

Passo a passo - CLASSIFICAÇÃO DE SÍTIO

Josiane Siqueira

Utilizando o modelo de Bailey e Clutter Polimórfico

- Antes de usar o modelo, tem que parear os dados:

Se for no Excel:

1) Na base de dados, onde se tem parcela, idade, e hdom, renomear para parcela, id1 e hd1. Copiar as três colunas do lado, nomeando parcela, id2, hd2. Agora, excluir a primeira linha de parcela, id2, hd2. Então ficaria assim:

parcela	id1	hd1	parcela	id2	hd2
1	20,76	12,55	1	32,33	18,26
1	32,33	18,26	1	45,57	24,2
1	45,57	24,2	1	70,66	30,6
1	70,66	30,6	2	20,76	11,32

2) Criar uma coluna chamada remover e colocar a fórmula: =SE(A2=D2;"fica";"sai"), em que A2 e D2 é a parcela (primeira) e a parcela (segunda). Se forem iguais, vai ficar escrito fica, se forem diferentes, vai ficar escrito sai;

3) Selecionar todos os dados, colocar para filtrar. Na coluna remover, filtrar só os “sai”. Selecione todos os dados que ficaram, que são todos os que tem “sai”. Excluí-los. Depois, volte no filtro de remover e deixe para filtrar só os “fica”. Tirar o filtro. Seus dados estão pareados agora.

Se for no R:

1) Importe os dados a serem pareados e coloque as funções:

```
dados=read.csv2('sitio.csv');
```

```
library(cmrinvflor);
```

```
dados=parear_seqmed(dados); #pareando os dados
```

```
View(dados);
```

Pronto!

- Agora, ajustando o modelo de Bailey e Clutter Polimórfico no R:

O modelo é
$$Hd2 = \beta_0 \left(\frac{hd1}{\beta_0} \right)^{\left(\frac{id1}{id2} \right)^{\beta_2}}$$

A hd2 está em função da hd1, id1 e id2, por isso pareia os dados;

2) Importar os dados, colocar o modelo e ajustar. O modelo fica assim:

```
bc_poli='hd2~b0*((hd1/b0)^((id1/id2)^b2))';
```

```
ajbcp=nls(bc_poli,dados,start=list(b0=40,b2=0.5));#ajuste do modelo não  
linear
```

```
summary(ajbcp); #devolve os valores dos betas, erros, coeficiente de  
determinação, se é significativo, etc. No caso, ficou assim:
```

Formula: $hd2 \sim b0 * ((hd1/b0)^{((id1/id2)^{b2})})$

Parameters:

	Estimate	Std. Error	t value	Pr(> t)
b0	46.35225	2.37292	19.53	<2e-16 ***
b2	0.82359	0.04794	17.18	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.266 on 549 degrees of freedom

Number of iterations to convergence: 5

Achieved convergence tolerance: 7.578e-07

3) Se os erros são tendenciosos, nem adianta fazer curvas anamórficas ou polimórficas.

Para isso, tem que ver os gráficos de dispersão de erros. O script fica assim:

```
x11();
```

```
plot(predict(ajbcp),residuals(ajbcp),pch='*',col='green',xlab='hdest  
(m)',ylab='resíduos (m)',ylim=c(-5,5));
```

```
abline(h=0);
```

```

sumario=summary(ajbcp);#para conseguir pegar o erro padrão

dados$respad=residuals(ajbcp)/sumario$sigma;#para fazer gráfico de resíduos
padronizados

#gráfico de resíduos padronizados (segundo escala de t)

x11();

plot(predict(ajbcp),dados$respad,pch='*',col='red',xlab='hdest
(m)',ylab='resíduos padronizados',ylim=c(-5,5));

abline(h=0);

tvalor=qt(0.995,nrow(dados)-1);#valor de t para 1%

#qt = inverso da distribuição de t

abline(h=tvalor,lty=2); #o lty = 2 é para deixar tracejado.

abline(h=-tvalor,lty=2);

```

4) Se quiser tirar os outliers, tem que fazer um procedimento. Primeiro, o gráfico dos resíduos padronizados deve estar aberto. Aí cria-se um objeto chamado ii que vai receber o comando identify. Com isso, você pode olhar no gráfico os pontos que estão além da linha tracejada, clicar neles (vai aparecer os números em cima de cada ponto), e ao final, clicar em parar, no canto esquerdo, e parar localizador. Volte para o script e coloca o restante dos comandos. No script, ficaria assim:

```

ii=identify(predict(ajbcp),dados$respad);#depois disso, vai no gráfico e
marque os pontos outliers, etc., como explicado.

ii;#mostra os números (outliers) marcados

dados=dados[-ii,];#dados recebe todo o arquivo dados, exceto as linhas ii [-ii
(menos ii) linha,coluna não remove]

#aí tem que rodar tudo de novo, desde:

ajbcp<-nls(bc_poli,dados,start=list(b0=40,b2=0.5));#ajuste do modelo não
linear

```

5) Guarde os betas gerados com o ajuste, que vão ser utilizados na aplicação do modelo.

- Agora, é bom abrir outro script para não bagunçar os dados a serem gerados. Será a aplicação de curvas de sítio pelo método da diferença algébrica, utilizando a biblioteca cmrinvlflor:

1) Acionar a biblioteca, importar os dados, renomear para idamostra, idade1, hdom1, criar um data.frame que guarde os parâmetros (betas) que foram gerados no ajuste não linear, colocar a idade de referência (Iref) = 72 (meses);

2) Escreve o modelo: `modelo='b0*((hdom1/b0)^((idade1/idade2)^b2))'`;

`#escreve hdom1 porque o modelo faz uma previsão de hdom2, etc.;`

3) Digite `x11()` para abrir um gráfico. Depois, coloque:

`cls=class_sitio_dif_alg(dados,modelo,parms,iref,graf_curvas = T);#segundo a sequência do comando`

Isso é um comando criado na biblioteca. Ao dar `ctrl + enter`, olhe lá no console. Apareceu a pergunta:

Os limites, inferior e superior, das curvas na idade de referência são respectivamente:

22.06 e 32.71.

Deseja alterar estes valores? `Afirmativo(T)` ou `Negativo(F)`

Responda `T` no console e dê `enter`. Vai aparecer outra pergunta e você responde conforme desejar que seja seu limite. No caso, ficou assim:

Informe os novos limites, inferior e superior respectivamente, das curvas na idade de referência.

1: 22

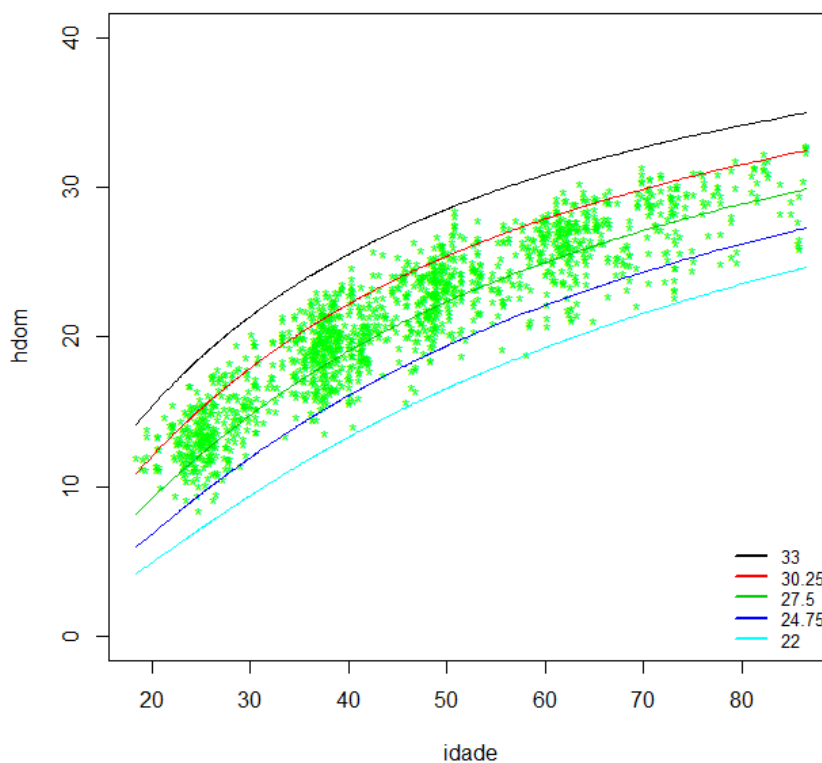
2: 33

Outra pergunta aparece ao dar `enter`. Esta é sobre o número de classes, que é arbitrário, mas não se costuma usar mais que 4 classes:

Informe o número de classes desejado:

1: 4

Pronto! Dê uma olhada no gráfico. As curvas apareceram! Deve ficar mais ou menos assim:



4) Agora volte para a parte de cima e coloque: `cls$estabilidade`. Isso vai mostrar no console:

```
migracao n_ocorrencia percmudanca
0         282         58.39
1         187         38.72
2          14          2.90
```

Isso mostra a estabilidade das suas parcelas nesse modelo quanto à classificação de sítio. Não é desejável que as parcelas migrem muito de classe, isso pode acontecer caso a parcela tenha sido medida errada. Mas elas estão migrando apenas no modelo que foi utilizado, pode ser que em outro elas sejam mais estáveis. Migrar significa mudar de sítio, o que não pode acontecer, já que o sítio é para ser constante.

Olhando os resultados, vemos que 282 parcelas não migraram nenhuma vez, 187 migraram 1 vez, e 14 migraram 2 vezes. O `percmudanca` quer dizer a porcentagem do quanto essas parcelas migraram.

5) Digite `cls$classes` para saber os limites das classes que foram geradas. Vai aparecer algo do tipo no console:

```
limite hdiref_72
lsc11    33.00
```

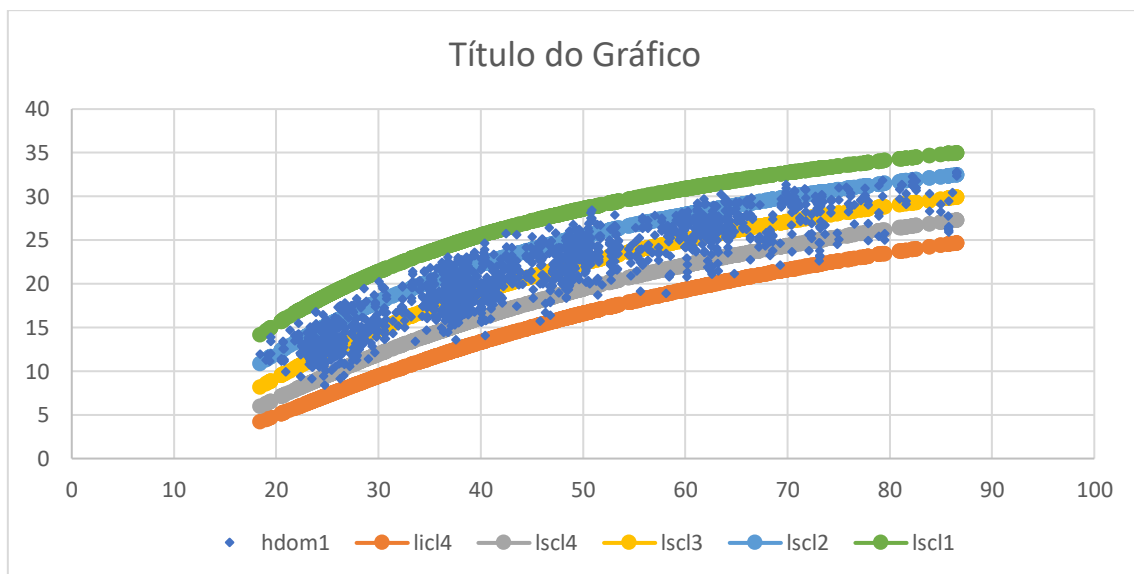
lsc12	30.25
lsc13	27.50
lsc14	24.75
lic14	22.00

6) Agora visualize como ficou a classificação de sítio, e salve em Excel, pois a parte de reclassificação só dá para fazer “à mão”:

```
View(cls$amostras);
write.csv2(cls$amostras,'reclassificacao.csv',row.names = F);#reclassificar
não dá pra fazer no R, aí arquiva em excel
shell.exec('reclassificacao.csv');#abre o arquivo no excel
```

- Agora é preciso reclassificar. Abra o arquivo que foi criado a partir do R no Excel, e você vai ver as colunas idamostra, idade1, hdom1, lic14, lsc14, lsc13, lsc12, lsc11, is, nmudou e nis.

- 1) Selecione todos os seus dados e personalize a classificação por idade1;
- 2) Selecione seus dados desde idade1 a lsc11. Insira um gráfico de dispersão com linhas suaves e marcadores. Formate o gráfico como preferir, desde que hdom fique como marcadores apenas, por cima de todas as linhas, para ser fácil de visualizar. As linhas são os limites das classes. Ficar assim, mais ou menos:



3) Coloque um filtro e selecione uma parcela de cada vez. Para cada parcela, olhe o gráfico, veja em que intervalo de limites a hdom está, e dê a nota/classificação na célula nis (mude os valores agora conforme o que você vê);

4) Crie uma coluna chamada s, que será o sítio final, da reclassificação. Como você já deu a nota para cada parcela, vai atribuir então a altura em que ela está encaixada, de acordo com os limites. Lembrando dos limites que o R te mostrou:

limite	hdiréf_72
lsc11	33.00
lsc12	30.25
lsc13	27.50
lsc14	24.75
lic14	22.00

O que importa para o sítio final é o valor central entre um limite e outro. Então, ficaria $(24,75+22)/2 = 23,375$, entre os limites inferior da classe 4 e superior da classe 4, $(27,50+24,75)/2 = 26,125$, entre limites superior da classe 4 e superior da classe 3, $(30,25+27,50)/2 = 28,875$, entre limites superior da classe 3 e superior da classe 2, e $(33+30,25)/2 = 31,625$, entre limites superior da classe 2 e superior da classe 1.

Isso quer dizer que se a nota da parcela for 4, seu sítio será 23,375, se for 3, será 26,125, se for 2, será 28,875, se for 1, seu sítio será 31,625. Então, as menores notas são dos melhores sítios.

Para fazer isso no Excel, coloque a fórmula:

=SE(K2=4;23,375;SE(K2=3;26,125;SE(K2=2;28,875;31,625)))

Pronto! A parcela inteira deve ter um mesmo valor de sítio, visto que foi realizada uma reclassificação a partir dos gráficos.

#FIM ----- ☺