

Exercício 4: Análise de sobrevida com dados da coorte TCGA-LIHC

Professor: Mauro Castro

Estudante: Clístenes Grizafis Bento

Para responder a questão abaixo, considere o conjunto de dados disponível no seguinte endereço da internet:

https://github.com/csgroen/RTN_example_TCGA_LIHC

Este conjunto de dados foi pré-processado a partir do estudo TCGA et al. (2017), que descreve uma coorte de câncer de fígado. Neste conjunto de dados está incluindo uma matriz de valores numéricos e metadados correspondentes, disponibilizado em formato .RData no arquivo:

"tcgaLIHCdata_preprocessed.RData"

Faça o download do arquivo .RData, e observe o tipo de objeto carregado no RStudio. Trata-se de um objeto da classe SummarizedExperiment, que representa um "container", contendo uma matriz numérica juntamente com os metadados, os quais incluem dados de sobrevida para n=371 casos. Em seguida, escolha uma variável de interesse para agrupar os casos, e execute uma análise de sobrevida usando o pacote survival.

FORMA DE ENVIO: Arquivo PDF

REFERÊNCIAS

The Cancer Genome Atlas Research Network (2017) Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell, **169**, 1327–1341.e23.

Example of data preprocessing for RTN and RTNsurvival using the TCGA-LIHC cohort

< https://github.com/csgroen/RTN_example_TCGA_LIHC >

Exemplo de análise de sobrevida com os pacotes "survfit" e "survminer"

Importando bibliotecas

```
In [11]: library(survival)
library(survminer)
library(SummarizedExperiment)
```

```
Warning message:
"package 'SummarizedExperiment' was built under R version 3.6.2"Loading required pack
age: GenomicRanges
Loading required package: stats4
Loading required package: BiocGenerics
Loading required package: parallel

Attaching package: 'BiocGenerics'

The following objects are masked from 'package:parallel':

    clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
    clusterExport, clusterMap, parApply, parCapply, parLapply,
    parLapplyLB, parRapply, parSapply, parSapplyLB

The following objects are masked from 'package:stats':

    IQR, mad, sd, var, xtabs

The following objects are masked from 'package:base':

    anyDuplicated, append, as.data.frame, basename, cbind, colnames,
    dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,
    grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,
    order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
    rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,
    union, unique, unsplit, which, which.max, which.min

Loading required package: S4Vectors
Warning message:
"package 'S4Vectors' was built under R version 3.6.3"
Attaching package: 'S4Vectors'

The following object is masked from 'package:base':

    expand.grid

Loading required package: IRanges
Warning message:
"package 'IRanges' was built under R version 3.6.2"
Attaching package: 'IRanges'

The following object is masked from 'package:grDevices':

    windows

Loading required package: GenomeInfoDb
Warning message:
"package 'GenomeInfoDb' was built under R version 3.6.3"Loading required package: Bio
base
Welcome to Bioconductor

    Vignettes contain introductory material; view with
    'browseVignettes()'. To cite Bioconductor, see
    'citation("Biobase")', and for packages 'citation("pkgname")'.

Loading required package: DelayedArray
Warning message:
"package 'DelayedArray' was built under R version 3.6.3"Loading required package: mat
rixStats
```

```
Warning message:
"package 'matrixStats' was built under R version 3.6.3"
Attaching package: 'matrixStats'

The following objects are masked from 'package:Biobase':

    anyMissing, rowMedians

Loading required package: BiocParallel
Warning message:
"package 'BiocParallel' was built under R version 3.6.2"
Attaching package: 'DelayedArray'

The following objects are masked from 'package:matrixStats':

    colMaxs, colMins, colRanges, rowMaxs, rowMins, rowRanges

The following objects are masked from 'package:base':

    aperm, apply, rowsum
```

Carregando pacote

```
In [8]: load("./data/tcgaLIHCdata_preprocessed.RData")
tcgaLIHCdata

class: RangedSummarizedExperiment
dim: 29885 371
metadata(1): data_release
assays(1): HTSeq - FPKM
rownames(29885): ENSG00000000003 ENSG00000000005 ... ENSG00000281910
    ENSG00000281912
rowData names(3): ENSEMBL SYMBOL OG_ENSEMBL
colnames(371): TCGA-DD-A3A3-01A-11R-A22L-07
    TCGA-DD-A1EF-01A-11R-A131-07 ... TCGA-DD-AACL-01A-11R-A41C-07
    TCGA-DD-AADW-01A-11R-A39D-07
colData names(22): bcr_patient_barcode gender ... mRNA4 mRNA5
```

Extração de dados da matrix e metadados

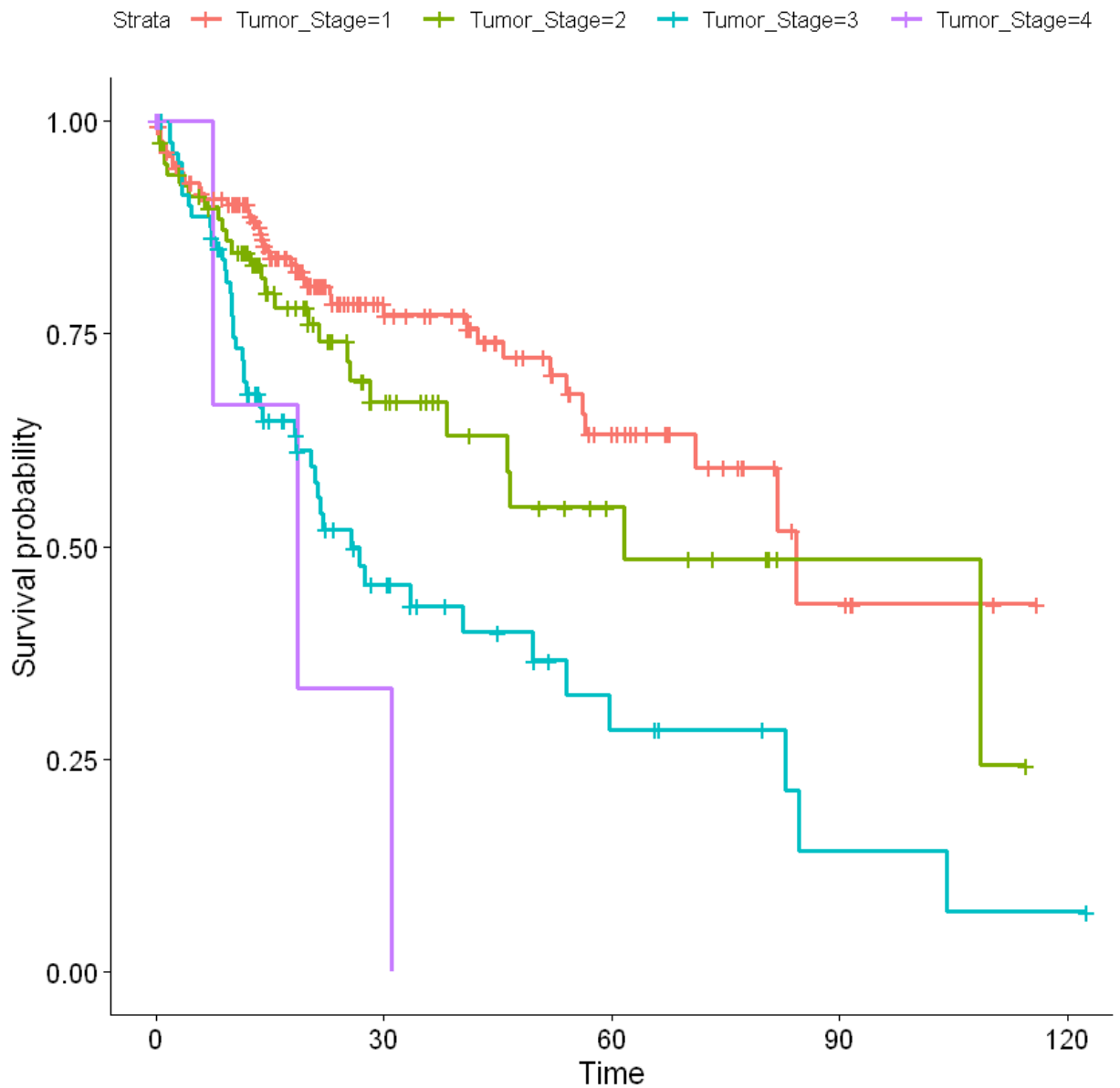
```
In [12]: gexp <- assay(tcgaLIHCdata)
rowAnnotation <- rowData(tcgaLIHCdata)
colAnnotation <- colData(tcgaLIHCdata)
```

Separando curvas de sobrevivência

```
In [25]: fit <- survfit(Surv(OS.time.months, OS) ~ Tumor_Stage, data = colAnnotation)
```

Plotagem básica de curvas

```
In [27]: gg survplot(fit, data = colAnnotation)
```



Plotagem customizada de curvas

```
In [31]: # Customized survival curves
ggsurvplot(fit, data = colAnnotation,
            surv.median.line = "hv", # Add medians survival

            # Change legends: title & labels
            legend.title = "Tumor Stage",
            legend.labs = c("I", "II", "III", "IV"),
            # Add p-value and tervals
            pval = TRUE,

            conf.int = TRUE,
            # Add risk table
            risk.table = TRUE,
            tables.height = 0.2,
            tables.theme = theme_cleantable(),

            # Color palettes. Use custom color: c("#E7B800", "#2E9FDF"),
```

```
# or brewer color (e.g.: "Dark2"), or ggsci color (e.g.: "jco")
palette = c("#E7B800", "#2E9FDF", "#FF0000", "#00FF00"),
ggtheme = theme_bw() # Change ggplot2 theme
)
```

