

## Análise de diagnóstico e medidas corretivas (Análise de resíduos)

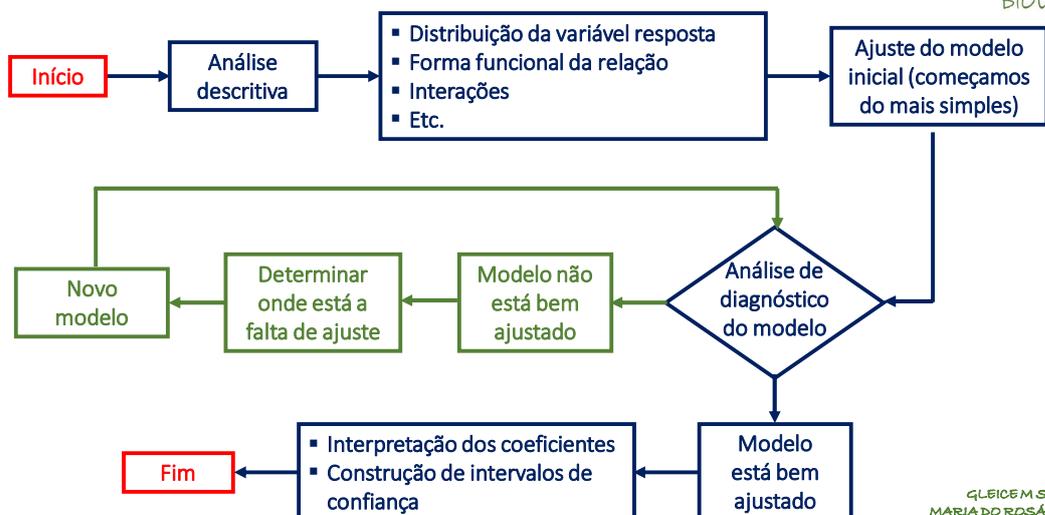


Na prática, durante um processo de modelagem, procedemos do seguinte modo:

- 1) Escolhemos o do modelo a ser ajustado
- 2) Examinamos se as suposições do modelo escolhido estão satisfeitas
- 3) Se não, consideramos o uso de transformações nos dados ou a escolha de outro modelo
- 4) Ajustamos o modelo
- 5) Avaliamos se o modelo está bem ajustado
- 6) Se não, voltamos ao passo 1.

GLEICEM S CONCEIÇÃO  
MARIA DO ROSÁRIO D D LATORRE  
FSP - USP

## Processo de Modelagem



GLEICEM S CONCEIÇÃO  
MARIA DO ROSÁRIO D D LATORRE  
FSP - USP

## Análise de diagnóstico e medidas corretivas



Para avaliar se as suposições do modelo escolhido estão satisfeitas:

### Antes de ajustar o modelo

- ✓ Gráficos
- ✓ Testes de normalidade
- ✓ Teste de homogeneidade de variâncias

### Depois de ajustar o modelo

- ✓ Análise de Resíduos

GLEICEM S CONCEIÇÃO  
MARIA DO ROSÁRIO D D LATORRE  
FSP - USP

## Resíduo

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$\hat{Y}_i = b_0 + b_1 X_i$$

O resíduo  $e_i$  é definido como a diferença entre o valor observado  $Y_i$  e o correspondente valor ajustado  $\hat{Y}_i$ .

$$e_i = Y_i - \hat{Y}_i$$

O resíduo  $e_i$  pode ser visto como o erro observado, em contrapartida ao verdadeiro erro  $\varepsilon_i$ , que é desconhecido e não observável.

O resíduos são extremamente úteis para avaliar se um determinado modelo é apropriado aos dados.



GLEICEM S CONCEIÇÃO  
MARIA DO ROSÁRIO D D LATORRE  
FSP - USP

## Algumas propriedades do resíduo



$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$\hat{Y}_i = b_0 + b_1 X_i$$

$$e_i = Y_i - \hat{Y}_i$$

Segundo o modelo de regressão, os erros  $\varepsilon_i$  são variáveis aleatórias independentes e identicamente distribuídas com distribuição  $N(0, \sigma^2)$ .

Se o modelo ajustado for realmente apropriado para os dados, os resíduos observados  $e_i$  devem refletir as propriedades assumidas para os  $\varepsilon_i$ , ou seja, devem ter distribuição normal com média igual a zero e variância constante.

GLEICEM S CONCEIÇÃO  
MARIA DO ROSÁRIO D D LATORRE  
FSP - USP

## Algumas propriedades do resíduo



- ✓ A média dos resíduos  $e_i$  é igual a zero.
- ✓ A variância dos resíduos  $e_i$  é dada pelo QMR.
- ✓ Os resíduos  $e_i$  não são independentes porque eles envolvem os valores ajustados  $\hat{Y}_i$ , que são baseados na mesma função de regressão ajustada.
- ✓ Se o tamanho da amostra é grande em comparação com o número de variáveis explicativas, o efeito da dependência dos resíduos  $e_i$  é insignificante e pode ser ignorado.

GLEICEM S CONCEIÇÃO  
MARIA DO ROSÁRIO D D LATORRE  
FSP - USP

## Tipos de Resíduos

### 1. Semistudentized residuals

$$e_i^* = \frac{e_i - \bar{e}}{\sqrt{QMR}} = \frac{e_i}{\sqrt{QMR}}$$

### 2. Studentized residuals

### 3. Deleted residuals

### 4. Studentized deleted residuals



GLEICEM S CONCEIÇÃO  
MARIA DO ROSÁRIO D D LATORRE  
FSP - USP

## Possíveis desvios das suposições do modelo

1. A função de regressão não é linear
2. Os erros  $\varepsilon_i$  não tem variância constante
3. Os erros  $\varepsilon_i$  não são independentes
4. O modelo ajusta bem mas há alguns *outliers*
5. Os erros  $\varepsilon_i$  não tem distribuição Normal
6. Uma ou várias variáveis explicativas importantes foram omitidas do modelo



GLEICEM S CONCEIÇÃO  
MARIA DO ROSÁRIO D D LATORRE  
FSP - USP

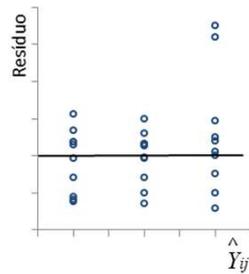




## Desvios das suposições do modelo



### 4. O modelo se ajusta, mas existem outliers



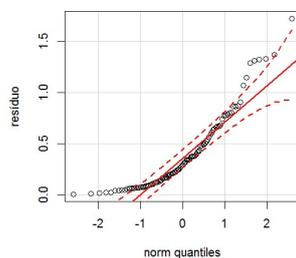
- ✓ Deletar a observação somente se houve erro de medida, problema no equipamento, etc.
- ✓ Avaliar o quanto o outlier afeta as estimativas do modelo
- ✓ Métodos de estimação mais robustos

GLEICEM S CONCEIÇÃO  
MARIA DO ROSÁRIO D D LATORRE  
FSP - USP

## Desvios das suposições do modelo



### 5. Os erros $\varepsilon_i$ não têm distribuição Normal



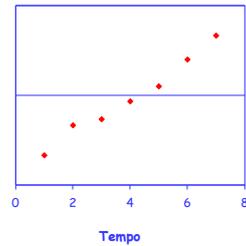
- ✓ Normalmente ocorre junto com a heterocedasticidade
- ✓ Transformações
- ✓ MLG permitem outras distribuições para a variável resposta

GLEICEM S CONCEIÇÃO  
MARIA DO ROSÁRIO D D LATORRE  
FSP - USP

## Desvios das suposições do modelo



5. Uma ou mais variáveis explicativas foram omitidas do modelo



- ✓ Verificar se os resíduos se adequam quando introduzimos uma nova variável

GLEICEM S CONCEIÇÃO  
MARIA DO ROSÁRIO D LATORRE  
FSP - USP