

PROJECT - DPSS

Alice Brunazzi

2025-05-02

ANALISI

```
library(skimr)

## Warning: package 'skimr' was built under R version 4.3.3

library(readxl)
data <- read_excel("C:/Users/Utente/Downloads/pone.0200893.s001.xls",
  col_types = c("skip", "numeric", "text",
    "text", "numeric", "numeric", "numeric",
    "numeric", "numeric", "numeric",
    "numeric", "numeric", "numeric",
    "numeric", "numeric", "text", "numeric",
    "numeric", "numeric", "numeric",
    "numeric", "numeric", "numeric",
    "numeric", "numeric", "numeric"))
#View(pone_0200893_s001)
skimr::skim_without_charts(data)
```

Data summary

Name	data
Number of rows	999
Number of columns	31

Column type frequency:

character	3
numeric	28

Group variables	None
-----------------	------

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Gender	0	1	1	1	0	2	0
ASA	0	1	1	1	0	5	0
AJCC	0	1	2	2	0	2	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Age	0	1.00	65.18	13.71	18.00	55.00	65.00	77.00	98.00
ASA3	0	1.00	0.39	0.49	0.00	0.00	0.00	1.00	1.00
DM	0	1.00	0.20	0.40	0.00	0.00	0.00	0.00	1.00
CAD	0	1.00	0.07	0.26	0.00	0.00	0.00	0.00	1.00
HF	0	1.00	0.04	0.21	0.00	0.00	0.00	0.00	1.00
CVA	0	1.00	0.06	0.23	0.00	0.00	0.00	0.00	1.00
CKD	0	1.00	0.14	0.35	0.00	0.00	0.00	0.00	1.00
CEA	19	0.98	258.98	991.10	0.44	4.00	18.90	91.32	15126.00
LogCEA	19	0.98	1.39	0.90	-0.36	0.60	1.28	1.96	4.18
Laparoscopic	0	1.00	0.04	0.19	0.00	0.00	0.00	0.00	1.00
TumorLOC	0	1.00	0.31	0.46	0.00	0.00	0.00	1.00	1.00
EA	0	1.00	0.17	0.37	0.00	0.00	0.00	0.00	1.00
AnesTime	0	1.00	338.40	121.01	45.00	255.00	315.00	390.00	960.00
Log2AT	0	1.00	8.31	0.52	5.49	7.99	8.30	8.61	9.91
RBC	0	1.00	0.54	0.70	0.00	0.00	0.00	1.00	2.00
Liver_Only	0	1.00	0.37	0.48	0.00	0.00	0.00	1.00	1.00
Cell_diff	55	0.94	1.12	0.35	0.00	1.00	1.00	1.00	2.00
Mucin_TYPE	57	0.94	0.08	0.27	0.00	0.00	0.00	0.00	1.00
SignetRING	57	0.94	0.04	0.21	0.00	0.00	0.00	0.00	1.00
Lymphovascularinvasion	54	0.95	0.51	0.50	0.00	0.00	1.00	1.00	1.00
perineural	57	0.94	0.22	0.42	0.00	0.00	0.00	0.00	1.00
CT	0	1.00	0.89	0.31	0.00	1.00	1.00	1.00	1.00
RT	0	1.00	0.11	0.31	0.00	0.00	0.00	0.00	1.00
NACTRT	0	1.00	0.16	0.36	0.00	0.00	0.00	0.00	1.00
Death	0	1.00	0.43	0.49	0.00	0.00	0.00	1.00	1.00
Interval	0	1.00	24.23	24.40	0.03	7.34	17.48	31.74	135.79
Progress	0	1.00	0.78	0.42	0.00	1.00	1.00	1.00	1.00
IntervalR	0	1.00	11.27	17.24	0.03	2.96	5.75	12.60	134.21

colnames(data)

## [1]	"Age"	"Gender"	"ASA"
## [4]	"ASA3"	"DM"	"CAD"
## [7]	"HF"	"CVA"	"CKD"
## [10]	"CEA"	"LogCEA"	"Laparoscopic"
## [13]	"TumorLOC"	"EA"	"AnesTime"
## [16]	"Log2AT"	"RBC"	"AJCC"
## [19]	"Liver_Only"	"Cell_diff"	"Mucin_TYPE"
## [22]	"SignetRING"	"Lymphovascularinvasion"	"perineural"
## [25]	"CT"	"RT"	"NACTRT"
## [28]	"Death"	"Interval"	"Progress"
## [31]	"IntervalR"		

VARIABILI DEMOGRAFICHE E CLINICHE: età, genere, ASA, comorbidità tipo: diabete, malattie cardio vascolari, insufficienza renale VARIABILI TUMORALI: CEA, localizzazione, grading, tipo istologico (mucinoso, a cellule ad anello con castone) , invasione linfovaskolare e perineurale. VARIABILI TRATTAMENTO: tipo di intervento (laparoscopico, anestesia epidurale), tempo di anestesia, trasfusioni RBC, chemioterapia CT, radioterapia RT e terapia neoadiuvante NACTRT TARGE: morte, progressione, tempi di followup (interval e intervalR)

Variabile	Descrizione
Age	Età del paziente (in anni)
Gender	Sesso del paziente (Male, Female)
ASA	Classificazione ASA fisica pre-operatoria (da 1 a 5)
ASA3	Indicatore binario: ASA ≥ 3 (1 = grave compromissione, 0 = no)
DM	Diabete mellito (1 = presente, 0 = assente)
CAD	Cardiopatia ischemica (Coronary Artery Disease)
HF	Insufficienza cardiaca (Heart Failure)
CVA	Ictus (Cerebrovascular Accident)
CKD	Malattia renale cronica (Chronic Kidney Disease)
CEA	Valore dell'antigene carcinoembrionario (marker tumorale)
LogCEA	Logaritmo naturale del CEA (per normalizzazione)
Laparoscopic	Intervento laparoscopico (1 = sì, 0 = no)
TumorLOC	Localizzazione del tumore (0 = colon, 1 = retto)
EA	Uso di anestesia epidurale (1 = sì, 0 = no)
AnesTime	Durata dell'anestesia (in minuti)
Log2AT	Logaritmo in base 2 della durata dell'anestesia
RBC	Numero di unità di sangue trasfuse durante l'intervento
AJCC	Stadio AJCC del tumore (tipicamente IV)
Liver_Only	Metastasi limitate al fegato (1 = sì, 0 = no)
Cell_diff	Grado di differenziazione cellulare (1 = ben differenziato, 2 = moderato, ecc.)

Variabile	Descrizione
Mucin_TYPE	Tipo mucinoso del tumore (1 = presente, 0 = no)
SignetRING	Istotipo a cellule ad anello con castone (1 = sì, 0 = no)
Lymphovascularinvasion	Invasione linfovascolare (1 = presente)
perineural	Invasione perineurale (1 = presente)
CT	Chemioterapia post-operatoria (1 = sì, 0 = no)
RT	Radioterapia post-operatoria (1 = sì, 0 = no)
NACTRT	Terapia neoadiuvante (radio o chemio prima dell'intervento)
Death	Evento morte durante il follow-up (1 = morto, 0 = vivo)
Interval	Tempo in mesi fino alla morte o censura
Progress	Progressione del tumore (1 = sì, 0 = no)
IntervalR	Tempo in mesi fino alla progressione o censura

PREPROCESSING:

0. INTERVALS

```
cor(data$Interval, data$IntervalR, use="complete.obs")

## [1] 0.6614045

data$IntervalOD<- data$Interval
data$Interval <- NULL

data$IntervalOR <- data$IntervalR
data$IntervalR <- NULL

colnames(data)

## [1] "Age" "Gender" "ASA"
## [4] "ASA3" "DM" "CAD"
## [7] "HF" "CVA" "CKD"
## [10] "CEA" "LogCEA" "Laparoscopic"
## [13] "TumorLOC" "EA" "AnesTime"
## [16] "Log2AT" "RBC" "AJCC"
## [19] "Liver_Only" "Cell_diff" "Mucin_TYPE"
## [22] "SignetRING" "Lymphovascularinvasion" "perineural"
## [25] "CT" "RT" "NACTRT"
## [28] "Death" "Progress" "IntervalOD"
## [31] "IntervalOR"
```

1. ASA

indica, rispettivamente, i pazienti sani= 0, lievemente malati =1, gravi ma non a rischio di morte = 3, con rischio costante di morte = 4 e morente =5. ASA3, invece, binarizza la variabile, mostrando come 0 i pazienti a rischio accettabile e 1 i pazienti a rischio grave.

SI TEME MULTICOLLINEARITÀ IN SECONDO MOMENTO.

Si plottano gli istogrammi al fine di vedere quale variabile possa essere più informativa.

Prima della verifica, serve controllare che la variabile ASA3 non contenga errori.

```
# CONVERSIONE ASA

str(data$ASA)

## chr [1:999] "3" "2" "2" "2" "2" "2" "2" "2" "3" "2" "2" "3" "2" "2" "3" "2" ...

data$ASA_num <- as.numeric(as.character(data$ASA))

# Step 2: Creazione manuale della variabile ASA3 (gold standard)
data$ASA3_check <- ifelse(data$ASA_num >= 3, 1, 0)

# Step 3: Confronto tra ASA3 originale e ASA3 calcolato
table(Originale = data$ASA3, Calcolato = data$ASA3_check)

##           Calcolato
## Originale    0    1
##           0 612    0
##           1    0 387
```

Il valore 387 corrisponde con quello della documentazione fornita. Si testa per multicollinearità controllando la correlazione tra variabili. Letteratura suggerisce non più di 0.7

```
cor(data$ASA_num, data$ASA3)

## [1] 0.9018298

model <- lm(data$Death ~ ASA_num + ASA3, data = data)
car::vif(model)

## ASA_num      ASA3
## 5.356097 5.356097
```

La correlazione mostra conferma per multicollinearità, molto alta. DECISIONE: non tenerle entrambe

Il variance index inflator non aiuta nella decisione su quale rimuovere. Si confronta AIC dei modelli fittati sulle variabili per verificare quale tenere.

```
# Modelli a confronto
model1 <- glm(Death ~ ASA_num, data = data, family = "binomial")
model2 <- glm(Death ~ ASA3, data = data, family = "binomial")

AIC(model1, model2) # Confronta la bontà del fit

##      df      AIC
## model1  2 1366.750
## model2  2 1363.609
```

AIC più basso per ASA3 -> miglior fit, anche se la differenza è modesta ($\Delta AIC > 2$ è comunque considerata significativa, in questo caso è -3.14).

```
# Rimozione della variabile ASA_num e ASA non modificata
data$ASA_num <- NULL
data$ASA3_check <- NULL
data$ASA <- NULL

cat("Numero di record nel dataset:", nrow(data), "\n")

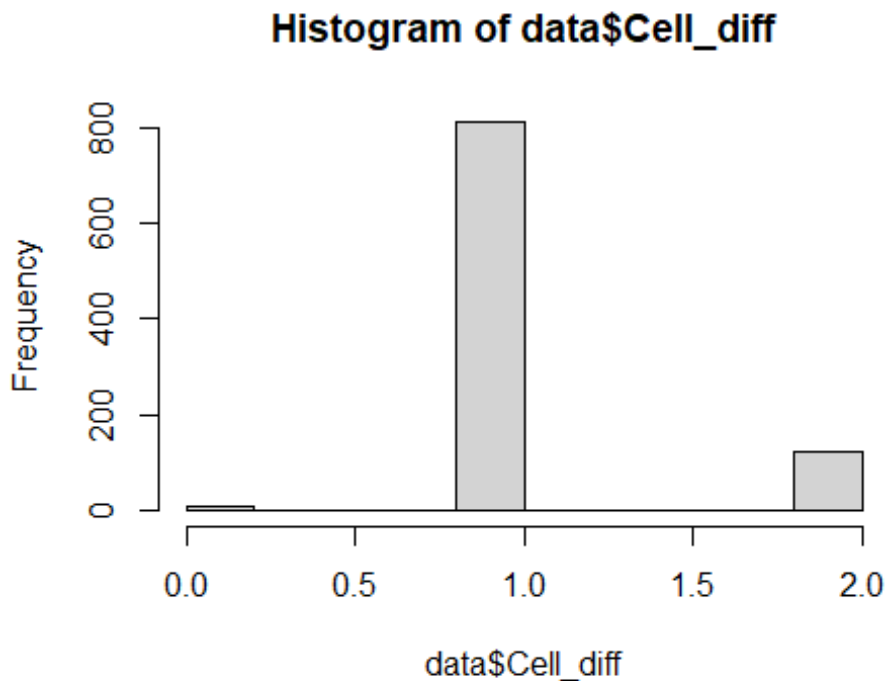
## Numero di record nel dataset: 999
```

##2- Cell_Diff

VARIABILE GRADO DI DIFFERENZIAZIONE CELLULARE

Le convenzioni standard di classificazione istologica danno alla variabile valori 1= ben differenziato, 2= non differenziato

```
hist(data$Cell_diff)
```



La variabile presenta dei valori non previsti nella scala 1-2, si provvede a verificarli manualmente e poi graficamente

```
sum(data$Cell_diff == 0, na.rm = TRUE)

## [1] 10
```

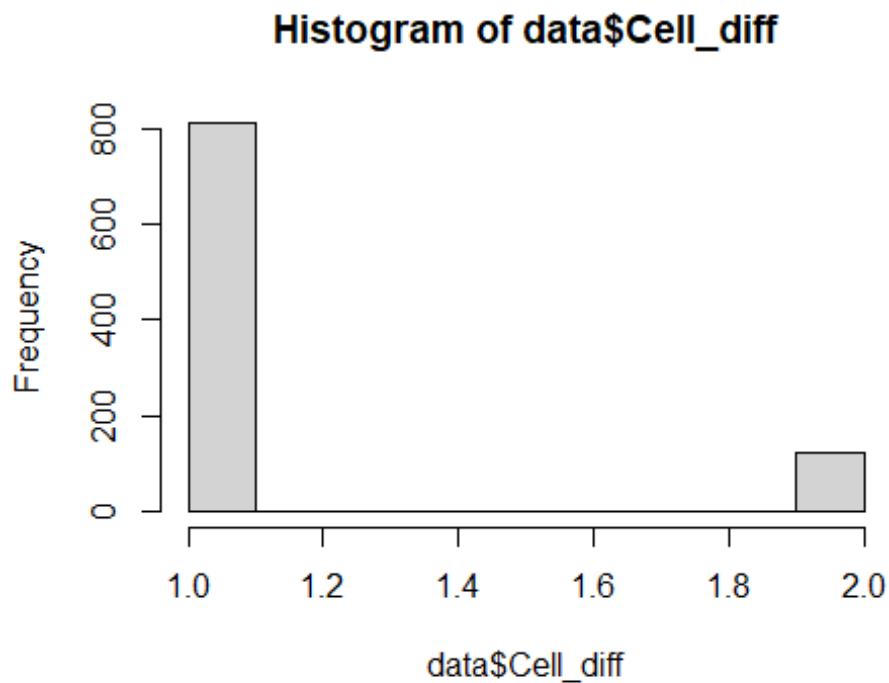
```
record_zero_diff <- data[!is.na(data$Cell_diff) & data$Cell_diff == 0, ]

record_zero_diff

## # A tibble: 10 × 30
##   Age Gender ASA3 DM CAD HF CVA CKD CEA LogCEA Laparoscopic
##   <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 64 1 0 0 0 0 0 0 11.6 1.06 0
## 2 51 2 0 0 0 0 0 0 24.2 1.38 0
## 3 83 1 1 1 0 0 1 1 581 2.76 0
## 4 61 2 0 1 0 0 0 0 2.29 0.360 0
## 5 56 2 0 0 0 0 0 0 1.47 0.167 0
## 6 58 1 0 0 0 0 0 0 94.8 1.98 0
## 7 57 1 1 0 0 0 0 0 32.2 1.51 0
## 8 41 1 0 0 0 0 0 0 2.69 0.430 0
## 9 57 1 0 0 0 0 0 1 1108 3.04 1
## 10 60 2 0 0 0 0 0 0 1.78 0.250 0
## # i 19 more variables: TumorLOC <dbl>, EA <dbl>, AnesTime <dbl>, Log2AT <dbl>,
## # RBC <dbl>, AJCC <chr>, Liver_Only <dbl>, Cell_diff <dbl>, Mucin_TYPE <dbl>,
## # SignetRING <dbl>, Lymphovascularinvasion <dbl>, perineural <dbl>, CT <dbl>,
## # RT <dbl>, NACTRT <dbl>, Death <dbl>, Progress <dbl>, IntervalOD <dbl>,
## # IntervalOR <dbl>
```

I record problematici snon solo 10, sono sicuramente mal codificati e/ o danneggiati. Si reputa che la la loro eliminazione non comprometta l'efficacia statistica. si poteva imputare i 10 record nulli alla variabile 1 ma dato il numero quali nullo di records, l'eliminazione non impatta sulla variabile. La media e la sd corrispondono a quella fornita dalla documentazione

```
data <- data[!(data$Cell_diff == 0 | is.na(data$Cell_diff)), ]
hist(data$Cell_diff)
```



```
sum(data$Cell_diff > 1.5)
## [1] 121
sum(data$Cell_diff < 1.5)
## [1] 813
cat("Numero di record nel dataset:", nrow(data), "\n")
## Numero di record nel dataset: 934
```

Che corrispondono con i 121 (non differenziati) record e gli 823 (differenziati) nella documentazione riportata

3 - Binary

Si verificano tutte le variabili binarie (codificate 0 No 1 si) per verificare se presentano valori fuori dal range.

```
binary_vars <- c("DM", "CAD", "HF", "CVA", "CKD", "Laparoscopic", "EA",
                 "Liver_Only", "Lymphovascularinvasion", "perineural",
                 "CT", "RT", "NACTRT", "Death", "Progress")

# VALORI UNICI, DEVONO ESSERE 0 E 1
for (var in binary_vars) {
  cat("\nVariabile:", var, "\n")
  print(unique(data[[var]]))
}
```



```
##
## Variabile: DM
## [1] 0 1
##
## Variabile: CAD
## [1] 0 1
##
## Variabile: HF
## [1] 0 1
##
## Variabile: CVA
## [1] 0 1
##
## Variabile: CKD
## [1] 0 1
##
## Variabile: Laparoscopic
## [1] 0 1
##
## Variabile: EA
## [1] 0 1
##
## Variabile: Liver_Only
## [1] 0 1
##
## Variabile: Lymphovascularinvasion
## [1] 1 0
##
## Variabile: perineural
## [1] 0 1 NA
##
## Variabile: CT
## [1] 1 0
##
## Variabile: RT
## [1] 0 1
##
## Variabile: NACTRT
## [1] 0 1
##
## Variabile: Death
## [1] 1 0
##
## Variabile: Progress
## [1] 1 0
```

Si nota che Perineural contiene NA, si procede alla verifica

```
# Mostra tutti i record dove perineural è NA
subset(data, is.na(perineural))
```

```
## # A tibble: 3 × 30
##   Age Gender  ASA3    DM    CAD    HF    CVA    CKD    CEA LogCEA Laparoscopic
##   <dbl> <chr>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>      <dbl>
## 1    53 1      0      0      0      0      0      0  2.61  0.417      0
## 2    71 2      0      0      1      0      0      0  23.4  1.37      0
## 3    52 1      0      0      0      0      0      0  21.1  1.32      0
## # i 19 more variables: TumorLOC <dbl>, EA <dbl>, AnesTime <dbl>, Log2AT <dbl>,
## #   RBC <dbl>, AJCC <chr>, Liver_Only <dbl>, Cell_diff <dbl>, Mucin_TYPE <dbl>,
## #   SignetRING <dbl>, Lymphovascularinvasion <dbl>, perineural <dbl>, CT <dbl>,
## #   RT <dbl>, NACTRT <dbl>, Death <dbl>, Progress <dbl>, IntervalOD <dbl>,
## #   IntervalOR <dbl>
```

```
data <- data[!is.na(data$perineural), ]
unique(data$perineural)
```

```
## [1] 0 1
```

```
cat("Numero di record nel dataset:", nrow(data), "\n")
```

```
## Numero di record nel dataset: 931
```

4- Gender

```
data$Gender <- as.numeric(as.character(data$Gender))
```

```
cat("Numero di record nel dataset:", nrow(data), "\n")
```

```
## Numero di record nel dataset: 931
```

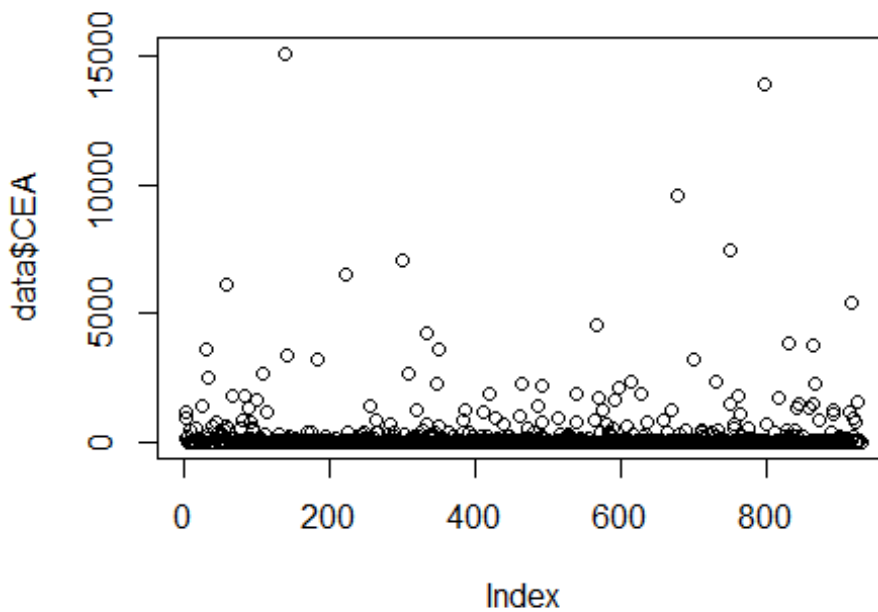
5- CEA

La variabile CEA contiene missing values 15, e di conseguenza LogCEA

```
summary(data$CEA)
```

```
##      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.      NA's
##    0.440      4.008      18.170     265.620      87.405    15126.000         15
```

```
plot(data$CEA)
```



Essendo una variabile di profondo senso clinico (valore del'carcogene ~ marker tumorale), si procede in ordine: 1- scelta della variabile da tenere tra CEA e Log CEA, 2- VALUTARE SE imputare O eliminare dei valori mancanti

Si inizia controllando che LogCEA corrisponda al Logaritmo di CEA, senza errori

```
# 1. Logaritmo naturale di (CEA + 1) - standard
mean(log(data$CEA + 1), na.rm = TRUE)

## [1] 3.328466

# 2. Logaritmo naturale
mean(log(data$CEA), na.rm = TRUE)

## [1] 3.200846

# 3. Logaritmo in base 10
mean(log10(data$CEA), na.rm = TRUE)

## [1] 1.39011

# 4. Logaritmo in base 10 di (CEA + 1)
mean(log10(data$CEA + 1), na.rm = TRUE)

## [1] 1.445534
```

Si trova dunque che la variabile é il logaritmo in base 10, si procede con l'analisi

Si testa per multicollinearità controllando la correlazione tra variabili. Letteratura suggerisce non più di 0.7

```

# Ricreiamo LogCEA come Log10(CEA)
data$LogCEA_check <- log10(data$CEA)

# Differenza assoluta tra le due
diff <- abs(data$LogCEA - data$LogCEA_check)

# Controlliamo quante differenze sono significative (> 0.001)
sum(diff > 0, na.rm = TRUE)

## [1] 0

```

Si verifica ora la multicollinearità tra le variabili

```

cor(data$CEA, data$LogCEA, use= "complete.obs")

## [1] 0.5459278

# Regressione + VIF

model <- lm(Death ~ CEA + LogCEA, data = data)
library(car)

## Warning: package 'car' was built under R version 4.3.3
## Loading required package: carData
## Warning: package 'carData' was built under R version 4.3.3

vif(model)

##      CEA      LogCEA
## 1.424577 1.424577

# Modelli alternativi
model1 <- glm(Death ~ CEA, data = data, family = "binomial")
model2 <- glm(Death ~ LogCEA, data = data, family = "binomial")

# Confronto AIC
AIC(model1, model2)

##      df      AIC
## model1  2 1245.873
## model2  2 1230.633

```

La correlazione non è sufficientemente alta per giustificare l'ipotesi di multicollinearità problematica. Il VIF ci sostiene nella scelta di mantenere entrambe le variabili.

Sapendo che entrambe verranno tenute, si manterrà a mente che LogCEA offre più potere predittivo. Si valuterà in un secondo momento se scartare CEA.

Imputazione dei NA

Prima si verifica che i records corrispondano tra le due variabili problematiche

```
# Verifica se gli NA sono nelle stesse righe
```

```
na_cea <- is.na(data$CEA)
```

```
na_logcea <- is.na(data$LogCEA)
```

```
# Rimuove i record con LogCEA ≤ 0 o NA
```

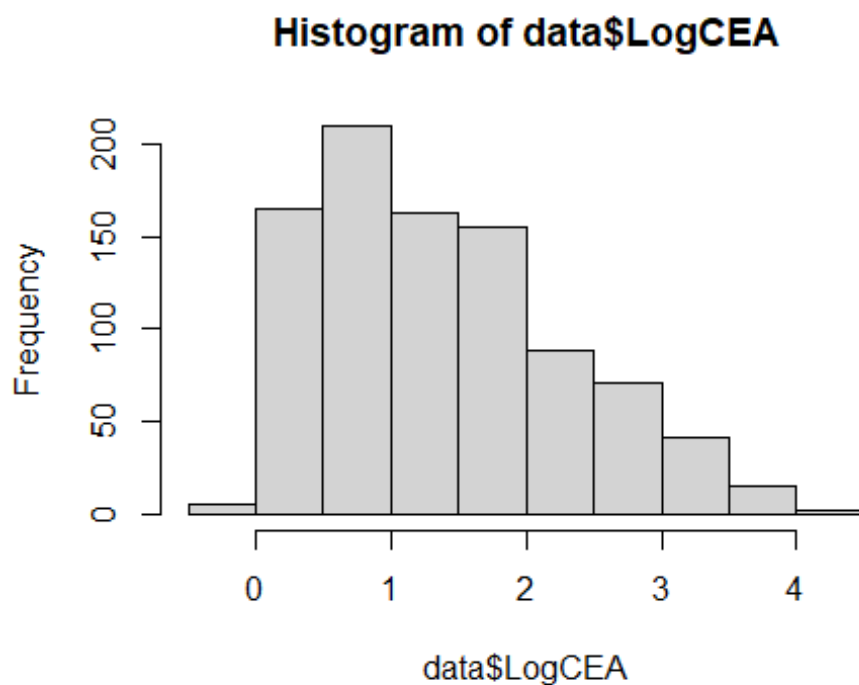
```
data <- data[!is.na(data$CEA), ]
```

```
#data$CEA_imputed <- NULL
```

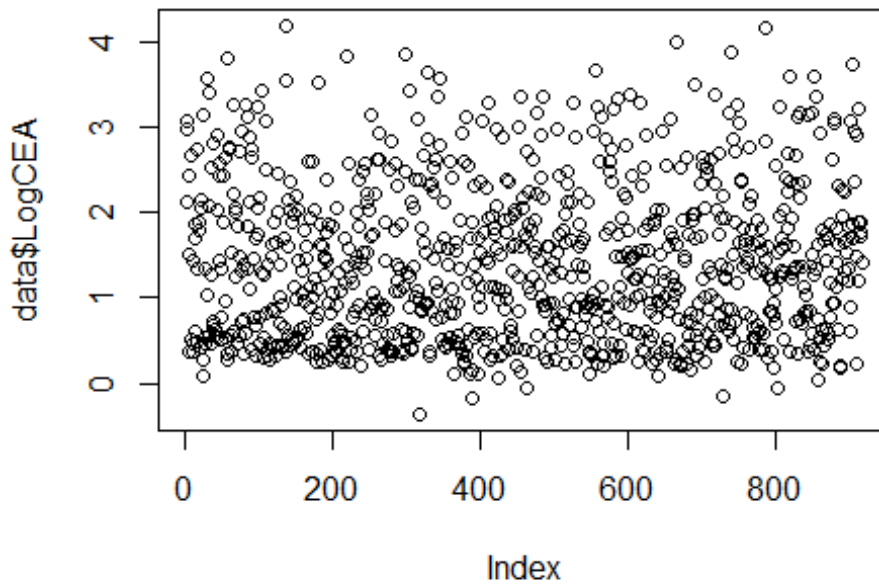
```
data$LogCEA_check <- NULL
```

Si verifica la distribuzione della variabile OG per scegliere che tipo di imputazione utilizzare

```
hist(data$LogCEA)
```



```
plot(data$LogCEA)
```



Variaile ESTREMAMENTE SKEWED, mediana scelta più prodente

```
data$LogCEA[is.na(data$LogCEA)] <- 1.26435
```

```
summary(data$CEA)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
##  0.440    4.008    18.170   265.620   87.405  15126.000
```

```
summary(data$LogCEA)
```

```
##      Min. 1st Qu.  Median     Mean 3rd Qu.     Max.
## -0.3565  0.6029  1.2594  1.3901  1.9415  4.1797
```

```
cat("Numero di record nel dataset:", nrow(data), "\n")
```

```
## Numero di record nel dataset: 916
```

```
data$LogCEA <- log10(data$CEA)
```

———— AGGIUNTA POST: conviene eliminare la variabile CEA, molto multicollineare in fare di analisi successiva:

```
data$CEA <- NULL
```

6 - AnesTime

Presenti outliers, il range INTERQUANTILE deve essere tra 255 e 390 minuti, presenti valori maggiori

```

summary(data$AnesTime)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      50.0   270.0   315.0   342.8   390.0   960.0

cat("Numero di record nel dataset:", nrow(data), "\n")

## Numero di record nel dataset: 916

cor(data$AnesTime, data$Log2AT, use= "complete.obs")

## [1] 0.966106

# Regressione + VIF

model <- lm(Death ~ AnesTime + Log2AT, data = data)
library(car)
vif(model)

## AnesTime    Log2AT
## 15.00617 15.00617

# Modelli alternativi
model1 <- glm(Death ~ AnesTime, data = data, family = "binomial")
model2 <- glm(Death ~ Log2AT, data = data, family = "binomial")

# Confronto AIC
AIC(model1, model2)

##           df          AIC
## model1    2 1251.090
## model2    2 1250.935

data$AnesTime <- NULL

```

7- Outliers

si controlla ora il dataset completo, per analizzare possibili outliers e valori mancanti

```
skimr::skim_without_charts(data)
```

Data summary

Name	data
Number of rows	916
Number of columns	28

Column type frequency:

character	1
numeric	27

Group variables None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
AJCC	0	1	2	2	0	2	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Age	0	1	65.23	13.62	18.00	55.00	65.00	77.00	95.00
Gender	0	1	1.38	0.49	1.00	1.00	1.00	2.00	2.00
ASA3	0	1	0.38	0.48	0.00	0.00	0.00	1.00	1.00
DM	0	1	0.21	0.41	0.00	0.00	0.00	0.00	1.00
CAD	0	1	0.07	0.26	0.00	0.00	0.00	0.00	1.00
HF	0	1	0.04	0.20	0.00	0.00	0.00	0.00	1.00
CVA	0	1	0.06	0.23	0.00	0.00	0.00	0.00	1.00
CKD	0	1	0.13	0.34	0.00	0.00	0.00	0.00	1.00
LogCEA	0	1	1.39	0.90	-0.36	0.60	1.26	1.94	4.18
Laparoscopic	0	1	0.04	0.20	0.00	0.00	0.00	0.00	1.00
TumorLOC	0	1	0.31	0.46	0.00	0.00	0.00	1.00	1.00
EA	0	1	0.16	0.37	0.00	0.00	0.00	0.00	1.00
Log2AT	0	1	8.34	0.48	5.64	8.08	8.30	8.61	9.91
RBC	0	1	0.53	0.68	0.00	0.00	0.00	1.00	2.00
Liver_Only	0	1	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Cell_diff	0	1	1.13	0.33	1.00	1.00	1.00	1.00	2.00
Mucin_TYPE	0	1	0.08	0.26	0.00	0.00	0.00	0.00	1.00
SignetRING	0	1	0.04	0.21	0.00	0.00	0.00	0.00	1.00
Lymphovascularinvasion	0	1	0.52	0.50	0.00	0.00	1.00	1.00	1.00
perineural	0	1	0.22	0.42	0.00	0.00	0.00	0.00	1.00
CT	0	1	0.90	0.30	0.00	1.00	1.00	1.00	1.00
RT	0	1	0.11	0.31	0.00	0.00	0.00	0.00	1.00
NACTRT	0	1	0.16	0.36	0.00	0.00	0.00	0.00	1.00
Death	0	1	0.42	0.49	0.00	0.00	0.00	1.00	1.00
Progress	0	1	0.79	0.41	0.00	1.00	1.00	1.00	1.00
IntervalOD	0	1	24.6	24.2	0.03	8.11	18.0	31.9	135.7

skim_variable	n_missi ng	complete_ra te	mea n	sd	p0	p25	p50	p75	p100
			1	5			5	5	9
IntervalOR	0	1	11.3	17.1	0.03	3.05	5.98	12.6	134.2
			2	0				2	1

La variabile AJCC non contiene nulli, si binarizza

```
data$AJCC_bin <- ifelse(data$AJCC == '4a', 0,1 )
```

```
data$AJCC <- NULL
```

8- analisi e controllo info perse

```
print(paste('Diabete Mellito (DM):', sum(data$DM == 1), 'su 203, TOT',  
length(data$DM)))
```

```
## [1] "Diabete Mellito (DM): 191 su 203, TOT 916"
```

```
print(paste('Cardiopatía ischemica (CAD):', sum(data$CAD == 1), 'su 72, TOT',  
length(data$CAD)))
```

```
## [1] "Cardiopatía ischemica (CAD): 68 su 72, TOT 916"
```

```
print(paste('Scompenso cardiaco (HF):', sum(data$HF == 1), 'su 44, TOT',  
length(data$HF)))
```

```
## [1] "Scompenso cardiaco (HF): 40 su 44, TOT 916"
```

```
print(paste('Ictus / Stroke (CVA):', sum(data$CVA == 1), 'su 57 TOT',  
length(data$CVA)))
```

```
## [1] "Ictus / Stroke (CVA): 53 su 57 TOT 916"
```

```
print(paste('Malattia renale cronica (CKD):', sum(data$CKD == 1), 'su 138, TOT',  
length(data$CKD)))
```

```
## [1] "Malattia renale cronica (CKD): 123 su 138, TOT 916"
```

```
print(paste('Sesso: Maschi (TOT 612)', sum(data$Gender == 1), '- Femmine (TOT  
387)', sum(data$Gender == 2), ', IN TOTALE ', length(data$Gender)))
```

```
## [1] "Sesso: Maschi (TOT 612) 568 - Femmine (TOT 387) 348 , IN TOTALE 916"
```

```
print(paste('Localizzazione Tumore - Colon:', sum(data$TumorLOC == 0), 'su 712 ,  
TOTALE', length(data$TumorLOC)))
```

```
## [1] "Localizzazione Tumore - Colon: 635 su 712 , TOTALE 916"
```

```
print(paste('Localizzazione Tumore - Retto:', sum(data$TumorLOC == 1), 'su 313,  
TOTALE ', length(data$TumorLOC)))
```

```
## [1] "Localizzazione Tumore - Retto: 281 su 313, TOTALE 916"
```

```
print(paste('Grado cellulare G1:', sum(data$Cell_diff == 1), 'su 823'))
```

```

## [1] "Grado cellulare G1: 799 su 823"
print(paste('Grado cellulare G2:', sum(data$Cell_diff == 2), 'su 121' ))
## [1] "Grado cellulare G2: 117 su 121"
print(paste('Tipo Mucinoso:', sum(data$Mucin_TYPE == 1), 'su 73' ))
## [1] "Tipo Mucinoso: 69 su 73"
print(paste('Tipo Non-Mucinoso:', sum(data$Mucin_TYPE == 0), 'su 999-73 = 926' ))
## [1] "Tipo Non-Mucinoso: 847 su 999-73 = 926"
print(paste('Cellule ad anello con castone:', sum(data$SignetRING == 1), 'su 42'
))
## [1] "Cellule ad anello con castone: 41 su 42"
print(paste('Chirurgia laparoscopica:', sum(data$Laparoscopic == 1)))
## [1] "Chirurgia laparoscopica: 37"
print(paste('Anestesia epidurale (EA):', sum(data$EA == 1), 'su 165' ))
## [1] "Anestesia epidurale (EA): 151 su 165"
print(paste('Metastasi solo liver (Liver_Only):', sum(data$Liver_Only == 1), 'su
370' ))
## [1] "Metastasi solo liver (Liver_Only): 357 su 370"
print(paste('Invasione linfovascolare:', sum(data$Lymphovascularinvasion == 1),
'su 485'))
## [1] "Invasione linfovascolare: 472 su 485"
print(paste('Invasione perineurale:', sum(data$perineural == 1), 'su 208'))
## [1] "Invasione perineurale: 206 su 208"
print(paste('Trasfusioni - Nessuna:', sum(data$RBC == 0), '/577',
'- Meno di 4 unita:', sum(data$RBC == 1), '/305',
'- Più di 4 unita:', sum(data$RBC == 2), '/117'))
## [1] "Trasfusioni - Nessuna: 530 /577 - Meno di 4 unita: 286 /305 - Più di 4
unita: 100 /117"
print(paste('Chemioterapia post:', sum(data$CT == 1), 'su 999' ))
## [1] "Chemioterapia post: 822 su 999"
print(paste('Radioterapia post:', sum(data$RT == 1), 'su 110' ))
## [1] "Radioterapia post: 98 su 110"
print(paste('Terapia pre operatoria (NACTRT):', sum(data$NACTRT == 1), 'su 155'))

```

```
## [1] "Terapia pre operatoria (NACTRT): 143 su 155"
print(paste('4a:', sum(data$AJCC_bin == 0), 'su 558'
            , '4:', sum(data$AJCC_bin == 1), 'su 441'))
## [1] "4a: 531 su 558 4: 385 su 441"
```

TARGET

```
print(paste('Decessi (Death):', sum(data$Death == 1)))
## [1] "Decessi (Death): 386"
print(paste("non decessi", sum(data$Death == 0)))
## [1] "non decessi 530"
print(paste('Progressioni tumorali (Progress):', sum(data$Progress == 1)))
## [1] "Progressioni tumorali (Progress): 725"
```

Si conclude che le INFERMAZIONI perse nel pre processing non impattano così tanto, ma si controllano le visualizzazioni delle distribuzioni delle variabili per controllare eventuali squilibri

DATA VIZUALIZATION

Distribuzioni

Si controllano inizialmente le distribuzioni generali di tutte le variabili.

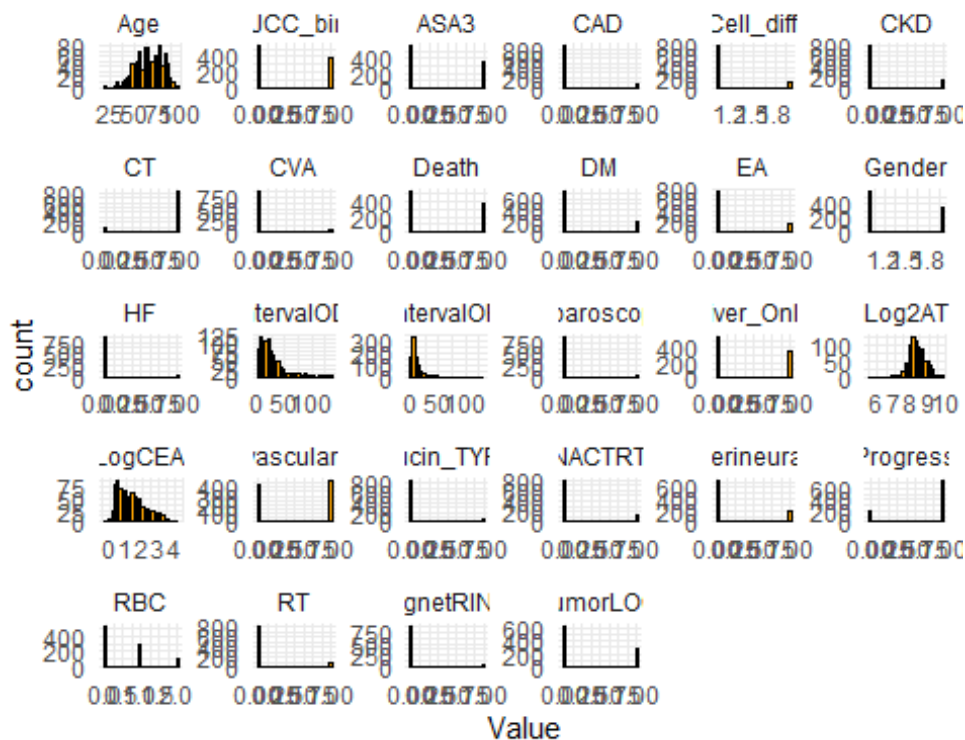
```
# Seleziona solo le colonne numeriche
numeric_vars <- data[, sapply(data, is.numeric)]

library(ggplot2)
## Warning: package 'ggplot2' was built under R version 4.3.3

library(tidyr)
## Warning: package 'tidyr' was built under R version 4.3.3

# Converti in formato Lungo
numeric_long <- pivot_longer(as.data.frame(numeric_vars), cols = everything(),
                             names_to = "Variable", values_to = "Value")

ggplot(numeric_long, aes(x = Value)) +
  geom_histogram(fill = "orange", color = "black", bins = 30) +
  facet_wrap(~ Variable, scales = "free") +
  theme_minimal()
```



Non si notano comportamenti preoccupanti

Outliers

Per verificare gli outliers:

```
library(ggplot2)
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.3.3

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:car':
##
##   recode

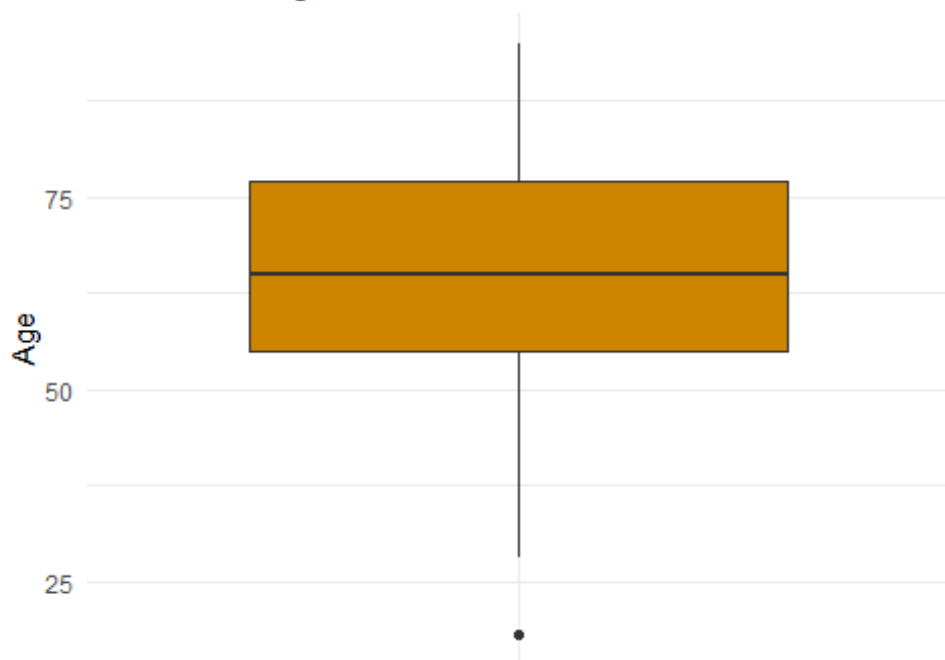
## The following objects are masked from 'package:stats':
##
##   filter, lag

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

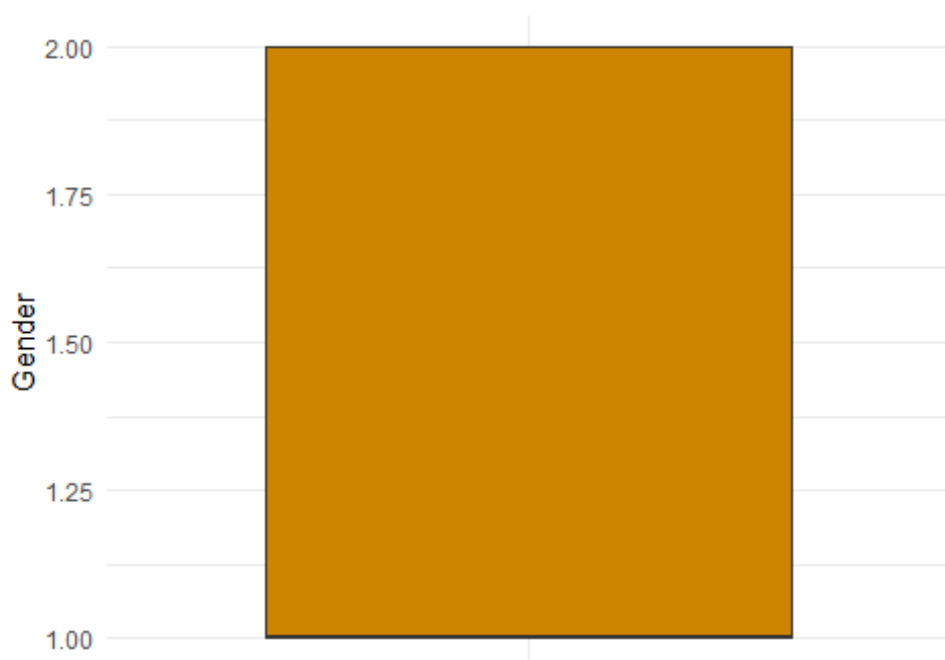
for (var in colnames(data)) {
  p <- ggplot(data, aes_string(x = "", y = var)) +
    geom_boxplot(fill = "orange3") +
    labs(title = paste("Box Plot di", var), y = var, x = "") +
```

```
    theme_minimal() +  
    theme(axis.text.x = element_blank(),  
          axis.ticks.x = element_blank())  
  
  print(p)  
}  
  
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.  
## i Please use tidy evaluation idioms with `aes()`.  
## i See also `vignette("ggplot2-in-packages")` for more information.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was  
## generated.
```

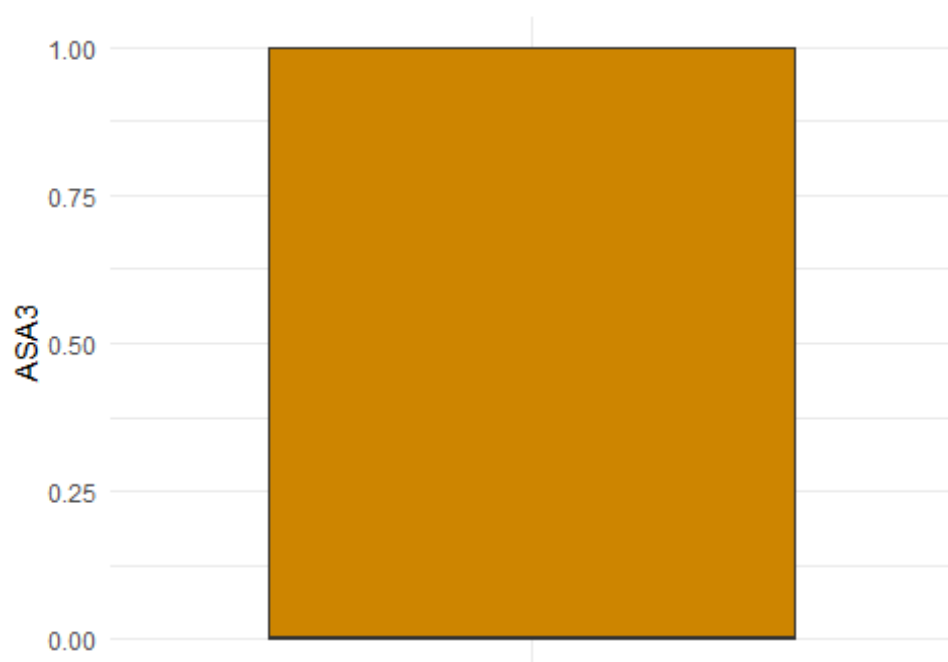
Box Plot di Age



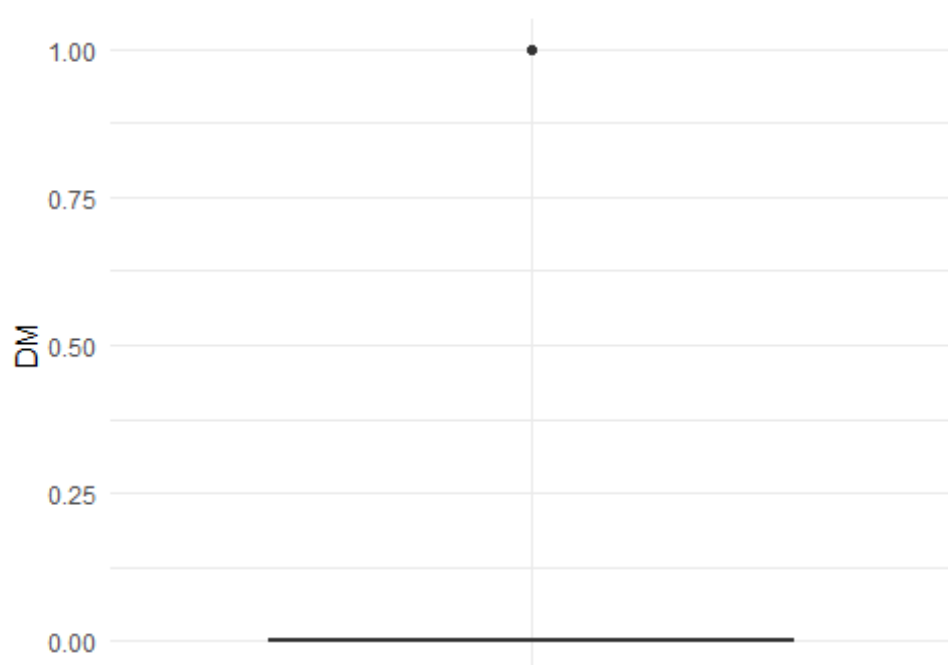
Box Plot di Gender



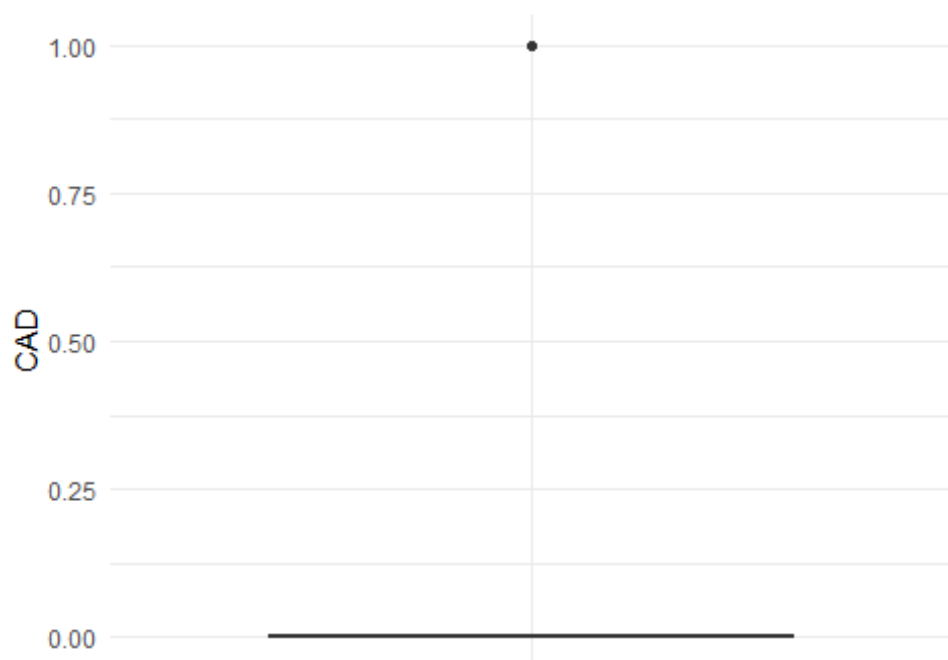
Box Plot di ASA3



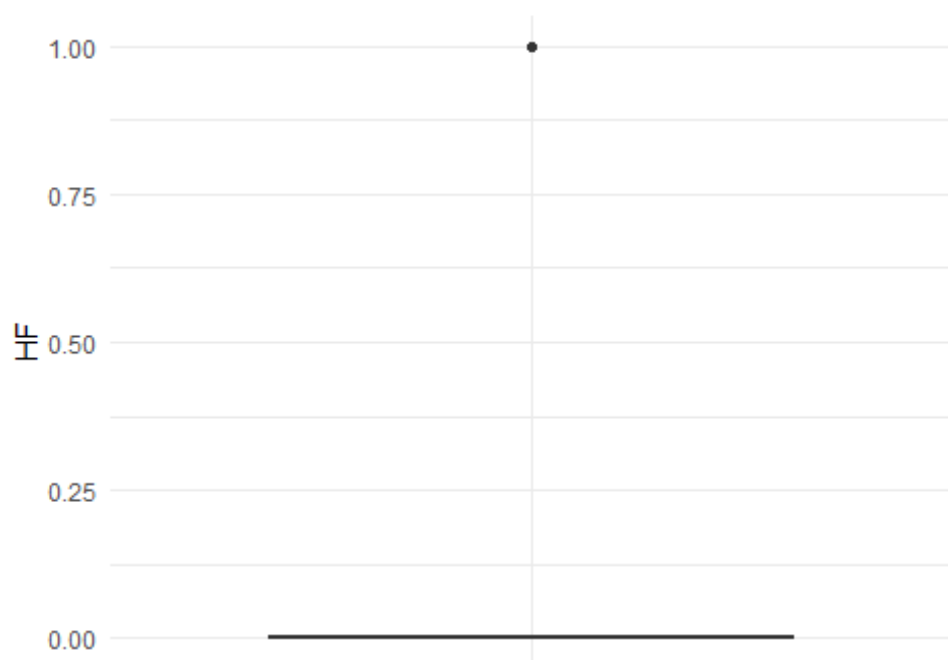
Box Plot di DM



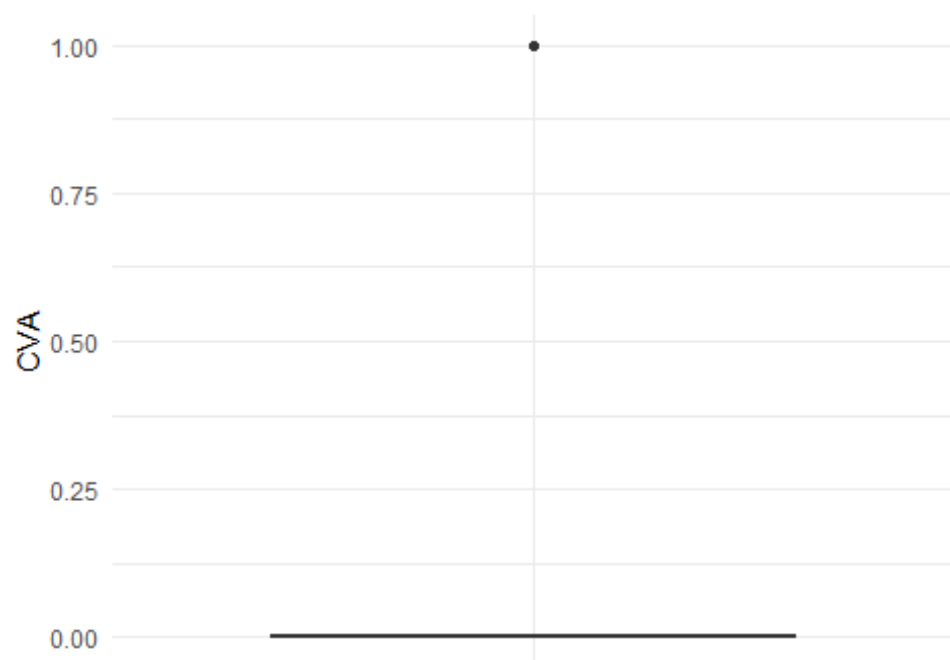
Box Plot di CAD



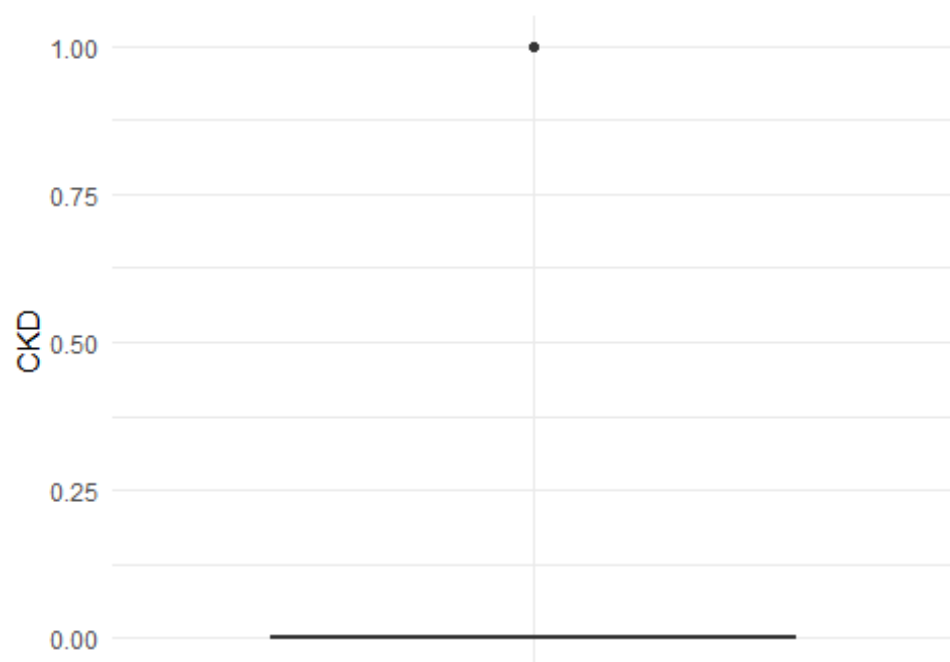
Box Plot di HF



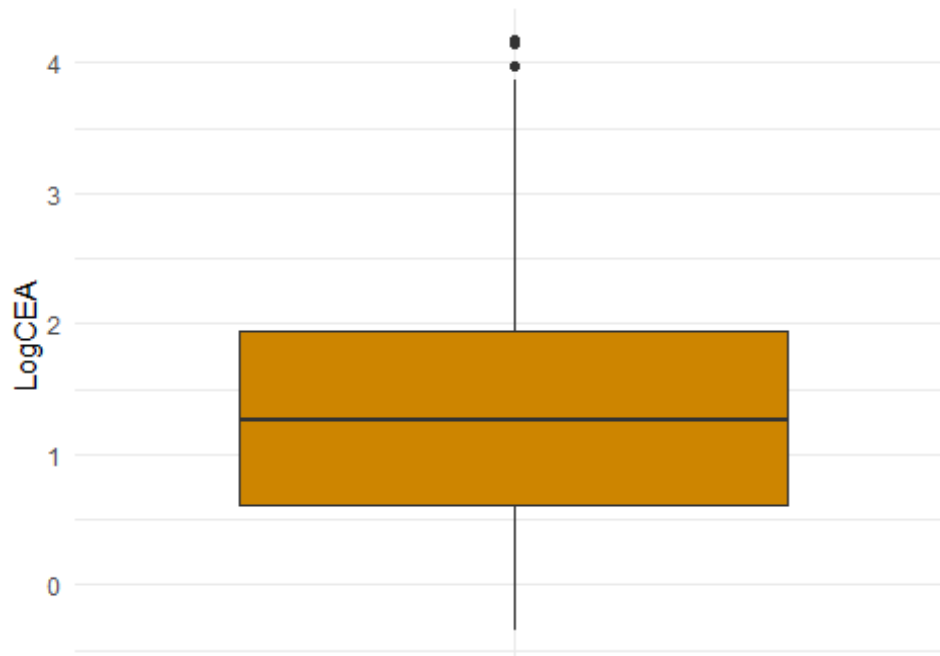
Box Plot di CVA



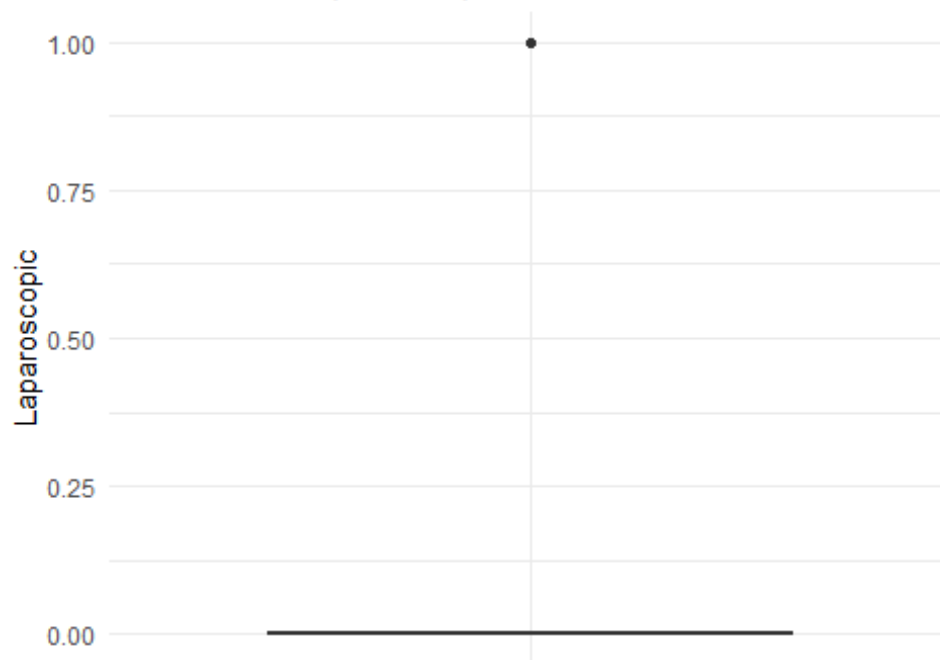
Box Plot di CKD



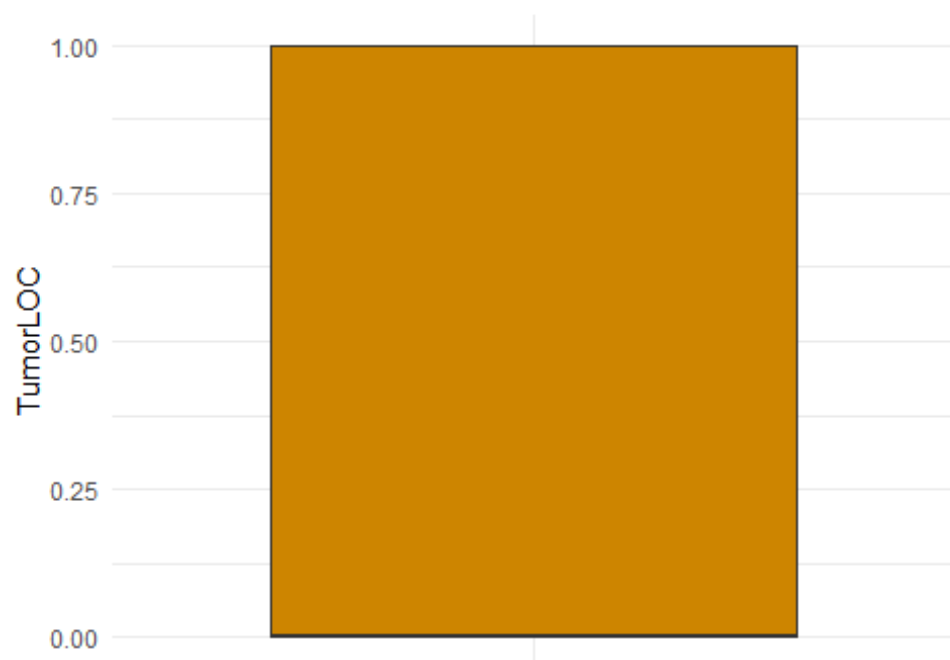
Box Plot di LogCEA



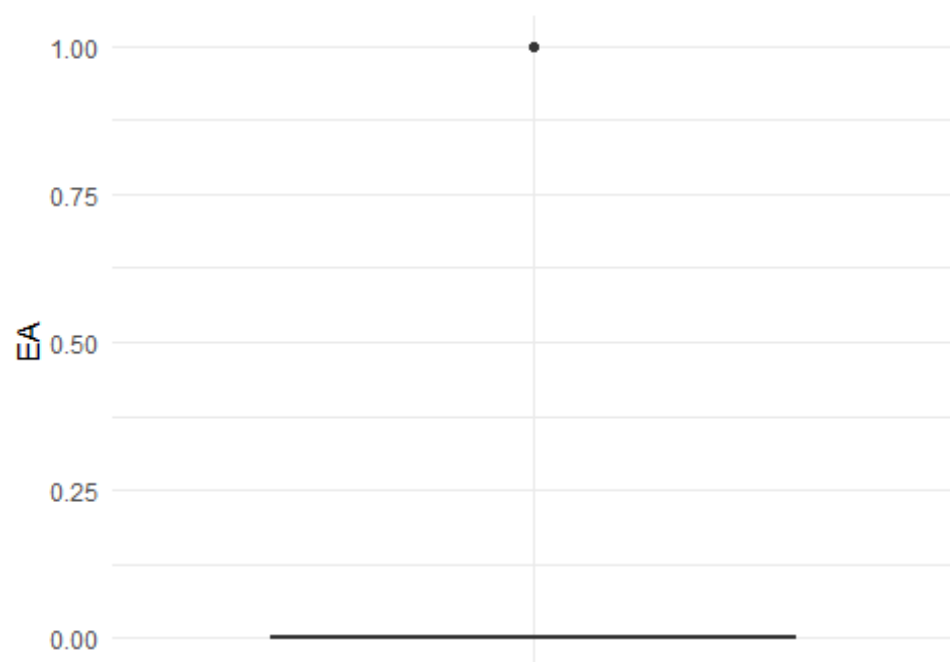
Box Plot di Laparoscopic



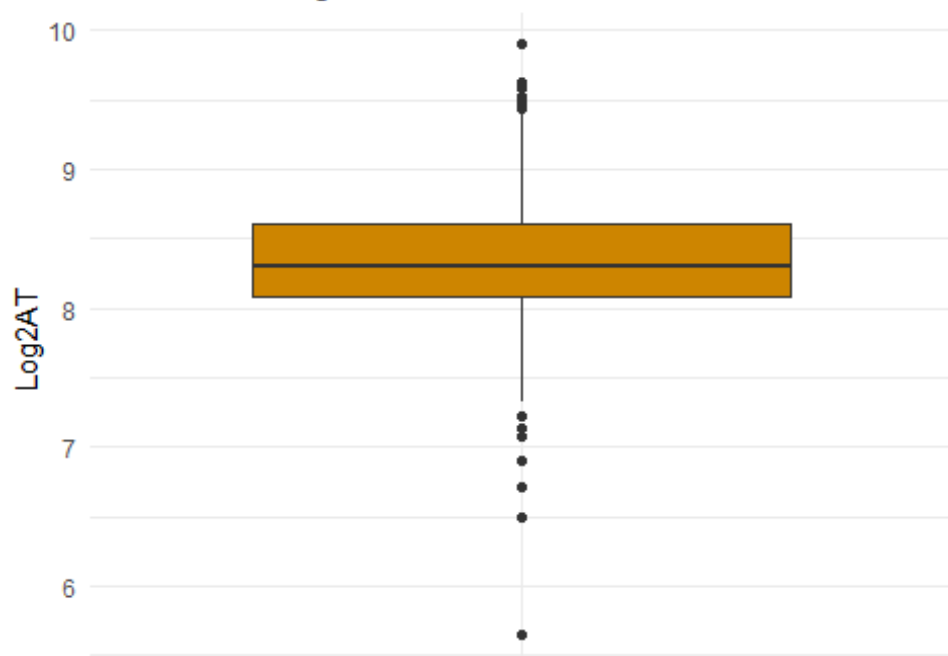
Box Plot di TumorLOC



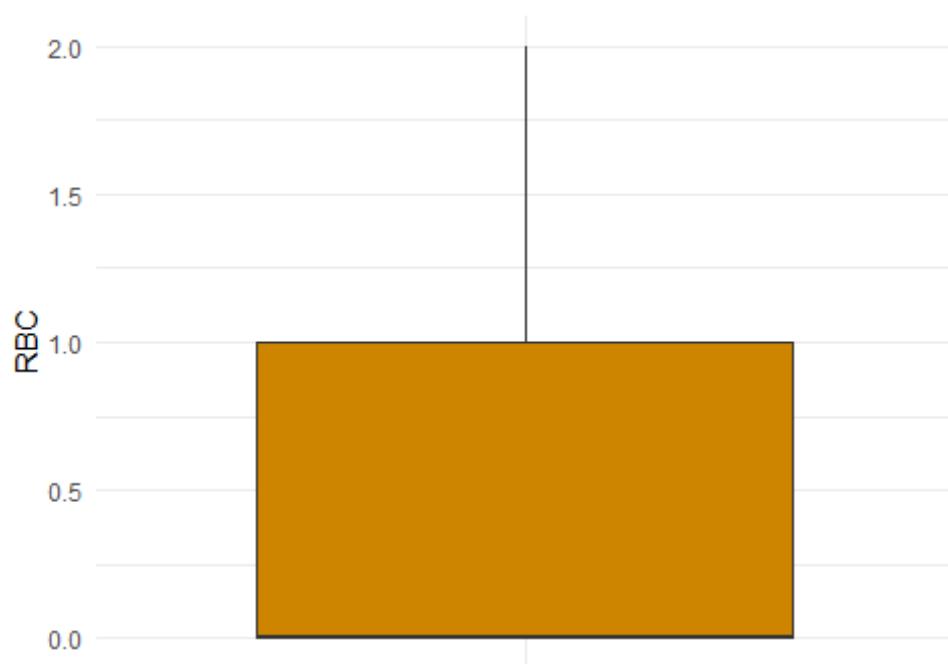
Box Plot di EA



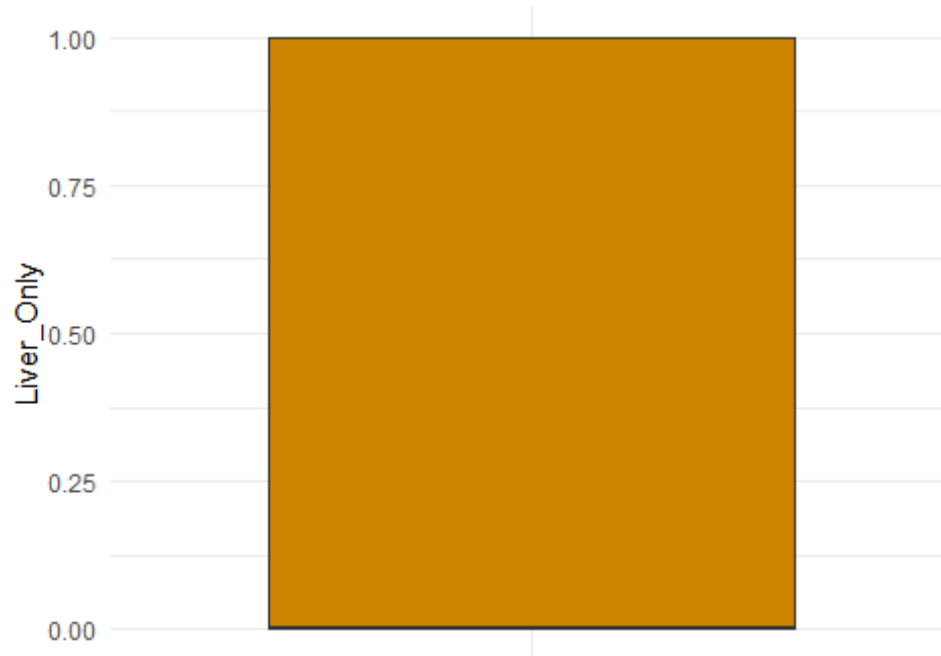
Box Plot di Log2AT



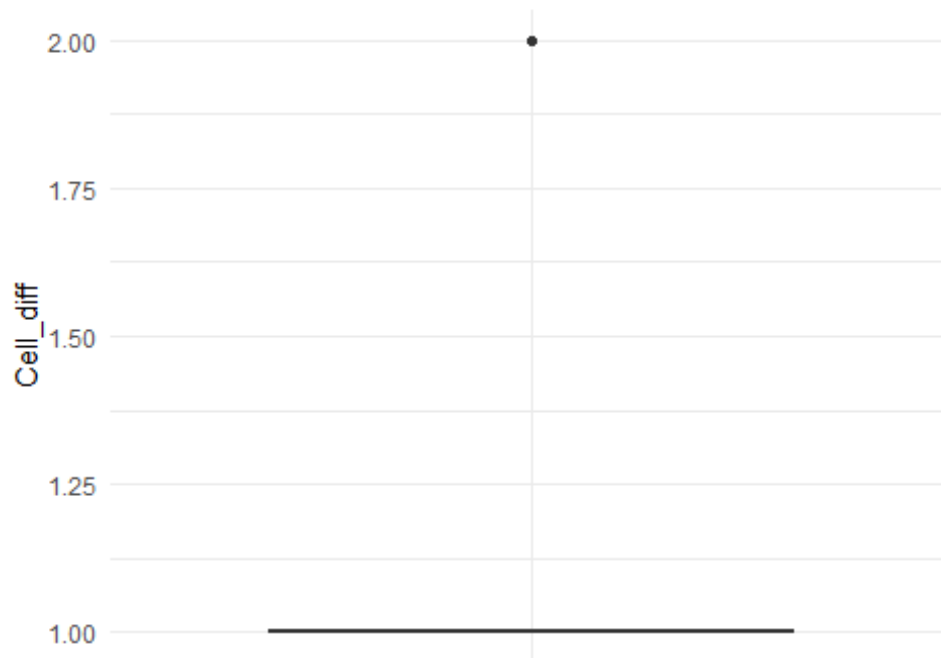
Box Plot di RBC



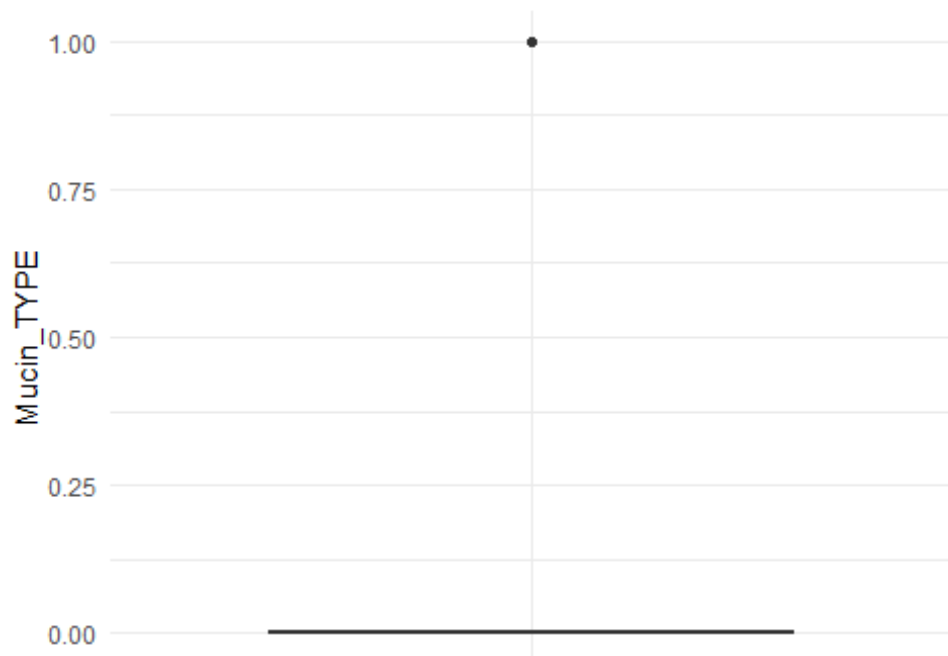
Box Plot di Liver_Only



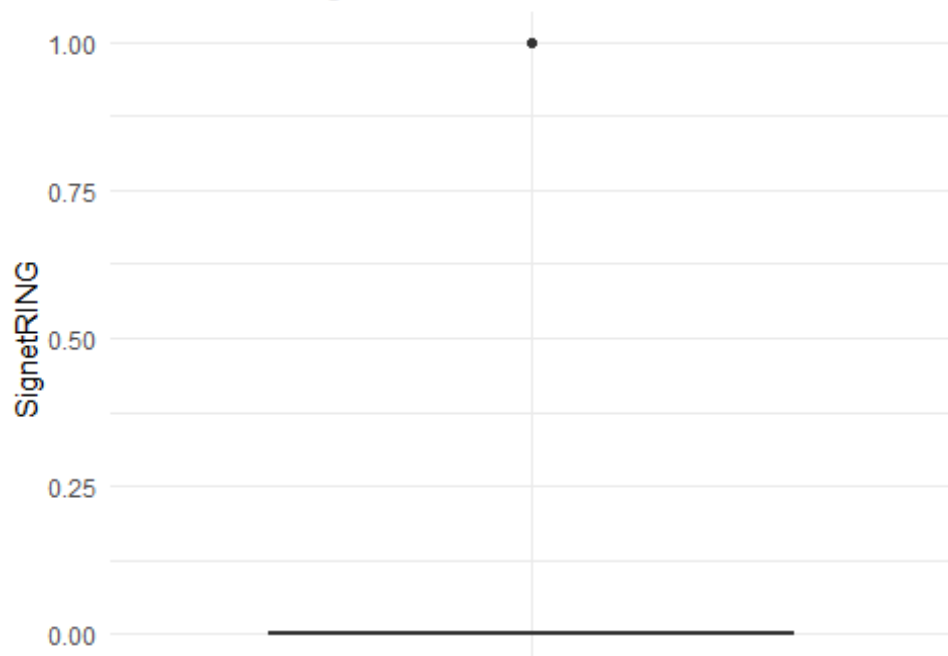
Box Plot di Cell_diff



Box Plot di Mucin_TYPE



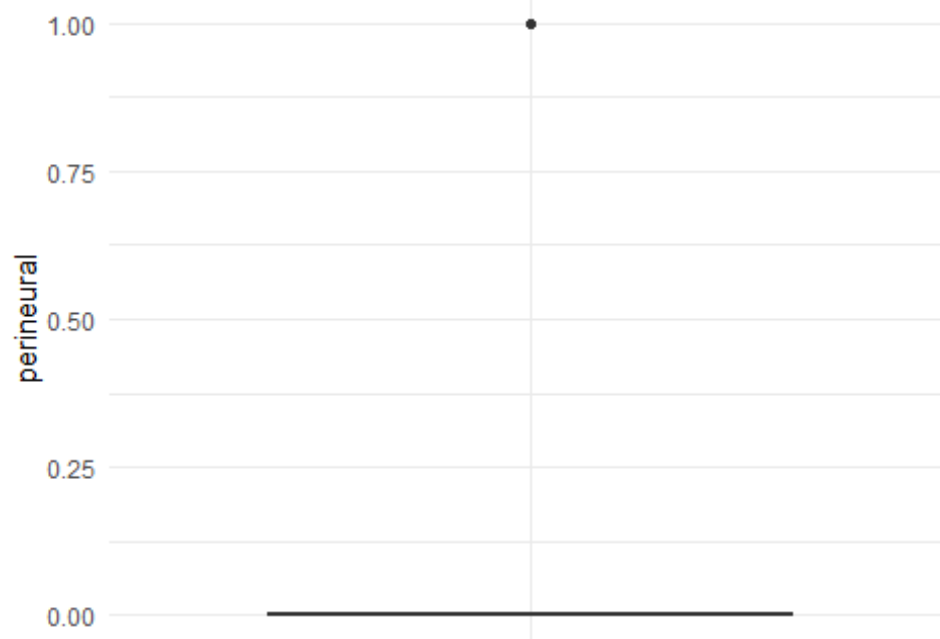
Box Plot di SignetRING



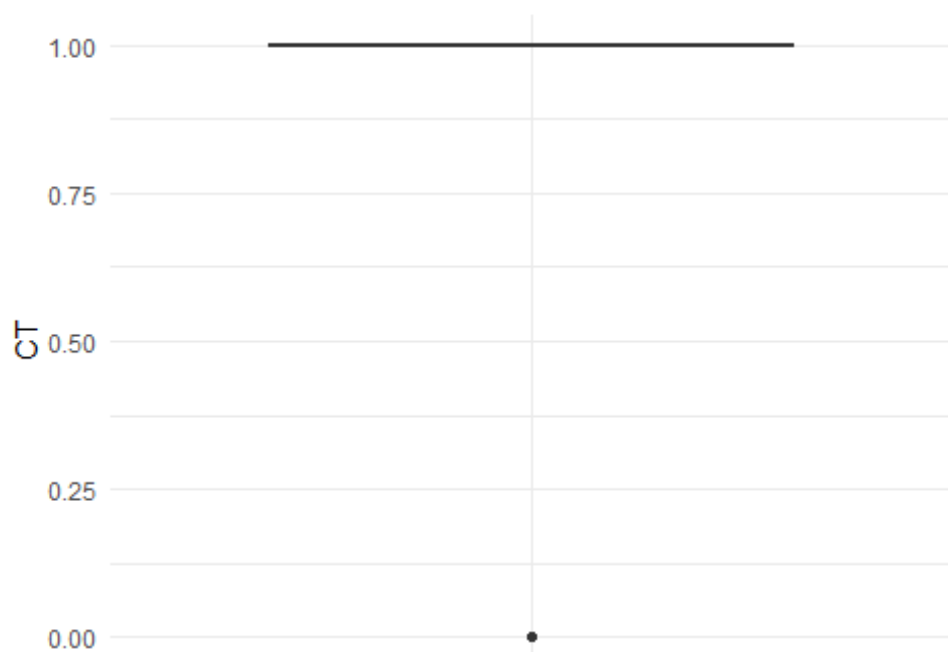
Box Plot di Lymphovascularinvasion



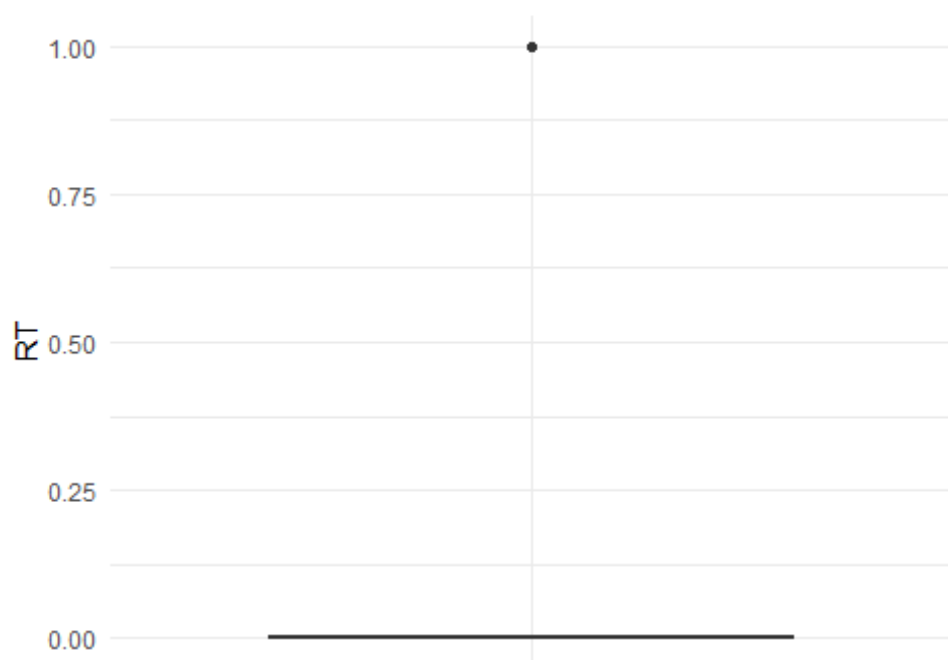
Box Plot di perineural



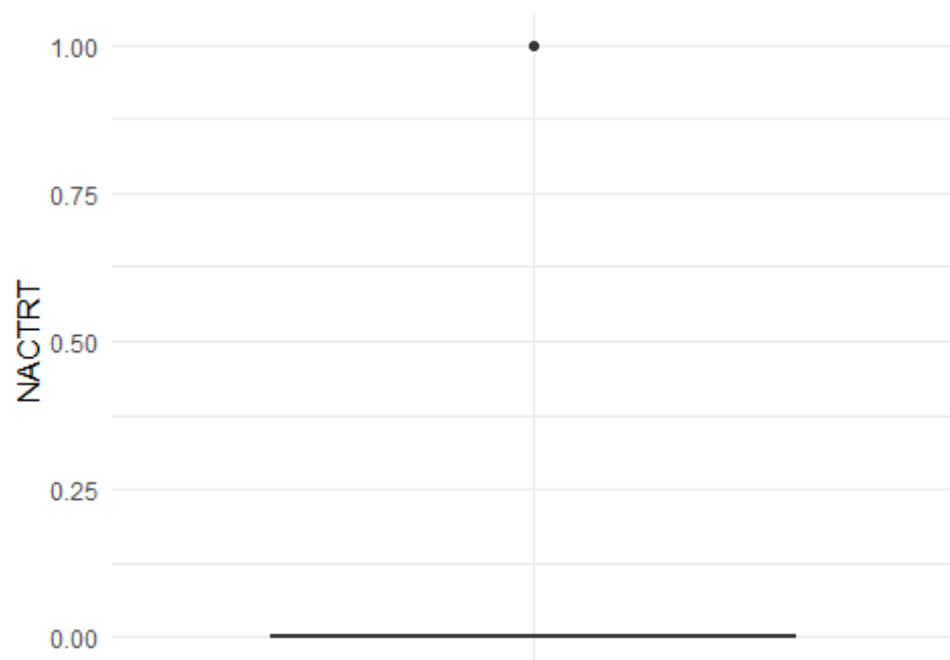
Box Plot di CT



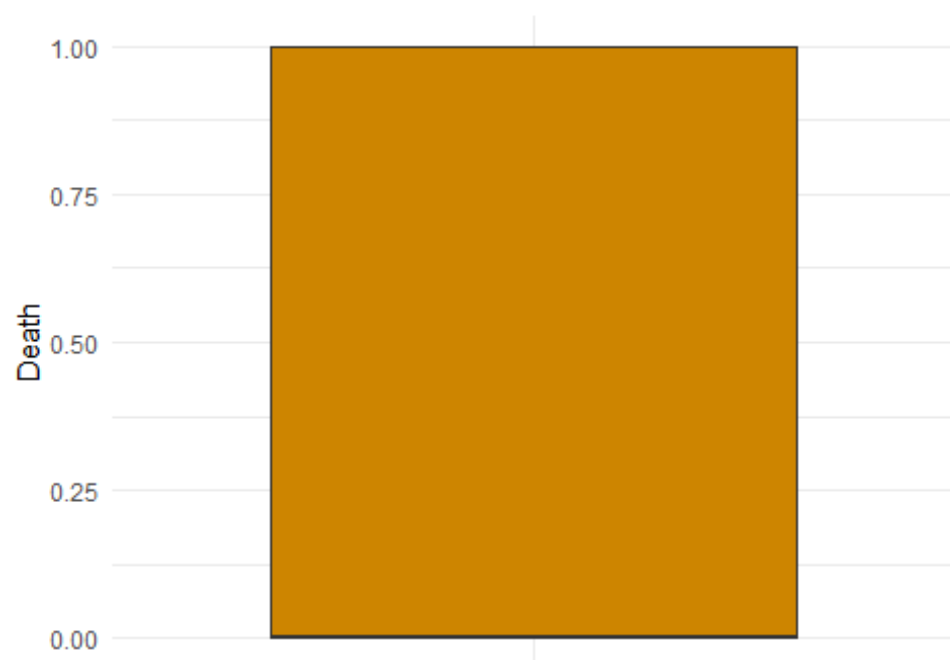
Box Plot di RT



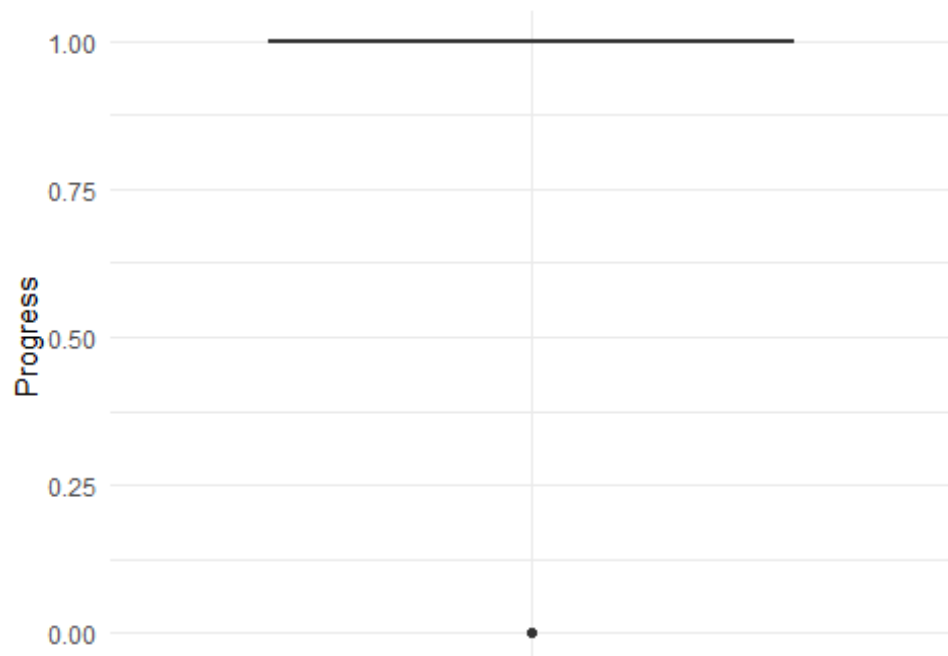
Box Plot di NACTRT



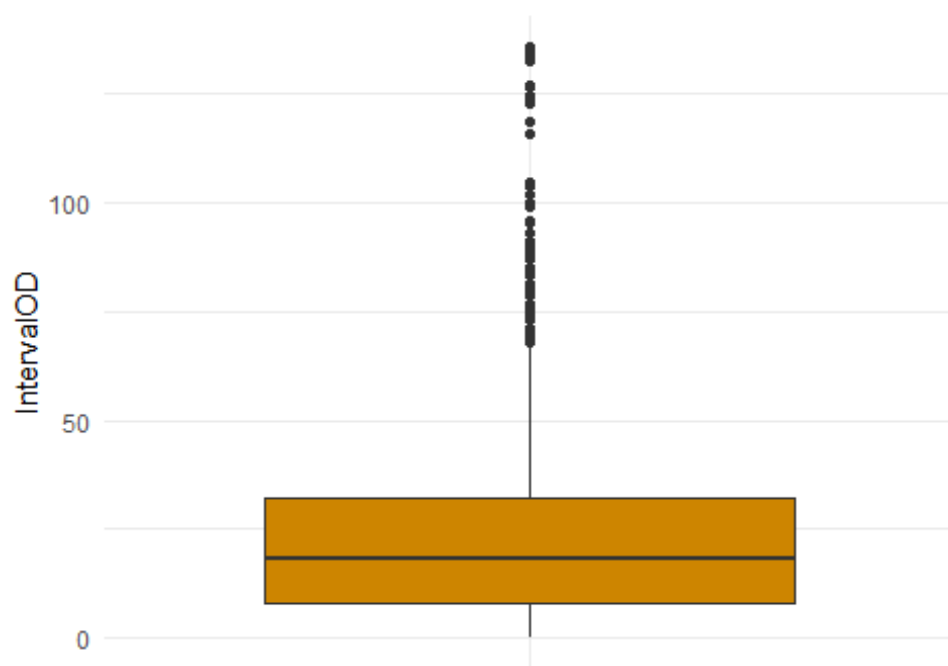
Box Plot di Death

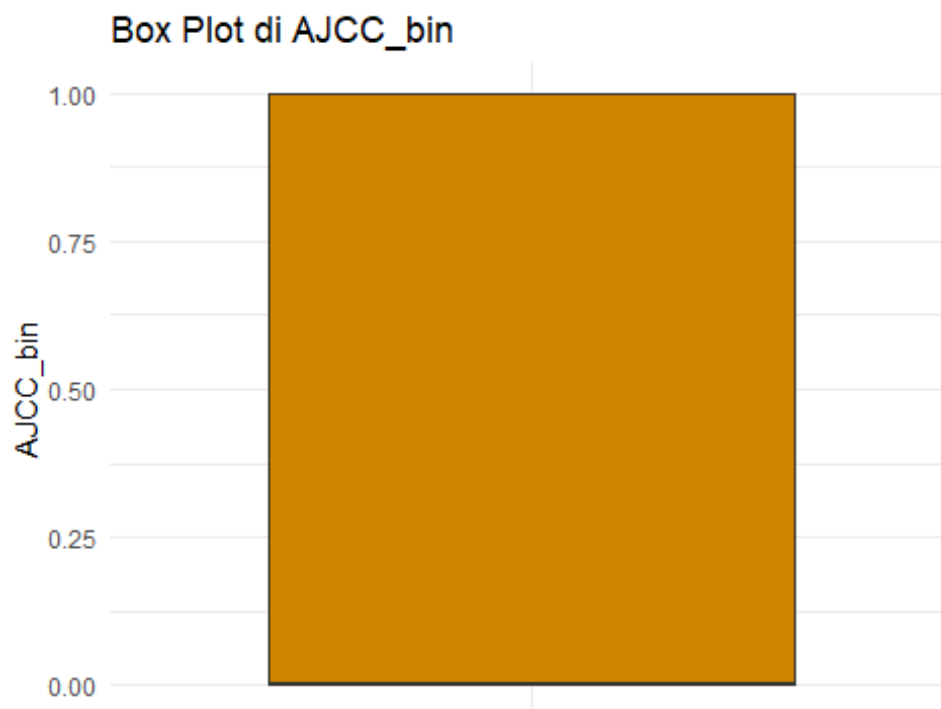
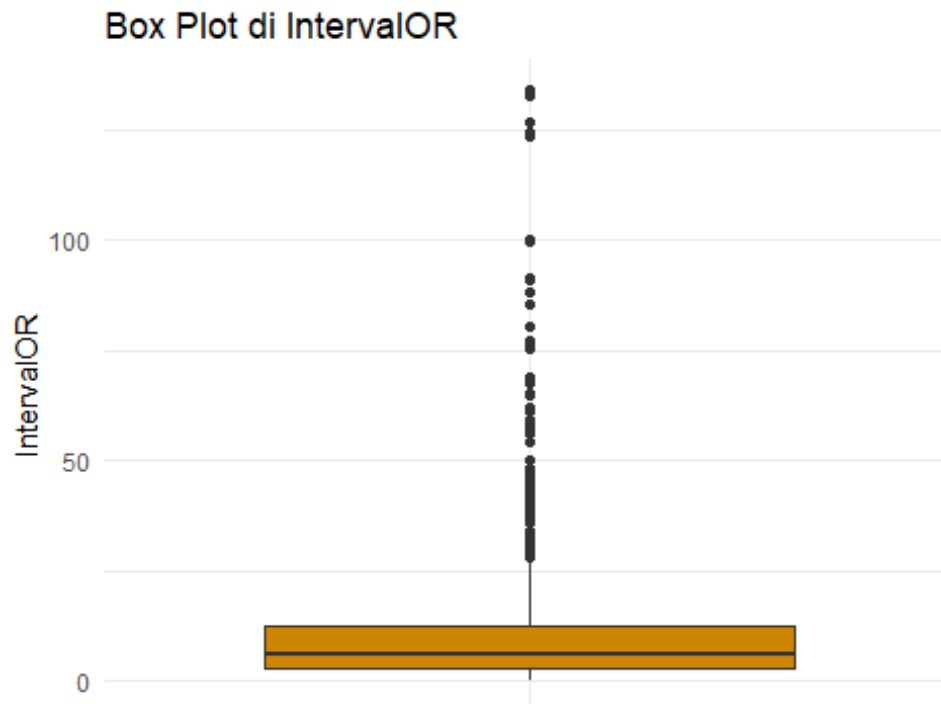


Box Plot di Progress



Box Plot di IntervalOD

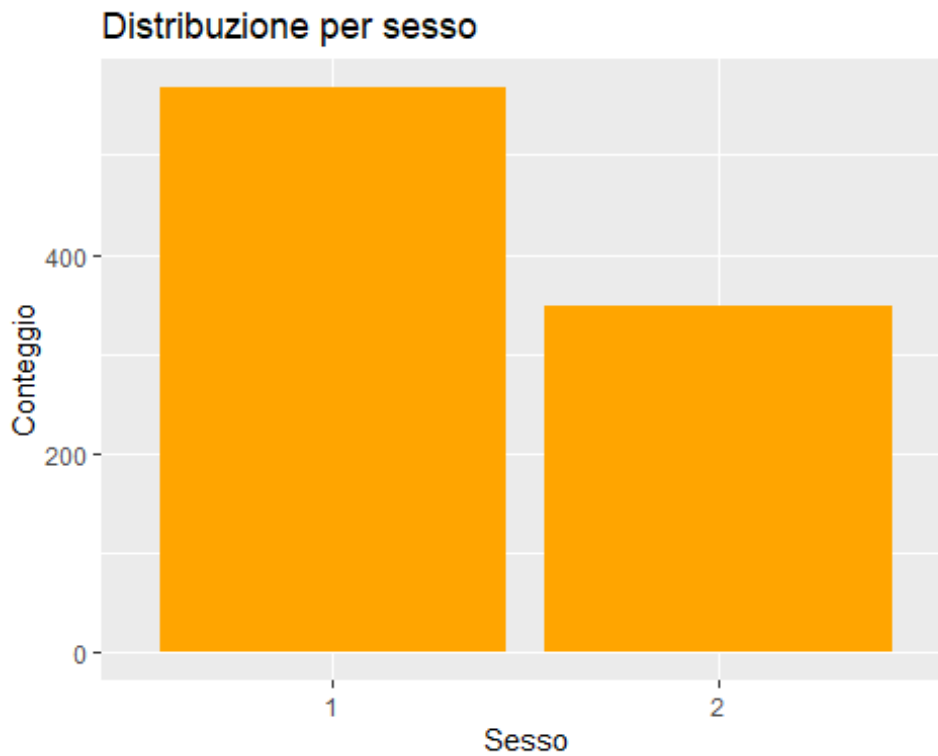




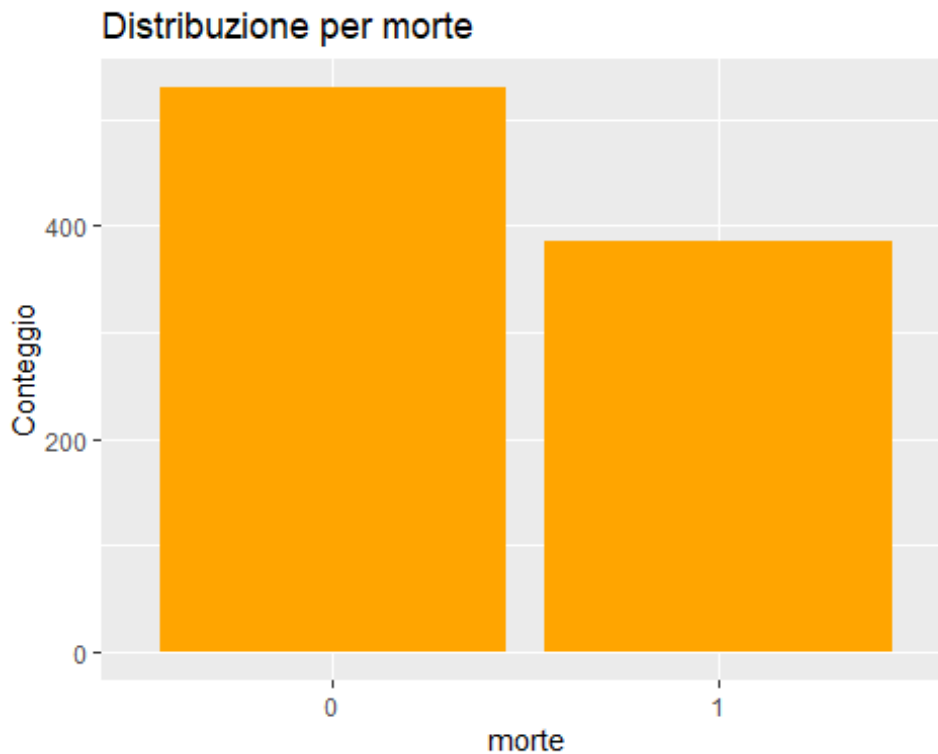
Variables

Si prova ad analizzare se i comportamenti delle variabili sono influenzati dal sesso o dall'eta

```
ggplot(data, aes(x = factor(Gender))) +  
  geom_bar(fill = "orange") +  
  labs(title = "Distribuzione per sesso", x = "Sesso", y = "Conteggio")
```



```
ggplot(data, aes(x = factor(Death))) +  
  geom_bar(fill = "orange") +  
  labs(title = "Distribuzione per morte", x = "morte", y = "Conteggio")
```



```
skim_without_charts(data$Age)
```

Data summary

Name data\$Age
 Number of rows 916
 Number of columns 1

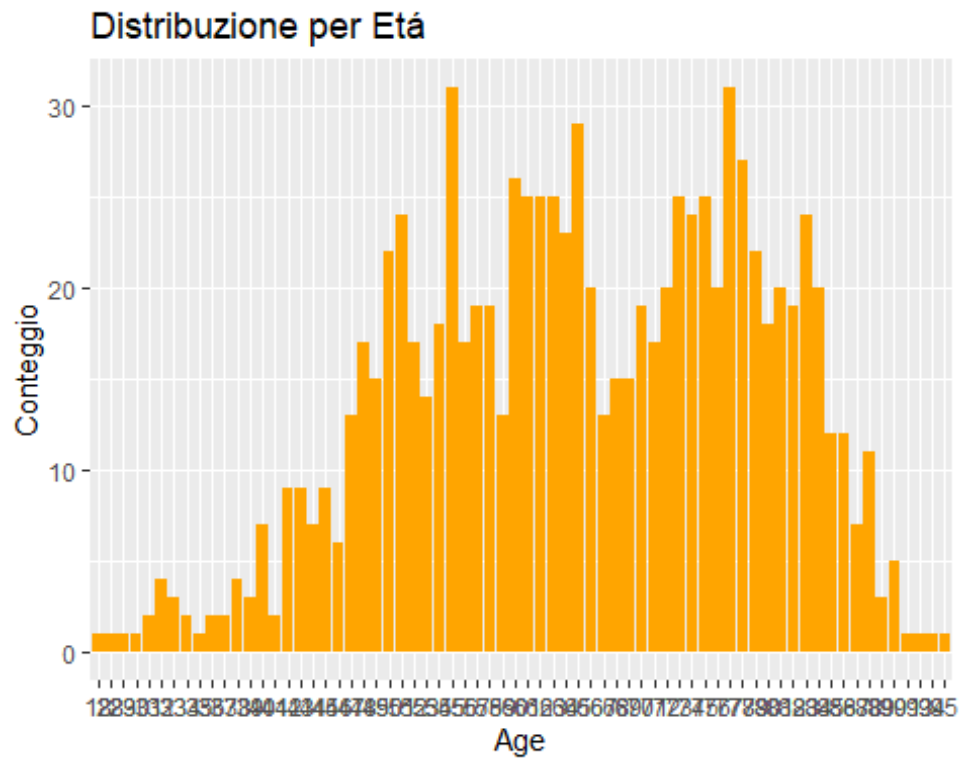
Column type frequency:
 numeric 1

Group variables None

Variable type: numeric

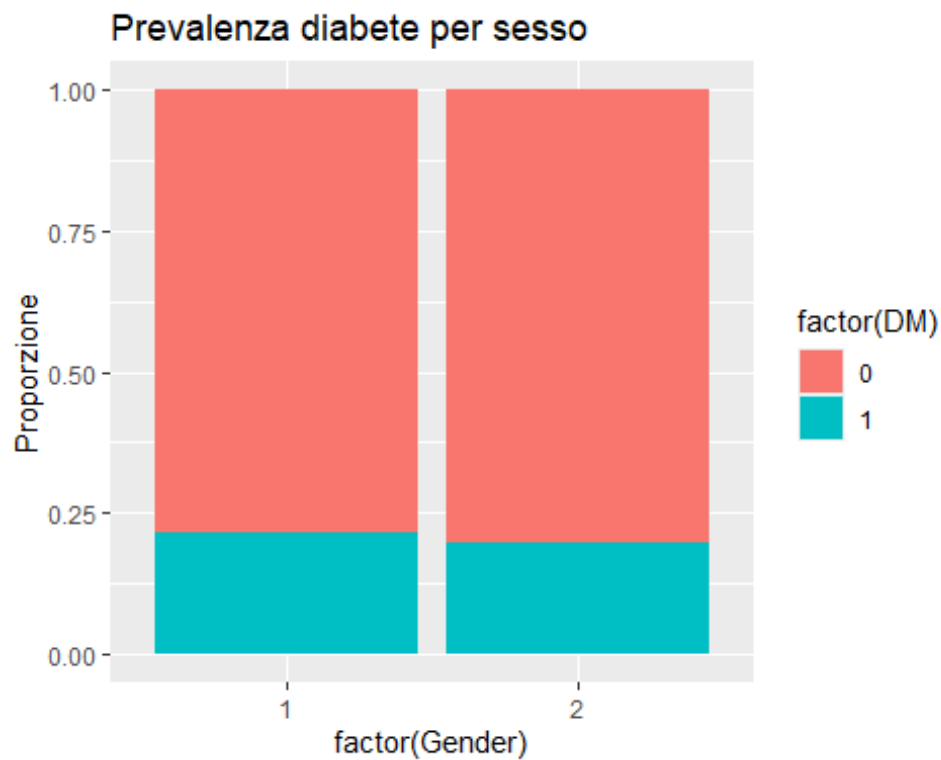
skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
data	0	1	65.23	13.62	18	55	65	77	95

```
ggplot(data, aes(x = factor(Age))) +  
  geom_bar(fill = "orange") +  
  labs(title = "Distribuzione per Etá", x = "Age", y = "Conteggio")
```

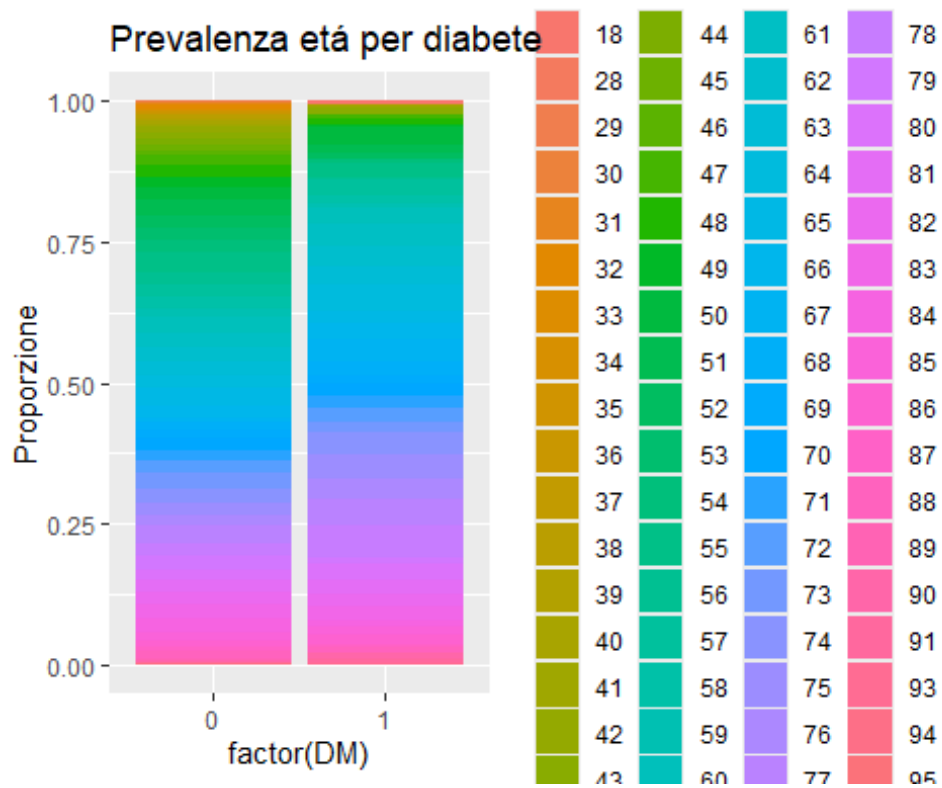


COMORBIDITÀ

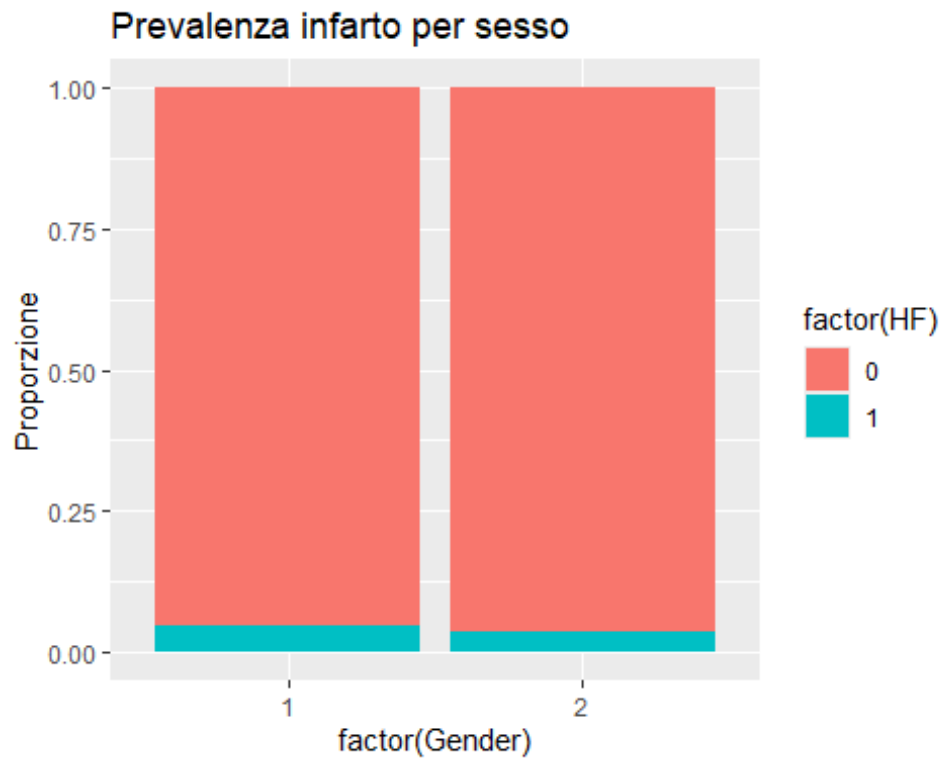
```
ggplot(data, aes(x = factor(Gender), fill = factor(DM))) +  
  geom_bar(position = "fill") +  
  labs(title = "Prevalenza diabete per sesso", y = "Proporzione")
```



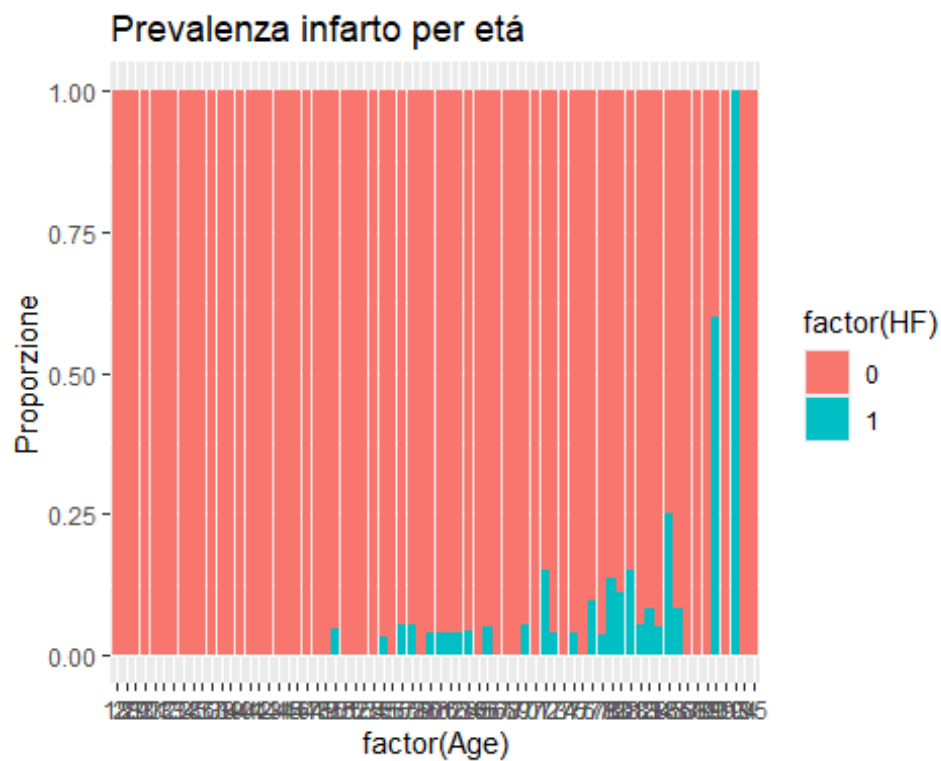
```
ggplot(data, aes(x = factor(DM), fill = factor(Age))) +
  geom_bar(position = "fill") +
  labs(title = "Prevalenza età per diabete ", y = "Proporzione")
```



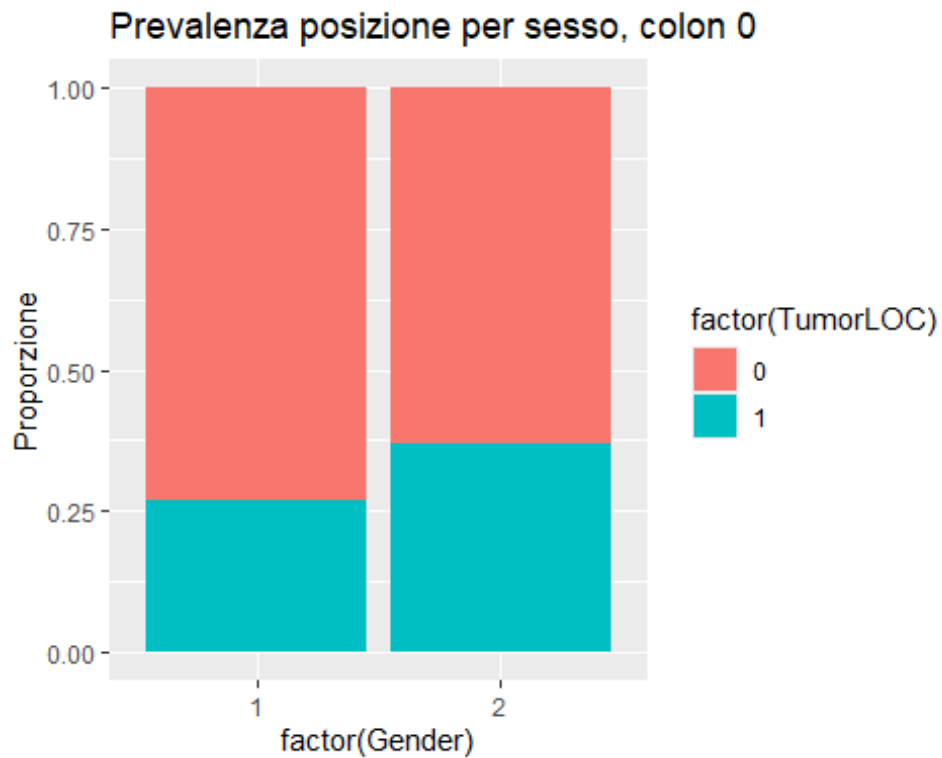
```
ggplot(data, aes(x = factor(Gender), fill = factor(HF))) +
  geom_bar(position = "fill") +
  labs(title = "Prevalenza infarto per sesso", y = "Proporzione")
```



```
ggplot(data, aes(x = factor(Age), fill = factor(HF))) +
  geom_bar(position = "fill") +
  labs(title = "Prevalenza infarto per età", y = "Proporzione")
```

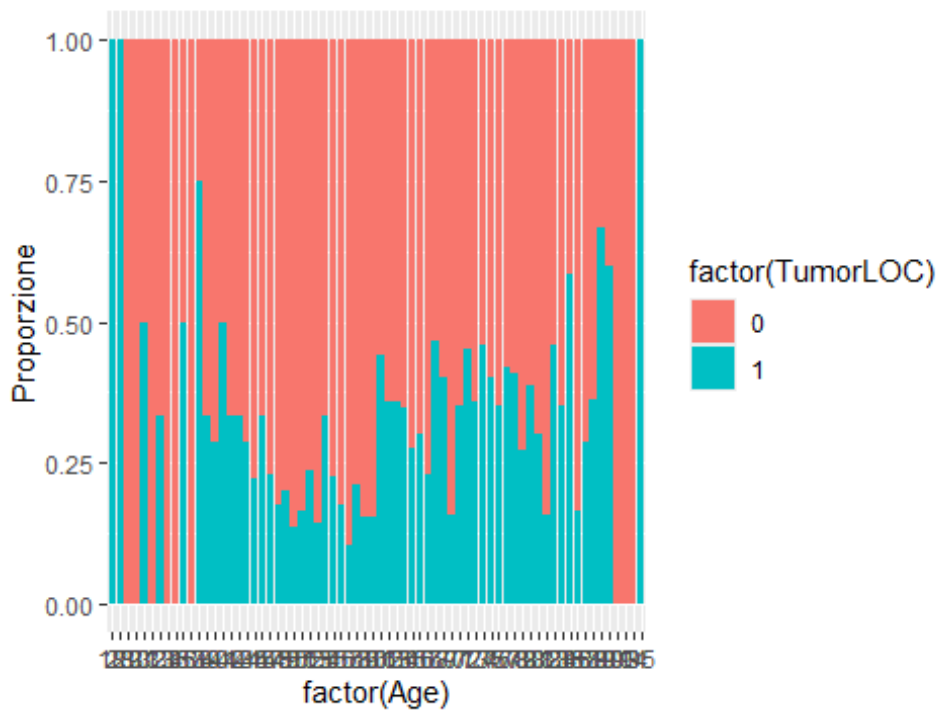



```
ggplot(data, aes(x = factor(Gender), fill = factor(TumorLOC))) +  
  geom_bar(position = "fill") +  
  labs(title = "Prevalenza posizione per sesso, colon 0", y = "Proporzione")
```

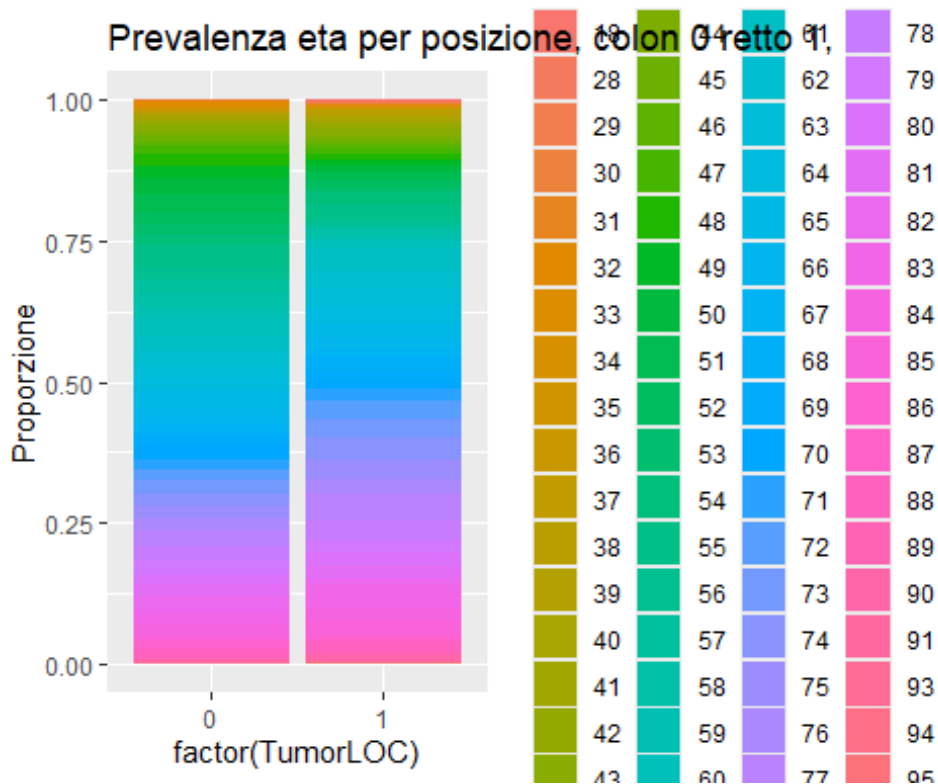


```
ggplot(data, aes(x = factor(Age), fill = factor(TumorLOC))) +  
  geom_bar(position = "fill") +  
  labs(title = "Prevalenza posizione per eta, colon 0 retto 1, ", y =  
"Proporzione")
```

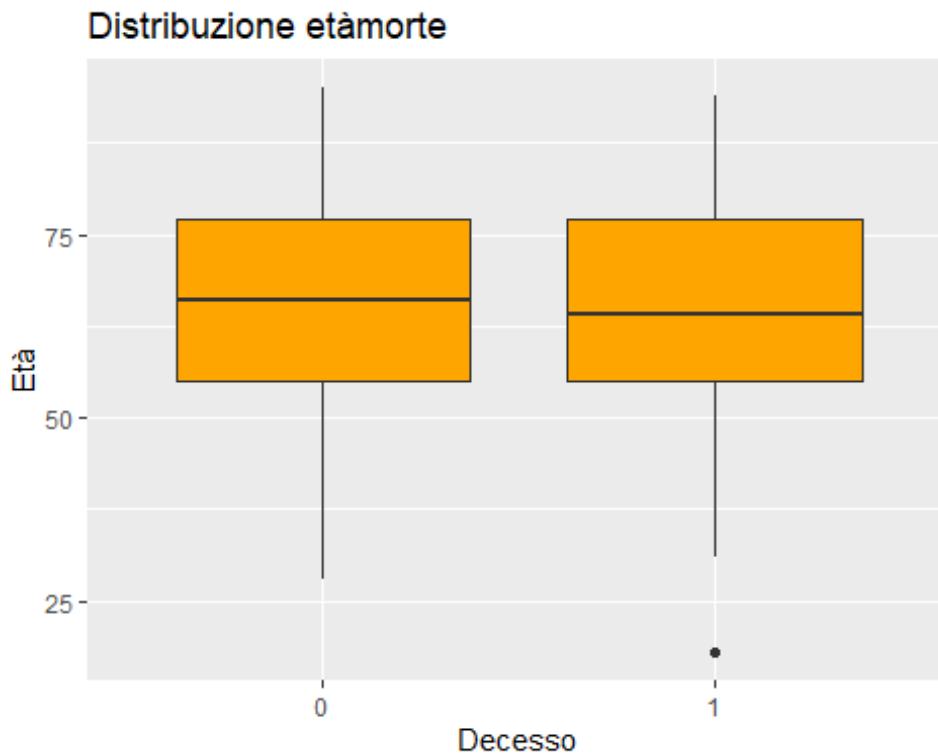
Prevalenza posizione per eta, colon 0 retto 1,



```
ggplot(data, aes(x = factor(TumorLOC), fill = factor(Age))) +
  geom_bar(position = "fill") +
  labs(title = "Prevalenza eta per posizione, colon 0 retto 1, ", y =
"Proporzione")
```

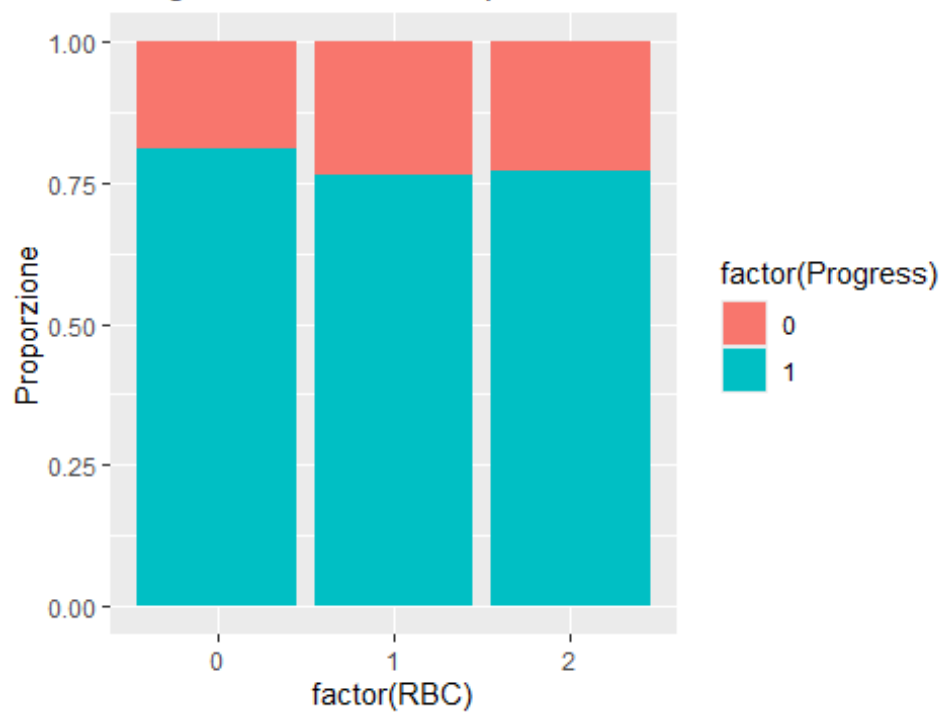


```
ggplot(data, aes(x = factor(Death), y = Age)) +
  geom_boxplot(fill = "orange") +
  labs(title = "Distribuzione etàmorte", x = "Decesso", y = "Età")
```

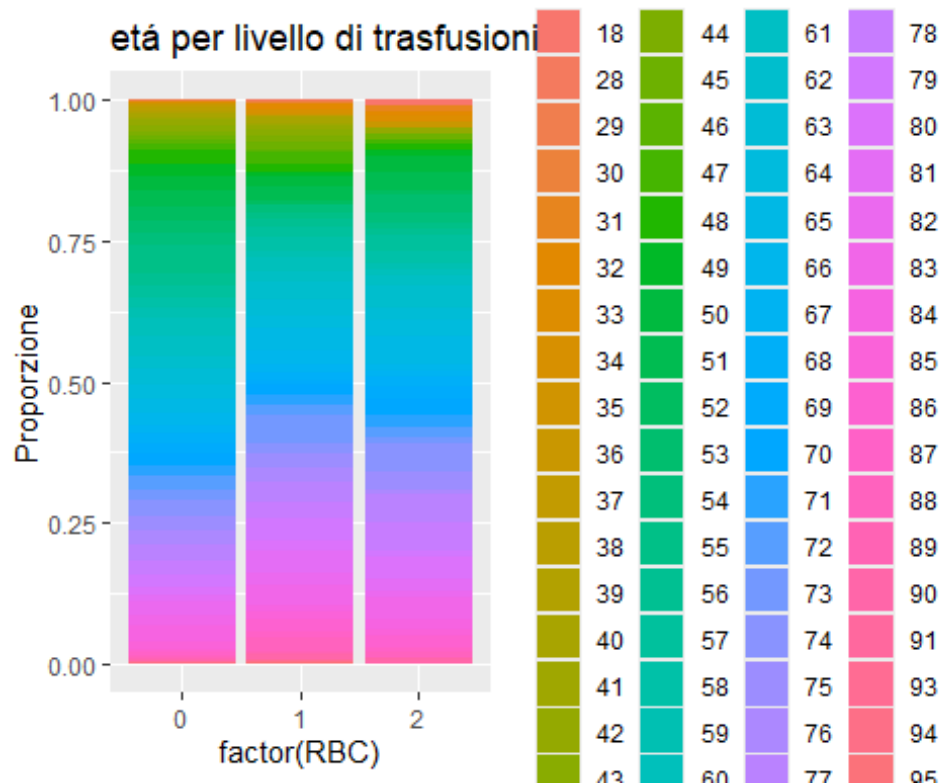


```
ggplot(data, aes(x = factor(RBC), fill = factor(Progress))) +
  geom_bar(position = "fill") +
  labs(title = "Progressione tumorale per livello di trasfusioni", y =
"Proporzione")
```

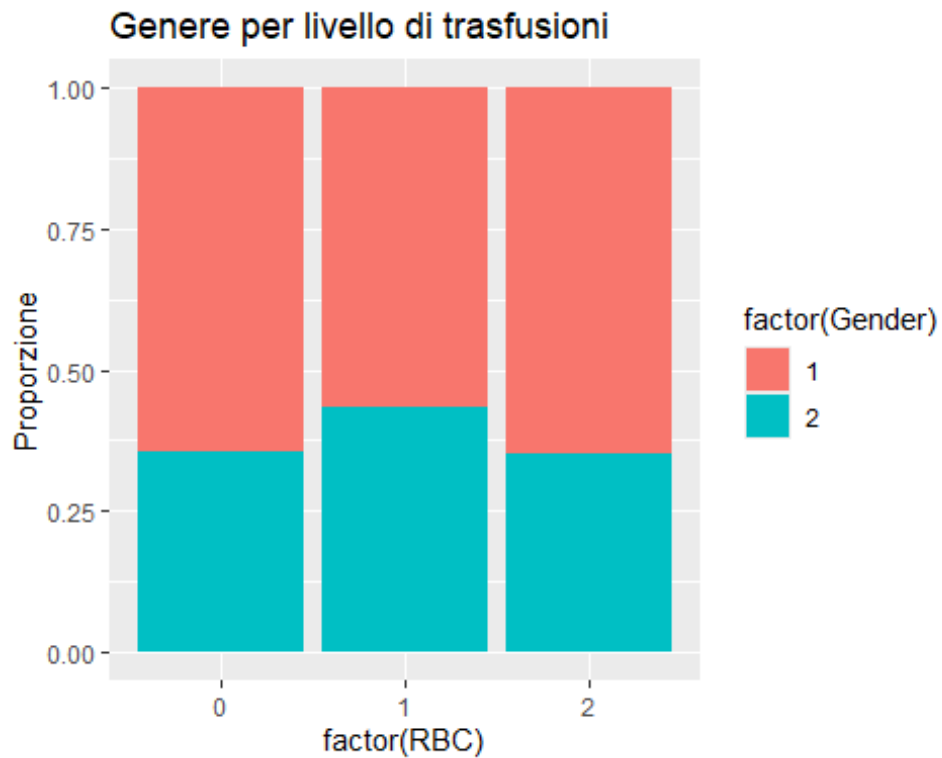
Progressione tumorale per livello di trasfusioni



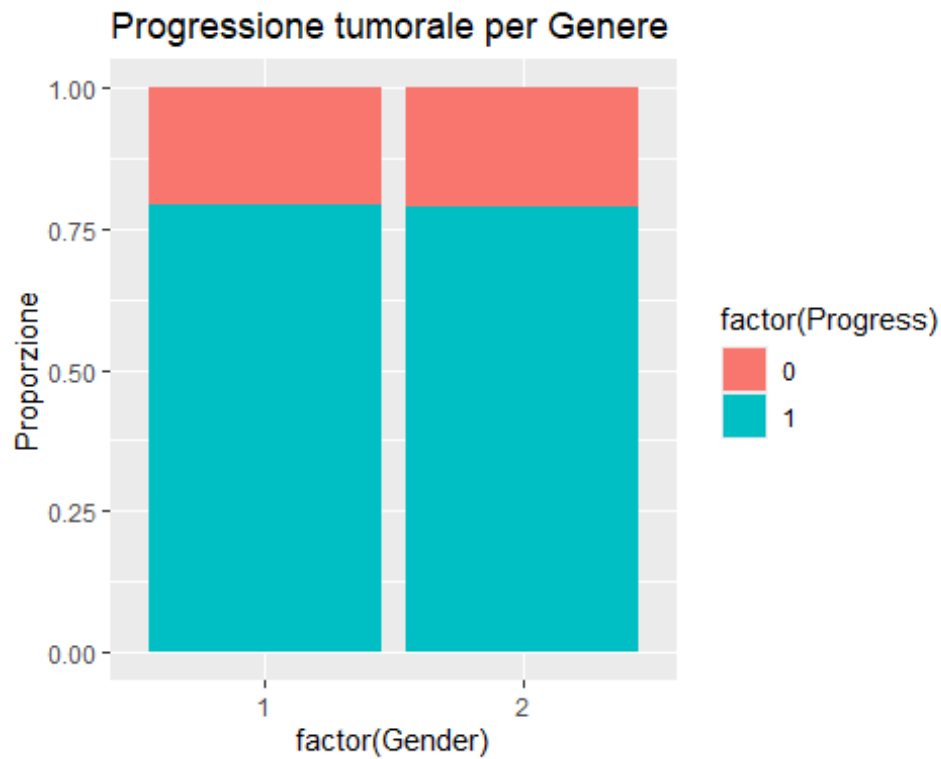
```
ggplot(data, aes(x = factor(RBC), fill = factor(Age))) +  
  geom_bar(position = "fill") +  
  labs(title = "età per livello di trasfusioni", y = "Proporzione")
```



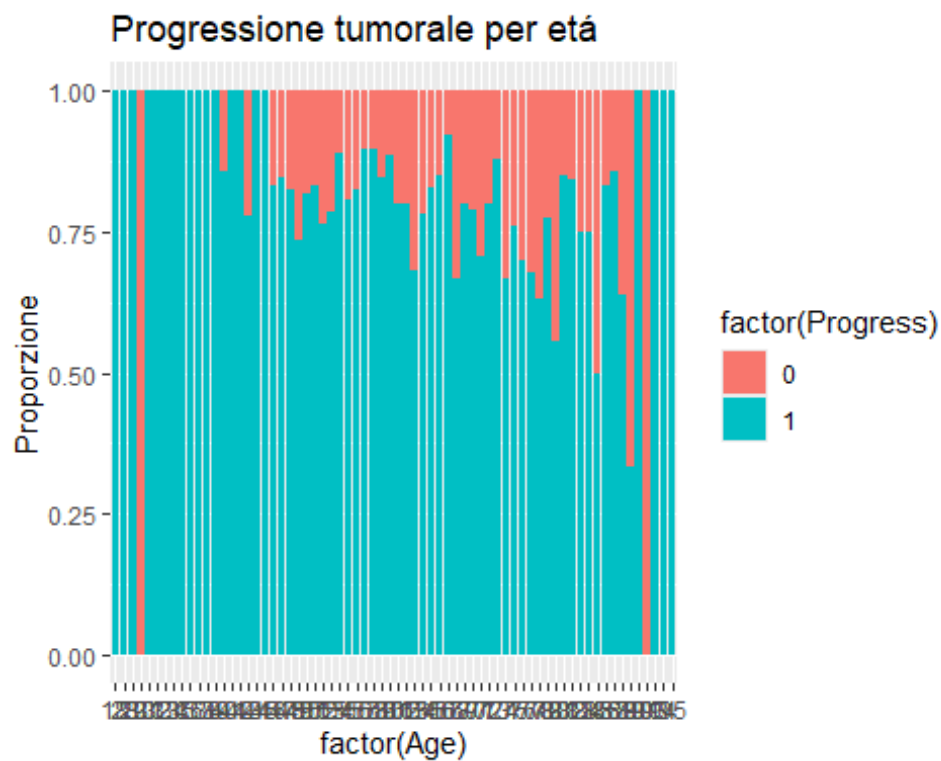
```
ggplot(data, aes(x = factor(RBC), fill = factor(Gender))) +  
  geom_bar(position = "fill") +  
  labs(title = "Genere per livello di trasfusioni", y = "Proporzione")
```



```
ggplot(data, aes(x = factor(Gender), fill = factor(Progress))) +  
  geom_bar(position = "fill") +  
  labs(title = "Progressione tumorale per Genere", y = "Proporzione")
```



```
ggplot(data, aes(x = factor(Age), fill = factor(Progress))) +
  geom_bar(position = "fill") +
  labs(title = "Progressione tumorale per età", y = "Proporzione")
```



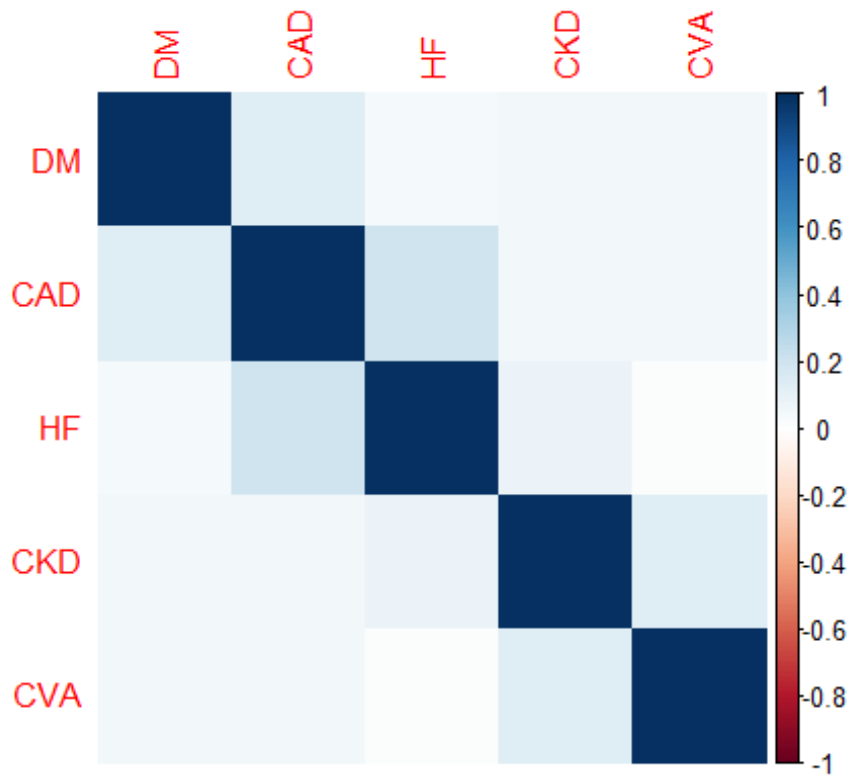
CORRELATION

COMORBIDITA CORRELAZIONE

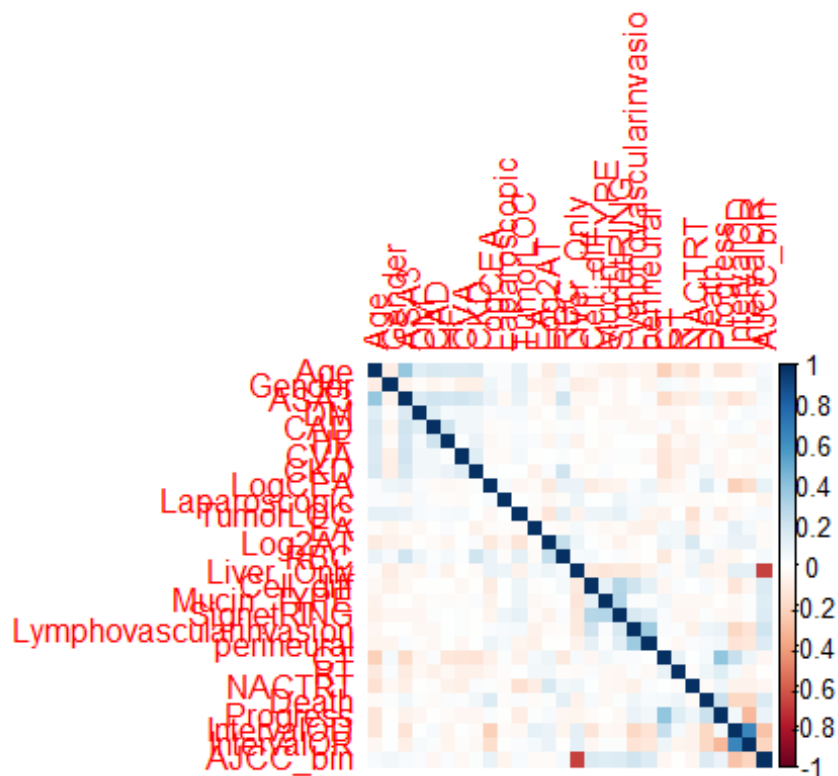
```
library(corrplot)

## Warning: package 'corrplot' was built under R version 4.3.3
## corrplot 0.92 loaded

corrplot(cor(data[, c("DM", "CAD", "HF", "CKD", "CVA")]), method = "color")
```



```
library(corrplot)
corrplot(cor(data), method = "color")
```



```
cor_matrix <- cor(data, use = "complete.obs")
```

Filtra le correlazioni con valore assoluto maggiore di 0.5 (escludendo la diagonale)

```
high_corr <- cor_matrix
```

```
high_corr[abs(high_corr) <= 0.35 | diag(ncol(high_corr)) == 1] <- NA
```

Mostra solo le correlazioni forti

```
print(high_corr)
```

[illegible]

## SignetRING	NA	NA		NA	NA	NA	NA	NA	NA	NA
## Lymphovascularinvasion	NA	NA		NA	NA	NA	NA	NA	NA	NA
## perineural	NA	NA		NA	NA	NA	NA	NA	NA	NA
## CT	NA	NA		NA	NA	NA	NA	NA	NA	NA
## RT	NA	NA		NA	NA	NA	NA	NA	NA	NA
## NACTRT	NA	NA		NA	NA	NA	NA	NA	NA	NA
## Death	NA	NA		NA	NA	NA	NA	NA	NA	NA
## Progress	NA	NA		NA	NA	NA	NA	NA	NA	NA
## IntervalOD	NA	NA		NA	NA	NA	NA	NA	NA	NA
## IntervalOR	NA	NA		NA	NA	NA	NA	NA	NA	NA
## AJCC_bin	NA	NA		NA	NA	NA	NA	NA	NA	NA
##	Laparoscopic	TumorLOC	EA	Log2AT	RBC	Liver_Only	Cell_diff			
## Age	NA	NA	NA	NA	NA	NA	NA			NA
## Gender	NA	NA	NA	NA	NA	NA	NA			NA
## ASA3	NA	NA	NA	NA	NA	NA	NA			NA
## DM	NA	NA	NA	NA	NA	NA	NA			NA
## CAD	NA	NA	NA	NA	NA	NA	NA			NA
## HF	NA	NA	NA	NA	NA	NA	NA			NA
## CVA	NA	NA	NA	NA	NA	NA	NA			NA
## CKD	NA	NA	NA	NA	NA	NA	NA			NA
## LogCEA	NA	NA	NA	NA	NA	NA	NA			NA
## Laparoscopic	NA	NA	NA	NA	NA	NA	NA			NA
## TumorLOC	NA	NA	NA	NA	NA	NA	NA			NA
## EA	NA	NA	NA	NA	NA	NA	NA			NA
## Log2AT	NA	NA	NA	NA	NA	NA	NA			NA
## RBC	NA	NA	NA	NA	NA	NA	NA			NA
## Liver_Only	NA	NA	NA	NA	NA	NA	NA			NA
## Cell_diff	NA	NA	NA	NA	NA	NA	NA			NA
## Mucin_TYPE	NA	NA	NA	NA	NA	NA	NA			NA
## SignetRING	NA	NA	NA	NA	NA	NA	NA			NA
## Lymphovascularinvasion	NA	NA	NA	NA	NA	NA	NA			NA
## perineural	NA	NA	NA	NA	NA	NA	NA			NA
## CT	NA	NA	NA	NA	NA	NA	NA			NA
## RT	NA	NA	NA	NA	NA	NA	NA			NA
## NACTRT	NA	NA	NA	NA	NA	NA	NA			NA
## Death	NA	NA	NA	NA	NA	NA	NA			NA
## Progress	NA	NA	NA	NA	NA	NA	NA			NA
## IntervalOD	NA	NA	NA	NA	NA	NA	NA			NA
## IntervalOR	NA	NA	NA	NA	NA	NA	NA			NA
## AJCC_bin	NA	NA	NA	NA	NA	-0.6804736				NA
##	Mucin_TYPE	SignetRING	Lymphovascularinvasion	perineural						
## Age	NA	NA				NA				NA
## Gender	NA	NA				NA				NA
## ASA3	NA	NA				NA				NA
## DM	NA	NA				NA				NA
## CAD	NA	NA				NA				NA
## HF	NA	NA				NA				NA
## CVA	NA	NA				NA				NA
## CKD	NA	NA				NA				NA
## LogCEA	NA	NA				NA				NA
## Laparoscopic	NA	NA				NA				NA

## TumorLOC	NA	NA	NA	NA
## EA	NA	NA	NA	NA
## Log2AT	NA	NA	NA	NA
## RBC	NA	NA	NA	NA
## Liver_Only	NA	NA	NA	NA
## Cell_diff	NA	NA	NA	NA
## Mucin_TYPE	NA	NA	NA	NA
## SignetRING	NA	NA	NA	NA
## Lymphovascularinvasion	NA	NA	NA	0.3550012
## perineural	NA	NA	0.3550012	NA
## CT	NA	NA	NA	NA
## RT	NA	NA	NA	NA
## NACTRT	NA	NA	NA	NA
## Death	NA	NA	NA	NA
## Progress	NA	NA	NA	NA
## IntervalOD	NA	NA	NA	NA
## IntervalOR	NA	NA	NA	NA
## AJCC_bin	NA	NA	NA	NA
##	CT	RT	NACTRT	Death
## Age	NA	NA	NA	NA
## Gender	NA	NA	NA	NA
## ASA3	NA	NA	NA	NA
## DM	NA	NA	NA	NA
## CAD	NA	NA	NA	NA
## HF	NA	NA	NA	NA
## CVA	NA	NA	NA	NA
## CKD	NA	NA	NA	NA
## LogCEA	NA	NA	NA	NA
## Laparoscopic	NA	NA	NA	NA
## TumorLOC	NA	NA	NA	NA
## EA	NA	NA	NA	NA
## Log2AT	NA	NA	NA	NA
## RBC	NA	NA	NA	NA
## Liver_Only	NA	NA	NA	NA
## Cell_diff	NA	NA	NA	NA
## Mucin_TYPE	NA	NA	NA	NA
## SignetRING	NA	NA	NA	NA
## Lymphovascularinvasion	NA	NA	NA	NA
## perineural	NA	NA	NA	NA
## CT	NA	NA	NA	0.3931775
## RT	NA	NA	NA	NA
## NACTRT	NA	NA	NA	NA
## Death	NA	NA	NA	NA
## Progress	0.3931775	NA	NA	NA
## IntervalOD	NA	NA	NA	NA
## IntervalOR	NA	NA	NA	0.6507935
## AJCC_bin	NA	NA	NA	NA
##	IntervalOR	AJCC_bin		
## Age	NA	NA		
## Gender	NA	NA		
## ASA3	NA	NA		

## DM	NA	NA
## CAD	NA	NA
## HF	NA	NA
## CVA	NA	NA
## CKD	NA	NA
## LogCEA	NA	NA
## Laparoscopic	NA	NA
## TumorLOC	NA	NA
## EA	NA	NA
## Log2AT	NA	NA
## RBC	NA	NA
## Liver_Only	NA	-0.6804736
## Cell_diff	NA	NA
## Mucin_TYPE	NA	NA
## SignetRING	NA	NA
## Lymphovascularinvasion	NA	NA
## perineural	NA	NA
## CT	NA	NA
## RT	NA	NA
## NACTRT	NA	NA
## Death	NA	NA
## Progress	NA	NA
## IntervalOD	0.6507935	NA
## IntervalOR	NA	NA
## AJCC_bin	NA	NA

Correlazioni emerse Age – ASA3: 0.4647 ► Interpretazione: C'è una correlazione positiva moderata tra età e lo score ASA3 (American Society of Anesthesiologists physical status classification). Ha senso clinicamente: all'aumentare dell'età, è più probabile che il paziente abbia una condizione clinica peggiore.

Liver_Only – AJCC_bin: -0.5995 ► Interpretazione: C'è una correlazione negativa moderata/forte tra la presenza di metastasi epatiche isolate e la variabile binaria AJCC_bin (probabilmente uno stadio AJCC categorizzato). Potrebbe indicare che i pazienti con solo metastasi epatiche tendono ad avere stadi AJCC più bassi rispetto a quelli con metastasi più estese.

Interval – IntervalR: 0.7079 ► Interpretazione: Correlazione fortemente positiva tra Interval e Interval

```
# Pearson
cor_pearson <- cor(data, use = "complete.obs", method = "pearson")

# Spearman
cor_spearman <- cor(data, use = "complete.obs", method = "spearman")

# Kendall
cor_kendall <- cor(data, use = "complete.obs", method = "kendall")

library(Hmisc)

## Warning: package 'Hmisc' was built under R version 4.3.3
```

```
##
## Attaching package: 'Hmisc'

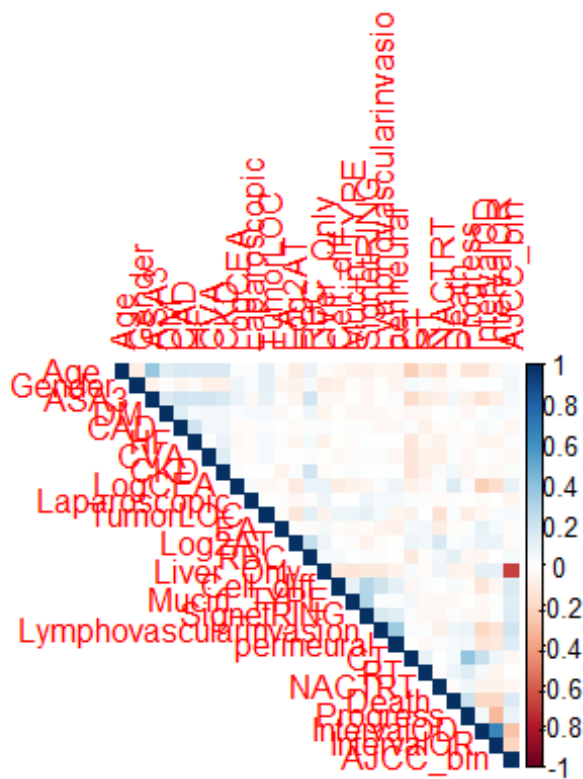
## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units

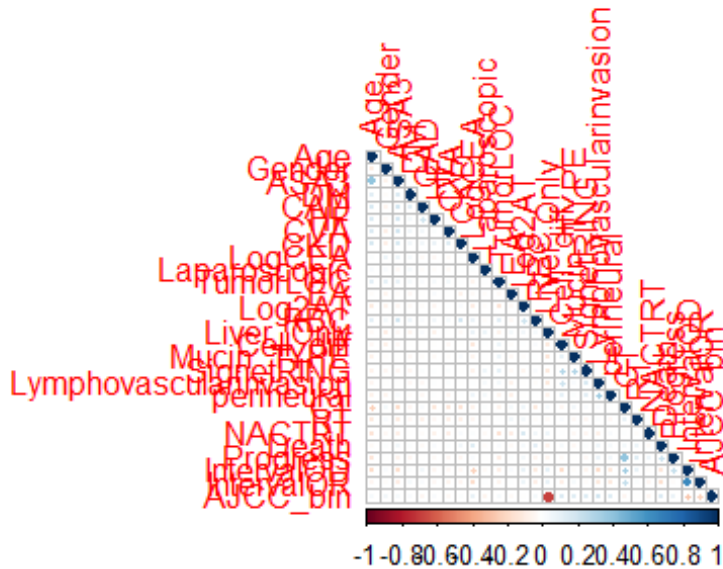
# Pearson con p-value
res_pearson <- rcorr(as.matrix(data), type = "pearson")
res_spearman <- rcorr(as.matrix(data), type = "spearman")

library(corrplot)

# Pearson
corrplot(cor_pearson, method = "color", type = "upper")
```



```
# Spearman
corrplot(cor_spearman, method = "circle", type = "lower")
```



```
methods <- c("pearson", "spearman", "kendall")
```

```
for (m in methods) {
  cat("\n==== Metodo:", m, "====\n")
  cor_matrix <- cor(data, use = "complete.obs", method = m)
  print(round(cor_matrix, 3))
}
```

```
##
```

```
## ===== Metodo: pearson =====
```

	Age	Gender	ASA3	DM	CAD	HF	CVA	CKD
## Age	1.000	-0.099	0.388	0.130	0.165	0.159	0.154	0.168
## Gender	-0.099	1.000	-0.070	-0.020	-0.050	-0.024	-0.088	-0.084
## ASA3	0.388	-0.070	1.000	0.156	0.192	0.187	0.193	0.170
## DM	0.130	-0.020	0.156	1.000	0.131	0.048	0.057	0.050
## CAD	0.165	-0.050	0.192	0.131	1.000	0.204	0.055	0.059
## HF	0.159	-0.024	0.187	0.048	0.204	1.000	0.016	0.088
## CVA	0.154	-0.088	0.193	0.057	0.055	0.016	1.000	0.136
## CKD	0.168	-0.084	0.170	0.050	0.059	0.088	0.136	1.000
## LogCEA	0.021	0.069	0.082	0.028	-0.027	-0.038	-0.024	0.017
## Laparoscopic	0.028	0.022	-0.034	0.031	-0.016	-0.017	-0.027	0.001
## TumorLOC	0.106	0.108	0.099	0.032	0.028	0.032	0.048	0.002
## EA	-0.016	-0.051	-0.017	-0.032	-0.002	0.020	-0.009	-0.063
## Log2AT	-0.107	-0.069	-0.050	-0.005	-0.024	-0.051	0.009	0.021
## RBC	0.088	0.031	0.194	0.065	0.018	0.045	0.020	0.191
## Liver_Only	-0.016	-0.100	-0.034	-0.058	0.030	-0.061	-0.006	-0.026
## Cell_diff	-0.032	-0.070	-0.034	-0.003	-0.009	-0.002	-0.039	-0.007
## Mucin_TYPE	-0.070	-0.019	-0.008	-0.034	-0.018	-0.021	-0.018	0.021

## SignetRING	-0.071	-0.050	-0.048	-0.046	0.019	0.005	0.014	-0.008
## Lymphovascularinvasion	-0.042	-0.001	-0.003	-0.040	-0.034	0.004	0.016	-0.009
## perineural	-0.059	-0.055	0.024	-0.019	-0.033	0.000	-0.044	0.003
## CT	-0.231	0.050	-0.183	-0.048	0.027	-0.121	-0.132	-0.141
## RT	-0.117	0.027	-0.109	-0.039	-0.071	-0.039	-0.071	-0.074
## NACTRT	-0.154	-0.033	-0.086	-0.006	-0.019	-0.062	-0.081	0.025
## Death	-0.027	0.011	0.058	0.019	-0.022	0.023	0.016	0.053
## Progress	-0.134	-0.008	-0.061	-0.047	-0.008	-0.022	-0.045	-0.042
## IntervalOD	-0.138	0.004	-0.146	0.003	0.031	-0.055	0.009	-0.024
## IntervalOR	-0.040	0.002	-0.084	0.031	0.014	-0.052	-0.012	-0.032
## AJCC_bin	0.031	0.053	0.032	0.026	-0.047	0.024	-0.031	0.015
##	LogCEA	Laparoscopic	TumorLOC	EA	Log2AT	RBC		
## Age	0.021		0.028	0.106	-0.016	-0.107	0.088	
## Gender	0.069		0.022	0.108	-0.051	-0.069	0.031	
## ASA3	0.082		-0.034	0.099	-0.017	-0.050	0.194	
## DM	0.028		0.031	0.032	-0.032	-0.005	0.065	
## CAD	-0.027		-0.016	0.028	-0.002	-0.024	0.018	
## HF	-0.038		-0.017	0.032	0.020	-0.051	0.045	
## CVA	-0.024		-0.027	0.048	-0.009	0.009	0.020	
## CKD	0.017		0.001	0.002	-0.063	0.021	0.191	
## LogCEA	1.000		0.030	-0.049	-0.026	-0.118	0.080	
## Laparoscopic	0.030		1.000	0.008	-0.016	-0.046	-0.013	
## TumorLOC	-0.049		0.008	1.000	-0.021	-0.131	0.114	
## EA	-0.026		-0.016	-0.021	1.000	-0.018	-0.026	
## Log2AT	-0.118		-0.046	-0.131	-0.018	1.000	0.235	
## RBC	0.080		-0.013	0.114	-0.026	0.235	1.000	
## Liver_Only	0.080		0.075	-0.061	0.043	0.071	0.031	
## Cell_diff	-0.106		0.005	0.079	-0.011	0.010	0.043	
## Mucin_TYPE	-0.069		0.004	0.070	-0.038	-0.018	0.002	
## SignetRING	-0.054		0.063	0.108	-0.025	-0.069	0.025	
## Lymphovascularinvasion	0.021		0.033	0.048	-0.064	-0.029	-0.017	
## perineural	0.079		0.049	-0.052	-0.077	-0.007	-0.009	
## CT	-0.073		0.015	-0.064	0.044	0.078	-0.116	
## RT	-0.040		-0.017	0.007	0.056	0.050	-0.010	
## NACTRT	0.013		-0.027	-0.156	-0.094	0.163	0.058	
## Death	0.150		-0.018	0.022	-0.028	0.014	0.117	
## Progress	0.027		-0.004	-0.096	0.011	0.034	-0.050	
## IntervalOD	-0.237		-0.028	-0.040	0.104	0.081	-0.060	
## IntervalOR	-0.174		-0.006	0.025	0.049	0.081	-0.003	
## AJCC_bin	0.096		-0.051	0.038	-0.021	-0.023	0.048	
##	Liver_Only	Cell_diff	Mucin_TYPE	SignetRING				
## Age	-0.016	-0.032	-0.070	-0.071				
## Gender	-0.100	-0.070	-0.019	-0.050				
## ASA3	-0.034	-0.034	-0.008	-0.048				
## DM	-0.058	-0.003	-0.034	-0.046				
## CAD	0.030	-0.009	-0.018	0.019				
## HF	-0.061	-0.002	-0.021	0.005				
## CVA	-0.006	-0.039	-0.018	0.014				
## CKD	-0.026	-0.007	0.021	-0.008				
## LogCEA	0.080	-0.106	-0.069	-0.054				
## Laparoscopic	0.075	0.005	0.004	0.063				

## TumorLOC	-0.061	0.079	0.070	0.108		
## EA	0.043	-0.011	-0.038	-0.025		
## Log2AT	0.071	0.010	-0.018	-0.069		
## RBC	0.031	0.043	0.002	0.025		
## Liver_Only	1.000	-0.118	-0.135	-0.119		
## Cell_diff	-0.118	1.000	0.101	0.281		
## Mucin_TYPE	-0.135	0.101	1.000	0.298		
## SignetRING	-0.119	0.281	0.298	1.000		
## Lymphovascularinvasion	-0.116	0.208	0.020	0.168		
## perineural	-0.061	0.107	0.015	0.124		
## CT	0.034	-0.043	0.056	0.021		
## RT	-0.023	0.026	-0.032	-0.041		
## NACTRT	0.069	0.007	-0.032	-0.049		
## Death	-0.065	0.038	-0.034	0.018		
## Progress	-0.058	-0.021	-0.006	-0.032		
## IntervalOD	0.118	-0.136	-0.031	-0.048		
## IntervalOR	0.113	-0.042	0.016	0.014		
## AJCC_bin	-0.680	0.151	0.134	0.147		
##		Lymphovascularinvasion	perineural	CT	RT	NACTRT
## Age		-0.042	-0.059	-0.231	-0.117	-0.154
## Gender		-0.001	-0.055	0.050	0.027	-0.033
## ASA3		-0.003	0.024	-0.183	-0.109	-0.086
## DM		-0.040	-0.019	-0.048	-0.039	-0.006
## CAD		-0.034	-0.033	0.027	-0.071	-0.019
## HF		0.004	0.000	-0.121	-0.039	-0.062
## CVA		0.016	-0.044	-0.132	-0.071	-0.081
## CKD		-0.009	0.003	-0.141	-0.074	0.025
## LogCEA		0.021	0.079	-0.073	-0.040	0.013
## Laparoscopic		0.033	0.049	0.015	-0.017	-0.027
## TumorLOC		0.048	-0.052	-0.064	0.007	-0.156
## EA		-0.064	-0.077	0.044	0.056	-0.094
## Log2AT		-0.029	-0.007	0.078	0.050	0.163
## RBC		-0.017	-0.009	-0.116	-0.010	0.058
## Liver_Only		-0.116	-0.061	0.034	-0.023	0.069
## Cell_diff		0.208	0.107	-0.043	0.026	0.007
## Mucin_TYPE		0.020	0.015	0.056	-0.032	-0.032
## SignetRING		0.168	0.124	0.021	-0.041	-0.049
## Lymphovascularinvasion		1.000	0.355	0.025	0.018	-0.070
## perineural		0.355	1.000	0.001	-0.034	0.020
## CT		0.025	0.001	1.000	0.094	0.007
## RT		0.018	-0.034	0.094	1.000	-0.013
## NACTRT		-0.070	0.020	0.007	-0.013	1.000
## Death		0.062	0.038	0.041	0.077	-0.026
## Progress		0.088	0.058	0.393	0.134	0.065
## IntervalOD		-0.203	-0.143	0.224	0.047	-0.061
## IntervalOR		-0.140	-0.127	0.075	-0.052	-0.070
## AJCC_bin		0.197	0.145	-0.040	-0.009	-0.049
##		Death	Progress	IntervalOD	IntervalOR	AJCC_bin
## Age	-0.027	-0.134	-0.138	-0.040	0.031	
## Gender	0.011	-0.008	0.004	0.002	0.053	
## ASA3	0.058	-0.061	-0.146	-0.084	0.032	

```

## DM          0.019  -0.047   0.003   0.031   0.026
## CAD         -0.022  -0.008   0.031   0.014  -0.047
## HF          0.023  -0.022  -0.055  -0.052   0.024
## CVA         0.016  -0.045   0.009  -0.012  -0.031
## CKD         0.053  -0.042  -0.024  -0.032   0.015
## LogCEA      0.150   0.027  -0.237  -0.174   0.096
## Laparoscopic -0.018  -0.004  -0.028  -0.006  -0.051
## TumorLOC    0.022  -0.096  -0.040   0.025   0.038
## EA         -0.028   0.011   0.104   0.049  -0.021
## Log2AT      0.014   0.034   0.081   0.081  -0.023
## RBC         0.117  -0.050  -0.060  -0.003   0.048
## Liver_Only  -0.065  -0.058   0.118   0.113  -0.680
## Cell_diff   0.038  -0.021  -0.136  -0.042   0.151
## Mucin_TYPE  -0.034  -0.006  -0.031   0.016   0.134
## SignetRING  0.018  -0.032  -0.048   0.014   0.147
## Lymphovascularinvasion 0.062   0.088  -0.203  -0.140   0.197
## perineural  0.038   0.058  -0.143  -0.127   0.145
## CT          0.041   0.393   0.224   0.075  -0.040
## RT          0.077   0.134   0.047  -0.052  -0.009
## NACTRT      -0.026   0.065  -0.061  -0.070  -0.049
## Death       1.000   0.215  -0.202  -0.208   0.156
## Progress    0.215   1.000   0.054  -0.323   0.083
## IntervalOD  -0.202   0.054   1.000   0.651  -0.285
## IntervalOR  -0.208  -0.323   0.651   1.000  -0.213
## AJCC_bin    0.156   0.083  -0.285  -0.213   1.000
##
## ===== Metodo: spearman =====
##
##           Age Gender  ASA3   DM   CAD   HF   CVA   CKD
## Age       1.000 -0.105  0.400  0.129  0.168  0.160  0.157  0.171
## Gender    -0.105  1.000 -0.070 -0.020 -0.050 -0.024 -0.088 -0.084
## ASA3      0.400 -0.070  1.000  0.156  0.192  0.187  0.193  0.170
## DM        0.129 -0.020  0.156  1.000  0.131  0.048  0.057  0.050
## CAD       0.168 -0.050  0.192  0.131  1.000  0.204  0.055  0.059
## HF        0.160 -0.024  0.187  0.048  0.204  1.000  0.016  0.088
## CVA       0.157 -0.088  0.193  0.057  0.055  0.016  1.000  0.136
## CKD       0.171 -0.084  0.170  0.050  0.059  0.088  0.136  1.000
## LogCEA    0.017  0.071  0.076  0.039 -0.029 -0.032 -0.024  0.012
## Laparoscopic 0.023  0.022 -0.034  0.031 -0.016 -0.017 -0.027  0.001
## TumorLOC  0.116  0.108  0.099  0.032  0.028  0.032  0.048  0.002
## EA       -0.015 -0.051 -0.017 -0.032 -0.002  0.020 -0.009 -0.063
## Log2AT    -0.109 -0.062 -0.042 -0.025 -0.019 -0.046  0.007  0.027
## RBC       0.115  0.043  0.193  0.061  0.023  0.051  0.031  0.189
## Liver_Only -0.024 -0.100 -0.034 -0.058  0.030 -0.061 -0.006 -0.026
## Cell_diff -0.036 -0.070 -0.034 -0.003 -0.009 -0.002 -0.039 -0.007
## Mucin_TYPE -0.067 -0.019 -0.008 -0.034 -0.018 -0.021 -0.018  0.021
## SignetRING -0.062 -0.050 -0.048 -0.046  0.019  0.005  0.014 -0.008
## Lymphovascularinvasion -0.048 -0.001 -0.003 -0.040 -0.034  0.004  0.016 -0.009
## perineural -0.067 -0.055  0.024 -0.019 -0.033  0.000 -0.044  0.003
## CT       -0.232  0.050 -0.183 -0.048  0.027 -0.121 -0.132 -0.141
## RT       -0.119  0.027 -0.109 -0.039 -0.071 -0.039 -0.071 -0.074
## NACTRT    -0.147 -0.033 -0.086 -0.006 -0.019 -0.062 -0.081  0.025

```


## Death	-0.025	0.011	0.058	0.019	-0.022	0.023	0.016	0.053
## Progress	-0.132	-0.008	-0.061	-0.047	-0.008	-0.022	-0.045	-0.042
## IntervalOD	-0.187	0.020	-0.173	0.018	0.051	-0.041	-0.032	-0.035
## IntervalOR	-0.043	-0.012	-0.093	0.036	0.023	-0.033	-0.054	-0.044
## AJCC_bin	0.034	0.053	0.032	0.026	-0.047	0.024	-0.031	0.015
##	LogCEA	Laparoscopic	TumorLOC	EA	Log2AT	RBC		
## Age	0.017	0.023	0.116	-0.015	-0.109	0.115		
## Gender	0.071	0.022	0.108	-0.051	-0.062	0.043		
## ASA3	0.076	-0.034	0.099	-0.017	-0.042	0.193		
## DM	0.039	0.031	0.032	-0.032	-0.025	0.061		
## CAD	-0.029	-0.016	0.028	-0.002	-0.019	0.023		
## HF	-0.032	-0.017	0.032	0.020	-0.046	0.051		
## CVA	-0.024	-0.027	0.048	-0.009	0.007	0.031		
## CKD	0.012	0.001	0.002	-0.063	0.027	0.189		
## LogCEA	1.000	0.016	-0.055	-0.022	-0.098	0.085		
## Laparoscopic	0.016	1.000	0.008	-0.016	-0.016	-0.016		
## TumorLOC	-0.055	0.008	1.000	-0.021	-0.142	0.127		
## EA	-0.022	-0.016	-0.021	1.000	-0.031	-0.030		
## Log2AT	-0.098	-0.016	-0.142	-0.031	1.000	0.209		
## RBC	0.085	-0.016	0.127	-0.030	0.209	1.000		
## Liver_Only	0.080	0.075	-0.061	0.043	0.053	0.031		
## Cell_diff	-0.126	0.005	0.079	-0.011	0.005	0.029		
## Mucin_TYPE	-0.059	0.004	0.070	-0.038	-0.016	0.009		
## SignetRING	-0.063	0.063	0.108	-0.025	-0.056	0.023		
## Lymphovascularinvasion	0.010	0.033	0.048	-0.064	0.000	-0.016		
## perineural	0.065	0.049	-0.052	-0.077	0.000	0.000		
## CT	-0.065	0.015	-0.064	0.044	0.065	-0.117		
## RT	-0.041	-0.017	0.007	0.056	0.043	-0.017		
## NACTRT	0.012	-0.027	-0.156	-0.094	0.164	0.051		
## Death	0.138	-0.018	0.022	-0.028	0.030	0.112		
## Progress	0.030	-0.004	-0.096	0.011	0.044	-0.054		
## IntervalOD	-0.229	-0.020	-0.083	0.023	0.096	-0.104		
## IntervalOR	-0.189	0.006	-0.050	0.009	0.079	-0.085		
## AJCC_bin	0.094	-0.051	0.038	-0.021	-0.003	0.045		
##	Liver_Only	Cell_diff	Mucin_TYPE	SignetRING				
## Age	-0.024	-0.036	-0.067	-0.062				
## Gender	-0.100	-0.070	-0.019	-0.050				
## ASA3	-0.034	-0.034	-0.008	-0.048				
## DM	-0.058	-0.003	-0.034	-0.046				
## CAD	0.030	-0.009	-0.018	0.019				
## HF	-0.061	-0.002	-0.021	0.005				
## CVA	-0.006	-0.039	-0.018	0.014				
## CKD	-0.026	-0.007	0.021	-0.008				
## LogCEA	0.080	-0.126	-0.059	-0.063				
## Laparoscopic	0.075	0.005	0.004	0.063				
## TumorLOC	-0.061	0.079	0.070	0.108				
## EA	0.043	-0.011	-0.038	-0.025				
## Log2AT	0.053	0.005	-0.016	-0.056				
## RBC	0.031	0.029	0.009	0.023				
## Liver_Only	1.000	-0.118	-0.135	-0.119				
## Cell_diff	-0.118	1.000	0.101	0.281				

## Mucin_TYPE	-0.135	0.101	1.000	0.298		
## SignetRING	-0.119	0.281	0.298	1.000		
## Lymphovascularinvasion	-0.116	0.208	0.020	0.168		
## perineural	-0.061	0.107	0.015	0.124		
## CT	0.034	-0.043	0.056	0.021		
## RT	-0.023	0.026	-0.032	-0.041		
## NACTRT	0.069	0.007	-0.032	-0.049		
## Death	-0.065	0.038	-0.034	0.018		
## Progress	-0.058	-0.021	-0.006	-0.032		
## IntervalOD	0.104	-0.182	-0.016	-0.055		
## IntervalOR	0.093	-0.090	0.033	-0.016		
## AJCC_bin	-0.680	0.151	0.134	0.147		
##		Lymphovascularinvasion	perineural	CT	RT	NACTRT
## Age		-0.048	-0.067	-0.232	-0.119	-0.147
## Gender		-0.001	-0.055	0.050	0.027	-0.033
## ASA3		-0.003	0.024	-0.183	-0.109	-0.086
## DM		-0.040	-0.019	-0.048	-0.039	-0.006
## CAD		-0.034	-0.033	0.027	-0.071	-0.019
## HF		0.004	0.000	-0.121	-0.039	-0.062
## CVA		0.016	-0.044	-0.132	-0.071	-0.081
## CKD		-0.009	0.003	-0.141	-0.074	0.025
## LogCEA		0.010	0.065	-0.065	-0.041	0.012
## Laparoscopic		0.033	0.049	0.015	-0.017	-0.027
## TumorLOC		0.048	-0.052	-0.064	0.007	-0.156
## EA		-0.064	-0.077	0.044	0.056	-0.094
## Log2AT		0.000	0.000	0.065	0.043	0.164
## RBC		-0.016	0.000	-0.117	-0.017	0.051
## Liver_Only		-0.116	-0.061	0.034	-0.023	0.069
## Cell_diff		0.208	0.107	-0.043	0.026	0.007
## Mucin_TYPE		0.020	0.015	0.056	-0.032	-0.032
## SignetRING		0.168	0.124	0.021	-0.041	-0.049
## Lymphovascularinvasion		1.000	0.355	0.025	0.018	-0.070
## perineural		0.355	1.000	0.001	-0.034	0.020
## CT		0.025	0.001	1.000	0.094	0.007
## RT		0.018	-0.034	0.094	1.000	-0.013
## NACTRT		-0.070	0.020	0.007	-0.013	1.000
## Death		0.062	0.038	0.041	0.077	-0.026
## Progress		0.088	0.058	0.393	0.134	0.065
## IntervalOD		-0.188	-0.129	0.348	0.092	-0.010
## IntervalOR		-0.118	-0.120	0.232	0.016	-0.041
## AJCC_bin		0.197	0.145	-0.040	-0.009	-0.049
##		Death	Progress	IntervalOD	IntervalOR	AJCC_bin
## Age	-0.025	-0.132	-0.187	-0.043	0.034	
## Gender	0.011	-0.008	0.020	-0.012	0.053	
## ASA3	0.058	-0.061	-0.173	-0.093	0.032	
## DM	0.019	-0.047	0.018	0.036	0.026	
## CAD	-0.022	-0.008	0.051	0.023	-0.047	
## HF	0.023	-0.022	-0.041	-0.033	0.024	
## CVA	0.016	-0.045	-0.032	-0.054	-0.031	
## CKD	0.053	-0.042	-0.035	-0.044	0.015	
## LogCEA	0.138	0.030	-0.229	-0.189	0.094	

```

## Laparoscopic      -0.018  -0.004  -0.020    0.006  -0.051
## TumorLOC          0.022  -0.096  -0.083   -0.050   0.038
## EA                -0.028   0.011   0.023    0.009  -0.021
## Log2AT             0.030   0.044   0.096    0.079  -0.003
## RBC                0.112  -0.054  -0.104   -0.085   0.045
## Liver_Only        -0.065  -0.058   0.104    0.093  -0.680
## Cell_diff          0.038  -0.021  -0.182   -0.090   0.151
## Mucin_TYPE        -0.034  -0.006  -0.016    0.033   0.134
## SignetRING         0.018  -0.032  -0.055   -0.016   0.147
## Lymphovascularinvasion 0.062   0.088  -0.188   -0.118   0.197
## perineural         0.038   0.058  -0.129   -0.120   0.145
## CT                 0.041   0.393   0.348    0.232  -0.040
## RT                 0.077   0.134   0.092    0.016  -0.009
## NACTRT            -0.026   0.065  -0.010   -0.041  -0.049
## Death              1.000   0.215  -0.137   -0.165   0.156
## Progress           0.215   1.000   0.235   -0.019   0.083
## IntervalOD        -0.137   0.235   1.000    0.640  -0.275
## IntervalOR        -0.165  -0.019   0.640    1.000  -0.228
## AJCC_bin           0.156   0.083  -0.275   -0.228   1.000
##
## ===== Metodo: kendall =====
##
## Age Gender  ASA3  DM  CAD  HF  CVA  CKD
## Age      1.000 -0.087  0.330  0.106  0.138  0.132  0.130  0.141
## Gender   -0.087  1.000 -0.070 -0.020 -0.050 -0.024 -0.088 -0.084
## ASA3      0.330 -0.070  1.000  0.156  0.192  0.187  0.193  0.170
## DM        0.106 -0.020  0.156  1.000  0.131  0.048  0.057  0.050
## CAD       0.138 -0.050  0.192  0.131  1.000  0.204  0.055  0.059
## HF        0.132 -0.024  0.187  0.048  0.204  1.000  0.016  0.088
## CVA       0.130 -0.088  0.193  0.057  0.055  0.016  1.000  0.136
## CKD       0.141 -0.084  0.170  0.050  0.059  0.088  0.136  1.000
## LogCEA    0.012  0.058  0.062  0.032 -0.024 -0.026 -0.020  0.010
## Laparoscopic 0.019  0.022 -0.034  0.031 -0.016 -0.017 -0.027  0.001
## TumorLOC   0.095  0.108  0.099  0.032  0.028  0.032  0.048  0.002
## EA        -0.012 -0.051 -0.017 -0.032 -0.002  0.020 -0.009 -0.063
## Log2AT     -0.075 -0.051 -0.035 -0.020 -0.016 -0.038  0.006  0.023
## RBC        0.091  0.042  0.186  0.059  0.022  0.049  0.030  0.182
## Liver_Only -0.020 -0.100 -0.034 -0.058  0.030 -0.061 -0.006 -0.026
## Cell_diff  -0.030 -0.070 -0.034 -0.003 -0.009 -0.002 -0.039 -0.007
## Mucin_TYPE -0.055 -0.019 -0.008 -0.034 -0.018 -0.021 -0.018  0.021
## SignetRING -0.051 -0.050 -0.048 -0.046  0.019  0.005  0.014 -0.008
## Lymphovascularinvasion -0.040 -0.001 -0.003 -0.040 -0.034  0.004  0.016 -0.009
## perineural -0.055 -0.055  0.024 -0.019 -0.033  0.000 -0.044  0.003
## CT        -0.192  0.050 -0.183 -0.048  0.027 -0.121 -0.132 -0.141
## RT        -0.099  0.027 -0.109 -0.039 -0.071 -0.039 -0.071 -0.074
## NACTRT     -0.122 -0.033 -0.086 -0.006 -0.019 -0.062 -0.081  0.025
## Death     -0.021  0.011  0.058  0.019 -0.022  0.023  0.016  0.053
## Progress  -0.109 -0.008 -0.061 -0.047 -0.008 -0.022 -0.045 -0.042
## IntervalOD -0.125  0.017 -0.141  0.015  0.041 -0.033 -0.027 -0.028
## IntervalOR -0.029 -0.010 -0.076  0.029  0.019 -0.027 -0.044 -0.036
## AJCC_bin   0.028  0.053  0.032  0.026 -0.047  0.024 -0.031  0.015
##
## LogCEA Laparoscopic TumorLOC EA Log2AT RBC

```

## Age	0.012	0.019	0.095	-0.012	-0.075	0.091
## Gender	0.058	0.022	0.108	-0.051	-0.051	0.042
## ASA3	0.062	-0.034	0.099	-0.017	-0.035	0.186
## DM	0.032	0.031	0.032	-0.032	-0.020	0.059
## CAD	-0.024	-0.016	0.028	-0.002	-0.016	0.022
## HF	-0.026	-0.017	0.032	0.020	-0.038	0.049
## CVA	-0.020	-0.027	0.048	-0.009	0.006	0.030
## CKD	0.010	0.001	0.002	-0.063	0.023	0.182
## LogCEA	1.000	0.013	-0.045	-0.018	-0.065	0.068
## Laparoscopic	0.013	1.000	0.008	-0.016	-0.013	-0.015
## TumorLOC	-0.045	0.008	1.000	-0.021	-0.118	0.123
## EA	-0.018	-0.016	-0.021	1.000	-0.025	-0.029
## Log2AT	-0.065	-0.013	-0.118	-0.025	1.000	0.168
## RBC	0.068	-0.015	0.123	-0.029	0.168	1.000
## Liver_Only	0.066	0.075	-0.061	0.043	0.044	0.030
## Cell_diff	-0.103	0.005	0.079	-0.011	0.004	0.028
## Mucin_TYPE	-0.048	0.004	0.070	-0.038	-0.013	0.008
## SignetRING	-0.051	0.063	0.108	-0.025	-0.047	0.022
## Lymphovascularinvasion	0.008	0.033	0.048	-0.064	0.000	-0.015
## perineural	0.053	0.049	-0.052	-0.077	0.000	0.000
## CT	-0.053	0.015	-0.064	0.044	0.054	-0.113
## RT	-0.033	-0.017	0.007	0.056	0.036	-0.016
## NACTRT	0.010	-0.027	-0.156	-0.094	0.136	0.050
## Death	0.113	-0.018	0.022	-0.028	0.025	0.108
## Progress	0.024	-0.004	-0.096	0.011	0.037	-0.052
## IntervalOD	-0.155	-0.017	-0.068	0.018	0.066	-0.082
## IntervalOR	-0.128	0.005	-0.041	0.007	0.054	-0.067
## AJCC_bin	0.077	-0.051	0.038	-0.021	-0.003	0.043
##	Liver_Only	Cell_diff	Mucin_TYPE	SignetRING		
## Age	-0.020	-0.030	-0.055	-0.051		
## Gender	-0.100	-0.070	-0.019	-0.050		
## ASA3	-0.034	-0.034	-0.008	-0.048		
## DM	-0.058	-0.003	-0.034	-0.046		
## CAD	0.030	-0.009	-0.018	0.019		
## HF	-0.061	-0.002	-0.021	0.005		
## CVA	-0.006	-0.039	-0.018	0.014		
## CKD	-0.026	-0.007	0.021	-0.008		
## LogCEA	0.066	-0.103	-0.048	-0.051		
## Laparoscopic	0.075	0.005	0.004	0.063		
## TumorLOC	-0.061	0.079	0.070	0.108		
## EA	0.043	-0.011	-0.038	-0.025		
## Log2AT	0.044	0.004	-0.013	-0.047		
## RBC	0.030	0.028	0.008	0.022		
## Liver_Only	1.000	-0.118	-0.135	-0.119		
## Cell_diff	-0.118	1.000	0.101	0.281		
## Mucin_TYPE	-0.135	0.101	1.000	0.298		
## SignetRING	-0.119	0.281	0.298	1.000		
## Lymphovascularinvasion	-0.116	0.208	0.020	0.168		
## perineural	-0.061	0.107	0.015	0.124		
## CT	0.034	-0.043	0.056	0.021		
## RT	-0.023	0.026	-0.032	-0.041		

## NACTRT	0.069	0.007	-0.032	-0.049	
## Death	-0.065	0.038	-0.034	0.018	
## Progress	-0.058	-0.021	-0.006	-0.032	
## IntervalOD	0.085	-0.148	-0.013	-0.045	
## IntervalOR	0.076	-0.073	0.027	-0.013	
## AJCC_bin	-0.680	0.151	0.134	0.147	
##	Lymphovascularinvasion	perineural	CT	RT	NACTRT
## Age		-0.040	-0.055	-0.192	-0.099
## Gender		-0.001	-0.055	0.050	0.027
## ASA3		-0.003	0.024	-0.183	-0.109
## DM		-0.040	-0.019	-0.048	-0.039
## CAD		-0.034	-0.033	0.027	-0.071
## HF		0.004	0.000	-0.121	-0.039
## CVA		0.016	-0.044	-0.132	-0.071
## CKD		-0.009	0.003	-0.141	-0.074
## LogCEA		0.008	0.053	-0.053	-0.033
## Laparoscopic		0.033	0.049	0.015	-0.017
## TumorLOC		0.048	-0.052	-0.064	0.007
## EA		-0.064	-0.077	0.044	0.056
## Log2AT		0.000	0.000	0.054	0.036
## RBC		-0.015	0.000	-0.113	-0.016
## Liver_Only		-0.116	-0.061	0.034	-0.023
## Cell_diff		0.208	0.107	-0.043	0.026
## Mucin_TYPE		0.020	0.015	0.056	-0.032
## SignetRING		0.168	0.124	0.021	-0.041
## Lymphovascularinvasion		1.000	0.355	0.025	0.018
## perineural		0.355	1.000	0.001	-0.034
## CT		0.025	0.001	1.000	0.094
## RT		0.018	-0.034	0.094	1.000
## NACTRT		-0.070	0.020	0.007	-0.013
## Death		0.062	0.038	0.041	0.077
## Progress		0.088	0.058	0.393	0.134
## IntervalOD		-0.153	-0.105	0.284	0.075
## IntervalOR		-0.096	-0.098	0.189	0.013
## AJCC_bin		0.197	0.145	-0.040	-0.009
##	Death	Progress	IntervalOD	IntervalOR	AJCC_bin
## Age	-0.021	-0.109	-0.125	-0.029	0.028
## Gender	0.011	-0.008	0.017	-0.010	0.053
## ASA3	0.058	-0.061	-0.141	-0.076	0.032
## DM	0.019	-0.047	0.015	0.029	0.026
## CAD	-0.022	-0.008	0.041	0.019	-0.047
## HF	0.023	-0.022	-0.033	-0.027	0.024
## CVA	0.016	-0.045	-0.027	-0.044	-0.031
## CKD	0.053	-0.042	-0.028	-0.036	0.015
## LogCEA	0.113	0.024	-0.155	-0.128	0.077
## Laparoscopic	-0.018	-0.004	-0.017	0.005	-0.051
## TumorLOC	0.022	-0.096	-0.068	-0.041	0.038
## EA	-0.028	0.011	0.018	0.007	-0.021
## Log2AT	0.025	0.037	0.066	0.054	-0.003
## RBC	0.108	-0.052	-0.082	-0.067	0.043
## Liver_Only	-0.065	-0.058	0.085	0.076	-0.680

## Cell_diff	0.038	-0.021	-0.148	-0.073	0.151
## Mucin_TYPE	-0.034	-0.006	-0.013	0.027	0.134
## SignetRING	0.018	-0.032	-0.045	-0.013	0.147
## Lymphovascularinvasion	0.062	0.088	-0.153	-0.096	0.197
## perineural	0.038	0.058	-0.105	-0.098	0.145
## CT	0.041	0.393	0.284	0.189	-0.040
## RT	0.077	0.134	0.075	0.013	-0.009
## NACTRT	-0.026	0.065	-0.008	-0.033	-0.049
## Death	1.000	0.215	-0.112	-0.135	0.156
## Progress	0.215	1.000	0.192	-0.015	0.083
## IntervalOD	-0.112	0.192	1.000	0.487	-0.225
## IntervalOR	-0.135	-0.015	0.487	1.000	-0.186
## AJCC_bin	0.156	0.083	-0.225	-0.186	1.000

A livello interpretativo, PEARSON - RELAZIONE LINEARE CONTINUA:

1- ETÁ: età correlata positivamente con HF (più anzianità = più infartuati) e ASA3.
Negativamente con chemioterapia 2- ASA3: correlato positivamente ma moderatamente con tutte le comorbidità: è un buon proxy dello stato comorbito dei pazienti?

PEARSON: MONOTONA

Age – ASA3: $\rho = 0.480 \rightarrow$ Età maggiore si associa a status ASA3 (logico: pazienti più anziani tendono ad avere comorbidità).

Age – Interval: $\rho = -0.345 \rightarrow$ Età maggiore correlata a un tempo inferiore tra diagnosi e trattamento? Potrebbe indicare priorità clinica.

Interval – ASA3: $\rho = -0.270 \rightarrow$ Anche qui, pazienti più gravi (ASA3) ricevono cure più tempestive.

CKD – CVA: $\rho = 0.264 \rightarrow$ Insufficienza renale cronica associata a pregressi eventi cerebrovascolari (coerenza clinica).

ASA3 – CAD/HF/CVA/CKD: tutte $>0.2 \rightarrow$ conferma che ASA3 sintetizza diverse comorbidità.

MOLTO INTERESSANTE: PROGRESS E TUMORLOC $\rho = -0.221$: sedi tumorali hanno minore tendenza alla progressione!!!!

Age vs ASA3: $\rho = 0.397 \rightarrow$ correlazione moderata e positiva. Conferma che l'età avanzata tende ad associarsi a uno stato fisico compromesso (ASA3 alto).

Age vs HF (scompenso cardiaco): $\rho = 0.176 \rightarrow$ relazione debole ma coerente con la letteratura.

Age vs Interval: $\rho = -0.236 \rightarrow$ età maggiore è debolmente associata a un intervallo più breve (probabile tra intervento e progressione).

CHEMIOTERAPIA invece è neg corr con ASA3 e età, cioè i pazienti giovani e meno fragili ricevono più spesso chemioterapia.

ASA3 È UN OTTIMO SURROGATO SINTETICO DELLE COMORBIDITÀ: STATO CLINICO

SE SPEARMAN È MAGGIORE DI PEARSON: POTREBBE ESSERCI UNA RELAZIONE MONOTONA NON LINEARE!!!! - AGE e INTERVAL - CKD e RBC

- HFe ASA3
- PROGRESS - TUMOR LOR

METODI PROIEZIONE - riduzione dimensionale spazi latenti

UMAP e t-SNE: sono solo metodi di proiezione (riduzione dimensionale). Non creano cluster di per sé, ma aiutano a visualizzare i dati.

I DATI VENGONO STANDARDIZZATI: per rendere comparabile ogni variabile, media 0 e sd 1. Lo richiedevano i metodi scelti - UMAP t-SNE e, dopo, PCA.

Viene effettuata la riduzione dimensionale per visualizzare dati ad alta dimensionalità in 2D.

t-SNE e UMAP sono tecniche non lineari (adatte a pattern complessi).

PCA è lineare, utile per interpretabilità e visualizzazione.

Non servono per clustering, ma per capire la distribuzione dei dati.

```
# install.packages(c("Rtsne", "umap"))

# Caricamento pacchetti
library(Rtsne)

## Warning: package 'Rtsne' was built under R version 4.3.3

library(umap)

## Warning: package 'umap' was built under R version 4.3.3

library(ggplot2)

# Standardizza i dati
data_scaled <- scale(data)

# t-SNE
set.seed(42)
tsne_out <- Rtsne(data_scaled, dims = 2, perplexity = 30, verbose = TRUE, max_iter = 500)

## Performing PCA
## Read the 916 x 28 data matrix successfully!
## OpenMP is working. 1 threads.
## Using no_dims = 2, perplexity = 30.000000, and theta = 0.500000
## Computing input similarities...
## Building tree...
## Done in 0.42 seconds (sparsity = 0.149802)!
## Learning embedding...
## Iteration 50: error is 68.038564 (50 iterations in 0.37 seconds)
## Iteration 100: error is 68.038521 (50 iterations in 1.00 seconds)
## Iteration 150: error is 68.038593 (50 iterations in 1.26 seconds)
## Iteration 200: error is 68.038593 (50 iterations in 1.33 seconds)
```

```
## Iteration 250: error is 68.038536 (50 iterations in 1.73 seconds)
## Iteration 300: error is 3.183366 (50 iterations in 1.18 seconds)
## Iteration 350: error is 1.441138 (50 iterations in 0.51 seconds)
## Iteration 400: error is 1.374193 (50 iterations in 0.49 seconds)
## Iteration 450: error is 1.341848 (50 iterations in 0.36 seconds)
## Iteration 500: error is 1.326467 (50 iterations in 0.41 seconds)
## Fitting performed in 8.63 seconds.
```

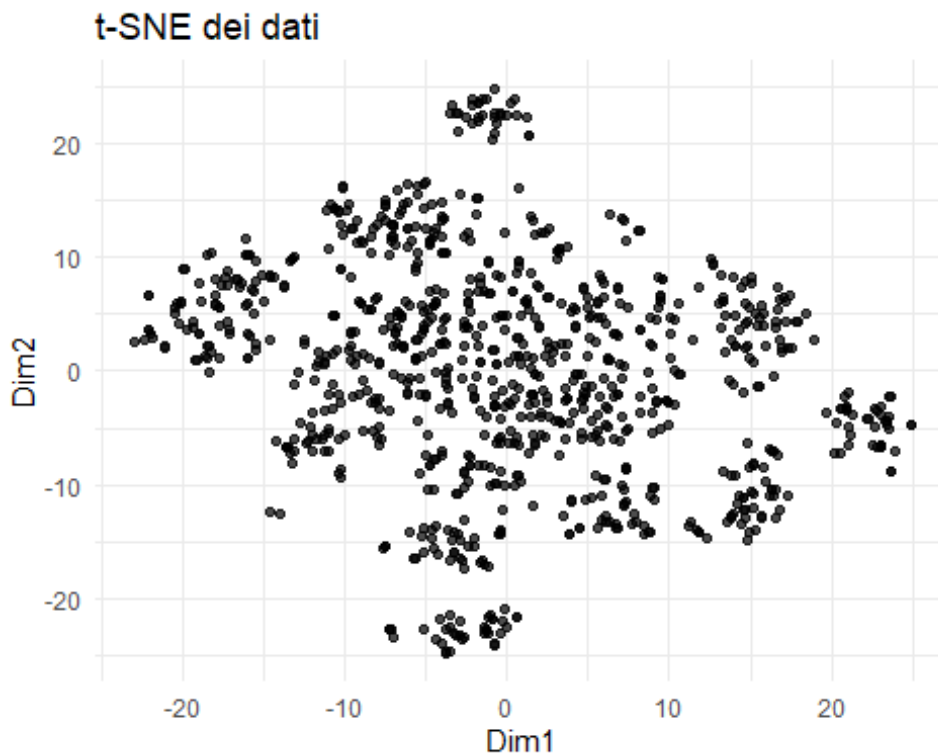
```
tsne_df <- data.frame(
  Dim1 = tsne_out$Y[,1],
  Dim2 = tsne_out$Y[,2]
)
```

```
# UMAP
```

```
umap_out <- umap(data_scaled)
umap_df <- data.frame(
  Dim1 = umap_out$layout[,1],
  Dim2 = umap_out$layout[,2]
)
```

```
# Grafico t-SNE
```

```
ggplot(tsne_df, aes(x = Dim1, y = Dim2)) +
  geom_point(alpha = 0.7) +
  labs(title = "t-SNE dei dati") +
  theme_minimal()
```

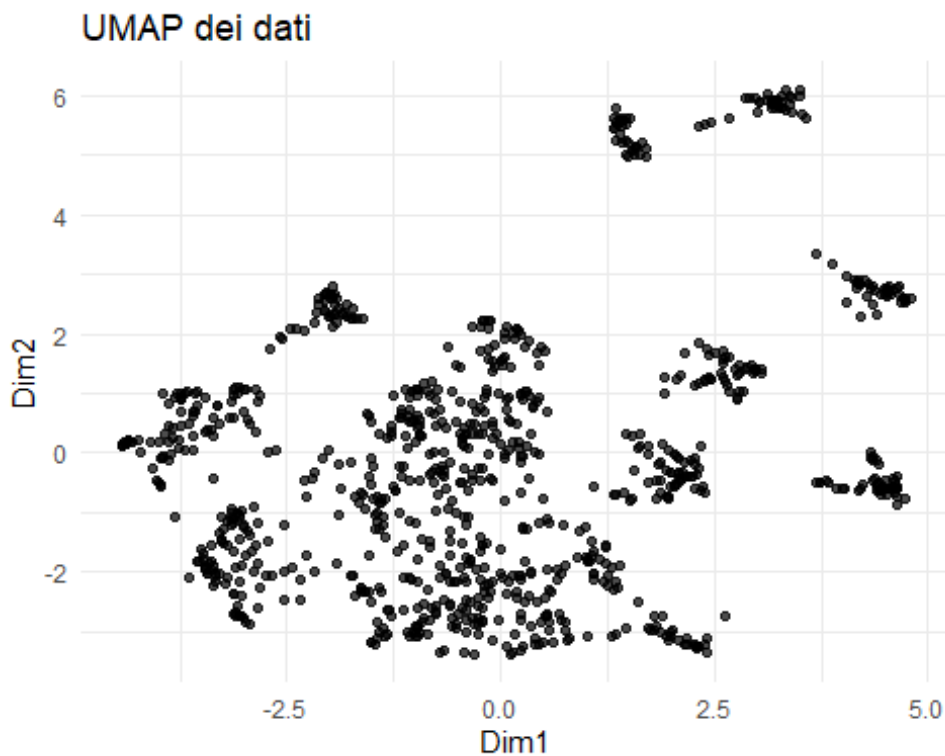


```
# Grafico UMAP
```

```
ggplot(umap_df, aes(x = Dim1, y = Dim2)) +
```



```
geom_point(alpha = 0.7) +
labs(title = "UMAP dei dati") +
theme_minimal()
```



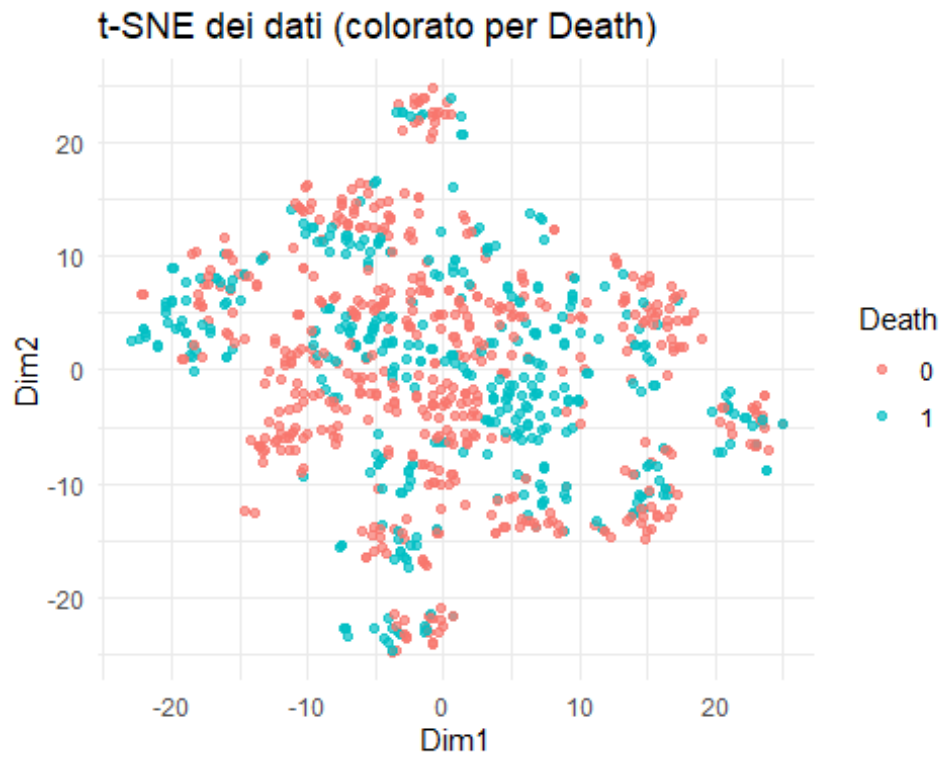
INTERPRETAZIONE

Serve per vedere se la variabile target separa i dati nella proiezione.

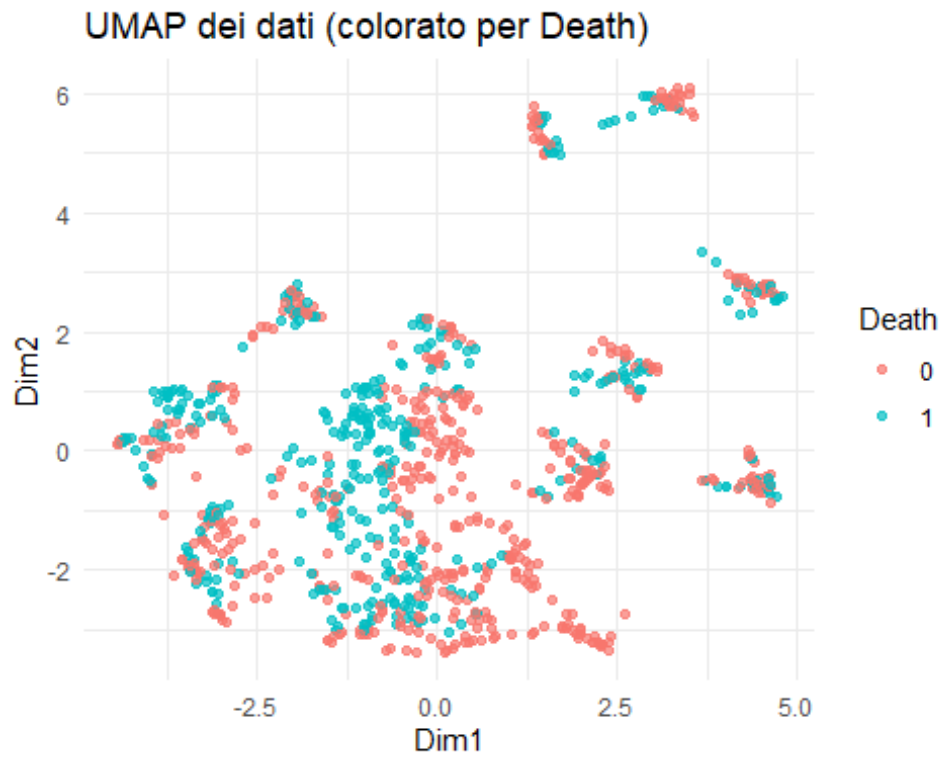
Se i punti colorati per “Death” o chi per lei si raggruppano, significa che quella variabile ha un effetto latente.

```
# Aggiungiamo la variabile target (assumendo sia "Death")
tsne_df$Death <- data$Death
umap_df$Death <- data$Death

# t-SNE con colorazione
ggplot(tsne_df, aes(x = Dim1, y = Dim2, color = as.factor(Death))) +
  geom_point(alpha = 0.7) +
  labs(title = "t-SNE dei dati (colorato per Death)", color = "Death") +
  theme_minimal()
```

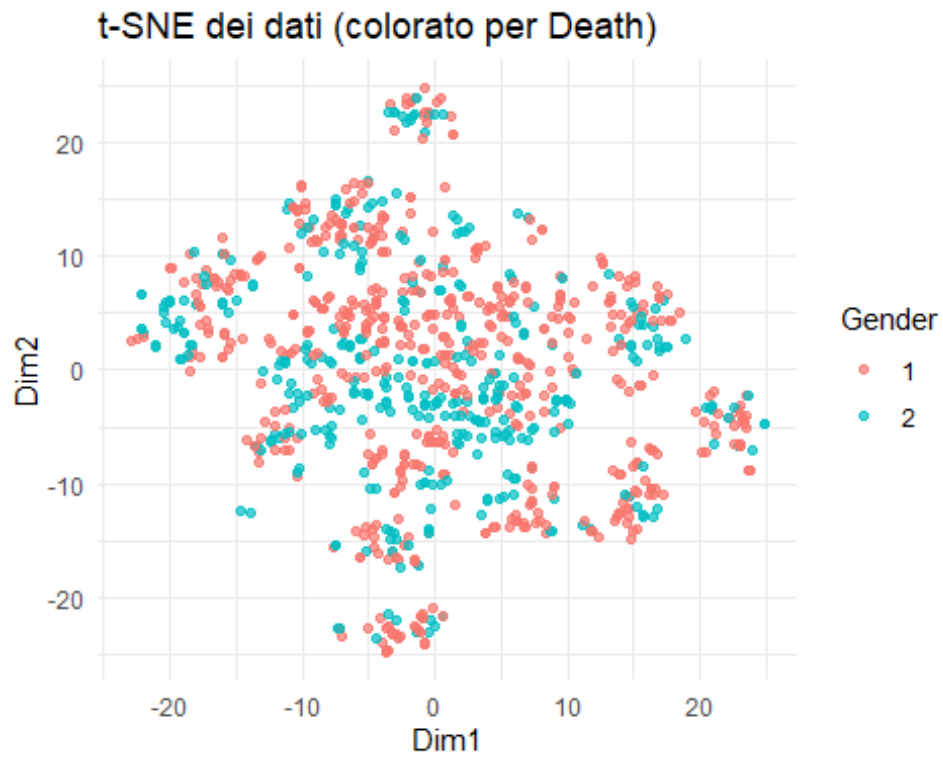


```
# UMAP con colorazione
ggplot(umap_df, aes(x = Dim1, y = Dim2, color = as.factor(Death))) +
  geom_point(alpha = 0.7) +
  labs(title = "UMAP dei dati (colorato per Death)", color = "Death") +
  theme_minimal()
```

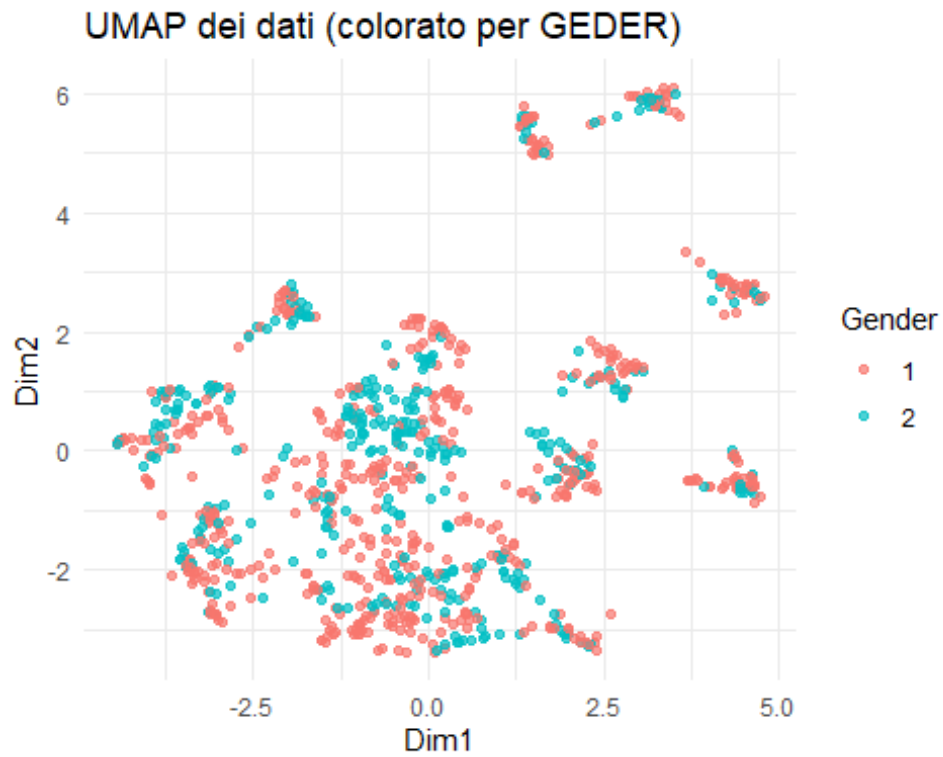


```
tsne_df$Gender <- data$Gender
umap_df$Gender <- data$Gender

# t-SNE con colorazione
ggplot(tsne_df, aes(x = Dim1, y = Dim2, color = as.factor(Gender))) +
  geom_point(alpha = 0.7) +
  labs(title = "t-SNE dei dati (colorato per Death)", color = "Gender") +
  theme_minimal()
```

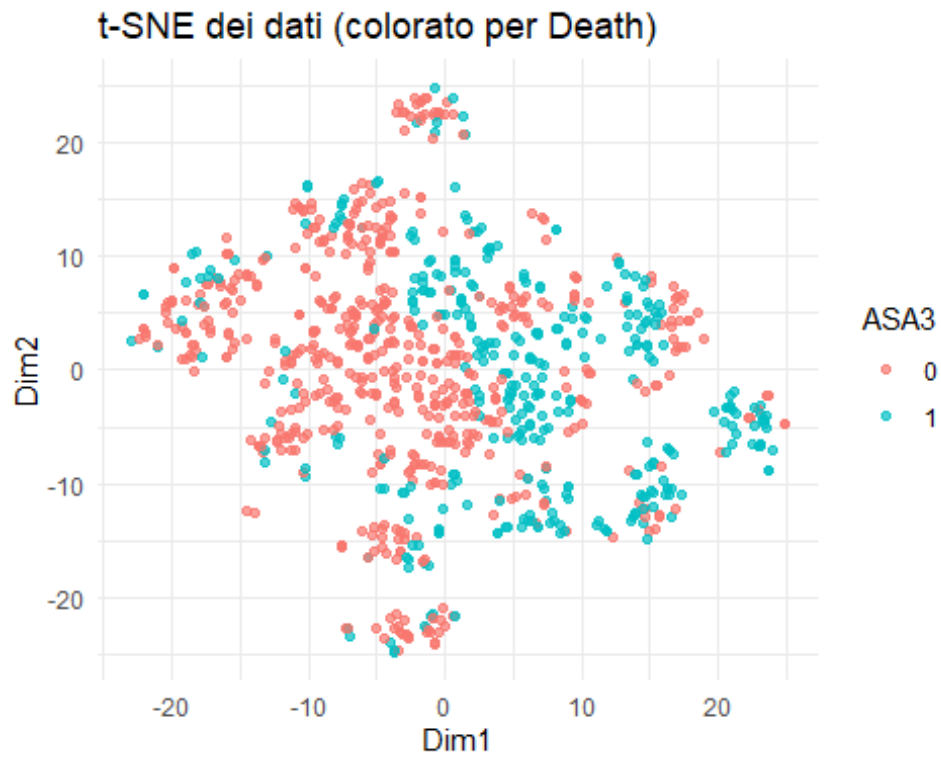


```
# UMAP con colorazione
ggplot(umap_df, aes(x = Dim1, y = Dim2, color = as.factor(Gender))) +
  geom_point(alpha = 0.7) +
  labs(title = "UMAP dei dati (colorato per GEDER)", color = "Gender") +
  theme_minimal()
```

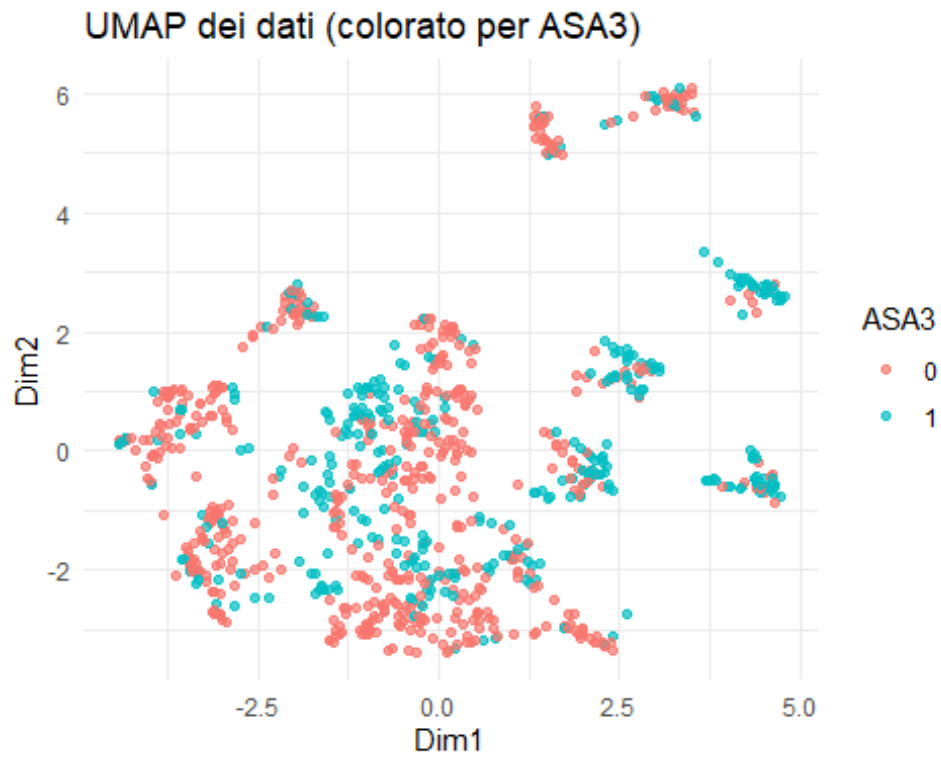


```
# Aggiungiamo la variabile target (assumendo sia "Death")
tsne_df$ASA3 <- data$ASA3
umap_df$ASA3 <- data$ASA3

# t-SNE con colorazione
ggplot(tsne_df, aes(x = Dim1, y = Dim2, color = as.factor(ASA3))) +
  geom_point(alpha = 0.7) +
  labs(title = "t-SNE dei dati (colorato per Death)", color = "ASA3") +
  theme_minimal()
```

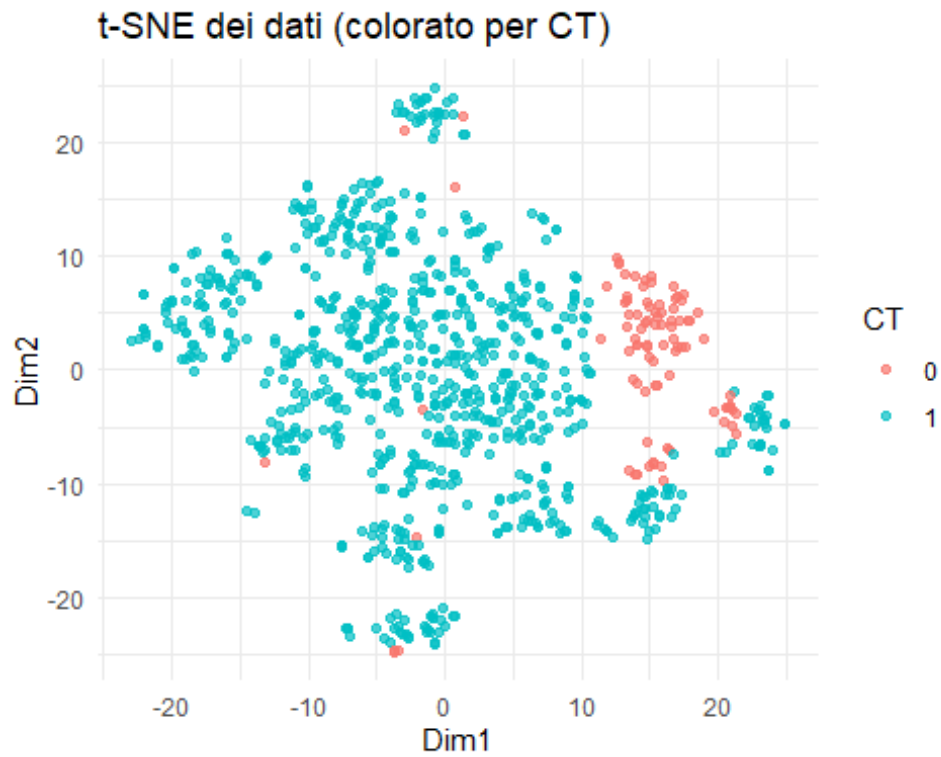


```
# UMAP con colorazione
ggplot(umap_df, aes(x = Dim1, y = Dim2, color = as.factor(ASA3))) +
  geom_point(alpha = 0.7) +
  labs(title = "UMAP dei dati (colorato per ASA3)", color = "ASA3") +
  theme_minimal()
```

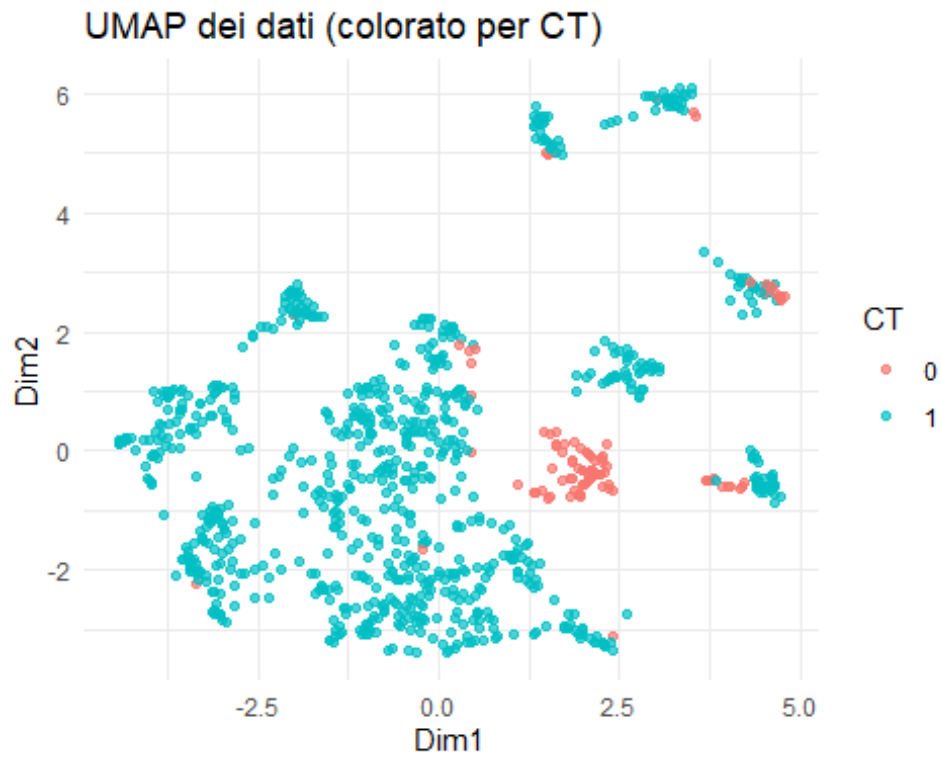


```
tsne_df$CT <- data$CT
umap_df$CT <- data$CT

# t-SNE con colorazione
ggplot(tsne_df, aes(x = Dim1, y = Dim2, color = as.factor(CT))) +
  geom_point(alpha = 0.7) +
  labs(title = "t-SNE dei dati (colorato per CT)", color = "CT") +
  theme_minimal()
```

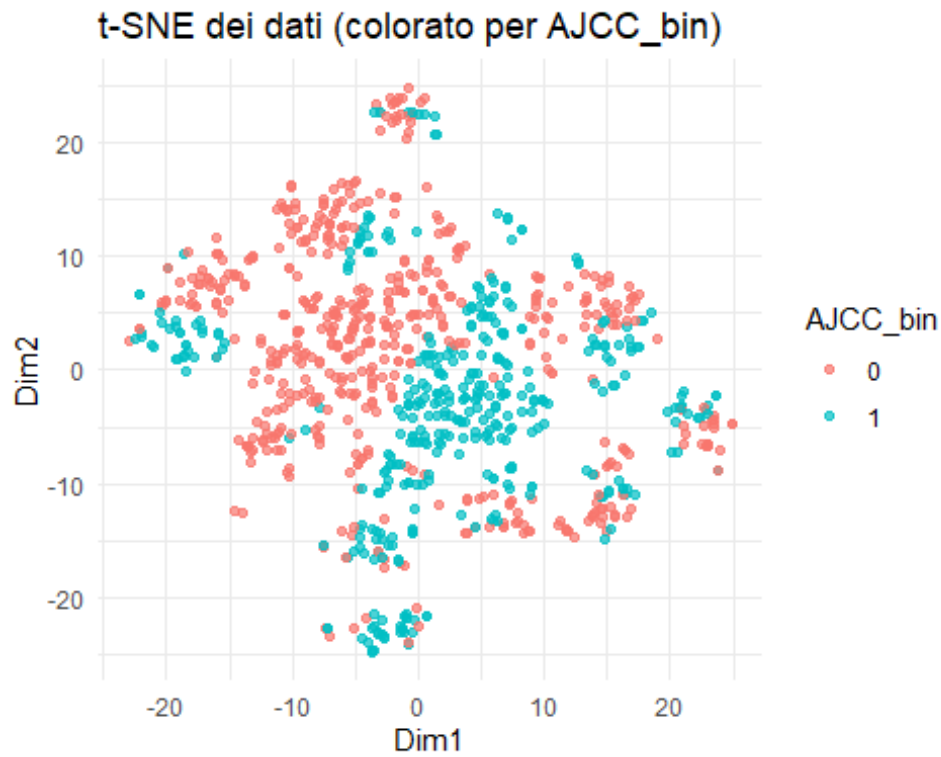


```
# UMAP con colorazione
ggplot(umap_df, aes(x = Dim1, y = Dim2, color = as.factor(CT))) +
  geom_point(alpha = 0.7) +
  labs(title = "UMAP dei dati (colorato per CT)", color = "CT") +
  theme_minimal()
```

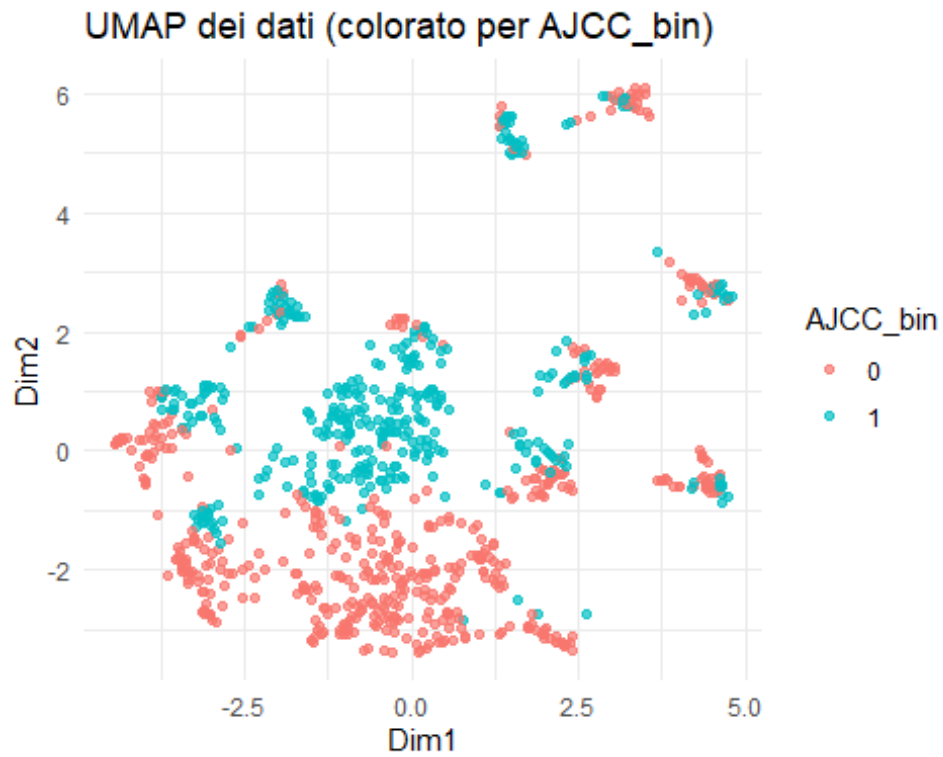



```
tsne_df$AJCC_bin <- data$AJCC_bin
umap_df$AJCC_bin <- data$AJCC_bin

# t-SNE con colorazione
ggplot(tsne_df, aes(x = Dim1, y = Dim2, color = as.factor(AJCC_bin))) +
  geom_point(alpha = 0.7) +
  labs(title = "t-SNE dei dati (colorato per AJCC_bin)", color = "AJCC_bin") +
  theme_minimal()
```

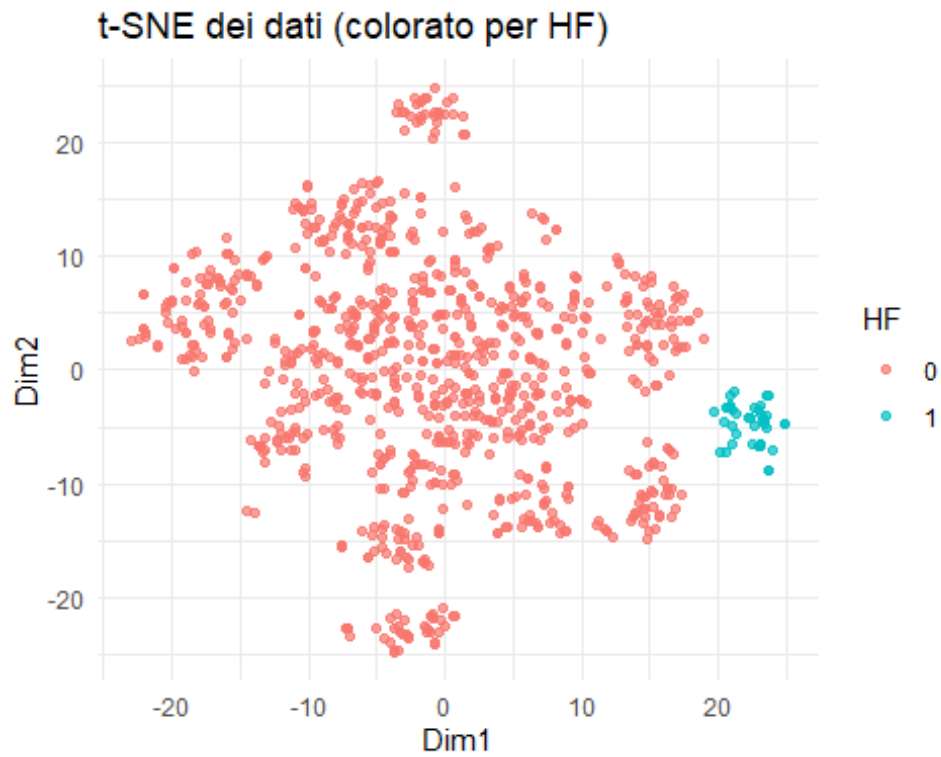


```
# UMAP con colorazione
ggplot(umap_df, aes(x = Dim1, y = Dim2, color = as.factor(AJCC_bin))) +
  geom_point(alpha = 0.7) +
  labs(title = "UMAP dei dati (colorato per AJCC_bin)", color = "AJCC_bin") +
  theme_minimal()
```

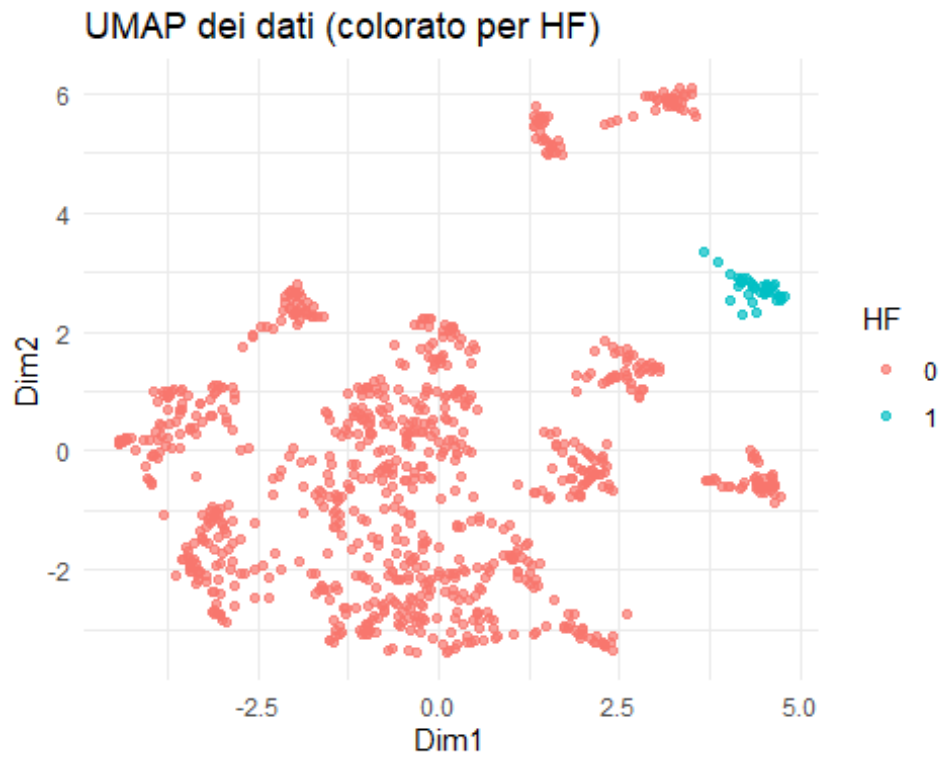


```
tsne_df$HF <- data$HF
umap_df$HF <- data$HF

# t-SNE con colorazione
ggplot(tsne_df, aes(x = Dim1, y = Dim2, color = as.factor(HF))) +
  geom_point(alpha = 0.7) +
  labs(title = "t-SNE dei dati (colorato per HF)", color = "HF") +
  theme_minimal()
```

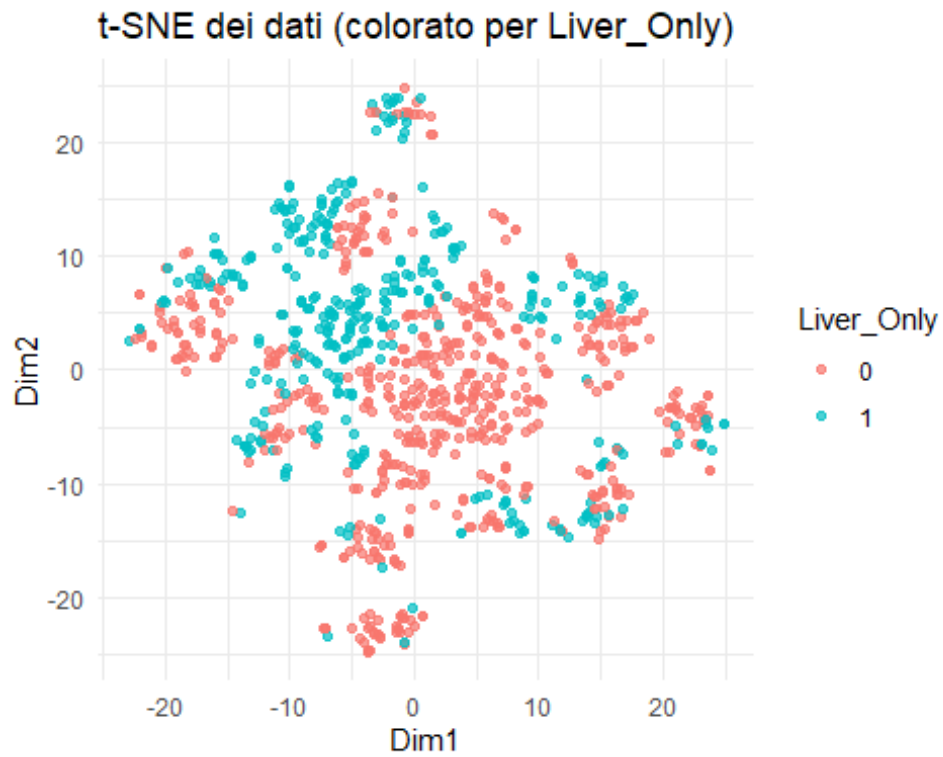


```
# UMAP con colorazione
ggplot(umap_df, aes(x = Dim1, y = Dim2, color = as.factor(HF))) +
  geom_point(alpha = 0.7) +
  labs(title = "UMAP dei dati (colorato per HF)", color = "HF") +
  theme_minimal()
```

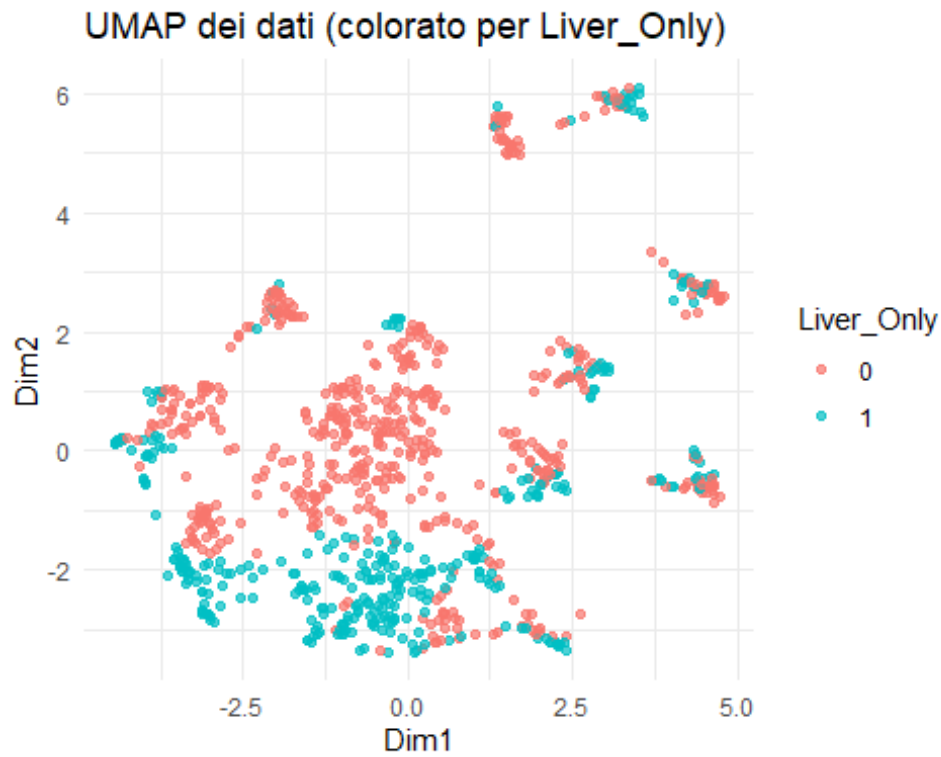


```
tsne_df$Liver_Only <- data$Liver_Only
umap_df$Liver_Only <- data$Liver_Only

# t-SNE con colorazione
ggplot(tsne_df, aes(x = Dim1, y = Dim2, color = as.factor(Liver_Only))) +
  geom_point(alpha = 0.7) +
  labs(title = "t-SNE dei dati (colorato per Liver_Only)", color = "Liver_Only") +
  theme_minimal()
```

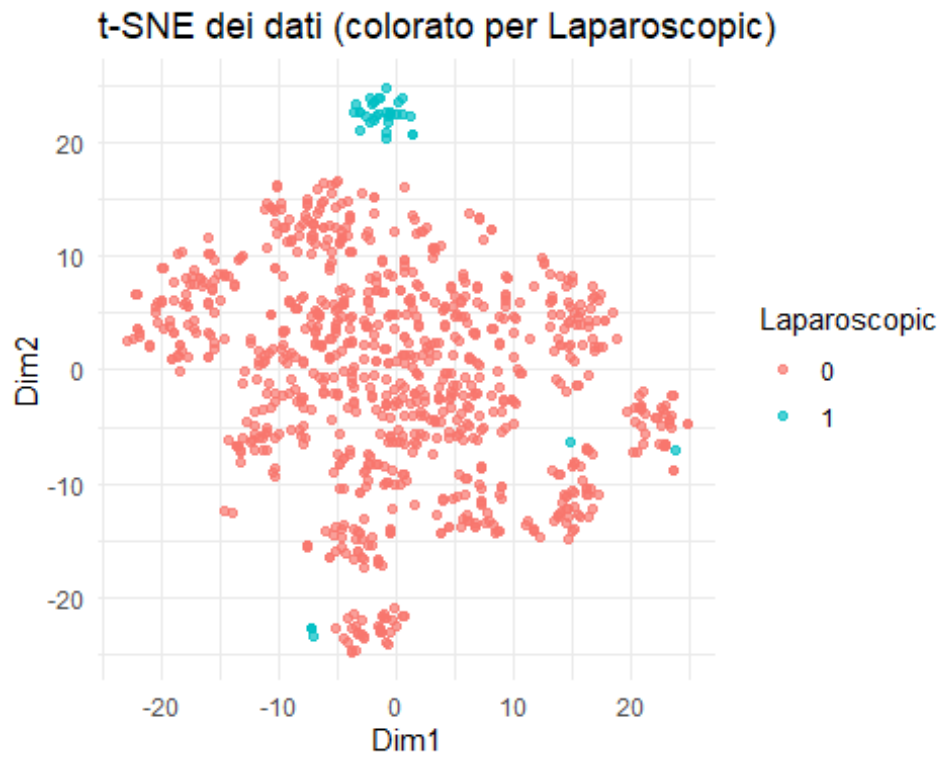


```
# UMAP con colorazione
ggplot(umap_df, aes(x = Dim1, y = Dim2, color = as.factor(Liver_Only))) +
  geom_point(alpha = 0.7) +
  labs(title = "UMAP dei dati (colorato per Liver_Only)", color = "Liver_Only") +
  theme_minimal()
```

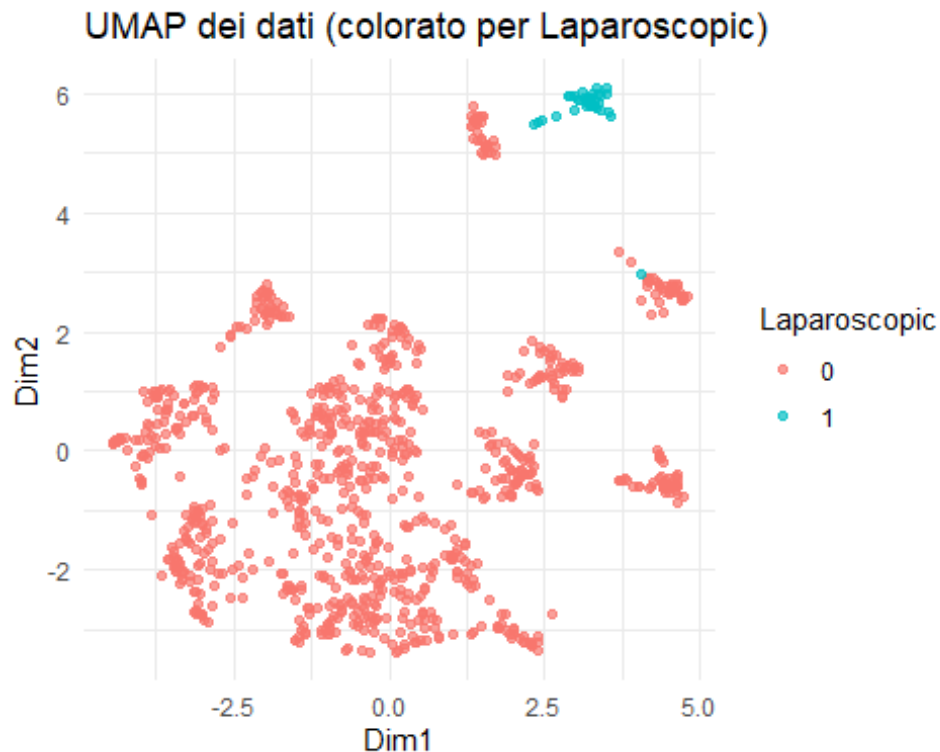


```
tsne_df$Laparoscopic <- data$Laparoscopic
umap_df$Laparoscopic <- data$Laparoscopic

# t-SNE con colorazione
ggplot(tsne_df, aes(x = Dim1, y = Dim2, color = as.factor(Laparoscopic))) +
  geom_point(alpha = 0.7) +
  labs(title = "t-SNE dei dati (colorato per Laparoscopic)", color =
"Laparoscopic") +
  theme_minimal()
```



```
# UMAP con colorazione
ggplot(umap_df, aes(x = Dim1, y = Dim2, color = as.factor(Laparoscopic))) +
  geom_point(alpha = 0.7) +
  labs(title = "UMAP dei dati (colorato per Laparoscopic)", color =
"Laparoscopic") +
  theme_minimal()
```

t-test SU UMAP

```
t.test(umap_df$Dim1 ~ data$Laparoscopic)
```

```
##
## Welch Two Sample t-test
##
## data:  umap_df$Dim1 by data$Laparoscopic
## t = -34.694, df = 285.68, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group 0 and group 1 is
## not equal to 0
## 95 percent confidence interval:
## -3.492608 -3.117587
## sample estimates:
## mean in group 0 mean in group 1
## -0.1335028      3.1715947
```

```
t.test(umap_df$Dim1 ~ data$HF)
```

```
##
## Welch Two Sample t-test
##
## data:  umap_df$Dim1 by data$HF
## t = -53.982, df = 492.31, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group 0 and group 1 is
## not equal to 0
## 95 percent confidence interval:
## -4.767586 -4.432722
## sample estimates:
```

```
## mean in group 0 mean in group 1
##      -0.2008801      4.3992735
```

LAPAROSCOPIA

Gruppo 0 (non laparoscopici): ≈ -0.14

Gruppo 1 (laparoscopici): $\approx 3.3 \rightarrow$ forte separazione = variabile determinante nei pattern latenti della popolazione

INFARTO GRUPPO 0 , NO HF, = -0.21 GRUPPO 1 , YES HF = 4.8

Si verificano tutte le variabili, per vedere che cosa comunica la U

T TEST SU DIMENSIONE 1 DI UMAP

Lista delle variabili da testare

```
variabili <- c("Age", "Gender", "ASA3", "DM", "CAD", "HF", "CVA",
              "CKD", "CEA", "LogCEA", "Laparoscopic", "TumorLOC", "EA",
              "AnesTime",
              "Log2AT", "RBC", "Liver_Only", "Cell_diff", "Mucin_TYPE",
              "SignetRING",
              "Lymphovascularinvasion", "perineural", "CT", "RT", "NACTRT",
              "Interval", "Progress", "IntervalR", "AJCC_bin")
```

Dataframe per salvare i risultati

```
risultati <- data.frame(
  Variabile = character(),
  p_value = numeric(),
  t_stat = numeric(),
  mean_group0 = numeric(),
  mean_group1 = numeric(),
  stringsAsFactors = FALSE
)
```

Loop sulle variabili

```
for (v in variabili) {
  if (v %in% names(data)) {
    # Isola la variabile target e la pulizia dei NA
    tmp <- data.frame(
      x = umap_df$Dim1,
      group = data[[v]]
    )
    tmp <- tmp[complete.cases(tmp), ]
  }
}
```

Converti character/factor in numerico binario se possibile

```
if (is.character(tmp$group) || is.factor(tmp$group)) {
  levels <- unique(tmp$group)
  if (length(levels) == 2) {
    tmp$group <- as.numeric(tmp$group == levels[2])
  } else {
    next # Salta variabili con più di 2 classi
  }
}
```

```

}

# Esegui t-test solo se gruppo binario
if (length(unique(tmp$group)) == 2) {
  t <- t.test(x ~ group, data = tmp)
  risultati <- rbind(risultati, data.frame(
    Variabile = v,
    p_value = t$p.value,
    t_stat = t$statistic,
    mean_group0 = mean(tmp$x[tmp$group == 0]),
    mean_group1 = mean(tmp$x[tmp$group == 1])
  ))
}
}
}

```

```

# Ordina per significatività
risultati <- risultati[order(risultati$p_value), ]

```

```

# Visualizza
print(risultati)

```

##	Variabile	p_value	t_stat	mean_group0	mean_group1
## t4	HF	6.398122e-209	-53.9822595	-0.20088007	4.39927349
## t7	Laparoscopic	1.936245e-104	-34.6936676	-0.13350285	3.17159469
## t5	CVA	1.393998e-68	-36.1143917	-0.25625941	4.17267680
## t3	CAD	2.699733e-58	-25.1980433	-0.24057574	3.00012099
## t16	CT	5.111489e-33	15.3779506	2.31053838	-0.26422215
## t17	RT	6.314969e-29	14.7199282	0.34978814	-2.91966015
## t19	Progress	8.426028e-25	11.0794002	1.40176979	-0.36929383
## t13	SignetRING	2.060273e-22	-13.1634029	-0.08301878	1.77174218
## t18	NACTRT	7.261617e-22	10.8598525	0.34233706	-1.85053532
## t1	ASA3	1.756279e-12	-7.1941937	-0.43971215	0.72775546
## t6	CKD	9.743247e-06	-4.5720872	-0.14596002	0.94102676
## t8	TumorLOC	2.669602e-04	-3.6694274	-0.18777152	0.42432354
## t2	DM	1.029026e-02	-2.5830349	-0.10644638	0.40405038
## t12	Mucin_TYPE	1.446632e-01	1.4725149	0.02937571	-0.36059753
## t	Gender	1.541788e-01	1.4263213	NaN	0.08556258
## t10	Liver_Only	2.128323e-01	1.2469389	0.07927630	-0.12413292
## t11	Cell_diff	2.273680e-01	-1.2118135	NaN	-0.03363744
## t14	Lymphovascularinvasion	2.521048e-01	-1.1459861	-0.09229925	0.08682387
## t15	perineural	4.620611e-01	0.7362460	0.02945362	-0.10151492
## t20	AJCC_bin	6.075282e-01	0.5137879	0.03334808	-0.04599436
## t9	EA	7.222358e-01	-0.3559367	-0.01220246	0.06182042

T TEST DI DIM2 DI UMAP

```

# Lista delle variabili binarie o categoriali 2 livelli

```

```

vars <- c("Age", "Gender", "ASA3", "DM", "CAD", "HF", "CVA",
  "CKD", "CEA", "LogCEA", "Laparoscopic", "TumorLOC", "EA",
  "AnesTime", "Log2AT", "RBC", "Liver_Only", "Cell_diff",
  "Mucin_TYPE", "SignetRING", "Lymphovascularinvasion", "perineural",

```

```

      "CT", "RT", "NACTRT", "Interval", "Progress", "IntervalR", "AJCC_bin")

# Data frame vuoto per salvare i risultati
results <- data.frame(Variabile=character(),
                      p_value=double(),
                      t_stat=double(),
                      mean_group0=double(),
                      mean_group1=double(),
                      stringsAsFactors=FALSE)

# Loop per ogni variabile
for (var in vars) {
  # Estrai la variabile target
  group <- data[[var]]

  # Verifica che ci siano solo due gruppi (t-test richiede questo)
  if (length(unique(na.omit(group))) == 2) {
    test <- try(t.test(umap_df$Dim2 ~ group), silent=TRUE)
    if (!inherits(test, "try-error")) {
      means <- tapply(umap_df$Dim2, group, mean, na.rm=TRUE)
      results <- rbind(results, data.frame(
        Variabile = var,
        p_value = test$p.value,
        t_stat = test$statistic,
        mean_group0 = means[1],
        mean_group1 = means[2]
      ))
    }
  } else {
    # Se la variabile non è binaria, saltiamo
    warning(paste("Variabile non binaria:", var))
  }
}

## Warning: Variabile non binaria: Age
## Warning: Variabile non binaria: CEA
## Warning: Variabile non binaria: LogCEA
## Warning: Variabile non binaria: AnesTime
## Warning: Variabile non binaria: Log2AT
## Warning: Variabile non binaria: RBC
## Warning: Variabile non binaria: Interval
## Warning: Variabile non binaria: IntervalR

# Ordina per p-value crescente
results <- results[order(results$p_value), ]

```

Visualizza i primi risultati

```
print(results)
```

##	Variabile	p_value	t_stat	mean_group0	mean_group1
## t4	HF	4.240648e-157	-34.5270195	-0.124409554	2.72456923
## t7	Laparoscopic	1.631970e-78	-56.8650397	-0.242756619	5.76710995
## t20	AJCC_bin	3.150395e-38	-13.5509611	-0.774713667	1.06850119
## t13	SignetRING	1.519383e-23	-19.1795691	-0.230563551	4.92056359
## t10	Liver_Only	1.193670e-22	10.1495307	0.597933643	-0.93626024
## t12	Mucin_TYPE	7.882359e-20	-12.1426103	-0.240759776	2.95541349
## t3	CAD	6.786351e-18	-10.3533042	-0.130674420	1.62958689
## t11	Cell_diff	4.539633e-16	-9.1202721	-0.257428308	1.75799331
## t14	Lymphovascularinvasion	3.102630e-11	-6.7255243	-0.510698681	0.48040300
## t17	RT	3.605453e-08	-5.7495865	-0.095378280	0.79611667
## t18	NACTRT	5.648979e-06	4.6482514	0.130434509	-0.70507605
## t15	perineural	6.041506e-06	-4.6099173	-0.206312319	0.71107644
## t8	TumorLOC	7.443072e-04	-3.3937558	-0.175536017	0.39667392
## t1	ASA3	7.379853e-03	-2.6857344	-0.149821673	0.24796573
## t16	CT	1.997188e-02	2.3549279	0.418365638	-0.04784230
## t9	EA	2.964267e-02	2.1892779	0.072290091	-0.36623788
## t2	DM	1.391424e-01	-1.4827086	-0.054992193	0.20874000
## t6	CKD	1.423395e-01	-1.4739216	-0.041014314	0.26442562
## t	Gender	4.151029e-01	-0.8153791	-0.047349085	0.07728241
## t19	Progress	4.472368e-01	-0.7610865	-0.116766381	0.03076190
## t5	CVA	6.882853e-01	0.4029654	0.005581875	-0.09088978

ASSI LATENTI DELLA UMPA: INTERPRETAZIONE CLINICA

ha imparato a comprimere l'informazione clinica in due dimensioni che riflettono cluster clinici. pattern che emerge:

dimensione1 è guidata soprattutto da: COMORBIDITÀ

Comorbidità cliniche e fragilità: CAD, CVA, HF, ASA3, CKD.

Tipo di chirurgia (Laparoscopic), radioterapia (RT), chemioterapia (CT).

Riflette quindi lo stato clinico generale e decisioni terapeutiche.

dimensione2 è guidata soprattutto da: CARATTERISTICHE TUMORALI, STAGING E TRATTAMENTO

Parametri istopatologici e prognostici: Signet Ring, Mucin Type, AJCC_bin, Cell_diff, Lymphovascular invasion.

Trattamenti avanzati: NACTRT, chirurgia laparoscopica.

Presenza di metastasi (Liver_Only).

Riflette l'aggressività tumorale e il tipo istologico.

CRITERI DI SEPARAZIONE DIM 2 ha un significato clinico focalizzato sull'approccio chirurgico e allo stadio del tumore DIM 1 si focalizza principalmente sulle comorbidità e alle caratteristiche del tumore.

CLUSTER

Il clustering è eseguito sia sui dati originali ("raw") che su quelli ridotti tramite tecniche come PCA e UMAP per verificare la presenza di strutture di gruppo intrinseche. Sui dati raw, il clustering cerca pattern naturali nello spazio delle variabili originali. Tuttavia, le alte dimensioni possono rendere difficile identificare cluster chiari. La riduzione dimensionale (PCA lineare o UMAP non lineare) permette di comprimere l'informazione in poche dimensioni interpretabili, evidenziando pattern latenti e migliorando la separabilità visiva. Clustering su dati ridotti può rivelare strutture che non emergono nei dati originali. Confrontare i risultati nei due spazi aiuta a determinare se i cluster sono reali o artefatti della proiezione. È un approccio integrato per validare la solidità dei raggruppamenti e l'informazione utile presente nei dati. ## CLUSTER UMAP

Per visualizzare la separazione tra gruppi trovati.

```
head(umap_df)
```

```
##           Dim1          Dim2 Death Gender ASA3 CT AJCC_bin HF Liver_Only
## 1 -1.06827569  0.52201078      1      1      1  1          1  0          0
## 2  2.58056259  1.33869973      1      2      0  1          0  0          1
## 3 -1.14504555 -2.64269915      1      2      0  1          0  0          1
## 4 -0.96934509  0.42800496      1      2      0  1          1  0          0
## 5 -1.38344301 -1.08368099      1      1      0  1          0  0          0
## 6  0.08707221  0.06347988      0      2      0  1          1  0          0
##  Laparoscopic
## 1              0
## 2              0
## 3              0
## 4              0
## 5              0
## 6              0
```

```
set.seed(42)
```

```
kmeans_result <- kmeans(umap_df, centers = 3)
```

```
umap_df$cluster <- as.factor(kmeans_result$cluster)
```

```
library(cluster)
```

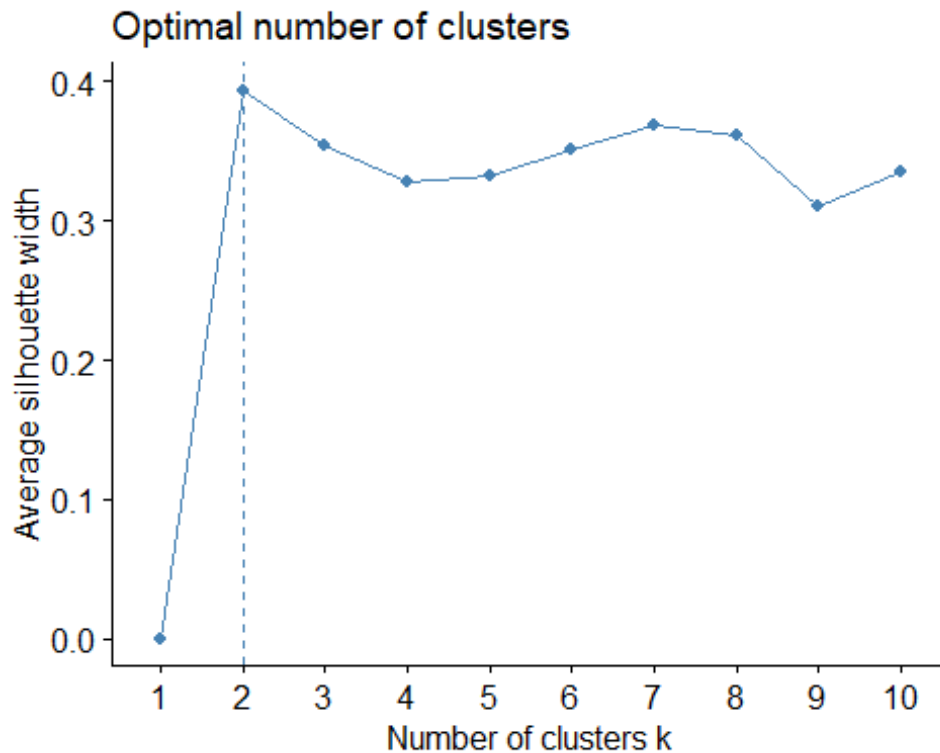
```
## Warning: package 'cluster' was built under R version 4.3.3
```

```
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.3.3
```

Welcome! Want to learn more? See two factoextra-related books at <https://goo.gl/ve3WBa>

```
fviz_nbclust(umap_df, kmeans, method = "silhouette")
```



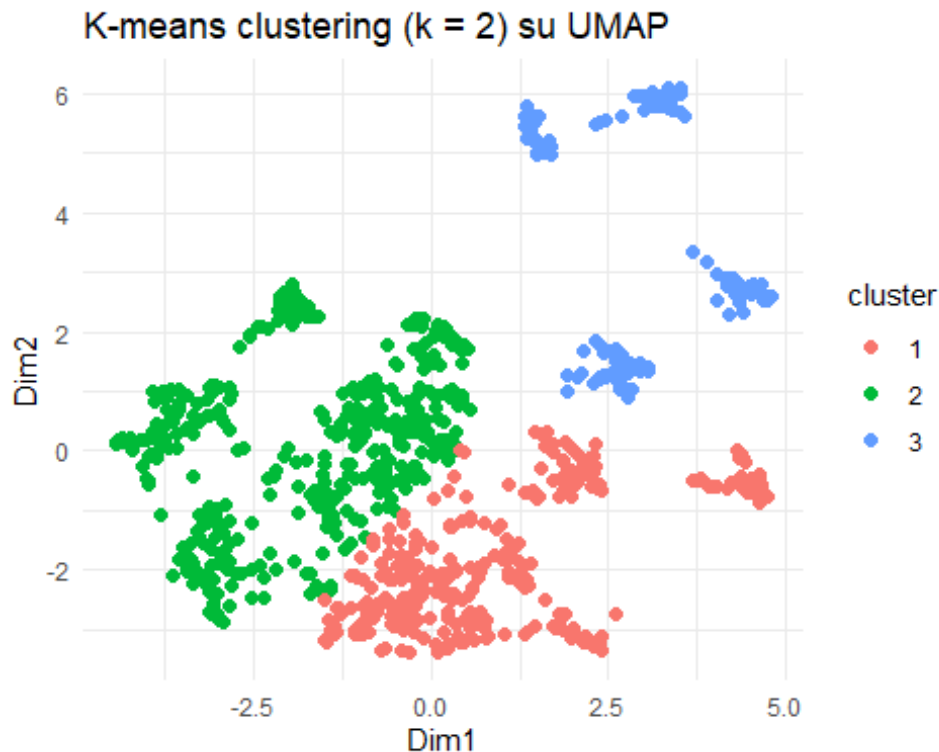
Si prova sia con 3 che con 2

```
set.seed(123) # Per riproducibilità
kmeans_res <- kmeans(umap_df, centers = 3)

# Aggiungi i cluster al data frame
umap_df$cluster <- as.factor(kmeans_res$cluster)

library(ggplot2)

ggplot(umap_df, aes(x = Dim1, y = Dim2, color = cluster)) +
  geom_point(size = 2) +
  theme_minimal() +
  labs(title = "K-means clustering (k = 2) su UMAP")
```



```
data$cluster <- umap_df$cluster
```

```
# Variabili continue
```

```
aggregate(. ~ cluster, data = data, FUN = mean, na.rm = TRUE)
```

##	cluster	Age	Gender	ASA3	DM	CAD	HF	CVA
## 1	1	67.09816	1.343558	0.3680982	0.1625767	0.02760736	0.0000000	0.14417178
## 2	2	62.39269	1.422374	0.3356164	0.2100457	0.00000000	0.0000000	0.00000000
## 3	3	69.37500	1.335526	0.5131579	0.3026316	0.38815789	0.2631579	0.03947368
##	CKD	LogCEA	Laparoscopic	TumorLOC	EA	Log2AT	RBC	
## 1	0.1717791	1.351321	0.0000000	0.3404908	0.1809816	8.293515	0.5306748	
## 2	0.0913242	1.435982	0.0000000	0.2488584	0.1575342	8.407882	0.5091324	
## 3	0.1776316	1.341117	0.2434211	0.4013158	0.1513158	8.248017	0.5921053	
##	Liver_Only	Cell_diff	Mucin_TYPE	SignetRING	Lymphovascularinvasion	perineural		
## 1	0.5950920	1.058282	0.01533742	0.006134969		0.4171779	0.1226994	
## 2	0.2557078	1.148402	0.09360731	0.000000000		0.5570776	0.2762557	
## 3	0.3355263	1.217105	0.15131579	0.256578947		0.6052632	0.2960526	
##	CT	RT	NACTRT	Death	Progress	IntervalOD	IntervalOR	
## 1	0.7852761	0.009202454	0.04601227	0.3006135	0.6441718	30.90140	16.497046	
## 2	0.9840183	0.200913242	0.26027397	0.5091324	0.9109589	20.93001	7.710688	
## 3	0.8881579	0.046052632	0.09210526	0.4276316	0.7631579	21.69999	10.611694	
##	AJCC_bin							
## 1	0.1349693							
## 2	0.6073059							
## 3	0.4934211							

```
results <- lapply(names(data)[names(data) != "cluster"], function(var) {
  formula <- as.formula(paste(var, "~ cluster"))
})
```



```

model <- aov(formula, data = data)
pval <- summary(model)[[1]][["Pr(>F)"]][1]

means <- tapply(data[[var]], data$cluster, mean, na.rm = TRUE)

data.frame(
  Variable = var,
  p_value = round(pval, 5),
  mean_cluster1 = round(means[1], 2),
  mean_cluster2 = round(means[2], 2),
  mean_cluster3 = round(means[3], 2)
)
})

results_df <- do.call(rbind, results)
results_df <- results_df[order(results_df$p_value), ]
print(results_df)

```

	Variable	p_value	mean_cluster1	mean_cluster2	mean_cluster3
## 1	Age	0.00000	67.10	62.39	69.38
## 14	CAD	0.00000	0.03	0.00	0.39
## 15	HF	0.00000	0.00	0.00	0.26
## 16	CVA	0.00000	0.14	0.00	0.04
## 19	Laparoscopic	0.00000	0.00	0.00	0.24
## 114	Liver_Only	0.00000	0.60	0.26	0.34
## 115	Cell_diff	0.00000	1.06	1.15	1.22
## 116	Mucin_TYPE	0.00000	0.02	0.09	0.15
## 117	SignetRING	0.00000	0.01	0.00	0.26
## 119	perineural	0.00000	0.12	0.28	0.30
## 120	CT	0.00000	0.79	0.98	0.89
## 121	RT	0.00000	0.01	0.20	0.05
## 122	NACTRT	0.00000	0.05	0.26	0.09
## 123	Death	0.00000	0.30	0.51	0.43
## 124	Progress	0.00000	0.64	0.91	0.76
## 125	IntervalOD	0.00000	30.90	20.93	21.70
## 126	IntervalOR	0.00000	16.50	7.71	10.61
## 127	AJCC_bin	0.00000	0.13	0.61	0.49
## 118	Lymphovascularinvasion	0.00003	0.42	0.56	0.61
## 112	Log2AT	0.00017	8.29	8.41	8.25
## 12	ASA3	0.00046	0.37	0.34	0.51
## 110	TumorLOC	0.00052	0.34	0.25	0.40
## 17	CKD	0.00123	0.17	0.09	0.18
## 13	DM	0.00205	0.16	0.21	0.30
## 11	Gender	0.03968	1.34	1.42	1.34
## 18	LogCEA	0.33378	1.35	1.44	1.34
## 113	RBC	0.43646	0.53	0.51	0.59
## 111	EA	0.61079	0.18	0.16	0.15

```

library(cluster)
library(factoextra)

```

```

# 1. Prepara i dati da clustering (solo le coordinate UMAP, non il factor cluster)
coords <- umap_df[, c("Dim1", "Dim2")]

# 2. Calcola la matrice di distanza (euclidea di default)
diss <- dist(coords)

# 3. Calcola i silhouette
sil <- silhouette(as.integer(umap_df$cluster), diss)

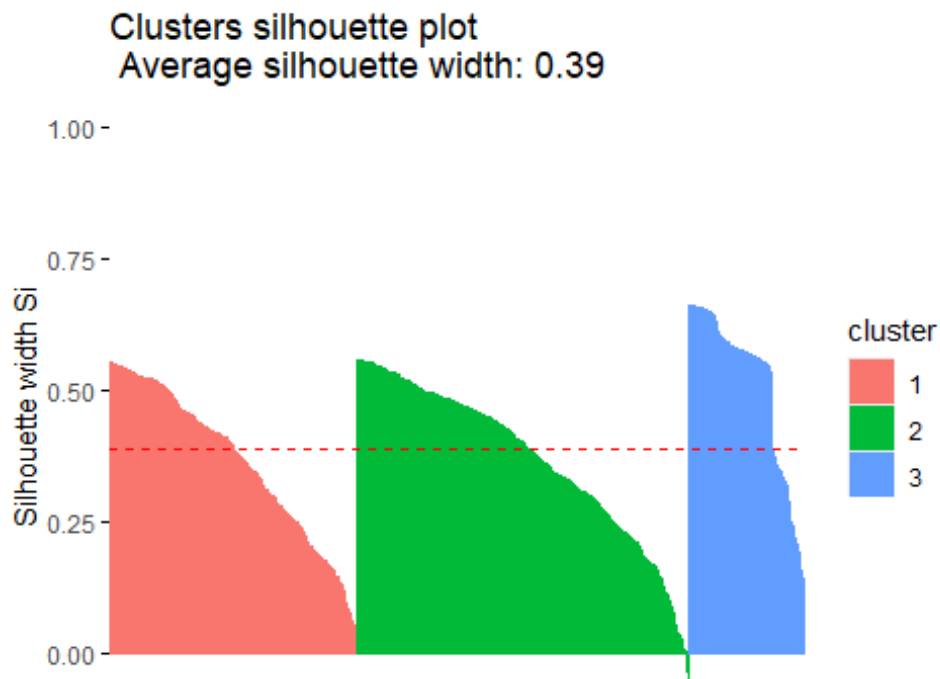
# 4. Statistica complessiva media
avg_sil_width <- mean(sil[, "sil_width"])
cat("Average silhouette width:", round(avg_sil_width, 3), "\n")

## Average silhouette width: 0.389

# 5. Visualizza il plot delle silhouette
fviz_silhouette(sil)

##   cluster size ave.sil.width
## 1         1  326         0.36
## 2         2  438         0.37
## 3         3  152         0.51

```



CLUSTER UMAP con k=2

Per visualizzare la separazione tra gruppi trovati.

```
head(umap_df)
```

##	Dim1	Dim2	Death	Gender	ASA3	CT	AJCC_bin	HF	Liver_Only
## 1	-1.06827569	0.52201078	1	1	1	1	1	0	0
## 2	2.58056259	1.33869973	1	2	0	1	0	0	1
## 3	-1.14504555	-2.64269915	1	2	0	1	0	0	1
## 4	-0.96934509	0.42800496	1	2	0	1	1	0	0
## 5	-1.38344301	-1.08368099	1	1	0	1	0	0	0
## 6	0.08707221	0.06347988	0	2	0	1	1	0	0

##	Laparoscopic	cluster
## 1	0	2
## 2	0	3
## 3	0	1
## 4	0	2
## 5	0	2
## 6	0	2

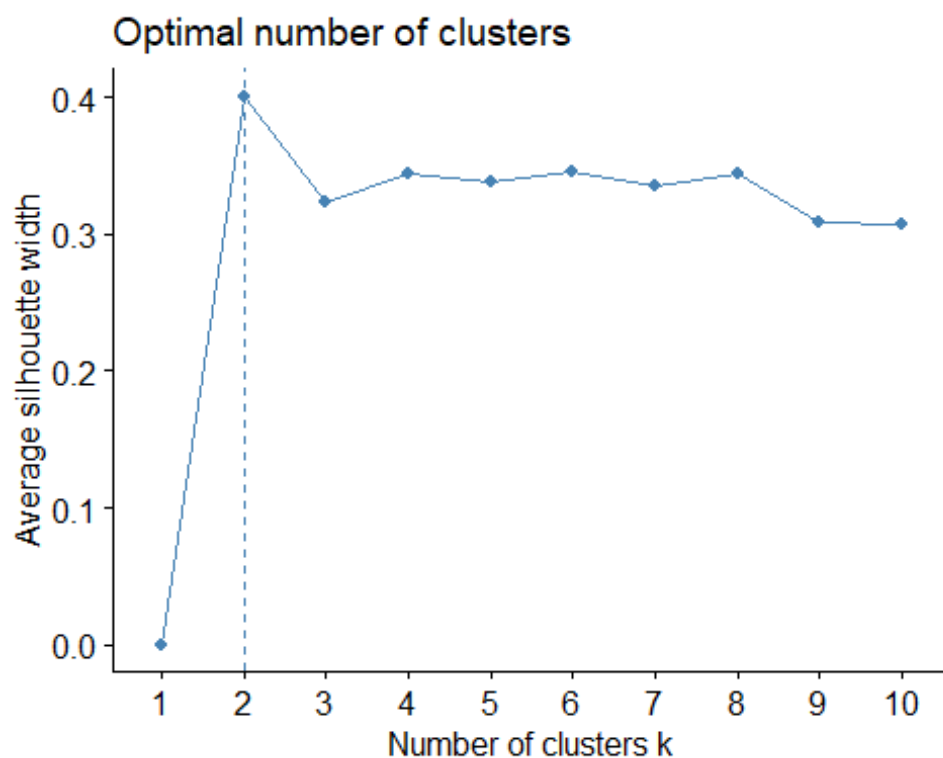
```

set.seed(42)
kmeans_result <- kmeans(umap_df, centers = 2)
umap_df$cluster <- as.factor(kmeans_result$cluster)

library(cluster)
library(factoextra)

fviz_nbclust(umap_df, kmeans, method = "silhouette")

```



```

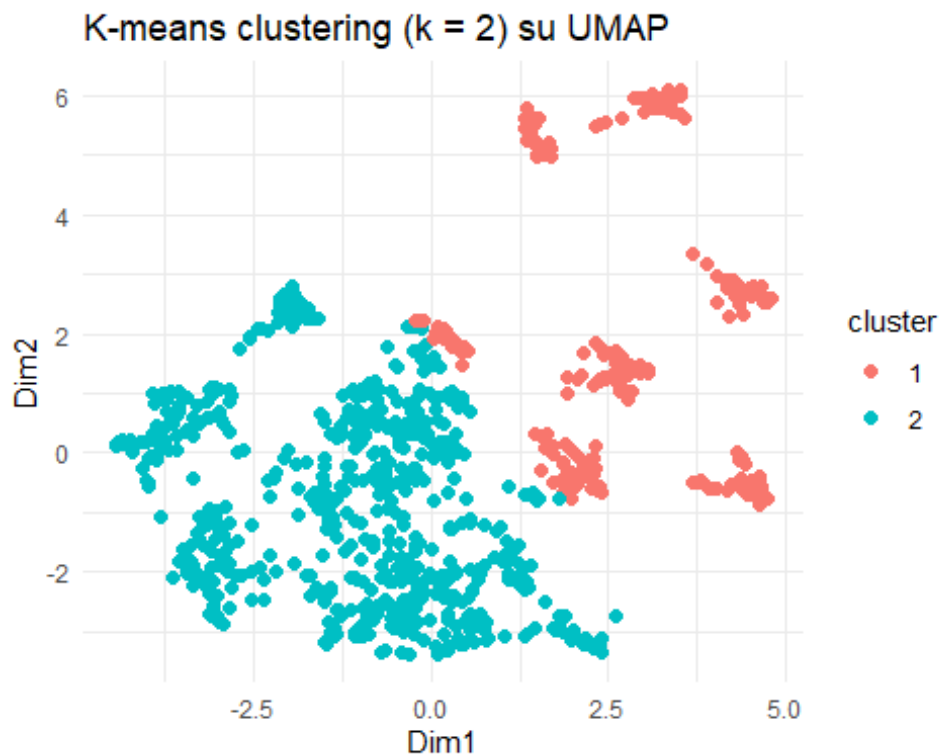
set.seed(123) # Per riproducibilità
kmeans_res <- kmeans(umap_df, centers = 2)

```

```
# Aggiungi i cluster al data frame
umap_df$cluster <- as.factor(kmeans_res$cluster)
```

```
library(ggplot2)
```

```
ggplot(umap_df, aes(x = Dim1, y = Dim2, color = cluster)) +
  geom_point(size = 2) +
  theme_minimal() +
  labs(title = "K-means clustering (k = 2) su UMAP")
```



```
data$cluster <- umap_df$cluster
```

```
# Variabili continue
```

```
aggregate(. ~ cluster, data = data, FUN = mean, na.rm = TRUE)
```

##	cluster	Age	Gender	ASA3	DM	CAD	HF	CVA
##	1	70.19245	1.313208	0.5358491	0.2792453	0.2566038	0.1509434	0.192452830
##	2	63.20430	1.407066	0.3118280	0.1797235	0.0000000	0.0000000	0.003072197
##		CKD	LogCEA	Laparoscopic	TumorLOC	EA	Log2AT	RBC
##	1	0.2000000	1.369588	0.1396226	0.3773585	0.1433962	8.263979	0.5471698
##	2	0.1075269	1.398464	0.0000000	0.2780338	0.1735791	8.371862	0.5238095
##		Liver_Only	Cell_diff	Mucin_TYPE	SignetRING	Lymphovascularinvasion	perineural	
##	1	0.3471698	1.226415	0.09811321	0.150943396		0.5924528	0.2830189
##	2	0.4070661	1.087558	0.06605223	0.001536098		0.4838710	0.2012289
##		CT	RT	NACTRT	Death	Progress	IntervalOD	IntervalOR
##	1	0.6943396	0.03396226	0.08679245	0.3962264	0.6528302	18.48973	8.90783
##	2	0.9800307	0.13671275	0.18433180	0.4316436	0.8479263	27.09650	12.30065

```
##      AJCC_bin
## 1 0.4528302
## 2 0.4070661

results <- lapply(names(data)[names(data) != "cluster"], function(var) {
  formula <- as.formula(paste(var, "~ cluster"))
  model <- aov(formula, data = data)
  pval <- summary(model)[[1]][["Pr(>F)"]][1]

  means <- tapply(data[[var]], data$cluster, mean, na.rm = TRUE)

  data.frame(
    Variable = var,
    p_value = round(pval, 5),
    mean_cluster1 = round(means[1], 2),
    mean_cluster2 = round(means[2], 2))
})

results_df <- do.call(rbind, results)
results_df <- results_df[order(results_df$p_value), ]
print(results_df)
```

```
##      Variable p_value mean_cluster1 mean_cluster2
## 1      Age 0.00000      70.19      63.20
## 12     ASA3 0.00000       0.54       0.31
## 14     CAD 0.00000       0.26       0.00
## 15     HF 0.00000       0.15       0.00
## 16     CVA 0.00000       0.19       0.00
## 19  Laparoscopic 0.00000       0.14       0.00
## 115    Cell_diff 0.00000       1.23       1.09
## 117   SignetRING 0.00000       0.15       0.00
## 120     CT 0.00000       0.69       0.98
## 121     RT 0.00000       0.03       0.14
## 124   Progress 0.00000       0.65       0.85
## 125  IntervalOD 0.00000      18.49      27.10
## 17     CKD 0.00019       0.20       0.11
## 122   NACTRT 0.00022       0.09       0.18
## 13     DM 0.00076       0.28       0.18
## 112   Log2AT 0.00213       8.26       8.37
## 118 Lymphovascularinvasion 0.00284       0.59       0.48
## 110     TumorLOC 0.00309       0.38       0.28
## 126  IntervalOR 0.00640       8.91      12.30
## 119  perineural 0.00714       0.28       0.20
## 11     Gender 0.00792       1.31       1.41
## 114  Liver_Only 0.09209       0.35       0.41
## 116  Mucin_TYPE 0.09569       0.10       0.07
## 127     AJCC_bin 0.20366       0.45       0.41
## 111     EA 0.26476       0.14       0.17
## 123    Death 0.32548       0.40       0.43
## 113     RBC 0.63957       0.55       0.52
## 18    LogCEA 0.65975       1.37       1.40
```

CLUSTER 1: ANZIANI CON COMORBIDITÁ Significativamente piú anziano, cono ASA piu alto, con molto piu presenti variabili come CAD, HF e CVA. In questo cluster sono anche presenti valori per SignetRing. Meno aderenze estese. Meno femmine, piu chirurgie laparoscopiche, tumori piu retto. mAGGIORE DIFFERENZIAZIONE CELLULARE. INTERVALLO MINORE TRA OPERAZIONE CHIRURGICA E MORTE / RECIDIVA. I TUMORI PROGREDISCONO DI MENO, IL GRADO 4A4B É UGUALE

CLUSTER 2: GIOVANI CON TRATTAMENTI AGGRESSIVI, PEGGIOR OUTCOME

Valori molto piu alti di CT, RT e NACTRT. Tasso di progressione del tumore molto maggiore, quasi doppio. Tasso di morte maggiore. Marker tumoreali piú elevati, interventi piu lunghi e piu complessi (meno laparoscopia, tumori piu colon che retto). Tempi piu lunghi prima delle recidive o della mote

```
library(cluster)
library(factoextra)

# 1. Prepara i dati da clustering (solo le coordinate UMAP, non il factor cluster)
coords <- umap_df[, c("Dim1", "Dim2")]

# 2. Calcola la matrice di distanza (euclidea di default)
diss <- dist(coords)

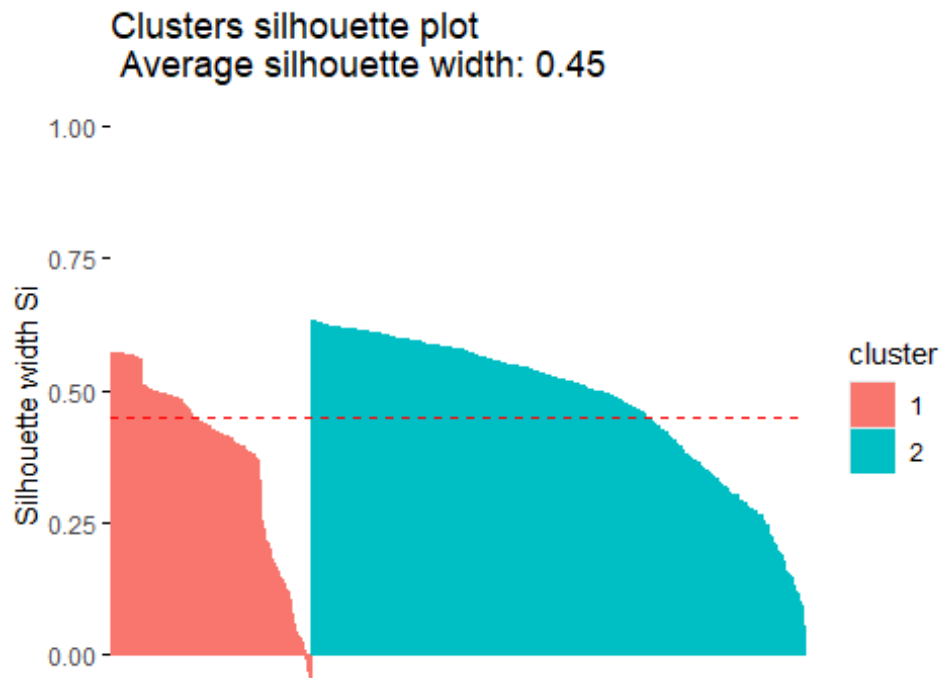
# 3. Calcola i silhouette
sil <- silhouette(as.integer(umap_df$cluster), diss)

# 4. Statistica complessiva media
avg_sil_width <- mean(sil[, "sil_width"])
cat("Average silhouette width:", round(avg_sil_width, 3), "\n")

## Average silhouette width: 0.448

# 5. Visualizza il plot delle silhouette
fviz_silhouette(sil)

##   cluster size ave.sil.width
## 1         1  265          0.38
## 2         2  651          0.48
```

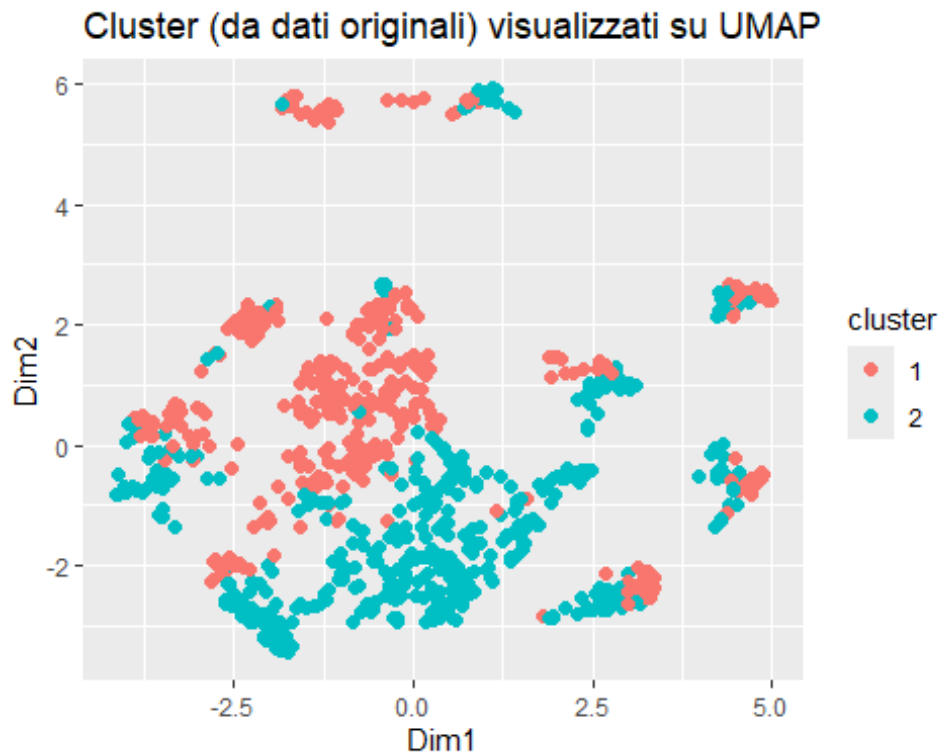


CLUSTER RAW con viz UMAP

```
# 2. Clustering
set.seed(42)
kmeans_res <- kmeans(data_scaled, centers = 2, nstart = 25)

# 3. UMAP per visualizzare
umap_out <- umap(data_scaled)
umap_df <- data.frame(Dim1 = umap_out$layout[,1], Dim2 = umap_out$layout[,2],
                     cluster = as.factor(kmeans_res$cluster))

# 4. Visualizzazione
ggplot(umap_df, aes(x = Dim1, y = Dim2, color = cluster)) +
  geom_point(size = 2) +
  labs(title = "Cluster (da dati originali) visualizzati su UMAP")
```



```
# 1. Calcola la matrice delle distanze (euclidea)
dist_matrix <- dist(data_scaled)

# 2. Calcola l'indice di silhouette
library(cluster)
sil <- silhouette(kmeans_res$cluster, dist_matrix)

# 3. Calcola la silhouette media
silhouette_score <- mean(sil[, 3])
print(paste("Indice di Silhouette medio:", round(silhouette_score, 3)))

## [1] "Indice di Silhouette medio: 0.069"
```

CLUSTER PCA

```
# PCA
pca <- prcomp(data_scaled)
# Scegli le prime N componenti che spiegano >80% varianza
summary(pca)

## Importance of components:
##              PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation  1.62086 1.5143 1.33523 1.20570 1.19565 1.15820 1.10167
## Proportion of Variance 0.09383 0.0819 0.06367 0.05192 0.05106 0.04791 0.04335
## Cumulative Proportion 0.09383 0.1757 0.23940 0.29132 0.34237 0.39028 0.43363
##              PC8      PC9      PC10     PC11     PC12     PC13     PC14
## Standard deviation  1.06569 1.05116 1.03781 0.9997 0.99144 0.97252 0.9391
## Proportion of Variance 0.04056 0.03946 0.03847 0.0357 0.03511 0.03378 0.0315
## Cumulative Proportion 0.47419 0.51365 0.55212 0.5878 0.62292 0.65670 0.6882
```


##	PC15	PC16	PC17	PC18	PC19	PC20	PC21
## Standard deviation	0.92065	0.91061	0.90472	0.87670	0.85766	0.84411	0.82224
## Proportion of Variance	0.03027	0.02961	0.02923	0.02745	0.02627	0.02545	0.02415
## Cumulative Proportion	0.71846	0.74808	0.77731	0.80476	0.83103	0.85648	0.88063

##	PC22	PC23	PC24	PC25	PC26	PC27	PC28
## Standard deviation	0.78506	0.77350	0.75823	0.73717	0.7159	0.52551	0.47004
## Proportion of Variance	0.02201	0.02137	0.02053	0.01941	0.0183	0.00986	0.00789
## Cumulative Proportion	0.90264	0.92400	0.94454	0.96394	0.9822	0.99211	1.00000

Prendiamo per esempio le prime 10

```
pca_data <- as.data.frame(pca$x[, 1:10])
```

K-means su PCA

```
kmeans_res <- kmeans(pca_data, centers = 2)
```

Riduci i dati a 18 componenti principali

```
pca_data <- as.data.frame(pca$x[, 1:17])
```

K-means clustering sui dati ridotti

```
set.seed(42)
```

```
kmeans_res <- kmeans(pca_data, centers = 2, nstart = 25)
```

Visualizza i cluster su UMAP

```
umap_out <- umap(pca_data)
```

```
umap_df <- data.frame(Dim1 = umap_out$layout[,1], Dim2 = umap_out$layout[,2],
                     cluster = as.factor(kmeans_res$cluster))
```

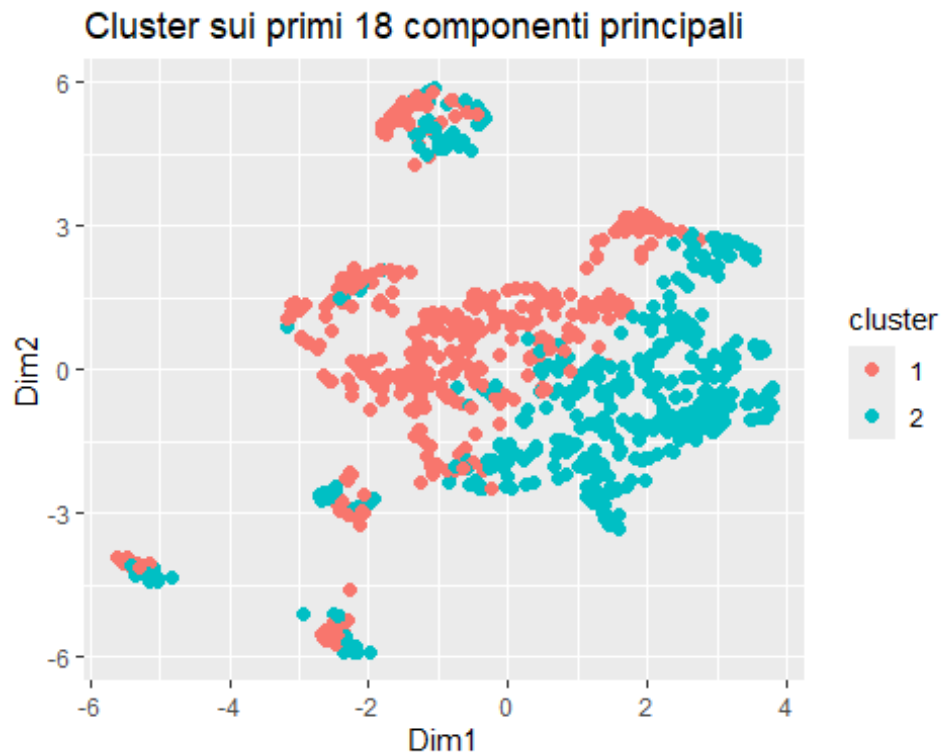
Visualizzazione

```
library(ggplot2)
```

```
ggplot(umap_df, aes(x = Dim1, y = Dim2, color = cluster)) +
```

```
  geom_point(size = 2) +
```

```
  labs(title = "Cluster sui primi 18 componenti principali")
```



```
library(cluster)
sil <- silhouette(kmeans_res$cluster, dist(pca_data))
mean(sil[, 3]) # silhouette score medio
## [1] 0.08753149
skim_without_charts(data)
```

Data summary

Name	data
Number of rows	916
Number of columns	29

Column type frequency:

factor	1
numeric	28

Group variables	None
-----------------	------

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
cluster	0	1	FALSE	2	2: 651, 1: 265

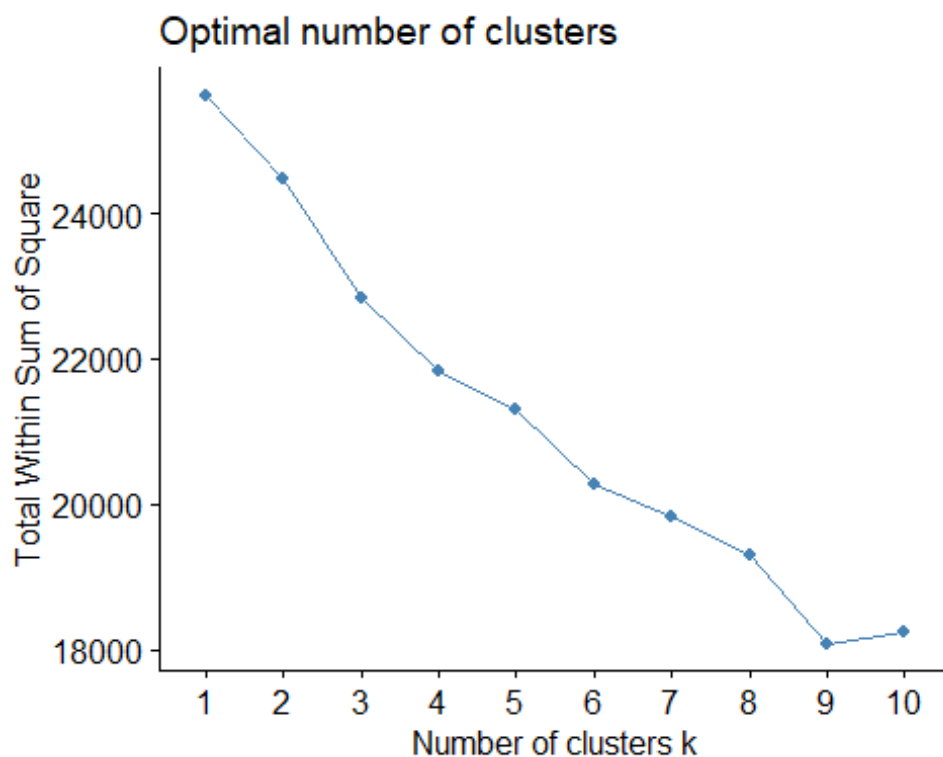
Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Age	0	1	65.23	13.62	18.00	55.00	65.00	77.00	95.00
Gender	0	1	1.38	0.49	1.00	1.00	1.00	2.00	2.00
ASA3	0	1	0.38	0.48	0.00	0.00	0.00	1.00	1.00
DM	0	1	0.21	0.41	0.00	0.00	0.00	0.00	1.00
CAD	0	1	0.07	0.26	0.00	0.00	0.00	0.00	1.00
HF	0	1	0.04	0.20	0.00	0.00	0.00	0.00	1.00
CVA	0	1	0.06	0.23	0.00	0.00	0.00	0.00	1.00
CKD	0	1	0.13	0.34	0.00	0.00	0.00	0.00	1.00
LogCEA	0	1	1.39	0.90	-0.36	0.60	1.26	1.94	4.18
Laparoscopic	0	1	0.04	0.20	0.00	0.00	0.00	0.00	1.00
TumorLOC	0	1	0.31	0.46	0.00	0.00	0.00	1.00	1.00
EA	0	1	0.16	0.37	0.00	0.00	0.00	0.00	1.00
Log2AT	0	1	8.34	0.48	5.64	8.08	8.30	8.61	9.91
RBC	0	1	0.53	0.68	0.00	0.00	0.00	1.00	2.00
Liver_Only	0	1	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Cell_diff	0	1	1.13	0.33	1.00	1.00	1.00	1.00	2.00
Mucin_TYPE	0	1	0.08	0.26	0.00	0.00	0.00	0.00	1.00
SignetRING	0	1	0.04	0.21	0.00	0.00	0.00	0.00	1.00
Lymphovascularinvasion	0	1	0.52	0.50	0.00	0.00	1.00	1.00	1.00
perineural	0	1	0.22	0.42	0.00	0.00	0.00	0.00	1.00
CT	0	1	0.90	0.30	0.00	1.00	1.00	1.00	1.00
RT	0	1	0.11	0.31	0.00	0.00	0.00	0.00	1.00
NACTRT	0	1	0.16	0.36	0.00	0.00	0.00	0.00	1.00
Death	0	1	0.42	0.49	0.00	0.00	0.00	1.00	1.00
Progress	0	1	0.79	0.41	0.00	1.00	1.00	1.00	1.00
IntervalOD	0	1	24.61	24.25	0.03	8.11	18.05	31.95	135.79
IntervalOR	0	1	11.32	17.10	0.03	3.05	5.98	12.62	134.21
AJCC_bin	0	1	0.42	0.49	0.00	0.00	0.00	1.00	1.00

K MEANS raw

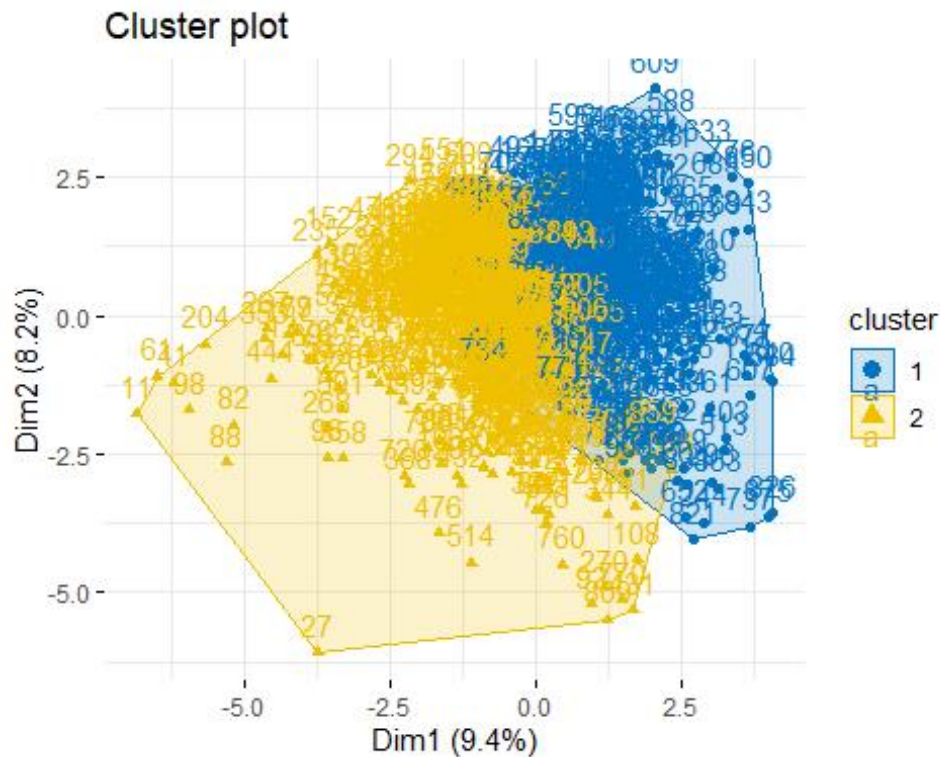
```
data$cluster <- NULL
data_scaled <- scale(data)
```

```
# Elbow plot per scegliere k
fviz_nbclust(data_scaled, kmeans, method = "wss")
```



```
set.seed(123)
kmeans_res <- kmeans(data_scaled, centers = 2, nstart = 25)

# Visualizzazione
fviz_cluster(kmeans_res, data = data_scaled,
  ellipse.type = "convex",
  palette = "jco",
  ggtheme = theme_minimal())
```



```
kmeans_res$centers
```

```
##          Age      Gender      ASA3      DM      CAD      HF
## 1  0.05192375  0.08058595  0.06292251  0.05520002 -0.04701169  0.05425701
## 2 -0.04396769 -0.06823810 -0.05328116 -0.04674196  0.03980828 -0.04594344
##          CVA      CKD      LogCEA  Laparoscopic  TumorLOC      EA
## 1 -0.02345529  0.06004137  0.09128604  0.0004222486  0.06273396 -0.04640629
## 2  0.01986133 -0.05084148 -0.07729866 -0.0003575492 -0.05312150  0.03929565
##          Log2AT      RBC  Liver_Only  Cell_diff  Mucin_TYPE  SignetRING
## 1 -0.01228065  0.07713773 -0.7889546  0.2163986  0.2016332  0.209462
## 2  0.01039893 -0.06531824  0.6680664 -0.1832408 -0.1707378 -0.177367
##  Lymphovascularinvasion perineural      CT      RT      NACTRT
## 1          0.2694111  0.1911987 -0.05410417 -0.007194450 -0.05617134
## 2          -0.2281304 -0.1619022  0.04581402  0.006092075  0.04756444
##          Death  Progress IntervalOD IntervalOR  AJCC_bin
## 1  0.2217476  0.1205294 -0.3623106 -0.2694170  0.9857446
## 2 -0.1877701 -0.1020612  0.3067953  0.2281353 -0.8347031
```

```
data$cluster <- as.factor(kmeans_res$cluster)
```

```
# Riassunto per cluster
```

```
aggregate(. ~ cluster, data = data, mean)
```

```
##   cluster      Age      Gender      ASA3      DM      CAD      HF
## 1      1  65.93333  1.419048  0.4071429  0.2309524  0.06190476  0.05476190
## 2      2  64.62702  1.346774  0.3508065  0.1895161  0.08467742  0.03427419
##          CVA      CKD      LogCEA  Laparoscopic  TumorLOC      EA  Log2AT
## 1  0.05238095  0.1547619  1.472213  0.04047619  0.3357143  0.1476190  8.334722
```

```
## 2 0.06250000 0.1169355 1.320587 0.04032258 0.2822581 0.1794355 8.345672
##      RBC Liver_Only Cell_diff Mucin_TYPE SignetRING Lymphovascularinvasion
## 1 0.5833333 0.004761905 1.200000 0.12857143 0.088095238 0.6500000
## 2 0.4858871 0.715725806 1.066532 0.03024194 0.008064516 0.4012097
## perineural CT RT NACTRT Death Progress IntervalOD
## 1 0.3047619 0.8809524 0.1047619 0.1357143 0.5309524 0.8404762 15.82094
## 2 0.1572581 0.9112903 0.1088710 0.1733871 0.3286290 0.7500000 32.04597
## IntervalOR AJCC_bin
## 1 6.712741 0.907142857
## 2 15.219646 0.008064516
```

```
summary(data$cluster)
```

```
## 1 2
## 420 496
```

```
library(dplyr)
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 4.3.3
```

```
##
## Attaching package: 'reshape2'
```

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

```
##
## smiths
```

```
library(ggplot2)
```

```
# Elenco delle variabili di interesse (includi qui quelle che vuoi nel grafico)
```

```
vars_of_interest <- c(
  "cluster", "Liver_Only", "Lymphovascularinvasion",
  "perineural", "Death", "EA", "CT", "RT", "NACTRT",
  "Age", "LogCEA", "IntervalOR", "IntervalOD", "Progress", "Cell_diff",
  "Mucin_TYPE"
)
```

```
# Seleziona solo queste variabili
```

```
data_subset <- data[, vars_of_interest]
```

```
# Calcola media per cluster
```

```
cluster_means <- data_subset %>%
  group_by(cluster) %>%
  summarise(across(everything(), mean, na.rm = TRUE)) %>%
  ungroup()
```

```
## Warning: There was 1 warning in `summarise()`.
```

```
## i In argument: `across(everything(), mean, na.rm = TRUE)`.
```

```
## i In group 1: `cluster = 1`.
```

```
## Caused by warning:
```

```
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
```

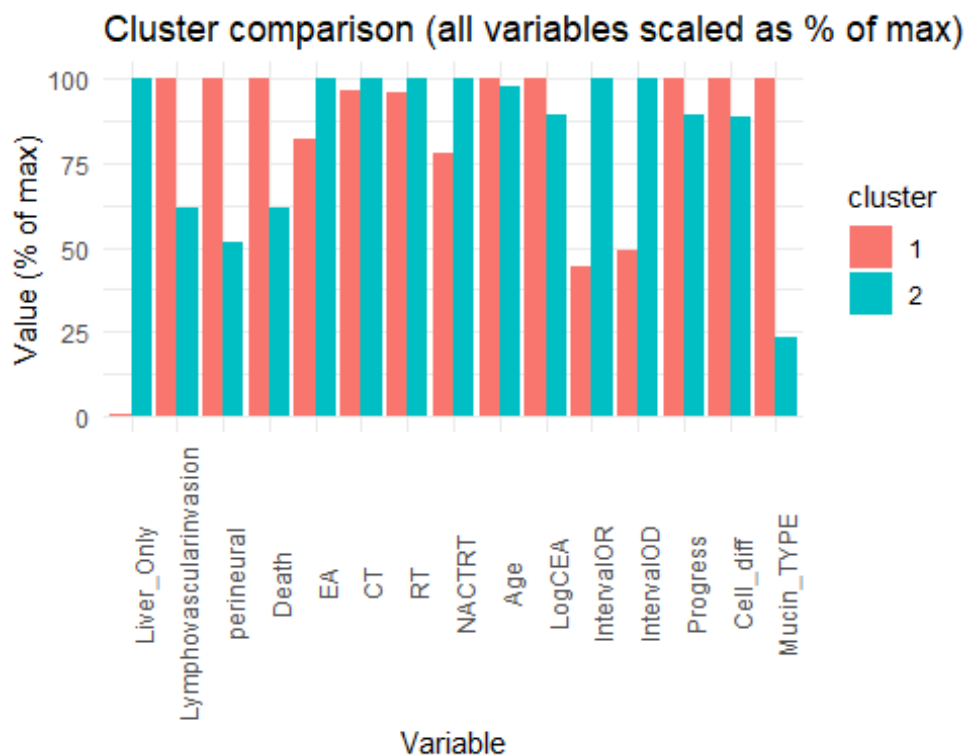
```
## Supply arguments directly to `.fns` through an anonymous function instead.
```

```
##
## # Previously
## across(a:b, mean, na.rm = TRUE)
##
## # Now
## across(a:b, \(x) mean(x, na.rm = TRUE))

# Normalizza ciascuna colonna (esclusa 'cluster') in percentuale rispetto al suo
massimo
cluster_means_scaled <- cluster_means
cluster_means_scaled[-1] <- lapply(cluster_means_scaled[-1], function(x) 100 * x /
max(x, na.rm = TRUE))

# Melting per ggplot
df_melt <- melt(cluster_means_scaled, id = "cluster")

# Grafico
ggplot(df_melt, aes(x = variable, y = value, fill = cluster)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 0.5)) +
  labs(
    title = "Cluster comparison (all variables scaled as % of max)",
    x = "Variable",
    y = "Value (% of max)"
  )
)
```



HAC

```
# Pacchetti
```

```
library(factoextra)
```

```
library(dplyr)
```

```
# Calcola distanza euclidea
```

```
dist_mat <- dist(data, method = "euclidean")
```

```
# Clustering gerarchico con metodo Ward
```

```
hc <- hclust(dist_mat, method = "ward.D2")
```

```
# Dendrogramma con taglio a k = 2
```

```
fviz_dend(hc, k = 2,  
  rect = TRUE,  
  rect_border = "jco",  
  rect_fill = TRUE,  
  show_labels = FALSE)
```

```
## Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none"  
## instead as
```

```
## of ggplot2 3.3.4.
```

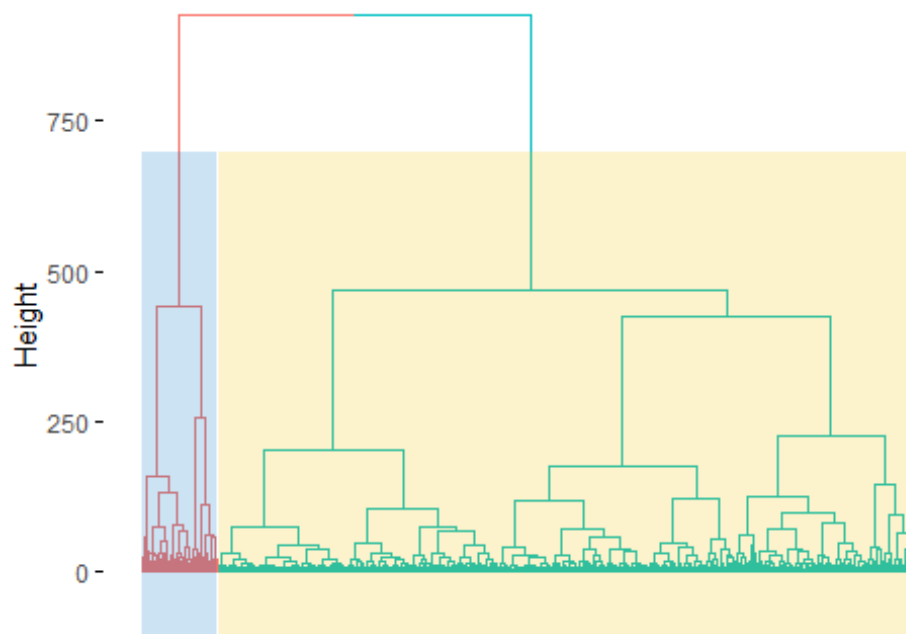
```
## i The deprecated feature was likely used in the factoextra package.
```

```
## Please report the issue at <https://github.com/kassambara/factoextra/issues>.
```

```
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was  
## generated.
```

Cluster Dendrogram




```
# Assegna i cluster al dataset
data$cluster_hc <- cutree(hc, k = 2)
```

METRICHE INTERNE per Kmeans e HAC

```
library(cluster)      # silhouette
```

```
library(clusterSim)   # Davies-Bouldin
```

```
## Warning: package 'clusterSim' was built under R version 4.3.3
```

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

```
##
```

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

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

```
##
```

```
##      select
```

```
library(factoextra)   # fviz_silhouette
```

```
library(dplyr)        # per operazioni su dataframe
```

```
# -----
```

```
# SILHOUETTE SCORE
```

```
# -----
```

```
# HAC
```

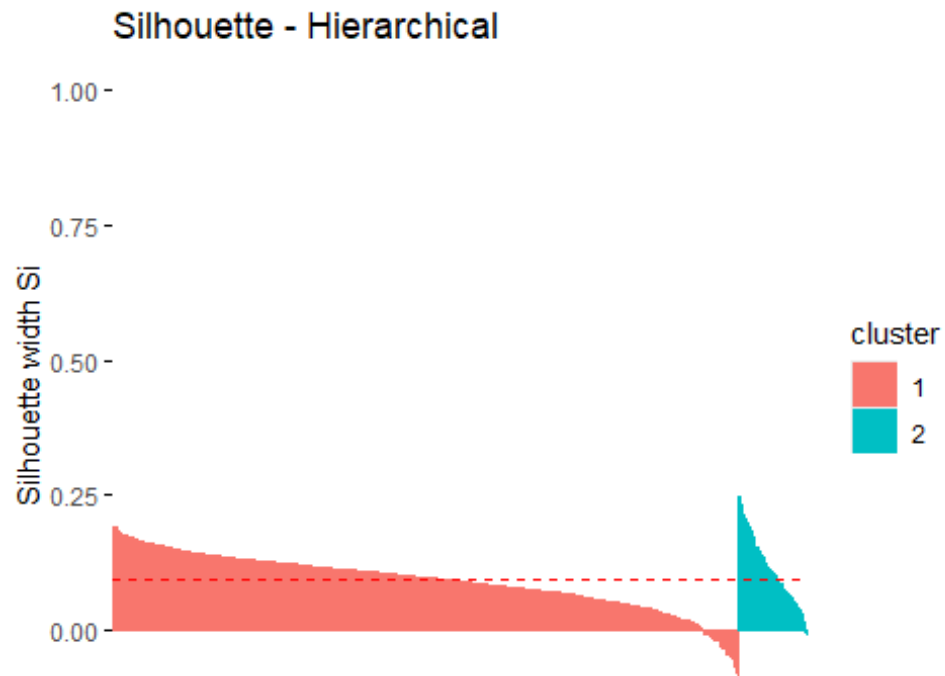
```
sil_hc <- silhouette(data$cluster_hc, dist(data_scaled))
```

```
fviz_silhouette(sil_hc) + ggtitle("Silhouette - Hierarchical")
```

```
##      cluster size ave.sil.width
```

```
## 1          1  826          0.09
```

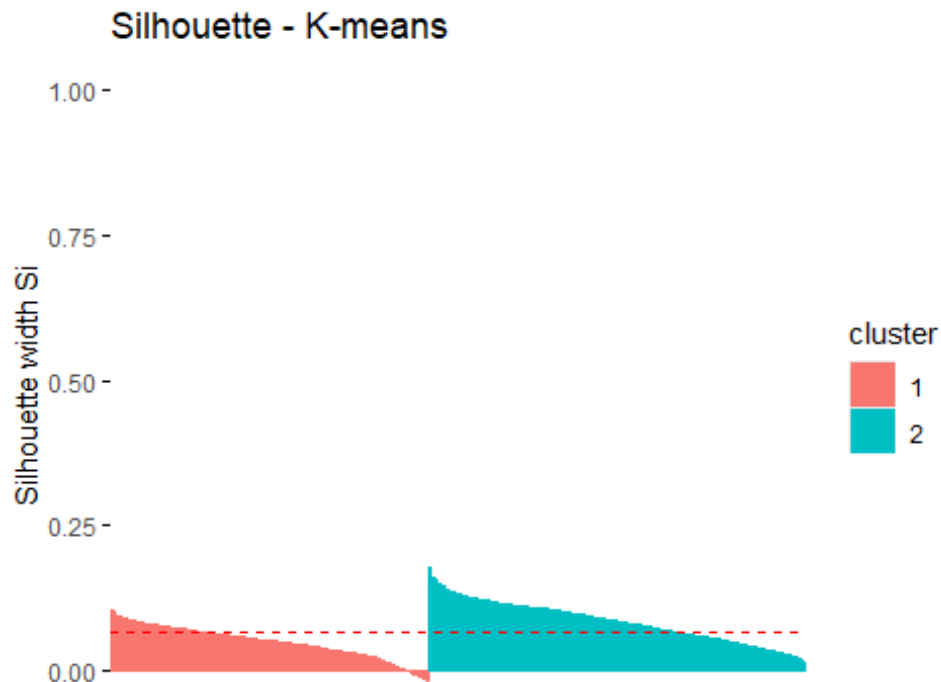
```
## 2          2   90          0.11
```



```
mean(sil_hc[, 3]) # Average silhouette for HAC
## [1] 0.09398352

# K-means
sil_km <- silhouette(kmeans_res$cluster, dist(data_scaled))
fviz_silhouette(sil_km) + ggtitle("Silhouette - K-means")

##   cluster size ave.sil.width
## 1      1  420         0.05
## 2      2  496         0.08
```



```
mean(sil_km[, 3]) # Average silhouette for K-means
## [1] 0.06879705

# -----
# DAVIES-BOULDIN INDEX
# -----

# HAC
db_hc <- index.DB(data_scaled, cl = data$cluster_hc, centrotypes = "centroids")$DB

# K-means
db_km <- index.DB(data_scaled, cl = kmeans_res$cluster, centrotypes =
"centroids")$DB

# -----
# RISULTATI
# -----

cat("Silhouette HAC: ", mean(sil_hc[,3]), "\n")
## Silhouette HAC: 0.09398352

cat("Silhouette K-means: ", mean(sil_km[,3]), "\n\n")
## Silhouette K-means: 0.06879705

cat("Davies-Bouldin HAC: ", db_hc, "\n")
```

```
## Davies-Bouldin HAC: 2.792395
cat("Davies-Bouldin K-means: ", db_km, "\n")
## Davies-Bouldin K-means: 3.777504
```

Si conferma che i cluster non sono ben separati nello spazio originale ad alta dimensionalità.

RISULTATI CLUSTERING:

Si conclude che il dataset non presenta strutture di cluster. Le metriche Silhouette e Davies-Bouldin indicano che i cluster non sono né ben separati né ben definiti, senza coesione interna. A prescindere dall'algoritmo scelto, si conclude l'analisi inconcludente. Si sottolinea anche, nonostante il valore accettabile di 0.5 della UMAP + k-means, UMAP deforma lo spazio, quindi non è detto che i cluster trovati riflettano correttamente la separazione dei dati originali. Il dataset non mostra una struttura di cluster ben definita nello spazio originale o nello spazio ridotto da PCA. Solo dopo proiezioni non lineari (UMAP), è possibile forzare una divisione con significato visivo, ma non necessariamente supportata da buona separabilità intrinseca.

```
data$cluster2 <- NULL
data$cluster_hc <- NULL
data$cluster <- NULL
```

LINEAR REGRESSION

```
summary(step(lm(Death ~., data = data)))

## Start: AIC=-1370.54
## Death ~ Age + Gender + ASA3 + DM + CAD + HF + CVA + CKD + LogCEA +
## Laparoscopic + TumorLOC + EA + Log2AT + RBC + Liver_Only +
## Cell_diff + Mucin_TYPE + SignetRING + Lymphovascularinvasion +
## perineural + CT + RT + NACTRT + Progress + IntervalOD + IntervalOR +
## AJCC_bin
##
##
```

	Df	Sum of Sq	RSS	AIC
## - Gender	1	0.0013	193.00	-1372.5
## - IntervalOR	1	0.0118	193.01	-1372.5
## - Log2AT	1	0.0194	193.01	-1372.4
## - perineural	1	0.0386	193.03	-1372.3
## - Cell_diff	1	0.0400	193.03	-1372.3
## - EA	1	0.0554	193.05	-1372.3
## - CAD	1	0.0603	193.06	-1372.2
## - Lymphovascularinvasion	1	0.0632	193.06	-1372.2
## - Laparoscopic	1	0.0762	193.07	-1372.2
## - HF	1	0.0765	193.07	-1372.2
## - SignetRING	1	0.0814	193.08	-1372.2
## - CT	1	0.0840	193.08	-1372.1
## - TumorLOC	1	0.0968	193.09	-1372.1
## - Liver_Only	1	0.1006	193.09	-1372.1
## - DM	1	0.1014	193.10	-1372.0
## - ASA3	1	0.1396	193.13	-1371.9

```

## - CVA 1 0.1527 193.15 -1371.8
## <none> 193.00 -1370.5
## - CKD 1 0.4523 193.45 -1370.4
## - Mucin_TYPE 1 0.6257 193.62 -1369.6
## - Age 1 0.6837 193.68 -1369.3
## - NACTRT 1 0.7449 193.74 -1369.0
## - RT 1 0.7650 193.76 -1368.9
## - AJCC_bin 1 1.0005 194.00 -1367.8
## - LogCEA 1 1.4824 194.48 -1365.5
## - RBC 1 1.5679 194.56 -1365.1
## - IntervalOD 1 3.3085 196.30 -1357.0
## - Progress 1 6.4634 199.46 -1342.4
##
## Step: AIC=-1372.53
## Death ~ Age + ASA3 + DM + CAD + HF + CVA + CKD + LogCEA + Laparoscopic +
## TumorLOC + EA + Log2AT + RBC + Liver_Only + Cell_diff + Mucin_TYPE +
## SignetRING + Lymphovascularinvasion + perineural + CT + RT +
## NACTRT + Progress + IntervalOD + IntervalOR + AJCC_bin
##
## Df Sum of Sq RSS AIC
## - IntervalOR 1 0.0119 193.01 -1374.5
## - Log2AT 1 0.0201 193.02 -1374.4
## - perineural 1 0.0378 193.03 -1374.3
## - Cell_diff 1 0.0412 193.04 -1374.3
## - EA 1 0.0545 193.05 -1374.3
## - CAD 1 0.0599 193.06 -1374.2
## - Lymphovascularinvasion 1 0.0635 193.06 -1374.2
## - HF 1 0.0765 193.07 -1374.2
## - Laparoscopic 1 0.0769 193.07 -1374.2
## - SignetRING 1 0.0828 193.08 -1374.1
## - CT 1 0.0833 193.08 -1374.1
## - TumorLOC 1 0.0955 193.09 -1374.1
## - DM 1 0.1018 193.10 -1374.0
## - Liver_Only 1 0.1041 193.10 -1374.0
## - ASA3 1 0.1408 193.14 -1373.9
## - CVA 1 0.1552 193.15 -1373.8
## <none> 193.00 -1372.5
## - CKD 1 0.4574 193.45 -1372.4
## - Mucin_TYPE 1 0.6246 193.62 -1371.6
## - Age 1 0.6835 193.68 -1371.3
## - NACTRT 1 0.7438 193.74 -1371.0
## - RT 1 0.7649 193.76 -1370.9
## - AJCC_bin 1 1.0017 194.00 -1369.8
## - LogCEA 1 1.4836 194.48 -1367.5
## - RBC 1 1.5692 194.56 -1367.1
## - IntervalOD 1 3.3100 196.31 -1359.0
## - Progress 1 6.4799 199.48 -1344.3
##
## Step: AIC=-1374.47
## Death ~ Age + ASA3 + DM + CAD + HF + CVA + CKD + LogCEA + Laparoscopic +
## TumorLOC + EA + Log2AT + RBC + Liver_Only + Cell_diff + Mucin_TYPE +

```

```

##      SignetRING + Lymphovascularinvasion + perineural + CT + RT +
##      NACTRT + Progress + IntervalOD + AJCC_bin
##
##
##      Df Sum of Sq    RSS      AIC
## - Log2AT      1    0.0219 193.03 -1376.4
## - perineural   1    0.0396 193.05 -1376.3
## - Cell_diff    1    0.0435 193.05 -1376.3
## - EA           1    0.0559 193.06 -1376.2
## - CAD          1    0.0612 193.07 -1376.2
## - Lymphovascularinvasion 1    0.0615 193.07 -1376.2
## - HF          1    0.0755 193.08 -1376.1
## - Laparoscopic 1    0.0763 193.08 -1376.1
## - SignetRING   1    0.0844 193.09 -1376.1
## - CT          1    0.0905 193.10 -1376.0
## - TumorLOC     1    0.0968 193.10 -1376.0
## - DM          1    0.1035 193.11 -1376.0
## - Liver_Only   1    0.1062 193.11 -1376.0
## - ASA3         1    0.1410 193.15 -1375.8
## - CVA          1    0.1517 193.16 -1375.8
## <none>                193.01 -1374.5
## - CKD          1    0.4523 193.46 -1374.3
## - Mucin_TYPE   1    0.6196 193.63 -1373.5
## - Age          1    0.6777 193.69 -1373.3
## - NACTRT       1    0.7459 193.75 -1372.9
## - RT          1    0.7559 193.76 -1372.9
## - AJCC_bin     1    1.0052 194.01 -1371.7
## - LogCEA       1    1.4850 194.49 -1369.5
## - RBC          1    1.5757 194.58 -1369.0
## - IntervalOD   1    5.4653 198.47 -1350.9
## - Progress     1    7.8893 200.90 -1339.8
##
## Step:  AIC=-1376.37
## Death ~ Age + ASA3 + DM + CAD + HF + CVA + CKD + LogCEA + Laparoscopic +
##      TumorLOC + EA + RBC + Liver_Only + Cell_diff + Mucin_TYPE +
##      SignetRING + Lymphovascularinvasion + perineural + CT + RT +
##      NACTRT + Progress + IntervalOD + AJCC_bin
##
##
##      Df Sum of Sq    RSS      AIC
## - perineural   1    0.0392 193.07 -1378.2
## - Cell_diff    1    0.0446 193.07 -1378.2
## - EA           1    0.0575 193.09 -1378.1
## - Lymphovascularinvasion 1    0.0617 193.09 -1378.1
## - CAD          1    0.0617 193.09 -1378.1
## - HF          1    0.0735 193.10 -1378.0
## - SignetRING   1    0.0787 193.11 -1378.0
## - Laparoscopic 1    0.0792 193.11 -1378.0
## - TumorLOC     1    0.0865 193.12 -1378.0
## - CT          1    0.0967 193.13 -1377.9
## - DM          1    0.1037 193.13 -1377.9
## - Liver_Only   1    0.1146 193.15 -1377.8
## - ASA3         1    0.1384 193.17 -1377.7

```

```

## - CVA 1 0.1578 193.19 -1377.6
## <none> 193.03 -1376.4
## - CKD 1 0.4510 193.48 -1376.2
## - Mucin_TYPE 1 0.6196 193.65 -1375.4
## - Age 1 0.6963 193.73 -1375.1
## - NACTRT 1 0.7263 193.76 -1374.9
## - RT 1 0.7662 193.80 -1374.7
## - AJCC_bin 1 1.0284 194.06 -1373.5
## - LogCEA 1 1.4639 194.49 -1371.5
## - RBC 1 1.8061 194.84 -1369.8
## - IntervalOD 1 5.4453 198.47 -1352.9
## - Progress 1 7.8791 200.91 -1341.7
##
## Step: AIC=-1378.18
## Death ~ Age + ASA3 + DM + CAD + HF + CVA + CKD + LogCEA + Laparoscopic +
## TumorLOC + EA + RBC + Liver_Only + Cell_diff + Mucin_TYPE +
## SignetRING + Lymphovascularinvasion + CT + RT + NACTRT +
## Progress + IntervalOD + AJCC_bin
##
## Df Sum of Sq RSS AIC
## - Cell_diff 1 0.0429 193.11 -1380.0
## - EA 1 0.0528 193.12 -1379.9
## - CAD 1 0.0599 193.13 -1379.9
## - SignetRING 1 0.0718 193.14 -1379.8
## - HF 1 0.0731 193.14 -1379.8
## - Laparoscopic 1 0.0836 193.15 -1379.8
## - TumorLOC 1 0.0952 193.16 -1379.7
## - CT 1 0.0978 193.17 -1379.7
## - DM 1 0.1041 193.17 -1379.7
## - Lymphovascularinvasion 1 0.1055 193.18 -1379.7
## - Liver_Only 1 0.1107 193.18 -1379.7
## - ASA3 1 0.1307 193.20 -1379.6
## - CVA 1 0.1656 193.24 -1379.4
## <none> 193.07 -1378.2
## - CKD 1 0.4488 193.52 -1378.1
## - Mucin_TYPE 1 0.6135 193.68 -1377.3
## - Age 1 0.6806 193.75 -1377.0
## - NACTRT 1 0.7344 193.80 -1376.7
## - RT 1 0.7804 193.85 -1376.5
## - AJCC_bin 1 1.0073 194.08 -1375.4
## - LogCEA 1 1.4450 194.51 -1373.3
## - RBC 1 1.8170 194.89 -1371.6
## - IntervalOD 1 5.4208 198.49 -1354.8
## - Progress 1 7.8583 200.93 -1343.6
##
## Step: AIC=-1379.98
## Death ~ Age + ASA3 + DM + CAD + HF + CVA + CKD + LogCEA + Laparoscopic +
## TumorLOC + EA + RBC + Liver_Only + Mucin_TYPE + SignetRING +
## Lymphovascularinvasion + CT + RT + NACTRT + Progress + IntervalOD +
## AJCC_bin
##

```

```

##          Df Sum of Sq    RSS    AIC
## - EA          1      0.0512 193.16 -1381.7
## - CAD          1      0.0590 193.17 -1381.7
## - HF          1      0.0713 193.18 -1381.6
## - Laparoscopic 1      0.0849 193.20 -1381.6
## - Lymphovascularinvasion 1      0.0889 193.20 -1381.6
## - CT          1      0.0927 193.21 -1381.5
## - TumorLOC     1      0.1001 193.21 -1381.5
## - SignetRING   1      0.1040 193.22 -1381.5
## - DM          1      0.1070 193.22 -1381.5
## - Liver_Only   1      0.1098 193.22 -1381.5
## - ASA3         1      0.1266 193.24 -1381.4
## - CVA          1      0.1587 193.27 -1381.2
## <none>                193.11 -1380.0
## - CKD          1      0.4479 193.56 -1379.9
## - Mucin_TYPE   1      0.6106 193.72 -1379.1
## - Age          1      0.6860 193.80 -1378.7
## - NACTRT       1      0.7260 193.84 -1378.5
## - RT          1      0.7932 193.91 -1378.2
## - AJCC_bin     1      1.0288 194.14 -1377.1
## - LogCEA       1      1.4049 194.52 -1375.3
## - RBC          1      1.8417 194.95 -1373.3
## - IntervalOD   1      5.5669 198.68 -1356.0
## - Progress     1      7.8463 200.96 -1345.5
##
## Step:  AIC=-1381.74
## Death ~ Age + ASA3 + DM + CAD + HF + CVA + CKD + LogCEA + Laparoscopic +
##      TumorLOC + RBC + Liver_Only + Mucin_TYPE + SignetRING +
Lymphovascularinvasion +
##      CT + RT + NACTRT + Progress + IntervalOD + AJCC_bin
##
##          Df Sum of Sq    RSS    AIC
## - CAD          1      0.0584 193.22 -1383.5
## - HF          1      0.0677 193.23 -1383.4
## - Lymphovascularinvasion 1      0.0819 193.25 -1383.3
## - Laparoscopic 1      0.0831 193.25 -1383.3
## - CT          1      0.0906 193.25 -1383.3
## - Liver_Only   1      0.1019 193.26 -1383.2
## - TumorLOC     1      0.1036 193.27 -1383.2
## - SignetRING   1      0.1047 193.27 -1383.2
## - DM          1      0.1113 193.28 -1383.2
## - ASA3         1      0.1251 193.29 -1383.1
## - CVA          1      0.1589 193.32 -1383.0
## <none>                193.16 -1381.7
## - CKD          1      0.4642 193.63 -1381.5
## - Mucin_TYPE   1      0.6001 193.76 -1380.9
## - Age          1      0.6836 193.85 -1380.5
## - NACTRT       1      0.6966 193.86 -1380.4
## - RT          1      0.7763 193.94 -1380.1
## - AJCC_bin     1      1.0084 194.17 -1379.0
## - LogCEA       1      1.4123 194.58 -1377.1

```



```

## - RBC          1      1.8461 195.01 -1375.0
## - IntervalOD   1      5.6857 198.85 -1357.2
## - Progress     1      7.8502 201.01 -1347.2
##
## Step:  AIC=-1383.46
## Death ~ Age + ASA3 + DM + HF + CVA + CKD + LogCEA + Laparoscopic +
##      TumorLOC + RBC + Liver_Only + Mucin_TYPE + SignetRING +
Lymphovascularinvasion +
##      CT + RT + NACTRT + Progress + IntervalOD + AJCC_bin
##
##              Df Sum of Sq    RSS    AIC
## - HF          1      0.0490 193.27 -1385.2
## - CT          1      0.0790 193.30 -1385.1
## - Lymphovascularinvasion 1      0.0795 193.30 -1385.1
## - Laparoscopic 1      0.0802 193.30 -1385.1
## - DM          1      0.0968 193.32 -1385.0
## - SignetRING   1      0.0974 193.32 -1385.0
## - Liver_Only   1      0.0983 193.32 -1385.0
## - TumorLOC     1      0.1024 193.32 -1385.0
## - ASA3         1      0.1072 193.33 -1385.0
## - CVA          1      0.1565 193.38 -1384.7
## <none>                193.22 -1383.5
## - CKD          1      0.4590 193.68 -1383.3
## - Mucin_TYPE   1      0.5942 193.82 -1382.7
## - NACTRT       1      0.7056 193.93 -1382.1
## - Age          1      0.7270 193.95 -1382.0
## - RT          1      0.7959 194.02 -1381.7
## - AJCC_bin     1      1.0170 194.24 -1380.7
## - LogCEA       1      1.4222 194.64 -1378.7
## - RBC          1      1.8625 195.08 -1376.7
## - IntervalOD   1      5.7334 198.96 -1358.7
## - Progress     1      7.8586 201.08 -1348.9
##
## Step:  AIC=-1385.23
## Death ~ Age + ASA3 + DM + CVA + CKD + LogCEA + Laparoscopic +
##      TumorLOC + RBC + Liver_Only + Mucin_TYPE + SignetRING +
Lymphovascularinvasion +
##      CT + RT + NACTRT + Progress + IntervalOD + AJCC_bin
##
##              Df Sum of Sq    RSS    AIC
## - CT          1      0.0699 193.34 -1386.9
## - Lymphovascularinvasion 1      0.0796 193.35 -1386.8
## - Laparoscopic 1      0.0819 193.35 -1386.8
## - Liver_Only   1      0.0919 193.36 -1386.8
## - DM          1      0.0986 193.37 -1386.8
## - SignetRING   1      0.1005 193.37 -1386.8
## - TumorLOC     1      0.1018 193.37 -1386.7
## - ASA3         1      0.1277 193.40 -1386.6
## - CVA          1      0.1486 193.42 -1386.5
## <none>                193.27 -1385.2
## - CKD          1      0.4737 193.74 -1385.0

```

```

## - Mucin_TYPE          1      0.6034 193.87 -1384.4
## - Age                  1      0.7040 193.97 -1383.9
## - NACTRT               1      0.7238 193.99 -1383.8
## - RT                   1      0.7899 194.06 -1383.5
## - AJCC_bin             1      1.0045 194.28 -1382.5
## - LogCEA               1      1.3969 194.67 -1380.6
## - RBC                  1      1.8651 195.14 -1378.4
## - IntervalOD           1      5.7614 199.03 -1360.3
## - Progress             1      7.9156 201.19 -1350.5
##
## Step:  AIC=-1386.9
## Death ~ Age + ASA3 + DM + CVA + CKD + LogCEA + Laparoscopic +
##         TumorLOC + RBC + Liver_Only + Mucin_TYPE + SignetRING +
Lymphovascularinvasion +
##         RT + NACTRT + Progress + IntervalOD + AJCC_bin
##
##              Df Sum of Sq    RSS    AIC
## - Lymphovascularinvasion  1      0.0760 193.42 -1388.5
## - Laparoscopic            1      0.0796 193.42 -1388.5
## - Liver_Only              1      0.0991 193.44 -1388.4
## - DM                      1      0.1004 193.44 -1388.4
## - TumorLOC                1      0.1006 193.44 -1388.4
## - SignetRING              1      0.1032 193.44 -1388.4
## - ASA3                    1      0.1181 193.46 -1388.3
## - CVA                     1      0.1319 193.47 -1388.3
## <none>                     193.34 -1386.9
## - CKD                     1      0.4480 193.79 -1386.8
## - Mucin_TYPE              1      0.5814 193.92 -1386.1
## - NACTRT                  1      0.7386 194.08 -1385.4
## - Age                     1      0.7614 194.10 -1385.3
## - RT                      1      0.7953 194.14 -1385.1
## - AJCC_bin                1      1.0085 194.35 -1384.1
## - LogCEA                  1      1.3761 194.72 -1382.4
## - RBC                     1      1.8307 195.17 -1380.3
## - IntervalOD              1      5.7163 199.06 -1362.2
## - Progress                 1      9.8422 203.18 -1343.4
##
## Step:  AIC=-1388.54
## Death ~ Age + ASA3 + DM + CVA + CKD + LogCEA + Laparoscopic +
##         TumorLOC + RBC + Liver_Only + Mucin_TYPE + SignetRING + RT +
##         NACTRT + Progress + IntervalOD + AJCC_bin
##
##              Df Sum of Sq    RSS    AIC
## - SignetRING              1      0.0813 193.50 -1390.2
## - Laparoscopic            1      0.0839 193.50 -1390.1
## - TumorLOC                1      0.0955 193.51 -1390.1
## - Liver_Only              1      0.0958 193.51 -1390.1
## - DM                      1      0.1051 193.52 -1390.0
## - ASA3                    1      0.1179 193.53 -1390.0
## - CVA                     1      0.1264 193.54 -1389.9
## <none>                     193.42 -1388.5

```

```

## - CKD      1      0.4450 193.86 -1388.4
## - Mucin_TYPE 1      0.5601 193.98 -1387.9
## - NACTRT    1      0.7060 194.12 -1387.2
## - Age       1      0.7342 194.15 -1387.1
## - RT        1      0.7894 194.21 -1386.8
## - AJCC_bin  1      0.9620 194.38 -1386.0
## - LogCEA    1      1.3972 194.81 -1383.9
## - RBC       1      1.8515 195.27 -1381.8
## - IntervalOD 1      5.6622 199.08 -1364.1
## - Progress  1      9.7663 203.18 -1345.4
##
## Step: AIC=-1390.15
## Death ~ Age + ASA3 + DM + CVA + CKD + LogCEA + Laparoscopic +
##      TumorLOC + RBC + Liver_Only + Mucin_TYPE + RT + NACTRT +
##      Progress + IntervalOD + AJCC_bin
##
##              Df Sum of Sq    RSS    AIC
## - Laparoscopic 1      0.0727 193.57 -1391.8
## - Liver_Only   1      0.0945 193.59 -1391.7
## - DM           1      0.0986 193.60 -1391.7
## - ASA3         1      0.1101 193.61 -1391.6
## - TumorLOC     1      0.1114 193.61 -1391.6
## - CVA          1      0.1338 193.63 -1391.5
## <none>                193.50 -1390.2
## - CKD          1      0.4416 193.94 -1390.1
## - Mucin_TYPE   1      0.4866 193.99 -1389.8
## - NACTRT       1      0.7235 194.22 -1388.7
## - Age          1      0.7691 194.27 -1388.5
## - RT           1      0.7698 194.27 -1388.5
## - AJCC_bin     1      1.0127 194.51 -1387.4
## - LogCEA       1      1.3692 194.87 -1385.7
## - RBC          1      1.8780 195.38 -1383.3
## - IntervalOD   1      5.7015 199.20 -1365.5
## - Progress     1      9.7176 203.22 -1347.3
##
## Step: AIC=-1391.81
## Death ~ Age + ASA3 + DM + CVA + CKD + LogCEA + TumorLOC + RBC +
##      Liver_Only + Mucin_TYPE + RT + NACTRT + Progress + IntervalOD +
##      AJCC_bin
##
##              Df Sum of Sq    RSS    AIC
## - Liver_Only   1      0.0868 193.66 -1393.4
## - DM           1      0.0920 193.66 -1393.4
## - TumorLOC     1      0.1097 193.68 -1393.3
## - ASA3         1      0.1199 193.69 -1393.2
## - CVA          1      0.1393 193.71 -1393.2
## <none>                193.57 -1391.8
## - CKD          1      0.4387 194.01 -1391.7
## - Mucin_TYPE   1      0.4935 194.06 -1391.5
## - NACTRT       1      0.7087 194.28 -1390.5
## - RT           1      0.7775 194.35 -1390.1

```

```

## - Age      1      0.7869 194.36 -1390.1
## - AJCC_bin 1      1.0227 194.59 -1389.0
## - LogCEA   1      1.3570 194.93 -1387.4
## - RBC      1      1.8875 195.46 -1384.9
## - IntervalOD 1      5.6597 199.23 -1367.4
## - Progress 1      9.6998 203.27 -1349.0
##
## Step: AIC=-1393.4
## Death ~ Age + ASA3 + DM + CVA + CKD + LogCEA + TumorLOC + RBC +
## Mucin_TYPE + RT + NACTRT + Progress + IntervalOD + AJCC_bin
##
##           Df Sum of Sq    RSS    AIC
## - DM      1      0.0819 193.74 -1395.0
## - TumorLOC 1      0.1025 193.76 -1394.9
## - ASA3    1      0.1117 193.77 -1394.9
## - CVA     1      0.1348 193.79 -1394.8
## <none>                193.66 -1393.4
## - CKD     1      0.4257 194.08 -1393.4
## - Mucin_TYPE 1      0.5130 194.17 -1393.0
## - NACTRT  1      0.6964 194.35 -1392.1
## - RT      1      0.7588 194.42 -1391.8
## - Age     1      0.7770 194.43 -1391.7
## - AJCC_bin 1      1.2260 194.88 -1389.6
## - LogCEA  1      1.5229 195.18 -1388.2
## - RBC     1      1.9763 195.63 -1386.1
## - IntervalOD 1      5.7744 199.43 -1368.5
## - Progress 1      9.6962 203.35 -1350.6
##
## Step: AIC=-1395.01
## Death ~ Age + ASA3 + CVA + CKD + LogCEA + TumorLOC + RBC + Mucin_TYPE +
## RT + NACTRT + Progress + IntervalOD + AJCC_bin
##
##           Df Sum of Sq    RSS    AIC
## - TumorLOC 1      0.1046 193.84 -1396.5
## - ASA3     1      0.1340 193.87 -1396.4
## - CVA      1      0.1392 193.88 -1396.3
## <none>                193.74 -1395.0
## - CKD     1      0.4282 194.17 -1395.0
## - Mucin_TYPE 1      0.5261 194.26 -1394.5
## - NACTRT  1      0.6858 194.43 -1393.8
## - Age     1      0.7455 194.49 -1393.5
## - RT      1      0.7529 194.49 -1393.5
## - AJCC_bin 1      1.2516 194.99 -1391.1
## - LogCEA  1      1.5388 195.28 -1389.8
## - RBC     1      2.0011 195.74 -1387.6
## - IntervalOD 1      5.7212 199.46 -1370.3
## - Progress 1      9.6486 203.39 -1352.5
##
## Step: AIC=-1396.51
## Death ~ Age + ASA3 + CVA + CKD + LogCEA + RBC + Mucin_TYPE +
## RT + NACTRT + Progress + IntervalOD + AJCC_bin

```

```

##
##           Df Sum of Sq    RSS      AIC
## - ASA3      1    0.1439 193.99 -1397.8
## - CVA        1    0.1446 193.99 -1397.8
## - CKD        1    0.4128 194.26 -1396.6
## <none>                193.84 -1396.5
## - Mucin_TYPE 1    0.4984 194.34 -1396.2
## - Age        1    0.7196 194.56 -1395.1
## - RT         1    0.7692 194.61 -1394.9
## - NACTRT     1    0.7779 194.62 -1394.8
## - AJCC_bin   1    1.2644 195.11 -1392.6
## - LogCEA     1    1.4954 195.34 -1391.5
## - RBC        1    2.1326 195.98 -1388.5
## - IntervalOD 1    5.7766 199.62 -1371.6
## - Progress   1    9.5535 203.40 -1354.5
##
## Step:  AIC=-1397.83
## Death ~ Age + CVA + CKD + LogCEA + RBC + Mucin_TYPE + RT + NACTRT +
##         Progress + IntervalOD + AJCC_bin
##
##           Df Sum of Sq    RSS      AIC
## - CVA        1    0.1911 194.18 -1398.9
## <none>                193.99 -1397.8
## - CKD        1    0.4502 194.44 -1397.7
## - Mucin_TYPE 1    0.4901 194.48 -1397.5
## - Age        1    0.5881 194.58 -1397.1
## - RT         1    0.7341 194.72 -1396.4
## - NACTRT     1    0.8123 194.80 -1396.0
## - AJCC_bin   1    1.2499 195.24 -1394.0
## - LogCEA     1    1.5458 195.53 -1392.6
## - RBC        1    2.3701 196.36 -1388.7
## - IntervalOD 1    5.9881 199.98 -1372.0
## - Progress   1    9.5871 203.57 -1355.7
##
## Step:  AIC=-1398.93
## Death ~ Age + CKD + LogCEA + RBC + Mucin_TYPE + RT + NACTRT +
##         Progress + IntervalOD + AJCC_bin
##
##           Df Sum of Sq    RSS      AIC
## <none>                194.18 -1398.9
## - Mucin_TYPE 1    0.4979 194.68 -1398.6
## - Age        1    0.5203 194.70 -1398.5
## - CKD        1    0.5260 194.71 -1398.5
## - RT         1    0.6989 194.88 -1397.6
## - NACTRT     1    0.8683 195.05 -1396.8
## - AJCC_bin   1    1.2217 195.40 -1395.2
## - LogCEA     1    1.5209 195.70 -1393.8
## - RBC        1    2.3623 196.54 -1389.9
## - IntervalOD 1    5.9644 200.14 -1373.2
## - Progress   1    9.5579 203.74 -1356.9

```

```
##
## Call:
## lm(formula = Death ~ Age + CKD + LogCEA + RBC + Mucin_TYPE +
##      RT + NACTRT + Progress + IntervalOD + AJCC_bin, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7706 -0.4167 -0.1564  0.4701  0.9464
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.2888106  0.0975717   2.960 0.003157 **
## Age         -0.0018497  0.0011878  -1.557 0.119774
## CKD          0.0726880  0.0464258   1.566 0.117773
## LogCEA       0.0470387  0.0176678   2.662 0.007896 **
## RBC          0.0763361  0.0230060   3.318 0.000942 ***
## Mucin_TYPE  -0.0899301  0.0590375  -1.523 0.128040
## RT           0.0910109  0.0504252   1.805 0.071428 .
## NACTRT       -0.0869646  0.0432289  -2.012 0.044545 *
## Progress     0.2581485  0.0386781   6.674 4.33e-11 ***
## IntervalOD  -0.0036237  0.0006873  -5.272 1.68e-07 ***
## AJCC_bin     0.0786215  0.0329491   2.386 0.017230 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4632 on 905 degrees of freedom
## Multiple R-squared:  0.1306, Adjusted R-squared:  0.121
## F-statistic: 13.59 on 10 and 905 DF, p-value: < 2.2e-16
```

R^2 0.136, basso. solo il 13.6 percento della varianza di morte é spiegato dalle variabili indipendenti. AIC buono, piú basso meglio é. F statistics: 13.59 con pvalue nullo, il modello complessivo é statisticamente molto significativo. con 905 dof il modello é robusto.

INTERPRETAZIONE COEFFICIENTI Intercept 0.289 ** Valore atteso di Death quando tutti i predittori sono a zero. LogCEA 0.047 ** Aumento del logaritmo del CEA (log10): associato ad aumento del rischio di morte. RBC 0.076 *** Più unità trasfuse, maggiore è la probabilità di morte. NACTRT -0.087 * Aver fatto chemio/radioterapia neoadiuvante è associato a minor rischio di morte. Progress 0.258 *** Malattia in progressione: fortemente associata a mortalità più alta. IntervalOD -0.0036 *** Intervallo tra intervento e morte: più lungo = rischio minore. AJCC_bin 0.0786 * Stadio AJCC binarizzato: stadio più avanzato → rischio maggiore.

LOGISTIC REGRESSION

```
# GLM Logistica binaria
glm_model <- glm(
  Death ~ .,
  data = data,
  family = binomial(link = "logit")
)

final <- (step(glm_model))
```

```
## Start: AIC=1159.51
## Death ~ Age + Gender + ASA3 + DM + CAD + HF + CVA + CKD + LogCEA +
## Laparoscopic + TumorLOC + EA + Log2AT + RBC + Liver_Only +
## Cell_diff + Mucin_TYPE + SignetRING + Lymphovascularinvasion +
## perineural + CT + RT + NACTRT + Progress + IntervalOD + IntervalOR +
## AJCC_bin
```

	Df	Deviance	AIC
## - Gender	1	1103.5	1157.5
## - Cell_diff	1	1103.6	1157.6
## - Log2AT	1	1103.7	1157.7
## - perineural	1	1103.7	1157.7
## - Liver_Only	1	1103.8	1157.8
## - Laparoscopic	1	1103.8	1157.8
## - CAD	1	1103.8	1157.8
## - TumorLOC	1	1103.8	1157.8
## - HF	1	1103.9	1157.9
## - EA	1	1103.9	1157.9
## - SignetRING	1	1104.0	1158.0
## - CT	1	1104.0	1158.0
## - DM	1	1104.0	1158.0
## - Lymphovascularinvasion	1	1104.0	1158.0
## - CVA	1	1104.1	1158.1
## - ASA3	1	1104.1	1158.1
## - CKD	1	1105.5	1159.5
## <none>		1103.5	1159.5
## - IntervalOR	1	1106.4	1160.4
## - Mucin_TYPE	1	1106.4	1160.4
## - Age	1	1106.5	1160.5
## - AJCC_bin	1	1106.8	1160.8
## - RT	1	1107.0	1161.0
## - NACTRT	1	1107.2	1161.2
## - LogCEA	1	1109.6	1163.6
## - RBC	1	1109.9	1163.9
## - IntervalOD	1	1117.5	1171.5
## - Progress	1	1134.7	1188.7

```
## Step: AIC=1157.51
## Death ~ Age + ASA3 + DM + CAD + HF + CVA + CKD + LogCEA + Laparoscopic +
## TumorLOC + EA + Log2AT + RBC + Liver_Only + Cell_diff + Mucin_TYPE +
## SignetRING + Lymphovascularinvasion + perineural + CT + RT +
## NACTRT + Progress + IntervalOD + IntervalOR + AJCC_bin
```

	Df	Deviance	AIC
## - Cell_diff	1	1103.6	1155.6
## - Log2AT	1	1103.7	1155.7
## - perineural	1	1103.7	1155.7
## - Liver_Only	1	1103.8	1155.8
## - CAD	1	1103.8	1155.8
## - Laparoscopic	1	1103.8	1155.8
## - TumorLOC	1	1103.8	1155.8

```

## - HF 1 1103.9 1155.9
## - EA 1 1103.9 1155.9
## - SignetRING 1 1104.0 1156.0
## - CT 1 1104.0 1156.0
## - DM 1 1104.0 1156.0
## - Lymphovascularinvasion 1 1104.0 1156.0
## - CVA 1 1104.2 1156.2
## - ASA3 1 1104.2 1156.2
## <none> 1103.5 1157.5
## - CKD 1 1105.5 1157.5
## - IntervalOR 1 1106.4 1158.4
## - Mucin_TYPE 1 1106.4 1158.4
## - Age 1 1106.5 1158.5
## - AJCC_bin 1 1106.8 1158.8
## - RT 1 1107.0 1159.0
## - NACTRT 1 1107.2 1159.2
## - LogCEA 1 1109.6 1161.6
## - RBC 1 1109.9 1161.9
## - IntervalOD 1 1117.5 1169.5
## - Progress 1 1134.8 1186.8
##
## Step: AIC=1155.63
## Death ~ Age + ASA3 + DM + CAD + HF + CVA + CKD + LogCEA + Laparoscopic +
## TumorLOC + EA + Log2AT + RBC + Liver_Only + Mucin_TYPE +
## SignetRING + Lymphovascularinvasion + perineural + CT + RT +
## NACTRT + Progress + IntervalOD + IntervalOR + AJCC_bin
##
## Df Deviance AIC
## - Log2AT 1 1103.8 1153.8
## - perineural 1 1103.8 1153.8
## - Liver_Only 1 1103.9 1153.9
## - CAD 1 1103.9 1153.9
## - Laparoscopic 1 1103.9 1153.9
## - TumorLOC 1 1104.0 1154.0
## - HF 1 1104.0 1154.0
## - EA 1 1104.0 1154.0
## - CT 1 1104.1 1154.1
## - Lymphovascularinvasion 1 1104.1 1154.1
## - DM 1 1104.1 1154.1
## - SignetRING 1 1104.2 1154.2
## - CVA 1 1104.2 1154.2
## - ASA3 1 1104.2 1154.2
## - CKD 1 1105.6 1155.6
## <none> 1103.6 1155.6
## - IntervalOR 1 1106.5 1156.5
## - Mucin_TYPE 1 1106.5 1156.5
## - Age 1 1106.6 1156.6
## - AJCC_bin 1 1107.0 1157.0
## - RT 1 1107.2 1157.2
## - NACTRT 1 1107.3 1157.3
## - LogCEA 1 1109.6 1159.6

```



```

## - RBC 1 1110.1 1160.1
## - IntervalOD 1 1118.1 1168.1
## - Progress 1 1135.0 1185.0
##
## Step: AIC=1153.77
## Death ~ Age + ASA3 + DM + CAD + HF + CVA + CKD + LogCEA + Laparoscopic +
## TumorLOC + EA + RBC + Liver_Only + Mucin_TYPE + SignetRING +
## Lymphovascularinvasion + perineural + CT + RT + NACTRT +
## Progress + IntervalOD + IntervalOR + AJCC_bin
##
## Df Deviance AIC
## - perineural 1 1103.9 1151.9
## - TumorLOC 1 1104.1 1152.1
## - CAD 1 1104.1 1152.1
## - Liver_Only 1 1104.1 1152.1
## - Laparoscopic 1 1104.1 1152.1
## - HF 1 1104.1 1152.1
## - EA 1 1104.2 1152.2
## - Lymphovascularinvasion 1 1104.2 1152.2
## - CT 1 1104.2 1152.2
## - DM 1 1104.3 1152.3
## - SignetRING 1 1104.3 1152.3
## - ASA3 1 1104.4 1152.4
## - CVA 1 1104.4 1152.4
## - CKD 1 1105.8 1153.8
## <none> 1103.8 1153.8
## - IntervalOR 1 1106.5 1154.5
## - Mucin_TYPE 1 1106.7 1154.7
## - Age 1 1106.9 1154.9
## - AJCC_bin 1 1107.2 1155.2
## - RT 1 1107.3 1155.3
## - NACTRT 1 1107.3 1155.3
## - LogCEA 1 1109.6 1157.6
## - RBC 1 1111.4 1159.4
## - IntervalOD 1 1118.2 1166.2
## - Progress 1 1135.2 1183.2
##
## Step: AIC=1151.94
## Death ~ Age + ASA3 + DM + CAD + HF + CVA + CKD + LogCEA + Laparoscopic +
## TumorLOC + EA + RBC + Liver_Only + Mucin_TYPE + SignetRING +
## Lymphovascularinvasion + CT + RT + NACTRT + Progress + IntervalOD +
## IntervalOR + AJCC_bin
##
## Df Deviance AIC
## - CAD 1 1104.2 1150.2
## - Liver_Only 1 1104.2 1150.2
## - TumorLOC 1 1104.3 1150.3
## - Laparoscopic 1 1104.3 1150.3
## - HF 1 1104.3 1150.3
## - EA 1 1104.3 1150.3
## - CT 1 1104.4 1150.4

```

```

## - SignetRING          1    1104.4 1150.4
## - DM                  1    1104.5 1150.5
## - ASA3                1    1104.5 1150.5
## - CVA                 1    1104.6 1150.6
## - Lymphovascularinvasion 1    1104.7 1150.7
## - CKD                 1    1105.9 1151.9
## <none>                 1    1103.9 1151.9
## - IntervalOR          1    1106.6 1152.6
## - Mucin_TYPE          1    1106.8 1152.8
## - Age                 1    1107.0 1153.0
## - AJCC_bin            1    1107.3 1153.3
## - NACTRT              1    1107.6 1153.6
## - RT                  1    1107.6 1153.6
## - LogCEA              1    1109.7 1155.7
## - RBC                 1    1111.6 1157.6
## - IntervalOD          1    1118.4 1164.4
## - Progress            1    1135.4 1181.4
##
## Step:  AIC=1150.23
## Death ~ Age + ASA3 + DM + HF + CVA + CKD + LogCEA + Laparoscopic +
##         TumorLOC + EA + RBC + Liver_Only + Mucin_TYPE + SignetRING +
##         Lymphovascularinvasion + CT + RT + NACTRT + Progress + IntervalOD +
##         IntervalOR + AJCC_bin
##
##              Df Deviance    AIC
## - HF          1    1104.5 1148.5
## - Liver_Only   1    1104.5 1148.5
## - TumorLOC     1    1104.6 1148.6
## - Laparoscopic 1    1104.6 1148.6
## - EA          1    1104.6 1148.6
## - CT          1    1104.7 1148.7
## - DM          1    1104.7 1148.7
## - SignetRING  1    1104.7 1148.7
## - ASA3        1    1104.7 1148.7
## - CVA         1    1104.9 1148.9
## - Lymphovascularinvasion 1    1104.9 1148.9
## - CKD         1    1106.2 1150.2
## <none>         1    1104.2 1150.2
## - IntervalOR  1    1106.9 1150.9
## - Mucin_TYPE  1    1107.1 1151.1
## - Age         1    1107.5 1151.5
## - AJCC_bin    1    1107.7 1151.7
## - NACTRT      1    1107.9 1151.9
## - RT          1    1108.0 1152.0
## - LogCEA      1    1110.0 1154.0
## - RBC         1    1111.8 1155.8
## - IntervalOD  1    1118.9 1162.9
## - Progress    1    1135.8 1179.8
##
## Step:  AIC=1148.5
## Death ~ Age + ASA3 + DM + CVA + CKD + LogCEA + Laparoscopic +

```

```

##      TumorLOC + EA + RBC + Liver_Only + Mucin_TYPE + SignetRING +
##      Lymphovascularinvasion + CT + RT + NACTRT + Progress + IntervalOD +
##      IntervalOR + AJCC_bin
##
##              Df Deviance      AIC
## - Liver_Only          1    1104.8 1146.8
## - TumorLOC            1    1104.8 1146.8
## - Laparoscopic        1    1104.8 1146.8
## - CT                  1    1104.9 1146.9
## - EA                  1    1104.9 1146.9
## - DM                  1    1105.0 1147.0
## - SignetRING          1    1105.0 1147.0
## - ASA3                1    1105.1 1147.1
## - CVA                 1    1105.1 1147.1
## - Lymphovascularinvasion 1    1105.2 1147.2
## <none>                 1    1104.5 1148.5
## - CKD                 1    1106.5 1148.5
## - IntervalOR          1    1107.2 1149.2
## - Mucin_TYPE          1    1107.4 1149.4
## - Age                 1    1107.7 1149.7
## - AJCC_bin            1    1107.9 1149.9
## - RT                  1    1108.2 1150.2
## - NACTRT              1    1108.3 1150.3
## - LogCEA              1    1110.1 1152.1
## - RBC                 1    1112.2 1154.2
## - IntervalOD          1    1119.2 1161.2
## - Progress            1    1136.2 1178.2
##
## Step:  AIC=1146.76
## Death ~ Age + ASA3 + DM + CVA + CKD + LogCEA + Laparoscopic +
##      TumorLOC + EA + RBC + Mucin_TYPE + SignetRING + Lymphovascularinvasion +
##      CT + RT + NACTRT + Progress + IntervalOD + IntervalOR + AJCC_bin
##
##              Df Deviance      AIC
## - TumorLOC            1    1105.1 1145.1
## - Laparoscopic        1    1105.1 1145.1
## - EA                  1    1105.1 1145.1
## - CT                  1    1105.2 1145.2
## - DM                  1    1105.2 1145.2
## - SignetRING          1    1105.2 1145.2
## - ASA3                1    1105.3 1145.3
## - CVA                 1    1105.4 1145.4
## - Lymphovascularinvasion 1    1105.5 1145.5
## - CKD                 1    1106.8 1146.8
## <none>                 1    1104.8 1146.8
## - IntervalOR          1    1107.5 1147.5
## - Mucin_TYPE          1    1107.7 1147.7
## - Age                 1    1107.9 1147.9
## - RT                  1    1108.4 1148.4
## - NACTRT              1    1108.5 1148.5
## - AJCC_bin            1    1109.1 1149.1

```

```

## - LogCEA          1    1111.0 1151.0
## - RBC              1    1112.7 1152.7
## - IntervalOD      1    1119.8 1159.8
## - Progress        1    1136.4 1176.4
##
## Step:  AIC=1145.06
## Death ~ Age + ASA3 + DM + CVA + CKD + LogCEA + Laparoscopic +
##         EA + RBC + Mucin_TYPE + SignetRING + Lymphovascularinvasion +
##         CT + RT + NACTRT + Progress + IntervalOD + IntervalOR + AJCC_bin
##
##              Df Deviance    AIC
## - Laparoscopic      1    1105.4 1143.4
## - EA                 1    1105.4 1143.4
## - CT                 1    1105.4 1143.4
## - DM                 1    1105.5 1143.5
## - SignetRING         1    1105.6 1143.6
## - ASA3               1    1105.7 1143.7
## - CVA                1    1105.7 1143.7
## - Lymphovascularinvasion 1    1105.7 1143.7
## - CKD                1    1107.0 1145.0
## <none>                1    1105.1 1145.1
## - IntervalOR        1    1107.8 1145.8
## - Mucin_TYPE        1    1108.0 1146.0
## - Age                1    1108.1 1146.1
## - RT                 1    1108.8 1146.8
## - NACTRT             1    1109.1 1147.1
## - AJCC_bin           1    1109.4 1147.4
## - LogCEA             1    1111.1 1149.1
## - RBC                1    1113.3 1151.3
## - IntervalOD         1    1120.3 1158.3
## - Progress           1    1136.5 1174.5
##
## Step:  AIC=1143.37
## Death ~ Age + ASA3 + DM + CVA + CKD + LogCEA + EA + RBC + Mucin_TYPE +
##         SignetRING + Lymphovascularinvasion + CT + RT + NACTRT +
##         Progress + IntervalOD + IntervalOR + AJCC_bin
##
##              Df Deviance    AIC
## - EA                 1    1105.7 1141.7
## - CT                 1    1105.8 1141.8
## - DM                 1    1105.8 1141.8
## - SignetRING         1    1105.8 1141.8
## - ASA3               1    1106.0 1142.0
## - CVA                1    1106.0 1142.0
## - Lymphovascularinvasion 1    1106.1 1142.1
## - CKD                1    1107.3 1143.3
## <none>                1    1105.4 1143.4
## - IntervalOR        1    1108.1 1144.1
## - Mucin_TYPE        1    1108.2 1144.2
## - Age                1    1108.5 1144.5
## - RT                 1    1109.1 1145.1

```

```

## - NACTRT          1    1109.4 1145.4
## - AJCC_bin        1    1109.9 1145.9
## - LogCEA          1    1111.3 1147.3
## - RBC             1    1113.7 1149.7
## - IntervalOD      1    1120.5 1156.5
## - Progress        1    1136.8 1172.8
##
## Step:  AIC=1141.74
## Death ~ Age + ASA3 + DM + CVA + CKD + LogCEA + RBC + Mucin_TYPE +
##      SignetRING + Lymphovascularinvasion + CT + RT + NACTRT +
##      Progress + IntervalOD + IntervalOR + AJCC_bin
##
##              Df Deviance    AIC
## - CT          1    1106.1 1140.1
## - DM          1    1106.2 1140.2
## - SignetRING   1    1106.2 1140.2
## - Lymphovascularinvasion 1    1106.4 1140.4
## - ASA3        1    1106.4 1140.4
## - CVA         1    1106.4 1140.4
## - CKD         1    1107.7 1141.7
## <none>        1    1105.7 1141.7
## - IntervalOR  1    1108.4 1142.4
## - Mucin_TYPE  1    1108.6 1142.6
## - Age         1    1108.8 1142.8
## - RT          1    1109.4 1143.4
## - NACTRT      1    1109.5 1143.5
## - AJCC_bin    1    1110.2 1144.2
## - LogCEA      1    1111.8 1145.8
## - RBC         1    1114.1 1148.1
## - IntervalOD  1    1121.1 1155.1
## - Progress    1    1137.2 1171.2
##
## Step:  AIC=1140.1
## Death ~ Age + ASA3 + DM + CVA + CKD + LogCEA + RBC + Mucin_TYPE +
##      SignetRING + Lymphovascularinvasion + RT + NACTRT + Progress +
##      IntervalOD + IntervalOR + AJCC_bin
##
##              Df Deviance    AIC
## - DM          1    1106.5 1138.5
## - SignetRING   1    1106.6 1138.6
## - CVA         1    1106.7 1138.7
## - ASA3        1    1106.7 1138.7
## - Lymphovascularinvasion 1    1106.7 1138.7
## - CKD         1    1108.0 1140.0
## <none>        1    1106.1 1140.1
## - IntervalOR  1    1108.7 1140.7
## - Mucin_TYPE  1    1108.8 1140.8
## - Age         1    1109.5 1141.5
## - RT          1    1109.8 1141.8
## - NACTRT      1    1110.0 1142.0
## - AJCC_bin    1    1110.5 1142.5

```

```

## - LogCEA          1    1112.1 1144.1
## - RBC              1    1114.3 1146.3
## - IntervalOD       1    1121.2 1153.2
## - Progress         1    1146.2 1178.2
##
## Step:  AIC=1138.55
## Death ~ Age + ASA3 + CVA + CKD + LogCEA + RBC + Mucin_TYPE +
##      SignetRING + Lymphovascularinvasion + RT + NACTRT + Progress +
##      IntervalOD + IntervalOR + AJCC_bin
##
##              Df Deviance    AIC
## - SignetRING      1    1107.0 1137.0
## - CVA              1    1107.1 1137.1
## - Lymphovascularinvasion 1    1107.2 1137.2
## - ASA3             1    1107.3 1137.3
## - CKD              1    1108.5 1138.5
## <none>              1    1106.5 1138.5
## - IntervalOR      1    1109.0 1139.0
## - Mucin_TYPE       1    1109.4 1139.4
## - Age              1    1109.8 1139.8
## - RT               1    1110.2 1140.2
## - NACTRT           1    1110.4 1140.4
## - AJCC_bin         1    1111.1 1141.1
## - LogCEA           1    1112.6 1142.6
## - RBC              1    1114.8 1144.8
## - IntervalOD       1    1121.6 1151.6
## - Progress         1    1146.5 1176.5
##
## Step:  AIC=1137.01
## Death ~ Age + ASA3 + CVA + CKD + LogCEA + RBC + Mucin_TYPE +
##      Lymphovascularinvasion + RT + NACTRT + Progress + IntervalOD +
##      IntervalOR + AJCC_bin
##
##              Df Deviance    AIC
## - Lymphovascularinvasion 1    1107.5 1135.5
## - CVA                    1    1107.6 1135.6
## - ASA3                   1    1107.7 1135.7
## - CKD                    1    1108.9 1136.9
## <none>                    1    1107.0 1137.0
## - Mucin_TYPE             1    1109.4 1137.4
## - IntervalOR             1    1109.5 1137.5
## - Age                    1    1110.4 1138.4
## - RT                     1    1110.5 1138.5
## - NACTRT                 1    1111.0 1139.0
## - AJCC_bin               1    1111.8 1139.8
## - LogCEA                 1    1112.9 1140.9
## - RBC                    1    1115.4 1143.4
## - IntervalOD             1    1122.2 1150.2
## - Progress               1    1146.8 1174.8
##
## Step:  AIC=1135.53

```

```

## Death ~ Age + ASA3 + CVA + CKD + LogCEA + RBC + Mucin_TYPE +
##      RT + NACTRT + Progress + IntervalOD + IntervalOR + AJCC_bin
##
##           Df Deviance    AIC
## - CVA      1   1108.1 1134.1
## - ASA3     1   1108.2 1134.2
## - CKD      1   1109.4 1135.4
## <none>      1   1107.5 1135.5
## - Mucin_TYPE 1   1109.9 1135.9
## - IntervalOR 1   1110.0 1136.0
## - Age      1   1110.7 1136.7
## - RT       1   1111.0 1137.0
## - NACTRT   1   1111.3 1137.3
## - AJCC_bin 1   1112.0 1138.0
## - LogCEA   1   1113.6 1139.6
## - RBC      1   1116.1 1142.1
## - IntervalOD 1   1122.2 1148.2
## - Progress  1   1146.8 1172.8
##
## Step:  AIC=1134.1
## Death ~ Age + ASA3 + CKD + LogCEA + RBC + Mucin_TYPE + RT + NACTRT +
##      Progress + IntervalOD + IntervalOR + AJCC_bin
##
##           Df Deviance    AIC
## - ASA3     1   1109.0 1133.0
## <none>      1   1108.1 1134.1
## - CKD      1   1110.2 1134.2
## - Mucin_TYPE 1   1110.5 1134.5
## - IntervalOR 1   1110.8 1134.8
## - Age      1   1111.2 1135.2
## - RT       1   1111.5 1135.5
## - NACTRT   1   1112.0 1136.0
## - AJCC_bin 1   1112.5 1136.5
## - LogCEA   1   1114.0 1138.0
## - RBC      1   1116.5 1140.5
## - IntervalOD 1   1122.5 1146.5
## - Progress  1   1147.0 1171.0
##
## Step:  AIC=1132.96
## Death ~ Age + CKD + LogCEA + RBC + Mucin_TYPE + RT + NACTRT +
##      Progress + IntervalOD + IntervalOR + AJCC_bin
##
##           Df Deviance    AIC
## <none>      1   1109.0 1133.0
## - Mucin_TYPE 1   1111.3 1133.3
## - Age      1   1111.3 1133.3
## - CKD      1   1111.3 1133.3
## - IntervalOR 1   1111.7 1133.7
## - RT       1   1112.1 1134.1
## - NACTRT   1   1113.1 1135.1
## - AJCC_bin 1   1113.3 1135.3

```

```
## - LogCEA      1    1115.1 1137.1
## - RBC         1    1118.5 1140.5
## - IntervalOD  1    1123.8 1145.8
## - Progress    1    1147.8 1169.8
```

summary(final)

```
##
## Call:
## glm(formula = Death ~ Age + CKD + LogCEA + RBC + Mucin_TYPE +
##      RT + NACTRT + Progress + IntervalOD + IntervalOR + AJCC_bin,
##      family = binomial(link = "logit"), data = data)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.880876   0.468445  -1.880 0.060050 .
## Age         -0.008475   0.005552  -1.526 0.126900
## CKD          0.333746   0.218141   1.530 0.126027
## LogCEA       0.202194   0.082029   2.465 0.013705 *
## RBC          0.337265   0.109720   3.074 0.002113 **
## Mucin_TYPE  -0.427789   0.282598  -1.514 0.130084
## RT           0.410487   0.231822   1.771 0.076610 .
## NACTRT       -0.408753   0.203099  -2.013 0.044159 *
## Progress     1.301195   0.217872   5.972 2.34e-09 ***
## IntervalOD  -0.017607   0.004852  -3.628 0.000285 ***
## IntervalOR  -0.015871   0.009828  -1.615 0.106346
## AJCC_bin     0.314574   0.151810   2.072 0.038251 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1247.1  on 915  degrees of freedom
## Residual deviance: 1109.0  on 904  degrees of freedom
## AIC: 1133
##
## Number of Fisher Scoring iterations: 5
```

ODDS RATIO

exp(**cbind**(OR= **coef**(final), **confint**(final)))

Waiting for profiling to be done...

```
##              OR      2.5 %    97.5 %
## (Intercept) 0.4144195 0.1645328 1.0341418
## Age         0.9915613 0.9807857 1.0023860
## CKD         1.3961891 0.9103082 2.1439434
## LogCEA      1.2240849 1.0427580 1.4387055
## RBC         1.4011099 1.1309504 1.7395012
## Mucin_TYPE  0.6519492 0.3707811 1.1272715
## RT         1.5075518 0.9580658 2.3821487
## NACTRT      0.6644785 0.4444137 0.9864588
```



```

## Progress      3.6736842 2.4138468 5.6790901
## IntervalOD    0.9825476 0.9728936 0.9916305
## IntervalOR    0.9842540 0.9647861 1.0027095
## AJCC_bin      1.3696754 1.0170982 1.8448146

# Calcola OR e intervalli di confidenza
or_ci <- exp(cbind(OR= coef(final), confint(final)))

## Waiting for profiling to be done...

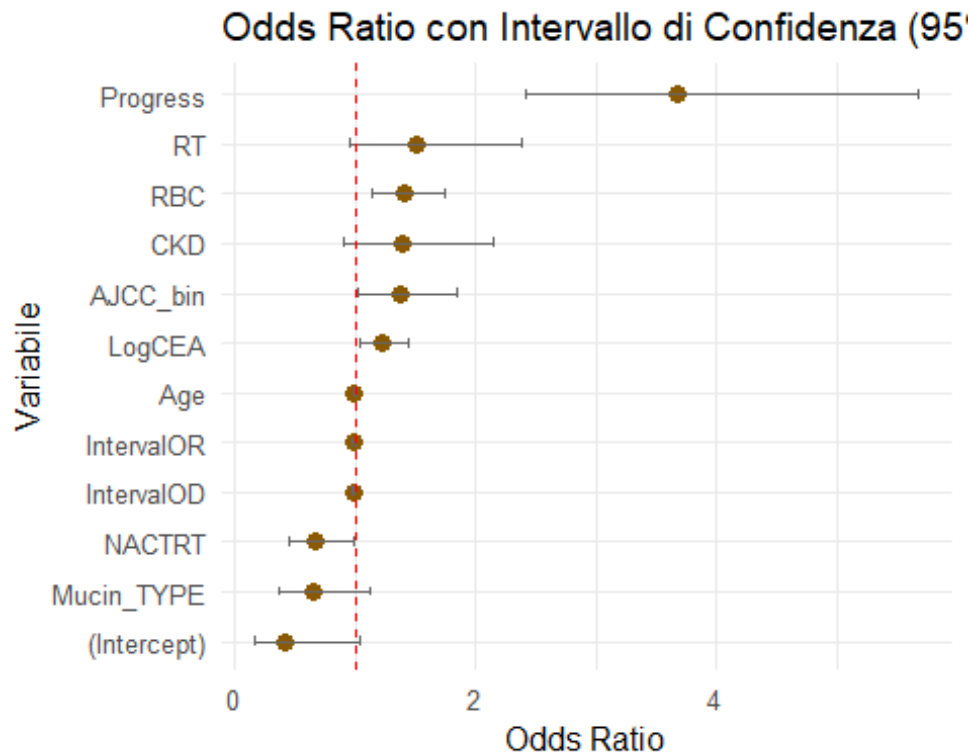
# Crea un data frame ordinato
or_df <- as.data.frame(or_ci)
or_df$Variable <- rownames(or_df)
colnames(or_df) <- c("OR", "CI_low", "CI_high", "Variable")

# Ordina per OR o per significatività, se preferisci
or_df <- or_df[order(or_df$OR, decreasing = TRUE), ]

# Carica ggplot2
library(ggplot2)

# Plot
ggplot(or_df, aes(x = reorder(Variable, OR), y = OR)) +
  geom_point(color = "orange4", size = 3) +
  geom_errorbar(aes(ymin = CI_low, ymax = CI_high), width = 0.2, color = "grey40")
+
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
  coord_flip() + # Inverte gli assi per leggibilità
  labs(title = "Odds Ratio con Intervallo di Confidenza (95%)",
       x = "Variabile",
       y = "Odds Ratio") +
  theme_minimal(base_size = 12)

```



OR = 1 → Nessuna associazione con l'outcome. CEA, ID, ANESTIME, INTERVAL

OR > 1 → La variabile è associata a un aumento della probabilità dell'evento (es. progressione). ANESTIME, PROGRESS, RADIOTHERAPY, DIABETES, AJCC

OR < 1 → La variabile è associata a una diminuzione della probabilità dell'evento. INTERCETTA

SURVIVAL ANALYSIS

PREPARAZIONE DEL DATASET

```
library(survival)

## Warning: package 'survival' was built under R version 4.3.3

library(survminer)

## Warning: package 'survminer' was built under R version 4.3.3

## Loading required package: ggpubr

## Warning: package 'ggpubr' was built under R version 4.3.3

##
## Attaching package: 'survminer'

## The following object is masked from 'package:survival':
##
##      myeloma
```

```
library(dplyr)
```

```
# senza le variabili tempo ed evento
```

```
data_covariates <- subset(data, select = -c(IntervalOD, Death))
```

```
# Creazione dell'oggetto Surv: tempo = Interval (in mesi), evento = Death  
(1=evento, 0=censura)
```

```
surv_object <- Surv(time = data$IntervalOD, event = data$Death)
```

```
summary(surv_object)
```

```
##           time           status  
## Min.      : 0.03285   Min.      :0.0000  
## 1st Qu.:  8.11499   1st Qu.:0.0000  
## Median : 18.05339   Median :0.0000  
## Mean     : 24.60655   Mean     :0.4214  
## 3rd Qu.: 31.95072   3rd Qu.:1.0000  
## Max.     :135.78645   Max.     :1.0000
```

KAPLAN MEIER CURVE - CURVA COMPLESSIVA SOPRAVVIVENZA

```
library(survival)
```

```
library(survminer)
```

```
# Opzione 1 - Con Surv e formula (raccomandata)
```

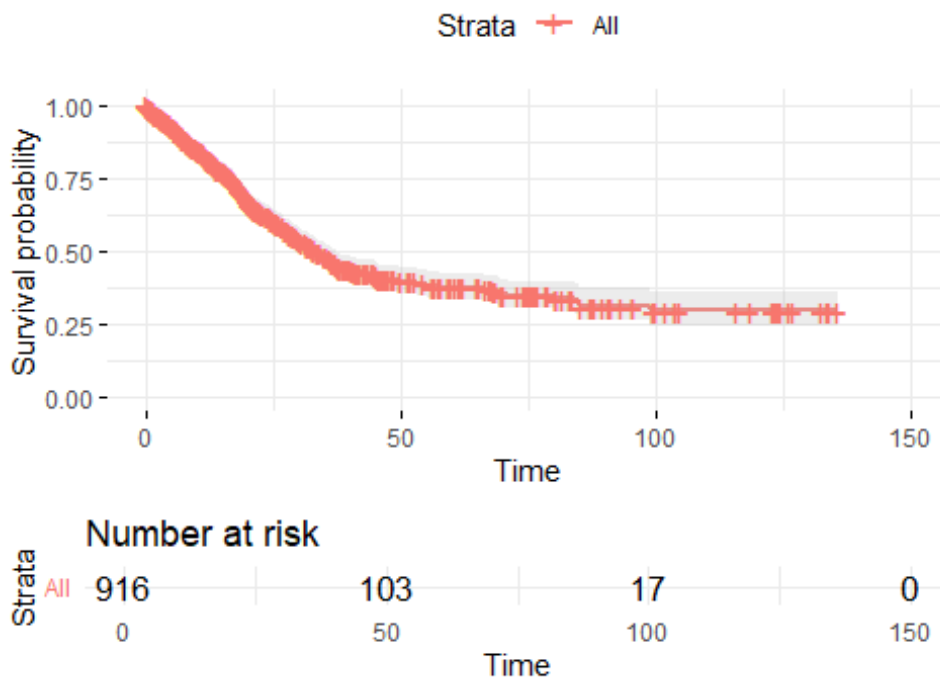
```
km_fit <- survfit(Surv(IntervalOD, Death) ~ 1, data = data)
```

```
ggsurvplot(km_fit, data = data, conf.int = TRUE, risk.table = TRUE, pval = TRUE,  
            ggtheme = theme_minimal(), title = "Kaplan-Meier curve - Overall  
Survival")
```

```
## Warning in .pvalue(fit, data = data, method = method, pval = pval, pval.coord =  
pval.coord, : There are no survival curves to be compared.
```

```
## This is a null model.
```

Kaplan-Meier curve - Overall Survival



KAPLAN MEIER - variabili

```
library(survival)
```

```
library(survminer)
```

```
# Controlla che la variabile sia presente e corretta
```

```
colnames(data)
```

```
## [1] "Age" "Gender" "ASA3"
## [4] "DM" "CAD" "HF"
## [7] "CVA" "CKD" "LogCEA"
## [10] "Laparoscopic" "TumorLOC" "EA"
## [13] "Log2AT" "RBC" "Liver_Only"
## [16] "Cell_diff" "Mucin_TYPE" "SignetRING"
## [19] "Lymphovascularinvasion" "perineural" "CT"
## [22] "RT" "NACTRT" "Death"
## [25] "Progress" "IntervalOD" "IntervalOR"
## [28] "AJCC_bin"
```

```
str(data$ASA3)
```

```
## num [1:916] 1 0 0 0 0 0 1 0 1 0 ...
```

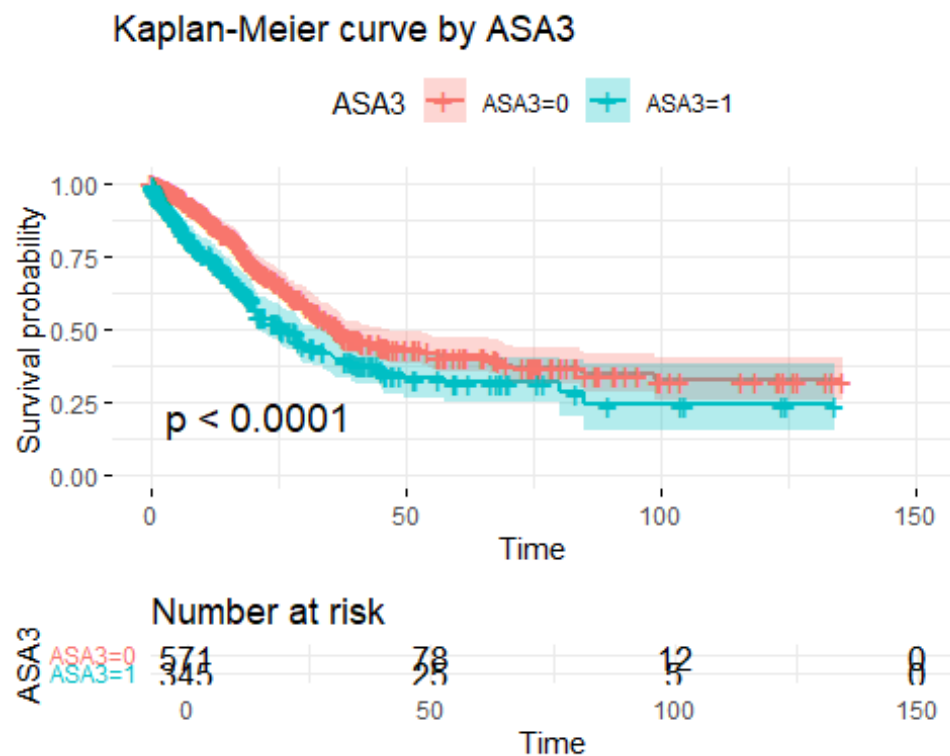
```
# Se necessario, converti ASA3 in fattore
```

```
data$ASA3 <- as.factor(data$ASA3)
```

```
# Creazione formula e fit
```

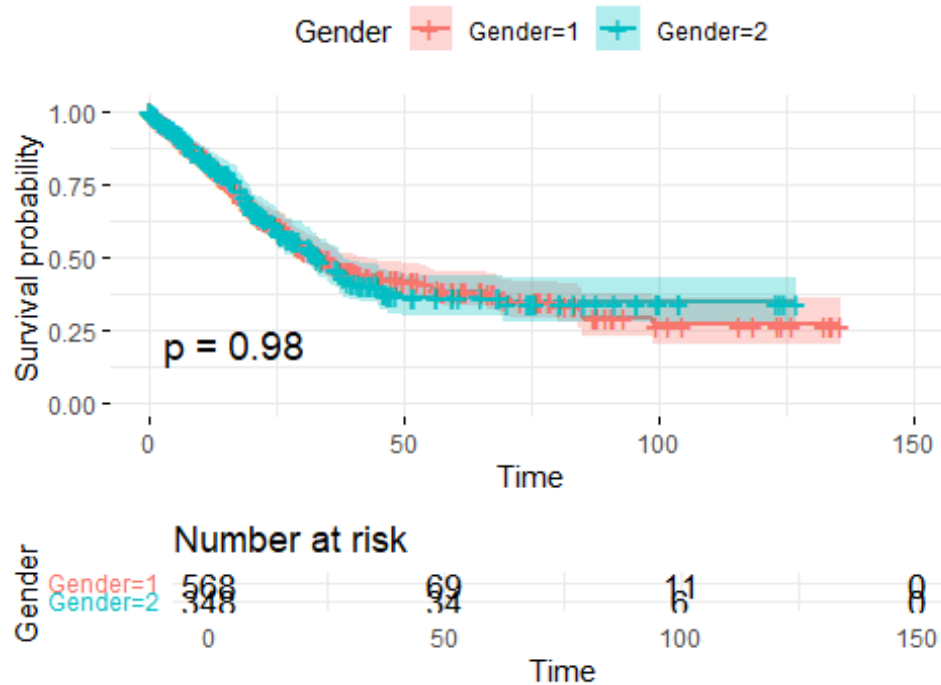
```
km_fit_ASA3 <- survfit(Surv(IntervalOD, Death) ~ ASA3, data = data)
```

```
# Plot Kaplan-Meier
ggsurvplot(km_fit_ASA3, data = data,
  conf.int = TRUE, risk.table = TRUE, pval = TRUE,
  ggtheme = theme_minimal(),
  title = "Kaplan-Meier curve by ASA3",
  legend.title = "ASA3")
```



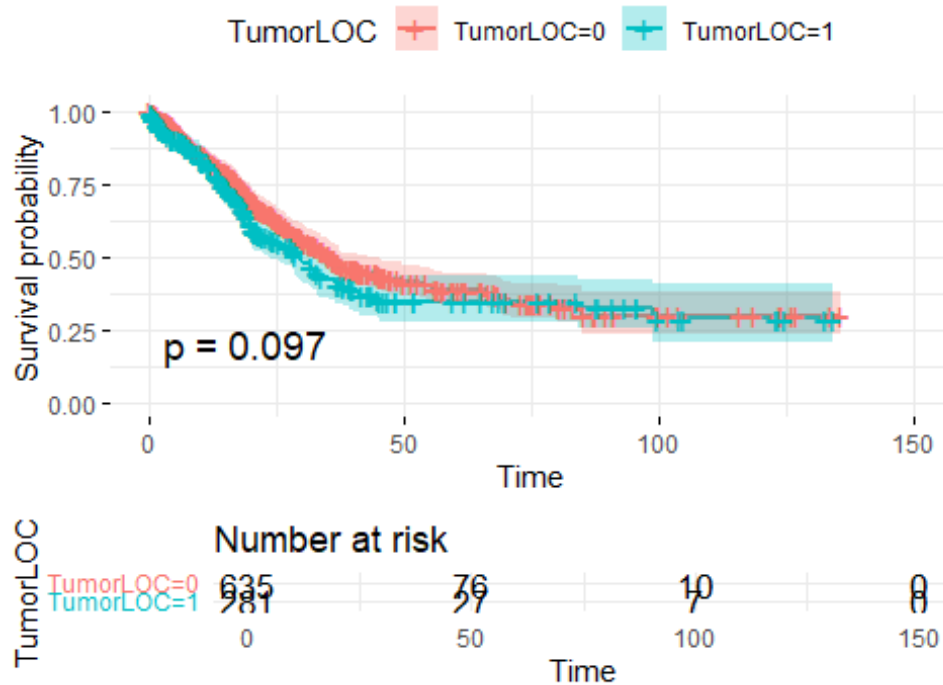
```
data$Gender <- as.factor(data$Gender)
km_fit_Gender <- survfit(Surv(IntervalOD, Death) ~ Gender, data = data)
ggsurvplot(km_fit_Gender, data = data,
  conf.int = TRUE, risk.table = TRUE, pval = TRUE,
  ggtheme = theme_minimal(),
  title = "Kaplan-Meier curve by Gender",
  legend.title = "Gender")
```

Kaplan-Meier curve by Gender



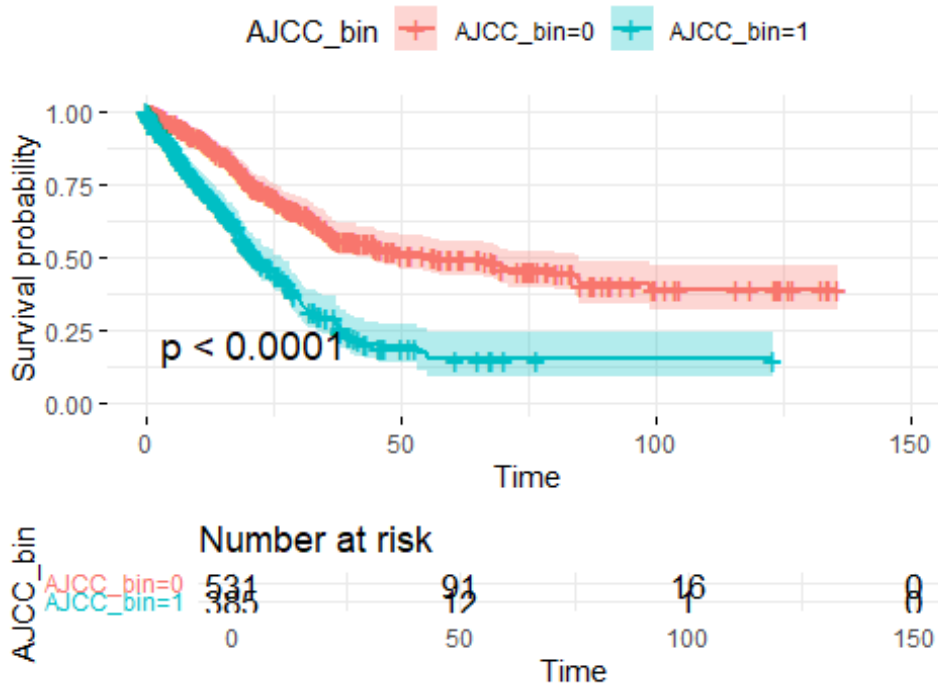
```
data$TumorLOC <- as.factor(data$TumorLOC)
km_fit_TumorLOC <- survfit(Surv(IntervalOD, Death) ~ TumorLOC, data = data)
ggsurvplot(km_fit_TumorLOC, data = data,
  conf.int = TRUE, risk.table = TRUE, pval = TRUE,
  ggtheme = theme_minimal(),
  title = "Kaplan-Meier curve by Tumor Location",
  legend.title = "TumorLOC")
```

Kaplan-Meier curve by Tumor Location



```
data$AJCC_bin <- as.factor(data$AJCC_bin)
km_fit_AJCC <- survfit(Surv(IntervalOD, Death) ~ AJCC_bin, data = data)
ggsurvplot(km_fit_AJCC, data = data,
  conf.int = TRUE, risk.table = TRUE, pval = TRUE,
  ggtheme = theme_minimal(),
  title = "Kaplan-Meier curve by AJCC_bin",
  legend.title = "AJCC_bin")
```

Kaplan-Meier curve by AJCC_bin



AJCC_bin

P-value < 0.0001 → differenza statisticamente significativa tra i gruppi 0 e 1.

Sopravvivenza marcatamente migliore per AJCC_bin=0. Interpretazione: Lo stadio AJCC è fortemente associato alla sopravvivenza, come atteso.

TumorLOC

P-value = 0.097 → non significativo.

Differenze modeste tra localizzazione colon/retto. Interpretazione: La sede del tumore sembra avere un impatto limitato sulla sopravvivenza.

Gender

P-value = 0.98 → non significativo.

Curve sovrapponibili. Interpretazione: Nessuna differenza di sopravvivenza tra uomini e donne.

ASA3

P-value < 0.0001 → differenza significativa.

Sopravvivenza inferiore nei pazienti ASA3=1 (compromessi). Interpretazione: Lo stato clinico pre-operatorio è un predittore importante.

Overall survival

Curve uniche, senza confronto. Interpretazione: Quadro generale di sopravvivenza della popolazione.

LOG RANK TEST

```
library(survival)

# ASA3
cat("\n=== Log-Rank Test per ASA3 ===\n")

##
## === Log-Rank Test per ASA3 ===

survdifff(Surv(IntervalOD, Death) ~ ASA3, data = data)

## Call:
## survdifff(formula = Surv(IntervalOD, Death) ~ ASA3, data = data)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## ASA3=0  571      228      265      5.12      16.4
## ASA3=1  345      158      121     11.19      16.4
##
## Chisq= 16.4  on 1 degrees of freedom, p= 5e-05

# Gender
cat("\n=== Log-Rank Test per Gender ===\n")

##
## === Log-Rank Test per Gender ===

survdifff(Surv(IntervalOD, Death) ~ Gender, data = data)

## Call:
## survdifff(formula = Surv(IntervalOD, Death) ~ Gender, data = data)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Gender=1  568      237      237  0.000204  0.000528
## Gender=2  348      149      149  0.000323  0.000528
##
## Chisq= 0  on 1 degrees of freedom, p= 1

# TumorLOC
cat("\n=== Log-Rank Test per TumorLOC ===\n")

##
## === Log-Rank Test per TumorLOC ===

survdifff(Surv(IntervalOD, Death) ~ TumorLOC, data = data)

## Call:
## survdifff(formula = Surv(IntervalOD, Death) ~ TumorLOC, data = data)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## TumorLOC=0 635      263      278      0.77      2.75
```

```
## TumorLOC=1 281      123      108      1.97      2.75
##
## Chisq= 2.7  on 1 degrees of freedom, p= 0.1

# AJCC_bin
cat("\n=== Log-Rank Test per AJCC_bin ===\n")

##
## === Log-Rank Test per AJCC_bin ===

survdifff(Surv(IntervalOD, Death) ~ AJCC_bin, data = data)

## Call:
## survdifff(formula = Surv(IntervalOD, Death) ~ AJCC_bin, data = data)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## AJCC_bin=0 531      189      269      23.6      80.4
## AJCC_bin=1 385      197      117      53.9      80.4
##
## Chisq= 80.4  on 1 degrees of freedom, p= <2e-16
```

LOG RANK TEST ANALISI:

1. ASA3 Chi-squared = 16.4, $p = 5e-05$ (molto significativo).

Interpretazione: La differenza di sopravvivenza tra pazienti ASA3=0 (basso rischio) e ASA3=1 (alto rischio) è significativa. I pazienti ASA3=1 hanno una sopravvivenza inferiore.

2. Gender Chi-squared ≈ 0 , $p = 1$ (non significativo).

Interpretazione: Nessuna differenza di sopravvivenza tra uomini (Gender=1) e donne (Gender=2).

3. TumorLOC Chi-squared = 2.7, $p \approx 0.1$ (non significativo).

Interpretazione: La localizzazione del tumore (colon vs retto) mostra una differenza non significativa. Potrebbe esserci una tendenza, ma non è statisticamente rilevante.

4. AJCC_bin Chi-squared = 80.4, $p < 2e-16$ (altamente significativo).

Interpretazione: Lo stadio AJCC è fortemente associato alla sopravvivenza. I pazienti con AJCC_bin=1 (stadio avanzato) hanno una sopravvivenza significativamente inferiore rispetto a quelli con AJCC_bin=0.

COX MODEL MULTIVARIATO BASE

```
library(survival)
# Costruisci il modello di Cox correttamente
cox_model <- survival::coxph(Surv(time = data$IntervalOD, event = data$Death) ~ .,
data = data_covariates)

# Riepilogo
summary(cox_model)
```

```
## Call:
## survival::coxph(formula = Surv(time = data$IntervalOD, event = data$Death) ~
##      ., data = data_covariates)
##
##      n= 916, number of events= 386
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Age           0.003871  1.003879  0.004420   0.876 0.381109
## Gender        -0.051848  0.949474  0.111083  -0.467 0.640682
## ASA3           0.332491  1.394438  0.122490   2.714 0.006639 **
## DM            -0.102926  0.902194  0.128998  -0.798 0.424938
## CAD           -0.257736  0.772799  0.217288  -1.186 0.235563
## HF            -0.006674  0.993348  0.254941  -0.026 0.979115
## CVA           -0.097148  0.907421  0.226383  -0.429 0.667827
## CKD           0.004420  1.004430  0.148534   0.030 0.976258
## LogCEA        0.291469  1.338393  0.059650   4.886 1.03e-06 ***
## Laparoscopic  0.136251  1.145970  0.278392   0.489 0.624543
## TumorLOC      0.050862  1.052178  0.119640   0.425 0.670745
## EA           -0.133389  0.875124  0.146984  -0.908 0.364137
## Log2AT        -0.070096  0.932305  0.111543  -0.628 0.529728
## RBC           0.276332  1.318285  0.081833   3.377 0.000733 ***
## Liver_Only    0.310046  1.363487  0.171486   1.808 0.070607 .
## Cell_diff     0.699494  2.012734  0.161757   4.324 1.53e-05 ***
## Mucin_TYPE    -0.100489  0.904395  0.219563  -0.458 0.647183
## SignetRING    0.095302  1.099991  0.270049   0.353 0.724160
## Lymphovascularinvasion 0.267522  1.306722  0.115471   2.317 0.020515 *
## perineural    -0.080879  0.922306  0.133749  -0.605 0.545375
## CT            -0.922426  0.397553  0.235829  -3.911 9.18e-05 ***
## RT            0.096490  1.101299  0.157235   0.614 0.539432
## NACTRT        -0.018240  0.981925  0.154773  -0.118 0.906186
## Progress      -1.332024  0.263942  0.229387  -5.807 6.36e-09 ***
## IntervalOR    -0.098992  0.905750  0.008579 -11.539 < 2e-16 ***
## AJCC_bin      0.798808  2.222889  0.167169   4.778 1.77e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## Age           1.0039      0.9961    0.9952    1.0126
## Gender         0.9495      1.0532    0.7637    1.1804
## ASA3           1.3944      0.7171    1.0968    1.7728
## DM             0.9022      1.1084    0.7006    1.1617
## CAD            0.7728      1.2940    0.5048    1.1831
## HF             0.9933      1.0067    0.6027    1.6372
## CVA            0.9074      1.1020    0.5823    1.4142
## CKD            1.0044      0.9956    0.7507    1.3439
## LogCEA         1.3384      0.7472    1.1907    1.5044
## Laparoscopic   1.1460      0.8726    0.6641    1.9776
## TumorLOC       1.0522      0.9504    0.8322    1.3302
## EA             0.8751      1.1427    0.6561    1.1673
## Log2AT         0.9323      1.0726    0.7492    1.1601
## RBC            1.3183      0.7586    1.1229    1.5476
```

```

## Liver_Only          1.3635      0.7334      0.9743      1.9082
## Cell_diff           2.0127      0.4968      1.4659      2.7636
## Mucin_TYPE          0.9044      1.1057      0.5881      1.3908
## SignetRING          1.1000      0.9091      0.6479      1.8675
## Lymphovascularinvasion 1.3067      0.7653      1.0421      1.6386
## perineural          0.9223      1.0842      0.7096      1.1987
## CT                  0.3976      2.5154      0.2504      0.6312
## RT                  1.1013      0.9080      0.8092      1.4988
## NACTRT              0.9819      1.0184      0.7250      1.3299
## Progress            0.2639      3.7887      0.1684      0.4138
## IntervalOR          0.9057      1.1041      0.8906      0.9211
## AJCC_bin            2.2229      0.4499      1.6019      3.0847
##
## Concordance= 0.804 (se = 0.011 )
## Likelihood ratio test= 432.5 on 26 df, p=<2e-16
## Wald test            = 320.9 on 26 df, p=<2e-16
## Score (logrank) test = 324.6 on 26 df, p=<2e-16

library(car)
vif(cox_model)

## Warning in vif.default(cox_model): No intercept: vifs may not be sensible.

##              Age              Gender              ASA3
##          1.314046          1.125176          1.365642
##              DM              CAD              HF
##          1.089694          1.136872          1.162036
##              CVA              CKD              LogCEA
##          1.139972          1.104450          1.156460
## Laparoscopic              TumorLOC              EA
##          1.043613          1.190184          1.055914
##              Log2AT              RBC              Liver_Only
##          1.291902          1.288758          2.544038
##              Cell_diff              Mucin_TYPE              SignetRING
##          1.185748          1.121536          1.305092
## Lymphovascularinvasion              perineural              CT
##          1.241031          1.255464          1.422660
##              RT              NACTRT              Progress
##          1.104047          1.138847          1.595181
##              IntervalOR              AJCC_bin
##          1.283020          2.544302

```

- AJCC_bin e ASA3 sono predittori forti e confermano le curve Kaplan-Meier e il log-rank test.
- LogCEA e Cell_diff forniscono informazioni cliniche aggiuntive.
- CT e Progress mostrano effetti protettivi, da interpretare in chiave clinica.

Variabile	HR	IC 95%	p-value	Interpretazione
ASA3	1.39	(1.10 – 1.77)	0.0066 (**)	Rischio +39% per ASA3=1 (alto rischio clinico)
LogCEA	1.34	(1.19 –	1e-06	Rischio +34% per unità di logCEA

Variabile	HR	IC 95%	p-value	Interpretazione
		1.50)	(***)	
RBC	1.32	(1.12 – 1.55)	0.0007 (***)	Rischio +32% per trasfusioni più alte
Cell_diff	2.01	(1.47 – 2.76)	1.5e-05 (***)	Rischio raddoppiato per differenziazione sfavorevole
Lymphovascularinvasion	1.31	(1.04 – 1.64)	0.0205 (*)	Rischio +31% con invasione linfovaskolare
CT (chemioterapia)	0.40	(0.25 – 0.63)	9e-05 (***)	Rischio -60%, effetto protettivo
Progress (progressione)	0.26	(0.17 – 0.41)	6e-09 (***)	Rischio -74%, effetto protettivo (attenzione: variabile temporale)
IntervalOR (tempo)	0.91	(0.89 – 0.92)	<2e-16 (***)	Effetto del tempo: con l'aumentare del follow-up, rischio decrescente
AJCC_bin	2.22	(1.60 – 3.08)	1.7e-06 (***)	Rischio più che raddoppiato per AJCC_bin=1

BONTA DEL MODELLO

```
summary(cox_model)$concordance
```

```
##           C           se(C)
## 0.80442454 0.01054447
```

modello discrimina molto bene, predice bene la sopravvivenza rispetto ai tempi osservati al 80%

TEST DI PROPORZIONALITA DEI RISCHI nel tempo

```
cox.zph(cox_model)
```

```
##               chisq df      p
## Age          2.03e-01  1 0.6527
## Gender       4.71e-02  1 0.8282
## ASA3         4.66e+00  1 0.0310
## DM           3.28e-01  1 0.5667
## CAD          2.82e-01  1 0.5957
## HF           1.05e-01  1 0.7459
## CVA          6.88e+00  1 0.0087
## CKD          4.47e-01  1 0.5037
## LogCEA       1.04e+00  1 0.3076
## Laparoscopic 1.04e+00  1 0.3067
## TumorLOC     4.81e-01  1 0.4882
## EA           9.23e-01  1 0.3368
## Log2AT       5.32e-03  1 0.9419
## RBC          1.20e+00  1 0.2725
## Liver_Only   7.78e-04  1 0.9777
## Cell_diff    3.12e+00  1 0.0773
## Mucin_TYPE   8.43e-01  1 0.3586
## SignetRING   1.55e+00  1 0.2133
```

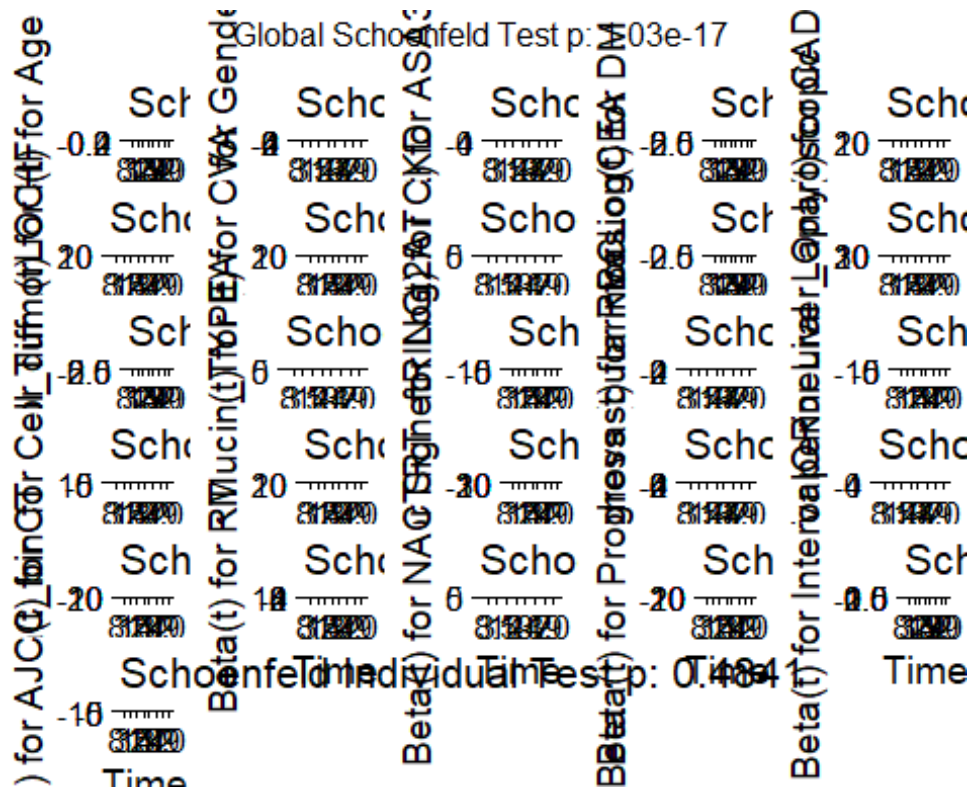
```
## Lymphovascularinvasion 7.66e+00 1 0.0057
## perineural 1.90e+00 1 0.1683
## CT 2.02e+00 1 0.1548
## RT 8.12e+00 1 0.0044
## NACTRT 1.40e+00 1 0.2363
## Progress 7.87e+00 1 0.0050
## IntervalOR 7.58e+01 1 <2e-16
## AJCC_bin 4.90e-01 1 0.4841
## GLOBAL 1.41e+02 26 <2e-16
```

GLOBAL: il modello, nel suo complesso, viola la ipotesi di proporzionalità dei rischi :(

Si visualizzana la curva di SCHOENFELD per una diagnosi grafica

```
cox.zph_test <- cox.zph(cox_model)
```

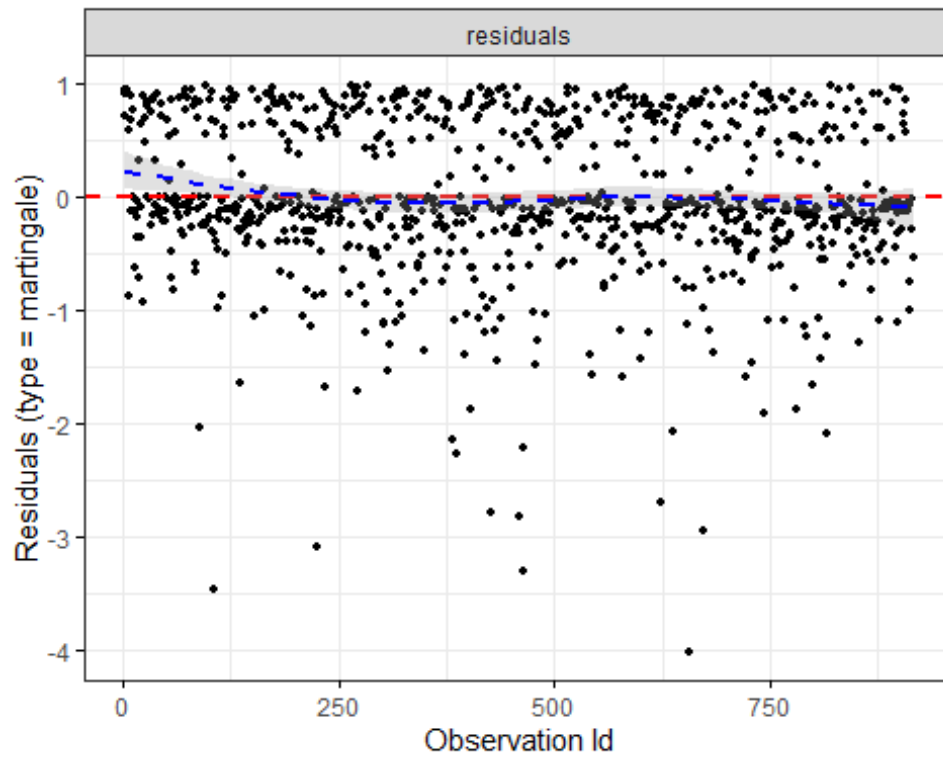
```
ggcoxzph(cox.zph_test)
```



```
# Residui martingala (per outliers)
```

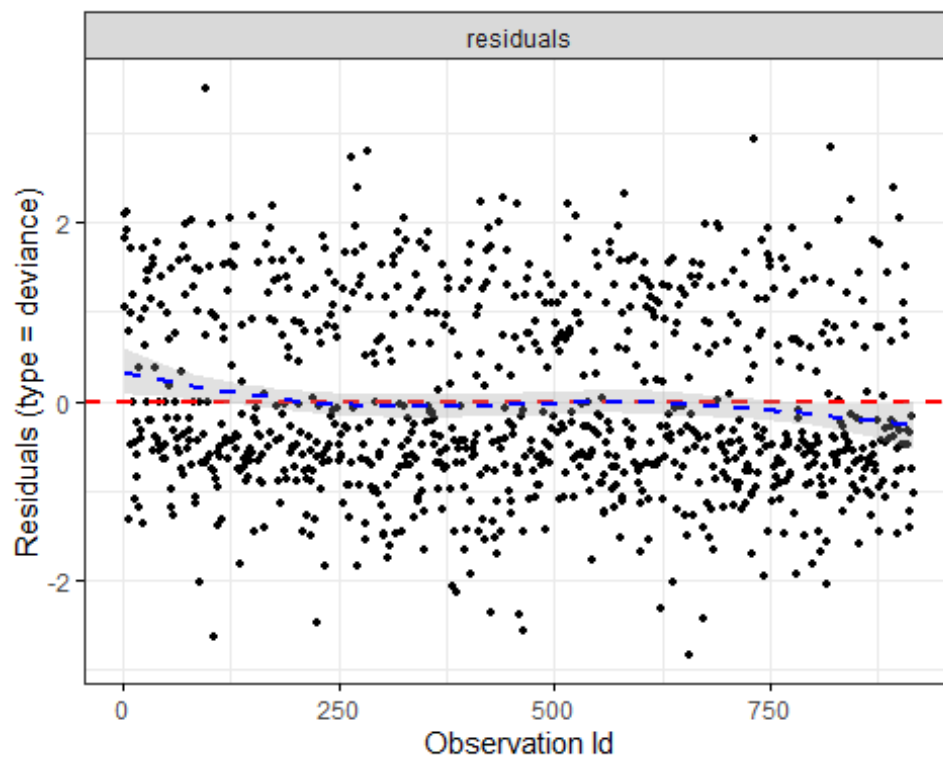
```
ggcoxdiagnostics(cox_model, type = "martingale", linear.predictions = FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

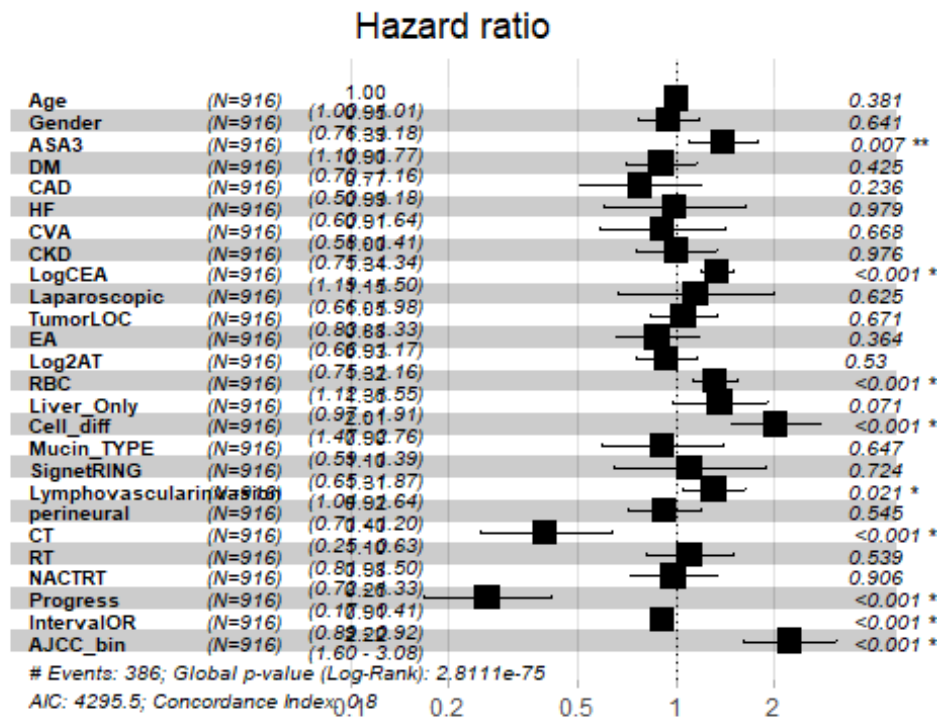


```
# Residui deviance (per influenza)
ggcoxdiagnostics(cox_model, type = "deviance", linear.predictions = FALSE)

## `geom_smooth()` using formula = 'y ~ x'
```



```
# Forest plot
ggforest(cox_model, data = data)
```



IPOTESI DI PROPORZIONALITÀ VIOLATA PER ALMENO UNA VARIABILE! Martingale residuals mostrano punti fuori dalla fascia ± 3 : possibili outlier o influenze estreme. Deviance residuals confermano questa tendenza: controlla osservazioni con residui estremi.

Variabili significative (es. con $p < 0.05$) mostrano:

- $HR > 1 \rightarrow$ aumento del rischio (ASA3, LogCEA, RBC, Cell_diff, AJCC_bin);
- $HR < 1 \rightarrow$ effetto protettivo (CT, Progress, IntervalOR).

Variabili come Gender, TumorLOC non significative.

SEMPLIFICAZIONE COX - SIGNIFICATIVE VARIABLES

```
cox_simplified <- coxph(Surv(IntervalOD, Death) ~ ASA3 + LogCEA + RBC +
  Cell_diff + Lymphovascularinvasion + CT +
  Progress + IntervalOR + AJCC_bin, data = data)
```

```
summary(cox_simplified)
```

```
## Call:
```

```
## coxph(formula = Surv(IntervalOD, Death) ~ ASA3 + LogCEA + RBC +
##   Cell_diff + Lymphovascularinvasion + CT + Progress + IntervalOR +
##   AJCC_bin, data = data)
```

```
##
```

```
## n= 916, number of events= 386
```



```
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## ASA31          0.286480  1.331732  0.108910   2.630 0.008528 **
## LogCEA         0.316300  1.372042  0.056465   5.602 2.12e-08 ***
## RBC            0.275427  1.317093  0.074254   3.709 0.000208 ***
## Cell_diff      0.683911  1.981614  0.155899   4.387 1.15e-05 ***
## Lymphovascularinvasion 0.278332  1.320924  0.106982   2.602 0.009277 **
## CT            -0.973865  0.377621  0.226980  -4.291 1.78e-05 ***
## Progress       -1.300823  0.272308  0.226292  -5.748 9.01e-09 ***
## IntervalOR     -0.097307  0.907277  0.008486 -11.466 < 2e-16 ***
## AJCC_bin1      0.589113  1.802389  0.107473   5.482 4.22e-08 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##               exp(coef) exp(-coef) lower .95 upper .95
## ASA31          1.3317      0.7509   1.0758   1.6486
## LogCEA         1.3720      0.7288   1.2283   1.5326
## RBC            1.3171      0.7592   1.1387   1.5234
## Cell_diff      1.9816      0.5046   1.4599   2.6898
## Lymphovascularinvasion 1.3209      0.7570   1.0711   1.6291
## CT            0.3776      2.6482   0.2420   0.5892
## Progress       0.2723      3.6723   0.1748   0.4243
## IntervalOR     0.9073      1.1022   0.8923   0.9225
## AJCC_bin1      1.8024      0.5548   1.4600   2.2250
```

```
##
## Concordance= 0.804 (se = 0.011 )
## Likelihood ratio test= 422.4 on 9 df, p=<2e-16
## Wald test              = 316.2 on 9 df, p=<2e-16
## Score (logrank) test = 314 on 9 df, p=<2e-16
```

```
exp(coef(cox_simplified))          # HR
```

```
##              ASA31              LogCEA              RBC
##              1.3317320              1.3720417              1.3170933
##              Cell_diff Lymphovascularinvasion              CT
##              1.9816136              1.3209243              0.3776206
##              Progress              IntervalOR              AJCC_bin1
##              0.2723076              0.9072773              1.8023894
```

```
exp(confint(cox_simplified))          # IC 95%
```

```
##              2.5 %      97.5 %
## ASA31          1.0757527  1.6486225
## LogCEA         1.2283000  1.5326048
## RBC            1.1387057  1.5234268
## Cell_diff      1.4598809  2.6898033
## Lymphovascularinvasion 1.0710616  1.6290762
## CT            0.2420194  0.5891977
## Progress       0.1747590  0.4243067
## IntervalOR     0.8923114  0.9224942
## AJCC_bin1      1.4600500  2.2249976
```



```
summary(cox_stratified)

## Call:
## coxph(formula = Surv(IntervalOD, Death) ~ ASA3 + LogCEA + RBC +
##       Cell_diff + Lymphovascularinvasion + CT + strata(Progress) +
##       strata(AJCC_bin) + strata(IntervalOR), data = data)
##
##      n= 916, number of events= 386
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## ASA31          0.6342    1.8856   0.3054  2.077  0.03784 *
## LogCEA          0.4213    1.5239   0.1572  2.679  0.00738 **
## RBC             0.3061    1.3581   0.1925  1.590  0.11185
## Cell_diff       0.5225    1.6863   0.4722  1.107  0.26849
## Lymphovascularinvasion -0.1475    0.8628   0.2891 -0.510  0.60988
## CT             -0.7464    0.4741   0.5537 -1.348  0.17769
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## ASA31          1.8856    0.5303    1.0363    3.431
## LogCEA          1.5239    0.6562    1.1197    2.074
## RBC             1.3581    0.7363    0.9312    1.981
## Cell_diff       1.6863    0.5930    0.6683    4.255
## Lymphovascularinvasion 0.8628    1.1590    0.4896    1.521
## CT             0.4741    2.1093    0.1601    1.403
##
## Concordance= 0.626 (se = 0.059 )
## Likelihood ratio test= 23.16 on 6 df,  p=7e-04
## Wald test              = 18.62 on 6 df,  p=0.005
## Score (logrank) test = 21.61 on 6 df,  p=0.001

cox.zph(cox_stratified)

##              chisq df      p
## ASA3          7.26e-04 1 0.979
## LogCEA         1.04e+00 1 0.309
## RBC            4.51e+00 1 0.034
## Cell_diff      1.93e-01 1 0.660
## Lymphovascularinvasion 5.32e+00 1 0.021
## CT             3.71e+00 1 0.054
## GLOBAL         1.47e+01 6 0.022
```

- **LogCEA** e **ASA3** rimangono significativi anche nel modello stratificato.
- La stratificazione su **AJCC_bin** e **IntervalOR** riduce l'effetto apparente di altre variabili, perché il modello assume che queste abbiano effetti di base separati.

NOTA BENE: C = 68.1 %, la concordanza é scesa ripstto al modello precedente perche la stratificazione elimina l'effetto delle variabili stratificate dal calcolo del rischio.

TIME DEPENDENT

IMPOSTANDO PROGRESSO 0 PRIMA DEL PROGRESSO E PROGRESSO 1 DOPO IL PROGRESSO.

Intervallo da O (operazione) a R (recidiva) é il tempo continuo di follow up prima del progresso

```
library(survival)
```

```
# Crea ID univoco per pazienti
```

```
data$id <- 1:nrow(data)
```

```
# Assicuriamoci che Progress sia 1=evento, 0=non avvenuto
```

```
data$Progress_event <- ifelse(data$Progress == 1, 1, 0)
```

```
# Step 1: dataset iniziale per morte
```

```
td_data <- tmerge(data1 = data, data2 = data,  
                  id = id,  
                  death = event(IntervalOD, Death))
```

```
# Step 2: aggiungi Progress come evento time-dependent
```

```
td_data <- tmerge(data1 = td_data, data2 = data,  
                  id = id,  
                  Progress_td = tdc(IntervalOR))
```

```
# Step 3: controlla il dataset risultante
```

```
head(td_data)
```

```
##   Age Gender ASA3 DM CAD HF CVA CKD   LogCEA Laparoscopic TumorLOC EA   Log2AT  
## 1  52      1    1  0  0  0  0  0 2.971740              0      0  0 8.491853  
## 2  52      1    1  0  0  0  0  0 2.971740              0      0  0 8.491853  
## 3  85      2    0  0  1  0  0  0 2.127105              0      0  1 8.714246  
## 4  85      2    0  0  1  0  0  0 2.127105              0      0  1 8.714246  
## 5  45      2    0  0  0  0  0  0 3.057666              0      0  0 8.154818  
## 6  45      2    0  0  0  0  0  0 3.057666              0      0  0 8.154818  
##   RBC Liver_Only Cell_diff Mucin_TYPE SignetRING Lymphovascularinvasion  
## 1   0           0          1           0           0                      1  
## 2   0           0          1           0           0                      1  
## 3   0           1          1           0           0                      0  
## 4   0           1          1           0           0                      0  
## 5   1           1          1           0           0                      1  
## 6   1           1          1           0           0                      1  
##   perineural CT RT NACTRT Death Progress IntervalOD IntervalOR AJCC_bin id  
## 1           0  1  0      0      1      1 11.761807  9.856263      1  1  
## 2           0  1  0      0      1      1 11.761807  9.856263      1  1  
## 3           0  1  0      0      1      1  8.377823  7.063655      0  2  
## 4           0  1  0      0      1      1  8.377823  7.063655      0  2  
## 5           0  1  0      0      1      1  4.763860  2.792608      0  3  
## 6           0  1  0      0      1      1  4.763860  2.792608      0  3  
##   Progress_event   tstart   tstop death Progress_td  
## 1               1 0.000000 9.856263    0          0  
## 2               1 9.856263 11.761807    1          1  
## 3               1 0.000000 7.063655    0          0
```

```
## 4          1 7.063655 8.377823      1      1
## 5          1 0.000000 2.792608      0      0
## 6          1 2.792608 4.763860      1      1

cox_td <- coxph(Surv(tstart, tstop, death) ~ ASA3 + LogCEA + RBC +
                Cell_diff + Lymphovascularinvasion + CT + Progress_td +
                strata(AJCC_bin), data = td_data)

summary(cox_td)

## Call:
## coxph(formula = Surv(tstart, tstop, death) ~ ASA3 + LogCEA +
##       RBC + Cell_diff + Lymphovascularinvasion + CT + Progress_td +
##       strata(AJCC_bin), data = td_data)
##
## n= 1624, number of events= 386
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## ASA31          0.27600   1.31785  0.10903  2.531  0.01136 *
## LogCEA          0.33009   1.39110  0.05639  5.854 4.80e-09 ***
## RBC             0.21258   1.23686  0.07460  2.850  0.00438 **
## Cell_diff       0.65427   1.92374  0.15529  4.213 2.52e-05 ***
## Lymphovascularinvasion 0.27871   1.32142  0.10809  2.579  0.00992 **
## CT             -1.49229   0.22486  0.20297 -7.352 1.95e-13 ***
## Progress_td      2.37913  10.79554  0.19383 12.275 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## ASA31          1.3178    0.75881    1.0643    1.6318
## LogCEA          1.3911    0.71886    1.2455    1.5537
## RBC             1.2369    0.80850    1.0686    1.4316
## Cell_diff       1.9237    0.51982    1.4189    2.6082
## Lymphovascularinvasion 1.3214    0.75676    1.0691    1.6332
## CT              0.2249    4.44729    0.1511    0.3347
## Progress_td     10.7955    0.09263    7.3834   15.7845
##
## Concordance= 0.772 (se = 0.013 )
## Likelihood ratio test= 355.2 on 7 df,  p=<2e-16
## Wald test              = 294.6 on 7 df,  p=<2e-16
## Score (logrank) test = 330.6 on 7 df,  p=<2e-16

table(td_data$Progress_td)

##
##    0    1
## 916 708

cox.zph(cox_td)

##              chisq df      p
## ASA3          6.051  1 0.0139
```

```
## LogCEA          0.604  1 0.4372
## RBC             0.170  1 0.6798
## Cell_diff       5.245  1 0.0220
## Lymphovascularinvasion 5.949  1 0.0147
## CT              1.813  1 0.1782
## Progress_td      5.743  1 0.0166
## GLOBAL          22.943  7 0.0017
```

ASA3 VIOLA, CELL_DIFF VIOLA, INVASIONE LINFO VIOLA, COMPLESSIVAMENTE VIOLATE

PENALIZED COX

Si utilizza glmnet per penalizzazione ridge per ridurre varianza gestire multicoll. COX base é instabile, i coef saranno estretti con exp() per ottenere HR

```
library(glmnet)
```

```
## Warning: package 'glmnet' was built under R version 4.3.3
```

```
## Loading required package: Matrix
```

```
##
```

```
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
```

```
##
```

```
##      expand, pack, unpack
```

```
## Loaded glmnet 4.1-8
```

```
library(survival)
```

```
# Prepara la matrice X e il vettore y (tempo + evento)
```

```
X <- model.matrix(~ ASA3 + LogCEA + RBC + Cell_diff + Lymphovascularinvasion + CT
+ Progress_td, data=td_data)[, -1]
```

```
y <- Surv(td_data$ststart, td_data$stop, td_data$death)
```

```
# Cox Lasso (alpha=1), puoi anche usare Ridge (alpha=0) o Elastic Net
```

```
cvfit <- cv.glmnet(X, y, family="cox", alpha=1)
```

```
## Warning: cox.fit: algorithm did not converge
```

```
## Warning: cox.fit: algorithm did not converge
```

```
## Warning: cox.fit: algorithm did not converge
```

```
## Warning: cox.fit: algorithm did not converge
```

```
## Warning: cox.fit: algorithm did not converge
```

```
## Warning: cox.fit: algorithm did not converge
```

```
## Warning: cox.fit: algorithm did not converge
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

## Warning: cox.fit: algorithm did not converge
## Warning: cox.fit: algorithm did not converge
## Warning: cox.fit: algorithm did not converge
## Warning: cox.fit: algorithm did not converge
## Warning: cox.fit: algorithm did not converge
## Warning: cox.fit: algorithm did not converge
## Warning: cox.fit: algorithm did not converge
## Warning: cox.fit: algorithm did not converge
## Warning: cox.fit: algorithm did not converge
## Warning: cox.fit: algorithm did not converge
# Visualizza i coefficienti penalizzati
coef(cvfit, s="lambda.min")

## 7 x 1 sparse Matrix of class "dgCMatrix"
##              1
## ASA31          0.2430918
## LogCEA          0.3399611
## RBC             0.2108039
## Cell_diff       0.7074687
## Lymphovascularinvasion 0.3425381
## CT             -1.5345704
## Progress_td      2.4104211

```

RIDGE PENALIZED COX: i coefficienti stimate rappresentano l'effetto delle covariate sul logaritmo del rischio di un evento. Il coefficiente positivo indica un aumento del rischio associato all'aumento della covariata, negativo viceversa.

COEFFICIENTI: ASA31: 0.243 \rightarrow HR $\approx \exp(0.243) \approx 1.275$

LogCEA: 0.340 \rightarrow HR $\approx \exp(0.340) \approx 1.405$

RBC: 0.211 \rightarrow HR $\approx \exp(0.211) \approx 1.235$

Cell_diff: 0.707 \rightarrow HR $\approx \exp(0.707) \approx 2.028$

Lymphovascularinvasion: 0.343 \rightarrow HR $\approx \exp(0.343) \approx 1.409$

CT: -1.535 \rightarrow HR $\approx \exp(-1.535) \approx 0.216$, coefficiente negativo, suggerendo un effetto protettivo.

Progress_td: 2.410 \rightarrow HR $\approx \exp(2.410) \approx 11.13$, effetto molto forte, con un hazard ratio (HR) superiore a 11, indicando che la progressione della malattia è fortemente associata a un aumento del rischio di morte.

Additive Aalen

permette di stimare effetti che variano nel tempo

```
library(timereg)

## Warning: package 'timereg' was built under R version 4.3.3

aalen_model <- aalen(Surv(tstart, tstop, death) ~ ASA3 + LogCEA + RBC + Cell_diff
+
                    Lymphovascularinvasion + CT + Progress_td, data=td_data)
summary(aalen_model)

## Additive Aalen Model
##
## Test for nonparametric terms
##
## Test for non-significant effects
##
##          Supremum-test of significance p-value H_0: B(t)=0
## (Intercept)                1.30                0.939
## ASA31                      3.73                0.006
## LogCEA                     5.63                0.000
## RBC                        3.67                0.012
## Cell_diff                   4.11                0.000
## Lymphovascularinvasion      3.96                0.003
## CT                         3.86                0.005
## Progress_td                 9.45                0.000
##
## Test for time invariant effects
##
##          Kolmogorov-Smirnov test
## (Intercept)                0.360
## ASA31                      0.171
## LogCEA                     0.262
## RBC                        0.139
## Cell_diff                   0.556
## Lymphovascularinvasion      0.358
## CT                         0.663
## Progress_td                 0.659
##
##          p-value H_0:constant effect
## (Intercept)                0.407
## ASA31                      0.614
## LogCEA                     0.021
## RBC                        0.146
## Cell_diff                   0.000
## Lymphovascularinvasion      0.001
## CT                         0.002
## Progress_td                 0.000
##
##          Cramer von Mises test
## (Intercept)                3.670
## ASA31                      0.427
## LogCEA                     2.310
## RBC                        0.361
```

```

## Cell_diff                9.100
## Lymphovascularinvasion   5.210
## CT                       17.600
## Progress_td              14.900
##                p-value H_0:constant effect
## (Intercept)             0.241
## ASA31                   0.691
## LogCEA                  0.017
## RBC                    0.192
## Cell_diff               0.000
## Lymphovascularinvasion  0.000
## CT                     0.002
## Progress_td             0.000
##
##
## Call:
## aalen(formula = Surv(tstart, tstop, death) ~ ASA3 + LogCEA +
##       RBC + Cell_diff + Lymphovascularinvasion + CT + Progress_td,
##       data = td_data)

```

1- TEST SIGNIFICATIVITÀ

test del supremum mostra che le seguenti variabili hanno effetti significativamente diversi da zero:

ASA31: $p = 0.013$

LogCEA: $p < 0.001$

RBC: $p = 0.008$

Cell_diff: $p < 0.001$

Lymphovascularinvasion: $p = 0.001$

CT: $p = 0.003$

Progress_td: $p < 0.001$

2 - TEST COSTANZA EFFETTO NEL TEMPO

Kolmodoroff smirnov, indica che LOGCEA, CELL DIFF, LINFO VASCOLARIZZAZIONE, CT E PROGRESS TD anno effetti che variano nel tempo

Progress_td mostra un effetto significativamente variabile nel tempo, suggerendo che l'impatto della progressione della malattia sul rischio di morte cambia durante il follow-up.

CT e Cell_diff presentano effetti temporali non costanti, indicando che il loro impatto sul rischio varia nel tempo.

Supervised learning SENZA BILANCIAMENTO

Logistic Regression

Random Forest

XGBoost

SVM

K-NN

DIVISIONE TRAIN/TEST

```
set.seed(123)
library(caret)

## Warning: package 'caret' was built under R version 4.3.3

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:survival':
##
##      cluster

train_index <- createDataPartition(data$Death, p = 0.8, list = FALSE)
train_data <- data[train_index, ]
test_data <- data[-train_index, ]

#-----aggiunta post
train_data$Death <- as.factor(train_data$Death)
test_data$Death <- as.factor(test_data$Death)
```

LOGISTIC REGRESSION

```
# Training della Logistic regression
set.seed(123)
logit_model <- train(
  Death ~ .,
  data = train_data,
  method = "glm",
  family = "binomial",
  trControl = trainControl(method = "none") # niente cross-validation per ora,
  step successivo
)

# Predizione delle classi
pred_class <- predict(logit_model, newdata = test_data)

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```

# Predizione delle probabilità (per AUC)
pred_prob <- predict(logit_model, newdata = test_data, type = "prob")

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

# Confusion Matrix
conf_mat <- confusionMatrix(pred_class, test_data$Death)
print(conf_mat)

## Confusion Matrix and Statistics
##
##              Reference
## Prediction  0   1
##           0 72 41
##           1 39 31
##
##              Accuracy : 0.5628
##              95% CI : (0.4877, 0.6359)
##      No Information Rate : 0.6066
##      P-Value [Acc > NIR] : 0.9003
##
##              Kappa : 0.0796
##
##  Mcnemar's Test P-Value : 0.9110
##
##              Sensitivity : 0.6486
##              Specificity : 0.4306
##      Pos Pred Value : 0.6372
##      Neg Pred Value : 0.4429
##      Prevalence : 0.6066
##      Detection Rate : 0.3934
##      Detection Prevalence : 0.6175
##      Balanced Accuracy : 0.5396
##
##      'Positive' Class : 0
##

# AUC con pROC
library(pROC)

## Warning: package 'pROC' was built under R version 4.3.3

## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##      cov, smooth, var

roc_obj <- roc(test_data$Death, pred_prob[, 2])

```

```
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

auc(roc_obj)

## Area under the curve: 0.6196

summary(logit_model$finalModel)

##
## Call:
## NULL
##
## Coefficients: (1 not defined because of singularities)
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.5597915   1.7240337   -0.325  0.745409
## Age           -0.0097777   0.0068488   -1.428  0.153393
## Gender2        0.1344741   0.1779651    0.756  0.449877
## ASA31          0.2218841   0.1952950    1.136  0.255895
## DM             0.2961600   0.2068710    1.432  0.152254
## CAD           -0.2474317   0.3456455   -0.716  0.474082
## HF            -0.0496266   0.3887257   -0.128  0.898414
## CVA            0.9000041   0.3824615    2.353  0.018613 *
## CKD            0.5046835   0.2503024    2.016  0.043769 *
## LogCEA         0.1537012   0.0967441    1.589  0.112119
## Laparoscopic  -0.3717732   0.4198880   -0.885  0.375935
## TumorLOC1      0.0604544   0.1897793    0.319  0.750067
## EA            -0.2909285   0.2479113   -1.174  0.240588
## Log2AT        -0.0680888   0.1878700   -0.362  0.717035
## RBC           0.3470738   0.1334364    2.601  0.009294 **
## Liver_Only     0.3794461   0.2512146    1.510  0.130930
## Cell_diff      0.2450865   0.2691399    0.911  0.362491
## Mucin_TYPE     -0.3372062   0.3443313   -0.979  0.327428
## SignetRING      0.2779922   0.4411310    0.630  0.528576
## Lymphovascularinvasion -0.1175056   0.1855072   -0.633  0.526454
## perineural      0.0316183   0.2145761    0.147  0.882854
## CT             0.1569898   0.3233707    0.485  0.627336
## RT             0.4139675   0.2627090    1.576  0.115080
## NACTRT         -0.2536589   0.2445127   -1.037  0.299547
## Progress       1.3646557   0.2769138    4.928  8.3e-07 ***
## IntervalOD     -0.0197915   0.0054928   -3.603  0.000314 ***
## IntervalOR     -0.0155006   0.0114291   -1.356  0.175025
## AJCC_bin1      0.4593299   0.2504911    1.834  0.066696 .
## id            -0.0008913   0.0003675   -2.425  0.015297 *
## Progress_event      NA          NA          NA          NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1001.06  on 732  degrees of freedom
```

```
## Residual deviance: 865.88 on 704 degrees of freedom
## AIC: 923.88
##
## Number of Fisher Scoring iterations: 5
```

Si nota che la maggior parte delle variabili non sono statisticamente significative. Si utilizza la funzione `step()` per fare una stepwise selection

```
# 1. Modello iniziale completo
```

```
full_model <- glm(Death ~ ., data = train_data, family = "binomial")
```

```
# 2. Selezione stepwise (AIC)
```

```
step_model <- step(full_model, direction = "both", trace = FALSE)
```

```
# 3. Sommario del modello selezionato
```

```
summary(step_model)
```

```
##
```

```
## Call:
```

```
## glm(formula = Death ~ CVA + CKD + LogCEA + RBC + RT + Progress +  
##      IntervalOD + IntervalOR + AJCC_bin + id, family = "binomial",  
##      data = train_data)
```

```
##
```

```
## Coefficients:
```

```
##      Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -1.1814318  0.3309172 -3.570 0.000357 ***  
## CVA          0.8725531  0.3646286  2.393 0.016712 *  
## CKD          0.4153874  0.2424035  1.714 0.086599 .  
## LogCEA       0.1896105  0.0910271  2.083 0.037250 *  
## RBC          0.3716770  0.1215958  3.057 0.002238 **  
## RT           0.3918939  0.2577716  1.520 0.128432  
## Progress     1.3520654  0.2450282  5.518 3.43e-08 ***  
## IntervalOD   -0.0192654  0.0052142 -3.695 0.000220 ***  
## IntervalOR   -0.0156049  0.0112493 -1.387 0.165384  
## AJCC_bin1     0.2468352  0.1696257  1.455 0.145621  
## id           -0.0007727  0.0003093 -2.498 0.012488 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
```

```
##      Null deviance: 1001.06 on 732 degrees of freedom
```

```
## Residual deviance: 878.75 on 722 degrees of freedom
```

```
## AIC: 900.75
```

```
##
```

```
## Number of Fisher Scoring iterations: 5
```

```
# Predizioni su test
```

```
pred_prob <- predict(step_model, newdata = test_data, type = "response")
```

```
pred_class <- ifelse(pred_prob > 0.5, "1", "0")
```

```
pred_class <- factor(pred_class, levels = levels(test_data$Death))
```

```

# Confusion matrix
confusionMatrix(pred_class, test_data$Death)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0   1
##           0  74  38
##           1  37  34
##
##               Accuracy : 0.5902
##               95% CI : (0.5152, 0.6622)
##       No Information Rate : 0.6066
##       P-Value [Acc > NIR] : 0.7031
##
##               Kappa : 0.1392
##
##  Mcnemar's Test P-Value : 1.0000
##
##       Sensitivity : 0.6667
##       Specificity : 0.4722
##       Pos Pred Value : 0.6607
##       Neg Pred Value : 0.4789
##       Prevalence : 0.6066
##       Detection Rate : 0.4044
##       Detection Prevalence : 0.6120
##       Balanced Accuracy : 0.5694
##
##       'Positive' Class : 0
##

# AUC
library(pROC)
roc_obj <- roc(test_data$Death, pred_prob)

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

auc(roc_obj)

## Area under the curve: 0.6418

```

RANDOM FOREST

Si inizia con una k fold

```

# Rinominare i Livelli della variabile target
train_data$Death <- factor(train_data$Death, levels = c("0", "1"), labels =
c("No", "Yes"))
test_data$Death <- factor(test_data$Death, levels = c("0", "1"), labels =
c("No", "Yes"))

```

```

library(caret)
set.seed(123)

rf_model <- train(
  Death ~ .,
  data = train_data,
  method = "rf",
  trControl = trainControl(
    method = "cv",
    number = 5,
    classProbs = TRUE,
    summaryFunction = twoClassSummary
  ),
  metric = "ROC"
)

# Predizioni
pred_class <- predict(rf_model, newdata = test_data)
pred_prob <- predict(rf_model, newdata = test_data, type = "prob")

# Confusion matrix
confusionMatrix(pred_class, test_data$Death)

## Confusion Matrix and Statistics
##
##              Reference
## Prediction No Yes
##          No  86  40
##          Yes  25  32
##
##              Accuracy : 0.6448
##              95% CI : (0.5708, 0.714)
##      No Information Rate : 0.6066
##      P-Value [Acc > NIR] : 0.16270
##
##              Kappa : 0.2275
##
##  Mcnemar's Test P-Value : 0.08248
##
##              Sensitivity : 0.7748
##              Specificity : 0.4444
##              Pos Pred Value : 0.6825
##              Neg Pred Value : 0.5614
##              Prevalence : 0.6066
##              Detection Rate : 0.4699
##      Detection Prevalence : 0.6885
##              Balanced Accuracy : 0.6096
##
##              'Positive' Class : No
##

```

```

# AUC
library(pROC)
roc_obj <- roc(test_data$Death, pred_prob[, "Yes"]) # o [, "1"] se usi 0/1

## Setting levels: control = No, case = Yes

## Setting direction: controls < cases

auc(roc_obj)

## Area under the curve: 0.6642

varImp(rf_model)

## rf variable importance
##
##    only 20 most important variables shown (out of 29)
##
##
##              Overall
## IntervalOR      100.000
## IntervalOD       99.214
## LogCEA          88.674
## id              88.407
## Age             75.054
## Log2AT          68.955
## RBC             29.784
## Progress        22.081
## Progress_event  19.164
## AJCC_bin1       16.530
## ASA31           14.997
## Gender2         11.822
## DM              11.093
## Lymphovascularinvasion 10.148
## TumorLOC1        9.808
## perineural       9.678
## Liver_Only       9.314
## CKD              8.606
## RT               7.986
## EA               6.942

```

Si procede facendo tuning degli iperparametri, con mtry, numero di alberi etc. dopo il fit di base.

```

set.seed(123)
mtry_grid <- expand.grid(mtry = c(2, 4, 6, 8, 10))

rf_model_tuned <- train(
  Death ~ .,
  data = train_data,
  method = "rf",
  trControl = trainControl(
    method = "cv",
    number = 5,
    classProbs = TRUE,

```

```
summaryFunction = twoClassSummary,  
verboseIter = TRUE  
)  
metric = "ROC",  
tuneGrid = mtry_grid,  
ntree = 500  
)
```

```
## + Fold1: mtry= 2  
## - Fold1: mtry= 2  
## + Fold1: mtry= 4  
## - Fold1: mtry= 4  
## + Fold1: mtry= 6  
## - Fold1: mtry= 6  
## + Fold1: mtry= 8  
## - Fold1: mtry= 8  
## + Fold1: mtry=10  
## - Fold1: mtry=10  
## + Fold2: mtry= 2  
## - Fold2: mtry= 2  
## + Fold2: mtry= 4  
## - Fold2: mtry= 4  
## + Fold2: mtry= 6  
## - Fold2: mtry= 6  
## + Fold2: mtry= 8  
## - Fold2: mtry= 8  
## + Fold2: mtry=10  
## - Fold2: mtry=10  
## + Fold3: mtry= 2  
## - Fold3: mtry= 2  
## + Fold3: mtry= 4  
## - Fold3: mtry= 4  
## + Fold3: mtry= 6  
## - Fold3: mtry= 6  
## + Fold3: mtry= 8  
## - Fold3: mtry= 8  
## + Fold3: mtry=10  
## - Fold3: mtry=10  
## + Fold4: mtry= 2  
## - Fold4: mtry= 2  
## + Fold4: mtry= 4  
## - Fold4: mtry= 4  
## + Fold4: mtry= 6  
## - Fold4: mtry= 6  
## + Fold4: mtry= 8  
## - Fold4: mtry= 8  
## + Fold4: mtry=10  
## - Fold4: mtry=10  
## + Fold5: mtry= 2  
## - Fold5: mtry= 2  
## + Fold5: mtry= 4
```



```

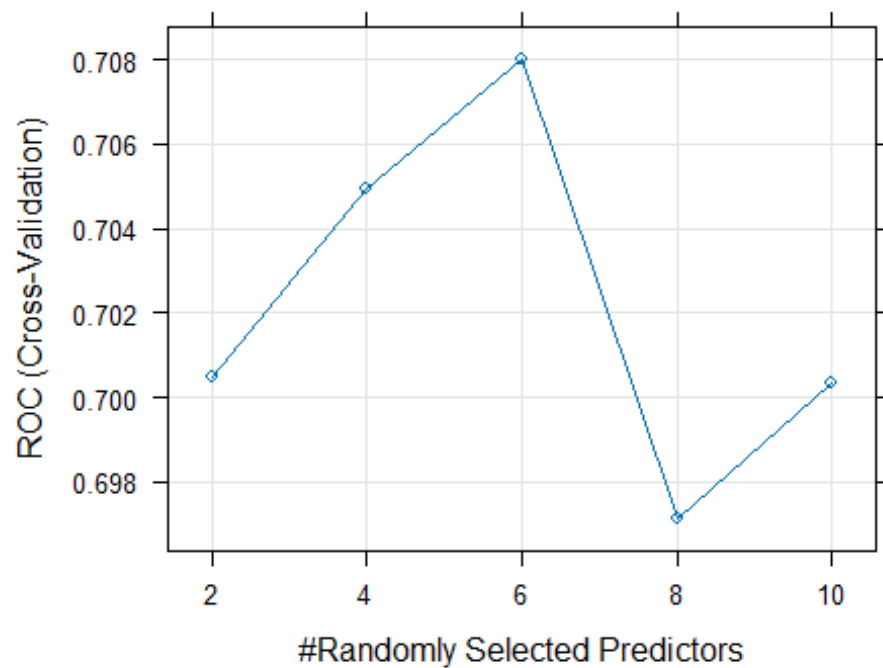
## - Fold5: mtry= 4
## + Fold5: mtry= 6
## - Fold5: mtry= 6
## + Fold5: mtry= 8
## - Fold5: mtry= 8
## + Fold5: mtry=10
## - Fold5: mtry=10
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 6 on full training set

# Visualizza la tabella con performance per ciascun mtry
print(rf_model_tuned)

## Random Forest
##
## 733 samples
## 29 predictor
## 2 classes: 'No', 'Yes'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 586, 586, 587, 587, 586
## Resampling results across tuning parameters:
##
##   mtry  ROC      Sens      Spec
##   2     0.7004742  0.7899885  0.4236047
##   4     0.7049238  0.7255881  0.5414747
##   6     0.7079994  0.7088640  0.5701485
##   8     0.6971334  0.7088067  0.5574501
##  10     0.7003616  0.7111876  0.5764465
##
## ROC was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 6.

plot(rf_model_tuned)

```



```
# Predizioni
pred_rf <- predict(rf_model_tuned, newdata = test_data)
probs_rf <- predict(rf_model_tuned, newdata = test_data, type = "prob")

# Confusion matrix
confusionMatrix(pred_rf, test_data$Death)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction No Yes
##      No   78   34
##      Yes  33   38
##
##              Accuracy : 0.6339
##              95% CI : (0.5596, 0.7037)
##      No Information Rate : 0.6066
##      P-Value [Acc > NIR] : 0.2489
##
##              Kappa : 0.231
##
##  Mcnemar's Test P-Value : 1.0000
##
##              Sensitivity : 0.7027
##              Specificity : 0.5278
##      Pos Pred Value : 0.6964
##      Neg Pred Value : 0.5352
##      Prevalence : 0.6066
```

```

##      Detection Rate : 0.4262
##      Detection Prevalence : 0.6120
##      Balanced Accuracy : 0.6152
##
##      'Positive' Class : No
##

# AUC
library(pROC)
roc_obj <- roc(test_data$Death, probs_rf[, "Yes"])

## Setting levels: control = No, case = Yes

## Setting direction: controls < cases

auc(roc_obj)

## Area under the curve: 0.6714

varImp(rf_model_tuned)

## rf variable importance
##
##      only 20 most important variables shown (out of 29)
##
##
##      Overall
## Interval0D      100.000
## Interval0R       92.153
## id              88.664
## LogCEA          87.326
## Age             73.749
## Log2AT          65.815
## RBC             21.316
## Progress_event  13.317
## AJCC_bin1       12.996
## Progress        10.728
## ASA31           10.635
## Gender2         10.185
## Liver_Only       9.065
## Lymphovascularinvasion 8.709
## DM              8.620
## TumorLOC1        8.598
## perineural       7.625
## NACTRT           6.439
## CKD              6.406
## RT               6.024

```

XGBoost

```

train_data$Death <- factor(train_data$Death, levels = c("No", "Yes"))
test_data$Death <- factor(test_data$Death, levels = c("No", "Yes"))

set.seed(123)

```

```

xgb_grid <- expand.grid(
  nrounds = c(100, 200),
  max_depth = c(3, 6),
  eta = c(0.01, 0.1, 0.3),
  gamma = 0,
  colsample_bytree = 1,
  min_child_weight = 1,
  subsample = 1
)

xgb_model <- train(
  Death ~ .,
  data = train_data,
  method = "xgbTree",
  trControl = trainControl(
    method = "cv",
    number = 5,
    classProbs = TRUE,
    summaryFunction = twoClassSummary,
    verboseIter = TRUE
  ),
  metric = "ROC",
  tuneGrid = xgb_grid
)

## + Fold1: eta=0.01, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:12] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:12] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold1: eta=0.01, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold1: eta=0.01, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:14] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:14] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold1: eta=0.01, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold1: eta=0.10, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:15] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:15] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold1: eta=0.10, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold1: eta=0.10, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200

```

```
## [21:16:16] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:16] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold1: eta=0.10, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold1: eta=0.30, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:17] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:17] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold1: eta=0.30, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold1: eta=0.30, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:19] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:19] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold1: eta=0.30, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold2: eta=0.01, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:20] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:20] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold2: eta=0.01, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold2: eta=0.01, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:22] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:22] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold2: eta=0.01, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold2: eta=0.10, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:23] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:23] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold2: eta=0.10, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold2: eta=0.10, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:24] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:24] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
```

```
`iteration_range` instead.
## - Fold2: eta=0.10, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold2: eta=0.30, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:25] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:25] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold2: eta=0.30, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold2: eta=0.30, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:27] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:27] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold2: eta=0.30, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold3: eta=0.01, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:28] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:28] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold3: eta=0.01, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold3: eta=0.01, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:29] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:29] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold3: eta=0.01, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold3: eta=0.10, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:30] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:30] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold3: eta=0.10, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold3: eta=0.10, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:32] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:32] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold3: eta=0.10, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
```

```
## + Fold3: eta=0.30, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:33] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:33] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold3: eta=0.30, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold3: eta=0.30, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:35] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:35] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold3: eta=0.30, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold4: eta=0.01, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:36] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:36] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold4: eta=0.01, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold4: eta=0.01, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:37] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:37] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold4: eta=0.01, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold4: eta=0.10, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:37] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:38] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold4: eta=0.10, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold4: eta=0.10, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:39] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:39] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold4: eta=0.10, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold4: eta=0.30, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:39] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
```

```
`iteration_range` instead.
## [21:16:39] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold4: eta=0.30, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold4: eta=0.30, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:40] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:40] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold4: eta=0.30, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold5: eta=0.01, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:41] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:41] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold5: eta=0.01, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold5: eta=0.01, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:42] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:42] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold5: eta=0.01, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold5: eta=0.10, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:43] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:43] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold5: eta=0.10, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold5: eta=0.10, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:44] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:44] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold5: eta=0.10, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold5: eta=0.30, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:44] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:44] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
```



```

## - Fold5: eta=0.30, max_depth=3, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## + Fold5: eta=0.30, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## [21:16:45] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## [21:16:45] WARNING: src/c_api/c_api.cc:935: `ntree_limit` is deprecated, use
`iteration_range` instead.
## - Fold5: eta=0.30, max_depth=6, gamma=0, colsample_bytree=1,
min_child_weight=1, subsample=1, nrounds=200
## Aggregating results
## Selecting tuning parameters
## Fitting nrounds = 200, max_depth = 6, eta = 0.01, gamma = 0, colsample_bytree =
1, min_child_weight = 1, subsample = 1 on full training set

# Predizioni
pred_xgb <- predict(xgb_model, newdata = test_data)
probs_xgb <- predict(xgb_model, newdata = test_data, type = "prob")

# Confusion matrix
confusionMatrix(pred_xgb, test_data$Death)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction No Yes
##           No  75  32
##           Yes 36  40
##
##           Accuracy : 0.6284
##           95% CI : (0.554, 0.6986)
##           No Information Rate : 0.6066
##           P-Value [Acc > NIR] : 0.2996
##
##           Kappa : 0.229
##
## Mcnemar's Test P-Value : 0.7160
##
##           Sensitivity : 0.6757
##           Specificity : 0.5556
##           Pos Pred Value : 0.7009
##           Neg Pred Value : 0.5263
##           Prevalence : 0.6066
##           Detection Rate : 0.4098
##           Detection Prevalence : 0.5847
##           Balanced Accuracy : 0.6156
##
##           'Positive' Class : No
##

```

```
# AUC
library(pROC)
roc_xgb <- roc(test_data$Death, probs_xgb[, "Yes"])

## Setting levels: control = No, case = Yes

## Setting direction: controls < cases

auc(roc_xgb)

## Area under the curve: 0.6607
```

SVM - radial kernel

```
# Carica la Libreria per SVM
library(e1071)

## Warning: package 'e1071' was built under R version 4.3.3

##
## Attaching package: 'e1071'

## The following object is masked from 'package:Hmisc':
##
##      impute

# SVM Model
svm_model <- svm(Death ~ ., data = train_data, kernel = "radial", probability =
TRUE)

# Predizioni per SVM
pred_svm <- predict(svm_model, newdata = test_data, probability = TRUE)
probs_svm <- attr(pred_svm, "probabilities")

# SVM Metrics
confusionMatrix(pred_svm, test_data$Death)

## Confusion Matrix and Statistics
##
##              Reference
## Prediction No Yes
##      No   80   45
##      Yes  31   27
##
##              Accuracy : 0.5847
##              95% CI   : (0.5097, 0.6569)
##      No Information Rate : 0.6066
##      P-Value [Acc > NIR] : 0.7529
##
##              Kappa   : 0.0991
##
##      Mcnemar's Test P-Value : 0.1359
##
##              Sensitivity : 0.7207
```

```
##           Specificity : 0.3750
##           Pos Pred Value : 0.6400
##           Neg Pred Value : 0.4655
##           Prevalence : 0.6066
##           Detection Rate : 0.4372
##           Detection Prevalence : 0.6831
##           Balanced Accuracy : 0.5479
##
##           'Positive' Class : No
##

roc(test_data$Death, probs_svm[, "Yes"])

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

##
## Call:
## roc.default(response = test_data$Death, predictor = probs_svm[, "Yes"])
##
## Data: probs_svm[, "Yes"] in 111 controls (test_data$Death No) < 72 cases
(test_data$Death Yes).
## Area under the curve: 0.6084
```

KNN

```
# Carica La Libreria per KNN
library(caret)

# KNN Model
knn_model <- train(Death ~ ., data = train_data, method = "knn",
                  trControl = trainControl(method = "none"), tuneGrid =
expand.grid(k = 5))

# Predizioni per KNN
pred_knn <- predict(knn_model, newdata = test_data)
probs_knn <- predict(knn_model, newdata = test_data, type = "prob")

# KNN Metrics
confusionMatrix(pred_knn, test_data$Death)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction No Yes
##           No  69  26
##           Yes 42  46
##
##           Accuracy : 0.6284
##           95% CI : (0.554, 0.6986)
##           No Information Rate : 0.6066
##           P-Value [Acc > NIR] : 0.29960
```

```

##
##          Kappa : 0.2507
##
## Mcnemar's Test P-Value : 0.06891
##
##          Sensitivity : 0.6216
##          Specificity : 0.6389
##          Pos Pred Value : 0.7263
##          Neg Pred Value : 0.5227
##          Prevalence : 0.6066
##          Detection Rate : 0.3770
##          Detection Prevalence : 0.5191
##          Balanced Accuracy : 0.6303
##
##          'Positive' Class : No
##

roc(test_data$Death, probs_knn[, "Yes"])

## Setting levels: control = No, case = Yes

## Setting direction: controls < cases

##
## Call:
## roc.default(response = test_data$Death, predictor = probs_knn[,      "Yes"])
##
## Data: probs_knn[, "Yes"] in 111 controls (test_data$Death No) < 72 cases
## (test_data$Death Yes).
## Area under the curve: 0.6532

# Carica La Libreria per KNN
library(caret)

# KNN Model
knn_model <- train(Death ~ ., data = train_data, method = "knn",
                  trControl = trainControl(method = "none"), tuneGrid =
expand.grid(k = 15))

# Predizioni per KNN
pred_knn <- predict(knn_model, newdata = test_data)
probs_knn <- predict(knn_model, newdata = test_data, type = "prob")

# KNN Metrics
confusionMatrix(pred_knn, test_data$Death)

## Confusion Matrix and Statistics
##
##          Reference
## Prediction No Yes
##          No  69  30
##          Yes 42  42

```

```

##
##          Accuracy : 0.6066
##          95% CI : (0.5318, 0.6778)
##    No Information Rate : 0.6066
##    P-Value [Acc > NIR] : 0.5323
##
##          Kappa : 0.1991
##
##    McNemar's Test P-Value : 0.1949
##
##          Sensitivity : 0.6216
##          Specificity : 0.5833
##          Pos Pred Value : 0.6970
##          Neg Pred Value : 0.5000
##          Prevalence : 0.6066
##          Detection Rate : 0.3770
##    Detection Prevalence : 0.5410
##          Balanced Accuracy : 0.6025
##
##          'Positive' Class : No
##
roc(test_data$Death, probs_knn[, "Yes"])
## Setting levels: control = No, case = Yes
## Setting direction: controls < cases
##
## Call:
## roc.default(response = test_data$Death, predictor = probs_knn[, "Yes"])
##
## Data: probs_knn[, "Yes"] in 111 controls (test_data$Death No) < 72 cases
## (test_data$Death Yes).
## Area under the curve: 0.6554

```

Supervised learning CON BILANCIAMENTO

Logistic Regression

Random Forest

XGBoost

SVM

K-NN

DIVISIONE TRAIN/TEST

```

# Controlla i livelli del target "Death"
table(data$Death)

```

```
##
## 0 1
## 530 386

# Controlla il bilanciamento
prop.table(table(y))

## y
## ( 0.0000000, 0.03285421] ( 0.0000000, 0.13141684]
## 0.0006157635 0.0006157635
## ( 0.0000000, 0.13141684+] ( 0.0000000, 0.16427105+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 0.22997947] ( 0.0000000, 0.22997947+]
## 0.0006157635 0.0012315271
## ( 0.0000000, 0.26283368] ( 0.0000000, 0.26283368+]
## 0.0006157635 0.0018472906
## ( 0.0000000, 0.29568789] ( 0.0000000, 0.29568789+]
## 0.0006157635 0.0018472906
## ( 0.0000000, 0.32854209+] ( 0.0000000, 0.39425051+]
## 0.0018472906 0.0006157635
## ( 0.0000000, 0.42710472+] ( 0.3942505, 0.49281314]
## 0.0006157635 0.0006157635
## ( 0.0000000, 0.49281314+] ( 0.0000000, 0.52566735+]
## 0.0012315271 0.0018472906
## ( 0.0000000, 0.55852156] ( 0.0000000, 0.55852156+]
## 0.0006157635 0.0018472906
## ( 0.0000000, 0.62422998+] ( 0.0000000, 0.65708419]
## 0.0006157635 0.0006157635
## ( 0.5256674, 0.65708419] ( 0.0000000, 0.65708419+]
## 0.0006157635 0.0012315271
## ( 0.0000000, 0.68993840] ( 0.0000000, 0.68993840+]
## 0.0006157635 0.0024630542
## ( 0.0000000, 0.72279261] ( 0.0000000, 0.72279261+]
## 0.0006157635 0.0024630542
## ( 0.0000000, 0.75564682] ( 0.0000000, 0.75564682+]
## 0.0006157635 0.0043103448
## ( 0.0000000, 0.78850103+] ( 0.0000000, 0.82135524+]
## 0.0030788177 0.0024630542
## ( 0.0000000, 0.88706366+] ( 0.0000000, 0.91991786]
## 0.0024630542 0.0006157635
## ( 0.0000000, 0.91991786+] ( 0.0000000, 0.95277207+]
## 0.0024630542 0.0030788177
## ( 0.0000000, 0.98562628] ( 0.0000000, 0.98562628+]
## 0.0006157635 0.0036945813
## ( 0.0000000, 1.01848049+] ( 0.0000000, 1.05133470+]
## 0.0024630542 0.0006157635
## ( 0.0000000, 1.08418891+] ( 0.0000000, 1.11704312]
## 0.0024630542 0.0006157635
## ( 0.0000000, 1.11704312+] ( 0.9199179, 1.14989733]
## 0.0024630542 0.0006157635
## ( 0.0000000, 1.14989733+] ( 0.0000000, 1.18275154+]
## 0.0043103448 0.0024630542
```

```
## ( 0.7227926, 1.21560575] ( 0.0000000, 1.24845996+]  
## 0.0006157635 0.0018472906  
## ( 0.0000000, 1.28131417] ( 0.0000000, 1.31416838]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 1.31416838+] ( 0.0000000, 1.34702259+]  
## 0.0006157635 0.0024630542  
## ( 0.0000000, 1.37987680+] ( 0.0000000, 1.41273101+]  
## 0.0018472906 0.0012315271  
## ( 0.0000000, 1.44558522+] ( 0.0000000, 1.51129363+]  
## 0.0012315271 0.0012315271  
## ( 0.0000000, 1.57700205] ( 0.0000000, 1.57700205+]  
## 0.0012315271 0.0012315271  
## ( 0.0000000, 1.60985626+] ( 0.0000000, 1.64271047]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 1.64271047+] ( 0.0000000, 1.67556468]  
## 0.0018472906 0.0006157635  
## ( 0.0000000, 1.67556468+] ( 0.7885010, 1.70841889]  
## 0.0006157635 0.0006157635  
## ( 1.1170431, 1.70841889] ( 0.0000000, 1.70841889+]  
## 0.0006157635 0.0018472906  
## ( 0.0000000, 1.74127310] ( 0.0000000, 1.74127310+]  
## 0.0012315271 0.0006157635  
## ( 0.8213552, 1.77412731] ( 0.0000000, 1.77412731+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 1.80698152+] ( 0.0000000, 1.83983573+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 1.87268994+] ( 0.0000000, 1.90554415]  
## 0.0006157635 0.0006157635  
## ( 1.6755647, 1.90554415] ( 0.0000000, 1.93839836+]  
## 0.0006157635 0.0006157635  
## ( 0.6899384, 1.93839836+] ( 0.0000000, 1.97125257+]  
## 0.0006157635 0.0012315271  
## ( 0.0000000, 2.03696099+] ( 0.0000000, 2.06981520]  
## 0.0018472906 0.0006157635  
## ( 0.0000000, 2.06981520+] ( 0.0000000, 2.10266940]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 2.10266940+] ( 0.0000000, 2.16837782+]  
## 0.0036945813 0.0012315271  
## ( 1.3470226, 2.16837782+] ( 0.0000000, 2.20123203+]  
## 0.0006157635 0.0012315271  
## ( 0.0000000, 2.23408624] ( 0.0000000, 2.23408624+]  
## 0.0006157635 0.0012315271  
## ( 0.0000000, 2.26694045+] ( 0.0000000, 2.29979466+]  
## 0.0012315271 0.0006157635  
## ( 0.0000000, 2.33264887+] ( 1.0184805, 2.33264887+]  
## 0.0024630542 0.0006157635  
## ( 0.0000000, 2.36550308+] ( 0.0000000, 2.39835729+]  
## 0.0024630542 0.0006157635  
## ( 0.0000000, 2.43121150+] ( 0.0000000, 2.46406571+]  
## 0.0024630542 0.0006157635  
## ( 0.0000000, 2.49691992+] ( 2.2012320, 2.52977413]
```

```
## 0.0006157635 0.0006157635
## ( 0.0000000, 2.52977413+] ( 1.7741273, 2.59548255]
## 0.0012315271 0.0006157635
## ( 0.0000000, 2.59548255+] ( 0.0000000, 2.62833676+]
## 0.0012315271 0.0006157635
## ( 0.0000000, 2.66119097+] ( 0.0000000, 2.69404517+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 2.72689938+] ( 2.4969199, 2.75975359]
## 0.0018472906 0.0006157635
## ( 0.0000000, 2.75975359+] ( 0.0000000, 2.79260780+]
## 0.0018472906 0.0030788177
## ( 2.5297741, 2.79260780+] ( 0.0000000, 2.82546201+]
## 0.0006157635 0.0036945813
## ( 0.0000000, 2.85831622+] ( 0.0000000, 2.89117043]
## 0.0018472906 0.0006157635
## ( 0.0000000, 2.89117043+] ( 0.0000000, 2.92402464+]
## 0.0012315271 0.0012315271
## ( 2.0369610, 2.95687885] ( 0.0000000, 2.95687885+]
## 0.0006157635 0.0018472906
## ( 0.0000000, 2.98973306+] ( 0.0000000, 3.02258727]
## 0.0006157635 0.0006157635
## ( 0.0000000, 3.02258727+] ( 0.0000000, 3.05544148+]
## 0.0043103448 0.0024630542
## ( 2.7597536, 3.05544148+] ( 0.0000000, 3.08829569]
## 0.0006157635 0.0006157635
## ( 0.0000000, 3.08829569+] ( 2.6611910, 3.12114990]
## 0.0024630542 0.0006157635
## ( 0.0000000, 3.12114990+] ( 2.1026694, 3.15400411]
## 0.0006157635 0.0006157635
## ( 0.0000000, 3.15400411+] ( 0.0000000, 3.18685832+]
## 0.0018472906 0.0030788177
## ( 0.0000000, 3.21971253+] ( 0.0000000, 3.25256674+]
## 0.0036945813 0.0018472906
## ( 0.0000000, 3.28542094+] ( 0.0000000, 3.31827515+]
## 0.0024630542 0.0024630542
## ( 1.2484600, 3.31827515+] ( 3.0554415, 3.35112936]
## 0.0006157635 0.0006157635
## ( 0.0000000, 3.35112936+] ( 0.0000000, 3.38398357+]
## 0.0024630542 0.0018472906
## ( 0.0000000, 3.41683778+] ( 2.8583162, 3.44969199]
## 0.0012315271 0.0006157635
## ( 0.0000000, 3.44969199+] ( 0.0000000, 3.48254620+]
## 0.0012315271 0.0030788177
## ( 0.0000000, 3.51540041+] ( 0.2956879, 3.54825462]
## 0.0012315271 0.0006157635
## ( 0.0000000, 3.54825462+] ( 3.0882957, 3.58110883]
## 0.0036945813 0.0006157635
## ( 0.0000000, 3.58110883+] ( 0.0000000, 3.61396304+]
## 0.0018472906 0.0018472906
## ( 0.0000000, 3.64681725+] ( 0.0000000, 3.67967146+]
## 0.0024630542 0.0036945813
```



```
## ( 0.0000000, 3.71252567+] ( 0.0000000, 3.74537988]
## 0.0036945813 0.0006157635
## ( 0.0000000, 3.74537988+] ( 0.0000000, 3.77823409+]
## 0.0018472906 0.0012315271
## ( 1.7084189, 3.77823409+] ( 3.1868583, 3.81108830]
## 0.0006157635 0.0006157635
## ( 0.0000000, 3.81108830+] ( 0.9527721, 3.81108830+]
## 0.0012315271 0.0006157635
## ( 0.0000000, 3.84394251+] ( 0.0000000, 3.87679671+]
## 0.0024630542 0.0030788177
## ( 0.0000000, 3.90965092+] ( 0.0000000, 3.94250513+]
## 0.0024630542 0.0012315271
## ( 3.8110883, 3.94250513+] ( 0.0000000, 3.97535934+]
## 0.0006157635 0.0012315271
## ( 0.3285421, 4.00821355] ( 0.0000000, 4.00821355+]
## 0.0006157635 0.0012315271
## ( 2.8254620, 4.04106776] ( 0.0000000, 4.04106776+]
## 0.0006157635 0.0006157635
## ( 3.0882957, 4.04106776+] ( 0.0000000, 4.07392197+]
## 0.0006157635 0.0036945813
## ( 0.0000000, 4.10677618+] ( 3.2525667, 4.13963039]
## 0.0024630542 0.0006157635
## ( 0.0000000, 4.13963039+] ( 0.0000000, 4.17248460]
## 0.0012315271 0.0006157635
## ( 2.7268994, 4.17248460] ( 0.0000000, 4.17248460+]
## 0.0006157635 0.0036945813
## ( 0.0000000, 4.20533881] ( 0.0000000, 4.20533881+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 4.23819302+] ( 2.4312115, 4.23819302+]
## 0.0030788177 0.0006157635
## ( 3.8767967, 4.23819302+] ( 1.1498973, 4.27104723]
## 0.0006157635 0.0006157635
## ( 0.0000000, 4.27104723+] ( 2.7268994, 4.30390144]
## 0.0006157635 0.0006157635
## ( 0.0000000, 4.30390144+] ( 0.0000000, 4.33675565+]
## 0.0018472906 0.0030788177
## ( 0.0000000, 4.36960986+] ( 2.3655031, 4.36960986+]
## 0.0012315271 0.0006157635
## ( 4.1067762, 4.36960986+] ( 0.0000000, 4.40246407+]
## 0.0006157635 0.0024630542
## ( 0.0000000, 4.43531828+] ( 2.3326489, 4.43531828+]
## 0.0030788177 0.0006157635
## ( 0.0000000, 4.46817248+] ( 2.2340862, 4.46817248+]
## 0.0018472906 0.0006157635
## ( 0.0000000, 4.50102669+] ( 0.0000000, 4.56673511+]
## 0.0012315271 0.0006157635
## ( 2.9897331, 4.56673511+] ( 0.0000000, 4.59958932+]
## 0.0006157635 0.0036945813
## ( 0.0000000, 4.63244353] ( 0.0000000, 4.63244353+]
## 0.0006157635 0.0018472906
## ( 0.0000000, 4.66529774+] ( 0.0000000, 4.69815195+]
```

```
## 0.0012315271 0.0018472906
## ( 0.0000000, 4.73100616+] ( 2.7926078, 4.76386037]
## 0.0018472906 0.0006157635
## ( 0.0000000, 4.76386037+] ( 3.8439425, 4.76386037+]
## 0.0006157635 0.0006157635
## ( 3.4496920, 4.79671458] ( 0.0000000, 4.79671458+]
## 0.0006157635 0.0012315271
## ( 2.8911704, 4.79671458+] ( 0.0000000, 4.82956879+]
## 0.0006157635 0.0006157635
## ( 4.1724846, 4.86242300] ( 4.4353183, 4.86242300]
## 0.0006157635 0.0006157635
## ( 0.0000000, 4.86242300+] ( 0.0000000, 4.89527721+]
## 0.0018472906 0.0006157635
## ( 0.0000000, 4.92813142+] ( 0.0000000, 4.96098563]
## 0.0006157635 0.0006157635
## ( 0.0000000, 4.99383984] ( 0.2628337, 4.99383984]
## 0.0006157635 0.0006157635
## ( 3.0225873, 4.99383984+] ( 3.1540041, 4.99383984+]
## 0.0006157635 0.0006157635
## ( 3.5482546, 5.02669405+] ( 0.0000000, 5.05954825+]
## 0.0006157635 0.0006157635
## ( 4.2381930, 5.09240246] ( 4.7967146, 5.09240246]
## 0.0006157635 0.0006157635
## ( 0.0000000, 5.09240246+] ( 0.0000000, 5.15811088+]
## 0.0012315271 0.0012315271
## ( 4.4681725, 5.15811088+] ( 0.0000000, 5.19096509+]
## 0.0006157635 0.0018472906
## ( 0.0000000, 5.22381930+] ( 1.3470226, 5.32238193]
## 0.0006157635 0.0006157635
## ( 0.0000000, 5.32238193+] ( 3.5482546, 5.32238193+]
## 0.0012315271 0.0006157635
## ( 0.0000000, 5.35523614+] ( 3.2197125, 5.38809035+]
## 0.0006157635 0.0006157635
## ( 4.3367556, 5.38809035+] ( 0.0000000, 5.42094456+]
## 0.0006157635 0.0012315271
## ( 0.0000000, 5.45379877+] ( 0.0000000, 5.48665298]
## 0.0012315271 0.0006157635
## ( 0.0000000, 5.48665298+] ( 0.0000000, 5.51950719+]
## 0.0012315271 0.0018472906
## ( 0.0000000, 5.55236140+] ( 2.1026694, 5.55236140+]
## 0.0006157635 0.0006157635
## ( 4.1724846, 5.58521561] ( 0.0000000, 5.58521561+]
## 0.0006157635 0.0006157635
## ( 3.0554415, 5.61806982] ( 0.0000000, 5.61806982+]
## 0.0006157635 0.0012315271
## ( 0.0000000, 5.65092402+] ( 0.0000000, 5.68377823+]
## 0.0006157635 0.0012315271
## ( 3.2854209, 5.74948665] ( 0.0000000, 5.74948665+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 5.78234086] ( 1.0841889, 5.78234086+]
## 0.0006157635 0.0006157635
```

```
## ( 0.0000000, 5.81519507+] ( 2.7597536, 5.81519507+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 5.84804928+] ( 4.0739220, 5.84804928+]
## 0.0024630542 0.0006157635
## ( 0.0000000, 5.88090349+] ( 0.8870637, 5.91375770]
## 0.0006157635 0.0006157635
## ( 5.6180698, 5.94661191] ( 0.0000000, 5.94661191+]
## 0.0006157635 0.0006157635
## ( 2.0369610, 5.94661191+] ( 3.9425051, 6.01232033]
## 0.0006157635 0.0006157635
## ( 5.6509240, 6.01232033] ( 0.0000000, 6.01232033+]
## 0.0006157635 0.0012315271
## ( 0.0000000, 6.07802875+] ( 0.0000000, 6.11088296+]
## 0.0012315271 0.0012315271
## ( 2.9568789, 6.11088296+] ( 1.4127310, 6.14373717]
## 0.0006157635 0.0006157635
## ( 0.0000000, 6.14373717+] ( 2.7597536, 6.17659138]
## 0.0006157635 0.0006157635
## ( 3.7782341, 6.17659138] ( 0.0000000, 6.17659138+]
## 0.0006157635 0.0024630542
## ( 0.0000000, 6.20944559+] ( 3.6796715, 6.20944559+]
## 0.0024630542 0.0006157635
## ( 3.0882957, 6.24229979] ( 0.0000000, 6.24229979+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 6.27515400+] ( 0.0000000, 6.37371663+]
## 0.0006157635 0.0024630542
## ( 2.4312115, 6.40657084] ( 0.0000000, 6.40657084+]
## 0.0006157635 0.0012315271
## ( 0.0000000, 6.43942505+] ( 0.5256674, 6.47227926]
## 0.0012315271 0.0006157635
## ( 0.0000000, 6.47227926+] ( 0.7227926, 6.47227926+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 6.50513347+] ( 4.0082136, 6.50513347+]
## 0.0006157635 0.0006157635
## ( 2.3655031, 6.53798768] ( 0.0000000, 6.53798768+]
## 0.0006157635 0.0012315271
## ( 0.9199179, 6.53798768+] ( 0.0000000, 6.60369610+]
## 0.0006157635 0.0006157635
## ( 3.3839836, 6.60369610+] ( 6.2422998, 6.63655031]
## 0.0006157635 0.0006157635
## ( 0.0000000, 6.63655031+] ( 0.0000000, 6.66940452]
## 0.0012315271 0.0006157635
## ( 0.0000000, 6.66940452+] ( 4.4353183, 6.70225873]
## 0.0006157635 0.0006157635
## ( 0.0000000, 6.70225873+] ( 3.0225873, 6.70225873+]
## 0.0006157635 0.0006157635
## ( 2.6940452, 6.73511294] ( 0.0000000, 6.73511294+]
## 0.0006157635 0.0012315271
## ( 0.0000000, 6.80082136+] ( 0.0000000, 6.83367556+]
## 0.0006157635 0.0006157635
## ( 3.6468172, 6.83367556+] ( 6.6365503, 6.86652977]
```

```

## 0.0006157635 0.0006157635
## ( 0.0000000, 6.86652977+] ( 3.3511294, 6.86652977+]
## 0.0018472906 0.0006157635
## ( 0.0000000, 6.89938398+] ( 0.0000000, 6.93223819+]
## 0.0012315271 0.0012315271
## ( 0.0000000, 6.96509240+] ( 1.6427105, 6.96509240+]
## 0.0018472906 0.0006157635
## ( 3.7125257, 6.99794661+] ( 0.0000000, 6.99794661+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 7.03080082+] ( 4.6981520, 7.06365503]
## 0.0012315271 0.0006157635
## ( 0.0000000, 7.06365503+] ( 3.6139630, 7.06365503+]
## 0.0006157635 0.0006157635
## ( 1.9383984, 7.09650924] ( 0.0000000, 7.09650924+]
## 0.0006157635 0.0006157635
## ( 0.9856263, 7.09650924+] ( 4.6324435, 7.12936345]
## 0.0006157635 0.0006157635
## ( 6.5379877, 7.12936345] ( 0.0000000, 7.12936345+]
## 0.0006157635 0.0012315271
## ( 5.3223819, 7.12936345+] ( 3.8110883, 7.16221766]
## 0.0006157635 0.0006157635
## ( 0.0000000, 7.19507187+] ( 3.4168378, 7.22792608]
## 0.0006157635 0.0006157635
## ( 6.5051335, 7.22792608] ( 0.0000000, 7.26078029+]
## 0.0006157635 0.0006157635
## ( 2.1026694, 7.29363450] ( 0.0000000, 7.29363450+]
## 0.0006157635 0.0006157635
## ( 7.1293634, 7.29363450+] ( 0.0000000, 7.32648871+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 7.35934292+] ( 7.2936345, 7.39219713]
## 0.0006157635 0.0006157635
## ( 0.0000000, 7.39219713+] ( 6.2094456, 7.42505133]
## 0.0006157635 0.0006157635
## ( 0.0000000, 7.42505133+] ( 0.0000000, 7.45790554+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 7.49075975+] ( 6.2094456, 7.49075975+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 7.55646817+] ( 0.0000000, 7.58932238+]
## 0.0018472906 0.0018472906
## ( 0.0000000, 7.62217659+] ( 0.0000000, 7.68788501+]
## 0.0024630542 0.0012315271
## ( 3.4825462, 7.72073922+] ( 0.0000000, 7.78644764+]
## 0.0006157635 0.0012315271
## ( 6.2094456, 7.78644764+] ( 7.3593429, 7.78644764+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 7.88501027+] ( 3.4496920, 7.91786448+]
## 0.0012315271 0.0006157635
## ( 0.7885010, 7.95071869] ( 0.0000000, 7.95071869+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 7.98357290+] ( 0.0000000, 8.01642710+]
## 0.0006157635 0.0006157635

```

```
## ( 1.7084189, 8.08213552] ( 0.0000000, 8.08213552+]  
## 0.0006157635 0.0012315271  
## ( 2.2669405, 8.11498973] ( 0.0000000, 8.11498973+]  
## 0.0006157635 0.0030788177  
## ( 4.6652977, 8.14784394+ ] ( 4.5010267, 8.18069815]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 8.18069815+ ] ( 5.1909651, 8.21355236]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 8.24640657+ ] ( 0.0000000, 8.27926078+]  
## 0.0006157635 0.0012315271  
## ( 2.5954825, 8.31211499] ( 7.3921971, 8.31211499]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 8.31211499+ ] ( 5.8809035, 8.31211499+]  
## 0.0030788177 0.0006157635  
## ( 2.7926078, 8.34496920] ( 7.0636550, 8.37782341]  
## 0.0006157635 0.0006157635  
## ( 2.8254620, 8.37782341+ ] ( 3.2197125, 8.41067762]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 8.41067762+ ] ( 4.4024641, 8.41067762+]  
## 0.0024630542 0.0006157635  
## ( 0.0000000, 8.44353183+ ] ( 0.0000000, 8.47638604+]  
## 0.0012315271 0.0018472906  
## ( 0.0000000, 8.50924025+ ] ( 3.7125257, 8.50924025+]  
## 0.0006157635 0.0006157635  
## ( 7.6221766, 8.50924025+ ] ( 4.1724846, 8.54209446]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 8.54209446+ ] ( 3.1868583, 8.54209446+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 8.57494867+ ] ( 0.0000000, 8.60780287+]  
## 0.0012315271 0.0012315271  
## ( 8.3121150, 8.64065708] ( 0.0000000, 8.64065708+]  
## 0.0006157635 0.0018472906  
## ( 0.9527721, 8.64065708+ ] ( 3.6796715, 8.64065708+]  
## 0.0006157635 0.0006157635  
## ( 2.4312115, 8.67351129] ( 3.0225873, 8.70636550]  
## 0.0006157635 0.0006157635  
## ( 2.8911704, 8.77207392] ( 0.0000000, 8.77207392+]  
## 0.0006157635 0.0006157635  
## ( 6.6694045, 8.77207392+ ] ( 0.0000000, 8.80492813+]  
## 0.0006157635 0.0006157635  
## ( 1.7084189, 8.83778234] ( 3.3839836, 8.83778234]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 8.83778234+ ] ( 0.0000000, 8.87063655+]  
## 0.0006157635 0.0018472906  
## ( 2.0369610, 8.90349076] ( 0.0000000, 8.90349076+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 8.93634497+ ] ( 0.0000000, 9.00205339+]  
## 0.0012315271 0.0018472906  
## ( 6.2751540, 9.00205339+ ] ( 1.5112936, 9.03490760]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 9.03490760+ ] ( 0.0000000, 9.06776181+]
```

```
## 0.0006157635 0.0006157635
## ( 0.0000000, 9.10061602+] ( 0.0000000, 9.13347023+]
## 0.0006157635 0.0018472906
## ( 0.0000000, 9.16632444+] ( 5.9466119, 9.16632444+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 9.19917864+] ( 0.0000000, 9.23203285+]
## 0.0006157635 0.0006157635
## ( 3.4825462, 9.23203285+] ( 4.0739220, 9.26488706]
## 0.0006157635 0.0006157635
## ( 0.0000000, 9.36344969+] ( 0.0000000, 9.39630390+]
## 0.0012315271 0.0006157635
## ( 0.0000000, 9.42915811+] ( 0.0000000, 9.46201232+]
## 0.0006157635 0.0006157635
## ( 8.6406571, 9.49486653] ( 7.0308008, 9.52772074+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 9.59342916] ( 0.0000000, 9.65913758+]
## 0.0006157635 0.0006157635
## ( 4.0410678, 9.65913758+] ( 0.0000000, 9.69199179+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 9.75770021+] ( 3.9753593, 9.75770021+]
## 0.0012315271 0.0006157635
## ( 0.0000000, 9.79055441+] ( 2.3983573, 9.82340862]
## 0.0012315271 0.0006157635
## ( 2.9240246, 9.85626283] ( 3.2525667, 9.85626283]
## 0.0006157635 0.0006157635
## ( 0.0000000, 9.85626283+] ( 0.0000000, 9.92197125+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 9.95482546+] ( 4.4681725, 9.98767967]
## 0.0006157635 0.0006157635
## ( 0.0000000, 9.98767967+] ( 0.0000000, 10.02053388+]
## 0.0018472906 0.0006157635
## ( 0.0000000, 10.05338809+] ( 3.8439425, 10.08624230]
## 0.0012315271 0.0006157635
## ( 0.0000000, 10.08624230+] ( 0.0000000, 10.11909651+]
## 0.0006157635 0.0012315271
## ( 0.0000000, 10.15195072+] ( 0.0000000, 10.18480493+]
## 0.0012315271 0.0006157635
## ( 8.1149897, 10.18480493+] ( 0.0000000, 10.21765914+]
## 0.0006157635 0.0012315271
## ( 0.0000000, 10.25051335+] ( 0.0000000, 10.28336756+]
## 0.0006157635 0.0012315271
## ( 4.1067762, 10.31622177] ( 1.3798768, 10.31622177+]
## 0.0006157635 0.0006157635
## ( 3.0882957, 10.34907598] ( 6.4394251, 10.34907598+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 10.38193018+] ( 0.0000000, 10.41478439+]
## 0.0018472906 0.0006157635
## ( 6.8336756, 10.48049281] ( 6.9650924, 10.48049281+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 10.51334702+] ( 0.0000000, 10.54620123+]
## 0.0012315271 0.0012315271
```

```
## ( 2.7926078, 10.61190965] ( 0.0000000, 10.61190965+]  
## 0.0006157635 0.0024630542  
## ( 3.1868583, 10.61190965+] ( 4.2710472, 10.61190965+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 10.64476386+] ( 3.5811088, 10.71047228]  
## 0.0006157635 0.0006157635  
## ( 6.3737166, 10.71047228] ( 0.0000000, 10.71047228+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 10.77618070+] ( 4.1067762, 10.84188912+]  
## 0.0018472906 0.0006157635  
## ( 0.0000000, 10.87474333+] ( 0.0000000, 10.90759754+]  
## 0.0006157635 0.0024630542  
## ( 0.0000000, 10.94045175+] ( 3.2525667, 10.97330595+]  
## 0.0006157635 0.0006157635  
## ( 3.1540041, 11.00616016] ( 8.3121150, 11.00616016+]  
## 0.0006157635 0.0006157635  
## ( 8.2792608, 11.03901437] ( 4.7310062, 11.07186858]  
## 0.0006157635 0.0006157635  
## ( 6.8993840, 11.07186858+] ( 0.0000000, 11.10472279+]  
## 0.0006157635 0.0012315271  
## ( 1.8726899, 11.10472279+] ( 7.9835729, 11.13757700]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 11.13757700+] ( 0.0000000, 11.17043121+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 11.20328542+] ( 0.0000000, 11.26899384+]  
## 0.0006157635 0.0006157635  
## ( 3.6796715, 11.30184805] ( 3.7125257, 11.30184805]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 11.30184805+] ( 4.1396304, 11.33470226+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 11.36755647+] ( 0.0000000, 11.43326489+]  
## 0.0006157635 0.0006157635  
## ( 3.0225873, 11.46611910] ( 0.0000000, 11.46611910+]  
## 0.0006157635 0.0006157635  
## ( 5.7494867, 11.49897331] ( 0.0000000, 11.49897331+]  
## 0.0006157635 0.0012315271  
## ( 3.1211499, 11.53182752] ( 0.0000000, 11.53182752+]  
## 0.0006157635 0.0006157635  
## ( 4.4353183, 11.53182752+] ( 2.9568789, 11.56468172]  
## 0.0006157635 0.0006157635  
## ( 5.5195072, 11.63039014] ( 0.0000000, 11.63039014+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 11.66324435+] ( 7.2607803, 11.72895277]  
## 0.0012315271 0.0006157635  
## ( 0.0000000, 11.72895277+] ( 9.0349076, 11.72895277+]  
## 0.0012315271 0.0006157635  
## ( 9.8562628, 11.76180698] ( 4.0739220, 11.76180698+]  
## 0.0006157635 0.0006157635  
## ( 6.9322382, 11.76180698+] ( 0.0000000, 11.79466119+]  
## 0.0006157635 0.0006157635  
## ( 6.1437372, 11.82751540] ( 6.9979466, 11.86036961]
```

```
## 0.0006157635 0.0006157635
## (11.1375770, 11.92607803] ( 5.8480493, 11.95893224+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 11.99178645+] ( 1.1498973, 11.99178645+]
## 0.0006157635 0.0006157635
## ( 3.2197125, 11.99178645+] ( 0.0000000, 12.02464066+]
## 0.0006157635 0.0006157635
## (11.9917864, 12.05749487] ( 0.0000000, 12.05749487+]
## 0.0006157635 0.0006157635
## ( 0.4928131, 12.05749487+] ( 0.0000000, 12.09034908+]
## 0.0006157635 0.0006157635
## ( 4.7310062, 12.12320329] ( 0.0000000, 12.12320329+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 12.15605749+] ( 1.1498973, 12.15605749+]
## 0.0006157635 0.0006157635
## ( 3.8767967, 12.15605749+] ( 4.9281314, 12.22176591]
## 0.0006157635 0.0006157635
## ( 0.0000000, 12.28747433+] ( 0.0000000, 12.35318275+]
## 0.0006157635 0.0006157635
## (10.5462012, 12.35318275+] ( 0.0000000, 12.38603696+]
## 0.0006157635 0.0012315271
## ( 0.6899384, 12.41889117] ( 0.0000000, 12.41889117+]
## 0.0006157635 0.0012315271
## ( 0.0000000, 12.45174538+] (11.3018480, 12.45174538+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 12.48459959+] ( 2.4640657, 12.51745380+]
## 0.0006157635 0.0006157635
## ( 5.1581109, 12.55030801+] ( 0.0000000, 12.58316222+]
## 0.0006157635 0.0006157635
## ( 1.1170431, 12.61601643] ( 3.8767967, 12.61601643]
## 0.0006157635 0.0006157635
## ( 0.0000000, 12.61601643+] ( 0.0000000, 12.64887064+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 12.68172485+] (12.3860370, 12.71457906]
## 0.0006157635 0.0006157635
## ( 0.0000000, 12.71457906+] ( 0.0000000, 12.78028747+]
## 0.0018472906 0.0012315271
## ( 4.0739220, 12.84599589] ( 0.0000000, 12.84599589+]
## 0.0006157635 0.0006157635
## ( 5.8480493, 12.87885010] ( 0.0000000, 12.87885010+]
## 0.0006157635 0.0006157635
## ( 9.1006160, 12.94455852+] ( 5.1909651, 12.97741273]
## 0.0006157635 0.0006157635
## ( 0.0000000, 12.97741273+] ( 6.3737166, 12.97741273+]
## 0.0006157635 0.0006157635
## ( 8.4106776, 12.97741273+] (10.1190965, 12.97741273+]
## 0.0006157635 0.0006157635
## ( 6.8665298, 13.01026694] ( 0.0000000, 13.01026694+]
## 0.0006157635 0.0012315271
## ( 0.0000000, 13.04312115+] ( 0.0000000, 13.07597536+]
## 0.0006157635 0.0006157635
```



```
## ( 3.9096509, 13.10882957] ( 4.8624230, 13.10882957]
## 0.0006157635 0.0006157635
## ( 0.0000000, 13.14168378+ ] ( 6.7351129, 13.14168378+ ]
## 0.0012315271 0.0006157635
## ( 0.0000000, 13.17453799+ ] ( 0.7556468, 13.20739220+ ]
## 0.0012315271 0.0006157635
## ( 0.0000000, 13.27310062+ ] ( 3.6796715, 13.27310062+ ]
## 0.0006157635 0.0006157635
## ( 5.1581109, 13.33880903+ ] ( 2.2340862, 13.37166324+ ]
## 0.0006157635 0.0006157635
## ( 6.4722793, 13.43737166 ] ( 0.0000000, 13.50308008+ ]
## 0.0006157635 0.0012315271
## ( 8.5749487, 13.53593429 ] ( 0.0000000, 13.53593429+ ]
## 0.0006157635 0.0006157635
## ( 9.9548255, 13.60164271 ] ( 0.0000000, 13.60164271+ ]
## 0.0006157635 0.0012315271
## ( 8.5420945, 13.60164271+ ] ( 0.0000000, 13.63449692+ ]
## 0.0006157635 0.0006157635
## ( 6.5379877, 13.63449692+ ] ( 3.2197125, 13.66735113+ ]
## 0.0006157635 0.0006157635
## ( 3.3511294, 13.66735113+ ] ( 0.0000000, 13.70020534+ ]
## 0.0006157635 0.0006157635
## ( 4.3696099, 13.70020534+ ] ( 0.0000000, 13.73305955+ ]
## 0.0006157635 0.0006157635
## ( 4.4024641, 13.76591376 ] ( 0.0000000, 13.76591376+ ]
## 0.0006157635 0.0006157635
## ( 2.5954825, 13.83162218 ] ( 3.2197125, 13.83162218+ ]
## 0.0006157635 0.0006157635
## (12.0246407, 13.89733060+ ] ( 1.6427105, 13.93018480 ]
## 0.0006157635 0.0006157635
## ( 5.6837782, 13.93018480 ] ( 4.4024641, 13.96303901 ]
## 0.0006157635 0.0006157635
## ( 0.0000000, 13.96303901+ ] ( 0.0000000, 13.99589322+ ]
## 0.0006157635 0.0006157635
## ( 1.1498973, 14.02874743 ] (10.9404517, 14.02874743 ]
## 0.0006157635 0.0006157635
## ( 4.5995893, 14.02874743+ ] ( 3.4825462, 14.06160164 ]
## 0.0012315271 0.0006157635
## ( 3.5482546, 14.06160164 ] ( 8.1806982, 14.06160164 ]
## 0.0006157635 0.0006157635
## ( 0.0000000, 14.06160164+ ] ( 7.5893224, 14.06160164+ ]
## 0.0006157635 0.0006157635
## ( 0.0000000, 14.09445585+ ] ( 7.7864476, 14.09445585+ ]
## 0.0006157635 0.0006157635
## ( 0.0000000, 14.12731006+ ] ( 0.9856263, 14.16016427 ]
## 0.0006157635 0.0006157635
## (10.5133470, 14.16016427+ ] ( 0.0000000, 14.22587269+ ]
## 0.0006157635 0.0006157635
## ( 0.0000000, 14.25872690+ ] ( 5.4209446, 14.25872690+ ]
## 0.0012315271 0.0006157635
## ( 0.0000000, 14.29158111+ ] ( 0.0000000, 14.32443532+ ]
```

```
## 0.0006157635 0.0006157635
## ( 0.0000000, 14.45585216+] ( 0.0000000, 14.48870637+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 14.52156057+] ( 4.1396304, 14.62012320]
## 0.0006157635 0.0006157635
## ( 0.0000000, 14.65297741+] ( 5.0595483, 14.71868583+]
## 0.0006157635 0.0006157635
## ( 0.9199179, 14.78439425] ( 9.7905544, 14.81724846+]
## 0.0006157635 0.0006157635
## ( 8.4435318, 14.85010267] ( 0.0000000, 14.91581109+]
## 0.0006157635 0.0006157635
## (10.6119097, 14.94866530] ( 0.0000000, 14.94866530+]
## 0.0006157635 0.0012315271
## ( 2.3655031, 14.94866530+] ( 4.4353183, 14.94866530+]
## 0.0006157635 0.0006157635
## (12.4517454, 14.94866530+] ( 0.0000000, 15.14579055+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 15.17864476+] ( 7.3264887, 15.21149897]
## 0.0006157635 0.0006157635
## ( 0.0000000, 15.21149897+] (13.5030801, 15.21149897+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 15.24435318+] (13.0102669, 15.27720739]
## 0.0006157635 0.0006157635
## (14.2587269, 15.31006160] ( 4.0082136, 15.37577002+]
## 0.0006157635 0.0006157635
## ( 6.3737166, 15.37577002+] ( 0.0000000, 15.40862423+]
## 0.0006157635 0.0018472906
## (12.7145791, 15.47433265] ( 0.0000000, 15.47433265+]
## 0.0006157635 0.0006157635
## ( 2.8254620, 15.47433265+] ( 0.0000000, 15.54004107+]
## 0.0006157635 0.0006157635
## ( 3.7125257, 15.54004107+] ( 0.0000000, 15.57289528+]
## 0.0006157635 0.0006157635
## ( 4.8295688, 15.60574949] ( 0.0000000, 15.60574949+]
## 0.0006157635 0.0012315271
## ( 0.0000000, 15.63860370+] ( 3.7125257, 15.67145791]
## 0.0006157635 0.0006157635
## ( 0.0000000, 15.67145791+] ( 5.8480493, 15.73716632]
## 0.0006157635 0.0006157635
## ( 0.0000000, 15.73716632+] ( 5.0924025, 15.77002053]
## 0.0006157635 0.0006157635
## ( 0.0000000, 15.80287474+] ( 0.0000000, 15.86858316+]
## 0.0006157635 0.0006157635
## (15.1457906, 15.93429158] ( 0.0000000, 15.93429158+]
## 0.0006157635 0.0006157635
## ( 8.1149897, 15.93429158+] (15.6714579, 15.96714579+]
## 0.0006157635 0.0006157635
## ( 9.2320329, 16.00000000] ( 0.0000000, 16.03285421+]
## 0.0006157635 0.0006157635
## ( 3.6796715, 16.13141684] ( 0.0000000, 16.13141684+]
## 0.0006157635 0.0006157635
```

```
## ( 7.0965092, 16.16427105] ( 0.0000000, 16.16427105+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 16.19712526+] (10.9075975, 16.32854209]  
## 0.0018472906 0.0006157635  
## ( 0.0000000, 16.32854209+] ( 0.0000000, 16.36139630+]  
## 0.0012315271 0.0012315271  
## (16.3285421, 16.36139630+] ( 0.0000000, 16.39425051+]  
## 0.0006157635 0.0006157635  
## ( 2.4312115, 16.39425051+] ( 0.0000000, 16.42710472+]  
## 0.0006157635 0.0006157635  
## ( 4.7638604, 16.49281314] ( 0.0000000, 16.52566735+]  
## 0.0006157635 0.0012315271  
## ( 4.8624230, 16.52566735+] ( 0.0000000, 16.55852156+]  
## 0.0006157635 0.0006157635  
## (11.3675565, 16.62422998] (16.1314168, 16.65708419+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 16.68993840+] (13.1416838, 16.68993840+]  
## 0.0006157635 0.0006157635  
## (16.3613963, 16.72279261] ( 0.0000000, 16.72279261+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 16.75564682+] ( 5.5523614, 16.85420945]  
## 0.0006157635 0.0006157635  
## (10.7761807, 16.95277207] ( 3.5482546, 16.95277207+]  
## 0.0006157635 0.0006157635  
## ( 2.1683778, 16.98562628] ( 0.0000000, 17.01848049+]  
## 0.0006157635 0.0006157635  
## ( 0.8870637, 17.05133470] ( 0.0000000, 17.05133470+]  
## 0.0006157635 0.0006157635  
## ( 8.8049281, 17.05133470+] (16.6899384, 17.11704312]  
## 0.0006157635 0.0006157635  
## ( 0.6242300, 17.14989733] ( 4.2381930, 17.14989733+]  
## 0.0006157635 0.0006157635  
## ( 9.4620123, 17.21560575] ( 0.0000000, 17.28131417+]  
## 0.0006157635 0.0006157635  
## ( 0.8870637, 17.31416838] ( 0.0000000, 17.31416838+]  
## 0.0006157635 0.0006157635  
## ( 3.7125257, 17.41273101] (11.4989733, 17.41273101]  
## 0.0006157635 0.0006157635  
## (11.6632444, 17.44558522] ( 7.9507187, 17.47843943]  
## 0.0006157635 0.0006157635  
## (10.2505133, 17.47843943] ( 0.0000000, 17.47843943+]  
## 0.0006157635 0.0012315271  
## ( 4.3367556, 17.47843943+] ( 8.4106776, 17.47843943+]  
## 0.0006157635 0.0006157635  
## ( 4.4681725, 17.51129363] ( 4.3039014, 17.54414784]  
## 0.0006157635 0.0006157635  
## ( 3.5482546, 17.57700205] ( 0.0000000, 17.64271047+]  
## 0.0006157635 0.0006157635  
## ( 4.6324435, 17.67556468] (11.1047228, 17.70841889]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 17.74127310+] ( 8.7720739, 17.74127310+]
```

```
## 0.0006157635 0.0006157635
## (10.5462012, 17.74127310+] ( 0.0000000, 17.77412731+]
## 0.0006157635 0.0012315271
## ( 1.5112936, 17.77412731+] (12.1560575, 17.83983573]
## 0.0006157635 0.0006157635
## (12.0903491, 17.83983573+] ( 6.8665298, 18.03696099]
## 0.0006157635 0.0006157635
## (10.2176591, 18.03696099] ( 0.0000000, 18.03696099+]
## 0.0006157635 0.0006157635
## ( 4.5667351, 18.06981520] ( 7.0308008, 18.13552361+]
## 0.0006157635 0.0006157635
## ( 4.3039014, 18.16837782] (17.4784394, 18.16837782]
## 0.0006157635 0.0006157635
## ( 3.9096509, 18.20123203] ( 8.6078029, 18.23408624]
## 0.0006157635 0.0006157635
## ( 0.0000000, 18.23408624+] ( 0.6570842, 18.26694045]
## 0.0006157635 0.0006157635
## ( 0.0000000, 18.26694045+] ( 0.0000000, 18.39835729+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 18.43121150] ( 7.5893224, 18.43121150]
## 0.0006157635 0.0006157635
## ( 0.0000000, 18.43121150+] ( 0.0000000, 18.52977413+]
## 0.0012315271 0.0012315271
## ( 1.5770021, 18.56262834] ( 0.0000000, 18.59548255+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 18.62833676+] ( 3.6468172, 18.66119097]
## 0.0018472906 0.0006157635
## ( 7.1950719, 18.66119097+] ( 7.5564682, 18.66119097+]
## 0.0006157635 0.0006157635
## (11.5318275, 18.66119097+] ( 0.0000000, 18.72689938]
## 0.0006157635 0.0006157635
## ( 2.1683778, 18.72689938+] ( 3.3182752, 18.79260780]
## 0.0006157635 0.0006157635
## ( 0.0000000, 18.85831622+] ( 9.4291581, 18.85831622+]
## 0.0006157635 0.0006157635
## ( 3.0554415, 18.89117043+] ( 0.9527721, 18.92402464]
## 0.0006157635 0.0006157635
## ( 8.8706366, 18.92402464] ( 4.1724846, 18.95687885]
## 0.0006157635 0.0006157635
## ( 9.1334702, 18.95687885] ( 0.0000000, 18.95687885+]
## 0.0006157635 0.0006157635
## (17.4784394, 19.05544148+] (18.2669405, 19.12114990+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 19.18685832+] ( 9.1991786, 19.25256674]
## 0.0006157635 0.0006157635
## (16.1971253, 19.28542094] ( 3.7453799, 19.28542094+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 19.44969199+] ( 3.5811088, 19.48254620+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 19.51540041+] ( 5.5852156, 19.51540041+]
## 0.0012315271 0.0006157635
```

```
## ( 0.0000000, 19.54825462+] ( 3.8439425, 19.58110883]
## 0.0006157635 0.0006157635
## ( 3.9425051, 19.61396304] ( 5.4537988, 19.61396304+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 19.67967146+] (13.7002053, 19.71252567]
## 0.0006157635 0.0006157635
## ( 1.3141684, 19.71252567+] (10.6119097, 19.71252567+]
## 0.0006157635 0.0006157635
## (19.5154004, 19.77823409] (11.7289528, 19.77823409+]
## 0.0006157635 0.0006157635
## ( 3.5154004, 19.84394251] ( 3.0225873, 19.87679671]
## 0.0006157635 0.0006157635
## (10.0533881, 19.87679671+] ( 3.9096509, 19.90965092+]
## 0.0006157635 0.0006157635
## ( 5.4866530, 19.94250513] ( 6.9650924, 19.97535934]
## 0.0006157635 0.0006157635
## (12.8788501, 19.97535934] ( 9.9219713, 20.00821355]
## 0.0006157635 0.0006157635
## (16.5585216, 20.00821355+] ( 0.0000000, 20.04106776+]
## 0.0006157635 0.0006157635
## ( 5.8480493, 20.07392197+] ( 4.7967146, 20.13963039]
## 0.0006157635 0.0006157635
## (15.4086242, 20.13963039] ( 9.0020534, 20.20533881]
## 0.0006157635 0.0006157635
## ( 4.8624230, 20.23819302] ( 0.0000000, 20.23819302+]
## 0.0006157635 0.0006157635
## ( 5.6837782, 20.27104723] ( 0.0000000, 20.27104723+]
## 0.0006157635 0.0006157635
## ( 0.8213552, 20.30390144] ( 0.0000000, 20.33675565+]
## 0.0006157635 0.0006157635
## ( 5.5195072, 20.40246407+] ( 1.8398357, 20.43531828]
## 0.0006157635 0.0006157635
## (14.9486653, 20.43531828] ( 0.3285421, 20.43531828+]
## 0.0006157635 0.0006157635
## ( 8.2792608, 20.46817248] ( 0.0000000, 20.46817248+]
## 0.0006157635 0.0006157635
## (15.4086242, 20.46817248+] ( 1.1498973, 20.50102669]
## 0.0006157635 0.0006157635
## ( 0.0000000, 20.56673511+] ( 7.5564682, 20.66529774]
## 0.0006157635 0.0006157635
## (10.2833676, 20.69815195+] ( 0.0000000, 20.82956879+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 20.89527721+] ( 3.4825462, 20.96098563]
## 0.0006157635 0.0006157635
## (10.4147844, 21.02669405] ( 2.7268994, 21.05954825]
## 0.0006157635 0.0006157635
## ( 2.6283368, 21.09240246+] (13.6016427, 21.22381930]
## 0.0006157635 0.0006157635
## (17.7412731, 21.22381930] (14.6529774, 21.22381930+]
## 0.0006157635 0.0006157635
## ( 0.7885010, 21.28952772] ( 0.0000000, 21.32238193+]
```

```
## 0.0006157635 0.0006157635
## ( 3.6139630, 21.32238193+] ( 0.0000000, 21.35523614+]
## 0.0006157635 0.0006157635
## ( 6.0123203, 21.45379877] ( 0.0000000, 21.45379877+]
## 0.0006157635 0.0006157635
## ( 2.5297741, 21.45379877+] ( 6.6365503, 21.58521561+]
## 0.0006157635 0.0006157635
## ( 1.0184805, 21.61806982+] ( 9.9876797, 21.61806982+]
## 0.0006157635 0.0006157635
## ( 3.7453799, 21.68377823+] (14.1273101, 21.68377823+]
## 0.0006157635 0.0006157635
## ( 8.6406571, 21.91375770] ( 7.6221766, 21.94661191+]
## 0.0006157635 0.0006157635
## (13.0102669, 22.01232033] ( 1.1170431, 22.01232033+]
## 0.0006157635 0.0006157635
## ( 5.8151951, 22.01232033+] ( 0.0000000, 22.04517454+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 22.07802875] ( 0.0000000, 22.11088296+]
## 0.0006157635 0.0006157635
## ( 3.3511294, 22.14373717+] (13.9630390, 22.27515400+]
## 0.0006157635 0.0006157635
## ( 8.9363450, 22.30800821] (22.1108830, 22.30800821+]
## 0.0006157635 0.0006157635
## ( 4.2053388, 22.34086242+] ( 8.8706366, 22.34086242+]
## 0.0006157635 0.0006157635
## (17.6427105, 22.40657084+] ( 0.9856263, 22.50513347]
## 0.0006157635 0.0006157635
## ( 5.2238193, 22.50513347] ( 0.0000000, 22.50513347+]
## 0.0006157635 0.0006157635
## ( 9.6919918, 22.57084189+] ( 6.3737166, 22.73511294]
## 0.0006157635 0.0006157635
## ( 0.0000000, 22.73511294+] (18.0369610, 22.76796715]
## 0.0012315271 0.0006157635
## ( 2.8254620, 22.76796715+] (19.4496920, 22.86652977+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 22.96509240+] (13.9958932, 23.03080082]
## 0.0006157635 0.0006157635
## (13.1416838, 23.19507187+] (10.3819302, 23.26078029]
## 0.0006157635 0.0006157635
## ( 3.6796715, 23.32648871] ( 0.0000000, 23.42505133+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 23.49075975] ( 0.0000000, 23.49075975+]
## 0.0006157635 0.0006157635
## (14.2258727, 23.68788501] ( 0.0000000, 23.68788501+]
## 0.0006157635 0.0006157635
## ( 4.1067762, 23.72073922+] ( 6.1765914, 23.72073922+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 23.85215606+] ( 0.0000000, 23.88501027+]
## 0.0012315271 0.0006157635
## ( 6.4394251, 23.88501027+] ( 3.4825462, 23.91786448+]
## 0.0006157635 0.0006157635
```

(5.4866530, 23.95071869+] (10.1519507, 23.98357290]
0.0006157635 0.0006157635
(0.0000000, 23.98357290+] (0.0000000, 24.04928131+]
0.0006157635 0.0006157635
(0.0000000, 24.11498973+] (9.0677618, 24.11498973+]
0.0006157635 0.0006157635
(7.4907598, 24.18069815+] (9.1334702, 24.24640657]
0.0006157635 0.0006157635
(2.9568789, 24.27926078+] (10.3819302, 24.27926078+]
0.0006157635 0.0006157635
(11.2689938, 24.31211499] (2.8583162, 24.37782341]
0.0006157635 0.0006157635
(0.0000000, 24.41067762+] (15.8028747, 24.44353183+]
0.0006157635 0.0006157635
(18.6283368, 24.47638604] (0.0000000, 24.47638604+]
0.0006157635 0.0006157635
(12.6160164, 24.47638604+] (3.0225873, 24.57494867]
0.0006157635 0.0006157635
(0.0000000, 24.67351129+] (7.6221766, 24.67351129+]
0.0006157635 0.0006157635
(3.6139630, 24.73921971+] (9.9876797, 24.73921971+]
0.0006157635 0.0006157635
(7.4250513, 24.77207392+] (15.9342916, 24.83778234]
0.0006157635 0.0006157635
(0.0000000, 24.83778234+] (0.0000000, 24.87063655+]
0.0006157635 0.0006157635
(8.1149897, 24.90349076+] (0.0000000, 24.93634497+]
0.0006157635 0.0006157635
(4.5995893, 24.96919918] (0.0000000, 25.10061602+]
0.0006157635 0.0006157635
(6.1765914, 25.13347023] (10.7761807, 25.16632444]
0.0006157635 0.0006157635
(7.8850103, 25.33059548+] (4.6324435, 25.36344969]
0.0006157635 0.0006157635
(14.0616016, 25.36344969+] (0.0000000, 25.42915811+]
0.0006157635 0.0006157635
(16.5256674, 25.42915811+] (14.4887064, 25.52772074]
0.0006157635 0.0006157635
(0.7556468, 25.52772074+] (13.0431211, 25.59342916+]
0.0006157635 0.0006157635
(13.5359343, 25.69199179] (3.3182752, 25.72484600]
0.0006157635 0.0006157635
(8.4435318, 25.72484600+] (4.2381930, 25.85626283+]
0.0006157635 0.0006157635
(1.1827515, 25.95482546+] (3.1868583, 25.95482546+]
0.0006157635 0.0006157635
(23.9835729, 25.98767967] (12.7145791, 25.98767967+]
0.0006157635 0.0006157635
(1.0184805, 26.05338809+] (10.9075975, 26.08624230+]
0.0006157635 0.0006157635
(0.0000000, 26.21765914+] (0.0000000, 26.34907598+]

```
## 0.0006157635 0.0006157635
## (15.6057495, 26.54620123] (12.2874743, 26.61190965]
## 0.0006157635 0.0006157635
## (18.5297741, 26.64476386] (21.3223819, 26.64476386]
## 0.0006157635 0.0006157635
## ( 6.4065708, 26.74332649] ( 0.0000000, 26.77618070+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 26.84188912+] (14.3244353, 26.84188912+]
## 0.0006157635 0.0006157635
## ( 1.1827515, 26.94045175+] ( 7.6878850, 27.07186858]
## 0.0006157635 0.0006157635
## (10.7104723, 27.07186858] (12.8459959, 27.10472279]
## 0.0006157635 0.0006157635
## ( 6.0123203, 27.13757700] (26.7761807, 27.17043121]
## 0.0006157635 0.0006157635
## (24.1149897, 27.23613963+] (12.4845996, 27.33470226+]
## 0.0006157635 0.0006157635
## (10.6119097, 27.40041068+] (17.3141684, 27.40041068+]
## 0.0006157635 0.0006157635
## ( 7.8850103, 27.66324435+] ( 6.7351129, 27.79466119]
## 0.0006157635 0.0006157635
## ( 3.2854209, 27.86036961+] (18.5954825, 27.92607803+]
## 0.0006157635 0.0006157635
## ( 2.2997947, 28.05749487] ( 7.1293634, 28.09034908]
## 0.0006157635 0.0006157635
## ( 0.0000000, 28.09034908+] ( 4.3696099, 28.12320329]
## 0.0006157635 0.0006157635
## ( 1.3470226, 28.15605749] ( 4.0739220, 28.15605749+]
## 0.0006157635 0.0006157635
## ( 9.3634497, 28.25462012] ( 0.0000000, 28.25462012+]
## 0.0006157635 0.0006157635
## ( 6.8993840, 28.25462012+] ( 2.7926078, 28.38603696+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 28.41889117+] (15.2114990, 28.41889117+]
## 0.0006157635 0.0006157635
## ( 9.7905544, 28.45174538] ( 0.0000000, 28.51745380+]
## 0.0006157635 0.0006157635
## (23.8521561, 28.55030801+] ( 7.4579055, 28.58316222]
## 0.0006157635 0.0006157635
## (22.9650924, 28.64887064] ( 0.0000000, 28.71457906+]
## 0.0006157635 0.0006157635
## ( 3.9753593, 28.78028747+] (15.5728953, 28.87885010]
## 0.0006157635 0.0006157635
## ( 0.0000000, 29.01026694+] ( 7.6878850, 29.01026694+]
## 0.0006157635 0.0006157635
## ( 4.5995893, 29.04312115+] ( 3.2854209, 29.07597536+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 29.10882957+] ( 4.3367556, 29.20739220+]
## 0.0006157635 0.0006157635
## (16.1971253, 29.24024641] (24.8377823, 29.24024641+]
## 0.0006157635 0.0006157635
```



```
## ( 8.9363450, 29.43737166] (10.0205339, 29.53593429+]  
## 0.0006157635 0.0006157635  
## (12.7145791, 29.63449692] ( 8.4763860, 29.70020534+]  
## 0.0006157635 0.0006157635  
## (13.1745380, 29.76591376] ( 8.6078029, 29.79876797]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 29.93018480+] (12.3531828, 30.02874743]  
## 0.0006157635 0.0006157635  
## ( 8.4763860, 30.22587269+] (14.4558522, 30.29158111]  
## 0.0006157635 0.0006157635  
## ( 8.5092402, 30.42299795] ( 0.7556468, 30.52156057]  
## 0.0006157635 0.0006157635  
## ( 1.1170431, 30.52156057] (13.2731006, 30.68583162+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 30.71868583+] ( 1.0841889, 30.98151951]  
## 0.0006157635 0.0006157635  
## ( 9.7577002, 31.37577002+] (11.4332649, 31.44147844]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 31.70431211+] (11.6632444, 31.73716632]  
## 0.0006157635 0.0006157635  
## ( 0.2956879, 31.73716632+] ( 9.1663244, 31.73716632+]  
## 0.0006157635 0.0006157635  
## (15.6057495, 31.83572895] ( 5.0924025, 31.86858316]  
## 0.0006157635 0.0006157635  
## ( 8.4106776, 31.86858316+] (19.5154004, 31.90143737+]  
## 0.0006157635 0.0006157635  
## ( 3.8767967, 31.93429158+] ( 4.3367556, 32.00000000]  
## 0.0006157635 0.0006157635  
## ( 1.2484600, 32.03285421] ( 3.1540041, 32.06570842+]  
## 0.0006157635 0.0006157635  
## (12.0574949, 32.06570842+] ( 8.9034908, 32.49281314+]  
## 0.0006157635 0.0006157635  
## ( 3.3182752, 32.72279261] ( 2.2012320, 32.78850103+]  
## 0.0006157635 0.0006157635  
## (21.4537988, 32.82135524] ( 0.0000000, 32.82135524+]  
## 0.0006157635 0.0006157635  
## ( 8.3121150, 32.82135524+] ( 0.0000000, 32.85420945+]  
## 0.0006157635 0.0006157635  
## ( 9.6591376, 32.88706366] (26.2176591, 32.88706366+]  
## 0.0006157635 0.0006157635  
## ( 8.5749487, 33.05133470] ( 5.4209446, 33.05133470+]  
## 0.0006157635 0.0006157635  
## ( 0.0000000, 33.08418891+] ( 4.1724846, 33.11704312+]  
## 0.0006157635 0.0006157635  
## (14.0944559, 33.28131417] ( 0.0000000, 33.31416838+]  
## 0.0006157635 0.0006157635  
## ( 0.8870637, 33.34702259] ( 0.7885010, 33.57700205]  
## 0.0006157635 0.0006157635  
## ( 0.1642710, 33.64271047] (15.1786448, 33.74127310+]  
## 0.0006157635 0.0006157635  
## (11.4989733, 33.77412731+] ( 3.5154004, 33.83983573]
```

```
## 0.0006157635 0.0006157635
## ( 0.0000000, 33.93839836+] (21.3552361, 33.97125257+]
## 0.0006157635 0.0006157635
## (11.1704312, 34.00410678+] (12.4188912, 34.06981520+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 34.10266940+] (24.9363450, 34.23408624+]
## 0.0006157635 0.0006157635
## ( 2.7926078, 34.52977413] (12.5831622, 34.75975359]
## 0.0006157635 0.0006157635
## ( 7.6221766, 34.79260780] (18.4312115, 35.15400411]
## 0.0006157635 0.0006157635
## ( 1.4127310, 35.21971253+] ( 2.1026694, 35.28542094+]
## 0.0006157635 0.0006157635
## ( 6.2094456, 35.31827515+] (12.1232033, 35.35112936+]
## 0.0006157635 0.0006157635
## (19.6796715, 35.44969199] ( 7.7864476, 35.48254620]
## 0.0006157635 0.0006157635
## ( 9.9876797, 35.48254620+] ( 1.2484600, 35.51540041]
## 0.0006157635 0.0006157635
## ( 1.0841889, 35.64681725] ( 0.0000000, 36.04106776+]
## 0.0006157635 0.0006157635
## ( 2.8254620, 36.10677618] (10.9075975, 36.13963039+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 36.36960986+] (36.3696099, 36.40246407]
## 0.0006157635 0.0006157635
## ( 0.0000000, 36.46817248+] ( 6.9650924, 36.53388090+]
## 0.0006157635 0.0006157635
## (14.2915811, 36.56673511] ( 1.3798768, 36.59958932]
## 0.0006157635 0.0006157635
## (29.1088296, 36.63244353+] ( 3.9096509, 36.79671458]
## 0.0006157635 0.0006157635
## ( 0.0000000, 36.89527721+] (16.5256674, 37.05954825+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 37.09240246+] ( 0.0000000, 37.25667351+]
## 0.0006157635 0.0006157635
## (20.2710472, 37.28952772] ( 5.5195072, 37.35523614]
## 0.0006157635 0.0006157635
## ( 6.7022587, 37.35523614] ( 5.3223819, 37.38809035+]
## 0.0006157635 0.0006157635
## ( 6.1765914, 37.45379877] (17.0184805, 37.45379877]
## 0.0006157635 0.0006157635
## (10.2176591, 37.65092402] (36.0410678, 37.81519507+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 37.91375770+] (18.6283368, 38.04517454]
## 0.0006157635 0.0006157635
## (20.0410678, 38.11088296] (10.2833676, 38.17659138+]
## 0.0006157635 0.0006157635
## ( 8.1149897, 38.20944559+] ( 8.0821355, 38.24229979+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 38.76796715+] (16.1971253, 38.86652977+]
## 0.0006157635 0.0006157635
```

(10.6119097, 38.99794661+] (26.8418891, 39.19507187+]
0.0006157635 0.0006157635
(20.3367556, 39.22792608+] (11.2032854, 39.35934292]
0.0006157635 0.0006157635
(0.0000000, 39.52361396+] (26.3490760, 39.52361396+]
0.0006157635 0.0006157635
(36.4681725, 39.75359343+] (18.8583162, 39.88501027+]
0.0006157635 0.0006157635
(0.0000000, 40.11498973+] (13.0759754, 40.14784394+]
0.0006157635 0.0006157635
(14.9486653, 40.21355236+] (4.6652977, 40.24640657]
0.0006157635 0.0006157635
(12.3860370, 40.44353183+] (20.8295688, 40.54209446+]
0.0006157635 0.0006157635
(10.6447639, 40.77207392+] (15.2443532, 40.93634497+]
0.0006157635 0.0006157635
(0.9199179, 40.96919918] (5.3552361, 40.96919918+]
0.0006157635 0.0006157635
(0.0000000, 41.03490760+] (14.9158111, 41.10061602]
0.0006157635 0.0006157635
(0.0000000, 41.13347023+] (3.7782341, 41.19917864]
0.0006157635 0.0006157635
(12.4188912, 41.19917864+] (0.0000000, 41.42915811+]
0.0006157635 0.0006157635
(10.8747433, 41.65913758+] (0.0000000, 41.82340862+]
0.0006157635 0.0006157635
(9.7577002, 41.88911704+] (0.0000000, 41.95482546+]
0.0006157635 0.0006157635
(16.0328542, 42.67761807] (41.9548255, 42.77618070+]
0.0006157635 0.0006157635
(0.0000000, 42.84188912+] (6.6036961, 43.03901437+]
0.0006157635 0.0006157635
(0.0000000, 43.10472279+] (10.3819302, 43.10472279+]
0.0006157635 0.0006157635
(7.5893224, 43.43326489+] (0.0000000, 43.49897331+]
0.0006157635 0.0006157635
(0.0000000, 43.92607803+] (0.0000000, 44.58316222+]
0.0006157635 0.0006157635
(14.2587269, 44.64887064+] (13.7330595, 44.81314168]
0.0006157635 0.0006157635
(22.0451745, 44.87885010] (23.4907598, 44.97741273+]
0.0006157635 0.0006157635
(0.0000000, 45.04312115+] (4.5010267, 45.07597536]
0.0012315271 0.0006157635
(13.7659138, 45.33880903] (0.9527721, 45.40451745+]
0.0006157635 0.0006157635
(5.4537988, 45.40451745+] (8.3121150, 45.50308008]
0.0006157635 0.0006157635
(0.0000000, 45.53593429+] (41.0349076, 45.86447639+]
0.0006157635 0.0006157635
(15.4743326, 45.93018480+] (0.0000000, 45.96303901+]

```
## 0.0006157635 0.0006157635
## ( 6.1765914, 46.19301848+] (15.8685832, 46.32443532+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 46.52156057+] ( 0.0000000, 46.94866530]
## 0.0006157635 0.0006157635
## ( 3.0554415, 46.98151951+] (17.0513347, 47.37577002+]
## 0.0006157635 0.0006157635
## (40.1149897, 47.47433265+] (18.2340862, 47.50718686+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 48.26283368+] ( 0.0000000, 48.42710472+]
## 0.0006157635 0.0006157635
## (20.2381930, 48.68993840+] ( 6.1108830, 49.41273101]
## 0.0006157635 0.0006157635
## ( 9.1334702, 49.74127310] ( 0.0000000, 50.03696099+]
## 0.0006157635 0.0006157635
## (12.6488706, 50.10266940+] (23.4250513, 51.48254620+]
## 0.0006157635 0.0006157635
## (25.1006160, 51.54825462+] (33.0841889, 52.00821355+]
## 0.0006157635 0.0006157635
## ( 4.7310062, 52.07392197+] ( 6.4065708, 52.86242300+]
## 0.0006157635 0.0006157635
## ( 6.1108830, 52.99383984] ( 0.0000000, 54.24229979+]
## 0.0006157635 0.0006157635
## (18.9568789, 54.30800821] (11.7289528, 55.03080082]
## 0.0006157635 0.0006157635
## (24.4106776, 55.85215606] ( 0.0000000, 56.14784394+]
## 0.0006157635 0.0006157635
## ( 8.0164271, 56.31211499+] ( 0.0000000, 56.87063655+]
## 0.0006157635 0.0006157635
## ( 1.4455852, 57.39630390] (16.1642710, 57.62628337+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 57.92197125+] ( 3.2854209, 57.98767967+]
## 0.0006157635 0.0006157635
## (19.1868583, 59.30184805+] ( 0.0000000, 59.43326489+]
## 0.0006157635 0.0006157635
## ( 5.6180698, 59.49897331+] (22.5051335, 59.95893224+]
## 0.0006157635 0.0006157635
## (30.7186858, 60.84599589+] ( 0.0000000, 61.37166324+]
## 0.0006157635 0.0006157635
## (17.7741273, 61.37166324+] ( 8.2464066, 61.96303901+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 62.02874743+] (24.0492813, 62.22587269+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 65.05133470+] ( 0.0000000, 65.08418891+]
## 0.0006157635 0.0006157635
## (10.0533881, 65.47843943+] (13.5030801, 66.03696099]
## 0.0006157635 0.0006157635
## (16.3285421, 66.62833676+] ( 8.4106776, 67.61396304+]
## 0.0006157635 0.0006157635
## ( 0.0000000, 67.67967146+] (34.1026694, 67.90965092+]
## 0.0006157635 0.0006157635
```

(43.9260780, 68.36960986] (12.9774127, 68.36960986+]
0.0006157635 0.0006157635
(0.0000000, 68.73100616+] (0.0000000, 68.76386037+]
0.0006157635 0.0006157635
(24.8706366, 69.02669405] (8.3121150, 69.32238193]
0.0006157635 0.0006157635
(18.4312115, 69.74948665+] (29.9301848, 69.84804928+]
0.0006157635 0.0006157635
(10.5133470, 70.27515400+] (24.6735113, 70.93223819]
0.0006157635 0.0006157635
(3.3182752, 73.00205339+] (32.8213552, 74.21765914+]
0.0006157635 0.0006157635
(4.2381930, 74.41478439+] (61.3716632, 74.94045175+]
0.0006157635 0.0006157635
(0.0000000, 75.46611910+] (0.0000000, 75.66324435+]
0.0006157635 0.0006157635
(3.6468172, 75.69609856+] (0.0000000, 76.12320329+]
0.0006157635 0.0006157635
(0.4271047, 76.55030801+] (13.1745380, 76.91170431+]
0.0006157635 0.0006157635
(0.0000000, 77.40451745+] (18.5297741, 78.45585216+]
0.0006157635 0.0006157635
(13.6016427, 78.91581109+] (11.1047228, 79.93429158]
0.0006157635 0.0006157635
(33.3141684, 80.39425051+] (0.8213552, 80.45995893+]
0.0006157635 0.0006157635
(0.0000000, 80.62422998+] (8.8377823, 81.54414784+]
0.0006157635 0.0006157635
(12.7802875, 83.12114990+] (1.0513347, 83.44969199+]
0.0006157635 0.0006157635
(23.8521561, 83.81108830+] (23.8850103, 84.17248460]
0.0006157635 0.0006157635
(77.4045175, 84.69815195] (10.1848049, 84.82956879]
0.0006157635 0.0006157635
(0.0000000, 85.28952772+] (9.0020534, 87.12936345+]
0.0006157635 0.0006157635
(46.5215606, 87.55646817+] (0.2299795, 87.68788501+]
0.0006157635 0.0006157635
(0.0000000, 88.21355236+] (12.6817248, 89.72484600+]
0.0006157635 0.0006157635
(0.0000000, 90.94045175+] (8.0821355, 91.10472279+]
0.0006157635 0.0006157635
(0.0000000, 91.20328542+] (3.5811088, 93.04312115+]
0.0006157635 0.0006157635
(2.8254620, 95.40862423+] (8.4763860, 95.77002053+]
0.0006157635 0.0006157635
(4.5995893, 98.85831622] (0.0000000, 99.61396304+]
0.0006157635 0.0006157635
(3.0225873, 99.67967146+] (3.6468172, 99.77823409+]
0.0006157635 0.0006157635
(0.0000000, 99.94250513+] (23.6878850,101.88090349+]

```
##          0.0006157635          0.0006157635
## (33.9383984,103.81930185+] ( 6.0780287,104.04928131+]
##          0.0006157635          0.0006157635
## ( 4.1724846,104.70636550+] (10.0862423,115.77823409+]
##          0.0006157635          0.0006157635
## (10.7761807,118.60369610+] (16.3942505,122.77618070+]
##          0.0006157635          0.0006157635
## ( 0.0000000,123.40041068+] (45.0431211,123.63039014+]
##          0.0006157635          0.0006157635
## ( 0.0000000,124.28747433+] ( 0.0000000,124.58316222+]
##          0.0006157635          0.0006157635
## (22.7351129,126.29158111+] ( 0.0000000,126.85010267+]
##          0.0006157635          0.0006157635
## ( 0.0000000,132.50102669+] ( 0.0000000,133.45379877+]
##          0.0006157635          0.0006157635
## ( 0.0000000,134.20944559+] (37.0924025,135.78644764+]
##          0.0006157635          0.0006157635
```

```
library(dplyr)
```

```
# Conta Le osservazioni per ciascuna classe
table(data$Death)
```

```
##
##    0    1
## 530 386
```

```
# Numero di osservazioni nella classe minoritaria (Death==1)
n_minority <- sum(data$Death == 1)
```

```
# Prima di combinare (opzione 1)
data_minority <- na.omit(data_minority)
```

```
## Error in eval(expr, envir, enclos): object 'data_minority' not found
```

```
data_majority_under <- na.omit(data_majority_under)
```

```
## Error in eval(expr, envir, enclos): object 'data_majority_under' not found
```

```
data_balanced <- bind_rows(data_minority, data_majority_under)
```

```
## Error in eval(expr, envir, enclos): object 'data_minority' not found
```

```
# Oppure dopo combinazione (opzione 2)
data_balanced <- bind_rows(data_minority, data_majority_under)
```

```
## Error in eval(expr, envir, enclos): object 'data_minority' not found
```

```
data_balanced <- na.omit(data_balanced)
```

```
## Error in eval(expr, envir, enclos): object 'data_balanced' not found
```

```
# Controllo bilanciamento
table(data_balanced$Death)
```

```
## Error in eval(expr, envir, enclos): object 'data_balanced' not found
# Controlla il nuovo bilanciamento
table(data_balanced$Death)

## Error in eval(expr, envir, enclos): object 'data_balanced' not found
# Visualizza la proporzione riequilibrata
prop.table(table(data_balanced$Death))

## Error in eval(expr, envir, enclos): object 'data_balanced' not found
summary(data_balanced)

## Error in h(simpleError(msg, call)): error in evaluating the argument 'object'
in selecting a method for function 'summary': object 'data_balanced' not found
set.seed(123)
library(caret)

# Partizione corretta basata su data_balanced
train_index <- createDataPartition(data_balanced$Death, p = 0.8, list = FALSE)

## Error in eval(expr, envir, enclos): object 'data_balanced' not found
train_data <- data_balanced[train_index, ]

## Error in eval(expr, envir, enclos): object 'data_balanced' not found
test_data <- data_balanced[-train_index, ]

## Error in eval(expr, envir, enclos): object 'data_balanced' not found
# Converti Death in fattore
train_data$Death <- as.factor(train_data$Death)
test_data$Death <- as.factor(test_data$Death)

# Controlla se ci sono ancora NA
sum(is.na(train_data))

## [1] 0
sum(is.na(test_data))

## [1] 0
```

```
summary(train data)
```

```

## Max. :95.00 Max. :1.0000 Max. :1.00000
## HF CVA CKD LogCEA
## Min. :0.00000 Min. :0.00000 Min. :0.0000 Min. : -0.3565
## 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.: 0.5911
## Median :0.00000 Median :0.00000 Median :0.0000 Median : 1.2420
## Mean :0.04911 Mean :0.05593 Mean :0.1364 Mean : 1.3821
## 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.: 1.9835
## Max. :1.00000 Max. :1.00000 Max. :1.0000 Max. : 4.1797
## Laparoscopic TumorLOC EA Log2AT RBC
## Min. :0.00000 0:510 Min. :0.0000 Min. :5.644 Min. :0.0000
## 1st Qu.:0.00000 1:223 1st Qu.:0.0000 1st Qu.:8.077 1st Qu.:0.0000
## Median :0.00000 Median :0.0000 Median :8.299 Median :0.0000
## Mean :0.04366 Mean :0.1583 Mean :8.341 Mean :0.5266
## 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:8.626 3rd Qu.:1.0000
## Max. :1.00000 Max. :1.0000 Max. :9.907 Max. :2.0000
## Liver_Only Cell_diff Mucin_TYPE SignetRING
## Min. :0.0000 Min. :1.000 Min. :0.00000 Min. :0.00000
## 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.:0.00000 1st Qu.:0.00000
## Median :0.0000 Median :1.000 Median :0.00000 Median :0.00000
## Mean :0.3888 Mean :1.128 Mean :0.06958 Mean :0.04502
## 3rd Qu.:1.0000 3rd Qu.:1.000 3rd Qu.:0.00000 3rd Qu.:0.00000
## Max. :1.0000 Max. :2.000 Max. :1.00000 Max. :1.00000
## Lymphovascularinvasion perineural CT RT
## Min. :0.000 Min. :0.0000 Min. :0.0000 Min. :0.0000
## 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:0.0000
## Median :1.000 Median :0.0000 Median :1.0000 Median :0.0000
## Mean :0.513 Mean :0.2251 Mean :0.8881 Mean :0.1105
## 3rd Qu.:1.000 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:0.0000
## Max. :1.000 Max. :1.0000 Max. :1.0000 Max. :1.0000
## NACTRT Death Progress IntervalOD
## Min. :0.0000 No :419 Min. :0.000 Min. : 0.03285
## 1st Qu.:0.0000 Yes:314 1st Qu.:1.000 1st Qu.: 8.08214
## Median :0.0000 Median :1.000 Median : 17.83984
## Mean :0.1623 Mean :0.794 Mean : 24.79991
## 3rd Qu.:0.0000 3rd Qu.:1.000 3rd Qu.: 31.86858
## Max. :1.0000 Max. :1.000 Max. :135.78645
## IntervalOR AJCC_bin id Progress_event
## Min. : 0.03285 0:424 Min. : 2.0 Min. :0.000
## 1st Qu.: 3.02259 1:309 1st Qu.:226.0 1st Qu.:1.000
## Median : 5.84805 Median :450.0 Median :1.000
## Mean : 11.08083 Mean :456.3 Mean :0.794
## 3rd Qu.: 12.61602 3rd Qu.:687.0 3rd Qu.:1.000
## Max. :133.45380 Max. :916.0 Max. :1.000

```

Training della Logistic regression

```

set.seed(123)
logit_model <- train(
  Death ~ .,
  data = train_data,
  method = "glm",
  family = "binomial",

```



```

trControl = trainControl(method = "none") # niente cross-validation per ora,
step successivo
)

# Predizione delle classi
pred_class <- predict(logit_model, newdata = test_data)

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

# Predizione delle probabilità (per AUC)
pred_prob <- predict(logit_model, newdata = test_data, type = "prob")

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

# Confusion Matrix
conf_mat <- confusionMatrix(pred_class, test_data$Death)
print(conf_mat)

## Confusion Matrix and Statistics
##
##              Reference
## Prediction No Yes
##      No   72   41
##      Yes  39   31
##
##              Accuracy : 0.5628
##              95% CI : (0.4877, 0.6359)
##      No Information Rate : 0.6066
##      P-Value [Acc > NIR] : 0.9003
##
##              Kappa : 0.0796
##
##  Mcnemar's Test P-Value : 0.9110
##
##              Sensitivity : 0.6486
##              Specificity : 0.4306
##              Pos Pred Value : 0.6372
##              Neg Pred Value : 0.4429
##              Prevalence : 0.6066
##              Detection Rate : 0.3934
##              Detection Prevalence : 0.6175
##              Balanced Accuracy : 0.5396
##
##              'Positive' Class : No
##

# AUC con pROC
library(pROC)

```

```

roc_obj <- roc(test_data$Death, pred_prob[, 2])

## Setting levels: control = No, case = Yes

## Setting direction: controls < cases

auc(roc_obj)

## Area under the curve: 0.6196

summary(logit_model$finalModel)

##
## Call:
## NULL
##
## Coefficients: (1 not defined because of singularities)
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.5597915   1.7240337  -0.325  0.745409
## Age           -0.0097777   0.0068488  -1.428  0.153393
## Gender2        0.1344741   0.1779651   0.756  0.449877
## ASA31          0.2218841   0.1952950   1.136  0.255895
## DM             0.2961600   0.2068710   1.432  0.152254
## CAD           -0.2474317   0.3456455  -0.716  0.474082
## HF            -0.0496266   0.3887257  -0.128  0.898414
## CVA            0.9000041   0.3824615   2.353  0.018613 *
## CKD            0.5046835   0.2503024   2.016  0.043769 *
## LogCEA         0.1537012   0.0967441   1.589  0.112119
## Laparoscopic  -0.3717732   0.4198880  -0.885  0.375935
## TumorLOC1      0.0604544   0.1897793   0.319  0.750067
## EA            -0.2909285   0.2479113  -1.174  0.240588
## Log2AT        -0.0680888   0.1878700  -0.362  0.717035
## RBC           0.3470738   0.1334364   2.601  0.009294 **
## Liver_Only     0.3794461   0.2512146   1.510  0.130930
## Cell_diff      0.2450865   0.2691399   0.911  0.362491
## Mucin_TYPE    -0.3372062   0.3443313  -0.979  0.327428
## SignetRING      0.2779922   0.4411310   0.630  0.528576
## Lymphovascularinvasion -0.1175056   0.1855072  -0.633  0.526454
## perineural     0.0316183   0.2145761   0.147  0.882854
## CT             0.1569898   0.3233707   0.485  0.627336
## RT             0.4139675   0.2627090   1.576  0.115080
## NACTRT        -0.2536589   0.2445127  -1.037  0.299547
## Progress       1.3646557   0.2769138   4.928  8.3e-07 ***
## IntervalOD    -0.0197915   0.0054928  -3.603  0.000314 ***
## IntervalOR    -0.0155006   0.0114291  -1.356  0.175025
## AJCC_bin1     0.4593299   0.2504911   1.834  0.066696 .
## id           -0.0008913   0.0003675  -2.425  0.015297 *
## Progress_event      NA          NA          NA          NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

```

## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1001.06 on 732 degrees of freedom
## Residual deviance: 865.88 on 704 degrees of freedom
## AIC: 923.88
##
## Number of Fisher Scoring iterations: 5

# 1. Modello iniziale completo
full_model <- glm(Death ~ ., data = train_data, family = "binomial")

# 2. Selezione stepwise (AIC)
step_model <- step(full_model, direction = "both", trace = FALSE)

# 3. Sommario del modello selezionato
summary(step_model)

##
## Call:
## glm(formula = Death ~ CVA + CKD + LogCEA + RBC + RT + Progress +
## IntervalOD + IntervalOR + AJCC_bin + id, family = "binomial",
## data = train_data)
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.1814318 0.3309172 -3.570 0.000357 ***
## CVA 0.8725531 0.3646286 2.393 0.016712 *
## CKD 0.4153874 0.2424035 1.714 0.086599 .
## LogCEA 0.1896105 0.0910271 2.083 0.037250 *
## RBC 0.3716770 0.1215958 3.057 0.002238 **
## RT 0.3918939 0.2577716 1.520 0.128432
## Progress 1.3520654 0.2450282 5.518 3.43e-08 ***
## IntervalOD -0.0192654 0.0052142 -3.695 0.000220 ***
## IntervalOR -0.0156049 0.0112493 -1.387 0.165384
## AJCC_bin1 0.2468352 0.1696257 1.455 0.145621
## id -0.0007727 0.0003093 -2.498 0.012488 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1001.06 on 732 degrees of freedom
## Residual deviance: 878.75 on 722 degrees of freedom
## AIC: 900.75
##
## Number of Fisher Scoring iterations: 5

# Predizioni su test
pred_prob <- predict(step_model, newdata = test_data, type = "response")
pred_class <- ifelse(pred_prob > 0.5, "1", "0")
pred_class <- factor(pred_class, levels = levels(test_data$Death))

```

```
# Confusion matrix
confusionMatrix(pred_class, test_data$Death)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction No Yes
##           No    0    0
##           Yes    0    0
##
##           Accuracy : NaN
##           95% CI : (NA, NA)
##           No Information Rate : NA
##           P-Value [Acc > NIR] : NA
##
##           Kappa : NaN
##
## Mcnemar's Test P-Value : NA
##
##           Sensitivity : NA
##           Specificity : NA
##           Pos Pred Value : NA
##           Neg Pred Value : NA
##           Prevalence : NaN
##           Detection Rate : NaN
##           Detection Prevalence : NaN
##           Balanced Accuracy : NA
##
##           'Positive' Class : No
##
```

```
# AUC
library(pROC)
roc_obj <- roc(test_data$Death, pred_prob)

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

auc(roc_obj)

## Area under the curve: 0.6418
```

K fold

```
x <- train_data[, setdiff(names(train_data), "Death")]
y <- train_data$Death
```

```
# Controllo per RFE con cross-validation
ctrl <- rfeControl(functions = lrFuncs, # Logistic regression
                   method = "cv",
                   number = 5) # 5-fold CV
```

```

# Esecuzione di RFE
set.seed(123)
rfe_results <- rfe(x = x, y = y,
                  sizes = c(1:10, 15, 20, 30),
                  rfeControl = ctrl)

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning: rfe is expecting 29 importance values but only has 28. This may be
## caused by having zero-variance predictors, excessively-correlated predictors,
## factor predictors that were expanded into dummy variables or you may have
## failed to drop one of your dummy variables.

## Warning: There were missing importance values. There may be linear dependencies
## in your predictor variables

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning: rfe is expecting 29 importance values but only has 28. This may be
## caused by having zero-variance predictors, excessively-correlated predictors,
## factor predictors that were expanded into dummy variables or you may have
## failed to drop one of your dummy variables.

## Warning: There were missing importance values. There may be linear dependencies
## in your predictor variables

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning: rfe is expecting 29 importance values but only has 28. This may be
## caused by having zero-variance predictors, excessively-correlated predictors,
## factor predictors that were expanded into dummy variables or you may have
## failed to drop one of your dummy variables.

## Warning: There were missing importance values. There may be linear dependencies
## in your predictor variables

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

```

```

## Warning: rfe is expecting 29 importance values but only has 28. This may be
## caused by having zero-variance predictors, excessively-correlated predictors,
## factor predictors that were expanded into dummy variables or you may have
## failed to drop one of your dummy variables.

## Warning: There were missing importance values. There may be linear dependencies
## in your predictor variables

## Error in {: task 1 failed - "Can't subset columns that don't exist.
## X Columns `AJCC_bin1` and `ASA31` don't exist."

# Visualizza Le variabili selezionate
rfe_results$optVariables

## Error in eval(expr, envir, enclos): object 'rfe_results' not found

# Plot delle performance in base al numero di variabili
plot(rfe_results)

## Error in eval(expr, envir, enclos): object 'rfe_results' not found

# Usa solo Le variabili selezionate
final_vars <- rfe_results$optVariables

## Error in eval(expr, envir, enclos): object 'rfe_results' not found

# Fit del modello su queste variabili
final_logit <- glm(
  as.formula(paste("Death ~", paste(final_vars, collapse = "+"))),
  data = train_data,
  family = "binomial"
)

## Error in eval(mf, parent.frame()): object 'final_vars' not found

# Summary
summary(final_logit)

## Error in h(simpleError(msg, call)): error in evaluating the argument 'object'
## in selecting a method for function 'summary': object 'final_logit' not found

# Predizione su test
pred_prob <- predict(final_logit, newdata = test_data, type = "response")

## Error in eval(expr, envir, enclos): object 'final_logit' not found

pred_class <- ifelse(pred_prob > 0.5, "1", "0")
pred_class <- factor(pred_class, levels = levels(test_data$Death))

# Confusion matrix
confusionMatrix(pred_class, test_data$Death)

## Confusion Matrix and Statistics
##

```

```

##           Reference
## Prediction No Yes
##           No    0    0
##           Yes   0    0
##
##           Accuracy : NaN
##           95% CI : (NA, NA)
##           No Information Rate : NA
##           P-Value [Acc > NIR] : NA
##
##           Kappa : NaN
##
## McNemar's Test P-Value : NA
##
##           Sensitivity : NA
##           Specificity : NA
##           Pos Pred Value : NA
##           Neg Pred Value : NA
##           Prevalence : NaN
##           Detection Rate : NaN
##           Detection Prevalence : NaN
##           Balanced Accuracy : NA
##
##           'Positive' Class : No
##
# AUC
roc_obj <- roc(test_data$Death, pred_prob)

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases
auc(roc_obj)

## Area under the curve: 0.6418

set.seed(123)
library(caret)
train_index <- createDataPartition(data_balanced$Death, p = 0.8, list = FALSE)

## Error in eval(expr, envir, enclos): object 'data_balanced' not found
train_data <- data_balanced[train_index, ]

## Error in eval(expr, envir, enclos): object 'data_balanced' not found
test_data <- data_balanced[-train_index, ]

## Error in eval(expr, envir, enclos): object 'data_balanced' not found
skim_without_charts(train_data)

```

Data summary

Name train_data
Number of rows 733
Number of columns 30

Column type frequency:

factor 5
numeric 25

Group variables None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
Gender	0	1	FALSE	2	1: 461, 2: 272
ASA3	0	1	FALSE	2	0: 457, 1: 276
TumorLOC	0	1	FALSE	2	0: 510, 1: 223
Death	0	1	FALSE	2	No: 419, Yes: 314
AJCC_bin	0	1	FALSE	2	0: 424, 1: 309

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Age	0	1	64.97	13.79	18.00	55.00	65.00	76.00	95.00
DM	0	1	0.21	0.41	0.00	0.00	0.00	0.00	1.00
CAD	0	1	0.07	0.26	0.00	0.00	0.00	0.00	1.00
HF	0	1	0.05	0.22	0.00	0.00	0.00	0.00	1.00
CVA	0	1	0.06	0.23	0.00	0.00	0.00	0.00	1.00
CKD	0	1	0.14	0.34	0.00	0.00	0.00	0.00	1.00
LogCEA	0	1	1.38	0.91	-0.36	0.59	1.24	1.98	4.18
Laparoscopic	0	1	0.04	0.20	0.00	0.00	0.00	0.00	1.00
EA	0	1	0.16	0.37	0.00	0.00	0.00	0.00	1.00
Log2AT	0	1	8.34	0.49	5.64	8.08	8.30	8.63	9.91
RBC	0	1	0.53	0.69	0.00	0.00	0.00	1.00	2.00
Liver_Only	0	1	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Cell_diff	0	1	1.13	0.33	1.00	1.00	1.00	1.00	2.00
Mucin_TYPE	0	1	0.07	0.25	0.00	0.00	0.00	0.00	1.00

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
SignetRING	0	1	0.05	0.21	0.00	0.00	0.00	0.00	1.00
Lymphovascularinvasion	0	1	0.51	0.50	0.00	0.00	1.00	1.00	1.00
perineural	0	1	0.23	0.42	0.00	0.00	0.00	0.00	1.00
CT	0	1	0.89	0.32	0.00	1.00	1.00	1.00	1.00
RT	0	1	0.11	0.31	0.00	0.00	0.00	0.00	1.00
NACTRT	0	1	0.16	0.37	0.00	0.00	0.00	0.00	1.00
Progress	0	1	0.79	0.40	0.00	1.00	1.00	1.00	1.00
IntervalOD	0	1	24.80	24.67	0.03	8.08	17.84	31.87	135.79
IntervalOR	0	1	11.08	16.66	0.03	3.02	5.85	12.62	133.45
id	0	1	456.31	265.73	2.00	226.00	450.00	687.00	916.00
Progress_event	0	1	0.79	0.40	0.00	1.00	1.00	1.00	1.00

RANDOM FOREST

Si inizia con una k fold

```
# Rinominare i livelli della variabile target
train_data$Death <- factor(train_data$Death, levels = c("0", "1"), labels =
c("No", "Yes"))
test_data$Death <- factor(test_data$Death, levels = c("0", "1"), labels =
c("No", "Yes"))

library(caret)
set.seed(123)

rf_model <- train(
  Death ~ .,
  data = train_data,
  method = "rf",
  trControl = trainControl(
    method = "cv",
    number = 5,
    classProbs = TRUE,
    summaryFunction = twoClassSummary
  ),
  metric = "ROC"
)

## Error in na.fail.default(structure(list(Death = structure(c(NA_integer_, :
missing values in object
```

```

# Predizioni
pred_class <- predict(rf_model, newdata = test_data)
pred_prob <- predict(rf_model, newdata = test_data, type = "prob")

# Confusion matrix
confusionMatrix(pred_class, test_data$Death)

## Confusion Matrix and Statistics
##
##              Reference
## Prediction No Yes
##      No    0    0
##      Yes    0    0
##
##              Accuracy : NaN
##              95% CI : (NA, NA)
##      No Information Rate : NA
##      P-Value [Acc > NIR] : NA
##
##              Kappa : NaN
##
##      McNemar's Test P-Value : NA
##
##              Sensitivity : NA
##              Specificity : NA
##              Pos Pred Value : NA
##              Neg Pred Value : NA
##              Prevalence : NaN
##              Detection Rate : NaN
##      Detection Prevalence : NaN
##              Balanced Accuracy : NA
##
##      'Positive' Class : No
##

# AUC
library(pROC)
roc_obj <- roc(test_data$Death, pred_prob[, "Yes"]) # o [, "1"] se usi 0/1

## Setting levels: control = No, case = Yes

## Error in roc.default(test_data$Death, pred_prob[, "Yes"]): No control
## observation.

auc(roc_obj)

## Area under the curve: 0.6418

varImp(rf_model)

## rf variable importance
##
##      only 20 most important variables shown (out of 29)

```

```
##
## Overall
## IntervalOR 100.000
## IntervalOD 99.214
## LogCEA 88.674
## id 88.407
## Age 75.054
## Log2AT 68.955
## RBC 29.784
## Progress 22.081
## Progress_event 19.164
## AJCC_bin1 16.530
## ASA31 14.997
## Gender2 11.822
## DM 11.093
## Lymphovascularinvasion 10.148
## TumorLOC1 9.808
## perineural 9.678
## Liver_Only 9.314
## CKD 8.606
## RT 7.986
## EA 6.942
```

Si procede facendo tuning degli iperparametri, con mtry, numero di alberi etc. dopo il fit di base.

```
set.seed(123)
mtry_grid <- expand.grid(mtry = c(2, 4, 6, 8, 10))

rf_model_tuned <- train(
  Death ~ .,
  data = train_data,
  method = "rf",
  trControl = trainControl(
    method = "cv",
    number = 5,
    classProbs = TRUE,
    summaryFunction = twoClassSummary,
    verboseIter = TRUE
  ),
  metric = "ROC",
  tuneGrid = mtry_grid,
  ntree = 500
)

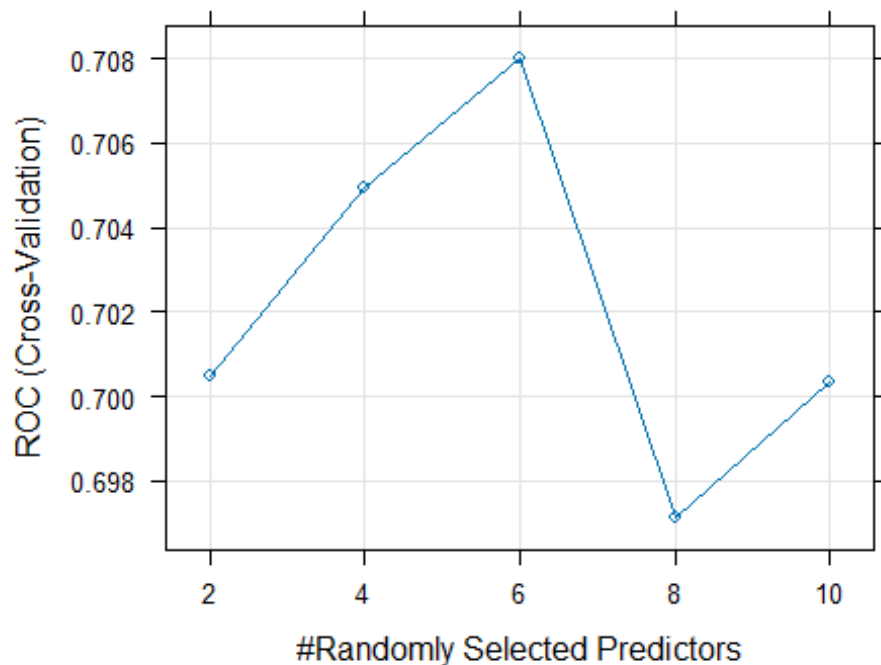
## Error in na.fail.default(structure(list(Death = structure(c(NA_integer_, :
missing values in object

# Visualizza la tabella con performance per ciascun mtry
print(rf_model_tuned)

## Random Forest
##
```

```
## 733 samples
## 29 predictor
## 2 classes: 'No', 'Yes'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 586, 586, 587, 587, 586
## Resampling results across tuning parameters:
##
## mtry ROC Sens Spec
## 2 0.7004742 0.7899885 0.4236047
## 4 0.7049238 0.7255881 0.5414747
## 6 0.7079994 0.7088640 0.5701485
## 8 0.6971334 0.7088067 0.5574501
## 10 0.7003616 0.7111876 0.5764465
##
## ROC was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 6.
```

```
plot(rf_model_tuned)
```



```
# Predizioni
pred_rf <- predict(rf_model_tuned, newdata = test_data)
probs_rf <- predict(rf_model_tuned, newdata = test_data, type = "prob")

# Confusion matrix
confusionMatrix(pred_rf, test_data$Death)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction No Yes
##           No    0    0
##           Yes   0    0
##
##           Accuracy : NaN
##           95% CI : (NA, NA)
##           No Information Rate : NA
##           P-Value [Acc > NIR] : NA
##
##           Kappa : NaN
##
## Mcnemar's Test P-Value : NA
##
##           Sensitivity : NA
##           Specificity : NA
##           Pos Pred Value : NA
##           Neg Pred Value : NA
##           Prevalence : NaN
##           Detection Rate : NaN
##           Detection Prevalence : NaN
##           Balanced Accuracy : NA
##
##           'Positive' Class : No
##
```

```
# AUC
```

```
library(pROC)
roc_obj <- roc(test_data$Death, probs_rf[, "Yes"])
```

```
## Setting levels: control = No, case = Yes
```

```
## Error in roc.default(test_data$Death, probs_rf[, "Yes"]): No control
observation.
```

```
auc(roc_obj)
```

```
## Area under the curve: 0.6418
```

```
varImp(rf_model_tuned)
```

```
## rf variable importance
```

```
##
```

```
## only 20 most important variables shown (out of 29)
```

```
##
```

```
##           Overall
## IntervalOD      100.000
## IntervalOR       92.153
## id              88.664
## LogCEA          87.326
## Age             73.749
```

```
## Log2AT          65.815
## RBC             21.316
## Progress_event  13.317
## AJCC_bin1       12.996
## Progress        10.728
## ASA31           10.635
## Gender2         10.185
## Liver_Only      9.065
## Lymphovascularinvasion  8.709
## DM              8.620
## TumorLOC1       8.598
## perineural      7.625
## NACTRT          6.439
## CKD             6.406
## RT              6.024
```

XGBoost

```
train_data$Death <- factor(train_data$Death, levels = c("No", "Yes"))
test_data$Death  <- factor(test_data$Death, levels = c("No", "Yes"))
```

```
set.seed(123)
```

```
xgb_grid <- expand.grid(
  nrounds = c(100, 200),
  max_depth = c(3, 6),
  eta = c(0.01, 0.1, 0.3),
  gamma = 0,
  colsample_bytree = 1,
  min_child_weight = 1,
  subsample = 1
)
```

```
xgb_model <- train(
  Death ~ .,
  data = train_data,
  method = "xgbTree",
  trControl = trainControl(
    method = "cv",
    number = 5,
    classProbs = TRUE,
    summaryFunction = twoClassSummary,
    verboseIter = TRUE
  ),
  metric = "ROC",
  tuneGrid = xgb_grid
)
```

```
## Error in na.fail.default(structure(list(Death = structure(c(NA_integer_, :
missing values in object
```

Predizioni

```
pred_xgb <- predict(xgb_model, newdata = test_data)
probs_xgb <- predict(xgb_model, newdata = test_data, type = "prob")
```

Confusion matrix

```
confusionMatrix(pred_xgb, test_data$Death)
```

```
## Confusion Matrix and Statistics
```

```
##
```

```
##           Reference
```

```
## Prediction No Yes
```

```
##           No    0    0
```

```
##           Yes   0    0
```

```
##
```

```
##           Accuracy : NaN
```

```
##           95% CI : (NA, NA)
```

```
## No Information Rate : NA
```

```
## P-Value [Acc > NIR] : NA
```

```
##
```

```
##           Kappa : NaN
```

```
##
```

```
## McNemar's Test P-Value : NA
```

```
##
```

```
##           Sensitivity : NA
```

```
##           Specificity : NA
```

```
##           Pos Pred Value : NA
```

```
##           Neg Pred Value : NA
```

```
##           Prevalence : NaN
```

```
##           Detection Rate : NaN
```

```
## Detection Prevalence : NaN
```

```
##           Balanced Accuracy : NA
```

```
##
```

```
##           'Positive' Class : No
```

```
##
```

AUC

```
library(pROC)
```

```
roc_xgb <- roc(test_data$Death, probs_xgb[, "Yes"])
```

```
## Setting levels: control = No, case = Yes
```

```
## Error in roc.default(test_data$Death, probs_xgb[, "Yes"]): No control observation.
```

```
auc(roc_xgb)
```

```
## Area under the curve: 0.6607
```

SVM - radial kernel

Carica la Libreria per SVM

```
library(e1071)
```

```

# SVM Model
svm_model <- svm(Death ~ ., data = train_data, kernel = "radial", probability =
TRUE)

## Error in if (any(co)) {: missing value where TRUE/FALSE needed

# Predizioni per SVM
pred_svm <- predict(svm_model, newdata = test_data, probability = TRUE)

## Error in matrix(ret$prob, nrow = nrow(newdata), byrow = TRUE, dimnames =
list(rowns, : length of 'dimnames' [2] not equal to array extent

probs_svm <- attr(pred_svm, "probabilities")

# SVM Metrics
confusionMatrix(pred_svm, test_data$Death)

## Confusion Matrix and Statistics
##
##              Reference
## Prediction No Yes
##      No      0      0
##      Yes      0      0
##
##              Accuracy : NaN
##              95% CI : (NA, NA)
##      No Information Rate : NA
##      P-Value [Acc > NIR] : NA
##
##              Kappa : NaN
##
##      McNemar's Test P-Value : NA
##
##              Sensitivity : NA
##              Specificity : NA
##      Pos Pred Value : NA
##      Neg Pred Value : NA
##              Prevalence : NaN
##      Detection Rate : NaN
##      Detection Prevalence : NaN
##      Balanced Accuracy : NA
##
##      'Positive' Class : No
##

roc(test_data$Death, probs_svm[, "Yes"])

## Setting levels: control = No, case = Yes

## Error in roc.default(test_data$Death, probs_svm[, "Yes"]): No control
observation.

```


KNN

```
# Carica La Libreria per KNN
library(caret)

# KNN Model
knn_model <- train(Death ~ ., data = train_data, method = "knn",
                  trControl = trainControl(method = "none"), tuneGrid =
expand.grid(k = 5))

## Error in na.fail.default(structure(list(Death = structure(c(NA_integer_, :
missing values in object

# Predizioni per KNN
pred_knn <- predict(knn_model, newdata = test_data)
probs_knn <- predict(knn_model, newdata = test_data, type = "prob")

# KNN Metrics
confusionMatrix(pred_knn, test_data$Death)

## Confusion Matrix and Statistics
##
##              Reference
## Prediction No  Yes
##      No      0    0
##      Yes     0    0
##
##              Accuracy : NaN
##              95% CI : (NA, NA)
##      No Information Rate : NA
##      P-Value [Acc > NIR] : NA
##
##              Kappa : NaN
##
##  Mcnemar's Test P-Value : NA
##
##      Sensitivity : NA
##      Specificity : NA
##      Pos Pred Value : NA
##      Neg Pred Value : NA
##      Prevalence : NaN
##      Detection Rate : NaN
##      Detection Prevalence : NaN
##      Balanced Accuracy : NA
##
##      'Positive' Class : No
##
roc(test_data$Death, probs_knn[, "Yes"])

## Setting levels: control = No, case = Yes
```

```
## Error in roc.default(test_data$Death, probs_knn[, "Yes"]): No control observation.
```

15 vicini non 5

```
# Carica La Libreria per KNN
```

```
library(caret)
```

```
# KNN Model
```

```
knn_model <- train(Death ~ ., data = train_data, method = "knn",  
                  trControl = trainControl(method = "none"), tuneGrid =  
expand.grid(k = 15))
```

```
## Error in na.fail.default(structure(list(Death = structure(c(NA_integer_, :  
missing values in object
```

```
# Predizioni per KNN
```

```
pred_knn <- predict(knn_model, newdata = test_data)  
probs_knn <- predict(knn_model, newdata = test_data, type = "prob")
```

```
# KNN Metrics
```

```
confusionMatrix(pred_knn, test_data$Death)
```

```
## Confusion Matrix and Statistics
```

```
##
```

```
##           Reference
```

```
## Prediction No Yes
```

```
##           No    0    0
```

```
##           Yes   0    0
```

```
##
```

```
##           Accuracy : NaN
```

```
##           95% CI : (NA, NA)
```

```
## No Information Rate : NA
```

```
## P-Value [Acc > NIR] : NA
```

```
##
```

```
##           Kappa : NaN
```

```
##
```

```
## McNemar's Test P-Value : NA
```

```
##
```

```
##           Sensitivity : NA
```

```
##           Specificity : NA
```

```
## Pos Pred Value : NA
```

```
## Neg Pred Value : NA
```

```
## Prevalence : NaN
```

```
## Detection Rate : NaN
```

```
## Detection Prevalence : NaN
```

```
## Balanced Accuracy : NA
```

```
##
```

```
## 'Positive' Class : No
```

```
##
```

```
roc(test_data$Death, probs_knn[, "Yes"])
```

```
## Setting levels: control = No, case = Yes
```

```
## Error in roc.default(test_data$Death, probs_knn[, "Yes"]): No control  
observation.
```