5. Análise de Resíduos

5.1 Resíduos “brutos”

$$Y_i \overset{ind}{\sim} N(\mu_i, \sigma^2), \quad i = 1, ..., n$$

onde $\mu \in L_1 = span\{X\}$, $X = (x_1, ..., x_k)$ matriz de posto completo.

Resíduos brutos: $R = Y - \widehat{\mu}$.

Como $\widehat{\mu} = HY \Rightarrow R = (I - H)Y$.

$\models I - H$ é simétrica e idempotente e, portanto, é uma matriz de projeção.

Além disso, é fácil ver que $(I - H)Y \in L_1^{\perp}$ e, portanto, é independente de $\widehat{\mu} = HY \in L_1$.

Como R é uma conbinação linear dos elementos de Y segue que R tem distribuição normal multivariada com

$$E[R] = (I - H)E[Y] = (I - H)\mu = \mu - H\mu = 0$$

$$Var(R) = (I - H)\sigma^2(I - H)^T = (I - H)\sigma^2$$

Assim, podemos escrever, $R \sim N_n\Big(0, \sigma^2(I - H)\Big)$.

Podemos verificar também que $\widehat{\mu} \sim N_n\Big(\mu, \sigma^2 H\Big)$.

Observe que tanto a distribuição de R como a distribuição de $\widehat{\mu}$ são normais singulares (degeneradas) pois as matrizes H e $I - H$ são singulares.

$\hookrightarrow$ Espera-se que os elementos de H sejam pequenos para grandes amostras.

$\hookrightarrow$ De alguma forma R é o estimador do erro ϵ quando escrevemos o modelo como

$$Y = \mu + \epsilon \quad \epsilon \sim N_n(0, \sigma^2 I)$$

Se $H \simeq 0$, a distribuição de R torna-se próxima da distribuição de ϵ.

Como $R \sim N_n\Big(0, \sigma^2(I - H)\Big)$, podemos escrever, $R_i \sim N(0, \sigma^2(1 - h_{ii})$, $i = 1, ..., n$, onde h_{ii} é o i-ésimo elemento da diagonal de H.

$\hookrightarrow$ os resíduos, em geral, não são independentes, nem identicamente distribuídos.

$$Cov(R_i, R_j) = -h_{ij}\sigma^2$$

$$\rho_{ij} = Corr(R_i, R_j) = \frac{-h_{ij}}{\sqrt{(1-h_{ii})(1-h_{jj})}}$$

5.1.2 Resíduos padronizados

De forma a se ter resíduos identicamente distribuídos (variância constante) definimos

$$S_i = \frac{R_i}{(1-h_{ii})^{1/2}}, \quad i = 1, ..., n$$

os resíduos padronizados. Assim, $S_i \sim N(0, \sigma^2)$, $i = 1, ..., n$, apeasr de não serem independentes $(Corr(S_i, S_j) = Corr(R_i, R_j))$.

Vantagem dos S_i's: se a correlação entre S_i e S_j for pequena $\forall\ i, j$, então os S_i's são quase independentes, além de serem identicamente distribuídos. Neste caso, faz sentido verificar a normalidade dos resíduos via normal plot de $s_1, s_2, ..., s_n$.

A interpretação de gráficos dos resíduos é imediata se estes são aproximadamente iid's em contraste com o caso das variâncias desiguais dos resíduos brutos onde um resíduo grande poderia tanto ser um outlier como poderia ser grande devido a uma variância grande.

Observe que $Var(\widehat{\mu}) = \sigma^2 H$ e

$Var(R) = \sigma^2(I - H)$ tal que

$Var(R_i) = \sigma^2(1 - h_{ii})$, $i = 1, ..., n \Rightarrow$

$0 \leq 1 - h_{ii} \leq 1 \Rightarrow 0 \leq h_{ii} \leq 1$, $i = 1, ..., n$.

Assim, a variância de R_i pode ser qualquer valor entre 0 e σ^2.

Usando os resíduos padronizados, corrigimos os resíduos brutos para o problema de variâncias desiguais. Porém, o fato de os resíduos terem variâncias diferentes é em si uma fonte de informação.

EXEMPLO 5.1: $Y_{ij} \overset{ind}{\sim} N(\beta_i, \sigma^2)$, $i = 1, ..., k; j = 1, ..., n_i$. Vimos que $\widehat{\beta} + i = \bar{Y}_{i+}$ e $R_{ij} = Y_{ij} - \bar{Y}_{i+}$.

Neste caso, a matriz H é bloco diagonal cujos k blocos são de dimensão $n_i \times n_i$ e cujas entradas são $1/n_i$. Assim,

$$R_{ij} \sim N\left(0, \sigma^2(1 - \frac{1}{n_i})\right), \quad i = 1, ..., k; j = 1, ..., n_i$$

$n_i = 1 \Rightarrow R_{ij}$ é identicamente igual a zero.

$n_i = 2 \Rightarrow R_{ij} \sim N(0, \sigma^2/2)$.

Por outro lado, $Var(R_{ij}) \rightarrow \sigma^2$ quando $n_i \rightarrow \infty$.

Em termos da correlação, observe que $R_{i_1 j_1}$ e $R_{i_2 j_2}$ são não correlacionados com $i_1 \neq i_2$.

Para resíduos do mesmo grupo temos

$$corr(R_{ij}, R_{il}) = \frac{-1}{n_i - 1}.$$

5.1.3 Influência (*Leverage*)

$$Var(\widehat{\mu}_i) = \sigma^2 h_{ii}$$

$\Rightarrow h_{ii}$ é a precisão com que μ_i é estimado relativo à σ^2.

Um pequeno valor de h_{ii} indica que o estimador de μ_i é baseado sobre a contribuição de muitas observações.

Por exemplo, no modelo de análise de variância a um fator temos $Var(\widehat{\mu}_i) = Var(\bar{Y}_{i+}) = \sigma^2/n_i$

Assim, se h_{ii} é pequeno, isto ocorre porque existem muitas observações que contribuem para a estimação de μ_i.

Se $h_{ii} \simeq 1 \Rightarrow Var(R_i) = Var(Y_i - \widehat{\mu}_i) \simeq 0$, caso no qual $\widehat{\mu}_i$ tende a ser próximo de Y_i.

Em outras palavras, $\widehat{\mu}_i$ é predominantemente determinado pela única observação Y_i.

$Cov(Y_i, \widehat{\mu}_i) = cov(Y_i, \sum_j h_{ij} Y_j) = \sum_j h_{ij} cov(Y_i, Y_j) = h_{ii}\sigma^2$
$\Rightarrow corr(Y_i, \widehat{\mu}_i) = h_{ii}^{1/2}$.

Assim, se $h_{ii} \simeq 1$, a correlação entre $\widehat{\mu}_i$ e Y_i é aproximadamente 1. Este fenômeno é chamado "influência" (*leverage*) significando que para $h_{ii} \simeq 1$, Y_i exerce grande influência sobre $\widehat{\mu}_i$.

EXEMPLO 5.2: $Y_i \overset{ind}{\sim} N(\beta_1 + \beta_2 x_i, \sigma^2)$,

$$i = 1, ..., n \text{ onde } x_i = \begin{cases} -1, & i = 1, ..., k \\ 0, & i = k+1, .., k+l \\ k, & i = k+l+1 (= n) \end{cases}$$

A regressão é pobremente determinada para valores de x próximos de k, pois só há uma observação neste ponto.

Para k grande, esperamos que Y_n exerça grande influência sobre $\widehat{\mu}_n = \widehat{\beta}_1 + k\widehat{\beta}_2$.

A matriz X deste modelo é

$$X = \begin{pmatrix} 1 & ... & 1 & 1 & ... & 1 & 1 \\ -1 & ... & -1 & 0 & ... & 0 & k \end{pmatrix}$$

Pode-se verificar neste caso que

$$h_{nn} = \frac{k + 1 + k(k + l + 1)}{k + l + 1 + k(k + l + 1)}$$

Se l é relativamente pequeno quando comparado com k, temos $h_{nn} \simeq 1$, caso no qual Y_n exerce grande influência sobre $\widehat{\mu}_n$.

5.2 Resíduos "Studentized"

Os resíduos podem ser usados para pelo menos menos três propósitos

(1) investigar a hipótese de variância constante

(2) investigar a hipótese de normalidade

(3) investigar a hipótese de linearidade

Consideraremos agora uma quarta questão, a questão dos outliers.

Por um OUTLIER, queremos nos referir a uma observação que não segue a distribuição especificado pelo ML dado.

Consideraremos aqui um caso mais restrito onde uma observação segue o ML dado exceto que sua média é diferente daquela especificada pelo modelo.

Suponha que o modelo nulo seja o ML dado por $Y = X\beta + \epsilon$.

Seja o modelo não-nulo correspondente à média de Y_i para algum i sendo não mais μ_i mas $E[Y_i] = \mu_i + \beta_{k+1}$, onde β_{k+1} é um parâmetro novo, a ser estimado junto com β.

Verifica-se que o modelo dado é um modelo linear $L_0 = span\{x_1, ..., x_k, x_{k+1}\}$ onde $x_1, ..., x_k$ são as colunas de X e x_{k+1} é o vetor

$(0, ..., 0, 1, 0, ..., 0)^T$ com um 1 na i-ésima entrada.

Suponha que $x_1, ..., x_k, x_{k+1}$ seja uma base para L_0.

O ML original dado por $L_1 = span\{x_1, ..., x_k\}$ é agora um submodelo de L_0 e podemos assim, testar L_1 sob L_0. (teste de $\beta_{k+1} = 0$)

De acordo com técnica estudada no Cap. 2, devemos modificar x_{k+1} de forma a termos uma base ortoginal para L_0 e simplificar o cálculo dos estimadores.

Assim, seja $z = x_{k+1} - p_1(x_{k+1}) = x_{k+1} - Hx_{k+1} = (I - H)x_{k+1}$.

Então, z é ortogonal à L_1 e devido ao fato de que $x_1, ..., x_k$ são li's, temos $z \neq 0$.

Pela ortogonalidade da base, a projeção sobre L_0 é

$p_0(y) = p_1(y) = \frac{z.y}{||z||^2}z$, $p_1(y) = X\widehat{\beta}$ onde $\widehat{\beta} = (X^TX)^{-1}X^TY$ é o estimador de β sob L_1 e,

$z.Y = x_{k+1}^T(I - H)Y = x_{k+1}^TR = R_i$ onde R é o vetor de resíduos brutos.

Além disso, $||z||^2 = x_{k+1}^T (I-H)^T (I-H) x_{k+1} = 1 - h_{ii}$.

Assim, o estimador de β_{k+1} é $\widehat{\beta}_{k+1} = \frac{R_i}{1-h_{ii}}$ e sob L_0, $\widehat{\beta}_{k+1} \sim N(0, \frac{\sigma^2}{1-h_{ii}})$.

Então, o teste t para a hip. $\beta_{k+1} = 0$ é

$$T_i = \frac{\widehat{\beta}_{k+1}}{\tilde{\sigma}_0/(1-h_{ii})^{1/2}} = \frac{R_i}{\tilde{\sigma}_0 (1-h_{ii})^{1/2}} = \frac{S_i}{\tilde{\sigma}_0}$$

onde $\tilde{\sigma}_0^2$ é o estimador de σ^2 sob L_0.

Como $dim L_0 = k+1$ temos, sob L_1,

$$\frac{S_i}{\tilde{\sigma}_0} \sim t_{(n-k-1)}.$$

Para encontrar $\tilde{\sigma}_0^2$, observe que $z \perp L_1$ e usando o TP junto com $p_0(y) = p_1(y) + \frac{z.y}{||z||^2} z$, segue

$$||Y - p_0(Y)||^2 = ||Y||^2 - ||p_0(Y)||^2 =$$

$$= ||Y - p_1(Y)||^2 - \underbrace{\frac{R_i^2}{1 - h_{ii}}}_{=S_i^2}$$

Como $\tilde{\sigma}_1^2 = \frac{||Y - p_1(Y)||^2}{n-k}$, obtemos

$$T_i = S_i \left\{ \frac{n-k-1}{n-k)\tilde{\sigma}_1^2} - S_i^2 \right\}^{1/2}$$

T_i é uma função monótona de S_i para $\tilde{\sigma}_1^2$ fixado.

Além disso, no caso extremo onde as $n - 1$ observações tirando Y_i ajustam o modelo perfeitamente tal que $\tilde{\sigma}_0^2 = 0$ ou $(n - k)\tilde{\sigma}_1^2 = S_i^2$, obtemos $T_i = \infty$.

Assim, T_i tenderá a enfatizar observações extremas. Isto também segue do fato de que no modelo L_0 β é estimado a partir das $n-1$ observações tal que se Y_i é um outlier, a diferença real entre Y_i e as observações restantes é capturada por T_i.

Chamamos os T_i's de resíduos "studentized".

5.3 A dsitribuição não-nula dos resíduos

A utilidade dos resíduos depende de sua habilidade em detectar desvios do modelo sob consideração. Consideraremos aqui a distribuição dos resíduos quando o verdadeiro modelo é diferente do modelo sob consideração.

Seja o modelo nulo dado por

$Y = \mu + \epsilon$, onde $\epsilon_i \overset{ind}{\sim} N(0, \sigma^2)$ e $\mu = X\beta$.

Suponha que a distribuição não-nula seja

$Y = \mu + u + \epsilon$ para algum vetor u e onde agora,

$\epsilon_i \overset{ind}{\sim} N(0, \sigma^2 w_i)$, $i = 1, ..., n$; $w_1, ..., w_n$

são positivos e chamados pesos (mais precisamente, são os inversos dos pesos).

$\Rightarrow$ A distribuição-nula é uma versão perturbada da dist. nula.

É possível considerar modelos não nulos mais gerais onde, por exemplo, os erros não são mais normalmente distribuídos, ou s̃ao dependentes de alguma forma.

Porém as suposições feitas aqui são suficientes para levantar alguns pontos importantes.

Considere agora a distribuição não-nula do vetor de resíduos.

Usando $R = (I - H)Y$ obtemos

$$E[R] = (I-H)E[Y] = (I-H)(\mu+u) = (I-H)u$$

onde aqui usamos o fato de que $H\mu = \mu$ devido ao fato de que $\mu \in L_1$, com $L_1 = span\{X\}$ denotando o modelo nulo.

Similarmente, a matriz de variância para R é, escrevendo $W = diag\{w_1, ..., w_n\}$ e usando as propriedades de H

$$Var(R) = \sigma^2(I - H)W(I - H) =$$
$$\sigma^2\{(I - H)W - WH + HWH\}$$

Assim, a variância de R_i é $Var(R_i) = \sigma^2\{(1 - h_{ii})w_i - w_i h_{ii} + \sum_{j=1}^{n} h_{ij} w_j h_{ji}\}$

$= \sigma^2\{(1 - h_{ii})w_i - w_i h_{ii} + \sum_j w_j h_{ij}^2\}$

Como $H^2 = H$ e $H^T = H$ obtemos $h_{ii} = \sum_j h_{ij}^2$ e, assim,

$Var(R_i) = \sigma^2\{(1 - h_{ii})w_i - w_i \sum_j h_{ij}^2 + \sum_j w_j h_{ij}^2\}$

$$= \sigma^2\{(1 - h_{ii} w_i + \sum_j h_{ij}^2 (w_j - w_i)\}$$

Assim, a distribuição não-nula de R é normal multivariada com vetor de média $(I - H)u$ e matriz de variância $\sigma^2 (I - H)W(I - H) = \sigma^2\{(I - H)W - WH + HWH\}$.

O vetor de média e a matriz de variância representam versões distorcidas da média e variância perturbadas, tornando possível, em princípio, descobrir, via análise de resíduos, se existe algo errado com o modelo suposto para os dados.

5.3.2 Estabilização da Variância

Os resíduos padronizados S_i, tendo variância comum sob o modelo nulo são adequados para mostrar a não homogeneidade da variância no modelo não-nulo.

Sob o modelo não-nulo temos

$$Var(S_i) = \frac{1}{1-h_{ii}} Var(R_i) =$$
$$\sigma^2 \{w_i + \frac{1}{1-h_{ii}} \sum_j h_{ij}^2 (w_j - w_i)\}$$

Assim, o resíduo padronizado S_i tem aproximadamente a mesma variância que ϵ_i no caso não-nulo, particularmente se os h_{ij}'s são pequenos (e assim h_{ij}^2 é bem menor), e se os w_i's não são muito diferentes.

O caso mais comum de não homogeneidade de variância ocorre quando os pesos têm a forma

$$w_i = V(\mu_i), \quad i = 1, ..., n$$

para alguma função V, chamada função de variância.

A distribuição nula corresponde ao caso de uma função de variância constante.

Se $u = 0$, um gráfico de S_i versus $\hat{\mu}_i$ revelará a forma de V.

Uma função de variância pode ser frequentemente eliminada por uma transformação não-linear de Y. Seja $Z_i = f(Y_i)$ para alguma função suave "smoth" f, e expanda f próxima de μ, para obter

$$Z_i \simeq f(\mu_i) + (Y_i - \mu_i)f'(\mu_i)$$

Assim, se σ^2 é pequeno, então como

$\epsilon_i \overset{ind}{\sim} N(0, \sigma^2 w_i)$ e $w_i = V(\mu_i)$ segue que

$$Var(Z_i) \simeq Var(Y_i) f'(\mu_i)^2 = \sigma^2 V(\mu_i) f'(\mu_i)^2$$

Sepodemos escolher f satisfazendo $f'(\mu) = V(\mu)^{-1/2}$, obtemos $Var(Z_i) \simeq \sigma^2$, tal queZ_i tem variância aproximadamente constante.

Se Z_i também é normalmente distribuído, obtemos uma nova variável que satisfaz as exigências básicas de um modelo linear, a saber, normalidade e variância constante.

Esta escolha de f é chamada transformação de estabilização da variância.

Em muitos casos, uma transformação de estabilização da variância também tenderá a simetrizar e normalizar a distribuição, apesar do fato de que estes propósitos podem, em princípio, exigir transformações diferentes.

Observe também que algumas vezes pode-se de fato assumir que ambos Z_i e Y_i são normalmente distribuídos aproximadamente. Isto pode acontecer se σ^2 é tão pequeno que f é aproximadamente linear sobre o conjunto onde a densidade normal é efetivamente positiva.

Uma função de variância frequentemente encontrada é $V(\mu) = \mu^2$, para a qual a função de estabilização da variância é determinada por $f'(\mu) = 1/\mu$, fornecendo $f(\mu) = \log \mu$. Neste caso, $Z_i = \log \mu_i$ tem variância constante aproximadamente. Esta transformação torna efeitos multiplicativos em efeitos aditivos, como já foi ilustrado no exemplo das árvores (Cap.4), e torna funções de poder em funções lineares.

Para a função de variância $V(\mu) = \mu^{\alpha}$, $\alpha \neq 2$, obtemos $f'(\mu) = \frac{1}{\mu^{\alpha/2}}$ e,

$$f(\mu) = \frac{1}{1-\alpha/2}\mu^{1-\alpha/2}$$

Por exemplo, se $\alpha = 1$, que é o caso da distribuição de Poisson, a transformação de estabilização da variância é $f(\mu) = 2\sqrt{\mu}$.

Para detectar a presença de não-homogeneidade da função de variância, fazemos um gráfico de s_i versus $\widehat{\mu}_i$. Se os valores de S_i variam de uma forma sistemática com $\widehat{\mu}_i$, tentamos adivinhar a forma de V fazendo a correspondente transformação de estabilização da variância e verificando se a variável transformada tem variância constante.

Se necessário, este procedimentopode ser repetido mais de uma vez até que uma função de estabilização adequada seja encontrada.

Se o gráfico de S_i versus $\widehat{\mu}_i$ tem uma forma de megafone, deve-se tentar a transformação logarítmica primeiro.

É importante atacar o problema de variância constante primeiro na análise de dados com modelos lineares. Isto porque uma transformação de estabilização é não-linear, e assim, geralmente afeta a relação entre a variável dependente e as variáveis independentes, enquanto que a linearização da relação entre Y e x's requer transformações dos x's, que em geral não afeta a constância da variância.

Por outro lado, um u não-nulo pode algumas vezes provocar uma distorção nos gráficos de variância tal que pode-se ter que verificar que a variância é ainda constante depois de tentar eliminar a não-linearidade no modelo.

5.3.3 Gráficos Parciais

Suponha que a variância tenha sido estabilizada. Assim, podemos continuar o processo e verificar a linearidade da regressão para cada uma das variáveis independentes. A forma mais simples de fazer isto é construir gráficos dos resíduos s_i contra cada uma das variáveis independentes no modelo. Os gráficos parciais que definiremos agora contêm essencialmente a mesma informação, mas têm a vantagem de representar aproximadamente a verdadeira relação entre $E[Y]$ e x para cada variável independente.

Defina $\widehat{Y}_{ij} = x_{ij}\widehat{\beta}_j + R_i = Y_i - \sim_{l \neq j} x_{il}\widehat{\beta}_l$

onde $\widehat{\beta} = (\widehat{\beta}_1, ..., \widehat{\beta}_k)^t$ é o estimador nulo de β. Sob o modelonulo temos $E(R_i) = 0$ e, portanto, $E[\widehat{Y}_{ij}] = x_{ij}\beta_j$.

$\Rightarrow$ um gráfico de $\widehat{y}_{ij}$ versus x_{ij}, $i = 1, ..., n$ mostrará uma relação linear através da origem, se o modelo for correto. Se a relação entre mu_i e x_{ij} não é linear, veremos que o gráfico mostrará, essencialmente, a verdadeira relação, sugerindo assim uma transformação de x para obter linearidade.

Suponha que a verdadeira relação entre μ e as variáveis independentes $x_1, ..., x_k$ seja linear, exceto para x_j onde a relação é dada pela função $f(x_{ij})$. Podemos então representar o modelo não-nulo por

$$f(x_{ij}) = x_{ij}\beta_j + u_i$$

para algumvetor u adequado. A média do vetor de resíduos brutos é então

$$E[R] = (I - H)(\mu + u) = (I - H)u$$

e a média do estimador $\widehat{\beta}$ é

$$E[\widehat{\beta}] = (X^T X)^{-1} X^T (\mu + u) = \beta + (X^T X)^{-1} X^T u$$

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

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

Assim, no caso especial onde $u \in L^{\perp}$, tal que $Hu = 0$ enconstramos

$$E[\widehat{Y}_{ij}] = x_{ij}\beta_j + u_i = f(x_{ij})$$

caso no qual o gráfico parcial dos resíduos mostra a verdadeira relação entre μ e x_j.

Em geral, porém, não podemos escolher β_j para obter $u \in L^{\perp}$ e portanto, o gráfico parcial dos resíduos representa uma versão distorcida da relação entre μ e x_j.

Além disso, pode existir a presença de não linearidade em outras variáveis de forma que devemos fazer um ciclo através das variáveis mais de uma vez para eliminar a não linearidade.

A principal vantagem dos gráficos parciais é que ele torna possível a distinção entre uma relação monótona porrém não linear tal como uma função de poder ou uma relação logarítmica e uma relação não monótona e não linear tal como uma relação quadrática ou cúbica.

No primeiro caso, a não linearidade pode ser removida transformando-se a variável independente em questão pela transformação sugerida pelo gráfico, enquanto que uma relação quadrática ou cúbica requer a adição de um termo quadrático ou cúbico no modelo.

(EXEMPLO)

5.4 Elementos de Análise de Resíduos

5.4.1 Discussão Geral

Vimos diversas definições de resíduos e que os mesmos servem para diferentes propósitos. Como uma recomendação geral, sugerimos que qualquer análise de regressão múltipla deve incluir pelo menos um dos seguintes gráficos na ordem indicada:

(1) s_i versus $\widehat{\mu}_i$

(ii) gráficos parciais dos resíduos para cada variável independente

(iii) t_i versus h_{ii}

(iv) normal plot para s_i

A função de cada um destes gráficos é

(i) revelar a não homogeneidade da função de variância. Se a função de variância não é constante, tenta-se uma tranformação de estabilização de variância para Y.

(ii) revelar a não linearidade na relação entre Y e cada uma das variáveis independentes x_i, $i, 1..., k$. Na presença de não linearidade, tenta-se tranformar x pela tranformação sugerida pelo gráfico. Se a relação é não monótona, adiciona-se um termo quadrático βx^2 no modelo.

(iii) revelar outliers (t_i grande) ou influência (h_{ii} grande). Deve-se prestar atenção especial às observações onde ambos t_i e h_{ii} são grandes. O t_i deve ser comparado com uma distribuição $t_{(n-k-1)}$. O tamanho médio de h_{ii} é k/n e um valor $h_{ii} > 2k/n$ é considerado grande.

(iv) confirmar a normalidade de S_i e de Y_i. Este gráfico deve ser feito somente uma vez, após a resolução das questões principais na análise (não linearidade, variância não constante).

Vale lembrar que os resíduos são combinaçõeslineares das observações e, portanto, pelo Teorema Central do Limite, tendem a ser normais, mesmo que as observações não sejam normais. Uma distribuição fortemente não normal se revelará ela própria via uma forma sigmóide ou curvada no gráfico. Neste caso, pode-se aplicar uma tranformação de estabilização da variância ou transformação de normalização para Y, em particular, escolhendo uma transformação diferente caso alguma transformação já tenha sido usada.