

Relatório

Limpeza dos dados. Os pacientes 773, 489 e 212 aparentam apresentar erros no registro da data de evolução.

```
data <- readRDS('dados_att.rds') %>% group_by(pac, exame) %>%  
mutate(DT_EVOL=replace(DT_EVOL, pac==773, '2017-05-26')) %>%  
  mutate(DT_EVOL=replace(DT_EVOL, pac==489, '2017-04-14')) %>%  
  mutate(DT_EVOL=replace(DT_EVOL, pac==212, '2017-02-16')) %>%  
  mutate(DATA_FINAL =difftime(DT_EVOL,DT_IS,units='days') ) %>%  
  mutate(dt_sup=DIAS,dt_inic =lag(DIAS) ) %>% select(-DIAS) %>%  
  mutate(dt_ate_inter =difftime(DT_INTERN,DT_IS,units='days') )
```

Estruturação dos dados

```
dados <- data %>%  
  select(pac,EVOLUCAO,exame,valor,SEXO,DT_NASC,FEBRE:DIARREIA,  
    dt_inic,dt_sup,DATA_FINAL,OBITO,dt_ate_inter) %>%  
  mutate(Idade = year(as.Date('2019-12-30'))-year(DT_NASC)) %>%  
  select(-DT_NASC) %>%  
  spread(exame,valor) %>% mutate(TARGET = ifelse(OBITO =='OBITO',1,0)) %>%  
  select(-OBITO,-EVOLUCAO)  
  
##### Seleção de variáveis  
data_inic <- dados %>% group_by(pac) %>% filter(is.na(dt_inic)) %>% ungroup(pac) %>%  
  select(Idade,SEXO,TGO,TGP,HEMORRAGIA,DOR_ABDOM,MIALGIA,Leucocitos,  
    Bastoes,Creatinina,Linfocitos,Ureia,DATA_FINAL,TARGET)
```

Estruturação dos dados

```
dados <- data %>%  
  select(pac,EVOLUCAO,exame,valor,SEXO,DT_NASC,FEBRE:DIARREIA,  
    dt_inic,dt_sup,DATA_FINAL,OBITO,dt_ate_inter) %>%  
  mutate(Idade = year(as.Date('2019-12-30'))-year(DT_NASC)) %>%  
  select(-DT_NASC) %>%  
  spread(exame,valor) %>% mutate(TARGET = ifelse(OBITO =='OBITO',1,0)) %>%  
  select(-OBITO,-EVOLUCAO)
```

Seleção de variáveis

```
data_inic <- dados %>% group_by(pac) %>% filter(is.na(dt_inic)) %>% ungroup(pac) %>%  
  select(Idade,SEXO,TGO,TGP,HEMORRAGIA,DOR_ABDOM,MIALGIA,Leucocitos,  
    Bastoes,Creatinina,Linfocitos,Ureia,DATA_FINAL,ICTERICIA,HEMORRAGIA,TARGET)
```

Correção de dados faltantes

```
dados2 <- dados %>%
  select(pac,dt_inic,dt_sup,DATA_FINAL,TGO,TGP,TARGET,Idade,Ht,BD,BI,TAP,INR,
         Leucocitos,Bastoes,Creatinina,ICTERICIA,HEMORRAGIA,SEXO,dt_ate_inter
         ,Ureia,GamaGT,Fosfatase_alcal,Plaquetas,SINAL_FAGET) %>%
  mutate(DATA_FINAL = ifelse(DATA_FINAL<0,max(dt_sup),DATA_FINAL))

dados_completos <- dados2 %>% group_by(pac)%>% filter(any(DATA_FINAL == dt_sup))

dados_faltantes <- dados2 %>% filter(!(pac %in% (dados_completos$pac %>% unique)))

dados_faltantes_2 <- dados_faltantes %>%
  group_by(pac) %>%
  summarise(dt_inic = max(dt_sup),dt_sup =max(DATA_FINAL)) %>%
  bind_rows(dados_faltantes,.) %>% arrange(pac)

dados_final <- bind_rows(dados_completos,dados_faltantes_2) %>%
  group_by(pac) %>% mutate(TARGET = ifelse(dt_sup==max(dt_sup),max(TARGET,na.rm = TRUE),0))

dados_input <- dados_final %>% group_by(pac) %>%
  mutate_at(.vars=vars(DATA_FINAL,TGO,TGP,Ht,Leucocitos,Creatinina,Ht,Idade,SEXO),
            .funs=list(~na.locf(., na.rm = FALSE))) %>% filter(dt_sup !=0) %>%
  mutate_at(.vars=vars(DATA_FINAL,TGO,TGP,Ht,Leucocitos,Creatinina,Ht,Idade,SEXO),
            .funs=list(~na.locf(., na.rm = FALSE,fromLast=TRUE)))

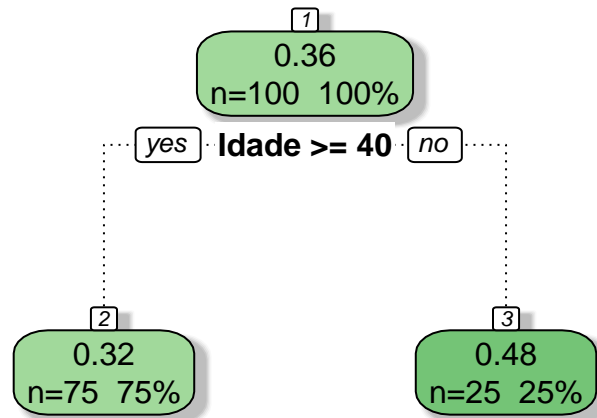
dados_input <- dados_input%>%
  group_by(pac) %>% mutate(TARGET = ifelse(dt_sup==max(dt_sup),
      max(TARGET,na.rm = TRUE),0)) %>% ungroup %>% mutate(dt_inic = ifelse(is.na(dt_inic),0,dt_inic))

dados_class <- dados_input %>% group_by(pac) %>%
  mutate(TARGET = max(TARGET)) %>%
  mutate(dt_sup = ifelse(dt_sup <7,'<7','>=7')) %>%
  mutate(dt_ate_inter=max(dt_ate_inter,na.rm = TRUE)) %>%
  mutate(dt_ate_inter = ifelse(dt_ate_inter <3,'<3','>=3'))
```

Mudança dos dados para classificação

```
require(rpart)
require(rattle)

dados_idade <- dados_class %>% select(pac,Idade,TARGET) %>% unique
Arvore_idade = rpart(TARGET~ Idade, data=dados_idade,maxdepth=1)
fancyRpartPlot(Arvore_idade)
```



Rattle 2019–nov–27 09:03:30 gusta

```

dados_class <- dados_input %>% group_by(pac) %>%
  mutate(TARGET = max(TARGET)) %>%
  mutate(dt_sup = ifelse(dt_sup < 7, '<7', '>=7')) %>%
  mutate(dt_ate_inter = max(dt_ate_inter, na.rm = TRUE)) %>%
  mutate(dt_ate_inter = ifelse(dt_ate_inter < 3, '<3', '>=3')) %>%
  mutate(Idade = ifelse(Idade < 40, '<40', '>=40'))
  
```

Nessa estrutura de dados estou considerando cada observação de um indivíduo como uma amostra do fenômeno. Para o indivíduo que sofreu óbito considere que houve óbito em todos os dias de doença, além disso, para a regressão logística categorizei o tempo de doença (dt_sup) em primeira semana de doença ou não.

Esse é um ajuste preliminar para ilustrar as técnicas e suas interpretações, a inclusão de mais variáveis fica a seu critério.

Regressão Logística

Procedimento step começando com uma variável

Variáveis testadas:

- log10(TGO)
- dt_sup
- Ht
- Creatinina
- ICTERICIA
- HEMORRAGIA

- log10(Leucocitos)
- dt_ate_inter
- BD
- BI
- TAP
- INR
- Plaquetas

Dados Faltantes

```
na_count <- apply(dados_class %>% ungroup %>%
  select(-c(pac,dt_inic,dt_sup,DATA_FINAL,TARGET)), function(y) sum(length(which(is.na(y)))))
na_count <- data.frame(Faltantes = na_count)
na_count
```

```
##           Faltantes
## TGO                0
## TGP                0
## Idade              0
## Ht                 0
## BD                172
## BI                174
## TAP               229
## INR               225
## Leucocitos         0
## Bastoes           142
## Creatinina         2
## ICTERICIA          155
## HEMORRAGIA         144
## SEXO               0
## dt_ate_inter       0
## Ureia              164
## GamaGT             259
## Fosfatase_alcal    290
## Plaquetas          119
## SINAL_FAGET        189
```

Preciso lidar esses valores faltantes, não é um procedimento trivial.

```
ajuste_logistica <- step(glm(TARGET ~ log10(TGO)+dt_sup+Ht+Creatinina+ICTERICIA+
  HEMORRAGIA+log10(Leucocitos)+dt_ate_inter+BD+BI+TAP+INR+Plaquetas,
  data=dados_class, family=binomial(link='logit')),direction='both')
```

```
## Start:  AIC=161.33
## TARGET ~ log10(TGO) + dt_sup + Ht + Creatinina + ICTERICIA +
##          HEMORRAGIA + log10(Leucocitos) + dt_ate_inter + BD + BI +
##          TAP + INR + Plaquetas
##
##           Df Deviance    AIC
## - BI        1   133.34 159.34
## - log10(Leucocitos) 1   133.57 159.57
## - Creatinina      1   133.64 159.64
## - log10(TGO)      1   134.69 160.69
## <none>           133.33 161.33
```

```

## - ICTERICIA          1   135.87 161.87
## - HEMORRAGIA         1   135.90 161.90
## - dt_ate_inter       1   136.37 162.37
## - TAP                1   137.43 163.43
## - Plaquetas          1   142.87 168.87
## - BD                 1   143.94 169.94
## - dt_sup             1   147.01 173.01
## - Ht                 1   150.99 176.99
## - INR                1   161.57 187.57
##
## Step:  AIC=159.34
## TARGET ~ log10(TGO) + dt_sup + Ht + Creatinina + ICTERICIA +
##          HEMORRAGIA + log10(Leucocitos) + dt_ate_inter + BD + TAP +
##          INR + Plaquetas
##
##              Df Deviance    AIC
## - log10(Leucocitos) 1   133.57 157.57
## - Creatinina         1   133.64 157.64
## - log10(TGO)         1   134.69 158.69
## <none>                133.34 159.34
## - HEMORRAGIA        1   135.90 159.90
## - ICTERICIA         1   135.90 159.90
## - dt_ate_inter      1   136.41 160.41
## + BI                1   133.33 161.33
## - TAP               1   137.43 161.43
## - Plaquetas         1   142.92 166.92
## - BD               1   144.17 168.17
## - dt_sup           1   147.02 171.02
## - Ht               1   151.97 175.97
## - INR              1   161.63 185.63
##
## Step:  AIC=157.57
## TARGET ~ log10(TGO) + dt_sup + Ht + Creatinina + ICTERICIA +
##          HEMORRAGIA + dt_ate_inter + BD + TAP + INR + Plaquetas
##
##              Df Deviance    AIC
## - Creatinina        1   134.35 156.35
## - log10(TGO)        1   134.78 156.78
## <none>               133.57 157.57
## - ICTERICIA         1   135.95 157.95
## - HEMORRAGIA        1   136.74 158.74
## - dt_ate_inter      1   137.28 159.28
## + log10(Leucocitos) 1   133.34 159.34
## - TAP               1   137.50 159.50
## + BI                1   133.57 159.57
## - Plaquetas         1   143.00 165.00
## - BD               1   144.26 166.26
## - dt_sup           1   150.11 172.11
## - Ht               1   154.03 176.03
## - INR              1   162.15 184.15
##
## Step:  AIC=156.35
## TARGET ~ log10(TGO) + dt_sup + Ht + ICTERICIA + HEMORRAGIA +
##          dt_ate_inter + BD + TAP + INR + Plaquetas

```

```
##
##           Df Deviance    AIC
## - log10(TGO)      1   135.12 155.12
## <none>              134.35 156.35
## - ICTERICIA       1   136.97 156.97
## - HEMORRAGIA      1   137.12 157.12
## - dt_ate_inter    1   137.55 157.55
## + Creatinina      1   133.57 157.57
## + log10(Leucocitos) 1   133.64 157.64
## - TAP             1   138.18 158.18
## + BI              1   134.32 158.32
## - Plaquetas       1   143.01 163.01
## - BD              1   144.48 164.48
## - dt_sup          1   150.34 170.34
## - Ht              1   155.74 175.74
## - INR             1   162.29 182.29
##
## Step:  AIC=155.12
## TARGET ~ dt_sup + Ht + ICTERICIA + HEMORRAGIA + dt_ate_inter +
##         BD + TAP + INR + Plaquetas
##
##           Df Deviance    AIC
## <none>              135.12 155.12
## - ICTERICIA       1   137.81 155.81
## - HEMORRAGIA      1   138.14 156.14
## + log10(TGO)      1   134.35 156.35
## + Creatinina      1   134.78 156.78
## + log10(Leucocitos) 1   134.81 156.81
## - TAP             1   139.06 157.06
## + BI              1   135.07 157.07
## - dt_ate_inter    1   139.70 157.70
## - Plaquetas       1   143.25 161.25
## - BD              1   145.29 163.29
## - dt_sup          1   155.12 173.12
## - Ht              1   155.81 173.81
## - INR             1   178.44 196.44
```

Utilizando as variáveis selecionadas no procedimento step, o melhor modelo é, pelo critério do AIC:

```
ajuste_logistica <- glm(TARGET~ dt_sup + Ht + Creatinina + Idade+dt_ate_inter + BD +
  TAP + INR + Ureia+ Plaquetas+Idade,
  data=dados_class, family=binomial(link='logit'))
```

```
summary(ajuste_logistica)
```

```
##
## Call:
## glm(formula = TARGET ~ dt_sup + Ht + Creatinina + Idade + dt_ate_inter +
##     BD + TAP + INR + Ureia + Plaquetas + Idade, family = binomial(link = "logit"),
##     data = dados_class)
```

```
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2440  -0.2409  -0.0948  -0.0193   3.6124
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    7.856e-01  1.668e+00   0.471 0.637654
## dt_sup>=7      -2.457e+00  6.387e-01  -3.847 0.000119 ***
## Ht             -1.021e-01  3.103e-02  -3.289 0.001006 **
## Creatinina     -4.726e-01  2.274e-01  -2.079 0.037646 *
## Idade>=40      -1.867e+00  5.469e-01  -3.414 0.000641 ***
## dt_ate_inter>=3 1.164e+00  5.703e-01   2.041 0.041271 *
## BD             2.134e-01  6.508e-02   3.279 0.001041 **
## TAP            -3.299e-02  9.454e-03  -3.490 0.000484 ***
## INR            1.522e+00  3.388e-01   4.492 7.05e-06 ***
## Ureia          3.838e-02  1.145e-02   3.351 0.000806 ***
## Plaquetas      6.279e-06  3.705e-06   1.695 0.090118 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 437.54  on 393  degrees of freedom
## Residual deviance: 135.19  on 383  degrees of freedom
## (323 observations deleted due to missingness)
## AIC: 157.19
##
## Number of Fisher Scoring iterations: 7
```

Interpretação das variáveis

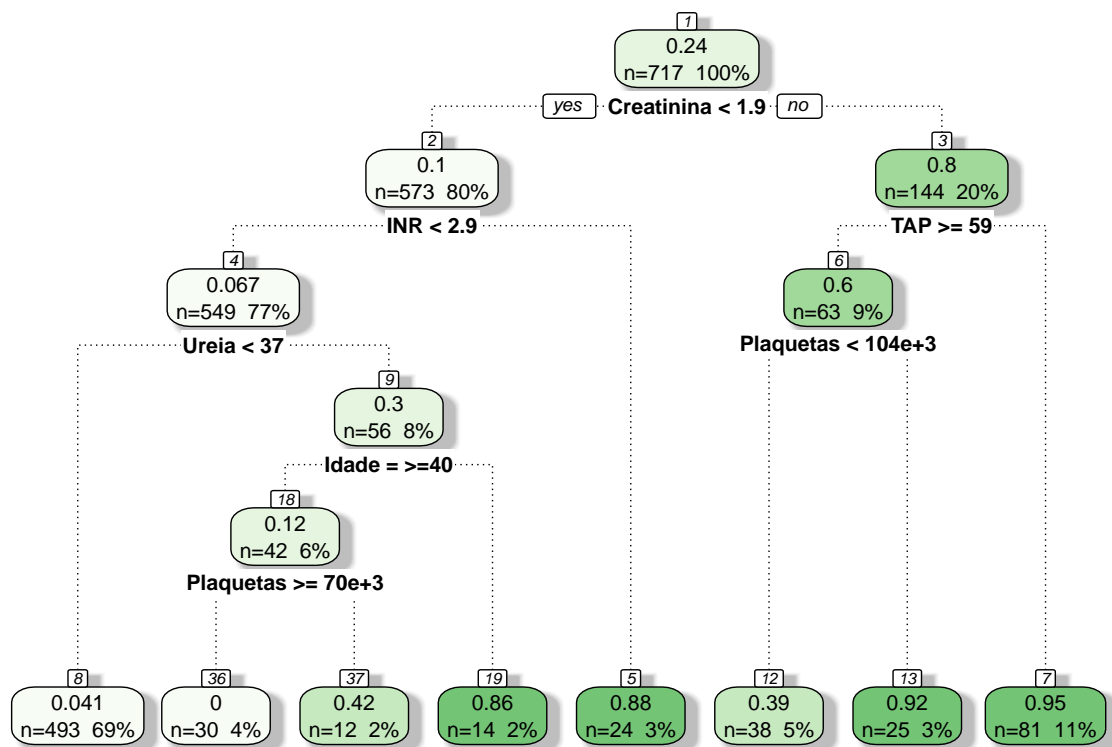
Estimativas Negativas indicam diminuição na chance de óbito e estimativas positivas indicam aumento.

- O indivíduo estar a partir do oitavo dia de doença aumenta a razão de chances de cura em $\exp(3.022)=20.53232$ vezes.
- O aumento de uma unidade de HT aumenta a razão de chances de cura em $\exp(-0.1169)=0.89$ vezes.
- A presença do sintoma de hemorragia aumenta a razão de chances de óbito em $\exp(1.160)=3.2$ vezes.
- O indivíduo ser internado após apenas após o terceiro dia de manifestação dos sintomas aumenta a razão de chances de óbito em $\exp(0.670)=2.63$

Árvore de Decisão

```
require(rpart)
require(rattle)

Arvore = rpart(TARGET~ dt_sup + Ht + Creatinina + Idade+dt_ate_inter + BD +
  TAP + INR + Ureia+ Plaquetas+Idade, data=dados_class,maxdepth=7)
fancyRpartPlot(Arvore)
```



Rattle 2019-nov-27 09:03:30 gusta

DESCONSIDERAR

Tentativas de aplicação de análise de sobrevivência

->