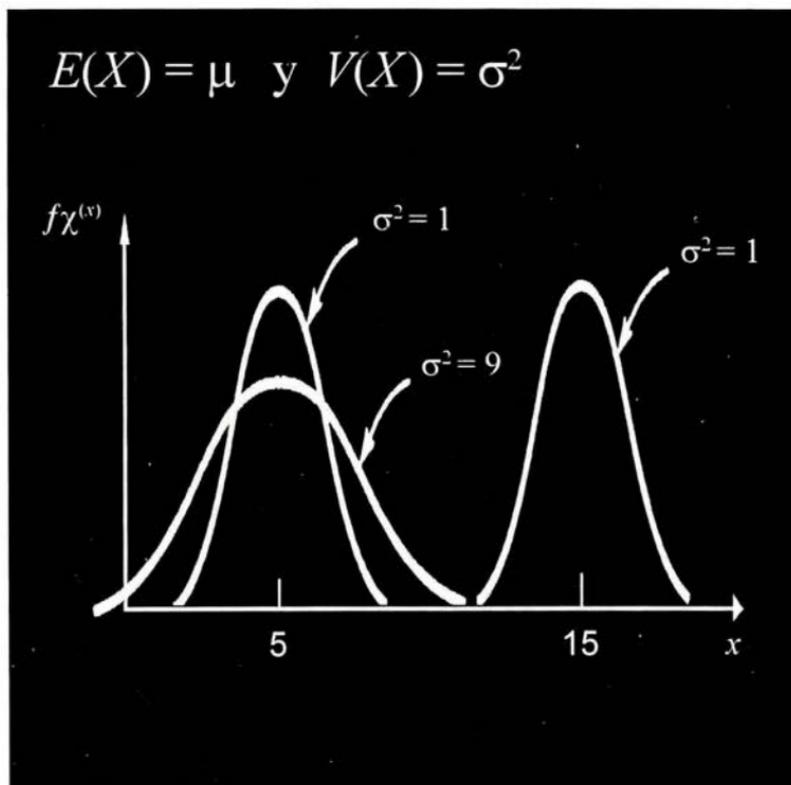


LA ESTADÍSTICA A^2

UNA ALTERNATIVA EN PRUEBAS DE NORMALIDAD

Clara Martha Adalid Diez de Urdanivia



**La estadística A^2 . Una alternativa
en pruebas de normalidad**

**Universidad
Autónoma
Metropolitana
Unidad Xochimilco**

**La estadística A^2 . Una alternativa
en pruebas de normalidad**

Clara Martha Adalid Diez de Urdanivia

Universidad Autónoma Metropolitana
Rector general, doctor José Luis Gázquez Mateos
Secretario general, licenciado Edmundo Jacobo Molina

Universidad Autónoma Metropolitana-Xochimilco
Rectora, doctora Patricia Elena Aceves Pastrana
Secretario de la Unidad, doctor Ernesto Soto Reyes Garmendia

División de Ciencias Sociales y Humanidades
Director, licenciado Gerardo Zamora Fernández de Lara
Secretario académico, maestro Roberto Constantino Toto

Comité editorial

Gerardo Ávalos Tenorio
Felipe Gálvez Cancino / Federico Manchón /
Roberto Manero Brito / Germán Monroy Alvarado /
Patricia Nettel

Edición: Salvador González Vilchis

Primera edición, febrero de 2000

DR. © 2000, Universidad Autónoma Metropolitana

Unidad Xochimilco
Calzada del Hueso 1100
Col. Villa Quietud, Coyoacán
C P. 04960, México DF.

ISBN 970-654-453-4 (obra completa)
970-654-584-0

Impreso y hecho en México / Printed and made in Mexico

Índice

Introducción	9
Capítulo 1. Generalidades;	15
La Función de distribución empírica, 17; Estadísticas basadas en la función de distribución empírica, 18; La estadística de Crámer von Mises, 19; La estadística de Anderson-Darling, 20; Una fórmula para el cálculo de la estadística A^2 , La transformación PIT, 21; La transformación PIT, 21; Los distintos casos: 0, 1, 2 y 3, para el caso normal, 22; La estadística A^2 para probar normalidad en las perturbaciones estocásticas en un modelo de regresión, 26; Uso de los residuos estudiantizados internos, 29; El uso de la CPIT en un modelo de regresión múltiple, 31.	
Capítulo 2. La estadística Jarque-Bera	35
La estadística Jarque-Bera para probar normalidad en las perturbaciones estocásticas, 36; La estadística W^c 36; La estadística W^c en una prueba de normalidad de las perturbaciones estocásticas en un modelo de regresión lineal, W'_{mod} , 38; Algunos comentarios acerca de las potencias de las estadísticas A^2 , Jarque-Bera y W^c . Potencia de las estadísticas A^2 cuando se estima los parámetros, 39; Potencia de las estadísticas W y W' , 40; Potencia de la estadística de Jarque-Bera, 41.	
Capítulo 3. Descripción del estudio	43
Las distribuciones alternativas, 44; los residuos, 45; Las alteraciones en la matriz X , 46; Las potencias empíricas, 46; Los puntos de rechazo, 47; Resultados, 49.	

Capítulo 4. Resumen y conclusiones	57
-------------------------------------------	----

Bibliografía	63
---------------------	----

Índice de tablas

Tabla 1.1. Puntos porcentuales para una prueba de normalidad con μ y σ^2 desconocidos	47
---------------------------------------------------------------------------------------------------------	----

Tabla 1.2. Normalidad en las observaciones; puntos significativos para pruebas LM_n de normalidad; 10 000 réplicas	48
-----------------------------------------------------------------------------------------------------------------------------	----

Tabla 1.3. Puntos porcentuales empíricos de la prueba W' aproximada	49
------------------------------------------------------------------------------	----

Tabla 2. Resultados de la simulación Porcentaje de rechazos en 200; $n = 20$, estadísticas A^2 , J-B y W'_{mod}	50
-----------------------------------------------------------------------------------------------------------------------------	----

Tabla 3. Resultados de la simulación Porcentaje de rechazos en 200; $n = 50$, para las estadísticas A^2 , J-B y W'_{mod}	54
--------------------------------------------------------------------------------------------------------------------------------------	----

Introducción

Sin temor a equivocarnos podemos decir que, en Estadística, la distribución más conocida y utilizada es la Normal.

Desde el siglo XIX, Laplace había probado que la distribución de probabilidad de la esperanza de vida para cualquier edad tendía a la Normal, y Quetelet había ajustado curvas normales a 5 738 mediciones del tórax de soldados escoceses y a 10 000 mediciones de la estatura de conscriptos franceses (Seal, 1967).

Cuando Gauss encontró que la distribución de los errores, en el modelo lineal planteado por Legendre

$$e_i = \sum_{j=1}^q \beta_j z_{ij} - x_i \quad (i = 1, 2, \dots, N; N > q) .$$

(en donde las z_{ij} son «coeficientes» conocidos, las x_i son mediciones, las β_j son las «incógnitas» y los e_i son los «errores») era necesariamente de forma normal, la ley Normal, como Galton la bautizó, o la ley de los errores por excelencia, cobró gran aceptación.

Gauss desarrolló el método de mínimos cuadrados, con el que pudo determinar los estimadores de las β_j 's y a través de ciertas relaciones demostró que éstos también poseen una distribución normal.

El modelo clásico de regresión lineal normal, como ahora lo conocemos:

$$y = XB + u$$

en donde y es un vector columna $n \times 1$ de la variable dependiente Y ; X es una matriz $n \times k$ de rango completo de observaciones; B es un vector columna $k \times 1$ de los parámetros desconocidos $\beta_1, \beta_2, \dots, \beta_k$, y u es un vector columna de $n \times 1$ de las perturbaciones (o errores) estocásticas, sobre el cual se basa la mayor parte de la teoría econométrica, tiene como uno de sus supuestos más importantes el que los errores

poseen una distribución normal. La justificación de este supuesto se apoya en el hecho de que las u_i representan la influencia combinada (en la variable dependiente) de un gran número de variables independientes que han sido omitidas en el modelo de regresión, y que esta influencia, aunque pequeña, es aleatoria.

Las pruebas de Bondad de Ajuste y la importancia de verificar el supuesto de normalidad

Entre las técnicas usadas para probar la suposición de normalidad en las perturbaciones estocásticas están las pruebas de bondad de ajuste, las que miden de alguna forma qué tanto se ajustan las observaciones de una muestra aleatoria a la distribución hipotética. Las mediciones que indican la consistencia o discrepancia son estadísticas de prueba. Diversos autores se han dedicado a la investigación y desarrollo de este tipo de pruebas, y en particular, a aquéllas relacionadas con la distribución normal.

En la literatura estadística pueden encontrarse muchos ejemplos: desde los más sencillos como los que utilizan técnicas gráficas, hasta aquéllos en donde se establece un procedimiento formal de una prueba de hipótesis. Entre las últimas están:

- 1) Las pruebas que utilizan la distribución Ji-cuadrada.
- 2) Las pruebas que se apoyan en la función de distribución empírica. Entre éstas podemos mencionar la prueba de Anderson-Darling.
- 3) Las pruebas basadas en los tercero y cuarto momentos de la distribución, el sesgo y la kurtosis.

En el caso particular del modelo de regresión

$$Y = X\beta + u$$

con residuos

$$e = y - \hat{y}$$

las pruebas mencionadas pueden utilizarse para probar la normalidad en las perturbaciones estocásticas u_i .

La importancia de verificar la normalidad es esencial, ya que cuando este supuesto se cumple las distribuciones de los estimadores de mínimos cuadrados de los coeficientes de regresión del modelo son también normales y son idénticos a los estimadores de máxima verosimilitud; la distribución de $(n-k) s^2/\sigma^2$ con

$$s^2 = \frac{(e'e)}{(n-k)}$$

es Ji -cuadrada con $n-k$ grados de libertad. El conocimiento de las distribuciones de los estimadores de β y de σ^2 nos permite establecer intervalos de confianza y llevar a cabo pruebas de hipótesis de dichos parámetros.

Por el contrario, cuando este supuesto no se cumple se invalidan algunos resultados. Por ejemplo, Spanos (1986: 451) demuestra que en la distribución asintótica de s^2 , el estimador de mínimos cuadrados de σ^2 , bajo el supuesto de que los errores poseen un cuarto momento finito, depende de la suposición de normalidad de los errores vía el coeficiente de kurtosis. Entonces, cualquier desviación de la normalidad afectará seriamente los resultados que se apoyen en este supuesto. En particular, el tamaño de α , y la potencia de algunas pruebas como la Ji -cuadrada, pueden ser muy diferentes de las que originalmente se postularon bajo el valor de α , afectando seriamente todas las pruebas que dependan de la distribución de este estimador. Un resultado similar puede encontrarse en Arnold (1980).

Objetivos

El propósito de este trabajo es utilizar la prueba de Anderson-Darling (A^2) (Anderson y Darling, 1952, 1954) para realizar una prueba de bondad de ajuste acerca de la normalidad de las perturbaciones estocásticas en un modelo de regresión lineal dado, y a la vez comparar su funcionamiento con otras dos pruebas que han sido propuestas para este caso de regresión.

La Anderson-Darling se menciona como una estadística que, al usar la tabla de puntos de rechazo correcta, reporta potencias comparables y en algunas ocasiones superiores a las de otras pruebas frente a diversas alternativas (Stephens, 1974), y que como estadística de bondad de ajuste, para el caso continuo de observaciones independientes e idénticamente distribuidas, se prefiere sobre otras por su sencillez de cálculo (Easterling, 1976: 6). A pesar de lo anterior, no se ha trabajado con ella en pruebas de bondad de ajuste en modelos de regresión para probar normalidad en las perturbaciones estocásticas.

En este trabajo intentaremos explorar si la estadística Anderson-Darling es una prueba comparable con las estadísticas de Jarque y Bera (1982) y la W' de Shapiro-Francia (modificada)¹ (Shapiro y Francia, 1972), en un modelo de regresión lineal múltiple para probar normalidad en las perturbaciones estocásticas.

Estructura

En el primer capítulo se presenta la función de distribución empírica, la estadística A^2 y su utilización en la prueba de bondad de ajuste correspondiente al caso de observaciones independientes e idénticamente distribuidas, según una distribución normal dada, con o sin parámetros desconocidos;

¹ La estadística W' propuesta por Shapiro-Francia, es una modificación de la estadística W de Shapiro-Wilk.

así como para la prueba de normalidad de las perturbaciones estocásticas en un modelo de regresión lineal múltiple. Este material se basa en el de D' Agostino y Stephens (1986, cap. 4).

En el segundo capítulo se hace una breve introducción a las estadísticas de Jarque y Bera y Shapiro y Francia, comentando algunos reportes acerca de su funcionamiento frente a algunas distribuciones alternativas.

En el tercer capítulo se exponen los resultados de un pequeño estudio de Monte Carlo —para comparar las tres estadísticas ya mencionadas con un modelo de regresión lineal múltiple— utilizando siete distribuciones alternativas para los errores, además de la Normal; éstas son: la Normal Heteroscedástica; t , (t de Student con cinco grados de libertad); Log-Normal; Exponencial; Uniforme $(0,1)$; Normal Truncada Derecha y, por último, la Normal Truncada Simétrica.

Finalmente, en el cuarto capítulo se mencionan algunas conclusiones obtenidas de las simulaciones referentes al comportamiento de las tres estadísticas; en particular con respecto a la de Jarque y Bera, que se usa en la prueba de normalidad de los errores en algunos paquetes conocidos de uso común en econometría.

Capítulo I

Generalidades

En Estadística es muy importante probar la concordancia entre la teoría y los hechos. Si disponemos de una muestra de n valores observados de cierta variable, desearemos saber si esta variable se comporta «razonablemente» como una variable aleatoria que tenga una distribución de probabilidad con ciertas propiedades dadas. Por ejemplo, podemos preguntarnos si es lógico suponer que los datos que han sido obtenidos de una población poseen una distribución normal, con una media $\mu = 10$ y una desviación estándar igual a 2. En este caso, la hipótesis estadística se plantea de la siguiente manera:

- (1) $H_0: x_1, x_2, \dots, x_n$ proviene de una población normal con media $\mu = 10$ y desviación estándar igual a 2

y la queremos contrastar con la hipótesis

- (2) $H_1: x_1, x_2, \dots, x_n$ no proviene de una población normal con media $\mu = 10$ y desviación estándar igual a 2

A la hipótesis establecida en (1) se le conoce como hipótesis nula, mientras que a la (2) como hipótesis alternativa. Observamos que bajo H_0 la distribución está completamente especificada, por lo que tenemos una hipótesis simple, no así bajo H_1 , que es una hipótesis compuesta. Si suponemos que la hipótesis H_0 es cierta, esperamos que el comportamiento muestral sea explicable por la función de distribución dada. Más adelante se verá cómo se pueden definir medidas para saber si el referido comportamiento muestral difiere mucho de la hipótesis nula H_0 .

Este esquema se presenta en inferencia estadística y es lo que se conoce como un planteamiento clásico de prueba de hipótesis.

En un problema de prueba de hipótesis como el anterior, la hipótesis que hay que probar precisa no sólo la distribución sino también los valores de los parámetros de la distribución especificada.

Formalmente, se podría considerar una hipótesis nula más general con el objetivo de probar que la función de distribución tiene una forma específica (por ejemplo normal) sin suponer que los valores de los parámetros son conocidos.

Por ejemplo podemos plantearnos probar:

H_0 : x_1, x_2, \dots, x_n proviene de una distribución normal con media $\mu = 0$ y desviación estándar desconocida.

contra

H_1 : x_1, x_2, \dots, x_n proviene de una distribución distinta a la normal.

Entonces, una prueba de bondad de ajuste examinará qué tan bien concuerda, o discrepa, una muestra de observaciones con la distribución especificada por la hipótesis nula. Al igual que en una prueba de hipótesis, la hipótesis nula puede ser simple o compuesta como ya se mencionó, pero en todos los problemas de bondad de ajuste la hipótesis alternativa H_1 es compuesta, ya que sólo establece que H_0 es falsa. Si en la prueba de bondad de ajuste correspondiente no se rechaza H_0 , entonces se puede utilizar el supuesto implicado en H_0 para elaborar inferencias sobre otros aspectos del modelo; por ejemplo, pruebas de hipótesis sobre los parámetros de la distribución.

Entre los métodos utilizados en la bondad de ajuste están aquéllos que hacen uso de la función de distribución empírica

que permite representar la «evidencia muestral» en forma de distribución.

A continuación se introduce la función de distribución empírica y las distintas estadísticas apoyadas en esta función, en particular la estadística de Anderson-Darling.

La función de distribución empírica

La función de distribución empírica (FDE) es una función escalonada, calculada con los valores de la muestra, que estima a la función de distribución de la población.

Suponga que se tiene una muestra aleatoria de tamaño n , X_1, X_2, \dots, X_n , y sean $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ las estadísticas de orden. Suponga, además, que la función de distribución de la variable X , $F(x)$ es continua.

La función de distribución empírica denotada por $F_n(x)$ se define como:

$$F_n(x) = \frac{\text{Número de observaciones } \leq x}{n}; \quad -\infty < x < \infty$$

es decir:

$$F_n(x) = 0 \quad \text{si } x < X_{(1)}$$

$$F_n(x) = \frac{i}{n} \quad \text{si } X_{(i)} \leq x < X_{(i+1)}, \quad i = 1, \dots, n-1$$

$$F_n(x) = 1 \quad \text{si } X_{(n)} \leq x$$

Para cada x , $F_n(x)$ representa la proporción de observaciones menores o iguales a x , mientras que $F(x)$ representa la

probabilidad de que la variable aleatoria tome un valor menor o igual al valor de x .

La función de distribución empírica, $F_n(x)$ es un estimador consistente de la función de distribución verdadera $F(x)$ (Rényi, 1970).

$$\text{Así: } \lim_{n \rightarrow \infty} P(|F_n(x) - F(x)| < \epsilon) = 1$$

La FDE difiere, en general, de la función de distribución F por ser aleatoria; si «difiere mucho» de la $F(x)$ supuesta, entonces podemos cuestionar la suposición de que $F(x)$ es la función de distribución verdadera, es decir, la diferencia entre las funciones de distribución empírica y la supuesta se utiliza como «estadística» para determinar si se rechaza o no la hipótesis de que la función de distribución supuesta es la verdadera.

Estadísticas basadas en la función de distribución empírica

Las estadísticas que se apoyan en la FDE miden la «discrepancia» existente entre la FDE y la función de distribución supuesta.

Algunas estadísticas que miden la diferencia vertical que existe entre $F_n(x)$ y $F(x)$ son las siguientes:

1. Las estadísticas D^+ y D^- :

$$D^+ = \sup_x (F_n(x) - F(x))$$

$$D^- = \sup_x (F(x) - F_n(x))$$

Estas estadísticas son, respectivamente, la diferencia vertical más grande cuando $F_n(x)$ es mayor que $F(x)$, y la diferencia vertical más grande cuando $F_n(x)$ es menor que $F(x)$.

2. La estadística de Kolmogorov-Smirnov

$$D = \sup_x |F_n(x) - F(x)| = \max(D^+, D^-)$$

Esta estadística es la diferencia vertical más grande sin importar su signo.

3. La estadística de Kuiper V :

$$V = D^+ + D^-$$

4. La estadística de Watson, U^2

$$U^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F(x)\} \int_{-\infty}^{\infty} (F_n(x) - F(x)) dF(x) \}^2 dF(x)$$

Las estadísticas de Kuiper V y de Watson U^2 se utilizan para observaciones en un círculo, ya que los valores que toman son independientes del origen que se elija. Estas estadísticas pueden utilizarse también para observaciones en una recta (Stephens, 1974) y

5. Las estadísticas pertenecientes a la familia Crámer von Mises, que merecen una atención especial, ya que la estadística de Anderson-Darling es un miembro de ellas.

$$CVM = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 G(x) dF(x) \quad (1.1)$$

en donde $G(x)$ es una función no-negativa que pondera los valores de $[F_n(x) - F(x)]^2$ y permite una mejor sensibilidad en algunos puntos en donde se requiera.

La estadística de Crámer von Mises

Cuando la función que pondera $G(x) = 1$, la estadística (1.1) se conoce como la de Crámer von Mises (W^2).

La estadística de Anderson-Darling

Si en (1.1) $G(x) = [F(x) \cdot (1-F(x))]^{-1}$, la estadística se conoce como la de Anderson-Darling (A^2).

$$A^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 [F(x) (1-F(x))]^{-1} dF(x) \quad (1.2)$$

La estadística de Anderson-Darling fue propuesta por estos autores en 1952 (Anderson y Darling, 1952, 1954), cuando compararon la función de distribución empírica con la función de distribución hipotética a través de la discrepancia o «distancia» existente entre estas dos vistas como funciones de distribución utilizando una seudodistancia.

La estadística A^2 le da mayor peso a la discrepancia en las colas de la distribución, que es cuando $[F(x) (1-F(x))]^{-1}$ crece; y se espera que detecte discrepancias en las colas mejor que la W^2 (Stephens, 1986).

En general, la distribución exacta de la estadística A^2 es muy difícil de calcular para una n finita. En Anderson y Darling (1954) se calculan los puntos de rechazo asintóticos para los niveles de significancia 0.10, 0.05 y 0.01, y también en Lewis (1961); con estudios de Monte Carlo se calcula

$$F(z; n) = Pr [A^2 \leq z]$$

con un rango de valores para z igual a $0 \leq z \leq 8$ y valores de $n = 1, 2, \dots, 8$, así como sus valores asintóticos, encontrándose que para $z = 8$ la probabilidad es igual a 1.000 para todos los valores de n . Se verifican también los puntos de rechazo asintóticos encontrados por Anderson y Darling. La convergencia de la distribución de A^2 a su distribución asintótica es muy rápida, para $n = 8$ la diferencia máxima que existe entre los puntos de rechazo calculados y los puntos asintóticos es aproximadamente igual a sólo 0.006. Para $F(z) \geq 0.8$ es de tan sólo 0.001; de tal manera que en la práctica, la distribución asintótica puede ser usada para valores de $n > 8$.

*Una fórmula para el cálculo de la estadística A^2 ,
la transformación PIT*

Una fórmula operacional para el cálculo de la estadística A^2 en (1.2), resultante de evaluar la integral es la siguiente (Anderson y Darling, 1954: 765):

$$A^2 = -n^{-1} \sum_{i=1}^n \{ (2i-1) 1nF(x_{(i)}) + (2n+1-2i) \ln(1-F(X_{(i)})) \} \quad (1.3)$$

en donde las $X_{(i)}$ son las estadísticas de orden de la muestra de tamaño n , y $F(X_{(i)})$ es la transformación de $X_{(i)}$ con la distribución hipotética F .

La transformación PIT

Si X es una variable aleatoria con una función de distribución $F(x)$ continua, entonces la variable aleatoria transformada $U = F(X)$ está distribuida en forma uniforme en $(0, 1)$.

Si $0 < u < 1$ entonces,

$$Pr(U \leq u) = Pr(F(X) \leq u) = Pr(X \leq F^{-1}(u)).$$

Pero

$$Pr(X \leq x) = F(x)$$

Por lo tanto

$$Pr(U \leq u) = F(F^{-1}(u)) = u$$

Es decir, U está distribuida en forma uniforme en $(0, 1)$.
Si a cada una de las observaciones $X_{(i)}$ se las transforma

$$U_{(i)} = F(X_{(i)})$$

entonces, bajo H_0 , las $U_{(i)}$ son las estadísticas de orden de una muestra independiente U_1, U_2, \dots, U_n de la distribución uniforme $(0,1)$, y (1.3) puede reescribirse como

$$A^2 = -n \cdot \frac{1}{n} \sum_{i=1}^n \{ (2i-1) \ln U_{(i)} + (2n+1-2i) \ln (1 - U_{(i)}) \} \quad (1.4)$$

Es decir, bajo H_0 , las U_i son una muestra independiente de la distribución uniforme $(0,1)$.

Entonces, la estadística (1.4) calculada con la FDE de las $U_{(i)}$ comparada con la distribución uniforme tendrá exactamente el mismo valor que la estadística (1.3) calculada de la FDE de las $X_{(i)}$, comparada con la función de distribución $F(x)$.

La transformación anterior se conoce como la transformación integral de probabilidad (del inglés *probability integral transformation* (PIT)).

El valor encontrado de la estadística se compara con el valor de la tabla asintótico, de acuerdo al nivel de significancia escogido.

Existen distintas tablas según el conocimiento que se tenga de la distribución $F(x)$, los distintos casos se clasifican como: casos 0, 1, 2 y 3, cuando F tiene hasta dos parámetros (Stephens, 1974: 730).

Los distintos casos: 0, 1, 2 y 3, para el caso normal

Caso 0: $F(x, \theta)$ es continua, y está completamente especificada. Las dos componentes del vector son conocidos.

En este caso, las $U_{(i)} = F(X_{(i)})$ forman un conjunto de observaciones que, bajo H_0 , son uniformes, ordenadas.

Caso 1: $F(x, \theta)$ es la distribución normal, con $\theta = (\mu, \sigma^2)$; σ^2 es conocida y μ se estima con \bar{x} la media muestral.

Caso 2: $F(x, \theta)$ es la distribución normal, con μ conocida
 σ^2 se estima con

$$\sum_i \frac{(x_i - \mu)^2}{n}$$

Caso 3: $F(x, \theta)$ es la distribución normal, con μ y σ^2 desconocidos. Se estiman con \bar{x} , y

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{(n-1)}$$

la media y la varianza muestrales, respectivamente, (la última corregida para ser insesgada).

El caso 0 es el más desarrollado teóricamente para las estadísticas FDE, aun cuando el tamaño de la muestra sea finito.

Cuando en el vector de parámetros uno, o los dos parámetros son desconocidos deben reemplazarse por sus estimadores respectivos. La fórmula (1.4) de la estadística sigue siendo válida con $U_{(i)} = F(X_{(i)}, \hat{\theta})$ pero ahora cuando H_0 es cierta, las transformadas U_i no son una muestra uniforme e independiente; y la distribución asintótica de la estadística FDE es diferente a la distribución correspondiente al caso 0, ya que puede depender de la distribución bajo consideración y del método de estimación de los parámetros. De esta forma, existen tablas para cada uno de los casos (1, 2 y 3) y para cada distribución F (Normal, Exponencial, Weibull, etc.). La estadística calculada debe compararse con el punto de rechazo de la tabla correspondiente, según sea el caso; además de hacersele una modificación que «corrije» por ser n finito. La A^2 modificada, para el caso 3 que prueba normalidad, es la siguiente:

$$A^2_{\text{mod}} = A^2 \left(\frac{1.0 + 0.75}{n} + \frac{2.25}{n^2} \right)$$

La estadística de Anderson-Darling converge tan rápidamente a su valor asintótico que no se requieren modificaciones para n si $n \geq 5$, esto para los casos 0, 1 y 2, sin embargo se hace la modificación de la fórmula para el caso 3.

En resumen, el procedimiento para probar

H_0 : la muestra aleatoria X_i proviene de $F(x; \theta)$ la distribución normal, $N(\mu, \sigma^2)$ en donde, uno o ambos parámetros de $\theta = (\mu, \sigma^2)$ pueden ser desconocidos.

es el siguiente:

1. Ordenar los valores de la muestra X_i , en orden ascendente, es decir, encontrar las estadísticas de orden $X_{(i)}$.
 - a) Si es necesario, estimar los parámetros con los usuales máximo verosímiles.
2. Calcular $U_{(i)} = F(X_{(i)}, \theta)$, $i = 1, 2, \dots, n$
3. Calcular la estadística A^2 con las $U_{(1)}, U_{(2)}, \dots, U_{(n)}$ y, si estamos en el caso 3, efectuar la modificación para n finito.
4. Para hacer la prueba, comparar el valor encontrado en la fórmula (1.4) con el valor de la tabla correspondiente, dependiendo de si el caso es 0, 1, 2 o 3. Si la estadística es mayor que el percentil respectivo, para un nivel de significancia determinado, se rechaza H_0 al nivel escogido.

Para una opción diferente del uso de la PIT, O'Reilly y Quesenberry (1973) y O' Reilly (1992) proponen para transformar las observaciones en el intervalo (0,1), cuando la distribución que se prueba contiene parámetros desconocidos, utilizar la función de distribución condicional de parte de la muestra, dada la estadística suficiente minimal (distribución conocida como el estimador de Rao-Blackwell), si esta distribución condicional es continua. Esta transformación, aplicada de manera secuencial, se conoce como transformación de probabilidad integral condicional (CPIT), y origina transformadas independientes e idénticamente distribuidas uniformemente en (0,1) reduciendo, efectivamente, el problema a una

prueba de uniformidad. Sin embargo, en su aplicación *no* secuencial, la CPIT consiste en transformar al (0,1) cada observación con el estimador Rao-Blackwell de su distribución. Esta aplicación (no secuencial) da resultados asintóticos equivalentes en las estadísticas basadas en la FDE, al procedimiento descrito previamente de estimar θ con $\hat{\theta}$, y sustituir las transformadas en la fórmula. La aplicación de la CPIT (no secuencial) se verá más adelante para el caso de la regresión.

Como un ejemplo de la utilización de la estadística A^2 (estimando los parámetros) con las primeras 50 observaciones de los datos que aparecen en el apéndice bajo el título de UNI50, tomados del libro de D'Agostino y Stephens (1986: 530), se va a comprobar normalidad; en donde la media y la varianza son desconocidas. Sabemos que, en realidad, esos datos fueron simulados de una distribución Uniforme en el intervalo (0,10).

La media y la varianza se estiman con \bar{x} y s^2

$$\bar{x} = \sum_i \frac{x_i}{n} \quad \text{y} \quad s^2 = \sum_i \frac{(x_i - \bar{x})^2}{(n-1)}$$

respectivamente.

Los resultados obtenidos son los siguientes:

$$\begin{aligned} \bar{x} &= 5.537999 & s &= 3.0355 \\ A^2 &= 1.03938 \\ A^2_{\text{mod}} &= 1.055906 \end{aligned}$$

Comparando estos valores con los porcentiles de la tabla para el caso 3, se rechaza H_0 si el valor calculado de la estadística es mayor que .631, .752, .873, 1.035 que corresponden a valores de alfa de .10, .05, .025 y .001, respectivamente.

Por lo tanto, y como era de esperarse, se rechaza H_0 , (aun con $\alpha = 0.001$).

Si por equivocación se hubiera considerado la tabla del caso 0 la conclusión habría sido muy distinta, porque entonces,

la hipótesis nula no se habría rechazado, ya que el valor obtenido con la estadística está muy por debajo del percentil correspondiente a un alfa de 0.25. Esto muestra lo importante que es usar la distribución asintótica correcta.

La estadística A^2 para probar normalidad en las perturbaciones estocásticas en un modelo de regresión

En el caso de un modelo de regresión, el supuesto de que las perturbaciones estocásticas poseen una distribución que es normal es muy importante por las implicaciones que tiene en las inferencias que se hacen acerca de los parámetros del modelo.

El modelo propuesto es el siguiente:

$$y = X\beta + u$$

en donde:

y es un vector columna de observaciones de la variable dependiente, $n \times 1$

X es una matriz $n \times k$ de observaciones de las $k-1$ variables explicativas: X_2, X_3, \dots, X_k , en donde $X_1 = 1$, es decir, la primer columna de 1 's para representar el término de intersección.

β es el vector columna $k \times 1$ de los parámetros desconocidos $\beta_j, j = 1, 2, \dots, k$.

u es el vector columna $n \times 1$ de perturbaciones estocásticas con $E(u) = 0$ y $\text{Var}(u) = \sigma^2 I$.

En el modelo propuesto nos interesa comprobar que la suposición de normalidad hecha acerca de los errores se cumpla.

Debido a que las perturbaciones estocásticas no son observables, no se puede aquí proponer el uso de la FDE calculada con ellas, por eso se propone utilizar los residuos. En la función de regresión ajustada:

$$y = X b + e$$

con

$$b = (X' X)^{-1} X' y$$

el estimador de mínimos cuadrados de β , de donde

$$e = y - X b = X \beta + u - X (X' X)^{-1} X' y$$

$$= X \beta + u - X (X' X)^{-1} X' (X \beta + u)$$

$$= (I - X (X' X)^{-1} X') u$$

$$e = (I - H) u$$

con

$$H = X (X' X)^{-1} X'$$

e es el vector de residuos, su vector de medias es 0 y su matriz de covarianza es $\sigma^2 (I - X (X' X)^{-1} X')$.

Esta identidad demuestra que la relación entre las perturbaciones u_i y los residuales e_i depende exclusivamente de la matriz H , que usualmente se le conoce como matriz «sombrero». Esta matriz es muy importante para hacer diagnósticos acerca del modelo² (Hoaglin y Welsch, 1978, para su aplicación).

Bajo el supuesto de que las $u_i, i = 1, \dots, n$ son independientes e idénticamente distribuidas, se quiere probar la hipótesis de que su distribución común es una $N(0, \sigma^2)$. El problema así expuesto, a primera vista podría intentarse resolver con el

² Para detectar observaciones discrepantes del conjunto que se está estudiando, por ejemplo.

procedimiento de prueba general visto anteriormente, refiriéndolo al caso 2, en donde la media es conocida y la varianza se estima. Eso sería un error, pues en general los residuos no son independientes y se distribuyen en forma diferente (sus varianzas también son, en general, distintas). Entonces, la FDE construida a partir de éstos no tendría las propiedades ya expuestas para el caso 2.

De todas formas, se ha demostrado en literatura que el comportamiento asintótico de la FDE construida con los residuos estudentizados

$$\frac{e_1}{\hat{\sigma}}, \frac{e_2}{\hat{\sigma}}, \dots, \frac{e_n}{\hat{\sigma}}$$

es el mismo que el de la FDE calculada con observaciones normales, independientes e idénticamente distribuidas, v_i con media y varianza *desconocidas* (Mukantseva, 1977). Por lo tanto, para probar que los errores u_i , $i = 1, \dots, n$ se distribuyen en forma normal se pueden utilizar las estadísticas construidas a partir de la FDE como las de Crámer von Mises, A^2 o la de Kolomogorov, teniendo cuidado de consultar en las tablas adecuadas.

Pierce y Kopecky (1979) y Loynes (1980), también demuestran que todas las estadísticas de prueba calculadas con los residuos de una regresión simple o múltiple tienen la misma distribución asintótica, bajo la hipótesis nula, que para el caso idénticamente distribuido con media y varianza desconocidas.

Es importante («crucial», lo denotan Pierce y Kopecky en su artículo) subrayar que el modelo de regresión debe incluir un término constante, es decir, el primer elemento de cada vector x_i debe ser 1, ya que de otra forma los resultados asintóticos no son válidos.

Por lo tanto, en una prueba de bondad de ajuste, en donde se quiere contrastar la hipótesis de normalidad de los errores con un modelo de regresión (cuando se utiliza la estadística A^2 como estadística de prueba) *la tabla que debe consultarse*

es la del caso 3 en donde se supone que tanto la media como la varianza deben ser estimadas.

Uso de los residuos estudentizados internos

Una observación cuidadosa de los residuos ordinarios

$$e_i = y_i - \hat{y}_i$$

nos permite detectar anomalías en el modelo. Algunos problemas de especificación pueden descubrirse con una simple gráfica, aunque existen técnicas más sofisticadas. Sin embargo, en ocasiones se presentan situaciones que el examen de los residuales no permite distinguir.

Los residuos ordinarios tienen una distribución que depende de los parámetros de escala, ya que son funciones tanto de σ^2 como de $(1-h_{ii})$, en donde h_{ii} son los elementos de la diagonal de la matriz H o «sombbrero». En procedimientos de diagnósticos es conveniente modificarlos. Cook y Weisberg (1982: 18) sugieren diferentes transformaciones como la de estudentización; estudentización interna y externa que permiten calcular residuos independientes de los parámetros de escala.

Se usa el término de estudentización para describir el cociente que resulta de dividir una estadística que depende de algún parámetro de escala entre el estimador del parámetro, de tal forma que este cociente posea una distribución libre de dichos parámetros. Si la estadística y el estimador son dependientes se conoce como estudentización interna; en caso contrario tenemos la estudentización externa.

En la regresión de mínimos cuadrados los residuos estudentizados internos se definen como:

$$\hat{e}_i = \frac{e_i}{s\sqrt{1-h_{ii}}}$$

en donde e_i es el residuo ordinario de mínimos cuadrados;

$$s^2 = \sum_i \frac{e_i^2}{(n-k)}$$

es el cuadrado medio del residuo, y $1-b_{ii}$ es el elemento diagonal de la matriz $I-H$.

De los mismos resultados, ya mencionados de Pierce y Kopecky y Mukantseva, se tiene que la FDE calculada con estos nuevos residuos tiene las mismas propiedades asintóticas.

En el caso de una regresión simple:

$$y_i = B_0 + B_1 x_i + u_i$$

los pasos a seguir son los siguientes:

1. Obtener los residuos de la regresión

$$e_i = y_i - \hat{y}_i$$

2. Calcular los residuos estudentizados (Pierce y Gray, 1982), que se obtienen con la fórmula:

$$\hat{z}_i = \frac{e_i}{\sqrt{\frac{s^2 \left\{ 1 - \frac{1}{n} \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right\}}}}$$

en donde

$$s^2 = \sum_i \frac{e_i^2}{(n-2)}$$

3. Calcular $U_i = F(x_i)$ en donde $F(\cdot)$ es la función de distribución normal, estándar.
4. Obtener las estadísticas de orden de las variables transformadas $U_i, U_{(i)}$.

- Calcular \mathcal{A}^2 con las $U_{(i)}^*$, $i = 1, \dots, n$
- Realizar la prueba: *consultando en la tabla del caso 3* (en este caso no se hace la modificación para n finita, pues no se conoce).

Rechazar H_0 al nivel especificado si \mathcal{A}^2 es mayor que el percentil de la tabla.

El uso de la CPIT en un modelo de regresión múltiple

En el caso de un modelo de regresión múltiple, la transformación CPIT (no secuencial) lleva a un procedimiento equivalente al uso de unos «nuevos residuos» e_j^* (de tipo predictivo), en el que

$$\tilde{U} = G_{n-k-1}(e_j^*)$$

sustituye a la $F(\mathcal{Z})$ del procedimiento anterior, en donde k es el número de variables en el modelo, y $G_\gamma(\cdot)$ es la función de distribución t de Student con γ grados de libertad, evaluada en \cdot , y e_j^* , son los residuos predictivos. Estos se obtienen a partir de los e_j con la siguiente fórmula (O'Reilly, 1993):

$$e_i^* = \frac{e_i}{\sqrt{\frac{n}{n-k-1} (1-x_i' (X'X)^{-1} x_i) \tilde{\sigma}^2 \frac{1}{n-k-1} e_i^2}}$$

con

$$e_i = y_i - \hat{y}_i$$

en donde

$$\tilde{\sigma}^2 = \frac{SCE}{n}$$

es el estimador máximo verosímil sin corrección para el sesgo.

Los residuos e_i^* , se conocen como los residuos predictivos o CPIT.

La transformación CPIT será utilizada a lo largo del presente trabajo.

Para ejemplificar el uso de la estadística A^2 en la prueba de normalidad de los errores en un modelo de regresión lineal múltiple, utilizaremos los datos que gentilmente me proporcionaron las profesoras Lilia Domínguez y Flor Braun.

Los datos aparecen en el apéndice bajo el título de PROB84, y consisten de 84 observaciones para cada una de 10 variables independientes.

El problema consiste en analizar la productividad entre empresas grandes y pequeñas.

Los cálculos realizados se hicieron en un programa de Fortran, que la autora del presente trabajo escribió.

Este programa recibe como entrada la matriz de observaciones independientes X y el vector de observaciones dependientes y . Obtiene el vector de los coeficientes de regresión estimados; los residuos; estandariza estos últimos; calcula los elementos de la diagonal de la matriz $I - X(X'X)^{-1}X'$, que sirven para estudentizar los residuos; transforma cada uno de los residuos con la CPIT; los ordena, y por último, calcula la estadística Anderson-Darling.

Los coeficientes de regresión estimados son los siguientes:

$$\beta_0 = -1.0337310; \beta_1 = 0.02944115; \beta_2 = -0.0175557$$

$$\beta_3 = 0.5965663; \beta_4 = -0.04134133; \beta_5 = -0.01439357;$$

$$\beta_6 = 0.1413104; \beta_7 = 0.6104409; \beta_8 = 0.04368152;$$

$$\beta_9 = -0.04007852$$

Se presentan, como ejemplo de cada uno de los elementos antes mencionados, los primeros 10 resultados tal y como se obtuvieron de la salida del programa.

Los diez primeros residuos estandarizados son:

-1.093073 -0.7844011 -1.163355 -1.469098 0.4263015
0.859099 0.7142838 -0.5398913 0.2117695 0.4479944

Los elementos de la diagonal son:

0.88273 0.9062845 0.8728446 0.8583438 0.834879
0.8385147 0.8109857 0.8081338 0.7912683 0.8097361

Los residuos estudentizados internos:

-1.093414 -0.7712405 -1.171686 -1.500863 0.4355022
0.8979219 0.7421956 -0.561075 0.2220086 0.4647985

Los residuos ordenados son:

-2.378733 -2.244148 -1.820073 -1.722173 -1.651665
-1.58481 -1.500863 -1.403008 -1.373298 -1.257801

Los residuos CPIT:

0.009994694 0.0139275 0.03642355 0.04463642 0.05144921
0.05866619 0.06885199 0.08242671 0.08693153 0.10622362

La estadística A^2 calculada es: 0.2053746

De acuerdo al resultado obtenido *no se rechaza* H_0 .

Por lo tanto podemos suponer que los errores se comportan en forma normal.

Capítulo 2

Para representar algún fenómeno existen diversos procedimientos para probar qué tan apropiada es la distribución normal como modelo, a partir de un conjunto de observaciones. Uno de éstos consiste en utilizar las estadísticas que se apoyan en la función de distribución empírica, tal es el caso de la estadística Anderson-Darling o A^2 , la que fue expuesta en el capítulo anterior. Entre otras alternativas propuestas en literatura que han mostrado un buen funcionamiento podemos mencionar las estadísticas W' de Shapiro y Francia y la $J-B$ de Jarque y Bera.

Debido a que en el capítulo tres se comparan estas estadísticas, nos parece conveniente explicar en forma breve en qué consisten las pruebas de Jarque Bera y de Shapiro-Francia.

La estadística Jarque-Bera

La estadística que proponen Jarque y Bera (1987) se construye a partir del método del multiplicador de Lagrange, en donde suponen que la función de densidad de las perturbaciones estocásticas (μ) pertenece a una «familia general»; por ejemplo, la familia Pearson, de la que la distribución bajo H_0 es un miembro en particular.³ Algunos representantes de esta familia son las distribuciones Normal, Beta, Gamma, t de Student y F .

³ En su artículo, Jarque y Bera (1987) comentan que las pruebas obtenidas a través del multiplicador de Lagrange poseen potencias óptimas con n grande para los representantes de esta «familia general» especificada, sin embargo, de esto no se concluye que para alternativas que no pertenecen a esta familia tendrán potencias malas.

La estadística Jarque-Bera para probar normalidad en las perturbaciones estocásticas

En este caso y bajo los supuestos de que:

- (1) u_1, u_2, \dots, u_n son independientes e idénticamente distribuidas, con media cero y varianza constante
- (2) la densidad de probabilidad de las perturbaciones estocásticas u_i pertenece a la familia Pearson

La estadística de prueba propuesta es la siguiente:

$$LM=N [(\sqrt{\hat{b}_1})^2 / 6 + (\hat{b}_2 - 3)^2 / 24]$$

en donde:

$$\hat{b}_1 = \hat{\mu}_3^2 / \hat{\mu}_2^3 \quad y \quad \hat{b}_2 = \hat{\mu}_4 / \hat{\mu}_2^2$$

y

$$\hat{\mu}_j = \frac{\sum_i e_i^j}{N}$$

los e_i son los residuos ordinarios ya mencionados.

Bajo H_0 , LM se distribuye asintóticamente como una Ji-cuadrada con 2 grados de libertad y es asintóticamente eficiente.

La estadística W'

La estadística W' de Shapiro y Francia (1972) es una modificación de la estadística W de Shapiro y Wilk (1965), diseñada para el caso de observaciones independientes e idénticamente distribuidas.

La prueba W' es la siguiente:

Sea $m' = (m_1, m_2, \dots, m_n)$ el vector de valores esperados de las estadísticas de orden de una normal estándar, y sea $V = (v_{ij})$ la matriz $n \times n$ de covarianza correspondiente.

Es decir, si $x_{(1)}^* \leq x_{(2)}^* \leq \dots \leq x_{(n)}^*$ representa una muestra aleatoria ordenada de tamaño n , de una distribución normal, con media cero y desviación estándar uno, entonces

$$E(x_{(i)}^*) = m_i \quad (i = 1, \dots, n),$$

y

$$\text{Cov}(x_{(i)}^*, x_{(j)}^*) = v_{ij} \quad (i, j = 1, \dots, n)$$

Sea X_1, X_2, \dots, X_n una muestra independiente, y se desea probar que proviene de una normal con media y varianza desconocidas. Sea $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ la muestra ordenada. Esto es: $x_i = \mu + \sigma^2 x_i^*$

La estadística de prueba W' se define como

$$W' = \frac{(\sum_i a_i X_{(i)})^2}{\sum_i (X_{(i)} - \bar{X})^2}$$

en donde

$$a' = (a_1, a_2, \dots, a_n) = \frac{m' I^{-1}}{(m' I^{-1} r^{-1} M)^{\frac{1}{2}}}$$

Sin embargo, los elementos de la matriz I^{-1} están calculados solamente para tamaños de muestra menores o iguales a 20. Valores aproximados para el cómputo de los coeficientes a_i se encuentran en Shapiro y Wilk (1965), y se usa la W' hasta para un valor de n igual a 50.

Una prueba aproximada para probar normalidad, es la W'' de Shapiro-Francia, que es una modificación de la anterior y se define por

$$W'' = \frac{(\sum_i b_i X_{(i)})^2}{\sum_i (X_{(i)} - \bar{X})^2}$$

con

$$b' = (b_1, b_2, \dots, b_n) = \frac{m'}{\sqrt{(m'm)}}$$

En donde $m' = (m_1, m_2, \dots, m_n)$ denota el vector de valores esperados de las estadísticas de orden de una normal estándar. Los valores para m están calculados en Harter (1961).

La estadística W' en una prueba de normalidad de las perturbaciones estocásticas en un modelo de regresión lineal, W'_{mod} .

White y Macdonald (1980) demuestran las condiciones bajo las cuales algunas estadísticas pueden ser utilizadas para probar normalidad de las perturbaciones estocásticas en un modelo de regresión: se reemplaza el error estocástico (no observable) por el residuo de mínimos cuadrados, modificándolas para tamaños de muestra suficientemente grandes.

La fórmula para W' en este caso se denotará como W'_{mod} y se calcula de la siguiente forma:

$$W'_{mod} = \left(\frac{-\frac{1}{2} \sum_i^n \left[\frac{m_i}{(m'_n m'_n)^{\frac{1}{2}}} \right] e_i \right)^2 / s^2$$

en donde:

e_i son los residuos de mínimos cuadrados ordinarios

$$s^2 = n^{-1} \sum_i^n (e_i - \bar{e})^2$$

y $e_{(1)} < e_{(2)} < \dots < e_{(n)}$

son los residuos ordenados.

*Algunos comentarios acerca de las potencias
de las estadísticas A^2 , Jarque-Bera y W .
Potencia de la estadística A^2 cuando
se estiman los parámetros*

Las potencias de las pruebas A^2 , W y W' en general están documentadas para la prueba de normalidad con observaciones independientes e idénticamente distribuidas (iid). En White y Macdonald (1980) se comenta acerca de las potencias de W_{mod} y W'_{mod} en una prueba de normalidad en las perturbaciones estocásticas en un modelo de regresión. En Pierce y Gray (1982) se obtienen resultados de una simulación utilizando las estadísticas A^2 de Anderson-Darling y W de Shapiro Wilk, en un modelo de regresión lineal simple que supone errores iid, para $n = 20$ y $n = 40$, en el que se observa que ambas estadísticas son muy adecuadas. En Shapiro y Wilk (1965) se comparan las estadísticas W y A^2 (entre otras), para el caso de observaciones iid frente a diversas alternativas; los resultados al 5 por ciento y tamaño de muestra igual a 20 son que la W es muy sensible a las distribuciones asimétricas, de *colas largas* (Ji -cuadrada de 1 y 2 grados de libertad, Log-Normal y Cauchy) y en grado menor a las distribuciones de *colas cortas* (Ji -cuadrada con 4 grados de libertad), mientras que la Anderson-Darling tiene una potencia superior a la Crámer von Mises y a la de Kolmogorov-Smirnov, especialmente en el caso de distribuciones de *colas largas* como la Cauchy en la que posee la potencia más alta entre todas las pruebas. Otros comentarios acerca de estas estadísticas aparecen en Jarque y Bera (1987), los que se mencionan párrafos más adelante.

En el caso iid existen teoría asintótica y métodos, a través de los cuales pueden obtenerse resultados de la potencia asintótica de las estadísticas FDE cuando los parámetros se estiman, al menos para el caso de la W^2 , U^2 y A^2 . Stephens (1976b)⁴ encuentra que cuando la prueba es de normalidad o Exponencialidad, con parámetros desconocidos, la estadística A^2 es

⁴ Citado en D'Agostino y Stephens (1986: 169).

ligeramente superior a la de Crámer von Mises (W^2) frente a las alternativas Weibull y Gamma. Otros estudios sobre el mismo tema han sido desarrollados con simulaciones de Monte Carlo; Stephens (1976) encontró, en un análisis de este tipo, que cuando las estadísticas de la FDE se comparan contra los valores percentiles del caso 3 (parámetros desconocidos y estimados), poseen potencias comparables con la prueba W de Shapiro Wilk. La prueba de Kolmogorov ofrece el resultado más pobre. De las estadísticas FDE las mejores son la de Anderson-Darling y la de Crámer von Mises.

Confrontándolas contra las pruebas W y W' , éstas poseen potencias ligeramente superiores a las de A^2 y W^2 , aunque su cálculo es bastante más complicado por requerir el uso de muchos coeficientes.

En este mismo estudio, Stephens calcula las potencias de distintas pruebas, entre éstas la de Anderson-Darling y Shapiro Wilk, para diferentes alternativas; los resultados para las pruebas W y A^2 son los mejores y sugieren una fuerte correlación entre ellas.

Potencia de las estadísticas W y W'

Shapiro y Francia (1972) proponen la estadística W' , que es una aproximación a W . Ellos comparan la sensibilidad de las estadísticas W y W' para detectar no-normalidad, con distribuciones alternativas que no especifican y que, argumentan, son de formas diversas: sesgadas y simétricas, así como distribuciones continuas y discretas. También comentan en su artículo que, en general, la prueba W' es más sensible que la W cuando la distribución alternativa es continua y simétrica, con momento de cuarto orden grande en comparación con la normal cuando se parece a ésta y cuando es discreta y sesgada. Son equivalentes para alternativas continuas asimétricas y discretas asimétricas. La prueba W es superior para todas las demás alternativas. Ambas pruebas poseen potencias muy similares.

Jarque y Bera (1987), con simulaciones de Monte Carlo presentan algunos resultados sobre la prueba de normalidad, tanto para un conjunto de observaciones en el caso independiente e idénticamente distribuido como para el caso de errores estocásticos en un modelo de regresión lineal.

Para el caso de la regresión comparan diversas estadísticas y, por interés del trabajo sólo mencionaremos la \hat{W} , W''_{mod} y la misma Jarque-Bera, la notación «gorro» es porque en su cálculo se utilizan los residuos. El modelo de regresión propuesto tiene un término constante y tres regresores, en donde los residuos son los ordinarios de mínimos cuadrados; utilizan diferentes tamaños de muestra. Los resultados obtenidos son todos a un nivel de significancia pretendido de 10 por ciento.

En el trabajo mencionado, los puntos de rechazo porcentuales que utilizan para las estadísticas mencionadas párrafos arriba, no son los del caso iid simulados en 10 000 réplicas. Los de este ejemplo de regresión fueron generados en 250 réplicas bajo errores normales. Aunque no se mencionan los puntos de rechazo así obtenidos.

Para las cuatro pruebas alternativas, Beta; t de Student; Gamma y Lognormal los resultados que consiguen son: con un tamaño de muestra $n = 20, 35$ y 50 , «probablemente» las mejores pruebas son la $J-B$ y la \hat{W} seguida por la W''_{mod} . Con tamaños de muestra $n = 100, 200$ y 300 , la más potente es la $J-B$ para todas las alternativas, seguida muy de cerca por las W y W''_{mod} ; por ejemplo, para $n = 200$ observan que la potencia $J-B$ fue de $0.928, 0.828, 1.0$ y 1.0 , comparada con $0.924, 0.820, 1.0$ y 1.0 para \hat{W} , para las alternativas antes mencionadas, respectivamente, sin hacer notar que basadas en 250 simulaciones estas proporciones resultan equivalentes.

Asimismo, calculan las potencias de las estadísticas para los tres conjuntos de observaciones del comentario de Weisberg (1980), con un tamaño de muestra igual a 20. Como

resultado encuentran que las estadísticas más potentes son la J - B , la \hat{W} y la W'_{mod} , en ese orden. En el artículo mencionado no se realizó la prueba con estadísticas como la Anderson-Darling, se argumentó que para el caso, iid tiene una potencia inferior a W , aseveración basada en una comparación equivocada de Shapiro Wilk y Chen (1968), corregida posteriormente en Stephens (1974).

Capítulo 3

El propósito de este capítulo es comparar la estadística A^2 de Anderson-Darling con la W'_{mod} de Shapiro-Francia y la $J-B$ de Jarque-Bera cuando la hipótesis que se prueba es la de normalidad de los errores en un modelo de regresión lineal múltiple. La comparación se realiza frente a distintas distribuciones en la alternativa.

Descripción del estudio

En el modelo de regresión:

$$y = X\beta + u$$

en donde: y es un vector de $n \times 1$ observaciones y_i , de la variable dependiente, X es una matriz $n \times k$, de rango completo k , con $X_{ii} = 1$ ($i = 1, \dots, n$), β es un vector de parámetros $k \times 1$ desconocidos, y u es un vector $n \times 1$ de perturbaciones estocásticas, no observables u_i , que se suponen normal e idénticamente distribuidas.

Se quiere comparar el comportamiento de las estadísticas A^2 , $J-B$ y W'_{mod} en la prueba de hipótesis:

H_0 : Las perturbaciones estocásticas u_i poseen una distribución normal con media cero y varianza desconocida.

H_1 : Las perturbaciones estocásticas u_i no poseen una distribución normal

Para efectuar la prueba se consideran en total cinco conjuntos de observaciones, que dependen del tamaño de la muestra (n) y el número de variables independientes (k) en el modelo.

El estudio se organizó de la siguiente forma: cuando $n = 20$, los conjuntos de observaciones en X son: el CONJUNTO1,

CONJUNTO2 y el CONJUNTO3. En cada uno de ellos se trabajan cuatro valores para el número de variables independientes, $k = 4, 6, 8$ y 10 , y para cada uno de estos valores se examina el comportamiento de las estadísticas A^2 , $J-B$ y W'_{mod} que generan las perturbaciones, tanto de la distribución normal como de siete distribuciones alternativas.

Si $n = 50$ el conjunto de observaciones en X se titula W50R, y de la misma forma que para el caso $n = 20$; se utilizan los cuatro valores de k y las mismas distribuciones para las perturbaciones.

Por último, para $n = 90$, la matriz de observaciones X se denota por W90R, en este caso sólo se trabaja con el valor de $k = 10$ y las mismas distribuciones para las perturbaciones que en los casos anteriores. La manera de variar el número k se consigue removiendo las últimas columnas de cada matriz de observaciones.

Los distintos conjuntos de datos aparecen en el apéndice bajo sus títulos respectivos. Las simulaciones de Monte Carlo fueron realizadas en un programa de Fortran.

Como lo hacen notar White y Macdonald (1980) y Weisberg (1980), las pruebas de normalidad dependen del número de variables independientes en el modelo del tipo de «datos» que se tengan para la matriz X , así como del número n de observaciones. Por esta razón, en el estudio de MonteCarlo realizado, las simulaciones se hacen con distintas matrices X , tamaños de muestra n y número de variables independientes, m .

Las distribuciones alternativas

Para los tres tamaños de muestra y sus respectivas variaciones en el valor de m , se examinan ocho distribuciones distintas para los errores: la Normal (N), la Normal Truncada a la derecha (NTD) (valores mayores a 1.64 se hacen iguales a 1.64); la Normal Truncada Simétrica (NTS) (valores menores (mayores) que -1.64 (1.64) se hacen igual a -1.64 (1.64));

la Log-Normal (LN); la Uniforme (0,1) (U); la t , (t de Student con cinco grados de libertad); la Exponencial (E) y la Normal Heteroscedástica (NH), la cual se consigue con una variable aleatoria de distribución Ji-cuadrada con 25 grados de libertad y una variable aleatoria normal estandarizada que transforma a ésta última para que tenga una varianza igual al valor de la variable Ji-cuadrada con 25 grados de libertad.

Para cada corrida y cada distribución se generan, con las rutinas GGUBS, MDSTI, MDNOR, MDTD, GGAMR del IMSL, n errores aleatorios, estandarizándolos a tener una media igual a cero y una varianza igual a 25, con lo cual se respeta lo hecho por White y Macdonald en su artículo.

Los residuos

Una vez que se han generado los errores aleatorios, el siguiente paso consiste en calcular los residuos para cada uno de los modelos considerados que, como se vio en el capítulo 1, se pueden obtener del siguiente modo:

$$e = (I - X(X'X)^{-1} X') u$$

Entonces, para obtener los residuos se multiplican los errores generados mediante las rutinas que ya se mencionaron por la matriz $I - X(X'X)^{-1} X'$, pues de esta forma no hay necesidad de especificar valores para los distintos parámetros de la regresión.

Ahora bien, cada una de las estadísticas A^2 , $J-B$ y W'_{mod} utilizan en su cálculo diferentes tipos de residuos:

Para las estadísticas $J-B$ y W'_{mod} los residuos son los e_i ordinarios.

$$e_i = y_i - \hat{y}_i \quad (1)$$

para la estadística A^2 los residuos estudentizados internos se calculan con los residuos obtenidos en (1), y la fórmula:

$$e_i^* = \frac{e_i}{\sqrt{\frac{n}{n-k-1} (1-x_i'(X'X)^{-1}x_i) \tilde{\sigma}^2 \frac{1}{n-k-1} e_1^2}}$$

en donde $\tilde{\sigma}$ es el estimador de σ de máxima verosimilitud, sin corrección para el sesgo.

Las observaciones en la matriz X

Los tres conjuntos de datos de la matriz X , para un tamaño de muestra igual a 20, se obtuvieron del comentario de Weisberg (1989: 29). El CONJUNTO1 es una matriz de 20 x 9, obtenida de una matriz de 24 x 11 descrita en Cook y Weisberg (1979); el CONJUNTO2, es una matriz de 20 x 9 generada de tal forma que cada entrada se comporte como si hubiera sido obtenida en forma independiente de una $U(0,1)$; por último, el CONJUNTO3 forma parte de una matriz de diseño 2 x 2 x 5.

El conjunto W50R, que corresponde a las observaciones en la matriz X para $n=50$, se obtuvo replicando una vez y media las observaciones de la matriz del CONJUNTO3, hasta obtener 50, de tal forma que la estructura de esa matriz se mantuviera en el conjunto W50R. Más adelante se presentan los resultados obtenidos para el CONJUNTO3 y valores de $k=8$ y $k=10$, que fueron sorprendentes, por lo que se decidió aumentar el tamaño de la muestra (denominando a este nuevo conjunto como W50R) y verificar estos resultados para un tamaño de muestra mayor.

Las potencias empíricas

La potencia de cada una de las pruebas se calcula dividiendo el número de rechazos obtenidos con las distribuciones alternativas a la normal, y entre el total de simulaciones realizadas.

Se efectuaron 200 simulaciones para cada tamaño de muestra, conjunto de observaciones, distribución del error y cada valor de k , registrándose el número de rechazos para las tres estadísticas: A^2 , Jarque-Bera y W''_{mod} a un nivel de significancia de 10 por ciento.

Los puntos de rechazo

Un aspecto muy importante que se debe considerar es la tabla de los valores de rechazo, contra los que se compara el resultado de cada una de las estadísticas calculadas.

Para la estadística A^2 , la tabla que se utilizó es la que aparece en la tabla 1.1. Los valores que aparecen en la tabla son los calculados para el caso 3 (D'Agostino y Stephens, 1986: 123) que corresponde a la tabla de puntos porcentuales para una normal con media y varianza desconocidas. Los puntos de rechazo son los de su distribución asintótica, y como se puede observar de la tabla, los valores son únicamente los que corresponden a los distintos niveles de significancia: no se requiere modificarlos de acuerdo al tamaño de la muestra, ya que esta estadística converge rápidamente en el caso independiente e idénticamente distribuido y se espera que aquí ocurra lo mismo

Tabla 1.1

Puntos porcentuales para una prueba de normalidad con μ y σ^2 desconocidos

	Nivel de significancia						
	.50	.25	.15	.10	.05	.025	.001
A^2	.341	.470	.561	.631	.752	.873	1.035

Para la estadística Jarque-Bera la tabla de percentiles⁵ se presenta en la tabla 1.2. De la misma forma que para la tabla de la A^2 , los puntos de rechazo de $J-B$ son para el caso de observaciones independientes e idénticamente distribuidas y no las de los residuos en la regresión. Estos valores fueron simulados en 10 000 réplicas para obtener, bajo normalidad en las observaciones, los puntos significativos.

Tabla 1.2

Normalidad en las observaciones; puntos significativos para pruebas LM_n de normalidad; 10 000 réplicas

N	0.10	0.05	N	0.10	0.05
20	2.13	3.26	200	3.48	4.43
30	2.49	3.71	250	3.54	4.51
40	2.70	3.99	300	3.68	4.60
50	2.90	4.26	400	3.76	4.74
75	3.09	4.27	500	3.91	4.82
100	3.14	4.29	800	4.32	5.46
125	3.31	4.34	∞	4.61	5.99
150	3.43	4.39			

⁵ Tomada de Jarque y Bera (1987: 169).

Por último, algunos de los percentiles para comparar a la estadística W''_{mod} aparecen en la tabla 1.3.⁶ Estos son valores aproximados mediante un estudio empírico.

Tabla 1.3
Puntos porcentuales empíricos de la prueba W' aproximada

n	P										
	.01	.05	.10	.15	.20	.50	.80	.85	.90	.95	.99
35	.918	.943	.952	.956	.964	.976	.982	.985	.987	.989	.992
50	.935	.953	.963	.968	.971	.981	.987	.988	.990	.991	.994
55	.940	.958	.966	.971	.973	.983	.988	.990	.991	.992	.994
65	.948	.965	.971	.974	.976	.985	.990	.981	.992	.993	.995
75	.956	.969	.973	.976	.979	.986	.991	.992	.993	.994	.995
85	.961	.972	.977	.980	.981	.988	.992	.992	.993	.994	.996
89	.961	.972	.977	.981	.982	.988	.992	.993	.994	.995	.996
91	.962	.973	.978	.981	.983	.989	.992	.993	.994	.995	.996
99	.967	.976	.980	.982	.984	.989	.993	.994	.994	.995	.996

Resultados

En la tabla 2 aparece el porcentaje en 200 simulaciones, con $n = 20$ para los que la hipótesis nula de normalidad fue rechazada a un nivel de significancia $\alpha = 0.10$.

La tabla está organizada de la siguiente manera: en la primera columna aparece el número del conjunto; en la segunda columna el número de variables independientes k , a la que nos referimos; en la tercera columna aparecen las estadísticas A^2 , $J-B$ y W''_{mod} , las columnas restantes tienen el nombre de las distribuciones contra las que se realiza la prueba: N para normal que corresponde a la hipótesis nula; NH para normal heteroscedástica; t_5 para la distribución t de Student con 5 grados de libertad; LN para la distribución Log-Normal; E para Exponencial; U para Uniforme; NTD para Normal Truncada Derecha y, por último, NTS para Normal Truncada Simétrica.

⁶ Tomados de Shapiro y Francia (1972: 216).

Tabla 2
Resultados de la simulación Porcentaje de
rechazos en 200; $n = 20$, estadísticas A^2 , J-B y W'_{mod}

conj.	k=No Col.		N	NH	t_5	LN	E	U	NTD	NTS
		A^2	13	13.5	21	80	59	16.5	12	12.5
	4	J-B	10.5	17.5	25	80	61.5	9	9	7
		W'_{mod}	10	10	18.5	91	51.5	8	7.5	4
		A^2	14.5	15	21	69	51.5	18	15.5	13
	6	J-B	12.5	14	24	73	58	8	12	7
		W'_{mod}	11	6.5	21.5	72.5	50.5	9.5	10	7.5
1		A^2	13	10.5	16	53	38	16	11	11
	8	J-B	13	12	23.5	58	43.5	6	12	10.5
		W'_{mod}	8	15.5	15	59	42.5	6.5	8	5.5
		A^2	12.5	11.5	12.5	39	25	11.5	15.5	14
	10	J-B	13.5	16	20.5	43.5	41.5	11.5	14	13.5
		W'_{mod}	8	11.5	17.5	56.5	29	8.5	9	7
		A^2	9	14.5	18.5	84	60	19	9.5	9
	4	J-B	9.5	16	28.5	81.5	57	5.5	7	7
		W'_{mod}	9	8	16.5	87	62	6	8.5	6.5
		A^2	8	9.5	16	66.5	36.5	13.5	7.5	7.5
	6	J-B	8	13.5	24.5	69.5	40.5	5.5	8	6.5
		W'_{mod}	10	13	18.5	70	49.5	6.5	9	5
2		A^2	8.5	7.5	14	51.5	28.5	9.5	7.5	9
	8	J-B	8.5	13.5	22.5	54.5	34.5	6	9	7.5
		W'_{mod}	9	5.5	16	52	33	3.5	7	6.5
		A^2	11	10.5	15.5	41.5	18.5	13	9.5	10.5
	10	J-B	11	11	18.5	41	26	11	11.5	10.5
		W'_{mod}	8	5.5	12.5	41	21	6.5	7.5	7.5
		A^2	7.5	15	19.5	79.5	59	10	7.5	7.5
	4	J-B	11.5	13.5	27.5	77	57.5	5.5	6	4.5
		W'_{mod}	10	11	19.5	86	52	13.5	8	5.5
		A^2	11.5	13	20.5	62	45.5	12	13.5	12
	6	J-B	13.5	11.5	22	63.5	50	7	10.5	7
		W'_{mod}	10.5	11	16.5	77	42.5	5	7	6.5
3		A^2	37	40.5	42	79.5	69	41	36.5	35.5
	8	J-B	21.5	26	31.5	59.5	50.5	11.5	19.5	15
		W'_{mod}	31.5	27	36.5	71.5	53.5	16	29	24
		A^2	37.5	42.5	46	74.5	57	40.5	40	40.5
	10	J-B	22.5	35.5	30	49	42.5	16	19	18
		W'_{mod}	30	28	36.5	63.5	43	4	28.5	26.5

Los resultados para la distribución normal en los tres conjuntos de observaciones de la matriz X son estimadores del error de Tipo I para las simulaciones de Monte Carlo. Debido a que el número de rechazos posee una distribución Binomial con parámetros p y N , en donde p es la probabilidad de rechazar la hipótesis nula de normalidad; la proporción

$$\frac{N_r \text{ de Rechazos}}{N}$$

tiene una media igual a p y desviación estándar

$$\sqrt{\frac{p(1-p)}{N}}$$

en donde $N =$ No. de simulaciones efectuadas.

Para un nivel de significancia $\alpha = 0.10$, esta desviación es igual a 0.02121, es decir, aproximadamente igual al 2 por ciento y, la cota del error, utilizando la aproximación normal al 95 por ciento es igual a $1.96(2) = \pm 3.92$ por ciento (aproximadamente igual a 4 %). Para las estadísticas \mathcal{A}^2 , $J-B$ y W'_{mod} , el alfa empírico de las pruebas se mantuvo en este intervalo para todos los conjuntos de datos y valores de k , excepto para el CONJUNTO3 con $k=8$ y $k=10$, en donde los valores respectivos ascendieron en \mathcal{A}^2 a 37 y 37.5 por ciento, respectivamente; con la estadística $J-B$ a 21.5 por ciento y 22.5 por ciento; mientras que para la W'_{mod} estos valores fueron del 31.5 y 30 por ciento, respectivamente.

Es decir, que los puntos de rechazo usados para las tres estadísticas parecen ser una buena aproximación, con excepción de los casos mencionados del CONJUNTO3, en los que evidentemente, la forma de la matriz X tiene un efecto muy grande.

Se subraya que el valor del alfa empírico (es decir, bajo normalidad) de la estadística \mathcal{A}^2 se mantuvo en el nivel correspondiente a la cota de 4 por ciento, aun cuando los

valores de rechazo consultados corresponden al de su distribución asintótica y, como ya se mencionó, esta estadística converge muy rápidamente. Mientras que para las estadísticas $J-B$ y W'_{mod} , los puntos de rechazo han sido simulados y varían de acuerdo al tamaño de la muestra. De haber utilizado el valor de rechazo asintótico para la $J-B$, igual a 4.61, que comparado con el valor de rechazo utilizado para $n = 20$ igual a 2.13 es mucho mayor, los resultados para esta última serían diferentes.

Para las alternativas, los valores de la tabla ofrecen un estimador de la potencia para cada prueba esperando que posean valores mayores que el reportado bajo la normal. La comparación para las tres pruebas se hace con respecto al valor del alfa empírico, tomando en consideración la cota del error, siempre y cuando este valor se encuentre contenido en el intervalo al que ya nos referimos.

La Normal Heteroscedástica. Con esta alternativa observamos un comportamiento parecido al de la Normal. Las tres pruebas presentan una potencia baja, con un valor semejante al del tamaño de la prueba que es igual al 10 por ciento, no se observa una clara predominancia de alguna de las tres sobre las demás. A continuación tenemos la distribución t_3 , de *colas* moderadas, para esta opción, la potencia varía entre 12.5-21 para la A^2 ; 18.5-28.5 para la $J-B$ y 12.5-21.5 para la W'_{mod} , dependiendo del valor de k y de las observaciones en X ; se puede advertir que la estadística $J-B$ tiene una potencia superior a las otras dos para todos los valores de k (número de variables explicativas) y X , seguida de la A^2 y la W'_{mod} . Las siguientes dos alternativas, Log-Normal y Exponencial se caracterizan por poseer *colas* muy alargadas. Las tres estadísticas presentan una buena potencia, con resultados equivalentes frente a estas alternativas, mejor para la Log-Normal que para la Exponencial, variando las potencias en la primera entre 39-80 para la A^2 , 41-81 para la $J-B$ y 41-91 para la W'_{mod} , con los distintos valores de k y X . Frente a la alternativa LN la W'_{mod} es la de mayor potencia, en general, seguida

de la $J-B$ y muy de cerca la A^2 . Para la Exponencial la variación de las potencias se encuentra entre 18.5-60 para la A^2 , 26-61.5 para la $J-B$ y 21-62 para la W'_{mod} , con la $J-B$ como la más potente, seguida de la W'_{mod} y la A^2 . La alternativa Uniforme posee una distribución de *colas* muy cortas. Frente a ésta la mejor prueba resulta la A^2 , con una potencia baja, que varía entre 9.5-19, las otras dos estadísticas funcionan muy por debajo de ésta con una potencia para casi todos los casos, por debajo del error de tipo I empírico: 5.5-11.5 la $J-B$; 3.5-13.5 la W'_{mod} . En las distribuciones Normales Truncadas, las potencias son bajas para las tres estadísticas, variando entre 7.5-15.5 para la A^2 ; 4.5- 14 la $J-B$ y 4-10 la W'_{mod} , siendo la primera la «mejor».

En la tablas 3 y 4 aparecen los resultados de las simulaciones en las estadísticas A^2 , $J-B$ y W'_{mod} , para los tamaños de muestra $n = 50$ y $n = 90$; el conjunto de observaciones W50R corresponde al tamaño de muestra $n = 50$, en donde también se trabajan los cuatro valores de k , mientras que el conjunto W90R tiene 90 observaciones y sólo se utiliza $k = 10$.

Los valores del alfa empírico para la las estadísticas A^2 y W'_{mod} se encuentran en el rango establecido por la cota del error, no sucede lo mismo con la estadística $J-B$ cuyo valor está por encima de ésta en los dos conjuntos de observaciones y todos los valores de k . Esto puede sugerir que requiera cierto ajuste con respecto a los grados de libertad utilizados.

Tabla 3
Resultados de la simulación Porcentaje de rechazos
en 200; $n=50$, para las estadísticas A^2 , J-B y W'_{mod} .

conj	k=NO COL		N	NH	t_s	LN	E	U	NTD	NTS
		A^2	9.5	9	36.5	100	99	46	11.5	11.5
	4	J-B	14.5	13.5	53.5	100	99	35	14.5	7.5
		W'_{mod}	13	13	44.5	100	99.5	40	15.5	8.5
		A^2	10.5	11	36	99	96.5	35	12.5	9.5
	6	J-B	14	14	51	100	98.5	22	11.5	4
		W'_{mod}	13	13	46	100	98.5	25.5	11	7.5
W50R		A^2	9	11.5	31.5	98.5	95.5	31.5	10.5	10
	8	J-B	14	15	50	99	96.5	15	9.5	4.5
		W'_{mod}	13	12.5	44.5	99.5	96.5	16	11.5	5.5
		A^2	10	8.5	29.5	97.5	91.5	24.5	9	9.5
	10	J-B	13	18.5	49.5	98	95	9.5	11	2
		W'_{mod}	12	14.5	43	98.5	93	10.5	8.5	3

Tabla 4
Resultados de la simulación Porcentajes de rechazos
en 200 $n=90$, para las estadísticas A^2 , J-B y W'_{mod} .

Conj.	K= No. Col.		N	NH	t^s	LN	E	U	NTD	NTS
		A^2	8.5	15	47.5	100	99.5	71.5	14	18
W90R	10	J-B	14	18	64	100	100	76.5	17	21
		W'_{mod}	11.5	17	64.5	100	100	67.5	18	18

El comportamiento bajo la alternativa NH , como en el caso de $n = 20$, sigue un patrón parecido al de la distribución normal. En las alternativas t , Log-Normal, Exponencial y Uniforme sí se observa un incremento en las potencias para las tres estadísticas, sobre todo para la A^2 , tomando en cuenta que posee el alfa empírico más pequeño de las tres en los dos tamaños de muestra, conjunto de observaciones y distintos valores de k . En general, se puede decir que entre todas las alternativas, en la única en la que la estadística $J-B$ es la más potente es para la t , aunque insistimos, el valor del alfa empírico para esta estadística fue el más alto de las tres. Para las alternativas truncadas, y $n=50$, sólo se percibe un incremento poco significativo para la A^2 , las otras dos, $J-B$ y W'_{mod} , descendieron, en potencia, a un nivel por debajo del alfa empírico encontrado (observar las potencias de ambas en el conjunto W50R con $k=10$), aumentando en estas mismas alternativas, cuando el tamaño de muestra es igual a 90.

Es importante señalar que en el estudio de simulación efectuado, tanto la $J-B$ como la W'_{mod} reportaron potencias inferiores al alfa empírico. Esto ocurrió por ejemplo cuando el tamaño de muestra es igual a 20 en la alternativa Uniforme, en el CONJUNTO2 con $k=8$ y para casi todos los casos en la Normal Truncada Simétrica (véase tabla 2); para esta misma alternativa, cuando el tamaño de la muestra asciende a 50 esto se percibe más claramente como ya fue mencionado. Lo anterior muestra que resultaron ser pruebas sesgadas. Es decir, si se establece como condición que en la prueba de hipótesis

$H_0: x_1, x_2, \dots, x_n$ posee una distribución F

vs.

$H_1: x_1, x_2, \dots, x_n$ no posee una distribución F

para ninguna alternativa en H , la probabilidad de rechazo sea menor que el tamaño de la prueba (α). Tenemos que una prueba es insesgada si la potencia (POT) satisface que

$$\text{POT} \leq \alpha \text{ bajo la hipótesis nula}$$

y

$$\text{POT} \geq \alpha \text{ en la alternativa.}$$

Capítulo 4

Resumen y conclusiones

El propósito de este trabajo es explorar el comportamiento de la estadística de Anderson y Darling (A^2) en una prueba de bondad de ajuste para probar normalidad en las perturbaciones estocásticas, en un modelo de regresión lineal múltiple; comparando su potencia frente a siete alternativas, con otras dos estadísticas, la de Jarque y Bera ($J-B$) y la de Shapiro y Francia W'_{mod} .

La estadística Anderson-Darling pertenece a la familia de Crámer von Mises y tiene la característica de estar basada en la función de distribución empírica (FDE). Las pruebas de bondad de ajuste que utilizan a las estadísticas de este tipo comparan la función de distribución $F(x)$, bajo la hipótesis nula, con la función de distribución empírica $F_n(x)$.

Cuando la función $F(x)$ es continua y está completamente especificada, se sabe que las estadísticas FDE son más potentes bajo H_0 , que otras pruebas, por ejemplo la Ji-cuadrada, con la advertencia de poner atención en la tabla de puntos de rechazo cuando los parámetros de la distribución deben ser estimados (Stephens, 1974: 730).

Las tablas de puntos de rechazo se catalogan de acuerdo al conocimiento que se tenga de la distribución bajo H_0 :

Caso 0: $F(x)$ es continua y está completamente especificada.

Caso 1: $F(x)$ es la distribución normal, σ^2 es conocida y se estima μ .

Caso 2: $F(x)$ es la distribución normal, μ es conocida y se estima σ^2 .

Caso 3: $F(x)$ es la distribución normal y ambos parámetros son desconocidos y deben estimarse.

Para cada uno de estos casos se han calculado los puntos de rechazo de la distribución asintótica a partir de observaciones independientes e idénticamente distribuidas (D'Agostino y Stephens, 1986: 105, 122 y 123). Sin embargo, el problema que nos interesa es el siguiente: en el modelo de regresión

$$y = X\beta + v$$

queremos probar la hipótesis de que las perturbaciones estocásticas v , provienen de una distribución normal con media igual a cero y varianza desconocida.

En este caso, las perturbaciones estocásticas no son observables, además de no ser independientes y de poseer, en general, varianzas distintas. Mukantseva (1977) y Pierce y Kopecky (1979) demuestran que las pruebas de normalidad, basadas en la FDE calculadas con los residuos de una regresión simple o múltiple, tienen la misma distribución asintótica bajo la hipótesis nula que para el caso de observaciones independientes e idénticamente distribuidas con media y varianza desconocidas. *Lo anterior, siempre y cuando en la regresión exista un término constante.*

Este es un resultado muy importante que nos asegura que para el problema propuesto, la tabla de puntos de rechazo que debe consultarse es la del caso 3; consiste de un solo renglón con los percentiles correspondientes a los diferentes valores de significancia, y es de esperarse, que al igual que en el caso de observaciones independientes e idénticamente distribuidas, sea posible utilizarla para una n no exageradamente grande.

En cuanto a las tablas de percentiles de las estadísticas $J-B$ y W'_{mod} tenemos que los puntos de rechazo son los del caso de observaciones independientes e idénticamente distribuidas, calculados a partir de simulaciones, en réplicas bajo normalidad. Por ejemplo, de acuerdo con la tabla cuando

$n=20$ el punto de rechazo, a un nivel de significancia de 10 por ciento para $J-B$ es igual a 2.13 y su valor asintótico igual a 4.61; mientras que para W''_{mod} en la tabla consultada no existen valores asintóticos. Entonces, en el caso de $J-B$, si la prueba se realiza comparando contra los valores asintóticos, los resultados obtenidos seguramente serían diferentes y existe la posibilidad de que la potencia de la prueba se vería disminuida.

Los resultados de las simulaciones efectuadas aparecen en el cuerpo del trabajo, en las tablas 2, 3 y 4, que corresponden a los tamaños de muestra $n = 20$, $n = 50$ y $n = 90$, respectivamente.

El nivel de significancia empírico (es decir bajo normalidad), para la estadística A^2 se mantuvo para todos los tamaños de muestra, excepto para el CONJUNTO3 con $k=8$ y $k=10$ en los límites establecidos por la cota del error al 90 por ciento a pesar de que, como ya se mencionó, los puntos de rechazo contra los que se compara son los asintóticos. Para el CONJUNTO3 referido, en donde las tres estadísticas tuvieron resultados no esperados, parecen deberse a que la matriz X , el número de variables independientes en el modelo, así como el tamaño de la muestra influyeron en su comportamiento. Esto concuerda con los obtenidos en el comentario de Weisberg al artículo de White y Macdonald (1980).

Más aún, cuando el tamaño de la muestra aumenta, el nivel empírico de la estadística A^2 resultó más pequeño a los encontrados para las estadísticas $J-B$ y W''_{mod} , en todos los casos estudiados, y si la comparación se realiza tomando este punto en consideración resulta que conforme el tamaño de la muestra aumenta la estadística A^2 es la más potente, excepto para el caso de la alternativa t_c .

Lo que llama más la atención del estudio de simulación, fue descubrir que tanto la $J-B$ como la W''_{mod} reportaron potencias inferiores al alfa empírico. Esto ocurrió por ejemplo para la alternativa Normal Truncada Simétrica (NTS) en el

conjunto W50R con $k=10$. Lo anterior muestra que resultaron pruebas sesgadas.

Considerando que los resultados obtenidos en este trabajo están circunscritos a las observaciones, tamaños de muestra y número de variables independientes en el modelo de regresión, podemos decir que la estadística A^2 no sólo compite razonablemente con las estadísticas $J-B$ y W'_{mod} , sino que en algunos casos resulta más potente, como en el caso antes mencionado. Su alfa empírica, aun para tamaños de muestra pequeños ($n=20$) se mantuvo en los límites establecidos por la cota del error, a pesar de que sus puntos de rechazo son los asintóticos. Es la prueba más sencilla, en cuanto a la utilización de los percentiles para concluir en el rechazo o no de la hipótesis nula. En el estudio de Monte Carlo realizado no hubo necesidad de cambiar el punto de rechazo, ya que éste sólo se modifica si cambia el nivel de significancia de la prueba, que en este caso se mantuvo constante e igual al 0.10; mientras que para las otras dos estadísticas el punto de rechazo varía con el tamaño de muestra. Esto se debe considerar para el uso de la estadística $J-B$ en algunos paquetes de econometría como el Econometric View, cuyo valor de rechazo corresponde al de una Ji-cuadrada con 2 grados de libertad que coincide con el valor asintótico de 4.61 que los autores señalan, y no se estipula que para el adecuado comportamiento de la $J-B$ hay que utilizar los percentiles del caso de observaciones independientes e idénticamente distribuidas, o bien, como lo hacen Jarque y Bera (1987: 170), generarlos por Monte Carlo; ya que en algunos casos, este valor lleva al no rechazo, incorrecto, de la hipótesis nula.

Quizá la desventaja que pueda tener la A^2 frente a la $J-B$ sea la necesidad de transformar las observaciones bajo la hipótesis nula; aunque el cálculo de los tercero y cuartos momentos, necesarios para el cómputo de $J-B$ requiere también de una rutina en particular. La W'_{mod} es la más complicada de las tres en cuanto a su cálculo, ya que requiere

de una tabla específica de coeficientes que varían de acuerdo al tamaño de la muestra.

Las estadísticas de la FDE, en particular la A^2 , no han sido estudiadas en literatura, en regresión. Sirva este modesto trabajo como una forma de acercarnos a su comportamiento.

Bibliografía

- Anderson, T.W. y D. A. Darling (1952), «Asymptotic theory of certain goodness of fit criteria based on stochastic processes», en *Annals of Mathematical Statistics* 23, pp. 193-212.
- (1954), «A test of goodness fit», en *Journal American Statistics Association* 49, pp. 765-769.
- Arnold, S.F. (1980), «Asymptotic validity of F test for the ordinary linear model and multiple correlation model», en *Journal American Statistics Association* 75, pp. 890-894.
- Cook, R. Dennis y S. Weisberg (1982), *Residuals and Influence in Regression*, Chapman and Hall, cap. 2.
- D'Agostino, R. B. y M. A. Stephens (1986), *Goodness of fit techniques*, Marcel Dekker, cap. 4.
- Dimitry, D. y T. Mott, Jr. (1972), *Introducción a la programación mediante Fortran IV*, Interamericana.
- Easterling, R. G. «Goodness of fit and parameter estimation», en *Technometrics*, vol. 18, núm. 1.
- Harter, H.L. (1961), «Expected values of Normal statistics», en *Biometrika* 48, 1 y 2, pp. 151-165.
- H. William (1986), *Numerical recipes. The art of scientific computing*, Cambridge University Press.
- Hogg, R. y A. Craig (1978), *Introduction to Mathematical Statistics*, 4ª ed., Macmillan.
- Jarque, C. M. y A. K. Bera (1982), «Model specification tests», en *Journal of Econometrics* 20, pp. 59-82.
- (1987), «A test for Normality of Observations and Regression Residuals», en *International Statistical Review*, pp. 163-172.
- Lewis, P. A. W. (1961), «Distribution of the Anderson Darling Statistic», en *Annals of Mathematical Statistics* 32, pp. 1118-1124.
- Lindgren, B. W. (1968), *Statistical Theory*, Macmillan.
- Loynes, R. A. (1980), «The empirical distribution function of residuals from generalised regression», en *Annals of Mathematical Statistics* 8, pp. 285-298.

- Maddala, G. S. (1985), *Econometria*, Mc Graw-Hill.
- Mukantseva, L. A. (1977) «Testing normality in one dimensional and multidimensional linear regression», en *Theory in Probability Applications* 22, pp. 591-519
- O'Reilly, F. (1990), «Algunas consideraciones sobre la inferencia estadística», en *Ciencia* 41, pp. 29-43.
- (1992), *Proceedings Conference on Statistical Inference and Biostatistics*, pp. 179-200.
- y C. P. Quesenberry (1973) «The Conditional Probability Integral Transformation and Applications to obtain Composite Chi-Square Goodness-of-Fit-Tests», en *The Annals of Statistics*, pp. 74-83.
- Pierce, D. A. y R. J. Gray (1982), «Testing normality of errors in regression models», en *Biometrika* 69, pp. 233-236
- y K. J. Kopecky (1978), «Testing goodness-of-fit for the distribution of errors in regression models», en *Technical Report Symp. 16*, Department of Statistics, Stanford University.
- Rényi, A. (1970), *Probability Theory*, North Holland Amsterdam, Publishing, p. 400.
- Seal, H. L. (1967), «Studies in the history of Probability and Statistics, XV. The historical development of the Gauss linear Model», en *Biometrika* 54, 1 y 2, p. 1.
- Shapiro, S. S. y R. S. Francia (1972) «An approximate Analysis of Variance Test for Normality», en *Journal of the American Association* 67, pp. 215-216.
- y Wilk, M. B. (1965) «An analysis of variance test for Normality», en *Biometrika* 52, pp. 591-611.
- , Wilk, M. B. y H. J. Chen (1968), «A comparative study of various tests for normality», en *Journal of American Statistical Association* 63, pp. 1343-1372.
- Spanos, A. (1986), *Statistical Foundations of econometric modelling*, Cambridge University Press, p. 450.
- Stephens, M. A. (1969), «Test for Normality», en *Technical Report*, núm. 152, Department of Statistics, Stanford University.

- (1970), «Kolmogorov-Type Tests for Exponentiality when the scale parameter is unknown», en *Technical Report*, núm. 152, Department of Statistics, Stanford University.
- (1974), «EDF statistics for goodness of fit and some comparisons», en *Journal American of Statistics Association* 69, pp. 730-737.
- (1976), «Asymptotic results for goodness-of-fit statistics with unknown parameters», en *The Annals of Statistics* 4, pp. 357-369.
- (1976b), «Asymptotic power of EDF statistics for exponentiality against Gamma and Weibull alternatives», en *Technical Report*, núm. 297, Department of Statistics, Stanford University.
- Weisberg, S. (1980), «Comment», en *Journal of the American Statistics Association* 75, pp. 28-31.
- White, H. y G. M. Macdonald (1980), «Some large sample tests for non-normality in the regression model», en *Journal American Statistics Association* 75, pp. 16-28.



Para verificar que un conjunto de observaciones proviene de una distribución normal se utilizan diversos métodos. Desde los más sencillos como los métodos gráficos, las pruebas de bondad de ajuste más conocidas como Ji-cuadrada y la de Kolmogorov-Smirnoff, y otras como la de Jarque-Bera.

Las pruebas que se apoyan en la función de distribución empírica, como la estadística de Anderson-Darling, o A^2 , han demostrado ser más potentes que las ya mencionadas en determinadas situaciones. En este ensayo se estudia el comportamiento de la estadística A^2 al probar normalidad de los residuales en un problema de regresión, verificando su potencia frente a distintas alternativas; al mismo tiempo que se le compara, en las mismas circunstancias, con otras dos estadísticas: la de Jarque-Bera y la de Shapiro-Francia modificada.

CLARA MARTHA ADALID DIEZ DE URDANIVIA, maestra en ciencias con especialidad en estadística por la UNAM, es profesora titular del Departamento de Política y Cultura de la UAM-Xochimilco.