

Capítulo 11

Evaluación del ajuste. Diagnósticos.

Ya hemos visto en lo que precede estadísticos para evaluar la bondad de ajuste de un modelo, como \bar{R}^2 ; pero se trata de estadísticos que dan una idea global del ajuste. Puede ocurrir que un \bar{R}^2 encubra el hecho de que localmente —para unas ciertas observaciones— el ajuste es muy deficiente.

En lo que sigue abordaremos esta cuestión, considerando instrumentos para examinar el ajuste localmente (para observaciones individuales). Examinaremos también la cuestión íntimamente relacionada de cuándo una observación (o varias) son muy influyentes, en el sentido de condicionar de modo importante la estimación del modelo.

11.1. Análisis de residuos.

En general, como se ha indicado ya en el Capítulo 13, no conocemos la forma en que se generan los valores de la variable respuesta \vec{Y} . Todos los modelos que ajustemos son en alguna medida provisionales, y su adecuación a los datos debe ser objeto de análisis. El desarrollo que se hace a continuación sigue principalmente a Cook and Weisberg (1982). Otras referencias de utilidad son Hawkins (1980), Barnett and Lewis (1978), Belsley et al. (1980), Myers (1990) y Trocóniz (1987a).

La forma más natural de examinar el ajuste consiste en considerar los residuos

$$\hat{\epsilon} = \vec{y} - X\hat{\beta} = (I - X(X'X)^{-1}X')\vec{y} = (I - X(X'X)^{-1}X')\vec{\epsilon} \quad (11.1)$$

Podemos contemplar los $\hat{\epsilon}_i$ como “estimaciones” de las perturbaciones ϵ_i (inobservables) que han intervenido en la generación de las Y_i . Veremos sin embargo que, en general, sólo vagamente reproduce $\hat{\epsilon}$ el comportamiento de $\vec{\epsilon}$. En particular,

Teorema 11.1 *Bajo los supuestos habituales se verifica que:*

1. *Los residuos no son, en general, homoscedásticos, incluso cuando las perturbaciones lo son.*
2. *Los residuos no son, en general, incorrelados, incluso cuando las perturbaciones lo son.*

DEMOSTRACIÓN:

$$\Sigma_{\hat{\epsilon}} = E[(\hat{\epsilon} - E(\hat{\epsilon}))(\hat{\epsilon} - E(\hat{\epsilon}))'] \quad (11.2)$$

Como $E(\hat{\epsilon}) = \vec{0}$, (11.2) se reduce a:

$$E\hat{\epsilon}\hat{\epsilon}' = E[(I - X(X'X)^{-1}X')\vec{y}\vec{y}'(I - X(X'X)^{-1}X)'] \quad (11.3)$$

$$= (I - X(X'X)^{-1}X')\sigma^2 I \quad (11.4)$$

$$= \sigma^2(I - P), \quad (11.5)$$

que en general no tiene elementos iguales a lo largo de la diagonal principal. El apartado 2) del enunciado es inmediato a partir de (11.5), dado que $(I - P)$ es una matriz no diagonal.

Sea,

$$p_{ij} = \vec{x}_i'(X'X)^{-1}\vec{x}_j \quad (11.6)$$

un elemento genérico de la matriz P (\vec{x}_i' denota la i -ésima fila de X). De la igualdad (11.1) se deduce:

$$\hat{\epsilon}_i = (1 - p_{ii})\epsilon_i - \sum_{i \neq j} p_{ij}\epsilon_j \quad (11.7)$$

Por tanto, el residuo i -ésimo es un promedio ponderado de la perturbación correspondiente a dicha observación y las de todas las demás observaciones, con ponderaciones $(1 - p_{ii})$ y $(-p_{ij})$. Dependiendo de los valores que tomen estos coeficientes, $\hat{\epsilon}_i$ recogerá con desigual fidelidad el valor de ϵ_i .

Los valores p_{ij} dependen sólo de la matrix de diseño y son del mayor interés, como veremos más abajo.

11.1.1. Residuos internamente studentizados.

Los residuos MCO definidos en (12.1) son, por causa de su heterocedasticidad, desaconsejables para la detección de observaciones anormales o diagnóstico de modelos de regresión. Es sin embargo fácil corregir dicha heterocedasticidad. De (12.5) se deduce que una estimación de la varianza de $\hat{\epsilon}_i$ viene dada por $\hat{\sigma}^2(1 - p_{ii})$. Por tanto,

$$r_i = \frac{\hat{\epsilon}_i}{\sqrt{\hat{\sigma}^2(1 - p_{ii})}} \quad (11.8)$$

para $i = 1, \dots, N$ son residuos de varianza común. Se llama *studentización* a la eliminación del efecto de un parámetro de escala (aquí σ^2) mediante división por una estimación adecuada. Se denomina *internamente studentizados* a los residuos definidos en (12.8).

Es de notar que, a pesar de su denominación, los r_i no siguen una distribución t de Student, pues numerador y denominador no son independientes ($\hat{\epsilon}_i$ ha intervenido en el cómputo de $\hat{\sigma}^2$). Es fácil demostrar, sin embargo, que bajo los supuestos habituales más el de normalidad en las perturbaciones, $r_i^2/(N - p)$ sigue una distribución beta $B(\frac{1}{2}, \frac{1}{2}(N - p - 1))$.

Al tener los r_i la misma varianza, se prestan mejor a ser examinados gráficamente para identificar posibles observaciones anómalas o *outliers*.

11.1.2. Residuos externamente studentizados.

Definidos por:

$$t_i = \frac{\hat{\epsilon}_i}{\sqrt{\hat{\sigma}^2(i)(1 - p_{ii})}} \quad (11.9)$$

son formalmente idénticos a los r_i , con la única salvedad de haberse tomado en el denominador un estimador $\hat{\sigma}^2(i)$ de σ^2 que no hace uso de $\hat{\epsilon}_i$. Mediante una elección adecuada de $\hat{\sigma}^2(i)$ puede lograrse que t_i siga una distribución t de Student con $(N - p - 1)$ grados de libertad. Esto permite, entre otras cosas, hacer uso de la distribución del máximo de k variables t de Student con correlación por pares ρ (véase Sección 9.3, pág. 118) para contrastar la presencia de *outliers*. Tomaremos,

$$\hat{\sigma}^2(i) = \frac{\hat{\epsilon}'\hat{\epsilon} - \hat{\epsilon}_i(1 - p_{ii})^{-1}\hat{\epsilon}_i}{(N - p - 1)} \quad (11.10)$$

lo que permite probar el siguiente,

Teorema 11.2 Con $\hat{\sigma}^2(i)$ definido como en (12.10), bajo los supuestos habituales más el de normalidad en las perturbaciones, los residuos t_i definidos en (12.9) (externamente studentizados) siguen una distribución t de Student con $(N - p - 1)$ grados de libertad.

DEMOSTRACIÓN:

Podemos escribir $\hat{\epsilon}_i = G'_i(I - P)\vec{\epsilon}$ siendo G'_i de dimensión $1 \times N$, con un único “uno” en posición i -ésima y ceros en los demás lugares. Llamando $A = G'_i(I - P)$ tenemos que:

$$\hat{\epsilon}_i = A\vec{\epsilon} \quad (11.11)$$

Por otra parte, de (12.10) deducimos:

$$\begin{aligned} (N - p - 1)\hat{\sigma}^2(i) &= \hat{\epsilon}'[I - G_i[G'_i(I - P)G_i]^{-1}G'_i]\hat{\epsilon} \\ &= \vec{\epsilon}' \underbrace{(I - P)[I - G_i[G'_i(I - P)G_i]^{-1}G'_i](I - P)}_B \vec{\epsilon} \\ &= \vec{\epsilon}' B \vec{\epsilon} \end{aligned} \quad (11.12)$$

Es fácil comprobar que $AB = 0$, luego $\hat{\epsilon}_i$ y $\hat{\sigma}^2(i)$ son independientes (Lema 7.3, pág. 71). Por otra parte, es también fácil comprobar que B es idempotente, con rango (= traza) $(N - p - 1)$. Por consiguiente,

$$\frac{\hat{\epsilon}_i}{\sqrt{\hat{\sigma}^2(i)(1 - p_{ii})}} = \frac{\hat{\epsilon}_i/\sqrt{\sigma^2(1 - p_{ii})}}{\sqrt{\hat{\sigma}^2(i)/\sigma^2}} \quad (11.13)$$

$$= \frac{\hat{\epsilon}_i/\sqrt{\sigma^2(1 - p_{ii})}}{\sqrt{\vec{\epsilon}' B \vec{\epsilon}/(N - p - 1)\sigma^2}} \quad (11.14)$$

Pero en el numerador y denominador de (12.14) hay respectivamente una variable aleatoria $N(0, 1)$ y una χ^2 dividida entre sus grados de libertad, ambas independientes, lo que demuestra el Teorema.

Para contrastar la hipótesis de presencia de *outliers*, podemos comparar el mayor de los residuos externamente studentizados con el cuantil apropiado de la distribución del máximo valor absoluto de k variables aleatorias t de Student (Sección 9.3, pág. 118). Supondremos que son incorrelados, salvo que podamos calcular fácilmente su correlación por pares, como sucede a menudo en Análisis de Varianza. El texto Seber (1977) reproduce en su Apéndice E tablas adecuadas. Alternativamente, podemos comparar el mayor residuo internamente studentizado con los valores críticos en las tablas de Lund (1975), o emplear la desigualdad de Bonferroni.

11.1.3. Residuos BLUS.

La studentización, tanto interna como externa, elimina la heterocedasticidad de los residuos, pero no la mutua correlación. No es posible obtener un vector de N residuos incorrelados y ortogonales a las columnas de X . La razón se ve fácilmente: $\hat{\epsilon} \perp R(X)$ es un vector aleatorio de N coordenadas, pero constreñido a yacer en un subespacio $(N - p)$ dimensional. Su distribución en R^N es degenerada, y su matriz de covarianzas de rango $(N - p)$ (supuesta X de rango completo). Ninguna transformación ortogonal puede convertir tal matriz en diagonal de rango N .

Si es posible, sin embargo, obtener $(N - p)$ residuos incorrelados, homoscedásticos, y de media 0; de hecho, hay multitud de maneras de hacerlo¹, dependiendo del subconjunto de $(N - p)$ residuos que escojamos.

Tales residuos, denominados BLUS (o ELIO), son de utilidad para contrastar homoscedasticidad (suministrando una alternativa al conocido método de Goldfeld-Quandt), normalidad, etc. Un tratamiento detallado puede encontrarse en Theil (1971), Cap. 5.

11.1.4. Residuos borrados.

Sean $X_{(i)}$ e $\vec{Y}_{(i)}$ la matriz de diseño y vector respuesta desprovistos de la observación i -ésima. Sea $\hat{\beta}_{(i)}$ el vector de estimadores de los parámetros obtenido sin dicha observación, es decir, $\hat{\beta}_{(i)} = (X'_{(i)}X_{(i)})^{-1}X'_{(i)}\vec{Y}_{(i)}$. Se llama *residuos borrados (deleted residuals)* a los d_i definidos así²:

$$d_i = y_i - \vec{x}_i' \hat{\beta}_{(i)} \quad (11.15)$$

Un d_i muy pequeño o nulo indicaría que la observación i -ésima no se separa en su comportamiento del recogido por la regresión sobre las restantes $N - 1$ observaciones. Lo contrario es cierto si d_i es muy grande.

Hay una relación muy simple que permite calcular los d_i sin necesidad de realizar N regresiones diferentes sobre todos los conjuntos posibles de $N - 1$

¹Véase Theil (1971), pág. 202 y ss.

²Una denominación alternativa frecuente en la literatura es la de residuos PRESS (predictive sum of squares residuals).

observaciones. En efecto, de (12.15) se deduce que:

$$\begin{aligned} d_i &= y_i - \vec{x}_i' (X'_{(i)} X_{(i)})^{-1} X'_{(i)} \vec{Y}_{(i)} \\ &= y_i - \vec{x}_i' [(X'X) - \vec{x}_i \vec{x}_i']^{-1} X'_{(i)} \vec{Y}_{(i)} \end{aligned} \quad (11.16)$$

$$= y_i - \vec{x}_i' \left[(X'X)^{-1} + \frac{(X'X)^{-1} \vec{x}_i \vec{x}_i' (X'X)^{-1}}{1 - \vec{x}_i' (X'X)^{-1} \vec{x}_i} \right] X'_{(i)} \vec{Y}_{(i)} \quad (11.17)$$

$$= y_i - \vec{x}_i' \left[\frac{(1 - p_{ii})(X'X)^{-1} + (X'X)^{-1} \vec{x}_i \vec{x}_i' (X'X)^{-1}}{1 - p_{ii}} \right] X'_{(i)} \vec{Y}_{(i)}$$

$$= y_i - \left[\frac{(1 - p_{ii}) \vec{x}_i' (X'X)^{-1} + p_{ii} \vec{x}_i' (X'X)^{-1}}{1 - p_{ii}} \right] X'_{(i)} \vec{Y}_{(i)}$$

$$\begin{aligned} &= y_i - \frac{\vec{x}_i' (X'X)^{-1} X'_{(i)} \vec{Y}_{(i)}}{1 - p_{ii}} \\ &= \frac{(1 - p_{ii})y_i - \vec{x}_i' (X'X)^{-1} (X' \vec{Y} - \vec{x}_i y_i)}{1 - p_{ii}} \end{aligned} \quad (11.18)$$

$$\begin{aligned} &= \frac{y_i - \vec{x}_i' (X'X)^{-1} X' \vec{Y}}{1 - p_{ii}} \\ &= \frac{\hat{\epsilon}_i}{1 - p_{ii}} \end{aligned} \quad (11.19)$$

en que el paso de (12.16) a (12.17) hace uso del Teorema A.2, pág. 230. Veremos en lo que sigue que d_i está relacionado con la influencia que la observación i -ésima tiene sobre la estimación de los parámetros.

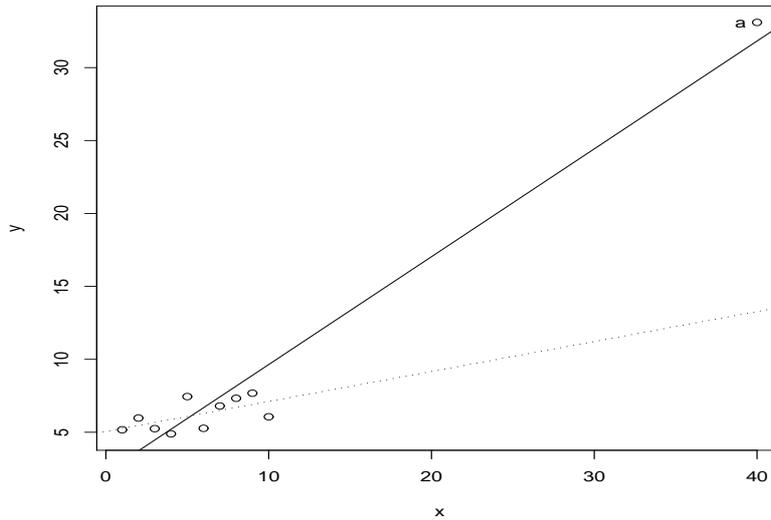
11.2. Análisis de influencia.

Es en general indeseable que la estimación de un parámetro dependa de modo casi exclusivo de una sola observación o de unas pocas, de manera que su eliminación conduzca a resultados completamente diferentes. En general, cuando esto ocurre, es necesario particionar la muestra o replantear el modelo. En todo caso, es necesario saber hasta qué punto observaciones aisladas influyen las estimaciones de los parámetros para obrar en consecuencia.

Puede parecer que para determinar qué observaciones influyen más en el resultado de la estimación basta mirar los residuos, brutos o studentizados. Ello es verdad, pero sólo en parte: puede haber observaciones extraordinariamente influyentes que resulten muy bien ajustadas por la regresión, como el ejemplo de la Fig. 12.1 pone de manifiesto.

Claramente, el punto a tiene una notable influencia en la estimación de la pendiente de la recta, hasta el punto de que su omisión daría lugar a un

Figura 11.1: Una observación como a tiene residuo borrado muy grande, y gran influencia en la pendiente de la recta de regresión.



resultado completamente diferente (la recta dibujada con trazo discontinuo). Sin embargo, su residuo MCO es muy pequeño; un exámen de los residuos MCO —o incluso de los residuos *studentizados*— difícilmente delataría ninguna anormalidad.

El examen de los residuos borrados detectaría una situación como la mencionada: a tendría un residuo borrado grande. Pero todavía es posible un análisis más sofisticado, que tenga en cuenta, en particular, los parámetros sobre los que una observación es muy influyente. Abordamos este análisis a continuación.

11.2.1. La curva de influencia muestral.

La forma obvia de examinar la influencia de la observación i -ésima consiste en comparar los vectores de estimadores obtenidos con y sin dicha observación: $\hat{\beta}$ y $\hat{\beta}_{(i)}$ respectivamente. En consecuencia, definimos la *curva de influencia muestral* (SIC) así:

$$\text{SIC}_i = (N - 1)(\hat{\beta} - \hat{\beta}_{(i)}). \quad (11.20)$$

El factor $(N - 1)$ tiene por misión corregir el efecto del tamaño muestral: en igualdad de todo lo demás, una observación altera la estimación tanto menos cuanto más grande sea la muestra.

La expresión (12.20) es vector-valorada: recoge, debidamente amplificadas por $(N - 1)$, por la razón apuntada, las diferencias que introduce la inclusión de la observación i -ésima sobre cada uno de los p parámetros estimados. Podemos relacionar (12.20) con el residuo borrado i -ésimo haciendo uso del siguiente lema.

Lema 11.1 *Se verifica que*

$$(\hat{\beta} - \hat{\beta}_{(i)}) = \frac{(X'X)^{-1}\vec{x}_i\hat{\epsilon}_i}{(1 - p_{ii})} = (X'X)^{-1}\vec{x}_id_i. \quad (11.21)$$

DEMOSTRACIÓN:

$$\begin{aligned} (\hat{\beta} - \hat{\beta}_{(i)}) &= (X'X)^{-1}X'\vec{Y} - ((X'X) - \vec{x}_i\vec{x}_i')^{-1}(X'\vec{Y} - \vec{x}_iy_i) \\ &= (X'X)^{-1}X'\vec{Y} \\ &\quad - \left[(X'X)^{-1} + \frac{(X'X)^{-1}\vec{x}_i\vec{x}_i'(X'X)^{-1}}{1 - \vec{x}_i'(X'X)^{-1}\vec{x}_i} \right] (X'\vec{Y} - \vec{x}_iy_i) \\ &= (X'X)^{-1}\vec{x}_iy_i - \frac{(X'X)^{-1}\vec{x}_i\vec{x}_i'(X'X)^{-1}X'\vec{Y}}{1 - p_{ii}} \\ &\quad + \frac{(X'X)^{-1}\vec{x}_i\vec{x}_i'(X'X)^{-1}\vec{x}_iy_i}{1 - p_{ii}} \\ &= \frac{(X'X)^{-1}\vec{x}_i}{1 - p_{ii}} \left[(1 - p_{ii})y_i - \vec{x}_i'\hat{\beta} + p_{ii}y_i \right] \\ &= (X'X)^{-1}\vec{x}_i\frac{\hat{\epsilon}_i}{1 - p_{ii}} \end{aligned}$$

En consecuencia,

$$\text{SIC}_i = (N - 1)(\hat{\beta} - \hat{\beta}_{(i)}) = (N - 1)(X'X)^{-1}\vec{x}_i\frac{\hat{\epsilon}_i}{1 - p_{ii}}$$

y el cálculo de la curva de influencia muestral SIC_i correspondiente a la observación i no requiere realizar una regresión para cada i ; todos los cálculos se se pueden hacer con ayuda de los residuos ordinarios y diagonal de la matriz de proyección correspondientes a la matriz de proyección $X(X'X)^{-1}X'$.

Diferentes versiones de la curva de influencia disponibles en regresión lineal puede encontrarse en Cook and Weisberg (1982) y Belsley et al. (1980). Alternativas como la *curva de influencia empírica EIC* y otras, difieren de la curva de influencia muestral presentada en el grado en que se corrige $\hat{\epsilon}_i$ (en la EIC se divide entre $(1 - p_{ii})^2$, en lugar de entre $(1 - p_{ii})$ como en (12.22).

11.2.2. Distancia de Cook.

Tal y como se indica más arriba, la curva de influencia en cualquiera de sus versiones es, en nuestro caso, un vector $p \times 1$ ($p =$ número de parámetros). La coordenada k -ésima de SIC_i proporciona información sobre la influencia de la observación i -ésima en la estimación de $\hat{\beta}_k$. Aunque esta información pormenorizada sea útil, en ocasiones queremos una única medida resumen de la influencia de una observación.

Sea $\hat{\beta}_{(i)}$ el vector de estimadores obtenido sin hacer uso de la observación i -ésima, y $\hat{\beta}$ el computado con la muestra completa. Una posibilidad es ponderar las discrepancias en una única expresión como:

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})' S (\hat{\beta} - \hat{\beta}_{(i)})}{c} \quad (11.22)$$

siendo S una matriz definida no negativa y c una constante positiva. Puesto que $\hat{\beta} \sim (\vec{\beta}, \sigma^2(X'X)^{-1})$, una elección posible que aproximadamente “normaliza” (12.22) es: $S = (X'X)$ y $c = p\hat{\sigma}^2$. Con esta elección, la expresión (12.22) se denomina *distancia de Cook* y es una medida global de la influencia de la observación (\vec{x}_i, y_i) . Hay otras posibles elecciones de S y c con diferencias, en general, sólo de matiz³.

Haciendo uso del Lema 12.1 tenemos que la distancia de Cook puede escribirse así:

$$D_i = \frac{\hat{\epsilon}_i \vec{x}_i' (X'X)^{-1} (X'X) (X'X)^{-1} \vec{x}_i \hat{\epsilon}_i}{p\hat{\sigma}^2(1 - p_{ii})^2} \quad (11.23)$$

$$= \frac{1}{p} r_i^2 \frac{p_{ii}}{1 - p_{ii}} \quad (11.24)$$

siendo r_i el i -ésimo residuo internamente studentizado.

11.2.3. DFFITS.

Se definen así:

$$DFFIT_i = t_i \sqrt{\frac{p_{ii}}{1 - p_{ii}}} \quad (11.25)$$

Se suele considerar observaciones inusuales a aquéllas con

$$|DFFIT_i| > 2\sqrt{\frac{p}{N}} \quad (11.26)$$

³Una relación de las mismas puede verse en Cook and Weisberg (1982), p. 124.

11.2.4. DFBETAS.

Se definen por:

$$\text{DFBETA}_{ij} = \frac{\hat{\beta}_j - \hat{\beta}_{j,(i)}}{\hat{\sigma} \sqrt{(X'X)^{-1}_{jj}}}; \quad (11.27)$$

Los estadísticos DFBETA permiten evaluar la influencia de la observación i -ésima sobre el parámetro j -ésimo. En cierto modo desglosan la información que la distancia de Cook resume en un único estadístico por observación. La motivación de la expresión (12.27) es clara: la diferencia entre la estimación de β_j -ésimo con y sin la observación i -ésima se divide por una estimación de la desviación típica de $\hat{\beta}_j$.

El criterio que se sigue es el de comparar $|\text{DFBETA}_{ij}|$ con $2/\sqrt{N}$. Más detalles en Belsley et al. (1980).

11.3. Análisis gráfico de residuos

Al margen del uso que pueda hacerse de los residuos en cualquiera de sus variedades para, por ejemplo, contrastar hipótesis de presencia de *outliers*, etc., con frecuencia será conveniente construir algunos gráficos. Es mucha, en efecto, la información que cabe obtener de ellos. Presentamos a continuación algunos de estos gráficos; otros aparecerán en contexto en los capítulos dedicados a selección de modelos (Capítulo 13) y transformaciones de las variables (capítulo 14). Referencias útiles para ampliar lo que se expone a continuación incluyen Trocóniz (1987a), Myers (1990), Ryan (1997) o Atkinson (1985).

11.3.1. Gráficos de residuos frente a índice de observación $(i, \hat{\epsilon}_i)$

Frecuentemente, el índice de cada observación es el tiempo, es decir, las observaciones han sido tomadas secuencialmente una después de otra. El representar $\hat{\epsilon}_i$ frente a i nos podría poner de manifiesto rupturas temporales —por ejemplo, una brusca disminución del tamaño de los residuos a partir de un cierto i —. En ocasiones podemos ver también en un gráfico de esta naturaleza pautas como agrupamiento de residuos, que puede convenir investigar.

Pueden emplearse residuos ordinarios o *studentizados* en cualquiera de sus variedades.

11.3.2. Gráficos de residuos frente a variables incluidas ($x_{ij}, \hat{\epsilon}_i$)

Los residuos ordinarios son por construcción ortogonales a cualquiera de los regresores. No obstante, un gráfico de esta naturaleza puede aportar información acerca del modo en que un regresor interviene en la generación de la respuesta: por ejemplo, podríamos ver una pauta de relación no lineal entre $\hat{\epsilon}_i$ y x_{ij} , sugiriendo que x_{ij} debe suplementarse con un término cuadrático, entrar como función exponencial, etc.

11.3.3. Gráficos de residuos frente a variables excluidas ($x_{ij}^*, \hat{\epsilon}_i$)

La idea es similar a la del apartado precedente, pero x_{ij}^* son ahora los valores de una variable no incluida (y candidato a serlo) en la regresión. Un gráfico de esta naturaleza permitiría ver si la parte no explicada de la respuesta (los residuos) tiene alguna relación evidente con la nueva variable. En su caso, dependiendo de la pauta que dibujaran los residuos, tendríamos pistas acerca de si dicha variable \vec{x}_j^* ha de incluirse tal cual o tras alguna transformación funcional.

11.3.4. Gráficos de variable añadida ($\hat{\epsilon}_{Y|X_{-j}}, \hat{\epsilon}_{X_j|X_{-j}}$)

La idea es similar a la del apartado anterior. Se dibujan los residuos de la regresión de Y sobre todas las variables *menos* X_j sobre los residuos de regresar dicha variable sobre todas las demás. Los residuos de ambas regresiones recogen, respectivamente, las partes de Y y X_j ortogonales al subespacio generado por las restantes variables.

Si hubiera alguna pauta en dicha gráfica, podríamos interpretarla como relación entre Y y X_j eliminado en ambas el efecto de las restantes variables.

11.3.5. Gráficos de normalidad de residuos

Aunque, como se ha visto (Sección 12.1.1 y siguiente), los residuos *studentizados* no siguen una distribución normal, a efectos prácticos y para tamaños muestrales moderados (Trocóniz (1987a), pág. 174, indica que suele bastar $N > 20$) la aproximación a la normalidad es muy buena, si las perturbaciones son a su vez normales.

Hay multitud de pruebas utilizables para contrastar ajuste a una distribución. La de Kolmogorov-Smirnov (véase Trocóniz (1987b), pág. 255) es de uso general con muestras grandes y distribuciones continuas —lo que incluye

a la normal—. Hay contrastes como el de Shapiro-Wilk descrito en Shapiro and Wilk (1965) y Shapiro and Francia (1972), especializados en el contraste de la hipótesis de normalidad.

Tan útil como pueda ser una prueba estadística convencional de normalidad, en ocasiones es útil un instrumento que permita visualizar la naturaleza y alcance de la desviación respecto a la normalidad, si existe. Los gráficos en papel normal cumplen esta finalidad.

El principio es muy simple: dada una muestra $\{x_i\}_{i=1}^N$, si procede de una distribución normal los puntos $(x_i, \Phi^{-1}(F_*(x_i)))$, en que $F_*(x_i)$ es la función de distribución empírica de la muestra, deben estar aproximadamente alineados. Véase por ejemplo Trocóniz (1987b), pág. 270.

El gráfico puede hacerse manualmente sobre papel especial (“papel normal”) en que la escala vertical absorbe la transformación $\Phi^{-1}(\cdot)$; o puede hacerse mediante ordenador en cuyo caso basta facilitar los datos y verificar la linealidad del gráfico resultante.

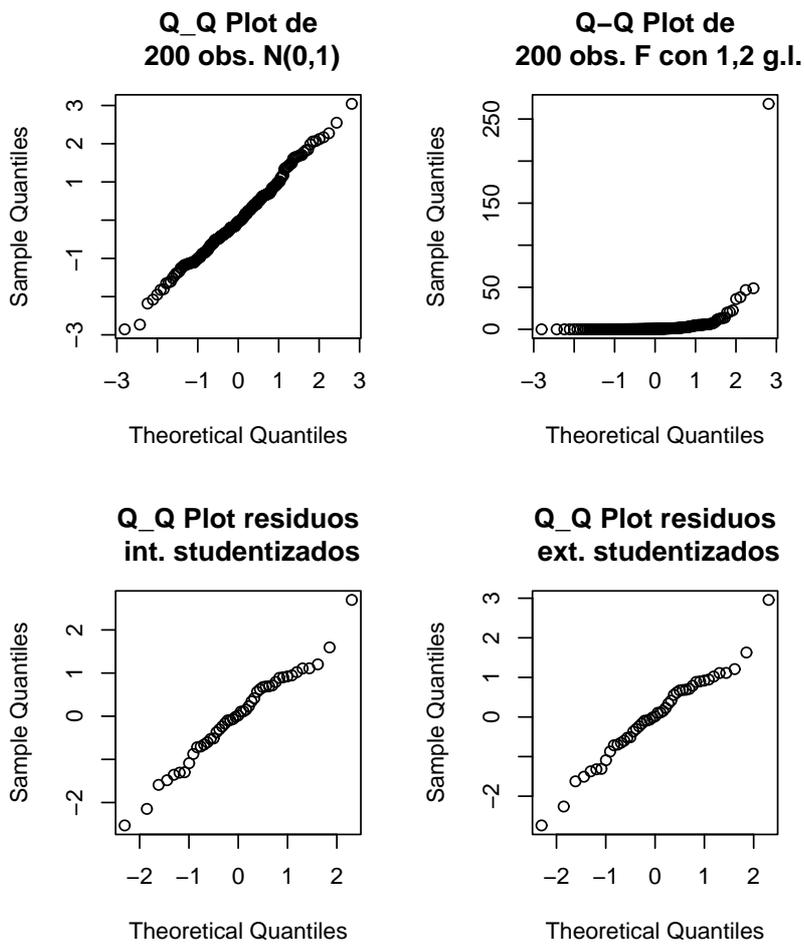
En cualquiera de los casos se cuenta con un instrumento que permite no sólo apreciar si hay desviaciones respecto de la normalidad, sino también de qué naturaleza son y a qué puntos afectan.

R: Ejemplo 11.1 (*gráficos para contraste de normalidad de residuos*)

La Figura 12.2 se genera mediante el fragmento de código reproducido a continuación. Los dos primeros paneles recogen sendos gráficos de normalidad para una muestra normal y una muestra procedente de una $\mathcal{F}_{1,2}$; puede verse la llamativa desviación de la normalidad en este último caso.

```
> par(mfrow = c(2, 2))
> muestra <- rnorm(200)
> qqnorm(muestra, main = "Q_Q Plot de\n 200 obs. N(0,1)")
> muestra <- rf(200, 1, 2)
> qqnorm(muestra, main = "Q-Q Plot de\n 200 obs. F con 1,2 g.l.")
> rm(muestra)
> library(MASS)
> data(UScrime)
> modelo <- lm(y ~ M + Ed +
+           Pol + M.F + U1 + U2 +
+           Prob + Ineq, data = UScrime)
> qqnorm(stdres(modelo), main = "Q_Q Plot residuos\n int. studentiza")
> qqnorm(studres(modelo), main = "Q_Q Plot residuos\n ext. studentiza")
```

Figura 11.2: Gráficos para contraste de normalidad



X11cairo

2

Los siguientes dos paneles muestran los gráficos de normalidad correspondientes a los residuos interna y externamente *studentizados* de un mismo modelo. Puede constatar que son casi idénticos y que sugieren un buen ajuste de la muestra a la hipótesis de normalidad.

FIN DEL EJEMPLO ■

11.3.6. Gráficos de residuos ordinarios frente a residuos borrados $(d_i, \hat{\epsilon}_i)$

Un residuo borrado no necesariamente es indicativo de que una observación sea muy influyente. Lo realmente sintomático es una gran divergencia entre el residuo ordinario y el residuo borrado, pues ello indica que al omitir la observación correspondiente los resultados varían mucho, al menos en el ajuste de la observación i -ésima.

Por ello se propone como gráfico útil en el diagnóstico de un modelo el de $\hat{\epsilon}_i$ frente a d_i . En general, deberíamos observar puntos aproximadamente sobre la bisectriz: $d_i \approx \hat{\epsilon}_i$. Puntos muy separados de la bisectriz corresponderían a observaciones que alteran sustancialmente la regresión.

COMPLEMENTOS Y EJERCICIOS

11.1 Demuéstrese que $r_i^2/(N-p)$, bajo los supuestos habituales más normalidad, sigue una distribución beta, $B(\frac{1}{2}, \frac{1}{2}(N-p-1))$.

