

survie

2025-01-14

```
options(repos = c(CRAN = "https://cran.rstudio.com"))
```

```
install.packages(c("KMsurv", "dplyr", "ggplot2", "survival", "survminer", "corrplot"))
```

```
## Installation des packages dans 'C:/Users/icoco/AppData/Local/R/win-library/4.4'
## (car 'lib' n'est pas spécifié)

## le package 'KMsurv' a été décompressé et les sommes MD5 ont été vérifiées avec succès
## le package 'dplyr' a été décompressé et les sommes MD5 ont été vérifiées avec succès

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

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problème lors de la
## copie de
## C:\Users\icoco\AppData\Local\R\win-library\4.4\00LOCK\dplyr\libs\x64\dplyr.dll
## vers C:\Users\icoco\AppData\Local\R\win-library\4.4\dplyr\libs\x64\dplyr.dll :
## Permission denied

## Warning: 'dplyr' restauré

## le package 'ggplot2' a été décompressé et les sommes MD5 ont été vérifiées avec succès
## le package 'survival' a été décompressé et les sommes MD5 ont été vérifiées avec succès

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

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problème lors de la
## copie de
## C:\Users\icoco\AppData\Local\R\win-library\4.4\00LOCK\survival\libs\x64\survival.dll
## vers
## C:\Users\icoco\AppData\Local\R\win-library\4.4\survival\libs\x64\survival.dll :
## Permission denied

## Warning: 'survival' restauré

## le package 'survminer' a été décompressé et les sommes MD5 ont été vérifiées avec succès
## le package 'corrplot' a été décompressé et les sommes MD5 ont été vérifiées avec succès
##
## Les packages binaires téléchargés sont dans
## C:\Users\icoco\AppData\Local\Temp\RtmpElPnMo\downloaded_packages
```

```
library(dplyr)
```

```
##
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:stats':
##
##     filter, lag

## Les objets suivants sont masqués depuis 'package:base':
##
```

```

##      intersect, setdiff, setequal, union
library(ggplot2)
library(survival)
library(survminer)

## Le chargement a nécessité le package : ggpubr
##
## Attachement du package : 'survminer'
## L'objet suivant est masqué depuis 'package:survival':
##
##      myeloma
library(corrplot)

## corrplot 0.95 loaded
library(KMsurv)

#Présentation des données
data(pneumon)
?pneumon

## démarrage du serveur d'aide httpd ... fini
head(pneumon)

##      chldage hospital mthage urban alcohol smoke region poverty bweight race
## 1      12        0      22     1        0    0      1        1        1    1
## 2      12        0      20     1        1    0      1        1        0    1
## 3       3        0      24     1        3    0      1        1        0    1
## 4       2        0      22     1        2    2      1        1        0    1
## 5       4        0      21     1        1    2      1        1        1    1
## 6      12        0      20     1        0    0      1        1        0    1
##      education nsibs wmonth sfmonth agepn
## 1          10     1      1        1      1
## 2          12     1      2        2     12
## 3          12     2      1        0      3
## 4           9     0      0        0      2
## 5          12     0      0        0      4
## 6          12     0      0        0     12
names(pneumon)

## [1] "chldage" "hospital" "mthage" "urban" "alcohol" "smoke"
## [7] "region" "poverty" "bweight" "race" "education" "nsibs"
## [13] "wmonth" "sfmonth" "agepn"
pneumon = pneumon %>% dplyr::select(-agepn)

pneumon = pneumon %>% mutate(urban = as.factor(urban) ,
                             alcohol = as.factor(alcohol),
                             smoke = as.factor(smoke) ,
                             region = as.factor(region) ,
                             poverty = as.factor(poverty) ,
                             race = as.factor(race))

```

```
)
glimpse(pneumon)

## Rows: 3,470
## Columns: 14
## $ chldage <dbl> 12.0, 12.0, 3.0, 2.0, 4.0, 12.0, 7.0, 3.0, 7.0, 12.0, 12.0, ~
## $ hospital <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ mthage <int> 22, 20, 24, 22, 21, 20, 24, 24, 26, 21, 24, 27, 20, 25, 19, ~
## $ urban <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, ~
## $ alcohol <fct> 0, 1, 3, 2, 1, 0, 0, 3, 2, 1, 0, 0, 0, 2, 0, 4, 0, 2, 1, 1, ~
## $ smoke <fct> 0, 0, 0, 2, 2, 0, 0, 0, 2, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, ~
## $ region <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ poverty <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ bweight <int> 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, ~
## $ race <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ education <int> 10, 12, 12, 9, 12, 12, 12, 14, 12, 12, 16, 16, 12, 12, 11, 9~
## $ nsibs <int> 1, 1, 2, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, ~
## $ wmonth <int> 1, 2, 1, 0, 0, 0, 0, 4, 1, 3, 0, 0, 0, 4, 0, 1, 9, 10, 9, 4, ~
## $ sfmonth <int> 1, 2, 0, 0, 0, 0, 0, 2, 1, 2, 0, 0, 0, 2, 0, 1, 2, 3, 9, 3, ~
```

```
colSums(is.na(pneumon))
```

```
##   chldage  hospital   mthage   urban  alcohol   smoke   region  poverty
##         0         0         0         0         0         0         0         0
##   bweight      race education   nsibs   wmonth   sfmonth
##         0         0         0         0         0         0
```

Aucune valeur n'est manquante dans notre dataset

```
# Identification des doublons
#duplicated_rows <- pneumon[duplicated(pneumon), ]
#print(duplicated_rows)
```

```
#Suppression des doublons
#pneumon <- pneumon[!duplicated(pneumon), ]
```

```
#Identification des valeurs aberrantes
#boxplot(pneumon$mthage, main = "Age de la mère", ylab = "Années")
#boxplot(pneumon$bweight, main = "Poids de naissance", ylab = "lbs")
```

#Représentation de la fonction de survie

```
table(pneumon$hospital)
```

```
##
##    0    1
## 3397   73
```

```
table(pneumon$chldage[pneumon$hospital == 1])
```

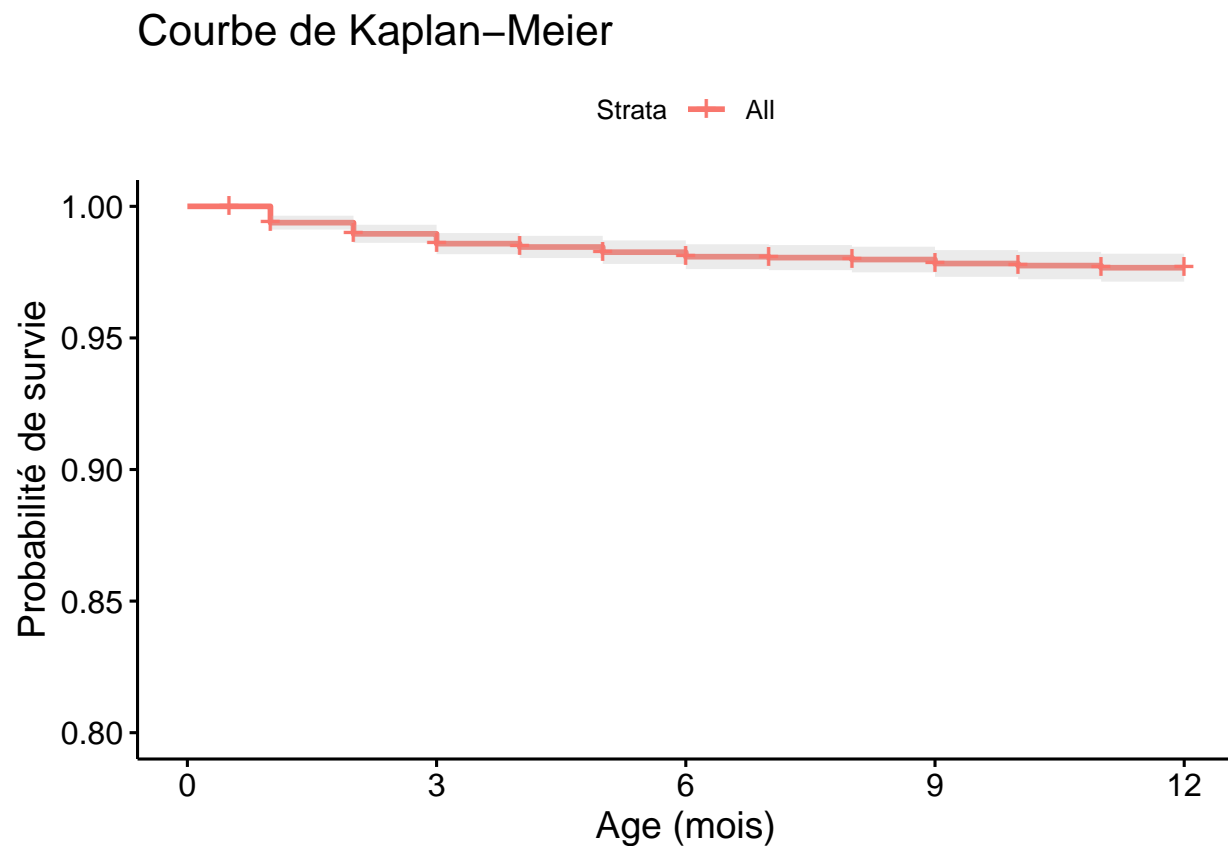
```
##
##  1  2  3  4  5  6  7  8  9 10 11
## 21 14 12  4  6  5  1  2  4  2  2
```

```
KM_estimator = survfit(Surv(chldage,hospital) ~ 1,data = pneumon)
ggsurvplot(
  KM_estimator,
  data = pneumon,
```

```

conf.int = TRUE,          # Ajout de l'intervalle de confiance
xlab = "Age (mois)",      # axe des abscisses
ylab = "Probabilité de survie", # axe des ordonnées
ylim = c(0.8, 1),
title = "Courbe de Kaplan-Meier"
)

```

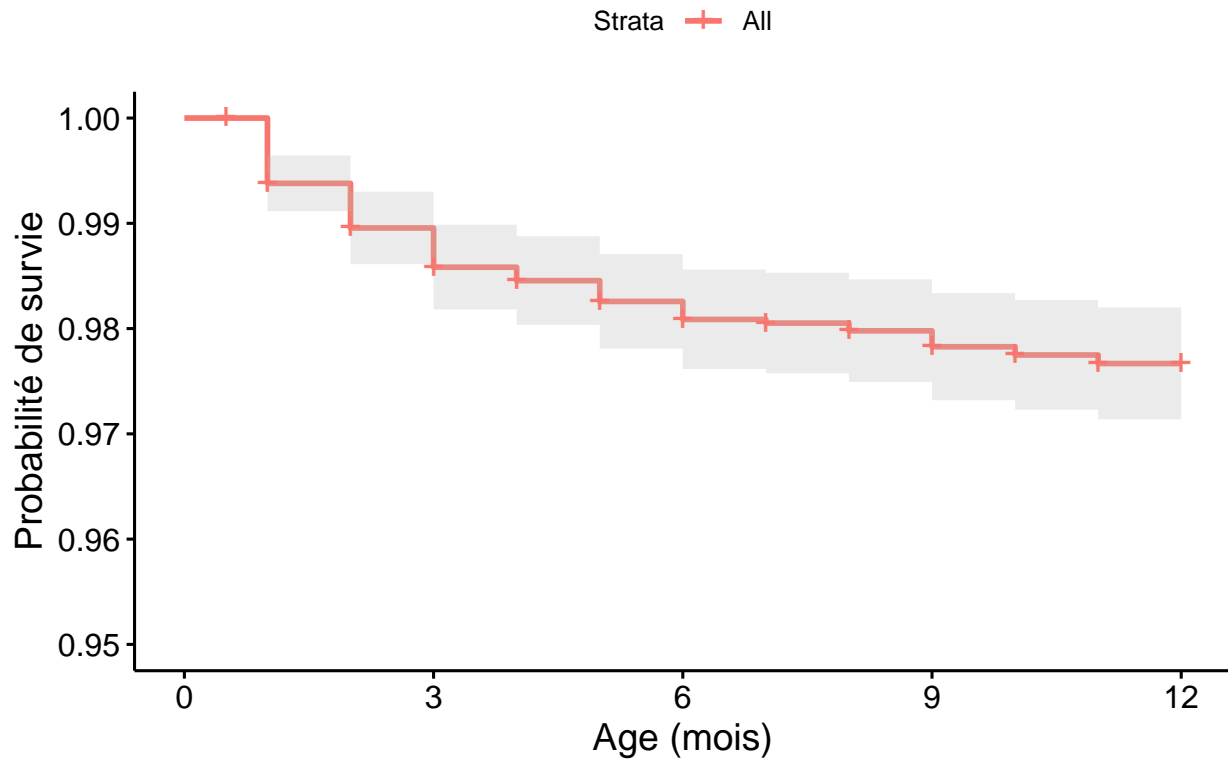


```

ggsurvplot(
  KM_estimator,
  data = pneumon,
  conf.int = TRUE,          # Ajout de l'intervalle de confiance
  xlab = "Age (mois)",      # axe des abscisses
  ylab = "Probabilité de survie", # axe des ordonnées
  ylim = c(0.95, 1),
  title = "Courbe de Kaplan-Meier"
)

```

Courbe de Kaplan–Meier



```
#Estimation de Kaplan-Meier à 6 mois
summary_km <- summary(KM_estimator, times = 6)
# Probabilité de survie à 6 mois
prob_6_months <- summary_km$surv
ci_lower <- summary_km$lower
ci_upper <- summary_km$upper

# Affichage
cat("La probabilité de ne pas avoir de pneumonie à 6 mois est de:", prob_6_months, "\n")

## La probabilité de ne pas avoir de pneumonie à 6 mois est de: 0.9808731
cat("Intervalle de confiance :", "[", ci_lower, ",", ci_upper, "]\n ")

## Intervalle de confiance : [ 0.9761631 , 0.9856058 ]
##

names(pneumon)

## [1] "chldage" "hospital" "mthage" "urban" "alcohol" "smoke"
## [7] "region" "poverty" "bweight" "race" "education" "nsibs"
## [13] "wmonth" "sfmonth"

pneumon <- pneumon %>%
  mutate(Z = ifelse(wmonth > 0, 1, 0))
```

Cox Model

```
surv_object <- Surv(time = pneumon$chldage, event = pneumon$hospital)
cox_model <- coxph(surv_object ~ Z, data = pneumon)
```

```
# Résumé
summary(cox_model)
```

```
## Call:
## coxph(formula = surv_object ~ Z, data = pneumon)
##
##      n= 3470, number of events= 73
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## Z -1.0970    0.3339    0.2973 -3.69 0.000224 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## Z    0.3339      2.995    0.1864    0.5979
##
## Concordance= 0.614 (se = 0.023 )
## Likelihood ratio test= 16.59 on 1 df,  p=5e-05
## Wald test               = 13.62 on 1 df,  p=2e-04
## Score (logrank) test = 15.04 on 1 df,  p=1e-04
```

Les enfants allaités ont 66% moins de risque de développer une pneumonie

```
# Identification des colonnes factorielles
factor_columns <- sapply(pneumon, is.factor)
```

```
# Affichage des noms des colonnes factorielles
names(factor_columns[factor_columns])
```

```
## [1] "urban" "alcohol" "smoke" "region" "poverty" "race"
```

```
# Transformation des colonnes factorielles en numériques
pneumon[factor_columns] <- lapply(pneumon[factor_columns], function(x) as.numeric(as.character(x)))
```

```
# Calcul de la matrice de corrélation
cor_matrix1 <- cor(pneumon, use = "complete.obs", method = "pearson") # Méthode de Pearson par défaut

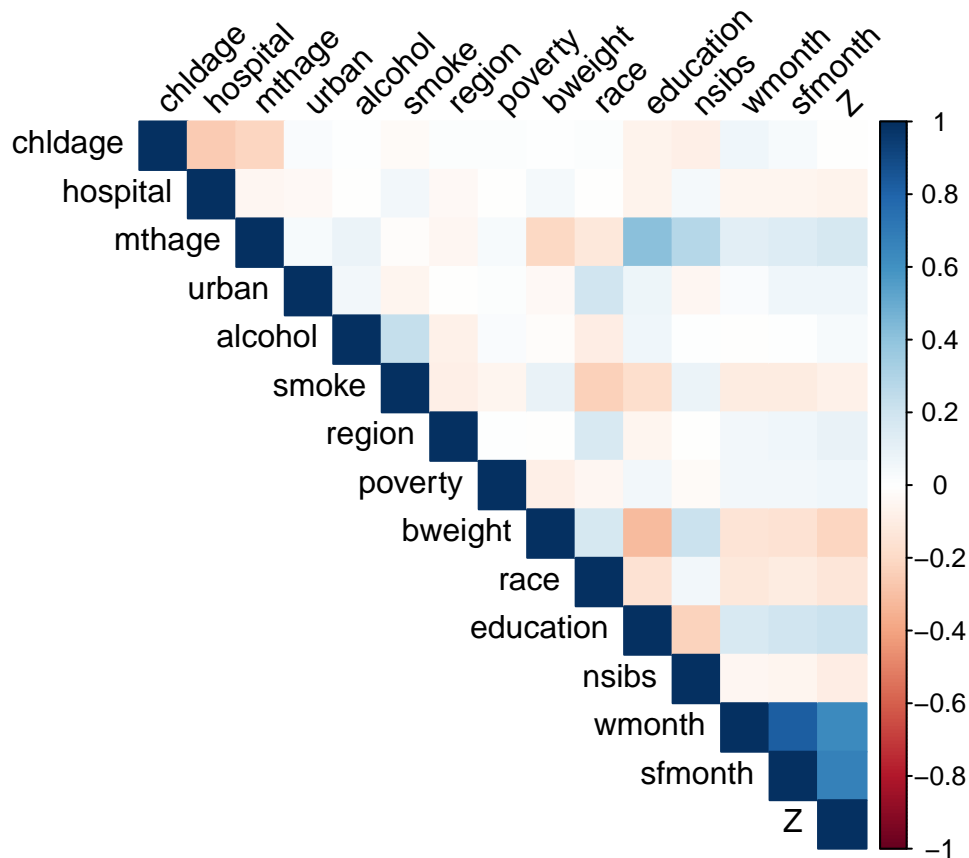
print(cor_matrix1)
```

```
##           chldage      hospital      mthage      urban      alcohol
## chldage      1.000000000 -0.253114792 -0.21556912 0.020903895 0.001286559
## hospital    -0.253114792  1.000000000 -0.04419288 -0.030669422 -0.006508275
## mthage      -0.215569124 -0.044192877  1.00000000 0.033410055 0.083981244
## urban       0.020903895 -0.030669422  0.03341006  1.000000000 0.057476346
## alcohol     0.001286559 -0.006508275  0.08398124  0.057476346  1.000000000
## smoke      -0.021004639  0.053400406 -0.01150465 -0.058724228 0.238315159
## region      0.012623649 -0.034459792 -0.04341037 -0.004563522 -0.075832142
## poverty     0.015333041 -0.009894626  0.03920641  0.018503414 0.022265520
## bweight     0.006711565  0.044963206 -0.20818595 -0.035357139 -0.014053533
## race        0.012379223 -0.001423781 -0.12446400  0.194596806 -0.092757874
## education  -0.068550486 -0.061352119  0.41182697  0.079763955 0.067635331
## nsibs       -0.082168602  0.045702280  0.28967512 -0.047958539 0.002674330
## wmonth      0.064864711 -0.055454658  0.12818182  0.023587550 -0.003909600
```

```
## sfmonth      0.038443804 -0.055421224  0.14465866  0.060974981  0.005563199
## Z            -0.002773125 -0.065933431  0.17871853  0.067766430  0.034577685
##              smoke      region      poverty      bweight      race
## chldage      -0.02100464  0.012623649  0.015333041  0.006711565  0.012379223
## hospital      0.05340041 -0.034459792 -0.009894626  0.044963206 -0.001423781
## mthage        -0.01150465 -0.043410375  0.039206414 -0.208185954 -0.124463998
## urban         -0.05872423 -0.004563522  0.018503414 -0.035357139  0.194596806
## alcohol        0.23831516 -0.075832142  0.022265520 -0.014053533 -0.092757874
## smoke          1.00000000 -0.089109419 -0.054406667  0.099754739 -0.235826854
## region         -0.08910942  1.000000000  0.001810924 -0.008590425  0.167805397
## poverty        -0.05440667  0.001810924  1.000000000 -0.080465097 -0.044483510
## bweight        0.09975474 -0.008590425 -0.080465097  1.000000000  0.174346223
## race           -0.23582685  0.167805397 -0.044483510  0.174346223  1.000000000
## education      -0.17972248 -0.056135227  0.051082726 -0.315518403 -0.152141308
## nsibs           0.08138380 -0.007396747 -0.025146320  0.218528171  0.059763082
## wmonth          -0.10066011  0.059056519  0.051673026 -0.140189405 -0.126745706
## sfmonth         -0.10226337  0.063369781  0.050159503 -0.150477769 -0.109649476
## Z              -0.07803857  0.091571789  0.060254794 -0.210663791 -0.138386910
##              education      nsibs      wmonth      sfmonth      Z
## chldage        -0.06855049 -0.082168602  0.06486471  0.038443804 -0.002773125
## hospital        -0.06135212  0.045702280 -0.05545466 -0.055421224 -0.065933431
## mthage           0.41182697  0.289675121  0.12818182  0.144658664  0.178718532
## urban           0.07976395 -0.047958539  0.02358755  0.060974981  0.067766430
## alcohol          0.06763533  0.002674330 -0.00390960  0.005563199  0.034577685
## smoke           -0.17972248  0.081383802 -0.10066011 -0.102263372 -0.078038574
## region           -0.05613523 -0.007396747  0.05905652  0.063369781  0.091571789
## poverty          0.05108273 -0.025146320  0.05167303  0.050159503  0.060254794
## bweight          -0.31551840  0.218528171 -0.14018940 -0.150477769 -0.210663791
## race             -0.15214131  0.059763082 -0.12674571 -0.109649476 -0.138386910
## education        1.00000000 -0.225044780  0.16896277  0.190921203  0.213630238
## nsibs            -0.22504478  1.000000000 -0.04407073 -0.058281512 -0.097174989
## wmonth           0.16896277 -0.044070735  1.00000000  0.823116038  0.630028229
## sfmonth          0.19092120 -0.058281512  0.82311604  1.000000000  0.672374257
## Z                0.21363024 -0.097174989  0.63002823  0.672374257  1.000000000
```

```
# Visualisation de la matrice de corrélation
```

```
corrplot(cor_matrix1, method = "color", type = "upper", tl.col = "black", tl.srt = 45)
```



```
# Identifions les corrélations fortes
high_corr <- which(abs(cor_matrix1) > 0.8 & cor_matrix1 != 1, arr.ind = TRUE)

# Affichage des paires de variables avec corrélation forte
high_corr_pairs <- data.frame(
  Var1 = rownames(cor_matrix1)[high_corr[, 1]],
  Var2 = colnames(cor_matrix1)[high_corr[, 2]],
  Correlation = cor_matrix1[high_corr]
)
print(high_corr_pairs)
```

```
##      Var1    Var2 Correlation
## 1 sfmonth wmonth    0.823116
## 2 wmonth sfmonth    0.823116
```

Les variables “sfmonth” et “wmonth” sont fortement corrélées. Nous allons donc supprimer “sfmonth” pour notre analyse.

```
# Objet de survie
surv_object <- Surv(time = pneumon$chldage, event = pneumon$hospital)

# Modèle de Cox avec la variable Z
cox_model <- coxph(surv_object ~ Z, data = pneumon)

# Résumé des résultats
summary(cox_model)
```

```
## Call:
```



```
## coxph(formula = surv_object ~ Z, data = pneumon)
##
## n= 3470, number of events= 73
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## Z -1.0970    0.3339    0.2973 -3.69 0.000224 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## Z      0.3339      2.995    0.1864    0.5979
##
## Concordance= 0.614 (se = 0.023 )
## Likelihood ratio test= 16.59 on 1 df,  p=5e-05
## Wald test              = 13.62 on 1 df,  p=2e-04
## Score (logrank) test = 15.04 on 1 df,  p=1e-04
```

```
# Modèle nul
cox_null <- coxph(surv_object ~ 1, data = pneumon)

# Test du rapport de vraisemblance
anova(cox_null, cox_model, test = "LRT")
```

```
## Analysis of Deviance Table
## Cox model: response is  surv_object
## Model 1: ~ 1
## Model 2: ~ Z
##      loglik  Chisq Df Pr(>|Chi|)
## 1 -587.06
## 2 -578.76 16.585  1  4.651e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On remarque que le modèle 2 a une meilleure vraisemblance(587) que le modèle 1. De plus la p value est vraiment inférieure à 0.05 ce qui signifie que l'effet de l'allaitement (variable Z) sur le risque de la pneumonie est significatif.

```
co_variables <- names(pneumon)[3:14]

# Fonction pour ajuster un modèle et résumer les résultats
model_resultat <- lapply(co_variables, function(var) {
  formule <- as.formula(paste("Surv(chldage, hospital) ~ Z +", var))
  summary(coxph(formule, data = pneumon))
})

model_aic <- sapply(co_variables, function(var) {
  formule <- as.formula(paste("Surv(chldage, hospital) ~ Z +", var))
  model <- coxph(formule, data = pneumon)
  AIC(model)
})

# Identification du modèle avec le plus petit AIC
best_model <- co_variables[which.min(model_aic)]
cat("Le meilleur modèle est avec la variable :", best_model, "\n")
```

```
## Le meilleur modèle est avec la variable : smoke
```

```

cat("AIC le plus faible :", min(model_aic), "\n")

## AIC le plus faible : 1154.283
best_model <- coxph(Surv(chldage, hospital) ~ Z + smoke, data = pneumon)

summary(best_model)

## Call:
## coxph(formula = Surv(chldage, hospital) ~ Z + smoke, data = pneumon)
##
##      n= 3470, number of events= 73
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Z          -1.0501    0.3499   0.2979 -3.525 0.000424 ***
## smoke    0.4224     1.5257   0.1510  2.797 0.005156 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## Z              0.3499      2.8579    0.1951    0.6274
## smoke         1.5257      0.6554    1.1348    2.0513
##
## Concordance= 0.665  (se = 0.031 )
## Likelihood ratio test= 23.83  on 2 df,   p=7e-06
## Wald test               = 21.62  on 2 df,   p=2e-05
## Score (logrank) test = 23.42  on 2 df,   p=8e-06

```

On peut donc dire que la variable “smoke” est la variable la plus importante parmi les covariables pour améliorer le modèle. Le modèle avec “smoke” ayant l’AIC le plus faible, celui ci est donc statistiquement préférable. Aussi le coefficient positif (0.4224) indique que l’exposition au tabac(le fait que la mère fume) est associée à une augmentation du risque de pneumonie. Ceux ci ont un risque de pneumonie 52,57% plus élevé par rapport aux enfants non exposés. La valeur de la p value indique que cette variable est significative.

```

# Liste des covariables à intégrer dans le modèle
facteurs <- c("mthage", "urban", "alcohol", "smoke", "region", "bweight", "poverty", "race", "nsibs")
#Formule dynamique pour stocker les résultats
resultats <- lapply(facteurs, function(factor) {

  formula <- as.formula(paste("Surv(chldage, hospital) ~ Z +", factor))

  # Ajustons le modèle de Cox
  model <- coxph(formula, data = pneumon)

  #Extraction des résultats du test de Wald pour l'allaitement
  wald_resultats <- summary(model)$coefficients["Z", ]

  # Détails du test de Wald
  data.frame(
    Factor = factor,
    Coefficient = wald_resultats["coef"],
    Exp_Coefficient = wald_resultats["exp(coef)"],
    Std_Error = wald_resultats["se(coef)"],
    z = wald_resultats["z"],
    P_Value = wald_resultats["Pr(>|z|)"]
  )
})

```

```

)
})

# Combine results into a single data frame
resultats_df <- do.call(rbind, resultats)

# View the results
print(resultats_df)

```

| ## | Factor | Coefficient | Exp_Coefficient | Std_Error | z | P_Value |
|----------|---------|-------------|-----------------|-----------|-----------|--------------|
| ## coef | mothage | -1.026509 | 0.3582556 | 0.3009579 | -3.410804 | 0.0006477155 |
| ## coef1 | urban | -1.071991 | 0.3423264 | 0.2978153 | -3.599516 | 0.0003188105 |
| ## coef2 | alcohol | -1.094903 | 0.3345719 | 0.2973969 | -3.681623 | 0.0002317540 |
| ## coef3 | smoke | -1.050098 | 0.3499036 | 0.2979123 | -3.524855 | 0.0004237147 |
| ## coef4 | region | -1.065794 | 0.3444541 | 0.2976207 | -3.581049 | 0.0003422169 |
| ## coef5 | bweight | -1.008689 | 0.3646967 | 0.3017797 | -3.342469 | 0.0008303666 |
| ## coef6 | poverty | -1.091894 | 0.3355804 | 0.2977203 | -3.667514 | 0.0002449200 |
| ## coef7 | race | -1.121234 | 0.3258775 | 0.2995133 | -3.743518 | 0.0001814612 |
| ## coef8 | nsibs | -1.045388 | 0.3515553 | 0.2982553 | -3.505012 | 0.0004565872 |

-Dans tous les modèles, l'allaitement est significativement associé à une réduction du risque de pneumonie nécessitant une hospitalisation chez le nourrisson. Les Hazard Ratios (HR) associés à l'allaitement varient entre 0,3259 et 0,3647, indiquant une réduction de 63% à 67% du risque de pneumonie pour les enfants allaités par rapport aux enfants non allaités. L'effet protecteur de l'allaitement reste robuste après ajustement pour chaque facteur. Cela suggère que l'allaitement exerce un effet protecteur indépendant du statut socio-économique, des habitudes de vie de la mère (tabac, alcool), ou d'autres caractéristiques démographiques. Bien que les résultats montrent que chaque facteur peut éventuellement influencer le risque, l'effet de l'allaitement est toujours significatif, indiquant qu'il s'agit d'un déterminant clé. -Pour la variable "mothage" par exemple le coefficient est de -1.02. Donc plus la mère est âgée, moins l'enfant a de risque d'attraper une pneumonie. Ceci peut s'expliquer par le fait que la mère ait déjà de l'expérience avec les enfants ou une meilleure stabilité économique. -Les enfants vivant en milieu urbain semblent avoir un risque plus faible de développer une pneumonie que ceux vivant en milieu rural. Cela pourrait s'expliquer par un meilleur accès aux soins de santé et une meilleure hygiène en milieu urbain. -Les enfants avec un faible poids de naissance (< 5.5 lbs) ont un risque accru de pneumonie. Un faible poids à la naissance est souvent lié à un système immunitaire affaibli. -Les enfants vivant dans des ménages pauvres ont un risque accru de pneumonie, probablement en raison d'un accès limité aux soins de santé, une nutrition insuffisante et de conditions de vie défavorables. -Les enfants issus de certaines races (par exemple, les mères blanches dans notre cas) pourraient être associées à un risque réduit de pneumonie, ce qui peut refléter des inégalités socio-économiques ou des différences dans les comportements de santé. -Un plus grand nombre de frères et sœurs augmente le risque de pneumonie. -Les enfants de mères qui consomment peu ou pas d'alcool ont un risque plus faible de développer une pneumonie. Une consommation excessive d'alcool peut indiquer des comportements à risque ou un environnement familial moins favorable.

```

# Modèle de Cox multivariée
full_model <- coxph(Surv(chldage, hospital) ~ Z + mothage + urban + alcohol + smoke + region + bweight +
                    data = pneumon)

summary(full_model)

## Call:
## coxph(formula = Surv(chldage, hospital) ~ Z + mothage + urban +
##       alcohol + smoke + region + bweight + poverty + race + nsibs,
##       data = pneumon)
##
##      n= 3470, number of events= 73

```

```
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## Z          -0.78952   0.45406  0.30808 -2.563  0.01039 *
## mthage      -0.11381   0.89243  0.05174 -2.200  0.02784 *
## urban       -0.33483   0.71546  0.26099 -1.283  0.19952
## alcohol     -0.08549   0.91807  0.11124 -0.768  0.44220
## smoke        0.39091   1.47833  0.16471  2.373  0.01763 *
## region      -0.22669   0.79717  0.13074 -1.734  0.08293 .
## bweight      0.13956   1.14976  0.25572  0.546  0.58525
## poverty     -0.01730   0.98285  0.40015 -0.043  0.96552
## race         0.01991   1.02011  0.18563  0.107  0.91459
## nsibs        0.35245   1.42255  0.12955  2.721  0.00652 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## Z           0.4541      2.2023   0.2482   0.8305
## mthage       0.8924      1.1205   0.8064   0.9877
## urban        0.7155      1.3977   0.4290   1.1933
## alcohol      0.9181      1.0892   0.7382   1.1417
## smoke        1.4783      0.6764   1.0705   2.0416
## region       0.7972      1.2544   0.6170   1.0300
## bweight      1.1498      0.8697   0.6965   1.8979
## poverty      0.9829      1.0174   0.4486   2.1533
## race         1.0201      0.9803   0.7090   1.4677
## nsibs        1.4226      0.7030   1.1036   1.8338
##
## Concordance= 0.709 (se = 0.027 )
## Likelihood ratio test= 39.93 on 10 df,  p=2e-05
## Wald test               = 37.73 on 10 df,  p=4e-05
## Score (logrank) test = 39.85 on 10 df,  p=2e-05

# model nul (en enlevant la variable Z)
null_model <- coxph(Surv(chldage, hospital) ~ mthage + urban + alcohol + smoke + region + bweight + poverty + race + nsibs,
  data = pneumon)

# Test d'Anova
anova(null_model, full_model, test = "LRT")

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, hospital)
## Model 1: ~ mthage + urban + alcohol + smoke + region + bweight + poverty + race + nsibs
## Model 2: ~ Z + mthage + urban + alcohol + smoke + region + bweight + poverty + race + nsibs
##      loglik  Chisq Df Pr(>|Chi|)
## 1 -570.83
## 2 -567.09 7.4747 1 0.006257 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Le test du rapport de vraisemblance montre que l'ajout de l'allaitement (Variable Z) améliore considérablement l'ajustement du modèle ($p=0,006257$). Ce résultat appuie la conclusion selon laquelle l'allaitement maternel a un effet protecteur statistiquement significatif contre la pneumonie.

```
modele_final <- coxph(Surv(chldage, hospital) ~ wmonth + mthage + urban + alcohol + smoke + region + poverty + race + nsibs,
  data = pneumon,
  Z=0,
```

```

wmonth=0,
mthage = 27,
urban = 1,
alcohol = 3,
smoke = 0,
region = 2,
poverty = 1,
bweight = 0,
race = 1,
education = 12,
nsibs = 1
)
# Prediction de la probabilité de Survie
surv_predictions <- survfit(modele_final, newdata = nouveau_ne)

# Probabilité de Survie à 6 mois

survie_a_6_mois <- summary(surv_predictions, times = 6)$surv
print(survie_a_6_mois)

## [1] 0.9910291

```

Pour un nouveau-né avec une mère blanche âgée de 27 ans, vivant en milieu urbain, ne consommant pas d'alcool ni de cigarettes, né dans une région du Nord Centre, vivante en situation de pauvreté, avec un poids à la naissance inférieur à 5.5 lbs, ayant une mère avec 12 ans d'éducation, et ayant un frère ou une soeur, la probabilité de ne pas avoir développé une pneumonie à 6 mois est estimée à 99,10%