

Índice general

5. El modelo de regresión lineal simple univariante. Análisis de residuos. Transformaciones en regresión lineal simple.	3
5.1. Introducción	3
5.2. Residuos y normalidad.	6
5.2.1. Gráfica probabilística normal.	6
5.2.2. Contraste de Shapiro-Wilks	9
5.2.3. Transformaciones en busca de normalidad	12
5.3. Residuos y heterocedasticidad.	17
5.3.1. Caso especial de heterocedasticidad. Mínimos Cuadrados Ponderados. . .	19
5.4. Residuos y autocorrelación.	21
5.4.1. Procedimientos gráficos.	22
5.4.2. Contraste de autocorrelación de primer orden: Test de Durbin-Watson. . .	23
5.4.3. Estimación bajo autocorrelación de primer orden: procedimiento de Cochran-Orcutt.	26

5.5. Residuos y datos anómalos.	28
5.5.1. Test de valores atípicos	30

Capítulo 5

**El modelo de regresión lineal simple univariante. Análisis de residuos.
Transformaciones en regresión lineal simple.**

5.1. Introducción

Hasta este momento hemos trabajado con el modelo de regresión lineal simple univariante pero suponiendo siempre el hecho de que las hipótesis básicas del mismo son ciertas, por lo que surge de inmediato la cuestión de cómo intentar comprobar si dichas hipótesis se verifican y, caso de no verificarse, cómo actuar ante ese hecho.

Como sabemos, las principales hipótesis sobre el modelo se expresan en términos de las perturbaciones aleatorias ϵ_i , $i = 1, \dots, N$. No es de extrañar, por lo tanto, que sean los residuos los que nos van a ayudar a clarificar un poco la situación.

Para ponernos un tanto en ambiente veamos un ejemplo propuesto por Anscombe en 1973:

x	y_1	y_2	y_3	x_d	y_d
10	8,04	9,14	7,46	8	6,58
8	6,95	8,14	6,77	8	5,76
13	7,58	8,74	12,7	8	7,71
9	8,81	8,77	7,11	8	8,84
11	8,33	9,26	7,81	8	8,47
14	9,96	8,1	8,84	8	7,04
6	7,24	6,13	6,08	8	5,25
4	4,26	3,1	5,39	19	12,5
12	10,84	9,13	8,15	8	5,56
7	4,82	7,26	6,42	8	7,91
5	5,68	4,74	5,73	8	6,89

Se puede comprobar que las regresiones de Y_1 , Y_2 e Y_3 sobre X y la de Y_d sobre X_d producen la misma recta

$$\hat{y} = 3 + 0,5x$$

con un análisis de la varianza igual en los cuatro casos. A la vista de ello se podría deducir que las cuatro regresiones parecen ser idénticas. No obstante, estudiando los residuos de cada regresión, que son diferentes obviamente, se observa que cada caso es bien distinto.

Por ejemplo, observando las gráficas de los valores estimados frente a los residuos podemos comentar lo siguiente:

- El primer modelo no ofrece ninguna clara evidencia de anomalías importantes.
- El segundo modelo refleja una falta clara de linealidad pues se observa una tendencia curvilínea.
- El tercer modelo presenta una observación atípica que seguramente determina la estimación de la recta.
- El cuarto modelo presenta también un valor atípico pero, a diferencia del tercer modelo, se puede observar que la pendiente de la recta viene determinada únicamente por un punto, lo cual muestra muy poca fiabilidad en el empleo de dicha recta con fines predictivos.

El análisis de los residuos tiene por objeto contrastar a posteriori las hipótesis del modelo lineal. Su uso se hace más necesario cuando, al tener un único valor de la variable dependiente para cada valor de la explicativa, los contrastes básicos de homocedasticidad, normalidad e independencia no pueden ser realizados a priori. Así pues este análisis va encaminado a comprobar:

- Si hay algún motivo para suponer no linealidad en el modelo.

- Si la distribución de los errores es aproximadamente normal.
- Si la varianza de los errores es constante (homocedasticidad).
- Si existe evidencia para rechazar la incorrelación de los errores.
- Si existe alguna observación anómala que está alterando la estimación y que, tal vez, debiera ser eliminada en el estudio.

5.2. Residuos y normalidad.

5.2.1. Gráfica probabilística normal.

Para verificar la hipótesis de normalidad existen toda una serie de contrastes propuestos por diversos autores. Algunos de ellos son de carácter general (contrastos de bondad de ajuste) y otros son específicos. De estos últimos sólo vamos a hacer referencia al contraste de Shapiro y Wilks. No obstante, en primer lugar vamos a exponer un procedimiento gráfico muy usual en la práctica y que trata de verificar, de forma gráfica, si existe evidencia para intuir la violación de la hipótesis de normalidad y que recibe el nombre de gráfica probabilística normal.

Expondremos a continuación la idea general de este tipo de gráfica y posteriormente la aplicaremos al caso concreto de la regresión.

Sea X una variable de la cual se poseen n observaciones y se desea investigar la posibilidad de que dichos valores procedan de una distribución normal de media μ y varianza σ^2 .

Suponiendo que dichos valores procedieran de una distribución normal con media μ y varianza σ^2 , los valores tipificados y posteriormente ordenados

$$\frac{x_{(1)} - \mu}{\sigma} \leq \frac{x_{(2)} - \mu}{\sigma} \leq \dots \leq \frac{x_{(n)} - \mu}{\sigma}$$

constituirán una muestra ordenada procedente de una distribución normal de media cero y varianza uno, cuyos valores esperados están tabulados. Sea $C_{i,n}$ el valor esperado del término que ocupa el lugar i en una muestra de tamaño n de una población normal, o sea

$$C_{i,n} = E \left[\frac{x_{(i)} - \mu}{\sigma} \right]$$

con lo cual es evidente que

$$E[x_{(i)}] = \mu + \sigma C_{i,n}$$

y por lo tanto el gráfico de $x_{(i)}$ frente a los valores $C_{i,n}$ será una recta cuya ordenada en el origen estimará el valor de μ y la pendiente estimará el valor de σ . Así se podría emplear dicha gráfica para estimar la media y la desviación típica. Para ello se ajustaría una recta de regresión a la nube de puntos $\{C_{i,n}, x_{(i)}\}_{i=1,\dots,n}$.

Evidentemente este ajuste presenta una serie de problemas y es que los estadísticos ordenados pierden la independencia que tenía la muestra original.

Por otro lado, el cálculo de los coeficientes $C_{i,n}$ es complicado. Se han propuesto varias formas de aproximar sus valores, siguiendo todos el esquema

$$C_{i,n} = \Phi^{-1} \left(\frac{i - c}{n - 2c + 1} \right)$$

donde Φ es la función de distribución de la normal estándar.

Hoy día no se emplea esta gráfica para estimar los parámetros μ y σ^2 . Por ello es frecuente representar los valores $C_{i,n}$ frente a los $x_{(i)}$. Statgraphics representa en su lugar la nube de puntos $\{x_{(i)}, \Phi(C_{i,n}) \cdot 100\}$, donde

$$C_{i,n} = \Phi^{-1} \left(\frac{i - \frac{3}{8}}{n + \frac{1}{4}} \right) \quad ; \quad c = \frac{3}{8}$$

y la recta teórica sobre la cual deberían estar los puntos si procedieran de una distribución normal de media μ y varianza σ^2 . Dicha recta se obtiene, obviamente, como

$$E[x_{(i)}] = \mu + \sigma C_{i,n} ,$$

con los valores μ y σ estimados previamente a partir de los puntos $x_{(i)}$ y $\Phi \left(\frac{x_{(i)} - \mu}{\sigma} \right) \cdot 100$.

Así pues, sea cual sea la gráfica empleada, en tanto en cuanto la nube de puntos representada se ajuste a la recta teórica, menos evidencia tendremos para suponer la violación de la hipótesis de normalidad de los residuos.

5.2.2. Contraste de Shapiro-Wilks

Como ya se comentó, son múltiples los contrastes de hipótesis que se han desarrollado para verificar la hipótesis de normalidad ya sean paramétricos o no paramétricos.

Nosotros expondremos el contraste de Shapiro-Wilks puesto que se basa en la representación probabilística normal anterior. Más concretamente lo que pretende es medir la bondad del ajuste de la nube de puntos a la recta teórica, lo cual será realizado por medio del coeficiente de correlación al cuadrado entre los valores considerados y los valores esperados bajo la hipótesis de normalidad. De esta manera se construye el estadístico

$$r^2 = \frac{\left[\sum_{i=1}^n x_{(i)} C_{i,n} \right]^2}{nS_x^2 \left[\sum_{i=1}^n C_{i,n}^2 \right]} .$$

Observemos que en la anterior expresión se ha tenido en cuenta que la media de los valores $C_{i,n}$, para n fijo, es cero, lo cual se deduce porque dichos valores son simétricos y, caso de ser n impar, el término central vale cero. En efecto

- Consideremos los términos $C_{i,n}$ y $C_{n+1-i,n}$. Como

$$\frac{i - \frac{3}{8}}{n + \frac{1}{4}} + \frac{n + 1 - i - \frac{3}{8}}{n + \frac{1}{4}} = 1$$

entonces

$$\Phi(C_{i,n}) + \Phi(C_{n+1-i,n}) = 1$$

de donde se deduce que $C_{i,n}$ y $C_{n+1-i,n}$ son términos simétricos, o sea, $C_{i,n} = -C_{n+1-i,n}$.

- Si $n = 2k + 1$ y consideramos el término central $C_{k+1,n}$ entonces

$$\frac{k + 1 - \frac{3}{8}}{2k + 1 + \frac{1}{4}} = \frac{k + \frac{5}{8}}{2(k + \frac{5}{8})} = \frac{1}{2},$$

de donde se deduce

$$C_{k+1,n} = \Phi^{-1}\left(\frac{1}{2}\right) = 0$$

y, por tanto, se verificará

$$\sum_{i=1}^n C_{i,n} = 0.$$

Con ello el estadístico anterior se puede expresar en la forma

$$W = \frac{1}{nS_x^2} \left[\sum_{i=1}^h a_{i,n} (x_{(n+1-i)} - x_{(i)}) \right]^2$$

donde

$$a_{i,n} = \frac{|C_{i,n}|}{\sqrt{\sum_{j=1}^n C_{j,n}^2}}$$

y

$$h = \begin{cases} \frac{n}{2} & \text{si } n \text{ es par} \\ \frac{n-1}{2} & \text{si } n \text{ es impar} \end{cases} .$$

Los valores $a_{i,n}$ fueron tabulados por Shapiro y Wilks sin necesidad de emplear la aproximación anterior dada para los valores $C_{i,n}$. Asimismo fueron calculados, para distintos tamaños muestrales, los percentiles de este estadístico. Si el valor que toma el estadístico, para una muestra ordenada concreta, es menor que el valor crítico dado en las tablas, para un nivel de

significación impuesto de antemano, se rechazará la hipótesis de normalidad (notemos que dicho estadístico toma valores entre 0 y 1).

En el caso concreto que nos atañe, los valores $x_{(i)}$ empleados serán los correspondientes a los residuos, ordenados, de la regresión.

5.2.3. Transformaciones en busca de normalidad

Si en una aplicación práctica concreta se observa que la hipótesis de normalidad no es admisible surge de inmediato la cuestión de qué hacer en tal situación. Evidentemente una de las *alternativas* sería ignorar este hecho y proceder con el estudio como si la hipótesis fuera cierta. Esta práctica, evidentemente, no es recomendable puesto que puede conducir a conclusiones incorrectas. Una segunda alternativa pasa por intentar *transformar* dichos datos para conseguir la normalidad. Múltiples son los procedimientos que a tal fin se han desarrollado, siendo uno de los más empleados en la práctica el uso de la familia Box-Cox de transformaciones, familia que se define de la siguiente forma

$$y^{(\lambda)} = \begin{cases} \frac{(y + m)^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(y + m) & \lambda = 0 \end{cases}$$

donde el parámetro λ ha de estimarse a partir de los datos y la constante m se elige de forma que $y + m$ sea siempre positiva. Por lo tanto m será cero si se trabaja con datos positivos y, en otro caso, igual en valor absoluto al valor más negativo observado. Notemos que, vista como función de λ , la función $y^{(\lambda)}$ es una función continua ya que

$$\lim_{\lambda \rightarrow 0} \frac{(y + m)^\lambda - 1}{\lambda} = \lim_{\lambda \rightarrow 0} (y + m)^\lambda \ln(y + m) = \ln(y + m) .$$

Apliquemos esta familia de transformaciones al caso de la regresión lineal simple univariante. Para ello hemos de obtener la estimación del parámetro λ , cuestión que será abordada por máxima verosimilitud.

Suponiendo que la transformación

$$y_i^{(\lambda)} = \begin{cases} \frac{(y_i + m)^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(y_i + m) & \lambda = 0 \end{cases}$$

conduce, para cada x_i , a una distribución normal de media $\beta_0 + \beta_1 x_i$ y varianza σ^2 , entonces la función de densidad de y_i , para cada x_i , vendrá dada por

$$h(y_i) = f(g^{-1}(y_i)) = f(y_i^{(\lambda)}) \left| \frac{dg^{-1}(y_i)}{dy_i} \right| =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} \left(\frac{(y_i + m)^\lambda - 1}{\lambda} - \beta_0 - \beta_1 x_i\right)^2\right) (y_i + m)^{\lambda-1}$$

Así la función de verosimilitud asociada a $\{y_1, \dots, y_N\}$ es

$$\begin{aligned} \mathbb{L}(y_1, y_2, \dots, y_N, \beta_0, \beta_1, \sigma^2, \lambda) &= \\ &= \frac{1}{(\sigma^2)^{\frac{N}{2}} (2\pi)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N \left[\frac{(y_i + m)^\lambda - 1}{\lambda} - \beta_0 - \beta_1 x_i\right]^2\right) \prod_{i=1}^N (y_i + m)^{\lambda-1}. \end{aligned}$$

Y el logaritmo de dicha verosimilitud está dado por

$$\begin{aligned} \log(\mathbb{L}(y_1, y_2, \dots, y_N, \beta_0, \beta_1, \sigma^2, \lambda)) &= -\frac{N}{2} \log(\sigma^2) - \frac{N}{2} \log(2\pi) - \\ &-\frac{1}{2\sigma^2} \sum_{i=1}^N \left[\frac{(y_i + m)^\lambda - 1}{\lambda} - \beta_0 - \beta_1 x_i\right]^2 + (\lambda - 1) \sum_{i=1}^N \log(y_i + m). \end{aligned}$$

Ahora, para cada valor de λ fijo, se obtienen los estimadores máximo verosímiles de los

parámetros, resultando

$$\widehat{\sigma}_u^2(\lambda) = \frac{\sum_{i=1}^N \left[\frac{(y_i + m)^\lambda - 1}{\lambda} - \widehat{\beta}_0(\lambda) - \widehat{\beta}_1(\lambda)x_i \right]^2}{N}$$

donde $\widehat{\beta}_0(\lambda)$ y $\widehat{\beta}_1(\lambda)$ son los estimadores de la recta de regresión de $Y^{(\lambda)}$ sobre X , o sea,

$$\widehat{\beta}_1(\lambda) = \frac{S_{xy^{(\lambda)}}}{S_x^2},$$

$$\widehat{\beta}_0(\lambda) = \bar{y}^{(\lambda)} - \widehat{\beta}_1(\lambda)\bar{x}.$$

Con ello, para cada valor de λ fijo, el valor máximo del logaritmo de la verosimilitud viene dado por

$$L(\lambda) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\widehat{\sigma}_u^2(\lambda)) + (\lambda - 1) \sum_{i=1}^N \log(y_i + m)$$

El procedimiento para obtener $\widehat{\lambda}$ consistirá en calcular $L(\lambda)$ para distintos valores de λ , de tal forma que el valor que maximice dicha función será el estimador máximo verosímil de la transformación. Este cálculo puede implementarse fácilmente en ordenador, si bien es

bastante útil representar gráficamente $L(\lambda)$. Asimismo es bastante aconsejable disponer de una tabla con los valores $(\lambda, L(\lambda))$ para estudiar el comportamiento de los distintos valores que estén próximos al valor $\hat{\lambda}$. Este procedimiento se debe al hecho de que en ocasiones se puede elegir un valor distinto de $\hat{\lambda}$ pero próximo a él y que produzca un valor $L(\lambda)$ parecido a $L(\hat{\lambda})$ si ello conduce a una mayor simplicidad en las expresiones resultantes.

Además, este procedimiento proporciona intervalos de confianza para el valor de λ y un contraste para verificar la normalidad de los datos originales.

En efecto, si $\lambda = 1$ entonces la variable original sólo se ve modificada por un cambio de origen. En tal caso, si la variable transformada es normal también lo será la de partida. Así pues el contraste que se desea estudiar es $H_0 : \lambda = 1$ frente a $H_1 : \lambda \neq 1$.

Para su resolución nos basaremos en que menos dos veces el logaritmo del cociente de verosimilitudes se distribuye de forma asintótica como una χ^2 con tantos grados de libertad como parámetros se hayan estimado. En nuestro caso, el cociente de verosimilitudes es

$$\Lambda = \frac{\sup_{\beta_0, \beta_1, \sigma^2} \mathbb{L}(y_1, y_2, \dots, y_N, \beta_0, \beta_1, \sigma^2, 1)}{\sup_{\beta_0, \beta_1, \sigma^2, \lambda} \mathbb{L}(y_1, y_2, \dots, y_N, \beta_0, \beta_1, \sigma^2, \lambda)}$$

y

$$-2 \log(\Lambda) = -2 \left[L(1) - L(\hat{\lambda}) \right] = 2 \left[L(\hat{\lambda}) - L(1) \right] \rightsquigarrow \chi_1^2$$

donde el grado de libertad procede de la estimación del parámetro λ .

Por tanto, si $\chi_{1,\alpha}^2$ es el valor de una χ^2 con un grado de libertad que deja a su derecha una probabilidad α , se tiene

$$1 - \alpha = P \left[2 \left[L(\hat{\lambda}) - L(1) \right] \leq \chi_{1,\alpha}^2 \right] = P \left[L(1) \geq L(\hat{\lambda}) - \frac{1}{2} \chi_{1,\alpha}^2 \right]$$

Cortando $L(\lambda)$ con la ordenada $L(\hat{\lambda}) - \frac{1}{2} \chi_{1,\alpha}^2$ se obtienen dos valores para el parámetro que definirán un intervalo de confianza para λ . Si el valor $\lambda = 1$ está incluido en dicho intervalo se aceptará la hipótesis de normalidad de los datos con un nivel de significación α .

5.3. Residuos y heterocedasticidad.

Como sabemos la heterocedasticidad consiste en la violación de la hipótesis de igualdad de varianzas en el modelo. Su detección es bastante complicada y aún más su tratamiento. El hecho es que uno puede observar, o intuir, una posible heterocedasticidad pero, ¿de qué tipo? Dicho con otras palabras, a la hora de contrastar la heterocedasticidad presente en un modelo hay que suponer previamente un modelo para dicha heterocedasticidad. Esto lleva a la

existencia de múltiples pruebas conducentes a contrastar la existencia de este problema atendiendo a cómo sea éste, existiendo incluso algún contraste que intenta detectar su presencia, sea del tipo que sea.

Ya comentamos en el tema segundo un ejemplo en el que se podía intuir la violación de la hipótesis de homocedasticidad sólo con estudiar la naturaleza del problema planteado. En dicho ejemplo comentábamos cómo dicha heterocedasticidad podía venir dada en función de la propia variable explicativa. Veamos otro ejemplo en el que se plantea otro esquema distinto al ya comentado.

Supongamos un conjunto de N observaciones distribuidas en m grupos con n_i el tamaño del grupo i -ésimo. No obstante no se dispone de información sobre cada una de las observaciones sino promedios de las variables estudiadas en cada grupo. Suponiendo que el modelo original es homocedástico, al tener datos agrupados debemos considerar un nuevo modelo de la forma

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + \bar{\epsilon}_i ; \quad i = 1, \dots, m$$

Es claro que si $\text{Var}[\epsilon_i] = \sigma^2$ entonces $\text{Var}[\bar{\epsilon}_i] = \frac{\sigma^2}{n_i}$, con lo cual el modelo que hemos de estudiar no verifica la hipótesis de igualdad de varianzas.

5.3.1. Caso especial de heterocedasticidad. Mínimos Cuadrados Ponderados.

Un procedimiento gráfico interesante es representar los residuos de la regresión en función de los valores previstos \hat{y}_i o, equivalentemente, en función de la variable explicativa. Con ello se puede observar si la variable independiente está influyendo en algún sentido en la variabilidad de las perturbaciones, haciendo que ésta aumente o disminuya.

Incluso se puede tener una información a priori de que las perturbaciones o, lo que es lo mismo, la variable dependiente, pueda estar influida en su variabilidad por la presencia de una cierta variable (aunque no sea ésta la explicativa), aunque, eso sí, dicha variable (para el tratamiento posterior) debiera ser observable por el investigador. En resumidas cuentas lo que se está proponiendo es la posible aparición de heterocedasticidad en la forma

$$E[\epsilon_i^2] = \sigma^2 z_i^2$$

donde la variable Z pudiera o no ser la variable explicativa.

En este ambiente no sería recomendable el empleo del método de mínimos cuadrados ordinarios ya que perdería eficiencia. En efecto, los puntos muestrales con varianza menor son más fiables y deben tener un peso mayor que los que tengan una variabilidad mayor, puesto que así cuanto menor sea su varianza menos se desviarán del valor medio que queremos estimar. Esta es la idea del método de mínimos cuadrados ponderados que pasamos a desarrollar a continuación.

Supongamos que la variabilidad de las perturbaciones está afectada por la variable explicativa, de tal forma que

$$E[\epsilon_i^2] = \sigma^2 x_i^2 ; i = 1, \dots, N$$

Bajo estas premisas el método de mínimos cuadrados ponderados lo que hará es tomar el modelo inicial

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i ; i = 1, \dots, N$$

y transformarlo en otro

$$y_i^* = \beta_0 x_i^* + \beta_1 + \epsilon_i^* ; i = 1, \dots, N$$

donde

$$y_i^* = \frac{y_i}{x_i} , x_i^* = \frac{1}{x_i} , \epsilon_i^* = \frac{\epsilon_i}{x_i} ; i = 1, \dots, N$$

En este modelo transformado se verifica

$$E[\epsilon_i^{*2}] = E\left[\frac{\epsilon_i^2}{x_i^2}\right] = \frac{1}{x_i^2} E[\epsilon_i^2] = \frac{\sigma^2 x_i^2}{x_i^2} = \sigma^2$$

$$E[\epsilon_i^*] = 0$$

$$E[\epsilon_i^* \epsilon_j^*] = 0 \quad i \neq j$$

es decir, este nuevo modelo, cuyos parámetros son los mismos del primitivo, ya sí verifica las hipótesis del modelo estándar y se le puede aplicar el método de mínimos cuadrados ordinarios.

5.4. Residuos y autocorrelación.

Otro de los problemas que puede surgir en el estudio de un modelo de regresión es el de la autocorrelación de los residuos, o sea, la violación de la hipótesis estándar

$$E[\epsilon_i \epsilon_j] = 0 \quad ; \quad i \neq j$$

En primer lugar hay que notar que cabe esperar una dependencia entre los residuos en un contexto donde los datos procedan de series temporales. Así, los residuos, y por ende los datos de la variable dependiente, dependen entre sí a través del tiempo.

Asimismo, hay que comentar, al igual que ocurría con la heterocedasticidad, que pueden ser múltiples los esquemas de autocorrelación que pueden seguir los residuos y por lo tanto hay que presuponer o investigar, en algún sentido, dicho tipo de relación para ser tratado con posterioridad de forma debida.

5.4.1. Procedimientos gráficos.

Son múltiples las representaciones gráficas que pueden hacernos sospechar la presencia de autocorrelación en los términos de error del modelo. Ningún contraste de autocorrelación debe excluir un examen riguroso de los residuos generados en la estimación del modelo. Dicho examen debe incluir el gráfico de la serie de residuos en el cual habrá que buscar rachas de residuos de igual signo, así como una tendencia creciente o decreciente en los mismos, lo cual sería un claro indicio de autocorrelación.

Otra gráfica de gran interés es aquel en el que se representan los residuos e_i frente a los residuos e_{i-1} . Si la mayoría de los puntos en dicho gráfico se hallan en el primer y tercer cuadrantes dará un indicio de autocorrelación de primer orden positiva, mientras que si se hallan en el segundo y cuarto cuadrantes será un indicio de autocorrelación negativa.

También se podría observar la gráfica de los residuos frente a los valores predichos, intentando ver una posible correlación positiva o negativa entre los residuos. Por ejemplo si se observa que valores por encima de la media tienden a ir seguidos por valores por debajo, ello puede ser un indicio de correlación negativa, mientras que si lo que se observa es que valores por encima de la media tienden a ir seguidos por valores también por encima de ella, esto puede ser un indicio de autocorrelación positiva.

5.4.2. Contraste de autocorrelación de primer orden: Test de Durbin-Watson.

Se define el coeficiente de autocorrelación de orden k entre residuos como

$$r_k = \frac{\sum_{i=k+1}^N e_i e_{i-k}}{\sum_{i=1}^N e_i^2} .$$

Ante tal definición es obvio que una forma de estudiar el problema de la autocorrelación sería contrastar si dichos coeficientes son o no nulos. Para fijar ideas consideraremos sólo la posibilidad de la existencia de autocorrelación de primer orden, o sea, $r_1 \neq 0$, ya que éste coeficiente suele ser el de mayor importancia dentro de la secuencia de coeficientes de autocorrelación anteriormente definidos.

Durbin y Watson propusieron para tal cuestión el estadístico

$$d = \frac{\sum_{i=2}^N (e_i - e_{i-1})^2}{\sum_{i=1}^N e_i^2}$$

y dado que

$$\sum_{i=1}^N e_i^2 \approx \sum_{i=2}^N e_i^2 \approx \sum_{i=2}^N e_{i-1}^2$$

se puede comprobar que

$$d \approx 2(1 - r_1) .$$

Luego,

- Si r_1 es cero, el valor de d es próximo a dos.
- Si $0 < r_1 < 1$, el estadístico d estará entre cero y dos.
- Si $-1 < r_1 < 0$, el estadístico d estará entre dos y cuatro.

El problema que plantea el contraste de autocorrelación de primer orden radica en la imposibilidad de calcular la distribución exacta del estadístico bajo la hipótesis nula, que es la de autocorrelación serial de primer orden nula.

Durbin y Watson estudiaron este tema y construyeron cotas para dicho estadístico, d_L y d_U , en función del número de observaciones (en este caso N) y del número de variables explicativas (en este caso 1). El contraste se lleva a cabo de la siguiente forma:

La hipótesis nula es la ausencia de autocorrelación de primer orden frente a la alternativa de autocorrelación de primer orden positiva (veremos posteriormente cómo tratar la correlación de sentido negativo). De esta forma, bajo la hipótesis nula, el valor crítico de d ha de estar comprendido entre las cotas d_L y d_U .

Por lo tanto, si el valor de d es menor que d_L es probable que dicho valor sea demasiado bajo para que venga explicado únicamente por la variación muestral, con lo cual se rechaza la hipótesis nula de independencia serial frente a la alternativa de autocorrelación de primer orden positiva.

De forma análoga, si el valor de d está comprendido entre d_U y 2, se acepta la hipótesis nula, ya que una distancia tan pequeña respecto a 2 puede ser atribuida al error muestral en un modelo con errores independientes.

No obstante, si el valor de d está entre los valores d_L y d_U , el contraste no es concluyente, si bien existen variantes en esta cuestión que no trataremos.

Si lo que interesa es la posibilidad de que haya autocorrelación negativa, los puntos significativos tabulados deben restarse de 4. De esta manera d varía entre 2 y 4.

Cuando el valor de d esté entre 2 y $4 - d_U$ se acepta la hipótesis nula, y cuando d esté entre $4 - d_L$ y 4 hay evidencia de que existe autocorrelación negativa significativa.

Finalmente, el contraste no es concluyente si el valor del estadístico está comprendido entre $4 - d_U$ y $4 - d_L$.

5.4.3. Estimación bajo autocorrelación de primer orden: procedimiento de Cochran-Orcutt.

En la sección anterior hemos descrito un procedimiento para contrastar la violación de la hipótesis de incorrelación de las perturbaciones aleatorias del modelo. Concretamente se ha estudiado la forma de comprobar la existencia de autocorrelación de primer orden. Supuesta dicha existencia es evidente que el método de mínimos cuadrados pierde fiabilidad y no es recomendable su empleo.

A continuación vamos a exponer un procedimiento mediante el cual se pueden estimar los parámetros del modelo. Hay que insistir en que dicho método está desarrollado para ser aplicado en el caso de autocorrelación de primer orden, por lo que supondremos un esquema del tipo

$$\epsilon_i = \rho\epsilon_{i-1} + \eta_i \quad ; \quad i = 1, \dots, N$$

donde $|\rho| < 1$ y η_i es una variable error que verifica las mismas hipótesis del modelo estándar.

Considerando el modelo original podemos escribir la expresión anterior en la forma

$$y_i - \beta_0 - \beta_1 x_i = \rho(y_{i-1} - \beta_0 - \beta_1 x_{i-1}) + \eta_i$$

o lo que es lo mismo

$$y_i - \rho y_{i-1} = \beta_0(1 - \rho) + \beta_1(x_i - \rho x_{i-1}) + \eta_i \quad ; \quad i = 1, \dots, N$$

expresión que es conocida como modelo en cuasidiferencias.

La expresión anterior motivó el siguiente procedimiento iterativo debido a Cochranne y Orcutt:

1. Estimar los parámetros del modelo inicial por mínimos cuadrados ordinarios, ignorando la autocorrelación de primer orden. Calcular los valores residuales e_i .
2. Estimar ρ mediante la regresión de e_i sobre e_{i-1} .
3. Sustituir dicha estimación en el modelo en cuasidiferencias.
4. Estimar el modelo en cuasidiferencias por mínimos cuadrados ordinarios. Notemos que el término independiente que se estima es $\widehat{\beta_0(1 - \rho)}$, por lo que $\widehat{\beta}_0$ se obtendrá despejando en la estimación anterior, o sea

$$\widehat{\beta}_0 = \frac{\widehat{\beta_0(1 - \rho)}}{1 - \widehat{\rho}}.$$

5. Sustituir las estimaciones $\widehat{\beta}_1$ y $\widehat{\beta}_0$ en el modelo inicial para generar una nueva serie de residuos.
6. Repetir el proceso desde el paso 2.

El criterio de parada para este proceso iterativo puede ser planteado de dos formas:

- Imponiendo una tolerancia sobre $\hat{\rho}$ tras dos aplicaciones sucesivas del método.
- En cada aplicación se realiza un contraste de autocorrelación de primer orden para comprobar la presencia o no de la misma en los términos de error.

5.5. Residuos y datos anómalos.

Las observaciones atípicas o heterogéneas pueden tener gran importancia puesto que una sola de ellas puede determinar la estimación de la recta de regresión (ver el ejemplo de la introducción de este tema). Una observación de tal naturaleza nos es detectable durante el proceso de estimación del modelo y es examinando los residuos cuando se puede hacer relevante su existencia.

En primer lugar hay que notar que el hecho de la existencia de un dato anómalo respecto a los demás no significa necesariamente que dicha observación sea influyente en la estimación del modelo. Un dato puede considerarse como anómalo en el conjunto de los datos y sin embargo no es influyente en el sentido de que si se prescindiera de él en el análisis, la recta de regresión no varía de forma sustancial, mientras que, en otros casos, puede no sólo ser atípico respecto de los demás, sino que es muy influyente ya que está haciendo cambiar la recta de regresión

tirando de ella hacia sí, por lo que podría ser una buena idea el suprimirlo en el estudio y así dar una mayor fiabilidad a la estimación.

Para detectar estas cuestiones es muy importante nuestro propio conocimiento acerca de las variables y del rango donde éstas se mueven, aunque un examen de residuos puede completar este estudio.

Recordemos que

$$\text{Var}[e_i] = \sigma^2 \left[1 - \frac{1}{N} - \frac{(x_i - \bar{x})^2}{NS_x^2} \right] = \sigma^2 [1 - v_i]$$

por lo que podemos considerar para su estudio los residuos estandarizados

$$r_i = \frac{e_i}{\hat{\sigma}_u \sqrt{1 - v_i}} .$$

Notemos que el numerador y el denominador de esa expresión no son independientes puesto que e_i se ha usado en el cálculo de $\hat{\sigma}_u^2$. Este problema se puede solucionar eliminando el punto (x_i, y_i) de los cálculos y estimando el modelo con los $N - 1$ puntos restantes. Con dichos cálculos se puede demostrar que

$$\hat{\sigma}_{u_i}^2 = \frac{(N - 2)\hat{\sigma}_u^2 - \frac{e_i^2}{1 - v_i}}{N - 3} .$$

Esto da lugar a definir los residuos *estudentizados* como

$$t_i = \frac{e_i}{\widehat{\sigma}_{u_i} \sqrt{1 - v_i}}$$

que se distribuyen según una t de Student con $N - 3$ grados de libertad, puesto que ahora el numerador y el denominador sí son independientes.

No obstante notemos que si el tamaño muestral N es grande y si los datos no contienen valores extremos, el comportamiento de los tres tipos de residuos, e_i , r_i y t_i es análogo. En otro caso los valores r_i y t_i suelen ser más informativos para detectar deficiencias en el modelo.

5.5.1. Test de valores atípicos

Para contrastar que un residuo es atípico se emplean los residuos *estudentizados* que, como sabemos, se distribuyen según una t de Student con $N - 3$ grados de libertad. Si decidimos que un valor es atípico cuando tenga una probabilidad menor del 5 % de provenir de la distribución de referencia, en este caso la t de Student, y aplicamos, por ejemplo, este contraste a 100 residuos, la probabilidad de que al menos uno resulte atípico será

$$1 - 0,95^{100} = 0,994 ,$$

que es el problema común a la inferencia simultánea.

Una solución pasa por fijar un nivel global de significación, α , y *repartirlo* entre cada uno de los residuos estudentizados. Para ello haremos uso de la desigualdad de Bonferroni

$$P \left[\bigcup_{i=1}^N A_i \right] \leq \sum_{i=1}^N P[A_i] .$$

Si A_i es el suceso "el i -ésimo residuo es atípico", o sea, $A_i = \{|t_i| \geq t_\alpha\}$, entonces, aplicando la desigualdad de Bonferroni, se tiene

$$P \left[\bigcap_{i=1}^N \bar{A}_i \right] = 1 - P \left[\bigcup_{i=1}^N A_i \right] \geq 1 - \sum_{i=1}^N P[A_i] = 1 - \sum_{i=1}^N \alpha_i .$$

Así, basta tomar $\alpha_i = \bar{\alpha} = \frac{\alpha}{N}$, $\forall i$, para que

$$P \left[\bigcap_{i=1}^N \bar{A}_i \right] \geq 1 - \alpha .$$

A la vista de lo anterior, el procedimiento para detectar si un dato es atípico será fijar el nivel de significación α y calcular $\bar{\alpha}$. Y, una vez calculado el valor $t_{\bar{\alpha}}$ de una t de Student tal que $P[|t| > t_{\bar{\alpha}}] = \bar{\alpha}$, se concluye que un residuo estudentizado será atípico si $|t_i| > t_{\bar{\alpha}}$.

En la práctica no será usual encontrar en las tablas de la t de Student los percentiles para $\bar{\alpha}$, por lo que será de utilidad el empleo de la siguiente aproximación de la distribución t de Student con n grados de libertad mediante la distribución normal:

$$t_{\alpha} \simeq z_{\alpha} \left(1 - \frac{z_{\alpha} + 1}{4n} \right)^{-1} .$$