

Modelos Lineares Generalizados

Unidade III

Terezinha K. A. Ribeiro

terezinha.ribeiro@unb.br

Instituto de Ciências Exatas
Departamento de Estatística
Universidade de Brasília

- ▶ Introdução
- ▶ Resíduos e suas propriedades
- ▶ Gráfico de probabilidade normal
- ▶ Envelope simulado
- ▶ Alavancagem
- ▶ Medida de influência

Após o ajuste de um MLG, se faz necessário realizar uma análise de diagnóstico para verificar a qualidade de ajuste do modelo aos dados.

Devemos avaliar se existem afastamentos importantes das suposições feitas:

- ▶ distribuição da resposta Y dadas as covariáveis;
- ▶ independência dos Y_i 's;
- ▶ estrutura de variância da resposta;
- ▶ função de ligação escolhida;
- ▶ linearidade.

Também, será necessário avaliar o impacto de possíveis observações discrepantes:

- ▶ pontos aberrantes são pontos que possuem ajustes ruins para a variável resposta, e conseqüentemente, possuem resíduos discrepantes. Usualmente pontos aberrantes distorcem o intercepto do modelo ajustado;
- ▶ pontos alavanca são pontos que possuem peso desproporcional no próprio valor ajustado \hat{y} . Estes pontos geralmente possuem perfil diferente dos demais no que diz respeito aos valores das covariáveis.
- ▶ pontos influentes são aqueles pontos que exercem influência desproporcional nas estimativas dos parâmetros do modelo e podem causar mudança inferencial.

Considere o modelo normal linear definido por

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i,$$

com $\mu_i = E(Y_i) = \mathbf{x}_i^\top \boldsymbol{\beta}$, e $\varepsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$.

Matricialmente, o modelo é definido por

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

em que \mathbf{Y} é o vetor de respostas, \mathbf{X} é a matriz do modelo não aleatória de dimensão $n \times p$ tal que $r(\mathbf{X}) = p$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\varepsilon}$ é o vetor de erros tal que $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

Sob este modelo, devemos verificar se os erros possuem distribuição normal com média zero, variância constante e se estes são não correlacionados. Entretanto, os erros aleatórios não são observáveis.

O que podemos fazer?

Uma solução é avaliar uma medida calculada sob o modelo ajustado que possua características parecidas com erro aleatório.

Se esta medida sob o modelo ajustado apresentar tais características, então teremos indícios de que as suposições feitas sobre o erro estão adequadas.

Tal medida é o resíduo. A principal função de qualquer resíduo é indicar se o modelo suposto para os dados produz um ajuste adequado.

O resíduo ordinário $\hat{\varepsilon}_i$ é definido por

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \hat{\mu}_i.$$

Este resíduo é o mais simples que pode ser calculado em modelo ajustado.

O resíduo ordinário $\hat{\varepsilon}_i$ possui as características do erro aleatório ε_i ?

Matricialmente, temos que

$$\hat{\varepsilon} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Y} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = (\mathbf{I}_n - \mathbf{H}) \mathbf{Y},$$

em que $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$.

Segue que

$$\begin{aligned} \mathbf{E}(\hat{\varepsilon}) &= (\mathbf{I}_n - \mathbf{H})\mathbf{E}(\mathbf{Y}) \\ &= [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \mathbf{X} \boldsymbol{\beta} \\ &= [\mathbf{X} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}] \boldsymbol{\beta} \\ &= (\mathbf{X} - \mathbf{X}) \boldsymbol{\beta} \\ &= \mathbf{0}. \end{aligned}$$

Portanto, $\mathbf{E}(\hat{\varepsilon}_i) = 0$, para todo $i = 1, 2, \dots, n$.

Também,

$$\begin{aligned}\text{Var}(\hat{\varepsilon}) &= (I_n - H)\text{Var}(\mathbf{Y})(I_n - H)^\top \\ &= (I_n - H)\sigma^2 I_n (I_n - H) \\ &= \sigma^2 (I_n - H)^2 \\ &= \sigma^2 (I_n - H),\end{aligned}$$

em que este resultado decorre do fato de $(I_n - H)$ ser uma matriz simétrica e idempotente.

Logo,

$$\text{Var}(\hat{\varepsilon}_i) = \sigma^2(1 - h_{ii}), \quad \text{Cov}(\hat{\varepsilon}_i, \hat{\varepsilon}_j) = -\sigma^2 h_{ij}, \quad i \neq j,$$

com $i, j = 1, \dots, n$, e h_{ij} sendo o elemento (i, j) da matriz H .

A variância dos resíduos ordinários depende da i -ésima observação, e portanto, não é constante com i .

Ainda,

$$\text{Corr}(\hat{\varepsilon}_i, \hat{\varepsilon}_j) = -\frac{\sigma^2 h_{ij}}{\sqrt{\sigma^2(1-h_{ii})}\sqrt{\sigma^2(1-h_{jj})}} = -\frac{h_{ij}}{\sqrt{(1-h_{ii})(1-h_{jj})}}.$$

Assim, os resíduos ordinários são correlacionados pois $\text{Cov}(\hat{\varepsilon}_i, \hat{\varepsilon}_j) \neq 0$ para $i \neq j$.

Entretanto, veremos que a medida h_{ij} tende a ser um valor pequeno quando n cresce. Portanto, espera-se que $\text{Cov}(\hat{\varepsilon}_i, \hat{\varepsilon}_j)$ seja um valor pequeno.

Por fim, perceba que

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} = y_i - \mathbf{x}_i^\top (X^\top X)^{-1} X^\top \mathbf{Y},$$

em que $\mathbf{x}_i^\top = (x_{i1} \ x_{i2} \ \dots \ x_{ip})^\top$. Então, o resíduo ordinário $\hat{\varepsilon}_i$ é uma combinação linear dos elementos de $\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, \sigma^2 I_n)$.

Como qualquer combinação linear dos elementos de um vetor normalmente distribuído segue uma distribuição normal univariada, temos que

$$\hat{\varepsilon}_i \sim N(0, \sigma^2(1 - h_{ii})), \quad i = 1, 2, \dots, n.$$

Resumindo, obtivemos que

- ▶ os resíduos ordinários $\hat{\varepsilon}_i$ possuem média zero e são normalmente distribuídos, assim como os erros aleatórios ε_i ;
- ▶ os resíduos ordinários $\hat{\varepsilon}_i$ possuem variância não constante e são correlacionados, diferente dos erros aleatórios ε_i .

Portanto, o resíduo ordinário $\hat{\varepsilon}_i$ não reproduz todas as características do erro aleatório.

Em particular, como avaliaremos se a suposição de homoscedasticidade do modelo normal linear está adequada através de $\hat{\varepsilon}_i$ se, sob o modelo, este tem variância não constante?

Podemos definir um novo resíduo através da padronização de $\hat{\varepsilon}_i$.

Sabemos que $\text{Var}(\hat{\varepsilon}_i) = \sigma^2(1 - h_{ii})$. Porém, σ^2 é um parâmetro desconhecido, então devemos substituí-lo pela sua estimativa $\hat{\sigma}^2$, obtendo assim o seguinte resíduo:

$$t_i = \frac{\hat{\varepsilon}_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}}.$$

O resíduo t_i é chamado de resíduo padronizado ou de resíduo studentizado internamente.

Mostra-se que

$$\frac{t_i^2}{n-p} \sim \text{Beta}\left(\frac{1}{2}, \frac{n-p}{2}\right).$$

em que $\text{Beta}(a,b)$ denota a distribuição beta de parâmetros $a > 0$ e $b > 0$. Daí, prova-se que $E(t_i) = 0$ e $\text{Var}(t_i) = 1$ para $i = 1, 2, \dots, n$, e

$$\text{Corr}(t_i, t_j) = -\frac{h_{ij}}{\sqrt{(1-h_{ii})(1-h_{jj})}}, \quad i \neq j.$$

Veja que apesar de t_i possuir variância constante, estes resíduos são correlacionados.

Qual a distribuição do resíduo padronizado t_i ? Note que

$$t_i = \frac{\hat{\varepsilon}_i}{\sqrt{\hat{\sigma}^2(1-h_{ii})}} = \frac{\frac{\hat{\varepsilon}_i}{\sqrt{\sigma^2(1-h_{ii})}}}{\sqrt{\frac{\text{SQRes}}{\sigma^2(n-p)}}}.$$

Entretanto, $\hat{\varepsilon}_i$ não é independente de SQRes (ou de $\hat{\sigma}^2$), e consequentemente, o resíduo t_i não segue distribuição t de Student com $(n-p)$ graus de liberdade.

A dependência entre $\hat{\varepsilon}_i$ e $\hat{\sigma}^2$ pode ser contornada substituindo $\hat{\sigma}^2$ por $\hat{\sigma}_{(i)}^2$, que é a variância σ^2 estimada correspondente ao modelo ajustado excluindo a i -ésima observação do conjunto de dados.

Note que para obter $\hat{\sigma}_{(i)}^2$ com $i = 1, 2, \dots, n$, precisamos ajustar n modelos de regressão. Isto pode não ser razoável se o conjunto de dados tiver um número muito grande de observações.

Pode-se mostrar que $\hat{\sigma}_{(i)}^2$ pode ser calculada através de

$$\hat{\sigma}_{(i)}^2 = \hat{\sigma}^2 \left(\frac{n - p - t_i^2}{n - p - 1} \right).$$

Com este resultado, não será necessário ajustar n modelos de regressão.

O resíduo final, denominado de resíduo studentizado (ou resíduo studentizado externamente), é definido por

$$t_i^* = \frac{\hat{\varepsilon}_i}{\sqrt{\hat{\sigma}_{(i)}^2 (1 - h_{ii})}}.$$

Substituindo a expressão de $\hat{\sigma}_{(i)}^2$, o resíduo studentizado t_i^* pode ser reescrito na forma

$$t_i^* = t_i \sqrt{\left(\frac{n - p - 1}{n - p - t_i^2} \right)}.$$

Mostra-se que $t_i^* \sim t_{n-p-1}$, e portanto, para n grande, $t_i^* \stackrel{n \rightarrow \infty}{\sim} N(0,1)$.

Se para n grande, os resíduos studentizados seguem uma distribuição normal padrão, isto é indicação de que a suposição de normalidade dos erros está adequada.

Como podemos avaliar a normalidade assintótica dos resíduos? Pode-se construir o gráfico de probabilidade normal dos resíduos studentizados t_i^* .

Para identificar fugas da suposição de homoscedasticidade do modelo, construímos o gráfico dos resíduos (t_i^*) versus os valores ajustados (\hat{Y}_i).

Observações com $|t_i^*| > 2$ devem ser investigados na análise.

Observações com $|t_i^*| > 3$ são chamados de aberrantes.

Gráfico dos resíduos studentizados versus valores ajustados

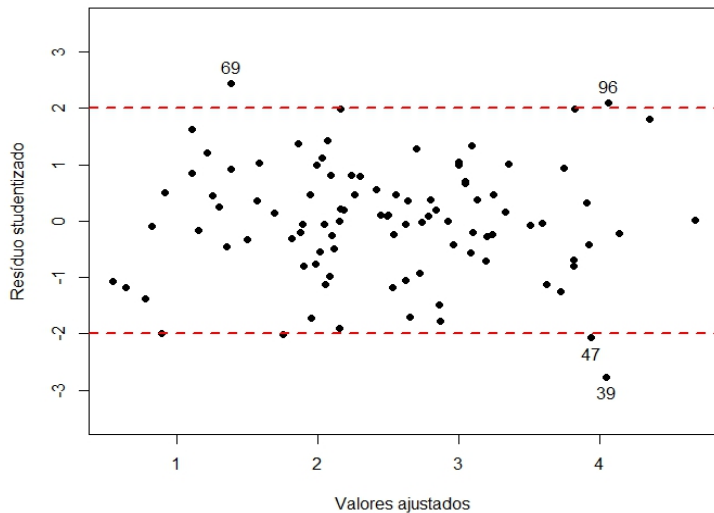


Gráfico dos resíduos studentizados versus valores ajustados

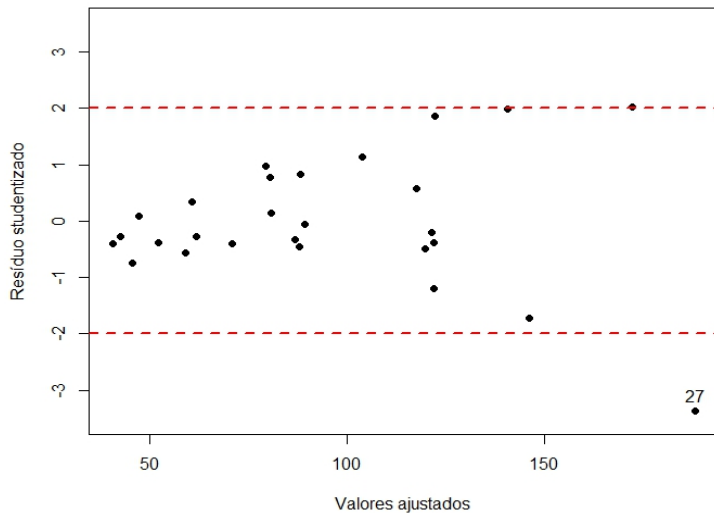


Gráfico de probabilidade normal dos resíduos

Sejam t_i^* , $i = 1, \dots, n$, os n resíduos studentizados calculados sob o modelo de regressão ajustado. Para avaliar se a distribuição de probabilidades suposta para a resposta está adequada, construiremos inicialmente o gráfico de probabilidades normal dos resíduos t_i^* .

Este gráfico será nada mais do que um gráfico dos resíduos ordenados $t_{[i]}^*$ (i -ésimo menor resíduo studentizado) versus os valores esperados das estatísticas de ordem $Z_{[i]}$, com $Z \sim N(0,1)$. Portanto, construiremos o gráfico de $(E(Z_{[i]}), t_{[i]}^*)$.

Mostra-se que

$$E(Z_{[i]}) \cong \Phi^{-1} \left(\frac{i - 3/8}{n + 1/4} \right), \quad i = 1, 2, \dots, n,$$

em que $\Phi^{-1}(\cdot)$ é a inversa da função de distribuição acumulada da distribuição normal padrão.

Gráfico de probabilidade normal dos resíduos

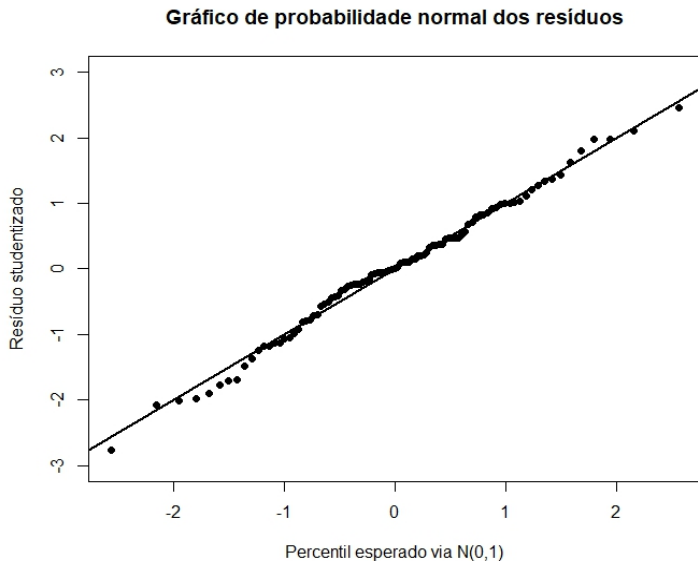


Gráfico de probabilidade normal dos resíduos

Gráfico de probabilidade normal dos resíduos

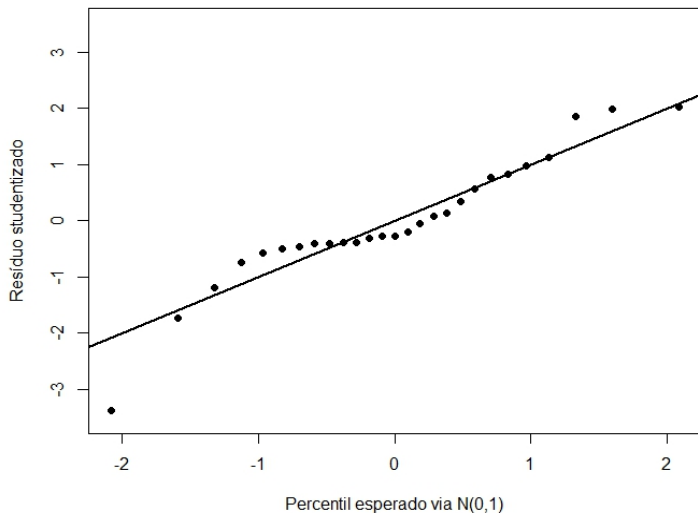


Gráfico de probabilidade normal do resíduos

A partir da análise do gráfico de probabilidade normal do resíduos não será possível avaliar com clareza até que ponto certas irregularidades no padrão linear dos resíduos studentizados podem ser consideradas aceitáveis.

Além disto, lembre-se que os resíduos studentizados são correlacionados e não possuem distribuição normal padrão para amostras finitas (n pequeno).

Por estas razões, se torna relevante construir juntamente com o gráfico normal de probabilidade dos resíduos t_j^* , um referencial para a flutuação dos pontos. Este referencial pode ser construído via simulação, e é comumente conhecido por envelope simulado ou bandas de confiança simuladas.

Envelope simulado

Para construir o envelope simulado, considere os passos a seguir:

- $P_1)$ Considerando a amostra (original), estime os parâmetros do modelo de regressão suposto e calcule os n resíduos t_i^* .
- $P_2)$ Sob o modelo de regressão avaliado nas estimativas dos parâmetros obtidas no Passo P_1 , gere uma nova amostra (réplica da amostra original) de tamanho n da variável resposta Y (com X fixada). Esta réplica de Y será denotada por $Y^{(1)} = (Y_1^1, Y_2^1, \dots, Y_n^1)^\top$.
- $P_3)$ Ajuste o modelo de regressão para a variável $Y^{(1)}$ dado x , e calcule os n resíduos (n réplicas) denotados por $t_{i,1}^*$, com $i = 1, 2, \dots, n$.

- P_4) Repita os passos P_2 e P_3 um número B suficiente de vezes. Por exemplo, tome $B = 100$. Ao fazer isto obteremos as quantidades apresentadas na tabela abaixo.

Tabela 1: Resíduos originais e simulados.

Resíduo Original	Réplica 1	Réplica 2	...	Réplica B
t_1^*	$t_{1,1}^*$	$t_{1,2}^*$...	$t_{1,B}^*$
t_2^*	$t_{2,1}^*$	$t_{2,2}^*$...	$t_{2,B}^*$
t_3^*	$t_{3,1}^*$	$t_{3,2}^*$...	$t_{3,B}^*$
\vdots	\vdots	\vdots	\vdots	\vdots
t_n^*	$t_{n,1}^*$	$t_{n,2}^*$...	$t_{n,B}^*$

Envelope simulado

P_5) Ordene (crescente) cada grupo de n resíduos, obtendo assim $t_{[i],j}^*$, $i = 1, \dots, n$, $j = 1, \dots, B$, conforme a tabela abaixo.

Tabela 2: Resíduos originais e simulados em ordem crescente.

Resíduo Original	Réplica 1	Réplica 2	...	Réplica B
$t_{[1]}^*$	$t_{[1],1}^*$	$t_{[1],2}^*$...	$t_{[1],B}^*$
$t_{[2]}^*$	$t_{[2],1}^*$	$t_{[2],2}^*$...	$t_{[2],B}^*$
$t_{[3]}^*$	$t_{[3],1}^*$	$t_{[3],2}^*$...	$t_{[3],B}^*$
\vdots	\vdots	\vdots	\vdots	\vdots
$t_{[n]}^*$	$t_{[n],1}^*$	$t_{[n],2}^*$...	$t_{[n],B}^*$

P_6) Para cada resíduo $t_{[i]}^*$, $i = 1, 2, \dots, n$, calcule um limite inferior (LI) e um limite superior (LS) usando as B réplicas.

Pode-se definir LI e LS como os percentis empíricos de ordem $(\alpha/2) \cdot 100\%$ e $(1 - \alpha/2) \cdot 100\%$, respectivamente, $0 < \alpha < 1$.

Denotando $t_{[i], [p \cdot 100]}^*$ o percentil de ordem $p \cdot 100\%$ obtido via as B réplicas de $t_{[i]}^*$, os limites empíricos de $t_{[i]}^*$ são

$$\text{LI}(t_{[i]}^*) = t_{[i], [(\alpha/2) \cdot 100]}^*, \quad \text{LS}(t_{[i]}^*) = t_{[i], [(1-\alpha/2) \cdot 100]}^*.$$

Por exemplo, fixando $\alpha = 0,10$ tem-se

$$\text{LI}(t_{[i]}^*) = t_{[i], [5]}^*, \quad \text{LS}(t_{[i]}^*) = t_{[i], [95]}^*.$$

P_7) Calcule a mediana empírica

$$\text{Med}(t_{[i]}^*) = t_{[i], [50]}^*$$

para cada resíduo $t_{[i]}^*$ com base nas respectivas B réplicas.

Se $B = 100$, então a mediana empírica para cada resíduo será a média entre as réplicas que ocupam as posições 50 e 51.

A mediana empírica também deve ser destacada no gráfico normal de probabilidades.

Os resíduos originais t_i^* devem se distribuir em torno desta mediana empírica.

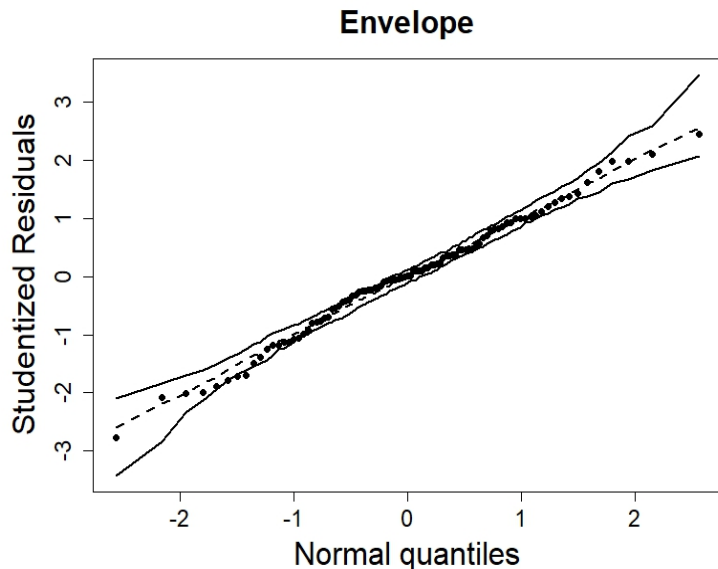
Envelope simulado

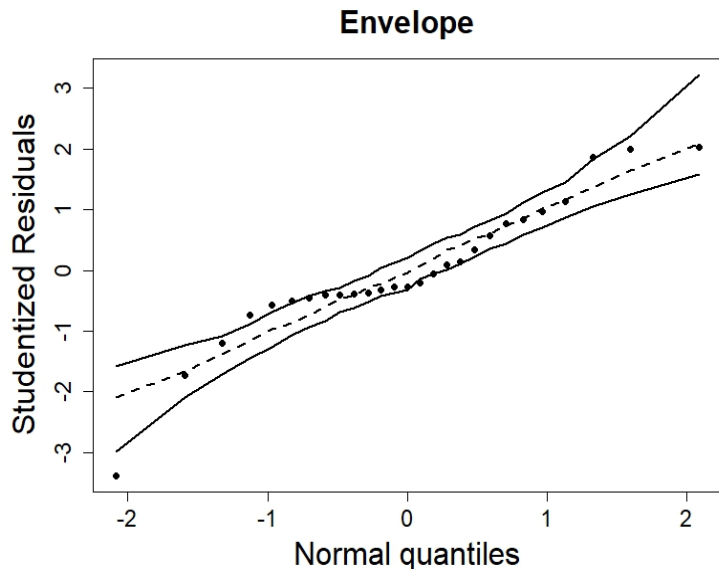
Para a construção do envelope utilizaremos as quantidades apresentadas na tabela abaixo.

Tabela 3: Resíduos originais, LI, LS e mediana empírica com $\alpha = 0,10$.

Resíduo Original	Limite Inferior	Mediana	Limite Superior
$t_{[1]}^*$	$t_{[1],[5]}^*$	$t_{[1],[50]}^*$	$t_{[1],[95]}^*$
$t_{[2]}^*$	$t_{[2],[5]}^*$	$t_{[2],[50]}^*$	$t_{[2],[95]}^*$
$t_{[3]}^*$	$t_{[3],[5]}^*$	$t_{[3],[50]}^*$	$t_{[3],[95]}^*$
\vdots	\vdots	\vdots	\vdots
$t_{[n]}^*$	$t_{[n],[5]}^*$	$t_{[n],[50]}^*$	$t_{[n],[95]}^*$

Para $\alpha = 0,10$, espera-se que 90% dos resíduos estejam distribuídos dentro do envelope simulado (bandas de “confiança”).





Extensão dos resíduos para os MLGs

Uma versão do resíduo studentizado para os MLGs pode ser definida. Entretanto, não necessariamente as propriedades que tínhamos sob regressão normal continuam valendo.

Assim, torna-se importante a definição de outros tipos de resíduo cujas propriedades sejam conhecidas.

O resíduo de Pearson é definido por

$$r_{p_i} = \frac{y_i - \mu_i}{\sqrt{\text{Var}(Y_i)}} = \frac{\sqrt{\phi}(y_i - \mu_i)}{\sqrt{V_i}}, \quad i = 1, 2, \dots, n,$$

desde que $\text{Var}(Y_i) = \phi^{-1} V_i$.

Extensão dos resíduos para os MLGs

Observe que $E(r_{pi}) = 0$ e

$$\text{Var}(r_{pi}) = \frac{\phi}{V_i} \text{Var}(Y_i) = \frac{\phi}{V_i} \phi^{-1} V_i = 1.$$

Note que o resíduo de Pearson é equivalente ao resíduo ordinário após padronização pela variância.

Na prática, deve-se calcular

$$\hat{r}_{pi} = \frac{\sqrt{\hat{\phi}}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i}} \quad i = 1, 2, \dots, n.$$

Assim, $\text{Var}(\hat{r}_{pi}) \approx 1$.

Empiricamente, observa-se que o resíduo de Pearson \hat{r}_{pi} não segue distribuição normal.

Extensão dos resíduos para os MLGs

O resíduo de Pearson studentizado é definido por

$$\hat{t}_{Si} = \frac{\sqrt{\hat{\phi}}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1 - \hat{h}_{ii})}} \quad i = 1, 2, \dots, n,$$

em que

$$\hat{h}_{ii} = \hat{\omega}_i \mathbf{x}_i^\top (\mathbf{X}^\top \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i$$

é a medida de alavancagem dos MLGs (veremos mais à frente).

A adição do termo $(1 - \hat{h}_{ii})$ faz com que a variância deste resíduo esteja mais próxima de um. Apesar disto, o resíduo de Pearson studentizado possui distribuição assimétrica mesmo para grandes amostras.

Extensão dos resíduos para os MLGs

O resíduo componente do desvio é definido por

$$\hat{t}_{Di} = \frac{d^*(y_i; \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}} = \frac{\sqrt{\phi} d(y_i; \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}},$$

em que

$$d(y_i; \hat{\mu}_i) = \pm \sqrt{2} \{y_i(\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i)\}^{1/2},$$

com \pm corresponde ao sinal de $(y_i - \hat{\mu}_i)$.

Alguns utilizam o resíduo do componente do desvio sem a correção por $\sqrt{1 - \hat{h}_{ii}}$. Este termo é incluído para que a variância de \hat{t}_{Di} seja mais próxima de um.

Extensão dos resíduos para os MLGs

O resíduo do componente do desvio é um dos mais utilizados sob os MLGs. Dentre os resíduos aqui apresentados, este é o que se aproxima mais da normalidade.

Entretanto, não é possível mostrar que este resíduo possui distribuição normal para amostras grandes.

De forma geral, não conhecemos a distribuição exata ou aproximada do resíduo \hat{t}_{Di} .

Portanto, para avaliar se a distribuição de probabilidade suposta para $Y_i | \mathbf{x}_i$ está adequada, precisamos necessariamente construir o envelope simulado.

Extensão dos resíduos para os MLGs

Seja $F(y_i; \mu_i, \phi)$ a função de distribuição acumulada de Y_i , isto é,

$$F(y_i; \mu_i, \phi) = P(Y_i \leq y_i).$$

Se F é contínua, então

- ▶ $F(y_i; \mu_i, \phi)$ é uniformemente distribuída no intervalo contínuo $(0,1)$, ou seja, $F(y_i; \mu_i, \phi) \sim U(0,1)$;
- ▶ $\Phi^{-1}(F(y_i; \mu_i, \phi)) \sim N(0,1)$;
- ▶ Y_1, \dots, Y_n são variáveis aleatórias independentes, então as variáveis $F(Y_1; \mu_1, \phi), \dots, F(Y_n; \mu_n, \phi)$ também são independentes.
- ▶ $\Phi^{-1}(F(y_1; \mu_1, \phi)), \dots, \Phi^{-1}(F(y_n; \mu_n, \phi))$ são independentes,

em que Φ^{-1} denota a função quantílica da distribuição normal padrão.

Extensão dos resíduos para os MLGs

Se a distribuição suposta para a resposta é contínua, o resíduo quantílico é definido por

$$r_{qi} = \Phi^{-1}(F(y_i; \hat{\mu}_i, \hat{\phi})), \quad i = 1, 2, \dots, n.$$

Observe que, para n finito, os resíduos quantílicos não são normais padrão independentes desde que calcula-se r_{qi} sob as estimativas $\hat{\mu}_i, \hat{\phi}$.

Entretanto, se $\hat{\mu}_i$ e $\hat{\phi}$ são estimadores consistentes de μ_i e ϕ , respectivamente, então $r_{qi} \stackrel{\text{iid}}{\sim} N(0, 1)$ para n grande.

Se y_i for um valor discrepante com relação ao modelo postulado F , o valor de $F(y_i; \hat{\mu}_i, \hat{\phi})$ será próximo de zero ou próximo de um. Consequentemente, $|r_{qi}|$ assumirá um valor maior ou igual que 3, indicando assim y_i como aberrante.

Extensão dos resíduos para os MLGs

Para o caso em que a resposta não é contínua, pode-se definir o resíduo quantílico aleatorizado:

$$r_{qi} = \Phi^{-1}(u_i), \quad i = 1, 2, \dots, n,$$

em que

$$u_i \sim U[F(y_i - 1; \hat{\mu}_i, \hat{\phi}), F(y_i; \hat{\mu}_i, \hat{\phi})].$$

A partir desta definição, o resíduo quantílico se torna contínuo para caso em que a resposta é discreta . Além disto, para n grande,

$$r_{qi} \stackrel{\text{iid}}{\sim} N(0, 1).$$

Até o momento conhecemos apenas um tipo de observação atípica que é o ponto aberrante. Entretanto, existem outros dois tipos de atipicidade que devem ser investigadas em um modelo de regressão ajustado: alavanca e influente.

Os pontos alavanca são pontos que possuem peso desproporcional no próprio valor ajustado \hat{Y} . Estes pontos geralmente possuem perfil diferente dos demais no que diz respeito aos valores das covariáveis.

O ponto de alavanca, além de influenciar no próprio valor ajustado da resposta, pode também exercer influência desproporcional nas estimativas dos coeficientes da regressão. No caso normal linear, o ponto de alavanca “puxa” a reta de regressão ajustada para próximo a si pois usualmente interfere no coeficiente angular estimado da reta.

Como podemos identificar pontos de alavanca?

Alavancagem

Para introduzir o conceito de alavancagem sob os MLGs, revisaremos este conceito sob regressão linear normal.

A medida de alavancagem da i -ésima observação é definida como sendo a taxa de variação instantânea de \hat{Y}_i no ponto Y_i , ou seja, é a derivada de \hat{Y}_i em relação a Y_i :

$$\frac{\partial \hat{Y}_i}{\partial Y_i} = \lim_{\epsilon \rightarrow 0} \frac{\hat{Y}_i^\epsilon - \hat{Y}_i}{\epsilon},$$

em que

- ▶ \hat{Y}_i é o ajuste de Y_i baseado em Y_1, Y_2, \dots, Y_n ;
- ▶ \hat{Y}_i^ϵ é o ajuste de Y_i baseado em $Y_1, Y_2, \dots, Y_i + \epsilon, \dots, Y_n$.

O valor \hat{Y}_i^ϵ é o ajuste de Y_i quando introduzimos uma contaminação infinitesimal no i -ésima resposta.

Portanto, a medida de alavancagem mensura o quanto o i -ésimo ajuste da resposta \hat{Y}_i é afetado por esta pequena contaminação.

Sob regressão linear normal, o vetor de valores ajustados para a resposta é

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = \mathbf{H}\mathbf{Y},$$

em que $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ é a matriz “chapéu” (*hat*).

Note que \mathbf{H} é uma matriz simétrica pois

$$\begin{aligned} \mathbf{H}^\top &= [\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top]^\top \\ &= [\mathbf{X}^\top]^\top [(\mathbf{X}^\top \mathbf{X})^{-1}]^\top \mathbf{X}^\top \\ &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \\ &= \mathbf{H}, \end{aligned}$$

desde que $\mathbf{X}^\top \mathbf{X}$ é simétrica, e portanto, sua inversa $(\mathbf{X}^\top \mathbf{X})^{-1}$ também será simétrica.

Além disso, a matriz H é idempotente desde que

$$\begin{aligned} H^2 &= X(X^\top X)^{-1}X^\top X(X^\top X)^{-1}X^\top \\ &= X(X^\top X)^{-1}X^\top. \end{aligned}$$

O elemento que ocupa a posição (i, j) é denotado por h_{ij} , com $i, j = 1, 2, \dots, n$.

Observe que a matriz do modelo satisfaz

$$X = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix} \Rightarrow X^\top = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n],$$

em que $\mathbf{x}_i^\top = (x_{i1} \ x_{i2} \ \cdots \ x_{ip})$, $i = 1, 2, \dots, n$.

Alavancagem

De $H = X(X^\top X)^{-1}X^\top$, segue que

$$\begin{aligned} H &= \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix} (X^\top X)^{-1} [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n] \\ &= \begin{bmatrix} \mathbf{x}_1^\top (X^\top X)^{-1} \mathbf{x}_1 & \mathbf{x}_1^\top (X^\top X)^{-1} \mathbf{x}_2 & \cdots & \mathbf{x}_1^\top (X^\top X)^{-1} \mathbf{x}_n \\ \mathbf{x}_2^\top (X^\top X)^{-1} \mathbf{x}_1 & \mathbf{x}_2^\top (X^\top X)^{-1} \mathbf{x}_2 & \cdots & \mathbf{x}_2^\top (X^\top X)^{-1} \mathbf{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_n^\top (X^\top X)^{-1} \mathbf{x}_1 & \mathbf{x}_n^\top (X^\top X)^{-1} \mathbf{x}_2 & \cdots & \mathbf{x}_n^\top (X^\top X)^{-1} \mathbf{x}_n \end{bmatrix}. \end{aligned}$$

Portanto, obtemos que

$$h_{ij} = \mathbf{x}_i^\top (X^\top X)^{-1} \mathbf{x}_j.$$

De $\hat{\mathbf{Y}} = H\mathbf{Y}$, obtemos que

$$\begin{aligned}\hat{Y}_i &= \mathbf{x}_i^\top (X^\top X)^{-1} \mathbf{x}_1 Y_1 + \cdots + \mathbf{x}_i^\top (X^\top X)^{-1} \mathbf{x}_n Y_n \\&= \sum_{j=1}^n \mathbf{x}_i^\top (X^\top X)^{-1} \mathbf{x}_j Y_j \\&= \sum_{j=1}^n h_{ij} Y_j \\&= h_{ii} Y_i + \sum_{j=1, j \neq i}^n h_{ij} Y_j.\end{aligned}$$

Isto implica que a medida de alavancagem é dada por

$$\frac{\partial \hat{Y}_i}{\partial Y_i} = h_{ii} = \mathbf{x}_i^\top (X^\top X)^{-1} \mathbf{x}_i.$$

Quais são as características da medida de alavancagem h_{ij} ?

Primeiro, observe que $h_{ii} > 0, \forall i$. Isto é válido desde que a matriz $(X^T X)^{-1}$ é positiva definida, isto é,

$$\mathbf{z}^T (X^T X)^{-1} \mathbf{z} > 0, \quad \forall \mathbf{z} \neq \mathbf{0},$$

desde que a matriz X seja de posto completo.

Então, temos que

$$h_{ii} = \mathbf{x}_i^T (X^T X)^{-1} \mathbf{x}_i > 0, \quad \forall \mathbf{x}_i \neq \mathbf{0}.$$

Também, como a matriz H é idempotente e simétrica, segue que

$$h_{ii} = \sum_{j=1}^n h_{ij} h_{ji} = \sum_{j=1}^n h_{ij}^2.$$

Este último resultado conduz a

$$h_{ii} = \sum_{j=1}^n h_{ij}^2 = h_{ii}^2 + \sum_{j=1, j \neq i}^n h_{ij}^2,$$

em que $\sum_{j=1, j \neq i}^n h_{ij}^2 > 0$.

Daí, tem-se que

$$h_{ii} - h_{ii}^2 > 0 \Rightarrow h_{ii}(1 - h_{ii}) > 0 \Rightarrow$$

$$1 - h_{ii} > 0 \Rightarrow h_{ii} < 1.$$

Logo, mostramos que $0 < h_{ii} < 1$.

Ainda, como o posto de uma matriz idempotente é igual a seu traço, segue que

$$\begin{aligned}r(H) &= \text{tr}(H) \\&= \sum_{i=1}^n h_{ii} \\&= \text{tr}\{X(X^\top X)^{-1}X^\top\} \\&= \text{tr}\{X^\top X(X^\top X)^{-1}\} \\&= \text{tr}(I_p) \\&= p.\end{aligned}$$

Assim, a média amostral das medidas de alavancagem é

$$\bar{h} = \frac{1}{n} \sum_{i=1}^n h_{ii} = \frac{p}{n} \Rightarrow \bar{h} = \frac{p}{n}.$$

Para o caso de regressão linear normal simples ($p = 2$), a medida de alavancagem da i -ésima observação fica expressa por

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2},$$

Observe que mensuramos o quanto o i -ésimo ajuste \hat{Y}_i é afetado pelo ponto i através do quão discrepante é o valor de x_i com relação a média amostral \bar{x} .

Sabendo que $0 < h_{ii} < 1$, concluímos que a medida h_{ii} assumirá um valor próximo a 1 se, para n fixo, $(x_i - \bar{x})^2$ for um valor alto, ou seja, se x_i é distante de \bar{x} .

Por esta razão que pontos de alavanca estão associados com discrepâncias nos valores das covariáveis.

Resumindo, sob regressão linear normal, a medida de alavancagem é calculada através de

$$h_{ii} = \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i.$$

Esta quantidade é um valor entre 0 e 1, assumindo valor próximo a 1 quando o i -ésimo ponto é de alavanca.

Em média, espera-se que h_{ii} seja próximo de p/n . Por isso, deve-se investigar os pontos que possuem alavancagem $h_{ii} > 2p/n$ (duas vezes a média amostral).

Para tamanhos amostrais grandes, investigar o efeito dos pontos que possuem alavancagem $h_{ii} > 3p/n$ (três vezes a média amostral).

Alavancagem para os MLGs

Vimos que o processo iterativo (na convergência) para β é dado por

$$\hat{\beta} = (X^\top \widehat{W} X)^{-1} X^\top \widehat{W} \hat{\mathbf{z}},$$

em que \mathbf{z} é o vetor das respostas modificadas definido por

$$\mathbf{z} = \boldsymbol{\eta} + \left\{ W^{-1/2} \right\} \left\{ V^{-1/2} \right\} (\mathbf{y} - \boldsymbol{\mu}).$$

Observe que $\hat{\beta}$ pode ser visto como o estimador de mínimos quadrados da regressão linear com vetor de respostas $\widehat{W}^{1/2} \hat{\mathbf{z}}$ e matriz do modelo $\widehat{W}^{1/2} X$.

Alavancagem para os MLGs

Neste caso, a matriz “chapéu” fica expressa por

$$\hat{H} = \hat{W}^{1/2} X (X^\top \hat{W} X)^{-1} X^\top \hat{W}^{1/2}.$$

A medida de alavancagem utilizada nos MLGs é obtida a partir da diagonal principal de \hat{H} , isto é, define-se

$$\hat{h}_{ij} = \hat{\omega}_i \mathbf{x}_i^\top (X^\top \hat{W} X)^{-1} \mathbf{x}_j.$$

Ainda, veja que

$$\hat{\eta} = X\hat{\beta} = X(X^\top \hat{W} X)^{-1} X^\top \hat{W}\hat{\mathbf{z}} = \hat{P}\hat{\mathbf{z}}, \quad (1)$$

com

$$\hat{P} = X(X^\top \hat{W} X)^{-1} X^\top \hat{W}.$$

As matrizes \hat{P} e \hat{H} não coincidem, entretanto possuem os mesmos elementos na diagonal principal dados por \hat{h}_{ij} .

Alavancagem para os MLGs

Devido a relação em (1), a medida \hat{h}_{ii} pode ser vista como a variação causada em $\hat{\eta}_i$ quando a resposta modificada \hat{z}_i é acrescida de uma contaminação infinitesimal, isto é,

$$\frac{\partial \hat{\eta}_i}{\partial \hat{z}_i} = \hat{h}_{ii} = \hat{\omega}_i \mathbf{x}_i^\top (X^\top \widehat{W} X)^{-1} \mathbf{x}_i.$$

Note que a matriz \widehat{P} é idempotente, e

$$\begin{aligned} r(\widehat{P}) &= \text{tr}\{X(X^\top \widehat{W} X)^{-1} X^\top \widehat{W}\} \\ &= \text{tr}\{X^\top \widehat{W} X (X^\top \widehat{W} X)^{-1}\} \\ &= \text{tr}\{I_p\} \\ &= p \end{aligned}$$

Alavancagem para os MLGs

Também,

$$\begin{aligned}\text{tr}\{X(X^\top \widehat{W}X)^{-1}X^\top \widehat{W}\} &= \text{tr}\{\widehat{W}^{1/2}X(X^\top \widehat{W}X)^{-1}X^\top \widehat{W}^{1/2}\} \\ &= \text{tr}(\widehat{H}).\end{aligned}$$

Segue que

$$\bar{\widehat{h}} = \frac{1}{n} \sum_{i=1}^n \widehat{h}_{ii} = \frac{p}{n} \Rightarrow \bar{\widehat{h}} = \frac{p}{n}.$$

Analogamente, mostra-se que $0 < \widehat{h}_{ii} < 1$.

Qual será o ponto de corte para \widehat{h}_{ii} ?

Alavancagem para os MLGs

A escolha do ponto de corte para a medida \hat{h}_{ij} é difícil, pois apesar de esperarmos que \hat{h}_{ij} seja próximo de p/n , esta medida depende de $\hat{\mu}_i$.

Portanto, uma forma de identificar possíveis pontos de alavanca é fazer o gráfico de \hat{h}_{ij} versus a média ajustada $\hat{\mu}_i$, com ponto de corte $2p/n$ ou $3p/n$.

É possível mostrar que se a ligação $g(\cdot)$ é canônica, então

$$\hat{h}_{ij} = \frac{\partial \hat{Y}_i}{\partial \hat{Y}_i}.$$

Se a ligação $g(\cdot)$ não é canônica, então

$$\hat{h}_{ij} \approx \frac{\partial \hat{Y}_i}{\partial \hat{Y}_i}.$$

Suponha ϕ conhecido. O logaritmo da função de verossimilhança perturbado é definido como sendo

$$\ell_{\delta}(\boldsymbol{\beta}) = \sum_{j=1}^n \delta_j \ell_j(\boldsymbol{\beta}),$$

em que $\ell_j(\boldsymbol{\beta})$ corresponde ao logaritmo da função de verossimilhança da j -ésima observação, e δ_j é um tipo de perturbação introduzida tal que $0 \leq \delta_j \leq 1$.

Quando $\delta_j = 1$, $\forall j$, não há perturbação no modelo, e quando $\delta_j = 0$, significa que a j -ésima observação foi excluída.

Uma medida que verifica a influência da i -ésima observação em $\hat{\beta}$ é o afastamento pela verossimilhança definido por

$$LD_i = 2\{\ell(\hat{\beta}) - \ell(\hat{\beta}_{(i)})\},$$

em que $\hat{\beta}_{(i)}$ é o EMV de β sem a i -ésima observação.

Observe que LD_i é definida considerando que o vetor de perturbação é $\delta = (1 \ 1 \ \dots \ 0 \ \dots \ 1)^\top \in \mathbb{R}^n$ com o valor zero ocupando a i -ésima entrada deste vetor.

A medida $LD_i \geq 0$ desde que $\ell(\hat{\beta}) \geq \ell(\hat{\beta}_{(i)})$ pois $\ell(\hat{\beta})$ é o máximo da função $\ell(\beta)$ para todo $\beta \in \mathbb{R}^p$.

Na maioria dos casos não será possível obter uma forma analítica para LD_i . Por esta razão, iremos aproximar esta medida.

Considere a expansão da função $\ell(\beta)$ em série de Taylor de 2a ordem em torno de $\beta = \hat{\beta}$:

$$\begin{aligned}\ell(\beta) &\approx \ell(\hat{\beta}) + (\beta - \hat{\beta})^\top \ell'(\hat{\beta}) + \frac{1}{2}(\beta - \hat{\beta})^\top \ell''(\hat{\beta})(\beta - \hat{\beta}) \\ &= \ell(\hat{\beta}) + (\beta - \hat{\beta})^\top U_\beta(\hat{\beta}) + \frac{1}{2}(\beta - \hat{\beta})^\top J_{\beta\beta}(\hat{\beta})(\beta - \hat{\beta}) \\ &= \ell(\hat{\beta}) + \frac{1}{2}(\beta - \hat{\beta})^\top J_{\beta\beta}(\hat{\beta})(\beta - \hat{\beta}),\end{aligned}$$

desde que $U_\beta(\hat{\beta}) = \mathbf{0}$.

Segue que

$$\begin{aligned}2\{\ell(\boldsymbol{\beta}) - \ell(\hat{\boldsymbol{\beta}})\} &\approx (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top J_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \Leftrightarrow \\2\{\ell(\hat{\boldsymbol{\beta}}) - \ell(\boldsymbol{\beta})\} &\approx (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \left\{ -J_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}) \right\} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).\end{aligned}$$

Para aproximar LD_i , substituiremos $-J_{\boldsymbol{\beta}\boldsymbol{\beta}}$ por $K_{\boldsymbol{\beta}\boldsymbol{\beta}}$ desde que $-J_{\boldsymbol{\beta}\boldsymbol{\beta}} \approx K_{\boldsymbol{\beta}\boldsymbol{\beta}}$. Também, substituiremos $\boldsymbol{\beta}$ por $\hat{\boldsymbol{\beta}}_{(i)}$. Assim, ficamos com

$$LD_i \approx (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^\top \left\{ K_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}) \right\} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}).$$

Para os MLGs tem-se que $K_{\boldsymbol{\beta}\boldsymbol{\beta}} = \phi X^\top W X$. Então, obtemos que

$$LD_i \approx \phi(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^\top (X^\top \widehat{W} X)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}).$$

Para calcular LD_i a partir desta expressão precisamos de $\hat{\boldsymbol{\beta}}_{(i)}$ para $i = 1, 2, \dots, n$.

Para o caso normal linear, temos

$$\begin{aligned}\hat{\beta}_{(i)} &= \{X_{(i)}^\top X_{(i)}\}^{-1} X_{(i)}^\top Y_{(i)} \\ &= \{X^\top X - \mathbf{x}_i \mathbf{x}_i^\top\}^{-1} \{X^\top Y - \mathbf{x}_i Y_i\},\end{aligned}$$

em que

- ▶ $X_{(i)}$ é a matriz do modelo de dimensão $(n-1) \times p$ sem a i -ésima observação;
- ▶ $Y_{(i)}$ é o vetor de respostas de dimensão $(n-1) \times 1$ sem a i -ésima observação;
- ▶ $\mathbf{x}_i = (x_{i1} \ x_{i2} \ \cdots \ x_{ip}) \in \mathbb{R}^p$ é o vetor com os p valores fixados das covariáveis para i -ésima observação;
- ▶ Y_i é a i -ésima resposta.

Medida de influência

Já vimos que $X^\top = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n]$. Daí, temos que

$$X^\top X = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n] \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix} = \mathbf{x}_1 \mathbf{x}_1^\top + \mathbf{x}_2 \mathbf{x}_2^\top + \cdots + \mathbf{x}_n \mathbf{x}_n^\top.$$

Assim, a matriz $X^\top X - \mathbf{x}_i \mathbf{x}_i^\top$ corresponde a matriz $X^\top X$ excluindo a i -ésima observação (i -ésima linha de X).

Ainda, veja que

$$X^\top \mathbf{Y} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n] \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{j=1}^n \mathbf{x}_j Y_j.$$

Então, o vetor $X^\top \mathbf{Y} - \mathbf{x}_i Y_i$ corresponde ao vetor $X^\top \mathbf{Y}$ excluindo a i -ésima observação (i -ésima linha de X e i -ésima entrada de \mathbf{Y}).

Resultado. Sejam $A_{p \times p}$ matriz não singular, \mathbf{u} e \mathbf{v} vetores p -dimensionais. É válido que

$$(A + \mathbf{u}\mathbf{v}^\top)^{-1} = A^{-1} - \frac{(A^{-1}\mathbf{u})(\mathbf{v}^\top A^{-1})}{1 + \mathbf{v}^\top A^{-1}\mathbf{u}}.$$

Considere $A = X^\top X$, $\mathbf{u} = -\mathbf{x}_i$ e $\mathbf{v} = \mathbf{x}_i$. Assim, temos que

$$\begin{aligned}\{X^\top X - \mathbf{x}_i \mathbf{x}_i^\top\}^{-1} &= (X^\top X)^{-1} - \frac{(X^\top X)^{-1} \cdot -\mathbf{x}_i \mathbf{x}_i^\top (X^\top X)^{-1}}{1 + \mathbf{x}_i^\top (X^\top X)^{-1} \cdot -\mathbf{x}_i} \\ &= (X^\top X)^{-1} + \frac{(X^\top X)^{-1} \mathbf{x}_i \mathbf{x}_i^\top (X^\top X)^{-1}}{1 - h_{ii}},\end{aligned}$$

em que $h_{ii} = \mathbf{x}_i^\top (X^\top X)^{-1} \mathbf{x}_i$ é a medida de alavancagem.

$$\begin{aligned}\hat{\beta}_{(i)} &= \{X^T X - \mathbf{x}_i \mathbf{x}_i^T\}^{-1} \{X^T \mathbf{Y} - \mathbf{x}_i Y_i\} \\&= \left\{ (X^T X)^{-1} + \frac{(X^T X)^{-1} \mathbf{x}_i \mathbf{x}_i^T (X^T X)^{-1}}{1 - h_{ii}} \right\} \{X^T \mathbf{Y} - \mathbf{x}_i Y_i\} \\&= (X^T X)^{-1} X^T \mathbf{Y} - (X^T X)^{-1} \mathbf{x}_i Y_i \\&\quad + \frac{(X^T X)^{-1} \mathbf{x}_i \mathbf{x}_i^T (X^T X)^{-1} X^T \mathbf{Y}}{1 - h_{ii}} - \frac{(X^T X)^{-1} \mathbf{x}_i \mathbf{x}_i^T (X^T X)^{-1} \mathbf{x}_i Y_i}{1 - h_{ii}} \\&= \hat{\beta} - (X^T X)^{-1} \mathbf{x}_i Y_i + \frac{(X^T X)^{-1} \mathbf{x}_i \mathbf{x}_i^T \hat{\beta}}{1 - h_{ii}} - \frac{(X^T X)^{-1} \mathbf{x}_i h_{ii} Y_i}{1 - h_{ii}} \\&= \hat{\beta} - \frac{(X^T X)^{-1} \mathbf{x}_i [(1 - h_{ii}) Y_i - \hat{Y}_i + h_{ii} Y_i]}{1 - h_{ii}} \\&= \hat{\beta} - \frac{\hat{\varepsilon}_i}{1 - h_{ii}} (X^T X)^{-1} \mathbf{x}_i.\end{aligned}$$

Substituindo $\hat{\beta}_{(i)}$ em LD_i , obtemos

$$\begin{aligned}LD_i &= \frac{1}{\sigma^2} (\hat{\beta} - \hat{\beta}_{(i)})^\top (X^\top X) (\hat{\beta} - \hat{\beta}_{(i)}) \\&= \frac{1}{\sigma^2} \left[\frac{\hat{\varepsilon}_i}{1 - h_{ii}} (X^\top X)^{-1} \mathbf{x}_i \right]^\top (X^\top X) \left[\frac{\hat{\varepsilon}_i}{1 - h_{ii}} (X^\top X)^{-1} \mathbf{x}_i \right] \\&= \frac{1}{\sigma^2} \left(\frac{\hat{\varepsilon}_i}{1 - h_{ii}} \right)^2 \mathbf{x}_i^\top (X^\top X)^{-1} (X^\top X) (X^\top X)^{-1} \mathbf{x}_i \\&= \frac{1}{\sigma^2} \left(\frac{\hat{\varepsilon}_i}{1 - h_{ii}} \right)^2 \mathbf{x}_i^\top (X^\top X)^{-1} \mathbf{x}_i \\&= \frac{1}{\sigma^2} \left(\frac{\hat{\varepsilon}_i}{\sqrt{1 - h_{ii}}} \right)^2 \frac{h_{ii}}{1 - h_{ii}}.\end{aligned}$$

Substituindo σ^2 por seu estimador $\hat{\sigma}^2$, ficamos com

$$\begin{aligned} LD_i &= \left(\frac{\hat{\varepsilon}_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}} \right)^2 \frac{h_{ii}}{1 - h_{ii}} \\ &= \frac{t_i^2 h_{ii}}{1 - h_{ii}}. \end{aligned}$$

Observe que LD_i assumirá um valor grande se a i -ésima observação for ponto de alavanca (h_{ii} alto) ou/e for ponto aberrante ($|t_i|$ alto).

A distância de Cook no modelo normal linear é definida por

$$D_i = \frac{t_i^2 h_{ii}}{1 - h_{ii}} \frac{1}{p}.$$

Uma definição análoga da distância de Cook é dada por

$$D_i = \frac{1}{p\hat{\sigma}^2} \sum_{j=1}^n \left(\hat{Y}_j - \hat{Y}_j^{(i)} \right)^2,$$

em que $\hat{Y}_j^{(i)}$ é o ajuste da resposta Y_j com base nas estimativas dos parâmetros sem a i -ésima observação.

O numerador da distância de Cook mede o efeito que a exclusão da i -ésima observação causa nos n ajustes da resposta.

Em geral, não conseguimos forma fechada para $\hat{\beta}_{(i)}$ sob os MLGs. Entretanto, pode-se utilizar uma aproximação para $\hat{\beta}_{(i)}$.

A ideia consiste em tomar a primeira iteração do processo iterativo escore de Fisher quando este é iniciado em $\hat{\beta}$. Isto faz sentido pois ao iniciar o processo iterativo em $\hat{\beta}$, o valor de $\hat{\beta}_{(i)}$ não estará muito distante, e portanto, o valor inicial (“chute”) será muito próximo do valor correto.

Mostra-se que

$$\hat{\beta}_{(i)}^1 = \hat{\beta} - \frac{\hat{r}_{pi} \sqrt{\hat{\omega}_i \phi^{-1}}}{(1 - \hat{h}_{ii})} (X^\top \hat{W} X)^{-1} \mathbf{x}_i.$$

Substituindo $\hat{\beta}_{(i)}$ por $\hat{\beta}_{(i)}^1$ em LD_i , obtemos

$$\begin{aligned}LD_i &\approx \phi(\hat{\beta} - \hat{\beta}_{(i)})^\top (X^\top \widehat{W} X) (\hat{\beta} - \hat{\beta}_{(i)}) \\&\approx \phi \left[\frac{\hat{r}_{pi} \sqrt{\hat{\omega}_i \phi^{-1}}}{(1 - \hat{h}_{ii})} (X^\top \widehat{W} X)^{-1} \mathbf{x}_i \right]^\top (X^\top \widehat{W} X) \left[\frac{\hat{r}_{pi} \sqrt{\hat{\omega}_i \phi^{-1}}}{(1 - \hat{h}_{ii})} (X^\top \widehat{W} X)^{-1} \mathbf{x}_i \right] \\&= \phi \left(\frac{\hat{r}_{pi} \sqrt{\hat{\omega}_i \phi^{-1}}}{(1 - \hat{h}_{ii})} \right)^2 \mathbf{x}_i^\top (X^\top \widehat{W} X)^{-1} (X^\top \widehat{W} X) (X^\top \widehat{W} X)^{-1} \mathbf{x}_i \\&= \left(\frac{\hat{r}_{pi}}{1 - \hat{h}_{ii}} \right)^2 \hat{\omega}_i \mathbf{x}_i^\top (X^\top \widehat{W} X)^{-1} \mathbf{x}_i \\&= \left(\frac{\hat{r}_{pi}}{\sqrt{1 - \hat{h}_{ii}}} \right)^2 \frac{\hat{h}_{ii}}{1 - \hat{h}_{ii}} \\&= \hat{t}_{Si}^2 \frac{\hat{h}_{ii}}{1 - \hat{h}_{ii}}.\end{aligned}$$

Medida de influência

- ▶ A medida LD_i assumirá um valor alto se o ponto for de alavanca e/ou aberrante;
- ▶ A medida LD_i sob os MLGs é obtida a partir de várias aproximações;
- ▶ Empiricamente, observa-se que a aproximação feita subestima o verdadeiro valor da medida LD_i ;
- ▶ Apesar da subestimação, a aproximação apresentada para LD_i é suficiente para destacar pontos influentes do ajuste de um MLG;
- ▶ Alguns autores utilizam a medida LD_i sob os MLGs como a medida de Cook;
- ▶ Outros autores definem a medida de Cook como sendo $D_i = LD_i / p$ (generalizando o caso normal linear).

Adequação da função de ligação

Na convergência do processo iterativo escore de Fisher

$$\hat{\beta} = (X^{\top} \widehat{W} X)^{-1} X^{\top} \widehat{W} \hat{\mathbf{z}},$$

com

$$\hat{\mathbf{z}} = \hat{\boldsymbol{\eta}} + \widehat{W}^{-1/2} \widehat{V}^{-1/2} (\mathbf{y} - \hat{\boldsymbol{\mu}})$$

Note que a quantidade $\widehat{W}^{-1/2} \widehat{V}^{-1/2} (\mathbf{y} - \hat{\boldsymbol{\mu}})$ é um vetor de “resíduos”.

Assim, se a ligação $g(\cdot)$ estiver adequada espera-se uma relação linear entre $\hat{\mathbf{z}}$ e $\hat{\boldsymbol{\eta}} = X \hat{\boldsymbol{\beta}}$.

Portanto, pode ser construído o gráfico de dispersão de \hat{z}_i contra $\hat{\eta}_i$ para avaliar a adequação da função de ligação.