

Terceira Lista - Análise de Sobrevivência

Gabriel D'assumpção de Carvalho

2025-08-07

Introdução

Este documento apresenta uma análise exploratória de sobrevivência com dados sobre hepatite, reincidência de tumor sólido e malária. A análise utilizará os estimadores de Kaplan-Meier, Nelson-Aalen e Atuarial e o teste de log-rank para comparar as curvas de sobrevivência entre os grupos de cada conjunto de dados.

Bibliotecas

```
# Carregamento dos pacotes necessários

# Pacote para análise de sobrevivência
# install.packages("survival")
library(survival)

# Pacote para visualização dos dados
# install.packages("ggplot2")
library(ggplot2)

# Pacote para manipulação de dados
# install.packages("dplyr")
library(dplyr)
```

Funções de análise de sobrevivência

A seguir vamos estruturar as funções de sobrevivência não paramétricas, onde vão ser utilizados os estimadores Atuarial, Kaplan-Meier e Nelson-Aalen, além da função do teste log-rank. Essas funções serão utilizadas para calcular as curvas de sobrevivência e os intervalos de confiança.

```
##### Estimador Atuarial #####
eaa <- function(time, evento, classEvento = 1, alpha = 0.05, conf = "normal") {
  tempo_evento <- sort(unique(time[evento == classEvento]))

  data <- data.frame(
    time = tempo_evento,
    n_risk = NA,
    n_event = NA,
    censored = NA,
    n_risk_adj = NA,
    q_i = NA,
    p_i = NA,
    S_t_i = NA,
    std_error = NA,
    lower = NA,
    upper = NA
  )

  S_t_back <- 1

  for (i in seq_along(data$time)) {
    t <- data$time[i]

    data$n_risk[i] <- sum(time >= t)
```

```

data$n_event[i] <- sum(time == t & evento == classEvento)
data$censored[i] <- sum(time == t & evento != classEvento)
data$n_risk_adj[i] <- data$n_risk[i] - data$censored[i] / 2

if (data$n_risk_adj[i] > 0) {
  data$q_i[i] <- data$n_event[i] / data$n_risk_adj[i]
  data$p_i[i] <- 1 - data$q_i[i]
  data$S_t_i[i] <- S_t_back * data$p_i[i]
  S_t_back <- data$S_t_i[i]
} else {
  data$S_t_i[i] <- S_t_back
}

# Variância (usada nos dois métodos)
var_S_t <- data$S_t_i[i]^2 * sum(data$q_i[1:i] / (data$n_risk_adj[1:i] *
data$p_i[1:i]))
data$std_error[i] <- sqrt(var_S_t)

if (conf == "normal") {
  data$upper[i] <- data$S_t_i[i] + qnorm(1 - alpha / 2) * data$std_error[i]
  data$lower[i] <- data$S_t_i[i] - qnorm(1 - alpha / 2) * data$std_error[i]
}

if (conf == "log-log") {
  if (data$S_t_i[i] > 0 && data$S_t_i[i] < 1) {
    loglog_S <- log(-log(data$S_t_i[i]))
    se_loglog <- sqrt(sum(data$q_i[1:i] / (data$n_risk_adj[1:i] * data$p_i[1:i])))
    z <- qnorm(1 - alpha / 2)
    lower_loglog <- loglog_S - z * se_loglog
    upper_loglog <- loglog_S + z * se_loglog
    data$lower[i] <- exp(-exp(upper_loglog))
    data$upper[i] <- exp(-exp(lower_loglog))
  } else {
    data$lower[i] <- NA
    data$upper[i] <- NA
  }
}
}
cols_to_round <- c("n_risk_adj", "q_i", "p_i", "S_t_i", "std_error", "lower", "upper")
data[cols_to_round] <- round(data[cols_to_round], 3)

return(data)
}

##### Estimador de Kaplan-Meier #####
ekm <- function(time, evento, classEvento = 1, alpha = 0.05, conf = "normal") {
  tempo_evento <- sort(unique(time[evento == classEvento]))

  data <- data.frame(
    time = tempo_evento,
    n_risk = NA,
    n_event = NA,

```

```

    censored = NA,
    q_i = NA,
    p_i = NA,
    S_t_i = NA,
    std_error = NA,
    lower = NA,
    upper = NA
  )

  S_t_back <- 1

  for (i in seq_along(data$time)) {
    t <- data$time[i]

    data$n_risk[i] <- sum(time >= t)
    data$n_event[i] <- sum(time == t & evento == classEvento)
    data$censored[i] <- sum(time == t & evento != classEvento)

    if (data$n_risk[i] > 0) {
      data$q_i[i] <- data$n_event[i] / data$n_risk[i]
      data$p_i[i] <- 1 - data$q_i[i]
      data$S_t_i[i] <- S_t_back * data$p_i[i]
      S_t_back <- data$S_t_i[i]
    } else {
      data$S_t_i[i] <- S_t_back
    }

    term <- data$n_event[1:i] / (data$n_risk[1:i] * (data$n_risk[1:i] -
data$n_event[1:i]))
    term[is.infinite(term)] <- 0
    var_S_t <- data$S_t_i[i]^2 * sum(term, na.rm = TRUE)

    data$std_error[i] <- sqrt(var_S_t)

    if (conf == "log-log") {
      if (data$S_t_i[i] > 0 && data$S_t_i[i] < 1) {
        z <- qnorm(1 - alpha / 2)
        var_loglog <- sum(term, na.rm = TRUE)
        loglog_S <- log(-log(data$S_t_i[i]))

        lower_loglog <- loglog_S - z * sqrt(var_loglog)
        upper_loglog <- loglog_S + z * sqrt(var_loglog)

        data$lower[i] <- exp(-exp(upper_loglog))
        data$upper[i] <- exp(-exp(lower_loglog))
      } else {
        data$lower[i] <- NA
        data$upper[i] <- NA
      }
    }

    if (conf == "normal") {
      if (data$S_t_i[i] > 0 && data$S_t_i[i] < 1) {

```

```

    data$upper[i] <- data$S_t_i[i] + qnorm(1 - alpha / 2) * data$std_error[i]
    data$lower[i] <- data$S_t_i[i] - qnorm(1 - alpha / 2) * data$std_error[i]
  } else {
    data$upper[i] <- NA
    data$lower[i] <- NA
  }
}
}
cols_to_round <- c("q_i", "p_i", "S_t_i", "std_error", "lower", "upper")
data[cols_to_round] <- round(data[cols_to_round], 3)

return(data)
}

##### Estimador de Nelson-Aalen #####
ena <- function(time, evento, classEvento = 1, alpha = 0.05, conf = "normal") {
  tempo_evento <- sort(unique(time[evento == classEvento]))

  data <- data.frame(
    time = tempo_evento,
    n_risk = NA,
    n_event = NA,
    censored = NA,
    q_i = NA,
    p_i = NA,
    S_t_i = NA,
    std_error = NA,
    lower = NA,
    upper = NA
  )

  for (i in seq_along(data$time)) {
    t <- data$time[i]

    data$n_risk[i] <- sum(time >= t)
    data$n_event[i] <- sum(time == t & evento == classEvento)

    if (data$n_risk[i] > 0) {
      data$q_i[i] <- data$n_event[i] / data$n_risk[i]
      data$p_i[i] <- 1 - data$q_i[i]
      data$censored[i] <- sum(time == t & evento != classEvento)
      estimatorNelson <- sum(data$n_event[1:i] / data$n_risk[1:i])
      data$S_t_i[i] <- exp(-estimatorNelson)
    } else {
      data$S_t_i[i] <- 1
    }
  }
  var_S_t <- data$S_t_i[i]**2 * sum(data$n_event[1:i] / data$n_risk[1:i]**2)
  data$std_error[i] <- sqrt(var_S_t)

  if (conf == "log-log") {
    z <- qnorm(1 - alpha / 2)
    if (data$S_t_i[i] > 0 && data$S_t_i[i] < 1) {

```

```

loglog_S <- log(-log(data$S_t_i[i]))
var_loglog <- sum(data$n_event[1:i] / (data$n_risk[1:i]^2))

lower_loglog <- loglog_S - z * sqrt(var_loglog)
upper_loglog <- loglog_S + z * sqrt(var_loglog)

data$lower[i] <- exp(-exp(upper_loglog))
data$upper[i] <- exp(-exp(lower_loglog))
} else {
  data$lower[i] <- NA
  data$upper[i] <- NA
}
}

if (conf == "normal") {
  if (data$S_t_i[i] > 0 && data$S_t_i[i] < 1) {
    data$upper[i] <- data$S_t_i[i] + qnorm(1 - alpha / 2) * data$std_error[i]
    data$lower[i] <- data$S_t_i[i] - qnorm(1 - alpha / 2) * data$std_error[i]
  } else {
    data$upper[i] <- NA
    data$lower[i] <- NA
  }
}
}

cols_to_round <- c("q_i", "p_i", "S_t_i", "std_error", "lower", "upper")
data[cols_to_round] <- round(data[cols_to_round], 3)

return(data)
}

##### Teste Log-Rank #####
logrank <- function(dadosGrupo1, dadosGrupo2) {
  time <- sort(unique(c(dadosGrupo1$time, dadosGrupo2$time)))

  O1_total <- 0 # Observado no grupo 1
  O2_total <- 0 # Observado no grupo 2
  E1_total <- 0 # Esperado no grupo 1
  E2_total <- 0 # Esperado no grupo 2
  V_total <- 0 # Variância acumulada

  for (t in time) {
    n1 <- sum(dadosGrupo1$time >= t)
    n2 <- sum(dadosGrupo2$time >= t)

    d1 <- sum(dadosGrupo1$time == t & dadosGrupo1$evento == 1)
    d2 <- sum(dadosGrupo2$time == t & dadosGrupo2$evento == 1)

    d <- d1 + d2
    n <- n1 + n2

    if (d > 0 && n > 1) {

```

```

E1 <- d * (n1 / n)
E2 <- d * (n2 / n)
V <- d * (n1 / n) * (n2 / n) * ((n - d) / (n - 1))

O1_total <- O1_total + d1
O2_total <- O2_total + d2
E1_total <- E1_total + E1
E2_total <- E2_total + E2
V_total <- V_total + V
}
}

Z <- (O1_total - E1_total)^2 / V_total
p_value <- 1 - pchisq(Z, df = 1)

tabela <- data.frame(
  Grupo = c("Grupo 1", "Grupo 2"),
  Observado = c(O1_total, O2_total),
  Esperado = c(E1_total, E2_total)
)

print(tabela)
cat("\nVariância total:", round(V_total, 4))
cat("\nEstatística do teste (Chi²):", round(Z, 4))
cat("\nnp-valor:", round(p_value, 4), "\n")

invisible(list(tabela = tabela, chi2 = Z, p = p_value))
}

```

Hepatite

Neste estudo, foram analisados dados de pacientes com hepatite viral aguda. O objetivo foi comparar a taxa de sobrevivência entre dois grupos, foram escolhidos 29 pacientes com hepatite viral aguda e separados aleatoriamente em ambos os conjuntos. Os pacientes foram acompanhados durante 16 semanas (4 meses) ou até a falha ou censura do paciente.

- **Tipo de estudo:** Estudo clínico randomizado;
- **Grupo de tratamento:** Pacientes que receberam terapia com esteroides;
- **Grupo controle:** Pacientes que receberam placebo.

```

##### Hepatite #####
hControle <- data.frame(
  time = c(
    1, 2, 3, 3, 5, 5, 16, 16, 16, 16,
    16, 16, 16, 16
  ),
  evento = c(0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
)

hCasos <- data.frame(
  time = c(
    1, 1, 1, 1, 4, 5, 7, 8, 10, 10, 12, 16, 16,

```

```

16
),
evento = c(1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0)
)

```

Como os tempos até a falha são registrados em semanas (uma variável discreta), um gráfico de barras é a escolha ideal para visualizar a distribuição dos eventos ao longo do tempo. Isso nos permitirá comparar diretamente a frequência de falhas em cada semana entre os grupos controle e casos.

```

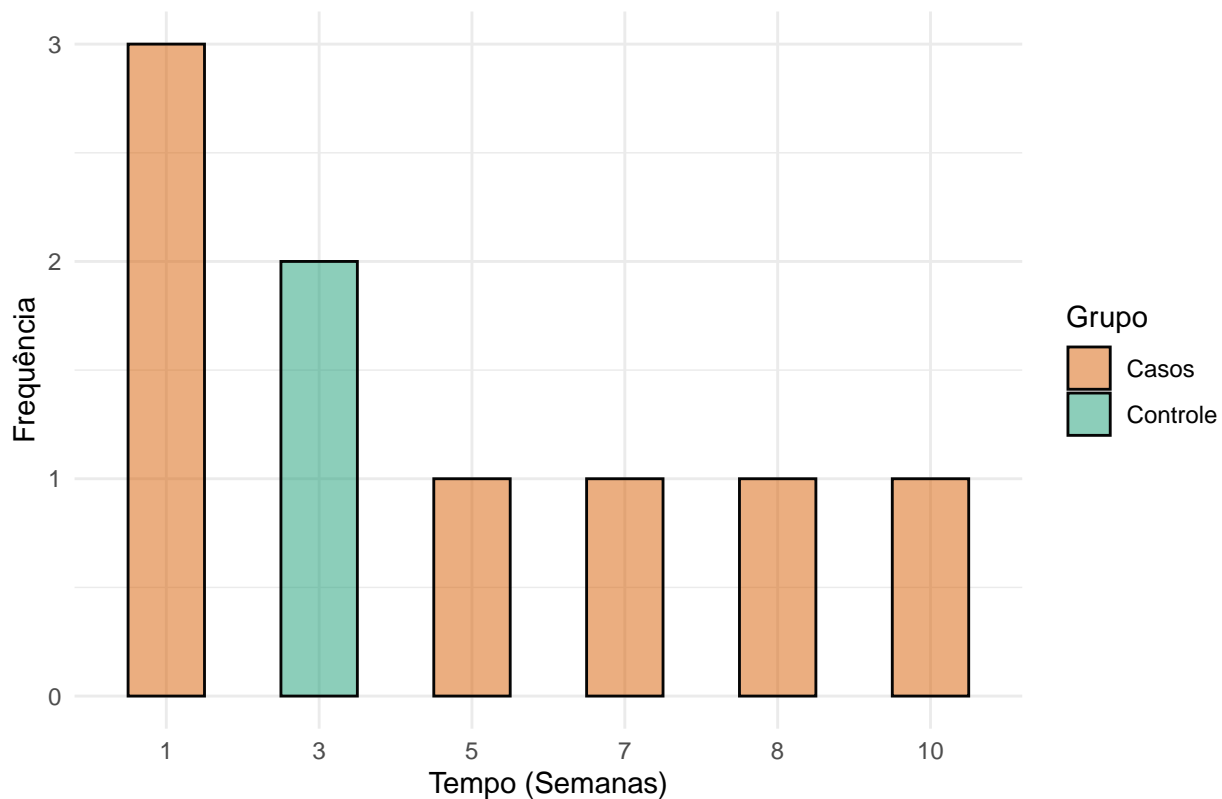
dados <- rbind(
  data.frame(time = hControle$time[hControle$evento == 1], grupo = "Controle"),
  data.frame(time = hCasos$time[hCasos$evento == 1], grupo = "Casos")
)

dados$time <- factor(dados$time, levels = sort(unique(dados$time)))

ggplot(dados, aes(x = time, fill = grupo)) +
  geom_bar(position = "identity", alpha = 0.5, color = "#050505", width = 0.5) +
  labs(
    title = "Gráfico de Barras dos Tempos até o Evento - Hepatite",
    x = "Tempo (Semanas)",
    y = "Frequência",
    fill = "Grupo"
  ) +
  scale_fill_manual(values = c("Controle" = "#1b9e77", "Casos" = "#d95f02")) +
  theme_minimal()

```

Gráfico de Barras dos Tempos até o Evento – Hepatite




```
print(length(hControle$time))
```

```
## [1] 15
```

```
print(length(hCasos$time))
```

```
## [1] 14
```

A análise do gráfico de barras revela diferenças marcantes na ocorrência de falhas entre os grupos. O grupo Controle, composto por 15 indivíduos, registrou apenas dois eventos de falha, ambos na terceira semana. Em contrapartida, o grupo Casos, com 14 indivíduos, apresentou sete eventos, distribuídos ao longo do estudo: três na primeira semana e os demais nas semanas 5, 7, 8 e 10.

A alta proporção de censura no grupo Controle, onde a maioria dos pacientes completou o estudo sem apresentar o evento, sugere uma taxa de sobrevivência consideravelmente maior em comparação ao grupo Casos. Esta disparidade será o foco da análise subsequente com as curvas de sobrevivência.

Estimador de Kaplan-Meier

A seguir, vamos calcular as curvas de sobrevivência para os grupos controle e casos utilizando o estimador Kaplan-Meier. As curvas serão plotadas para visualização das diferenças entre os grupos.

```
ekmControle <- ekm(
  time = hControle$time,
  evento = hControle$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
ekmCasos <- ekm(
  time = hCasos$time,
  evento = hCasos$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
```

```
print(ekmControle)
```

```
##   time n_risk n_event censored   q_i   p_i S_t_i std_error lower upper
## 1     3     13        2         1 0.154 0.846 0.846      0.1  0.81 0.876
```

Ao analisar a tabela gerada pela função criada no início deste relatório, observa-se que o grupo controle apresenta uma taxa de sobrevivência estimada em aproximadamente 0,846 na terceira semana, com desvio padrão em torno de 0,1. O intervalo de confiança calculado pelo método log-log situa-se entre aproximadamente 0,81 e 0,876.

```
print(ekmCasos)
```

```
##   time n_risk n_event censored   q_i   p_i S_t_i std_error lower upper
## 1     1     14        3         1 0.214 0.786 0.786      0.110 0.728 0.832
## 2     5      9        1         0 0.111 0.889 0.698      0.128 0.598 0.778
## 3     7      8        1         0 0.125 0.875 0.611      0.138 0.464 0.729
## 4     8      7        1         0 0.143 0.857 0.524      0.143 0.331 0.685
## 5    10      6        1         1 0.167 0.833 0.437      0.144 0.206 0.647
```

Para o grupo de casos, a taxa de sobrevivência estimada cai para aproximadamente 0,786 já na primeira semana. A sobrevida continua a diminuir, atingindo cerca de 0,436 na décima semana, indicando que menos da metade dos indivíduos deste grupo sobreviveu até esse ponto. O intervalo de confiança para esta estimativa (0,206 a 0,648) é amplo, refletindo o aumento da incerteza estatística ao longo do tempo devido ao acúmulo de variância e à redução do número de indivíduos em risco.

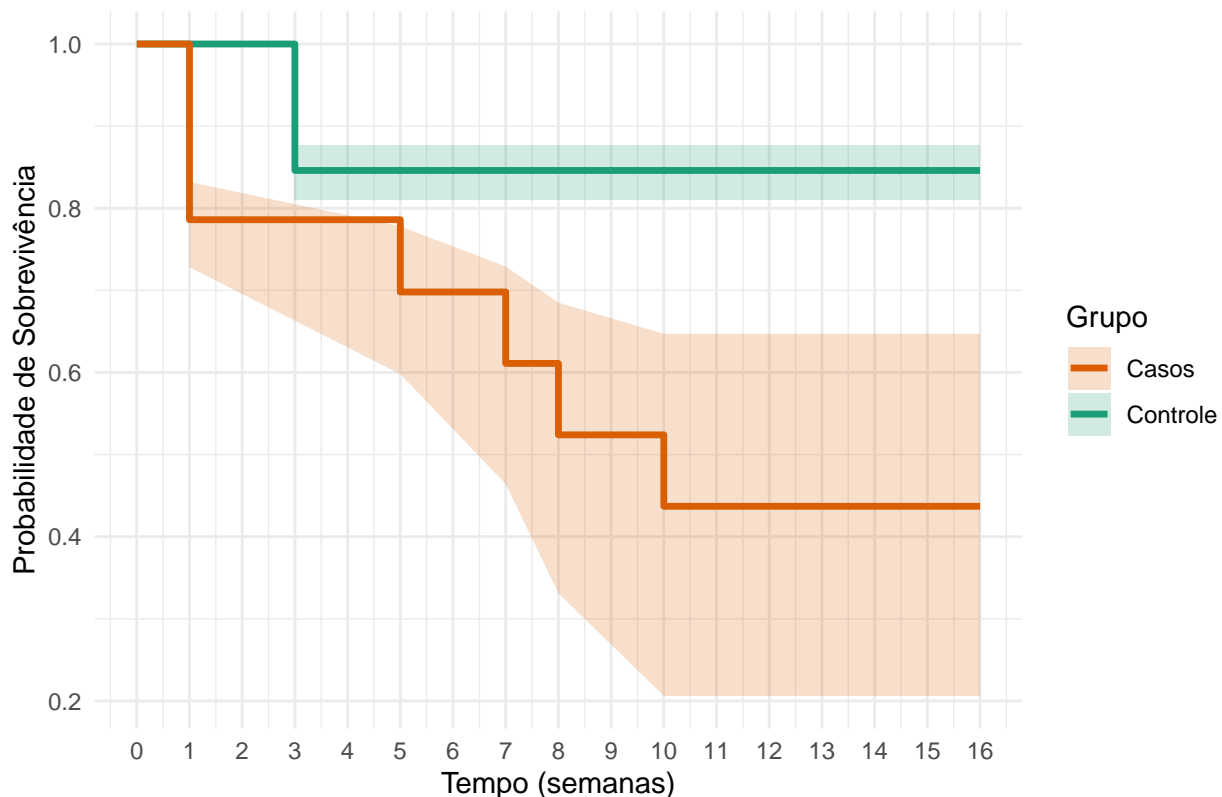
A seguir, vamos plotar as curvas de sobrevivência Kaplan-Meier para os grupos controle e casos. Para isso, adicionaremos manualmente os pontos iniciais e finais das curvas, garantindo que ambas comecem em 1 (100% de sobrevivência) e terminem no último tempo observado.

```
# Adiciona manualmente os pontos iniciais e finais
ekmControle_ext <- ekmControle %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 16,
    S_t_i = tail(ekmControle$S_t_i, 1),
    lower = tail(ekmControle$lower, 1),
    upper = tail(ekmControle$upper, 1)
  )

ekmCasos_ext <- ekmCasos %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 16,
    S_t_i = tail(ekmCasos$S_t_i, 1),
    lower = tail(ekmCasos$lower, 1),
    upper = tail(ekmCasos$upper, 1)
  )

# Gráfico de escada Kaplan-Meier
ggplot() +
  geom_step(data = ekmControle_ext, aes(x = time, y = S_t_i, color = "Controle"), size =
    1.2) +
  geom_step(data = ekmCasos_ext, aes(x = time, y = S_t_i, color = "Casos"), size = 1.2) +
  geom_ribbon(
    data = ekmControle_ext,
    aes(x = time, ymin = lower, ymax = upper, fill = "Controle"), alpha = 0.2
  ) +
  geom_ribbon(
    data = ekmCasos_ext,
    aes(x = time, ymin = lower, ymax = upper, fill = "Casos"), alpha = 0.2
  ) +
  scale_color_manual(values = c("Controle" = "#1b9e77", "Casos" = "#d95f02")) +
  scale_fill_manual(values = c("Controle" = "#1b9e77", "Casos" = "#d95f02")) +
  labs(
    title = "Curvas de Sobrevivência de Hepatite - Estimador Kaplan-Meier",
    x = "Tempo (semanas)",
    y = "Probabilidade de Sobrevivência",
    color = "Grupo",
    fill = "Grupo"
  ) +
  theme_minimal() +
  scale_x_continuous(breaks = seq(0, 16, 1))
```

Curvas de Sobrevivência de Hepatite – Estimador Kaplan–Meier



O gráfico de Kaplan-Meier ilustra uma nítida separação entre as curvas de sobrevivência dos grupos controle e casos, com o grupo controle mantendo consistentemente uma taxa de sobrevivência mais elevada. A distância entre as curvas é menor no início do acompanhamento, particularmente entre a terceira e a quinta semana, mas se acentua marcadamente a partir deste ponto, indicando uma divergência crescente nos desfechos. Adicionalmente, é notável o alargamento do intervalo de confiança para o grupo de casos ao longo do tempo. Esse aumento da incerteza reflete a diminuição do número de indivíduos em risco, o que torna a estimativa da taxa de sobrevivência progressivamente menos precisa.

Estimador Atuarial

A diante, utilizamos o estimador atuarial para calcular as curvas de sobrevivência dos grupos controle e casos. As curvas resultantes serão plotadas para permitir uma comparação visual entre os grupos, evidenciando possíveis diferenças nos padrões de sobrevivência ao longo do tempo.

```
eaaControle <- eaa(
  time = hControle$time,
  evento = hControle$evento,
  classEvento = 1,
  conf = "log-log"
)
eaaCasos <- eaa(
  time = hCasos$time,
  evento = hCasos$evento,
  classEvento = 1,
  conf = "log-log"
)
```

```
print(eaaControle)
```

```
##   time n_risk n_event censored n_risk_adj  q_i  p_i S_t_i std_error lower upper
## 1     3     13        2         1      12.5 0.16 0.84  0.84    0.104 0.801 0.872
```

O método atuarial, que ajusta o denominador de risco ao assumir que as censuras ocorrem uniformemente ao longo de cada intervalo, oferece uma perspectiva ligeiramente diferente. Para o grupo controle, a sobrevida estimada é de aproximadamente 0,84, um valor um pouco menor que o estimado por Kaplan-Meier. Essa abordagem também resulta em um intervalo de confiança mais amplo (aproximadamente 0,800 a 0,872), refletindo uma maior incerteza na estimativa.

```
print(eaaCasos)
```

```
##   time n_risk n_event censored n_risk_adj  q_i  p_i S_t_i std_error lower
## 1     1     14        3         1      13.5 0.222 0.778 0.778    0.113 0.716
## 2     5      9        1         0       9.0 0.111 0.889 0.691    0.129 0.587
## 3     7      8        1         0       8.0 0.125 0.875 0.605    0.139 0.454
## 4     8      7        1         0       7.0 0.143 0.857 0.519    0.144 0.323
## 5    10      6        1         1       5.5 0.182 0.818 0.424    0.145 0.187
##   upper
## 1 0.828
## 2 0.774
## 3 0.726
## 4 0.683
## 5 0.645
```

No grupo de casos, a estimativa atuarial de sobrevida na primeira semana é de 0,778, caindo para 0,424 na décima semana, o que corrobora a observação de que menos da metade dos indivíduos sobreviveu até esse ponto. O intervalo de confiança para a décima semana é notavelmente amplo (0,187 a 0,645), destacando a crescente incerteza à medida que o número de indivíduos em risco diminui.

Para representar corretamente a curva de sobrevivência pelo método atuarial, o ideal é usar um gráfico de linha, já que esse método assume distribuição uniforme de eventos dentro dos intervalos.

```
# Adiciona manualmente os pontos iniciais e finais
eaaControle_ext <- eaaControle %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 16,
    S_t_i = tail(eaaControle$S_t_i, 1),
    lower = tail(eaaControle$lower, 1),
    upper = tail(eaaControle$upper, 1)
  )

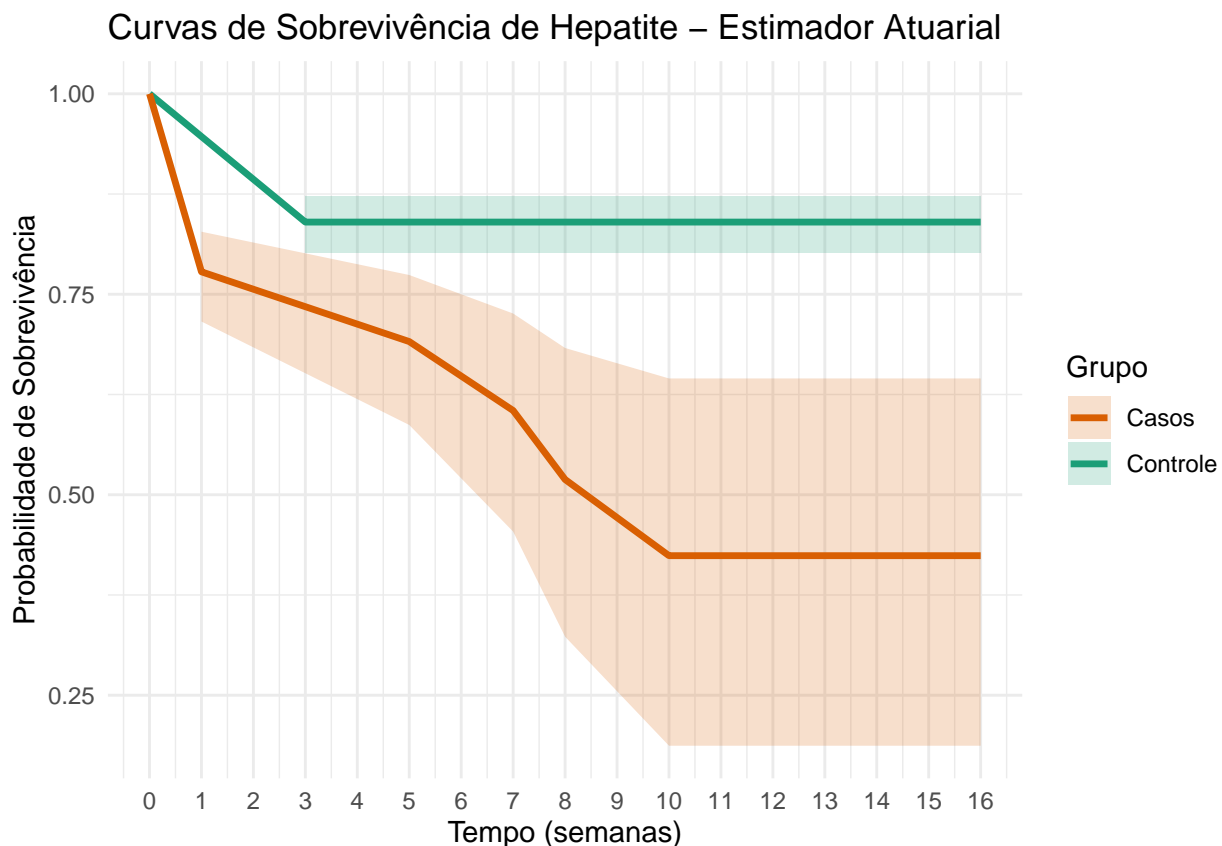
eaaCasos_ext <- eaaCasos %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 16,
    S_t_i = tail(eaaCasos$S_t_i, 1),
    lower = tail(eaaCasos$lower, 1),
    upper = tail(eaaCasos$upper, 1)
  )

# Gráfico de sobrevivência atuarial com linhas suaves
ggplot() +
  geom_line(data = eaaControle_ext, aes(x = time, y = S_t_i, color = "Controle"), size =
    1.2) +
```

```

geom_line(data = eaaCasos_ext, aes(x = time, y = S_t_i, color = "Casos"), size = 1.2) +
geom_ribbon(
  data = eaaControle_ext,
  aes(x = time, ymin = lower, ymax = upper, fill = "Controle"), alpha = 0.2
) +
geom_ribbon(
  data = eaaCasos_ext,
  aes(x = time, ymin = lower, ymax = upper, fill = "Casos"), alpha = 0.2
) +
scale_color_manual(values = c("Controle" = "#1b9e77", "Casos" = "#d95f02")) +
scale_fill_manual(values = c("Controle" = "#1b9e77", "Casos" = "#d95f02")) +
labs(
  title = "Curvas de Sobrevivência de Hepatite - Estimador Atuarial",
  x = "Tempo (semanas)",
  y = "Probabilidade de Sobrevivência",
  color = "Grupo",
  fill = "Grupo"
) +
theme_minimal() +
scale_x_continuous(breaks = seq(0, 16, 1))

```



O gráfico de sobrevivência atuarial revela uma tendência semelhante à observada nas curvas de Kaplan-Meier, com o grupo controle apresentando uma taxa de sobrevivência mais alta ao longo do tempo. A curva do grupo controle inicia em 1 (100% de sobrevivência) e diminui gradualmente, enquanto a curva do grupo casos mostra uma queda mais acentuada, especialmente nas primeiras semanas. O intervalo de confiança para o grupo controle é relativamente estreito, indicando maior precisão na estimativa, enquanto o intervalo para o grupo casos é mais amplo, refletindo a incerteza crescente à medida que o número de indivíduos em risco

diminui.

Estimador de Nelson-Aalen

A seguir, aplicaremos o estimador de Nelson-Aalen, que se baseia na função de risco acumulado para estimar a sobrevivência. Frequentemente, este método produz resultados muito próximos aos do Kaplan-Meier, embora as estimativas de sobrevivência tendam a ser ligeiramente superiores, oferecendo uma perspectiva complementar para a comparação entre os grupos.

```
enaControle <- ena(
  time = hControle$time,
  evento = hControle$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
enaCasos <- ena(
  time = hCasos$time,
  evento = hCasos$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
```

```
print(enaControle)
```

```
##   time n_risk n_event censored   q_i   p_i S_t_i std_error lower upper
## 1     3     13        2         1 0.154 0.846 0.857    0.093 0.827 0.883
```

A análise com o estimador de Nelson-Aalen para o grupo controle indica uma taxa de sobrevivência de aproximadamente 0,857 na terceira semana, onde ocorreu o único evento. Este valor é ligeiramente superior ao estimado por Kaplan-Meier. Adicionalmente, o estimador de Nelson-Aalen resultou em um desvio padrão menor (0,093), o que se reflete em um intervalo de confiança mais estreito, de aproximadamente 0,827 a 0,883.

```
print(enaCasos)
```

```
##   time n_risk n_event censored   q_i   p_i S_t_i std_error lower upper
## 1     1     14        3         1 0.214 0.786 0.807    0.100 0.761 0.845
## 2     5      9        1         0 0.111 0.889 0.722    0.120 0.637 0.791
## 3     7      8        1         0 0.125 0.875 0.637    0.133 0.508 0.741
## 4     8      7        1         0 0.143 0.857 0.553    0.139 0.378 0.696
## 5    10      6        1         1 0.167 0.833 0.468    0.141 0.253 0.657
```

Podemos notar também que o estimação da função de sobrevida para Nelson-Aalen é ligeiramente maior que a probabilidade p_i , nota-se essa diferença para o primeiro tempo, onde o estimador de Nelson-Aalen é 0,802, enquanto p_1 é 0,786. Diferente do Kaplan-Meier e o estimador atuarial, onde a função de sobrevida é o produto dos p_i . Além disso, podemos ver que o estimador de Nelson-Aalen é mais suave, tendo um probabilidade de sobreviver além da décima semana de aproximadamente 0.468, mesmo assim corroborando para a observação de que menos da metade dos indivíduos sobreviveu até esse ponto. O intervalo de confiança para a décima semana é amplo (0,253 a 0,657), refletindo a crescente incerteza à medida que o número de indivíduos em risco diminui.

Abaixo conseguimos visualizar esse aumento da suavidade do estimador de Nelson-Aalen, e também o aumento do intervalo de confiança para o grupo de casos ao longo do tempo.

```
# Adiciona manualmente os pontos iniciais e finais
enaControle_ext <- enaControle %>%
```

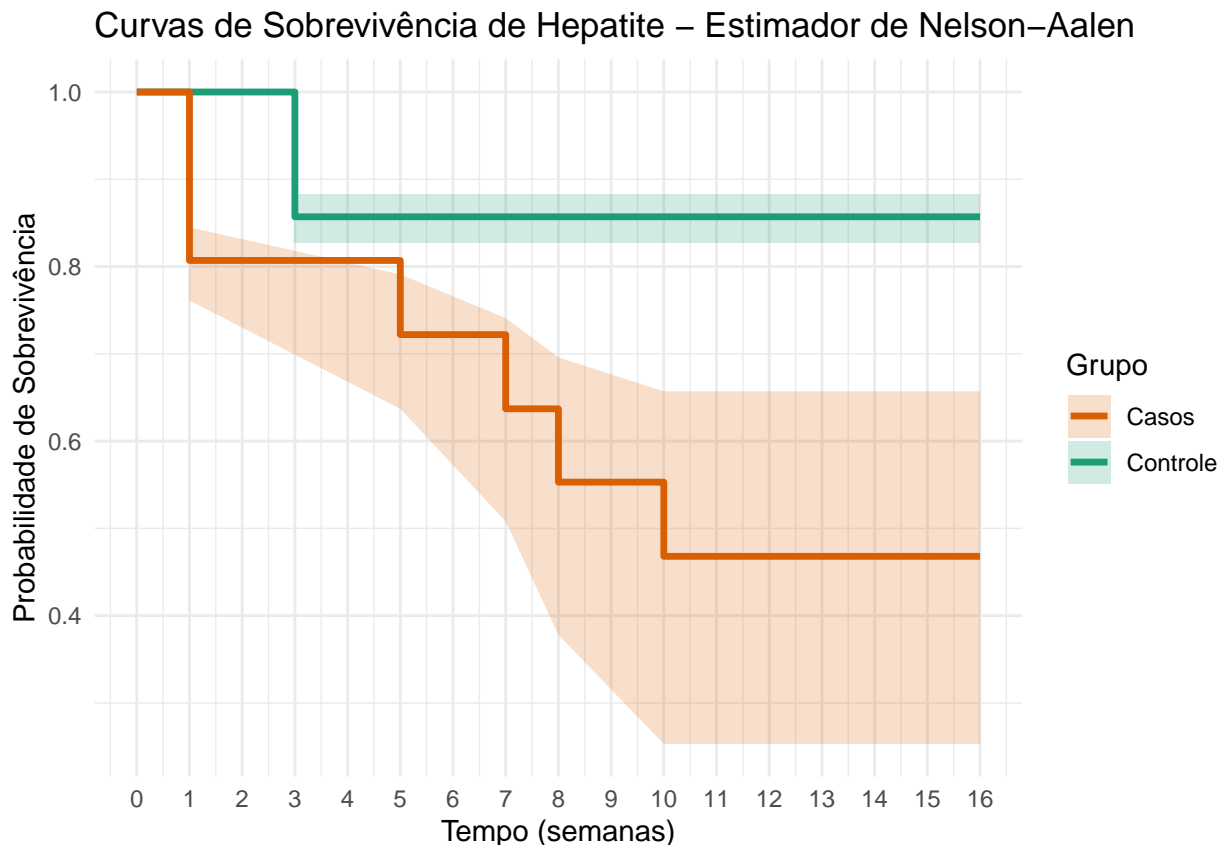
```

add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
add_row(
  time = 16,
  S_t_i = tail(enaControle$S_t_i, 1),
  lower = tail(enaControle$lower, 1),
  upper = tail(enaControle$upper, 1)
)

enaCasos_ext <- enaCasos %>%
add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
add_row(
  time = 16,
  S_t_i = tail(enaCasos$S_t_i, 1),
  lower = tail(enaCasos$lower, 1),
  upper = tail(enaCasos$upper, 1)
)

# Gráfico de escada Kaplan-Meier
ggplot() +
  geom_step(data = enaControle_ext, aes(x = time, y = S_t_i, color = "Controle"), size =
1.2) +
  geom_step(data = enaCasos_ext, aes(x = time, y = S_t_i, color = "Casos"), size = 1.2) +
  geom_ribbon(
    data = enaControle_ext,
    aes(x = time, ymin = lower, ymax = upper, fill = "Controle"), alpha = 0.2
  ) +
  geom_ribbon(
    data = enaCasos_ext,
    aes(x = time, ymin = lower, ymax = upper, fill = "Casos"), alpha = 0.2
  ) +
  scale_color_manual(values = c("Controle" = "#1b9e77", "Casos" = "#d95f02")) +
  scale_fill_manual(values = c("Controle" = "#1b9e77", "Casos" = "#d95f02")) +
  labs(
    title = "Curvas de Sobrevida de Hepatite - Estimador de Nelson-Aalen",
    x = "Tempo (semanas)",
    y = "Probabilidade de Sobrevida",
    color = "Grupo",
    fill = "Grupo"
  ) +
  theme_minimal() +
  scale_x_continuous(breaks = seq(0, 16, 1))

```



A visualização das curvas de Nelson-Aalen corrobora os achados dos métodos anteriores, confirmando o padrão de sobrevivência diferencial entre os grupos. A trajetória para o grupo controle demonstra um declínio suave, enquanto o grupo de casos experimenta uma redução mais abrupta na probabilidade de sobrevivência, principalmente no início do estudo. A precisão da estimativa, indicada pela largura da faixa de confiança, é notavelmente maior para o grupo controle. Em contrapartida, a faixa de confiança para o grupo de casos se alarga progressivamente, refletindo a maior incerteza associada à diminuição do número de indivíduos sob risco ao longo do tempo.

Teste Log-Rank

Para avaliar formalmente se a diferença observada entre as curvas de sobrevivência é estatisticamente significativa, empregaremos o teste log-rank. Este teste de hipóteses não paramétrico compara as distribuições de sobrevivência de dois ou mais grupos, testando a hipótese nula de que não há diferença entre elas ao longo de todo o período de acompanhamento.

```
logrank(hCasos, hControle)
```

```
##      Grupo Observado Esperado
## 1 Grupo 1          7 4.186763
## 2 Grupo 2          2 4.813237
##
## Variância total: 2.1578
## Estatística do teste (Chi²): 3.6677
## p-valor: 0.0555
```

Como pode ser observado, o Grupo 1 (casos) apresentou 7 eventos observados, enquanto o número esperado sob a hipótese de igualdade entre os grupos era de aproximadamente 4,18. Já o Grupo 2 (controle) teve apenas 2 eventos observados, com uma expectativa de cerca de 4,81 eventos. Isso indica que o grupo de casos

teve mais eventos do que o esperado, e o grupo controle, menos.

Apesar dessa diferença, a estatística do teste qui-quadrado foi de aproximadamente 3,67, resultando em um p-valor de 0,0555 (5,55%). Esse valor está ligeiramente acima do limiar convencional de 0,05, o que sugere que não há evidência estatística suficiente, ao nível de 5%, para rejeitar a hipótese nula de que as curvas de sobrevivência são iguais.

No entanto, como observado nas curvas estimadas pelos três métodos (Kaplan-Meier, Atuarial e Nelson-Aalen), a sobrevivência no grupo controle é consistentemente maior em todas as faixas de tempo. Com isso, pode-se adotar uma abordagem mais flexível e considerar a rejeição da hipótese nula em um nível de significância de 10%, reconhecendo uma tendência à diferença entre os grupos.

Malária

Esse estudo experimental foi conduzido com 44 camundongos, com o intuito de identificar a eficácia da imunização contra a malária. Os camundongos foram divididos em três grupos, cada um recebendo um tratamento diferente:

- **Grupo 1 (Controle):** Camundongos foram imunizados 3 dias antes da infecção pela esquistossomose.
- **Grupo 2:** Não especificado.
- **Grupo 3 (Tratamento B):** Só foi especificado que os camundongos foram infectados pela esquistossomose.

```
##### Malária #####
mG1 <- data.frame(
  time = c(7, 8, 8, 8, 8, 12, 12, 17, 18, 22, 30, 30, 30, 30, 30),
  evento = c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0),
  grupo = "Grupo 1"
)
mG2 <- data.frame(
  time = c(8, 8, 9, 10, 10, 14, 15, 15, 18, 19, 21, 22, 22, 23, 25),
  evento = c(rep(1, 15)),
  grupo = "Grupo 2"
)
mG3 <- data.frame(
  time = c(8, 8, 8, 8, 8, 9, 10, 10, 10, 11, 17, 19),
  evento = c(rep(1, 13)),
  grupo = "Grupo 3"
)
```

Como podemos ver acima, o tempo de sobrevivência dos camundongos foi registrado em dias, portanto, um gráfico de barras é a escolha ideal para a visualização da distribuição dos eventos ao longo do tempo entre os 3 grupos.

```
# Junta os três grupos
dadosMalaria <- bind_rows(mG1, mG2, mG3)

# Filtra apenas os eventos (evento == 1)
dadosEventos <- dadosMalaria %>% filter(evento == 1)

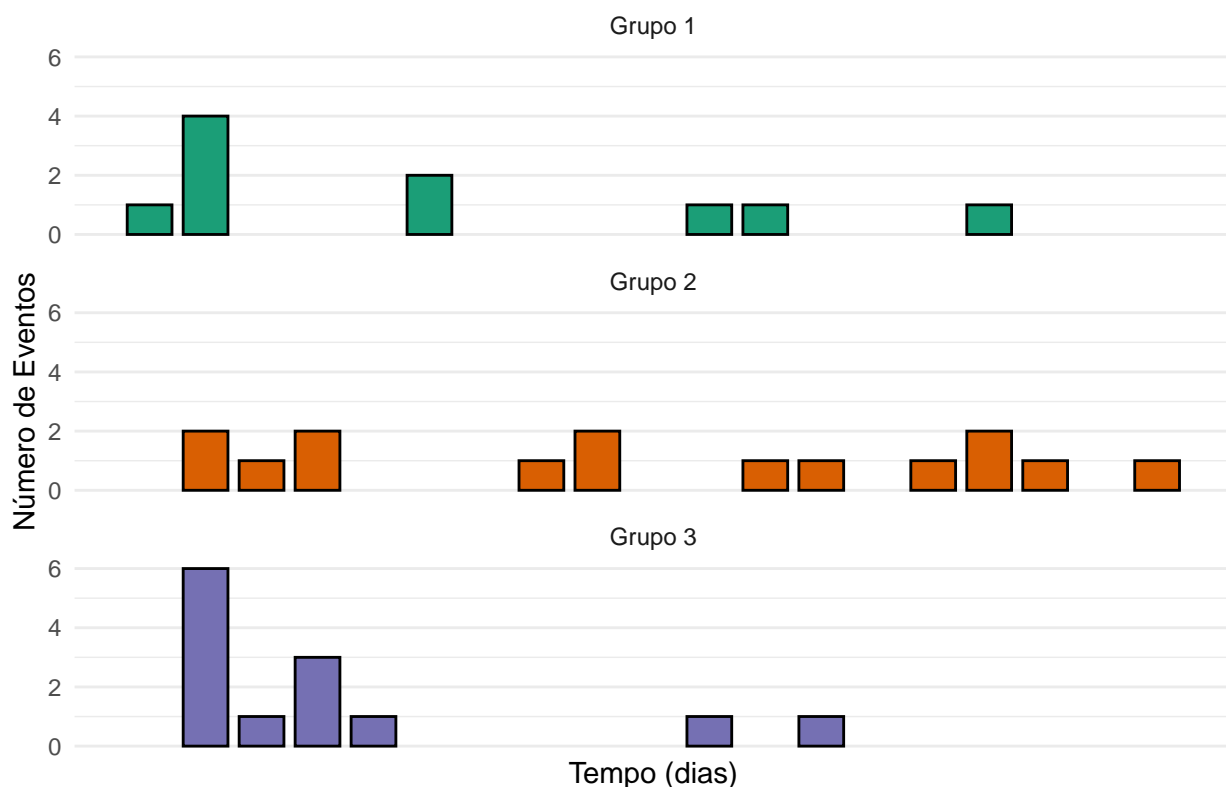
# Gráfico com facetas e cores diferentes
ggplot(dadosEventos, aes(x = time, fill = grupo)) +
  geom_bar(
    color = "black", width = 0.8
  ) +
```

```

facet_wrap(~grupo, ncol = 1) + # Mude para nrow = 1 para mostrar lado a lado
labs(
  title = "Frequência dos Tempos até o Evento - Malária",
  x = "Tempo (dias)",
  y = "Número de Eventos"
) +
scale_x_continuous(breaks = sort(unique(dadosEventos$tempo))) +
scale_fill_manual(
  values = c("Grupo 1" = "#1b9e77", "Grupo 2" = "#d95f02", "Grupo 3" = "#7570b3")
) +
theme_minimal() +
theme(legend.position = "none")

```

Frequência dos Tempos até o Evento – Malária



```

print(paste("Número de camundongos no Grupo 1:", length(mG1$tempo), "| Número de falhas:", sum(mG1$evento)))

```

```
## [1] "Número de camundongos no Grupo 1: 0 | Número de falhas: 10"
```

```

print(paste("Número de camundongos no Grupo 2:", length(mG2$tempo), "| Número de falhas:", sum(mG2$evento)))

```

```
## [1] "Número de camundongos no Grupo 2: 0 | Número de falhas: 15"
```

```

print(paste("Número de camundongos no Grupo 3:", length(mG3$tempo), "| Número de falhas:", sum(mG3$evento)))

```

```
## [1] "Número de camundongos no Grupo 3: 0 | Número de falhas: 13"
```

A análise da distribuição de eventos revela desfechos distintos para cada grupo. Notavelmente, os grupos 2 e 3 apresentaram 100% de mortalidade, com todos os 15 e 13 camundongos, respectivamente, vindo a óbito durante o estudo. Em contrapartida, o Grupo 1, que recebeu imunização prévia, registrou 10 falhas em 16 camundongos, indicando que parte do grupo sobreviveu até o final do período de observação.

As falhas no Grupo 1 se concentraram principalmente no dia 8 (4 eventos), enquanto o Grupo 3 apresentou um pico de mortalidade ainda mais acentuado nesse mesmo dia (6 eventos) e no dia 10 (3 eventos). Já o Grupo 2 exibiu uma distribuição de falhas mais dispersa ao longo do tempo. Esses resultados preliminares sugerem que a imunização prévia conferiu ao Grupo 1 uma vantagem em termos de sobrevivência em relação aos demais.

A seguir, serão estimadas as curvas de sobrevivência para cada grupo utilizando os estimadores de Kaplan-Meier e Nelson-Aalen. Como as censuras ocorreram apenas após a ocorrência de todos os eventos em cada grupo, o estimador atuarial coincide com o de Kaplan-Meier nesse caso. Por fim, será aplicado o teste log-rank para comparar as curvas de sobrevivência entre os grupos.

Estimador de Kaplan-Meier

```
ekmMalariaG1 <- ekm(
  time = mG1$time,
  evento = mG1$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
ekmMalariaG2 <- ekm(
  time = mG2$time,
  evento = mG2$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
ekmMalariaG3 <- ekm(
  time = mG3$time,
  evento = mG3$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
```

```
print(ekmMalariaG1)
```

##	time	n_risk	n_event	censored	q_i	p_i	S_t_i	std_error	lower	upper
## 1	7	16	1	0	0.062	0.938	0.938	0.061	0.929	0.945
## 2	8	15	4	0	0.267	0.733	0.688	0.116	0.594	0.764
## 3	12	11	2	0	0.182	0.818	0.562	0.124	0.412	0.688
## 4	17	9	1	0	0.111	0.889	0.500	0.125	0.323	0.654
## 5	18	8	1	0	0.125	0.875	0.438	0.124	0.237	0.622
## 6	22	7	1	0	0.143	0.857	0.375	0.121	0.158	0.594

A análise do Grupo 1 (imunizado) revela uma alta taxa de sobrevida inicial (0,938) até o dia 7, com um intervalo de confiança estreito (0,929 a 0,945). Contudo, uma queda acentuada ocorre no dia 8, onde se concentra o maior número de falhas, reduzindo a probabilidade de sobrevivência para 0,688. A mediana de sobrevivência para este grupo é alcançada no dia 17, quando a taxa atinge 0,500, indicando que metade dos camundongos imunizados sobreviveu além deste ponto. A última falha registrada ocorre no dia 22, com

uma sobrevida estimada de 0,385 e um intervalo de confiança mais amplo (0,158 a 0,594), refletindo a maior incerteza devido ao menor número de indivíduos em risco.

```
print(ekmMalariaG2)
```

##	time	n_risk	n_event	censored	q_i	p_i	S_t_i	std_error	lower	upper
## 1	8	15	2	0	0.133	0.867	0.867	0.088	0.840	0.889
## 2	9	13	1	0	0.077	0.923	0.800	0.103	0.750	0.841
## 3	10	12	2	0	0.167	0.833	0.667	0.122	0.560	0.753
## 4	14	10	1	0	0.100	0.900	0.600	0.126	0.462	0.713
## 5	15	9	2	0	0.222	0.778	0.467	0.129	0.270	0.642
## 6	18	7	1	0	0.143	0.857	0.400	0.126	0.182	0.611
## 7	19	6	1	0	0.167	0.833	0.333	0.122	0.106	0.584
## 8	21	5	1	0	0.200	0.800	0.267	0.114	0.047	0.565
## 9	22	4	2	0	0.500	0.500	0.133	0.088	0.001	0.574
## 10	23	2	1	0	0.500	0.500	0.067	0.064	0.000	0.665
## 11	25	1	1	0	1.000	0.000	0.000	0.000	NA	NA

O Grupo 2, por sua vez, exibe uma dispersão de falhas mais gradual ao longo do tempo. A sobrevida inicial no dia 8 é de 0,867, mas decai progressivamente até que todos os camundongos venham a óbito no dia 25. A mediana de sobrevivência deste grupo é atingida no dia 15 (sobrevida de 0,467), dois dias antes do Grupo 1, indicando uma mortalidade geral mais rápida, apesar do declínio menos abrupto.

```
print(ekmMalariaG3)
```

##	time	n_risk	n_event	censored	q_i	p_i	S_t_i	std_error	lower	upper
## 1	8	13	6	0	0.462	0.538	0.538	0.138	0.359	0.688
## 2	9	7	1	0	0.143	0.857	0.462	0.138	0.249	0.651
## 3	10	6	3	0	0.500	0.500	0.231	0.117	0.019	0.581
## 4	11	3	1	0	0.333	0.667	0.154	0.100	0.001	0.593
## 5	17	2	1	0	0.500	0.500	0.077	0.074	0.000	0.677
## 6	19	1	1	0	1.000	0.000	0.000	0.000	NA	NA

O Grupo 3 demonstrou ser o mais vulnerável, sofrendo um impacto severo e imediato. Já no dia 8, a probabilidade de sobrevivência despenca para 0,538, com quase metade dos camundongos falhando. A mortalidade continua de forma acelerada, e no dia 19, todos os indivíduos do grupo já haviam morrido, destacando este como o pior desfecho entre os três.

Para visualizar e comparar formalmente essas trajetórias, as curvas de sobrevivência de Kaplan-Meier para os três grupos são plotadas a seguir.

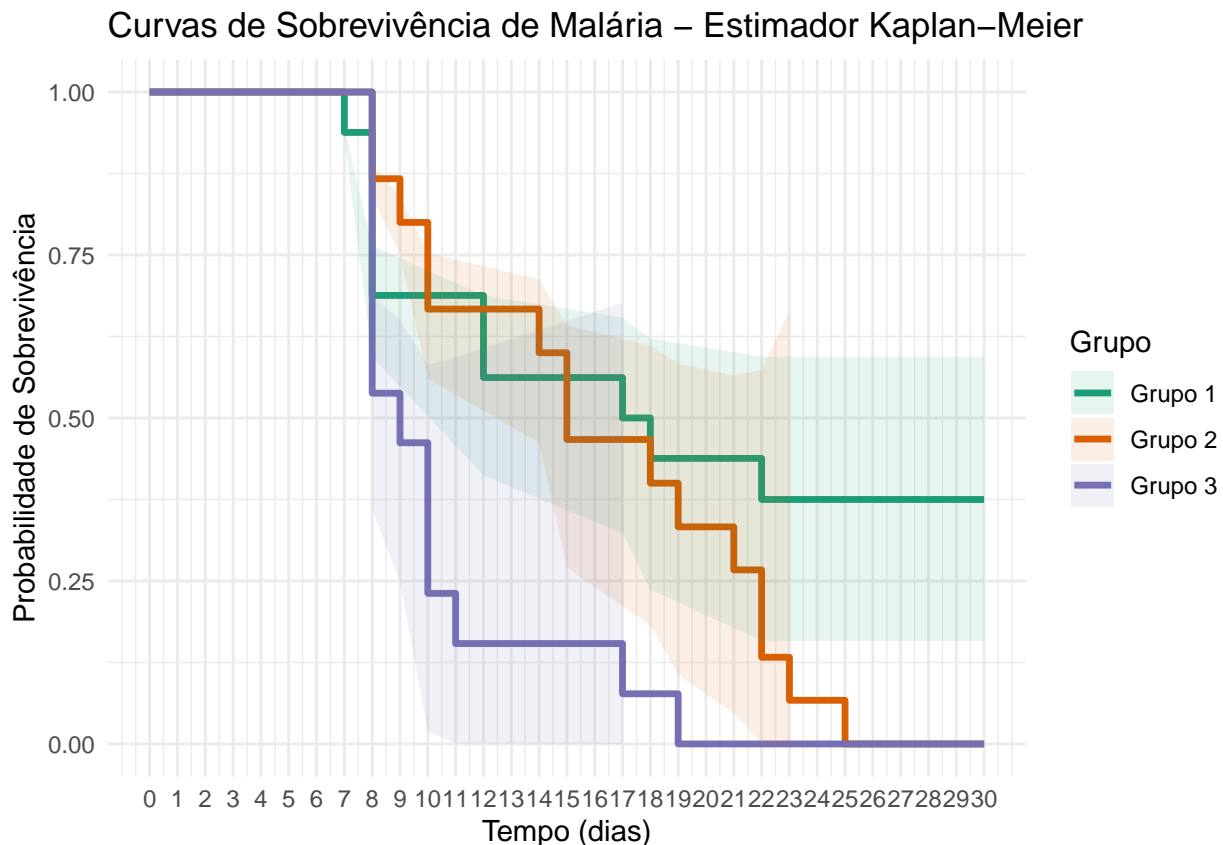
```
# Adiciona manualmente os pontos iniciais e finais
ekmMalariaG1_ext <- ekmMalariaG1 %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 30,
    S_t_i = tail(ekmMalariaG1$S_t_i, 1),
    lower = tail(ekmMalariaG1$lower, 1),
    upper = tail(ekmMalariaG1$upper, 1)
  )
ekmMalariaG2_ext <- ekmMalariaG2 %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 30,
    S_t_i = tail(ekmMalariaG2$S_t_i, 1),
    lower = tail(ekmMalariaG2$lower, 1),
    upper = tail(ekmMalariaG2$upper, 1)
  )
```

```

)
ekmMalariaG3_ext <- ekmMalariaG3 %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 30,
    S_t_i = tail(ekmMalariaG3$S_t_i, 1),
    lower = tail(ekmMalariaG3$lower, 1),
    upper = tail(ekmMalariaG3$upper, 1)
  )

ggplot() +
  geom_step(data = ekmMalariaG1_ext, aes(x = time, y = S_t_i, color = "Grupo 1"), size =
1.2) +
  geom_step(data = ekmMalariaG2_ext, aes(x = time, y = S_t_i, color = "Grupo 2"), size =
1.2) +
  geom_step(data = ekmMalariaG3_ext, aes(x = time, y = S_t_i, color = "Grupo 3"), size =
1.2) +
  geom_ribbon(
    data = ekmMalariaG1_ext,
    aes(x = time, ymin = lower, ymax = upper, fill = "Grupo 1"), alpha = 0.1
  ) +
  geom_ribbon(
    data = ekmMalariaG2_ext,
    aes(x = time, ymin = lower, ymax = upper, fill = "Grupo 2"), alpha = 0.1
  ) +
  geom_ribbon(
    data = ekmMalariaG3_ext,
    aes(x = time, ymin = lower, ymax = upper, fill = "Grupo 3"), alpha = 0.1
  ) +
  scale_color_manual(values = c("Grupo 1" = "#1b9e77", "Grupo 2" = "#d95f02", "Grupo 3" =
"#7570b3")) +
  scale_fill_manual(values = c("Grupo 1" = "#1b9e77", "Grupo 2" = "#d95f02", "Grupo 3" =
"#7570b3")) +
  labs(
    title = "Curvas de Sobrevida de Malária - Estimador Kaplan-Meier",
    x = "Tempo (dias)",
    y = "Probabilidade de Sobrevida",
    color = "Grupo",
    fill = "Grupo"
  ) +
  theme_minimal() +
  scale_x_continuous(breaks = seq(0, 30, 1))

```



O gráfico de Kaplan-Meier ilustra vividamente as distintas trajetórias de sobrevivência. A curva do Grupo 3 se destaca pela queda drástica e precoce, confirmando seu desfecho desfavorável. Em contraste, os Grupos 1 e 2 apresentam curvas inicialmente próximas, mas suas trajetórias divergem a partir do dia 18. Nesse ponto, o Grupo 2 entra em um declínio contínuo até a mortalidade total no dia 25. O Grupo 1, por outro lado, mantém uma taxa de sobrevivência mais estável, com a última falha ocorrendo no dia 22 e uma sobrevida final de aproximadamente 38,5%.

Para validar estatisticamente as diferenças observadas visualmente, daremos continuidade à análise utilizando o estimador de Nelson-Aalen. Em seguida, aplicaremos o teste de Log-Rank para comparações par a par entre os grupos, com correção de Bonferroni, a fim de controlar o aumento do risco de erro tipo I decorrente das múltiplas comparações.

Estimador de Nelson-Aalen

```
enaMalariaG1 <- ena(
  time = mG1$time,
  evento = mG1$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
enaMalariaG2 <- ena(
  time = mG2$time,
  evento = mG2$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
```

```
)
enaMalariaG3 <- ena(
  time = mG3$time,
  evento = mG3$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
```

```
print(enaMalariaG1)
```

```
##   time n_risk n_event censored   q_i   p_i S_t_i std_error lower upper
## 1    7     16      1         0 0.062 0.938 0.939    0.059 0.932 0.946
## 2    8     15      4         0 0.267 0.733 0.720    0.106 0.644 0.781
## 3   12     11      2         0 0.182 0.818 0.600    0.117 0.473 0.706
## 4   17      9      1         0 0.111 0.889 0.537    0.121 0.380 0.670
## 5   18      8      1         0 0.125 0.875 0.474    0.122 0.290 0.637
## 6   22      7      1         0 0.143 0.857 0.411    0.121 0.205 0.607
```

A aplicação do estimador de Nelson-Aalen confirma as tendências gerais observadas, apresentando estimativas de sobrevivência ligeiramente mais otimistas que as do Kaplan-Meier. Por exemplo, no Grupo 1, a sobrevida estimada pelo Nelson-Aalen no dia 17 é de 0,537, enquanto pelo Kaplan-Meier é de 0,500.

```
print(enaMalariaG2)
```

```
##   time n_risk n_event censored   q_i   p_i S_t_i std_error lower upper
## 1     8     15      2         0 0.133 0.867 0.875    0.083 0.852 0.895
## 2     9     13      1         0 0.077 0.923 0.810    0.099 0.766 0.847
## 3    10     12      2         0 0.167 0.833 0.686    0.116 0.591 0.763
## 4    14     10      1         0 0.100 0.900 0.621    0.122 0.496 0.723
## 5    15      9      2         0 0.222 0.778 0.497    0.125 0.318 0.653
## 6    18      7      1         0 0.143 0.857 0.431    0.125 0.227 0.620
## 7    19      6      1         0 0.167 0.833 0.365    0.122 0.144 0.592
## 8    21      5      1         0 0.200 0.800 0.299    0.116 0.075 0.569
## 9    22      4      2         0 0.500 0.500 0.181    0.095 0.008 0.544
## 10   23      2      1         0 0.500 0.500 0.110    0.080 0.000 0.587
## 11   25      1      1         0 1.000 0.000 0.040    0.050 0.000 0.752
```

No Grupo 2, a sobrevida estimada no dia 15 é de 0,497 pelo método de Nelson-Aalen, contra 0,467 pelo Kaplan-Meier. Vale destacar que o estimador de Nelson-Aalen nunca atinge zero, mesmo quando todos os indivíduos experimentam o evento, característica que resulta na manutenção de uma probabilidade de sobrevivência em torno de 0,497. Contudo, esse estimador apresenta intervalos de confiança mais amplos, refletindo a crescente incerteza à medida que o número de indivíduos em risco diminui.

```
print(enaMalariaG3)
```

```
##   time n_risk n_event censored   q_i   p_i S_t_i std_error lower upper
## 1     8     13      6         0 0.462 0.538 0.630    0.119 0.513 0.727
## 2     9      7      1         0 0.143 0.857 0.546    0.129 0.383 0.684
## 3    10      6      3         0 0.500 0.500 0.331    0.124 0.101 0.588
## 4    11      3      1         0 0.333 0.667 0.237    0.119 0.022 0.583
## 5    17      2      1         0 0.500 0.500 0.144    0.102 0.000 0.616
## 6    19      1      1         0 1.000 0.000 0.053    0.065 0.000 0.766
```

No Grupo 3, a diferença também é notável no dia 9, com uma sobrevida estimada de 0,546 pelo Nelson-Aalen, em comparação com 0,538 pelo Kaplan-Meier. Apesar dessas pequenas variações numéricas, a interpretação

geral dos resultados se mantém: o Grupo 1 apresenta melhor sobrevida, o Grupo 2 tem mortalidade intermediária, e o Grupo 3 mostra o desfecho mais rápido. A seguir, serão plotadas as curvas de sobrevivência de Nelson-Aalen para facilitar a comparação visual

```
# Adiciona manualmente os pontos iniciais e finais
enaMalariaG1_ext <- enaMalariaG1 %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 30,
    S_t_i = tail(enaMalariaG1$S_t_i, 1),
    lower = tail(enaMalariaG1$lower, 1),
    upper = tail(enaMalariaG1$upper, 1)
  )
enaMalariaG2_ext <- enaMalariaG2 %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 30,
    S_t_i = tail(enaMalariaG2$S_t_i, 1),
    lower = tail(enaMalariaG2$lower, 1),
    upper = tail(enaMalariaG2$upper, 1)
  )
enaMalariaG3_ext <- enaMalariaG3 %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 30,
    S_t_i = tail(enaMalariaG3$S_t_i, 1),
    lower = tail(enaMalariaG3$lower, 1),
    upper = tail(enaMalariaG3$upper, 1)
  )
ggplot() +
  geom_step(data = enaMalariaG1_ext, aes(x = time, y = S_t_i, color = "Grupo 1"), size =
1.2) +
  geom_step(data = enaMalariaG2_ext, aes(x = time, y = S_t_i, color = "Grupo 2"), size =
1.2) +
  geom_step(data = enaMalariaG3_ext, aes(x = time, y = S_t_i, color = "Grupo 3"), size =
1.2) +
  geom_ribbon(
    data = enaMalariaG1_ext,
    aes(x = time, ymin = lower, ymax = upper, fill = "Grupo 1"), alpha = 0.1
  ) +
  geom_ribbon(
    data = enaMalariaG2_ext,
    aes(x = time, ymin = lower, ymax = upper, fill = "Grupo 2"), alpha = 0.1
  ) +
  geom_ribbon(
    data = enaMalariaG3_ext,
    aes(x = time, ymin = lower, ymax = upper, fill = "Grupo 3"), alpha = 0.1
  ) +
  scale_color_manual(values = c("Grupo 1" = "#1b9e77", "Grupo 2" = "#d95f02", "Grupo 3" =
"#7570b3")) +
  scale_fill_manual(values = c("Grupo 1" = "#1b9e77", "Grupo 2" = "#d95f02", "Grupo 3" =
"#7570b3")) +
  labs(
    title = "Curvas de Sobrevivência de Malária - Estimador de Nelson-Aalen",
```

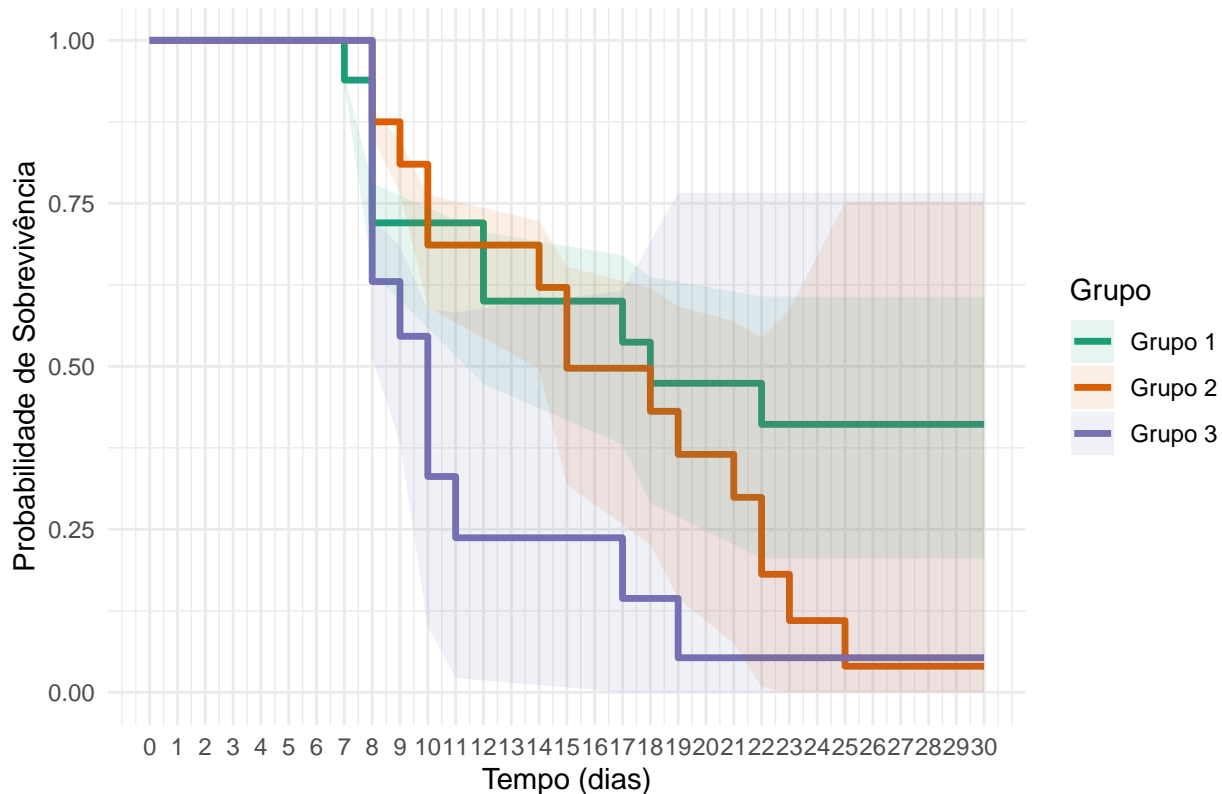


```

x = "Tempo (dias)",
y = "Probabilidade de Sobrevida",
color = "Grupo",
fill = "Grupo"
) +
theme_minimal() +
scale_x_continuous(breaks = seq(0, 30, 1))

```

Curvas de Sobrevida de Malária – Estimador de Nelson–Aalen



Podemos observar no gráfico acima que o estimador de Nelson-Aalen segue a mesma tendência apresentada pelo Kaplan-Meier, embora apresente intervalos de confiança mais amplos para os três grupos.

Teste Log-Rank

Como mencionado anteriormente, o teste log-rank será utilizado para comparar as curvas de sobrevivência entre os grupos. Dado que estamos realizando múltiplas comparações (Grupo 1 vs Grupo 2, Grupo 1 vs Grupo 3, e Grupo 2 vs Grupo 3), aplicaremos a correção de Bonferroni para ajustar o nível de significância e controlar o risco de erro tipo I.

```

# Correção de Bonferroni
print(round(0.05 / 3, 3))

```

```
## [1] 0.017
```

```
print(round(0.10 / 3, 3))
```

```
## [1] 0.033
```

Para um nível de significância de 5%, o valor ajustado é aproximadamente 0,017, e para 10%, é cerca de 0,033.

Esses valores serão usados como novos limiares para determinar a significância estatística nas comparações par a par entre os grupos.

Grupo 1 vs Grupo 2

```
logrank(mG1, mG2)
```

```
##      Grupo Observado Esperado
## 1 Grupo 1          10 13.71948
## 2 Grupo 2          15 11.28052
##
## Variância total: 5.4748
## Estatística do teste (Chi²): 2.527
## p-valor: 0.1119
```

Como mostrado acima, o Grupo 1 (imunizado) apresentou 10 falhas observadas, contra uma expectativa de aproximadamente 13,72 falhas sob a hipótese nula de igualdade entre os grupos. O Grupo 2 (não especificado) teve 15 falhas observadas, com uma expectativa de cerca de 11,28. Isso indica que o grupo imunizado teve menos falhas que o esperado, enquanto o grupo não especificado apresentou mais. No entanto, o p-valor foi de aproximadamente 0,112 (11,2%), superior aos limiares ajustados de 0,017 (5%) e 0,033 (10%), o que significa que não há evidência estatística suficiente para rejeitar a hipótese nula de igualdade das curvas de sobrevivência entre esses grupos.

Grupo 1 vs Grupo 3

```
logrank(mG1, mG3)
```

```
##      Grupo Observado Esperado
## 1 Grupo 1          10 15.339429
## 2 Grupo 2          13  7.660571
##
## Variância total: 3.6256
## Estatística do teste (Chi²): 7.8635
## p-valor: 0.005
```

Na tabela acima, o teste comparou os Grupos 1 e 3, com expectativas médias de falhas de aproximadamente 15,34 e 7,66, respectivamente. O Grupo 1 apresentou menos falhas que o esperado, enquanto o Grupo 3 teve mais. O p-valor de 0,005 (0,5%) está abaixo do limiar ajustado de 0,017 (5%), indicando evidência estatística suficiente para rejeitar a hipótese nula. Isso demonstra uma diferença significativa nas curvas de sobrevivência entre o grupo imunizado e o grupo infectado por esquistossomose sem imunização prévia.

Grupo 2 vs Grupo 3

```
logrank(mG2, mG3)
```

```
##      Grupo Observado Esperado
## 1 Grupo 1          14 19.525039
## 2 Grupo 2          13  7.474961
##
## Variância total: 3.8251
## Estatística do teste (Chi²): 7.9804
## p-valor: 0.0047
```

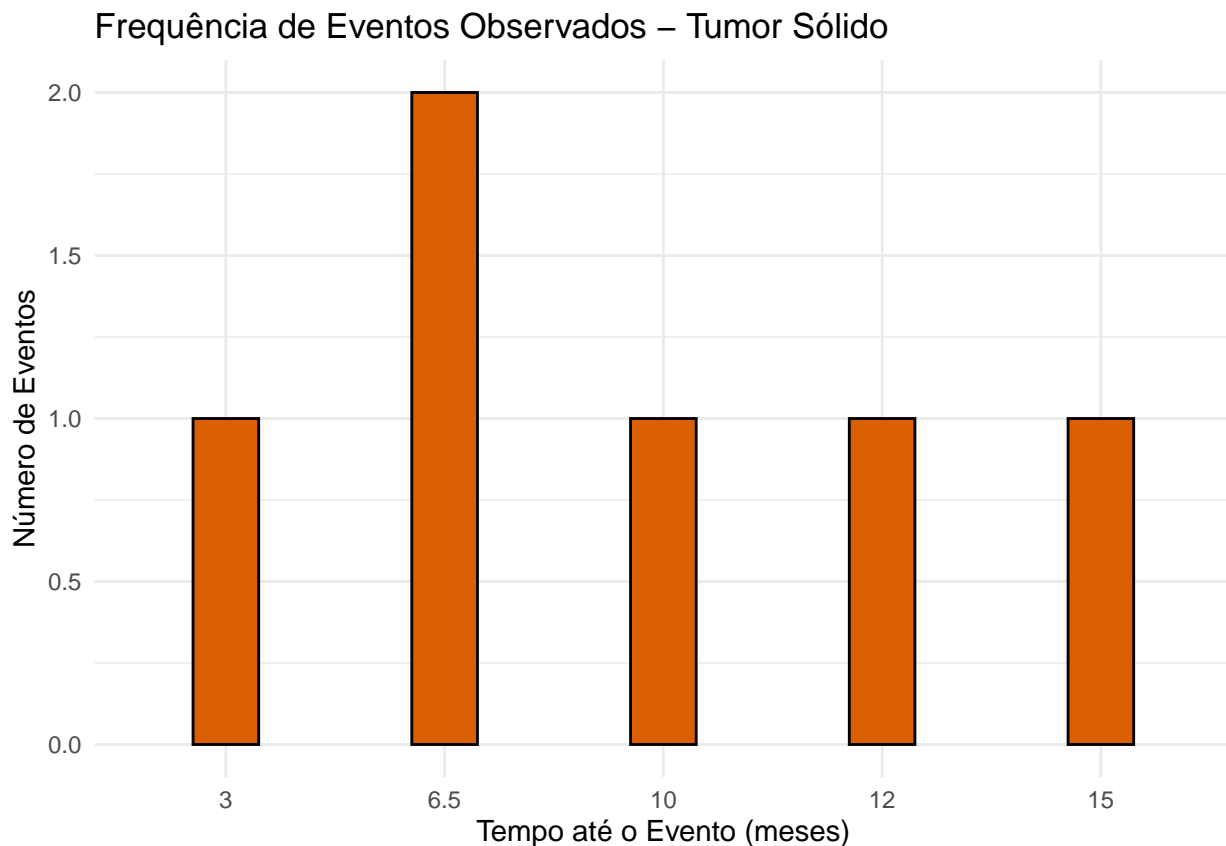
No teste entre os Grupos 2 e 3, o Grupo 2 apresentou 15 falhas observadas, com uma expectativa de aproximadamente 19,52 falhas sob a hipótese nula. Já o Grupo 3 teve 13 falhas observadas, contra uma expectativa de cerca de 7. Isso indica que o Grupo 2 teve mais falhas que o esperado, enquanto o Grupo 3 teve menos. O p-valor foi de aproximadamente 0,0047 (0,47%), bem abaixo do limiar ajustado de 0,017 (5%), evidenciando que a infecção pela esquistossomose sem imunização prévia (Grupo 3) está associada a uma sobrevivência significativamente menor.

Reincidência de tumor sólido

O presente estudo visa informar a respeito do tempo de reincidência de 10 pacientes com tumor sólido, tendo uma censura aleatória no paciente que foi acompanhado durante 8.4 meses e três pacientes que sofreram censura a direita com 4, 5.7 e 10 meses de acompanhamento. A baixo podemos ver o gráfico de barras para observar a distribuição da reincidência de tumor sólido ao longo dos meses.

```
##### Reincidência de tumor sólido #####
tumorSolido <- data.frame(
  time = c(3, 4, 5.7, 6.5, 6.5, 8.4, 10, 10, 12, 15),
  evento = c(1, 0, 0, 1, 1, 0, 1, 0, 1, 1)
)

# Gráfico com facetas e cores diferentes
tumorSolido %>%
  filter(evento == 1) %>%
  ggplot(aes(x = factor(time))) +
  geom_bar(fill = "#d95f02", color = "black", width = 0.3) +
  labs(
    title = "Frequência de Eventos Observados - Tumor Sólido",
    x = "Tempo até o Evento (meses)",
    y = "Número de Eventos"
  ) +
  theme_minimal()
```



O gráfico de barras ilustra a distribuição temporal das reincidências, com eventos únicos ocorrendo nos meses 3, 10, 12 e 15, e um pico de duas reincidências no mês 6.5. Para analisar a dinâmica da probabilidade de sobrevida livre de reincidência, a seguir serão aplicados os estimadores de Kaplan-Meier, Atuarial e

Nelson-Aalen.

Estimador de Kaplan-Meier

```
# Estimador de Kaplan-Meier
ekmTumor <- ekm(
  time = tumorSolido$time,
  evento = tumorSolido$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
```

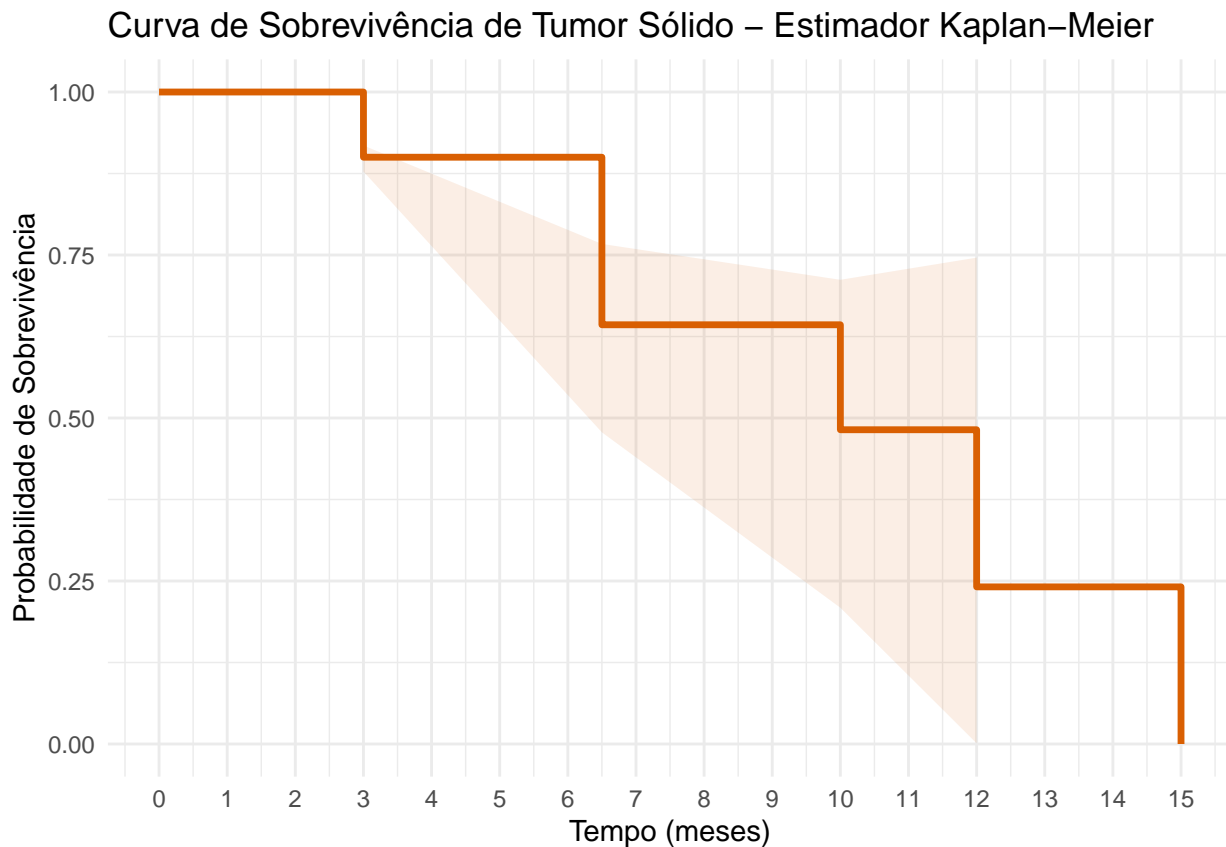
```
print(ekmTumor)
```

```
##   time n_risk n_event censored   q_i   p_i S_t_i std_error lower upper
## 1  3.0     10      1         0 0.100 0.900 0.900    0.095 0.878 0.918
## 2  6.5      7      2         0 0.286 0.714 0.643    0.168 0.478 0.767
## 3 10.0      4      1         1 0.250 0.750 0.482    0.188 0.209 0.712
## 4 12.0      2      1         0 0.500 0.500 0.241    0.195 0.001 0.746
## 5 15.0      1      1         0 1.000 0.000 0.000    0.000   NA   NA
```

A análise da tabela de Kaplan-Meier revela que, no terceiro mês, a probabilidade de um paciente permanecer livre de reincidência é de 90%. Contudo, observa-se uma queda acentuada no mês 6.5, quando a probabilidade de sobrevivência diminui para 0,643, reflexo da ocorrência de duas reincidências. O alargamento do intervalo de confiança (0,478 a 0,767) neste ponto, e ao longo dos meses subsequentes, evidencia a crescente incerteza nas estimativas, uma consequência direta do pequeno número de pacientes que permanecem em risco.

Para uma melhor visualização da evolução da probabilidade de reincidência ao longo do tempo, a curva de sobrevivência de Kaplan-Meier será plotada a seguir.

```
# Adiciona manualmente os pontos iniciais e finais
ekmTumor_ext <- ekmTumor %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 15,
    S_t_i = tail(ekmTumor$S_t_i, 1),
    lower = tail(ekmTumor$lower, 1),
    upper = tail(ekmTumor$upper, 1)
  )
ggplot() +
  geom_step(data = ekmTumor_ext, aes(x = time, y = S_t_i), color = "#d95f02", size = 1.2) +
  geom_ribbon(
    data = ekmTumor_ext,
    aes(x = time, ymin = lower, ymax = upper), fill = "#d95f02", alpha = 0.1
  ) +
  labs(
    title = "Curva de Sobrevivência de Tumor Sólido - Estimador Kaplan-Meier",
    x = "Tempo (meses)",
    y = "Probabilidade de Sobrevivência"
  ) +
  theme_minimal() +
  scale_x_continuous(breaks = seq(0, 16, 1))
```



O gráfico de Kaplan-Meier ilustra a queda progressiva na probabilidade de sobrevivência livre de reincidência ao longo do tempo. No marco de 12 meses, a probabilidade estimada de um paciente permanecer sem reincidência é de aproximadamente 24,1%. Contudo, a precisão desta estimativa é baixa, como evidenciado pelo intervalo de confiança extremamente amplo, que vai de 0,1% a 74,6%. Essa grande incerteza reflete o pequeno número de pacientes ainda em risco nos estágios mais avançados do acompanhamento.

Estimador Atuarial

```
# Estimador Atuarial
eaTumor <- eaa(
  time = tumorSolido$time,
  evento = tumorSolido$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
```

```
print(eaTumor)
```

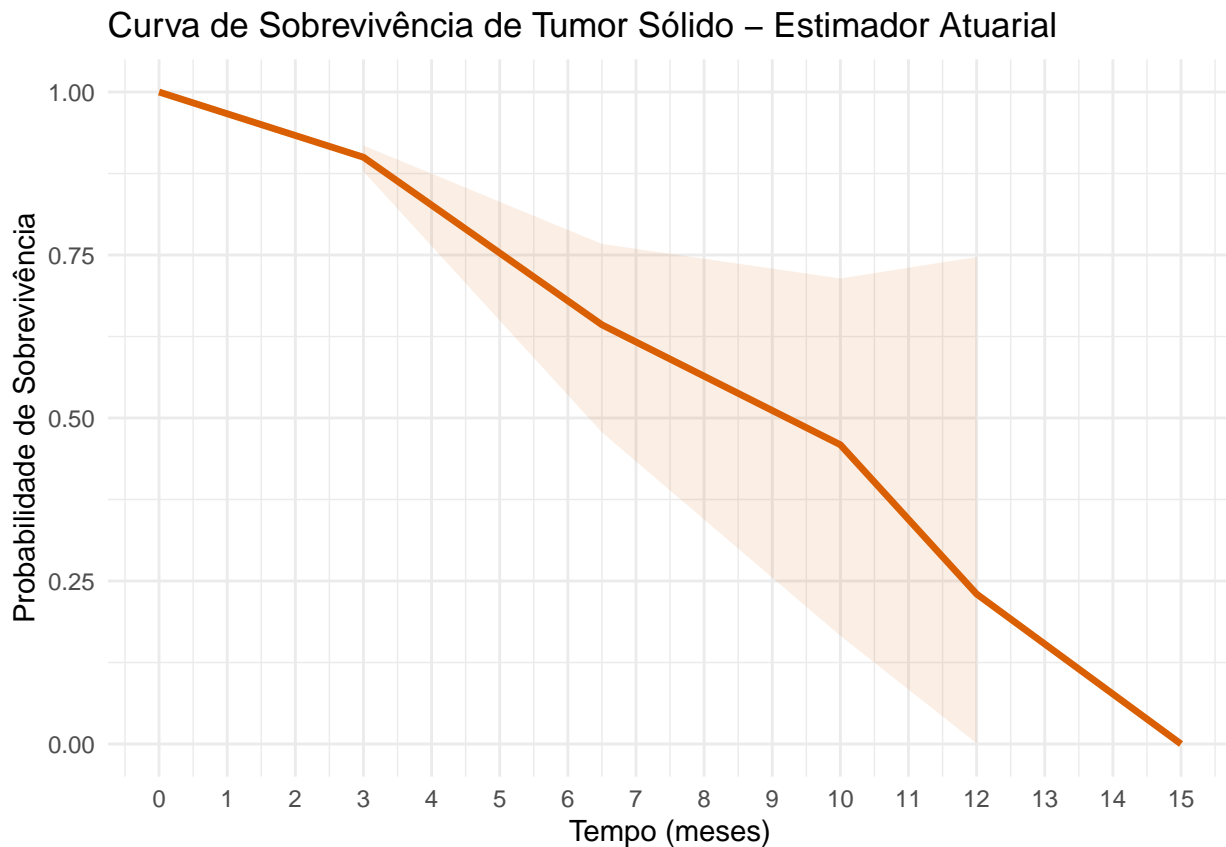
```
##   time n_risk n_event censored n_risk_adj  q_i  p_i S_t_i std_error lower
## 1  3.0    10      1        0      10.0 0.100 0.900 0.900    0.095 0.878
## 2  6.5     7      2        0       7.0 0.286 0.714 0.643    0.168 0.478
## 3 10.0     4      1        1       3.5 0.286 0.714 0.459    0.196 0.166
## 4 12.0     2      1        0       2.0 0.500 0.500 0.230    0.190 0.001
## 5 15.0     1      1        0       1.0 1.000 0.000 0.000      NaN   NA
##   upper
## 1 0.918
```

```
## 2 0.767
## 3 0.714
## 4 0.747
## 5    NA
```

Pelo fato de existir apenas uma censura aleatoria durante o estudo e nenhuma censura a esquerda, o estimador atuarial coincide com o Kaplan-Meier, tendo uma pequena diferença de na probabilidade de sobrevivencia no decimo e decimo segundo mês, sendo de 0,459 contra 0,482 no decimo mês e 0,230 e 0,241 no decimo segundo mês respectivamente.

Abaixo podemos ver a curva de sobrevivência Atuarial

```
# Adiciona manualmente os pontos iniciais e finais
eaTumor_ext <- eaTumor %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 15,
    S_t_i = tail(eaTumor$S_t_i, 1),
    lower = tail(eaTumor$lower, 1),
    upper = tail(eaTumor$upper, 1)
  )
ggplot() +
  geom_line(data = eaTumor_ext, aes(x = time, y = S_t_i), color = "#d95f02", size = 1.2) +
  geom_ribbon(
    data = eaTumor_ext,
    aes(x = time, ymin = lower, ymax = upper), fill = "#d95f02", alpha = 0.1
  ) +
  labs(
    title = "Curva de Sobrevivência de Tumor Sólido - Estimador Atuarial",
    x = "Tempo (meses)",
    y = "Probabilidade de Sobrevivência"
  ) +
  theme_minimal() +
  scale_x_continuous(breaks = seq(0, 16, 1))
```



Podemos verificar que o gráfico do estimador atuarial é praticamente idêntico ao de Kaplan-Meier, sendo estimado que um quarto da população de pacientes estará livre de reincidência após 12 meses, com um intervalo de confiança bastante amplo (0,1% a 74,7%), refletindo a incerteza associada ao pequeno número de pacientes em risco.

Estimador de Nelson-Aalen

```
# Estimador de Nelson-Aalen
enaTumor <- ena(
  time = tumorSolido$time,
  evento = tumorSolido$evento,
  classEvento = 1,
  alpha = 0.05,
  conf = "log-log"
)
```

```
print(enaTumor)
```

##	time	n_risk	n_event	censored	q_i	p_i	S_t_i	std_error	lower	upper
## 1	3.0	10	1	0	0.100	0.900	0.905	0.090	0.885	0.921
## 2	6.5	7	2	0	0.286	0.714	0.680	0.153	0.549	0.780
## 3	10.0	4	1	1	0.250	0.750	0.530	0.178	0.292	0.720
## 4	12.0	2	1	0	0.500	0.500	0.321	0.194	0.025	0.706
## 5	15.0	1	1	0	1.000	0.000	0.118	0.138	0.000	0.805

O estimador de Nelson-Aalen, caracterizado por sua abordagem baseada na função de risco acumulado, tende a produzir estimativas de sobrevivência ligeiramente mais otimistas. Essa característica é evidente na

probabilidade de sobrevivência no décimo segundo mês, que é de 0,321, um valor superior ao obtido com os métodos de Kaplan-Meier (0,241) e Atuarial (0,230). A seguir, a curva de Nelson-Aalen é plotada para visualizar essa trajetória de declínio mais suave.

```
# Adiciona manualmente os pontos iniciais e finais
enaTumor_ext <- enaTumor %>%
  add_row(time = 0, S_t_i = 1, lower = NA, upper = NA, .before = 1) %>%
  add_row(
    time = 15,
    S_t_i = tail(enaTumor$S_t_i, 1),
    lower = tail(enaTumor$lower, 1),
    upper = tail(enaTumor$upper, 1)
  )
ggplot() +
  geom_step(data = enaTumor_ext, aes(x = time, y = S_t_i), color = "#d95f02", size = 1.2) +
  geom_ribbon(
    data = enaTumor_ext,
    aes(x = time, ymin = lower, ymax = upper), fill = "#d95f02", alpha = 0.1
  ) +
  labs(
    title = "Curva de Sobrevivência de Tumor Sólido - Estimador de Nelson-Aalen",
    x = "Tempo (meses)",
    y = "Probabilidade de Sobrevivência"
  ) +
  theme_minimal() +
  scale_x_continuous(breaks = seq(0, 16, 1))
```

