formR.R

Mr ELYSEE

2025-06-18

options(repos = c(CRAN = "https://cloud.r-project.org/"))  
install.packages("tidyverse")

## le package 'tidyverse' a été décompressé et les sommes MD5 ont été vérifiées avec succés  
##   
## Les packages binaires téléchargés sont dans  
## C:\Users\Mr ELYSEE\AppData\Local\Temp\RtmpCyGE8P\downloaded\_packages

install.packages("ggplot2")

## le package 'ggplot2' a été décompressé et les sommes MD5 ont été vérifiées avec succés  
##   
## Les packages binaires téléchargés sont dans  
## C:\Users\Mr ELYSEE\AppData\Local\Temp\RtmpCyGE8P\downloaded\_packages

install.packages("labelled")

## le package 'labelled' a été décompressé et les sommes MD5 ont été vérifiées avec succés  
##   
## Les packages binaires téléchargés sont dans  
## C:\Users\Mr ELYSEE\AppData\Local\Temp\RtmpCyGE8P\downloaded\_packages

install.packages("GGally")

## le package 'GGally' a été décompressé et les sommes MD5 ont été vérifiées avec succés  
##   
## Les packages binaires téléchargés sont dans  
## C:\Users\Mr ELYSEE\AppData\Local\Temp\RtmpCyGE8P\downloaded\_packages

install.packages("lubridate")

## le package 'lubridate' a été décompressé et les sommes MD5 ont été vérifiées avec succés  
##   
## Les packages binaires téléchargés sont dans  
## C:\Users\Mr ELYSEE\AppData\Local\Temp\RtmpCyGE8P\downloaded\_packages

install.packages("forcats")

## le package 'forcats' a été décompressé et les sommes MD5 ont été vérifiées avec succés  
##   
## Les packages binaires téléchargés sont dans  
## C:\Users\Mr ELYSEE\AppData\Local\Temp\RtmpCyGE8P\downloaded\_packages

install.packages("gtsummary")

## installation de la dépendance 'cli'

## le package 'cli' a été décompressé et les sommes MD5 ont été vérifiées avec succés

## Warning: impossible de supprimer l'installation précédente du package 'cli'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problème lors de la  
## copie de D:\R\R-4.4.1\library\00LOCK\cli\libs\x64\cli.dll vers  
## D:\R\R-4.4.1\library\cli\libs\x64\cli.dll : Permission denied

## Warning: 'cli' restauré

## le package 'gtsummary' a été décompressé et les sommes MD5 ont été vérifiées avec succés  
##   
## Les packages binaires téléchargés sont dans  
## C:\Users\Mr ELYSEE\AppData\Local\Temp\RtmpCyGE8P\downloaded\_packages

library(tidyverse)

## Warning: le package 'tidyverse' a été compilé avec la version R 4.4.3

## Warning: le package 'ggplot2' a été compilé avec la version R 4.4.3

## Warning: le package 'purrr' a été compilé avec la version R 4.4.3

## Warning: le package 'forcats' a été compilé avec la version R 4.4.3

## Warning: le package 'lubridate' a été compilé avec la version R 4.4.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.4

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(car)

## Warning: le package 'car' a été compilé avec la version R 4.4.3

## Le chargement a nécessité le package : carData

## Warning: le package 'carData' a été compilé avec la version R 4.4.2

##   
## Attachement du package : 'car'  
##   
## L'objet suivant est masqué depuis 'package:dplyr':  
##   
## recode  
##   
## L'objet suivant est masqué depuis 'package:purrr':  
##   
## some

library(cardx)

## Warning: le package 'cardx' a été compilé avec la version R 4.4.3

library(ggplot2)  
library(labelled)

## Warning: le package 'labelled' a été compilé avec la version R 4.4.3

library(gtsummary)

## Warning: le package 'gtsummary' a été compilé avec la version R 4.4.3

library(ggstatsplot)

## Warning: le package 'ggstatsplot' a été compilé avec la version R 4.4.3

## You can cite this package as:  
## Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach.  
## Journal of Open Source Software, 6(61), 3167, doi:10.21105/joss.03167

library(Hmisc)

## Warning: le package 'Hmisc' a été compilé avec la version R 4.4.3

##   
## Attachement du package : 'Hmisc'  
##   
## Les objets suivants sont masqués depuis 'package:dplyr':  
##   
## src, summarize  
##   
## Les objets suivants sont masqués depuis 'package:base':  
##   
## format.pval, units

data("trial")  
theme\_gtsummary\_language("fr",decimal.mark = ",", big.mark = " ")

## Setting theme "language: fr"

View(trial)  
trial%>%  
 look\_for()

## pos variable label col\_type missing values  
## 1 trt Chemotherapy Treatment chr 0   
## 2 age Age dbl 11   
## 3 marker Marker Level (ng/mL) dbl 10   
## 4 stage T Stage fct 0 T1   
## T2   
## T3   
## T4   
## 5 grade Grade fct 0 I   
## II   
## III   
## 6 response Tumor Response int 7   
## 7 death Patient Died int 0   
## 8 ttdeath Months to Death/Censor dbl 0

trial|>nrow()

## [1] 200

str(trial)

## tibble [200 × 8] (S3: tbl\_df/tbl/data.frame)  
## $ trt : chr [1:200] "Drug A" "Drug B" "Drug A" "Drug A" ...  
## ..- attr(\*, "label")= chr "Chemotherapy Treatment"  
## $ age : num [1:200] 23 9 31 NA 51 39 37 32 31 34 ...  
## ..- attr(\*, "label")= chr "Age"  
## $ marker : num [1:200] 0.16 1.107 0.277 2.067 2.767 ...  
## ..- attr(\*, "label")= chr "Marker Level (ng/mL)"  
## $ stage : Factor w/ 4 levels "T1","T2","T3",..: 1 2 1 3 4 4 1 1 1 3 ...  
## ..- attr(\*, "label")= chr "T Stage"  
## $ grade : Factor w/ 3 levels "I","II","III": 2 1 2 3 3 1 2 1 2 1 ...  
## ..- attr(\*, "label")= chr "Grade"  
## $ response: int [1:200] 0 1 0 1 1 0 0 0 0 0 ...  
## ..- attr(\*, "label")= chr "Tumor Response"  
## $ death : int [1:200] 0 0 0 1 1 1 0 1 0 1 ...  
## ..- attr(\*, "label")= chr "Patient Died"  
## $ ttdeath : num [1:200] 24 24 24 17.6 16.4 ...  
## ..- attr(\*, "label")= chr "Months to Death/Censor"

#Recodage des variables et labellisation  
  
trial <- trial %>%  
 mutate(  
 trt = as.character(trt),  
 stage = as.character(stage),  
 grade = as.character(grade),  
 response=as.character(response),  
 death=as.character(death)  
 ) %>%  
 set\_variable\_labels(  
 trt = "Traitement attribué",  
 age = "Age du patient",  
 stage = "Stade de la maladie",  
 grade = "Gravité de la maladie",  
 response="Réponse de la tumeur",  
 death="En vie ou mort",  
 ttdeath="Total décès",  
 marker="Marqueur sanguin"  
 ) %>%  
 set\_value\_labels(  
 trt = c("A" = "Drug A", "B" = "Drug B"),  
 stage = c("0" = "T1", "1" = "T2", "2" = "T3", "3" = "T4"),  
 grade = c("Gravité1" = "I", "Gravité2" = "II", "Gravité3" = "III")  
 ) %>%  
 mutate(  
 trt = labelled::to\_factor(trt),  
 stage = labelled::to\_factor(stage),  
 grade = labelled::to\_factor(grade)  
 )  
  
# Forcer la mise à jour des étiquettes (facultatif)  
force(trial)

## # A tibble: 200 × 8  
## trt age marker stage grade response death ttdeath  
## <fct> <dbl> <dbl> <fct> <fct> <chr> <chr> <dbl>  
## 1 A 23 0.16 0 Gravité2 0 0 24   
## 2 B 9 1.11 1 Gravité1 1 0 24   
## 3 A 31 0.277 0 Gravité2 0 0 24   
## 4 A NA 2.07 2 Gravité3 1 1 17.6  
## 5 A 51 2.77 3 Gravité3 1 1 16.4  
## 6 B 39 0.613 3 Gravité1 0 1 15.6  
## 7 A 37 0.354 0 Gravité2 0 0 24   
## 8 A 32 1.74 0 Gravité1 0 1 18.4  
## 9 A 31 0.144 0 Gravité2 0 0 24   
## 10 B 34 0.205 2 Gravité1 0 1 10.5  
## # ℹ 190 more rows

trial %>%  
 tbl\_summary()%>%  
 modify\_header(label = "\*\*Variables\*\*")

| **Variables** | **N = 200***1* |
| --- | --- |
| Traitement attribué |  |
| A | 98 (49%) |
| B | 102 (51%) |
| Age du patient | 47 (38 – 57) |
| Manquant | 11 |
| Marqueur sanguin | 0,64 (0,22 – 1,41) |
| Manquant | 10 |
| Stade de la maladie |  |
| 0 | 53 (27%) |
| 1 | 54 (27%) |
| 2 | 43 (22%) |
| 3 | 50 (25%) |
| Gravité de la maladie |  |
| Gravité1 | 68 (34%) |
| Gravité2 | 68 (34%) |
| Gravité3 | 64 (32%) |
| Réponse de la tumeur |  |
| 0 | 132 (68%) |
| 1 | 61 (32%) |
| Manquant | 7 |
| En vie ou mort |  |
| 0 | 88 (44%) |
| 1 | 112 (56%) |
| Total décès | 22,4 (15,9 – 24,0) |
| *1*n (%); Médiane (Q1 – Q3) | |

trial %>%  
 tbl\_summary(by="trt")%>%  
 modify\_header(label = "\*\*Variables\*\*")

| **Variables** | **A** N = 98*1* | **B** N = 102*1* |
| --- | --- | --- |
| Age du patient | 46 (37 – 60) | 48 (39 – 56) |
| Manquant | 7 | 4 |
| Marqueur sanguin | 0,84 (0,23 – 1,60) | 0,52 (0,18 – 1,21) |
| Manquant | 6 | 4 |
| Stade de la maladie |  |  |
| 0 | 28 (29%) | 25 (25%) |
| 1 | 25 (26%) | 29 (28%) |
| 2 | 22 (22%) | 21 (21%) |
| 3 | 23 (23%) | 27 (26%) |
| Gravité de la maladie |  |  |
| Gravité1 | 35 (36%) | 33 (32%) |
| Gravité2 | 32 (33%) | 36 (35%) |
| Gravité3 | 31 (32%) | 33 (32%) |
| Réponse de la tumeur |  |  |
| 0 | 67 (71%) | 65 (66%) |
| 1 | 28 (29%) | 33 (34%) |
| Manquant | 3 | 4 |
| En vie ou mort |  |  |
| 0 | 46 (47%) | 42 (41%) |
| 1 | 52 (53%) | 60 (59%) |
| Total décès | 23,5 (17,4 – 24,0) | 21,2 (14,5 – 24,0) |
| *1*Médiane (Q1 – Q3); n (%) | | |

library(GGally)

## Warning: le package 'GGally' a été compilé avec la version R 4.4.3

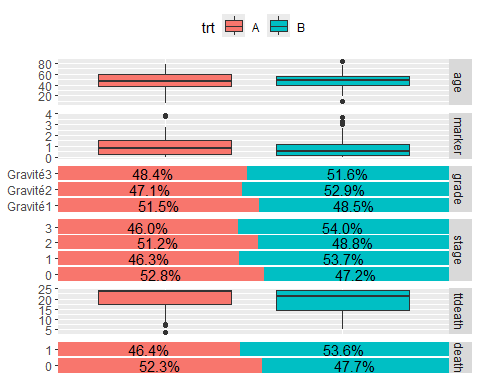
## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

trial|>ggbivariate(outcome = "trt", explanatory = c("age", "marker","grade", "stage","ttdeath","death"))

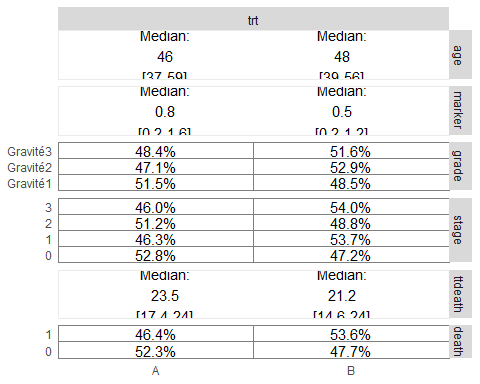
## Warning: Removed 11 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

## Warning: Removed 11 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

## Warning: Removed 10 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).



trial|>ggtable(columnsY = c("age", "marker","grade", "stage","ttdeath","death"), columnsX = "trt",  
 fill="std.resid",  
 cells="row.prop",  
 )

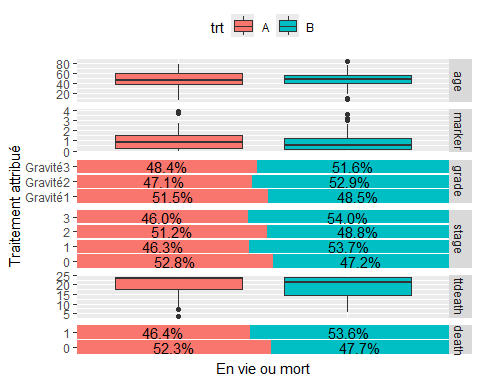


#library(tidyverse)  
#library(labelled)  
  
#data("trial")  
  
#trial\_labeled <- trial %>%  
# set\_variable\_labels(  
# trt = "Traitement attribué",  
# age = "Age du patient",  
# marker = "Marqueur sanguin",  
# grade = "Gravité de la maladie",  
# stage = "Stade de la maladie",  
# ttdeath = "Temps jusqu'au décès",  
# death = "Décès"  
# )  
  
apply\_labels\_ggbivariate <- function(p, data, outcome, explanatory) {  
 outcome\_label <- var\_label(data[[outcome]])  
 for (var in explanatory) {  
 var\_label\_text <- var\_label(data[[var]])  
 if (!is.null(var\_label\_text)) {  
 p <- p + labs(x = var\_label\_text, y = outcome\_label)  
 }  
 }  
 return(p)  
}  
  
plot <- trial %>%  
 ggbivariate(  
 outcome = "trt",  
 explanatory = c("age", "marker", "grade", "stage", "ttdeath", "death")  
 )  
  
plot <- apply\_labels\_ggbivariate(plot, trial, "trt", c("age", "marker", "grade", "stage", "ttdeath", "death"))  
  
print(plot)

## Warning: Removed 11 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

## Warning: Removed 11 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

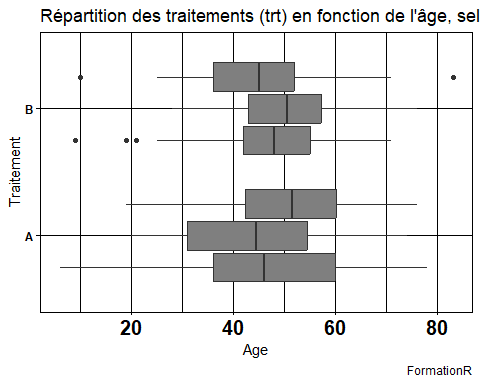
## Warning: Removed 10 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).



library(ggplot2)  
  
ggplot(trial) +  
 aes(x = age, y = trt, fill = grade) +  
 geom\_boxplot() +  
 scale\_fill\_manual(values = c(I = "#F8766D",   
 II = "#00C19F", III = "#FF61C3")) +  
 labs(x = "Age", y = "Traitement", title = "Répartition des traitements (trt) en fonction de l'âge, selon la gravité de la maladie (grade)",   
 caption = "FormationR", fill = "Gravité de la tumeur") +  
 theme\_linedraw() +  
 theme(axis.text.y = element\_text(face = "bold"),   
 axis.text.x = element\_text(face = "bold", size = 15L), legend.text = element\_text(face = "bold"),   
 legend.title = element\_text(face = "bold"))

## Warning: Removed 11 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

## Warning: No shared levels found between `names(values)` of the manual scale and the  
## data's fill values.  
## No shared levels found between `names(values)` of the manual scale and the  
## data's fill values.



#Modèle de regression linéaire simple et multiple.  
  
  
# Préparation des données : sélection des variables et suppression des valeurs manquantes  
trial\_regression <- trial %>%  
 select(age, marker,ttdeath) %>%  
 drop\_na()  
  
# Création du modèle  
modele\_regression <- lm(ttdeath ~ marker+age, data = trial\_regression)  
  
# Affichage du résumé du modèle  
summary(modele\_regression)

##   
## Call:  
## lm(formula = ttdeath ~ marker + age, data = trial\_regression)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.937 -3.534 3.144 4.007 4.594   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 20.64546 1.38933 14.860 <2e-16 \*\*\*  
## marker 0.24828 0.45000 0.552 0.582   
## age -0.01925 0.02694 -0.715 0.476   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.199 on 176 degrees of freedom  
## Multiple R-squared: 0.004624, Adjusted R-squared: -0.006687   
## F-statistic: 0.4088 on 2 and 176 DF, p-value: 0.6651

library(tidyverse)  
library(gtsummary)  
  
data("trial")  
  
# Conversion des variables catégorielles en facteurs  
trial$stage <- as.factor(trial$stage)  
trial$grade <- as.factor(trial$grade)  
trial$trt <- as.factor(trial$trt)  
trial$response <- as.factor(trial$response)  
  
# Gestion des valeurs manquantes  
trial <- trial %>% drop\_na()  
  
# Vérification des types de variables  
str(trial)

## tibble [173 × 8] (S3: tbl\_df/tbl/data.frame)  
## $ trt : Factor w/ 2 levels "Drug A","Drug B": 1 2 1 1 2 1 1 1 2 2 ...  
## $ age : num [1:173] 23 9 31 51 39 37 32 31 34 42 ...  
## ..- attr(\*, "label")= chr "Age"  
## $ marker : num [1:173] 0.16 1.107 0.277 2.767 0.613 ...  
## ..- attr(\*, "label")= chr "Marker Level (ng/mL)"  
## $ stage : Factor w/ 4 levels "T1","T2","T3",..: 1 2 1 4 4 1 1 1 3 1 ...  
## ..- attr(\*, "label")= chr "T Stage"  
## $ grade : Factor w/ 3 levels "I","II","III": 2 1 2 3 1 2 1 2 1 3 ...  
## ..- attr(\*, "label")= chr "Grade"  
## $ response: Factor w/ 2 levels "0","1": 1 2 1 2 1 1 1 1 1 1 ...  
## $ death : int [1:173] 0 0 0 1 1 0 1 0 1 0 ...  
## ..- attr(\*, "label")= chr "Patient Died"  
## $ ttdeath : num [1:173] 24 24 24 16.4 15.6 ...  
## ..- attr(\*, "label")= chr "Months to Death/Censor"

# Modèle de régression logistique  
modele\_response <- glm(response ~ age + marker + stage + grade + trt, data = trial, family = "binomial")  
  
# Affichage du résumé du modèle  
summary(modele\_response)

##   
## Call:  
## glm(formula = response ~ age + marker + stage + grade + trt,   
## family = "binomial", data = trial)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.85393 0.73976 -2.506 0.0122 \*  
## age 0.01901 0.01193 1.593 0.1112   
## marker 0.34596 0.19688 1.757 0.0789 .  
## stageT2 -0.80936 0.47127 -1.717 0.0859 .  
## stageT3 -0.14589 0.48377 -0.302 0.7630   
## stageT4 -0.43985 0.47316 -0.930 0.3526   
## gradeII 0.03926 0.42852 0.092 0.9270   
## gradeIII 0.03984 0.40810 0.098 0.9222   
## trtDrug B 0.29470 0.34084 0.865 0.3872   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 214.80 on 172 degrees of freedom  
## Residual deviance: 205.89 on 164 degrees of freedom  
## AIC: 223.89  
##   
## Number of Fisher Scoring iterations: 4

installed.packages("broom.helpers")

## Package LibPath Version Priority Depends Imports LinkingTo Suggests  
## Enhances License License\_is\_FOSS License\_restricts\_use OS\_type Archs  
## MD5sum NeedsCompilation Built

library(broom.helpers)

## Warning: le package 'broom.helpers' a été compilé avec la version R 4.4.3

##   
## Attachement du package : 'broom.helpers'  
##   
## Les objets suivants sont masqués depuis 'package:gtsummary':  
##   
## all\_categorical, all\_continuous, all\_contrasts, all\_dichotomous,  
## all\_interaction, all\_intercepts

modele\_response%>%  
 tbl\_regression(  
 exponentiate=TRUE, add\_estimate\_to\_reference\_rows = TRUE  
 )%>%  
 add\_global\_p()

| **Caractéristique** | **OR** | **95% IC** | **p-valeur** |
| --- | --- | --- | --- |
| Age | 1,02 | 1,00 – 1,04 | 0,11 |
| Marker Level (ng/mL) | 1,41 | 0,96 – 2,09 | 0,080 |
| T Stage |  |  | 0,3 |
| T1 | 1,00 | — |  |
| T2 | 0,45 | 0,17 – 1,11 |  |
| T3 | 0,86 | 0,33 – 2,22 |  |
| T4 | 0,64 | 0,25 – 1,62 |  |
| Grade |  |  | >0,9 |
| I | 1,00 | — |  |
| II | 1,04 | 0,45 – 2,42 |  |
| III | 1,04 | 0,47 – 2,32 |  |
| trt |  |  | 0,4 |
| Drug A | 1,00 | — |  |
| Drug B | 1,34 | 0,69 – 2,64 |  |
| Abréviations: IC = intervalle de confiance, OR = rapport de cotes | | | |

library(tidyverse)  
library(gtsummary)  
  
data("trial")  
  
# Préparation des données : sélection des variables et conversion en facteurs  
trial\_anova <- trial %>%  
 select(age, trt, grade) %>%  
 mutate(  
 trt = as.factor(trt),  
 grade = as.factor(grade)  
 ) %>%  
 drop\_na()  
  
  
# ANOVA à un facteur : age ~ trt  
modele\_anova\_1 <- aov(age ~ trt, data = trial\_anova)  
  
# Affichage du résumé du modèle  
summary(modele\_anova\_1)

## Df Sum Sq Mean Sq F value Pr(>F)  
## trt 1 9 9.05 0.044 0.834  
## Residuals 187 38499 205.88

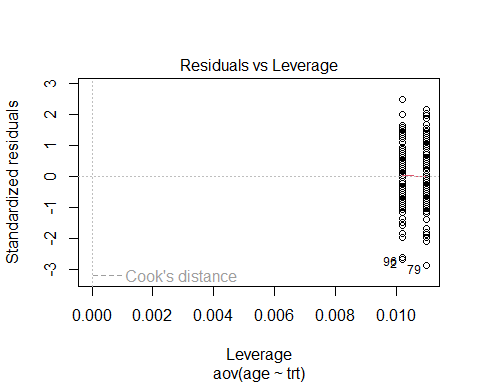
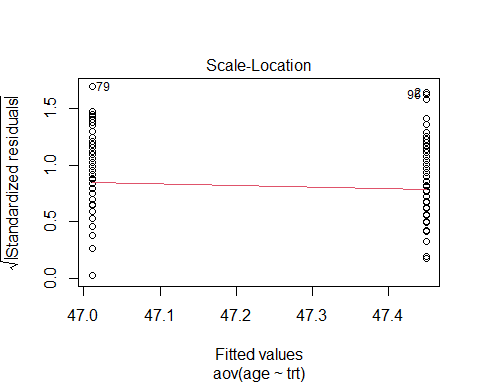
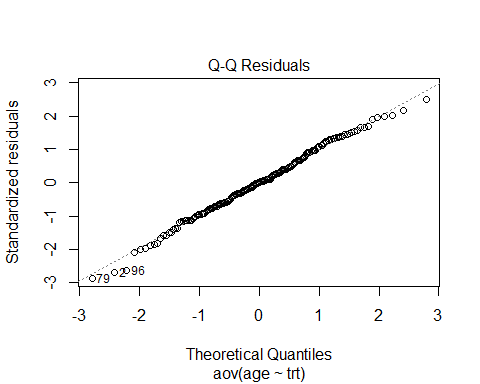
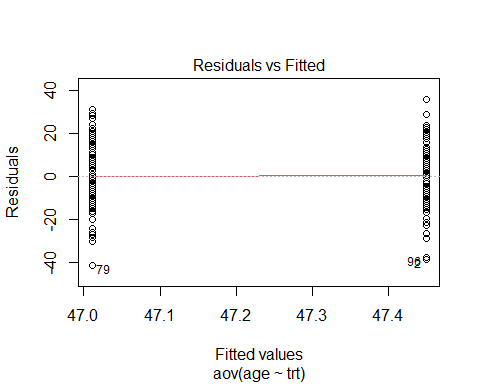
# ANOVA à deux facteurs : age ~ trt + grade  
modele\_anova\_2 <- aov(age ~ trt + grade, data = trial\_anova)  
  
# Affichage du résumé du modèle  
summary(modele\_anova\_2)

## Df Sum Sq Mean Sq F value Pr(>F)  
## trt 1 9 9.05 0.044 0.835  
## grade 2 128 64.07 0.309 0.735  
## Residuals 185 38371 207.41

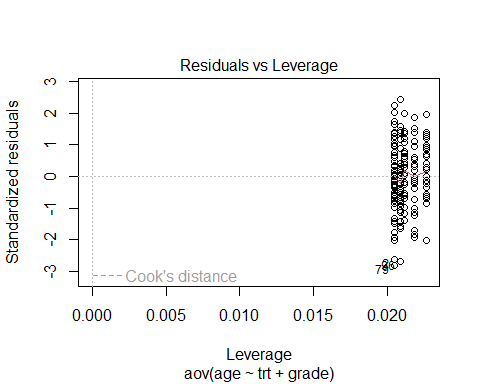
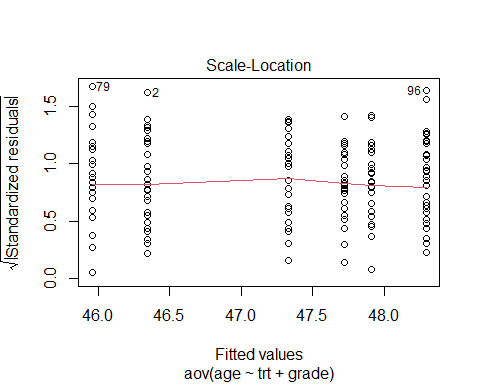
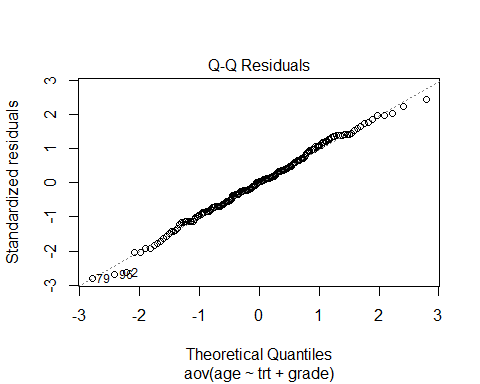
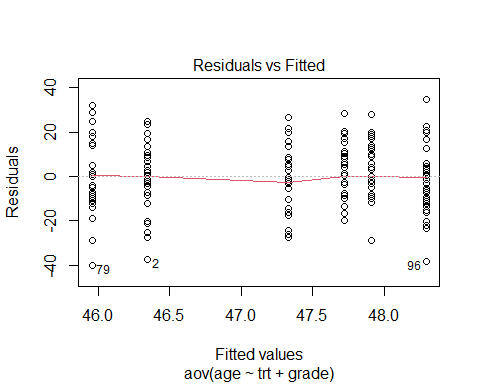
# ANOVA à deux facteurs avec interaction : age ~ trt \* grade  
modele\_anova\_inter <- aov(age ~ trt \* grade, data = trial\_anova)  
  
# Affichage du résumé du modèle  
summary(modele\_anova\_inter)

## Df Sum Sq Mean Sq F value Pr(>F)  
## trt 1 9 9.1 0.044 0.834  
## grade 2 128 64.1 0.313 0.732  
## trt:grade 2 929 464.7 2.271 0.106  
## Residuals 183 37442 204.6

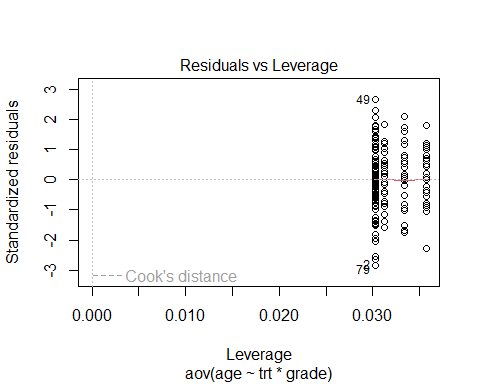
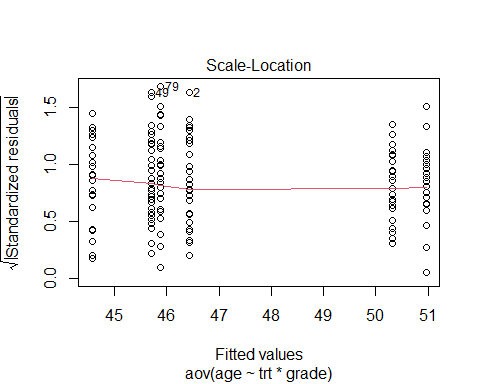
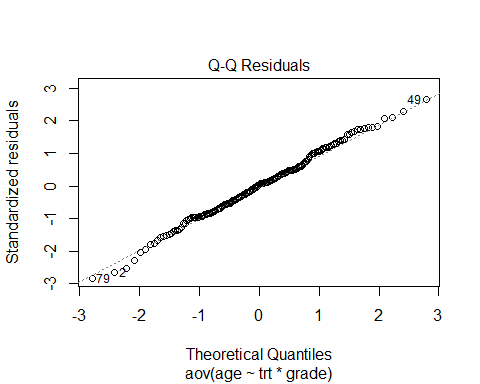
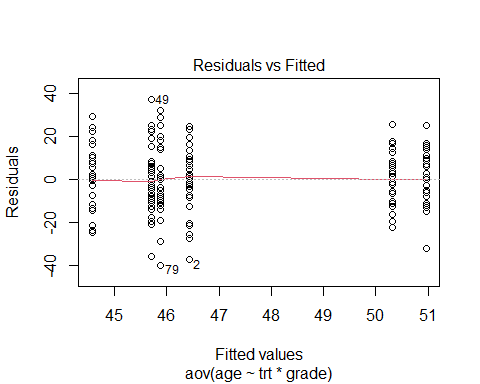
# Graphiques de diagnostic  
plot(modele\_anova\_1)



plot(modele\_anova\_2)



plot(modele\_anova\_inter)



# Test de Levene pour l'homogénéité des variances  
library(car)  
leveneTest(age ~ trt, data = trial\_anova)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 1 0.7053 0.4021  
## 187

leveneTest(age ~ trt \* grade, data = trial\_anova)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 5 0.4313 0.8264  
## 183

library(tidyverse)  
library(survival)  
library(survminer)

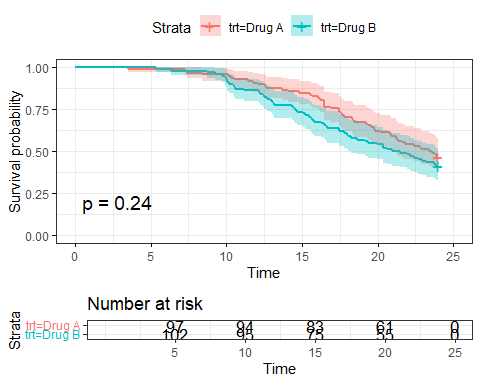
## Warning: le package 'survminer' a été compilé avec la version R 4.4.2

## Le chargement a nécessité le package : ggpubr

## Warning: le package 'ggpubr' a été compilé avec la version R 4.4.3

##   
## Attachement du package : 'survminer'  
##   
## L'objet suivant est masqué depuis 'package:survival':  
##   
## myeloma

library(gtsummary)  
  
data("trial")  
  
# Préparation des données : sélection des variables et conversion en facteurs  
trial\_survie <- trial %>%  
 select(ttdeath, death, trt) %>%  
 mutate(  
 trt = as.factor(trt)  
 ) %>%  
 drop\_na()  
  
  
# Création de l'objet de survie  
objet\_survie <- Surv(time = trial\_survie$ttdeath, event = trial\_survie$death)  
  
  
# Estimation des courbes de survie  
modele\_survie <- survfit(objet\_survie ~ trt, data = trial\_survie)  
  
  
# Visualisation des courbes de survie  
ggsurvplot(modele\_survie, data = trial\_survie,  
 pval = TRUE, conf.int = TRUE,  
 risk.table = TRUE,  
 ggtheme = theme\_bw())



# Test du log-rank  
test\_logrank <- survdiff(objet\_survie ~ trt, data = trial\_survie)  
test\_logrank

## Call:  
## survdiff(formula = objet\_survie ~ trt, data = trial\_survie)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## trt=Drug A 98 52 58.2 0.665 1.39  
## trt=Drug B 102 60 53.8 0.720 1.39  
##   
## Chisq= 1.4 on 1 degrees of freedom, p= 0.2

library(tidyverse)  
library(caret)

## Warning: le package 'caret' a été compilé avec la version R 4.4.3

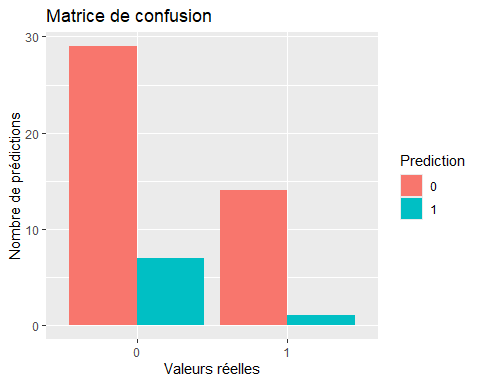
## Le chargement a nécessité le package : lattice  
##   
## Attachement du package : 'caret'  
##   
## L'objet suivant est masqué depuis 'package:survival':  
##   
## cluster  
##   
## L'objet suivant est masqué depuis 'package:purrr':  
##   
## lift

library(gtsummary)  
library(ggplot2)  
library(pROC)

## Warning: le package 'pROC' a été compilé avec la version R 4.4.3

## Type 'citation("pROC")' for a citation.  
##   
## Attachement du package : 'pROC'  
##   
## Les objets suivants sont masqués depuis 'package:stats':  
##   
## cov, smooth, var

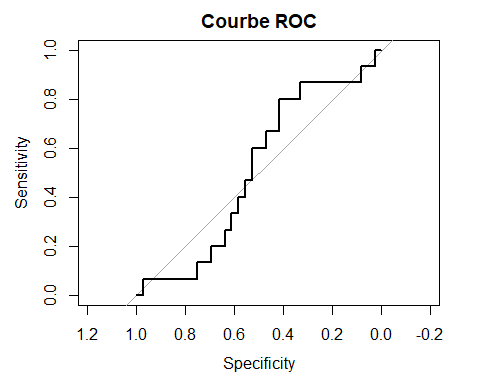
data("trial")  
  
# Préparation des données (comme précédemment)  
trial\_ml <- trial %>%  
 select(age, marker, stage, grade, trt, response) %>%  
 mutate\_if(is.character, as.factor) %>%  
 drop\_na()  
  
set.seed(123)  
trainIndex <- createDataPartition(trial\_ml$response, p = 0.7, list = FALSE)  
train\_data <- trial\_ml[trainIndex, ]  
test\_data <- trial\_ml[-trainIndex, ]  
  
# Modèle de régression logistique  
modele\_logistique <- glm(response ~ ., data = train\_data, family = "binomial")  
  
# Prédictions sur l'ensemble de test  
predictions\_logistique\_prob <- predict(modele\_logistique, newdata = test\_data, type = "response")  
predictions\_logistique <- ifelse(predictions\_logistique\_prob > 0.5, "1", "0")  
  
# Convertir les prédictions et les valeurs réelles en facteurs  
predictions\_logistique <- as.factor(predictions\_logistique)  
test\_data$response <- as.factor(test\_data$response)  
  
# Trouver et appliquer les mêmes niveaux  
unique\_levels <- unique(c(levels(predictions\_logistique), levels(test\_data$response)))  
  
predictions\_logistique <- factor(predictions\_logistique, levels = unique\_levels)  
test\_data$response <- factor(test\_data$response, levels = unique\_levels)  
  
# Matrice de confusion  
cm <- confusionMatrix(predictions\_logistique, test\_data$response)  
  
# Visualisation de la matrice de confusion (exemple avec ggplot2)  
cm\_data <- data.frame(  
 Prediction = predictions\_logistique,  
 Reference = test\_data$response  
)  
  
ggplot(cm\_data, aes(x = Reference, fill = Prediction)) +  
 geom\_bar(position = "dodge") +  
 labs(title = "Matrice de confusion", x = "Valeurs réelles", y = "Nombre de prédictions")



# Courbe ROC  
roc\_obj <- roc(test\_data$response, predictions\_logistique\_prob)

## Setting levels: control = 0, case = 1  
## Setting direction: controls < cases

plot(roc\_obj, main = "Courbe ROC")



# Importance des variables (exemple avec les forêts aléatoires)  
modele\_foret <- train(response ~ ., data = train\_data, method = "rf")

## Warning in train.default(x, y, weights = w, ...): You are trying to do  
## regression and your outcome only has two possible values Are you trying to do  
## classification? If so, use a 2 level factor as your outcome column.

## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?  
## Warning in randomForest.default(x, y, mtry = param$mtry, ...): The response has  
## five or fewer unique values. Are you sure you want to do regression?

importance <- varImp(modele\_foret)  
plot(importance, main = "Importance des variables")

