COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA

ESTADÍSTICA

PRUEBAS DE HETEROGENEIDAD DE VARIANZA PARA MODELOS ECONOMÉTRICOS

ARACELI MARTÍNEZ FRANCO

T E S I S

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE:

M A E S T R A E N C I E N C I A S

MONTECILLO, TEXCOCO, EDO. DE MÉXICO

2008

La presente tesis, titulada: **PRUEBAS DE HETEROGENEIDAD DE VARIANZA PARA MODELOS ECONOMÉTRICOS**, realizada por la alumna: **Araceli Martínez Franco**, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRA EN CIENCIAS

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA

CONSEJO PARTICULAR

CONSEJERO ______________________
Dr. Gustavo Ramírez Valverde

ASESOR ______________________
Dr. Sergio Pérez Elizalde

ASESOR ______________________
Dr. Humberto Vaquera Huerta

ASESOR ______________________
Dr. Benito Ramírez Valverde

Montecillo, Texcoco, Edo. de México, 2008

PRUEBAS DE HETEROGENEIDAD DE VARIANZA
PARA MODELOS ECONOMÉTRICOS

Araceli Martínez Franco, MC.

Colegio de Postgraduados, 2008

La regresión lineal es la herramienta estadística más usada para estudiar la relación entre variables; la inferencia basada en este modelo requiere homogeneidad de varianzas. La prueba de White (1980) para detectar heterogeneidad de varianzas es una prueba general que no requiere que se especifique estructura alguna a las varianzas, motivo que la hizo muy popular, al grado de ser la prueba que en forma automática realizan algunos paquetes estadístico (SAS, E-Views, Stata; y Gretl entre otros); sin embargo, algunos estudios empíricos sugieren que esta prueba no debería ser empleada dada su baja potencia (Lyon-Tsai, 1996 y Dios-Rodríguez, 1999). En este trabajo se propone una modificación a la prueba de White para mejorar su potencia. Mediante un estudio de simulación, la potencia y el tamaño de la prueba propuesta son evaluados y comparados con algunas pruebas alternativas.

Palabras clave: heterogeneidad de varianzas, poder de la prueba, tamaño de la prueba, prueba White.

TESTS OF HETEROGENEITY OF VARIANCE FOR ECONOMETRIC MODELS

Araceli Martínez Franco, MC.

Colegio de Postgraduados, 2008

Linear regression is the most used statistical tool to study the relationship among variables; the inference based in this model assumes homogeneity of variances. The White test (1980) for detecting heterocedasticity, is an overall test that does not ask for any structure to the variances tested. That is why, it has been a popular test, in a such a way, that it has been designed the automatic test for several statistical software (SAS, E-Views, Stata; and Gretl among others). However, some studies does not recommend its use, because its low power (Lyon-Tsai, 1996 y Dios-Rodríguez, 1999). In this study, a modification to the White test was proposed in order to improve its power. Using a simulation study the proposed test was evaluated and compared to alternative tests.

Key words: heterogeneity of variances, power of the test, size of the test, White test.

DEDICATORIA

Especialmente a mi abuelita Leonor Amaro Linares †

Por el amor incondicional que siempre recibí hasta el último día de su ser

Por elegir el momento preciso que me permitió acompañarla en su caminar

Por formar parte de este proyecto y que no vio terminar

Para ella con todo cariño donde quiera que este.

Con cariño

Araceli

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el apoyo económico brindado durante la realización de mis estudios.

Al Colegio de Postgraduados, por brindarme la oportunidad de llevar acabo mis estudios de Maestría.

A mi Consejo Particular:

Al Dr. Gustavo Ramírez Valverde, mi más sincero agradecimiento por el tiempo dedicado, consejos y sugerencias que fueron de gran utilidad en la realización del presente trabajo.

A los Drs. Sergio Pérez Elizalde, Humberto Vaquera Huerta y Benito Ramírez Valverde, por el paciente trabajo de revisión y acertados consejos.

A mis demás maestros por compartir desinteresadamente sus amplios conocimientos y experiencias.

A mis padres Jacinto Martínez y Lorenza Franco por darme la oportunidad, la confianza y el amor para cristalizar cada uno de mis anhelos.

A mi hermana Alejandra Martínez Franco por los alientos de superación.

A mi abuelito Ángel Martínez por mostrar siempre interés en mi bienestar.

Especialmente a Rigoberto Sifuentes Amaya.

A Martha Aguilar, Luís Colorado, Iván Pérez y Marcelo Quiterio, amigos que forme durante la maestría y quienes me brindaron su apoyo, motivación y tiempo para el logro de mis objetivos.

CONTENIDO

Página

1. INTRODUCCIÓN 1

2. OBJETIVOS 3

2.1. OBJETIVO GENERAL 3

2.2. OBJETIVOS PARTICULARES 3

3. CONCEPTOS BÁSICOS DE REGRESIÓN MÚLTIPLE 4

3.1. SUPUESTOS DEL MODELO DE REGRESIÓN LINEAL 5

3.2. ESTIMADORES DE MÍNIMOS CUADRADOS 6

4. HETEROGENEIDAD DE VARIANZAS 10

4.1. NATURALEZA DE LA HETEROCEDASTICIDAD 10

4.2. CAUSAS QUE PROVOCAN LA HETEROCEDASTICIDAD 11

4.3. CONSECUENCIAS DE LA HETEROCEDASTICIDAD 13

4.4. DETECCIÓN DE LA HETEROCEDASTICIDAD 14

4.4.1. Métodos gráficos 14

4.4.2. Métodos analíticos 16

4.4.2.1. Prueba de White 17

4.4.2.2. Prueba de Breusch-Pagan-Godfrey 18

4.4.2.3. Prueba de Score de Residuales 20

4.4.2.4. Prueba de Score Estudentizado 22

5. COMPARACIÓN DE PRUEBAS 24

6. COLINEALIDAD 26

6.1. NATURALEZA DE LA COLINEALIDAD 27

6.2. CAUSAS QUE PROVOCAN LA COLINEALIDAD 29

6.3. CONSECUENCIAS DE LA COLINEALIDAD 30

6.4. DETECCIÓN DE LA COLINEALIDAD 30

6.5. ANÁLISIS DE COMPONENTES PRINCIPALES 31

6.5.1. Propiedades de los componentes principales.. 32
6.5.2. Elección del número de componentes principales.. 33

7. PRUEBA WHITE MODIFICADA... 34

7.1. PLANTEAMIENTO DEL PROBLEMA 34
7.2. METODOLOGÍA PROPUESTA 35

8. ESTUDIO DE SIMULACIÓN 37

8.1. SIMULACIÓN PARA ESTIMAR EL PODER DE LA PRUEBA 39
8.2. SIMULACIÓN PARA ESTIMAR EL TAMAÑO DE LA PRUEBA 40

9. RESULTADOS Y DISCUSIÓN 41

10. CONCLUSIONES 48

11. BIBLIOGRAFÍA 49

ANEXOS 52

ANEXO A. CUADROS DEL PODER Y EL TAMAÑO DE LA PRUEBA....................... 53
ANEXO B. FIGURAS DEL PODER Y EL TAMAÑO DE LA PRUEBA 61
ANEXO C. PROGRAMAS EN R 65

1. INTRODUCCIÓN

La regresión lineal es la herramienta estadística más usada para estudiar la relación entre variables; la inferencia basada en este modelo requiere el cumplimiento de supuestos.
Los supuestos del modelo de regresión son:

a) Insesgamiento
b) Homogeneidad de Varianzas
c) Independencia
d) El modelo de regresión está correctamente especificado
e) Normalidad

En particular, cuando se tiene heterogeneidad de varianzas, los estimadores de mínimos cuadrados ordinarios (MCO) se mantienen consistentes e insesgados, pero no son eficientes. Por consiguiente, ya no son los mejores estimadores lineales e insesgados (MELI).
Para verificar el cumplimiento de homogeneidad de varianzas, se han diseñado varias pruebas entre las cuales podemos destacar: la prueba de White (White, 1980), la prueba de Breusch-Pagan- Godfrey (Breusch-Pagan, 1979 y Godfrey, 1978), la prueba de Score de Residuales (Lyon y Tsai, 1996) y la prueba de Score Estudentizado (Koenker, 1981).

La prueba de White ha mostrado tener poca potencia, sin embargo, no se tiene que suponer una relación funcional entre las varianzas y alguna variable independiente, motivo que la hizo muy popular, al grado de ser la prueba que en forma automática realizan algunos paquetes estadístico (SAS, E-Views y Stata, entre otros).

Por otro lado, la colinealidad entre las variables independientes tiene como efectos: varianzas grandes y consecuentemente poca potencia en las pruebas de hipótesis, lo que conduce a generar modelos con muy poco poder explicativo o de difícil interpretación.

La prueba de White, por construcción, genera problemas de colinealidad en las variables independientes debido a que se agregan sus términos cuadráticos y puede ocurrir que tengan dependencia casi lineal.

En este trabajo se propone una modificación a la prueba de White que disminuya los efectos de la colinealidad, tratando de lograr una prueba con mayor potencia que no tenga la necesidad de especificar la relación funcional entre las varianzas y las variables independientes. Mediante un trabajo de simulación se evalúa la potencia y el tamaño de la prueba propuesta y los resultados son comparados con los de otras pruebas que se utilizan con la misma finalidad.

2. OBJETIVOS

2.1. OBJETIVO GENERAL

- Proponer una modificación a la prueba de White que tenga mayor potencia y no dependa de la especificación de la relación funcional entre las varianzas y las variables independientes.

2.2. OBJETIVOS PARTICULARES

- Generar una modificación de la prueba de White tendiente a aminorar los posibles efectos de la colinealidad presente en la regresión auxiliar de la prueba.
- Evaluar las capacidades inferenciales de la prueba propuesta basadas en la potencia y el tamaño de la prueba.
- Comparar las capacidades inferenciales de la prueba propuesta con algunas pruebas alternativas.

3. CONCEPTOS BÁSICOS DE REGRESIÓN MÚLTIPLE

Definición. Modelo de Regresión Múltiple.

Es un modelo donde se estudia la relación que existe entre una variable dependiente y varias variables independientes. (Greene, 1999).

La forma genérica del modelo de regresión lineal múltiple es:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i \qquad (i = 1, \ldots, n)$$

Expresado en notación matricial, las *n* ecuaciones de la expresión anterior se escriben como:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

o

$$y = X\beta + \varepsilon \qquad (3.1)$$

donde:

y : es un vector de $(n \times 1)$, es la variable endógena o explicada cuyo comportamiento se quiere analizar.

X : es una matriz de $(n \times (k+1))$ con rango $k+1$, cada una de las variables exógenas o explicativas y que son consideradas como las causas que crean transformaciones en la variable endógena.

β : un vector de $(k+1) \times 1$, son los parámetros cuyos valores son desconocidos. A través de la estimación de los parámetros se obtiene una cuantificación de las relaciones existentes entre la y y cada una de las X .

ε: un vector de $(n \times 1)$, es la perturbación aleatoria que recoge el efecto conjunto de otras variables no directamente explicitadas en el modelo, cuyo efecto individual sobre la endógena no resulta relevante.

3.1. SUPUESTOS DEL MODELO DE REGRESIÓN LINEAL

La inferencia en el modelo de regresión lineal requiere del cumplimiento de varios supuestos, los cuales se resumen en seguida (Gujarati, 2004):

1. *Insesgamiento*. Toda perturbación aleatoria tiene media cero, o sea

$$E(\varepsilon_i) = 0 \qquad \forall\ i = 1, 2, \ldots, n$$

o matricialmente,

$$E(\varepsilon) = E\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \vdots \\ E(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$$

2. *Homocedasticidad*. Todas las perturbaciones aleatorias tienen la misma varianza, la cual denominaremos σ^2.

$$Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma_i^2 = \sigma^2 \qquad \forall\ i = 1, 2, \ldots, n$$

o matricialmente,

$$\begin{aligned} V(\varepsilon) &= E(\varepsilon'\varepsilon) - \left(E(\varepsilon)E(\varepsilon)\right) = E(\varepsilon'\varepsilon) - 0 \\ &= E(\varepsilon'\varepsilon) = \sigma^2 I \end{aligned}$$

Siendo I la matriz unitaria de orden n.

3. *Independencia de las y´s*. Las perturbaciones son independientes entre sí, lo que equivale a decir que las covarianzas son nulas, de tal forma que

$$E\left(\varepsilon_i, \varepsilon_j\right) = 0$$

Si las observaciones de la muestra son temporales esta condición expresa la ausencia de autocorrelación.

Consecuencia de los supuestos 2 y 3 es que la matriz de varianza y covarianza adopta la forma:

$$E\left(\varepsilon'\varepsilon\right) = E\begin{bmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \cdots & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \cdots & \varepsilon_2\varepsilon_n \\ \vdots & \vdots & & \vdots \\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \cdots & \varepsilon_n^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 I$$

Siendo I la matriz unitaria de orden *n*.

4. *El Modelo de regresión está correctamente especificado.* No hay un sesgo de especificación o error en el modelo utilizado en el análisis empírico.

$$E\left(y/x\right) = X\beta$$

Es decir, no se omiten variables independientes relevantes para explicar a la variable dependiente.

5. *Normalidad*. Consecuencia de los supuestos 1, 2 y 3, el supuesto de normalidad puede expresarse como:

$$\varepsilon_i \sim N\left(0, \sigma^2\right)$$

3.2. ESTIMADORES DE MÍNIMOS CUADRADOS

El método de mínimos cuadrados ordinarios (MCO) es el principal método de estimación en regresión lineal múltiple y puede ser expresado de la forma:

$$\hat{\beta}_{MCO} = (X'X)^{-1} X'y$$

El estimador de MCO existe en forma única siempre que existe la matriz $(X'X)^{-1}$; si la matriz $(X'X)$ es singular, el estimador de MCO no existe en forma única, coloquialmente se dice que no existe el estimador de MCO. (Montgomery, 2002).

Las propiedades estadísticas del estimador de mínimos cuadrados $\hat{\beta}$ son:

1. *Linealidad.* $\hat{\beta}$ es lineal: En este caso, $\hat{\beta} = AY$ en donde $A = (X'X)^{-1} X'$ es la matriz que transforma a *Y* linealmente en $\hat{\beta}$.

2. *Insesgamiento.* $E(\hat{\beta}) = \beta$, donde $\hat{\beta} = (X'X)^{-1} X'y$.

$$\begin{aligned} E(\hat{\beta}) &= E\left[(X'X)^{-1} X'y\right] = E\left[(X'X)^{-1} X'(X\beta+\varepsilon)\right] \\ &= E\left[(X'X)^{-1} X'X\beta + (X'X)^{-1} X'\varepsilon\right] = E\left[\beta + (X'X)^{-1} X'\varepsilon\right] \\ &= \beta + (X'X)^{-1} X'E(\varepsilon) = \beta \end{aligned}$$

Porque $E(\varepsilon) = 0$ y $(X'X)^{-1} X'X = I$.

3. *Varianza.* $V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

$$\begin{aligned} V(\hat{\beta}) &= V\left[(X'X)^{-1} X'y\right] = (X'X)^{-1} X'V(y) X (X'X)^{-1} \\ &= (X'X)^{-1} X'\sigma^2 \mathrm{I} X (X'X)^{-1} = \sigma^2 \mathrm{I} (X'X)^{-1} X'X (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} \end{aligned}$$

Siempre que $V(y) = \sigma^2 I$

4. T*eorema de Gauss-Markov*. Bajo los supuestos del modelo de regresión lineal, para cualquier estimador lineal e insesgado β^*,

$$V\left(\beta^*\right) - Var\left(\hat{\beta}\right) \geq 0$$

Esto es, el estimador de MCO es el estimador lineal de menor varianza dentro de los estimadores insesgados.

Prueba:

Que β^* sea lineal implica que existe una matriz no estocástica A de rango K, tal que $\beta^* = AY$. Bajo los supuestos del modelo de regresión lineal:

$$E\left(\beta^*\right) = AE\left(Y\right) = A\left(X\beta + E\left(\varepsilon\right)\right) = AX\beta \qquad (3.2)$$

Que β^* sea insesgado implica que

$$E\left(\beta^*\right) = \beta \qquad (3.3)$$

Consecuentemente, que β^* sea lineal e insesgado implica que (3.2) y (3.3) tiene que cumplirse simultáneamente, lo que ocurre si y solo si $AX = I$.

Trivialmente:

$$\beta^* = \hat{\beta} + \beta^* - \hat{\beta} = \hat{\beta} + \hat{\gamma} \qquad (3.4)$$

con $\hat{\gamma} = \beta^* - \hat{\beta}$. Note que:

$$V\left(\beta^*\right) = V\left(\hat{\beta}\right) + V\left(\hat{\gamma}\right) - 2Cov\left(\hat{\beta}, \hat{\gamma}\right) \qquad (3.5)$$

donde:

$$\hat{\gamma} = AY - \left(X'X\right)^{-1} X'Y = \left(A - \left(X'X\right)^{-1} X'\right)Y = \left(A - \left(X'X\right)^{-1} X'\right)\left(X\beta + \varepsilon\right)$$

$$= \left(AX\beta - \left(X'X\right)^{-1} X'X\beta + A\varepsilon - \left(X'X\right)^{-1} X'\varepsilon\right)$$

y como $AX = X'A' = I$, entonces:

$$E\left(\hat{\gamma}\right) = E\left(AX\beta - \left(X'X\right)^{-1} X'X\beta + A\varepsilon - \left(X'X\right)^{-1} X'\varepsilon\right) = 0$$

Adicionalmente:

$$
\begin{aligned}
Cov\left(\hat{\beta},\hat{\gamma}\right) &= E\left[\left(\hat{\beta}-\beta\right)\hat{\gamma}'\right] = E\left[\left(\left(X'X\right)^{-1}X'\right)\varepsilon\varepsilon'\left(A-\left(X'X\right)^{-1}X'\right)'\right] \\
&= \sigma^2\left[\left(X'X\right)^{-1}X'\left(A-X\left(X'X\right)^{-1}\right)\right] = \sigma^2\left[\left(X'X\right)^{-1}X'A-\left(X'X\right)^{-1}X'X\left(X'X\right)^{-1}\right] \\
&= 0
\end{aligned}
$$

De (3.5) se tiene entonces:

$$
V\left(\beta^*\right) = V\left(\hat{\beta}\right) + V\left(\hat{\gamma}\right) \tag{3.6}
$$

Entonces, despejando en (3.6):

$$
V\left(\beta^*\right) - V\left(\hat{\beta}\right) = V\left(\hat{\gamma}\right)
$$

que por definición es semidefinida positiva, por lo que $V\left(\beta^*\right) \geq V\left(\hat{\beta}\right)$.

4. HETEROGENEIDAD DE VARIANZAS

El problema de heterocedasticidad se presenta cuando es violado el supuesto de varianza constante de los errores en el modelo de regresión lineal. (Bonilla, 2006).

4.1. NATURALEZA DE LA HETEROCEDASTICIDAD

Uno de los supuestos importantes del modelo clásico de regresión lineal es que la varianza de cada término de perturbación ε_i, condicional a los valores seleccionados de las variables explicativas, es algún número constante igual a σ^2. Simbólicamente:

$$\begin{aligned} V(\varepsilon_i) &= E(\varepsilon_i^2) - (E(\varepsilon_i))^2 \qquad i = 1, 2, \ldots, n \\ &= E(\varepsilon_i^2) = \sigma^2 \end{aligned}$$

Si la varianza de cada término de la perturbación ε_i no es constante, la regresión es heterocedástica. Simbólicamente:

$$V(\varepsilon_i) = E(\varepsilon_i^2) - (E(\varepsilon_i))^2 = \sigma_i^2 \qquad i = 1, 2, \ldots, n$$

Matricialmente el resultado anterior se expresa como:

$$E(\varepsilon_i^2) = \sigma^2 \Sigma \qquad i = 1, 2, \ldots, n$$

donde $\Sigma = diag[\sigma_i]$ con $\sigma_i^2 \neq \sigma_j^2$, para algún $i \neq j$ (Σ es una matriz definida positiva).

Gráficamente, la homocedasticidad y la heterocedasticidad en el modelo de regresión pueden observarse en las siguientes figuras (Gujarati, 2004):

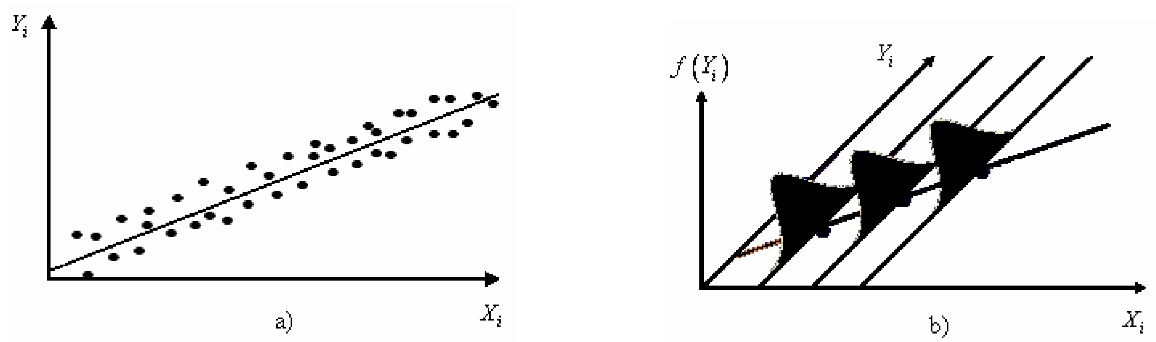


Figura 1. Perturbaciones homocedásticas

En la figura 1 se observa que la gráfica a) es empírica y la gráfica b) es teórica. Se indica que la varianza de Y_i (la cual es igual a la de ε_i), permanece igual sin importar los valores que tome la variable X.

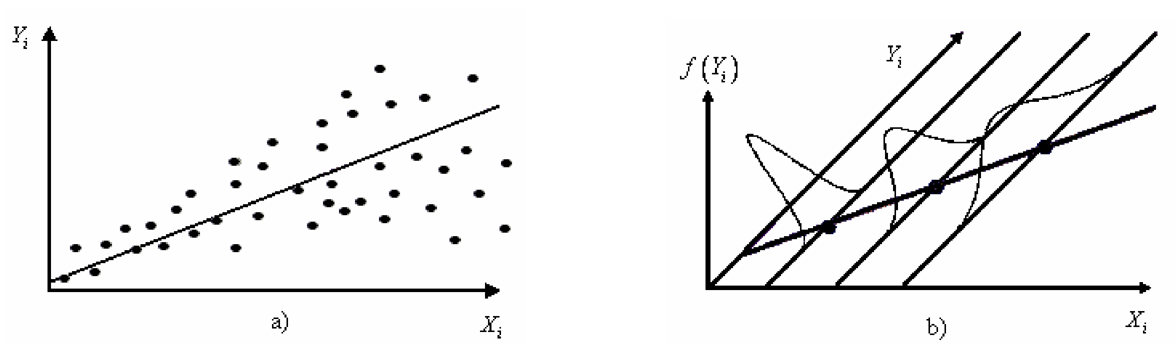


Figura 2. Perturbaciones heterocedásticas

En la figura 2 se observa que la gráfica a) es empírica y la gráfica b) es teórica. Se indica que la varianza de Y_i aumenta a medida que X aumenta.

4.2. CAUSAS QUE PROVOCAN LA HETEROCEDASTICIDAD

De acuerdo a Arce (2001) las causas más frecuentes de heterocedasticidad son:

a) *Variables explicativas cuyo recorrido tenga una gran dispersión respecto a su propia media.*

Esta situación se presenta cuando existen observaciones que tienen valores grandes en una determinada variable explicativa y valores pequeños en esta misma variable. Si esta variable es la que está produciendo la distorsión en el modelo de heterocedasticidad, dicha distorsión será probablemente mayor en aquellas observaciones que contengan una mayor carga de ésta y menor en las que su peso sea más pequeño. Por ello, la varianza de las perturbaciones aleatorias estimada por subperíodos distintos de la muestra sería diferente; es decir, habría heterocedasticidad.

b) *Omisión de variables relevantes en el modelo especificado.*

Cuando se omite una variable importante en el modelo, dicha variable quedará parcialmente recogida en el comportamiento de las perturbaciones aleatorias, pudiendo introducir en éstas su variación.

Recuérdese que la hipótesis inicial del modelo de regresión lineal de homocedasticidad hacía referencia a la varianza constante de las perturbaciones aleatorias, pero no obligaba a que las variables independientes tuvieran también varianza constante, hecho que, además, sería una restricción muy poco plausible.

c) *Asimetría en la distribución de una o más variables independientes incluidas en el modelo de regresión.*

Los ejemplos los constituyen variables económicas como el ingreso, la riqueza y la educación. Es bien sabido que la distribución del ingreso y la riqueza en la mayoría de las sociedades es desigual, pues la mayor parte de estos elementos les corresponden a unos cuantos individuos pertenecientes a los estratos superiores.

d) *Incorrecta transformación de los datos.*

Por ejemplo, se ha tomado la decisión de transformar la variable dependiente siguiendo alguna función (por ejemplo, un logaritmo) cuando no era apropiada; o se ha decidido transformar una o algunas variables independientes cuando esto no era necesario teórica o empíricamente (es decir, tomando primeras diferencias o rezagando).

Al revisar las causas frecuentes de heterocedasticidad, se puede deducir que la varianza no constante de las perturbaciones aleatorias viene casi siempre inducida por alguna variable, presente o no en el modelo, por lo que se podrían distinguir dos componentes en la varianza heterocedástica resultante del modelo: una cambiante, proveniente de esa variable que induce el problema, y una constante.

Matemáticamente escribimos esto del siguiente modo:

$$\sigma_i^2 = f\left(\sigma^2 Z_i\right)$$

donde σ^2 es el parámetro fijo o parte fija de la varianza, y Z_i es la matriz de variable o variables que está produciendo el comportamiento no constante de la varianza de las perturbaciones aleatorias.

4.3. CONSECUENCIAS DE LA HETEROCEDASTICIDAD

El problema de heterocedasticidad repercute directamente sobre la varianza de los parámetros estimados de la regresión (Maddala, 1985).

Los principales problemas de la heterocedasticidad son:

a) *Estimadores de MCO ineficientes.* Los estimadores de mínimos cuadrados se mantienen consistentes e insesgados, pero no son eficientes. Esto significa que existe algún otro estimador insesgado que tienen menor varianza.

   El estimador de Mínimos Cuadrados Generalizados $\left(\hat{\beta} = \left(X'\Sigma^{-1}X\right)^{-1} X'\Sigma^{-1}Y\right)$ es el MELI.

b) *Cálculo incorrecto de las varianzas.* Considere el modelo de regresión $y = X\beta + \varepsilon$, donde $\varepsilon \sim N\left(0, \sigma^2\Sigma\right)$ y $\hat{\beta} = \left(X'X\right)^{-1} X'Y$, entonces:

$$V\left(\hat{\beta}\right) = V\left[\left(X'X\right)^{-1} X'Y\right] = \left(X'X\right)^{-1} X'V\left(Y\right) X \left(X'X\right)^{-1} = \left(X'X\right)^{-1} X'\sigma^2\Sigma X \left(X'X\right)^{-1}$$

Por lo cual la fórmula tradicional de estimación de la matriz de varianzas de MCO $\left(\sigma^2\left(X'X\right)^{-1}\right)$ ya no resulta valida, por lo que las varianzas obtenidas en los paquetes estadísticos (SAS, Minitab, etc.) no es el correcto.

c) *Invalidación de las pruebas de significancia.* Las pruebas de hipótesis y los intervalos de confianza habituales en los paquetes estadísticos en general se basan en la varianza de los estimadores bajo los supuestos ($V\left(\hat{\beta}\right) = \sigma^2\left(X'X\right)^{-1}$, vea punto b), provocando que las pruebas de hipótesis no sean válidas.

4.4. DETECCIÓN DE LA HETEROCEDASTICIDAD

Para probar si un modelo presenta heterocedasticidad se utilizan los métodos gráficos y los métodos analíticos (Greene, 1999). A continuación se describen los métodos:

4.4.1. Métodos gráficos

Estos métodos permiten evaluar gráficamente si existe heterocedasticidad causada por una variable independiente en particular o por todo el conjunto de variables independientes. Para el primer caso se elabora un diagrama de dispersión entre x_i y ε_i^2 donde x_i es el regresor que el investigador supone genera la heterocedasticidad. En el segundo caso, se construye el diagrama de dispersión entre Y_i estimado y ε_i^2. Si estas gráficas muestran alguna tendencia específica, puede afirmarse que existe heterocedasticidad en el modelo de regresión.

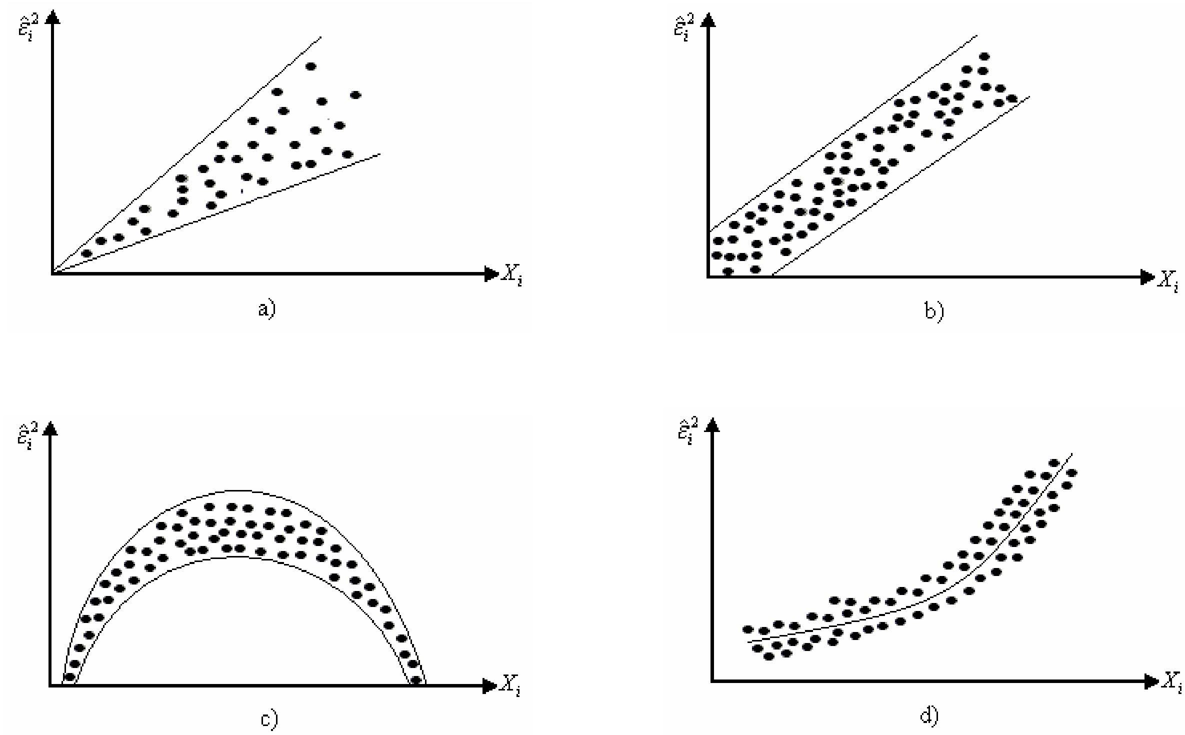
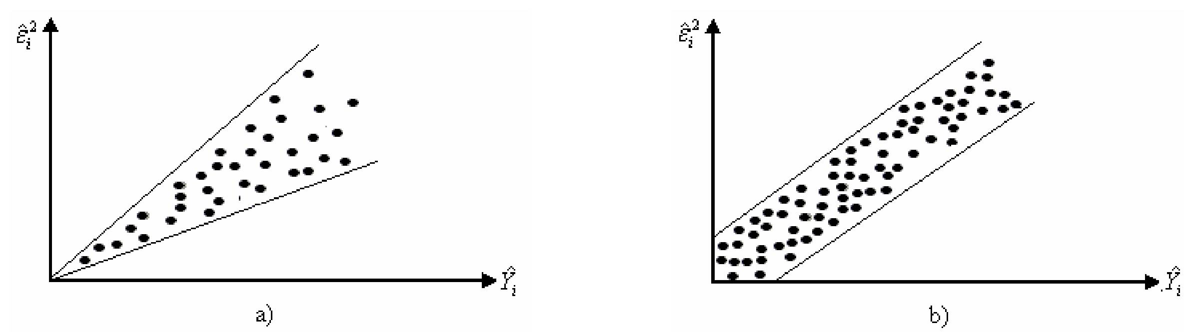


Figura 3. Diagrama de dispersión de los residuos estimados al cuadrado frente a *X*.

La figura 3 muestra patrones definidos, por ejemplo, la gráfica 3b sugiere una relación lineal, mientras que las gráficas 3c y 3d indican una relación cuadrática entre $\hat{\varepsilon}_i^2$ y *X*.



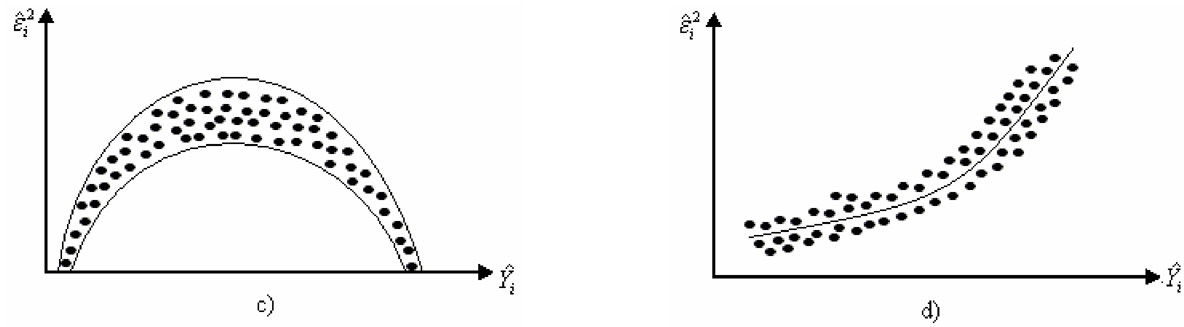


Figura 4. Patrones hipotéticos de los residuos estimados al cuadrado.

La figura 4 puede revelar patrones similares a aquellos dados en la figura 3, por ejemplo, la gráfica 4b sugiere que la varianza del término perturbación está relacionada linealmente con la variable $\hat{Y}_i$, mientras que las gráficas 4c y 4d indican una relación cuadrática entre $\hat{\varepsilon}_i^2$ y $\hat{Y}_i$

4.4.2. Métodos analíticos

Los métodos analíticos presentan una prueba de hipótesis formal que nos permiten evaluar potencias bajo un tamaño de prueba determinado. No se debe olvidar que todas las pruebas se construyen bajo la hipótesis nula de homocedasticidad contra la hipótesis alternativa de heterocedasticidad.

Se destacan algunas pruebas como: la de White (White, 1980), la de Breusch-Pagan-Godfrey (Breusch-Pagan, 1979 y Godfrey, 1978), la de Score de Residuales (Lyon y Tsai, 1996) y finalmente, la prueba de Score Estudentizado (Koenker, 1981).

La prueba de White no hace ningún supuesto, mientras que las pruebas Breusch-Pagan-Godfrey, Score de Residuales y Score Estudentizado suponen que la varianza depende de alguna variable, la cual interviene en la heterogeneidad de varianzas.

A continuación se describe el procedimiento para efectuar las pruebas.

4.4.2.1. Prueba de White

La prueba de White (1980) constituye una forma general de identificar un posible problema de heterocedasticidad causado por alguna de las variables explicativas del modelo, sin que se hagan supuestos sobre la incidencia de una variable en particular o sobre la distribución de los residuos. (Gujarati, 2004).

Sin pérdida de generalidad, se mostrará la prueba de White considerando el siguiente modelo:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \tag{4.1}$$

Para realizar la prueba de White, se procede de la siguiente forma:

Paso 1. Dada la información, estímese (4.1) por el método de mínimos cuadrados ordinarios (MCO) y obténganse los residuos $\hat{\varepsilon}_i$ $\left(\hat{\varepsilon}_i = Y_i - \hat{Y}_i\right)$.

Paso 2. Efectúese la siguiente regresión (auxiliar):

$$\hat{\varepsilon}_i^2 = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \alpha_4 x_{1i}^2 + \alpha_5 x_{2i}^2 + \alpha_6 x_{3i}^2 + \alpha_7 x_{1i} x_{2i} + \alpha_8 x_{1i} x_{3i} + \alpha_9 x_{2i} x_{3i} + v_i$$

Que incluye un término lineal y un término cuadrático para cada variable independiente y un término de interacción para cada par posible de variables independientes.

Es decir, se hace la regresión con los residuos cuadrados de la regresión original, las variables x originales, sus valores al cuadrado y los productos cruzados de las regresoras. Obténgase R^2 de esta regresión (auxiliar).

Paso 3. Bajo la hipótesis nula de que no hay heterocedasticidad, puede demostrarse que el tamaño de la muestra (n) multiplicada por R^2 obtenido de la regresión auxiliar, asintóticamente sigue la distribución χ^2 con grados de libertad igual al número de variables independientes en la regresión auxiliar. Es decir,

$$n \cdot R^2 \sim \chi^2_{gl}$$

donde $gl = grados\ de\ libertad$.

Paso 4. La prueba de hipótesis relacionada con el modelo auxiliar es:

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = \alpha_7 = \alpha_8 = \alpha_9 = 0$ (No hay heterocedasticidad)
$H_a :$ Al menos un $\alpha_i \neq 0 \quad \forall \quad i = 1, 2, \ldots, 9$ (Si hay heterocedasticidad)

La regla de decisión es:

Se rechaza H_0 si $nR^2 > \chi^2_{gl}$ y se concluye que hay heterocedasticidad

No se rechaza H_0 si $nR^2 \leq \chi^2_{gl}$ y se concluye que hay homocedasticidad

4.4.2.2. Prueba de Breusch-Pagan-Godfrey

La prueba de Breusch-Pagan-Godfrey (Breusch-Pagan, 1979 y Godfrey, 1978) admite la posibilidad de que la heterocedasticidad se pueda modelar como una función conjunta de un grupo de variables que sirven para explicar la evolución de la varianza de las perturbaciones aleatorias, estimada ésta a partir del cuadrado de los errores del modelo inicial, que pueden o no estar incluidas en el modelo. (Gujarati, 2004).

Considérese el siguiente modelo de regresión:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i \tag{4.2}$$

Supóngase que la varianza del error σ_i^2 se describe como:

$$\sigma_i^2 = f\left(\alpha_0 + \alpha_1 Z_{1i} + \ldots + \alpha_m Z_{mi}\right)$$

Es decir, σ_i^2 es algún tipo de función de las variables *Z* no estocásticas; alguna de las *X* o todas ellas pueden servir como *Z*. Sin pérdida de generalidad, supóngase que:

$$\sigma_i^2 = \alpha_0 + \alpha_1 Z_{1i} + \ldots + \alpha_m Z_{mi}$$

Es decir, σ_i^2 es una función lineal de las *Z*. Si $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$, $\sigma_i^2 = \alpha_0$ que es una constante. Por consiguiente, para probar si σ_i^2 es homoscedástica, se puede probar la hipótesis de que $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$. Ésta es la idea básica detrás de la prueba de Breusch-Pagan. El procedimiento de la prueba es el siguiente:

Paso 1. Estímese (4.2) mediante el método de mínimos cuadrados ordinarios (MCO) y obténganse los residuos $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$.

Paso 2. Obténgase $\tilde{\sigma}^2 = \sum \hat{\varepsilon}_i^2 / n$.

Paso 3. Constrúyanse las variables p_i definidas como:

$$p_i = \hat{\varepsilon}_i^2 / \tilde{\sigma}^2$$

Que es simplemente cada residuo elevado al cuadrado dividido por $\tilde{\sigma}^2$.

Paso 4. Hágase la regresión de los p_i así construidos sobre las *Z* como

$$p_i = \alpha_0 + \alpha_1 Z_{1i} + \ldots + \alpha_m Z_{mi} + v_i \qquad (4.3)$$

Donde v_i es el término de residuo para esta regresión.

Paso 5. Obténgase la SCR (suma cuadrados de la regresión) de (4.3) y defínase

$$\Theta = \frac{1}{2}(SCR)$$

Bajo la hipótesis nula de que no hay heterocedasticidad, el estadístico Θ se distribuye asintóticamente como una χ^2 con *m* grados de libertad, donde *m* es el número de variables independientes incluidas en la regresión. Es decir:

$$\Theta \sim \chi_m^2$$

Paso 6. La prueba de hipótesis es:

$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$ (No hay heterocedasticidad)

$H_a :$ Al menos un $\alpha_i \neq 0$ $\forall \quad i = 1, 2, \ldots, m$ (Si hay heterocedasticidad)

La regla de decisión es:

Se rechaza H_0 si $\Theta > \chi_m^2$ y se concluye que hay heterocedasticidad

No se rechaza H_0 si $\Theta \leq \chi_m^2$ y se concluye que hay homocedasticidad

4.4.2.3. Prueba de Score de Residuales

Breusch-Pagan (1979) y Cook-Weisberg (1983) propusieron la prueba de Score para probar la heterogeneidad de las varianzas. Verbyla (1993) obtiene la prueba de Score de Residuales para una función exponencial ponderada. Más tarde, Lyon y Tsai (1996) adoptaron ese enfoque y obtuvieron la prueba de Score de Residuales para una función general ponderada. (Lyon y Tsai, 1996).

Para realizar la prueba de Score de Residuales, se procede de la siguiente forma:

Paso 1. Estímese (4.2) por el método de mínimos cuadrados ordinarios (MCO).

Paso 2. Obténgase el modelo reducido

$$y_i = \alpha_0 + \alpha_1 Z_{1i} + \ldots + \alpha_m Z_{mi} + v_i$$

Donde v_i es el término de residuo para esta regresión.

Paso 3. Obtenga las siguientes variables:

a) $d = \left(Y - x\hat{\beta}\right)^2$

b) $\hat{\beta} = \left(X'X\right)^{-1} X'Y$

c) $\bar{\sigma}_0^2 = \dfrac{n\hat{\sigma}_0^2}{n-p}$

d) $\hat{\sigma}_0^2 = \left(\dfrac{1}{n}\right)\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)$

e) $V_0 = X\left(X'X\right)^{-1} X'$

f) $H = \Sigma^{-1/2} X\left(X'\Sigma^{-1} X\right)^{-1} X'\Sigma^{-1/2}$

g) $h = \left(h_{11}h_{22}\ldots h_{nn}\right)'$, son los elementos de la diagonal de H

h) $\Sigma = \sigma_i^2 = \sigma^2 X_i^{\lambda}$

i) $\sigma^2 = 1$

j) $\lambda =$ cualquier número

k) $\mathbb{Z} = X$

Paso 4. Obténgase el estadístico (Verbyla, 1993):

$$S_R = \frac{1}{2}\left(\frac{d}{\bar{\sigma}_0^2} - 1_n + h\right)' \mathbb{Z}\left(\mathbb{Z}'V_0\mathbb{Z}\right)^{-1}\mathbb{Z}'\left(\frac{d}{\bar{\sigma}_0^2} - 1_n + h\right)$$

Bajo la hipótesis nula de que no hay heterocedasticidad, el estadístico S_R se distribuye asintóticamente como una χ^2 con m grados de libertad, donde m es el número de variables independientes incluidas en la regresión. Es decir:

$$S_R \sim \chi_m^2$$

Paso 5. La prueba de hipótesis es:

$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$		(No hay heterocedasticidad)
H_a : Al menos un $\alpha_i \neq 0$	$\forall \quad i = 1, 2, \ldots, m$	(Si hay heterocedasticidad)

La regla de decisión es:

Se rechaza H_0 si $S_R > \chi^2_m$ y se concluye que hay heterocedasticidad

No se rechaza H_0 si $S_R \leq \chi^2_m$ y se concluye que hay homocedasticidad

4.4.2.4. Prueba de Score Estudentizado

Koenker (1981) propuso el estadístico de la prueba de Score Estudentizado.

Para realizar la prueba de Score Estudentizado, se procede de la siguiente forma:

Paso 1. Estímese (4.2) por el método de MCO.

Paso 2. Obténgase el modelo reducido

$$y_i = \alpha_0 + \alpha_1 Z_{1i} + \ldots + \alpha_m Z_{mi} + v_i$$

Donde v_i es el término de residuo para esta regresión.

Paso 3. Obtenga las siguientes variables:

$$d = \left(Y - x\hat{\beta}\right)^2 \quad , \quad \hat{\beta} = \left(X'X\right)^{-1} X'Y \quad , \quad \hat{\sigma}_0^2 = \left(\frac{1}{n}\right)\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)$$

Paso 4. Obténgase el estadístico de la prueba de Score (Verbyla, 1993):

$$S = \frac{1}{2}\left(\frac{d}{\hat{\sigma}_0^2} - 1_n\right)' \mathbb{Z}\left(\mathbb{Z}'\mathbb{Z}\right)^{-1}\mathbb{Z}'\left(\frac{d}{\hat{\sigma}_0^2} - 1_n\right)$$

Paso 5. Obténgase el estadístico de la prueba de Score Estudentizado (Godfrey, 1988):

$$S_{Est} = \frac{2n(S)}{\left(\frac{d}{\hat{\sigma}_0^2} - 1_n\right)'\left(\frac{d}{\hat{\sigma}_0^2} - 1_n\right)}$$

Bajo la hipótesis nula de que no hay heterocedasticidad, el estadístico S_{Est} se distribuye asintóticamente como una χ^2 con m grados de libertad, donde m es el número de variables independientes incluidas en la regresión. Es decir:

$$S_{Est} \sim \chi^2_m$$

Paso 6. La prueba de hipótesis es:

$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$ (No hay heterocedasticidad)
$H_a :$ Al menos un $\alpha_i \neq 0 \quad \forall \quad i = 1, 2, \ldots, m$ (Si hay heterocedasticidad)

La regla de decisión es:

Se rechaza H_0 si $S_{Est} > \chi^2_m$ y se concluye que hay heterocedasticidad

No se rechaza H_0 si $S_{Est} \leq \chi^2_m$ y se concluye que hay homocedasticidad

5. COMPARACIÓN DE PRUEBAS

En esta sección se presentan algunas comparaciones de pruebas de heterogeneidad de varianzas que han realizado algunos autores.

a) Lyon y Tsai (1996) comparan la prueba de White con la prueba de Score Estudentizado.

   Consideran el modelo de regresión lineal:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

   donde: $\beta_0 = \beta_1 = 1$, $x_1 \sim N(0,1)$ y $\varepsilon_i \sim N(0,1)$.

   Los resultados de la función ponderada del segundo orden polinomial son: $e^{\alpha_1 x_1 + \alpha_2 x_1^2}$ con valores de α_1 entre -4 y 4 y α_2 entre -2 y 2.

   Para 1000 simulaciones consideran un tamaño de muestra de 500, comparando cada simulación con la χ_2^2. Utilizan todas las combinaciones posibles para α_1 y α_2 excepto $\alpha_1 = \alpha_2 = 0$. Las diferencias en potencia fueron grandes y concluyen que debería usarse la prueba de Score Estudentizado en lugar de la prueba de White.

b) Dios y Rodríguez (1999) comparan las pruebas específicas: $\sigma_i^2 = z_i'\alpha$ (Prueba cuando la varianza es una función lineal de las variables exógenas), $\sigma_i = z_i'\alpha$ (Prueba cuando la desviación típica es una función lineal de las variables exógenas) y $\sigma_i^2 = exp(z_i'\alpha)$ (Prueba para la heterocedasticidad multiplicativa) y las pruebas: Breusch-Pagan, Goldfeld-Quandt, Harvey, Picos, Correlación por rangos y White.

   Presentan datos del gasto total en productos agroalimentarios (Y) y en renta familiar disponible (X) correspondientes a 47 provincias españolas en el año 1990. La especificación del modelo es $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ y los valores estimados por Mínimos Cuadrados Generalizados Factibles en el modelo real son $\hat{\beta}_0 = 2633.135$ y $\hat{\beta}_1 = 0.156$.

Por lo que todas las pruebas de heterocedasticidad utilizadas rechazan la hipótesis de homocedasticidad en el modelo.

Entonces, se establecen 2000 simulaciones para un tamaño de muestra 47, bajo los supuestos: $\sigma_i^2 = a + bX_i$, $\sigma_i^2 = (a + bX_i)^2$ y $\sigma_i^2 = exp(a + bX_i)$.

Para cada muestra generan un valor aleatorio de una distribución $N(0,1)$ y la variable ε_i a través de $\varepsilon_i = \sigma_i z_i$.

Concluyen que la prueba de Goldfeld-Quandt es bastante buena, presentando un buen comportamiento tanto para homogeneidad como para heterogeneidad. Si se sospecha la existencia de heterocedasticidad, las pruebas más potentes son las tres específicas para cada uno de los supuestos sobre la estructura de la varianza del error y la de Goldfeld-Quandt. Dios y Rodríguez (1999) concluyen que la prueba de White sólo es eficiente cuando la gravedad de la heterocedasticidad es extrema.

c) Simonoff y Tsai (1994) comparan las pruebas: prueba de Razón de Verosimilitud Modificada, prueba de Score, prueba de Score Modificada, prueba de Score Estudentizado y prueba de Score Estudentizado Modificada.

Para 1000 simulación consideraron los siguientes factores:

Tamaño de muestra: 20, 50 y 100

Variables *X*: $X \sim U(0,15)$ y $X \sim exp(7.5)$

Parámetros: $\beta_1 = 0$ y $\beta_2 = \cdots = \beta_p = 1$

Errores: distribución normal, uniforme y t_3 [1].

Concluyen que para errores de distribución no normales y para errores de distribución de colas largas, las pruebas Estudentizadas tienen un mejor tamaño. Bajo errores de distribución de colas pequeñas, las pruebas Razón de Verosimilitud y Score Estudentizado son las más potentes dependiendo del grado de heterocedasticidad.

1 La distribución t_3 se formo como la razón de una normal con la raíz cuadrada de una χ_3^2 variante dividida por sus grados de libertad, con la χ_3^2 variante generada como la suma de una normal cuadrada y el logaritmo natural de una uniforme multiplicada por -2.

6. COLINEALIDAD

Definición. Colinealidad

> Se refiere a una situación en la que dos o más variables independientes están fuertemente correlacionadas y, por tanto, resulta difícil medir sus efectos individuales sobre la variable dependiente. (Bonilla, 2006)

Definición. Colinealidad perfecta

> Es la relación lineal exacta entre dos o más variables independientes o columnas de la matriz *X*.

Si la colinealidad es perfecta, el determinante de la matriz $(X'X)$ es nulo, es decir no existe una solución única a las ecuaciones normales y no pueden obtenerse en forma única las estimaciones de MCO. Un ejemplo típico es el caso en el que el investigador introduce una transformación lineal de dos o más variables que están en el modelo. (Dhrymes, 1984).

Definición. Colinealidad imperfecta

> Se refiere a que dos o más variables independientes o columnas de la matriz *X* están relacionadas de forma aproximadamente lineal.

Si existe colinealidad imperfecta, al obtener la matriz $(X'X)^{-1}$ los valores de sus elementos serán muy elevados (dado que el determinante tomará un valor muy pequeño) y las varianzas estimadas para los parámetros del modelo serán, asimismo, muy grandes. (Dhrymes, 1984).

Definición. Ausencia de Colinealidad

Se refiere a que las variables independientes no tienen ningún tipo de correlación entre ellas. Esto ocurre cuando se cumple que $X_i X_j = 0$, $\forall\ i \neq j$. (Dhrymes, 1984).

6.1. NATURALEZA DE LA COLINEALIDAD

El término colinealidad se atribuye a Ragnar Frisch. Originalmente, significó la existencia de una relación lineal "perfecta" o exacta entre algunas o todas las variables explicativas de un modelo de regresión. Para la regresión con *k* variables que incluye las variables explicativas $X_1, X_2, \ldots, X_k$ (donde $X_1 = 1$ para todas las observaciones que den cabida al término intersección), se dice que existe una relación lineal exacta si se satisface la siguiente condición:

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k = 0 \tag{6.1}$$

donde $\lambda_1, \lambda_2, \ldots, \lambda_k$ son constantes tales que no todas ellas son simultáneamente iguales a cero.

En la actualidad, el término colinealidad se utiliza para incluir el caso de colinealidad perfecta como lo indica la ecuación (5.1), como también, en el caso en el cual hay *X* variables correlacionadas pero no en forma perfecta, de la siguiente manera:

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k + v_i = 0 \tag{6.2}$$

donde v_i es un término de residuos. (Gujarati, 2004).

Si la colinealidad es perfecta los coeficientes de la regresión de las variables *X* son indeterminados y sus errores estándar son infinitos. Si la colinealidad es menos que perfecta los coeficientes de regresión poseen grandes errores estándar, lo que hace que los coeficientes no pueden ser estimados con gran precisión.

En seguida se presentan algunos gráficos donde se muestra el grado de colinealidad (Gujarati, 2004):

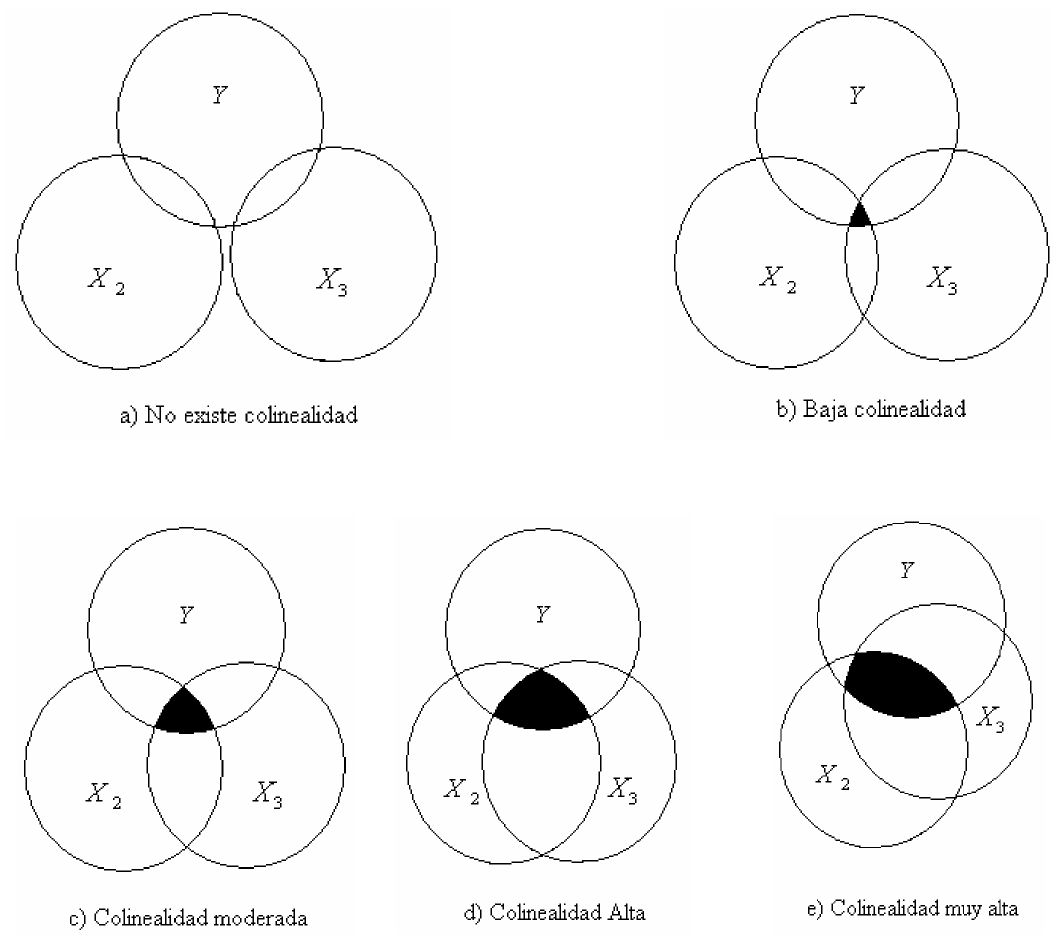


Figura 5. Gráficas de Colinealidad

En la figura 5 los círculos *Y*, X_2 y X_3 representan las variaciones en *Y* (variable dependiente) y en X_2 y X_3 (variables independientes), respectivamente. El grado de la colinealidad puede medirse por la magnitud del área sombreada de los círculos X_2 y X_3. En la gráfica 5a no hay área sombreada entre X_2 y X_3 y, por tanto, no hay colinealidad. En las gráficas 5b hasta 5e, el grado de colinealidad va de bajo a alto.

6.2. CAUSAS QUE PROVOCAN LA COLINEALIDAD

Como lo afirman Montgomery y Peck, la colinealidad puede deberse a los siguientes factores (Mongomery, 2002):

a) *El método de recolección de información empleado.* Por ejemplo, la obtención de muestras en un intervalo limitado de valores tomados por las variables de la población.
b) *Restricciones en el modelo o en la población objeto de muestreo.* Por ejemplo, en la regresión del consumo de electricidad sobre el ingreso y el tamaño de las viviendas. En este ejemplo, una restricción física en la población fue lo que causó este fenómeno: las familias que tienen ingresos mayores en general tienen casas grandes que las familias de menores ingresos, cuando hay restricciones físicas como ésta, habrá colinealidad independientemente del método de muestreo que se emplee.
c) *Especificación del modelo.* Por ejemplo, al agregar términos polinomiales a un modelo de regresión se produce un deterioramiento en $X'X$, además, si el rango de la variable x es pequeño, al agregar un término en x^2 puede producir una colinealidad importante. Con frecuencia se encuentran casos como ésos, cuando dos o más regresores tienen dependencia casi lineal, y el retener esos regresores puede contribuir a la colinealidad.
d) *Un modelo sobredeterminado.* Esto sucede cuando el modelo tiene más variables explicativas que el número de observaciones. Esto podría suceder en investigación médica, donde puede haber un número reducido de pacientes sobre quienes se reúne información respecto a un gran número de variables. También ocurre este suceso en las regresiones auxiliares que ayudan a evaluar la relación lineal existente entre un conjunto de variables independientes, en general, las pruebas de hipótesis sobre relevancia y dependencia estadística en la regresión auxiliar determinan si existe o no colinealidad.

Una razón adicional para la existencia de la colinealidad, sobre todo en los datos de series de tiempo, podría ser que las variables independientes incluidas en el modelo compartan una tendencia común; es decir, que todas aumenten o disminuyan a lo largo del tiempo.

6.3. CONSECUENCIAS DE LA COLINEALIDAD

En los casos de casi o alta colinealidad es probable que se presenten las siguientes consecuencias (Gujarati, 2004):

a) Aún cuando los estimadores MCO son MELI (solo bajo los supuestos usuales), éstos presentan varianzas y covarianzas grandes.
b) Debido a la consecuencia a), los intervalos de confianza tienden a ser mucho más amplios, lo cual propicia una aceptación más fácil de la "hipótesis nula cero" (es decir, que el verdadero coeficiente poblacional es cero.
c) También debido a la consecuencia a), la prueba t de uno o más coeficientes tiende a ser estadísticamente no significativa (falta de potencia).
d) Aun cuando la prueba t de uno o más coeficientes sea estadísticamente no significativa, R^2, la medida global de bondad de ajuste, puede ser muy alta.
e) Los estimadores MCO y sus errores estándar son sensibles a pequeños cambios en la información.

6.4. DETECCIÓN DE LA COLINEALIDAD

Algunas formas de detectar la colinealidad son: la matriz de correlaciones muestrales entre las variables explicatorias, la inversa de la matriz de correlaciones muestrales entre las variables explicatorias, el coeficiente de correlación múltiple, los factores de inflación de varianza, los factores de tolerancia, el determinante de $X^t X$ entre otros. Basados en argumentos numéricos

Belsley *et* al. (1980) sugiere el uso del número de condición de X para determinar el nivel de colinealidad. (Godinez, 2005).

El número de condición de X, η, se define como:

$$\eta = \sqrt{\lambda_1 / \lambda_p}$$

donde λ_1 y λ_p son el máximo y el mínimo de los valores propios de $X^t X$.

Si los regresores son ortogonales $\eta = 1$. Cuanto mayor sea la correlación entre las variables, mayor será el número de condición.
Según Belsley *et* al. (1980), si $\eta < 10$ entonces no hay problemas de colinealidad, si $10 \leq \eta < 30$ entonces hay colinealidad moderada y si $\eta \geq 30$ entonces hay colinealidad dañina o severa que puede afectar seriamente la estimación y las inferencias. (Greene, 1999).

Se han propuesto varias técnicas para atenuar los problemas causados por la colinealidad. Entre los métodos generales están el reunir más datos, la reespecificación del modelo, transformar los datos y el uso de métodos de estimación distintos de los mínimos cuadrados, diseñados en forma específica que posibilitan seguir explicando un porcentaje similar o mayor de la variabilidad de la variable respuesta.

6.5. ANÁLISIS DE COMPONENTES PRINCIPALES

El Análisis de Componentes Principales se emplea cuando deseamos obtener una representación más simple (y en menor dimensión) para un conjunto de variables correlacionadas. En este método, las variables no se consideran explicativas o de respuesta, sino que todas son tratadas de la misma manera. Para examinar las relaciones entre un conjunto de k variables correlacionadas, transformamos el conjunto original de variables a un nuevo conjunto no correlacionado usando una *rotación ortogonal* en el espacio k-dimensional.

Estas nuevas variables serán llamadas *componentes principales*, y son obtenidas en orden decreciente de importancia, de modo que las primeras componentes principales resuman la mayor cantidad posible de la variabilidad de los datos originales. Si un número pequeño de dichas primeras componentes resume la mayor parte de la variación de los datos, podría pensarse que la dimensionalidad real de los datos es menor que k. En tal caso, esperaríamos que estas componentes resulten intuitivamente significativas, y nos ayuden a entender mejor los datos y sean útiles para simplificar análisis posteriores (esto no siempre es cierto en la práctica). Nótese que si las variables son casi no correlacionadas, no tiene sentido llevar a cabo un Análisis de Componentes Principales. Observemos además que el Análisis de Componentes Principales es una técnica matemática, y por tanto no requiere el uso de una estructura probabilística para los datos. (Riba, 1989).

6.5.1. Propiedades de los componentes principales

Los componentes tienen las siguientes propiedades:

a) Conservan la variabilidad inicial: la suma de las varianzas de los componentes es igual a la suma de las varianzas de las variables originales, y la varianza generalizada de los componentes es igual a la original.
b) La proporción de variabilidad explicada por un componente es el cociente entre su varianza, el valor propio asociado al vector propio que lo define, y la suma de los valores propios de la matriz.
c) Las covarianzas entre cada componente principal y las variables X vienen dadas por el producto de las coordenadas del vector propio que define el componente por su valor propio.
d) Las correlaciones entre un componente principal y una variable X es proporcional al coeficiente de esa variable en la definición del componente, y el coeficiente de proporcionalidad es el cociente entre la desviación típica del componente y la desviación típica de la variable.

e) Los k' componentes principales $(k' < k)$ proporcionan la predicción lineal óptima con k' variables del conjunto de variables X.

6.5.2. Elección del número de componentes principales

Por lo general se usan las siguientes dos alternativas (Morrison, 1976):

a) Elegir el número de componentes hasta donde se ha acumulado por lo menos el 75% de la proporción determinada de la varianza. Esta regla es arbitraria y debe aplicarse con cierto cuidado. Por ejemplo es posible que un único componente alcance el 90% de la variabilidad y sin embargo pueden existir otros componentes que sean muy adecuados para explicar las variables.
b) Elegir hasta la componente cuyo valor propio sea mayor que 1.

7. PRUEBA WHITE MODIFICADA

La prueba White Modificada se propuso para disminuir el efecto de la colinealidad en la prueba de White, con el objetivo de comparar las capacidades inferenciales de esta prueba contra las de otras pruebas alternativas. A continuación se explica la construcción de la prueba.

7.1. PLANTEAMIENTO DEL PROBLEMA

Sin pérdida de generalidad, considérese el modelo (Harvey, 1976):

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \qquad (7.1)$$

donde:

$$\begin{aligned} \varepsilon_i &\sim N\left(0, \sigma_i^2\right) \\ \sigma_i^2 &= \sigma^2 X_i^\lambda \end{aligned} \qquad i = 1, 2, \ldots, n$$

donde:

x_1, x_2 y x_3 son los vectores de observaciones, β_0, β_1, β_2 y β_3 son los parámetros del modelo, $X = x_2 + x_3$ ó $X = x_1$ y $\lambda = 2,\ 2.5,\ 3\ y\ 3.5$.

La prueba de White se construye considerando $\hat{\varepsilon}_i^2 = \left(Y_i - \hat{Y}_i\right)^2$ del modelo de regresión (7.1) y se forma el nuevo modelo de regresión:

$$\hat{\varepsilon}_i^2 = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \alpha_4 x_{1i}^2 + \alpha_5 x_{2i}^2 + \alpha_6 x_{3i}^2 + \alpha_7 x_{1i} x_{2i} + \alpha_8 x_{1i} x_{3i} + \alpha_9 x_{2i} x_{3i} + v_i \qquad (7.2)$$

Esta prueba utiliza un polinomio entre las variables de regresión auxiliar, de tal manera que se induce colinealidad mediante la especificación del modelo.

El estimador de MCO en (7.2) es $\hat{\beta} = \left(X'X\right)^{-1} X'y$ y $V\left(\hat{\beta}\right) = \left(X'X\right)^{-1} \sigma^2$

Se tiene el interés en probar el siguiente juego de hipótesis:

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = \alpha_7 = \alpha_8 = \alpha_9 = 0 \qquad \text{(No hay heterocedasticidad)}$$
$$H_a: \text{ Al menos un } \alpha_i \neq 0 \qquad \forall \quad i = 1, 2, \ldots, 9 \qquad \text{(Si hay heterocedasticidad)}$$

7.2. METODOLOGÍA PROPUESTA

La regresión auxiliar se realizará con los componentes principales que presentan ausencia de colinealidad por ser ortogonales.

El objetivo del Análisis de Componentes Principales (Hotelling, 1933) es hacer una reducción de dimensionalidad. Es decir, la información contenida en *k* variables predictoras (en este trabajo, *k=9*) $X = (X_1, \ldots, X_k)$ reducirla a $Z = (Z_1, \ldots, Z_{k'})$, con $k' < k$; las nuevas variables Z_i son llamados los componentes principales y son no correlacionadas entre si.

Los criterios utilizado para seleccionar el número de componentes principales en este estudio será determinado por la proporción de la varianza acumulada que satisfaga el 75% y elegir hasta el componente cuyo valor propio sea mayor que 1 (Morrison, 1976).

Para determinar los componentes principales, hay que hallar una matriz ortogonal V tal que $Z = X^{*}\mathrm{V}$, para la cual

$$\begin{aligned} Z'Z &= \left(X^{*}\mathrm{V}\right)'\left(X^{*}\mathrm{V}\right) = \mathrm{V}'X^{*\prime}X^{*}\mathrm{V} \\ &= diag\left(\lambda_1, \ldots, \lambda_m\right) \end{aligned}$$

De tal forma que $\mathrm{VV'=V'V=I}$, donde los λ_j para $j = 1, 2, \ldots, m$ son los valores propios (autovalores) de la matriz de correlación $X^{*\prime}X^{*}$.

El número máximo *m* de componentes principales que se puede construir es igual a $m = mín(n, k)$, es el mínimo entre el número de observaciones y variables predictoras. Aplicando una propiedad de los valores propios se tiene:

$$tr\left(\hat{\Sigma}\right) = \sum_{i=1}^{k} \lambda_i$$

Por lo tanto, la *j*-ésima componente principal Z_j tiene desviación estándar igual a $\sqrt{\lambda_j}$ y puede ser escrita como:

$$Z_j = \mathrm{v}_{j1}X_1^* + \mathrm{v}_{j2}X_2^* + \ldots + \mathrm{v}_{jk}X_k^*$$

donde $\mathrm{v}_{j1}, \mathrm{v}_{j2}, \ldots, \mathrm{v}_{jk}$ son los elementos de la *j*-ésima fila de V.

Posteriormente, se compara el nivel de significancia de los componentes principales bajo los dos criterios antes mencionados (WM_1 se refiere al primer criterio y WM_2 se refiere al segundo criterio); el nivel de significancia será corregido de acuerdo al número de componentes principales seleccionados y al método de Bonferroni.

El método de Bonferroni es un método muy general (para cualquier modelo) tiene la ventaja de ser muy simple. Controla la tasa de error, dividiendo el nivel de significancia (α) entre el número de comparaciones (m) llevadas a cabo, para este caso m corresponde al número de componentes principales seleccionados para la regresión auxiliar. Cada comparación se evalúa utilizando un nivel de significancia $\alpha_c = \frac{\alpha}{m}$. (Hochberg, 1988).

Entonces hay que hacer ajustes a los p-value de las hipótesis individuales.

Las pruebas de hipótesis son:

$$\begin{aligned} H_{0_j} &= Homogeneidad\ de\ varianzas \\ H_{a_j} &= Heterogeneidad\ de\ varianzas \end{aligned} \qquad j = 1, \ldots, m$$

La regla de decisión es:

Se rechaza H_{0_j} si $(p-value)_j \leq \frac{\alpha}{m}$ y se concluye que hay heterocedasticidad

No se rechaza H_{0_j} si $(p-value)_j > \frac{\alpha}{m}$ y se concluye que hay homocedasticidad

8. ESTUDIO DE SIMULACIÓN

De acuerdo a los dos criterios utilizados para seleccionar el número de componentes principales en la prueba de White, resultaron dos pruebas para este estudio: la White Modificada 1 (WM_1) y White Modificada 2 (WM_2).

Para evaluar las capacidades inferenciales (poder y tamaño de la prueba) de las pruebas WM_1 y WM_2 se realizó un estudio de simulación. Dichos resultados son comparados con los de las pruebas White (W), Breusch-Pagan-Godfrey (BP), Score de Residuales (S_R) y Score Estudentizado (S_{Est}). El estudio de simulación estuvo formado por tres diseños, los cuales son:

a) Diseño I. Favorece a la prueba de White

   El $\varepsilon_i \sim NIID\left(0,\sigma_i^2\right)$ y la varianza del error $\left(\sigma_i^2 = \sigma^2\left(x_2 + x_3\right)^{\lambda},\ \lambda = 2,\ 2.5,\ 3 \text{ y } 3.5\right)$ depende de dos variables $\left(x_2 \text{ y } x_3\right)$; como en la prueba de White no se hace ningún supuesto, en las demás pruebas se supuso que la varianza dependía de x_1, la cual no interviene en la heterogeneidad de las varianzas.

b) Diseño II. Favorece a las otras pruebas.

   El $\varepsilon_i \sim NIID\left(0,\sigma_i^2\right)$ y la varianza del error $\left(\sigma_i^2 = \sigma^2 x_1^{\lambda},\ \text{donde } \lambda = 2,\ 2.5,\ 3 \text{ y } 3.5\right)$ depende de la variable x_1 y las pruebas competidoras supusieron que la varianza dependía de x_1, la cual interviene en la heterogeneidad de las varianzas.

c) Diseño III. No hay heterogeneidad de varianzas.

   El $\varepsilon_i \sim NIID\left(0,\sigma_i^2\right)$ y la varianza del error es constante $\left(\sigma_i^2 = 1\right)$.

Los diseños I, II y III, tienen ciertos factores en común, los cuales se describen enseguida:

1. Tamaños de muestra:
   $n = 20,\ 30,\ 40,\ 50,\ 60,\ 70,\ 80,\ 90 \text{ y } 100$
2. Grados de correlación muestral:
   Baja $(0.0001\ ,0.01)$, Media $(0.45\ ,0.55)$ y Alta (≥ 0.95).

Las condiciones que se establecen para los tres diseños son:

1. Variables explicatorias fijas:
   - x_1
     $x_1 \sim Unif(0,1)$
   - x_2
     $x_2 = x_1 + (U)K$, donde $U \sim Unif(0,1)$ y K toma valores que permite obtener la correlación alta entre x_1 y x_2.
     $x_2 = (x_1 J) + (U)K$, donde $U \sim Unif(0,1)$, K y J toman valores que permiten obtener la correlación media entre x_1 y x_2.
     $x_2 = x_1 + (U)L$, donde $U \sim Unif(0,1)$ y L toma valores que permite obtener la correlación baja entre x_1 y x_2.
   - x_3
     $x_3 \sim Unif(0,1)$

2. Los valores del parámetro β. Se usara la orientación favorable para estimar a β que corresponde al vector propio asociado con el mayor valor propio de $X^t X$.

8.1. SIMULACIÓN PARA ESTIMAR EL PODER DE LA PRUEBA

En esta sección se establece el diseño I y II de la simulación de Monte Carlo para obtener las estimaciones del poder de la prueba.

Definición. Poder de la Prueba (Mukhodhyay, 2000).

$\beta(\theta) = 1 - \pi(\theta)$ para valores de θ bajo H_a

La cual es equivalente a:

$\beta(\theta) = \Pr\{\text{Error tipo I}\} = \Pr\{\text{Rechazar } H_0 \mid H_a \text{ es verdadera}\}$

La hipótesis empleada en la simulación es:

$$H_0: \textit{Homogeneidad de varianzas}$$
$$vs$$
$$H_a: \textit{Heterogeneidad de varianzas}$$

El procedimiento de simulación para obtener las estimaciones de $\beta(\theta)$, denotado por $\hat{\beta}(\theta)$, consta de los siguientes pasos:

1. Generar 1000 muestras de tamaño *n* para los diseños I y II.
2. Para cada muestra que se genera en los diseños I y II, se calcula el valor del estadístico de las pruebas.
3. Cada valor de los estadísticos de prueba se compara con el valor crítico de la χ^2_{gl} $(\alpha = 0.05)$ y se toma la decisión de rechazar o no rechazar H_0, de acuerdo a la regla de decisión presentada en todas las pruebas de estudio.
4. Realizar un conteo de cuántas veces se rechaza H_0 de las 1000 muestras generadas, para todas las pruebas estudiadas.
5. Por último, la estimación del poder de la prueba, $\hat{\beta}(\theta)$, se calcula por medio de la fracción $\frac{r}{1000}$, donde r es el número de veces que se rechaza H_0.

8.2. SIMULACIÓN PARA ESTIMAR EL TAMAÑO DE LA PRUEBA

En esta sección se presenta el diseño III de la simulación de Monte Carlo para obtener las estimaciones del criterio: tamaño de la prueba.

Definición. Tamaño de la Prueba (Mukhodhyay, 2000).

$$\beta(\theta) = \alpha(\theta) = \Pr\{\text{Rechazar } H_0 \mid H_0 \text{ es verdadera}\}$$

Para valores de θ bajo H_0.

La hipótesis empleada para esta simulación es la misma que se utilizó para estimar el poder de la prueba.

El procedimiento de simulación para obtener las estimaciones de $\alpha(\theta)$, denotado por $\hat{\alpha}(\theta)$, consta de los siguientes pasos:

1. Generar 1000 muestras de tamaño *n* para el diseño III
2. Para cada muestra que se genera en el diseño III, se calcula el valor del estadístico de las pruebas.
3. Cada valor de los estadísticos de prueba se compara con el valor crítico de la χ^2_{gl} $(\alpha = 0.05)$ y se toma la decisión de rechazar o no rechazar H_0, de acuerdo a la regla de decisión presentada en todas las pruebas de estudio.
4. Realizar un conteo de cuántas veces se rechaza H_0 de las 1000 muestras generadas para todas las pruebas estudiadas.
5. Por último, la estimación del tamaño de la prueba, $\hat{\alpha}(\theta)$, se calcula por medio de la fracción $\frac{r}{1000}$, donde r es el número de veces que se rechaza H_0.

9. RESULTADOS Y DISCUSIÓN

A continuación se muestran los resultados obtenidos de las simulaciones correspondientes a los diseños I, II y III. Cabe mencionar que en la regla de decisión se utilizó el tamaño de la prueba $\alpha = 0.05$ en todas las simulaciones.

Resultado para el diseño I con correlación alta.

En la figura 6 se observa que al aumentar el tamaño de la muestra (n) aumenta la potencia de la prueba. La prueba White Modificada 1 (WM_1) y la prueba White Modificada 2 (WM_2) son las más potentes y la prueba de White (W) es la menos potente.

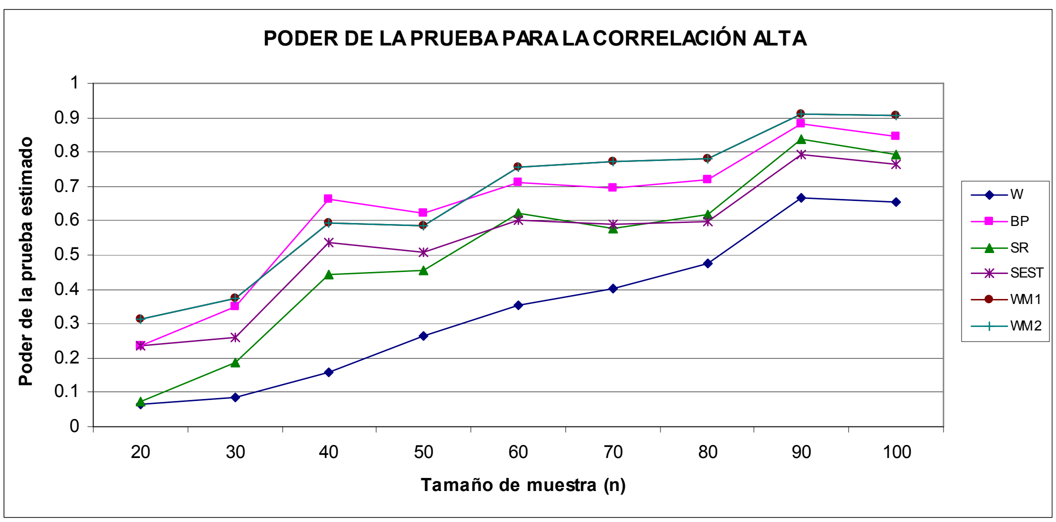


Figura 6. Potencia de las pruebas W, BP, S_R, S_{Est}, WM_1 y WM_2 para varios tamaños de muestra, nivel de significancia $\alpha = 0.05$ y $\varepsilon \sim N\left(0, \sigma^2 (x_2 + x_3)^{\lambda=2}\right)$.

Resultado para el diseño I con correlación media.

En la figura 7 se observan los resultados de las simulaciones para $\sigma_i^2 = \sigma^2 (x_2 + x_3)^{\lambda}$, donde $\lambda = 2$. Estos resultados representan las estimaciones del poder de la prueba. En la figura puede verse que al aumentar el tamaño de la muestra (n), aumenta la potencia de la prueba.

La prueba de Score Estudentizado (S_{Est}) presenta menor potencia y las pruebas White Modificada 1 (WM_1) y White Modificada 2 (WM_2) son las más potentes.

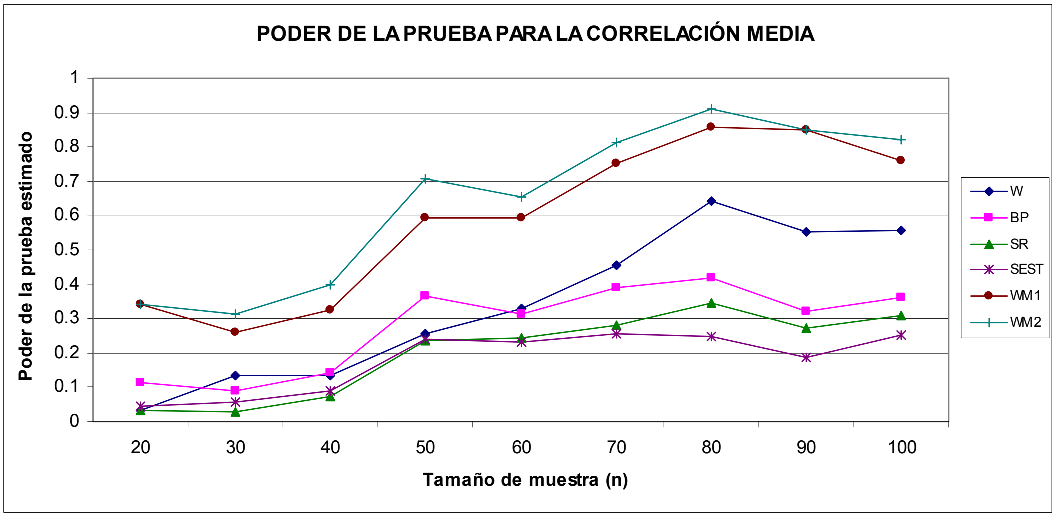


Figura 7. Potencia de las pruebas W, BP, S_R, S_{Est}, WM_1 y WM_2 para varios tamaños de muestra, nivel de significancia $\alpha = 0.05$ y $\varepsilon \sim N\left(0, \sigma^2 (x_2 + x_3)^{\lambda=2}\right)$.

Resultado para el diseño I con correlación baja.

En la figura 8 se muestran los resultados de las simulaciones para $\sigma_i^2 = \sigma^2 (x_2 + x_3)^{\lambda}$ y $\lambda = 2.5$. Puede verse que al aumentar el tamaño de la muestra (n) aumenta la potencia de la prueba.

Como ocurre en la correlación media también en la correlación baja la prueba de Score Estudentizado (S_{Est}) es la que presenta menor potencia, las pruebas White Modificada 1 (WM_1) y White Modificada 2 (WM_2) son las más potentes.

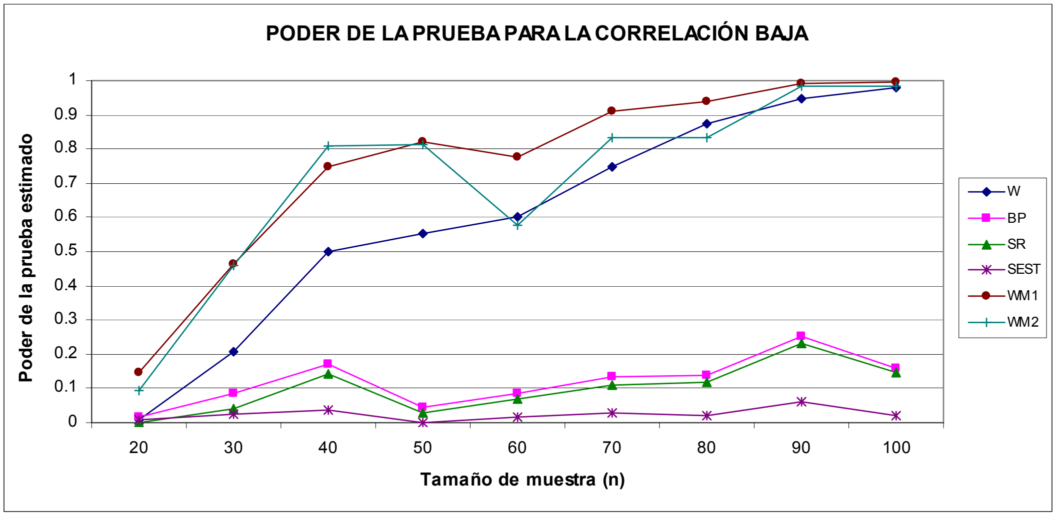


Figura 8. Potencia de las pruebas W, BP, S_R, S_{Est}, WM_1 y WM_2 para varios tamaños de muestra, nivel de significancia $\alpha = 0.05$ y $\varepsilon \sim N\left(0, \sigma^2 \left(x_2 + x_3\right)^{\lambda=2.5}\right)$.

Resultado para el diseño II con correlación alta.

En la figura 9 se muestran los resultados de las simulaciones para $\sigma_i^2 = \sigma^2 x_1^\lambda$ y $\lambda = 3$. La prueba de Breusch-Pagan-Godfrey (BP) presenta mayor potencia y la prueba de White (W) presenta menor potencia, esto se debe a que la prueba BP se ve favorecida por la simulación.

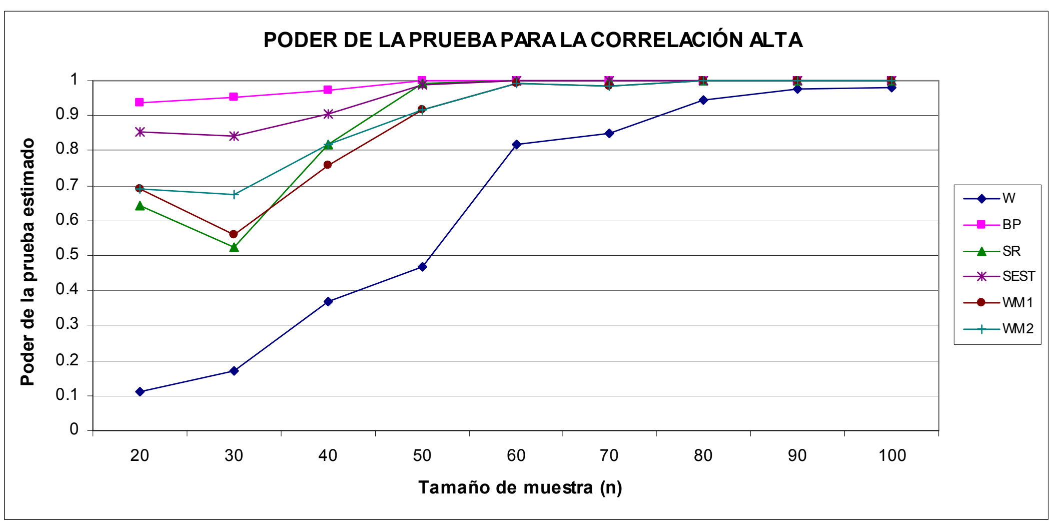


Figura 9. Potencia de las pruebas W, BP, S_R, S_{Est}, WM_1 y WM_2 para varios tamaños de muestra, nivel de significancia $\alpha = 0.05$ y $\varepsilon \sim N\left(0, \sigma^2 x_1^{\lambda=3}\right)$.

Resultado para el diseño II con correlación media.

En la figura 10 se muestran los resultados de las simulaciones para $\sigma_i^2 = \sigma^2 x_1^3$. En general, al aumentar el tamaño de la muestra (n) aumenta la potencia de las pruebas. La prueba de Breusch-Pagan-Godfrey (BP) es la más potente y la prueba menos potente es la de White (W).

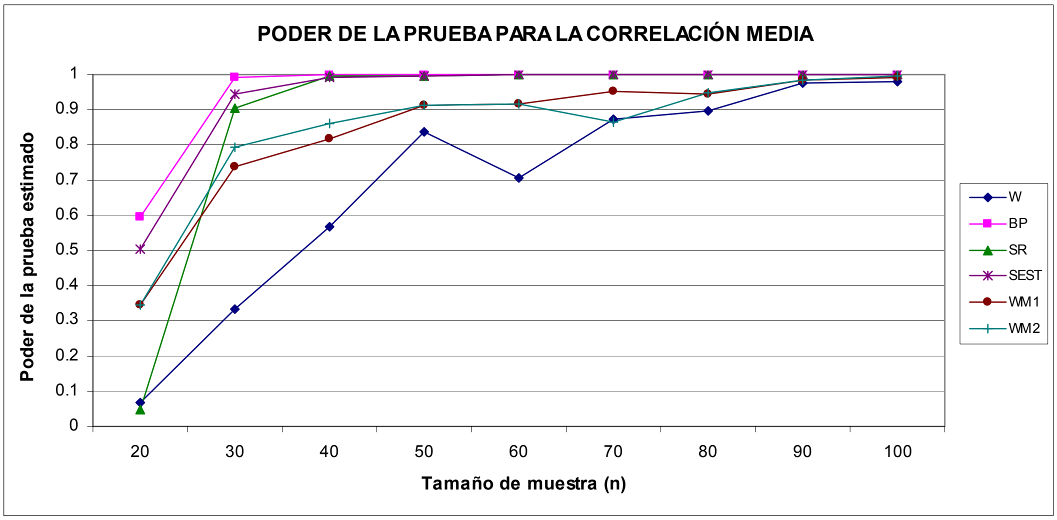


Figura 10. Potencia de las pruebas W, BP, S_R, S_{Est}, WM_1 y WM_2 para varios tamaños de muestra, nivel de significancia $\alpha = 0.05$ y $\varepsilon \sim N\left(0, \sigma^2 x_1^{\lambda=3}\right)$.

Resultado para el diseño II con correlación baja.

En la figura 11 se muestran los resultados de las simulaciones para $\sigma_i^2 = \sigma^2 x_1^{\lambda=3.5}$. Estos resultados representan las estimaciones del poder de la prueba. Al aumentar el tamaño de la muestra (n) aumenta la potencia de las pruebas. Para tamaños de muestra *n=20, 30,…, 100* la prueba de Breusch-Pagan-Godfrey (BP) es la más potente. En General, las pruebas White (W) y White Modificada 2 (WM_2) presentan menor potencia.

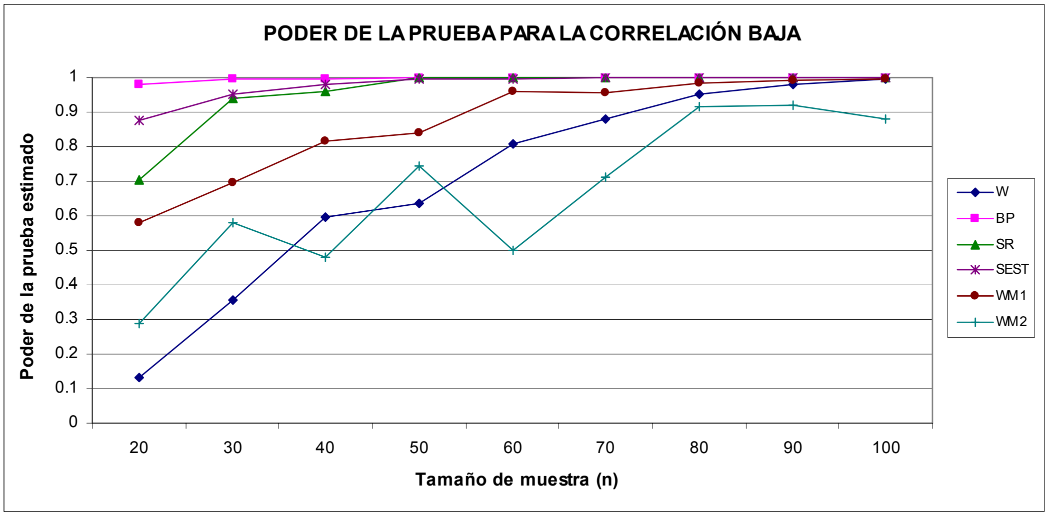


Figura 11. Potencia de las pruebas W, BP, S_R, S_{Est}, WM_1 y WM_2 para varios tamaños de muestra, nivel de significancia $\alpha = 0.05$ y $\varepsilon \sim N\left(0, \sigma^2 x_1^{\lambda=3.5}\right)$.

Resultado para el diseño III con correlación alta.

En la figura 12 se muestran los resultados para las estimaciones del tamaño de la prueba para $\sigma_i^2 = 1$. Para muestras pequeñas la prueba de Score de Residuales $\left(S_R\right)$ y la prueba de White $\left(W\right)$ presentan tamaños de prueba muy por debajo del nivel de significancia nominal $\left(\alpha = 0.05\right)$. La prueba de Score de Residuales $\left(S_R\right)$ se muestra ligeramente conservadora aún cuando se aumenta el tamaño de muestra. Las demás pruebas oscilan alrededor del nivel de significancia nominal.

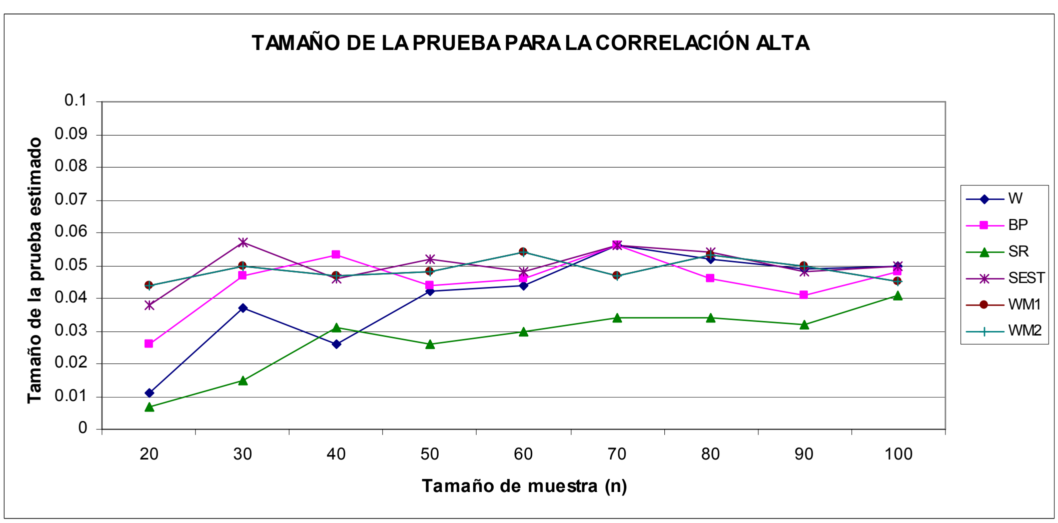


Figura 12. Tamaño de las pruebas W, BP, S_R, S_{Est}, WM_1 y WM_2 para varios tamaños de muestra, nivel de significancia $\alpha = 0.05$ y $\varepsilon \sim N(0,1)$.

Resultado para el diseño III con correlación media.

En la figura 13 se muestran los resultados para las estimaciones del tamaño de la prueba para $\sigma_i^2 = 1$. Para muestras pequeñas la prueba de Score de Residuales (S_R) y la prueba de White (W) presentan tamaños de prueba muy por debajo del nivel de significancia nominal $(\alpha = 0.05)$. La prueba de Score de Residuales (S_R) se muestra ligeramente conservadora aún cuando se aumenta el tamaño de muestra. Las demás pruebas oscilan alrededor del nivel de significancia nominal.

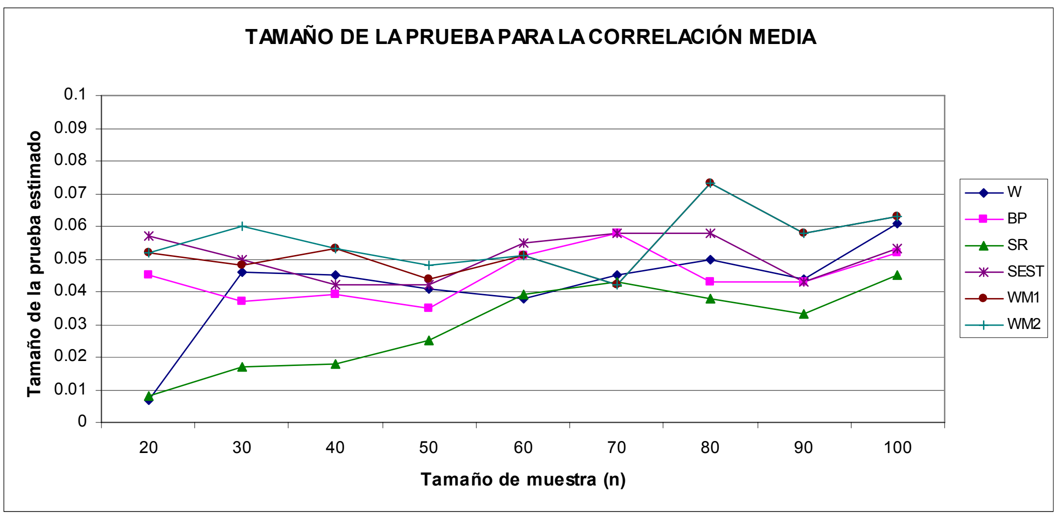


Figura 13. Tamaño de las pruebas W, BP, S_R, S_{Est}, WM_1 y WM_2 para varios tamaños de muestra, nivel de significancia $\alpha = 0.05$ y $\varepsilon \sim N(0,1)$.

Resultado para el diseño III con correlación baja.

En la figura 14 se muestran los resultados para las estimaciones del tamaño de la prueba para $\sigma_i^2 = 1$. Para muestras pequeñas la prueba de Score de Residuales (S_R) y la prueba de White (W) presentan tamaños de prueba muy por debajo del nivel de significancia nominal $(\alpha = 0.05)$.

La prueba de Score de Residuales $\left(S_R\right)$ se muestra ligeramente conservadora aún cuando se aumenta el tamaño de muestra. Las demás pruebas oscilan alrededor del nivel de significancia nominal.

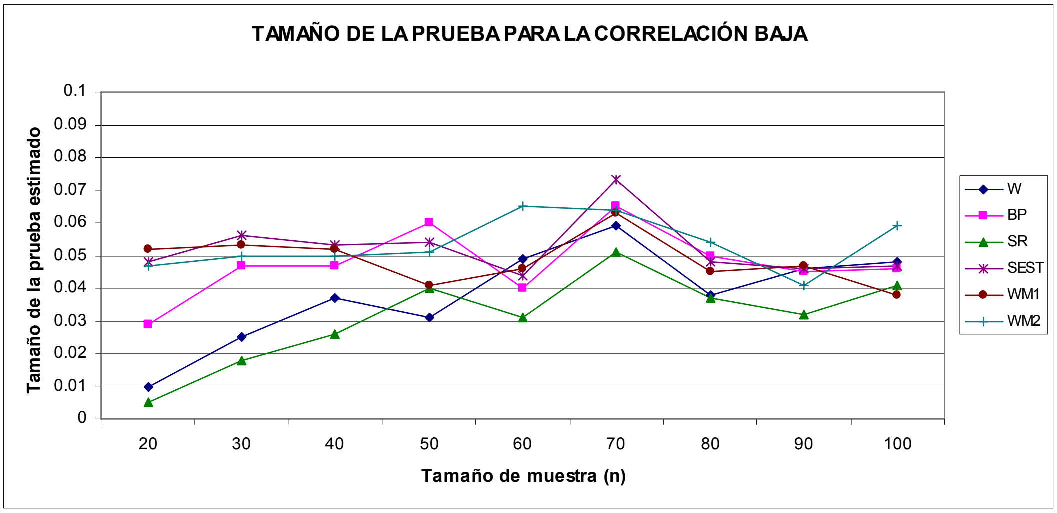


Figura 14. Tamaño de las pruebas W, BP, S_R, S_{Est}, WM_1 y WM_2 para varios tamaños de muestra, nivel de significancia $\alpha = 0.05$ y $\varepsilon \sim N\left(0,1\right)$.

Coincidiendo con algunos trabajos (Lyon y Tsai, 1996), las pruebas Score Estudentizado $\left(S_{Est}\right)$ y Score de Residuales $\left(S_R\right)$ son más potentes que la prueba de White $\left(W\right)$, aunque esto no siempre ocurre, un caso en particular es para el diseño I aplicado a la correlación media y baja entre las variables x_1 y x_2.

10. CONCLUSIONES

Con base en el método de simulación realizado para evaluar las capacidades inferenciales (el poder y el tamaño de la prueba) de la prueba White Modificada con respecto a las demás pruebas, se obtienen las siguientes conclusiones:

1. El poder de la prueba aumenta conforme aumentan el tamaño de la muestra n y el valor del parámetro λ, lo que da evidencia de que se trata de pruebas consistentes.
2. La prueba White Modificada $\left(WM_1 \text{ y } WM_2\right)$ resultó más potente que la prueba de White con excepción de pocos casos cuando la prueba de White es más potente que la prueba White Modificada 2 (correlación baja y tamaños de muestra grande).
3. En general, la prueba White Modificada 1 $\left(WM_1\right)$ es más potente que la prueba White Modificada 2 $\left(WM_2\right)$.
4. En la mayoría de los casos la prueba White Modificada $\left(WM_1 \text{ y } WM_2\right)$ presenta mayor potencia que las pruebas alternativas (Breusch-Pagan-Godfrey, Score de Residuales y Score Estudentizado).
5. Cuando las pruebas alternativas (Breusch-Pagan-Godfrey, Score de Residuales y Score Estudentizado) son favorecidas, la prueba Breusch-Pagan-Godfrey es la más potente y la diferencia de está con la prueba White Modificada no fue grande y siempre menor que la prueba de White.
6. Con respecto al tamaño de la prueba, la prueba de Breusch-Pagan-Godfrey y la de Score de Residuales fueron muy conservadoras en tamaños de muestra pequeños.
7. La prueba de Score de Residuales se mantuvo conservadora al aumentar el tamaño de muestra, sin embargo, la diferencia se fue acortando al aumentar el tamaño de muestra. El resto de las pruebas mantuvo su tamaño de prueba cerca del nivel de significancia nominal.
8. Los paquetes estadísticos deberían utilizar la prueba White Modificada 1 $\left(WM_1\right)$ en lugar de la prueba de White como la prueba automática en el software estadístico.

11. BIBLIOGRAFÍA

Arce, Rafael. (2001). Conceptos básicos sobre la heterocedasticidad en el modelo básico de regresión lineal. Universidad Autónoma de Madrid.

Belsley, D.A., E. Kuh and R. Welsh. (1980). Regression diagnostics: Identifying Influential Data and Source of Collinearity. John Wiley & Sons. New York.

Bonilla Londoño, Jorge A. (2006). Introducción a la Econometría. Centro de Estudios sobre Desarrollo Económico.

Breusch, T. S. and Pagan, A. R. (1979). A simple Test for Heteroscedasticity and Random coefficient Variation. *Econometrica*, Vol. 47, No. 5. Págs. 1287-1294.

Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, Vol. 70, Págs. 1-10.

Dios, Palomares R. y Rodríguez, Fonseca C. (1999). Análisis comparativo de estimadores pretest de heterocedasticidad en modelos económetricos. *Quesito,* Vol. 23, No. 3. Págs: 437-464.

Dhrymes, P. J. (1984). Econometría. Editorial AC

Georges Monette, John Fox. (1992). Generalizad Collinearity Diagnostics. *Journal of the American Statistical Association*, Vol. 87, No. 417. Págs. 178-183.

Glejser, J. (1969). A new test of heteroskedasticity. *Journal of the American Association*, Vol. 64, No. 325. Págs. 316-323.

Godinez Jaimes, Flaviano. (2005). Inferencia en el modelo de regresión logística en presencia de colinealidad entre las variables explicatorias y diferentes porcentajes de traslape. Tesis. Colegio de Postgraduados.

Godfrey, L. G. (1978). Testing for Multiplicative Heteroscedasticity. *Journal of Enocometrics*. Vol. 8, Págs. 227-236.

Godfrey, L. G. (1988). Misspecification tests in econometrics. Cambridge University.

Greene H. William. (1999). Análisis Econométrico. Tercera Edición. Prentice Hall.

Gujarati N. Damodar. (2004). Econometría. Cuarta Edición. Mc Graw-Hill.

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica,* Vol. 44, No. 3, Págs. 461-466.

Hochberg. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika.* Vol 75: Págs. 800-802.

Hotelling, Harold. (1933). Analysis of a Complex of Statistical Variables into Principal Componentes. *Journal of Educational Psychology*, Vol. 24, Págs. 417-441, 489-520.

Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. *J. Econometr.,* Vol. 17, Págs. 107-112.

Lyon, John D. and Tsai, Chih-Ling. (1996). A Comparison of Tests for Heteroscedasticity. *The Statistician.* Vol. 45, No. 3. Págs. 337-349.

Maddala .G. S. (1985). Econometría, Mc Graw-Hill.

Montgomery, Douglas C. (2002). Introducción al análisis de regresión lineal. Tercera edición. CECSA.

Morrison, D.F. (1976). Multivariate Statistical Methods. 2nd Edition. McGraw−Hill, Montreal

Mukhodhyay, Nitis. (2000). Probability and statistical inference. Editorial Marcel Dekker.

Ramírez Valverde, Gustavo y Ramírez Valverde, Benito. (2006). Colinealidad y mínimos cuadrados ponderados. *Análisis de Coyuntura,* Vol. 12, No. 1, Págs. 283-296.

Riba, M. D. (1989). Una panorámica de las técnicas estadísticas multivariantes. Barcelona: Univ. Autónoma.

Rutemiller, H. C. and Bowers, D. A. (1968). Estimation in heteroscedastic regression models. *J. Am. Statist. Ass,* Vol 63, Págs. 552-557.

Simonoff, Jeffrey S. y Tsai, Chih-Ling (1994). Use of modified profile likelihood for improved tests of constancy of variante in regresión. *Applied Statistics,* Vol. 43, No. 2. Págs: 357-370.

Stewart W, G. (1987). Collinearity and Least Squares Regression. *Statistical Science*, Vol. 2, No. 1, Págs. 68-84.

Verbyla, A. P. (1993). Modelling variance heterogeneity: residual maximum likelihood and diagnostics. *J.R. Statist. Soc. B,* Vol. 55, No. 2, Págs. 493-508.

White, Halbert. (1980). A Heteroskedasticity-Consistent Covariance Matriz Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, Vol. 48, No. 4. Págs. 817-838.

ANEXOS

ANEXO A. CUADROS DEL PODER Y EL TAMAÑO DE LA PRUEBA

Cuadro 1. Resultados de las simulaciones correspondientes al diseño I.

El $\varepsilon \sim N\left(0, \sigma^2 \left(x_2 + x_3\right)^{\lambda}\right)$. Se muestra $\hat{\beta}(\theta)$ = poder de la prueba estimado. El tamaño de la prueba utilizado en la regla de decisión es $\alpha = 0.05$.

CORRELACIÓN ALTA							
		Pruebas					
λ	n	W	BP	S_R	S_{Est}	White Modificada	
						WM_1	WM_2
2	20	0.065	0.236	0.073	0.236	0.313	0.313
	30	0.084	0.351	0.187	0.261	0.373	0.373
	40	0.16	0.664	0.443	0.537	0.592	0.592
	50	0.263	0.623	0.454	0.51	0.584	0.584
	60	0.353	0.711	0.62	0.602	0.756	0.756
	70	0.403	0.694	0.579	0.588	0.773	0.773
	80	0.474	0.721	0.617	0.596	0.78	0.78
	90	0.668	0.881	0.838	0.794	0.911	0.911
	100	0.656	0.846	0.793	0.765	0.905	0.905
2.5	20	0.044	0.293	0.027	0.242	0.281	0.281
	30	0.213	0.603	0.337	0.477	0.532	0.633
	40	0.223	0.594	0.39	0.477	0.555	0.555
	50	0.443	0.768	0.676	0.625	0.765	0.765
	60	0.505	0.817	0.711	0.626	0.858	0.858
	70	0.666	0.929	0.88	0.835	0.944	0.964
	80	0.627	0.841	0.768	0.733	0.901	0.901
	90	0.791	0.955	0.918	0.877	0.975	0.975
	100	0.844	0.961	0.934	0.886	0.986	0.986
3	20	0.109	0.492	0.152	0.44	0.505	0.505
	30	0.181	0.419	0.171	0.263	0.457	0.457
	40	0.29	0.689	0.446	0.483	0.668	0.668
	50	0.582	0.833	0.726	0.664	0.871	0.871
	60	0.623	0.895	0.815	0.714	0.941	0.941
	70	0.766	0.955	0.919	0.853	0.956	0.956
	80	0.773	0.977	0.961	0.905	0.984	0.984
	90	0.856	0.918	0.884	0.811	0.978	0.978
	100	0.956	0.99	0.984	0.964	0.998	0.998
3.5	20	0.047	0.307	0.046	0.186	0.31	0.31
	30	0.362	0.555	0.423	0.419	0.591	0.591
	40	0.566	0.91	0.785	0.776	0.87	0.87
	50	0.563	0.815	0.703	0.643	0.851	0.851
	60	0.644	0.873	0.764	0.692	0.913	0.913
	70	0.819	0.982	0.959	0.921	0.976	0.976
	80	0.911	0.981	0.968	0.934	0.993	0.993
	90	0.91	0.997	0.995	0.973	0.996	0.996
	100	0.938	0.998	0.991	0.94	0.997	0.997

Cuadro 2. Resultados de las simulaciones correspondientes al diseño I.

El $\varepsilon \sim N\left(0, \sigma^2 \left(x_2 + x_3\right)^{\lambda}\right)$. Se muestra $\hat{\beta}(\theta)$ = poder de la prueba estimado. El tamaño de la prueba utilizado en la regla de decisión es $\alpha = 0.05$.

CORRELACIÓN MEDIA							
		Pruebas					
λ	n	W	BP	S_R	S_{Est}	White Modificada	
						WM_1	WM_2
2	20	0.033	0.114	0.032	0.045	0.341	0.341
	30	0.133	0.091	0.029	0.058	0.262	0.312
	40	0.134	0.143	0.074	0.088	0.327	0.399
	50	0.255	0.367	0.237	0.238	0.592	0.709
	60	0.328	0.313	0.245	0.231	0.595	0.656
	70	0.457	0.39	0.281	0.258	0.752	0.813
	80	0.641	0.42	0.345	0.247	0.859	0.911
	90	0.553	0.321	0.274	0.186	0.851	0.851
	100	0.555	0.36	0.308	0.251	0.76	0.823
2.5	20	0.015	0.102	0.009	0.074	0.163	0.18
	30	0.235	0.134	0.05	0.072	0.35	0.363
	40	0.217	0.22	0.144	0.136	0.41	0.41
	50	0.386	0.511	0.393	0.363	0.648	0.741
	60	0.505	0.586	0.447	0.385	0.756	0.786
	70	0.623	0.679	0.593	0.469	0.828	0.839
	80	0.663	0.512	0.419	0.314	0.914	0.914
	90	0.842	0.091	0.073	0.017	0.979	0.979
	100	0.915	0.807	0.753	0.62	0.972	0.988
3	20	0.119	0.099	0.048	0.062	0.396	0.396
	30	0.32	0.53	0.392	0.279	0.758	0.836
	40	0.449	0.332	0.246	0.218	0.531	0.569
	50	0.5	0.388	0.291	0.221	0.737	0.796
	60	0.54	0.398	0.315	0.199	0.861	0.861
	70	0.617	0.134	0.103	0.043	0.899	0.493
	80	0.869	0.7	0.638	0.466	0.95	0.971
	90	0.878	0.734	0.673	0.541	0.952	0.974
	100	0.949	0.534	0.489	0.249	0.98	0.98
3.5	20	0.023	0.163	0.056	0.141	0.16	0.232
	30	0.149	0.322	0.158	0.204	0.432	0.569
	40	0.295	0.195	0.092	0.09	0.494	0.591
	50	0.445	0.273	0.225	0.091	0.652	0.652
	60	0.563	0.387	0.281	0.197	0.768	0.768
	70	0.836	0.5	0.35	0.19	0.938	0.971
	80	0.908	0.742	0.715	0.487	0.978	0.945
	90	0.934	0.813	0.731	0.56	0.979	0.986
	100	0.971	0.379	0.347	0.167	0.996	0.8

n = tamaño de la muestra

Cuadro 3. Resultados de las simulaciones correspondientes al diseño I.

El $\varepsilon \sim N\left(0, \sigma^2 \left(x_2 + x_3\right)^{\lambda}\right)$. Se muestra $\hat{\beta}(\theta)$ = poder de la prueba estimado. El tamaño de la prueba utilizado en la regla de decisión es $\alpha = 0.05$.

CORRELACIÓN BAJA							
		Pruebas					
λ	n	W	BP	S_R	S_{Est}	White Modificada WM_1	White Modificada WM_2
2	20	0.026	0.035	0.002	0.013	0.229	0.131
	30	0.158	0.07	0.031	0.029	0.294	0.296
	40	0.565	0.156	0.118	0.032	0.752	0.386
	50	0.654	0.115	0.098	0.019	0.84	0.743
	60	0.545	0.082	0.066	0.018	0.838	0.629
	70	0.595	0.107	0.082	0.032	0.856	0.635
	80	0.822	0.142	0.122	0.028	0.944	0.664
	90	0.872	0.121	0.108	0.021	0.954	0.852
	100	0.882	0.118	0.098	0.023	0.934	0.88
2.5	20	0.009	0.015	0.002	0.009	0.146	0.095
	30	0.206	0.085	0.04	0.024	0.465	0.459
	40	0.501	0.171	0.144	0.037	0.749	0.809
	50	0.553	0.044	0.028	0	0.82	0.811
	60	0.601	0.084	0.068	0.018	0.776	0.578
	70	0.747	0.134	0.11	0.027	0.911	0.833
	80	0.873	0.14	0.117	0.019	0.939	0.833
	90	0.949	0.253	0.231	0.061	0.993	0.984
	100	0.978	0.16	0.145	0.021	0.997	0.982
3	20	0.022	0.074	0.02	0.022	0.223	0.22
	30	0.147	0.04	0.021	0.006	0.326	0.234
	40	0.549	0.151	0.128	0.036	0.787	0.499
	50	0.731	0.139	0.109	0.016	0.887	0.827
	60	0.86	0.205	0.19	0.027	0.949	0.872
	70	0.904	0.184	0.168	0.027	0.955	0.484
	80	0.891	0.085	0.068	0.007	0.965	0.495
	90	0.944	0.139	0.127	0.017	0.983	0.646
	100	0.991	0.223	0.206	0.037	0.999	0.859
3.5	20	0.167	0.242	0.139	0.079	0.503	0.436
	30	0.299	0.133	0.117	0.045	0.636	0.285
	40	0.704	0.163	0.145	0.024	0.89	0.678
	50	0.628	0.127	0.096	0.014	0.861	0.421
	60	0.839	0.222	0.192	0.025	0.963	0.591
	70	0.937	0.278	0.25	0.04	0.974	0.746
	80	0.941	0.161	0.138	0.016	0.987	0.961
	90	0.99	0.291	0.276	0.035	0.999	0.941
	100	0.984	0.245	0.231	0.029	0.995	0.816

n = tamaño de la muestra

Cuadro 4. Resultados de las simulaciones correspondientes al diseño II. El $\varepsilon \sim N\left(0, \sigma^2 x_1^{\lambda}\right)$.

Se muestra $\hat{\beta}(\theta)$ = poder de la prueba estimado. El tamaño de la prueba utilizado en la regla de decisión es $\alpha = 0.05$.

CORRELACIÓN ALTA							
		Pruebas					
						White Modificada	
λ	n	W	BP	S_R	S_{Est}	WM_1	WM_2
	20	0.086	0.737	0.239	0.632	0.457	0.457
	30	0.062	0.791	0.202	0.641	0.387	0.387
	40	0.413	0.992	0.947	0.952	0.792	0.792
	50	0.402	0.989	0.947	0.957	0.855	0.892
2	60	0.715	1	0.999	0.995	0.966	0.966
	70	0.655	0.999	0.996	0.994	0.965	0.965
	80	0.892	1	1	1	0.99	0.99
	90	0.896	1	1	1	0.995	0.995
	100	0.879	1	1	1	0.995	0.995
	20	0.03	0.745	0.118	0.61	0.391	0.391
	30	0.309	0.991	0.926	0.939	0.814	0.814
	40	0.379	0.985	0.823	0.936	0.755	0.755
	50	0.576	1	0.996	0.993	0.938	0.938
2.5	60	0.767	1	0.999	0.993	0.984	0.969
	70	0.74	1	1	1	0.979	0.979
	80	0.781	1	0.999	1	0.974	0.974
	90	0.969	1	1	1	0.999	0.999
	100	0.954	1	1	1	0.999	0.999
	20	0.11	0.935	0.641	0.852	0.691	0.691
	30	0.169	0.954	0.522	0.841	0.561	0.676
	40	0.37	0.974	0.818	0.906	0.758	0.817
	50	0.467	0.999	0.991	0.989	0.915	0.915
3	60	0.817	1	1	1	0.993	0.993
	70	0.85	1	1	0.999	0.983	0.983
	80	0.946	1	1	1	1	1
	90	0.976	1	1	1	1	1
	100	0.979	1	1	1	0.999	0.999
	20	0.069	0.894	0.411	0.785	0.538	0.538
	30	0.143	0.983	0.626	0.891	0.642	0.642
	40	0.58	0.999	0.995	0.988	0.887	0.887
	50	0.663	1	0.998	0.996	0.95	0.95
3.5	60	0.839	1	1	0.999	0.983	0.983
	70	0.88	1	1	1	0.997	0.997
	80	0.948	1	1	1	0.997	0.997
	90	0.976	1	1	1	1	1
	100	0.988	1	1	1	1	1

n = tamaño de la muestra

Cuadro 5. Resultados de las simulaciones correspondientes al diseño II. El $\varepsilon \sim N\left(0,\sigma^2 x_1^{\lambda}\right)$. Se muestra $\hat{\beta}(\theta)$ = poder de la prueba estimado. El tamaño de la prueba utilizado en la regla de decisión es $\alpha = 0.05$.

CORRELACIÓN MEDIA							
		Pruebas					
λ	n	W	BP	S_R	S_{Est}	White Modificada	
						WM_1	WM_2
2	20	0.071	0.751	0.302	0.679	0.452	0.452
	30	0.375	0.849	0.647	0.782	0.532	0.602
	40	0.452	0.95	0.843	0.907	0.686	0.686
	50	0.463	0.987	0.912	0.957	0.77	0.77
	60	0.592	0.999	0.995	0.99	0.784	0.784
	70	0.754	0.999	0.995	0.994	0.812	0.816
	80	0.751	1	0.999	0.997	0.889	0.932
	90	0.907	1	1	1	0.965	0.973
	100	0.953	1	1	1	0.986	0.988
2.5	20	0.05	0.658	0.184	0.587	0.383	0.439
	30	0.175	0.871	0.53	0.745	0.415	0.388
	40	0.359	0.993	0.892	0.955	0.759	0.759
	50	0.642	1	0.999	0.998	0.931	0.975
	60	0.557	0.998	0.99	0.991	0.691	0.779
	70	0.842	1	0.999	0.999	0.951	0.979
	80	0.863	1	1	1	0.943	0.943
	90	0.952	1	1	1	0.986	0.967
	100	0.968	1	1	1	0.979	0.979
3	20	0.067	0.596	0.047	0.502	0.346	0.346
	30	0.333	0.993	0.904	0.943	0.739	0.794
	40	0.568	1	0.998	0.994	0.817	0.862
	50	0.839	1	0.998	0.996	0.912	0.912
	60	0.707	1	1	0.999	0.917	0.917
	70	0.872	1	1	1	0.952	0.866
	80	0.898	1	1	1	0.944	0.949
	90	0.978	1	1	1	0.985	0.985
	100	0.98	1	1	1	0.991	0.996
3.5	20	0.087	0.816	0.115	0.69	0.48	0.48
	30	0.329	0.999	0.895	0.947	0.598	0.613
	40	0.74	1	1	0.997	0.914	0.914
	50	0.533	1	0.998	0.992	0.818	0.901
	60	0.815	1	1	1	0.919	0.919
	70	0.858	1	1	0.999	0.967	0.967
	80	0.953	1	1	1	0.988	0.986
	90	0.971	1	1	1	0.994	0.999
	100	0.996	1	1	1	0.997	0.997

n = tamaño de la muestra

Cuadro 6. Resultados de las simulaciones correspondientes al diseño II. El $\varepsilon \sim N\left(0, \sigma^2 x_1^{\lambda}\right)$. Se muestra $\hat{\beta}(\theta)$ = poder de la prueba estimado. El tamaño de la prueba utilizado en la regla de decisión es $\alpha = 0.05$.

CORRELACIÓN BAJA							
λ	n	***Pruebas***					
						White Modificada	
		W	BP	S_R	S_{Est}	WM_1	WM_2
2	20	0.048	0.63	0.223	0.549	0.239	0.304
	30	0.291	0.859	0.545	0.761	0.471	0.27
	40	0.341	0.891	0.72	0.828	0.498	0.341
	50	0.369	0.974	0.866	0.931	0.618	0.327
	60	0.62	1	0.997	0.992	0.842	0.81
	70	0.695	1	0.99	0.993	0.907	0.522
	80	0.788	1	0.998	0.998	0.899	0.646
	90	0.832	1	1	1	0.934	0.69
	100	0.936	1	1	1	0.969	0.89
2.5	20	0.077	0.461	0.115	0.439	0.32	0.183
	30	0.182	0.959	0.596	0.821	0.477	0.383
	40	0.408	0.986	0.918	0.949	0.601	0.456
	50	0.69	1	1	0.999	0.859	0.753
	60	0.711	1	0.996	0.994	0.881	0.576
	70	0.841	1	1	1	0.945	0.718
	80	0.927	1	1	1	0.992	0.608
	90	0.95	1	1	1	0.992	0.796
	100	0.981	1	1	1	0.993	0.942
3	20	0.037	0.633	0.155	0.53	0.276	0.06
	30	0.275	0.99	0.896	0.939	0.59	0.622
	40	0.46	0.996	0.942	0.97	0.765	0.514
	50	0.671	1	0.999	0.997	0.826	0.848
	60	0.708	1	0.997	0.994	0.898	0.845
	70	0.729	1	1	0.998	0.893	0.283
	80	0.924	1	1	1	0.987	0.936
	90	0.974	1	1	1	0.996	0.769
	100	0.983	1	1	1	0.995	0.721
3.5	20	0.133	0.979	0.705	0.877	0.579	0.29
	30	0.356	0.997	0.939	0.954	0.695	0.58
	40	0.596	0.998	0.961	0.979	0.815	0.48
	50	0.635	1	1	0.997	0.841	0.745
	60	0.809	1	1	0.998	0.96	0.5
	70	0.882	1	1	0.999	0.955	0.714
	80	0.952	1	1	1	0.985	0.915
	90	0.979	1	1	1	0.992	0.92
	100	0.996	1	1	1	0.998	0.88

n = tamaño de la muestra

Cuadro 7. Resultados de las simulaciones correspondientes al diseño III. El $\varepsilon \sim N(0,1)$. Se muestra $\hat{\alpha}(\theta)$ = tamaño de la prueba estimado. El tamaño de la prueba utilizado en la regla de decisión es $\alpha = 0.05$.

CORRELACIÓN ALTA						
	Pruebas					
n	W	BP	S_R	S_{Est}	White Modificada	
					WM_1	WM_2
20	0.011	0.026	0.007	0.038	0.044	0.044
30	0.037	0.047	0.015	0.057	0.05	0.05
40	0.026	0.053	0.031	0.046	0.047	0.047
50	0.042	0.044	0.026	0.052	0.048	0.048
60	0.044	0.046	0.03	0.048	0.054	0.054
70	0.056	0.056	0.034	0.056	0.047	0.047
80	0.052	0.046	0.034	0.054	0.053	0.053
90	0.049	0.041	0.032	0.048	0.05	0.05
100	0.05	0.048	0.041	0.05	0.045	0.045

n = tamaño de la muestra

Cuadro 8. Resultados de las simulaciones correspondientes al diseño III. El $\varepsilon \sim N(0,1)$. Se muestra $\hat{\alpha}(\theta)$ = tamaño de la prueba estimado. El tamaño de la prueba utilizado en la regla de decisión es $\alpha = 0.05$.

CORRELACIÓN MEDIA						
	Pruebas					
n	W	BP	S_R	S_{Est}	White Modificada	
					WM_1	WM_2
20	0.007	0.045	0.008	0.057	0.052	0.052
30	0.046	0.037	0.017	0.05	0.048	0.06
40	0.045	0.039	0.018	0.042	0.053	0.053
50	0.041	0.035	0.025	0.042	0.044	0.048
60	0.038	0.051	0.039	0.055	0.051	0.051
70	0.045	0.058	0.043	0.058	0.042	0.042
80	0.05	0.043	0.038	0.058	0.073	0.073
90	0.044	0.043	0.033	0.043	0.058	0.058
100	0.061	0.052	0.045	0.053	0.063	0.063

n = tamaño de la muestra

Cuadro 9. Resultados de las simulaciones correspondientes al diseño III. El $\varepsilon \sim N(0,1)$. Se muestra $\hat{\alpha}(\theta)$ = tamaño de la prueba estimado. El tamaño de la prueba utilizado en la regla de decisión es $\alpha = 0.05$.

CORRELACIÓN BAJA						
	Pruebas					
n	W	BP	S_R	S_{Est}	White Modificada	
					WM_1	WM_2
20	0.01	0.029	0.005	0.048	0.052	0.047
30	0.025	0.047	0.018	0.056	0.053	0.05
40	0.037	0.047	0.026	0.053	0.052	0.05
50	0.031	0.06	0.04	0.054	0.041	0.051
60	0.049	0.04	0.031	0.044	0.046	0.065
70	0.059	0.065	0.051	0.073	0.063	0.064
80	0.038	0.05	0.037	0.048	0.045	0.054
90	0.046	0.045	0.032	0.046	0.047	0.041
100	0.048	0.046	0.041	0.047	0.038	0.059

n = tamaño de la muestra

ANEXO B. FIGURAS DEL PODER Y EL TAMAÑO DE LA PRUEBA

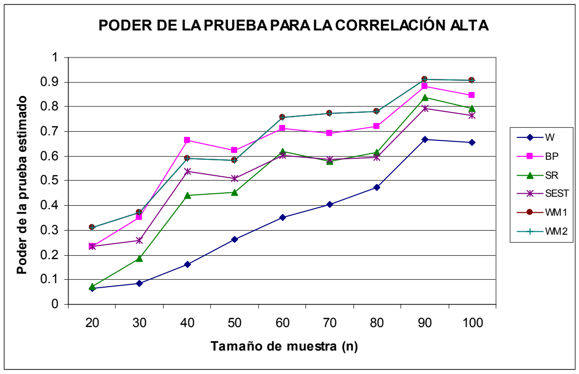


Figura B.1. Corresponde al cuadro 1 con $\lambda = 2$

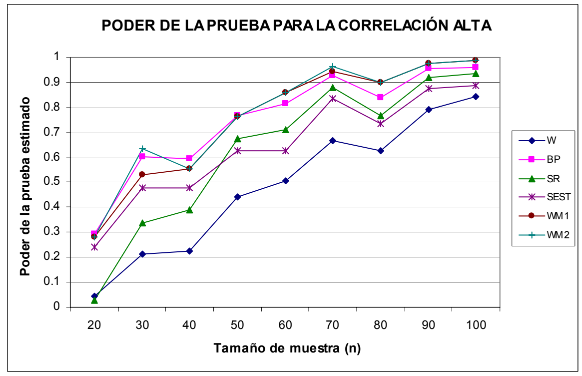


Figura B.2. Corresponde al cuadro 1 con $\lambda = 2.5$

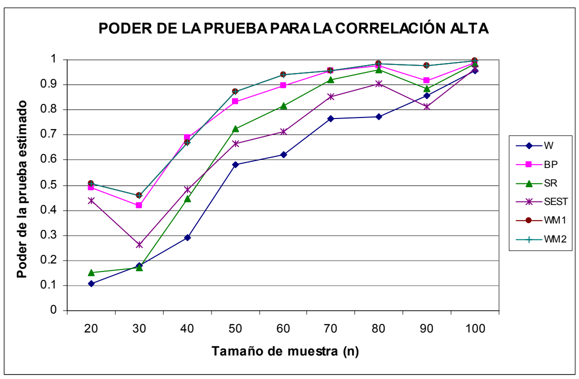


Figura B.3. Corresponde al cuadro 1 con $\lambda = 3$

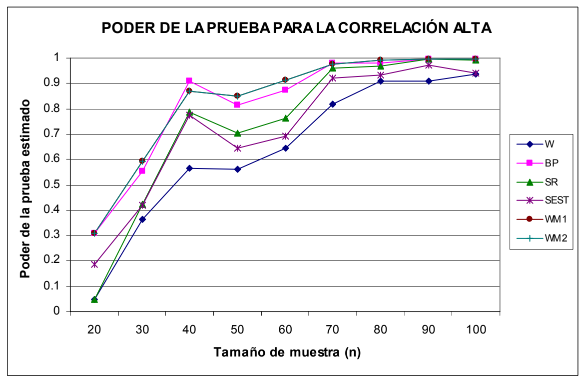


Figura. B.4. Corresponde al cuadro 1 con $\lambda = 3.5$

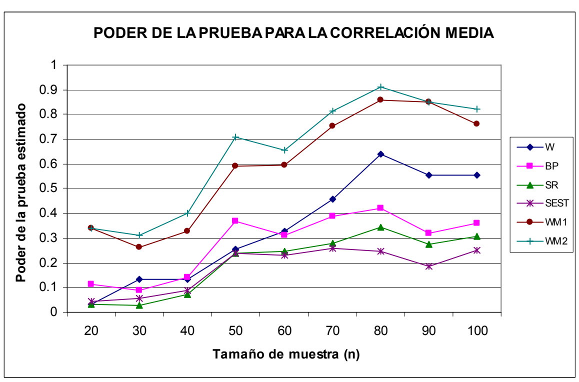


Figura B.5. Corresponde al cuadro 2 con $\lambda = 2$

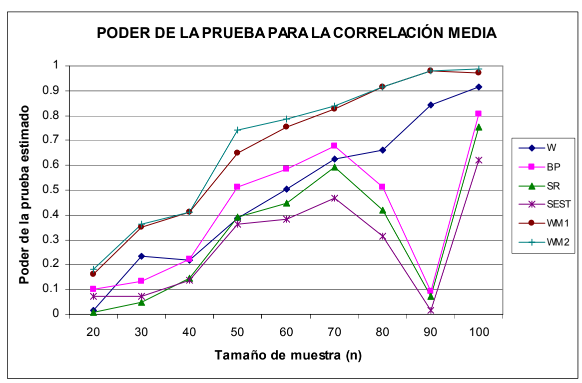


Figura B.6. Corresponde al cuadro 2 con $\lambda = 2.5$

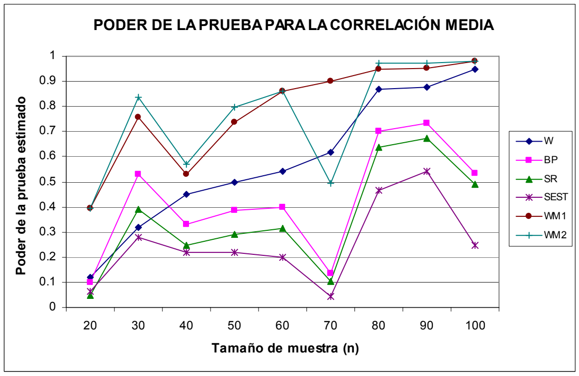


Figura B.7. Corresponde al cuadro 2 con $\lambda = 3$

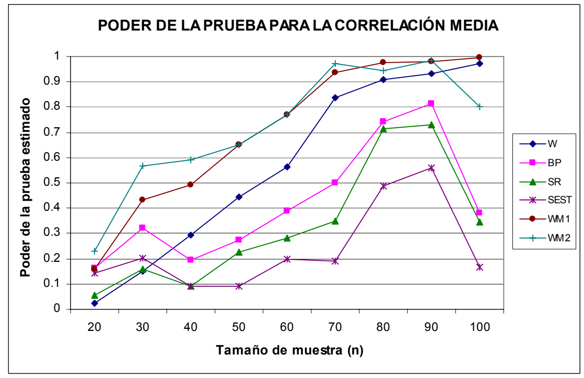


Figura B.8. Corresponde al cuadro 2 con $\lambda = 3.5$

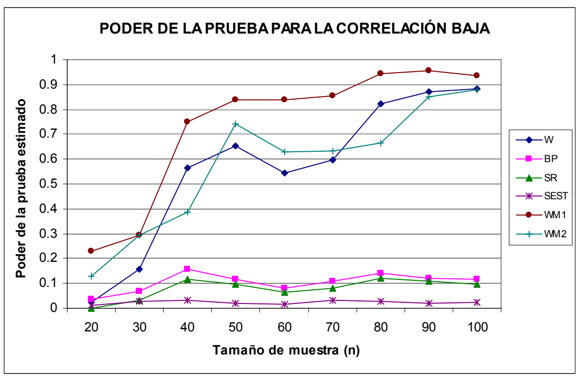


Figura B.9. Corresponde al cuadro 3 con $\lambda = 2$

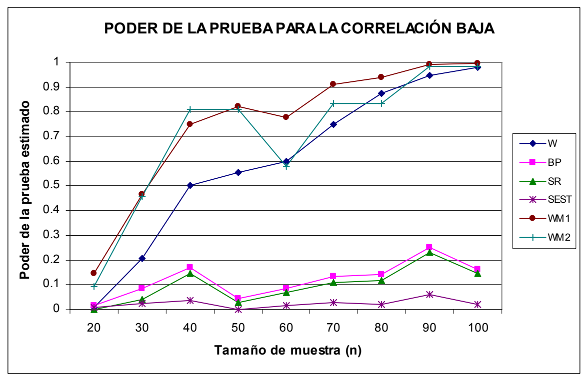


Figura B.10. Corresponde al cuadro 3 con $\lambda = 2.5$

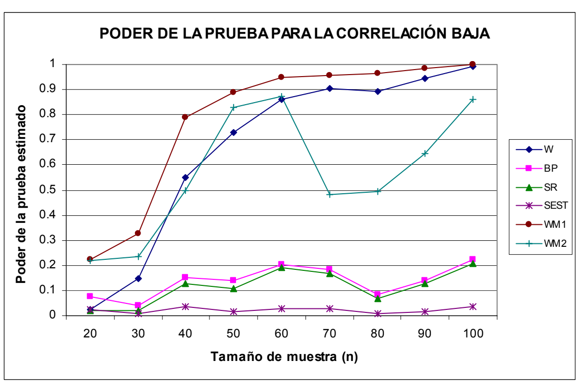


Figura B.11. Corresponde al cuadro 3 con $\lambda = 3$

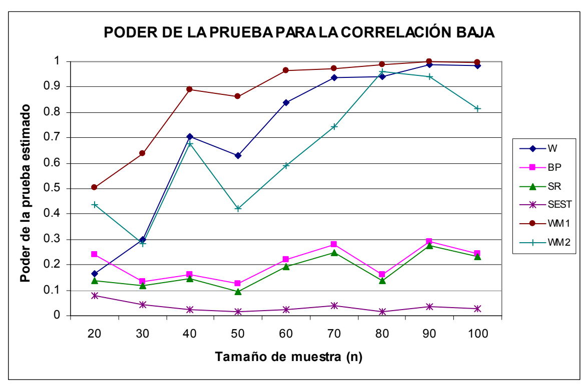


Figura B.12. Corresponde al cuadro 3 con $\lambda = 3.5$

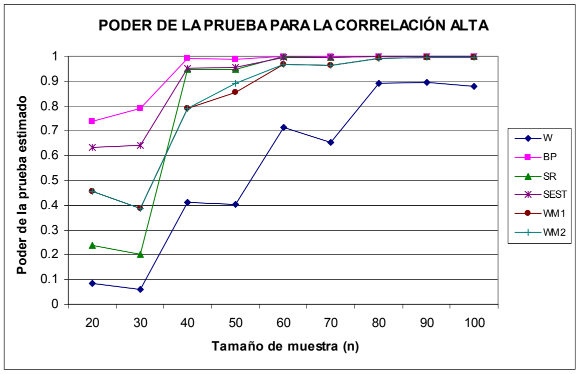


Figura B.13. Corresponde al cuadro 4 con $\lambda = 2$

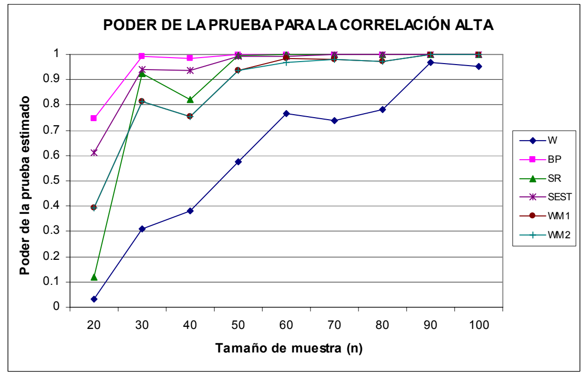


Figura B.14. Corresponde al cuadro 4 con $\lambda = 2.5$

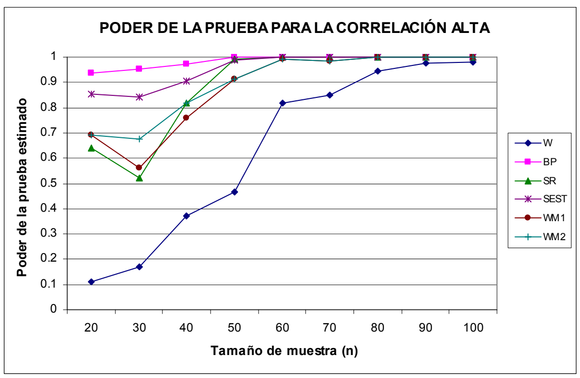


Figura B.15. Corresponde al cuadro 4 con $\lambda = 3$

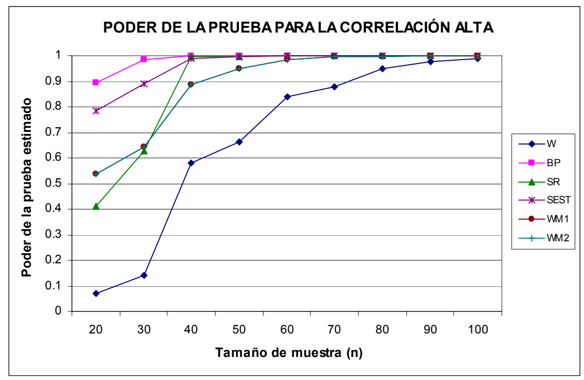


Figura B.16. Corresponde al cuadro 4 con $\lambda = 3.5$

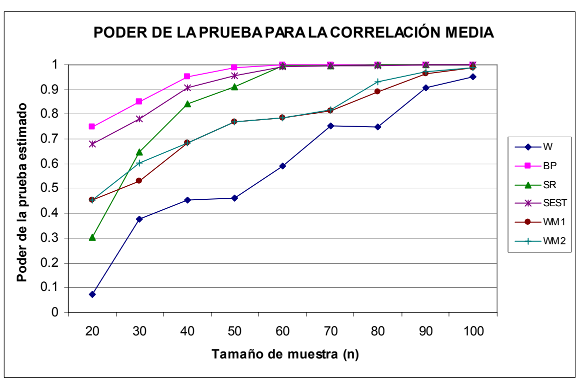


Figura B.17. Corresponde al cuadro 5 con $\lambda = 2$

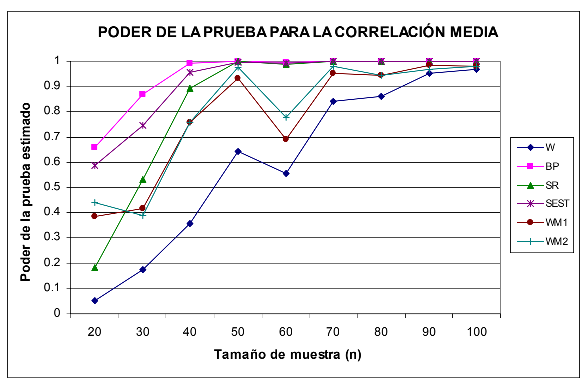


Figura B.18. Corresponde al cuadro 5 con $\lambda = 2.5$

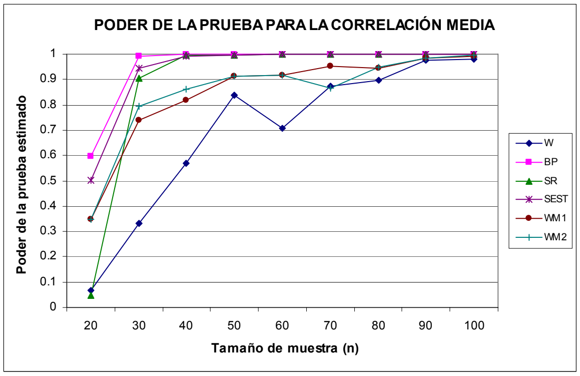


Figura B.19. Corresponde al cuadro 5 con $\lambda = 3$

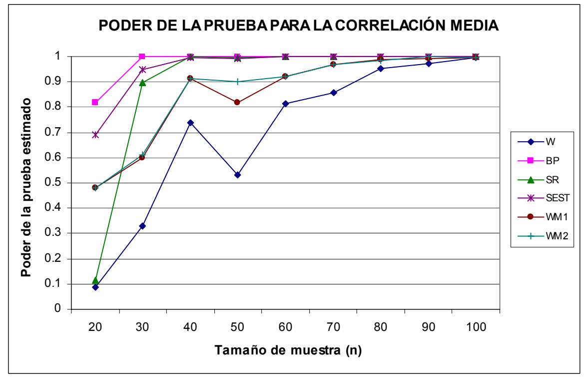


Figura B.20. Corresponde al cuadro 5 con $\lambda = 3.5$

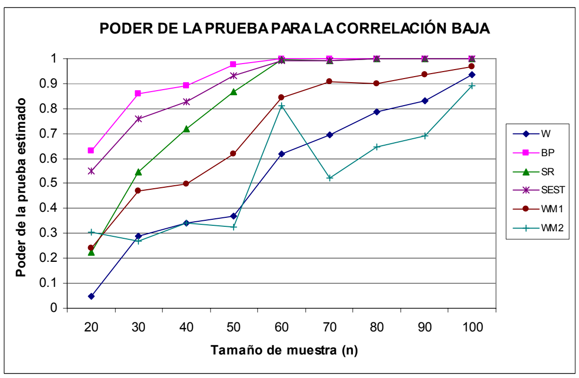


Figura B.21. Corresponde al cuadro 6 con $\lambda = 2$

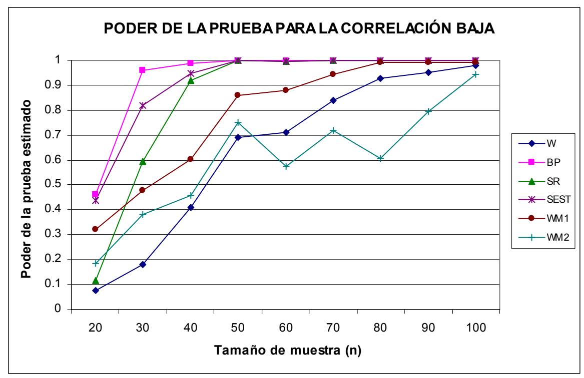


Figura B.22. Corresponde al cuadro 6 con $\lambda = 2.5$

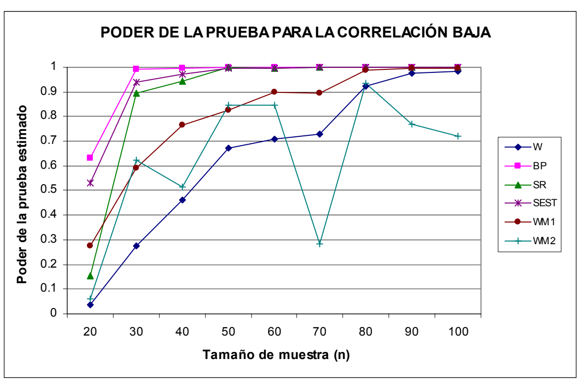


Figura B.23. Corresponde al cuadro 6 con $\lambda = 3$

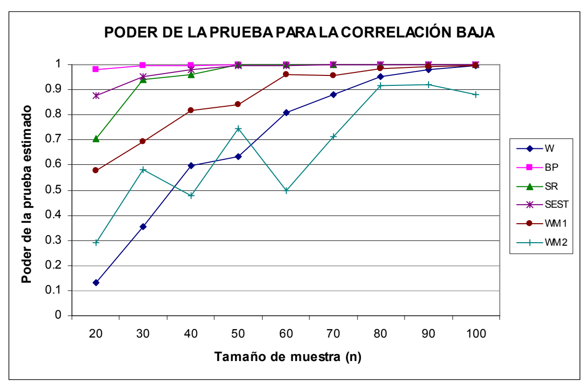


Figura B.24. Corresponde al cuadro 6 con $\lambda = 3.5$

ANEXO C. PROGRAMAS EN R

Programa 1: Rutinas en R para la obtención de las estimaciones del poder y el tamaño de la prueba para una correlación alta entre x_1 y x_2 para el diseño I, II y III.

```
#----------------------------------------------------------------------------------------------------------------------#
#            DISEÑO: I, II Y III                                                                                        #
#            TAMAÑO DE MUESTRA: 20,30,40,50,60,70,80,90 Y 100                                                           #
#            CORRELACIÓN: ALTA                                                                                          #
#            PARAMETROS: ORIENTACION FAVORABLE                                                                          #
#----------------------------------------------------------------------------------------------------------------------#

        # ************************************************************** #
        #                          SIMULACIONES                            #
        # ************************************************************** #

simulacion       <- 1000                # Número total de Simulaciones
columnas         <- 9                   # Número de columnas de la tabla " pruebas.bf "
N                <- 0                   # Tamaño de la muestra
x1               <- 0                   # Variable x1
x2               <- 0                   # Variable x2
x3               <- 0                   # Variable x3
K                <- 0                   # K que permite obtener una correlación alta entre (x1,x2)
contar           <- 0                   # Un contador que comienza en 0
U                <- 0                   # Es la variable uniforme
matrizx          <- 0                   # Es la matriz [x1,x2,x3]
matrizxcon1s     <- 0                   # Son los datos (Intercepto,x1,x2,x3)
x                <- 0                   # Es la matriz [Intercepto,x1,x2,x3]

##### EIGENVALORES DE LA MATRIZ DE CORRELACIÓN #####

correlacion       <- 0                  # Matriz de correlaciones de X´X
eigen.corr        <- 0                  # Eigenvalores de la matriz de correlación
eigenvalor.corre  <- 0                  # Eigenvalores expresados en matriz
criterio.uno      <- 0                  # Proporción de la varianza acumulativa para los componentes

##### ELEMENTOS PARA DEFINIR EL ERROR #####

L                <- 2     # Es valor de L= 2, 2.5, 3 y 3.5
unos.columna     <- 0     # Es un vector de 1's
X                <- 0     # X = x2+x3 para el diseño I, X = x1 para el diseño II ó X = 1 para el diseño III
e1               <- 0     # Es el error

### ================================###
### PRUEBAS PARA BETAS FAVORABLES ###
### ================================###

Y1.bf            <- 0                   # Es el vector de Y´s estimadas obtenido con los betas favorables
res.bf           <- 0                   # Resumen del análisis de varianza del modelo de regresión
residuos.bf      <- 0                   # Es el vector de los residuales con los betas favorables
```

```
y.bf            <- 0                       # Es el vector de los residuales al cuadrado de los betas favorables
r2.mobf         <- 0                       # R^2 del modelo original
pruebas.bf      <- matrix(0,13,columnas, byrow=TRUE, dimnames = list(c("W", "BP", "SR", "SEST",
                   "WM1", "WM2", "WM3", "WM4", "WM5", "WM6", "WM7", "WM8", "WM9"),
                   c("n=20","n=30","n=40","n=50", "n=60","n=70","n=80","n=90","n=100")))
                                           # Es la tabla donde se guardan los resultados de todas las pruebas

# Prueba White #

min.cuad.bf     <- 0                       # Modelo de Regresión
resumen.bf      <- 0                       # Resumen del análisis de varianza del modelo de regresión
r2.bf           <- 0                       # Es el R^2 del modelo de regresión de los Betas favorables
white.bf        <- 0                       # Es el resultado de la prueba de White para betas favorables
ji.w            <- 16.9190                 # Es la ji-cuadrada con 9 gl = 16.9190
p.white_bf      <- matrix(,simulacion,1)   # Se guardan los valores de la Prueba W para un tamaño de muestra
resultado1.bf   <- matrix(,simulacion,1)   # Es una matriz de 0 y 1 que compara todos los valores obtenidos
resultados_white <- matrix(0,1,9)          # Se guardan los valores de la Prueba W para 9 tamaños de muestra

# Prueba Breusch-Pagan #

sigma.cuad_bf   <- 0                       # Es el valor de sigma^2
p.bf            <- 0                       # p = y/sigma cuadrada
modelo.bf       <- 0                       # Modelo de regresión con los betas favorables
resum.bf        <- 0                       # Resumen del análisis de varianza del modelo de regresión
coefi.bf        <- 0                       # Los coeficientes del análisis de varianza
yest.bf         <- 0                       # y's estimadas
scr.bf          <- 0                       # Es la SCR para cada muestra de los betas favorables
theta.bf        <- 0                       # Es el valor de theta para cada muestra de los betas favorables
ji.b            <- 3.84146                 # Es la ji-cuadrada con 1 gl = 3.84146
p.breusch_bf    <- matrix(,simulacion,1)   # Se guardan los valores de la Prueba BP para un tamaño de muestra
resultado2.bf   <- matrix(,simulacion,1)   # Es una matriz de 0 y 1 que compara todos los valores obtenidos
resultados_bp   <- matrix(0,1,9)           # Se guardan los valores de la Prueba BP para 9 tamaños de muestra

# Prueba de Score de residuales #

d.bf            <- 0                       # Es el vector de residuales^2 del modelo completo
sigma0.2_bf     <- 0                       # Es una parte del estadístico para cada muestra obtenida
sigmai          <- 0                       # Es el producto de z'a
dimension       <- 0                       # Es el tamaño de la muestra
sigma1.1        <- 0                       # Convierte la matriz sigmai en una matriz cuadrada
sigma1          <- 0                       # Es la inversa de sigma1.1
sigma1.2        <- 0                       # Es la matriz sigmai cuadrada
sigma2          <- 0                       # Es la inversa de sigma1.2
g               <- 0                       # Es el producto de algunas matrices
H               <- 0                       # Es la matriz H
V0              <- 0                       # Es el producto de matrices por el modelo reducido
Z.bf            <- 0                       # Es la matriz de N X 2
SR.bf           <- 0                       # Es el resultado de la prueba para cada muestra obtenida
ji.ps           <- 3.84146                 # Es la ji-cuadrada con 1 gl = 3.84146
p.residuales_bf <- matrix(,simulacion,1)   # Se guardan los valores de la Prueba SR para un tamaño de muestra
resultado3.bf   <- matrix(,simulacion,1)   # Es una matriz de 0 y 1 que compara los valores obtenidos
resultados_sr   <- matrix(0,1,9)           # Se guardan los valores de la Prueba SR para 9 tamaños de muestra

# Prueba de Score Estudentizado #

S.bf            <- 0                       # Es el resultado de la prueba de Score para cada muestra obtenida
```

```
S.STUD_bf        <- 0                        # Es el resultado de la prueba de SEst para cada muestra obtenida
p.scorest_bf     <- matrix(,simulacion,1)    # Se guardan los valores de la Prueba SEst para un tamaño de
muestra
resultado4.bf    <- matrix(,simulacion,1)    # Es una matriz de 0 y 1 que compara todos los valores obtenidos
resultados_sestud<- matrix(0,1,9)            # Se guardan los valores de la Prueba SEst para 9 tamaños de
muestra

# Prueba White Modificada (Componentes Principales con el Método de Bonferroni) #

datos_cp             <- 0                      # La desviación estándar de los 9 componentes principales
compon.bf            <- 0                      # Es el modelo de los residuales^2 del modelo completo con
                                               # todos los componentes principales
coeficientes_cp.bf   <- matrix(,9,1)           # Son los coeficientes para cada muestra
alfa                 <- 0.05                   # Es el valor de Alfa
alfa1                <- 0                      # Valor que se utiliza para comparar el método de Bonferroni
componentes.prin_bf  <- matrix(0,9,9)          # Es el resultado de la prueba de Componentes Principales
todas.una_bf         <- matrix(0,9,1)          # Es una matriz donde se guarda los valores de la Prueba de
                                               # CP utilizando el método de Bonferroni
CP                   <- matrix(0,13,columnas)  # Es una matriz donde se van guardando uno a uno los CP
resultado5.bf        <- matrix(,9,1)           # Es una matriz de 0 y 1 que compara los valores obtenidos
b                    <- matrix(0,9,9)

for (k in 1:columnas)
{
        N <- (k+1)*10                          # Tamaño de la muestra
        x3 <- runif(N,min=0,max=1)             # La variable x3 uniforme
        corr12 <- 0                            # La correlación de (x1,x2) = 0
        constante <- 1                         # El valor de Constantes =1

        ### CORRELACIÓN ALTA ###

        while (corr12<=0.95)
        {
                K <- 5/constante
                x1 <- runif(N,min=0,max=1)
                U  <- runif(N,min=0,max=1)
                x2 <- x1+(U*K)
                corr12 <- cor(x1,x2)
                constante <- constante+1
                contar <- contar+1
        }
        matrizx <- cbind(x1,x2,x3)
        Intercepto <- rep(1, N)
        matrizxcon1s <- cbind(Intercepto,x1,x2,x3)
        x <- as.matrix(matrizxcon1s)

        # *********************************************************************** #
        # BETAS ESTIMADOS POR MEDIO DE LOS EIGENVECTORES: Orientación favorable #
        # *********************************************************************** #

        XTX <- t(x)%*%x
        eigenvalores <- eigen(XTX)
        eigenvectores <- eigenvalores$vectors
        betas.favorables <- eigenvectores[,1]
        betas.desfavor <- eigenvectores[,4]
```

```
# NUEVAS VARIABLES #

x4<- (x1)^2
x5<- (x2)^2
x6<- (x3)^2
x7<- (x1)*(x2)
x8<- (x1)*(x3)
x9<- (x2)*(x3)
datos.x <- cbind(x1,x2,x3,x4,x5,x6,x7,x8,x9)

# EIGENVALORES DE LA MATRIZ DE CORRELACIONES Y PRIMER CRITERIO DE LOS
COMPONENTES PRINCIPALES #

correlacion <- cor(t(datos.x)%*%datos.x)
eigen.corr <- eigen(correlacion, only.values=TRUE)
eigenvalor.corre <- eigen.corr$values
HTML(eigenvalor.corre,file="G:/Correlacion_Alta/cac2x135.html")

# SEGUNDO CRITERIO DE LOS COMPONENTES PRINCIPALES #

todas.una_bf <- matrix(0,9,1)
criterio.uno <- summary(datos_cp<-princomp(datos.x,cor=TRUE))
HTML(criterio.uno,file="G:/Correlacion_Alta/cac1x135.html")

##### PRUEBAS #####

unos.columna <- matrix(1,N,1)
X <- x2+x3 # X = x1 ó X = matrix(1,N,1)

for (i in 1:simulacion)
{
	e1 <- rnorm(N,0,(sqrt(X^L)))

	# ********************************************************************** #
	#                          MODELO DE REGRESIÓN LINEAL                          #
	# ********************************************************************** #

	betas.favorables
	Y1.bf <-(x%*%betas.favorables)+ e1
	analisis.bf <- lm(Y1.bf~x1+x2+x3)
	coeficientes.bf <- analisis.bf$coefficients
	matriz.coef_bf <- as.matrix(coeficientes.bf)
	res.bf <-summary(analisis.bf)
	residuos.bf <- res.bf$residuals
	y.bf <- residuos.bf^2
	datos.x
	r2.mobf <- res.bf$r.squared

	### PRUEBA WHITE ###
	min.cuad.bf <- lm(y.bf~x1+x2+x3+x4+x5+x6+x7+x8+x9)
	resumen.bf <- summary(min.cuad.bf)
	r2.bf <-resumen.bf$r.squared
	white.bf <- N*r2.bf
	resultado1.bf[i,] <- white.bf
	if (white.bf > ji.w)
		p.white_bf[i,] <- 1			# 1 = Heterogeneidad
```

```
else
        p.white_bf[i,] <- 0                    # 0 = Homogeneidad

### PRUEBA BREUSCH-PAGAN ###
sigma.cuad_bf <- (sum(y.bf))/(N)
p.bf <- y.bf/sigma.cuad_bf
modelo.bf <- lm(p.bf~x1)
resum.bf <- summary(modelo.bf)
coefi.bf <- resum.bf$coefficients[,1]
x.uno <- x[,1:2]
yest.bf <- (x.uno)%*%coefi.bf
ymedia.bf <- mean(p.bf)
scr.bf <- sum((yest.bf-ymedia.bf)^2)
theta.bf <- (scr.bf)/2
resultado2.bf[i,]<- theta.bf
if (theta.bf>ji.b)
        p.breusch_bf[i,] <- 1                  # 1 = Heterogeneidad
else
        p.breusch_bf[i,] <- 0                  # 0 = Homogeneidad

### PRUEBA DE SCORE DE RESIDUALES ###
d.bf <- y.bf
P <- 4
sigmai <- X^L
sigma0.2_bf <- (N/(N-P))*(sigma.cuad_bf)
dimension <- length(sigmai)
sigma1.1 <- diag(sigmai,ncol=dimension)
sigma1 <- solve(sigma1.1)
sigma1.2 <- diag((sigmai^(1/2)),ncol=dimension)
sigma2 <- solve(sigma1.2)
g <- (t(x)%*%sigma1%*%x)
H <- diag(sigma2%*%x%*%solve(g)%*%t(x)%*%sigma2)
V0 <- x%*%solve(t(x)%*%x)%*%t(x)
Z.bf <- cbind(Intercepto,x1)
SR.bf <- (1/2)*(t((d.bf/sigma0.2_bf)-unos.columna + H)%*%Z.bf%*%solve(t(Z.bf)
 %*%V0%*%Z.bf)%*%t(Z.bf)%*%((d.bf/sigma0.2_bf)-unos.columna + H))
resultado3.bf[i,] <- SR.bf
if (SR.bf > ji.ps)
        p.residuales_bf[i,] <- 1                          # 1 = Heterogeneidad
else
        p.residuales_bf[i,] <- 0                          # 0 = Homogeneidad

### PRUEBA DE SCORE ESTUDENTIZADO ###
S.bf <- (1/2)*(t((d.bf/sigma.cuad_bf)-unos.columna)%*%Z.bf%*%
 solve(t(Z.bf)%*%Z.bf)%*%t(Z.bf)%*%((d.bf/sigma.cuad_bf)-unos.columna))
S.STUD_bf <- (2*N)*(S.bf/(t((d.bf/sigma.cuad_bf)-unos.columna)%*%
 ((d.bf/sigma.cuad_bf)-unos.columna)))
resultado4.bf[i,] <- S.STUD_bf
if (S.STUD_bf > ji.ps)
        p.scorest_bf[i,] <- 1                        # 1 = Heterogeneidad
else
        p.scorest_bf[i,] <- 0                        # 0 = Homogeneidad

### PRUEBA WHITE MODIFICADA (CP1-CP9 aplicando el método de Bonferroni) ###

compon.bf <- summary(lm(y.bf~datos_cp$scores[,1]+datos_cp$scores[,2]+
```

```
                    datos_cp$scores[,3]+datos_cp$scores[,4]+datos_cp$scores[,5]+datos_cp$scores[,6]+
                    datos_cp$scores[,7]+datos_cp$scores[,8]+datos_cp$scores[,9]))
                    coeficientes_cp.bf <-compon.bf$coefficients[2:10,4]
                    coeficientes_cp.bf <-as.matrix(coeficientes_cp.bf)
                    for (u in 1:9)
                    {
                              for(v in 1:u)
                              {
                                        alfa1<-alfa/u                  # Valor del alfa1 de acuerdo al número de CP
                                        if (coeficientes_cp.bf[v,1]<alfa1)
                                                  componentes.prin_bf[v,u] <-1              # 1 = Heterogeneidad
                                        else
                                                  componentes.prin_bf[v,u] <-0              # 0 = Homogeneidad
                                        if (sum(componentes.prin_bf[,u])>0)
                                                  resultado5.bf[u,1]<-1
                                        else
                                                  resultado5.bf[u,1]<-0
                              }
                    }
                    todas.una_bf<- todas.una_bf+resultado5.bf
                    resultados_white[,k] <- apply(p.white_bf,2,sum)
                    resultados_bp[,k] <- apply(p.breusch_bf,2,sum)
                    resultados_sr[,k] <- apply(p.residuales_bf,2,sum)
                    resultados_sestud[,k] <- apply(p.scorest_bf,2,sum)
          }
          CP[5:13,k] <- CP[5:13,k]+todas.una_bf
          pruebas.bf[5:13,] <- (1/simulacion)*(CP[5:13,])
          b[,k]<- coeficientes_cp.bf
}
WHITE <- (1/simulacion)* resultados_white
BP <- (1/simulacion)* resultados_bp
SR <- (1/simulacion)* resultados_sr
SEST <- (1/simulacion)* resultados_sestud
pruebas.bf[1,] <- WHITE
pruebas.bf[2,] <- BP
pruebas.bf[3,] <- SR
pruebas.bf[4,] <- SEST
pruebas.bf
```

Programa 2: Rutinas en R para la obtención de las estimaciones del poder y el tamaño de la prueba para una correlación media entre x_1 y x_2 para el diseño I, II y III.

```
#----------------------------------------------------------------------------------------------------------------#
#            DISEÑO: I, II Y III                                                                                 #
#            TAMAÑO DE MUESTRA: 20,30,40,50,60,70,80,90 Y 100                                                    #
#            CORRELACIÓN: MEDIA                                                                                  #
#            PARAMETROS: ORIENTACION FAVORABLE                                                                   #
#----------------------------------------------------------------------------------------------------------------#

         # ************************************************************** #
         #                          SIMULACIONES                          #
         # ************************************************************** #

simulacion       <- 1000          # Número total de Simulaciones
columnas         <- 9             # Número de columnas de la tabla " pruebas.bf "
N                <- 0             # Tamaño de la muestra
x1               <- 0             # Variable x1
x2               <- 0             # Variable x2
x3               <- 0             # Variable x3
K                <- 0             # K permite obtener una correlación media entre (x1,x2)
J                <- 0             # J permite obtener una correlación media entre (x1,x2)
contar           <- 0             # Un contador que comienza en 0
U                <- 0             # Es la variable uniforme
matrizx          <- 0             # Es la matriz [x1,x2,x3]
matrizxcon1s     <- 0             # Son los datos (Intercepto,x1,x2,x3)
x                <- 0             # Es la matriz [Intercepto,x1,x2,x3]

##### EIGENVALORES DE LA MATRIZ DE CORRELACIÓN #####

correlacion      <- 0             # Matriz de correlaciones de X´X
eigen.corr       <- 0             # Eigenvalores de la matriz de correlación
eigenvalor.corre <- 0             # Eigenvalores expresados en matriz
criterio.uno     <- 0             # Proporción de la varianza acumulativa para los componentes

##### ELEMENTOS PARA DEFINIR EL ERROR #####

L                <- 2     # Es valor de L= 2, 2.5, 3 y 3.5
unos.columna     <- 0     # Es un vector de 1's
X                <- 0     # X = x2+x3 para el diseño I, X = x1 para el diseño II ó X = 1 para el diseño III
e1               <- 0     # Es el error

### ================================###
### PRUEBAS PARA BETAS FAVORABLES ###
### ================================###

Y1.bf            <- 0             # Es el vector de Y´s estimadas obtenido con los betas favorables
res.bf           <- 0             # Resumen del análisis de varianza del modelo de regresión
residuos.bf      <- 0             # Es el vector de los residuales con los betas favorables
y.bf             <- 0             # Es el vector de los residuales al cuadrado de los betas favorables
r2.mobf          <- 0             # R^2 del modelo original
```

```
pruebas.bf          <- matrix(0,13,columnas, byrow=TRUE, dimnames = list(c("W", "BP", "SR", "SEST",
                        "WM1", "WM2", "WM3", "WM4", "WM5", "WM6", "WM7", "WM8", "WM9"),
                        c("n=20","n=30","n=40","n=50", "n=60","n=70","n=80","n=90","n=100")))
                                        # Es la tabla donde se guardan los resultados de todas las pruebas

# Prueba White #

min.cuad.bf         <- 0                # Modelo de Regresión
resumen.bf          <- 0                # Resumen del análisis de varianza del modelo de regresión
r2.bf               <- 0                # Es el R^2 del modelo de regresión de los Betas favorables
white.bf            <- 0                # Es el resultado de la prueba de White para betas favorables
ji.w                <- 16.9190          # Es la ji-cuadrada con 9 gl = 16.9190
p.white_bf          <- matrix(,simulacion,1)  # Se guardan los valores de la Prueba W para un tamaño de muestra
resultado1.bf       <- matrix(,simulacion,1)  # Es una matriz de 0 y 1 que compara todos los valores obtenidos
resultados_white    <- matrix(0,1,9)    # Se guardan los valores de la Prueba W para 9 tamaños de muestra

# Prueba Breusch-Pagan #

sigma.cuad_bf       <- 0                # Es el valor de sigma^2
p.bf                <- 0                # p = y/sigma cuadrada
modelo.bf           <- 0                # Modelo de regresión con los betas favorables
resum.bf            <- 0                # Resumen del análisis de varianza del modelo de regresión
coefi.bf            <- 0                # Los coeficientes del análisis de varianza
yest.bf             <- 0                # y's estimadas
scr.bf              <- 0                # Es la SCR para cada muestra de los betas favorables
theta.bf            <- 0                # Es el valor de theta para cada muestra de los betas favorables
ji.b                <- 3.84146          # Es la ji-cuadrada con 1 gl = 3.84146
p.breusch_bf        <- matrix(,simulacion,1)  # Se guardan los valores de la Prueba BP para un tamaño de muestra
resultado2.bf       <- matrix(,simulacion,1)  # Es una matriz de 0 y 1 que compara todos los valores obtenidos
resultados_bp       <- matrix(0,1,9)    # Se guardan los valores de la Prueba BP para 9 tamaños de muestra

# Prueba de Score de residuales #

d.bf                <- 0                # Es el vector de residuales^2 del modelo completo
sigma0.2_bf         <- 0                # Es una parte del estadístico para cada muestra obtenida
sigmai              <- 0                # Es el producto de z'a
dimension           <- 0                # Es el tamaño de la muestra
sigma1.1            <- 0                # Convierte la matriz sigmai en una matriz cuadrada
sigma1              <- 0                # Es la inversa de sigma1.1
sigma1.2            <- 0                # Es la matriz sigmai cuadrada
sigma2              <- 0                # Es la inversa de sigma1.2
g                   <- 0                # Es el producto de algunas matrices
H                   <- 0                # Es la matriz H
V0                  <- 0                # Es el producto de matrices por el modelo reducido
Z.bf                <- 0                # Es la matriz de N X 2
SR.bf               <- 0                # Es el resultado de la prueba para cada muestra obtenida
ji.ps               <- 3.84146          # Es la ji-cuadrada con 1 gl = 3.84146
p.residuales_bf     <- matrix(,simulacion,1)  # Se guardan los valores de la Prueba SR para un tamaño de muestra
resultado3.bf       <- matrix(,simulacion,1)  # Es una matriz de 0 y 1 que compara los valores obtenidos
resultados_sr       <- matrix(0,1,9)    # Se guardan los valores de la Prueba SR para 9 tamaños de muestra

# Prueba de Score Estudentizado #

S.bf                <- 0                # Es el resultado de la prueba de Score para cada muestra obtenida
S.STUD_bf           <- 0                # Es el resultado de la prueba de SEst para cada muestra obtenida
```

```
p.scorest_bf        <- matrix(,simulacion,1)   # Se guardan los valores de la Prueba SEst para un tamaño de
muestra
resultado4.bf       <- matrix(,simulacion,1)   # Es una matriz de 0 y 1 que compara todos los valores obtenidos
resultados_sestud<- matrix(0,1,9)              # Se guardan los valores de la Prueba SEst para 9 tamaños de
muestra

# Prueba White Modificada (Componentes Principales con el Método de Bonferroni) #

datos_cp                <- 0                    # La desviación estándar de los 9 componentes principales
compon.bf               <- 0                    # Es el modelo de los residuales^2 del modelo completo con
                                                # todos los componentes principales
coeficientes_cp.bf      <- matrix(,9,1)         # Son los coeficientes para cada muestra
alfa                    <- 0.05                 # Es el valor de Alfa
alfa1                   <- 0                    # Valor que se utiliza para comparar el método de Bonferroni
componentes.prin_bf     <- matrix(0,9,9)        # Es el resultado de la prueba de Componentes Principales
todas.una_bf            <- matrix(0,9,1)        # Es una matriz donde se guarda los valores de la Prueba de
                                                # CP utilizando el método de Bonferroni
CP                      <- matrix(0,13,columnas) # Es una matriz donde se van guardando uno a uno los CP
resultado5.bf           <- matrix(,9,1)         # Es una matriz de 0 y 1 que compara los valores obtenidos
b                       <- matrix(0,9,9)

for (k in 1:columnas)
{
        N <- (k+1)*10                           # Tamaño de la muestra
        x3 <- runif(N,min=0,max=1)              # La variable x3 uniforme
        corr12 <- 0                             # La correlación de (x1,x2) = 0
        constante1 <- 1                         # El valor de Constantes =1
        constante2 <- 1

        ### CORRELACIÓN MEDIA ###

        while (0.45>corr12 | corr12>0.55)
        {
                K <- 5/constante1
                J <- 2/constante2
                x1 <- runif(N,min=0,max=1)
                U  <- runif(N,min=0,max=1)
                x2 <- (x1*J)+(U*K)
                corr12 <- cor(x1,x2)
                constante1 <- constante1+1
                constante2 <- constante2+1
                contar <- contar+1
        }
        matrizx <- cbind(x1,x2,x3)
        Intercepto <- rep(1, N)
        matrizxcon1s <- cbind(Intercepto,x1,x2,x3)
        x <- as.matrix(matrizxcon1s)

        # ************************************************************************ #
        # BETAS ESTIMADOS POR MEDIO DE LOS EIGENVECTORES: Orientación favorable #
        # ************************************************************************ #

        XTX <- t(x)%*%x
        eigenvalores <- eigen(XTX)
        eigenvectores <- eigenvalores$vectors
        betas.favorables <- eigenvectores[,1]
```

```
betas.desfavor <- eigenvectores[,4]

# NUEVAS VARIABLES #

x4<- (x1)^2
x5<- (x2)^2
x6<- (x3)^2
x7<- (x1)*(x2)
x8<- (x1)*(x3)
x9<- (x2)*(x3)
datos.x <- cbind(x1,x2,x3,x4,x5,x6,x7,x8,x9)

# EIGENVALORES DE LA MATRIZ DE CORRELACIONES Y PRIMER CRITERIO DE LOS
COMPONENTES PRINCIPALES #

correlacion <- cor(t(datos.x)%*%datos.x)
eigen.corr <- eigen(correlacion, only.values=TRUE)
eigenvalor.corre <- eigen.corr$values
HTML(eigenvalor.corre,file="G:/Correlacion_Media/cmc2x135.html")

# SEGUNDO CRITERIO DE LOS COMPONENTES PRINCIPALES #

todas.una_bf <- matrix(0,9,1)
criterio.uno <- summary(datos_cp<-princomp(datos.x,cor=TRUE))
HTML(criterio.uno,file="G:/Correlacion_Media/cmc1x135.html")

##### PRUEBAS #####

unos.columna <- matrix(1,N,1)
X <- x2+x3 # X = x1 ó X = matrix(1,N,1)

for (i in 1:simulacion)
{
        e1 <- rnorm(N,0,(sqrt(X^L)))

        # ****************************************************************** #
        #                    MODELO DE REGRESIÓN LINEAL                      #
        # ****************************************************************** #

        betas.favorables
        Y1.bf <-(x%*%betas.favorables)+ e1
        analisis.bf <- lm(Y1.bf~x1+x2+x3)
        coeficientes.bf <- analisis.bf$coefficients
        matriz.coef_bf <- as.matrix(coeficientes.bf)
        res.bf <-summary(analisis.bf)
        residuos.bf <- res.bf$residuals
        y.bf <- residuos.bf^2
        datos.x
        r2.mobf <- res.bf$r.squared

        ### PRUEBA WHITE ###
        min.cuad.bf <- lm(y.bf~x1+x2+x3+x4+x5+x6+x7+x8+x9)
        resumen.bf <- summary(min.cuad.bf)
        r2.bf <-resumen.bf$r.squared
        white.bf <- N*r2.bf
        resultado1.bf[i,] <- white.bf
```

```
if (white.bf > ji.w)
        p.white_bf[i,] <- 1                    # 1 = Heterogeneidad
else
        p.white_bf[i,] <- 0                    # 0 = Homogeneidad

### PRUEBA BREUSCH-PAGAN ###
sigma.cuad_bf <- (sum(y.bf))/(N)
p.bf <- y.bf/sigma.cuad_bf
modelo.bf <- lm(p.bf~x1)
resum.bf <- summary(modelo.bf)
coefi.bf <- resum.bf$coefficients[,1]
x.uno <- x[,1:2]
yest.bf <- (x.uno)%*%coefi.bf
ymedia.bf <- mean(p.bf)
scr.bf <- sum((yest.bf-ymedia.bf)^2)
theta.bf <- (scr.bf)/2
resultado2.bf[i,]<- theta.bf
if (theta.bf>ji.b)
        p.breusch_bf[i,] <- 1                  # 1 = Heterogeneidad
else
        p.breusch_bf[i,] <- 0                  # 0 = Homogeneidad

### PRUEBA DE SCORE DE RESIDUALES ###
d.bf <- y.bf
P <- 4
sigmai <- X^L
sigma0.2_bf <- (N/(N-P))*(sigma.cuad_bf)
dimension <- length(sigmai)
sigma1.1 <- diag(sigmai,ncol=dimension)
sigma1 <- solve(sigma1.1)
sigma1.2 <- diag((sigmai^(1/2)),ncol=dimension)
sigma2 <- solve(sigma1.2)
g <- (t(x)%*%sigma1%*%x)
H <- diag(sigma2%*%x%*%solve(g)%*%t(x)%*%sigma2)
V0 <- x%*%solve(t(x)%*%x)%*%t(x)
Z.bf <- cbind(Intercepto,x1)
SR.bf <- (1/2)*(t((d.bf/sigma0.2_bf)-unos.columna + H)%*%Z.bf%*%solve(t(Z.bf)
 %*%V0%*%Z.bf)%*%t(Z.bf)%*%((d.bf/sigma0.2_bf)-unos.columna + H))
resultado3.bf[i,] <- SR.bf
if (SR.bf > ji.ps)
        p.residuales_bf[i,] <- 1                              # 1 = Heterogeneidad
else
        p.residuales_bf[i,] <- 0                              # 0 = Homogeneidad

### PRUEBA DE SCORE ESTUDENTIZADO ###
S.bf <- (1/2)*(t((d.bf/sigma.cuad_bf)-unos.columna)%*%Z.bf%*%
 solve(t(Z.bf)%*%Z.bf)%*%t(Z.bf)%*%((d.bf/sigma.cuad_bf)-unos.columna))
S.STUD_bf <- (2*N)*(S.bf/(t((d.bf/sigma.cuad_bf)-unos.columna)%*%
 ((d.bf/sigma.cuad_bf)-unos.columna)))
resultado4.bf[i,] <- S.STUD_bf
if (S.STUD_bf > ji.ps)
        p.scorest_bf[i,] <- 1                        # 1 = Heterogeneidad
else
        p.scorest_bf[i,] <- 0                        # 0 = Homogeneidad
```

```
        ### PRUEBA WHITE MODIFICADA (CP1-CP9 aplicando el método de Bonferroni) ###

        compon.bf <- summary(lm(y.bf~datos_cp$scores[,1]+datos_cp$scores[,2]+
         datos_cp$scores[,3]+datos_cp$scores[,4]+datos_cp$scores[,5]+datos_cp$scores[,6]+
         datos_cp$scores[,7]+datos_cp$scores[,8]+datos_cp$scores[,9]))
        coeficientes_cp.bf <-compon.bf$coefficients[2:10,4]
        coeficientes_cp.bf <-as.matrix(coeficientes_cp.bf)
        for (u in 1:9)
        {
            for(v in 1:u)
            {
                alfa1<-alfa/u          # Valor del alfa1 de acuerdo al número de CP
                if (coeficientes_cp.bf[v,1]<alfa1)
                    componentes.prin_bf[v,u] <-1          # 1 = Heterogeneidad
                else
                    componentes.prin_bf[v,u] <-0          # 0 = Homogeneidad
                if (sum(componentes.prin_bf[,u])>0)
                    resultado5.bf[u,1]<-1
                else
                    resultado5.bf[u,1]<-0
            }
        }
        todas.una_bf<- todas.una_bf+resultado5.bf
        resultados_white[,k] <- apply(p.white_bf,2,sum)
        resultados_bp[,k] <- apply(p.breusch_bf,2,sum)
        resultados_sr[,k] <- apply(p.residuales_bf,2,sum)
        resultados_sestud[,k] <- apply(p.scorest_bf,2,sum)
    }
    CP[5:13,k] <- CP[5:13,k]+todas.una_bf
    pruebas.bf[5:13,] <- (1/simulacion)*(CP[5:13,])
    b[,k]<- coeficientes_cp.bf
}
WHITE <- (1/simulacion)* resultados_white
BP <- (1/simulacion)* resultados_bp
SR <- (1/simulacion)* resultados_sr
SEST <- (1/simulacion)* resultados_sestud
pruebas.bf[1,] <- WHITE
pruebas.bf[2,] <- BP
pruebas.bf[3,] <- SR
pruebas.bf[4,] <- SEST
pruebas.bf
```

Programa 3: Rutinas en R para la obtención de las estimaciones del poder y el tamaño de la prueba para una correlación baja entre x_1 y x_2 para el diseño I, II y III.

```
#-------------------------------------------------------------------------------------------------------------#
#            DISEÑO: I, II Y III                                                                              #
#            TAMAÑO DE MUESTRA: 20,30,40,50,60,70,80,90 Y 100                                                 #
#            CORRELACIÓN: BAJA                                                                                #
#            PARAMETROS: ORIENTACION FAVORABLE                                                                #
#-------------------------------------------------------------------------------------------------------------#

        # ************************************************************ #
        #                        SIMULACIONES                          #
        # ************************************************************ #

simulacion      <- 1000              # Número total de Simulaciones
columnas        <- 9                 # Número de columnas de la tabla " pruebas.bf "
N               <- 0                 # Tamaño de la muestra
x1              <- 0                 # Variable x1
x2              <- 0                 # Variable x2
x3              <- 0                 # Variable x3
K               <- 0                 # K que permite obtener una correlación baja entre (x1,x2)
contar          <- 0                 # Un contador que comienza en 0
U               <- 0                 # Es la variable uniforme
matrizx         <- 0                 # Es la matriz [x1,x2,x3]
matrizxcon1s    <- 0                 # Son los datos (Intercepto,x1,x2,x3)
x               <- 0                 # Es la matriz [Intercepto,x1,x2,x3]

##### EIGENVALORES DE LA MATRIZ DE CORRELACIÓN #####

correlacion       <- 0               # Matriz de correlaciones de X´X
eigen.corr        <- 0               # Eigenvalores de la matriz de correlación
eigenvalor.corre  <- 0               # Eigenvalores expresados en matriz
criterio.uno      <- 0               # Proporción de la varianza acumulativa para los componentes

##### ELEMENTOS PARA DEFINIR EL ERROR #####

L               <- 2       # Es valor de L= 2, 2.5, 3 y 3.5
unos.columna    <- 0       # Es un vector de 1's
X               <- 0       # X = x2+x3 para el diseño I, X = x1 para el diseño II ó X = 1 para el diseño III
e1              <- 0       # Es el error

### ================================###
### PRUEBAS PARA BETAS FAVORABLES ###
### ================================###

Y1.bf           <- 0                 # Es el vector de Y´s estimadas obtenido con los betas favorables
res.bf          <- 0                 # Resumen del análisis de varianza del modelo de regresión
residuos.bf     <- 0                 # Es el vector de los residuales con los betas favorables
y.bf            <- 0                 # Es el vector de los residuales al cuadrado de los betas favorables
r2.mobf         <- 0                 # R^2 del modelo original
```

```
pruebas.bf          <- matrix(0,13,columnas, byrow=TRUE, dimnames = list(c("W", "BP", "SR", "SEST",
                       "WM1", "WM2", "WM3", "WM4", "WM5", "WM6", "WM7", "WM8", "WM9"),
                       c("n=20","n=30","n=40","n=50", "n=60","n=70","n=80","n=90","n=100")))
                                            # Es la tabla donde se guardan los resultados de todas las pruebas

# Prueba White #

min.cuad.bf         <- 0                    # Modelo de Regresión
resumen.bf          <- 0                    # Resumen del análisis de varianza del modelo de regresión
r2.bf               <- 0                    # Es el R^2 del modelo de regresión de los Betas favorables
white.bf            <- 0                    # Es el resultado de la prueba de White para betas favorables
ji.w                <- 16.9190              # Es la ji-cuadrada con 9 gl = 16.9190
p.white_bf          <- matrix(,simulacion,1) # Se guardan los valores de la Prueba W para un tamaño de muestra
resultado1.bf       <- matrix(,simulacion,1) # Es una matriz de 0 y 1 que compara todos los valores obtenidos
resultados_white    <- matrix(0,1,9)        # Se guardan los valores de la Prueba W para 9 tamaños de muestra

# Prueba Breusch-Pagan #

sigma.cuad_bf       <- 0                    # Es el valor de sigma^2
p.bf                <- 0                    # p = y/sigma cuadrada
modelo.bf           <- 0                    # Modelo de regresión con los betas favorables
resum.bf            <- 0                    # Resumen del análisis de varianza del modelo de regresión
coefi.bf            <- 0                    # Los coeficientes del análisis de varianza
yest.bf             <- 0                    # y's estimadas
scr.bf              <- 0                    # Es la SCR para cada muestra de los betas favorables
theta.bf            <- 0                    # Es el valor de theta para cada muestra de los betas favorables
ji.b                <- 3.84146              # Es la ji-cuadrada con 1 gl = 3.84146
p.breusch_bf        <- matrix(,simulacion,1) # Se guardan los valores de la Prueba BP para un tamaño de muestra
resultado2.bf       <- matrix(,simulacion,1) # Es una matriz de 0 y 1 que compara todos los valores obtenidos
resultados_bp       <- matrix(0,1,9)        # Se guardan los valores de la Prueba BP para 9 tamaños de muestra

# Prueba de Score de residuales #

d.bf                <- 0                    # Es el vector de residuales^2 del modelo completo
sigma0.2_bf         <- 0                    # Es una parte del estadístico para cada muestra obtenida
sigmai              <- 0                    # Es el producto de z'a
dimension           <- 0                    # Es el tamaño de la muestra
sigma1.1            <- 0                    # Convierte la matriz sigmai en una matriz cuadrada
sigma1              <- 0                    # Es la inversa de sigma1.1
sigma1.2            <- 0                    # Es la matriz sigmai cuadrada
sigma2              <- 0                    # Es la inversa de sigma1.2
g                   <- 0                    # Es el producto de algunas matrices
H                   <- 0                    # Es la matriz H
V0                  <- 0                    # Es el producto de matrices por el modelo reducido
Z.bf                <- 0                    # Es la matriz de N X 2
SR.bf               <- 0                    # Es el resultado de la prueba para cada muestra obtenida
ji.ps               <- 3.84146              # Es la ji-cuadrada con 1 gl = 3.84146
p.residuales_bf     <- matrix(,simulacion,1) # Se guardan los valores de la Prueba SR para un tamaño de muestra
resultado3.bf       <- matrix(,simulacion,1) # Es una matriz de 0 y 1 que compara los valores obtenidos
resultados_sr       <- matrix(0,1,9)        # Se guardan los valores de la Prueba SR para 9 tamaños de muestra

# Prueba de Score Estudentizado #

S.bf                <- 0                    # Es el resultado de la prueba de Score para cada muestra obtenida
S.STUD_bf           <- 0                    # Es el resultado de la prueba de SEst para cada muestra obtenida
```

```
p.scorest_bf        <- matrix(,simulacion,1)   # Se guardan los valores de la Prueba SEst para un tamaño de
muestra
resultado4.bf       <- matrix(,simulacion,1)   # Es una matriz de 0 y 1 que compara todos los valores obtenidos
resultados_sestud<- matrix(0,1,9)              # Se guardan los valores de la Prueba SEst para 9 tamaños de
muestra

# Prueba White Modificada (Componentes Principales con el Método de Bonferroni) #

datos_cp               <- 0                         # La desviación estándar de los 9 componentes principales
compon.bf              <- 0                         # Es el modelo de los residuales^2 del modelo completo con
                                                     # todos los componentes principales
coeficientes_cp.bf     <- matrix(,9,1)              # Son los coeficientes para cada muestra
alfa                   <- 0.05                      # Es el valor de Alfa
alfa1                  <- 0                         # Valor que se utiliza para comparar el método de Bonferroni
componentes.prin_bf    <- matrix(0,9,9)             # Es el resultado de la prueba de Componentes Principales
todas.una_bf           <- matrix(0,9,1)             # Es una matriz donde se guarda los valores de la Prueba de
                                                     # CP utilizando el método de Bonferroni
CP                     <- matrix(0,13,columnas)     # Es una matriz donde se van guardando uno a uno los CP
resultado5.bf          <- matrix(,9,1)              # Es una matriz de 0 y 1 que compara los valores obtenidos
b                      <- matrix(0,9,9)

for (k in 1:columnas)
{
        N <- (k+1)*10                              # Tamaño de la muestra
        x3 <- runif(N,min=0,max=1)                 # La variable x3 uniforme
        corr12 <- 0                                # La correlación de (x1,x2) = 0
        constante <- 1                             # El valor de Constantes =1

        ### CORRELACIÓN BAJA ###

        while (0.0001>corr12 | corr12>0.01)
        {
                K <- 2*constante
                x1 <- runif(N,min=0,max=1)
                U  <- runif(N,min=0,max=1)
                x2 <- x1+(U*K)
                corr12 <- cor(x1,x2)
                constante <- constante+1
                contar <- contar+1
        }
        matrizx <- cbind(x1,x2,x3)
        Intercepto <- rep(1, N)
        matrizxcon1s <- cbind(Intercepto,x1,x2,x3)
        x <- as.matrix(matrizxcon1s)

        # ********************************************************************** #
        # BETAS ESTIMADOS POR MEDIO DE LOS EIGENVECTORES: Orientación favorable #
        # ********************************************************************** #

        XTX <- t(x)%*%x
        eigenvalores <- eigen(XTX)
        eigenvectores <- eigenvalores$vectors
        betas.favorables <- eigenvectores[,1]
        betas.desfavor <- eigenvectores[,4]
```

```
# NUEVAS VARIABLES #

x4<- (x1)^2
x5<- (x2)^2
x6<- (x3)^2
x7<- (x1)*(x2)
x8<- (x1)*(x3)
x9<- (x2)*(x3)
datos.x <- cbind(x1,x2,x3,x4,x5,x6,x7,x8,x9)

# EIGENVALORES DE LA MATRIZ DE CORRELACIONES Y PRIMER CRITERIO DE LOS
COMPONENTES PRINCIPALES #

correlacion <- cor(t(datos.x)%*%datos.x)
eigen.corr <- eigen(correlacion, only.values=TRUE)
eigenvalor.corre <- eigen.corr$values
HTML(eigenvalor.corre,file="G:/Correlacion_Baja/cbc2x135.html")

# SEGUNDO CRITERIO DE LOS COMPONENTES PRINCIPALES #

todas.una_bf <- matrix(0,9,1)
criterio.uno <- summary(datos_cp<-princomp(datos.x,cor=TRUE))
HTML(criterio.uno,file="G:/Correlacion_Baja/cbc1x135.html")

##### PRUEBAS #####

unos.columna <- matrix(1,N,1)
X <- x2+x3 # X = x1 ó X = matrix(1,N,1)

for (i in 1:simulacion)
{
        e1 <- rnorm(N,0,(sqrt(X^L)))

        # ******************************************************************* #
        #                    MODELO DE REGRESIÓN LINEAL                        #
        # ******************************************************************* #

        betas.favorables
        Y1.bf <-(x%*%betas.favorables)+ e1
        analisis.bf <- lm(Y1.bf~x1+x2+x3)
        coeficientes.bf <- analisis.bf$coefficients
        matriz.coef_bf <- as.matrix(coeficientes.bf)
        res.bf <-summary(analisis.bf)
        residuos.bf <- res.bf$residuals
        y.bf <- residuos.bf^2
        datos.x
        r2.mobf <- res.bf$r.squared

        ### PRUEBA WHITE ###
        min.cuad.bf <- lm(y.bf~x1+x2+x3+x4+x5+x6+x7+x8+x9)
        resumen.bf <- summary(min.cuad.bf)
        r2.bf <-resumen.bf$r.squared
        white.bf <- N*r2.bf
        resultado1.bf[i,] <- white.bf
        if (white.bf > ji.w)
                p.white_bf[i,] <- 1                # 1 = Heterogeneidad
```

```
else
        p.white_bf[i,] <- 0                    # 0 = Homogeneidad

### PRUEBA BREUSCH-PAGAN ###
sigma.cuad_bf <- (sum(y.bf))/(N)
p.bf <- y.bf/sigma.cuad_bf
modelo.bf <- lm(p.bf~x1)
resum.bf <- summary(modelo.bf)
coefi.bf <- resum.bf$coefficients[,1]
x.uno <- x[,1:2]
yest.bf <- (x.uno)%*%coefi.bf
ymedia.bf <- mean(p.bf)
scr.bf <- sum((yest.bf-ymedia.bf)^2)
theta.bf <- (scr.bf)/2
resultado2.bf[i,]<- theta.bf
if (theta.bf>ji.b)
        p.breusch_bf[i,] <- 1                  # 1 = Heterogeneidad
else
        p.breusch_bf[i,] <- 0                  # 0 = Homogeneidad

### PRUEBA DE SCORE DE RESIDUALES ###
d.bf <- y.bf
P <- 4
sigmai <- X^L
sigma0.2_bf <- (N/(N-P))*(sigma.cuad_bf)
dimension <- length(sigmai)
sigma1.1 <- diag(sigmai,ncol=dimension)
sigma1 <- solve(sigma1.1)
sigma1.2 <- diag((sigmai^(1/2)),ncol=dimension)
sigma2 <- solve(sigma1.2)
g <- (t(x)%*%sigma1%*%x)
H <- diag(sigma2%*%x%*%solve(g)%*%t(x)%*%sigma2)
V0 <- x%*%solve(t(x)%*%x)%*%t(x)
Z.bf <- cbind(Intercepto,x1)
SR.bf <- (1/2)*(t((d.bf/sigma0.2_bf)-unos.columna + H)%*%Z.bf%*%solve(t(Z.bf)
 %*%V0%*%Z.bf)%*%t(Z.bf)%*%((d.bf/sigma0.2_bf)-unos.columna + H))
resultado3.bf[i,] <- SR.bf
if (SR.bf > ji.ps)
        p.residuales_bf[i,] <- 1                         # 1 = Heterogeneidad
else
        p.residuales_bf[i,] <- 0                         # 0 = Homogeneidad

### PRUEBA DE SCORE ESTUDENTIZADO ###
S.bf <- (1/2)*(t((d.bf/sigma.cuad_bf)-unos.columna)%*%Z.bf%*%
 solve(t(Z.bf)%*%Z.bf)%*%t(Z.bf)%*%((d.bf/sigma.cuad_bf)-unos.columna))
S.STUD_bf <- (2*N)*(S.bf/(t((d.bf/sigma.cuad_bf)-unos.columna)%*%
 ((d.bf/sigma.cuad_bf)-unos.columna)))
resultado4.bf[i,] <- S.STUD_bf
if (S.STUD_bf > ji.ps)
        p.scorest_bf[i,] <- 1                    # 1 = Heterogeneidad
else
        p.scorest_bf[i,] <- 0                    # 0 = Homogeneidad

### PRUEBA WHITE MODIFICADA (CP1-CP9 aplicando el método de Bonferroni) ###

compon.bf <- summary(lm(y.bf~datos_cp$scores[,1]+datos_cp$scores[,2]+
```

```
                    datos_cp$scores[,3]+datos_cp$scores[,4]+datos_cp$scores[,5]+datos_cp$scores[,6]+
                    datos_cp$scores[,7]+datos_cp$scores[,8]+datos_cp$scores[,9]))
                    coeficientes_cp.bf <-compon.bf$coefficients[2:10,4]
                    coeficientes_cp.bf <-as.matrix(coeficientes_cp.bf)
                    for (u in 1:9)
                    {
                            for(v in 1:u)
                            {
                                    alfa1<-alfa/u                # Valor del alfa1 de acuerdo al número de CP
                                    if (coeficientes_cp.bf[v,1]<alfa1)
                                            componentes.prin_bf[v,u] <-1          # 1 = Heterogeneidad
                                    else
                                            componentes.prin_bf[v,u] <-0          # 0 = Homogeneidad
                                    if (sum(componentes.prin_bf[,u])>0)
                                            resultado5.bf[u,1]<-1
                                    else
                                            resultado5.bf[u,1]<-0
                            }
                    }
                    todas.una_bf<- todas.una_bf+resultado5.bf
                    resultados_white[,k] <- apply(p.white_bf,2,sum)
                    resultados_bp[,k] <- apply(p.breusch_bf,2,sum)
                    resultados_sr[,k] <- apply(p.residuales_bf,2,sum)
                    resultados_sestud[,k] <- apply(p.scorest_bf,2,sum)
            }
            CP[5:13,k] <- CP[5:13,k]+todas.una_bf
            pruebas.bf[5:13,] <- (1/simulacion)*(CP[5:13,])
            b[,k]<- coeficientes_cp.bf
}
WHITE <- (1/simulacion)* resultados_white
BP <- (1/simulacion)* resultados_bp
SR <- (1/simulacion)* resultados_sr
SEST <- (1/simulacion)* resultados_sestud
pruebas.bf[1,] <- WHITE
pruebas.bf[2,] <- BP
pruebas.bf[3,] <- SR
pruebas.bf[4,] <- SEST
pruebas.bf
```