Death following patient's 1st admission to hospital for heart failure

Amira Ibrahim

November 26, 2020

Import Data :

```
library(readr)
HF_Data <- read_csv(file.choose())
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double()
## )
## i Use `spec()` for the full column specifications.
```

```
attach(HF_Data)
```

load libraries :

```
library(survival)
```

```
##
## Attaching package: 'survival'
```

```
## The following object is masked from 'HF_Data':
##
##     cancer
```

```
library(survminer)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: ggpubr
```

```
library(ggplot2)
library(dplyr)
```

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

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

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

About the dataset :

The data are simulated based on real hospital administrative data for England called Hospital Episodes Statistics.

Your simulated extract contains a random sample of emergency (unplanned) admissions for heart failure (ICD10 code I50).

Here's a list of the fields and an explanation for some of them.

Many of the fields are comorbidities coded as 0/1, where 1 indicates that the patient had it recorded.

All comorbidities are recorded in HES's secondary diagnosis fields, of which there are currently 19.

There are 24 fields to capture procedures and operations.

#death (0/1)

#los (hospital length of stay in nights)

#age (in years)

#gender (1=male, 2=female)

#cancer

#cabg (previous heart bypass)

#crt (cardiac resynchronisation device - a treatment for heart failure)

#defib (defibrillator implanted)

#dementia

#diabetes (any type)

#hypertension

#ihd (ischaemic heart disease)

#mental_health (any mental illness)

#arrhythmias

#copd (chronic obstructive lung disease)

#obesity

#pvd (peripheral vascular disease)

#renal_disease

#valvular_disease (disease of the heart valves)

#metastatic_cancer

#pacemaker

#pneumonia

#prior_appts_attended (number of outpatient appointments attended in the previous year)

#prior_dnas (number of outpatient appointments missed in the previous year)

#pci (percutaneous coronary intervention)

#stroke (history of stroke)

#senile

#quintile (socio-economic status for neighbourhood of patients, from 1 (most affluent) to 5 (poorest))

#ethnicgroup (see below for categories)

#fu_time (follow-up time, i.e. time in days since admission to hospital)

#Ethnic group has the following categories in this extract:

```
#1=white

#2=black

#3=Indian subcontinent

#8=not known

#9=other
```

Explore Data :

```
dim(HF_Data)
```

```
## [1] 1000   31
```

```
head(HF_Data)
```

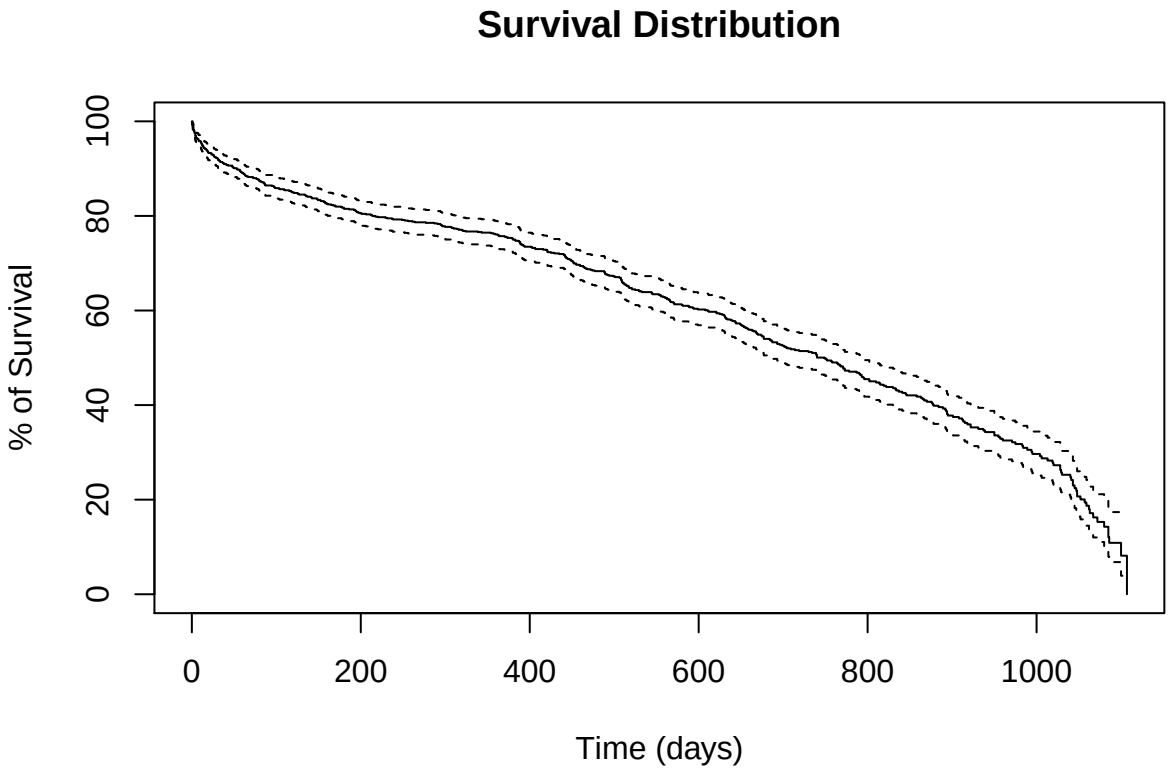
```
## # A tibble: 6 x 31
##      id death   los   age gender cancer  cabg   crt defib dementia diabetes
##   <dbl> <dbl> <dbl> <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl>    <dbl>    <dbl>
## 1     1     0     2    90      2      0     0     0     0        0        0
## 2     2     0    10    74      1      0     0     0     0        0        0
## 3     3     0     3    83      2      0     0     0     0        0        0
## 4     4     0     1    79      1      0     0     0     0        0        1
## 5     5     0    17    94      2      0     0     0     0        0        1
## 6     6     0    47    89      1      0     0     0     0        0        0
## # ... with 20 more variables: hypertension <dbl>, ihd <dbl>,
## #   mental_health <dbl>, arrhythmias <dbl>, copd <dbl>, obesity <dbl>,
## #   pvd <dbl>, renal_disease <dbl>, valvular_disease <dbl>,
## #   metastatic_cancer <dbl>, pacemaker <dbl>, pneumonia <dbl>,
## #   prior_appts_attended <dbl>, prior_dnas <dbl>, pci <dbl>, stroke <dbl>,
## #   senile <dbl>, quintile <dbl>, ethnicgroup <dbl>, fu_time <dbl>
```

Create a Surv object

```
survobj <- with(HF_Data , Surv(fu_time, death))
```

Run km plot :

```
km_fit <- survfit(survobj ~ 1,data=HF_Data)

plot(km_fit,main="Survival Distribution",xlab = "Time (days)",yscale = 100 , ylab = "% of Survival")
```



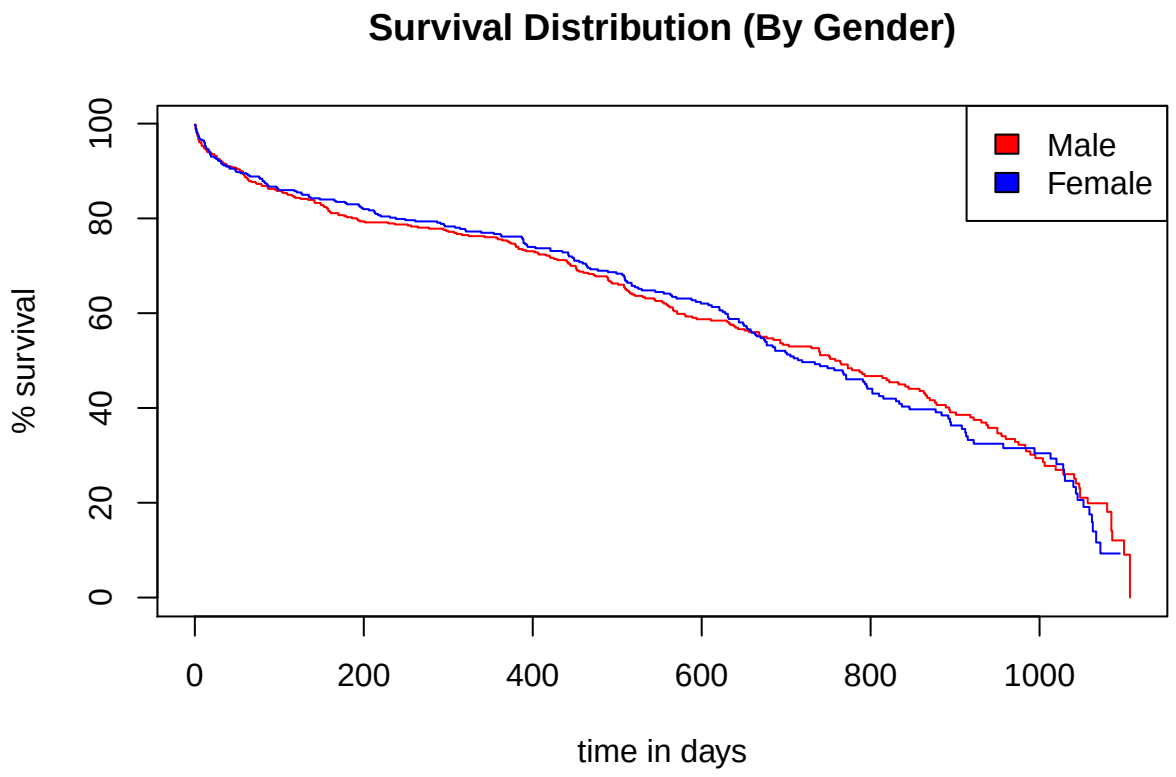

```
summary(km_fit, times = c(1:7,30,60,90*(1:10)))
```

```
## Call: survfit(formula = survobj ~ 1, data = HF_Data)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1    992      12    0.988 0.00346        0.981        0.995
##     2    973       7    0.981 0.00435        0.972        0.989
##     3    963       5    0.976 0.00489        0.966        0.985
##     4    954       6    0.970 0.00546        0.959        0.980
##     5    945       5    0.964 0.00590        0.953        0.976
##     6    938       1    0.963 0.00598        0.952        0.975
##     7    933       1    0.962 0.00606        0.951        0.974
##    30    865      39    0.921 0.00865        0.905        0.939
##    60    809      28    0.891 0.01010        0.871        0.911
##    90    770      24    0.864 0.01117        0.843        0.887
##   180    698      43    0.815 0.01282        0.790        0.841
##   270    653      24    0.787 0.01363        0.760        0.814
##   360    619      21    0.761 0.01428        0.733        0.789
##   450    525      44    0.705 0.01554        0.675        0.736
##   540    429      47    0.639 0.01681        0.607        0.673
##   630    362      32    0.589 0.01765        0.556        0.625
##   720    266      43    0.514 0.01876        0.479        0.552
##   810    190      31    0.448 0.01979        0.411        0.488
##   900    126      26    0.378 0.02098        0.339        0.421
```

At the first day of admission the survival probability is 98.8% , at 900 days after a first emergency admission for heart failure, the probability of surviving is just 38%.

Survival Probability according to gender:

```
## Splitting the curve by gender:

km_gender_fit <- survfit(survobj ~ gender,data=HF_Data)

plot(km_gender_fit , xlab="time in days" , ylab="% survival" , yscale=100 ,
     main="Survival Distribution (By Gender)" ,col = c("red","blue") )
legend("topright" , c("Male","Female"),fill = c("red","blue"))
```




```
survfit(survobj~gender,data= HF_Data)
```

```
## Call: survfit(formula = survobj ~ gender, data = HF_Data)
##
##             n events median 0.95LCL 0.95UCL
## gender=1 548    268    758     685     845
## gender=2 452    224    719     670     802
```

median survival time for males was 25 months

median survival time for females was 23.9 months

Test for difference between male , female:

#(logrank test)

```
survdiff(survobj ~ gender, rho=0,data = HF_Data)
```

```
## Call:
## survdiff(formula = survobj ~ gender, data = HF_Data, rho = 0)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## gender=1 548      268      271    0.0365     0.082
## gender=2 452      224      221    0.0448     0.082
##
##  Chisq= 0.1  on 1 degrees of freedom, p= 0.8
```

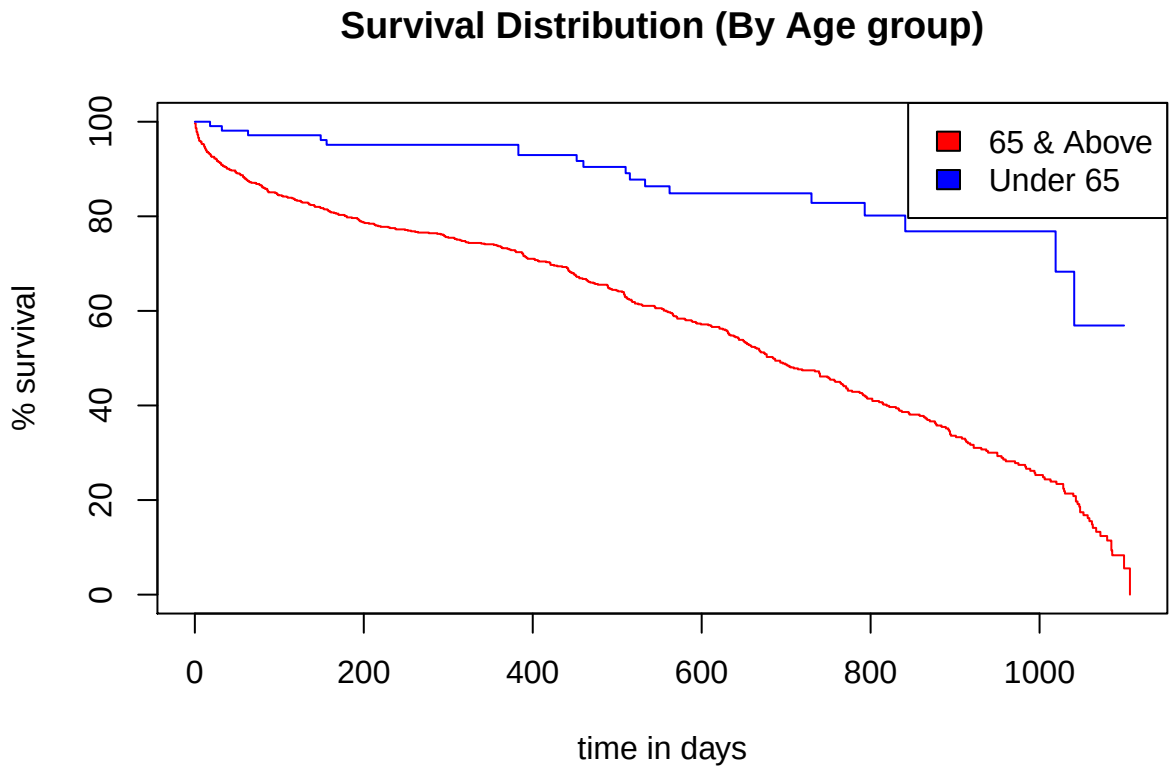
p value = 0.8 , There's no good evidence of a difference between the genders in their survival times from the first admission for heart failure.

Survival Probability according to age group:

those aged 65 and above versus those aged under 65.

```
## Prepare Data :
  HF_Data <- HF_Data %>%
  mutate(Age = ifelse(age <65 , "Under 65","65 & Above"))
```

```
## Splitting km curve by age group:
km_Age_fit <- survfit(Surv(fu_time, death) ~ Age,data=HF_Data)

plot(km_Age_fit, xlab="time in days" , ylab="% survival" , yscale=100 ,
     main="Survival Distribution (By Age group)" ,col = c("red","blue") )
legend("topright" , c("65 & Above","Under 65"),fill = c("red","blue"))
```



Test for difference between age groups:

#(logrank test)

```
survdiff(Surv(fu_time, death) ~ Age, rho=0,data = HF_Data)
```

```
## Call:
## survdiff(formula = Surv(fu_time, death) ~ Age, data = HF_Data,
##     rho = 0)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## Age=65 & Above 885      474      425      5.65      41.7
```

```
## Age=Under 65   115       18       67     35.85      41.7
##
##  Chisq= 41.7  on 1 degrees of freedom, p= 1e-10
```

p value$<$ 0.001 , There's significant difference between the age groups in their survival times from the first admission for heart failure.

Younger patients (under 65 years) live significantly longer after hospital admission than older ones do.

Run Cox regression model with age as predictor (continuous variable)

#Asuume that linear relation with hazard

```
##Generate model
cox <- coxph(Surv(fu_time, death) ~ age, data = HF_Data)
```

```
##Summarise model
summary(cox)
```

```
## Call:
## coxph(formula = Surv(fu_time, death) ~ age, data = HF_Data)
##
##   n= 1000, number of events= 492
##
##         coef exp(coef) se(coef)     z Pr(>|z|)
## age 0.056005  1.057602 0.005193 10.78   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## age     1.058     0.9455     1.047     1.068
##
## Concordance= 0.651  (se = 0.013 )
## Likelihood ratio test= 138  on 1 df,   p=<2e-16
## Wald test            = 116.3  on 1 df,   p=<2e-16
## Score (logrank) test = 115.7  on 1 df,   p=<2e-16
```

For each increase of 1 year in age ,the hazard of death at any given time point goes up by 6% .

Older people have a higher hazard of death than young people .

P-value is very tiny (<2e-16) , Age is highly statisticaly significant.

Run Cox regression model with ethnic group as predictor (categorical variable)

```
# Summarise variable
table(ethnicgroup, exclude = NULL)
```

```
## ethnicgroup
##    1    2    3    9 <NA>
##  889   17   34   17   43
```

```
# Generate and summarise model
cox_ethinic <- coxph(Surv(fu_time, death) ~ as.factor(ethnicgroup),data = HF_Data)
summary(cox_ethinic)
```

```
## Call:
## coxph(formula = Surv(fu_time, death) ~ as.factor(ethnicgroup),
##     data = HF_Data)
##
##   n= 957, number of events= 471
##    (43 observations deleted due to missingness)
##
##                              coef exp(coef) se(coef)      z Pr(>|z|)
## as.factor(ethnicgroup)2 -0.06428   0.93774  0.32000 -0.201  0.84078
## as.factor(ethnicgroup)3 -1.19586   0.30244  0.41108 -2.909  0.00362 **
## as.factor(ethnicgroup)9  0.07394   1.07674  0.35706  0.207  0.83596
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                          exp(coef) exp(-coef) lower .95 upper .95
## as.factor(ethnicgroup)2    0.9377     1.0664    0.5008    1.7558
## as.factor(ethnicgroup)3    0.3024     3.3064    0.1351    0.6769
## as.factor(ethnicgroup)9    1.0767     0.9287    0.5348    2.1679
##
## Concordance= 0.516  (se = 0.006 )
## Likelihood ratio test= 12.99  on 3 df,   p=0.005
## Wald test            = 8.55  on 3 df,   p=0.04
## Score (logrank) test = 9.61  on 3 df,   p=0.02
```

43 observations were excluded because of missing data.

check the standard errors. They're between about 0.3 and 0.4, so not too big to worry about but not exactly small either.

For ethnic group 2, black people, the hazard ratio ("exp(coef)") is 0.94 to two decimal places, with a 95% CI of 0.50 to 1.75. That's pretty wide.

The hazard for all-cause mortality for black people is 0.94 times the hazard for white people, but p=0.837 and the CI is so wide that you can't reliably conclude anything except that you don't have any evidence to reject the null hypothesis.

You'd therefore conclude that black and white people appear to have the same hazard.

For ethnic group 3, those from the Indian subcontinent, however, their hazard ratio is 0.30, 95% CI 0.14 to 0.68, p=0.004. That's a statistically significant difference in favour of these

patients compared with white patients. There's quite a bit of uncertainty about just how much lower their hazard is, but you've good evidence to suggest that it's lower.

For ethnic group 9 (other) you fail to reject the null, just like for black people.

Missing ethnic group : make an "unknown" category for them, give the value 8.

```
levels(HF_Data$ethnicgroup) <- c(levels(HF_Data$ethnicgroup),"8") # add level 8 to the factor
HF_Data$ethnicgroup[is.na(HF_Data$ethnicgroup)] <- "8" # Change NA to "None"
```

Run this model after modifications :

```
cox_ethinic1<- coxph(Surv(fu_time, death) ~ as.factor(ethnicgroup), data = HF_Data)
summary(cox_ethinic1)
```

```
## Call:
## coxph(formula = Surv(fu_time, death) ~ as.factor(ethnicgroup),
##     data = HF_Data)
##
##   n= 1000, number of events= 492
##
##                              coef exp(coef) se(coef)      z Pr(>|z|)
## as.factor(ethnicgroup)2 -0.06573   0.93638  0.31999 -0.205  0.83725
## as.factor(ethnicgroup)3 -1.19368   0.30310  0.41107 -2.904  0.00369 **
## as.factor(ethnicgroup)8 -0.02353   0.97675  0.22363 -0.105  0.91621
## as.factor(ethnicgroup)9  0.08160   1.08502  0.35706  0.229  0.81923
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                          exp(coef) exp(-coef) lower .95 upper .95
## as.factor(ethnicgroup)2    0.9364     1.0679    0.5001    1.7532
## as.factor(ethnicgroup)3    0.3031     3.2992    0.1354    0.6784
## as.factor(ethnicgroup)8    0.9767     1.0238    0.6301    1.5140
## as.factor(ethnicgroup)9    1.0850     0.9216    0.5389    2.1846
##
## Concordance= 0.518  (se = 0.008 )
## Likelihood ratio test= 12.95  on 4 df,   p=0.01
## Wald test            = 8.53  on 4 df,   p=0.07
## Score (logrank) test = 9.58  on 4 df,   p=0.05
```

For ethnic group 8. People with unknown ethnicity have a very similar hazard (0.98, CI 0.63 to 1.51, p=0.916) to white people.

All the results for the other three categories are very close to what they were before.

Investigating variables in order to best perform a Cox model with multiple predictors :

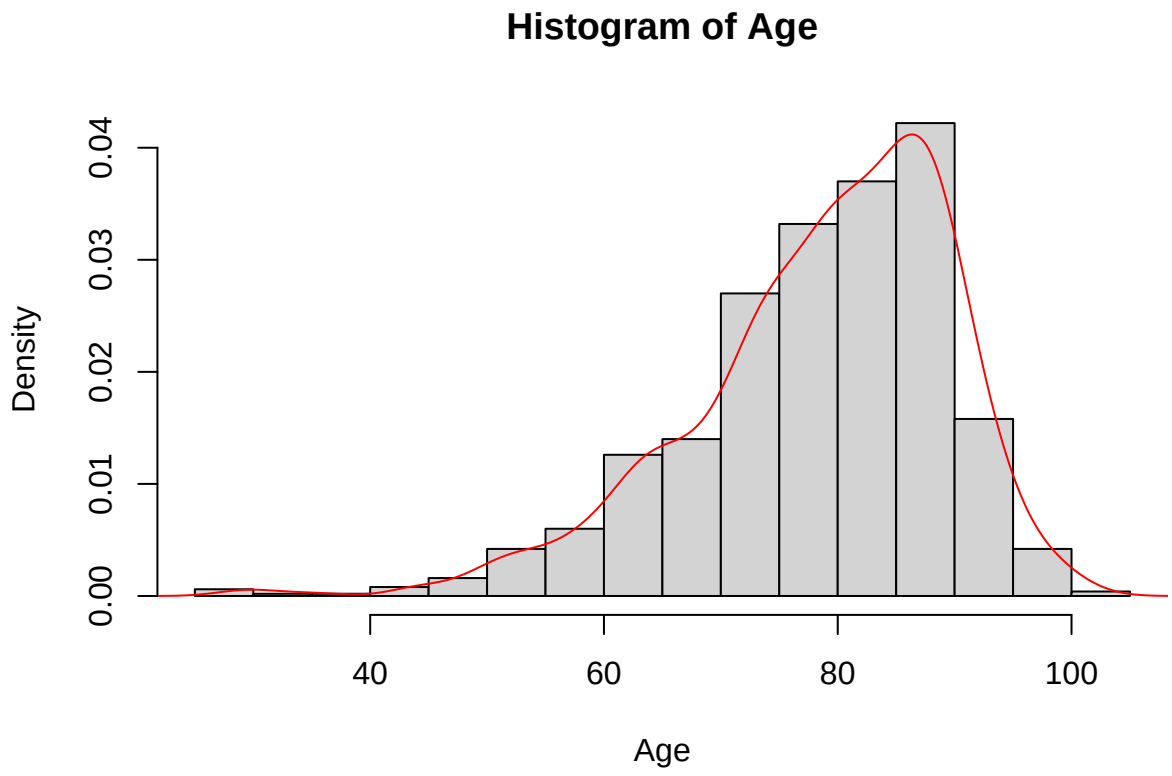
Check variables :

Age

```
summary(age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   29.00   73.00   80.00   78.73   87.00  102.00
```

```
hist(age,main = "Histogram of Age",xlab = "Age" ,freq = FALSE)
lines(density(age),col="red",lwd=1)
```



age is nearly normally distributed

Gender

```
table_gender <- table(as.factor(gender),exclude = NULL)
addmargins(table_gender)
```

```
##
##    1    2  Sum
##  548  452 1000
```

```
round(100*prop.table(table_gender) , digits = 1)
```

```
##
##    1    2
## 54.8 45.2
```

54.8% are males with no missing values .

prior dnas

```
table_prior_dnas <- table(as.factor(prior_dnas),exclude = NULL)
addmargins(table_prior_dnas)
```

```
##
##    0    1    2    3    4    5    6    7    8   10  Sum
##  732  156   50   34   17    3    3    2    1    2 1000
```

```
round(100*prop.table(table_prior_dnas) , digits = 1)
```

```
##
##    0    1    2    3    4    5    6    7    8   10
## 73.2 15.6  5.0  3.4  1.7  0.3  0.3  0.2  0.1  0.2
```

So nearly 74% of patients had missed no appointments, # but nearly three percent had missed five or more, with a maximum of ten.

Ethinic group

```
table_ethnic_gp <- table(as.factor(ethnicgroup),exclude = NULL)
addmargins(table_ethnic_gp)
```

```
##
##    1    2    3    9 <NA>  Sum
##  889   17   34   17   43 1000
```

```
round(100*prop.table(table_ethnic_gp) , digits = 1)
```

```
##
##    1    2    3    9 <NA>
## 88.9  1.7  3.4  1.7  4.3
```

there is 4.3% of data is missed .

Missing ethnic group : make an "unknown" category for them, give the value 8.

```
levels(HF_Data$ethnicgroup)<-c(levels(ethnicgroup),"8") # add level 8 to the factor
HF_Data$ethnicgroup[is.na(HF_Data$ethnicgroup)] <- "8" # Change NA to "None"
```

```
table_ethnic_gp <- table(as.factor(HF_Data$ethnicgroup),exclude = NULL)
addmargins(table_ethnic_gp)
```

```
##
##    1    2    3    8    9  Sum
##  889   17   34   43   17 1000
```

```
round(100*prop.table(table_ethnic_gp) , digits = 1)
```

```
##
##    1    2    3    8    9
## 88.9  1.7  3.4  4.3  1.7
```

COPD

```
table_copd <- table(as.factor(HF_Data$copd),exclude = NULL)
addmargins(table_copd)
```

```
##
##    0    1  Sum
##  758  242 1000
```

```
round(100*prop.table(table_copd) , digits = 1)
```

```
##
##    0    1
## 75.8 24.2
```

24% of patients had COPD, with no missing values.

may there is missing values masquerading as regular values.

It's actually likely that some patients have COPD but haven't been recorded as having it.

Such underrecording of comorbidities is common with administrative data for various reasons.

Run Cox Model :

```
cox1 <- coxph(Surv(fu_time, death) ~ age + as.factor(gender) + as.factor(copd)
              + prior_dnas + ethnicgroup ,data = HF_Data)

summary(cox1)
```

```
## Call:
## coxph(formula = Surv(fu_time, death) ~ age + as.factor(gender) +
##     as.factor(copd) + prior_dnas + ethnicgroup, data = HF_Data)
##
##   n= 1000, number of events= 492
##
##                          coef exp(coef)  se(coef)      z Pr(>|z|)
## age                  0.061999  1.063961  0.005516 11.241  < 2e-16 ***
## as.factor(gender)2 -0.253460  0.776111  0.094349 -2.686  0.00722 **
## as.factor(copd)1    0.136649  1.146425  0.103880  1.315  0.18836
## prior_dnas          0.163461  1.177579  0.039832  4.104 4.07e-05 ***
## ethnicgroup2       -0.307915  0.734978  0.353009 -0.872  0.38307
## ethnicgroup3       -0.823643  0.438830  0.414301 -1.988  0.04681 *
## ethnicgroup8       -0.045372  0.955642  0.225204 -0.201  0.84033
## ethnicgroup9        0.408255  1.504190  0.360737  1.132  0.25775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                    exp(coef) exp(-coef) lower .95 upper .95
## age                   1.0640     0.9399    1.0525    1.0755
## as.factor(gender)2    0.7761     1.2885    0.6451    0.9338
## as.factor(copd)1      1.1464     0.8723    0.9352    1.4053
## prior_dnas            1.1776     0.8492    1.0891    1.2732
```

```
## ethnicgroup2          0.7350     1.3606    0.3680    1.4681
## ethnicgroup3          0.4388     2.2788    0.1948    0.9884
## ethnicgroup8          0.9556     1.0464    0.6146    1.4859
## ethnicgroup9          1.5042     0.6648    0.7417    3.0504
##
## Concordance= 0.667  (se = 0.013 )
## Likelihood ratio test= 168.4  on 8 df,   p=<2e-16
## Wald test            = 141.7  on 8 df,   p=<2e-16
## Score (logrank) test = 140  on 8 df,   p=<2e-16
```

Femals(gender=2) have hazard ratio = 0.78 compared to males , with 95% CI between 0.65 to 0.93 , with p = 0.007 , So there is clear evidence of association between females and lower hazard of mortality.

females have 22% lower hazard than males after addmision .

Patients with COPD (copd=1) have hazard ratio = 1.15 compared to patients without COPD ,

Patients with COPD have higher Hazard by 15% than patients without COPD, but p = 0.188 , So we don't have evidence of association between copd and death in this sample.

Only ethnic group 3 has lower hazard with p = 0.047 which just sneak under 0.05 threshold.

people with indian subcontient live longer than white people in this sample after their addmision.

For one year increase in age , the hazard increase by 6% .

Ihere is clear and strong evidence of associatio between age and hazard of mortality after addmision .

For each appointment missed , the hazard increase by 18% .

Testing for proportional hazards assumption (with gender as predictor variable) :

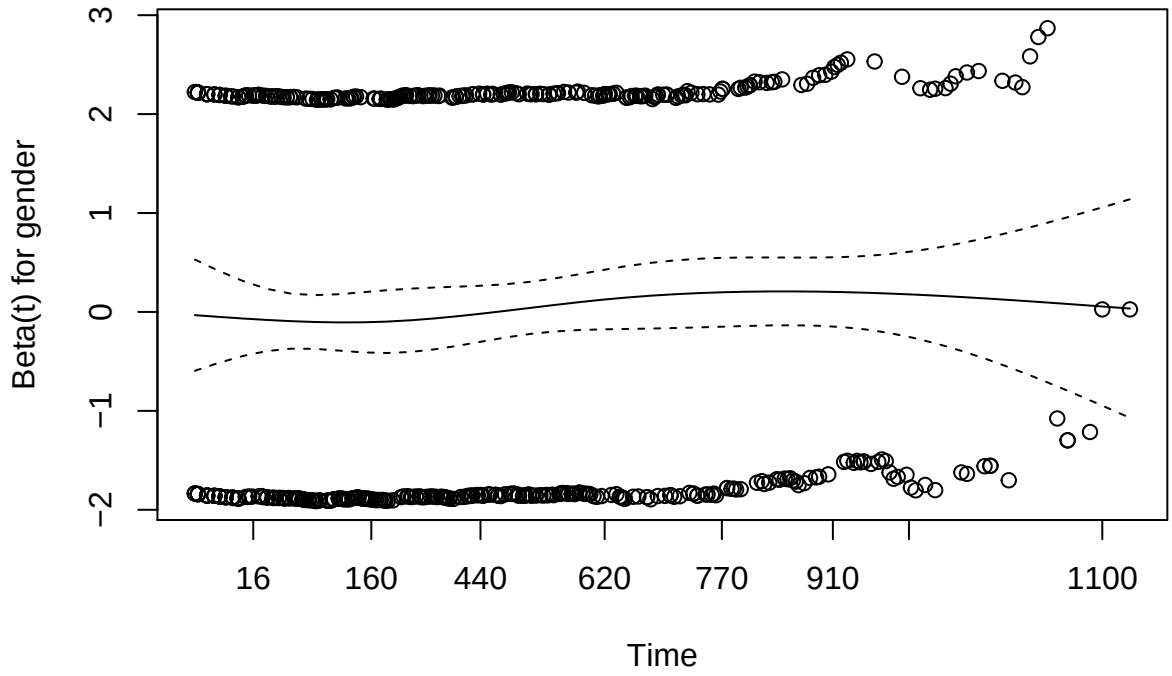
```
# Generate model fit
fit <- coxph(Surv(fu_time, death) ~ gender)
```

```
# Apply the test to the model
temp <- cox.zph(fit)
```

```
# Display results
print(temp)
```

```
##        chisq df    p
## gender  1.24  1 0.26
## GLOBAL  1.24  1 0.26
```

```r
# Plot the curves
plot(temp)
```

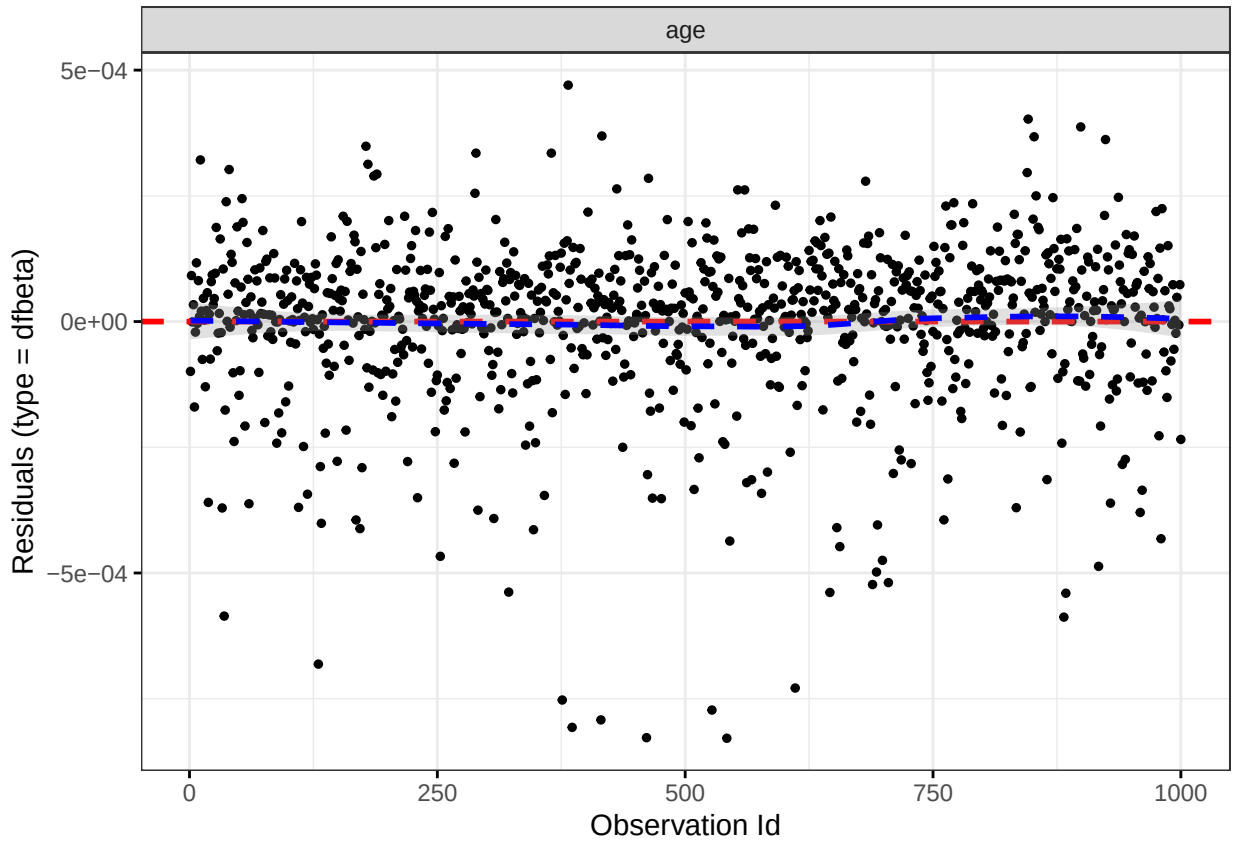


####P-value > 0.05 , there is no evidence for relation between residuals and time , so it is good model .

Generating other diagnostic plots for Cox Proportional Hazards model

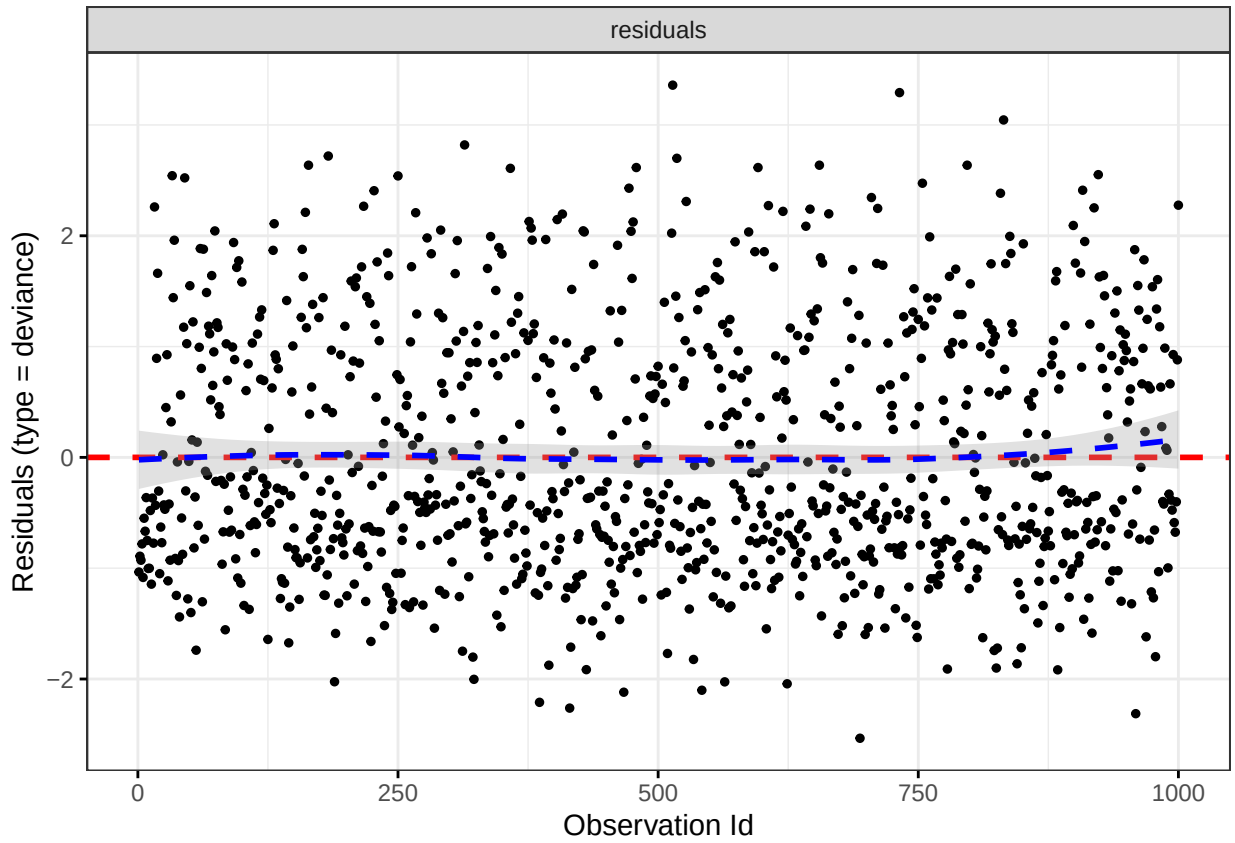
```r
# Define model
res.cox <- coxph(Surv(fu_time, death) ~ age)

# Generate diagnostic plots
 #Plotting the estimated changes in the regression coefficients on deleting each patient
ggcoxdiagnostics(res.cox, type = "dfbeta",
                 linear.predictions = FALSE, ggtheme = theme_bw())
```

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

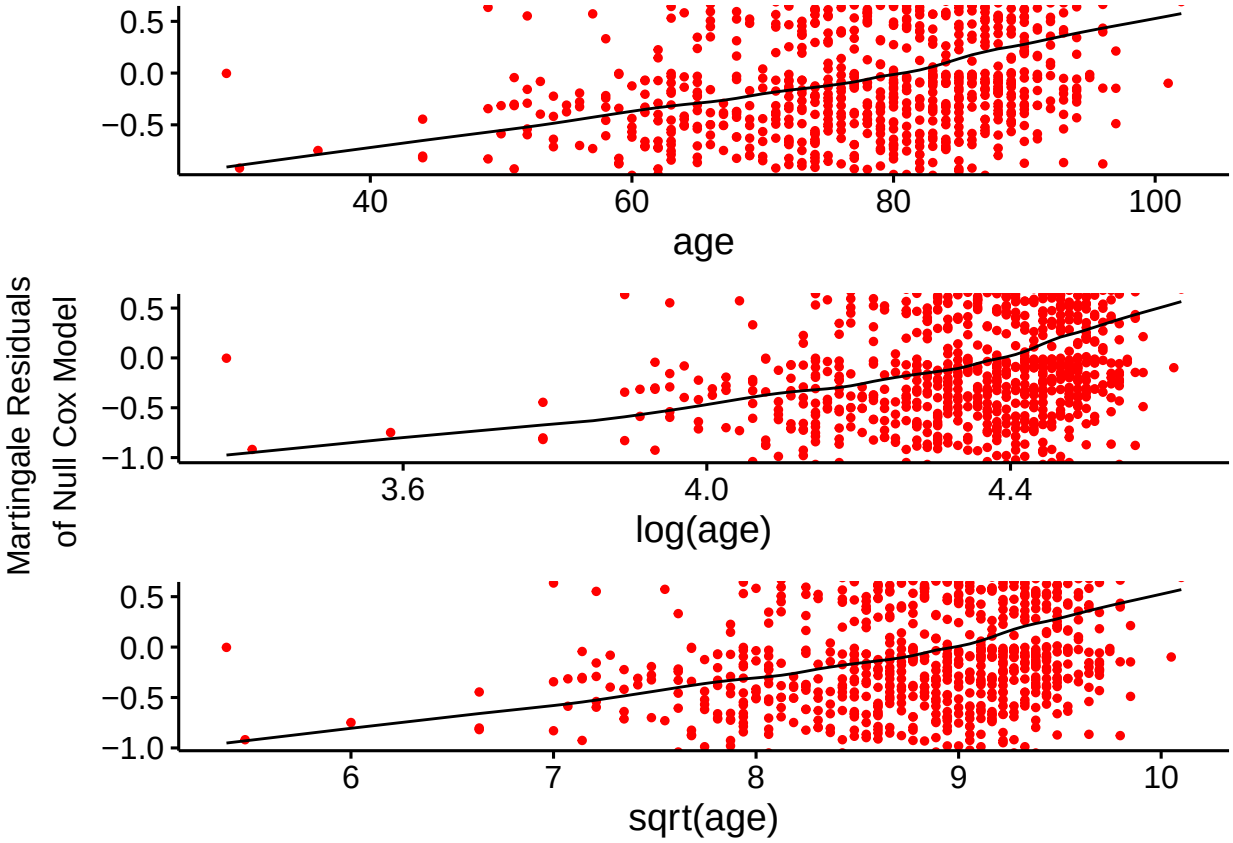



```
 #Plotting deviance residuals
ggcoxdiagnostics(res.cox, type = "deviance",
                 linear.predictions = FALSE, ggtheme = theme_bw())
```

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




```
 # Plotting Martingale residuals
fit <- coxph(Surv(fu_time, death) ~ age + log(age) + sqrt(age))
ggcoxfunctional(fit, data = HF_Data)
```


Martingale Residuals of Null Cox Model
0.5
0.0
−0.5
40
60
80
100
age
0.5
0.0
−0.5
−1.0
3.6
4.0
4.4
log(age)
0.5
0.0
−0.5
−1.0
6
7
8
9
10
sqrt(age)

Backwards elimination to choose predictors for Cox regression :

```
# Run the full model with all of your predictors
cox <- coxph(Surv(fu_time, death) ~ age + as.factor(gender) + as.factor(ethnicgroup) + as.factor(ihd) +
as.factor(valvular_disease) + as.factor(pvd) + as.factor(stroke) + as.factor(copd) +
as.factor(pneumonia) + as.factor(hypertension) + as.factor(renal_disease )+
as.factor(cancer) + as.factor(metastatic_cancer) + as.factor(mental_health) +
as.factor(senile) + los + prior_dnas,data=HF_Data)

summary(cox)
```

```
## Call:
## coxph(formula = Surv(fu_time, death) ~ age + as.factor(gender) +
##     as.factor(ethnicgroup) + as.factor(ihd) + as.factor(valvular_disease) +
##     as.factor(pvd) + as.factor(stroke) + as.factor(copd) + as.factor(pneumonia) +
##     as.factor(hypertension) + as.factor(renal_disease) + as.factor(cancer) +
##     as.factor(metastatic_cancer) + as.factor(mental_health) +
##     as.factor(senile) + los + prior_dnas, data = HF_Data)
##
##   n= 1000, number of events= 492
##
##                                    coef exp(coef)  se(coef)      z Pr(>|z|)
## age                             0.059699  1.061517  0.005789 10.313  < 2e-16 ***
## as.factor(gender)2             -0.198217  0.820192  0.097189 -2.040 0.041400 *
## as.factor(ethnicgroup)2        -0.157243  0.854497  0.352742 -0.446 0.655761
## as.factor(ethnicgroup)3        -0.734118  0.479929  0.416323 -1.763 0.077844 .
## as.factor(ethnicgroup)8        -0.037522  0.963173  0.229929 -0.163 0.870370
## as.factor(ethnicgroup)9         0.408204  1.504114  0.363881  1.122 0.261944
## as.factor(ihd)1                 0.179237  1.196304  0.096547  1.856 0.063387 .
## as.factor(valvular_disease)1    0.185810  1.204193  0.108644  1.710 0.087218 .
## as.factor(pvd)1                 0.023403  1.023679  0.161776  0.145 0.884976
## as.factor(stroke)1             -0.015915  0.984211  0.301721 -0.053 0.957933
## as.factor(copd)1                0.099938  1.105102  0.107259  0.932 0.351471
## as.factor(pneumonia)1           0.323954  1.382583  0.140773  2.301 0.021378 *
## as.factor(hypertension)1       -0.052464  0.948888  0.096222 -0.545 0.585585
## as.factor(renal_disease)1       0.149920  1.161741  0.109718  1.366 0.171812
## as.factor(cancer)1              0.264898  1.303298  0.206096  1.285 0.198683
## as.factor(metastatic_cancer)1  2.196144  8.990277  0.399115  5.503 3.74e-08 ***
## as.factor(mental_health)1      -0.038571  0.962164  0.181910 -0.212 0.832083
## as.factor(senile)1              0.151203  1.163233  0.191318  0.790 0.429337
## los                             0.011774  1.011844  0.003277  3.593 0.000327 ***
## prior_dnas                      0.112348  1.118902  0.042008  2.674 0.007486 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                exp(coef) exp(-coef) lower .95 upper .95
## age                               1.0615     0.9420    1.0495    1.0736
## as.factor(gender)2                0.8202     1.2192    0.6779    0.9923
## as.factor(ethnicgroup)2           0.8545     1.1703    0.4280    1.7059
## as.factor(ethnicgroup)3           0.4799     2.0836    0.2122    1.0853
## as.factor(ethnicgroup)8           0.9632     1.0382    0.6137    1.5115
## as.factor(ethnicgroup)9           1.5041     0.6648    0.7371    3.0691
## as.factor(ihd)1                   1.1963     0.8359    0.9901    1.4455
## as.factor(valvular_disease)1      1.2042     0.8304    0.9732    1.4900
## as.factor(pvd)1                   1.0237     0.9769    0.7455    1.4056
```

```
## as.factor(stroke)1                0.9842     1.0160    0.5448    1.7779
## as.factor(copd)1                  1.1051     0.9049    0.8956    1.3636
## as.factor(pneumonia)1             1.3826     0.7233    1.0492    1.8219
## as.factor(hypertension)1          0.9489     1.0539    0.7858    1.1458
## as.factor(renal_disease)1         1.1617     0.8608    0.9370    1.4405
## as.factor(cancer)1                1.3033     0.7673    0.8702    1.9520
## as.factor(metastatic_cancer)1     8.9903     0.1112    4.1119   19.6562
## as.factor(mental_health)1         0.9622     1.0393    0.6736    1.3743
## as.factor(senile)1                1.1632     0.8597    0.7995    1.6925
## los                               1.0118     0.9883    1.0054    1.0184
## prior_dnas                        1.1189     0.8937    1.0305    1.2149
##
## Concordance= 0.704  (se = 0.012 )
## Likelihood ratio test= 222  on 20 df,   p=<2e-16
## Wald test            = 212.2  on 20 df,   p=<2e-16
## Score (logrank) test = 228.1  on 20 df,   p=<2e-16
```

```
# Run the model with only significant predictors
cox1 <- coxph(Surv(fu_time, death) ~ age + as.factor(gender) + as.factor(ihd) +
as.factor(pneumonia) + as.factor(metastatic_cancer) + los + prior_dnas,data=HF_Data)
summary(cox1)
```

```
## Call:
## coxph(formula = Surv(fu_time, death) ~ age + as.factor(gender) +
##     as.factor(ihd) + as.factor(pneumonia) + as.factor(metastatic_cancer) +
##     los + prior_dnas, data = HF_Data)
##
##   n= 1000, number of events= 492
##
##                                    coef exp(coef)  se(coef)      z Pr(>|z|)
## age                            0.061954  1.063914  0.005590 11.082  < 2e-16 ***
## as.factor(gender)2            -0.247746  0.780558  0.094699 -2.616  0.00889 **
## as.factor(ihd)1                0.185005  1.203224  0.093495  1.979  0.04784 *
## as.factor(pneumonia)1          0.343260  1.409535  0.137329  2.500  0.01244 *
## as.factor(metastatic_cancer)1  2.428248 11.339001  0.364561  6.661 2.72e-11 ***
## los                            0.012648  1.012729  0.003202  3.950 7.82e-05 ***
## prior_dnas                     0.115996  1.122991  0.036755  3.156  0.00160 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                exp(coef) exp(-coef) lower .95 upper .95
## age                               1.0639    0.93993    1.0523    1.0756
## as.factor(gender)2                0.7806    1.28114    0.6483    0.9398
## as.factor(ihd)1                   1.2032    0.83110    1.0018    1.4452
## as.factor(pneumonia)1             1.4095    0.70945    1.0769    1.8449
## as.factor(metastatic_cancer)1    11.3390    0.08819    5.5496   23.1681
## los                               1.0127    0.98743    1.0064    1.0191
## prior_dnas                        1.1230    0.89048    1.0449    1.2069
##
## Concordance= 0.699  (se = 0.012 )
## Likelihood ratio test= 208.6  on 7 df,   p=<2e-16
## Wald test            = 200.9  on 7 df,   p=<2e-16
## Score (logrank) test = 212.9  on 7 df,   p=<2e-16
```

For each increase of 1 year in age ,the hazard of death at any given time point goes up by 6% , while other variables are adjusted .

Older people have a higher hazard of death than young people .

P-value is very tiny (<2e-16) , Age is highly statisticaly significant.

Femals(gender=2) have hazard ratio = 0.78 compared to males , with 95% CI between 0.65 to 0.93 ,

with p very tiny (<2e-16) , So there is clear evidence of association between females and lower hazard of mortality while other variables are adjusted .

patients with ischaemic heart disease (ihd=1) have hazard ratio = 1.2 compared to patients without ischaemic heart disease ,

patients with ischaemic heart disease have higher Hazard by 20% than patients without ischaemic heart disease,

p = 0.04 , So there is evidence of association between ihd and death in this sample while other variables are adjusted.

patients with pneumonia (pneumonia=1) have hazard ratio = 1.4 compared to patients without pneumonia,

patients with pneumonia have higher Hazard by 40% than patients without pneumonia,

p = 0.012 , So there is evidence of association between pneumonia and death in this sample while other variables are adjusted .

patients with metastatic cancer (metastatic_cancer=1) have hazard ratio = 11.33 compared to patients without metastatic cancer,

patients with metastatic cancer have higher Hazard by 33% than patients without metastatic cancer,

p value is very tiny , So there is strong evidence of association between metastatic cancer and death in this sample while other variables are adjusted .

For each night increase of hospital stay ,the hazard of death at any given time point goes up by 1% , while other variables are adjusted .

P-value is very tiny (<2e-16) , it is highly statisticaly significant.

For each appointment missed , the hazard increase by 12% .

Test proportionality assumption on these predictors

```
temp <- cox.zph(cox1)
print(temp)
```

```
##                                chisq df      p
## age                           0.5207  1 0.4705
## as.factor(gender)             1.7044  1 0.1917
## as.factor(ihd)                7.6788  1 0.0056
## as.factor(pneumonia)          1.0371  1 0.3085
## as.factor(metastatic_cancer)  0.4083  1 0.5228
## los                           0.1518  1 0.6969
## prior_dnas                    0.0246  1 0.8754
## GLOBAL                       11.0225  7 0.1376
```