

Notas: Regressão Stepwise no MINITAB

Mônica Barros

O procedimento stepwise permite selecionar variáveis a partir de um conjunto inicial de variáveis explicativas. A escolha de variáveis se baseia num procedimento heurístico que, apesar de intuitivamente razoável, não garante que a regressão encontrada possui o maior R^2 , nem que o modelo encontrado é o melhor, do ponto de vista prático. Entretanto, o procedimento é útil nos estágios iniciais de análise, especialmente quando existe um número muito grande de possíveis variáveis explicativas.

A escolha das variáveis é feita a partir da estatística F de cada variável, onde F é o quadrado do " t-ratio " da variável.

Características do procedimento

É necessário especificar 2 valores positivos, FENTER e REMOVE. Uma variável será incluída no modelo se a estatística F correspondente exceder FENTER, obedecendo ao procedimento mostrado a seguir. O valor REMOVE é usado para remover variáveis ao longo do procedimento, e uma variável será excluída se sua estatística F for menor que REMOVE.

Note que $FENTER \geq REMOVE$ sempre.

Passo 1 - ajustar todos os modelos com apenas 1 variável explicativa, e escolher aquela com o maior valor da estatística F, desde que $F > FENTER$.

Passo 2 - ajustar todos os modelos com 2 variáveis, sendo uma delas a variável incluída no passo 1. Incluir a variável com maior valor de F, desde que $F > FENTER$.

Passo 3 - verificar se alguma variável (exceto a incluída no passo 2) pode ser retirada do modelo. A variável com menor valor de F é excluída, desde que $F < REMOVE$.

Passo 4 - se nenhuma variável é removida, STEPWISE tenta inserir uma nova variável. O procedimento se repete, e variáveis são incluídas se $F > FENTER$, e removidas se $F < REMOVE$. O procedimento continua até que nenhuma variável seja incluída ou removida do modelo.

Sintaxe do comando STEPWISE no MINITAB

STEPWISE y na coluna C, variáveis nas colunas C, C , , C

Subcomandos:

FENTER = k (o default é $k = 4$)

REMOVE = k (o default é $k = 4$)

FORCE C, C , , C (incluir estas variáveis no primeiro passo e não excluí-las posteriormente).

ENTER C, C , , C (incluir estas variáveis no primeiro passo, permitindo uma posterior exclusão).

Casos Especiais

1) Forward Selection

As variáveis são inseridas da mesma maneira que no comando STEPWISE, mas nunca são removidas. O procedimento termina quando nenhuma variável tem $F > FENTER$. A implementação no MINITAB é:

STEPWISE y na coluna C, variáveis nas colunas C, C , , C
FREMOVE = 0 (nenhuma variável será removida)

2) Backward Elimination

A partir de todas as variáveis disponíveis, elimina-se uma variável a cada passo. Não se permite o reingresso de variáveis. O procedimento termina quando nenhuma variável pode mais ser removida, isto é, quando todas as variáveis no modelo apresentam $F > FREMOVE$.

A implementação no MINITAB é:

STEPWISE y na coluna C, variáveis nas colunas C, C , , C
ENTER C, C, , C (para forçar a inclusão de todas as variáveis no primeiro passo).
FENTER = 10000 (ou outro número grande, para impedir a inclusão de variáveis no procedimento STEPWISE)