Document title

Subtitle

Davi Wentrick Feijó -200016806, Micael Papa - 000000000

November 28, 2023

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean ut elit odio. Donec fermentum tellus neque, vitae fringilla orci pretium vitae. Fusce maximus finibus facilisis. Donec ut ullamcorper turpis. Donec ut porta ipsum. Nullam cursus mauris a sapien ornare pulvinar. Aenean malesuada molestie erat quis mattis. Praesent scelerisque posuere faucibus. Praesent nunc nulla, ullamcorper ut ullamcorper sed, molestie ut est. Donec consequat libero nisi, non semper velit vulputate et. Quisque eleifend tincidunt ligula, bibendum finibus massa cursus eget. Curabitur aliquet vehicula quam non pulvinar. Aliquam facilisis tortor nec purus finibus, sit amet elementum eros sodales. Ut porta porttitor vestibulum.

1 Introdução

2 Metodologia

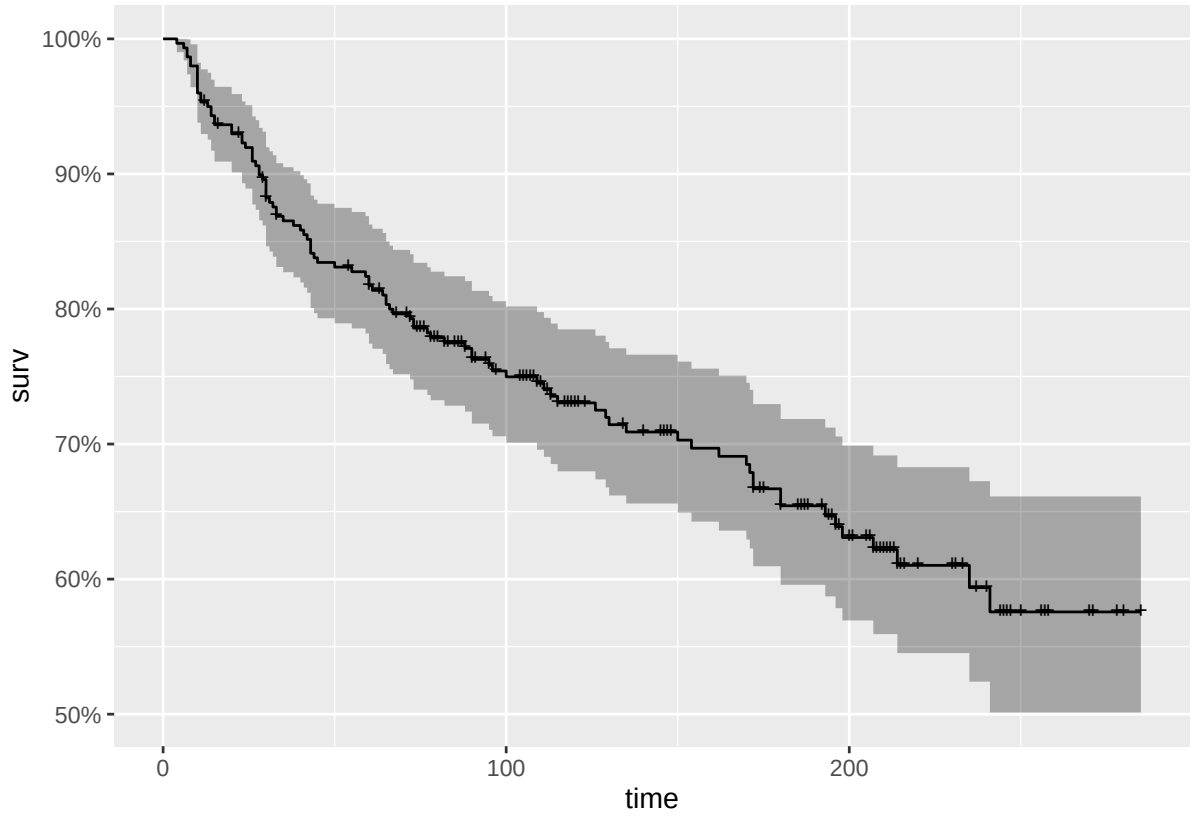
3 Resultados

3.1 Analise exploratoria

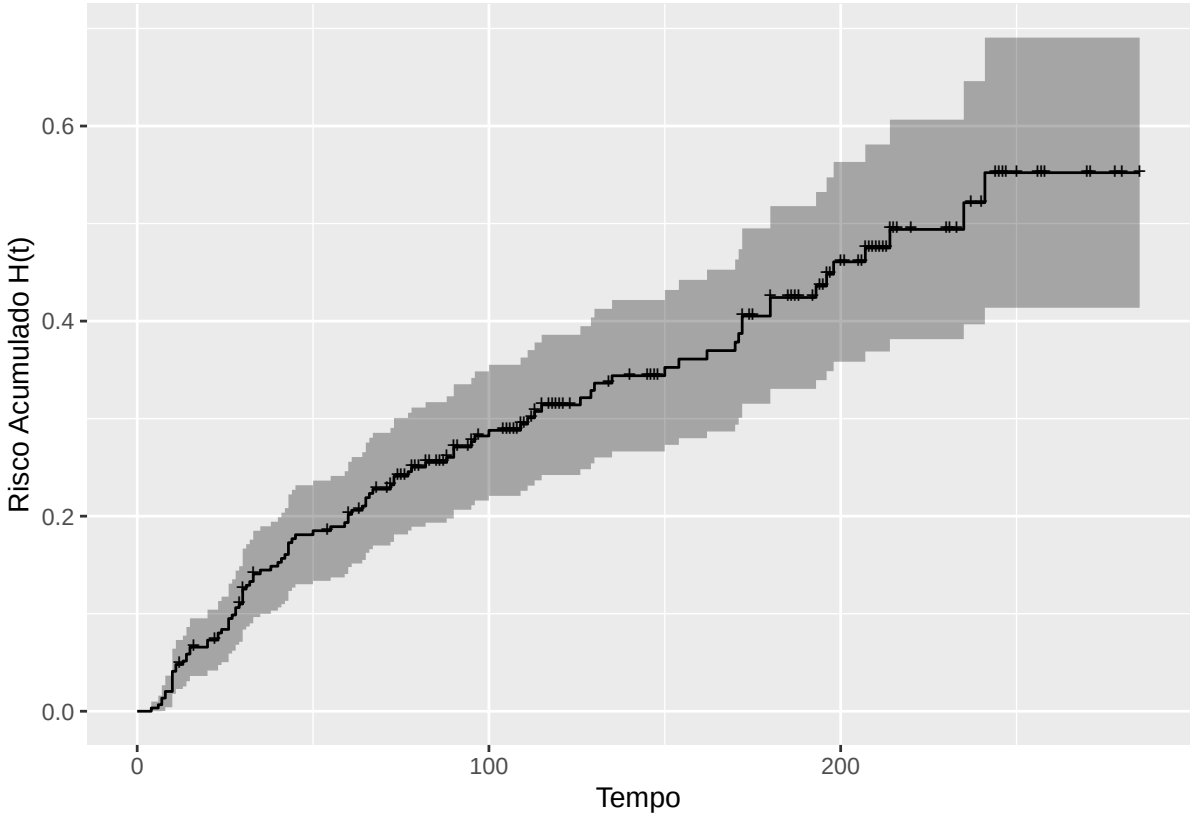
Antes de ajustar o modelo,é necessario estudarmos o comportamento dos dados antes para pode identificar qual distribuicao sera mais adequada e ja realizar uma seleção das variaveis categoricas que sao significativas para o modelo final.

```
## # A tibble: 6 x 13
##     age anaemia creatinine_phosphokinase diabetes ejection_fraction
##   <dbl>   <dbl>                    <dbl>    <dbl>              <dbl>
## 1    75       0                      582        0                 20
## 2    55       0                     7861        0                 38
## 3    65       0                      146        0                 20
## 4    50       1                      111        0                 20
## 5    65       1                      160        1                 20
## 6    90       1                       47        0                 40
## # i 8 more variables: high_blood_pressure <dbl>, platelets <dbl>,
## #   serum_creatinine <dbl>, serum_sodium <dbl>, sex <dbl>, smoking <dbl>,
## #   censura <dbl>, tempo <dbl>
```

3.1.1 Modelo de sobrevivencia nao parametrico de Kaplan-Meier



3.1.2 A funcao de risco acumulado



3.1.3 Curva TTT

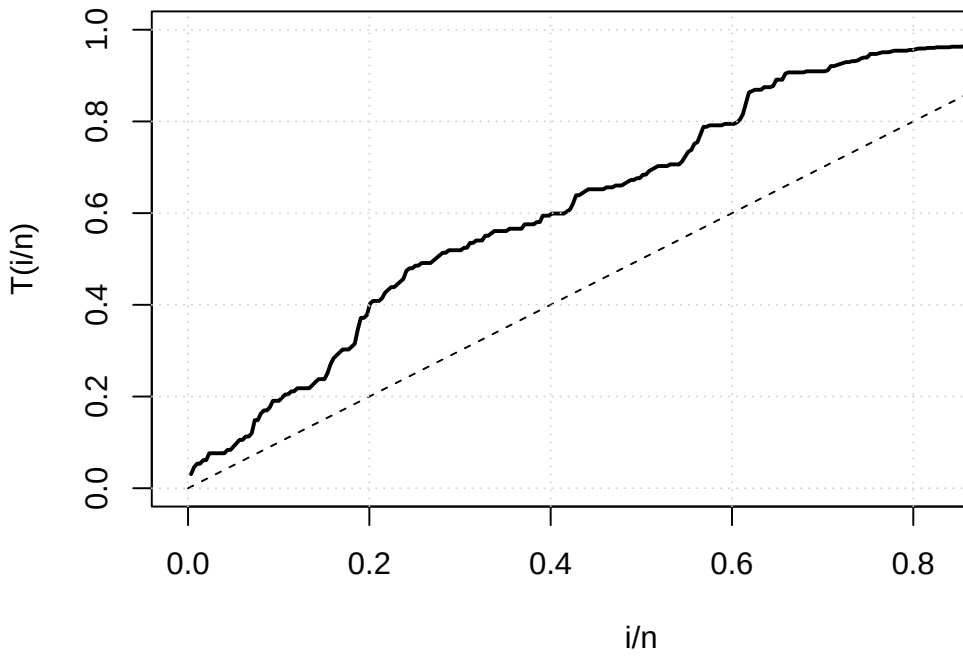
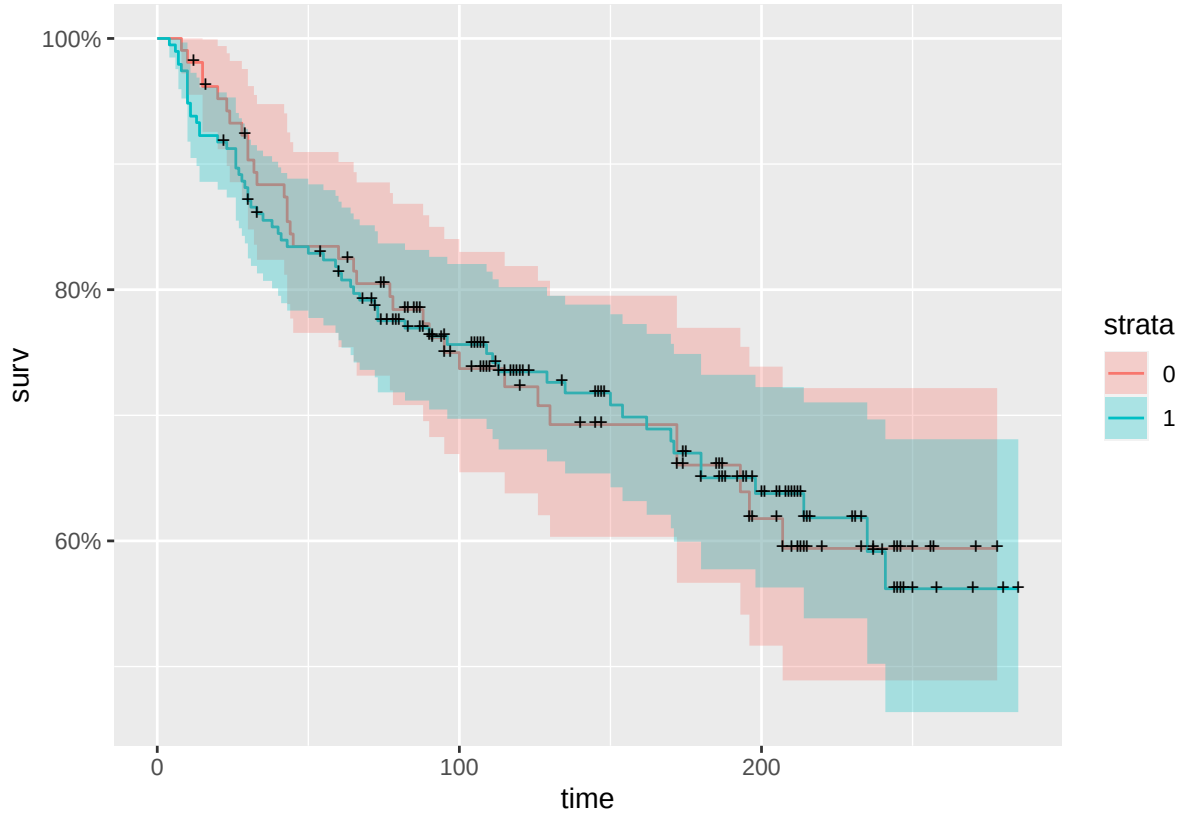


Grafico do Tempo Total sobre Teste

3.1.4 Analise das Variaveis Categoricas

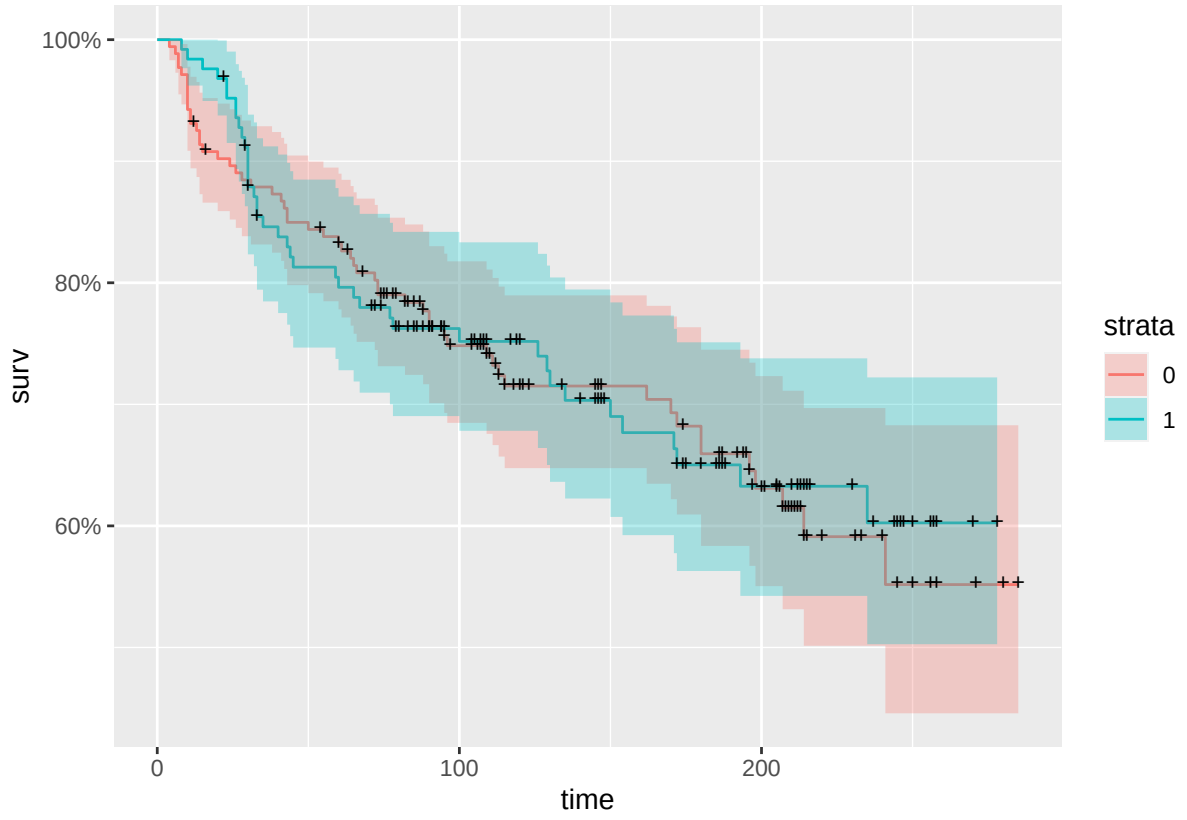
3.1.4.1 Variavel Sex Vamos comparar as curvas de sobrevivencias divididas por Sexo, com o objetivo de ver se essa variavel influencia na curva de sobrevivencia. Em seguida iremos fazer um teste para verificar a diferenca entre as curvas.




```
## Call:
## survdiff(formula = Surv(tempo, censura) ~ sex, data = dados,
##     rho = 1)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=0 105     27.9     28.5   0.01467    0.0271
## sex=1 194     52.0     51.4   0.00814    0.0271
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.9
```

Podemos notar que tanto pelo grafico quanto pelo teste, com p-valor = 0.9, que a variavel Sexo nao parece influenciar nas curvas de sobrevivencia

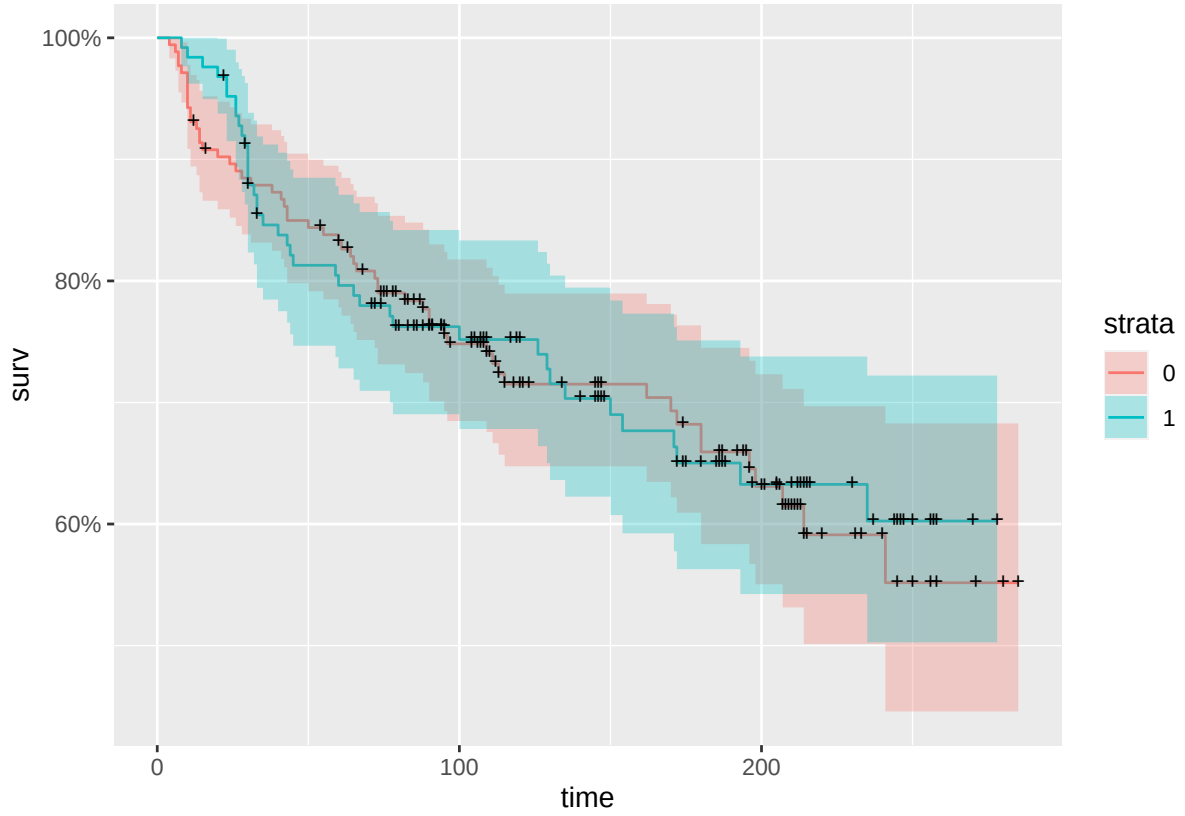
3.1.4.2 **Variavel Diabetes** Comparacao entre curvas de sobrevivencia da variavel Dibetes




```
## Call:
## survdiff(formula = Surv(tempo, censura) ~ diabetes, data = dados,
##     rho = 1)
##
##                N Observed Expected (O-E)^2/E (O-E)^2/V
## diabetes=0 174      46.7     45.9    0.0125    0.0349
## diabetes=1 125      33.2     34.0    0.0168    0.0349
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.9
```

Pelo p-valor de 0.9 pode assumir que nao existe diferenca entre as curvas

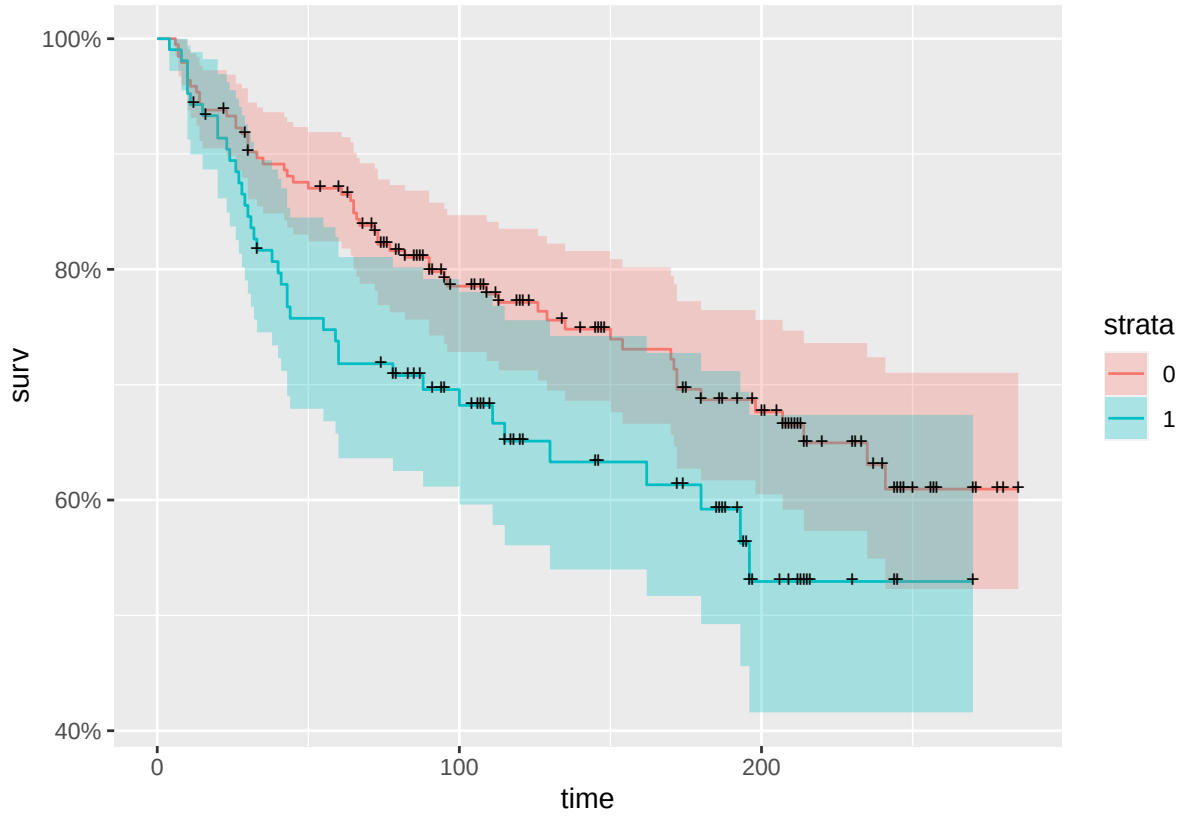
3.1.4.3 Variavel Anaemia Comparacao das curvas entre os grupo que tem Anemia e os que nao tem.




```
## Call:
## survdiff(formula = Surv(tempo, censura) ~ diabetes, data = dados,
##     rho = 1)
##
##                N Observed Expected (O-E)^2/E (O-E)^2/V
## diabetes=0 174     46.7     45.9    0.0125    0.0349
## diabetes=1 125     33.2     34.0    0.0168    0.0349
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.9
```

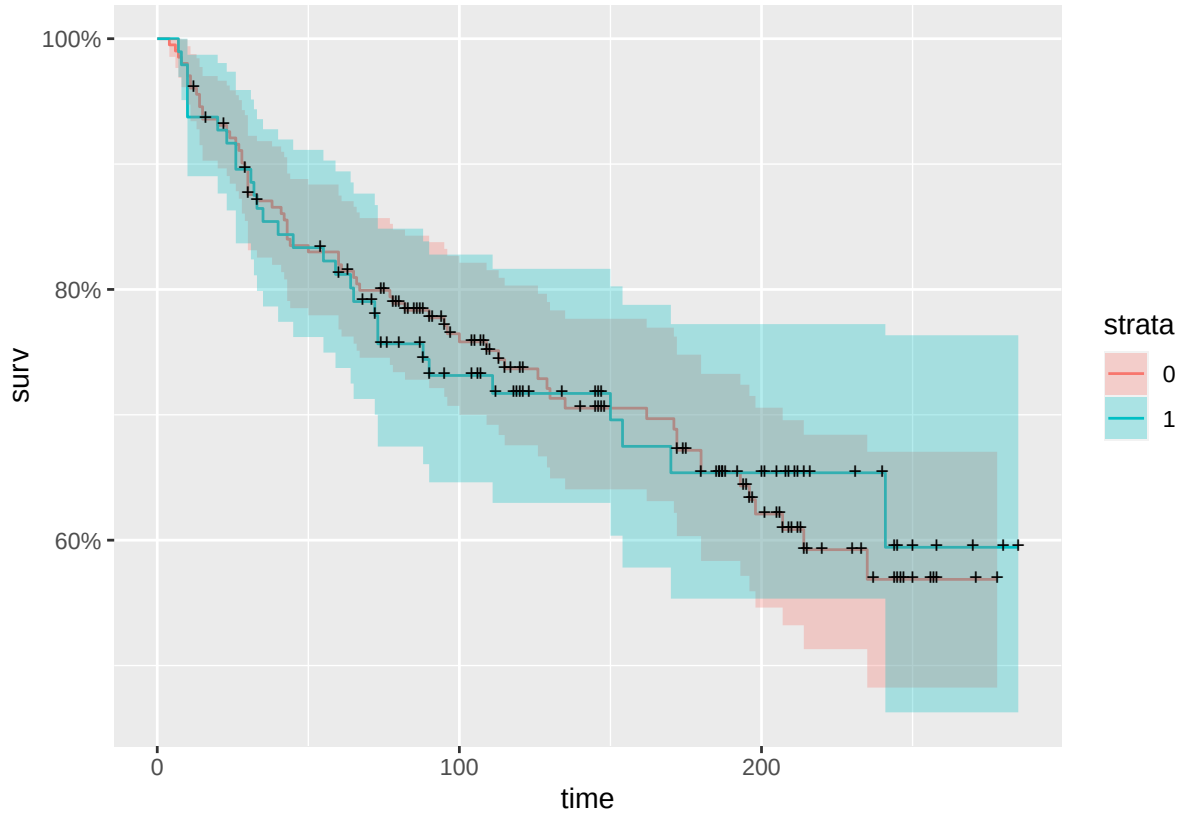
Pelo p-valor de 0.9 podemos assumir que nao ha diferenca entre as categorias

3.1.4.4 Variavel High Blood Pressure




```
## Call:
## survdiff(formula = Surv(tempo, censura) ~ high_blood_pressure,
##     data = dados, rho = 1)
##
##                           N Observed Expected (O-E)^2/E (O-E)^2/V
## high_blood_pressure=0 194     46.6     54.8      1.25      4.71
## high_blood_pressure=1 105     33.3     25.1      2.72      4.71
##
##  Chisq= 4.7  on 1 degrees of freedom, p= 0.03
```

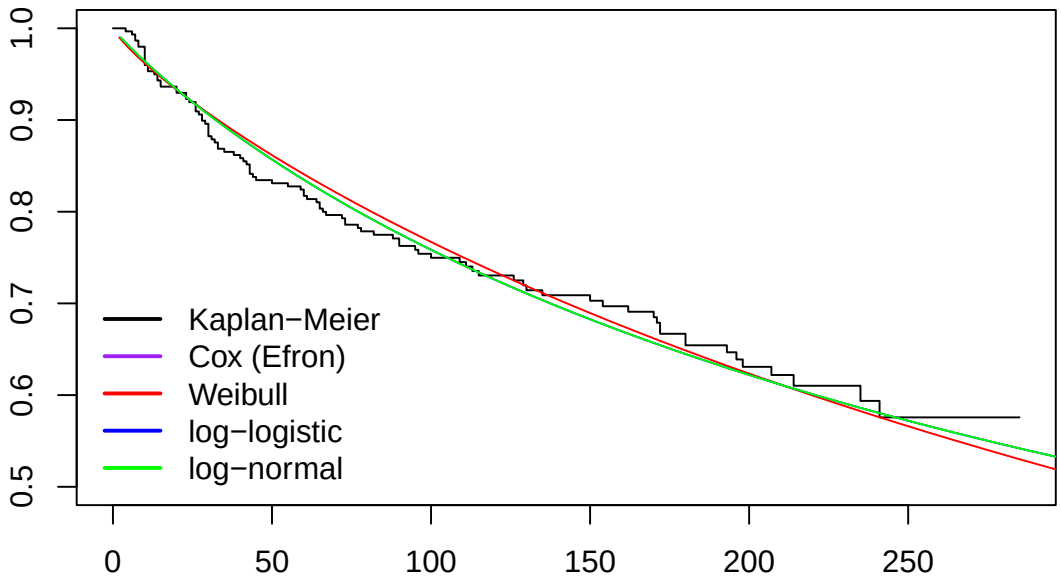
3.1.4.5 Variavel Smoking




```
## Call:
## survdiff(formula = Surv(tempo, censura) ~ smoking, data = dados,
##     rho = 1)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## smoking=0 203     54.5     54.7  0.000902   0.00339
## smoking=1  96     25.5     25.2  0.001955   0.00339
##
##  Chisq= 0  on 1 degrees of freedom, p= 1
```

3.1.5 Seleçao da distribuicao

Vamos ajustar algumas distribuicoes sobre o grafico de Kaplan-Meier para selecionar aquela que se adapta melhor a curva. Alem disso estaremos verificando os valores AIC, AIC corrigido e BIC para deicidir a distribuicao a ser utilizada.




```
## Weibull ~( 0.8333038 , 491.7358 )

##         AICws   AICcws    BICws
## [1,] 1344.876 1344.916 1352.277

##
## Call:
## survreg(formula = s ~ 1, data = dados, dist = "weibull")
##               Value Std. Error     z      p
## (Intercept) 6.1979     0.1638 37.83 <2e-16
## Log(scale)  0.1824     0.0923  1.98  0.048
##
## Scale= 1.2
##
## Weibull distribution
## Loglik(model)= -670.4   Loglik(intercept only)= -670.4
## Number of Newton-Raphson Iterations: 5
## n= 299

## Log-Normal ~( 5.914729 , 1.916652 )

##         AICln  AICcln   BICln
## [1,] 1336.546 1336.587 1343.947

##
```

```
## Call:
## survreg(formula = s ~ 1, data = dados, dist = "loglogistic")
##              Value Std. Error     z      p
## (Intercept) 5.8326     0.1613 36.17 <2e-16
## Log(scale)  0.0703     0.0897  0.78   0.43
##
## Scale= 1.07
##
## Log logistic distribution
## Loglik(model)= -669.2   Loglik(intercept only)= -669.2
## Number of Newton-Raphson Iterations: 4
## n= 299

## Log-Normal ~( 0.9320725 , 341.255 )

##        AIClls  AICclls   BIClls
## [1,] 1342.334 1342.375 1349.735

##
## Call:
## survreg(formula = s ~ 1, data = dados, dist = "loglogistic")
##              Value Std. Error     z      p
## (Intercept) 5.8326     0.1613 36.17 <2e-16
## Log(scale)  0.0703     0.0897  0.78   0.43
##
## Scale= 1.07
##
## Log logistic distribution
## Loglik(model)= -669.2   Loglik(intercept only)= -669.2
## Number of Newton-Raphson Iterations: 4
## n= 299

##         AICws   AICcws    BICws
## [1,] 1344.876 1344.916 1352.277

##        AIClns  AICclns   BIClns
## [1,] 1336.546 1336.587 1343.947

##        AIClls  AICclls   BIClls
## [1,] 1342.334 1342.375 1349.735
```

4 Conclusão