

**Taller #8**  
**Multicolinealidad Y Heteroscedasticidad**  
**Respuestas Sugeridas**  
**Econometría 06216**

**Julio César Alonso**

**Notas:**

- o Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- o Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 31 de agosto.

**INSTRUCCIONES:**

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

Un estudio en el sector de la industria textil, confección, cuero y calzado relaciona los gastos de personal (Gpersonal), en millones de dólares, con las siguientes variables explicativas:

- Ingresos de explotación (I), en millones de dólares.
- Cifra de negocios (Neg), en millones de dólares.
- Número de personas ocupadas (Ocup).

1. Responda alas siguientes preguntas empleando la información contenida en el archivo "T8-02-04.xls".
  - a) Escriba el modelo a estimar.
  - b) Estime por MCO el anterior modelo y reporte los resultados en una tabla.
  - c) Determine la significancia individual y conjunta de los coeficientes estimados.
2. Determine si existe o no multicolinealidad (muestre todo su trabajo y conclusiones).
3. Empleando la misma información, responda:
  - a) Corrija el problema de Multicolinealidad. Explique claramente su decisión y reporte el modelo estimado que considere adecuado. Compruebe que el problema ha sido eliminado.
  - b) Interprete los coeficientes estimados y discuta la significancia individual y conjunta.

Un empresario Colombiano dedicado a la exportación de chips de platanitos está interesado en analizar el consumo privado de los distintos países en función de su renta nacional a precios de mercado. Para tal fin, lo contrata a usted para que le ayude a hacer las estimaciones econométricas. El empresario, dispone de datos del año 1989 referentes a dos grupos de países, uno formado por once países europeos y otro compuesto por países del resto del mundo, datos en ambos casos medidos en unidades monetarias nacionales. Puesto que dispone de los tipos de cambio de cada una de las monedas respecto al dólar, le pide a usted que trabaje con las variables expresadas en dólares.

El modelo es el siguiente:

$$CP_i = \alpha + \beta RN_i + \varepsilon_i$$

donde:

CP: Consumo privado en miles de millones de dólares.

RN: Renta nacional a precios de mercado en miles de millones de dólares.

4. Responda alas siguientes preguntas empleando la información contenida en el archivo "T8-02-04.xls".
  - a) Estime el modelo
  - b) Grafique los residuos versus los valores estimados de la renta nacional en miles de millones de dólares. Comente.
5. Continuando con la pregunta anterior, responda
  - a) Teniendo en cuenta que los datos de que se dispone han sido agrupados, estudiar la presencia de heteroscedasticidad utilizando el contraste de Golfeld-Quandt.
  - b) Realice la prueba de Heteroscedasticidad de Breush-Pagan. (escriba claramente la hipótesis nula, alterna, el estadístico, el valor crítico y la decisión)
  - c) Realice la prueba de Heteroscedasticidad de White. (escriba claramente la hipótesis nula, alterna, el estadístico, el valor crítico y la decisión).
6. Continuando con la pregunta anterior, responda
  - a) Escriba un modelo que no tenga problemas de heteroscedasticidad. (demuestre que teóricamente, el modelo ya no poseería heteroscedasticidad).
  - b) Estime el modelo escrito en el numeral a) de esta pregunta. (reporte sus resultados en una tabla)
  - c) Interprete los coeficientes estimados.
  - d) Comente la significancia global e individual de los coeficientes estimados.

**Taller #8**  
**Multicolinealidad Y Heteroscedasticidad**  
**Respuestas Sugeridas**  
**Econometría 06216**

**Julio César Alonso**

**Notas:**

- o Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- o Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 31 de agosto.

**INSTRUCCIONES:**

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

Un estudio en el sector de la industria textil, confección, cuero y calzado relaciona los gastos de personal (Gpersonal), en millones de dólares, con las siguientes variables explicativas:

- Ingresos de explotación (I), en millones de dólares.
  - Cifra de negocios (Neg), en millones de dólares.
  - Número de personas ocupadas (Ocup).
1. Responda alas siguientes preguntas empleando la información contenida en el archivo "T8-02-04.xls".
    - a) Escriba el modelo a estimar.

El modelo a estimar es el que incluye como variable endógena los gastos de personal y como regresores, además de la constante, el resto de las variables indicadas. Es decir,

$$Gpersonal_i = \beta_0 + \beta_1 I_i + \beta_2 Neg_i + \beta_3 Ocup_i + \varepsilon_i \quad (1)$$

- b) Estime por MCO el anterior modelo y reporte los resultados en una tabla.

Los resultados se presentan en la Tabla 1

**Tabla 1. Resultados de la estimación de diferentes Modelos.**

	VARIABLE DEPENDIENTE: $EXP_i$		
	Estadísticos t entre paréntesis		
	Ecuación 1	Ecuación 2	Ecuación 3
	MCO	MCO	MCP
constante	-1,360.0456 (-0.45)	-1,523.9177 (-0.51)	-3,491.3225 (-1.31)
$I_i$	1.9643 (0.68)	0.0901 (1.35)	--
$Neg_i$	-1.8482 (-0.65)	--	--
$Ocup_i$	0.8924 (0.92)	1.2993 (1.79) *	2.2762 (26.35) ***
$R^2$	0.98189	0.98130	0.97886
F	235 ***	367,37***	694,56***
# de Obs.	17	17	17

(\*) nivel de significancia: 10%

(\*\*) nivel de significancia: 5%

(\*\*\*) nivel de significancia: 1%

MCO: Mínimos Cuadrados Ordinarios

MCO: Mínimos Cuadrados Ordinarios

NA: No Aplica

- c) Determine la significancia individual y conjunta de los coeficientes estimados.

De acuerdo a la estimación del modelo, ninguno de los coeficientes es significativo.

Como  $F_{calculado} = 235 > F_{(0,01),(3,14)} = 5.74$ , todos los coeficientes asociados a pendientes son significativos conjuntamente a un nivel de significancia del 1%.

- d) ¿Qué tan bueno es el fit del modelo?

El modelo tiene un  $R^2$  alto, por lo que el modelo aparenta ser muy bueno. Sin embargo, todos lo síntomas de multicolinealidad están presentes (F global alto, t calculados bajos), por lo que este  $R^2$  alto también puede ser debido a la presencia de este problema econométrico.

2. Determine si existe o no multicolinealidad (muestre todo su trabajo y conclusiones).

Como se mencionó en el punto anterior, todos los síntomas de multicolinealidad están presentes. Por lo tanto, se emplearán las pruebas para corroborar la presencia de este problema econométrico.

Matriz de correlación de las X's

El determinante de las matriz de correlaciones es:

$$|R| = 2.990877 \cdot 0.0091192 \cdot 0.0000038 = 1.036E - 7$$

Como este determinante es muy cercano a cero, el problema de multicolinealidad es muy serio.

Medida de Multicoñealidad de Belsley, Kuck y Welsch (1980)

Se tiene que:

$$k(X) = \frac{\sqrt{\lambda_{Max}}}{\sqrt{\lambda_{Min}}} = \frac{\sqrt{2.990877}}{\sqrt{0.0000038}} = 887.17$$

Como  $k(X) > 30$ , se tiene un problema de multicolinealidad serio.

Matriz de correlaciones de los coeficientes estimados

En este caso se tiene la siguiente matriz de correlaciones de lo valores estimados:

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_0$
$\hat{\beta}_1$	1			
$\hat{\beta}_2$	-0,999719677	1		
$\hat{\beta}_3$	-0,660978691	0,643150628	1	
$\hat{\beta}_0$	0,093890196	-0,082378117	-0,465336115	1

Existe un elevada correlación entre  $\hat{\beta}_2$  y  $\hat{\beta}_1$ .

Conclusión: Todas las anteriores pruebas brindan evidencia que existe un problema de multicolinealidad en el modelo.

3. Empleando la misma información, responda:

- a) Corrija el problema de Multicolinealidad. Explique claramente su decisión y reporte el modelo estimado que considere adecuado. Compruebe que el problema ha sido eliminado.

En primer lugar, se suprime la variable  $Neg_i$  pues está fuertemente correlacionada con  $I_i$ , y es la que menos información aporta al modelo. Por lo tanto, el modelo a estimar queda:

$$Gpersonal_i = \beta_0 + \beta_1 I_i + \beta_3 Ocup_i + \varepsilon_i \quad (2)$$

Los resultados se presentan en la Tabla 1, ecuación 2.

Según estos, la variable  $I_i$  no es significativa, así que sería conveniente suprimirla del modelo. No obstante, antes de tomar esta decisión en base al t calculado se debe determinar si en este modelo (2) existe multicolinealidad, porque de ser así, este t calculado estaría sesgado.

Este modelo un  $R^2$  y F global alto, y algunos de los t calculados son bajos. Por lo tanto, presenta síntomas de multicolinealidad. Por consiguiente, se empleará de nuevo las pruebas para determinar si existe o no este problema econométrico.

Matriz de correlación de las X's

El determinante de las matriz de correlaciones es:

$$|R| = 1.9932973 \cdot 0.0067027 = 0.0134$$

Como este determinante es cercano a cero, existe un problema de multicolinealidad.

Medida de Multicoñealidad de Belsley, Kuck y Welsch (1980)

Se tiene que:

$$k(X) = \frac{\sqrt{\lambda_{Max}}}{\sqrt{\lambda_{Min}}} = \frac{\sqrt{1.9932973}}{\sqrt{0.0067027}} = 17.2449$$

Como  $10 < k(X) < 30$ , se tiene un problema de multicolinealidad entre moderado y fuerte.

Matriz de correlaciones de los coeficientes estimados

En este caso se tiene la siguiente matriz de correlaciones de lo valores estimados:

	$\hat{\beta}_1$	$\hat{\beta}_3$	$\hat{\beta}_0$
$\hat{\beta}_1$	1		
$\hat{\beta}_3$	-0,993319689	1	
$\hat{\beta}_0$	0,488876086	-0,540341445	1

Existe un elevada correlación entre  $\hat{\beta}_3$  y  $\hat{\beta}_1$ .

Conclusión: Todas las anteriores pruebas brindan evidencia que existe un problema de multicolinealidad en el modelo.

Así que se suprime la variable  $I_i$  del modelo (2) pues se considera que la variable Número de personas ocupadas, está relacionada más directamente con los Gastos del personal. El modelo a estimar queda así:

$$Gpersonal_i = \beta_0 + \beta_3 Ocup_i + \varepsilon_i \quad (2)$$

Así, queda solucionado el problema de multicolinealidad.

- b) Interprete los coeficientes estimados y discuta la significancia individual y conjunta.

$\hat{\beta}_0$ : -3491.3225 millones de dólares es el Gasto en personal (ahorro) que no depende del Número de personas ocupadas.

Este coeficiente no es significativamente diferente de cero.

$\hat{\beta}_3$ : 2.2762. Cuando aumenta en 1 el número de personas ocupadas, el gasto en personal aumenta en 2.2762 millones de dólares.

Este coeficiente es significativo a un nivel de significancia del 1%.

Un empresario Colombiano dedicado a la exportación de chips de platanitos está interesado en analizar el consumo privado de los distintos países en función de su renta nacional a precios de mercado. Para tal fin, lo contrata a usted para que le ayude a hacer las estimaciones econométricas. El empresario, dispone de datos del año 1989 referentes a dos grupos de países, uno formado por once países europeos y otro compuesto por países del resto del mundo, datos en ambos casos medidos en unidades monetarias nacionales. Puesto que dispone de los tipos de cambio de cada una de las monedas respecto al dólar, le pide a usted que trabaje con las variables expresadas en dólares.

El modelo que a estimar es el siguiente:

$$CP_i = \alpha + \beta RN_i + \varepsilon_i$$

donde:

CP: Consumo privado en miles de millones de dólares.

RN: Renta nacional a precios de mercado en miles de millones de dólares.

4. Responda a las siguientes preguntas empleando la información contenida en el archivo "T8-02-04.xls".

a) Estime el modelo

Los resultados se presentan en la Tabla 2.

**Tabla 2. Resultados de la estimación de diferentes Modelos.**

	VARIABLE DEPENDIENTE: $CP_i$	
	Estadísticos t entre paréntesis	
	Ecuación 1	Ecuación 2
	MCO	MCP
constante	0.7366 (73.22) ***	2.5657 (2.10) **
$RN_i$	-21.0424 (-1.77) *	0.6540 (37.48) ***
$R^2$	0.99628	0.18076
$R^2$ Ajustado	1.00	0.13979
F	5,361.20 ***	NA
# de Obs.	22	22

(\*) nivel de significancia: 10%

(\*\*) nivel de significancia: 5%

(\*\*\*) nivel de significancia: 1%

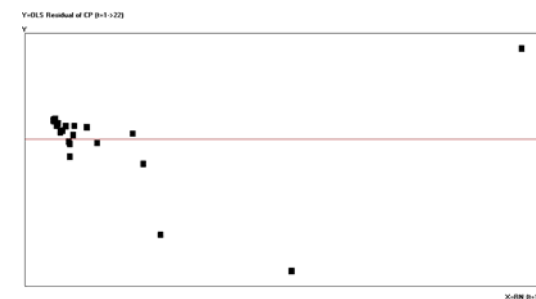
MCO: Mínimos Cuadrados Ordinarios

MCP: Mínimos Cuadrados Ponderados

NA: No Aplica

b) Grafique los residuos versus los valores estimados de la renta nacional en miles de millones de dólares. Comente.

**Gráfico 1. Errores estimados vs la variable Renta Nacional**



Se puede observar que la dispersión de los residuos aumenta a medida que la Renta Nacional es mayor. Por lo tanto, este gráfico lleva a intuir que existe un problema de heteroscedasticidad en la regresión.

5. Continuando con la pregunta anterior, responda

a) Teniendo en cuenta que los datos de que se dispone han sido agrupados, estudiar la presencia de heteroscedasticidad utilizando el contraste de Goldfeld-Quandt.

El contraste de Goldfeld y Quandt, adopta la siguiente hipótesis alternativa:

$$H_0 : \sigma_i^2 = \sigma^2$$

$$H_A : \sigma_i^2 = \sigma^2 RN_i^2$$

La aplicación de esta prueba implica tomar las siguientes decisiones:

- 1) Elegir la variable, o una combinación de variables, para la ordenación de las observaciones. En este caso, para la ordenación se ha utilizado la única variable del modelo **Renta Nacional**.
- 2) Indicar las observaciones que forman las muestras utilizadas para calcular el denominador (grupo 1) y en numerador (grupo 2) del estadístico  $GQ$ . En este caso, no se omiten variables centrales pues se conservarán los dos grupos extremos ya ordenados. Los países que conforman el grupo 1 son los once países europeos, y los países del resto del mundo corresponden al grupo 2.

El estadístico  $GQ$  obtenido es el siguiente

$$GQ = \frac{SSE2}{SSE1} = \frac{33833.056261}{5881.191} = 5.753$$

El valor crítico corresponde a la distribución *F* de Snedecor con 18 grados de libertad en el numerador y 18 en el denominador; es decir 3.128. Así, se puede rechazar la hipótesis nula de homoscedasticidad a favor de la hipótesis alterna de una heteroscedasticidad provocada por la renta nacional.

- b) Realice la prueba de Heteroscedasticidad de Breush-Pagan. (escriba claramente la hipótesis nula, alterna, el estadístico, el valor crítico y la decisión)

Las hipótesis nula y alternativa vendrán dadas por:

$$H_0 : \sigma_i^2 = \sigma^2$$

$$H_A : f_c(RN_i)$$

En éste caso tenemos que  $SSE = \hat{\varepsilon}^T \hat{\varepsilon} = 48246.500056$  y por tanto  $\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n} = \frac{48246.500056}{22} = 2193.023$ . Así en este caso la regresión auxiliar será

$\frac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2} = \gamma + \delta(RN_i) + \mu_i$ . Para esta regresión auxiliar tenemos,  $SSR = SST - SSE = 124.854641 - 59.725665 = 65.128976$ . Así,

$BP = \frac{SSR}{2} = \frac{65.128976}{2} = 32.5645$ . Para tomar una decisión, este estadístico BP lo comparamos con  $\chi_{m-1, \alpha}^2 = \chi_{1, 0.01}^2 = 6.63$ , y entonces podemos rechazar la hipótesis nula a favor de la presencia de heteroscedasticidad.

- c) Realice la prueba de Heteroscedasticidad de White. (escriba claramente la hipótesis nula, alterna, el estadístico, el valor crítico y la decisión).

El contraste de White tiene un carácter muy general, ya que la hipótesis alternativa recoge esquemas de heteroscedasticidad asociados a todas las variables explicativas y a combinaciones de las mismas. Para la aplicación de este contraste hay que efectuar una regresión auxiliar, tomando como regresando a los residuos al cuadrado obtenidos al estimar el modelo original, y como regresores a los regresores del modelo original, a sus cuadrados y a los productos cruzados, después de haber eliminado las posibles redundancias. Como en este caso sólo hay una variable independiente, la regresión se hace utilizando como regresores a *RN* y un único *RN*<sup>2</sup>, para evitar problemas de multicolinealidad.

Así, esta prueba implica:

$$H_0 : \sigma_i^2 = \sigma^2$$

$$H_A : \text{Hetero}(\text{causado por las } X_s)$$

El estadístico de White corresponde a:  $W_a = 22 * 0.720880 = 15.8594$ ; y se debe comparar con el valor crítico de la distribución Chi-cuadrado con 2 grados de libertad. En este caso este valor crítico corresponde a 9.21 con nivel de significancia del 1%. Así, la decisión es rechazar la Hipótesis nula, en favor de la alterna.

6. Continuando con la pregunta anterior, responda  
a) Escriba un modelo que no tenga problemas de heteroscedasticidad. (demuestre que teóricamente, el modelo ya no poseería heteroscedasticidad).

Puesto que existe heteroscedasticidad ligada a la variable Renta Nacional, es conveniente con objeto de obtener estimadores más eficientes, aplicar el método de *MCP*. Para ello se requiere en primer lugar determinar cual es la función concreta que relaciona la varianza de las perturbaciones con la variable *RN*. Es decir, se trata de especificar la siguiente función:

$$\sigma_i^2 = f(RN_i)$$

De acuerdo a la prueba de Goldfeld y Quandt, podemos concluir que el esquema de heteroscedasticidad aplicable a nuestro modelo sería el siguiente:

$$\sigma_i^2 = \sigma^2 RN_i^2$$

Bajo el anterior esquema, la aplicación de *MCP* (es decir, la aplicación de *MCG* en el caso de heteroscedasticidad) implica transformar el modelo original de la siguiente forma:

$$\frac{CP_i}{RN_i} = \alpha \frac{1}{RN_i} + \beta \frac{RN_i}{RN_i} + \frac{\varepsilon_i}{RN_i}$$

$$\frac{CP_i}{RN_i} = \alpha \frac{1}{RN_i} + \beta + \frac{\varepsilon_i}{RN_i}$$

$$CP_i^* = \alpha RN_i^* + \beta + \varepsilon_i^* \quad (2)$$

y aplicar *MCO* al modelo transformado. Es muy fácil demostrar que  $Var\left(\frac{\varepsilon_i}{RN_i}\right) = \sigma^2$ .

- b) Estime el modelo escrito en el numeral a) de esta pregunta. (reporte sus resultados en una tabla)

Los resultados de la regresión por *MCP* aparecen en la tabla 1, ecuación 2. Se puede verificar que este modelo transformado ya no tiene problemas de heteroscedasticidad.

- c) Interprete los coeficientes estimados.

$\hat{\alpha} = 2.56579$ ; 2.56579 miles de dólares corresponden al consumo autónomo privado.

$\hat{\beta} = 0.65396$ ; 0.65396 es la propensión marginal a consumir. Cuando la Renta Nacional aumenta en un mil de millón de dólares, el consumo privado aumenta en 0.65396 miles de millones de dólares.

d) Comente la significancia global e individual de los coeficientes estimados.

El intercepto es significativo al 5% y el coeficiente asociado a la variables RN es significativos a un nivel de significancia del 1%.

Para la significancia conjunta, se emplea la prueba de Wald, pues el F global no tiene sentido. Sin embargo, como sólo hay una pendiente, es suficiente el test de significancia individual.