Quiz1-Solution

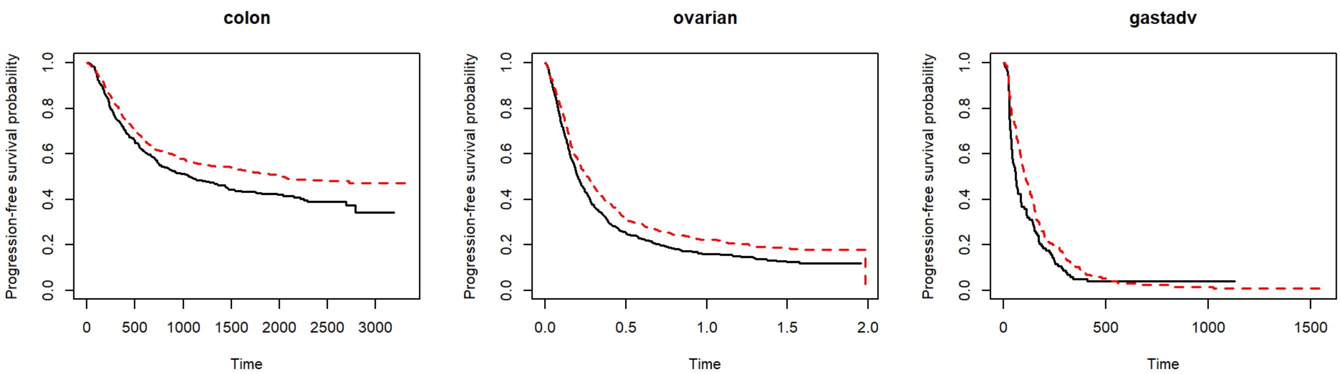
(a)

Plot the Kaplan-Meier curves for PFT by treatment group. Compare the survival distributions of two groups using the log-rank test.

```
#load packages
library(survRM2)
```

```
## Loading required package: survival
```

```
library(coin)

#load datasets
dataname=c("colon","ovarian","gastadv")
datalist=list()
for(i in 1:3){
      datalist[[i]]=read.csv(paste(dataname[i],".txt",sep=""))
  datalist[[i]]$group=as.factor(datalist[[i]]$group)
}

par(mfrow=c(1,3))
for(i in 1:3){
  mydata=datalist[[i]]
  # Get the survival curve for PFT
  km_fit<-survfit(Surv(PFT,status_PF)~group,data=mydata)
  plot(km_fit,xlab="Time",ylab="Progression-free survival probability",
       main=dataname[i],col=c(1,2),lwd=1.5,lty=c(1,2))
}
```



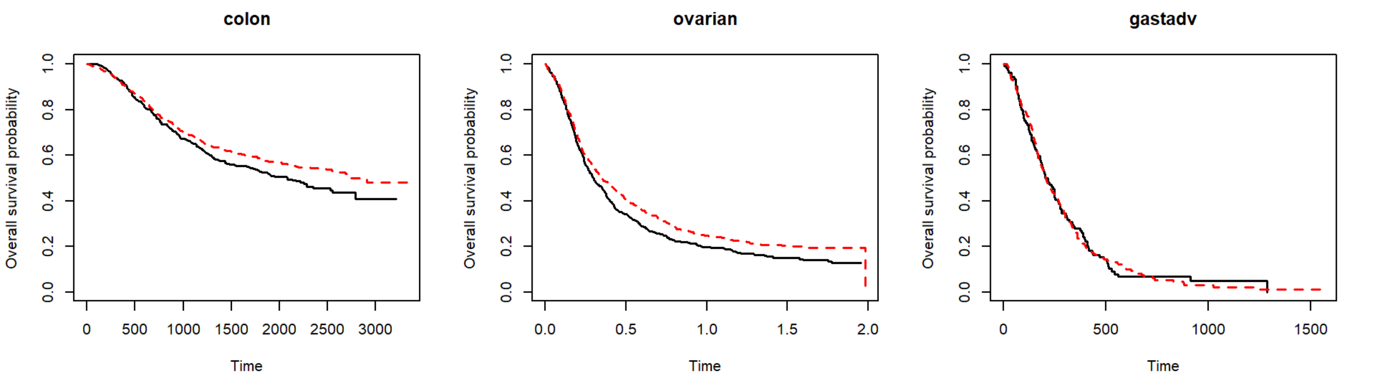

```
for(i in 1:3){
  mydata=datalist[[i]]
  PFS_logrank<-logrank_test(Surv(PFT,status_PF)~group,
                          data=mydata,
                          type="logrank",alternative="two.sided")
  print(paste("The p-value of the log-rank test for the",dataname[i],"dataset is",pvalue(PFS_lo
grank),sep=" "))

}
```

```
## [1] "The p-value of the log-rank test for the colon dataset is 0.00898408923748778"
## [1] "The p-value of the log-rank test for the ovarian dataset is 0.0012913750507516"
## [1] "The p-value of the log-rank test for the gastadv dataset is 0.0550528405500075"
```

(b)

Plot the Kaplan-Meier curves for OT by treatment group. Compare the survival distributions of two groups using the log-rank test.

```
par(mfrow=c(1,3))
for(i in 1:3){
  mydata=datalist[[i]]
  # Get the survival curve for OT
  km_fit<-survfit(Surv(OT,status_0)~group,data=mydata)
  plot(km_fit,xlab="Time",ylab="Overall survival probability",
       main=dataname[i],col=c(1,2),lwd=1.5,lty=c(1,2))
}
```




```
for(i in 1:3){
  mydata=datalist[[i]]
  OS_logrank<-logrank_test(Surv(OT,status_0)~group,
                           data=mydata,
                           type="logrank",alternative="two.sided")
  print(paste("The p-value of the log-rank test for the",dataname[i],"dataset is",pvalue(OS_log
rank),sep=" "))
}
```

```
## [1] "The p-value of the log-rank test for the colon dataset is 0.051705047068992"
## [1] "The p-value of the log-rank test for the ovarian dataset is 0.0114070649353004"
## [1] "The p-value of the log-rank test for the gastadv dataset is 0.945246388821234"
```

(c)

Present values of RMST_P.

```
RMST_P=list()
for(i in 1:3){
  mydata=datalist[[i]]
  maxmin_P=max(min(mydata$PFT[mydata$group==1]),min(mydata$PFT[mydata$group==0]))
  minmax_P=min(max(mydata$PFT[mydata$group==1]),max(mydata$PFT[mydata$group==0]))
      tau=seq(maxmin_P,minmax_P,length.out=22)
  RMST_P[[i]]=rep(0,22)
  for(j in 1:22){
    temp=rmst2(mydata$PFT,mydata$status_PF,mydata$group,tau=tau[j])
    RMST_P[[i]][j]=temp$unadjusted.result[1,1]
  }
  print(dataname[i])
  print(paste("maxmin_P=",maxmin_P,"minmax_P=",minmax_P,sep=" "))
  print("RMST_P is")
  print(RMST_P[[i]])
}
```

```
## [1] "colon"
## [1] "maxmin_P= 20 minmax_P= 3192"
## [1] "RMST_P is"
##  [1]  -0.03908795   1.47562715   8.56167259  16.24994892  23.77700531
##  [6]  31.31198238  42.14222188  51.77566096  62.38396672  73.55822162
## [11]  87.20433358 101.48193346 114.78930119 127.72627939 139.98623019
## [16] 152.48508181 166.51122411 179.95914123 193.97896375 211.79629179
## [21] 231.17133372 250.54637564
## [1] "ovarian"
## [1] "maxmin_P= 0.003174603175 minmax_P= 1.95555555555556"
## [1] "RMST_P is"
##  [1] 1.047724e-05 2.120841e-03 7.744034e-03 1.493369e-02 2.245122e-02
##  [6] 3.001411e-02 3.574922e-02 4.131424e-02 4.706777e-02 5.294919e-02
## [11] 5.887837e-02 6.462777e-02 7.020793e-02 7.522464e-02 8.016617e-02
## [16] 8.529233e-02 9.065728e-02 9.610224e-02 1.015693e-01 1.070260e-01
## [21] 1.124828e-01 1.179395e-01
## [1] "gastadv"
## [1] "maxmin_P= 5 minmax_P= 1130"
## [1] "RMST_P is"
##  [1]  0.00952381  5.41205858 14.98370229 20.42508629 23.60896485 26.95378640
##  [7] 30.40039141 33.28160920 34.78434771 35.74064944 36.02535840 35.67049429
## [13] 35.13993231 34.44598773 33.60754398 32.76910023 31.66902539 30.42007261
## [19] 29.17111982 27.92216704 26.27912559 24.61966377
```

(d)

Present values of RMST_O.

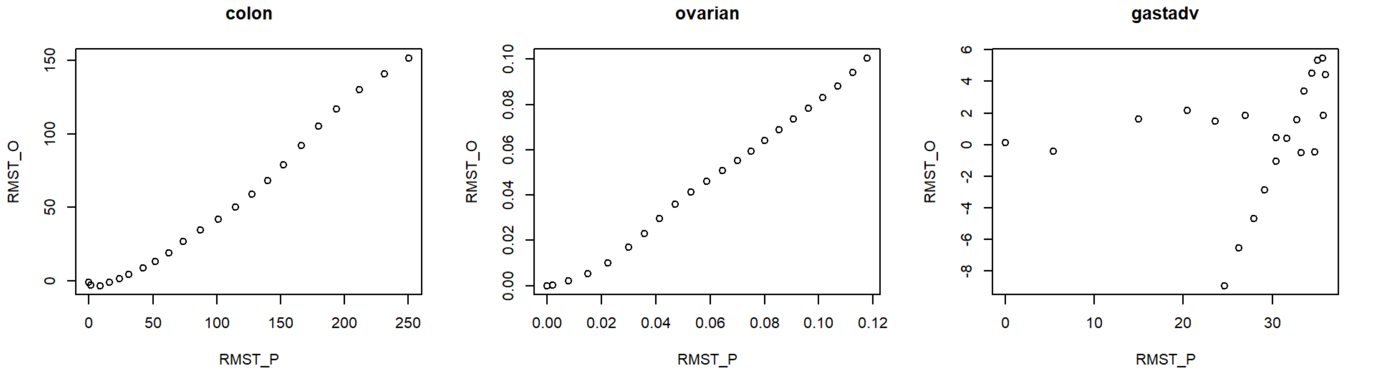
```
RMST_O=list()
for(i in 1:3){
  mydata=datalist[[i]]
  maxmin_O=max(min(mydata$OT[mydata$group==1]),min(mydata$OT[mydata$group==0]))
  minmax_O=min(max(mydata$OT[mydata$group==1]),max(mydata$OT[mydata$group==0]))
  tau=seq(maxmin_O,minmax_O,length.out=22)
  RMST_O[[i]]=rep(0,22)
  for(j in 1:22){
    temp=rmst2(mydata$OT,mydata$status_O,mydata$group,tau=tau[j])
    RMST_O[[i]][j]=temp$unadjusted.result[1,1]
  }
  print(dataname[i])
  print(paste("maxmin_O=",maxmin_O,"minmax_O=",minmax_O,sep=" "))
  print("RMST_O is")
  print(RMST_O[[i]])
}
```

```
## [1] "colon"
## [1] "maxmin_O= 113 minmax_O= 3214"
## [1] "RMST_O is"
##  [1]  -0.8110749  -2.4886829  -2.9758346  -0.9891148   1.5629042   4.7209279
##  [7]   9.1101255  13.4744126  19.1706963  26.9432655  34.9229338  42.1495462
## [13]  50.1002412  59.0602208  68.2857336  78.9164707  91.8122951 104.8919170
## [19] 116.7472892 129.8393738 140.5195136 151.1996534
## [1] "ovarian"
## [1] "maxmin_O= 0.003174603175 minmax_O= 1.95555555555556"
## [1] "RMST_O is"
##  [1] 9.822411e-06 1.743802e-04 2.086898e-03 5.292132e-03 9.843471e-03
##  [6] 1.682546e-02 2.309456e-02 2.964623e-02 3.584233e-02 4.132773e-02
## [11] 4.613987e-02 5.066377e-02 5.506878e-02 5.944838e-02 6.412563e-02
## [16] 6.864777e-02 7.358461e-02 7.834869e-02 8.310914e-02 8.803386e-02
## [21] 9.418190e-02 1.003753e-01
## [1] "gastadv"
## [1] "maxmin_O= 15 minmax_O= 1288"
## [1] "RMST_O is"
##  [1]  0.1333333 -0.4094289  1.6314517  2.1474600  1.5045057  1.8426379
##  [7]  0.4322693 -0.4921401 -0.4759481  1.8368255  4.4234585  5.4620111
## [13]  5.3340527  4.4937747  3.3810455  1.5926607  0.3836299 -1.0344376
## [19] -2.8608788 -4.6873199 -6.5137611 -8.9105672
```

(e)

Show the scatter plot for RMST_P and RMST_O.

```
par(mfrow=c(1,3))
for(i in 1:3){
  plot(RMST_P[[i]],RMST_O[[i]],xlab="RMST_P",ylab="RMST_O",main=dataname[i])
}
```



(f)

Calculate Spearman's rank correlation and Kendall's tau between RMST_P and RMST_O.

```
for(i in 1:3){
  sp=cor.test(RMST_P[[i]],RMST_O[[i]],alternative="two.sided",
         method="spearman")
  kd=cor.test(RMST_P[[i]],RMST_O[[i]],alternative="two.sided",
         method="kendall")
  print(dataname[i])
  print(sp)
  print(kd)
}
```

```
## [1] "colon"
##
##  Spearman's rank correlation rho
##
## data:  RMST_P[[i]] and RMST_O[[i]]
## S = 14, p-value = 2.719e-06
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.9920949
##
##
##  Kendall's rank correlation tau
##
## data:  RMST_P[[i]] and RMST_O[[i]]
## T = 227, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.965368
##
## [1] "ovarian"
##
##  Spearman's rank correlation rho
##
## data:  RMST_P[[i]] and RMST_O[[i]]
## S = 3.9324e-13, p-value = 2.438e-06
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
##   1
##
##
##  Kendall's rank correlation tau
##
## data:  RMST_P[[i]] and RMST_O[[i]]
## T = 231, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
## tau
##   1
##
## [1] "gastadv"
##
##  Spearman's rank correlation rho
##
## data:  RMST_P[[i]] and RMST_O[[i]]
## S = 952, p-value = 0.03156
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.4624506
##
##
##  Kendall's rank correlation tau
##
## data:  RMST_P[[i]] and RMST_O[[i]]
```

```
## T = 157, p-value = 0.01945
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##       tau
## 0.3593074
```

(g)

Restricted mean survival time provides a practical way as a general measure of the treatment effect in a randomized controlled trial.

For the colon and ovarian datasets, by the result of the log-rank test, there exists significant difference between the treatment and control groups for both PFT and OT. Furthermore, Spearman's rank correlation and Kendall's tau show that the sequences of RMST_P and RMST_O are highly correlated. In this case, PFT might be an appropriate surrogate endpoint for OT.

However, for the gastadv dataset, we observe a significant difference in PFT but no significant difference in OT by the result of the log-rank test. Spearman's rank correlation and Kendall's tau also show that the sequences of RMST_P and RMST_O are not highly correlated. Therefore, the PFT vs OT discrepancies deserve great attention.