

## Pràctica 2

Grup 3 - Clara Albert, Enric Sintes, Joel Baltasar, Sergi Orozco

Maig del 2021

### Importació de les dades

```
library(haven)
datos=read_sas("P2_baseline.sas7bdat", NULL)
```

### Qualitat de les dades

Veiem l'estructura de les dades.

```
str(datos)

## tibble [202 x 23] (S3: tbl_df/tbl/data.frame)
## $ PatNo          : chr [1:202] "12862484" "12881071" "12890790"
##                  : chr [1:202] "12901286" ...
## $ treat          : num [1:202] 0 0 0 0 0 0 0 0 0 0 ...
## ..- attr(*, "label")= chr "Treat"
## $ EDAD_DON       : num [1:202] 58 52 57 57 50 59 51 39 52 65 ...
## ..- attr(*, "label")= chr "Age (y) - Donor"
## ..- attr(*, "format.sas")= chr "F"
## $ Gender_Don     : num [1:202] 0 1 0 1 1 1 0 0 0 1 ...
## ..- attr(*, "label")= chr "Gender - Donor"
## ..- attr(*, "format.sas")= chr "GENDER"
## $ Cause_death_CVA : num [1:202] 0 0 0 1 1 1 0 0 0 0 ...
## ..- attr(*, "label")= chr "Cause of death: CVA"
## $ Cause_death_Anoxia : num [1:202] 0 1 1 0 0 0 1 1 1 1 ...
## ..- attr(*, "label")= chr "Cause of death: Anoxic brain injury"
## $ Cause_death_TBI  : num [1:202] 0 0 0 0 0 0 0 0 0 0 ...
## ..- attr(*, "label")= chr "Cause of death: Traumatic brain injury"
## $ Cause_death_Other : num [1:202] 1 0 0 0 0 0 0 0 0 0 ...
## ..- attr(*, "label")= chr "Cause of death: Others"
## $ TmUCILtsv       : num [1:202] 47 8 11 5 5 8 5 14 12 9 ...
## ..- attr(*, "label")= chr "ICU stay (days)"
## $ TmIsqCalTotAbd   : num [1:202] 22 18 25 20 21 25 21 20 41 22 ...
## ..- attr(*, "label")= chr "Total WITb (min)"
## $ TmIsqCalFxAAbd   : num [1:202] 20 17 18 15 19 20 16 12 11 12 ...
## ..- attr(*, "label")= chr "Functional WITc (min)"
## $ TIEMPOISQUEMIAFR_A_RETH: num [1:202] 350 450 480 150 365 370 405 435 300
##                  : num [1:202] 240 ...
## ..- attr(*, "label")= chr "CIT (min)"
## ..- attr(*, "format.sas")= chr "F"
## $ LIQUIDOPRESERV_UW_IGL1 : num [1:202] 0 0 0 0 1 0 0 0 0 0 ...
## ..- attr(*, "label")= chr "Preservation solution: UW or IGL-1"
## $ LIQUIDOPRESERV_HTK     : num [1:202] 0 0 0 0 0 0 1 0 0 0 ...
```

```
## ..- attr(*, "label")= chr "Preservation solution: HTK"
## $ LIQUIDOPRESERV_CELSIOR : num [1:202] 1 1 1 1 0 1 0 1 1 1 ...
## ..- attr(*, "label")= chr "Preservation solution: Celsior"
## $ EDAD_REC_TX : num [1:202] 54 66 59 45 56 60 57 46 65 66 ...
## ..- attr(*, "label")= chr "Age (yr) - Recipient"
## ..- attr(*, "format.sas")= chr "F"
## $ Gender_Rec : num [1:202] 0 1 0 1 1 1 1 1 0 1 ...
## ..- attr(*, "label")= chr "Gender - Recipient"
## ..- attr(*, "format.sas")= chr "GENDER"
## $ MELDScore : num [1:202] 23 27 18 34 26 11 8 21 17 8 ...
## ..- attr(*, "label")= chr "Laboratory MELD score"
## ..- attr(*, "format.sas")= chr "F"
## $ Volume : num [1:202] 0 1 1 0 1 0 0 1 0 0 ...
## ..- attr(*, "label")= chr "High-volume transplant centre"
## ..- attr(*, "format.sas")= chr "VOL"
## $ FIN_INDIC_reTx_fulmi : num [1:202] 0 0 0 0 0 0 0 0 0 0 ...
## ..- attr(*, "label")= chr "Transplant indication: Retransplantation or
fulminant liver failure"
## $ FIN_INDIC_HCC : num [1:202] 0 0 0 0 0 1 1 0 0 1 ...
## ..- attr(*, "label")= chr "Transplant indication: Hepatocellular carcinoma"
## $ FIN_INDIC_Cirr : num [1:202] 1 1 1 1 1 0 0 1 1 0 ...
## ..- attr(*, "label")= chr "Transplant indication: Cirrhosis"
## $ FIN_INDIC_Other : num [1:202] 0 0 0 0 0 0 0 0 0 0 ...
## ..- attr(*, "label")= chr "Transplant indication: Other"
## - attr(*, "label")= chr "P2_BASELINE"
```

Veiem que les dades ja tenen l'estructura que ens interessa. Aquelles que tenen dues categories ja estan factoritzades. Una vegada comprovat que el seu format és el correcte, realitzarem un resum estadístic per veure si hi ha alguna dada que no està preparada per l'anàlisi.

`summary(datos)`

```
##      PatNo          treat      EDAD_DON      Gender_Don
## Length:202      Min.   :0.0000      Min.   : 2.00      Min.   :0.0000
## Class :character 1st Qu.:0.0000      1st Qu.:47.00      1st Qu.:0.0000
## Mode  :character Median :0.0000      Median :56.00      Median :1.0000
##                      Mean   :0.4505      Mean   :54.39      Mean   :0.6683
##                      3rd Qu.:1.0000      3rd Qu.:64.00      3rd Qu.:1.0000
##                      Max.   :1.0000      Max.   :77.00      Max.   :1.0000
## Cause_death_CVA Cause_death_Anoxia Cause_death_TBI
Cause_death_Other
## Min.   :0.0000      Min.   :0.000      Min.   :0.00000      Min.   :0.00000
## 1st Qu.:0.0000      1st Qu.:0.000      1st Qu.:0.00000      1st Qu.:0.00000
## Median :0.0000      Median :0.000      Median :0.00000      Median :0.00000
## Mean   :0.4406      Mean   :0.401      Mean   :0.09406      Mean   :0.06436
## 3rd Qu.:1.0000      3rd Qu.:1.000      3rd Qu.:0.00000      3rd Qu.:0.00000
## Max.   :1.0000      Max.   :1.000      Max.   :1.00000      Max.   :1.00000
## TmUCILtsv      TmIsqCalTotAbd TmIsqCalFxAAbd
TIEMPOISQUEMIAFR_A_RETH
## Min.   : 0.000      Min.   : 7.00      Min.   : 5.00      Min.   :133.0
```

```
## 1st Qu.: 5.000 1st Qu.:16.00 1st Qu.:11.00 1st Qu.:275.5
## Median : 7.000 Median :20.00 Median :14.00 Median :330.5
## Mean : 9.144 Mean :21.17 Mean :14.75 Mean :337.5
## 3rd Qu.:11.000 3rd Qu.:25.00 3rd Qu.:18.00 3rd Qu.:380.0
## Max. :80.000 Max. :62.00 Max. :30.00 Max. :810.0
## LIQUIDOPRESERV_UW_IGL1 LIQUIDOPRESERV_HTK LIQUIDOPRESERV_CELSIOR
## Min. :0.0000 Min. :0.0000 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0.0000 Median :0.0000 Median :1.0000
## Mean :0.2327 Mean :0.1386 Mean :0.6287
## 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:1.0000
## Max. :1.0000 Max. :1.0000 Max. :1.0000
## EDAD_REC_TX Gender_Rec MELDSCORE Volume
## Min. : 1.00 Min. :0.0000 Min. : 5.0 Min. :0.0000
## 1st Qu.:53.00 1st Qu.:1.0000 1st Qu.: 9.0 1st Qu.:0.0000
## Median :57.00 Median :1.0000 Median :14.0 Median :1.0000
## Mean :56.26 Mean :0.8267 Mean :14.5 Mean :0.7426
## 3rd Qu.:62.00 3rd Qu.:1.0000 3rd Qu.:18.0 3rd Qu.:1.0000
## Max. :81.00 Max. :1.0000 Max. :37.0 Max. :1.0000
## FIN_INDIC_reTx_fulmi FIN_INDIC_HCC FIN_INDIC_Cirr
FIN_INDIC_Other
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min.
:0.00000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st
Qu.:0.00000
## Median :0.0000 Median :0.0000 Median :1.0000 Median
:0.00000
## Mean :0.0198 Mean :0.3317 Mean :0.6139 Mean
:0.03465
## 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd
Qu.:0.00000
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max.
:1.00000
```

En les variables categòriques veiem que no hi ha cap discrepància, ja que tots els valors són 0 o 1. Per a les variables contínues, veiem un outlier en l'edat del donant, ja que el mínim és de 2 anys. Tot i això, no ens interfereix al nostre estudi que realitzarem.

## Missing values

```
apply(is.na(datos),2,sum)
```

```
##          PatNo          treat
EDAD_DON
##          0          0
0
##          Gender_Don          Cause_death_CVA
Cause_death_Anoxia
```

```
##          0          0
0
##          Cause_death_TBI          Cause_death_Other
TmUCILtsv
##          0          0
0
##          TmIsqCalTotAbd          TmIsqCalFxAbd
TIEMPOISQUEMIAFR_A_RETH
##          0          0
0
##  LIQUIDOPRESERV_UW_IGL1          LIQUIDOPRESERV_HTK
LIQUIDOPRESERV_CELSIOR
##          0          0
0
##          EDAD_REC_TX          Gender_Rec
MELDSCORE
##          0          0
0
##          Volume          FIN_INDIC_reTx_fulmi
FIN_INDIC_HCC
##          0          0
0
##          FIN_INDIC_Cirr          FIN_INDIC_Other
##          0          0
```

Veiem que no hi ha cap dada amb missing values.

## Comparativa dades basals

Per comparar els dos grups utilitzarem la funció `compareGroups` on els descriptius que utilitzarem seran:

- Per a variables categòriques: N i percentatge
- Per a variables numèriques: mediana i primer i tercer quartil

A més a més, afegirem la diferència mitjana estandarditzada per veure si les dades entre els dos grups estan molt desequilibrades (aquest valor és superior a 0.2).

**library(dplyr)**

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```

## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union

library(tableone)

cols.to.factor <- sapply(datos, function(col) length(unique(col)) < 3);
head(cols.to.factor)

##           PatNo           treat           EDAD_DON
Gender_Don
##           FALSE           TRUE           FALSE
TRUE
## Cause_death_CVA Cause_death_Anoxia
##           TRUE           TRUE

datos[cols.to.factor] <- lapply(datos[ cols.to.factor] , factor)

vars = colnames(datos); vars = vars[-1]; vars

## [1] "treat"           "EDAD_DON"
## [3] "Gender_Don"         "Cause_death_CVA"
## [5] "Cause_death_Anoxia" "Cause_death_TBI"
## [7] "Cause_death_Other"  "TmUCILtsv"
## [9] "TmIsqCalTotAbd"     "TmIsqCalFxAAbd"
## [11] "TIEMPOISQUEMIAFR_A_RETH" "LIQUIDOPRESERV_UW_IGL1"
## [13] "LIQUIDOPRESERV_HTK"   "LIQUIDOPRESERV_CELSIOR"
## [15] "EDAD_REC_TX"         "Gender_Rec"
## [17] "MELDScore"          "Volume"
## [19] "FIN_INDIC_reTx_fulmi" "FIN_INDIC_HCC"
## [21] "FIN_INDIC_Cirr"      "FIN_INDIC_Other"

continues=names(datos[cols.to.factor==FALSE])

## Construct a table
tabUnmatched <- CreateTableOne(vars = vars, strata = "treat", data =
datos, test = TRUE)

## Show table with SMD
print(tabUnmatched, nonnormal = continues, smd = TRUE)

##           Stratified by treat
##           0
## n 111
## treat = 1 (%) 0 ( 0.0)
## EDAD_DON (median [IQR]) 56.00 [48.00, 63.00]
## Gender_Don = 1 (%) 75 (67.6)
## Cause_death_CVA = 1 (%) 48 (43.2)
## Cause_death_Anoxia = 1 (%) 45 (40.5)
## Cause_death_TBI = 1 (%) 11 ( 9.9)
## Cause_death_Other = 1 (%) 7 ( 6.3)
## TmUCILtsv (median [IQR]) 7.00 [5.00, 11.00]
## TmIsqCalTotAbd (median [IQR]) 22.00 [19.00, 26.00]

```

##	TmIsqCalFxAAbd (median [IQR])	15.00 [12.00, 20.00]	
##	TIEMPOISQUEMIAFR_A_RETH (median [IQR])	340.00 [285.00, 385.50]	
##	LIQUIDOPRESERV_UW_IGL1 = 1 (%)	11 ( 9.9)	
##	LIQUIDOPRESERV_HTK = 1 (%)	27 (24.3)	
##	LIQUIDOPRESERV_CELSIOR = 1 (%)	73 (65.8)	
##	EDAD_REC_TX (median [IQR])	59.00 [53.00, 62.50]	
##	Gender_Rec = 1 (%)	96 (86.5)	
##	MELDSCORE (median [IQR])	13.00 [9.00, 17.50]	
##	Volume = 1 (%)	84 (75.7)	
##	FIN_INDIC_reTx_fulmi = 1 (%)	2 ( 1.8)	
##	FIN_INDIC_HCC = 1 (%)	34 (30.6)	
##	FIN_INDIC_Cirr = 1 (%)	73 (65.8)	
##	FIN_INDIC_Other = 1 (%)	2 ( 1.8)	
##		Stratified by treat	
##		1	p
##	n	91	
##	treat = 1 (%)	91 (100.0)	<0.001
##	EDAD_DON (median [IQR])	57.00 [45.00, 65.00]	0.810
##	Gender_Don = 1 (%)	60 ( 65.9)	0.924
##	Cause_death_CVA = 1 (%)	41 ( 45.1)	0.908
##	Cause_death_Anoxia = 1 (%)	36 ( 39.6)	1.000
##	Cause_death_TBI = 1 (%)	8 ( 8.8)	0.977
##	Cause_death_Other = 1 (%)	6 ( 6.6)	1.000
##	TmUCILtsv (median [IQR])	7.00 [4.00, 10.50]	0.427
##	TmIsqCalTotAbd (median [IQR])	18.00 [13.00, 23.00]	<0.001
##	TmIsqCalFxAAbd (median [IQR])	12.00 [9.50, 15.00]	<0.001
##	TIEMPOISQUEMIAFR_A_RETH (median [IQR])	313.00 [264.00, 362.50]	0.109
##	LIQUIDOPRESERV_UW_IGL1 = 1 (%)	36 ( 39.6)	<0.001
##	LIQUIDOPRESERV_HTK = 1 (%)	1 ( 1.1)	<0.001
##	LIQUIDOPRESERV_CELSIOR = 1 (%)	54 ( 59.3)	0.427
##	EDAD_REC_TX (median [IQR])	56.00 [52.00, 61.50]	0.243
##	Gender_Rec = 1 (%)	71 ( 78.0)	0.163
##	MELDSCORE (median [IQR])	15.00 [11.00, 18.50]	0.111
##	Volume = 1 (%)	66 ( 72.5)	0.728
##	FIN_INDIC_reTx_fulmi = 1 (%)	2 ( 2.2)	1.000
##	FIN_INDIC_HCC = 1 (%)	33 ( 36.3)	0.487
##	FIN_INDIC_Cirr = 1 (%)	51 ( 56.0)	0.205
##	FIN_INDIC_Other = 1 (%)	5 ( 5.5)	0.298
##		Stratified by treat	
##		SMD	
##	n		
##	treat = 1 (%)	NaN	
##	EDAD_DON (median [IQR])	0.079	
##	Gender_Don = 1 (%)	0.035	
##	Cause_death_CVA = 1 (%)	0.036	
##	Cause_death_Anoxia = 1 (%)	0.020	
##	Cause_death_TBI = 1 (%)	0.038	
##	Cause_death_Other = 1 (%)	0.012	
##	TmUCILtsv (median [IQR])	0.016	
##	TmIsqCalTotAbd (median [IQR])	0.530	
##	TmIsqCalFxAAbd (median [IQR])	0.562	
##	TIEMPOISQUEMIAFR_A_RETH (median [IQR])	0.090	
##	LIQUIDOPRESERV_UW_IGL1 = 1 (%)	0.732	
##	LIQUIDOPRESERV_HTK = 1 (%)	0.744	
##	LIQUIDOPRESERV_CELSIOR = 1 (%)	0.133	
##	EDAD_REC_TX (median [IQR])	0.261	
##	Gender_Rec = 1 (%)	0.223	
##	MELDSCORE (median [IQR])	0.210	
##	Volume = 1 (%)	0.072	
##	FIN_INDIC_reTx_fulmi = 1 (%)	0.028	
##	FIN_INDIC_HCC = 1 (%)	0.120	

```
## FIN_INDIC_Cirr = 1 (%) 0.200
## FIN_INDIC_Other = 1 (%) 0.198
```

Les xifres de les variables FIN\_INDIC\_HCC, FIN\_INDIC\_Cirr, FIN\_INDIC\_Other, MELDScore, Gender\_Rec, EDAD\_REC\_TX, LIQUIDOPRESERV\_UW\_IGL1, LIQUIDOPRESERV\_HTK, LIQUIDOPRESERV\_CELSIOR, TmIsqCalTotAbd i TmIsqCalFxAbd són les més desproporcionades respecte a les restants.

```
library(survey)

## Loading required package: grid
## Loading required package: Matrix
## Loading required package: survival
##
## Attaching package: 'survey'

## The following object is masked from 'package:graphics':
##
## dotchart

#propensity score model
psmodel <- glm(treat ~ ., data=datos, family=binomial(link="logit"))

## Warning: glm.fit: algorithm did not converge

# valor del propensity score per cada individu.
ps <- predict(psmodel, type="response")

#Pesos per cada individu
weight<-ifelse(datos$treat==1,(1/ps),(1/(1-ps)))

#Apliquem els pesos a les dades
weightdata<-svydesign(ids=~1,data=datos,weights=weight)
weightedtable<-svyCreateTableOne(vars=vars,strata="treat",
data=weightdata,test=FALSE)
print(weightedtable,smd=TRUE)

##
## Stratified by treat
##      0      1      SMD
## n      111.0      91.0
## treat = 1 (%)      0.0 ( 0.0)      91.0 (100.0)      NaN
## EDAD_DON (mean (SD))      54.86 (11.18)      53.81 (15.28)      0.079
## Gender_Don = 1 (%)      75.0 (67.6)      60.0 ( 65.9)      0.035
## Cause_death_CVA = 1 (%)      48.0 (43.2)      41.0 ( 45.1)      0.036
## Cause_death_Anoxia = 1 (%)      45.0 (40.5)      36.0 ( 39.6)      0.020
## Cause_death_TBI = 1 (%)      11.0 ( 9.9)      8.0 (  8.8)      0.038
## Cause_death_Other = 1 (%)      7.0 ( 6.3)      6.0 (  6.6)      0.012
## TmUCILtsv (mean (SD))      9.21 (7.90)      9.07 (10.02)      0.016
## TmIsqCalTotAbd (mean (SD))      22.95 (6.62)      18.99 (8.24)      0.530
```

##	TmIsqCalFxAAbd (mean (SD))	16.07 (5.28)	13.13 (5.19)	0.562
##	TIEMPOISQUEMIAFR_A_RETH (mean (SD))	341.75 (95.81)	332.38 (111.46)	0.090
##	LIQUIDOPRESERV_UW_IGL1 = 1 (%)	11.0 ( 9.9)	36.0 ( 39.6)	0.732
##	LIQUIDOPRESERV_HTK = 1 (%)	27.0 (24.3)	1.0 ( 1.1)	0.744
##	LIQUIDOPRESERV_CELSIOR = 1 (%)	73.0 (65.8)	54.0 ( 59.3)	0.133
##	EDAD_REC_TX (mean (SD))	57.43 (7.15)	54.84 (12.12)	0.261
##	Gender_Rec = 1 (%)	96.0 (86.5)	71.0 ( 78.0)	0.223
##	MELDSCORE (mean (SD))	13.92 (6.15)	15.22 (6.25)	0.210
##	Volume = 1 (%)	84.0 (75.7)	66.0 ( 72.5)	0.072
##	FIN_INDIC_reTx_fulmi = 1 (%)	2.0 ( 1.8)	2.0 ( 2.2)	0.028
##	FIN_INDIC_HCC = 1 (%)	34.0 (30.6)	33.0 ( 36.3)	0.120
##	FIN_INDIC_Cirr = 1 (%)	73.0 (65.8)	51.0 ( 56.0)	0.200
##	FIN_INDIC_Other = 1 (%)	2.0 ( 1.8)	5.0 ( 5.5)	0.198

Amb el model amb totes les variables no aconseguim unes dades balancejades, per tant provarem amb altres models. Partint d'un model amb totes les variables que tenien un SMD major que 0.2 al model complet hem construït el següent model fent algunes modificacions perquè obtinguem totes les variables balancejades.

```
psmodel2 <- glm(treat ~ EDAD_DON + TmIsqCalTotAbd + sqrt(TmIsqCalFxAAbd) +
LIQUIDOPRESERV_UW_IGL1 + LIQUIDOPRESERV_HTK + EDAD_REC_TX + Gender_Rec +
FIN_INDIC_Cirr + Volume + TIEMPOISQUEMIAFR_A_RETH, data=datos,
family=binomial(link="logit"))
```

```
# valor del propensity score per cada individu.
datos$ps2 <- predict(psmodel2, type="response")
```

```
#Calulem les participacions de les variables
part <- sum(datos$treat==1)/length(datos$treat)
```

```
#Pesos per cada individu
datos$weight2 <- ifelse(datos$treat==1,(1/datos$ps2)*part,(1/(1-
datos$ps2))*(1-part))
```

```
#Apliquem els pesos a les dades
weightdata2 <- svydesign(ids=~1,data=datos,weights=datos$weight2)
weightedtable2<-svyCreateTableOne(vars=vars,strata="treat",
data=weightdata2,test=FALSE)
print(weightedtable2,smd=TRUE)
```

##	Stratified by treat			
##	0	1	SMD	
##	n	105.7	98.9	
##	treat = 1 (%)	0.0 ( 0.0)	98.9 (100.0)	NaN
##	EDAD_DON (mean (SD))	54.56 (12.13)	53.04 (16.47)	0.105
##	Gender_Don = 1 (%)	72.8 (68.9)	66.1 ( 66.9)	0.043
##	Cause_death_CVA = 1 (%)	46.5 (44.0)	39.1 ( 39.6)	0.090
##	Cause_death_Anoxia = 1 (%)	45.2 (42.8)	49.6 ( 50.1)	0.147
##	Cause_death_TBI = 1 (%)	9.2 ( 8.7)	6.5 ( 6.6)	0.080
##	Cause_death_Other = 1 (%)	4.8 ( 4.6)	3.7 ( 3.8)	0.039



##	TmUCILtsv (mean (SD))	8.77 (7.29)	9.14 (8.68)	0.046
##	TmIsqCalTotAbd (mean (SD))	21.96 (6.44)	23.70 (12.90)	0.171
##	TmIsqCalFxAbd (mean (SD))	15.00 (4.88)	16.10 (7.31)	0.177
##	TIEMPOISQUEMIAFR_A_RETH (mean (SD))	344.06 (92.05)	355.24 (127.38)	0.101
##	LIQUIDOPRESERV_UW_IGL1 = 1 (%)	22.0 (20.8)	21.2 ( 21.5)	0.016
##	LIQUIDOPRESERV_HTK = 1 (%)	15.4 (14.6)	11.3 ( 11.4)	0.093
##	LIQUIDOPRESERV_CELSIOR = 1 (%)	68.3 (64.6)	66.4 ( 67.1)	0.052
##	EDAD_REC_TX (mean (SD))	56.43 (7.57)	55.89 (9.42)	0.063
##	Gender_Rec = 1 (%)	87.5 (82.7)	84.0 ( 84.9)	0.059
##	MELDSCORE (mean (SD))	14.73 (6.87)	14.65 (5.53)	0.014
##	Volume = 1 (%)	78.7 (74.4)	72.6 ( 73.4)	0.023
##	FIN_INDIC_reTx_fulmi = 1 (%)	3.2 ( 3.0)	1.4 ( 1.4)	0.112
##	FIN_INDIC_HCC = 1 (%)	32.2 (30.5)	26.0 ( 26.3)	0.092
##	FIN_INDIC_Cirr = 1 (%)	68.7 (65.0)	68.7 ( 69.5)	0.095
##	FIN_INDIC_Other = 1 (%)	1.6 ( 1.5)	2.8 ( 2.9)	0.089

## Part 2 - Anàlisi amb inverse of probability of treatment weighting

### 1- Descripció de la supervivència de l'empelt i dels pacients

```
library(haven)
outcomes <- read_sas("outcomes.sas7bdat", NULL)
datos2 <- merge(datos, outcomes, by="PatNo")

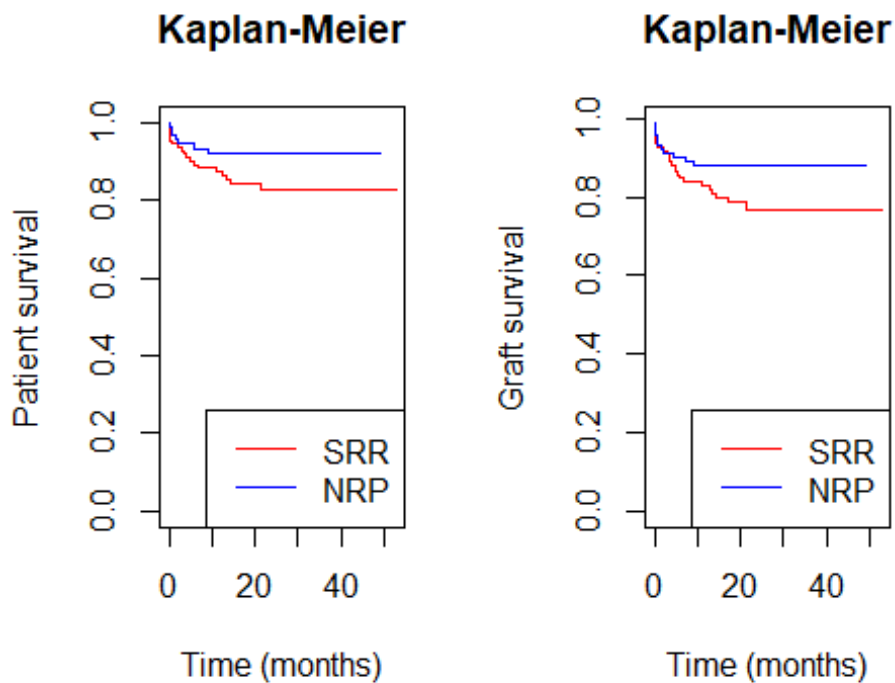
library(survival)

km.model <- survfit(Surv(DeathTm, DeathYN)~treat, type="kaplan-meier",
data=datos2)

km.model2 <- survfit(Surv(GraftNotFxTm, GraftNotFxYN)~treat,
type="kaplan-meier", data=datos2)

{par(mfrow=c(1,2))
plot(km.model, conf.int=F, xlab="Time (months)", ylab="Patient
survival", main="Kaplan-Meier", col=c("red", "blue"))
legend("bottomright", c("SRR", "NRP"), col=c("red", "blue"), lty=c(1,1))

plot(km.model2, conf.int=F, xlab="Time (months)", ylab="Graft
survival", main="Kaplan-Meier", col=c("red", "blue"))
legend("bottomright", c("SRR", "NRP"), col=c("red", "blue"), lty=c(1,1))
}
```



### Ho podríem fer amb IPTW?

Si que es podria fer la representació de la funció de supervivència amb IPTW, ja que tenim uns pesos per a cada individu que els podríem afegir a la base de dades i tindríem unes noves dades per poder aplicar el mètode de Kaplan-Meier.

## 2- Anàlisi inferencial cru i ajustat de comparació de riscos en anàlisi de supervivència

Calcularíem el log rank per veure diferències entre grups i saber si són o no iguals.

```
surv_diff <- survdiff(Surv(DeathTm,DeathYN)~treat, data = datos2)
surv_diff

## Call:
## survdiff(formula = Surv(DeathTm, DeathYN) ~ treat, data = datos2)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## treat=0 111      18      13.6      1.42      3.11
## treat=1  91       7      11.4      1.69      3.11
##
##  Chisq= 3.1  on 1 degrees of freedom, p= 0.08
```

Per calcular els riscos del model utilitzarem la funció coxph per calcular el Hazard Ratio

```

cox.mod<-coxph(Surv(DeathTm,DeathYN)~treat,data=datos2, weights=weight2,
method="breslow")
summary(cox.mod)

## Call:
## coxph(formula = Surv(DeathTm, DeathYN) ~ treat, data = datos2,
##       weights = weight2, method = "breslow")
##
##      n= 202, number of events= 25
##
##              coef exp(coef) se(coef) robust se      z Pr(>|z|)
## treat1 -0.6788    0.5072   0.4443    0.5295 -1.282    0.2
##
##      exp(coef) exp(-coef) lower .95 upper .95
## treat1    0.5072    1.971    0.1797    1.432
##
## Concordance= 0.574 (se = 0.061 )
## Likelihood ratio test= 2.49 on 1 df,  p=0.1
## Wald test               = 1.64 on 1 df,  p=0.2
## Score (logrank) test = 2.42 on 1 df,  p=0.1,  Robust = 1.91 p=0.2
##
## (Note: the likelihood ratio and score tests assume independence of
##       observations within a cluster, the Wald and robust score tests do
##       not).

```

### 3- Comparació una variable continua gaussiana entre els dos grups

Podem utilitzar la mitjana i la desviació, ja que sabem que les dades estan distribuïdes de forma normal. Per fer la comparació entre grups sense ajust emprariem tests paramètrics; com t-student, test ANOVA etc. I amb ajust utilitzaríem el test weighted t-test ja que tindríem en compte que les dades estan sota IPTW.

### 4 - Comparació una variable continua no-gaussiana entre els dos grups

En aquest cas, els estadístics descriptius són la mediana i el rang interquartilic (o el primer i tercer quartil), ja que les dades no són gaussianes. Els test que podríem usar seria el de Wilcoxon per a mostres relacionades i veure diferències entre variables.

### 5 - Comparació variable binària entre els dos grups

Per comprar variables binàries entre dos grups sense ajust podríem utilitzar el test de Fisher i amb ajust podríem utilitzar la prova de chi-quadrat. Els descriptius per a aquest tipus de variable són les freqüències i el percentatge d'aquestes.

## Part 3 - Anàlisis amb propensity score matching (PSM)

```
library(Matching)
```

```
## Loading required package: MASS
```

```
##
```

```
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##      select
```

```
## ##
```

```
## ## Matching (Version 4.9-9, Build Date: 2021-03-15)
```

```
## ## See http://sekhon.berkeley.edu/matching for additional  
documentation.
```

```
## ## Please cite software as:
```

```
## ## Jasjeet S. Sekhon. 2011. ``Multivariate and Propensity Score  
Matching
```

```
## ## Software with Automated Balance Optimization: The Matching  
package for R.''
```

```
## ## Journal of Statistical Software, 42(7): 1-52.
```

```
## ##
```

```
listMatch <- Match(Tr = datos2$treat==1, X=log(datos2$ps2/(1-  
datos2$ps2)),
```

```
  M = 1,
```

```
  caliper = 0.2,
```

```
  replace = FALSE,
```

```
  ties = TRUE,
```

```
  version = "fast")
```

```
datosMatched <-
```

```
datos2[unlist(listMatch[c("index.treated", "index.control")]), ]
```

```
tabMatched <- CreateTableOne(vars = vars, strata = "treat", data =  
datosMatched, test = FALSE)
```

```
print(tabMatched, smd = TRUE)
```

```
##                               Stratified by treat  
##                               0           1           SMD  
##    n                               50           50  
##    treat = 1 (%)                     0 ( 0.0)       50 (100.0)      NaN  
##    EDAD_DON (mean (SD))             55.20 (12.15)  53.38 (16.07)    0.128  
##    Gender_Don = 1 (%)                34 (68.0)     35 ( 70.0)    0.043  
##    Cause_death_CVA = 1 (%)           24 (48.0)     22 ( 44.0)    0.080  
##    Cause_death_Anoxia = 1 (%)        20 (40.0)     20 ( 40.0)   <0.001  
##    Cause_death_TBI = 1 (%)            5 (10.0)       4 (  8.0)    0.070  
##    Cause_death_Other = 1 (%)          1 ( 2.0)       4 (  8.0)    0.278  
##    TmUCILtsv (mean (SD))             8.66 (7.49)     8.00 (6.80)    0.092  
##    TmIsqCalTotAbd (mean (SD))        21.48 (6.47)    21.62 (9.12)    0.018
```

```
## TmIsqCalFxAAbd (mean (SD))          14.04 (4.38)    15.00 (5.36)    0.196
## TIEMPOISQUEMIAFR_A_RETH (mean (SD)) 341.38 (87.35) 340.54 (117.69) 0.008
## LIQUIDOPRESERV_UW_IGL1 = 1 (%)       11 (22.0)     12 ( 24.0)    0.048
## LIQUIDOPRESERV_HTK = 1 (%)           2 ( 4.0)      1 ( 2.0)     0.117
## LIQUIDOPRESERV_CELSIOR = 1 (%)       37 (74.0)     37 ( 74.0)   <0.001
## EDAD_REC_TX (mean (SD))              56.48 (7.70)   55.92 (11.51) 0.057
## Gender_Rec = 1 (%)                   42 (84.0)     41 ( 82.0)   0.053
## MELDSCORE (mean (SD))                15.32 (7.42)   15.12 (5.98)  0.030
## Volume = 1 (%)                       37 (74.0)     37 ( 74.0)   <0.001
## FIN_INDIC_reTx_fulmi = 1 (%)          1 ( 2.0)      2 ( 4.0)     0.117
## FIN_INDIC_HCC = 1 (%)                 18 (36.0)     20 ( 40.0)   0.082
## FIN_INDIC_Cirr = 1 (%)                30 (60.0)     26 ( 52.0)   0.162
## FIN_INDIC_Other = 1 (%)               1 ( 2.0)      2 ( 4.0)     0.117
```

*##No funciona como toca, hay qe cambiar algo*

## 1- Descripció de la supervivència de l'empelt i dels pacients

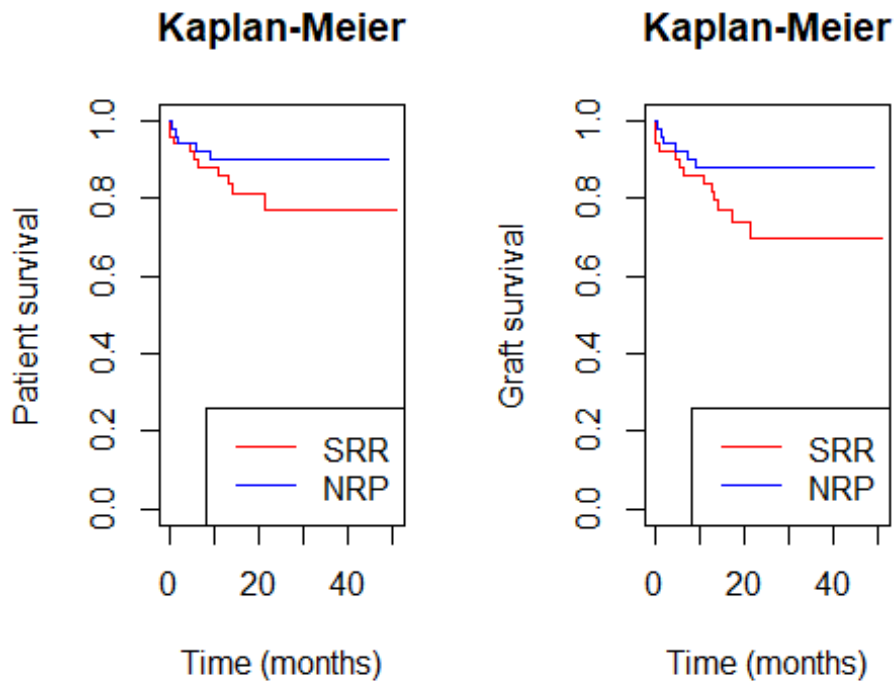
```
##
##Surten les corbes igual que l'apartat anterior. Hi ha canviar algo!!
##

km.model <- survfit(Surv(DeathTm,DeathYN)~treat, type="kaplan-meier",
data=datosMatched)

km.model2 <- survfit(Surv(GraftNotFxTm, GraftNotFxYN)~treat,
type="kaplan-meier", data=datosMatched)

{par(mfrow=c(1,2))
plot(km.model,conf.int=F,xlab="Time (months)",ylab="Patient
survival",main="Kaplan-Meier", col=c("red", "blue"))
legend("bottomright", c("SRR", "NRP"), col=c("red", "blue"), lty=c(1,1))

plot(km.model2,conf.int=F,xlab="Time (months)",ylab="Graft
survival",main="Kaplan-Meier", col=c("red", "blue"))
legend("bottomright", c("SRR", "NRP"), col=c("red", "blue"), lty=c(1,1))
}
```



### Diferència amb l'anàlisi IPTW

Els dos mètodes donen resultats molt semblants, tant com en el SMD com en els gràfics de la funció de supervivència. Tot i això, com les variables del model les introduïm per a que les SMD del IPTW siguin menors de 0.2 i utilitzem aquest mateix model per al PSM, les diferències si que són més petites amb el primer mètode.

### 2- Anàlisi inferencial de comparació de riscos en anàlisi de supervivència

No faria cap ajust addicional, ja que les dades les hem ajustat a partir del matching i ja hem minimitzat la diferència entre els grups.

### 3- Com comparàrieu una variable continua gaussiana entre els dos grups. Quins estadístics descriptius usàrieu per grup. Si coneixeu més d'una opció inferencial especifiqueu. Té sentit fer cap ajust addicional després del matching?

Podem utilitzar la mitjana i la desviació, ja que sabem que les dades estan distribuïdes de forma normal. Per fer la comparació entre grups sense ajust empràrieu tests paramètrics aparellats, com paired t-test. Ja que les dades finals contemplem l'aparellament, no té sentit fer cap ajust addicional.

**4- Com comparàrieu una variable continua no-gaussiana entre els dos grups. Quins estadístics descriptius usaríeu per grup. Si coneixeu més d'una opció inferencial especifiqueu. Té sentit fer cap ajust addicional després del matching?**

En aquest cas, els estadístics descriptius són la mediana i el rang interquartilic (o el primer i tercer quartil), ja que les dades no són gaussianes. Els test que podríem usar seria el de Wilcoxon per a mostres relacionades i veure diferències entre variables. Tal com passa amb les variables no-gaussianes, no té sentit cap ajust addicional.

**5- Com comparàrieu una variable binària entre els dos grups. Quins estadístics descriptius usaríeu per grup. Si coneixeu més d'una opció inferencial especifiqueu. Té sentit fer cap ajust addicional després del matching?**

Per comprar variables binàries entre dos grups sense ajust podríem utilitzar el test de Fisher o inclús el test de chi-quadrat. Els descriptius per a aquest tipus de variable són les freqüències i el percentatge d'aquestes.

**6- Compareu la interpretació dels anàlisis de la pràctica 1 (regressió logística condicional) amb l'anàlisi IPTW i PSM de la pràctica 2. Caldria haver fet per la pràctica 2 alguna mena d'anàlisi de dades dependents o estratificat addicional? Perquè?**

La pràctica 1 estàvem fent l'aparellament de les dades de forma manual a partir d'un indicador MatchCC i calculàvem la seva recta de regressió lineal condicionant a aquesta variable. En aquesta pràctica, l'objectiu era determinar uns pesos i a partir d'aquests minimitzar la diferència entre els grups. Al utilitzar els pesos no es necessari un anàlisi de dades dependents o estratificar, ja que les diferències es minimitzen amb aquests pesos.