

Script para estimação de capacidade produtiva no R

Sollano Rabelo Braga

Ana Carolina Araujo

Outubro, 2016

Contents

1) Carregar pacotes e dados	2
2) Verificar tendência dos dados	2
3) Estimar altura dominante	3
3.1) Ajuste do modelo linear	3
3.2) Estimação de HD	4
3.3) Análises da qualidade do ajuste	4
3.3) Análise de erro do ajuste	5
4) Gerar curva-guia - metodo dos fatores	6
5) Exportar graficos	12

1) Carregar pacotes e dados

Primeiro carrega-se os pacotes que serão utilizados:

```
library(tidyverse)
library(readxl)
```

Carrega-se os dados do excel utilizando readxl

```
dados <- read_excel("dados.xlsx")
```

Visualiza-se os dados com head e tail:

```
head(dados)
```

```
## # A tibble: 6 × 4
##   talhao parcela idade      HD
##   <dbl>   <dbl> <dbl>   <dbl>
## 1      1      2    30 12.29841
## 2      1      2    42 16.32000
## 3      1      2    54 17.58000
## 4      1      2    66 18.94000
## 5      1      2    78 21.04000
## 6      1      2    90 20.48000
```

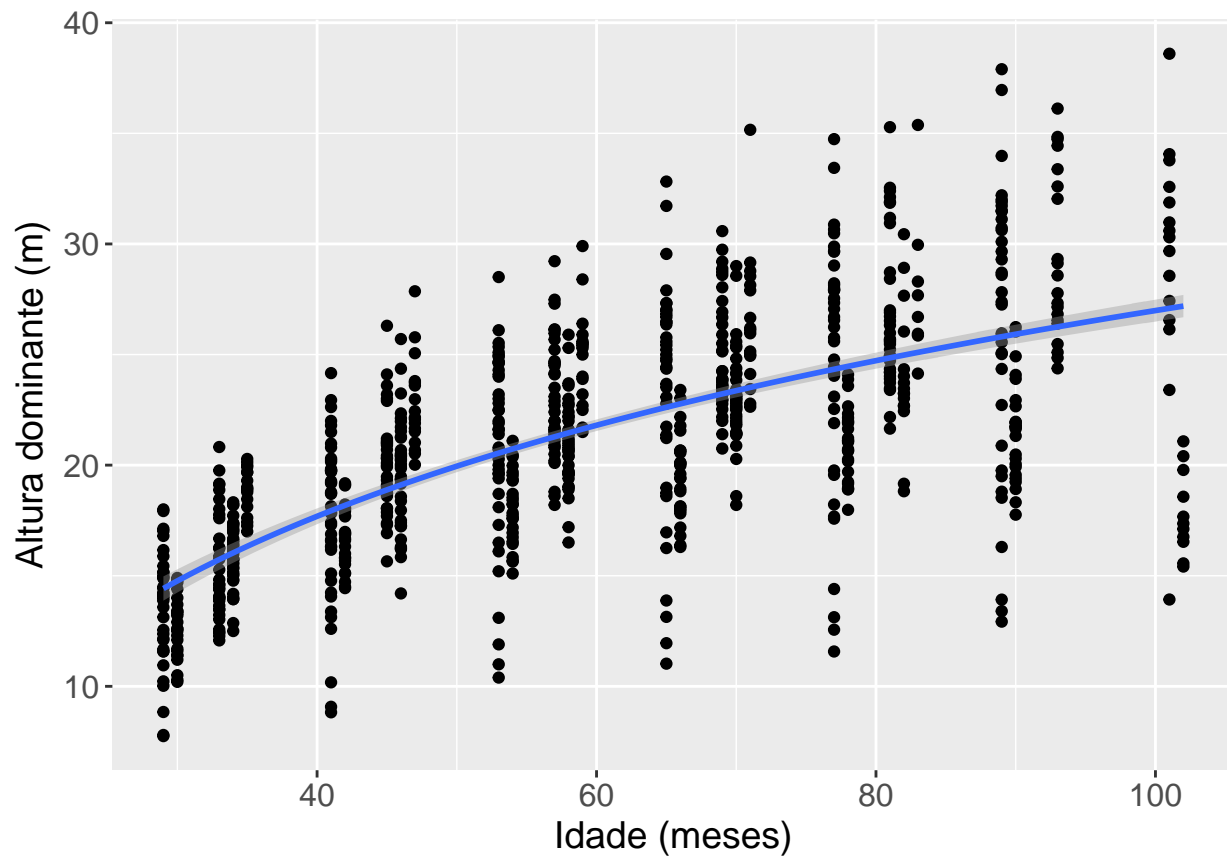
```
tail(dados)
```

```
## # A tibble: 6 × 4
##   talhao parcela idade      HD
##   <dbl>   <dbl> <dbl>   <dbl>
## 1     36     107    81 30.9400
## 2     36     107    93 33.3800
## 3     37     106    33 19.7600
## 4     37     106    45 23.0037
## 5     37     106    57 26.1200
## 6     37     106    69 28.0400
```

2) Verificar tendência dos dados

Verifica-se a tendência de altura, gerando um gráfico de idade em relação a altura dominante.

```
ggplot(dados, aes(idade, HD) ) +
  geom_point() +
  stat_smooth(method="lm", formula=y~log(x) ) +
  labs(x      = "Idade (meses)",
       y      = "Altura dominante (m)" ) +
  theme(axis.title = element_text(size = 14),
        axis.text  = element_text(size = 12) )
```



3) Estimar altura dominante

3.1) Ajuste do modelo linear

Primeiro, cria-se variáveis necessárias para regressão:

```
dados$LN_HD <- log(dados$HD)
dados$INV_I <- 1/dados$idade
```

Em seguida, ajusta-se o modelo, utilizando a função lm:

```
reg <- lm(LN_HD ~ INV_I, dados)
```

resumo sobre a regressão:

```
summary(reg)
```

```
##
## Call:
## lm(formula = LN_HD ~ INV_I, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.72718 -0.09336  0.01491  0.12147  0.41286
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.51767    0.01813   194.0  <2e-16 ***
## INV_I       -26.30623    0.89777   -29.3  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1836 on 857 degrees of freedom
## Multiple R-squared:  0.5005, Adjusted R-squared:  0.4999
## F-statistic: 858.6 on 1 and 857 DF,  p-value: < 2.2e-16
```

```
summary(reg)[[9]]
```

```
## [1] 0.4998807
```

3.2) Estimação de HD

Primeiro, cria-se objetos que contém os coeficientes

```
b0 <- coef(reg)[[1]]
b1 <- coef(reg)[[2]]
b0
```

```
## [1] 3.517669
```

```
b1
```

```
## [1] -26.30623
```

Em seguida, estima-se HD:

```
dados$HD_EST <- exp(b0 + b1 * dados$INV_I)
```

Visualiza-se o resultado com head:

```
head(dados)
```

```
## # A tibble: 6 × 7
##   talhao parcela idade      HD    LN_HD      INV_I    HD_EST
##   <dbl>   <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1     1     2    30 12.29841 2.509470 0.03333333 14.02434
## 2     1     2    42 16.32000 2.792391 0.02380952 18.01726
## 3     1     2    54 17.58000 2.866762 0.01851852 20.70792
## 4     1     2    66 18.94000 2.941276 0.01515152 22.62577
## 5     1     2    78 21.04000 3.046425 0.01282051 24.05660
## 6     1     2    90 20.48000 3.019449 0.01111111 25.16307
```

Em seguida, remove-se as variáveis utilizadas na estimação:

```
dados[c("LN_HD", "INV_I")] <- NULL
```

3.3) Análises da qualidade do ajuste

Correlação

```
cor(dados$HD, dados$HD_EST)
```

```
## [1] 0.6880623
```

Bias

```
sum(dados$HD - dados$HD_EST)/nrow(dados)
```

```
## [1] 0.3419279
```

RQEM - raiz quadrada do erro médio (RMSE - root mean squared error)

```
1/mean(dados$HD) * ( sqrt( sum((dados$HD - dados$HD_EST)^2)/nrow(dados) ) ) * 100
```

```
## [1] 18.10005
```

Visualiza-se o resultado com head:

```
head(dados)
```

```
## # A tibble: 6 × 5
##   talhao parcela idade      HD  HD_EST
##   <dbl>   <dbl> <dbl>   <dbl>   <dbl>
## 1     1     2    30 12.29841 14.02434
## 2     1     2    42 16.32000 18.01726
## 3     1     2    54 17.58000 20.70792
## 4     1     2    66 18.94000 22.62577
## 5     1     2    78 21.04000 24.05660
## 6     1     2    90 20.48000 25.16307
```

3.3) Análise de erro do ajuste

Primeiro, calcula-se o erro em porcentagem:

```
dados$erro <- ((dados$HD_EST - dados$HD)/dados$HD) * 100
```

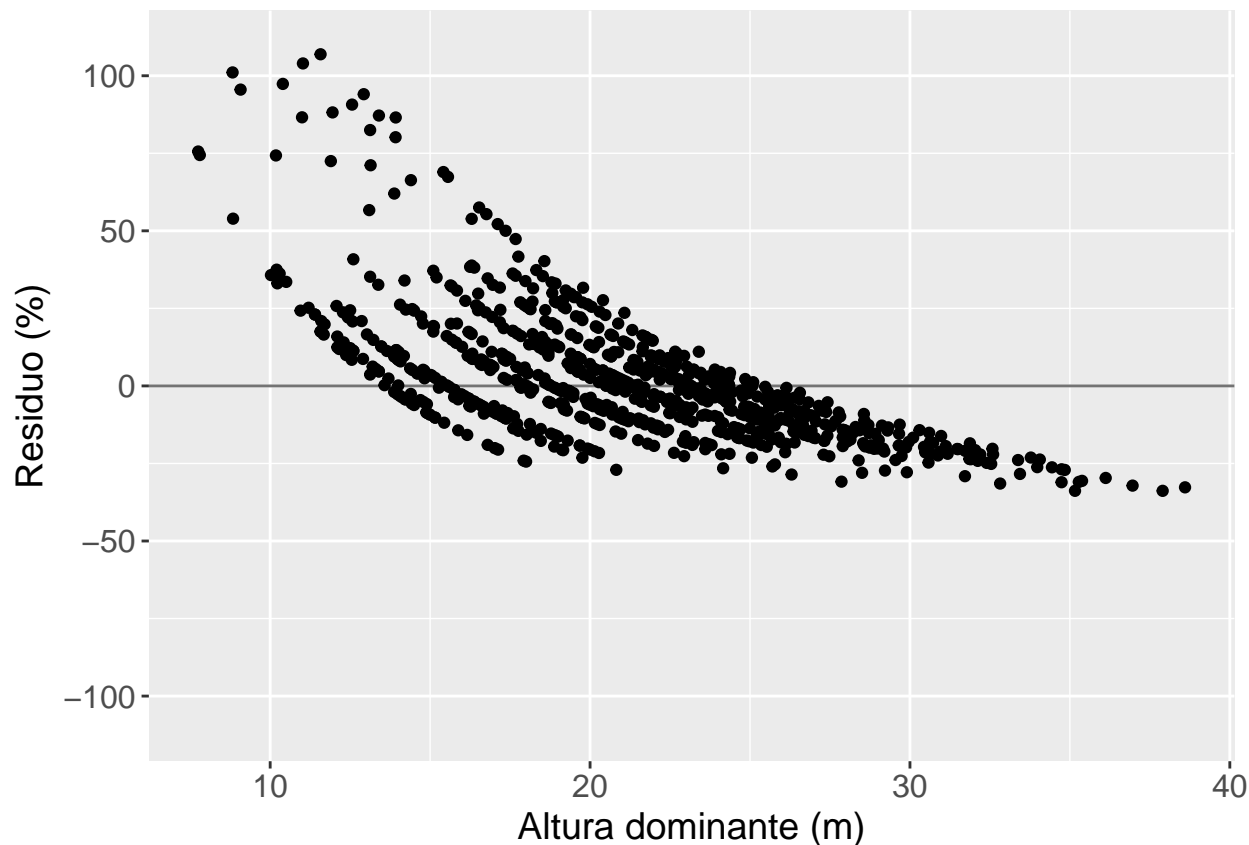
Visualiza-se o resultado com head:

```
head(dados)
```

```
## # A tibble: 6 × 6
##   talhao parcela idade      HD  HD_EST  erro
##   <dbl>   <dbl> <dbl>   <dbl>   <dbl>   <dbl>
## 1     1     2    30 12.29841 14.02434 14.03377
## 2     1     2    42 16.32000 18.01726 10.39985
## 3     1     2    54 17.58000 20.70792 17.79250
## 4     1     2    66 18.94000 22.62577 19.46024
## 5     1     2    78 21.04000 24.05660 14.33745
## 6     1     2    90 20.48000 25.16307 22.86653
```

Agora, gera-se o gráfico de resíduos:

```
ggplot(dados, aes(HD, erro)) +
  geom_hline(yintercept = 0, color = "gray45") +
  geom_point() +
  ylim(c(-110, 110)) +
  labs(x = "Altura dominante (m)",
       y = "Residuo (%)") +
  theme(axis.title = element_text(size = 14),
        axis.text = element_text(size = 12) )
```



4) Gerar curva-guia - metodo dos fatores

O primeiro passo é calcular a razão entre a altura observada e a altura estimada:

```
dados$fator <- dados$HD / dados$HD_EST
head(dados)
```

```
## # A tibble: 6 × 7
##   talhao parcela idade      HD  HD_EST      erro      fator
##   <dbl>   <dbl> <dbl>   <dbl>  <dbl>   <dbl>   <dbl>
## 1     1     2    30 12.29841 14.02434 14.03377 0.8769332
## 2     1     2    42 16.32000 18.01726 10.39985 0.9057984
## 3     1     2    54 17.58000 20.70792 17.79250 0.8489505
## 4     1     2    66 18.94000 22.62577 19.46024 0.8370986
## 5     1     2    78 21.04000 24.05660 14.33745 0.8746041
## 6     1     2    90 20.48000 25.16307 22.86653 0.8138913
```

Em seguida definimos a idade índice que será utilizada:

```
idade_indice <- 72
```

E então calcula-se a altura dominante na idade índice, para ser utilizada futuramente:

```
HD_EST_II <- exp( b0 + b1*( 1/ idade_indice ) )
```

Em seguida calcula-se os limites inferior e superior das classes de site; para isso multiplica-se os valores máximo e mínimo do fator pela altura dominante na idade índice:

```
lim_inf <- HD_EST_II * min( dados$fator )
lim_sup <- HD_EST_II * max( dados$fator )
```

```
lim_inf
```

```
## [1] 11.30362
```

```
lim_sup
```

```
## [1] 35.34519
```

O intervalo de classe pode ser calculado, com base no numero de classes:

```
nc <- 5
intervalo <- ceiling( (lim_sup - lim_inf ) / nc )
intervalo
```

```
## [1] 5
```

A seguir, serão calculados os limites inferior e superior para todas as classes; serão criados varios dataframes, um para cada classe, que serão unidos em seguida por rbind. os dataframes serão compostos por 3 colunas: nível, limites, e fator. Primeiro utiliza-se o limite inferior como base:

```
c1 <- data.frame(nivel = c("inf","sup"),

                limites = c(round(floor(lim_inf), -1) ,
                           round(floor(lim_inf), -1) + intervalo),

                fator = c(round(floor(lim_inf), -1) / HD_EST_II,
                         (round(floor(lim_inf), -1) + intervalo ) / HD_EST_II ) )

c1
```

```
##   nivel limites   fator
## 1   inf      10 0.4275349
## 2   sup      15 0.6413024
```

Agora o limite inferior da próxima classe será igual ao limite superior da classe anterior, portanto:

```
c2 <- data.frame(nivel = c("inf","sup"),

                limites = c(c1$limites[2],
                           c1$limites[2] + intervalo),

                fator = c(c1$limites[2] / HD_EST_II,
                         (c1$limites[2] + intervalo ) / HD_EST_II ) )

c2
```

```
##   nivel limites   fator
## 1   inf      15 0.6413024
## 2   sup      20 0.8550698
```

e assim sucessivamente, ate atingir-se o número de classes desejado:

```
c3 <- data.frame(nivel = c("inf","sup"),

                limites = c(c2$limites[2],
                           c2$limites[2] + intervalo),

                fator = c(c2$limites[2] / HD_EST_II,
```

```

(c2$limites[2] + intervalo ) / HD_EST_II ) )
c3

```

```

##  nivel limites      fator
## 1   inf        20 0.8550698
## 2   sup        25 1.0688373

```

```

c4 <- data.frame(nivel = c("inf","sup"),

                 limites = c(c3$limites[2],
                             c3$limites[2] + intervalo),

                 fator = c(c3$limites[2] / HD_EST_II,
                           (c3$limites[2] + intervalo ) / HD_EST_II ) )
c4

```

```

##  nivel limites      fator
## 1   inf        25 1.068837
## 2   sup        30 1.282605

```

```

c5 <- data.frame(nivel = c("inf","sup"),

                 limites = c(c4$limites[2],
                             c4$limites[2] + intervalo),

                 fator = c(c4$limites[2] / HD_EST_II,
                           (c4$limites[2] + intervalo ) / HD_EST_II ) )
c5

```

```

##  nivel limites      fator
## 1   inf        30 1.282605
## 2   sup        35 1.496372

```

Em seguida, utilizando rbind, cria-se a tabela de classes, unindo todas as classes em um unico objeto:

```

classes <- rbind(c1,c2,c3,c4,c5)
classes

```

```

##  nivel limites      fator
## 1   inf        10 0.4275349
## 2   sup        15 0.6413024
## 3   inf        15 0.6413024
## 4   sup        20 0.8550698
## 5   inf        20 0.8550698
## 6   sup        25 1.0688373
## 7   inf        25 1.0688373
## 8   sup        30 1.2826048
## 9   inf        30 1.2826048
## 10  sup        35 1.4963722

```

Uma forma mais pratica de realizar esta ação é utilizando um loop for.

Segue-se o mesmo raciocínio anteriormente, porem define-se que a operação será repetida por um numero de vezes, ou seja, um loop.

como será utilizado um loop for a seguir, cria-se uma lista que contém o número de entradas referente ao número de classes desejado:


```
nc <- 5
list <- vector("list", nc)
```

Cria-se uma lista vazia com 5 elementos:

```
list

## [[1]]
## NULL
##
## [[2]]
## NULL
##
## [[3]]
## NULL
##
## [[4]]
## NULL
##
## [[5]]
## NULL
```

Os primeiros limites são calculados separadamente do loop, e são salvos no primeiro elemento da lista: Isto é importante e é feito separadamente, pois os calculos feitos no loop irão se iniciar utilizando o primeiro elemento da lista.

```
list[[1]] <- data.frame(nivel = c("inf","sup"),

                        limites = c(round(floor(lim_inf), -1) ,
                                    round(floor(lim_inf), -1) + intervalo),

                        fator = c(round(floor(lim_inf), -1) / HD_EST_II,
                                  (round(floor(lim_inf), -1) + intervalo ) / HD_EST_II ) )
```

Agora aplica-se o loop. No loop for (para), se determina uma variável, chamada de “i” neste caso, que representa a posição onde será aplicado o loop. Entao neste caso o loop vai de 2 a 5, como pode-se ver abaixo. A operação será repetida i vezes, ate que se chegue ao numero final.

Anteriormente o mesmo processo foi repetido 5 vezes, mudando-se apenas a posição, pois primeiro foi feito para o segundo limite, depois para o terceiro, ate o quinto. Portanto pode-se aplicar este padrao aqui, substituindo estas posições que foram mudando por “i”, para que o loop faça o processo automaticamente.

Cada interção e salva em um elemento de uma lista. Portanto, a primeira operação, que anteriormente foi chamada de c1, e era uma matriz, agora é um elemento de uma lista, e pode ser chamado por:

```
list[[1]]

##   nivel limites   fator
## 1   inf      10 0.4275349
## 2   sup      15 0.6413024
```

Os próximos elementos serão salvos nas posições 2, 3, 4 e 5. Portanto, onde anteriormente foi utilizado c, utiliza-se list[[i]]. Como i representa a posição atual do loop, se o calculo for para a posição 2, i terá o valor 2. Substituindo abaixo, temos que: list terá o elemento 2 criado; e os elementos e os fatores serão criados utilizando list[[2 - 1]] ou seja, 1. Isto quando se determina os fatores e os limites sempre se utiliza a informação do limite anterior. O processo então é repetido para as demais classes.

```
for(i in 2:(nc)){
```

```
list[[i]] <- data.frame(nivel = c("inf","sup"),
                        limites = c(list[[i-1]] [[2]] [[2]],
                                   list[[i-1]] [[2]] [[2]] + intervalo),
                        fator = c(list[[i-1]] [[2]] [[2]] / HD_EST_II,
                                   (list[[i-1]] [[2]] [[2]]+ intervalo ) / HD_EST_II ) )
}
```

transforma-se a lista em matriz e em seguida em data frame:

```
classes2 <- do.call(rbind, list)
```

Os dois objetos criados são idênticos:

```
classes
```

```
##      nivel limites      fator
## 1      inf       10 0.4275349
## 2      sup       15 0.6413024
## 3      inf       15 0.6413024
## 4      sup       20 0.8550698
## 5      inf       20 0.8550698
## 6      sup       25 1.0688373
## 7      inf       25 1.0688373
## 8      sup       30 1.2826048
## 9      inf       30 1.2826048
## 10     sup       35 1.4963722
```

```
classes2
```

```
##      nivel limites      fator
## 1      inf       10 0.4275349
## 2      sup       15 0.6413024
## 3      inf       15 0.6413024
## 4      sup       20 0.8550698
## 5      inf       20 0.8550698
## 6      sup       25 1.0688373
## 7      inf       25 1.0688373
## 8      sup       30 1.2826048
## 9      inf       30 1.2826048
## 10     sup       35 1.4963722
```

O próximo passo é unir os dados originais de altura dominante e idade, com as classes criadas. Para uni-los mantendo a ordem dos limites, é necessário criar uma variável auxiliar que seja comum aos dois data frames:

```
dados$aux <- 1
classes$aux <- 1
```

em seguida, une-se os dados utilizando merge, com base na variável aux:

```
tab_curva <- merge( dados[c("HD", "HD_EST", "idade", "aux")] , classes , by = "aux")
```

Agora pode-se remover a variável auxiliar:

```
tab_curva$aux <- NULL
```

Visualiza-se o resultado com head:

```
head(tab_curva, 15)
```

```
##           HD   HD_EST idade nivel limites      fator
## 1  12.29841 14.02434   30   inf      10 0.4275349
## 2  12.29841 14.02434   30   sup      15 0.6413024
## 3  12.29841 14.02434   30   inf      15 0.6413024
## 4  12.29841 14.02434   30   sup      20 0.8550698
## 5  12.29841 14.02434   30   inf      20 0.8550698
## 6  12.29841 14.02434   30   sup      25 1.0688373
## 7  12.29841 14.02434   30   inf      25 1.0688373
## 8  12.29841 14.02434   30   sup      30 1.2826048
## 9  12.29841 14.02434   30   inf      30 1.2826048
## 10 12.29841 14.02434   30   sup      35 1.4963722
## 11 16.32000 18.01726   42   inf      10 0.4275349
## 12 16.32000 18.01726   42   sup      15 0.6413024
## 13 16.32000 18.01726   42   inf      15 0.6413024
## 14 16.32000 18.01726   42   sup      20 0.8550698
## 15 16.32000 18.01726   42   inf      20 0.8550698
```

Já que os limites serão utilizados para classificar os dados, é interessante transformá-los em fator:

```
# transforma-se os limites em fator, para que o gráfico possa ser plotado corretamente:
tab_curva$limites <- factor(tab_curva$limites)
class(tab_curva$limites)
```

```
## [1] "factor"
```

```
# Em seguida calcula-se do HD com base no fator calculado:
tab_curva$HD_CURVA <- tab_curva$HD_EST * tab_curva$fator
```

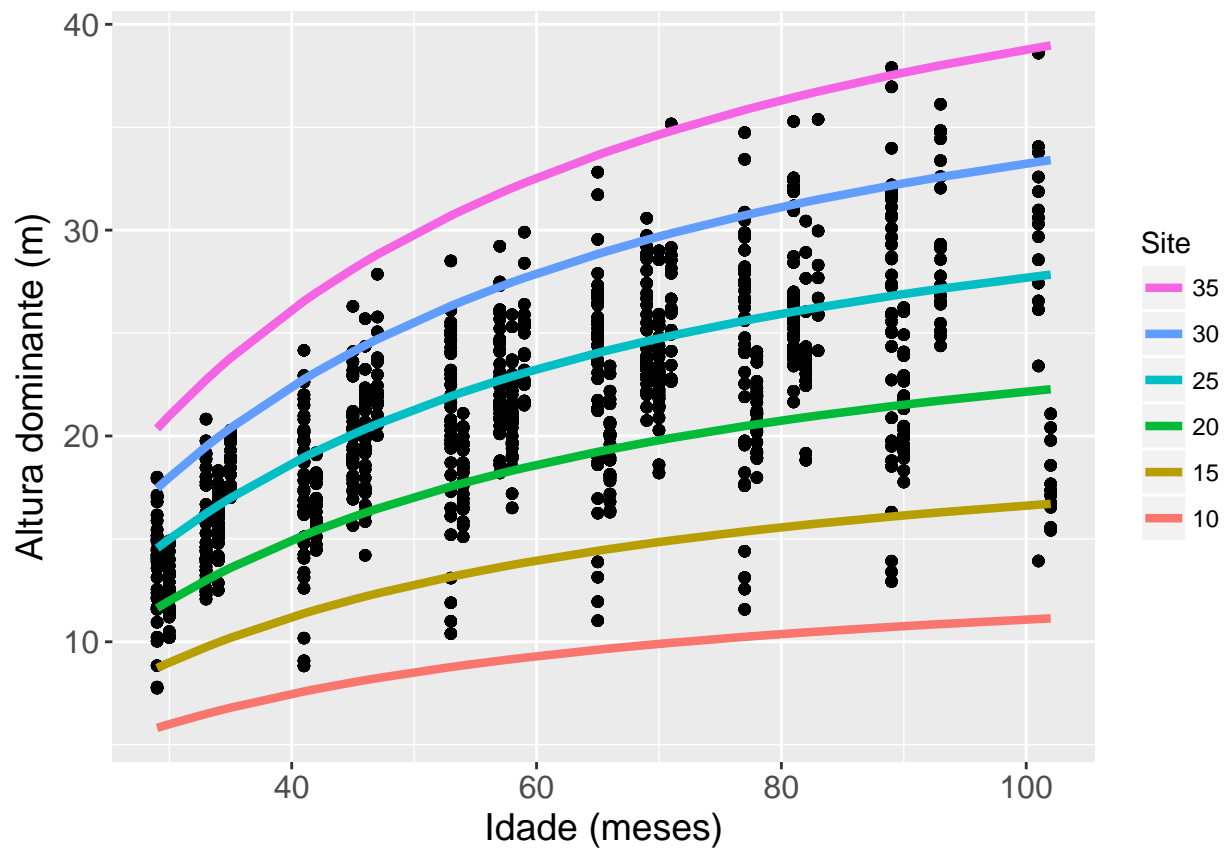
Visualiza-se o resultado com head:

```
head(tab_curva, 15)
```

```
##           HD   HD_EST idade nivel limites      fator  HD_CURVA
## 1  12.29841 14.02434   30   inf      10 0.4275349   5.995895
## 2  12.29841 14.02434   30   sup      15 0.6413024   8.993843
## 3  12.29841 14.02434   30   inf      15 0.6413024   8.993843
## 4  12.29841 14.02434   30   sup      20 0.8550698  11.991790
## 5  12.29841 14.02434   30   inf      20 0.8550698  11.991790
## 6  12.29841 14.02434   30   sup      25 1.0688373  14.989738
## 7  12.29841 14.02434   30   inf      25 1.0688373  14.989738
## 8  12.29841 14.02434   30   sup      30 1.2826048  17.987686
## 9  12.29841 14.02434   30   inf      30 1.2826048  17.987686
## 10 12.29841 14.02434   30   sup      35 1.4963722  20.985633
## 11 16.32000 18.01726   42   inf      10 0.4275349   7.703006
## 12 16.32000 18.01726   42   sup      15 0.6413024  11.554509
## 13 16.32000 18.01726   42   inf      15 0.6413024  11.554509
## 14 16.32000 18.01726   42   sup      20 0.8550698  15.406012
## 15 16.32000 18.01726   42   inf      20 0.8550698  15.406012
```

```
curvas <- ggplot(tab_curva ) + # cria-se a base para o grafico
  geom_point(aes(idade, HD)) + # plota-se os dados originais como pontos
  geom_line(aes( idade,
                HD_CURVA,
                color = limites ),
            size = 1.5 ) + # plota-se as linhas utilizando
```

```
labs(x = "Idade (meses)",
     y = "Altura dominante (m)",
     color = "Site") +
guides(color= guide_legend(reverse = T)) +
theme( #com theme muda-se o tamanho e estilo de letra de cada parte do grafico
      axis.title = element_text(size = 14),
      axis.text = element_text(size = 12))
curvas
```



5) Exportar graficos

```
ggsave("curvas.png", curvas, width = 12,height = 8)
```