadcoll	0.0404604	0.00523385	7.731	1.36e-014	***
cue80	0.00172507	0.000682255	2.528	0.0115	**
stwmfg80	-0.00342624	0.00140585	-2.437	0.0149	**
momcoll	0.0267018	0.00578988	4.612	4.12e-06	***
Media de la vble. dep.	2.618383	D.T. de la vble. dep.		0.128938	
Suma de cuad. residuos	45.13608	D.T. de la regresión		0.109216	
R-cuadrado	0.284598	R-cuadrado corregido		0.282518	
F(11, 3784)	136.8483	Valor p (de F)		2.4e-265	
Log-verosimilitud	3025.686	Criterio de Akaike		-6027.372	
Criterio de Schwarz	-5952.471	Crit. de Hannan-Quinn		-6000.751	

Si *dist* aumenta de 2 a 3,  $\ln(ed)$  disminuirá en 0.0022. Esto significa que los años de educación se estima que disminuyan en un 0.22%, si la distancia aumentada en 10 millas.

2. En primer lugar añadimos el cuadrado de la variable *dist* y posteriormente estimamos el nuevo modelo no lineal en variables. Para ello marcamos la variable *dist* y

#### AÑADIR-CUADRADO DE LA VARIABLE SELECCIONADA

Posteriormente estimamos el modelo:

gretl: modelo 8					
Archivo Editar Contraste Guardar Gráficos Análisis LaTeX					
Modelo 8: MCO, usando las observaciones 1-3796					
Variable dependiente: ed					
	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	8.95141	0.253671	35.29	4.68e-236	***
dist	-0.0748923	0.0253454	-2.955	0.0031	***
bytest	0.0920443	0.00317263	29.01	2.85e-167	***
female	0.143719	0.0504340	2.850	0.0044	***
black	0.337298	0.0716878	4.705	2.63e-06	***
hispanic	0.374506	0.0755410	4.958	7.45e-07	***
incomehi	0.364234	0.0606671	6.004	2.11e-09	***
ownhome	0.150390	0.0666574	2.256	0.0241	**
dadcoll	0.560174	0.0738515	7.585	4.16e-014	***
cue80	0.0216798	0.00970365	2.234	0.0255	**
stwmfg80	-0.0503283	0.0197936	-2.543	0.0110	**
momcoll	0.379089	0.0815179	4.650	3.43e-06	***
sq_dist	0.00460274	0.00231163	1.991	0.0465	**
Media de la vble. dep.	13.82929	D.T. de la vble. dep.	1.813969		
Suma de cuad. residuos	8944.896	D.T. de la regresión	1.537692		
R-cuadrado	0.283685	R-cuadrado corregido	0.281413		
F(12, 3783)	124.8499	Valor p (de F)	3.2e-263		
Log-verosimilitud	-7013.133	Criterio de Akaike	14052.27		
Criterio de Schwarz	14133.41	Crit. de Hannan-Quinn	14081.11		

Cuando *Dist* pasa de 2 a 3, el cambio estimado en *ED* es:

$$(-0.075 \times 3 + 0.0046 \times 3^2) - (-0.075 \times 2 + 0.0046 \times 2^2) = -0.052.$$

Esto significa que el número de años de educación completados se estima que disminuya en 0.052 años.

Cuando *dist* pasa de 6 a 7, el cambio estimado en *ED* es:

$$(-0.075 \times 7 + 0.0046 \times 7^2) - (-0.075 \times 6 + 0.0046 \times 6^2) = -0.015.$$

Esto significa que el número de años de educación completados se estima que disminuya en 0.015 años.

**3.** Añadimos la nueva variable y estimamos el modelo:

gretl: modelo 11					
Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX					
Modelo 11: MCO, usando las observaciones 1-3796					
Variable dependiente: ed					
	Coefficiente	Desv. Tipica	Estadístico t	Valor p	
const	8.94090	0.253576	35.26	1.01e-235	***
dist	-0.0746826	0.0253318	-2.948	0.0032	***
bytest	0.0919707	0.00317108	29.00	3.56e-167	***
female	0.141037	0.0504206	2.797	0.0052	***
black	0.333963	0.0716641	4.660	3.27e-06	***
hispanic	0.371202	0.0755141	4.916	9.22e-07	***
incomehi	0.357051	0.0607173	5.881	4.44e-09	***
ownhome	0.148384	0.0666272	2.227	0.0260	**
dadcoll	0.652418	0.0843367	7.736	1.31e-014	***
cue80	0.0214712	0.00969882	2.214	0.0269	**
stwmfg80	-0.0493682	0.0197874	-2.495	0.0126	**
momcoll	0.569933	0.117315	4.858	1.23e-06	***
sq_dist	0.00463833	0.00231042	2.008	0.0448	**
dadcoll_momcoll	-0.364918	0.161400	-2.261	0.0238	**
Media de la vble. dep.	13.82929	D.T. de la vble. dep.	1.813969		
Suma de cuad. residuos	8932.822	D.T. de la regresión	1.536857		
R-cuadrado	0.284652	R-cuadrado corregido	0.282193		
F(13, 3782)	115.7645	Valor p (de F)	2.9e-263		
Log-verosimilitud	-7010.570	Criterio de Akaike	14049.14		
Criterio de Schwarz	14136.52	Crit. de Hannan-Quinn	14080.20		

La nueva variable es significativa y el coeficiente estimado es -0.365 este es el efecto extra, además de los efectos de momcoll y dadcoll, cuando ambos padre y madre iban al mismo instituto.

#### 4. Creamos las dos nuevas variables de interacción:

$$\text{dist} * \text{incomehi} = \text{dist} * \text{incomehi}$$

$$\text{dist2\_incomehi} = \text{sq\_dist} * \text{incomehi}$$

y estimamos el modelo añadiendo estas dos variables:



gretl: modelo 12					
Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX					
Modelo 12: MCO, usando las observaciones 1-3796					
Variable dependiente: ed					
	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	8.97365	0.254012	35.33	1.70e-236	***
dist	-0.100759	0.0284067	-3.547	0.0004	***
bytest	0.0921020	0.00317084	29.05	1.29e-167	***
female	0.141856	0.0504089	2.814	0.0049	***
black	0.336611	0.0716630	4.697	2.73e-06	***
hispanic	0.368168	0.0755136	4.876	1.13e-06	***
incomehi	0.221647	0.0907049	2.444	0.0146	**
ownhome	0.151287	0.0666233	2.271	0.0232	**
dadcoll	0.660637	0.0844398	7.824	6.60e-015	***
cue80	0.0214007	0.00969731	2.207	0.0274	**
stwmfg80	-0.0503251	0.0197875	-2.543	0.0110	**
momcoll	0.568227	0.117309	4.844	1.32e-06	***
sq_dist	0.00630745	0.00249234	2.531	0.0114	**
dadcoll_momcoll	-0.354772	0.161528	-2.196	0.0281	**
dist_incomehi	0.115769	0.0640408	1.808	0.0707	*
dist2_incomehi	-0.00804633	0.00713511	-1.128	0.2595	
Media de la vble. dep.	13.82929	D.T. de la vble. dep.	1.813969		
Suma de cuad. residuos	8922.990	D.T. de la regresión	1.536418		
R-cuadrado	0.285440	R-cuadrado corregido	0.282604		
F(15, 3780)	100.6643	Valor p (de F)	4.3e-262		
Log-verosimilitud	-7008.479	Criterio de Akaike	14048.96		
Criterio de Schwarz	14148.83	Crit. de Hannan-Quinn	14084.45		
Sin considerar la constante, el valor p más alto fue el de la variable 24 (dist2_incomehi)					

Estudiantes sin renta alta ( $Incomehi = 0$ )

$$\widehat{ed} = -0.100dist + 0.0063 sq\_dist + \text{other factors}$$

Estudiantes con renta alta ( $Incomehi = 1$ )

$$\begin{aligned}\widehat{ed} &= (-0.100 + 0.115) dist + (0.0063 - 0.008) sq\_dist + \text{other factors} \\ &= 0.015 Dist - 0.0017 sq\_dist + \text{other factors}.\end{aligned}$$

Para contrastar si la diferencia entre rentas altas y no altas, del efecto de la distancia sobre los años de educación es significativo, hacemos un contraste:

#### CONTRASTES-RESTRICCIONES LINEALES

$$b_{15} = 0$$

$$b_{16} = 0$$

Conjunto de restricciones					
1: b[dist_incomehi] = 0					
2: b[dist2_incomehi] = 0					
Estadístico de contraste: F(2, 3780) = 2.08242, con valor p = 0.124771					
Estimaciones restringidas:					
	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	8.94090	0.253576	35.26	1.01e-235	***
dist	-0.0746826	0.0253318	-2.948	0.0032	***
bytest	0.0919707	0.00317108	29.00	3.56e-167	***
female	0.141037	0.0504206	2.797	0.0052	***
black	0.333963	0.0716641	4.660	3.27e-06	***
hispanic	0.371202	0.0755141	4.916	9.22e-07	***
incomehi	0.357051	0.0607173	5.881	4.44e-09	***
ownhome	0.148384	0.0666272	2.227	0.0260	**
dadcoll	0.652418	0.0843367	7.736	1.31e-014	***
cue80	0.0214712	0.00969882	2.214	0.0269	**
stwmfg80	-0.0493682	0.0197874	-2.495	0.0126	**
momcoll	0.569933	0.117315	4.858	1.23e-06	***
sq_dist	0.00463833	0.00231042	2.008	0.0448	**
dadcoll_momcoll	-0.364918	0.161400	-2.261	0.0238	**
dist_incomehi	0.00000	0.00000	NA	NA	
dist2_incomehi	0.00000	0.00000	NA	NA	
Desviación típica de la regresión = 1.53686					

El contraste de que las dos variables de interacción son cero tiene un p-valor de 0.124 por lo que no son significativos al 5%.

5. (1) Variables omitidas aunque se intenta medir el efecto de la riqueza de las familias con la variable *incomehi*, esto puede ser escaso y podemos tener omisión de variables relevantes
- (2) Error en la forma funcional: Hemos probado distintas formas funcionales
- (3) Errores en variables: La renta de las familias solo está medida con la ficticia y podría no estar correctamente definida incluso la forma de medir los años de educación.
- (4) Causalidad simultánea: el argumento podría ser que los padres que quieran enviar a lo hijos al instituto pueden ir a vivir más cerca del mismo. El efecto debe de ser pequeño
- (5) Inconsistencia de los errores estándar MCO: Puede que haya problemas de heterocedasticidad que no se ha contrastado, habría que hacerlo y/o estimar con estimación robusta a heterocedasticidad.

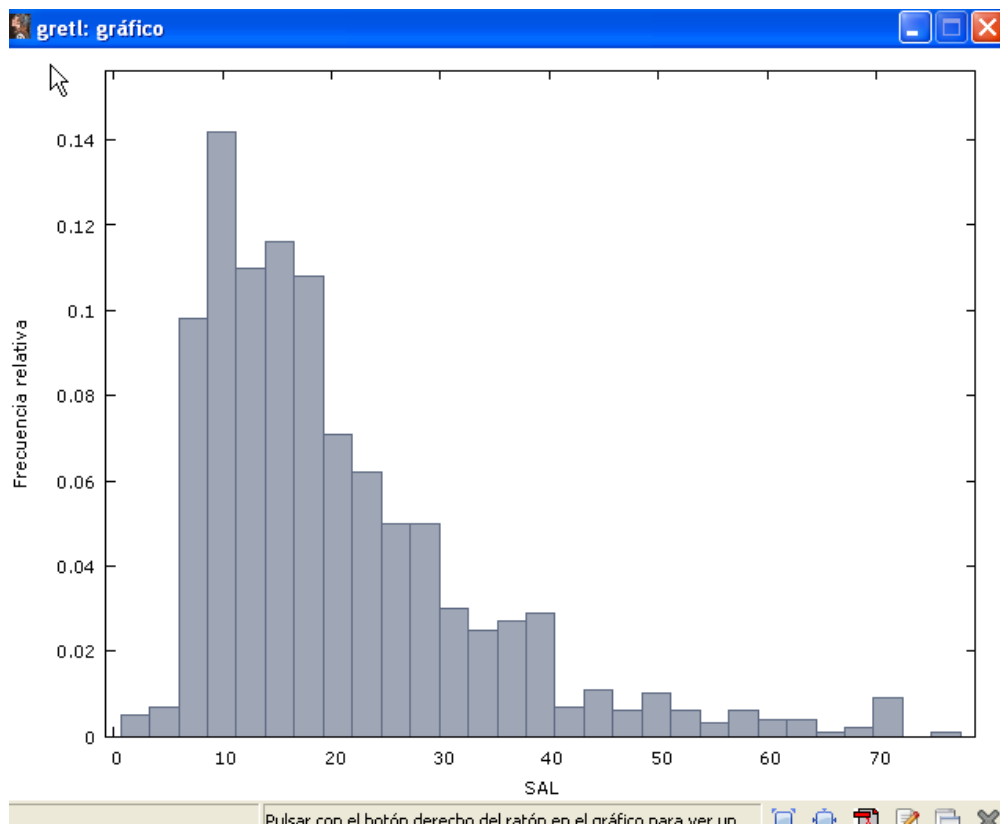
## Solución ejercicio empírico 2.4

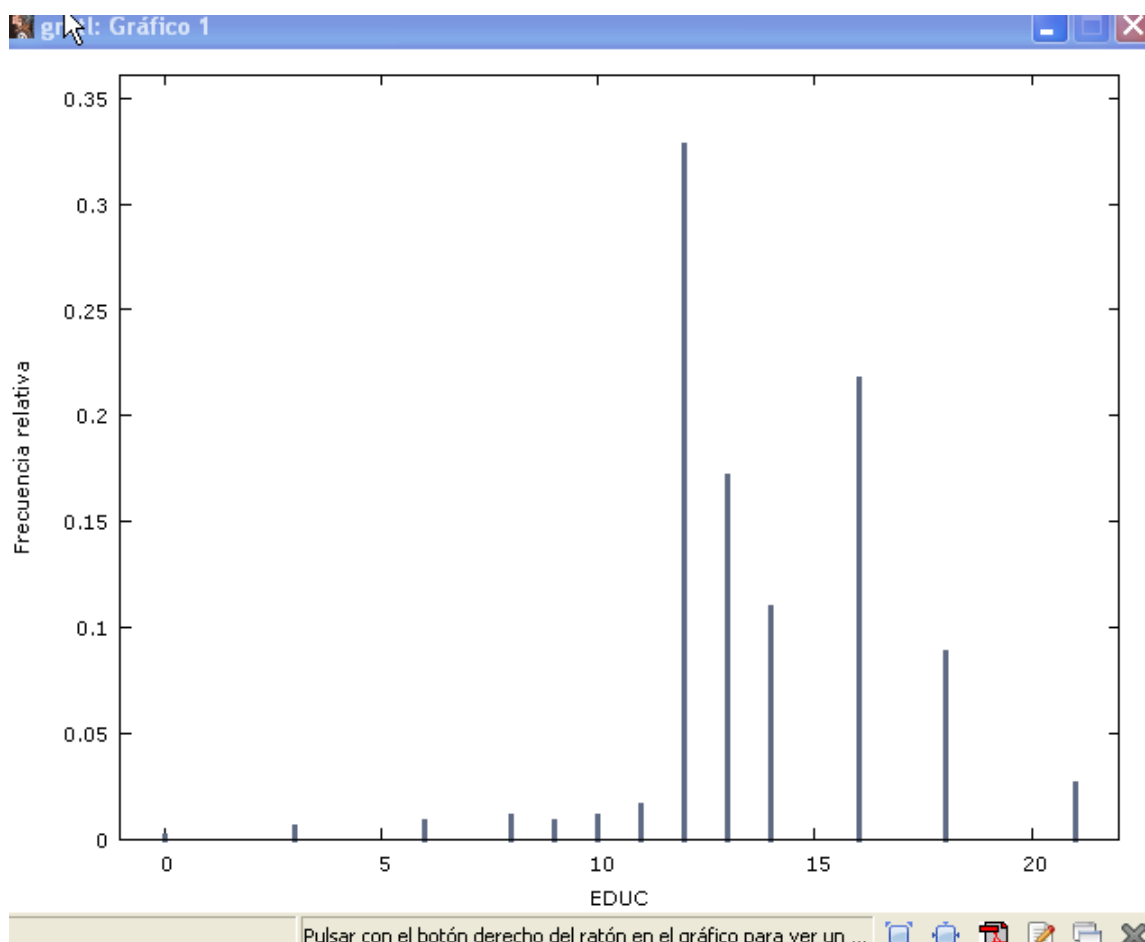
1)

Marcar Las variables e indicar con la parte derecha del ratón **ESTADISTICOS PRINCIPALES**

gretl: estadísticos principales				
	Media	Mediana	Mínimo	Máximo
SAL	20.616	17.300	1.9700	76.390
EDUC	13.799	13.000	0.0000	21.000
	Desv. Típica.	C.V.	Asimetría	Exc. de curtosis
SAL	12.835	0.62257	1.5839	2.9214
EDUC	2.7111	0.19647	-0.070932	2.0697
	Perc. 5%	Perc. 95%	Rango IQ	Observaciones ausentes
SAL	7.4905	47.595	14.380	0
EDUC	11.000	18.000	4.0000	0

Para el histograma, marcar la variable e ir a **VARIABLE-DISTRIBUCIÓN DE FRECUENCIAS**





## 2) Para estimar **MODELO-MINIMOS CUADRADOS ORDINARIOS**.

gretl: modelo 1

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

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

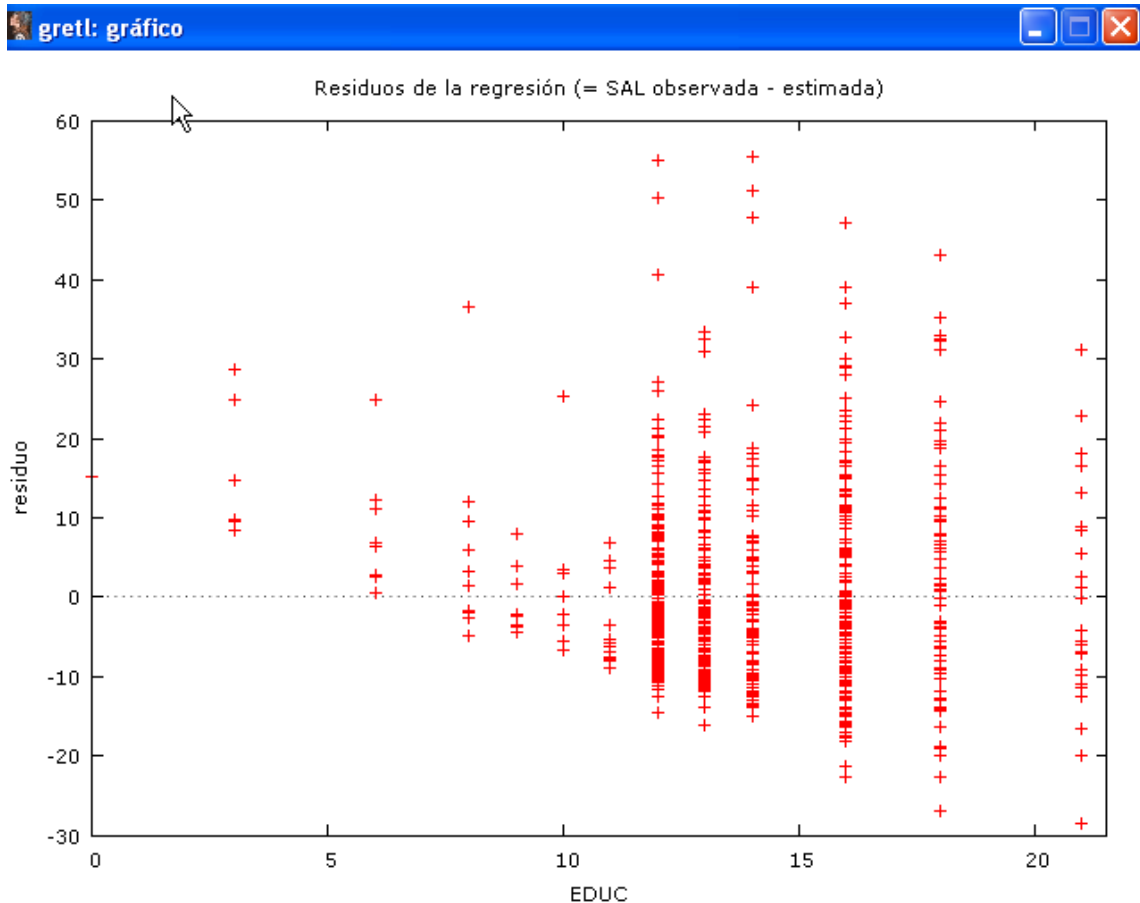
	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	-6.71033	1.91416	-3.506	0.0005	***
EDUC	1.98029	0.136117	14.55	1.25e-043	***

Media de la vble. dep.	20.61566	D.T. de la vble. dep.	12.83472
Suma de cuad. residuos	135771.1	D.T. de la regresión	11.66376
R-cuadrado	0.174972	R-cuadrado corregido	0.174145
F(1, 998)	211.6554	Valor p (de F)	1.25e-43
Log-verosimilitud	-3874.424	Criterio de Akaike	7752.848
Criterio de Schwarz	7762.663	Crit. de Hannan-Quinn	7756.578

En principio, suponiendo el cumplimiento de las hipótesis básicas la variable EDUC es significativa en la explicación del salario y la influencia es positiva. Por cada año más de educación el salario se incrementa en 1.98 unidades.

3) Para guardar los residuos en el modelo que hemos estimado le indicamos **GUARDAR- RESIDUOS** y para hacer el gráfico **GRAFICOS-RESIDUOS-CONTRA EDUC**



Parece que se incumple la hipótesis de homocedasticidad ya que a mayor número de años de educación mayor dispersión en los residuos.

4) Para seleccionar solo hombres hay que indicar **MUESTRA-RESTRINGIR A PARTIR DE CRITERIO** e indicar que para FEMALE = 0

HOMBRES

gretl: modelo 2

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

Modelo 2: MCO, usando las observaciones 1-486  
Variable dependiente: SAL

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	-3.05449	2.49353	-1.225	0.2212
EDUC	1.87530	0.181443	10.34	9.26e-023 ***
Media de la vble. dep.	22.14245	D.T. de la vble. dep.	12.74380	
Suma de cuad. residuos	64524.96	D.T. de la regresión	11.54626	
R-cuadrado	0.180803	R-cuadrado corregido	0.179111	
F(1, 484)	106.8226	Valor p (de F)	9.26e-23	
Log-verosimilitud	-1877.534	Criterio de Akaike	3759.067	
Criterio de Schwarz	3767.440	Crit. de Hannan-Quinn	3762.357	

## MUJERES

Hay que cambiar la muestra **MUESTRA-RECUPERAR EL RANGO COMPLETO** e indicar la nueva muestra **MUESTRA-RESTRINGIR A PARTIR DE CRITERIO-UTILIZAR LA VARIABLE FICTICIA FEMALE**.

gretl: modelo 3

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

Modelo 3: MCO, usando las observaciones 1-514  
Variable dependiente: SAL

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	-14.1680	2.89570	-4.893	1.33e-06 ***
EDUC	2.35752	0.201674	11.69	3.81e-028 ***
Media de la vble. dep.	19.17204	D.T. de la vble. dep.	12.76533	
Suma de cuad. residuos	65984.27	D.T. de la regresión	11.35234	
R-cuadrado	0.210669	R-cuadrado corregido	0.209127	
F(1, 512)	136.6507	Valor p (de F)	3.81e-28	
Log-verosimilitud	-1977.056	Criterio de Akaike	3958.112	
Criterio de Schwarz	3966.597	Crit. de Hannan-Quinn	3961.438	

## ASIATICOS

gretl: modelo 4

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

Modelo 4: MCO, usando las observaciones 1-43  
Variable dependiente: SAL

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	7.54635	10.5633	0.7144	0.4790
EDUC	1.02051	0.659640	1.547	0.1295
Media de la vble. dep.	23.61349	D.T. de la vble. dep.	12.86554	
Suma de cuad. residuos	6568.485	D.T. de la regresión	12.65729	
R-cuadrado	0.055157	R-cuadrado corregido	0.032112	
F(1, 41)	2.393439	Valor p (de F)	0.129530	
Log-verosimilitud	-169.1344	Criterio de Akaike	342.2688	
Criterio de Schwarz	345.7912	Crit. de Hannan-Quinn	343.5677	

## BLANCOS

Para ello hay que restringir la muestra a BLACK=0 && ASIAN=0

gretl: modelo 16

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

Modelo 16: MCO, usando las observaciones 1-845  
Variable dependiente: SAL

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	-6.55065	2.07641	-3.155	0.0017 ***
EDUC	1.99193	0.148160	13.44	1.79e-037 ***
Media de la vble. dep.	20.83902	D.T. de la vble. dep.	12.85092	
Suma de cuad. residuos	114773.8	D.T. de la regresión	11.66830	
R-cuadrado	0.176560	R-cuadrado corregido	0.175583	
F(1, 843)	180.7541	Valor p (de F)	1.79e-37	
Log-verosimilitud	-3274.062	Criterio de Akaike	6552.124	
Criterio de Schwarz	6561.602	Crit. de Hannan-Quinn	6555.756	

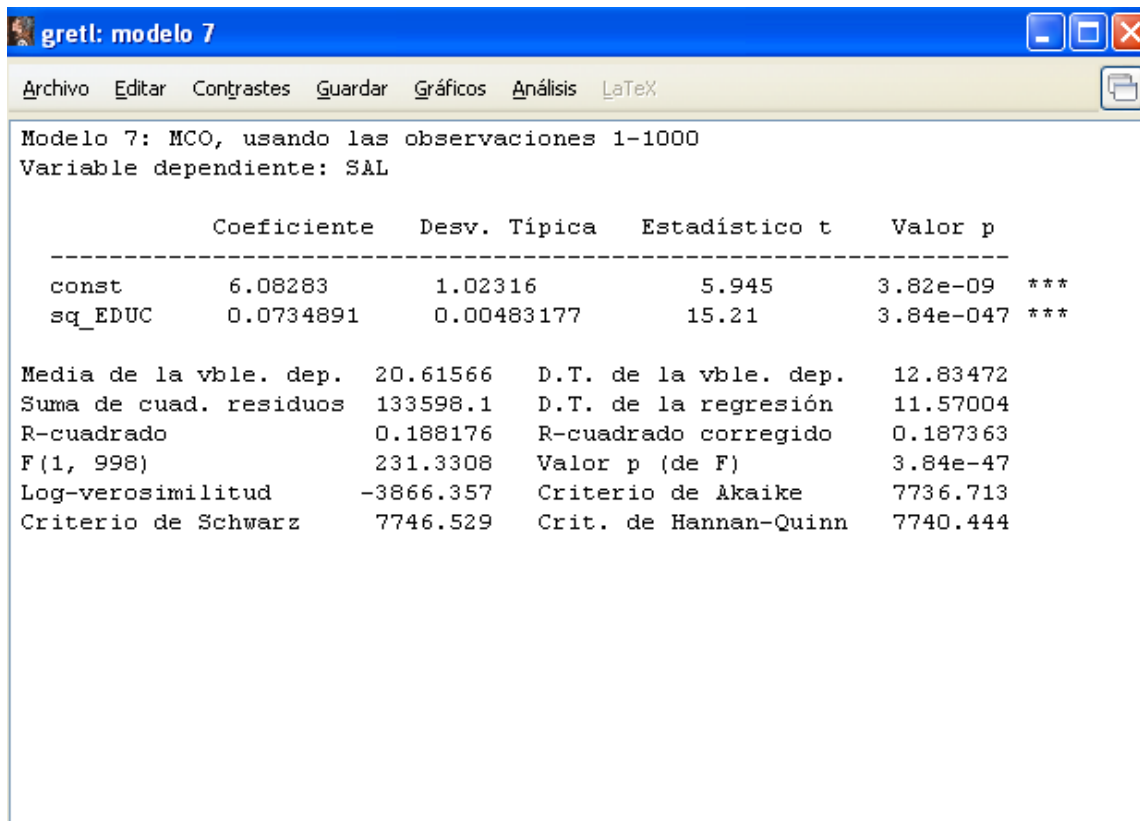
En todos los grupos la variable EDUC es significativa excepto en el grupo de asiáticos.

En las mujeres es donde cada año de educación tiene más efecto en el salario.

## RECORDAR: MUESTRA-RECUPERAR RANGO COMPLETO

### 5) Primero seleccionar la variable EDUC y AÑADIR-CUADRADO DE LA VARIABLE SELECCIONADA

Hacer la regresión.



Modelo 7: MCO, usando las observaciones 1-1000  
Variable dependiente: SAL

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	6.08283	1.02316	5.945	3.82e-09	***
sq_EDUC	0.0734891	0.00483177	15.21	3.84e-047	***
Media de la vble. dep.	20.61566	D.T. de la vble. dep.	12.83472		
Suma de cuad. residuos	133598.1	D.T. de la regresión	11.57004		
R-cuadrado	0.188176	R-cuadrado corregido	0.187363		
F(1, 998)	231.3308	Valor p (de F)	3.84e-47		
Log-verosimilitud	-3866.357	Criterio de Akaike	7736.713		
Criterio de Schwarz	7746.529	Crit. de Hannan-Quinn	7740.444		

Esto implica que el efecto de la educación en el salario depende del número de años de educación.

$$\frac{\partial SAL}{\partial EDUC} = 2 * \hat{\beta}_2 * EDUC$$

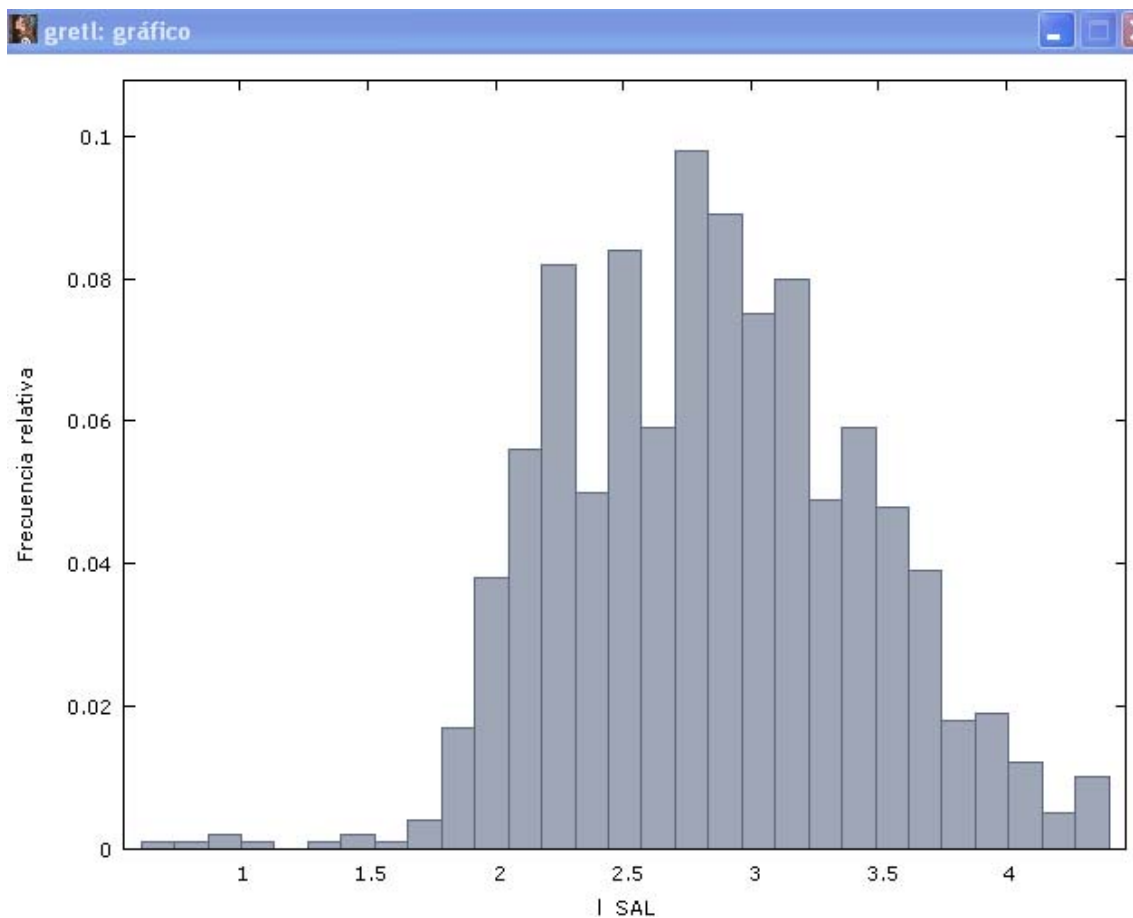
Para una persona que tiene 12 años de educación  $= 2 * 0.0735 * 12 = 1.76$ , por cada año más de educación el salario se incrementa en 1.76 unidades.

Para una persona con 16 años de educación  $= 2 * 0.0735 * 16 = 2.352$ , por cada año de educación en una persona con 16 años de educación, su salario se incrementa en 2.352 unidades.

En la regresión del apartado 2) el efecto era 1.98 para cualquier número de años de educación.



**6) Marcar la variable SAL y AÑADIR-LOGARITMO DE LA VARIABLE SELECCIONADA**



Esta se aproxima más a una normal porque es más simétrica.

7)

gretl: modelo 8				
Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX				
Modelo 8: MCO, usando las observaciones 1-1000				
Variable dependiente: l_SAL				
	Coefficiente	Desv. Típica	Estadístico t	Valor p
const	1.60944	0.0864229	18.62	1.15e-066 ***
EDUC	0.0904082	0.00614561	14.71	1.75e-044 ***
Media de la vble. dep.	2.856988	D.T. de la vble. dep.	0.580619	
Suma de cuad. residuos	276.7649	D.T. de la regresión	0.526611	
R-cuadrado	0.178205	R-cuadrado corregido	0.177381	
F(1, 998)	216.4141	Valor p (de F)	1.75e-44	
Log-verosimilitud	-776.6451	Criterio de Akaike	1557.290	
Criterio de Schwarz	1567.106	Crit. de Hannan-Quinn	1561.021	

En este modelo log-lin el efecto marginal de un año adicional de educación sobre el salario es:  $\frac{\partial SAL}{\partial EDUC} = \beta_2 e^{\beta_1 + \beta_2 EDUC}$  por que  $SAL = e^{\beta_1 + \beta_2 EDUC + u}$

Por lo que para una persona de 12 años de educación será:  $0.09 * e^{1.6 + 0.09 * 12} = 1.31$

Y para una persona de 16 años de educación es:  $0.09 * e^{1.6 + 0.09 * 16} = 1.88$

Con la regresión del 2) era 1.98 independientemente del nivel de EDUC y en el apartado 5) era para 12 años de EDUC: 1.76 y para 16 años de EDUC: 2.352.

8) El contraste de significatividad es:

$$H_0 : \beta_2 = 0 \quad |t| = \left| \frac{1.98}{0.136} \right| = 14.55 > 1.96; \quad p\text{-valor} = 0.000$$

$$H_1 : \beta_2 \neq 0$$

Se rechaza la hipótesis nula para cualquier nivel de significación y por lo tanto se acepta la significatividad individual.

La estimación por intervalo es:

$$\hat{\beta}_2 \pm t_{\varepsilon/2} SE(\hat{\beta}_2) = 1.98 \pm 1.96 * 0.136 = (1.71; 2.25)$$

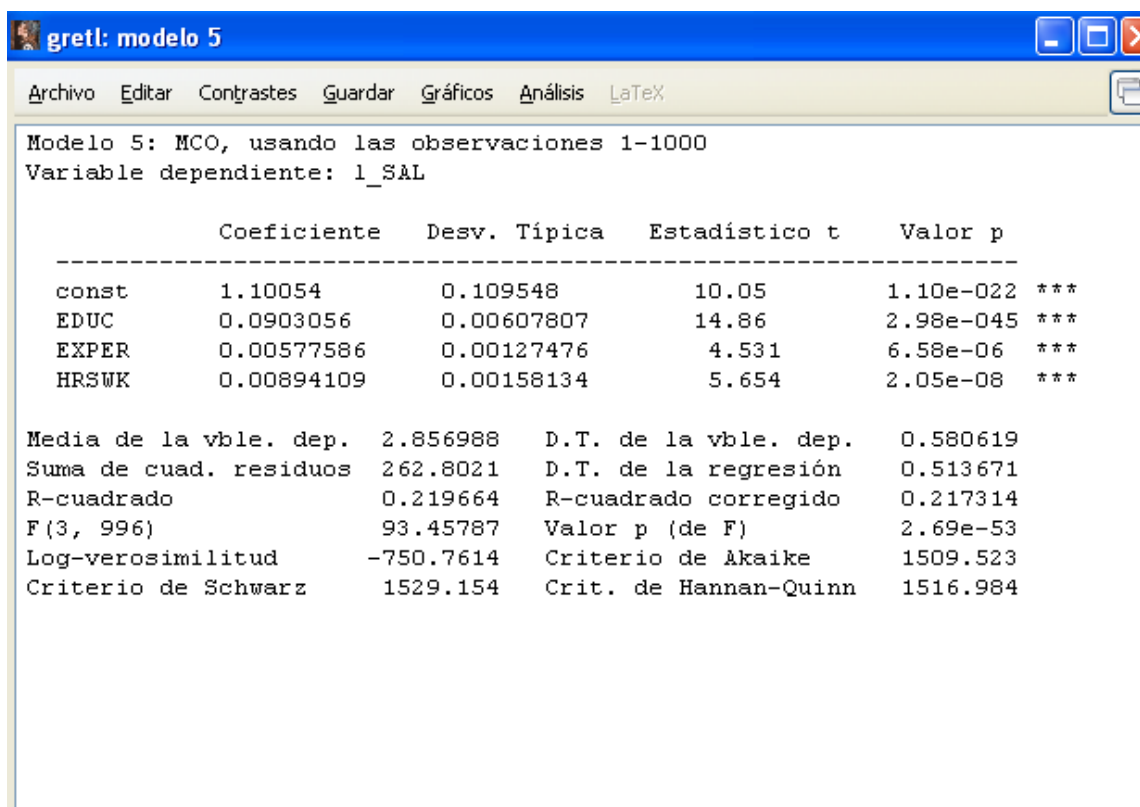
9)

$$H_0 : \beta_2 \geq 0 \quad t = \frac{1.98}{0.136} = 14.55 > -2.33;$$

$$H_1 : \beta_2 < 0$$

El punto crítico es el de la Normal (asintóticamente) pero la región crítica sólo tiene una cola, por ser una hipótesis nula compuesta unilateral,  $N_{0.01} = -2.33$ . Como  $14.55 > -2.33$  no se rechaza la hipótesis nula para un nivel de significación del 5%.

10)



	Coefficiente	Desv. Típica	Estadístico t	Valor p
const	1.10054	0.109548	10.05	1.10e-022 ***
EDUC	0.0903056	0.00607807	14.86	2.98e-045 ***
EXPER	0.00577586	0.00127476	4.531	6.58e-06 ***
HRSWK	0.00894109	0.00158134	5.654	2.05e-08 ***

Media de la vble. dep.	2.856988	D.T. de la vble. dep.	0.580619
Suma de cuad. residuos	262.8021	D.T. de la regresión	0.513671
R-cuadrado	0.219664	R-cuadrado corregido	0.217314
F(3, 996)	93.45787	Valor p (de F)	2.69e-53
Log-verosimilitud	-750.7614	Criterio de Akaike	1509.523
Criterio de Schwarz	1529.154	Crit. de Hannan-Quinn	1516.984

$\beta_2$  por cada año adicional de educación, el salario horario se incrementa en 9 por ciento

$\beta_3$  por cada año adicional de experiencia, el salario horario se incrementa en 0.57 por ciento

$\beta_4$  por cada hora adicional trabajada a la semana, el salario horario se incrementa en 0.89 por ciento.

Los tres son significativos individualmente como indican los asteriscos.

11)

$$H_0 : \beta_2 \geq 0.10 \quad t = \frac{0.09 - 0.10}{0.006} = -1.66 < -1.64$$

$$H_1 : \beta_2 < 0.10$$

Por ser la hipótesis alternativa compuesta unilateral, la región crítica está concentrada en una zona (0.05). Rechazo la hipótesis nula de que el efecto de un año adicional de educación es al menos de un crecimiento del salario del 10%.

12)

$$\left( \hat{\beta}_4 \pm N_{0.005} * SE(\hat{\beta}_4) \right) = (0.0089 \pm 2.58 * 0.0016) = (0.0047; 0.013)$$

13) En primer lugar hay que añadir las nuevas variables:

AÑADIR-DEFINIR NUEVA VARIABLE EDUC-EXPER=EDUC\*EXPER

Y añadir el cuadrado de la variable EXPER.

Estimar el modelo:

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	3.80993	7.55881	0.5040	0.6143
EDUC	-0.994310	0.821589	-1.210	0.2265
EXPER	0.651100	0.217122	2.999	0.0028 ***
EDUC_EXPER	0.00758525	0.0113255	0.6698	0.5032
sq_EDUC	0.102312	0.0247397	4.136	3.84e-05 ***
sq_EXPER	-0.0111861	0.00196206	-5.701	1.57e-08 ***

Media de la vble. dep.	20.61566	D.T. de la vble. dep.	12.83472
Suma de cuad. residuos	125163.6	D.T. de la regresión	11.22137
R-cuadrado	0.239429	R-cuadrado corregido	0.235604
F(5, 994)	62.58272	Valor p (de F)	8.35e-57
Log-verosimilitud	-3833.749	Criterio de Akaike	7679.499
Criterio de Schwarz	7708.945	Crit. de Hannan-Quinn	7690.691

Sin considerar la constante, el valor p más alto fue el de la variable 16 (EDUC\_EXPER)

$$14) \frac{\partial SAL}{\partial EDUC} = \beta_2 + \beta_4 * EXPER + 2 * \beta_5 * EDUC$$

$$\frac{\partial SAL}{\partial EXPER} = \beta_3 + \beta_4 * EDUC + 2 * \beta_6 * EXPER$$

$$15) \text{ Hill : } \frac{\partial SAL}{\partial EDUC} = -0.99 + 0.00758 * 2 + 2 * 0.1023 * 16 = 2.2987$$

$$\text{Wendy: } \frac{\partial SAL}{\partial EDUC} = -0.99 + 0.00758 * 2 + 2 * 0.1023 * 12 = 1.4804$$

El efecto depende de los años de educación y de los de experiencia.

16)

$$H_0 : \beta_5 \leq 0 \quad t = \frac{0.10}{0.025} = 4.136 > 1.64$$

$$H_1 : \beta_5 > 0$$

Como  $t = 4.136 > 1.64$  se rechaza la hipótesis nula de que el efecto marginal es menor o igual en Hill que en Wendy.

17) El efecto marginal es:

$$\beta_2 + \beta_4 * EXPER + 2 * \beta_5 * EDUC$$

La estimación es:  $-0.99 + 0.00758 * 2 + 2 * 0.1023 * 16 = 2.2987$

Para calcular el intervalo de confianza al 95% contrastamos la restricción  $\mathbf{b2+2*b4+32*b5=0}$  y como sólo hay una restricción  $F_{(1,T-k)} = (t_{T-k})^2$ , a partir de ahí como conocemos la estimación y la  $t$ , despejamos la desviación típica para poder calcular el intervalo de confianza.



Restricción:

$$b[EDUC] + 2 * b[EDUC\_EXPER] + 32 * b[sq\_EDUC] = 0$$

Estadístico de contraste:  $F(1, 994) = 54.8984$ , con valor  $p = 2.70824e-013$

Estimaciones restringidas:

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	33.4519	6.58474	5.080	4.50e-07	***
EDUC	-2.60800	0.813370	-3.206	0.0014	***
EXPER	-0.505615	0.154929	-3.264	0.0011	***
EDUC_EXPER	0.0774017	0.00645109	12.00	4.50e-031	***
sq_EDUC	0.0766623	0.0251511	3.048	0.0024	***
sq_EXPER	-0.00757638	0.00195141	-3.883	0.0001	***

Desviación típica de la regresión = 11.5213

$$t = \sqrt{54.8984} = 7.41; SE = estimación / t = 2.2987 / 7.41 = 0.31$$

Por lo que el intervalo de confianza al 95% será:  $(2.2987 \pm 1.96 * 0.31) = (1.6911; 2.9063)$

18)

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	-16.4432	2.44008	-6.739	2.70e-011	***
EDUC	2.01199	0.135384	14.86	2.85e-045	***
EXPER	0.143709	0.0283942	5.061	4.96e-07	***
HRSWK	0.137313	0.0352230	3.898	0.0001	***
Media de la vble. dep.	20.61566	D.T. de la vble. dep.	12.83472		
Suma de cuad. residuos	130386.0	D.T. de la regresión	11.44157		
R-cuadrado	0.207695	R-cuadrado corregido	0.205309		
F(3, 996)	87.03067	Valor p (de F)	5.13e-50		
Log-verosimilitud	-3854.188	Criterio de Akaike	7716.376		
Criterio de Schwarz	7736.007	Crit. de Hannan-Quinn	7723.837		

## CONTRASTES-RESTRICCIONES LINEALES

**b3 = 0**

**b4 = 0**

Conjunto de restricciones

1: b[EXPER] = 0

2: b[HRSWK] = 0

Estadística de contraste:  $F(2, 996) = 20.5683$ , con valor p =  $1.76547e-009$

Estimaciones restringidas:

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	-6.71033	1.91416	-3.506	0.0005	***
EDUC	1.98029	0.136117	14.55	1.25e-043	***
EXPER	0.00000	0.00000	NA	NA	
HRSWK	0.00000	0.00000	NA	NA	

Desviación típica de la regresión = 11.6638

Se rechaza la hipótesis conjunta. Luego los dos son significativas conjuntamente.

19)

gretl: modelo 11					
Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX					
Modelo 11: MCO, usando las observaciones 1-1000					
Variable dependiente: l_SAL					
	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	1.72208	0.0907083	18.98	8.46e-069	***
EDUC	0.0952991	0.00601246	15.85	1.32e-050	***
BLACK	-0.201247	0.0798684	-2.520	0.0119	**
FEMALE	-0.257876	0.0344216	-7.492	1.50e-013	***
BLACK_FEMALE	0.176613	0.103661	1.704	0.0887	*
SOUTH	-0.0370787	0.0458316	-0.8090	0.4187	
MID	-0.130664	0.0473796	-2.758	0.0059	***
WEST	0.0234945	0.0473872	0.4958	0.6201	
Media de la vble. dep.	2.856988	D.T. de la vble. dep.	0.580619		
Suma de cuad. residuos	256.9892	D.T. de la regresión	0.508981		
R-cuadrado	0.236924	R-cuadrado corregido	0.231540		
F(7, 992)	44.00031	Valor p (de F)	2.62e-54		
Log-verosimilitud	-739.5779	Criterio de Akaike	1495.156		
Criterio de Schwarz	1534.418	Crit. de Hannan-Quinn	1510.078		
Sin considerar la constante, el valor p más alto fue el de la variable					
10 (WEST)					

La interacción BLACK-FEMALE solo es significativa al 10%

Para contrastar que no hay efecto regional:

## CONTRASTES-RESTRICCIONES LINEALES

**b6 = 0**

**b7 = 0**

**b8 = 0**

Conjunto de restricciones				
1: b[SOUTH] = 0				
2: b[MID] = 0				
3: b[WEST] = 0				
Estadístico de contraste: $F(3, 992) = 4.22231$ , con valor $p = 0.00560455$				
Estimaciones restringidas:				
	Coefficiente	Desv. Típica	Estadístico t	Valor p
const	1.68619	0.0849721	19.84	4.28e-074 ***
EDUC	0.0953595	0.00603056	15.81	2.05e-050 ***
BLACK	-0.215265	0.0793495	-2.713	0.0068 ***
FEMALE	-0.259417	0.0345802	-7.502	1.39e-013 ***
BLACK_FEMALE	0.187663	0.104072	1.803	0.0717 *
SOUTH	0.00000	0.00000	NA	NA
MID	0.00000	0.00000	NA	NA
WEST	0.00000	0.00000	NA	NA
Desviación típica de la regresión = 0.511448				

Por lo que se rechaza la hipótesis nula de que los coeficientes de las variables que miden el efecto regional sean cero. Luego sí que hay efecto regional.

20) Marcar en la estimación que sea robusta a heterocedastidad.

gretl: modelo 13				
Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX				
Modelo 13: MCO, usando las observaciones 1-1000				
Variable dependiente: l_SAL				
Desviaciones típicas robustas ante heterocedasticidad, variante HC1				
	Coefficiente	Desv. Típica	Estadístico t	Valor p
const	0.529677	0.252825	2.095	0.0364 **
EDUC	0.127195	0.0169597	7.500	1.41e-013 ***
EXPER	0.0629807	0.0113775	5.536	3.97e-08 ***
sq_EXPER	-0.000713939	9.20134e-05	-7.759	2.11e-014 ***
EDUC_EXPER	-0.00132239	0.000636794	-2.077	0.0381 **
Media de la vble. dep.	2.856988	D.T. de la vble. dep.	0.580619	
Suma de cuad. residuos	254.4216	D.T. de la regresión	0.505668	
R-cuadrado	0.244548	R-cuadrado corregido	0.241511	
F(4, 995)	85.06746	Valor p (de F)	3.57e-62	
Log-verosimilitud	-734.5572	Criterio de Akaike	1479.114	
Criterio de Schwarz	1503.653	Crit. de Hannan-Quinn	1488.441	

21)



Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX					
Modelo 14: MCO, usando las observaciones 1-1000					
Variable dependiente: l_SAL					
Desviaciones típicas robustas ante heterocedasticidad, variante HC1					
	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	0.541061	0.254209	2.128	0.0335	**
EDUC	0.126120	0.0170564	7.394	3.02e-013	***
EXPER	0.0613731	0.0115877	5.296	1.45e-07	***
sq_EXPER	-0.000693346	9.55671e-05	-7.255	8.07e-013	***
EDUC_EXPER	-0.00130912	0.000638420	-2.051	0.0406	**
MARRIED	0.0402895	0.0339231	1.188	0.2352	
Media de la vble. dep.	2.856988	D.T. de la vble. dep.	0.580619		
Suma de cuad. residuos	254.0582	D.T. de la regresión	0.505561		
R-cuadrado	0.245627	R-cuadrado corregido	0.241833		
F(5, 994)	69.11228	Valor p (de F)	4.41e-62		
Log-verosimilitud	-733.8426	Criterio de Akaike	1479.685		
Criterio de Schwarz	1509.132	Crit. de Hannan-Quinn	1490.877		

La variable MARRIED no es significativa.

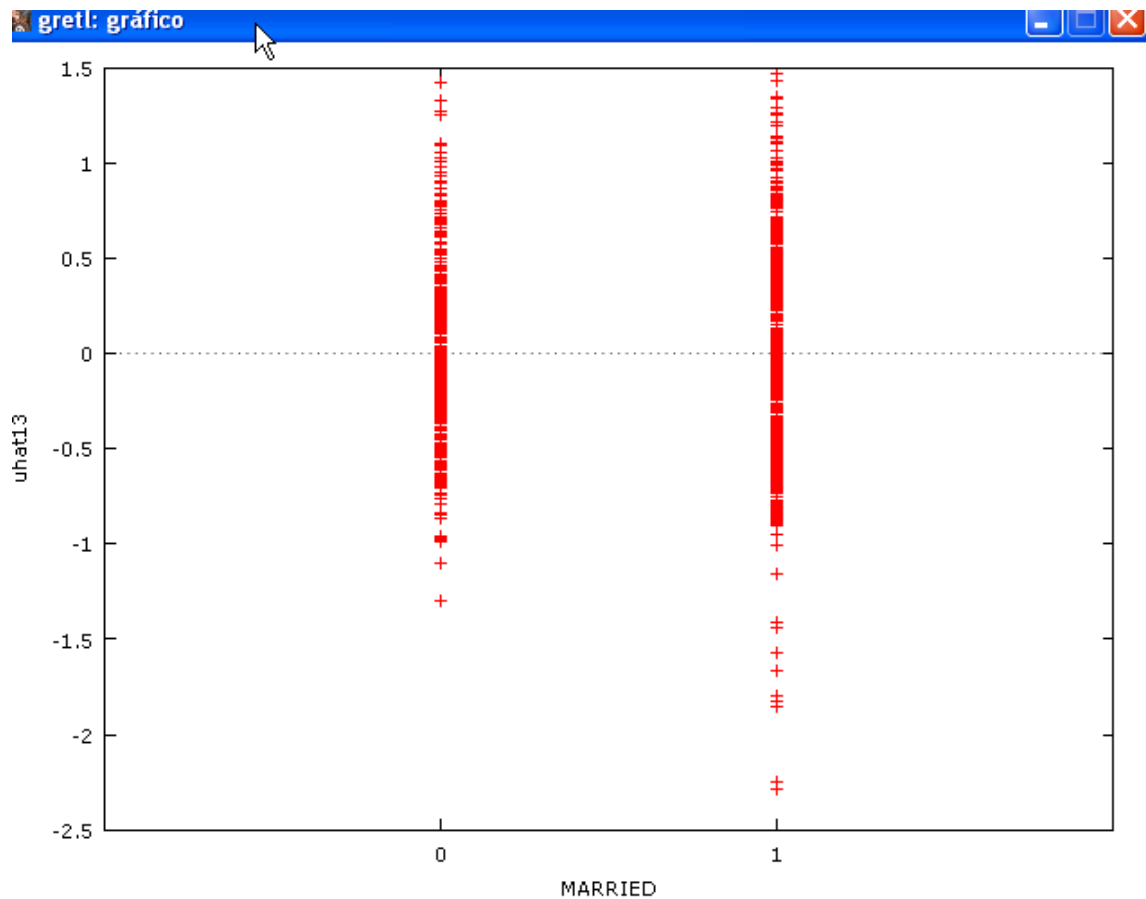
El contraste para un nivel de significación del 1% es:

$$H_0 : \beta_6 \leq 0 \quad t = \frac{0.0402}{0.0339} = 1.188 < 2.33 = N_{0.01}$$

$$H_1 : \beta_6 > 0$$

Y no se rechaza la hipótesis nula de que los salarios de los casados son menores que los de los solteros.

**22)** Guardar los residuos de la regresión y luego marcar las dos variables para las que queremos hacer el gráfico y GRAFICO DE DOS VARIABLES XY.



Hay evidencia de heterocedasticidad ya que para los casados la varianza de los residuos es mayor que para los no casados.

### 23) CASADOS

Recordar que hay que restringir el rango:

**MUESTRA-RESTRINGIR A PARTIR DE CRITERIO**

**MARRIED = 1**

**gretl: modelo 15**

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

Modelo 15: MCO, usando las observaciones 1-581  
Variable dependiente: l\_SAL

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	0.919696	0.355796	2.585	0.0100	***
EDUC	0.100828	0.0221957	4.543	6.77e-06	***
EXPER	0.0506938	0.0149271	3.396	0.0007	***
sq_EXPER	-0.000708813	0.000137875	-5.141	3.75e-07	***
EDUC_EXPER	-0.000461970	0.000747837	-0.6177	0.5370	
Media de la vble. dep.	2.922767	D.T. de la vble. dep.	0.600551		
Suma de cuad. residuos	165.0727	D.T. de la regresión	0.535336		
R-cuadrado	0.210872	R-cuadrado corregido	0.205392		
F(4, 576)	38.48002	Valor p (de F)	1.48e-28		
Log-verosimilitud	-458.8484	Criterio de Akaike	927.6968		
Criterio de Schwarz	949.5206	Crit. de Hannan-Quinn	936.2046		

Sin considerar la constante, el valor p más alto fue el de la variable 16 (EDUC\_EXPER)

## SOLTEROS

MARRIED = 0

**gretl: modelo 16**

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

Modelo 16: MCO, usando las observaciones 1-419  
Variable dependiente: l\_SAL

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
const	0.197488	0.294472	0.6707	0.5028	
EDUC	0.151292	0.0194232	7.789	5.48e-014	***
EXPER	0.0728360	0.0127057	5.733	1.91e-08	***
sq_EXPER	-0.000701383	0.000119277	-5.880	8.45e-09	***
EDUC_EXPER	-0.00214480	0.000653828	-3.280	0.0011	***
Media de la vble. dep.	2.765777	D.T. de la vble. dep.	0.539353		
Suma de cuad. residuos	88.11997	D.T. de la regresión	0.461357		
R-cuadrado	0.275312	R-cuadrado corregido	0.268310		
F(4, 414)	39.32007	Valor p (de F)	6.53e-28		
Log-verosimilitud	-267.8888	Criterio de Akaike	545.7775		
Criterio de Schwarz	565.9669	Crit. de Hannan-Quinn	553.7580		

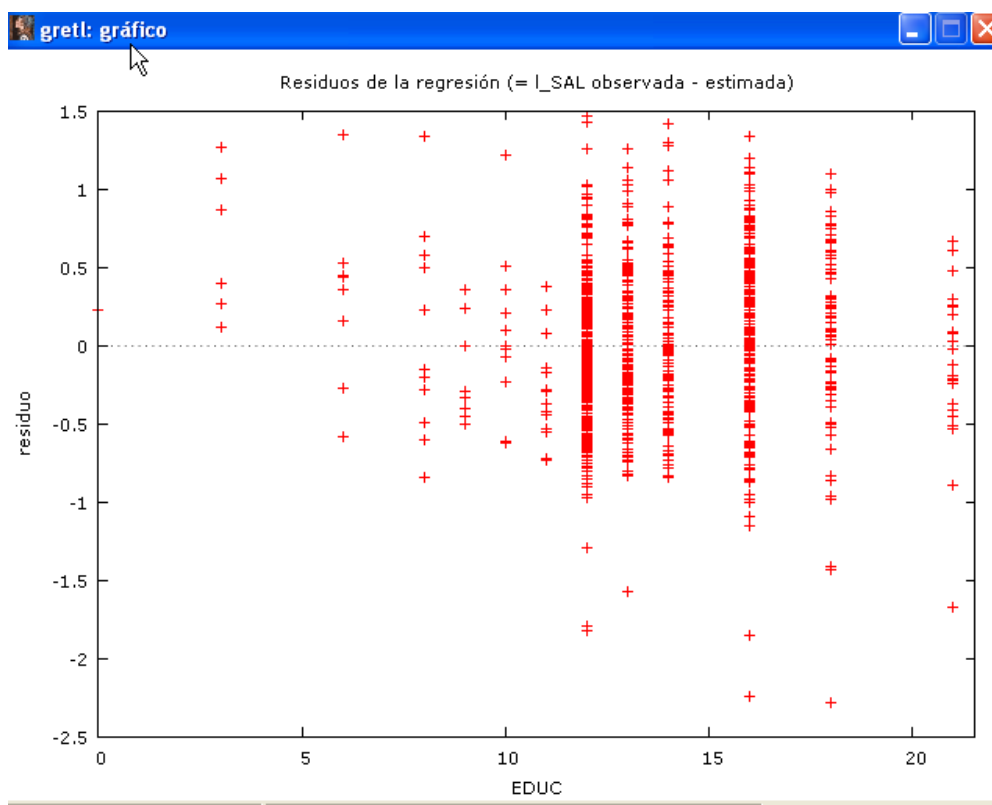
$$SR1 = 165.07; \hat{\sigma}_1^2 = 165.07 / (581 - 5) = 0.2866$$

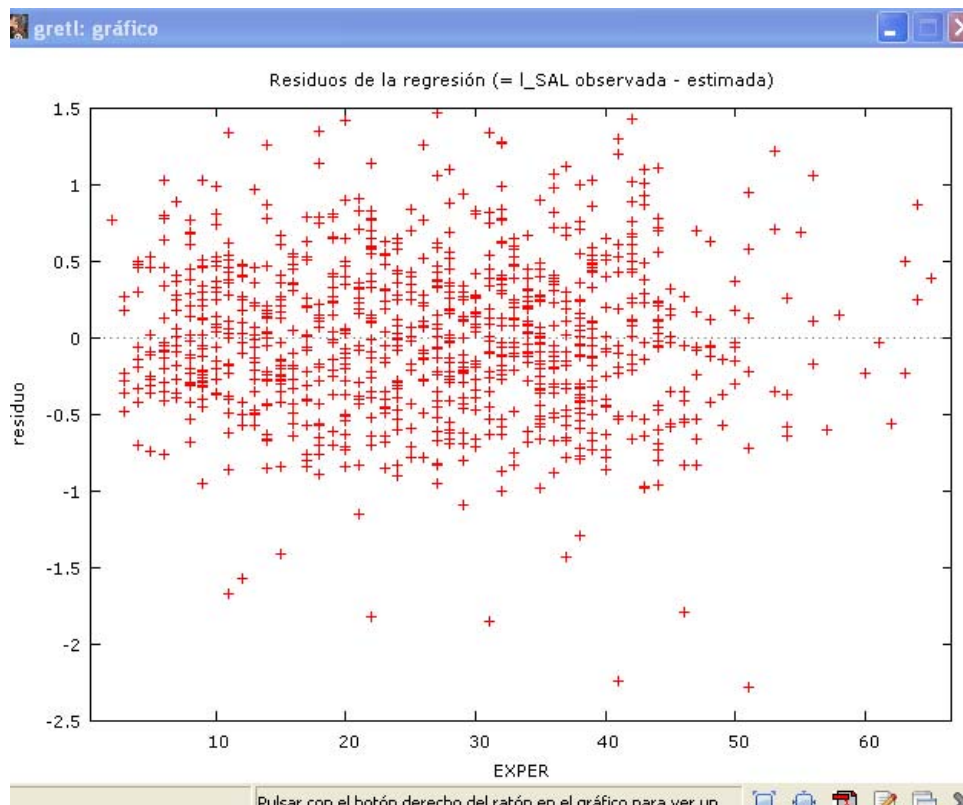
$$SR2 = 88.12; \hat{\sigma}_2^2 = 88.12 / (419 - 5) = 0.2128$$

$GQ = 0.2866/0.2128 = 1.3468 > 1.16$  Se rechaza la hipótesis nula de Homocedasticidad

$F_{(581-5,419-5)}(0.05) = 1.16$

24)





Parece que hay una mayor dispersión de los residuos cuando los años de educación aumentan.

**25)** Estimar el modelo, guardar los residuos y estimar la regresión auxiliar siguiente.

Recordar que la endógena son los residuos al cuadrado.

**gretl: modelo 17**

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

Modelo 17: MCO, usando las observaciones 1-1000  
Variable dependiente: sq\_uhat13

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	0.0675534	0.0801195	0.8432	0.3993
EDUC	0.00546997	0.00510853	1.071	0.2845
EXPER	0.00295293	0.00109105	2.707	0.0069 ***
MARRIED	0.0569910	0.0282794	2.015	0.0441 **
Media de la vble. dep.	0.254422	D.T. de la vble. dep.	0.434086	
Suma de cuad. residuos	185.4850	D.T. de la regresión	0.431544	
R-cuadrado	0.014645	R-cuadrado corregido	0.011677	
F(3, 996)	4.934491	Valor p (de F)	0.002090	
Log-verosimilitud	-576.5478	Criterio de Akaike	1161.096	
Criterio de Schwarz	1180.727	Crit. de Hannan-Quinn	1168.557	

Sin considerar la constante, el valor p más alto fue el de la variable 2 (EDUC)

$BP = T \cdot R^2 = 1000 \cdot 0.0146 = 14.6 > \chi^2_3 = 7.81$  Se rechaza la hipótesis de homocedasticidad.