

# Quarta Lista - Analise de Sobrevivência

Gabriel D'assumpção de Carvalho

2025-08-11

## Introdução

Esta atividade tem como objetivo ajustar modelos de regressão para dados de sobrevivência de pacientes com leucemia no grupo controle, que não apresenta o antígeno Ag-. As variáveis explicativas do estudo são a contagem de glóbulos brancos (WBC) e o seu logaritmo na base 10.

## Bibliotecas

```
# Carregamento dos pacotes necessários  
  
# Pacote para análise de sobrevivência  
# install.packages("survival")  
library(survival)
```

## Leucemia

Abaixo está o dataframe dos pacientes sem leucemia, que constituem o grupo controle do estudo, com 16 observações. Observa-se que a variável explicativa WBC é contínua, assumindo valores em  $\mathbb{Z}^+$ .

```
agN <- data.frame(  
  time = c(56, 65, 17, 7, 16, 22, 3, 4, 2, 3, 8, 4, 3, 30, 4, 43),  
  status = rep(1, 16),  
  WBC = c(4400, 3000, 4000, 1500, 9000, 5300, 10000, 19000, 27000, 28000, 31000, 26000,  
    21000, 79000, 100000, 100000),  
  log_10 = c(3.64, 3.48, 3.60, 3.18, 3.95, 3.72, 4.00, 4.28, 4.43, 4.45, 4.49, 4.41,  
    4.32, 4.90, 5.00, 5.00)  
)  
print(agN)
```

##	time	status	WBC	log_10
## 1	56	1	4400	3.64
## 2	65	1	3000	3.48
## 3	17	1	4000	3.60
## 4	7	1	1500	3.18
## 5	16	1	9000	3.95
## 6	22	1	5300	3.72
## 7	3	1	10000	4.00
## 8	4	1	19000	4.28
## 9	2	1	27000	4.43
## 10	3	1	28000	4.45
## 11	8	1	31000	4.49
## 12	4	1	26000	4.41
## 13	3	1	21000	4.32
## 14	30	1	79000	4.90
## 15	4	1	100000	5.00
## 16	43	1	100000	5.00

A seguir, ajustaremos o modelo de Cox para avaliar a relação entre as variáveis WBC,  $\log_{10}$  e o tempo de sobrevivência, com o objetivo de analisar o impacto da contagem de glóbulos brancos sobre o tempo de sobrevivência dos pacientes.

## Modelo de Cox

```
# Ajuste do modelo de Cox
modelo_cox <- coxph(Surv(time, status) ~ WBC + log_10, data = agN)
print(summary(modelo_cox))
```

```
## Call:
## coxph(formula = Surv(time, status) ~ WBC + log_10, data = agN)
##
##      n= 16, number of events= 16
##
##              coef exp(coef)    se(coef)      z Pr(>|z|)
## WBC      -6.004e-05  9.999e-01  2.544e-05 -2.360   0.0183 *
## log_10   4.076e+00  5.890e+01  1.616e+00  2.522   0.0117 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## WBC              0.9999    1.00006    0.9999      1
## log_10         58.8974    0.01698    2.4789    1399
##
## Concordance= 0.75 (se = 0.07 )
## Likelihood ratio test= 7.78 on 2 df,  p=0.02
## Wald test              = 6.36 on 2 df,  p=0.04
## Score (logrank) test = 7.12 on 2 df,  p=0.03
```

O modelo de Cox foi ajustado considerando as variáveis WBC e log(WBC). Os coeficientes estimados foram aproximadamente  $-6 \times 10^{-5}$  para WBC e 4.07 para log(WBC).

Para interpretar o efeito das variáveis sobre o risco, utiliza-se a exponencial dos coeficientes, que resultam em aproximadamente 1 para WBC e 58.66 para log(WBC). Isso significa que, para cada aumento unitário em WBC, o risco praticamente não se altera (razão de risco próxima de 1). Por outro lado, para cada aumento unitário no log(WBC), o risco aumenta em cerca de 58.66 vezes, reduzindo o tempo de sobrevivência.

Ambos os coeficientes apresentaram significância estatística, com valores-p abaixo de 2%, permitindo rejeitar a hipótese nula com 95% de confiança.

O intervalo de confiança para a razão de risco para WBC e log(WBC) é aproximadamente [0.9999, 1] e [2.47, 1399], respectivamente.

Portanto, apesar do WBC ter tido um valor-p baixo, seu intervalo de confiança inclui 1, indicando que ele não tem impacto significativo no tempo de sobrevivência. Já o log(WBC) é uma variável que tem impacto significativo no tempo de sobrevivência, considerando que ambas estão no modelo.

Também é possível verificar que a estatística de concordância (**Concordance**) foi 0.75, indicando que, ao comparar aleatoriamente dois indivíduos, o modelo acertou em 75% das vezes qual deles teria maior tempo de sobrevivência.

Além disso, os testes da razão de verossimilhança (likelihood ratio) e de Wald avaliam se o modelo é significativo, isto é, se pelo menos um coeficiente é diferente de zero, apresentando p-valores de 0.02 e 0.04, respectivamente, reforçando a validade do modelo.