

Survival analysis of ovarian cancer data

Rania Ahmed

November 22, 2019

Source: built in data in R studio

Discription:

futime: survival or censoring time

fustat: censoring status

age: in years

resid.ds: residual disease present (1=noprogression,2=progression)

rx: treatment group(1= placebo, 2= new drug)

ecog.ps: ECOG performance status (1 is better, see reference)

```
library(survival)
library(survminer)

## Loading required package: ggplot2
## Loading required package: ggpubr
## Loading required package: magrittr

library(dplyr)

##
## Attaching package: 'dplyr'

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

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

library(finalfit)
library(moments)
library(ggplot2)
library(knitr)

attach(ovarian)
New_ovarian <- mutate(ovarian, censoring_status=ifelse(fustat=="1",
```

```

"death", "nodeath"),
Treatment=ifelse(ovarian$rx=="2", "newdrug", "placebo"),
progression=ifelse(ovarian$resid.ds=="1", "no
progression", "progression"),
dummy_ecog=ifelse(ecog.ps=="1", "better", "see ref.))
summary(New_ovarian)

##      futime      fustat      age      resid.ds
## Min.   : 59.0   Min.   :0.0000   Min.   :38.89   Min.   :1.000
## 1st Qu.: 368.0   1st Qu.:0.0000   1st Qu.:50.17   1st Qu.:1.000
## Median : 476.0   Median :0.0000   Median :56.85   Median :2.000
## Mean    : 599.5   Mean    :0.4615   Mean    :56.17   Mean    :1.577
## 3rd Qu.: 794.8   3rd Qu.:1.0000   3rd Qu.:62.38   3rd Qu.:2.000
## Max.    :1227.0   Max.    :1.0000   Max.    :74.50   Max.    :2.000
##      rx      ecog.ps      censoring_status      Treatment
## Min.    :1.0   Min.    :1.000   Length:26   Length:26
## 1st Qu.:1.0   1st Qu.:1.000   Class :character   Class :character
## Median :1.5   Median :1.000   Mode  :character   Mode  :character
## Mean    :1.5   Mean    :1.462
## 3rd Qu.:2.0   3rd Qu.:2.000
## Max.    :2.0   Max.    :2.000
## progression      dummy_ecog
## Length:26        Length:26
## Class :character   Class :character
## Mode  :character   Mode  :character
##
##
##

```

Analysis of variance:

1- censoring_status:

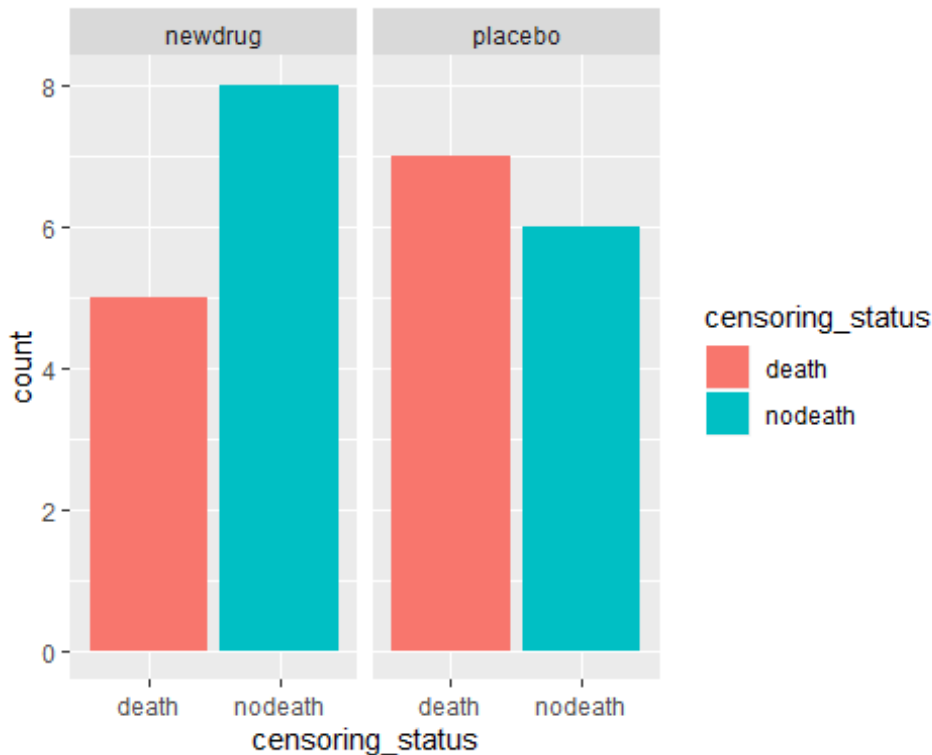
```

New_ovarian$censoring_status <- as.factor(New_ovarian$censoring_status)
New_ovarian$Treatment <- as.factor(New_ovarian$Treatment)
censoring_coun <- as.data.frame(New_ovarian %>%
summary_factorlist("censoring_status", "Treatment", p=F,
add_dependent_label = T, cont_cut = 1))
censoring_coun

##   Dependent: censoring_status      death  nodeath
## 1              Treatment newdrug 5 (38.5) 8 (61.5)
## 2              placebo 7 (53.8) 6 (46.2)

ggplot(New_ovarian, aes(censoring_status, fill=censoring_status))+geom_bar(
r(stat="count"))+
facet_wrap(~Treatment)+theme_get()

```



```
chisq.test(table(New_ovarian$censoring_status, New_ovarian$Treatment))

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(New_ovarian$censoring_status, New_ovarian$Treatment)
## X-squared = 0.15476, df = 1, p-value = 0.694
```

The number of dead patients(death) taking new drug is 5 while whose taking placebo is 7. The number of survived patient taking new drug is 8 while whose taking placebo is 6 patients.

In chi-square test, p- value is more than 0.05 so there is no significant difference between drugs and censoring status thus we will fail to reject null hypothesis.

2- Age:

```
Age <- group_by(New_ovarian, Treatment)%>%
summarise(count=n(),mean=mean(age),sd=sd(age), var=var(age))
Age

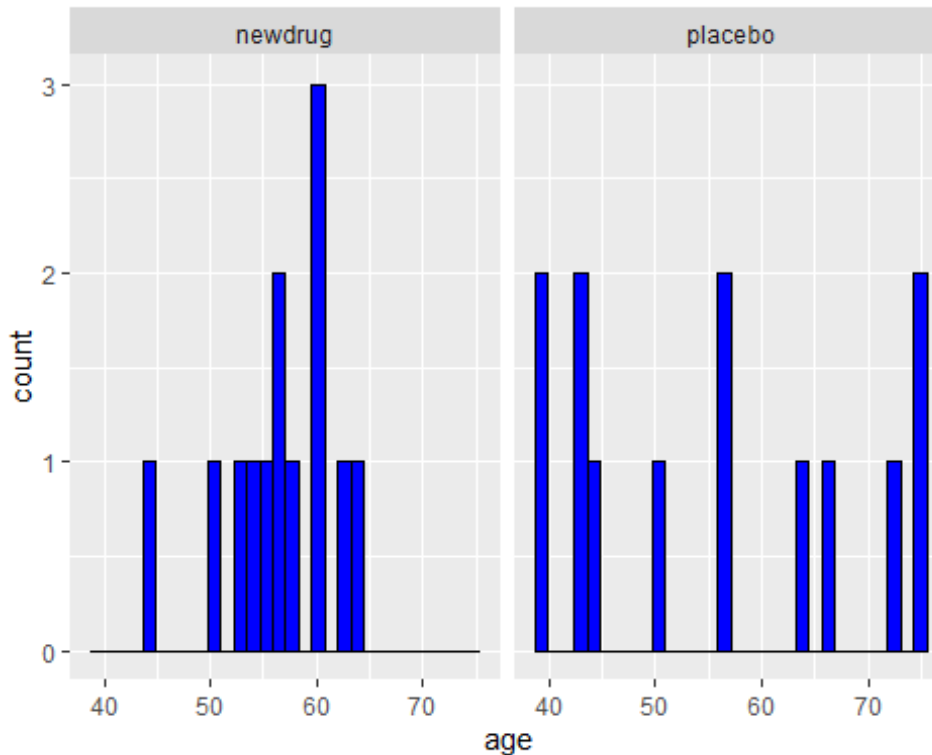
## # A tibble: 2 x 5
##   Treatment count  mean    sd   var
##   <fct>      <int> <dbl> <dbl> <dbl>
## 1 newdrug      13  56.6  5.42  29.4
## 2 placebo      13  55.7 13.5  183.
```

```
summary(New_ovarian$age)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  38.89   50.17   56.85   56.17   62.38   74.50

ggplot(New_ovarian, aes(age, fill= age))+geom_histogram(stat = "bin",
color="black", fill="blue")+
facet_wrap(~Treatment)+theme_get()

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
#check normality:

skewness(New_ovarian$age)

## [1] 0.04340272

kurtosis(New_ovarian$age)

## [1] 2.343786

shapiro.test(New_ovarian$age)

##
## Shapiro-Wilk normality test
##
## data:  New_ovarian$age
## W = 0.96169, p-value = 0.4259
```

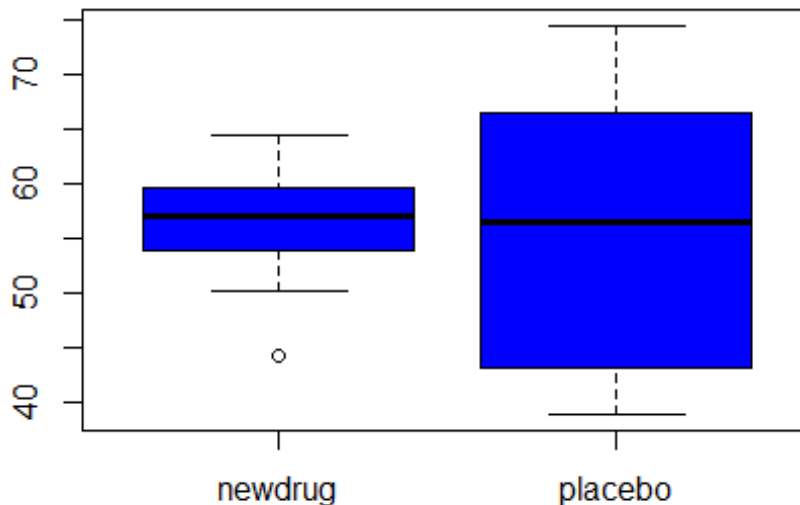
```

t.test(age~Treatment, mu=0, alternative="two.sided",
var.equal=F,conf.level= 0.95, data = New_ovarian)

##
##  Welch Two Sample t-test
##
## data:  age by Treatment
## t = 0.21443, df = 15.76, p-value = 0.833
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.707806  9.440160
## sample estimates:
## mean in group newdrug mean in group placebo
##           56.59853           55.73235

boxplot(age~ Treatment, col="blue", data = New_ovarian)

```



The mean of the age in sample is 56.17 and the range of age is from 38.89 to 74.5 years. the mean of the age which patients taking new drug is 56.6 and standard deviation is 5.4 while the mean of the age in patients taking placebo is 55.73 and standard deviation is 13.52.

Check normality:

in numerical test: sample is lower skewed and in acceptable range of kurtosis.

in statistical test: the sample is normally distributed. P-value is more than 0.05 so data is normally distributed.

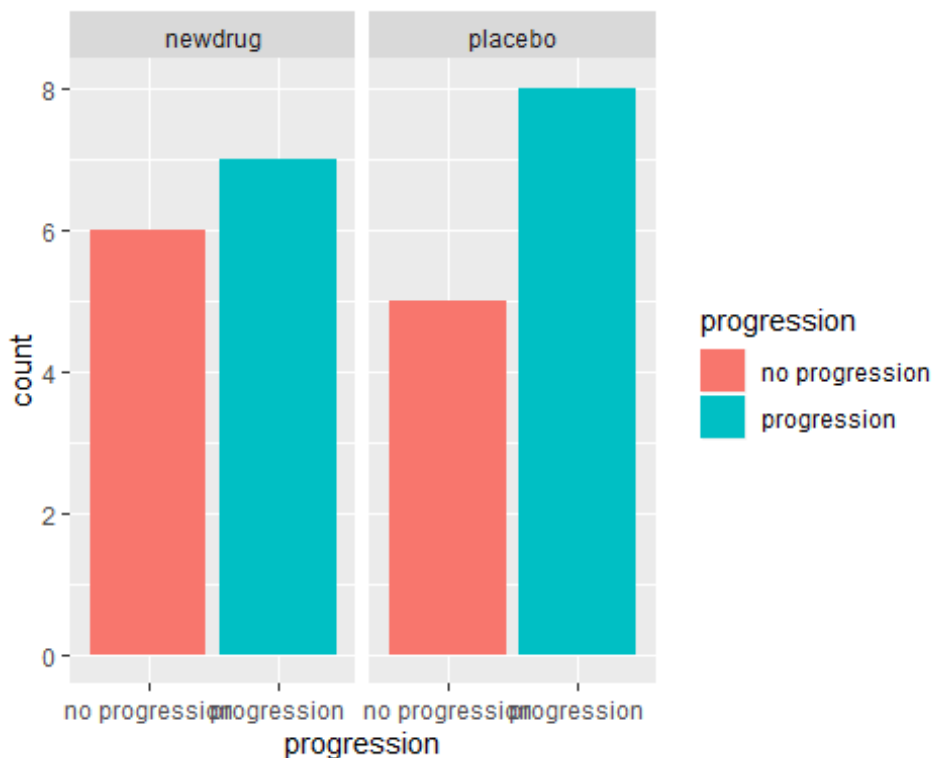
In t.test: p value is more than 0.05 so there is no significant difference between drugs and age so we fail to reject null hypothesis.

3- Progression:

```
progress <- as.data.frame(New_ovarian %>%
summary_factorlist("progression", "Treatment", p=F, add_dependent_label
= T, cont_cut = 1))
progress

##   Dependent: progression      no progression progression
## 1      Treatment newdrug         6 (46.2)      7 (53.8)
## 2              placebo         5 (38.5)      8 (61.5)

ggplot(New_ovarian, aes(progression,
fill=progression))+geom_bar(stat="count")+facet_wrap(~Treatment)+theme_
get()
```



```
chisq.test(table(New_ovarian$progression, New_ovarian$Treatment))

##
## Pearson's Chi-squared test with Yates' continuity correction
##
```

```
## data: table(New_ovarian$progression, New_ovarian$Treatment)
## X-squared = 0, df = 1, p-value = 1
```

The number of patient who having no progression for disease taking new drug is 6 and who taking placebo is 5. The number of patients who having progression of disease(worsen the case) taking new drug is 7 and who taking placebo is 8.

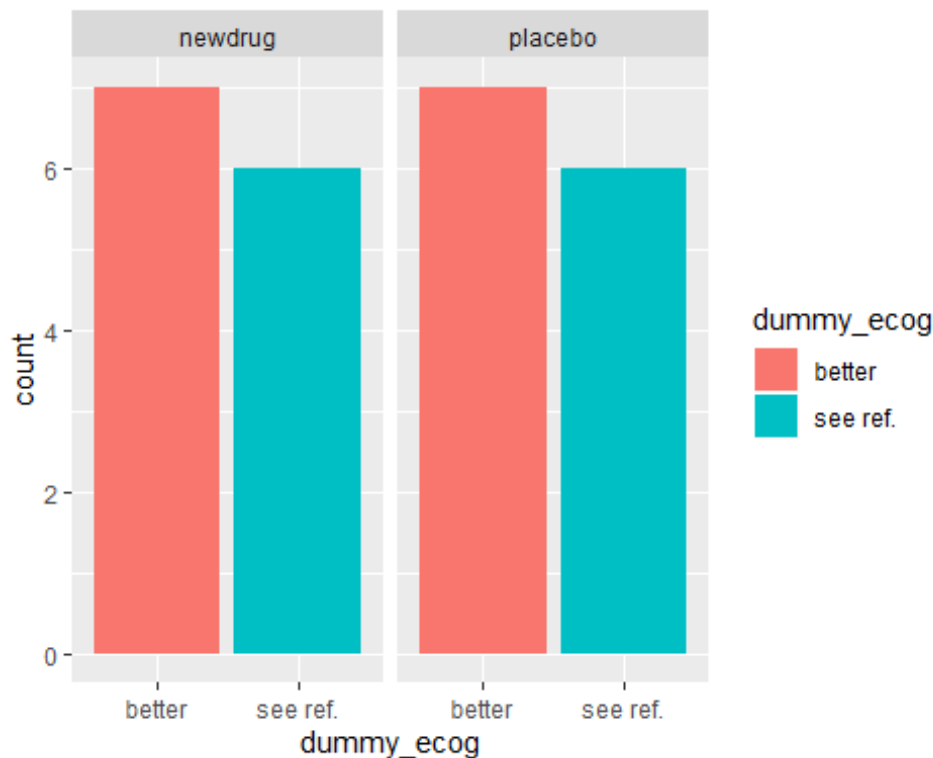
In chi square test : P value is more than 0.05 so there is no significant difference between drugs and progression . so we fail to reject null hypothesis.

4- ECOG performance status:

```
ecog <- as.data.frame(New_ovarian %>%
summary_factorlist("dummy_ecog","Treatment", p=F, add_dependent_label =
T, cont_cut = 1))
ecog
```

```
##   Dependent: dummy_ecog          better see ref.
## 1           Treatment newdrug 7 (53.8) 6 (46.2)
## 2           Treatment placebo 7 (53.8) 6 (46.2)
```

```
ggplot(New_ovarian, aes(dummy_ecog, fill=dummy_ecog))+geom_bar(stat =
"count")+
facet_wrap(~Treatment)+theme_get()
```



```
chisq.test(table(New_ovarian$dummy_ecog,New_ovarian$Treatment))
```

```
##
## Pearson's Chi-squared test
##
## data: table(New_ovarian$dummy_ecog, New_ovarian$Treatment)
## X-squared = 0, df = 1, p-value = 1
```

The number of patient with ECOG performance status showing better taking new drug is 7 while who showing other performance is 6 patients.

In ch square test: p value is more than 0.05 so there is no significant between drugs and ECOG status . so we fail to reject null hypothesis.

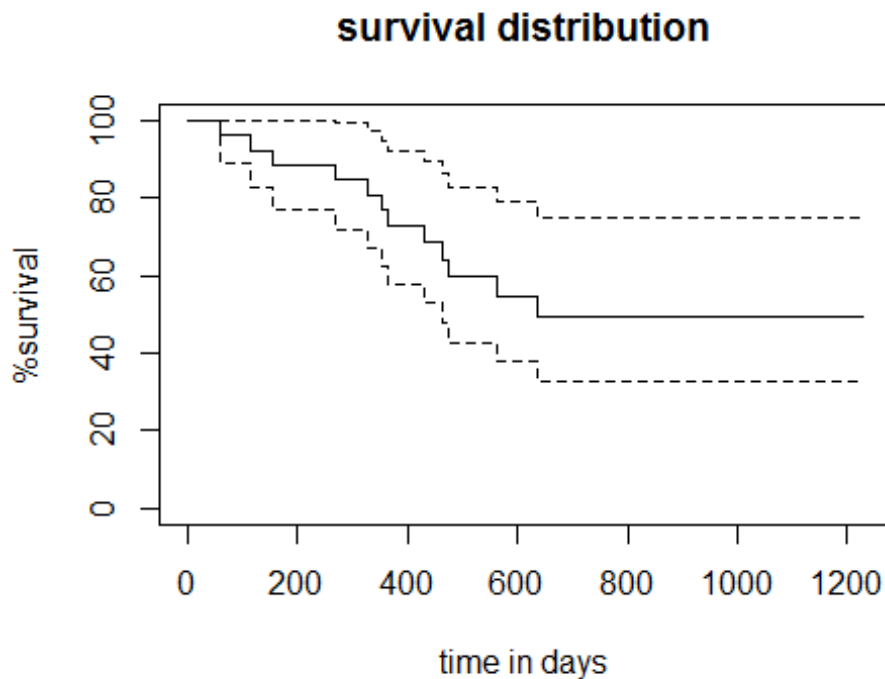
Kaplan_meier estimator:

```
survobj <- with(New_ovarian, Surv(futime,fustat))
fit0 <- survfit(survobj~1, data = New_ovarian)
summary(fit0)
```

```
## Call: survfit(formula = survobj ~ 1, data = New_ovarian)
##
```

##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
##	59	26	1	0.962	0.0377	0.890	1.000
##	115	25	1	0.923	0.0523	0.826	1.000
##	156	24	1	0.885	0.0627	0.770	1.000
##	268	23	1	0.846	0.0708	0.718	0.997
##	329	22	1	0.808	0.0773	0.670	0.974
##	353	21	1	0.769	0.0826	0.623	0.949
##	365	20	1	0.731	0.0870	0.579	0.923
##	431	17	1	0.688	0.0919	0.529	0.894
##	464	15	1	0.642	0.0965	0.478	0.862
##	475	14	1	0.596	0.0999	0.429	0.828
##	563	12	1	0.546	0.1032	0.377	0.791
##	638	11	1	0.497	0.1051	0.328	0.752

```
plot(fit0, xlab = "time in days", ylab = "%survival", yscale = 100,
      main="survival distribution")
```

The number of event (dead patient)= 12 patients begin at time 59 days till 638 days.

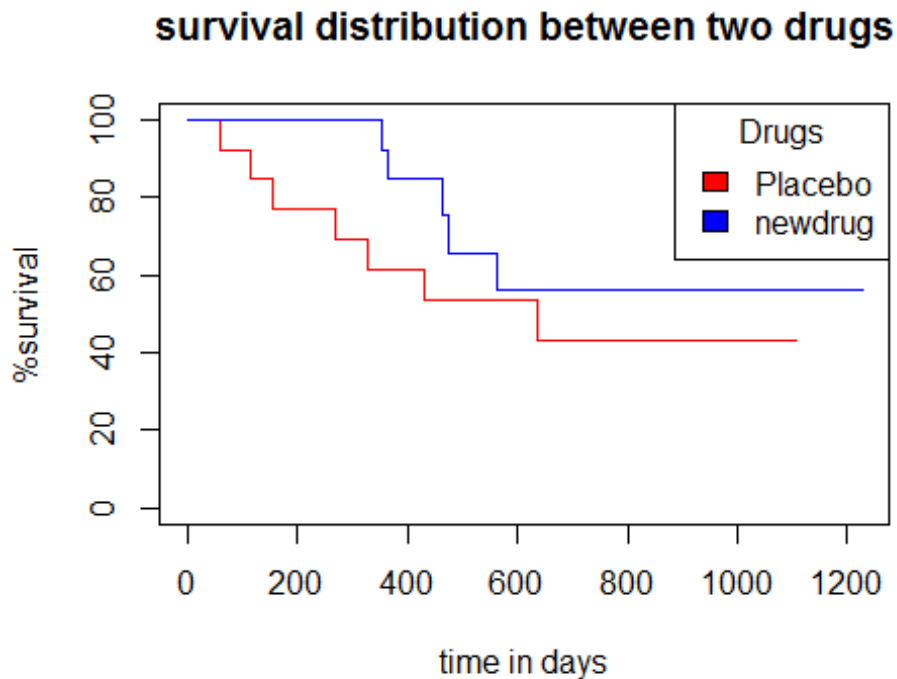
let's compare the survival distribution between two drugs(placebo and new drug):

```
fitdrug <- survfit(survobj~rx, data = New_ovarian)
summary(fitdrug)
```

```
## Call: survfit(formula = survobj ~ rx, data = New_ovarian)
##
##               rx=1
##   time  n.risk  n.event  survival  std.err  lower 95% CI  upper 95% CI
##    59      13        1    0.923   0.0739    0.789    1.000
##   115      12        1    0.846   0.1001    0.671    1.000
##   156      11        1    0.769   0.1169    0.571    1.000
##   268      10        1    0.692   0.1280    0.482    0.995
##   329       9        1    0.615   0.1349    0.400    0.946
##   431       8        1    0.538   0.1383    0.326    0.891
##   638       5        1    0.431   0.1467    0.221    0.840
##
##               rx=2
##   time  n.risk  n.event  survival  std.err  lower 95% CI  upper 95% CI
##   353      13        1    0.923   0.0739    0.789    1.000
##   365      12        1    0.846   0.1001    0.671    1.000
##   464       9        1    0.752   0.1256    0.542    1.000
```

```
##    475      8      1    0.658 0.1407      0.433      1.000
##    563      7      1    0.564 0.1488      0.336      0.946
```

```
plot(fitdrug, xlab = "time in days", ylab = "%survival", yscale = 100,
col = c("red", "blue"), main="survival distribution between two drugs")
legend("topright", title = "Drugs", c("Placebo", "newdrug"), fill =
c("red", "blue"))
```



```
survdifff(survobj~rx, data = New_ovarian)

## Call:
## survdifff(formula = survobj ~ rx, data = New_ovarian)
##
##      N Observed Expected (O-E)^2/E (O-E)^2/V
## rx=1 13        7      5.23    0.596    1.06
## rx=2 13        5      6.77    0.461    1.06
##
## Chisq= 1.1 on 1 degrees of freedom, p= 0.3
```

In survival distribution comparing between two drugs, it is found that the patients taking new drug survived more than taking placebo.

Number of dead patient with new drug=5 while taking placebo =7 patients

p-value is more than 0.05 means that median time of survival between two drugs is not significant.

let's compare between survival distribution and age:

```
fitage <- survfit(survobj~age, data = New_ovarian)
summary(fitage)

## Call: survfit(formula = survobj ~ age, data = New_ovarian)
##
##               age=38.8932
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##               age=39.2712
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##               age=43.1233
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##               age=43.137
##      time      n.risk      n.event      survival      std.err
##      329           1           1           0           NaN
## lower 95% CI upper 95% CI
##      NA           NA
##
##               age=44.2055
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##               age=44.6
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##               age=50.1096
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##               age=50.3397
##      time      n.risk      n.event      survival      std.err
##      431           1           1           0           NaN
## lower 95% CI upper 95% CI
##      NA           NA
##
##               age=53.3644
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##               age=53.9068
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##               age=55.1781
##      time      n.risk      n.event      survival      std.err
##      563           1           1           0           NaN
## lower 95% CI upper 95% CI
##      NA           NA
##
##               age=56.4301
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

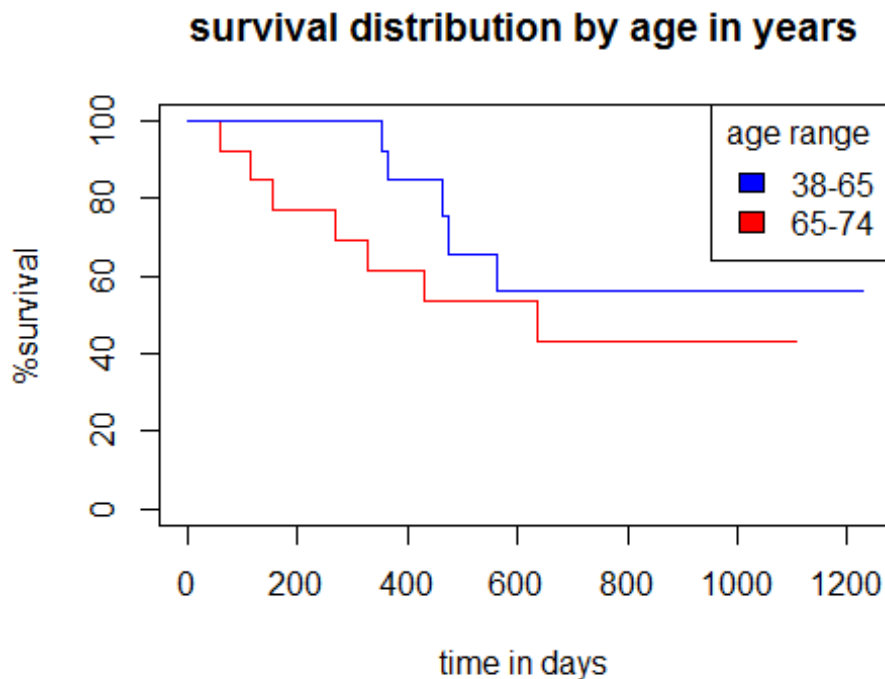
```

##
##           age=56.7562
##      time      n.risk      n.event      survival      std.err
##      638          1          1          0          NaN
## lower 95% CI upper 95% CI
##      NA          NA
##
##           age=56.937
##      time      n.risk      n.event      survival      std.err
##      464          1          1          0          NaN
## lower 95% CI upper 95% CI
##      NA          NA
##
##           age=57.0521
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##           age=58.3096
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##           age=59.589
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##           age=59.6301
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##           age=59.8548
##      time      n.risk      n.event      survival      std.err
##      475          1          1          0          NaN
## lower 95% CI upper 95% CI
##      NA          NA
##
##           age=63.2192
##      time      n.risk      n.event      survival      std.err
##      353          1          1          0          NaN
## lower 95% CI upper 95% CI
##      NA          NA
##
##           age=64.1753
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##           age=64.4247
##      time      n.risk      n.event      survival      std.err
##      365          1          1          0          NaN
## lower 95% CI upper 95% CI
##      NA          NA
##
##           age=66.4658
##      time      n.risk      n.event      survival      std.err
##      156          1          1          0          NaN
## lower 95% CI upper 95% CI

```

```
##           NA           NA
##
##           age=72.3315
##           time      n.risk      n.event      survival      std.err
##           59         1         1         0         NaN
## lower 95% CI upper 95% CI
##           NA           NA
##
##           age=74.4932
##           time      n.risk      n.event      survival      std.err
##           115        1         1         0         NaN
## lower 95% CI upper 95% CI
##           NA           NA
##
##           age=74.5041
##           time      n.risk      n.event      survival      std.err
##           268        1         1         0         NaN
## lower 95% CI upper 95% CI
##           NA           NA

plot(fitdrug, xlab = "time in days", ylab = "%survival", yscale =
100,col=c("red","blue"), main="survival distribution by age in years")
legend("topright", title="age range", c("38-65","65-74"),
fill=c("blue","red"))
```



```
survdifff(survobj~age, data = New_ovarian)
```

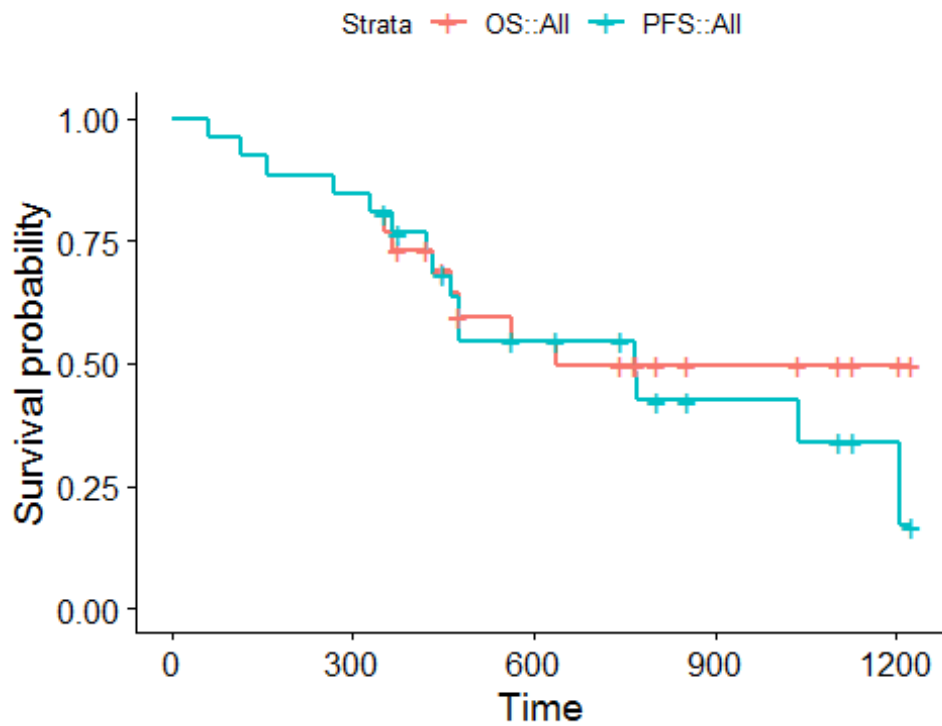
```
## Call:
## survdiff(formula = survobj ~ age, data = New_ovarian)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## age=38.8932 1          0   0.6778    0.678    0.722
## age=39.2712 1          0   0.6778    0.678    0.722
## age=43.1233 1          0   0.6778    0.678    0.722
## age=43.137  1          1   0.2091    2.992    3.123
## age=44.2055 1          0   0.6778    0.678    0.722
## age=44.6    1          0   0.6778    0.678    0.722
## age=50.1096 1          0   0.6778    0.678    0.722
## age=50.3397 1          1   0.3655    1.101    1.155
## age=53.3644 1          0   0.3067    0.307    0.321
## age=53.9068 1          0   0.6778    0.678    0.722
## age=55.1781 1          1   0.5869    0.291    0.308
## age=56.4301 1          0   0.3655    0.366    0.383
## age=56.7562 1          1   0.6778    0.153    0.163
## age=56.937  1          1   0.4322    0.746    0.785
## age=57.0521 1          0   0.6778    0.678    0.722
## age=58.3096 1          0   0.3067    0.307    0.321
## age=59.589  1          0   0.6778    0.678    0.722
## age=59.6301 1          0   0.6778    0.678    0.722
## age=59.8548 1          1   0.5036    0.489    0.517
## age=63.2192 1          1   0.2567    2.153    2.249
## age=64.1753 1          0   0.5036    0.504    0.532
## age=64.4247 1          1   0.3067    1.567    1.640
## age=66.4658 1          1   0.1201    6.445    6.714
## age=72.3315 1          1   0.0385   24.038   25.000
## age=74.4932 1          1   0.0785   10.824   11.266
## age=74.5041 1          1   0.1636    4.276    4.459
##
##  Chisq= 71.9  on 25 degrees of freedom, p= 2e-06
```

from the curve we found that as age increase the number of event increase by time.
from age range 38 -65 years nu. of dead patients is 7 pts while from 65-75 years no.
of event is 4pts.

so there is significant difference between age and median time of survival.

Progression free survival:

```
os1 <- survfit(Surv(futime, fustat)~1, data= New_ovarian)
pfs1 <- survfit(Surv(futime, resid.ds)~1, data = New_ovarian)
fit <- list(OS = os1, PFS= pfs1)
ggsurvplot_combine(fit,New_ovarian)
```



OS(overall survival) shows us less survival probability than PFS(progression free survival).

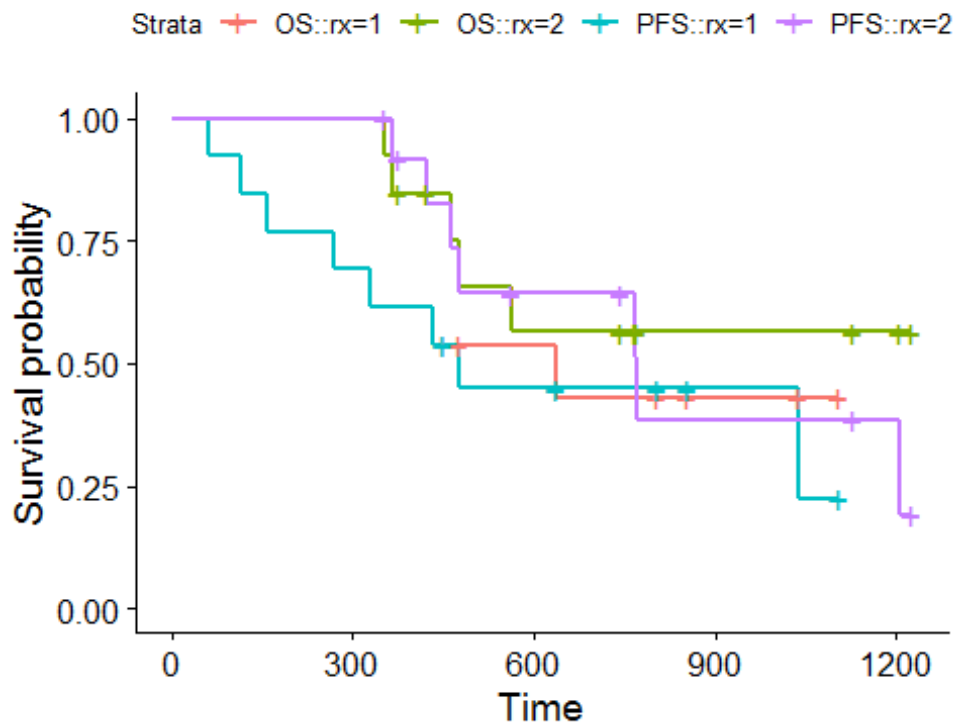
let's compare between PFS with treatments:

```
os2 <- survfit(Surv(futime, fustat)~rx, data = New_ovarian)

pfs2 <- survfit(Surv(futime, resid.ds)~rx, data = New_ovarian)

fit2 <- list(OS=os2,PFS=pfs2)

ggsurvplot_combine(fit2, New_ovarian)
```



In over all survival the new drug shows less survival probability than placebo .. while in progression free survival placebo shows less survival probability than new drug. this is mean patients taking new drugs, progression of the death occurs less than whose taking placebo.

cox regression:

to reporting hazard ratio

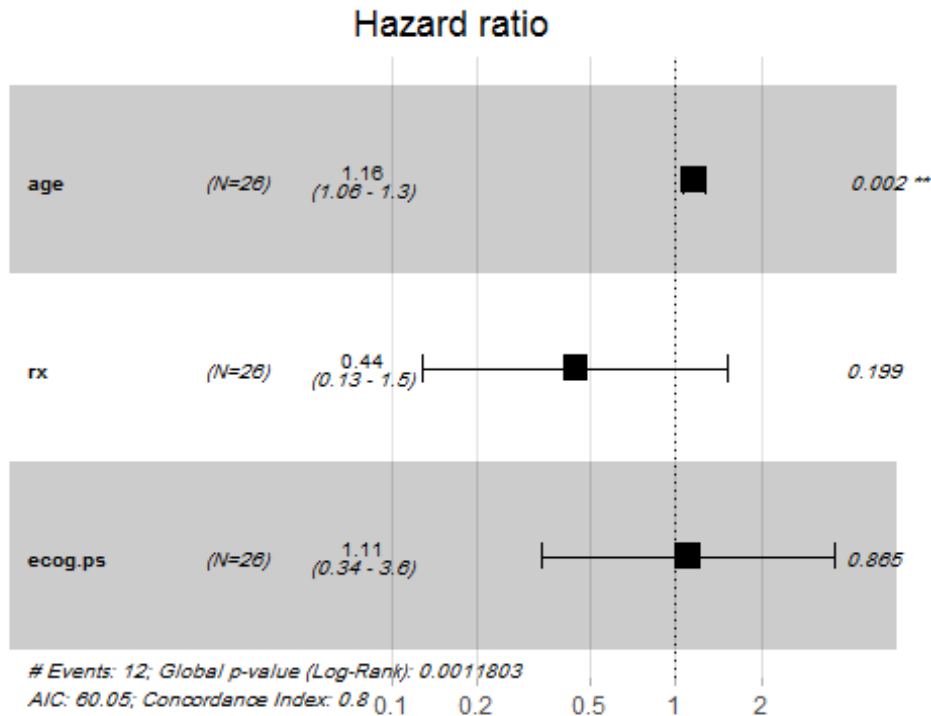
```
mod1 <- coxph(survobj~age+rx+ecog.ps, data = New_ovarian)
summary(mod1)
```

```
## Call:
## coxph(formula = survobj ~ age + rx + ecog.ps, data = New_ovarian)
##
## n= 26, number of events= 12
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## age      0.1470   1.1583  0.0463  3.175  0.0015 **
## rx      -0.8146   0.4428  0.6342 -1.285  0.1990
## ecog.ps  0.1032   1.1087  0.6064  0.170  0.8649
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## age          1.1583      0.8633    1.0579    1.268
```



```
## rx          0.4428    2.2582    0.1278    1.535
## ecog.ps     1.1087    0.9020    0.3378    3.639
##
## Concordance= 0.798 (se = 0.078 )
## Likelihood ratio test= 15.92 on 3 df,  p=0.001
## Wald test          = 13.32 on 3 df,  p=0.004
## Score (logrank) test = 18.7 on 3 df,  p=3e-04

ggforest(mod1, data = New_ovarian)
```



1-In age:

- i- as age increases the hazard ratio(risk of death) increases.
- ii-the effect size of age as covariate(hazard ratio) is 1.15
- iii-p-value is less than 0.05 so there is significant difference.

2-In treatment:

- i- taking new drug is lower hazard(risk of death) by 44.28%
- ii- the effect size of treatment as covariate= 0.55
- iii- p-value is more than 0.05 so there is no significant difference.

3-In ecog.ps:

- i- patients showing less performance increase hazard(risk of death)
- ii- the effect size of ecog as covariate is 1.10
- iii- p-value is more than 0.05 so there is no significance.

N.B: the global significant(likelihood ratio test)(0.001) is less than 0.05

Predication Risk model:

```
mod2 <- coxph(survobj~age, data = New_ovarian)
summary(mod2)

## Call:
## coxph(formula = survobj ~ age, data = New_ovarian)
##
##    n= 26, number of events= 12
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age 0.16162    1.17541  0.04974  3.249  0.00116 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## age      1.175      0.8508    1.066    1.296
##
## Concordance= 0.784  (se = 0.083 )
## Likelihood ratio test= 14.29  on 1 df,  p=2e-04
## Wald test               = 10.56  on 1 df,  p=0.001
## Score (logrank) test = 12.26  on 1 df,  p=5e-04
```

The age is significant with hazard of death considering other value of the data the p value of likelihood ratio test is decreased from 0.001 to 2e-04 and being highly significant which improve that the risk of death increases by age increases especially after 50 years.

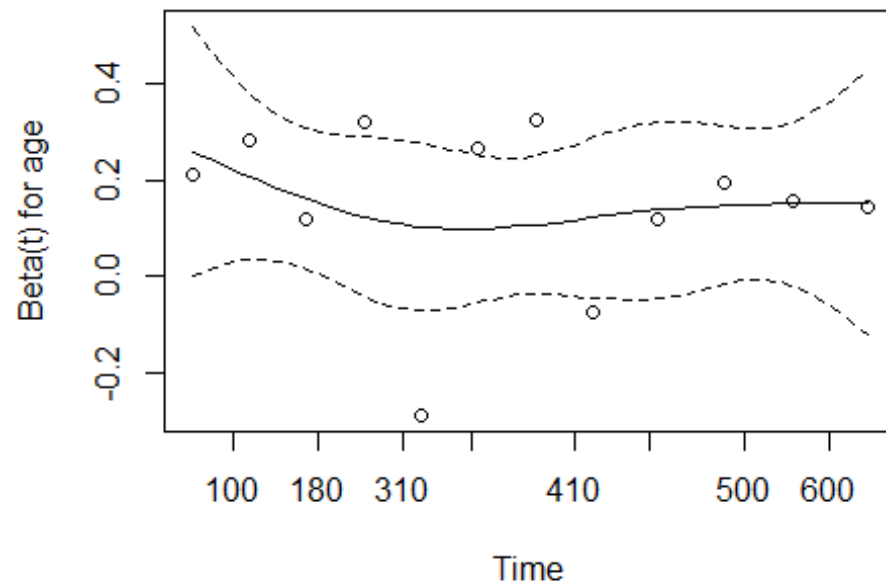
evaluation of propotional hazard asumption:

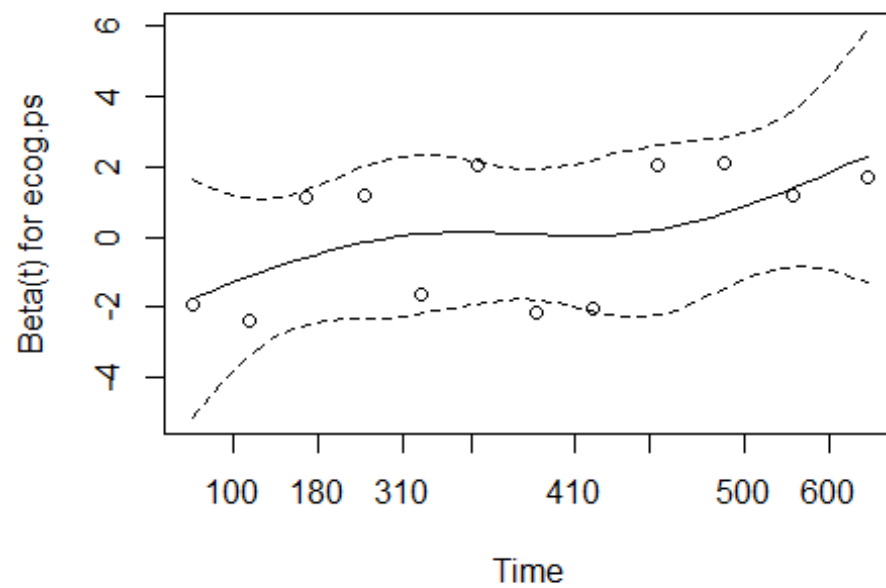
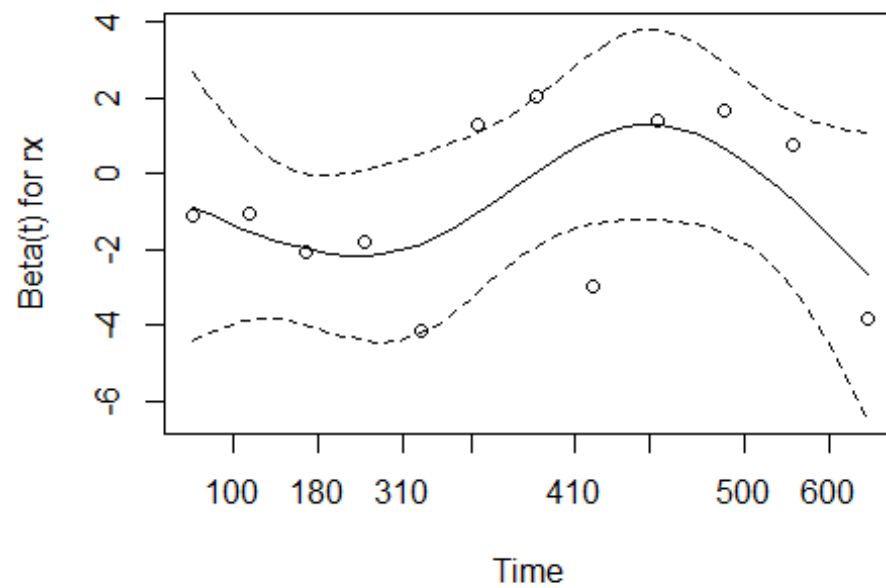
```
cox.zph(mod1)

##              rho chisq      p
## age      -0.111 0.162 0.687
```

```
## rx      0.170 0.318 0.573
## ecog.ps 0.515 2.427 0.119
## GLOBAL   NA 3.158 0.368

plot(cox.zph(mod1), data= New_ovarian)
```





P-value of all variances is non significant and also global Pvalue which means that all variances are non significant with time. All variances don't change by time . our modul is a good modul.