TESTE DE FALTA DE AJUSTE

Aplicação em Curvas de Sítio

Luan Fiorentin

UFPR - DEST

2019-06-26

Introdução

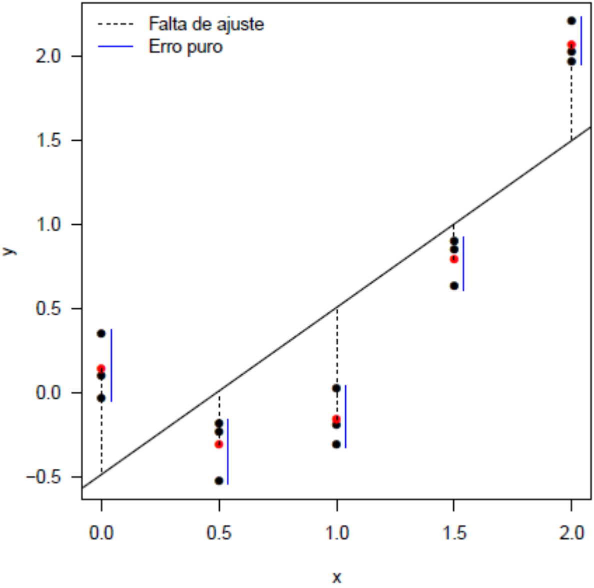
- O sítio é uma área geográfica com características homogêneas.
- Qualidade do sítio pode ser definida como a capacidade do local em produzir madeira.
- A classificação de sítios florestais por variáveis dendrométricas é a abordagem mais comum:
    - Permite expressar a capacidade de produção da espécie em função da própria dimensão da árvore.
    - A altura dominante é a variável mais utilizada.
- Curva de sítio é a curva da altura dominante em função da idade. Então, a curva de sítio é um modelo de crescimento.
- Índice de sítio é a altura dominante na idade de referência.

- Há dois tipos de curvas de sítio:
  - *Anamórficas*: supõem que o rítmo de crescimento da altura dominante das árvores é igual independente do sítio, exceto pelo valor absoluto em cada idade.
    Pontos de inflexão ocorrem na mesma idade.
  - *Polimórficas*: supõem que o rítmo de crescimento da altura dominante varia em função do sítio e da idade.
- Se pressupor que os índices de sítio são dependentes apenas das alturas dominantes das árvores e não da idade dos povoamentos florestais, tem-se curvas de sítio anamórficas.
- Nestes casos, espera-se que a altura dominante e os índices de sítio apresentem uma relação linear.
- Caso não há relação linear, temos evidências em favor de curvas polimórficas.

Teste de Falta de Ajuste

- O **Teste de Falta de Ajuste** (*Lack of Fit - LOF*) permite testar formalmente a adequação de um modelo de regresão linear.
- Assume-se que os pressupostos de normalidade, variância constante e independência são satisfeitos.
- A suposição do teste LOF é a de relação linear entre as variáveis.
- O teste é baseado na decomposição da variância dos resíduos em dois componentes:
  - Falta de Ajuste (FA).
  - Erro Puro (EP).

Figura 1: Ilustração da falta de ajuste do modelo de regressão linear



- O teste da falta de ajuste requer replicações independentes de y para ao menos um valor de x.
- A partir das réplicas de y em diferentes valores de x, tem-se condições de obter uma estimativa da variância do erro (σ^2) independente do modelo de regressão ajustado.

- Seja y_{ij} a j-ésima observação da variável resposta para um particular valor x_i, $i = 1, 2, ..., m$; $j = 1, 2, ..., n$; $n = \sum_{i=1}^{m} n_i$. Então,

$$\text{Resíduos} = \text{Erro Puro} + \text{Falta de Ajuste}$$

$$r_i = (y_{ij} - \hat{y}_i) = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \hat{y}_i),$$

em que $\bar{y}_i$ é a média das n_i observações tomadas em x_i.

- Tomando o quadrado de cada componente, tem-se que

$$SQRes = SQEP + SQFA$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{m}\sum_{j=1}^{n_i}(\bar{y}_i - \hat{y}_i)^2.$$

- Sob a suposição de variância constante, $SQEP$ é uma medida de dispersão dos erros independente do modelo, pois é calculada com base nas variações de y_i para cada valor de x_i.
- Cada valor de x_i contribui com $n_i - 1$ graus de liberdade para o erro puro e $(n-1)-(n-m)=(m-2)$ graus de liberdade para FA.
- Os resultados do teste podem ser apresentados em quadro de análise de variância.

Figura 2: Quadro de análise de variância para o teste de falta de ajuste

Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrados médios	F
Regressão	1	$\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2$	$QM_{Reg} = \frac{SQ_{Reg}}{1}$	$F = \frac{QM_{Reg}}{QM_{Res}}$
Resíduos	n-2	$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$QM_{Res} = \frac{SQ_{Res}}{n-2}$	
Falta de ajuste	m-2	$\sum_{i=1}^{m} n_i(\bar{y}_i - \hat{y}_i)^2$	$QM_{FA} = \frac{SQ_{FA}}{m-2}$	$F = \frac{QM_{FA}}{QM_{EP}}$
Erro puro	n-m	$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$	$QM_{EP} = \frac{SQ_{EP}}{n-m}$	
Total	n-1	$\sum_{i=1}^{n}(y_i - \bar{y})^2$		

- Se a função de regressão verdadeira de fato é linear, então tanto $QMEP$ quanto $QMFA$ são estimadores não viciados de σ^2.
- Caso a real função de regressão seja não linear, então $E[QMFA] > \sigma^2$.
- Sob a hipótese nula de que não há falta de ajuste (a função de regressão verdadeira é linear), então:

$$F_0 = \frac{SQFA/(m-2)}{SQEP/(n-m)} = \frac{QMFA}{QMEP}$$

tem distribuição F-Snedecor com graus de liberdade $m-2$ e $n-m$.

- A hipótese nula de que não há falta de ajuste (ou seja, a regressão é de fato linear) deverá ser rejeitada, ao nível de significância α, se $F_0 > F_{m-2;n-m;\alpha}$.
- O nível descritivo (p-valor) do teste deve ser calculado por $P(X > F_0)$, sendo $X \sim F_{m-2;n-m}$.
- Caso não há réplicas de y_i para testar a falta de ajuste, é possível agrupar indivíduos com valores próximos de x e proceder a análise.