<div align="center">

Chapter 7

Mary Lesperance

2018-10-24

</div>

Contents

Chapter 7

Section 7.1.1 martingale and deviance residuals

```r
library(asaur)
library(survival)
#attach(pharmacoSmoking)
summary(pharmacoSmoking)
```

```
##        id              ttr            relapse            grp
##  Min.   :  1.00   Min.   :  0.00   Min.   :0.000   combination:61
##  1st Qu.: 33.00   1st Qu.:  8.00   1st Qu.:0.000   patchOnly  :64
##  Median : 67.00   Median : 49.00   Median :1.000
##  Mean   : 66.15   Mean   : 77.44   Mean   :0.712
##  3rd Qu.: 99.00   3rd Qu.:182.00   3rd Qu.:1.000
##  Max.   :130.00   Max.   :182.00   Max.   :1.000
##       age           gender         race      employment  yearsSmoking
##  Min.   :22.00   Female:81   black   :38    ft   :72    Min.   : 9.00
##  1st Qu.:41.00   Male  :44   hispanic: 8    other:39    1st Qu.:22.00
##  Median :49.00               other   : 2    pt   :14    Median :30.00
##  Mean   :48.84               white   :77                Mean   :30.88
##  3rd Qu.:56.00                                          3rd Qu.:39.00
##  Max.   :86.00                                          Max.   :56.00
##  levelSmoking  ageGroup2   ageGroup4   priorAttempts     longestNoSmoke
##  heavy:89     21-49:66    21-34:16    Min.   :   0.00   Min.   :   0.0
##  light:36     50+  :59    35-49:50    1st Qu.:   1.00   1st Qu.:   7.0
##                           50-64:48    Median :   2.00   Median :  90.0
##                           65+  :11    Mean   :  12.68   Mean   : 539.7
##                                       3rd Qu.:   5.00   3rd Qu.: 365.0
##                                       Max.   :1000.00   Max.   :6205.0
```

```r
priorAttemptsT <- pharmacoSmoking$priorAttempts
sum(priorAttemptsT > 20)
```

```
## [1] 4
```

```
priorAttemptsT[priorAttemptsT > 20] <- 20

result.0.coxph <- coxph(Surv(ttr, relapse) ~ 1, data=pharmacoSmoking)
rr.0 <- residuals(result.0.coxph, type="martingale")
sum(rr.0)   #martingale residuals sum to zero!!
```

```
## [1] 4.884981e-15
```

```
summary(rr.0)   #martingale residuals are less than 1
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.2350 -1.2350  0.3029  0.0000  0.6949  0.9464
```

Motivation for the use of Martingale residuals to display covariate functional form
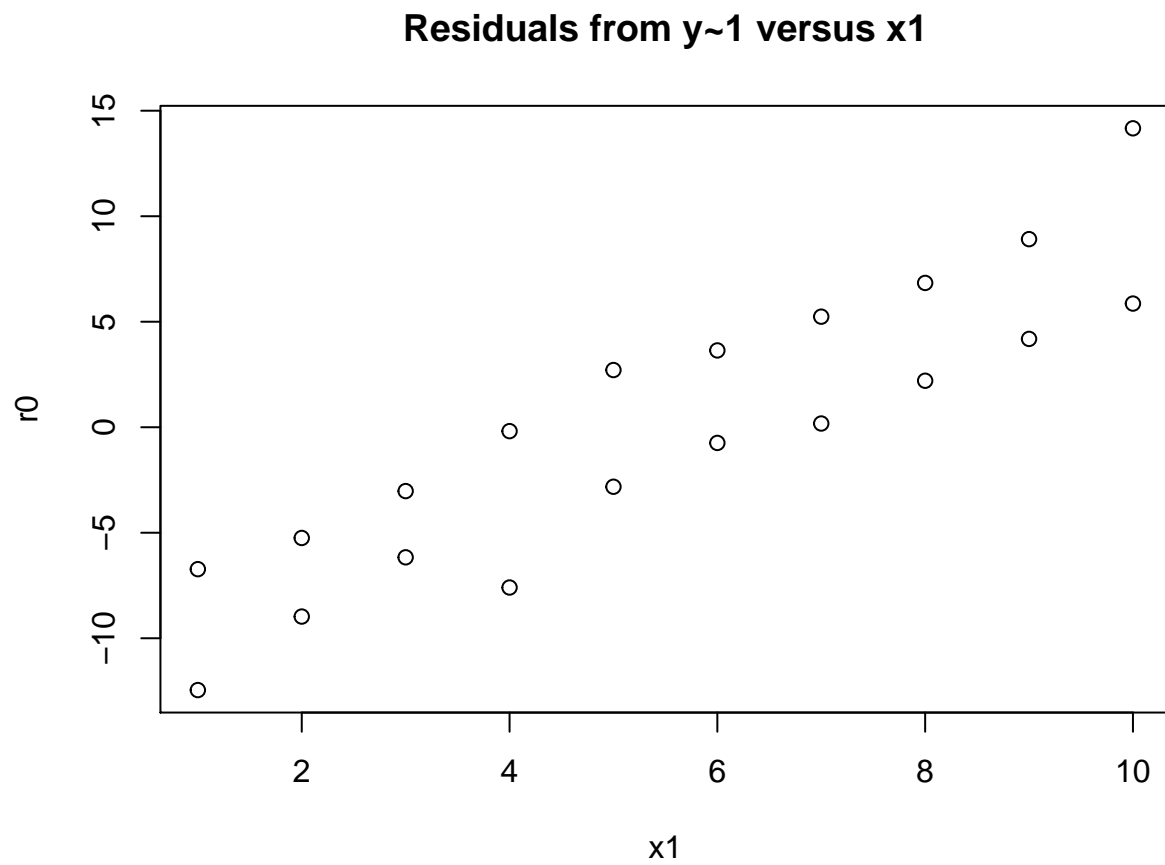
```
# simulate data for lm to see the idea
set.seed(1234)
x1 <- c(1:10, 1:10); x2 <- rep(c(1,2), each=10)   #covariates
beta0 <- 1; beta1 <- 2; beta2 <- 5                 #beta's
y <- rnorm(20, mean=beta0 + beta1*x1 + beta2*x2)   #generate y's

plot(y~x1); title('Y~X1')
```
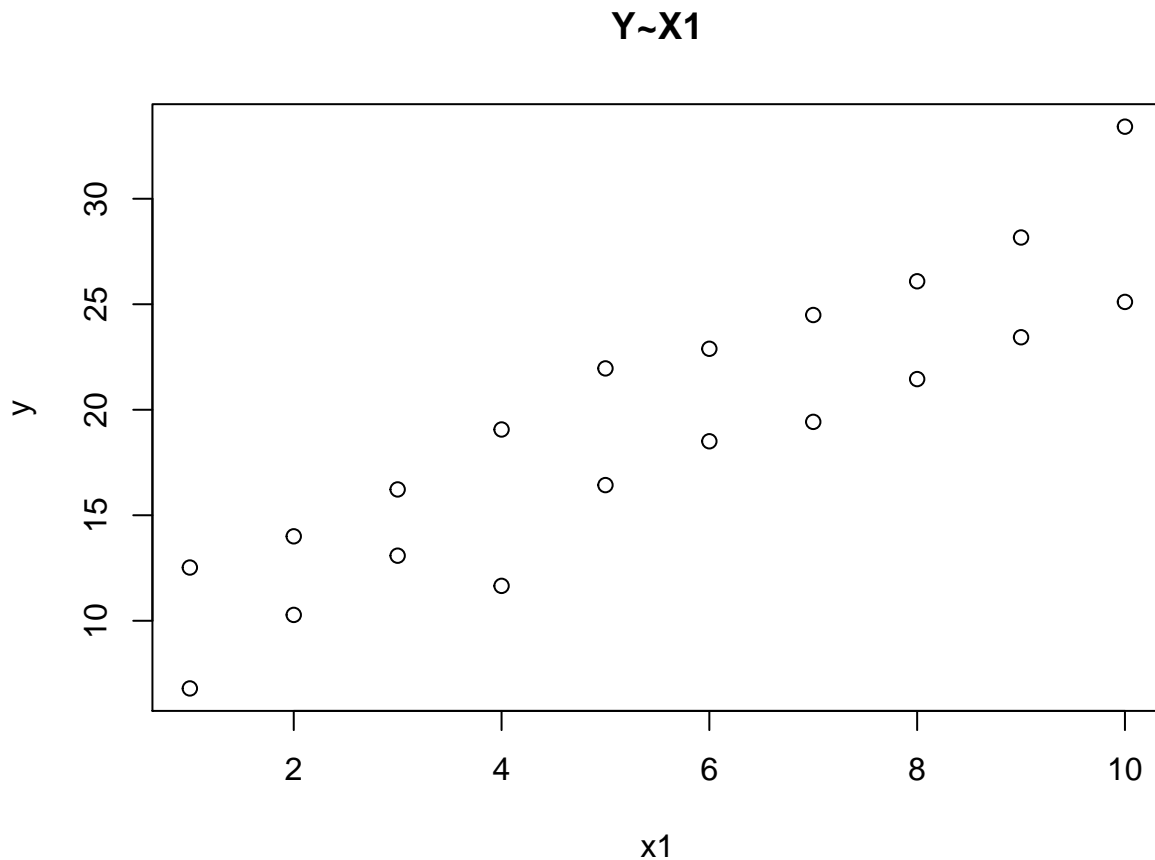


```
r0 <- resid(lm(y~1)) #residuals from model with no covariates
```

```
plot(r0~x1)
title('Residuals from y~1 versus x1')
```

Residuals from y~1 versus x1


```
plot(r0~x2)
title('Residuals from y~1 versus x2')
```
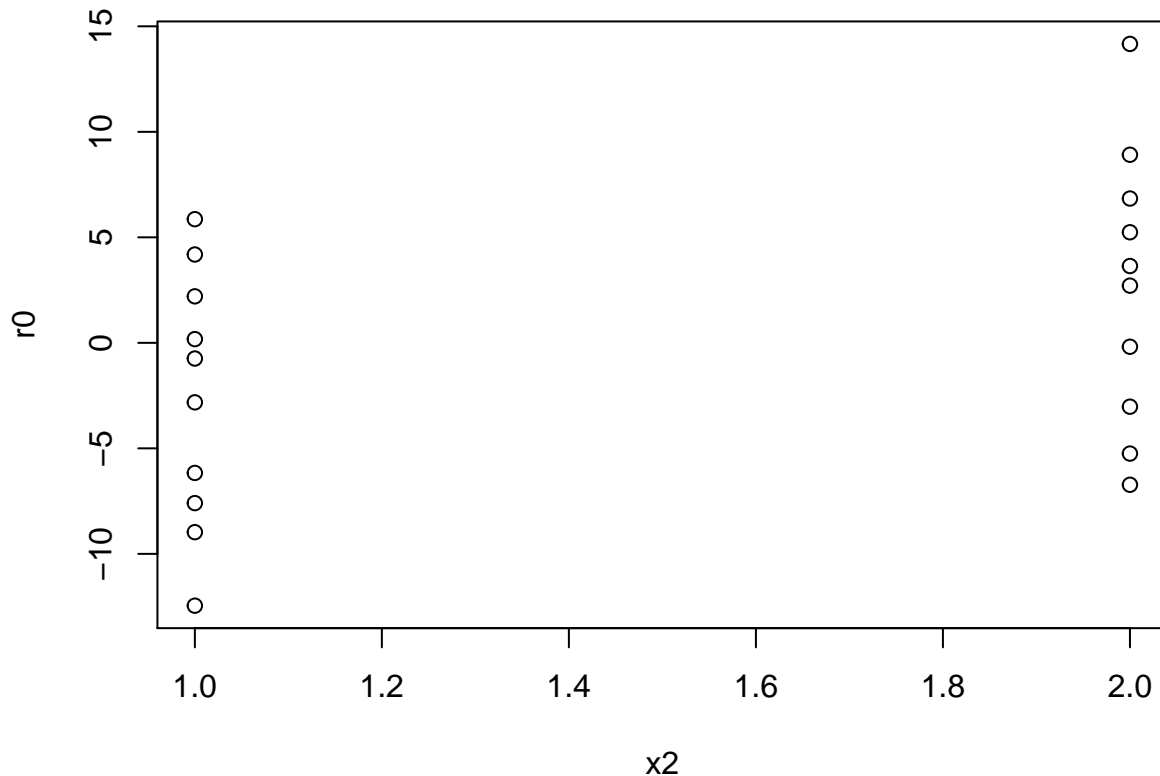
Residuals from y~1 versus x2

Figure 7.1

First, we need the "smoothSEcurve" function (also in the Appendix)

```r
smoothSEcurve <- function(yy, xx) {
  # use after a call to "plot"
  # fit a lowess curve and 95% confidence interval curve
  xx.list <- min(xx) + ((0:100)/100)*(max(xx) - min(xx)) # make list of x values

  # Then fit loess function through the points (xx, yy) at the listed values
  yy.xx <- predict(loess(yy ~ xx), se=T, newdata=data.frame(xx=xx.list))

  lines(yy.xx$fit ~ xx.list, lwd=2)
  lines(yy.xx$fit - qt(0.975, yy.xx$df)*yy.xx$se.fit ~ xx.list, lty=2)
  lines(yy.xx$fit + qt(0.975, yy.xx$df)*yy.xx$se.fit ~ xx.list, lty=2)
  }

# Now the plot
oldpar=par(mfrow=c(1,2))   #save old graph params

plot(rr.0 ~ age, data=pharmacoSmoking)
smoothSEcurve(rr.0, pharmacoSmoking$age)
# Note that "\n"  is a "newline" indicator
title("Martingale residuals\nversus age")
```
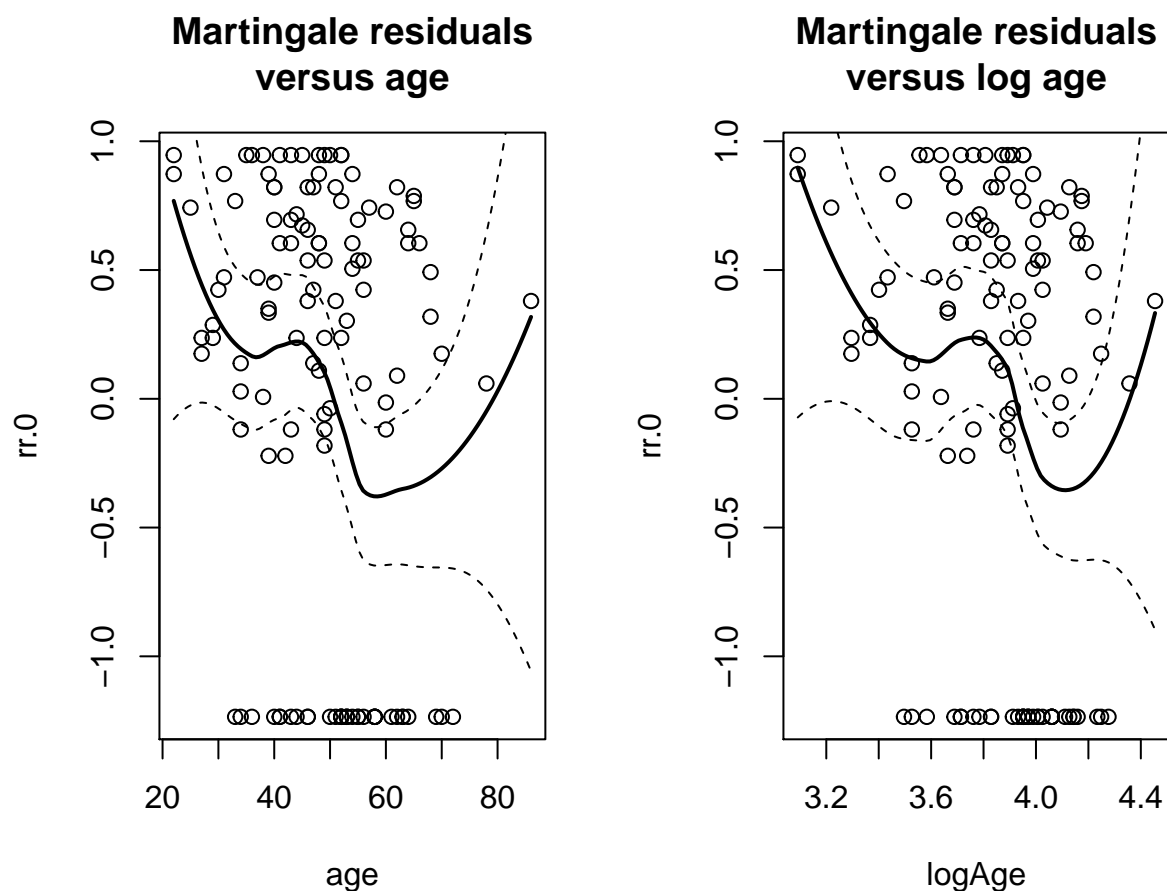
```
logAge <- log(pharmacoSmoking$age)
plot(rr.0 ~ logAge)
smoothSEcurve(rr.0, logAge)
title("Martingale residuals\nversus log age")
```



```
plot(rr.0 ~ priorAttemptsT)
smoothSEcurve(rr.0, priorAttemptsT)
title("Martingale residuals versus\nprior attempts")

logPriorAttemptsT <- log(priorAttemptsT + 1)
plot(rr.0 ~ logPriorAttemptsT)
smoothSEcurve(rr.0, logPriorAttemptsT)
title("Martingale residuals versus\nlog prior attempts")
```

Martingale residuals versus prior attempts
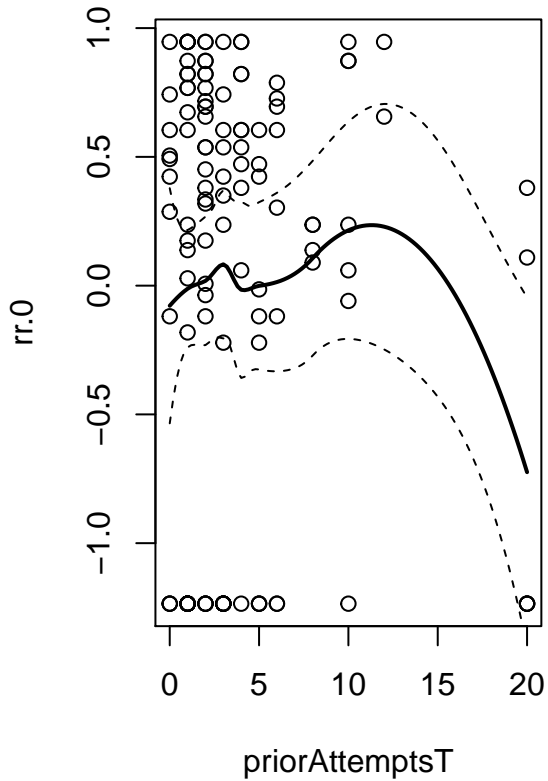
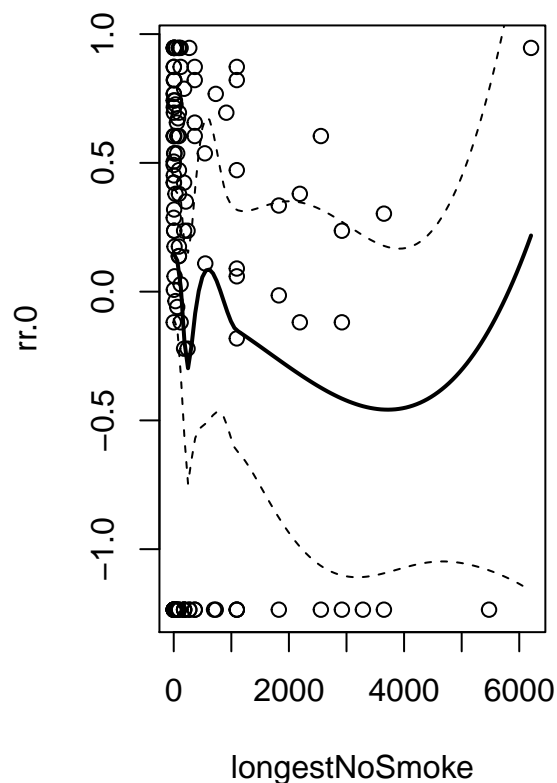## Martingale residuals versus log prior attempts
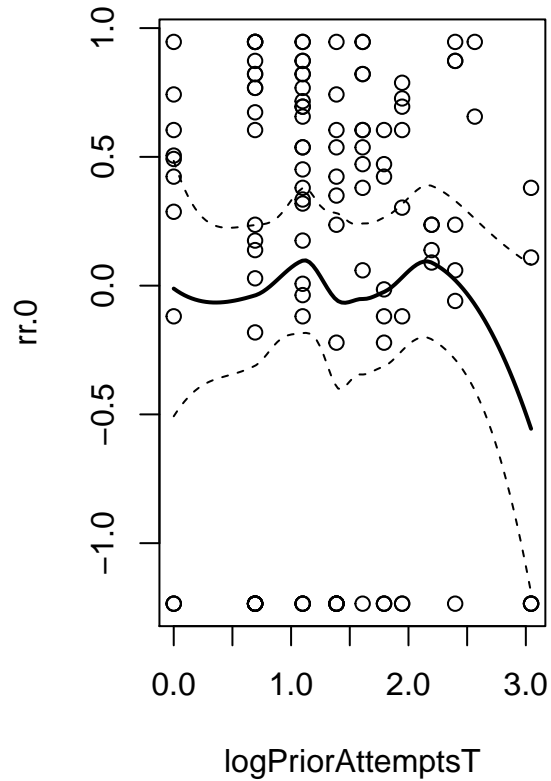


```
plot(rr.0 ~ longestNoSmoke, data=pharmacoSmoking)
smoothSEcurve(rr.0, pharmacoSmoking$longestNoSmoke)
title("Martingale residuals versus\nlongest period without smoking")

logLongestNoSmoke <- log(pharmacoSmoking$longestNoSmoke + 1)
plot(rr.0 ~ logLongestNoSmoke)
smoothSEcurve(rr.0, logLongestNoSmoke)
title("Martingale residuals versus\nlog of longest period without smoking")
```

Martingale residuals versus longest period without smoking

Martingale residuals versus log of longest period without smoke

longestNoSmoke

logLongestNoSmoke

```r
par(oldpar)
```
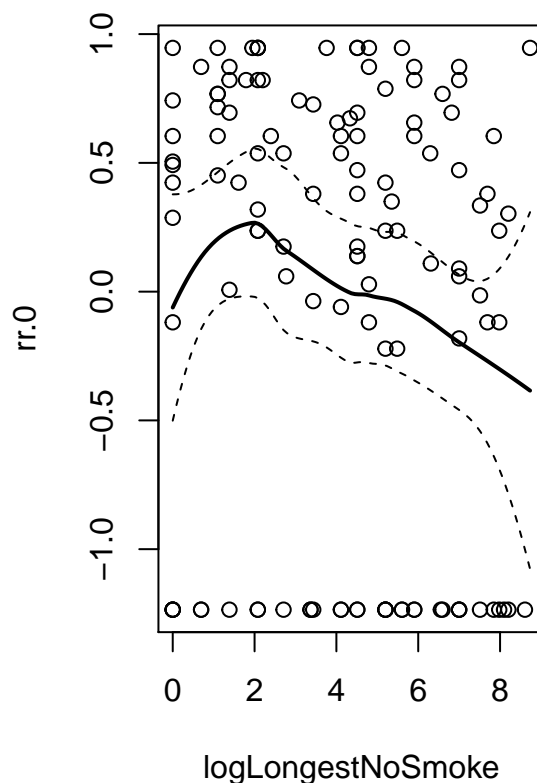
```r
#Use step to identify a good AIC model
#  initial model
result.grp.coxph <- coxph(Surv(ttr, relapse) ~ grp, data=pharmacoSmoking)
result.step <- step(result.grp.coxph, scope=list(upper=~ grp +
                gender + race + employment + yearsSmoking +
                levelSmoking + age + priorAttemptsT +
                logLongestNoSmoke, lower=~grp) )
```

```
## Start:  AIC=766.32
## Surv(ttr, relapse) ~ grp
##
##                    Df   AIC
## + age               1 762.48
## + logLongestNoSmoke 1 765.19
## + yearsSmoking      1 765.76
## <none>                766.32
## + employment        2 767.24
## + gender            1 767.43
## + priorAttemptsT    1 767.63
## + levelSmoking      1 768.27
## + race              3 770.72
##
## Step:  AIC=762.48
## Surv(ttr, relapse) ~ grp + age
##
```

7

```
##                         Df    AIC
## + employment          2 758.28
## + logLongestNoSmoke   1 761.99
## <none>                  762.48
## + yearsSmoking        1 764.12
## + gender              1 764.20
## + priorAttemptsT      1 764.28
## + levelSmoking        1 764.48
## - age                 1 766.32
## + race                3 766.52
##
## Step:  AIC=758.28
## Surv(ttr, relapse) ~ grp + age + employment
##
##                         Df    AIC
## <none>                  758.28
## + logLongestNoSmoke   1 758.55
## + yearsSmoking        1 759.80
## + priorAttemptsT      1 759.87
## + gender              1 760.27
## + levelSmoking        1 760.28
## + race                3 761.10
## - employment          2 762.48
## - age                 1 767.24
```

```
result.step
```

```
## Call:
## coxph(formula = Surv(ttr, relapse) ~ grp + age + employment,
##     data = pharmacoSmoking)
##
##                     coef exp(coef) se(coef)     z      p
## grppatchOnly      0.6079    1.8365   0.2184  2.78 0.0054
## age              -0.0353    0.9653   0.0108 -3.28 0.0010
## employmentother   0.7035    2.0208   0.2693  2.61 0.0090
## employmentpt      0.6537    1.9226   0.3273  2.00 0.0458
##
## Likelihood ratio test=22  on 4 df, p=0.000198
## n= 125, number of events= 89
```

Figure 7.2

```
rr.final <- residuals(result.step, type="martingale")

oldpar <- par(mfrow=c(1,2))

plot(rr.final ~ age, data=pharmacoSmoking)
smoothSEcurve(rr.final, pharmacoSmoking$age)
title("Martingale residuals\nversus age")

plot(rr.final ~ grp, data=pharmacoSmoking)
title("Martingale residuals\nversus treatment group")
```
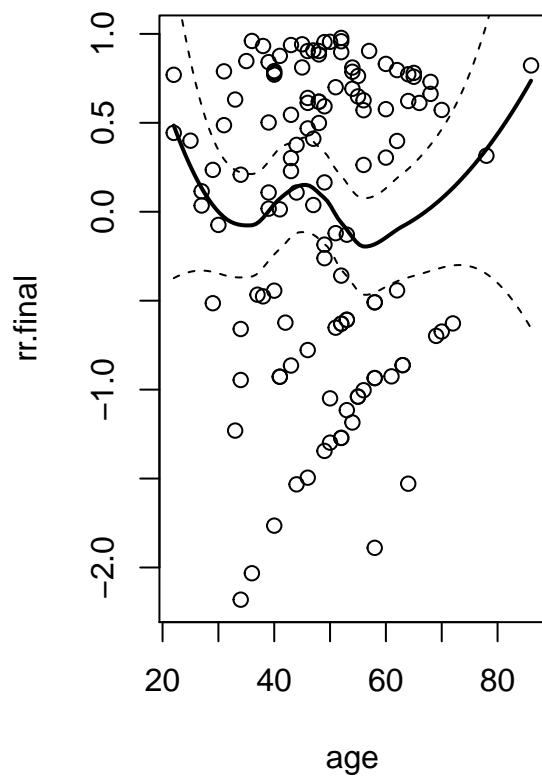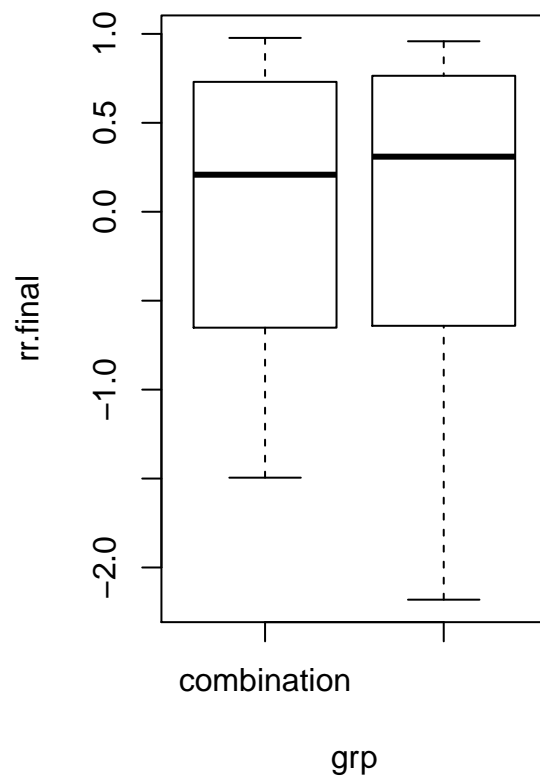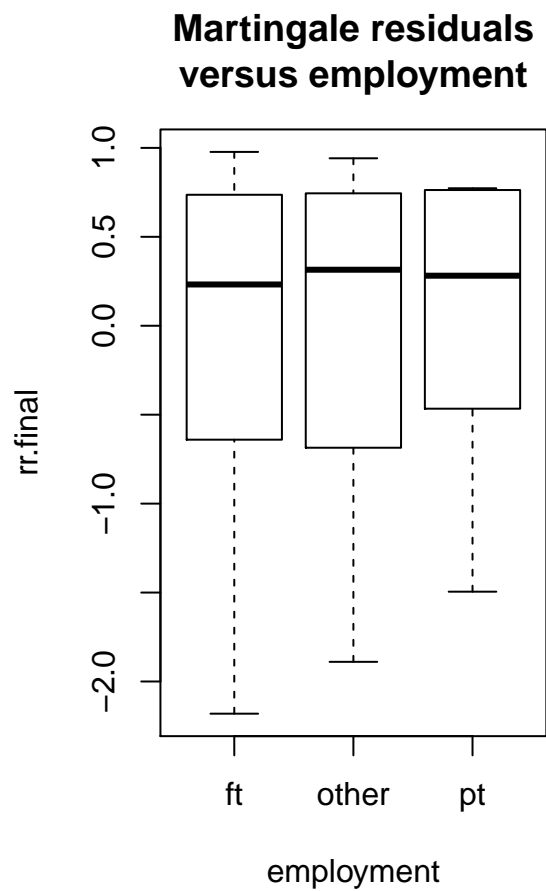
Martingale residuals versus age

Martingale residuals versus treatment group

```r
plot(rr.final ~ employment, data=pharmacoSmoking)
title("Martingale residuals\nversus employment")

par(oldpar)
```
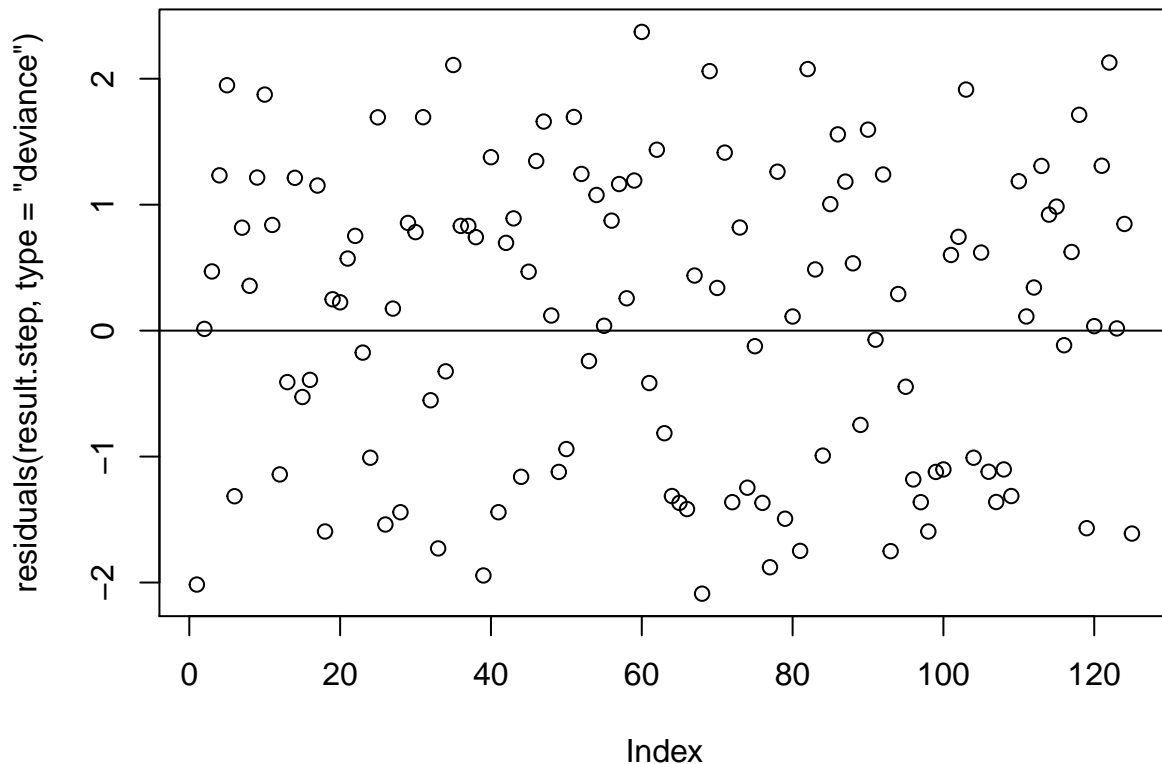
Martingale residuals versus employment

Deviance residuals are used for outlier detection. Plots of these should look like a random normal sample when there is light to moderate censoring.

```
plot(residuals(result.step, type="deviance") )
abline(h=0)
title('Deviance residuals for final model')
```

Deviance residuals for final model

Section 7.1.2 case deletion residuals

```
result.coxph <- coxph(Surv(ttr, relapse) ~ grp + employment + age, data=pharmacoSmoking)
coef.all <- result.coxph$coef[4]   #extract coef for age
coef.all
```

```
##          age
## -0.03528934
```

```
n.obs <- nrow(pharmacoSmoking)

# this is the way the book computes the jackknife resids
jkbeta.vec <- rep(NA, n.obs)
attach(pharmacoSmoking)     #this can lead to errors if you have variables with the same
                            # names in your workspace!!!
for (i in 1:n.obs) {
  tt.i <- ttr[-i]
  delta.i <- relapse[-i]
  grp.i <- grp[-i]
  employment.i <- employment[-i]
  age.i <- age[-i]
  result.coxph.i <- coxph(Surv(tt.i, delta.i) ~ grp.i +
        employment.i + age.i)
  coef.i <- result.coxph.i$coef[4]
  jkbeta.vec[i] <- (coef.all - coef.i)
```

11

```
}
detach(pharmacoSmoking)

# this is a more efficient way to compute the jackknife resids
jackbeta.vec <- rep(NA, n.obs)
for (i in 1:n.obs){
  result.coxph.i <- coxph(Surv(ttr, relapse) ~ grp + employment + age,
                          data=pharmacoSmoking, subset=(i != (1:n.obs)))
  jackbeta.vec[i] <- result.coxph.i$coef[4]
}
jackbeta.vec <- coef.all - jackbeta.vec

sum(jackbeta.vec!=jkbeta.vec)  # the two methods are equivalent
```

[1] 0

```
index.obs <- 1:n.obs
plot(jkbeta.vec ~ index.obs, type="h",
     xlab="Observation", ylab="Change in coefficient for age")
abline(h=0)
title('Change in coefficient for age')
```


**Change in coefficient for age**
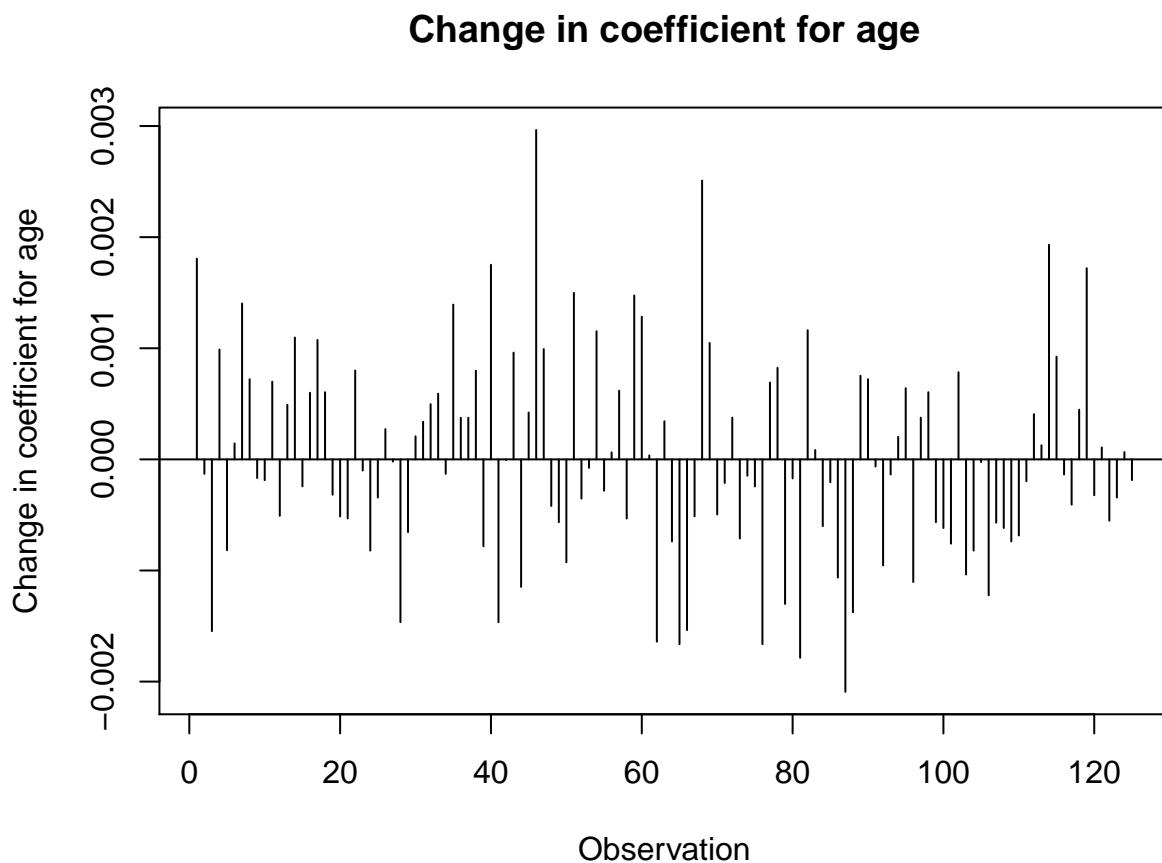
```
#identify(jkbeta.vec ~ index.obs)  # click on points to get their identifying label.
#   Press the "Esc" button to quit the "identify" mode.
```

dfbeta residuals (dfbetas are scaled by standard error of coefficients)

```
resid.dfbeta <- residuals(result.coxph,type='dfbeta')
plot(resid.dfbeta[,4]~index.obs, type='h', xlab='Observation', ylab='Change in Coefficient')
abline(h=0)
title('Approximate change in coefficient for age')
```

Approximate change in coefficient for age
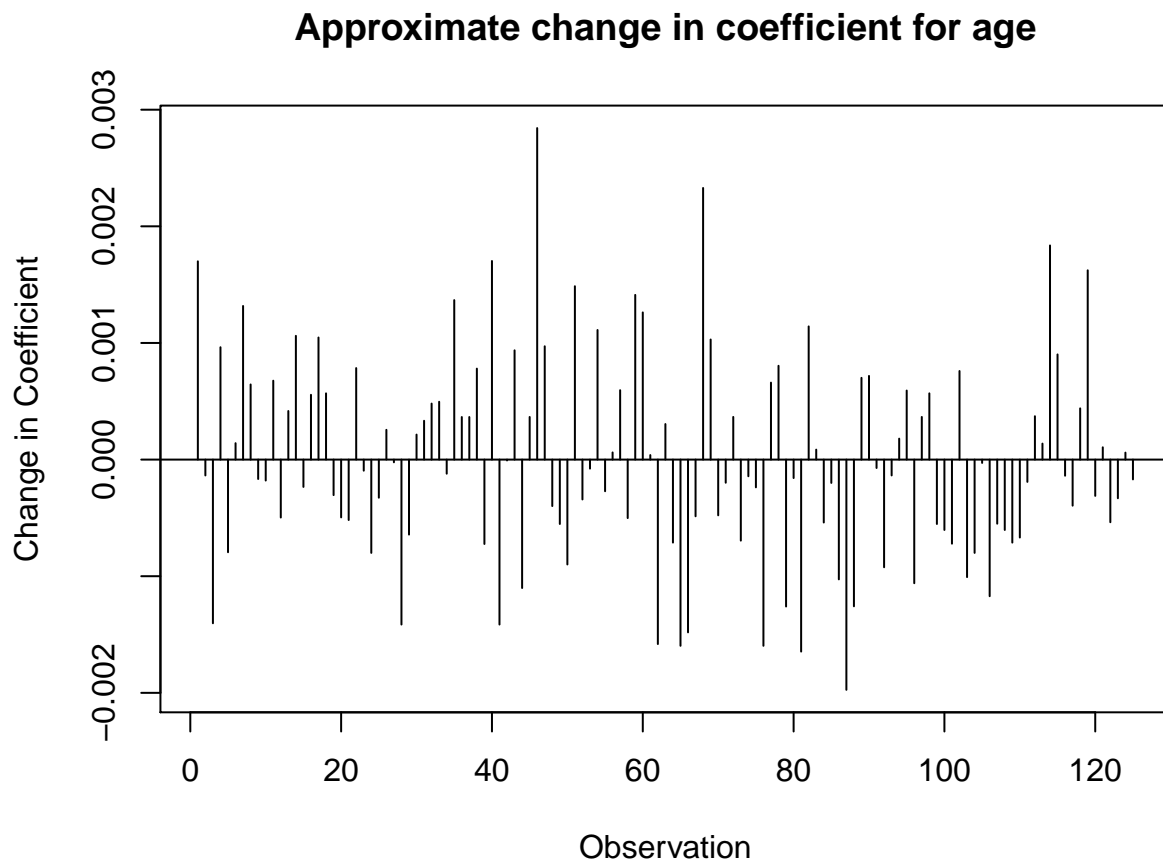
Checking the proportional hazards assumption

Section 7.2.1; log cumulative hazard plots

```
#attach(pancreatic2)
summary(pancreatic2)
```

```
##       pfs              os              status   stage
##  Min.   : 36.0    Min.   : 36.0    Min.   :1    LA: 8
##  1st Qu.: 57.0    1st Qu.:109.0    1st Qu.:1    M :33
##  Median :119.0    Median :192.0    Median :1
##  Mean   :178.6    Mean   :255.4    Mean   :1
##  3rd Qu.:185.0    3rd Qu.:361.0    3rd Qu.:1
##  Max.   :975.0    Max.   :975.0    Max.   :1
```

```
#pfs.month <- pancreatic2$pfs/30.25    #this creates a global variable in the workspace
#locally advanced

result.surv.LA <- survfit(Surv(pfs/30.25, status) ~ stage, subset={stage == "LA"}, data=pancreatic2)
time.LA <- result.surv.LA$time
```

```
surv.LA <- result.surv.LA$surv
surv.LA   #last value is zero!
```

```
## [1] 0.875 0.750 0.625 0.500 0.375 0.250 0.125 0.000
```

```
cloglog.LA <- log(-log(surv.LA))[-length(surv.LA)]
logtime.LA <- log(time.LA)[-length(surv.LA)]

#metastatic
result.surv.M <- survfit(Surv(pfs/30.25, status) ~ stage, subset={stage == "M"}, data=pancreatic2)
time.M <- result.surv.M$time
surv.M <- result.surv.M$surv
surv.M   #last value is zero!
```

```
##  [1] 0.96969697 0.93939394 0.90909091 0.87878788 0.81818182 0.78787879
##  [7] 0.75757576 0.72727273 0.69696970 0.63636364 0.60606061 0.54545455
## [13] 0.51515152 0.48484848 0.45454545 0.42424242 0.39393939 0.36363636
## [19] 0.33333333 0.30303030 0.27272727 0.24242424 0.21212121 0.18181818
## [25] 0.15151515 0.12121212 0.09090909 0.06060606 0.03030303 0.00000000
```

```
cloglog.M <- log(-log(surv.M))[-length(surv.M)]
logtime.M <- log(time.M)[-length(surv.M)]

# Fig 7.4 - ### the book's plot does not cover the entire graph!!

plot(cloglog.LA ~ logtime.LA, type="s", col=1,lty=1, xlim=c(min(logtime.LA,logtime.M),
        max(logtime.LA,logtime.M)), ylim=c(min(cloglog.LA,cloglog.M),max(cloglog.LA,cloglog.M)),
    xlab="Log time", ylab="Complementary log-log survival")
lines(cloglog.M ~ logtime.M, col=2,lty=2, type="s")
legend("bottomright", legend=c("Locally advanced", "Metastatic"),
        col=1:2, lty=1:2)
title('Cloglog of Survival curves for Locally advanced, Metastatic patients')
```
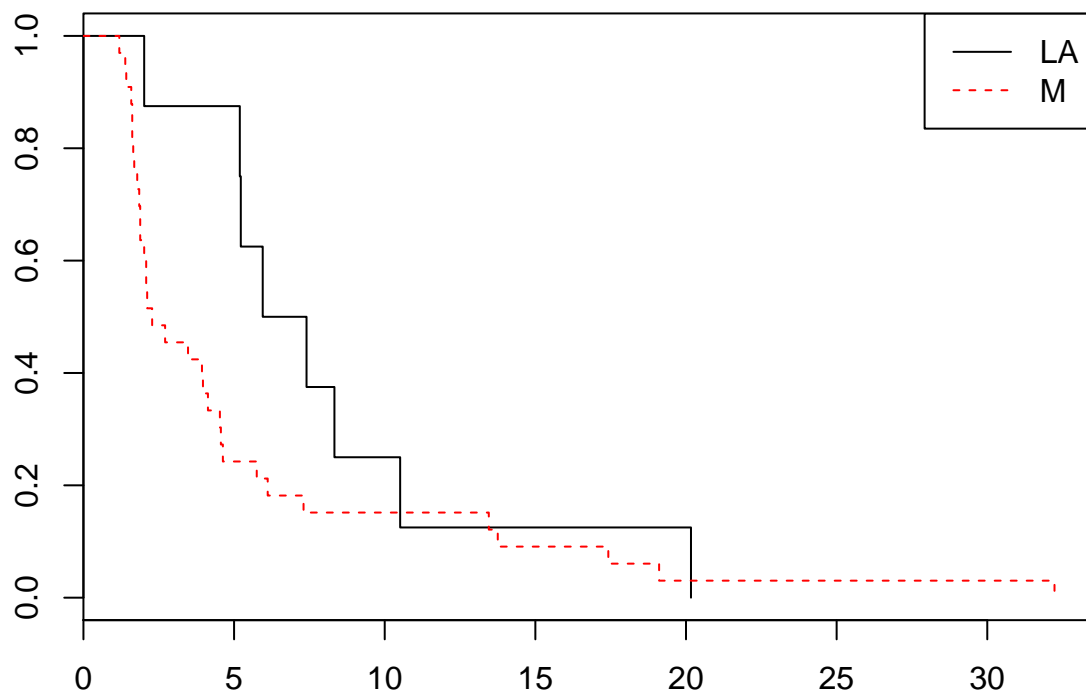
Cloglog of Survival curves for Locally advanced, Metastatic patients
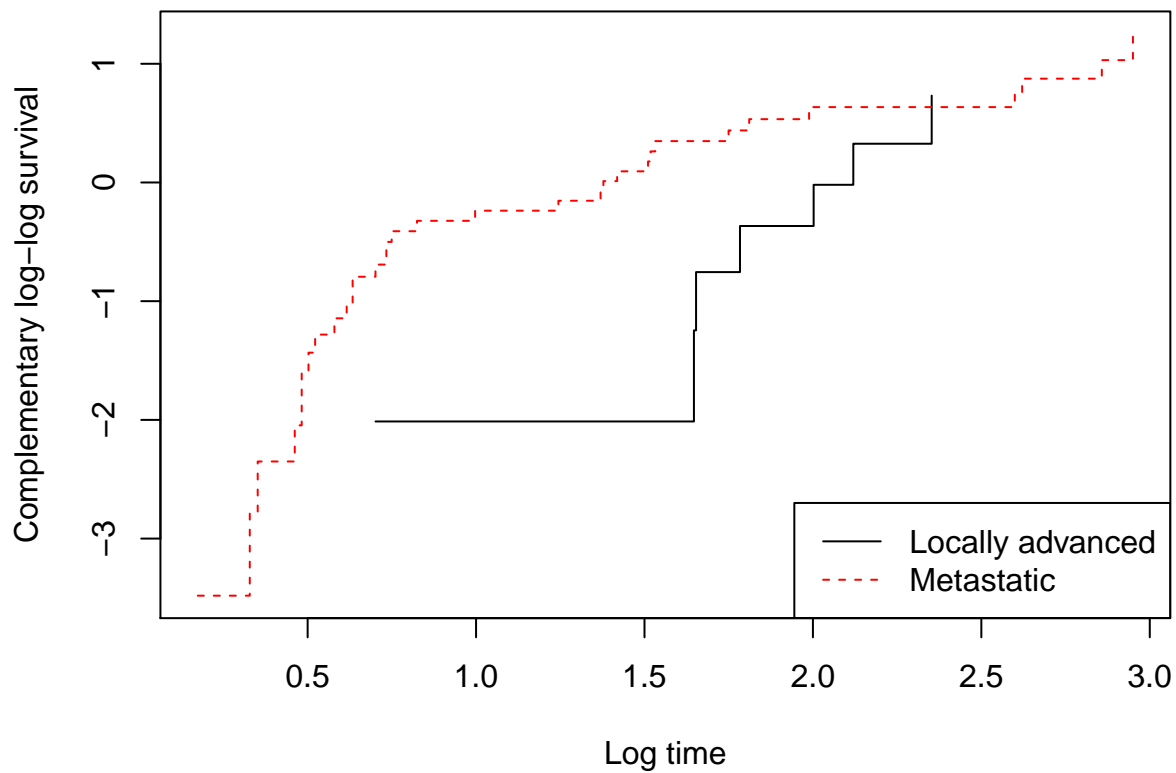

```
# plot survival curves using survfit
plot(survfit(Surv(pfs/30.25, status) ~ stage, data=pancreatic2),lty=1:2, col=1:2)
legend('topright',levels(pancreatic2$stage),lty=1:2, col=1:2)
title('Progression-free survival versus stage')
```
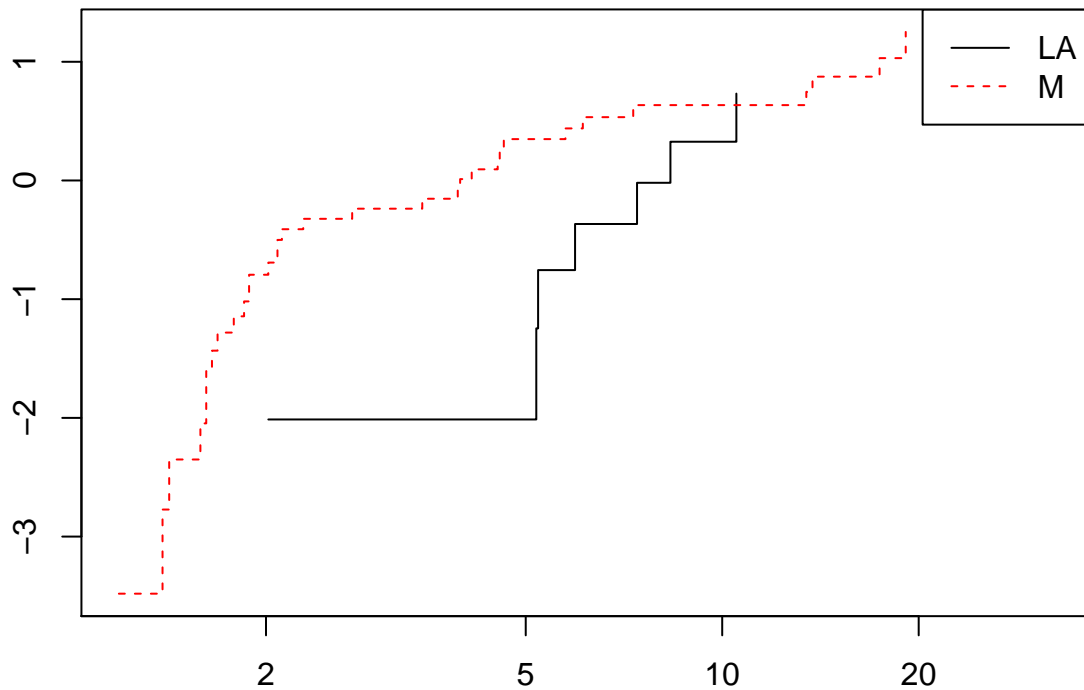
Progression–free survival versus stage


```
# This is a better way to get cloglog of survivor curves
plot(survfit(Surv(pfs/30.25, status) ~ stage, data=pancreatic2),fun='cloglog',
     lty=1:2, col=1:2)
legend('topright',levels(pancreatic2$stage),lty=1:2, col=1:2)
title('Cloglog of Progression-free survival versus stage')
```

Cloglog of Progression–free survival versus stage

```
#detach(pancreatic2)
```

Section 7.2.2 Schoenfeld residuals - Test proportional hazards assumption

```
tt <- c(6, 7, 10, 15, 19, 25)
delta <- c(1, 0, 1, 1, 0, 1)
trt <- c(0, 0, 1, 0, 1, 1)
result.coxph <- coxph(Surv(tt, delta) ~ trt)
result.coxph$coef
```

```
##        trt
## -1.326129
```

```
residuals(result.coxph, type="schoenfeld")
```

```
##          6         10         15         25
## -0.2098004  0.5566351 -0.3468347  0.0000000
```

```
resid.unscaled <- residuals(result.coxph, type="schoenfeld")
resid.scaled <- resid.unscaled * as.vector(result.coxph$var) * sum(delta)

resid.unscaled
```

```
##          6         10         15         25
## -0.2098004  0.5566351 -0.3468347  0.0000000
```

```
resid.scaled
```

```
##         6        10        15        25
## -1.313064  3.483776 -2.170712  0.000000
```

```
resid.scaled + result.coxph$coef
```

```
##         6        10        15        25
## -2.639193  2.157647 -3.496841 -1.326129
```

```
#obtain scaled Schoenfeld resids using cox.zph
resid.sch <- cox.zph(result.coxph)
resid.sch$y
```

```
##          trt
## 6  -2.639193
## 10  2.157647
## 15 -3.496841
## 25 -1.326129
```

```
resid.sch$x  #transformed time by 1-KM
```

```
## [1] 0.0000000 0.1666667 0.3750000 0.5833333
```

```
#attach(pancreatic2)
#pfs.month <- pfs/30.25
result.coxph <- coxph(Surv(pfs/30.25, status) ~ stage, data=pancreatic2)
summary(result.coxph)
```

```
## Call:
## coxph(formula = Surv(pfs/30.25, status) ~ stage, data = pancreatic2)
##
##   n= 41, number of events= 41
##
##          coef exp(coef) se(coef)    z Pr(>|z|)
## stageM 0.5931    1.8095   0.4007 1.48    0.139
##
##        exp(coef) exp(-coef) lower .95 upper .95
## stageM      1.81     0.5526    0.8251     3.969
##
## Concordance= 0.589  (se = 0.041 )
## Rsquare= 0.058   (max possible= 0.996 )
## Likelihood ratio test= 2.43  on 1 df,   p=0.1188
## Wald test            = 2.19  on 1 df,   p=0.1389
## Score (logrank) test = 2.25  on 1 df,   p=0.1335
```

```
#Do the KM curves cross?
plot(survfit(Surv(pfs/30.25, status) ~ stage, data=pancreatic2),
     mark.time=TRUE, lty=1:2, col=1:2)
legend('topright',levels(pancreatic2$stage),lty=1:2, col=1:2)
title('Progression-free survival versus stage')
```

Progression–free survival versus stage
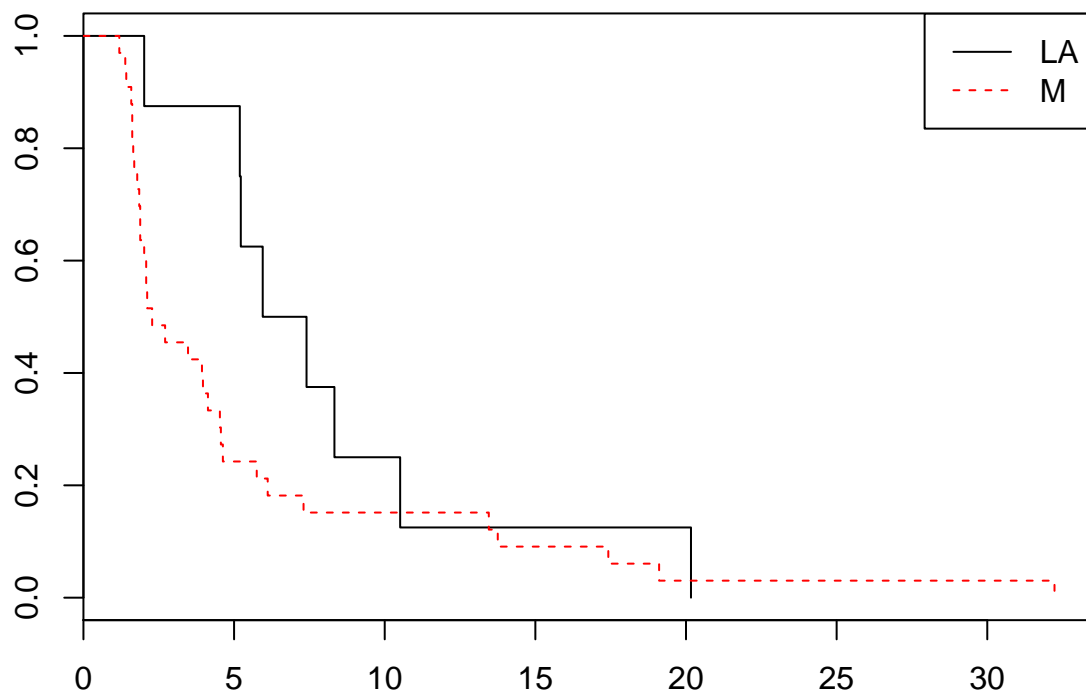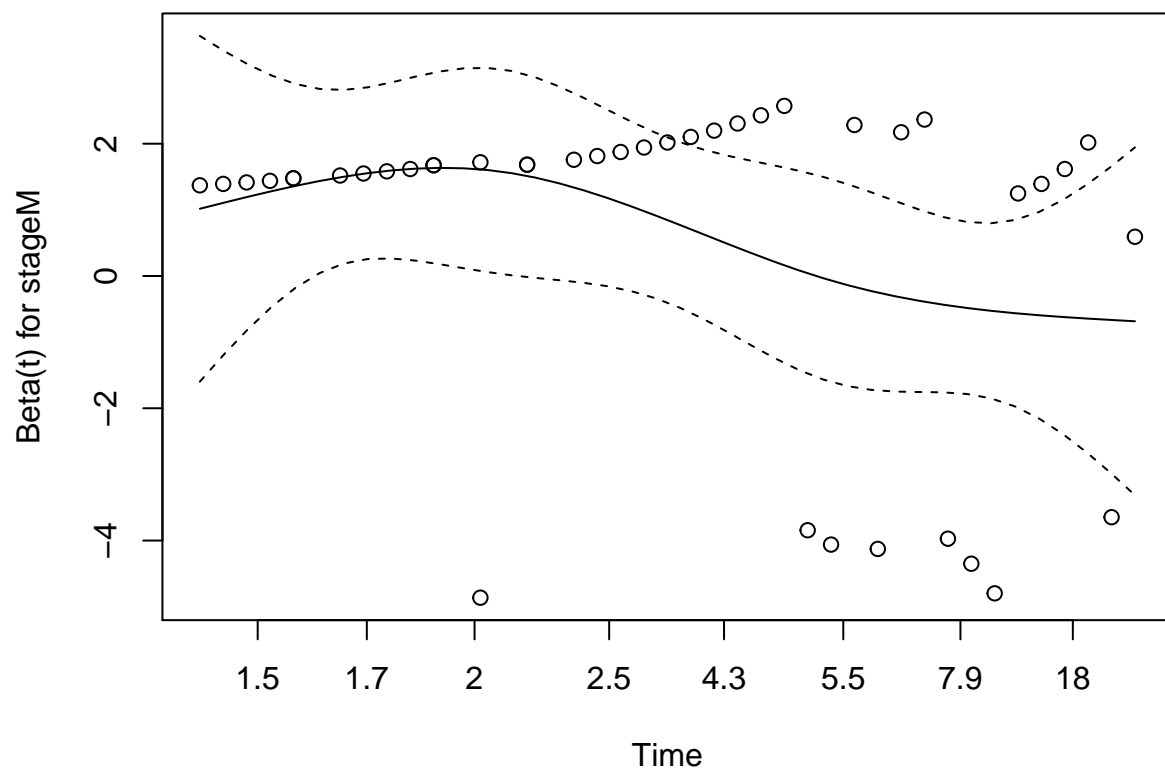


```
result.sch.resid <- cox.zph(result.coxph, transform="km")
result.sch.resid
```

```
##          rho chisq      p
## stageM -0.328  3.86 0.0496
```

```
# Fig 7.5
plot(result.sch.resid)
title('Schoenfeld residuals for pancreatic stage vs km time')
```

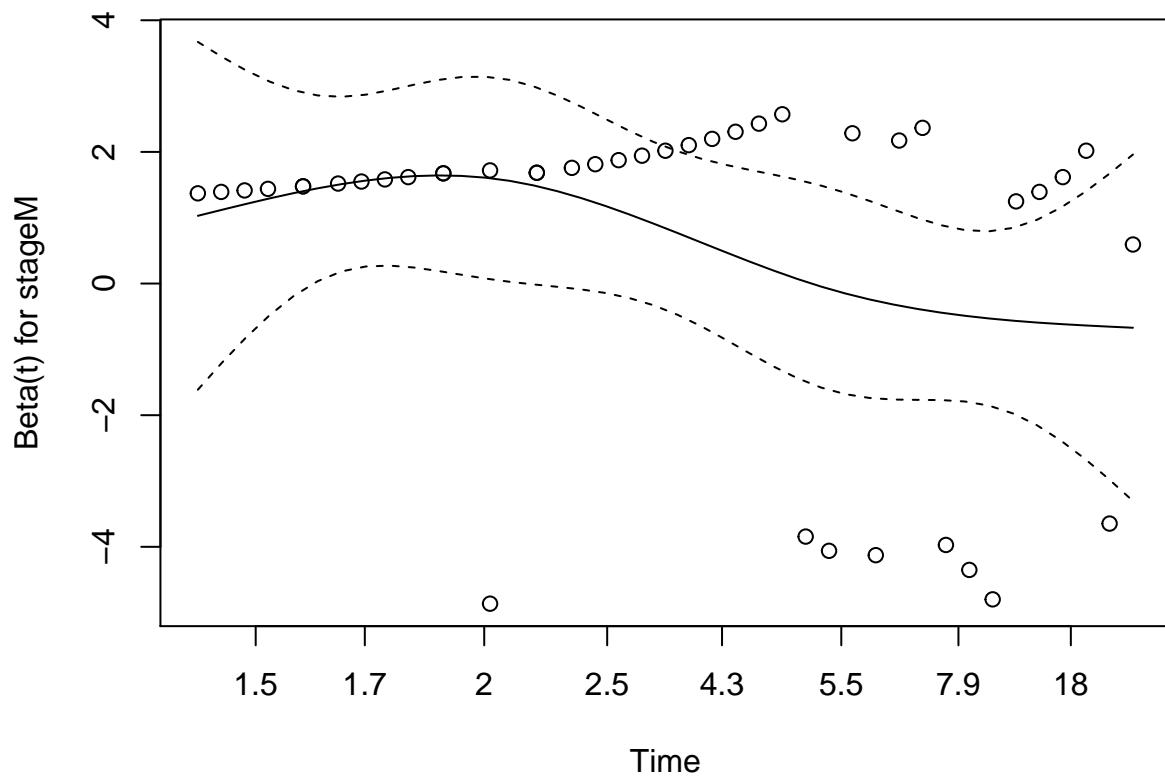**Schoenfeld residuals for pancreatic stage vs km time**


```
result.sch.rank <- cox.zph(result.coxph, transform="rank")
result.sch.rank
```

```
##          rho chisq      p
## stageM -0.33  3.89 0.0486
```

```
plot(result.sch.rank)
title('Schoenfeld residuals for pancreatic stage vs rank time')
```

Schoenfeld residuals for pancreatic stage vs rank time


```
#detach(pancreatic2)
```