HW_2

2023-02-23

```
#####
# 4.1. Using the pharmacoSmoking data, compare the two treatments using
 the Prentice modification of the Gehan test, and compare your results
to those from the log-rank test.
#####

library(asaur)
attach(pharmacoSmoking)
library(survival)

head(pharmacoSmoking)

##     id ttr relapse          grp age gender     race employment yearsSm
oking
## 1  21 182       0    patchOnly  36   Male    white         ft
  26
## 2 113  14       1    patchOnly  41   Male    white      other
  27
## 3  39   5       1 combination  25 Female    white      other
  12
## 4  80  16       1 combination  54   Male    white         ft
  39
## 5  87   0       1 combination  45   Male    white      other
  30
## 6  29 182       0 combination  43   Male hispanic         ft
  30
##   levelSmoking ageGroup2 ageGroup4 priorAttempts longestNoSmoke
## 1        heavy     21-49     35-49             0              0
## 2        heavy     21-49     35-49             3             90
## 3        heavy     21-49     21-34             3             21
## 4        heavy       50+     50-64             0              0
## 5        heavy     21-49     35-49             0              0
## 6        heavy     21-49     35-49             2           1825

survdiff(Surv(ttr, relapse) ~ grp)

## Call:
## survdiff(formula = Surv(ttr, relapse) ~ grp)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## grp=combination 61       37     49.9      3.36      8.03
## grp=patchOnly   64       52     39.1      4.29      8.03
##
##  Chisq= 8  on 1 degrees of freedom, p= 0.005
```
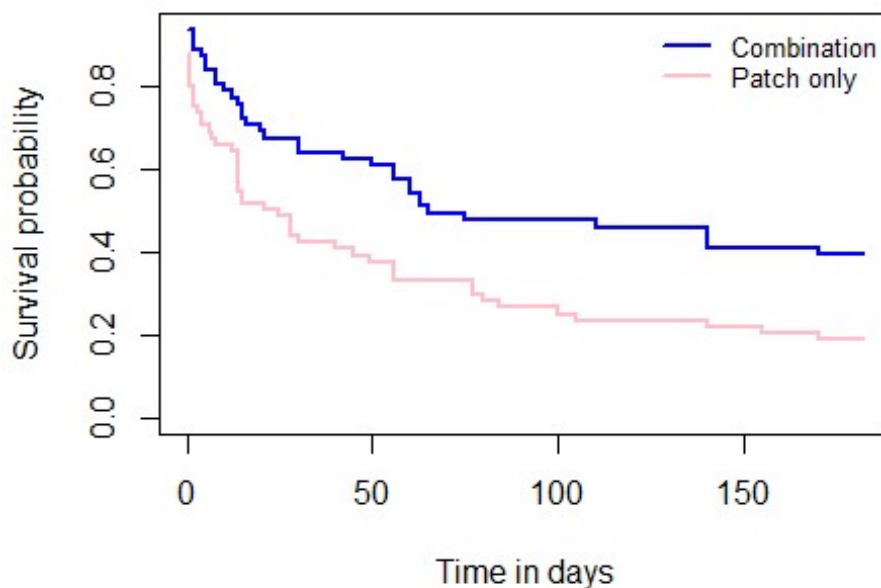
```
survdiff(Surv(ttr, relapse) ~ grp, rho=1)

## Call:
## survdiff(formula = Surv(ttr, relapse) ~ grp, rho = 1)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## grp=combination 61     23.1     32.1      2.53      8.01
## grp=patchOnly   64     35.8     26.8      3.04      8.01
##
##  Chisq= 8  on 1 degrees of freedom, p= 0.005

plot(survfit(Surv(ttr, relapse) ~ grp), xlab="Time in days", ylab="Surv
ival probability",
col=c("blue", "pink"), lwd=2)

legend("topright", legend=c("Combination", "Patch only"),
col=c("blue","pink"),bty = "n", lwd=2, cex=0.8)
```


The P-values is 0.005 for both tests, strongly rejecting the null hypothesis. However, the expected and observed outcome drops in the Prentice modified test. This is because Prentice modified test places a higher weight on earlier survival differences.

```
#####
# 4.2 Again using the pharmacoSmoking data, carry out a log-rank test

# comparing the two treatments stratifying on employment status.
#####
survdiff(Surv(ttr, relapse) ~ grp + strata(employment))

## Call:
## survdiff(formula = Surv(ttr, relapse) ~ grp + strata(employment))
##
##                    N Observed Expected (O-E)^2/E (O-E)^2/V
## grp=combination 61       37     50.3      3.50      8.58
## grp=patchOnly   64       52     38.7      4.54      8.58
##
##  Chisq= 8.6  on 1 degrees of freedom, p= 0.003
```

The p-value is 0.003. Therefore, the stratified log-rank test shows the difference between the group of patch only and the group of combination is highly significant. By the test, the combination is superior to patch only.

```
#####
# 5.1. Consider the data set "aml", which is included in the "survival"
 package. This is a study of whether or not maintenance therapy increas
es survival among patients with acute myelogenous leukemia, with surviv
al time measured in weeks. The basic Cox model may be fitted as follows
:
# result <- coxph(Surv(time, status) ~ x, data=aml)

# Regression Analysis Using the Proportional Hazards Model Create a coa
rser time variable by expressing it in months instead of weeks as follo
ws:
# time.months <- cut(aml$time, breaks=seq(0,161,4), labels=F)
# Now re-fit the model,modeling ties using the Breslow, Efron, and exac
t methods. Which approximate method gives a result closest to that from
 the exact method?
#####


head(aml)

##   time status        x
## 1    9      1 Maintained
## 2   13      1 Maintained
## 3   13      0 Maintained
## 4   18      1 Maintained
## 5   23      1 Maintained
## 6   28      0 Maintained
```

```
result.cox <- coxph(Surv(time, status) ~ x, data=aml)
summary(result.cox)

## Call:
## coxph(formula = Surv(time, status) ~ x, data = aml)
##
##   n= 23, number of events= 18
##
##                   coef exp(coef) se(coef)     z Pr(>|z|)
## xNonmaintained 0.9155    2.4981   0.5119 1.788   0.0737 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                exp(coef) exp(-coef) lower .95 upper .95
## xNonmaintained     2.498     0.4003    0.9159     6.813
##
## Concordance= 0.619  (se = 0.063 )
## Likelihood ratio test= 3.38  on 1 df,    p=0.07
## Wald test            = 3.2  on 1 df,    p=0.07
## Score (logrank) test = 3.42  on 1 df,    p=0.06

time.months <- cut(aml$time, breaks=seq(0,161,4), labels=F)
result.bres <- coxph(Surv(time.months, status) ~ x, data=aml)
summary(result.bres)

## Call:
## coxph(formula = Surv(time.months, status) ~ x, data = aml)
##
##   n= 22, number of events= 18
##    (1 observation deleted due to missingness)
##
##                   coef exp(coef) se(coef)     z Pr(>|z|)
## xNonmaintained 0.5955    1.8139   0.4889 1.218    0.223
##
##                exp(coef) exp(-coef) lower .95 upper .95
## xNonmaintained     1.814     0.5513    0.6958     4.729
##
## Concordance= 0.6  (se = 0.07 )
## Likelihood ratio test= 1.53  on 1 df,    p=0.2
## Wald test            = 1.48  on 1 df,    p=0.2
## Score (logrank) test = 1.53  on 1 df,    p=0.2

result.efron <- coxph(Surv(time.months, status) ~ x, data=aml, ties = "
efron")
summary(result.efron)

## Call:
## coxph(formula = Surv(time.months, status) ~ x, data = aml, ties = "e
fron")
##
##   n= 22, number of events= 18
```

```
##     (1 observation deleted due to missingness)
##
##                  coef exp(coef) se(coef)      z Pr(>|z|)
## xNonmaintained 0.5955    1.8139   0.4889 1.218    0.223
##
##                exp(coef) exp(-coef) lower .95 upper .95
## xNonmaintained     1.814     0.5513    0.6958     4.729
##
## Concordance= 0.6  (se = 0.07 )
## Likelihood ratio test= 1.53  on 1 df,    p=0.2
## Wald test            = 1.48  on 1 df,    p=0.2
## Score (logrank) test = 1.53  on 1 df,    p=0.2
```

```r
result.exact <- coxph(Surv(time.months, status) ~ x, data=aml, ties = "
exact")
summary(result.exact)
```

```
## Call:
## coxph(formula = Surv(time.months, status) ~ x, data = aml, ties = "e
xact")
##
##   n= 22, number of events= 18
##     (1 observation deleted due to missingness)
##
##                  coef exp(coef) se(coef)      z Pr(>|z|)
## xNonmaintained 0.6365    1.8898   0.5325 1.195    0.232
##
##                exp(coef) exp(-coef) lower .95 upper .95
## xNonmaintained      1.89     0.5292    0.6655     5.367
##
## Concordance= 0.6  (se = 0.07 )
## Likelihood ratio test= 1.48  on 1 df,    p=0.2
## Wald test            = 1.43  on 1 df,    p=0.2
## Score (logrank) test = 1.46  on 1 df,    p=0.2
```

The parameter estimates by the three methods are similar, with close standard errors, and the same p-values, suggesting that the number of ties in the data is not high enough to significantly impact the results.

```r
######
# 5.2. Consider again the synthetic data in Table 4.1, discussed in Exa
mple 5.1 in Sect. 5.2. Use the "basehaz" function to obtain an estimate
 of the baseline cumulative hazard function. Use this to compute the pr
edicted survival curves for the control and experimental groups based o
n the proportional hazards model we fitted in Sect. 5.2.
#####

tt <- c(6, 7, 10, 15, 19, 25)
delta <- c(1, 0, 1, 1, 0, 1)
```

```r
trt <- c(0, 0, 1, 0, 1, 1)


syndata <- data.frame(tt = c(6, 7, 10, 15, 19, 25), delta = c(1, 0, 1,
1, 0, 1), trt = c(0, 0, 1, 0, 1, 1))
syndata

##    tt delta trt
## 1  6     1   0
## 2  7     0   0
## 3 10     1   1
## 4 15     1   0
## 5 19     0   1
## 6 25     1   1

fit <- coxph(Surv(tt, delta) ~ trt, data = syndata)


bh <- basehaz(fit)
bh

##       hazard time
## 1 0.2633999    6
## 2 0.2633999    7
## 3 0.8200350   10
## 4 1.4732003   15
## 5 1.4732003   19
## 6 5.2396357   25

ctrl_data <- data.frame(trt = 0, tt = rep(mean(syndata$tt), 1), delta =
 rep(mean(syndata$delta), 1))
exp_data <- data.frame(trt = 1, tt = rep(mean(syndata$tt), 1), delta =
rep(mean(syndata$delta), 1))

ctrl_haz <- exp(-predict(fit, newdata = ctrl_data, type = "expected",ce
ntered="F") * bh$haz)

## Warning in Surv(tt, delta): Invalid status value, converted to NA

exp_haz <- exp(-predict(fit, newdata = exp_data, type = "expected", cen
tered="F") * bh$haz)

## Warning in Surv(tt, delta): Invalid status value, converted to NA

ctrl_surv <- cumprod(ctrl_haz)
exp_surv <- cumprod(exp_haz)
plot(bh$time, ctrl_surv, type = "l", xlab = "Time", ylab = "Survival Pr
obability", col = "pink",main ="Predicted Survival Curves for Control a
nd Experimental Group", cex.main = 0.9)
lines(bh$time, exp_surv, col = "blue")
```

```
legend("topright", legend = c("Control", "Experimental"), col = c("red"
, "blue"), bty = "n", lty = 1, cex=0.8)
```

Predicted Survival Curves for Control and Experimental Group