

SME0821 - Análise de Sobrevida - Atividade I

Francisco Rosa Dias de Miranda - 4402962
Lua

Heitor Carvalho Pinheiro - 11833351
Gusthavo

abril 2022

1) Introdução

O mieloma múltiplo é o câncer que afeta aos plasmócitos, células da medula óssea responsáveis pela produção de anticorpos. Nos indivíduos acometidos, os plasmócitos são anormais e se multiplicam rapidamente, comprometendo a produção das outras células do sangue.

Foram obtidas medidas de expressão gênica em indivíduos com mieloma múltiplo, a partir de bases disponíveis no GEO (Id: GSE4581), um repositório de dados genômicos públicos do NCBI (National Center for Biotechnology Information). Nesse estudo, foram coletados dados de uma amostra de 256 pacientes, consistindo nas 11 colunas descritas abaixo:

Variável	Descrição
molecular_group	Subgrupos moleculares dos pacientes
chr1q21_status	Status de amplificação do cromossomo 1q21
treatment	Todos os pacientes receberam o tratamento TT2
event	Status de sobrevivência, 0 = vivo, 1 = morto
time	Tempo de sobrevivência, em meses
CCND1, CRIM1, DEPDC1, IRF4 TP53, WHSC1	Nível de expressão dos respectivos genes

```
# Conjunto de dados utilizado
```

```
df <- survminer::myeloma %>% rownames_to_column %>% tibble  
head(df)
```

```
## # A tibble: 6 x 12  
##   rowname molecular_group chr1q21_status treatment event   time CCND1 CRIM1  
##   <chr>      <fct>          <fct>          <fct>    <int> <dbl> <dbl> <dbl>  
## 1 GSM50986 Cyclin D-1      3 copies      TT2          0  69.2  9908. 421.  
## 2 GSM50988 Cyclin D-2      2 copies      TT2          0  66.4 16699.  52  
## 3 GSM50989 MMSET          2 copies      TT2          0  66.5   294. 618.  
## 4 GSM50990 MMSET          3 copies      TT2          1  42.7   242. 11.9  
## 5 GSM50991 MAF            <NA>          TT2          0   65    473. 38.8  
## 6 GSM50992 Hyperdiploid    2 copies      TT2          0  65.2   664. 16.9  
## # ... with 4 more variables: DEPDC1 <dbl>, IRF4 <dbl>, TP53 <dbl>, WHSC1 <dbl>
```

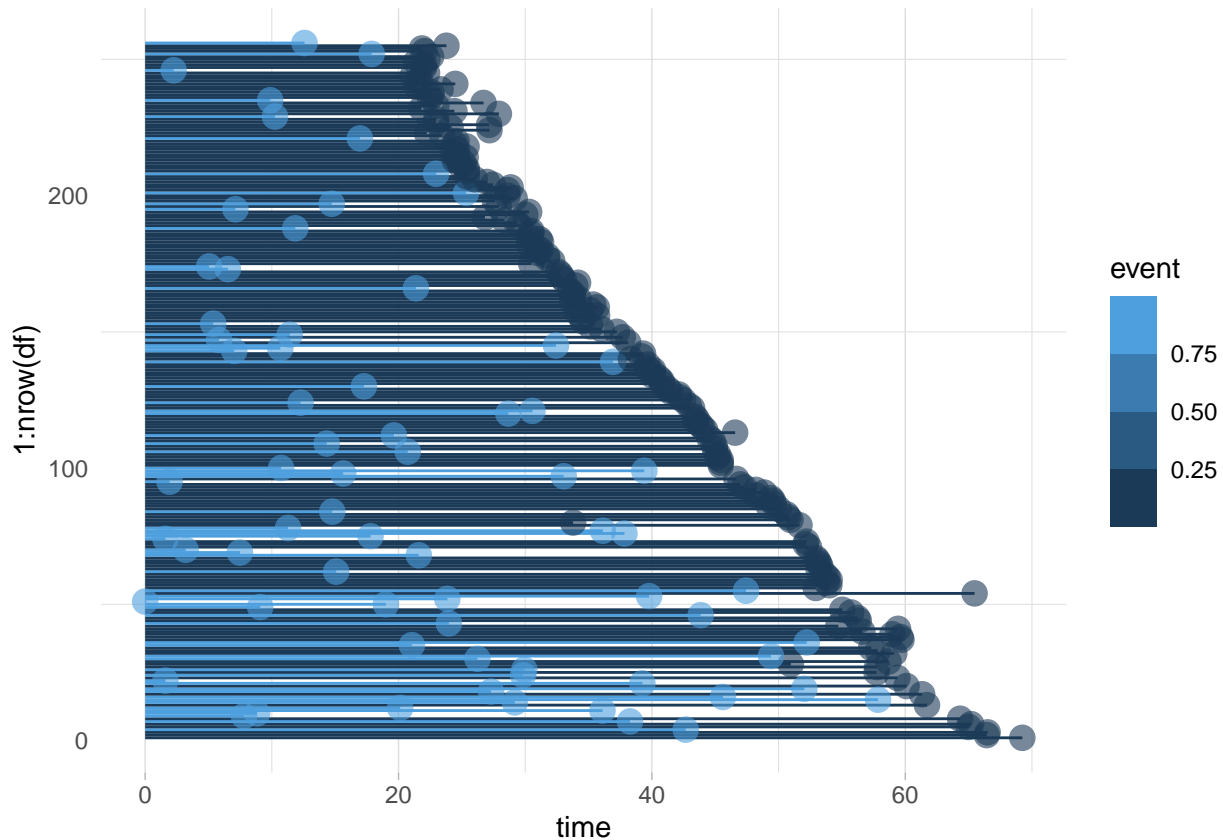
```
## trocar por tempo vs raiz quadrada da contagem
```

```
df %>%  
ggplot(aes(x=1:nrow(df), y=time)) +  
  geom_segment(aes(x=1:nrow(df), xend=1:nrow(df), y=0, yend=time, color=event)) +  
  geom_point(aes(color=event), size=4, alpha=0.6) +
```

```

scale_color_binned() +
theme_light() +
coord_flip() +
theme(
  panel.grid.major.y = element_blank(),
  panel.border = element_blank(),
  axis.ticks.y = element_blank()
)

```



2) Metodologia: Uma breve descrição da metodologia;

Nesse trabalho, nosso objetivo é a análise de dados de sobrevivência com censura a direita a partir de uma abordagem não-paramétrica, em que o interesse é identificar fatores de prognóstico para o mioma múltiplo a partir da amostra coletada.

A análise de sobrevivência tem como objetivo a identificação de Neste estudo, utilizamos metodologia não paramétrica a dados de sobrevivência em presença de censura

3) Análise de dados

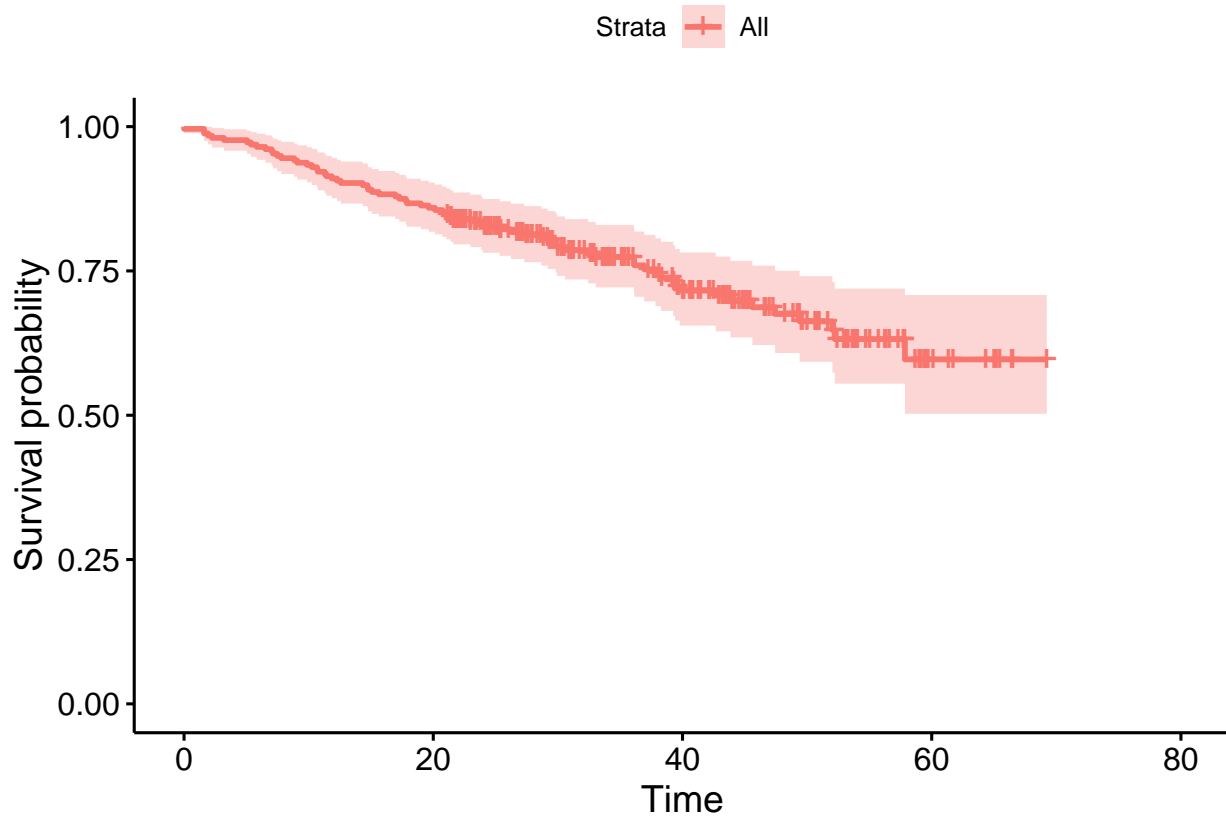
- [-] K-M
- ☐ Atuarial
- ☐ Nelson-Aalen
- ☐ EMV
- ☐ Newton-Rapson
- ☐ Testes de hipótese

#####

I) *K-M Algoritmo*

```
km_fit <- survfit(Surv(time, event) ~ 0, data = df)
```

```
ggsurvplot(km_fit)
```



```
summary(km_fit, times = c(0,1:70))
```

```
## Call: survfit(formula = Surv(time, event) ~ 0, data = df)
```

```
##
```

##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
##	0	256	1	0.996	0.00390	0.988	1.000
##	1	255	0	0.996	0.00390	0.988	1.000
##	2	252	3	0.984	0.00775	0.969	1.000
##	3	251	1	0.980	0.00865	0.964	0.998
##	4	250	1	0.977	0.00946	0.958	0.995
##	5	250	0	0.977	0.00946	0.958	0.995
##	6	247	3	0.965	0.01151	0.943	0.988
##	7	246	1	0.961	0.01211	0.937	0.985
##	8	242	4	0.945	0.01421	0.918	0.974
##	9	241	1	0.941	0.01468	0.913	0.971
##	10	239	2	0.934	0.01556	0.904	0.965
##	11	236	3	0.922	0.01677	0.890	0.955
##	12	233	3	0.910	0.01787	0.876	0.946
##	13	231	2	0.902	0.01855	0.867	0.939
##	14	231	0	0.902	0.01855	0.867	0.939
##	15	228	3	0.891	0.01951	0.853	0.930
##	16	226	2	0.883	0.02010	0.844	0.923

##	17	225	1	0.879	0.02039	0.840	0.920
##	18	222	3	0.867	0.02121	0.827	0.910
##	19	222	1	0.863	0.02147	0.822	0.906
##	20	220	1	0.859	0.02173	0.818	0.903
##	21	218	2	0.852	0.02222	0.809	0.896
##	22	208	3	0.840	0.02293	0.796	0.886
##	23	194	1	0.836	0.02321	0.791	0.882
##	24	189	2	0.827	0.02378	0.781	0.875
##	25	179	0	0.827	0.02378	0.781	0.875
##	26	172	1	0.822	0.02410	0.776	0.871
##	27	167	1	0.817	0.02443	0.771	0.867
##	28	160	1	0.812	0.02478	0.765	0.862
##	29	155	1	0.807	0.02516	0.759	0.858
##	30	149	3	0.791	0.02628	0.741	0.845
##	31	139	1	0.786	0.02667	0.735	0.840
##	32	134	0	0.786	0.02667	0.735	0.840
##	33	129	1	0.780	0.02712	0.728	0.835
##	34	121	1	0.774	0.02758	0.722	0.830
##	35	116	0	0.774	0.02758	0.722	0.830
##	36	111	0	0.774	0.02758	0.722	0.830
##	37	107	3	0.753	0.02939	0.697	0.813
##	38	104	1	0.745	0.02997	0.689	0.807
##	39	101	1	0.738	0.03054	0.681	0.801
##	40	94	3	0.716	0.03225	0.655	0.782
##	41	88	0	0.716	0.03225	0.655	0.782
##	42	86	0	0.716	0.03225	0.655	0.782
##	43	81	1	0.707	0.03299	0.645	0.775
##	44	74	1	0.698	0.03385	0.635	0.767
##	45	69	0	0.698	0.03385	0.635	0.767
##	46	63	1	0.687	0.03503	0.622	0.759
##	47	60	0	0.687	0.03503	0.622	0.759
##	48	58	1	0.675	0.03632	0.608	0.750
##	49	55	0	0.675	0.03632	0.608	0.750
##	50	50	1	0.663	0.03780	0.592	0.741
##	51	45	0	0.663	0.03780	0.592	0.741
##	52	44	0	0.663	0.03780	0.592	0.741
##	53	37	2	0.632	0.04187	0.555	0.719
##	54	30	0	0.632	0.04187	0.555	0.719
##	55	27	0	0.632	0.04187	0.555	0.719
##	56	25	0	0.632	0.04187	0.555	0.719
##	57	22	0	0.632	0.04187	0.555	0.719
##	58	17	1	0.597	0.05222	0.503	0.708
##	59	15	0	0.597	0.05222	0.503	0.708
##	60	10	0	0.597	0.05222	0.503	0.708
##	61	9	0	0.597	0.05222	0.503	0.708
##	62	7	0	0.597	0.05222	0.503	0.708
##	63	7	0	0.597	0.05222	0.503	0.708
##	64	7	0	0.597	0.05222	0.503	0.708
##	65	6	0	0.597	0.05222	0.503	0.708
##	66	3	0	0.597	0.05222	0.503	0.708
##	67	1	0	0.597	0.05222	0.503	0.708
##	68	1	0	0.597	0.05222	0.503	0.708
##	69	1	0	0.597	0.05222	0.503	0.708

```
res.cut <- surv_cutpoint(df, time = "time", event = "event",
                        variables = c("CRIM1", "DEPDC1", "WHSC1",
                                     "CCND1", "IRF4", "TP53"),
                        progressbar = FALSE)

summary(res.cut)
```

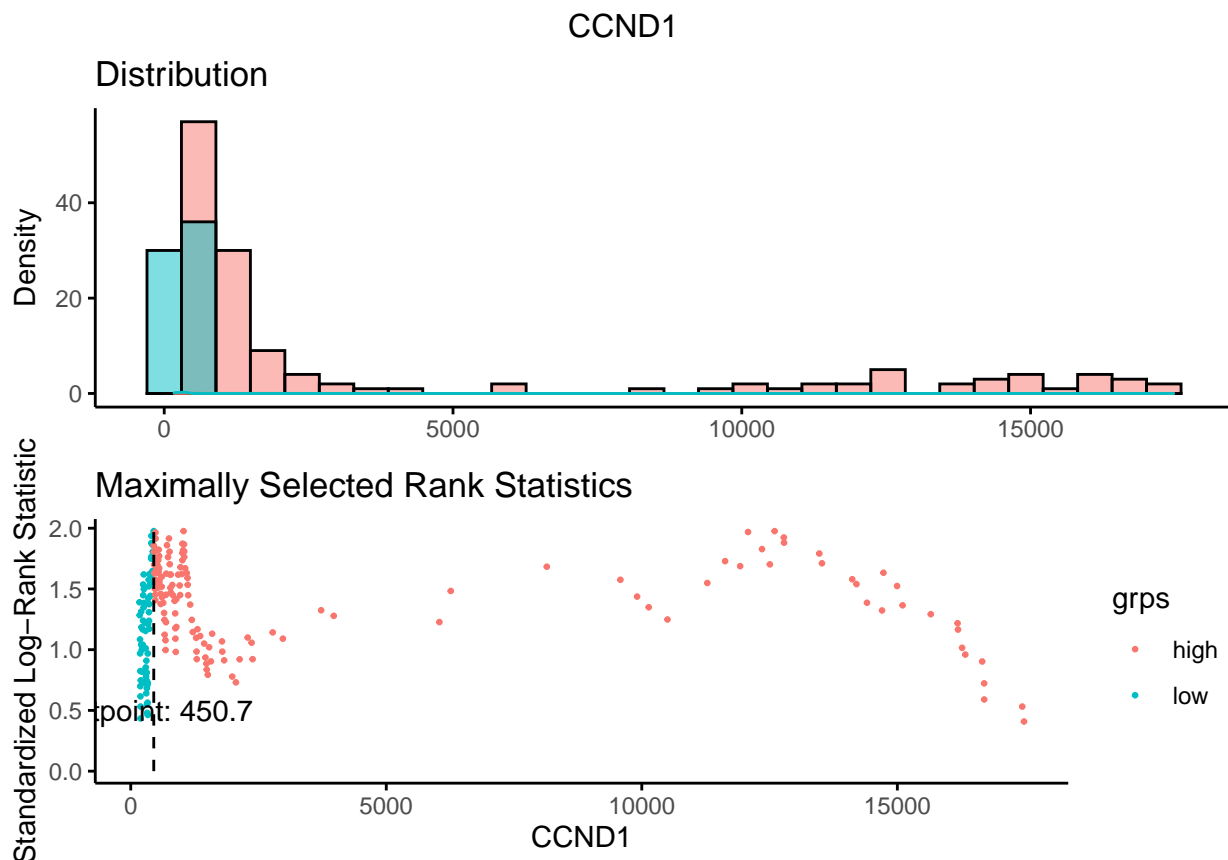
Determining the optimal cutpoint for each gene expression

```
##          cutpoint statistic
## CRIM1      82.3  1.968317
## DEPDC1    279.8  4.275452
## WHSC1    3205.6  3.361330
## CCND1     450.7  1.976398
## IRF4    12052.9  2.177788
## TP53     748.3  2.928906
```

Plot of each cutpoint

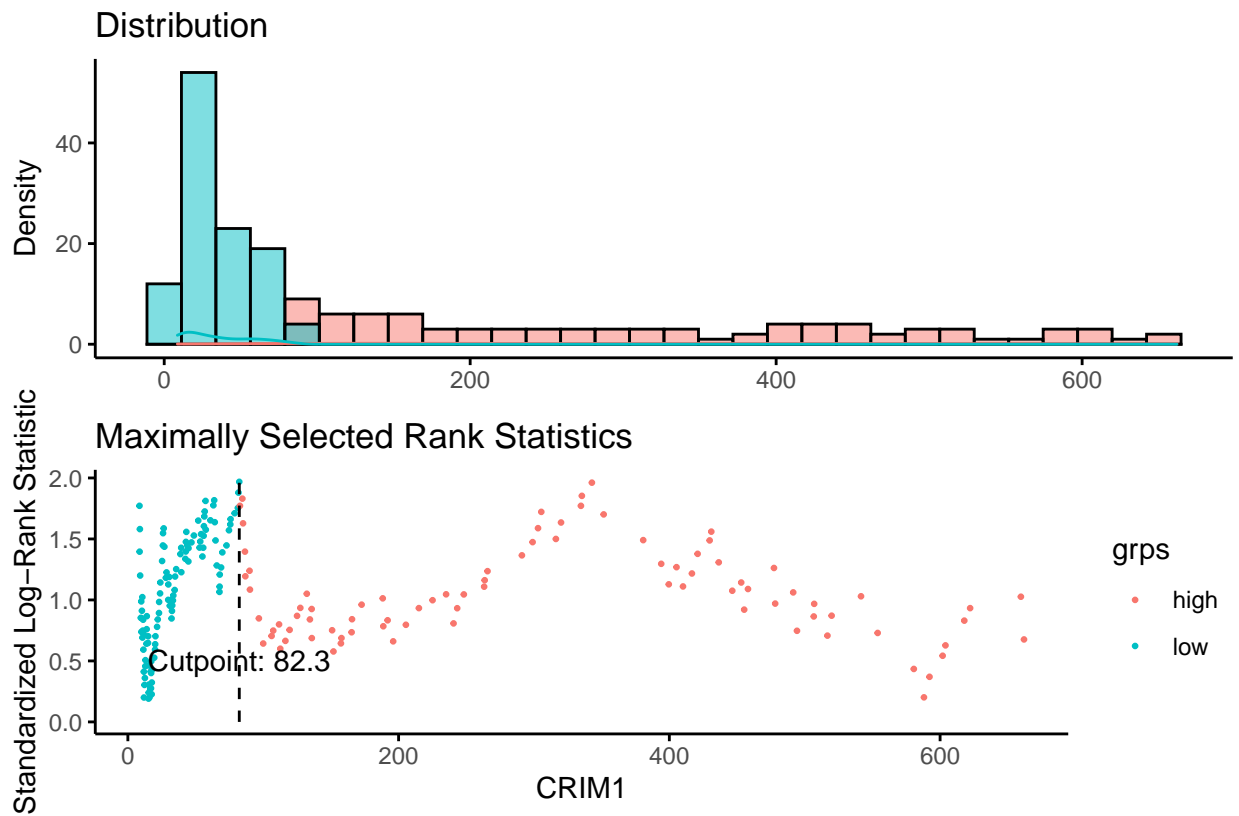
```
genes <- c("CCND1", "CRIM1", "DEPDC1", "IRF4", "TP53", "WHSC1")
#plotando a distribuicao de cada gene
for(gene in genes){
  print(plot(res.cut, gene, pallete = "npg"))
}
```

```
## $CCND1
```



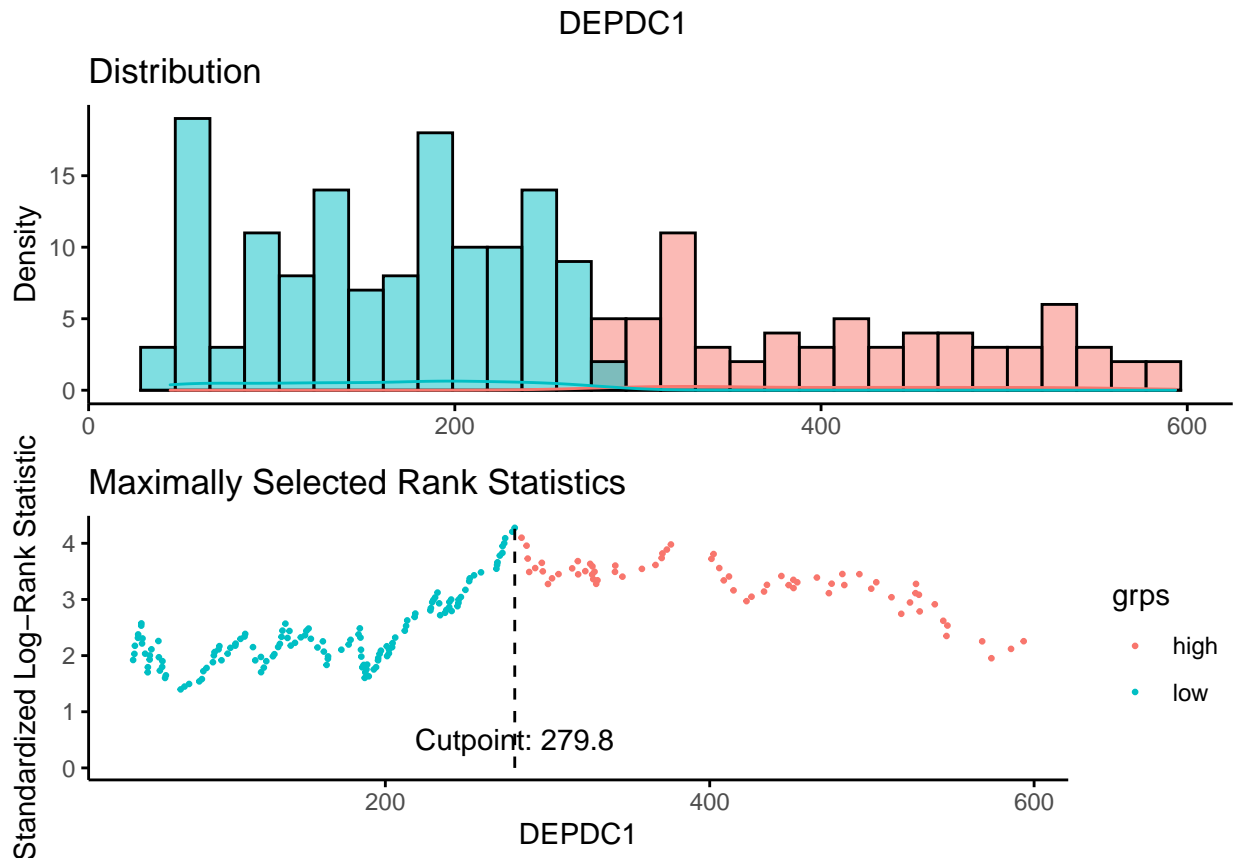
```
##
## $CRIM1
```

CRIM1

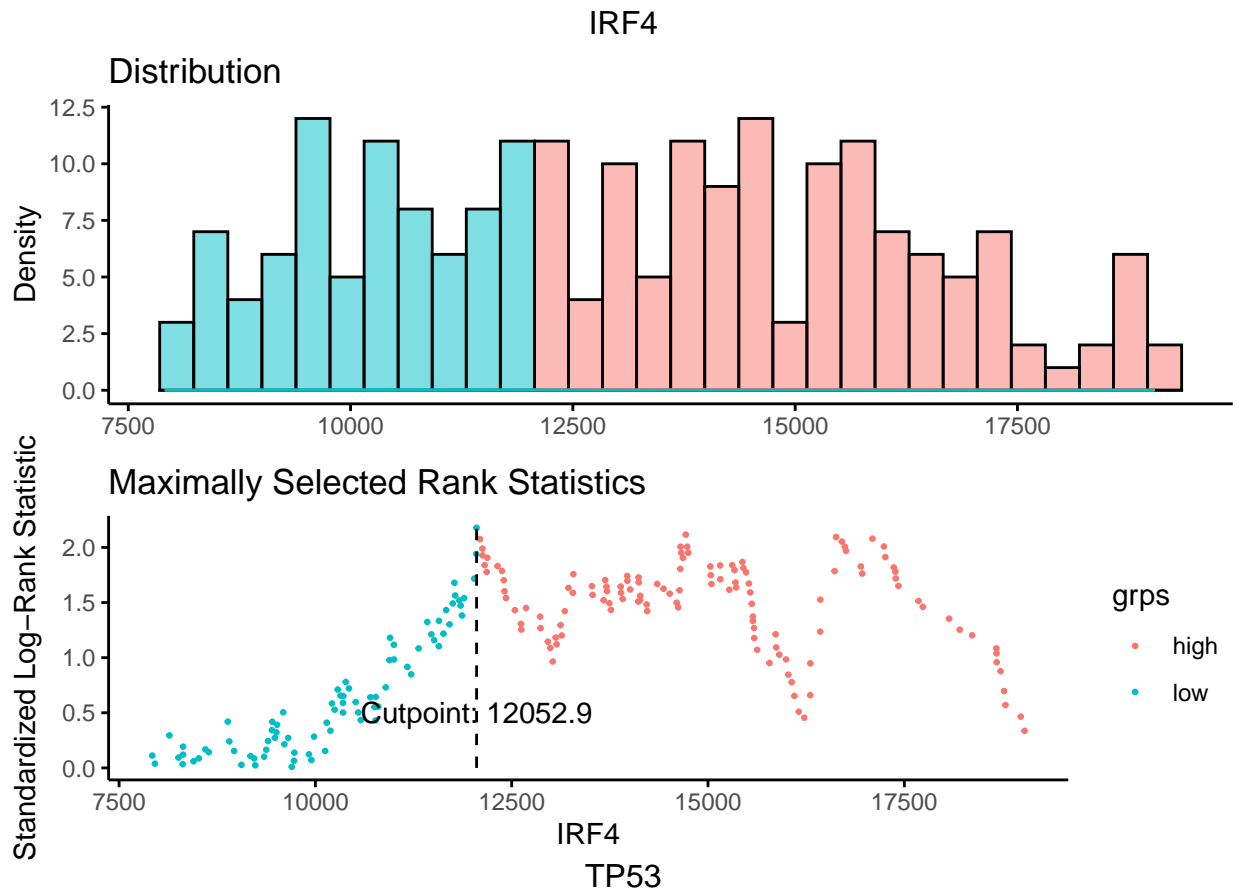


##

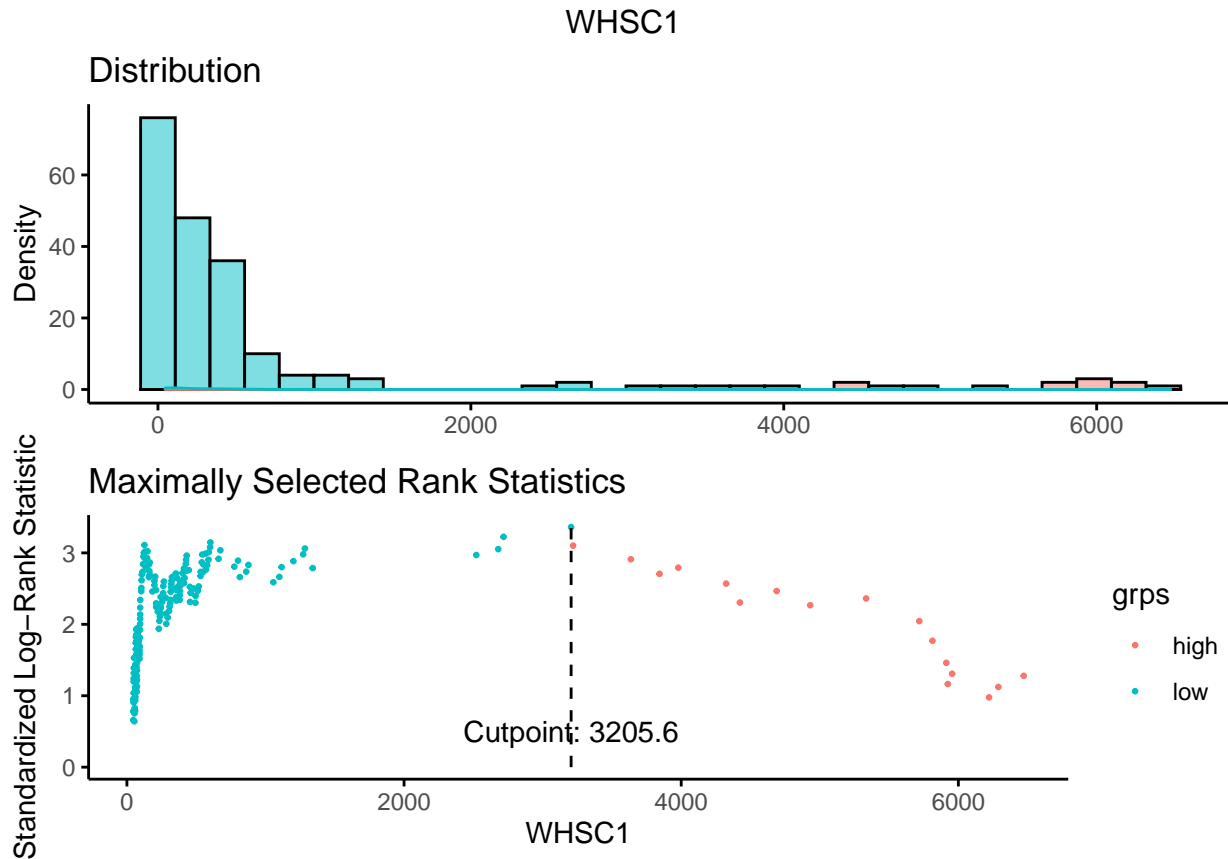
\$DEPDC1



```
##
## $IRF4
##
## $TP53
## Warning: Groups with fewer than two data points have been dropped.
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```




```
##
## $WHSC1
```



```
# plot(res.cut, "CRIM1", pallete = "npg")
```

Categorizing the variables

```
res.cat <- surv_categorize(res.cut)
head(res.cat)
```

```
##      time event CRIM1 DEPDC1 WHSC1 CCND1 IRF4 TP53
## 1 69.24     0   high   high   low   high high  low
## 2 66.43     0    low    low   low   high high  high
## 3 66.50     0   high    low   high    low low  high
## 4 42.67     1    low    low   high    low low  high
## 5 65.00     0    low    low   low   high low  low
## 6 65.20     0    low   high   low   high high  high
```

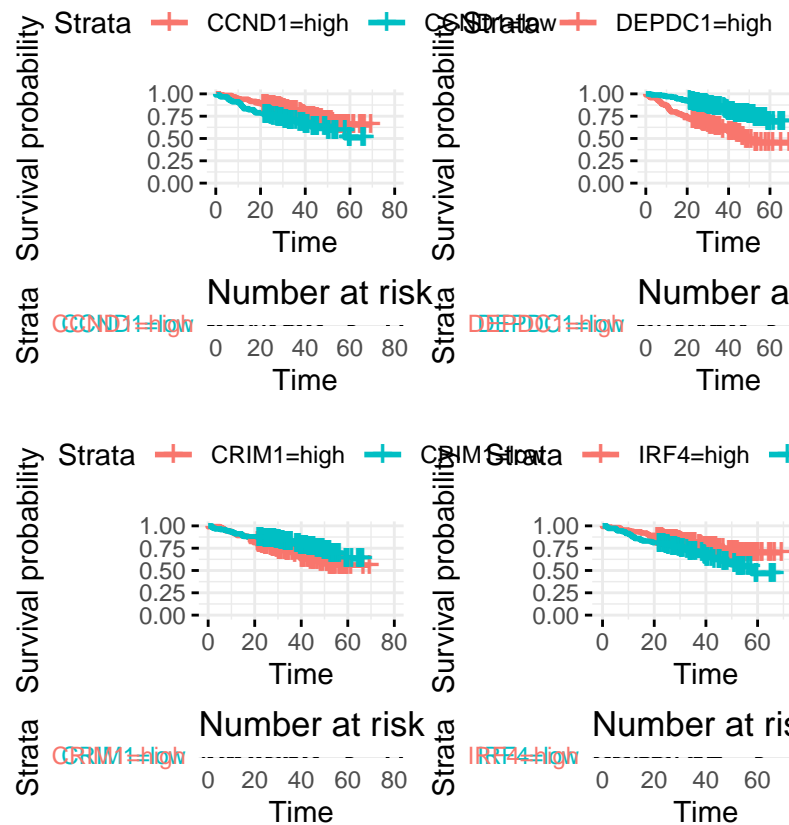
```
#plots
```

```
#defyning each fit for each gene
fit1 <- survfit(Surv(time, event) ~ CCND1, data = res.cat)
fit2 <- survfit(Surv(time, event) ~ CRIM1, data = res.cat)
fit3 <- survfit(Surv(time, event) ~ DEPDC1, data = res.cat)
fit4 <- survfit(Surv(time, event) ~ IRF4, data = res.cat)
fit5 <- survfit(Surv(time, event) ~ TP53, data = res.cat)
```

```

fit6 <- survfit(Surv(time, event) ~ WHSC1, data = res.cat)
#List of ggsurvplots
splots <- list()
splots[[1]] <- ggsurvplot(fit1, data = df, risk.table = TRUE, risk.table.height = 0.3,
                          ggtheme = theme_minimal())
splots[[2]] <- ggsurvplot(fit2, data = df, risk.table = TRUE, risk.table.height = 0.3,
                          ggtheme = theme_minimal())
splots[[3]] <- ggsurvplot(fit3, data = df, risk.table = TRUE, risk.table.height = 0.3,
                          ggtheme = theme_minimal())
splots[[4]] <- ggsurvplot(fit4, data = df, risk.table = TRUE, risk.table.height = 0.3,
                          ggtheme = theme_minimal())
splots[[5]] <- ggsurvplot(fit5, data = df, risk.table = TRUE, risk.table.height = 0.3,
                          ggtheme = theme_minimal())
splots[[6]] <- ggsurvplot(fit6, data = df, risk.table = TRUE, risk.table.height = 0.3,
                          ggtheme = theme_minimal())
#arrange multiple ggsurvplots
arrange_ggsurvplots(splots, print = TRUE,
                    ncol = 3, nrow = 2)

```



Fitting Survival Curves for each gene expression

```

if (TRUE) {
# Arrange and save into pdf file
res <- arrange_ggsurvplots(splots, print = FALSE)
ggsave("myfile.pdf", res)
}

```

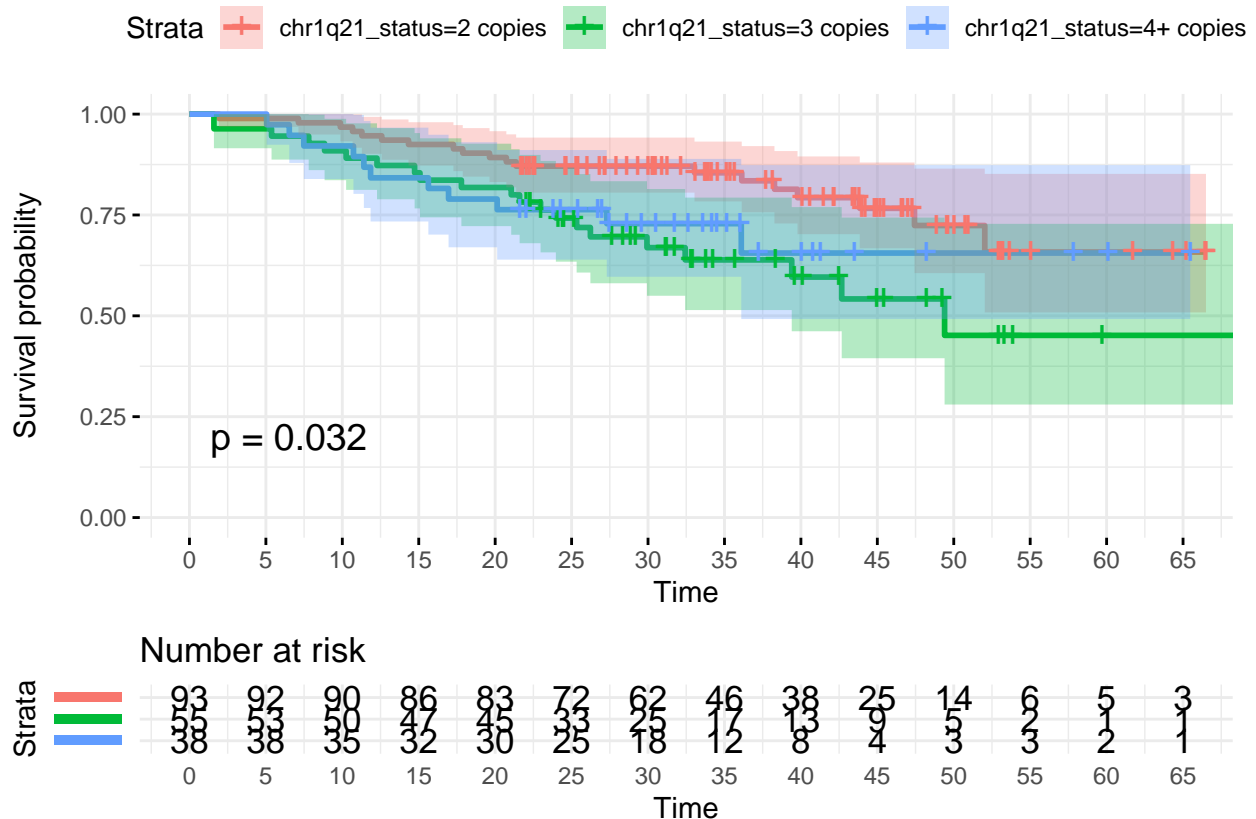
Saving 6.5 x 4.5 in image

Ratio of Distribution of Event Times

```
#surv <- ggsurvevents()
```

```
fit <- survfit(Surv(time, event) ~ chr1q21_status, data = df)
```

```
ggsurvplot(
  fit,                                # survfit object with calculated statistics.
  data = df, # data used to fit survival curves.
  risk.table = TRUE,                  # show risk table.
  pval = TRUE,                        # show p-value of log-rank test.
  conf.int = TRUE,                    # show confidence intervals for
                                     # point estimates of survival curves.
  xlim = c(0,65),                    # present narrower X axis, but not affect
                                     # survival estimates.
  break.time.by = 5,                 # break X axis in time intervals by 500.
  ggtheme = theme_minimal(),          # customize plot and risk table with a theme.
  risk.table.y.text.col = T,          # colour risk table text annotations.
  risk.table.y.text = FALSE           # show bars instead of names in text annotations
                                     # in legend of risk table
)
```



Referências

- Noll, J.E., Vandyke, K., Hewett, D.R. et al. PTTG1 expression is associated with hyperproliferative disease and poor prognosis in multiple myeloma. J Hematol Oncol 8, 106 (2015). <https://doi.org/10.1186/s13045-015-0209-2>
- João, Cristina et al. "Long-term survival in multiple myeloma." Clinical case reports vol. 2,5 (2014): 173-9. doi:10.1002/ccr3.76