

Lab 6: Survival Analysis

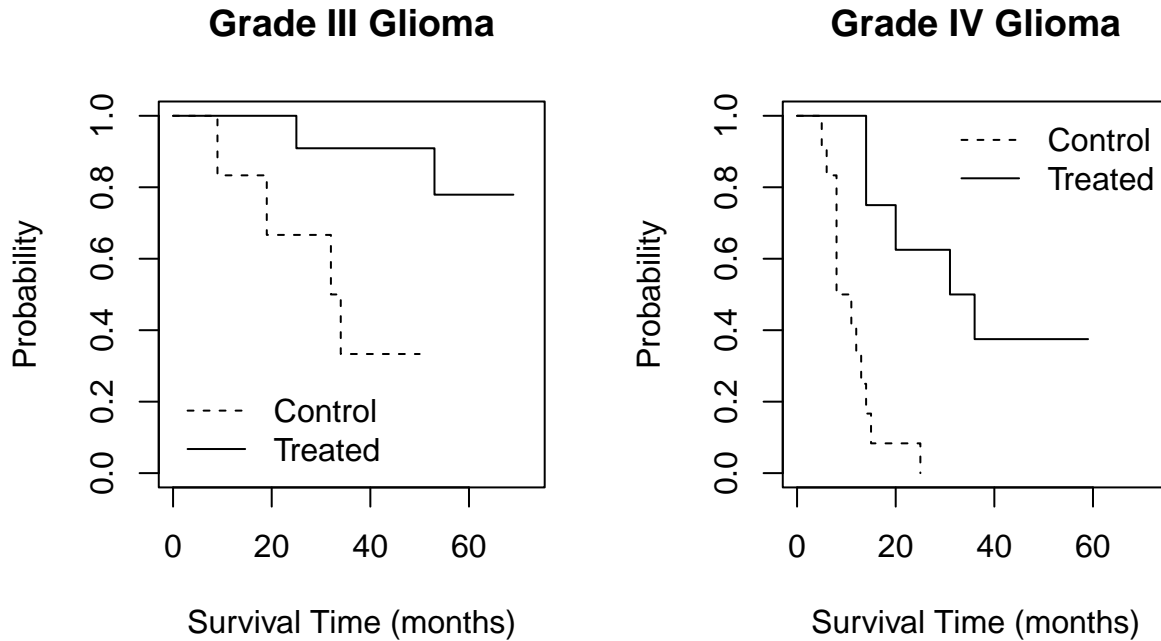
2022-03-22

Analysis of glioma data

```
# Load data and library  
data("glioma", package = "coin")  
library("survival")
```

We plot Kaplan-Meier curves (estimate of survival function) for both treated and control patient groups, within each grade (glioma type):

```
# Prepare plot layout  
layout(matrix(1:2, ncol = 2))  
par(pty = "s")      # makes square plots  
  
# Take subset for grade III, then plot  
g3 <- subset(glioma, histology == "Grade3")  
plot(survfit(Surv(time, event) ~ group, data = g3),  
     main="Grade III Glioma", lty = c(2,1),  
     ylab = "Probability", xlab = "Survival Time (months)",  
     xlim= c(0, max(glioma$time) * 1.05))  
legend("bottomleft", c("Control", "Treated"), lty = c(2,1), bty = "n")  
  
# Take subset grade IV, then plot  
g4 <- subset(glioma, histology == "GBM")  
plot(survfit(Surv(time, event) ~ group, data = g4),  
     main="Grade IV Glioma", lty = c(2,1),  
     ylab = "Probability", xlab = "Survival Time (months)",  
     xlim= c(0, max(glioma$time) * 1.05))  
legend("topright", c("Control", "Treated"), lty = c(2,1), bty = "n")
```



Kaplan-Meier curves between treatment and control groups look different. There appears to be an effect of treatment since for a given time the probability to survive is higher for treated patients (for both Grade III and IV glioma). Yet, it is difficult to say with certainty without confidence intervals on the curves.

We want to test equality of these survival curves using the log-rank test:

```
survdif(Surv(time, event) ~ group, data = g3)
```

```
## Call:
## survdiff(formula = Surv(time, event) ~ group, data = g3)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## group=Control  6         4   1.49     4.23     6.06
## group=RIT      11         2   4.51     1.40     6.06
##
## Chisq= 6.1 on 1 degrees of freedom, p= 0.01
```

```
survdif(Surv(time, event) ~ group, data = g4)
```

```
## Call:
## survdiff(formula = Surv(time, event) ~ group, data = g4)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## group=Control 12         12   5.93     6.23    12.6
## group=RIT      8          5  11.07     3.33    12.6
##
## Chisq= 12.6 on 1 degrees of freedom, p= 4e-04
```

For testing a treatment against cancer, one could argue that a significant level of 1% at least is good to assess statistical significance. With that level, it seems that there are significant differences between control and treatment groups for grade IV glioma ($p < 0.01$), and for grade III ($p = 0.01$).

One problem with using the log-rank test here is that the test is an *asymptotic* one but the number of patients is rather small ($n = 37$, here). In addition, there are *tied results* - equal survival times for some patients, which reduces variance. In this case, we can use an alternative approach: condition on the observed data and compute the distribution of the test statistics using a permutation test to get an *exact* log-rank test p-value:

```
library("coin")
logrank_test(Surv(time, event) ~ group, data = g3, distribution = "exact")
```

```
##
## Exact Two-Sample Logrank Test
##
## data: Surv(time, event) by group (Control, RIT)
## Z = -2.1711, p-value = 0.02877
## alternative hypothesis: true theta is not equal to 1
```

```
logrank_test(Surv(time, event) ~ group, data = g4, distribution = "exact")
```

```
##
## Exact Two-Sample Logrank Test
##
## data: Surv(time, event) by group (Control, RIT)
## Z = -3.2215, p-value = 0.0001588
## alternative hypothesis: true theta is not equal to 1
```

We see that p-values go to 0.028 from 0.01 and to 0.00015 from 0.0004. If significance level is 1%, then results change relative to asymptotic tests, and there are no significant difference between KM curves for grade III glioma.

Another question that we can address is whether the new treatment is superior for both groups (Grade III and Grade IV tumors) simultaneously. To do this, we *stratify* or block by tumor grade (adding `| histology`). Here, we only approximate the exact conditional distribution, since the exact distribution is hard to compute.

```
logrank_test(Surv(time, event) ~ group | histology, data = glioma,
             distribution = approximate(nresample=10000)) #Monte Carlo resampling
```

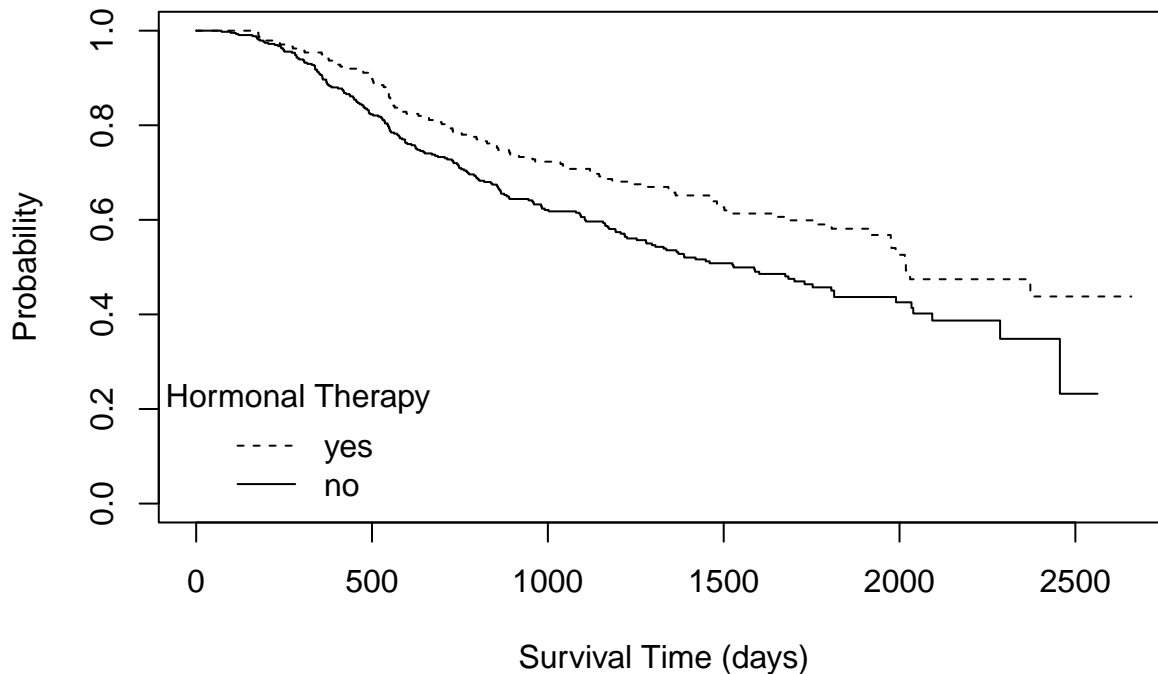
```
##
## Approximative Two-Sample Logrank Test
##
## data: Surv(time, event) by
## group (Control, RIT)
## stratified by histology
## Z = -3.6704, p-value < 1e-04
## alternative hypothesis: true theta is not equal to 1
```

Considering both glioma types, we observe a significant differences between KM curves between treatment and control patients, similarly to asymptotic tests.

Analysis of breast cancer data

Load data, look at the data help. Then, plot KM curves:

```
data("GBSG2", package = "TH.data")
plot(survfit(Surv(time, cens) ~ horTh, data = GBSG2), lty = 1:2, mark.time = FALSE, ylab = "Probability",
legend("bottomleft", legend = c("yes", "no"), lty = c(2,1), title = "Hormonal Therapy", bty = "n")
```



To as-

sess the impact of the covariates on patient survival time, we use Cox proportional hazards modeling. We fit a Cox model with all variables, then examine covariates coefficient.

```
GBSG2.coxph <- coxph(Surv(time, cens) ~ ., data = GBSG2)
summary(GBSG2.coxph)
```

```
## Call:
## coxph(formula = Surv(time, cens) ~ ., data = GBSG2)
##
##   n= 686, number of events= 299
##
##              coef exp(coef)  se(coef)      z Pr(>|z|)
## horThyes    -0.3462784  0.7073155  0.1290747 -2.683 0.007301 **
## age         -0.0094592  0.9905854  0.0093006 -1.017 0.309126
## menostatPost 0.2584448  1.2949147  0.1834765  1.409 0.158954
## tsize       0.0077961  1.0078266  0.0039390  1.979 0.047794 *
## tgrade.L    0.5512988  1.7355056  0.1898441  2.904 0.003685 **
## tgrade.Q   -0.2010905  0.8178384  0.1219654 -1.649 0.099199 .
## pnodes      0.0487886  1.0499984  0.0074471  6.551 5.7e-11 ***
## progrec     -0.0022172  0.9977852  0.0005735 -3.866 0.000111 ***
## estrec      0.0001973  1.0001973  0.0004504  0.438 0.661307
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## horThyes      0.7073      1.4138      0.5492      0.9109
## age           0.9906      1.0095      0.9727      1.0088
## menostatPost  1.2949      0.7723      0.9038      1.8553
## tsize         1.0078      0.9922      1.0001      1.0156
## tgrade.L     1.7355      0.5762      1.1963      2.5178
## tgrade.Q     0.8178      1.2227      0.6439      1.0387
## pnodes       1.0500      0.9524      1.0348      1.0654
## progrec      0.9978      1.0022      0.9967      0.9989
```

```
## estrec          1.0002      0.9998      0.9993      1.0011
##
## Concordance= 0.692 (se = 0.015 )
## Likelihood ratio test= 104.8 on 9 df, p=<2e-16
## Wald test          = 114.8 on 9 df, p=<2e-16
## Score (logrank) test = 120.7 on 9 df, p=<2e-16
```

It seems that the number of positive nodes, progesterone receptor, tumor grade, and presence of hormonal therapy are important predictors.

We estimate the relative risk (exponentiated coefficient) for the hormonal therapy.

```
ci <- confint(GBSG2.coxph)
exp(cbind(coef(GBSG2.coxph), ci))["horThyes",]
```

```
##          2.5 %    97.5 %
## 0.7073155 0.5492178 0.9109233
```

Relative risk change on survival (hazard ratio, HR) associated with presence of hormonal therapy, all other explanatory variables kept constant is $HR = 0.707 < 1$ thus hormonal therapy is protective (reduction in hazard).

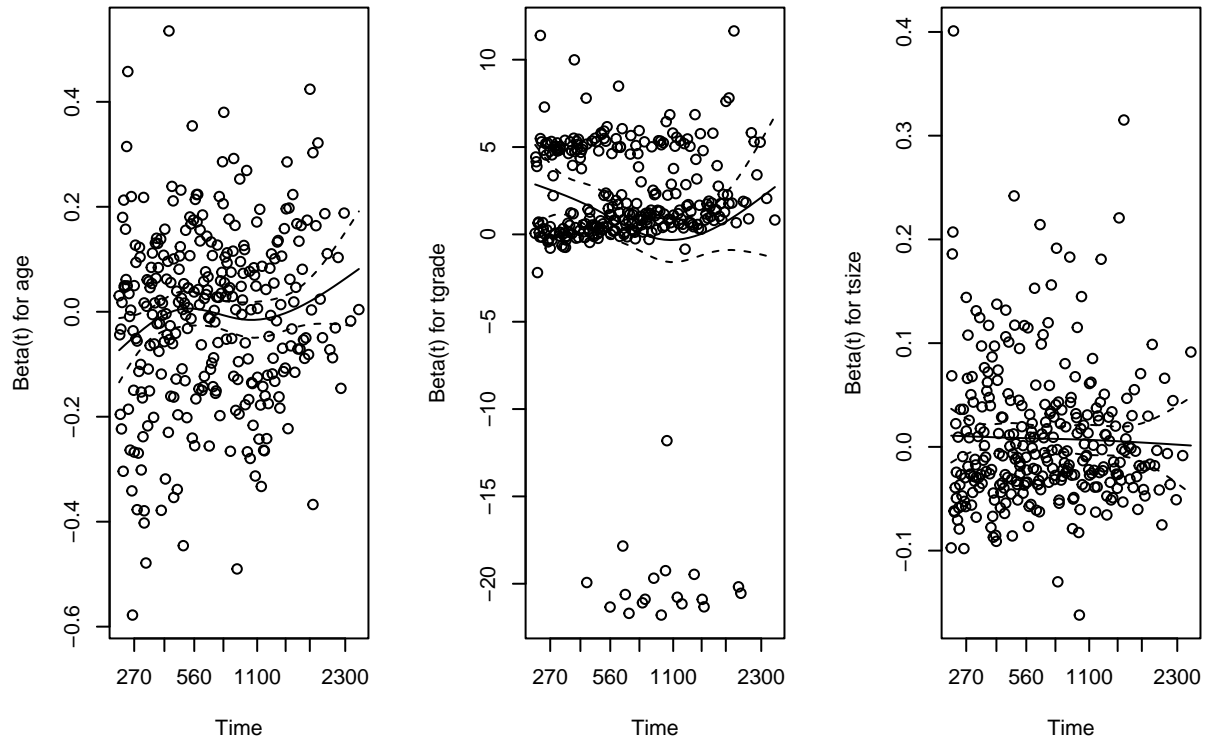
In general, model checking and selection for proportional hazards models is complicated. However, one way to check the proportional hazards assumption is by looking at the parameter estimates over time: if these don't vary much over time, then the proportional hazards assumption seems reasonable. To test for constant regression coefficients over time (2-sided test), both globally and for each variable separately, use the `cox.zph` function:

```
GBSG2.zph <- cox.zph(GBSG2.coxph) # To test proportional hazards assumption
GBSG2.zph
```

```
##          chisq df      p
## horTh      0.239  1 0.6253
## age       10.438  1 0.0012
## menostat   5.406  1 0.0201
## tsize       0.191  1 0.6620
## tgrade     10.712  2 0.0047
## pnodes      0.808  1 0.3688
## progrec     4.386  1 0.0362
## estrec      5.893  1 0.0152
## GLOBAL     24.421  9 0.0037
```

Looking at help for this function, we see the p-value corresponds to a test on the slope for the linear dependence of the coefficient on time. Maybe, it seems that for age and tgrade, and globally, there's evidence that the coefficient are not constant over time.

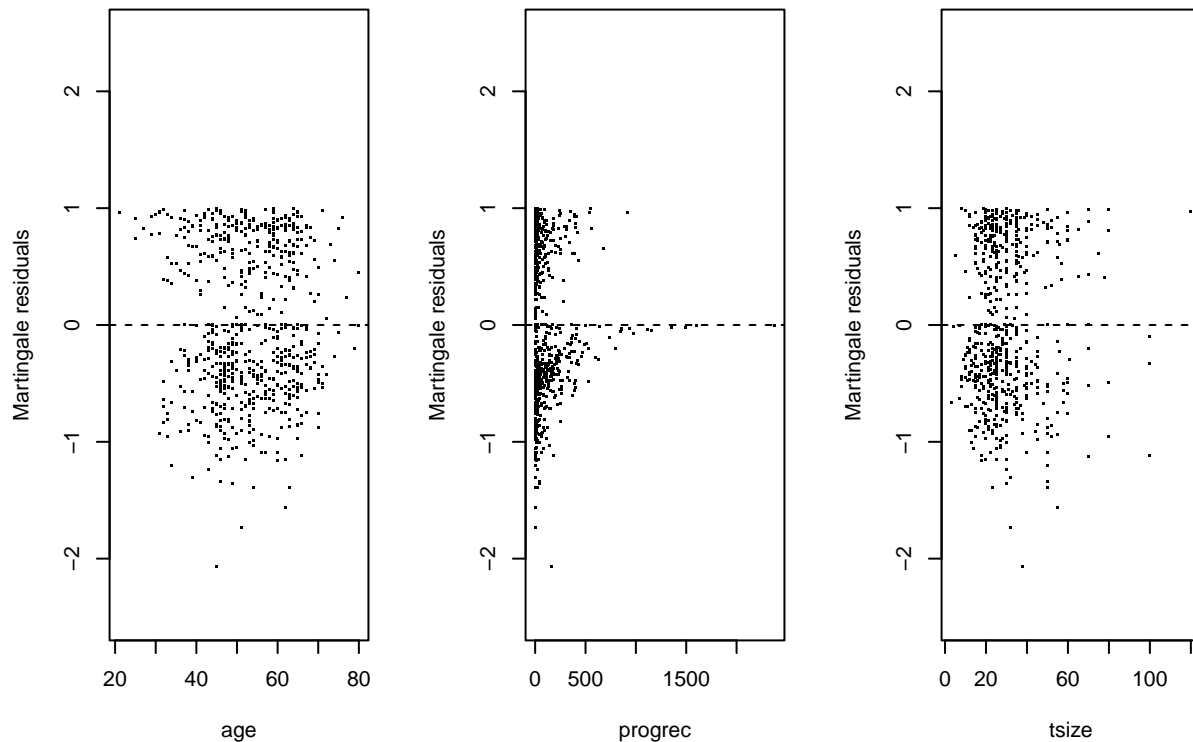
```
layout(matrix(1:3, ncol = 3))
plot(GBSG2.zph, var = "age") # Plot the other variables
plot(GBSG2.zph, var = "tgrade")
plot(GBSG2.zph, var = "tsize")
```



It seems that there are no obvious patterns with time, and therefore that the assumption of proportional hazards appears to be supported, even though numerical results would indicate otherwise (look at the scatter).

The residuals computed in an ordinary linear regression do not apply in the survival context. We check for systematic deviations from the assumed model by inspecting scatterplots of the martingale residuals against covariate values. To do this, we make some martingale residual plots:

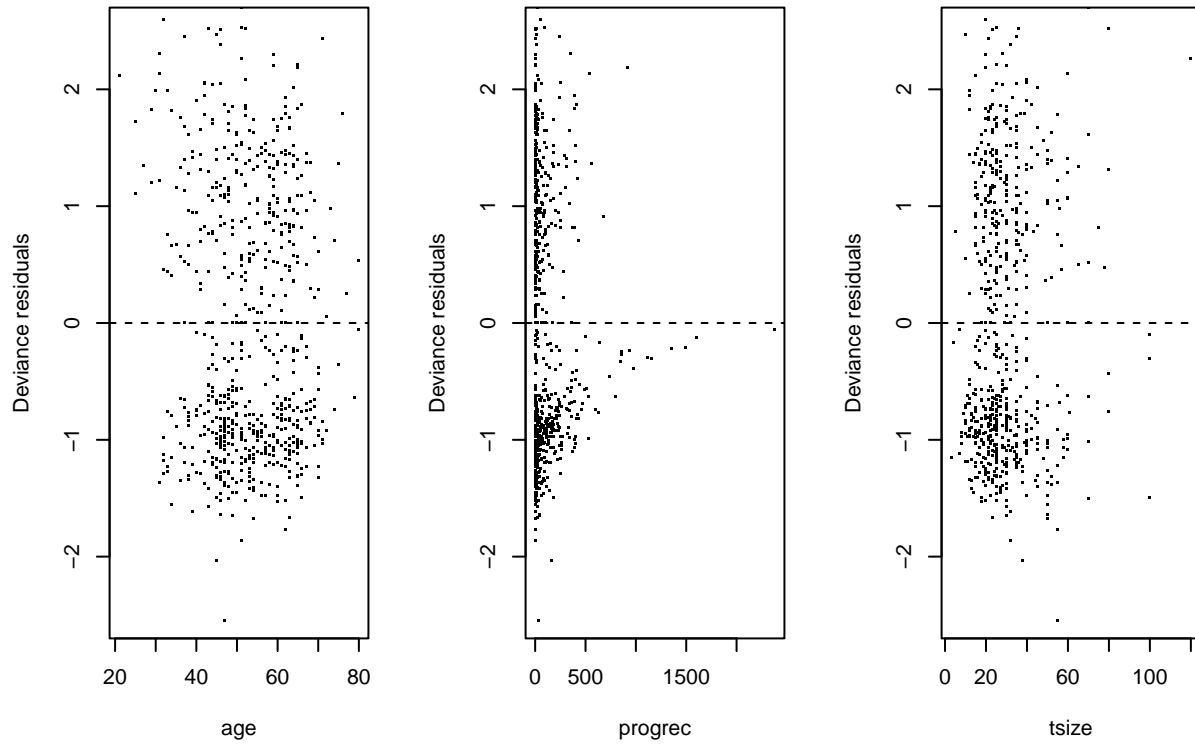
```
layout(matrix(1:3, ncol = 3))
res <- residuals(GBSG2.coxph)
plot(res ~ age, data = GBSG2, ylim = c(-2.5, 2.5), pch=".", ylab = "Martingale residuals")
abline(h=0, lty=2)
plot(res ~ progrec, data = GBSG2, ylim = c(-2.5, 2.5), pch=".", ylab = "Martingale residuals")
abline(h=0, lty=2)
plot(res ~ tsize, data = GBSG2, ylim = c(-2.5, 2.5), pch=".", ylab = "Martingale residuals")
abline(h=0, lty=2)
```



Here however, it seems that martingale residuals are scattered fairly evenly above and below 0, and in addition they do not seem to show any particular pattern. Thus on average, the expected martingale residuals seem to be close to zero. This indicates no systematic deviation from expectation and that therefore there is no major violation of the PH assumption.

We can also look at deviance residuals, which have the benefit to be symmetric:

```
layout(matrix(1:3, ncol = 3))
res_dev <- residuals(GBSG2.coxph, type="deviance")
plot(res_dev ~ age, data = GBSG2, ylim = c(-2.5, 2.5), pch=".", ylab = "Deviance residuals")
abline(h=0, lty=2)
plot(res_dev ~ progrec, data = GBSG2, ylim = c(-2.5, 2.5), pch=".", ylab = "Deviance residuals")
abline(h=0, lty=2)
plot(res_dev ~ tsize, data = GBSG2, ylim = c(-2.5, 2.5), pch=".", ylab = "Deviance residuals")
abline(h=0, lty=2)
```



From Cox regression modeling, it seems that women receiving treatment reduces hazard ($\text{coef} < 1$), which appears to be the most important factor for survival here.