

## TEST DE HOMOCEDASTICIDAD PARA OBSERVACIONES PREVIAMENTE CLASIFICADAS

*Alfredo Mario Baronio*[alfredomariobaronio@yahoo.com.ar](mailto:alfredomariobaronio@yahoo.com.ar)

Facultad de Ciencias Económicas - Universidad Nacional de Río Cuarto  
Instituto Académico Pedagógico de Ciencias Sociales – Universidad Nacional de Villa María  
Doctor en Ciencias Económicas. Profesor Titular Responsable de Econometría en UNRC y UNVM. Docente de posgrado: Econometría en UNR y UNSL, Producción y tratamiento de datos en UNRC. Investigador Categoría 1. Director del Proyecto "Producción de datos y Econometría Aplicada" en UNVM. Coautor de Cuadernos de Econometría publicado en sitio web [www.econometricos.com.ar](http://www.econometricos.com.ar). Director del Doctorado en Desarrollo Territorial Facultad de Ciencias Económicas en UNRC.

*Ana María Vianco*[anavianco@yahoo.com.ar](mailto:anavianco@yahoo.com.ar)

Facultad de Ciencias Económicas - Universidad Nacional de Río Cuarto  
Licenciado en Economía. Investigador categorizado 3. Profesor Adjunto afectado a Inferencia Estadística, Econometría y Econometría Avanzada en UNRC. Coautor de Herramientas para la investigación Regional. Universitas Editorial Científica Universitaria de Córdoba. Coautor de Cuadernos de Econometría disponible en [www.econometricos.com.ar](http://www.econometricos.com.ar). Dirección y codirección de trabajos finales de grado, posgrado y becas internas de investigación. Director de Producción de datos y econometría aplicada en UNRC.

Palabras claves: Heterocedasticidad – Clasificación – Exploratorio

En la estimación de modelos econométricos, es habitual la presencia de heterocedasticidad; esta diferencia en las varianzas puede ser el resultado de la existencia de grupos de individuos que se comportan de manera significativamente diferente y no ha sido advertido al momento de especificar el modelo. En este contexto, la exploración de la información y la aplicación de la clasificación jerárquica y reunión de las observaciones en clases mutuamente excluyentes son herramientas que posibilitan identificar esas diferencias. El objetivo de este trabajo es proponer la construcción de un test de homocedasticidad a partir de la información que brinda cada grupo, con la finalidad de contrastar la similitud de las varianzas y el número de clases existentes en la tabla de datos bajo estudio. La réplica de este test para todos los pares de grupos existentes en el estudio, además de informar la existencia eventual de heterocedasticidad, sugiere un camino para corregir la especificación del modelo.

## Introducción

Uno de los supuestos de los modelos econométricos es que  $E(\varepsilon\varepsilon') = \sigma^2\mathbf{I}$ . Conocida como la hipótesis de perturbaciones esféricas, expresa el doble supuesto de que la varianza de la perturbación es constante para cada unidad de observación (homocedasticidad) y que la covarianza de las perturbaciones para todos los pares de individuos es nula (no autocorrelación).

En la teoría econométrica, las pruebas existentes para verificar la presencia de homocedasticidad difieren en la formulación de la hipótesis alternativa, la que responde al supuesto particular de comportamiento de los errores; al tener que aplicar uno o varios de estos test, en general, no se menciona cuál de todos es más potente a la hora de ser utilizado. La detección del problema de heterocedasticidad conduce a estimar el modelo aplicando mínimos cuadrados generalizados o buscando una forma funcional para introducirla en la estimación.

Los test desarrollados hasta hoy no contemplan las diferencias naturales que puedan existir entre las observaciones y que inevitablemente conducen a errores heterocedásticos. Quizás, uno de los ejemplos más claro de esto se encuentre en el ámbito empresarial, donde las diferencias en los niveles de ventas entre empresas generan errores heterocedásticos provenientes de clusters diferenciados. La no incorporación de estas diferencias en el modelo da lugar al error de especificación, que es una de las causas de residuos heterocedásticos.

El objetivo de este trabajo es presentar un test para el contraste de homocedasticidad. A partir de la clasificación de los individuos en clases mutuamente excluyentes es posible identificar la estructura de heterocedasticidad, incorporar esa información en el modelo y obtener estimaciones homocedásticas considerando la aplicación de variables ficticias.

En lo que sigue, se describen los principales test utilizados en la actualidad para contrastar la hipótesis de homocedasticidad, se muestra un procedimiento para identificar heterocedasticidad y se propone un test para verificar el supuesto de homocedasticidad.

## Antecedentes teóricos

Si la varianza del término de perturbación del modelo de regresión lineal no es constante para todas las observaciones se dice que es heterocedástica. En numerosas aplicaciones económicas, la heterocedasticidad puede originarse en: especificación errónea del modelo, cambio estructural y alta dispersión, absoluta y relativa, a medida que crece el tamaño de la muestra. La consecuencia de esto es que la matriz de varianzas y covarianzas de las perturbaciones no es escalar. Bajo el supuesto de que no existe autocorrelación en las perturbaciones, la heterocedasticidad implica la siguiente estructura de la matriz de varianzas y covarianzas:

$$E(\varepsilon\varepsilon') = \Omega = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

En la práctica, habitualmente a priori no se sabe si hay o no problemas de heterocedasticidad en las perturbaciones, por esto es que la estimación se realiza bajo el supuesto de

homocedasticidad; luego, para confirmar el cumplimiento del supuesto, se realizan los contrastes.

Se ha desarrollado un gran número de métodos para contrastar la hipótesis nula de igualdad de varianzas u homocedasticidad. Esta variedad se debe a que la especificación de la hipótesis alternativa de heterocedasticidad suele no ser conocida, puede ser más o menos general y reflejar diferentes comportamientos de las perturbaciones. Entre los contrastes más utilizados en la literatura se encuentran el de Goldfeld y Quandt (1965), Breusch y Pagan (1979) y White (1980).

Goldfeld y Quandt ordenan las observaciones de menor a mayor en función de alguna variable que puede o no ser un regresor del modelo especificado, eliminan un número determinado de observaciones centrales y dividen en dos grupos para efectuar regresiones por separado y evaluar la constancia de la variabilidad en ambos grupos.

Explícitamente, se tienen sospechas de que las varianzas ( $\sigma_i^2$ ) mantienen una relación monótona con los valores de alguna variable  $\mathbf{Z}$ . Se supone que la hipótesis alternativa es  $\sigma_1^2 = \sigma_2^2 G(\mathbf{Z}_i, \gamma)$ , donde  $G(\cdot)$  es una función monótona creciente en  $\mathbf{Z}_i$  que puede ser uno de los regresores incluidos en el modelo. Para realizar el contraste, se ordenan las observaciones de la tabla de datos siguiendo el orden creciente de la variable  $\mathbf{Z}_i$ ; se eliminan  $p$  observaciones centrales dando lugar a dos bloques con  $(T - p)/2$  observaciones, reuniendo  $T_1$  y  $T_2$  observaciones respectivamente. Las observaciones centrales que se eliminan permiten mayor independencia entre los dos grupos; en cada uno de ellos se estima el modelo de regresión -el primer grupo concentra las observaciones con menor valor nominal de  $\mathbf{Z}$  y el segundo grupo las de mayor valor nominal- y se obtienen los errores de estimación. El estadístico de contraste se construye bajo la hipótesis nula de homocedasticidad  $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_T^2$  y suponiendo que la perturbación se distribuye normal de media cero y no correlacionada serialmente, el estadístico GQ sigue una distribución  $F$  de Snedecor:

$$GQ = \frac{e_2' e_2 / T_2 - k}{e_1' e_1 / T_1 - k} \sim F(T_1 - k, T_2 - k)$$

donde,  $e_2' e_2$  es la suma de cuadrados de residuos de la regresión de  $Y$  sobre  $X$  en el segundo grupo de observaciones, y  $e_1' e_1$  es la suma de cuadrados de residuos de la regresión  $Y$  sobre  $X$  utilizando el primer grupo de observaciones. Mientras que, bajo la hipótesis nula, las varianzas deben ser iguales; bajo la hipótesis alternativa, crecerán de un grupo a otro. Cuanto más difieran estas sumas de cuadrados, mayor será el valor del estadístico y, por lo tanto, mayor evidencia habrá en contra de la hipótesis nula. Se rechaza la  $H_0$ , a un nivel de significación  $\alpha$ , si  $GQ > F_{\alpha}(T_1 - k, T_2 - k)$ .

Algunos autores proponen una forma funcional para la hipótesis alternativa como es el caso de Breusch-Pagan quienes derivan un contraste de heterocedasticidad donde la hipótesis alternativa es  $H_A: \sigma_i^2 = \sigma^2 G(\alpha_0 + \gamma' \mathbf{Z}_i)$ , siendo la función  $G(\cdot)$  no especificada y  $\mathbf{Z}_i$  un vector de variables exógenas que pueden ser las explicativas del modelo. La hipótesis nula del contraste es la de homocedasticidad que, dada la alternativa, implica contrastar  $\gamma = 0$ .

Una forma operativa de realizar el contraste, es estimar el modelo  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + \varepsilon_i$  para obtener los errores; utilizar los residuos  $\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{MCO}}$  para construir la serie  $r_i = e_i^2 / \mathbf{e}' \mathbf{e} \quad \forall i = 1 \dots T$ ; especificar y estimar el modelo  $r_i = \alpha_0 + \alpha' \mathbf{Z}_i + \omega_i \quad \forall i = 1 \dots T$  de donde se obtiene la suma de cuadrados explicada ( $SCE$ ). El estadístico de contraste es  $SCE/2$ , que bajo hipótesis nula de homocedasticidad se distribuye asintóticamente  $\chi^2(S)$ , donde  $S$  son los grados de libertad igual al número de variables en  $\mathbf{Z}_i$ . Se rechaza la hipótesis nula a un nivel de significación ( $\alpha$ ), si el valor muestral del estadístico excede el cuantil  $\chi_{\alpha}^2(S)$ .

White ha derivado una forma funcional para la estimación de la varianza, de los estimadores del modelo mínimo cuadrático ordinario, que es aplicable aún en el caso de perturbaciones heterocedásticas. Define  $V_{\text{WHITE}}(\hat{\beta}) = (X'X)^{-1}(X'SX)(X'X)^{-1}$ ; donde,  $S$  es una matriz diagonal cuyos elementos son los residuos mínimo-cuadráticos ordinarios al cuadrado [ $S = \text{diag}(e_1^2, e_2^2, \dots, e_T^2)$ ]. El estimador  $V_{\text{WHITE}}(\hat{\beta})$  es consistente siempre que la matriz sea diagonal; es decir, en aquellas situaciones en las que el modelo tenga perturbaciones heterocedásticas pero no autocorrelacionadas. Bajo la hipótesis nula de homocedasticidad, tanto el estimado de White como el de mínimos cuadrados ( $\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$ ) son consistentes; mientras que, bajo la alternativa de heterocedasticidad, el estimador  $\hat{V}(\hat{\beta})$  no lo es.

La forma operativa de realizar el contraste consiste en estimar el modelo bajo estudio  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + \varepsilon_i$  para obtener los errores  $e_i = Y_i - \hat{Y}_i$ ; White supone que el cuadrado de los errores estimados se explica por las variables independientes del modelo, el cuadrado de ellas y su producto cruzado:  $e_i^2 = \alpha_0 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + \varphi_2 X_{2i}^2 + \varphi_3 X_{3i}^2 + \dots + \varphi_k X_{ki}^2 + \delta_{23} X_{2i} X_{3i} + \delta_{24} X_{2i} X_{4i} + \dots + \delta_{ki} X_{ki} X_{li} \quad \forall i = 1, 2 \dots T$ , de la estimación de este modelo se obtiene el coeficiente  $R^2$  con el que se construye el estadístico  $\lambda = TR^2$ . La hipótesis nula de homocedasticidad contrasta que todos los coeficientes de la regresión de los errores, exceptuando el intercepto, son conjuntamente cero, es decir:  $H_0: \beta_j, \varphi_j, \delta_{js} = 0 \quad \forall j, s; j < s; j = 2, 3, \dots, k; s = 3, 4, \dots, k$ . Bajo la hipótesis nula  $\lambda \sim \chi^2(p)$ , donde  $p$  es el número de regresores en el modelo especificado sin incluir el término constante, se rechaza la hipótesis nula si el valor muestral del estadístico excede el valor crítico de  $\chi^2$ , elegido un nivel de significación.

### Clasificación de las unidades de observación en clases mutuamente excluyentes

Siguiendo a Lebart, L. *et al* (1995), aplicar un método de clasificación a un conjunto dado de unidades de observación significa definir las "clases" entre las cuales se distribuyen los elementos del conjunto. Es así que, los métodos de clasificación fraccionan el conjunto de unidades de observación en subconjuntos homogéneos a partir de la medición de índices de disimilaridad.

La teoría estadística ofrece una abundante referencia bibliográfica para llevar a cabo la clasificación de los individuos observados. Entre ellos se encuentra Ward J.H. (1963), quien propone una clasificación jerárquica de los individuos a partir de la utilización de distancias cuadráticas con el criterio de minimizar la varianza dentro del grupo.

El método de Ward es una generalización multidimensional del modelo clásico de análisis de varianza. Su utilidad radica en que, el procedimiento, minimiza el crecimiento de la varianza intra-grupos resultante de la agregación de dos grupos en una nueva clase. Este criterio genera grupos homogéneos a su interior y hace máxima la distancia entre ellos.

A partir de una tabla de datos de tamaño  $(n \times p)$ ,  $n$  individuos sobre los que se observan  $p$  características, están representados en el espacio  $p$ -dimensional. Se denominan "clases" a los subconjuntos de individuos de ese espacio de representación, tales que agrupen individuos semejantes. Un algoritmo de clasificación depende de la selección de una distancia adecuada, para evaluar la semejanza entre los elementos y grupos de elementos a comparar. En la dispersión de la nube de puntos-individuos se pueden encontrar zonas con alta densidad y zonas de baja densidad.

La tabla (nxp) tiene más de una variable ( $p > 1$ ); el procedimiento de clasificación consiste en calcular las distancias euclidianas de los individuos, con respecto al centro de gravedad de la nube de puntos de la clase y, desde aquí, al centro de gravedad de la nube de puntos. En una primera etapa, se tiene tantas clases como individuos, los que se reúnen en grupos de acuerdo a su grado de similitud. En una segunda etapa, los grupos se agregan con el mismo criterio que se agregaron los individuos. Este procedimiento se repite hasta que se consigue minimizar la varianza intra-grupos y maximizar la varianza inter-grupos.

### Construcción del Test de Homocedasticidad

Para construir el test que permita evaluar la existencia de errores homocedásticos, en primer lugar, se especifica y estima el modelo

$$X_{1i} = \beta_1 + \sum_{j=2}^p \beta_j X_{ji} + \varepsilon_j; \quad \forall i = 1, \dots, n \quad j = 2, \dots, p$$

Luego, se utiliza el criterio de clasificar el conjunto de  $n$  observaciones de las  $p$  variables en  $K$  grupos, utilizando el método de clasificación jerárquica de Ward. El procedimiento de clasificación genera clases mutuamente excluyentes que no implica, necesariamente, que se esté en presencia de heterocedasticidad. La estimación del modelo en cada una de las clases provee información de la varianza a partir de conocer la suma de los errores cuadráticos estimados intraclass. Con las varianzas estimadas en las  $K$  clases se obtiene el promedio:

$$\bar{S}^2 = \frac{(n_1 - p)S_1^2 + (n_2 - p)S_2^2 + \dots + (n_K - p)S_K^2}{n_1 + n_2 + \dots + n_K - pK}$$

donde  $S_k^2 = \frac{e_k' e_k}{n_k - p}$   $\forall k = 1, \dots, K$ , es la varianza de la perturbación aleatoria para el  $k$ -ésimo grupo de observaciones;  $p$  es la cantidad de parámetros estimados y  $n_k$   $\forall k = 1, \dots, K$  es el tamaño de cada grupo.  $\bar{S}^2$  es una estimación centrada de  $\sigma^2$ ; si los errores de estimación en cada clase tienen distribuciones normales, la distribución muestral del estadístico de  $\bar{S}^2/\sigma^2$  es una  $\chi^2/n_1 + n_2 + \dots + n_K - pK$  grados de libertad. El intervalo de confianza para la estimación de  $\sigma^2$  es

$$P\left(\frac{(n_1 + n_2 + \dots + n_K - pK) \bar{S}^2}{\chi_{(1-\alpha/2), (n_1 + n_2 + \dots + n_K - pK)}^2} < \sigma^2 < \frac{(n_1 + n_2 + \dots + n_K - pK) \bar{S}^2}{\chi_{(\alpha/2), (n_1 + n_2 + \dots + n_K - pK)}^2}\right) = 1 - \alpha$$

Bajo hipótesis nula de varianza constante  $H_0: \sigma^2 = \sigma_k^2 \quad \forall k = 1, 2, \dots, K$ ; el estadístico es

$$\frac{\chi_{(n_1 + n_2 + \dots + n_K - pK)}^2}{(n_1 + n_2 + \dots + n_K - pK)} = \frac{\bar{S}^2}{\sigma^2}$$

Como se desconoce el valor de  $\sigma^2$ , para realizar el contraste se utiliza el valor de la varianza estimada en cada grupo  $S_k^2$ ; la dódima consiste en comparar el valor de la varianza intragrupo con el intervalo de confianza para  $\sigma^2$ . Se está en presencia de homocedasticidad cuando las varianzas de cada grupo se encuentren dentro del intervalo; en caso de que el valor de la varianza para un determinado grupo esté fuera del intervalo, se considera que ese grupo produce heterocedasticidad en el modelo.

Por lo tanto, la regla de decisión es:

- Si para todo  $k = 1, \dots, K$ ,  $\frac{(n_1 + n_2 + \dots + n_K - pK) \bar{S}^2}{\chi_{(1-\alpha/2), (n_1 + n_2 + \dots + n_K - pK)}^2} \leq S_k^2 \leq \frac{(n_1 + n_2 + \dots + n_K - pK) \bar{S}^2}{\chi_{(\alpha/2), (n_1 + n_2 + \dots + n_K - pK)}^2}$  se acepta la hipótesis nula de homocedasticidad

- Si para algún  $k = 1, \dots, K$ ,  $S_k^2 < \frac{(n_1+n_2+\dots+n_K-pK) \bar{S}^2}{\chi^2_{(1-\alpha/2), (n_1+n_2+\dots+n_K-pK)}}$  o  $S_k^2 > \frac{(n_1+n_2+\dots+n_K-pK) \bar{S}^2}{\chi^2_{(1-\alpha/2), (n_1+n_2+\dots+n_K-pK)}}$  se rechaza la hipótesis nula de homocedasticidad

El resultado de la prueba proporciona información que permite reorganizar a las unidades de observación; si solo se aceptan  $G < K$  grupos homocedásticos, se puede proceder a reagrupar las observaciones que dan lugar a errores homocedásticos en un solo grupo. Los grupos identificados como heterocedásticos permanecen aislados. La identificación de los grupos permitiría incorporar variables ficticias al modelo, salvando el error de especificación y posibilitando estimaciones homocedásticas.

### Aplicación

Se ha simulado una muestra aleatoria de tamaño  $n = 100$  para 3 variables, en la cual una de ellas es función lineal de las otras dos. Se especifica el modelo

$$X_{1i} = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i; \quad \forall i = 1, \dots, 100$$

Se clasificaron las 100 muestras en  $K = 3$  grupos; en cada grupo, se especifica y estima el modelo inicial produciendo las siguientes varianzas de los errores

$$S_1^2 = \frac{e_1' e_1}{n_1 - p} = \frac{85579,36}{32 - 3} = 2951,01$$

$$S_2^2 = \frac{e_2' e_2}{n_2 - p} = \frac{75897,51}{35 - 3} = 2371,80$$

$$S_3^2 = \frac{e_3' e_3}{n_3 - p} = \frac{76981,73}{33 - 3} = 2566,06$$

La estimación centrada puntual de la varianza es

$$\bar{S}^2 = \frac{(n_1 - p)S_1^2 + (n_2 - p)S_2^2 + \dots + (n_K - p)S_K^2}{n_1 + n_2 + \dots + n_K - pK}$$

$$\bar{S}^2 = \frac{(29)S_1^2 + (32)S_2^2 + (30)S_3^2}{32 + 35 + 33 - 9} = \frac{85579,29 + 75879,6 + 76981,8}{91} \cong 2620,23$$

El intervalo de confianza para la estimación de  $\sigma^2$  se construye a partir de la distribución chi cuadrado considerando 91 grados de libertad

$$P\left(\frac{(n_1 + n_2 + \dots + n_K - pK) \bar{S}^2}{\chi^2_{(1-\alpha/2), (n_1+n_2+\dots+n_K-pK)}} < \sigma^2 < \frac{(n_1 + n_2 + \dots + n_K - pK) \bar{S}^2}{\chi^2_{(\alpha/2), (n_1+n_2+\dots+n_K-pK)}}\right) = 1 - \alpha$$

$$P\left(\frac{91 * 2620,23}{119,28} < \sigma^2 < \frac{91 * 2620,23}{66,50}\right) = 0,95$$

$$P(1998,97 < \sigma^2 < 3585,54) = 0,95$$

Los límites de confianza para  $\sigma^2$ , al 0.95 de la distribución chi cuadrado, son 1998.97 y 3585.54. Bajo hipótesis nula de homocedasticidad, *el intervalo contiene las varianzas de cada grupo*, por lo que se acepta la hipótesis nula.

Para un conjunto de  $n = 156$  observaciones de 4 variables; se especifica y estima el modelo

$$X_{1i} = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i; \quad \forall i = 1, \dots, 156$$

Se clasifica la información dando por resultado la existencia de  $K = 3$  grupos; en cada grupo, se especifica y estima el modelo que proveen las varianzas:

$$S_1^2 = \frac{e_1' e_1}{n_1 - p} = \frac{544.10}{33 - 4} = 18.76$$

$$S_2^2 = \frac{e_2' e_2}{n_2 - p} = \frac{396.43}{66 - 4} = 6.39$$

$$S_3^2 = \frac{e_3' e_3}{n_3 - p} = \frac{119.55}{57 - 4} = 2.24$$

La estimación centrada puntual de la varianza es

$$\begin{aligned} \overline{S^2} &= \frac{(n_1 - p)S_1^2 + (n_2 - p)S_2^2 + (n_3 - p)S_3^2}{n_1 + n_2 + n_3 - pK} \\ \overline{S^2} &= \frac{(29)S_1^2 + (62)S_2^2 + (53)S_3^2}{33 + 66 + 57 - 12} = \frac{544.10 + 396.43 + 119.55}{144} \cong 7.36 \end{aligned}$$

El intervalo de confianza para la estimación de  $\sigma^2$  se construye a partir de la distribución chi cuadrado considerando 144 grados de libertad

$$P(5.91 < \sigma^2 < 9.40) = 0.95$$

Los límites de confianza para  $\sigma^2$ , al 0.95 de la distribución chi cuadrado, son 5.91 y 9.40. Bajo hipótesis nula de homocedasticidad, *el intervalo contiene la varianza del segundo grupo y no las correspondientes al grupo 1 y 3*, por lo que se rechaza la hipótesis nula. La información analizada proporciona errores heterocedásticos, cada uno de los grupos corresponde a una categoría o estrato en la población que debe ser individualizado en la especificación del modelo para no cometer errores.

### Conclusiones

El estudio se inspira inicialmente en el contraste de Goldfeld y Quandt y se complementa con las técnicas de clasificación de la información y partición. El test desarrollado permite construir grupos homogéneos con la utilización de todas las variables que intervienen en el modelo. En caso de que exista heterocedasticidad, informa convenientemente el procedimiento a seguir utilizando otras herramientas econométricas.

La prueba propuesta conserva toda la información, no transforma las variables y no utiliza resultados asintóticos. La limitante que puede suceder, es que haya que reagrupar clases cuando las observaciones clasificadas en una de ellas son inferiores o iguales a la cantidad de parámetros a estimar. La solución a este problema es usar procedimientos de partición no automáticos; es decir, reagrupar los miembros de dos o más grupos, tantas veces como sea necesario, hasta superar aquella limitante.

### Bibliografía

- Berndt, E. (1996). *The Practice of Econometrics*. Addison Wesley.
- Dixon, W. J. y Massey, F.J. (1957). *Introduction to Statistical Analysis*. Mc.Graw Hill.
- Green, W. (1999). *Econometrics Analysis*. Prentice Hall.
- Gujarati, D. (2004). *Econometría*. Mc.Graw Hill.
- Hair, J.F. (1995). *Multivariate dates analysis with readings*. Edit. Prentice International Hall
- Johnston and Dinardo. (1997). *Econometric Methods*. McGraw Hill.
- Lebart, L. Morineau, A. Piron, M. (1995) *Statistique exploratoire multidimensionnelle*. Dunod.
- Ward J.H. (1963). *Hierarchical grouping to optimize an objective function*. Journal of the American Statistical Association.
- Wooldridge, J.M. (2010) *Introducción a la econometría: Un enfoque moderno*, 4ta.edición. Cengage Learning Editores SA.