

SME0821 - Análise de Sobrevivência e Confiabilidade

Amanda Caroline de Oliveira Pires

April 2024

1 Questão 1

Demonstre que a variância do estimador de Kaplan-Meier é dada por

$$\text{Var} \left[\hat{S}(t) \right] = \left[\hat{S}(t) \right]^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - q_j)}$$

Em que d_j é o número de falhas em t_j e n_j é o número de indivíduos sob risco em t_j .

Como a probabilidade de se obter d_i falhas em n_i casos segue uma distribuição binomial com probabilidade de falha h_i .

$$\hat{h}_i = \frac{d_i}{n_i} \text{ e } E \left(\hat{h}_i \right) = h_i$$

$$\text{Var} \left(\hat{h}_i \right) = h_i (1 - h_i) / n_i$$

Tomamos o logaritmo da função de sobrevivência e usamos o Método Delta para retornar a escala original.

$$\text{Var}(\log \hat{S}(t)) \sim \frac{1}{\hat{S}(t)^2} \text{Var}(\hat{S}(t)) \Rightarrow$$

$$\text{Var}(\hat{S}(t)) \sim \hat{S}(t)^2 \text{Var}(\log \hat{S}(t))$$

$$\log \hat{S}(t) = \sum_{i:t_i \leq t} \log(1 - \hat{h}_i)$$

Podemos escrever

$$\text{Var}(\widehat{S}(t)) \sim \widehat{S}(t)^2 \text{Var} \left(\sum_{i:t_i \leq t} \log(1 - \widehat{h}_i) \right)$$

$$\text{Var}(\widehat{S}(t)) \sim \widehat{S}(t)^2 \sum_{i:t_i \leq t} \text{Var} \left(\log(1 - \widehat{h}_i) \right)$$

$$\text{Var}(\widehat{S}(t)) \sim \widehat{S}(t)^2 \sum_{i:t_i \leq t} \left(\frac{\partial \log(1 - \widehat{h}_i)}{\partial \widehat{h}_i} \right)^2 \text{Var}(\widehat{h}_i)$$

$$\text{Var}(\widehat{S}(t)) = \widehat{S}(t)^2 \sum_{i:t_i \leq t} \left(\frac{1}{1 - \widehat{h}_i} \right)^2 \frac{\widehat{h}_i (1 - \widehat{h}_i)}{n_i}$$

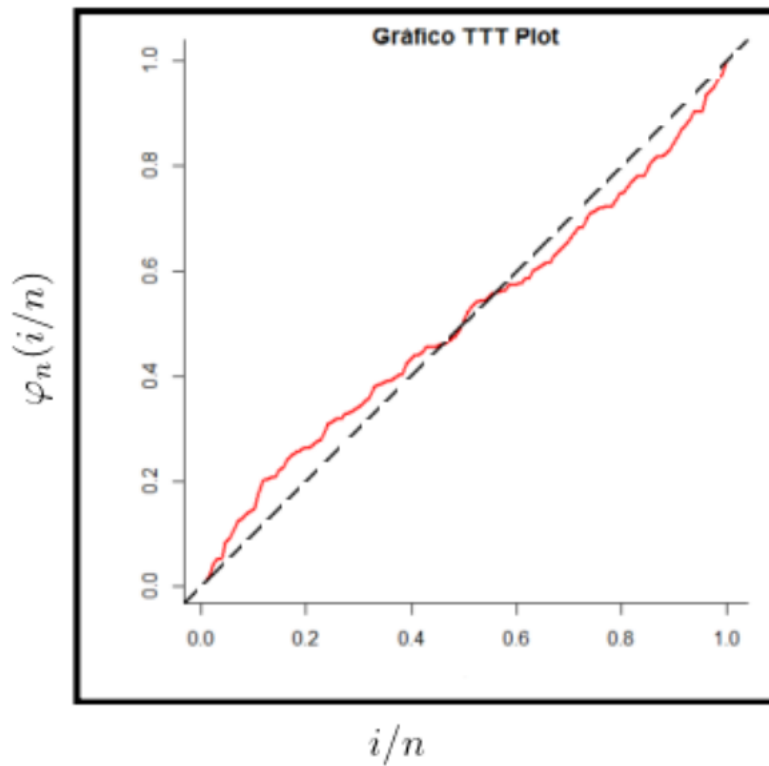
$$\text{Var}(\widehat{S}(t)) = \widehat{S}(t)^2 \sum_{i:t_i \leq t} \frac{\widehat{h}_i}{n_i (1 - \widehat{h}_i)}$$

E substituímos $\widehat{h}_i = \frac{d_i}{n_i}$

$$\text{Var}(\widehat{S}(t)) = \widehat{S}(t)^2 \sum_{i:t_i \leq t} \frac{d_i}{n_i (n_i - d_i)}$$

2 Questão 2

Com base na teoria e o seguinte gráfico, responda às seguintes perguntas.



2.1 a)

Qual é o objetivo principal do TTT plot na análise de sobrevivência?

É um gráfico empírico e independente de escala baseado em dados de falha. Tem o objetivo de fornecer análises gráficas a respeito do comportamento do tempo de sobrevivência, identifica quaisquer padrões ou tendências nos dados e se torna uma ferramenta útil quando há informações qualitativas sobre a curva de risco. Além disso, é usado para verificar quão bem um modelo se ajusta aos dados.

2.2 b)

Qual é o significado do eixo x no TTT plot? E o eixo y?

O eixo x representa o tempo acumulado de observação durante o período do estudo e o eixo y equivale ao número cumulativo de indivíduos em risco ou sob observação em cada ponto do tempo acumulado.

2.3 c)

O que representa a linha traçada no gráfico?

Representa a Função de risco constante.

2.4 d)

Como você interpretaria o ponto onde a linha atinge 0.7 no eixo y?

Cerca de 70% dos itens ou eventos foram observados até esse ponto específico no tempo.

2.5 e)

O que pode ser interpretado/inferido com base na inclinação da linha no TTT plot?

A linha do TTT apresenta forma de banheira. Podemos dizer que de 0 até aproximadamente 0.5 a curva se mantém crescente e côncava, de 0.5 até 1 apresenta comportamento decrescente e convexa. Por isso fica em formato de banheira.

3 Questão 3

Considere os dados de sobrevivência abaixo (em semanas) para 2 tratamentos. 6MP (novo tratamento) e placebo (controle). Os dados do grupo controle foram completos e os dados do grupo tratamentos foram censurados progressivamente.

Grupo	Tempos
Droga (6MP)	6, 6, 6, 6+, 7, 9+, 10, 10+, 11+, 13, 16, 17+, 19+, 20+, 22, 23, 25+, 32+, 32+, 34+, 35+
Placebo	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

3.1 a)

Encontrar os estimadores produto limite, Kaplan-Meier e Nelson Aalen para os dois grupos.

Estimador de Kaplan-Meier

$$\hat{S}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j}\right).$$

```

1 # Bibliotecas
2 library(survival)
3 library(survminer)
4 library(ggplot2)
5 library(tidyverse)
6 library(EstimationTools)
7
8 # Kaplan-Meier
9 Droga <- data.frame(
10   droga = c
11     (6,6,6,6,7,9,10,10,11,13,16,17,19,20,22,23,25,32,32,34,35),
12   cens1 = c(1,1,1,0,1,0,1,0,0,1,1,0,0,0,1,1,0,0,0,0,0)
13 )
14 Droga_km <- survfit(Surv(droga, cens1) ~ 1, data = Droga)
15
16 # Modelo de riscos proporcionais (Cox)
17 cox_model1 <- coxph(Surv(droga, cens1) ~ 1, data = Droga)
18 fit1_na <- survfit(cox_model1)
19
20 # Resultados
21 print(summary(Droga_km))

```

time	n.risk	n.event	survival	std.err	lower	95% CI upper	95% CI
6	21	3	0.857	0.0764	0.720	1.000	
7	17	1	0.807	0.0869	0.653	0.996	
10	15	1	0.753	0.0963	0.586	0.968	
13	12	1	0.690	0.1068	0.510	0.935	
16	11	1	0.627	0.1141	0.439	0.896	
22	7	1	0.538	0.1282	0.337	0.858	
23	6	1	0.448	0.1346	0.249	0.807	

```

1 # Kaplan-Meier
2 Placebo <- data.frame(
3   placebo = c(1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23),
4   cens2 = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)
5 )
6
7 Placebo_km <- survfit(Surv(placebo, cens2) ~ 1, data = Placebo)
8
9 # Modelo de riscos proporcionais (Cox)
10 cox_model2 <- coxph(Surv(placebo, cens2) ~ 1, data = Placebo)
11 fit2_na <- survfit(cox_model1)
12
13 # Resultados
14 print('Estimador de Kaplan-Meier Placebo:')
15 print(summary(Placebo_km))

```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	95% CI
1	21	2	0.9048			0.0641	0.78754 1.000
2	19	2	0.8095			0.0857	0.65785 0.996
3	17	1	0.7619			0.0929	0.59988 0.968
4	16	2	0.6667			0.1029	0.49268 0.902
5	14	2	0.5714			0.1080	0.39455 0.828
8	12	4	0.3810			0.1060	0.22085 0.657
11	8	2	0.2857			0.0986	0.14529 0.562
12	6	2	0.1905			0.0857	0.07887 0.460
15	4	1	0.1429			0.0764	0.05011 0.407
17	3	1	0.0952			0.0641	0.02549 0.356
22	2	1	0.0476			0.0465	0.00703 0.322
23	1	1	0.0000			NaN	NA NA

Nelson-Aalen

```
1 print(summary(fit1_na))
2 print(summary(fit2_na))
```

Droga

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	95% CI
6	21	3	0.860	0.0747		0.726	1.000
7	17	1	0.811	0.0851		0.661	0.996
10	15	1	0.759	0.0943		0.595	0.968
13	12	1	0.698	0.1045		0.521	0.936
16	11	1	0.638	0.1116		0.452	0.899
22	7	1	0.553	0.1249		0.355	0.861
23	6	1	0.468	0.1314		0.270	0.811

Placebo

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	95% CI
1	21	2	0.9070	0.0626		0.79219	1.000
2	19	2	0.8140	0.0839		0.66509	0.996
3	17	1	0.7675	0.0911		0.60821	0.968
4	16	2	0.6745	0.1010		0.50290	0.905
5	14	2	0.5815	0.1064		0.40631	0.832
8	12	4	0.3955	0.1054		0.23462	0.667
11	8	2	0.3026	0.0990		0.15936	0.575
12	6	2	0.2097	0.0877		0.09241	0.476
15	4	1	0.1633	0.0796		0.06286	0.424
17	3	1	0.1170	0.0691		0.03680	0.372
22	2	1	0.0710	0.0549		0.01558	0.323
23	1	1	0.0261	0.0330		0.00219	0.311

3.2 b)

Encontrar as variâncias dos estimadores de Kaplan-Meier e intervalos de confiança.

```
1 # Droga
2 Droga <- data.frame(
3   temp = c
4     (6,6,6,6,7,9,10,10,11,13,16,17,19,20,22,23,25,32,32,34,35),
5   cens = c(1,1,1,0,1,0,1,0,0,1,1,0,0,0,1,1,0,0,0,0,0)
6 )
7 # Placebo
8 Placebo <- data.frame(
9   temp = c(1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23),
10  cens = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)
11 )
12
13 varg <- function(grupos) {
14   km_fit = survfit(Surv(temp, cens) ~ 1, data = grupos)
15   variance = sum(km_fit$std.err^2)
16   ci_low = km_fit$lower
17   ci_high = km_fit$upper
18   return(list(variance = variance, ci_low = ci_low, ci_high = ci_
19     high))
20 }
21 grupo1resultado = varg(Droga)
22 grupo2resultado = varg(Placebo)
23
24 print(paste0("Variância: ", round(grupo1resultado$variance, 3)))
25 print(sprintf("Intervalo de Confiança (IC): [%s, %s]",
26   round(grupo1resultado$ci_low, 3),
27   round(grupo1resultado$ci_high, 3)))
28 print(paste0("Variância: ", round(grupo2resultado$variance, 3)))
29 print(sprintf("Intervalo de Confiança (IC): [%s, %s]",
30   round(grupo2resultado$ci_low, 3),
31   round(grupo2resultado$ci_high, 3)))
```

Droga

```
[1] "Variância: 0.728"
[1] "Intervalo de Confiança (IC): [0.72, 1]"
[2] "Intervalo de Confiança (IC): [0.653, 0.996]"
[3] "Intervalo de Confiança (IC): [0.653, 0.996]"
[4] "Intervalo de Confiança (IC): [0.586, 0.968]"
[5] "Intervalo de Confiança (IC): [0.586, 0.968]"
[6] "Intervalo de Confiança (IC): [0.51, 0.935]"
[7] "Intervalo de Confiança (IC): [0.439, 0.896]"
[8] "Intervalo de Confiança (IC): [0.439, 0.896]"
[9] "Intervalo de Confiança (IC): [0.439, 0.896]"
[10] "Intervalo de Confiança (IC): [0.439, 0.896]"
[11] "Intervalo de Confiança (IC): [0.337, 0.858]"
[12] "Intervalo de Confiança (IC): [0.249, 0.807]"
[13] "Intervalo de Confiança (IC): [0.249, 0.807]"
[14] "Intervalo de Confiança (IC): [0.249, 0.807]"
[15] "Intervalo de Confiança (IC): [0.249, 0.807]"
[16] "Intervalo de Confiança (IC): [0.249, 0.807]"
```

Placebo

```
"Variância: Inf"
"Intervalo de Confiança (IC): [0.788, 1]"
"Intervalo de Confiança (IC): [0.658, 0.996]"
"Intervalo de Confiança (IC): [0.6, 0.968]"
"Intervalo de Confiança (IC): [0.493, 0.902]"
"Intervalo de Confiança (IC): [0.395, 0.828]"
"Intervalo de Confiança (IC): [0.221, 0.657]"
"Intervalo de Confiança (IC): [0.145, 0.562]"
"Intervalo de Confiança (IC): [0.079, 0.46]"
"Intervalo de Confiança (IC): [0.05, 0.407]"
"Intervalo de Confiança (IC): [0.025, 0.356]"
"Intervalo de Confiança (IC): [0.007, 0.322]"
"Intervalo de Confiança (IC): [NA, NA]"
```

3.3 c)

Faça o gráfico das curvas de sobrevivência estimadas.

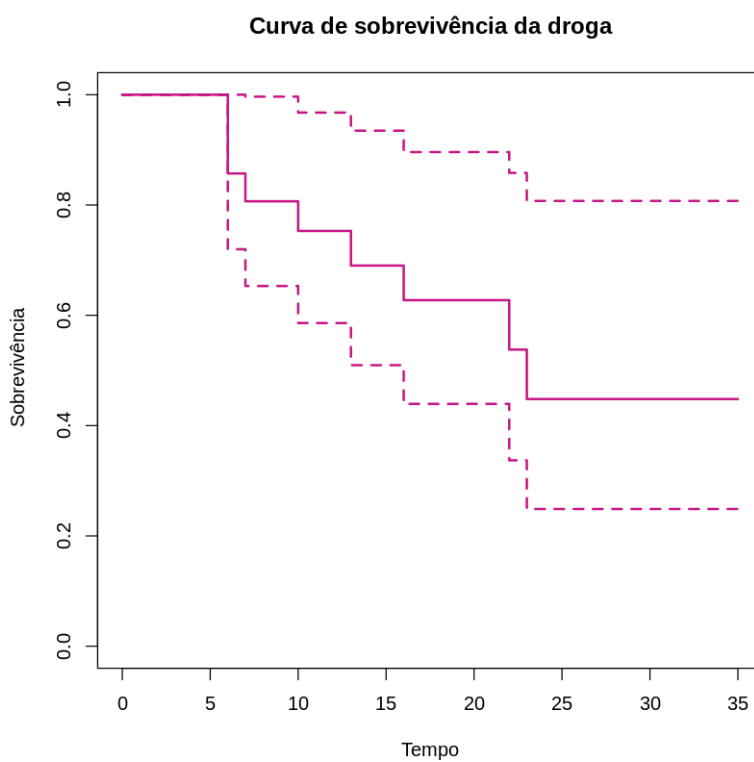
```
1 # Biblioteca survival
2 library(survival )
3
4 droga <- c
  (6,6,6,6,6,7,9,10,10,11,13,16,17,19,20,22,23,25,32,32,34,35)
```



```

5 cens1 <- c(1,1,1,0,1,0,1,0,0,1,1,0,0,0,1,1,0,0,0,0,0)
6
7 sobrevivencia1 <- Surv (droga , event = cens1 )
8 Surv(droga, cens1)
9 # Ajustar o modelo
10 modelo_sobrevivencia1 <- survfit(sobrevivencia1 ~ 1)
11 # Plotar a curva
12 plot(modelo_sobrevivencia1, xlab = " Tempo ", ylab = "
    Sobrevivência", main = "Curva de sobrevivência da droga", col
    = "#C71585",lwd = 2)
13 summary(modelo_sobrevivencia1)

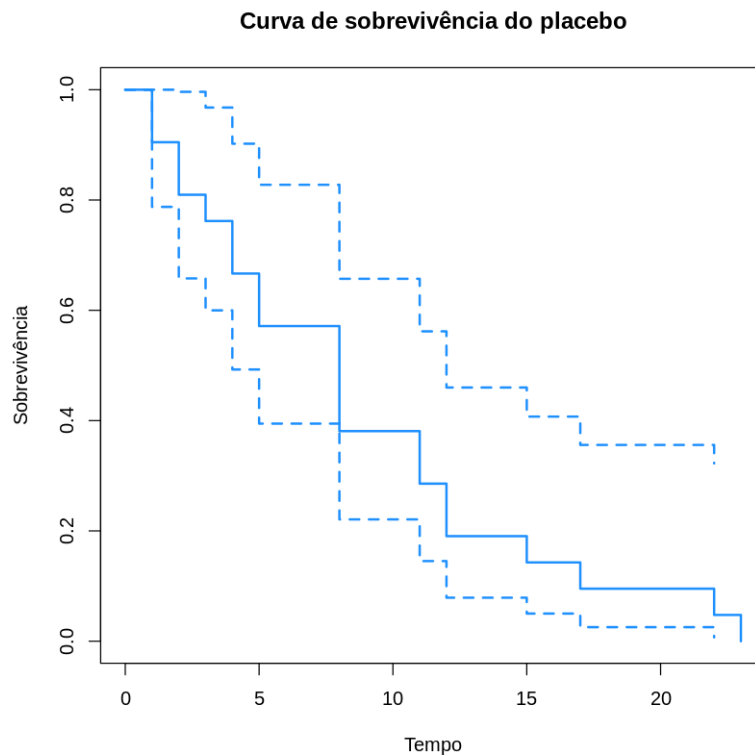
```



```

1 placebo=c(1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23)
2 cens2 = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)
3 sobrevivencia2 <- Surv (placebo , event = cens2)
4 Surv(placebo, cens2)
5 # Ajustar o modelo
6 modelo_sobrevivencia2 <- survfit(sobrevivencia2 ~ 1)
7 # Plotar a curva
8 plot(modelo_sobrevivencia2, xlab = "Tempo", ylab = "Sobrevivência",
9      , main = "Curva de sobrevivência do placebo", col = "#1E90FF",
10     , lwd = 2)
11 summary(modelo_sobrevivencia2)

```



3.4 d)

Calcule o tempo médio de vida dos pacientes para os dois grupos.

3.5 e)

Com apoio computacional obtenha todos os estimadores e faça os gráficos.

3.6 f)

Faça uma conclusão para os resultados encontrado.

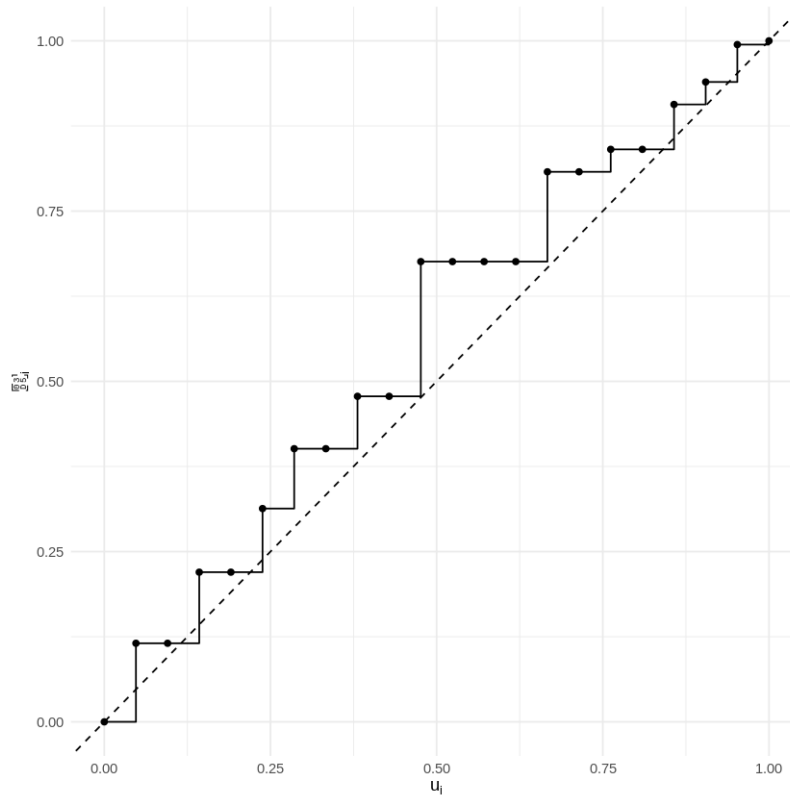
Observa-se que as curvas de sobrevivência e os intervalos de confiança da droga e do placebo são diferentes. O grupo que recebeu a droga apresenta um tempo de sobrevida maior em comparação com o grupo que recebeu apenas o placebo. Portanto, há indícios que o tratamento com a droga é eficaz.

3.7 g)

Construa o gráfico TTT plot considerando os grupos droga e placebo. Em relação à função de risco associada a cada um desses grupos, o que podemos concluir?

```
1 library(tidyverse)
2 library(EstimationTools)
3 # =====
4 # TTT PLOT PARA DADOS N O -CENSURADOS (Placebo)
5 # =====
6 t = c(1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23)
7 TTT = TTTE_Analytical(t~1, method='Barlow')
8
9 dadosTTT = tibble(
10   x = TTT$i/n`,
11   y = TTT$phi_n
12 )
13 G =
14   dadosTTT %>%
15   ggplot(aes(x=x, y=y)) +
16   geom_step() +
17   geom_point() +
18   geom_abline(slope=1, intercept=0, linetype="dashed") +
19   lims(y=c(0,1), x=c(0,1)) +
20   labs(x = bquote(u[i]), y = bquote(varphi[i])) +
21   theme_minimal(); G
```

Placebo



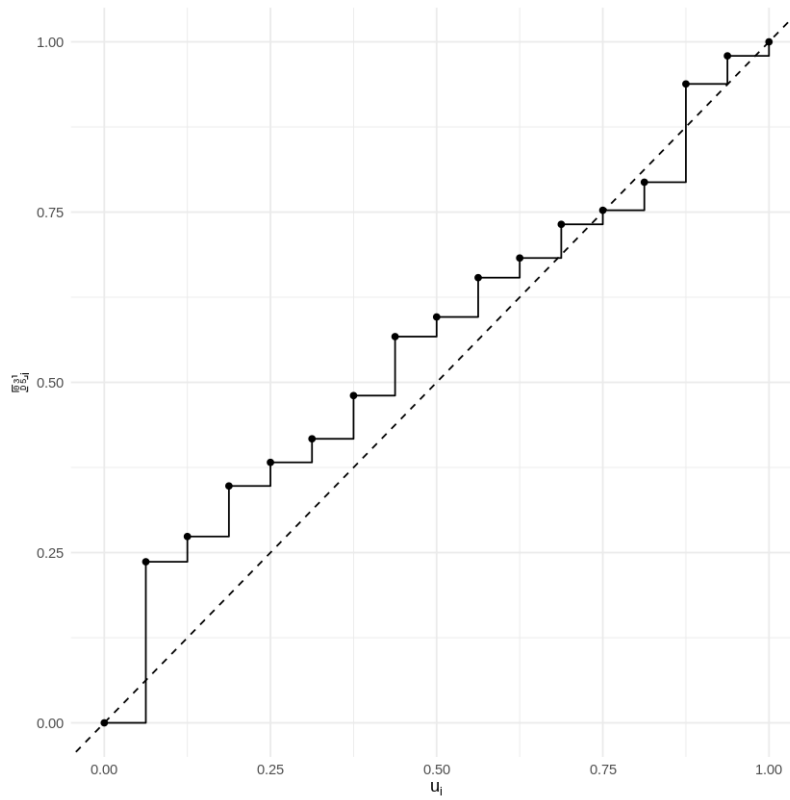
```

1 # =====
2 # TTT PLOT PARA DADOS CENSURADOS
3 # =====
4
5 droga = c(6,6,6,6,7,9,10,10,11,13,16,17,19,20,22,23,25,32,32,34,35)
6 cens = c(1,1,1,0,1,0,1,0,0,1,1,0,0,0,1,1,0,0,0,0,0)
7
8 TTT = TTTE_Analytical(Surv(droga,cens)~1, method='censored')
9
10 dadosTTT = tibble(
11   x = TTT$i/n,
12   y = TTT$phi_n
13 )
14
15 G =
16   dadosTTT %>%
17   ggplot(aes(x=x, y=y)) +
18   geom_step() +
19   geom_point() +
20   geom_abline(slope=1, intercept=0, linetype="dashed") +
21   lims(y=c(0,1), x=c(0,1)) +
22   labs(x = bquote(u[i]), y = bquote(varphi[i])) +

```

23 `theme_minimal(); G`

Droga



Conclusão: O gráfico TTT do placebo apresenta curva crescente, ou seja, tem um aumento na taxa de falha, isso pode indicar que mais pessoas estão morrendo conforme o tempo passa. O TTT da droga apresenta a curva quase em forma de banheira, indicando um crescimento nos tempos iniciais, um decaimento no meio e por fim um decrescimento no final, isso poderia indicar que a droga está surtindo efeito no tratamento contra certa doença.