SOLUTION OBLIGATORY STK4080/9080 SPRING 2021

First we read the data into R by:

```
popes = read.csv("https://folk.ntnu.no/bo/STK4080/2021/popes_25_December_2016.csv")
```

In order to have updated data, knowing that both Benedict XVI and Francis are still alive, we add the survival time from 25.12.2016 to 01.03.2021 for these popes, i.e., 4.18 years. This gives us the new data frame popes.new, obtained as follows:

```
popes.new = popes

popes.new$Survival[1]=3.79+4.18

popes.new$Survival[2]=11.70+4.18
```

As a check, you get the information of the last six popes by the command

```
head(popes.new)
```

We then introduce the working variables:

```
time = popes.new$Survival

delta = 1 - popes.new$Censored

ageelect = popes.new$Age.Election

yearelect = popes.new$Year.Elected
```

PROBLEM 1

(a)

```
Number of popes:  63
```

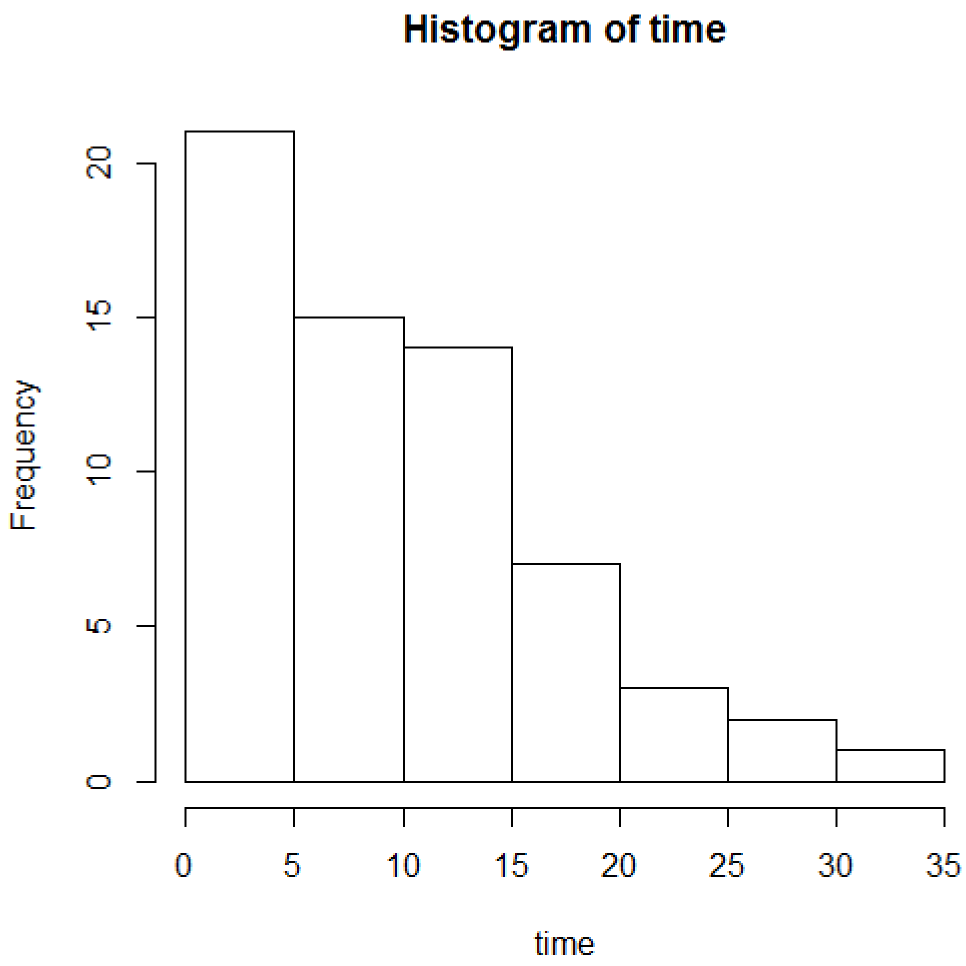
```
> summary(time)

   Min.    1st Qu.  Median  Mean  3rd Qu.   Max.

   0.03    5.00     9.00    9.89   14.50    31.00
```

```
> summary(ageelect)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  37.00   55.50   64.00   62.93   68.50   81.00
> summary(yearelect)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1404    1528    1623    1666    1788    2013
```

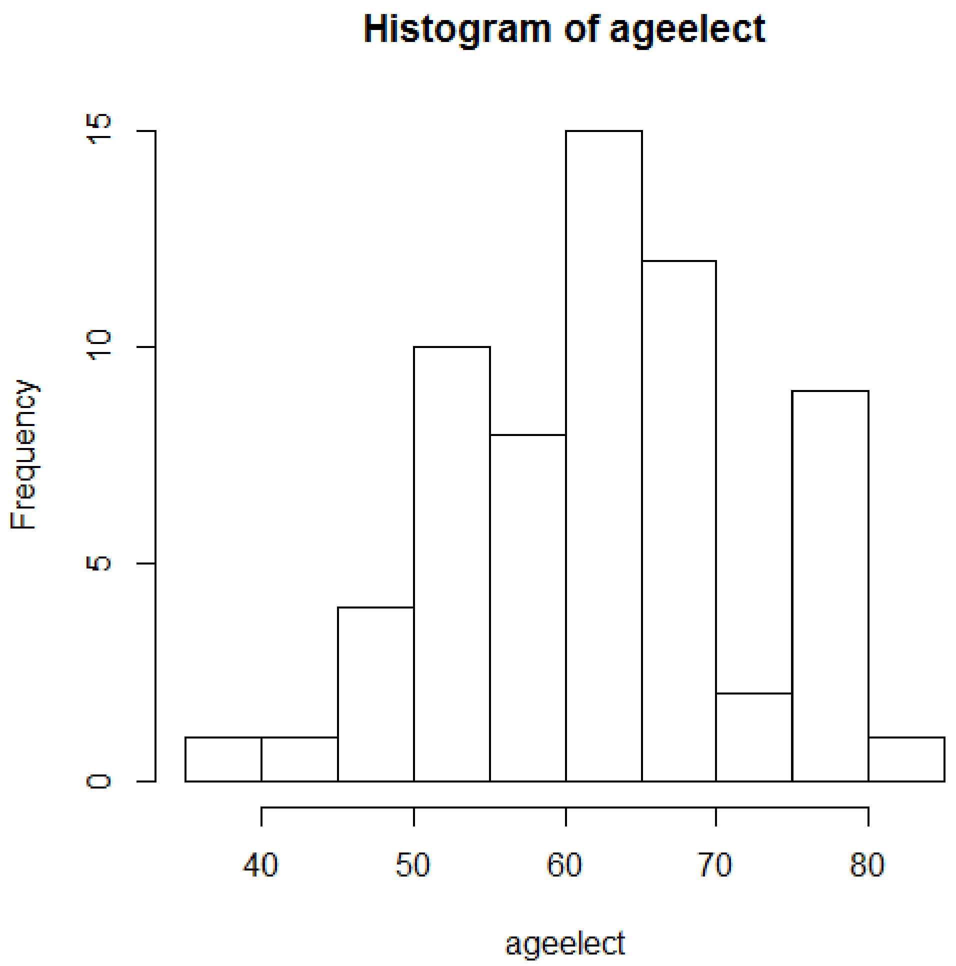
delta = 1 : 61 popes   delta = 0 : 2 popes

Plots:

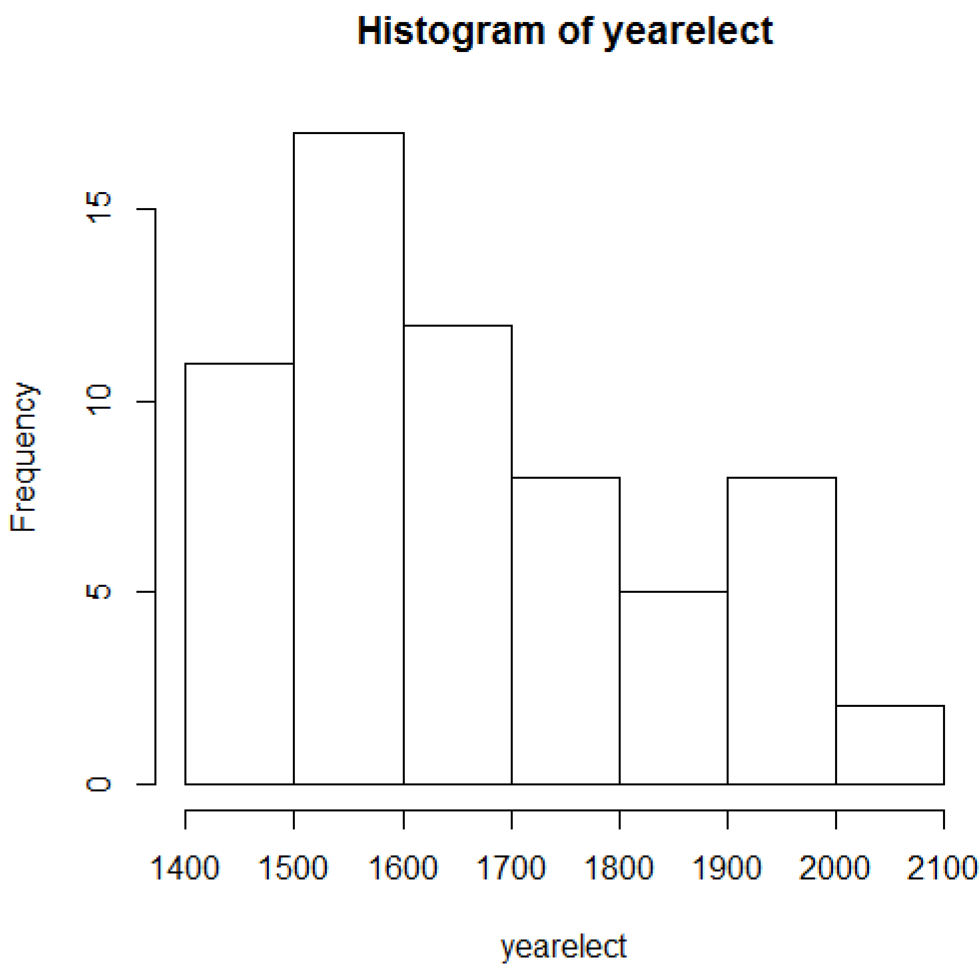
```
> hist(time)
```



```
> hist(ageelect)
```




```
> hist(yearelect)
```



Linear regression:

```
> summary(lm(ageelect~yearelect))

Call:
lm(formula = ageelect ~ yearelect)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 32.726215  11.342419   2.885  0.00540 **
```

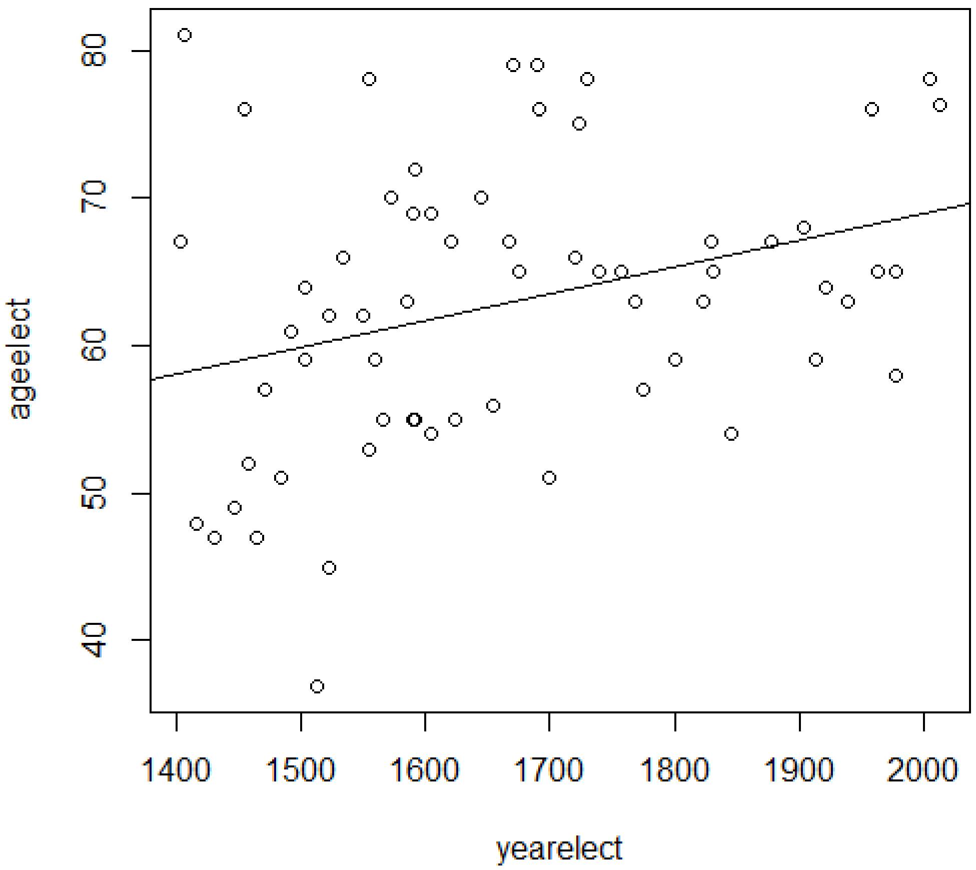
```
yearelect    0.018124   0.006771   2.677  0.00954 **
```

Estimated regression equation:

ageelect = 32.73 + 0.01812 * yearelect

Plotting:

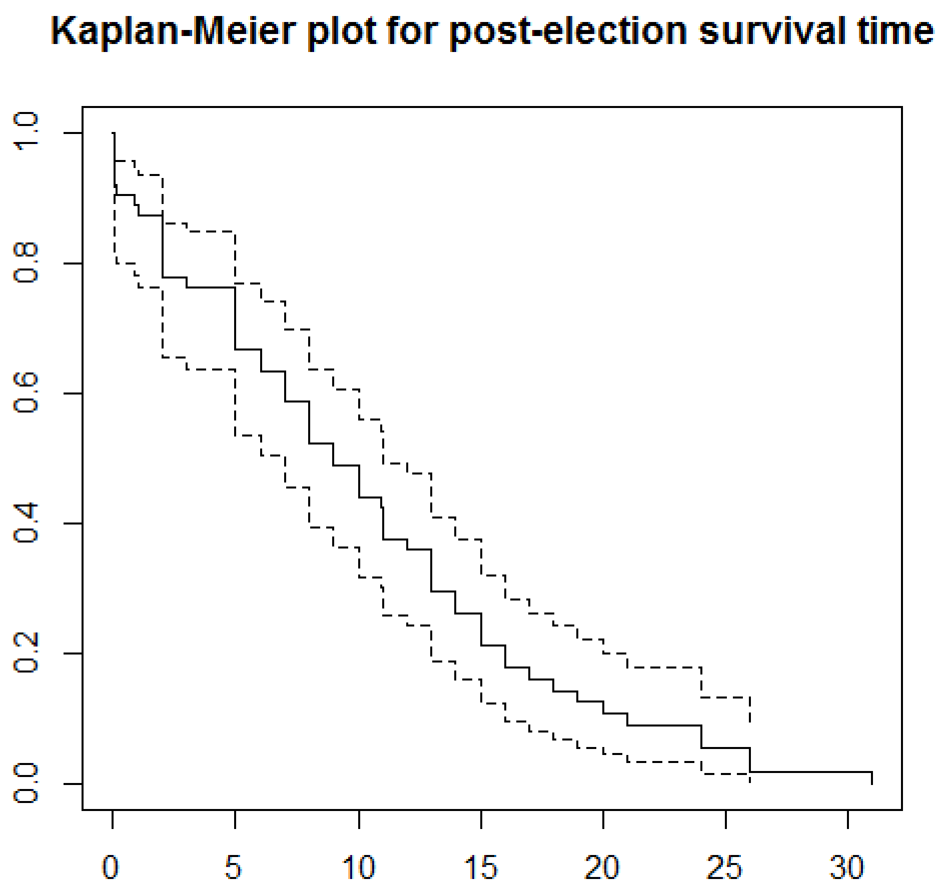
```
reg1 <- lm(ageelect~yearelect)
plot(yearelect,ageelect)
abline(reg1)
```



(b)

```
library(survival)

surv.pope = Surv(time,delta==1)


fit1=survfit(surv.pope~1, type="kaplan-meier", conf.type="log-log")

plot(fit1)
```

The confidence interval here is log-log, equation (3.30) in ABG. The method for ties is the default, i.e., “kaplan-meier”



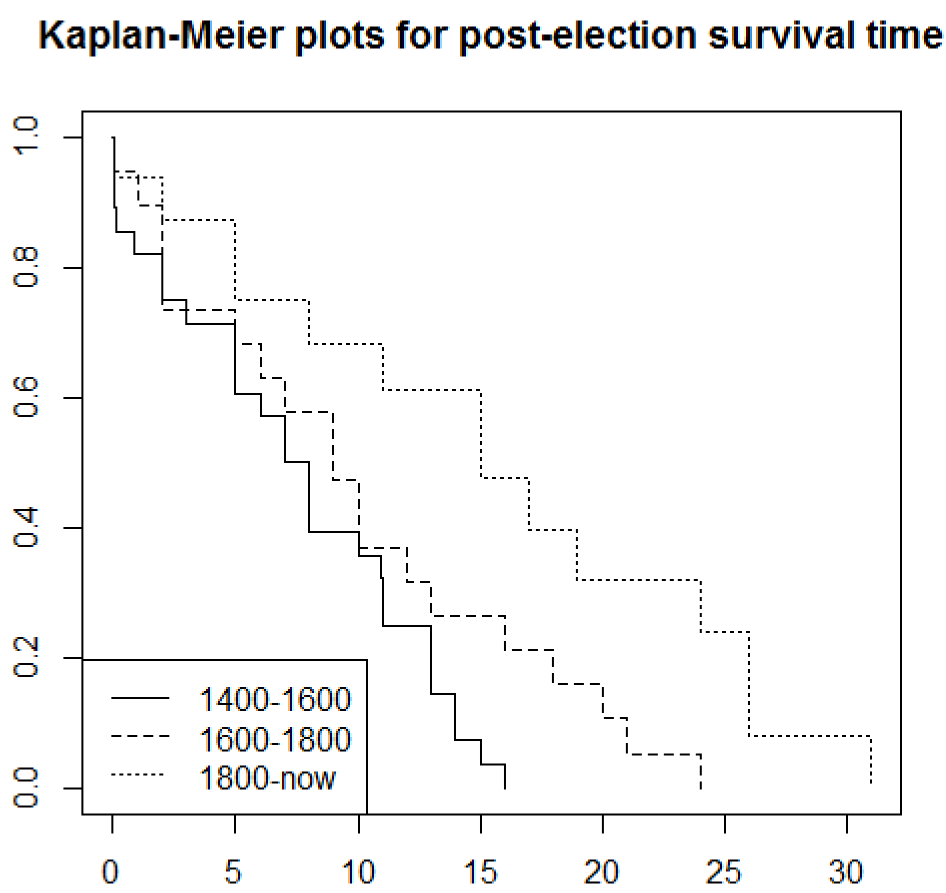

```
> quantile(fit1,probs=c(.25,.5,.75))
$quantile
25 50 75
 5  9 15

$lower
25 50 75
 2  7 12

$upper
25 50 75
 6 11 18
```

```
Thus the estimates are:

Median          9    95% conf.int.   (7,11)

Lower quartile  5    95% conf.int.   (2,6)

Upper quartile  15   95% conf.int.   (12,18)
```

(c)

Consider three groups: 1400-1599, 1600-1799, 1800-today,

```
> century = cut(yearelect, breaks = c(0,1599,1799,2021))

> fit2=survfit(surv.pope~century, conf.type="log-log")

> plot(fit2, main = "Kaplan-Meier plots for post-election survival
time",lty=1:3)

> legend("bottomleft",c("1400-1600","1600-1800","1800-now"),lty=1:3)
```



We perform a 3-sample logrank test:

```
> survdiff(surv.pope~century)

Call:

survdiff(formula = surv.pope ~ century)


                                  N Observed Expected (O-E)^2/E (O-E)^2/V

century=(0,1.6e+03]              28       28     18.2    5.3027     9.189

century=(1.6e+03,1.8e+03]        19       19     17.9    0.0711     0.111

century=(1.8e+03,2.02e+03]       16       14     24.9    4.8031    10.252


 Chisq= 13.1  on 2 degrees of freedom, p= 0.001
```

We therefore reject the null hypothesis that the survival times are equally distributed for these groups.

Looking at the Kaplan-Meier plots, it might also be of interest to do a separate comparison involving only the two periods 1400-1599, 1600-1799. There may be several ways of doing this. Below we create new vectors consisting of the popes elected before 1800.

```
> newtime = time[17:63]

> newdelta = delta[17:63]

> newyear = yearelect[17:63]

> newsurv = Surv(newtime,newdelta)

> newcentury = cut(newyear, breaks = c(0,1599,1800))

> survdiff(newsurv~newcentury)

Call:

survdiff(formula = newsurv ~ newcentury)


                                N Observed Expected (O-E)^2/E (O-E)^2/V

newcentury=(0,1.6e+03]         28       28     22.9      1.12      2.72

newcentury=(1.6e+03,1.8e+03]   19       19     24.1      1.07      2.72
```

```
 Chisq= 2.7  on 1 degrees of freedom, p= 0.1
```

There is thus no significant difference between the survival in these two groups .

PROBLEM 2

(a)

```
> surv.pope = Surv(time,delta==1)

> fit.pope=coxph(surv.pope ~ ageelect + yearelect)

> summary(fit.pope)

Call:

coxph(formula = surv.pope ~ ageelect + yearelect)


  n= 63, number of events= 61


                coef  exp(coef)   se(coef)      z Pr(>|z|)

ageelect   0.0428143  1.0437440  0.0139371  3.072  0.00213 **

yearelect -0.0034181  0.9965877  0.0008564 -3.991 6.57e-05 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


          exp(coef) exp(-coef) lower .95 upper .95

ageelect     1.0437     0.9581    1.0156    1.0726

yearelect    0.9966     1.0034    0.9949    0.9983


Concordance= 0.678  (se = 0.041 )

Likelihood ratio test= 19.72  on 2 df,   p=5e-05

Wald test            = 21.16  on 2 df,   p=3e-05

Score (logrank) test = 21.66  on 2 df,   p=2e-05
```

Model:

```
alpha(t|ageelect,yearelect) =

     alpha_0(t)*exp{beta1*ageeelect + beta2*yearelect)
```

Both covariates are highly significant, and the model is significant.

The hazard ratio for ageelect is 1.0437. This means that one additional year leads to a factor of 1.0437 in the hazard.

The hazard ratio for yearelect is 0.9966. This means, for example, that 100 additional years leads to a factor of 0.9966^100 = 0.71 in the hazard.

(b)

```
> fit.pope.inter=coxph(surv.pope ~ ageelect + yearelect +
ageelect:yearelect)

> summary(fit.pope.inter)

Call:

coxph(formula = surv.pope ~ ageelect + yearelect +
ageelect:yearelect)


  n= 63, number of events= 61


                         coef  exp(coef)   se(coef)      z  Pr(>|z|)

ageelect           -1.332e-01  8.753e-01  1.219e-01 -1.093   0.2744

yearelect          -1.100e-02  9.891e-01  5.319e-03 -2.067   0.0387
*

ageelect:yearelect  1.136e-04  1.000e+00  7.807e-05  1.456   0.1455

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


                   exp(coef) exp(-coef) lower .95 upper .95

ageelect              0.8753     1.1425    0.6893    1.1115

yearelect             0.9891     1.0111    0.9788    0.9994

ageelect:yearelect    1.0001     0.9999    1.0000    1.0003
```

```
Concordance= 0.657  (se = 0.042 )
Likelihood ratio test= 21.78  on 3 df,   p=7e-05
Wald test            = 20.27  on 3 df,   p=1e-04
Score (logrank) test = 21.8  on 3 df,   p=7e-05


> anova(fit.pope,fit.pope.inter)
Analysis of Deviance Table
 Cox model: response is  surv.pope
 Model 1: ~ ageelect + yearelect
 Model 2: ~ ageelect + yearelect + ageelect:yearelect
   loglik  Chisq Df P(>|Chi|)
1 -184.97
2 -183.94 2.0512  1    0.1521
```

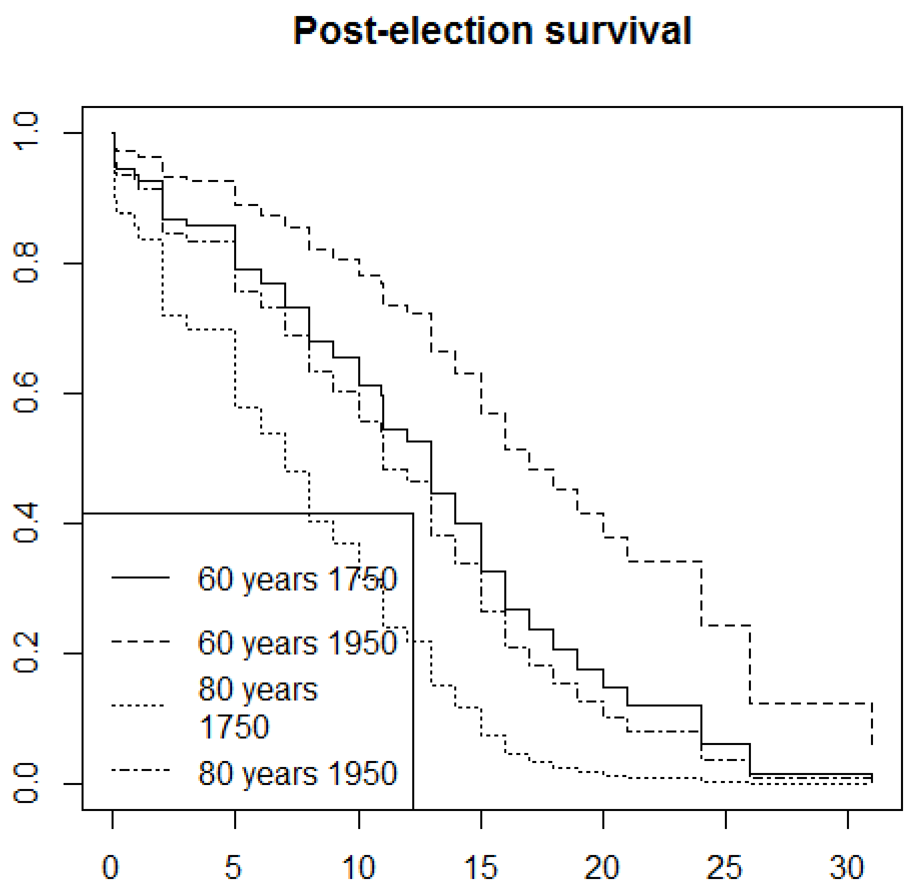
The above output says that a likelihood ratio test for the null hypothesis of no interaction has p-value 0.1521, which is not significant.

(c)

```
new.cov=data.frame(ageelect=c(60,60,80,80),yearelect=c(1750,1950,
1750,1950))
```

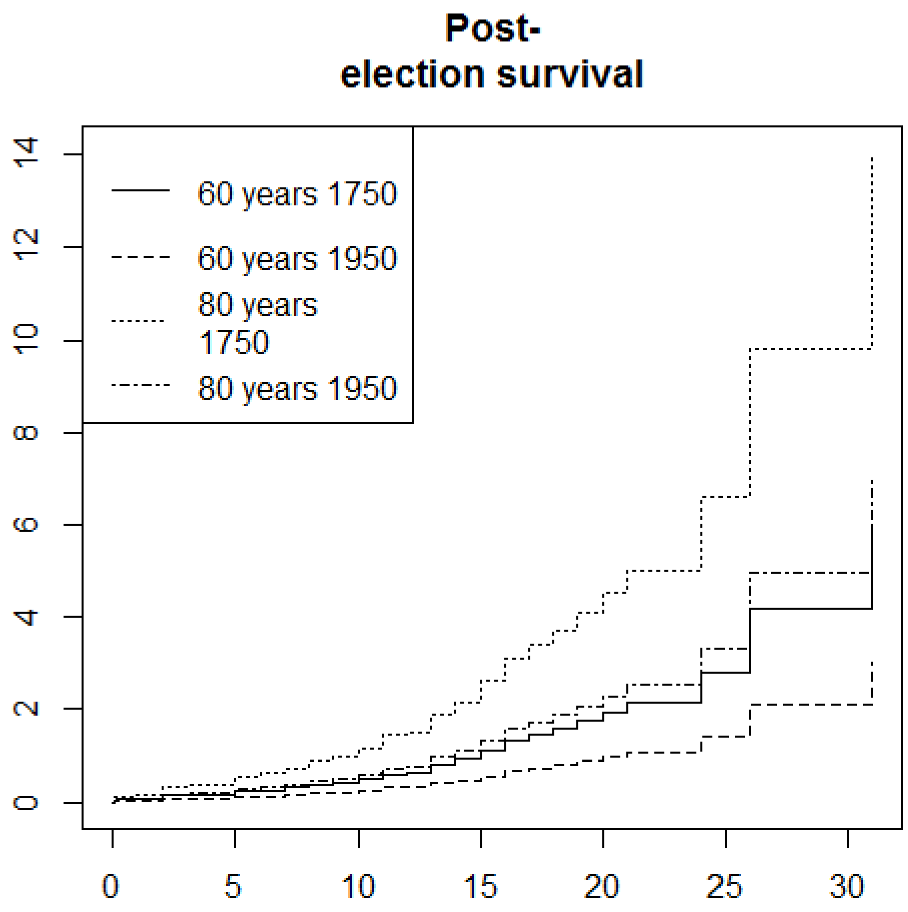
Kaplan-Meier:

```
surv.new=survfit(fit.pope,newdata=new.cov)
plot(surv.new,mark.time=F,lty=1:4, main="Post-election survival")
legend("bottomleft", c("60 years 1750","60 years 1950", "80 years
1750","80 years 1950"), lty=1:4)
```



Nelson-Aalen:

```
fit.na=coxph(Surv(time,delta==1)~ageelect + yearelect)

surv.na=survfit(fit.na,newdata=new.cov)

plot(surv.na, fun="cumhaz", mark.time=F, lty=1:4, main="Post-
election survival")

legend("topleft", c("60 years 1750","60 years 1950", "80 years
1750","80 years 1950"), lty=1:4)
```



(d)

Martingale residuals are defined in Slides 12.

We have from before:

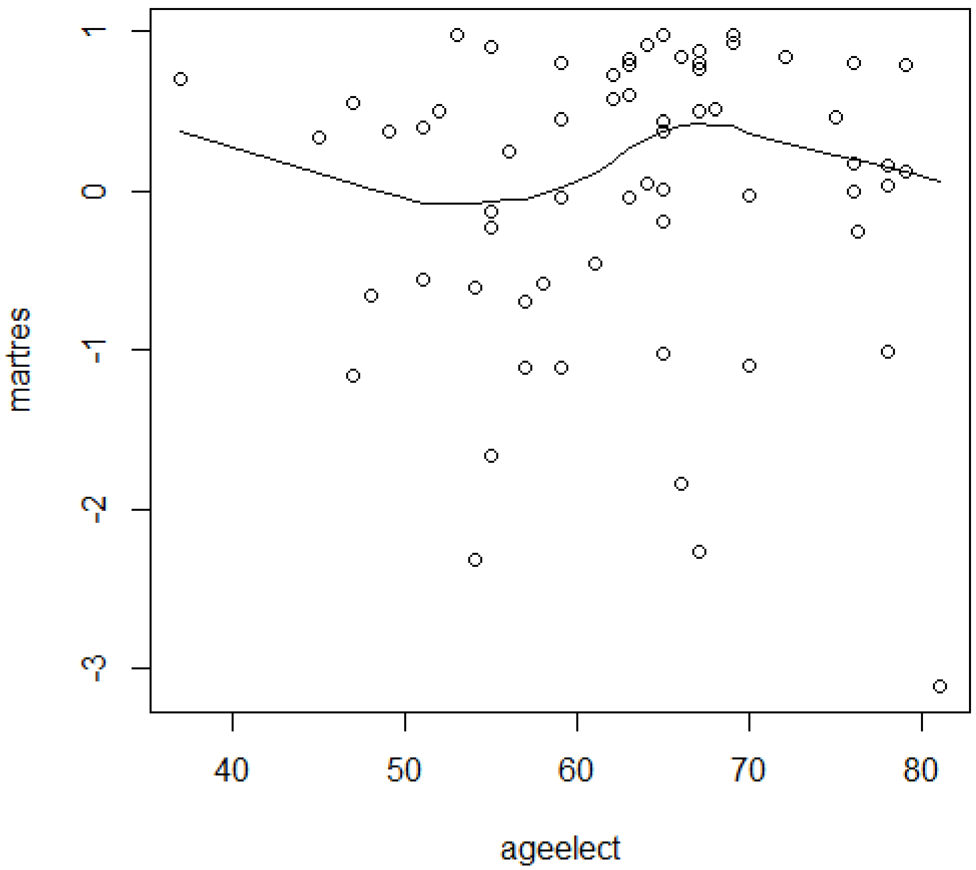
```
fit.pope=coxph(surv.pope ~ ageelect + yearelect)
```

Now we use:

```
martres = fit.pope$residuals
```

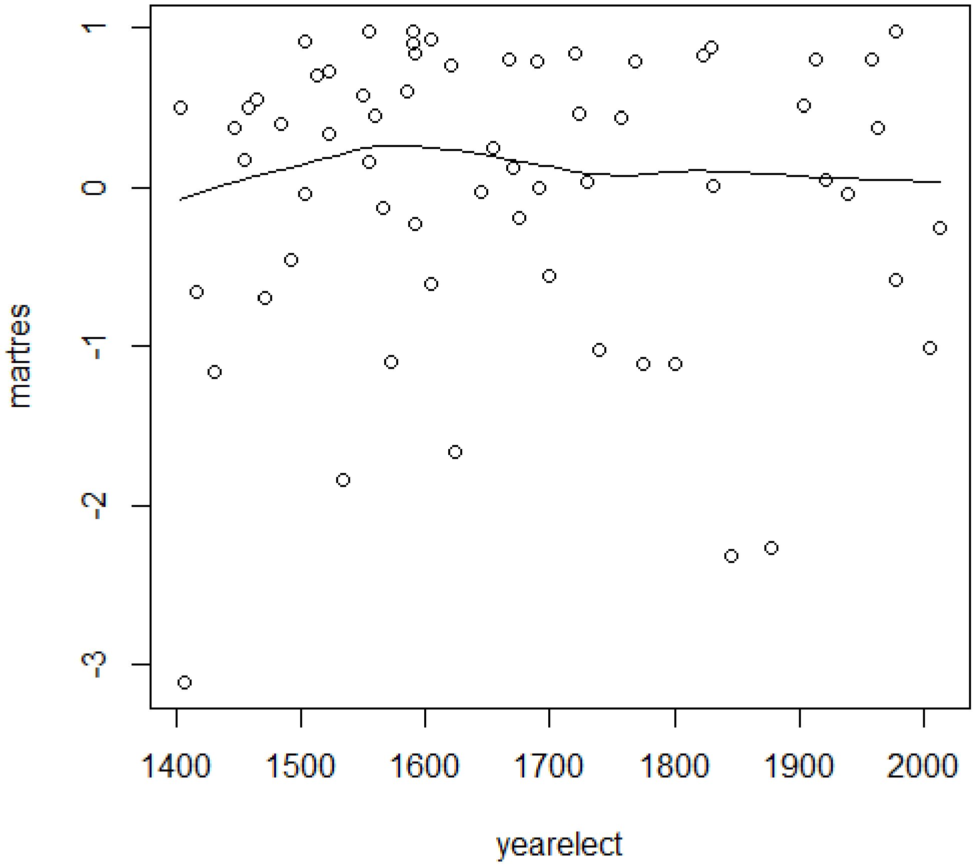
Residuals for ageelect:

```
plot(ageelect,martres)

lines(lowess(ageelect,martres))
```



Martingale residuals for yearelect:

```
plot(yearelect,martres)

lines(lowess(yearelect,martres))
```

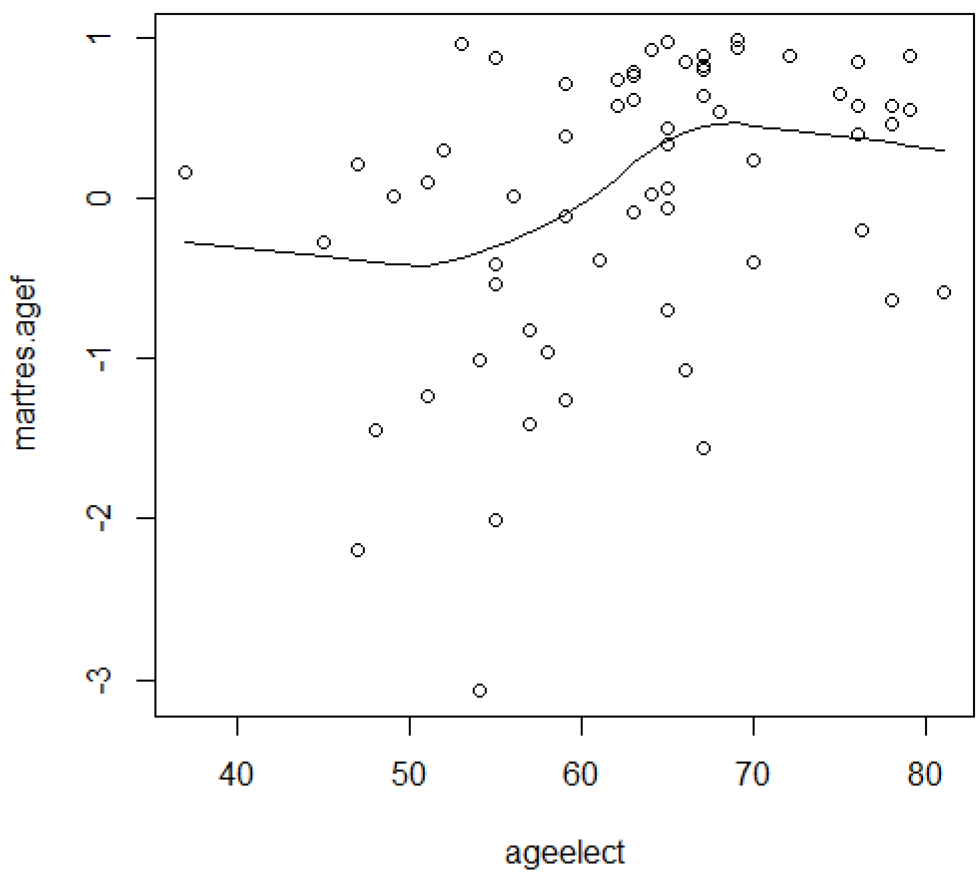



```
The plot for yeareelct looks fine, but there may be a slight problem
with ageelect (?)
```

(e)

In order to check for possible transformations of ageelect, we fit a model with only yearelect as covariate. Then we plot the martingale residual for ageelect.

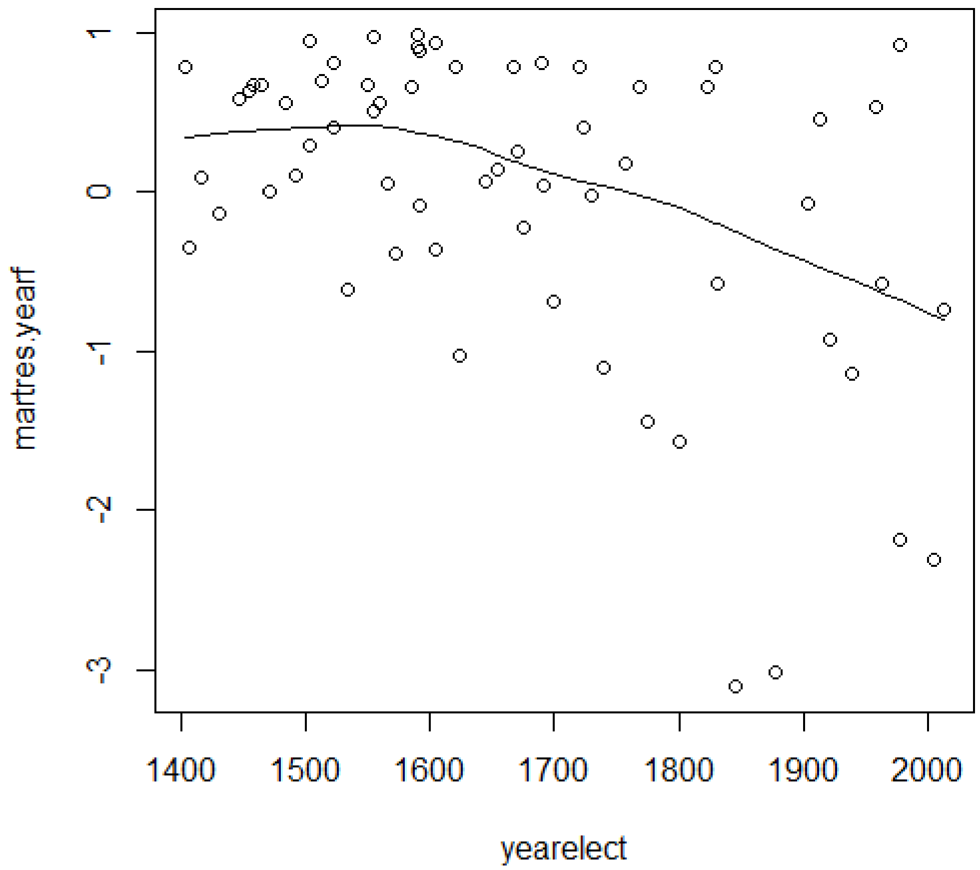
```
fit.agef=coxph(surv.pope ~ yearelect)

martres.agef = fit.agef$residuals

plot(ageelect,martres.agef)

lines(lowess(ageelect,martres.agef))
```



It might seem that the covariate ageelect shouold be transformed in some way. The behavior between 50 and 70 is OK, but the curve looks fairly horizontal up to 50 and after 70. If we ignore the part up to say 45, it might also seem that the underlying curve is logarithmic.

Below is the analysis for yearelect:

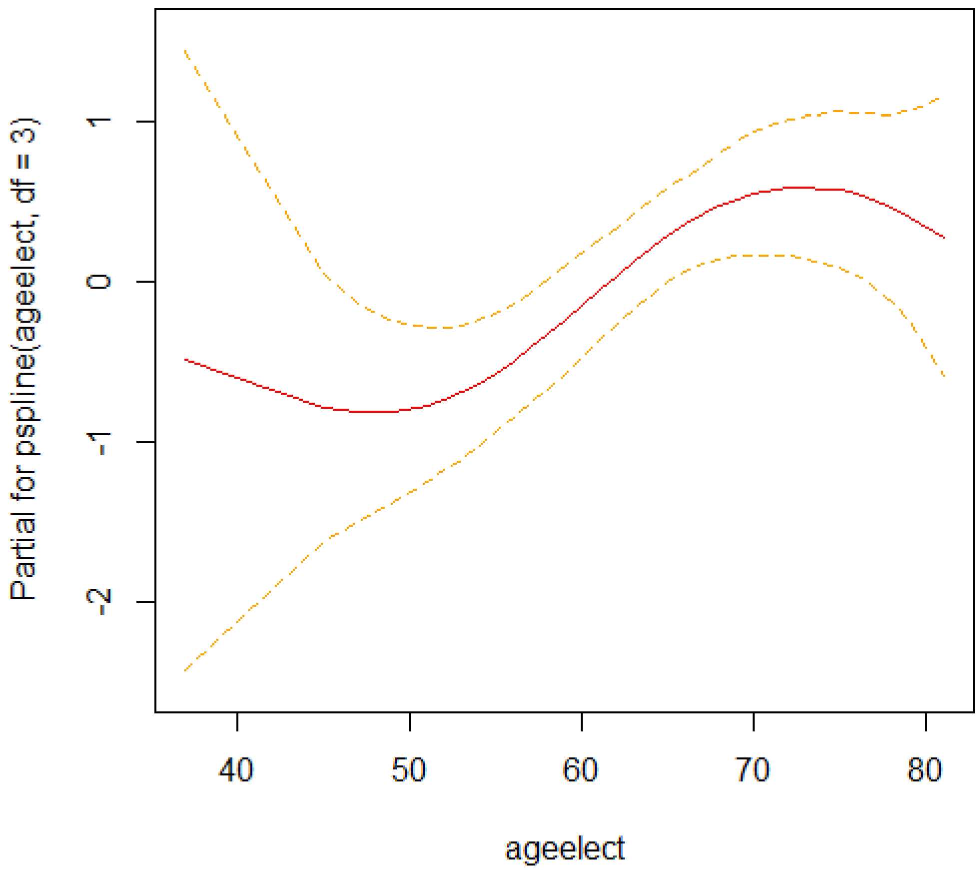
```
fit.yearf=coxph(surv.pope ~ ageelect)

martres.yearf = fit.yearf$residuals

plot(yearelect,martres.yearf)

lines(lowess(yearelect,martres.yearf))
```



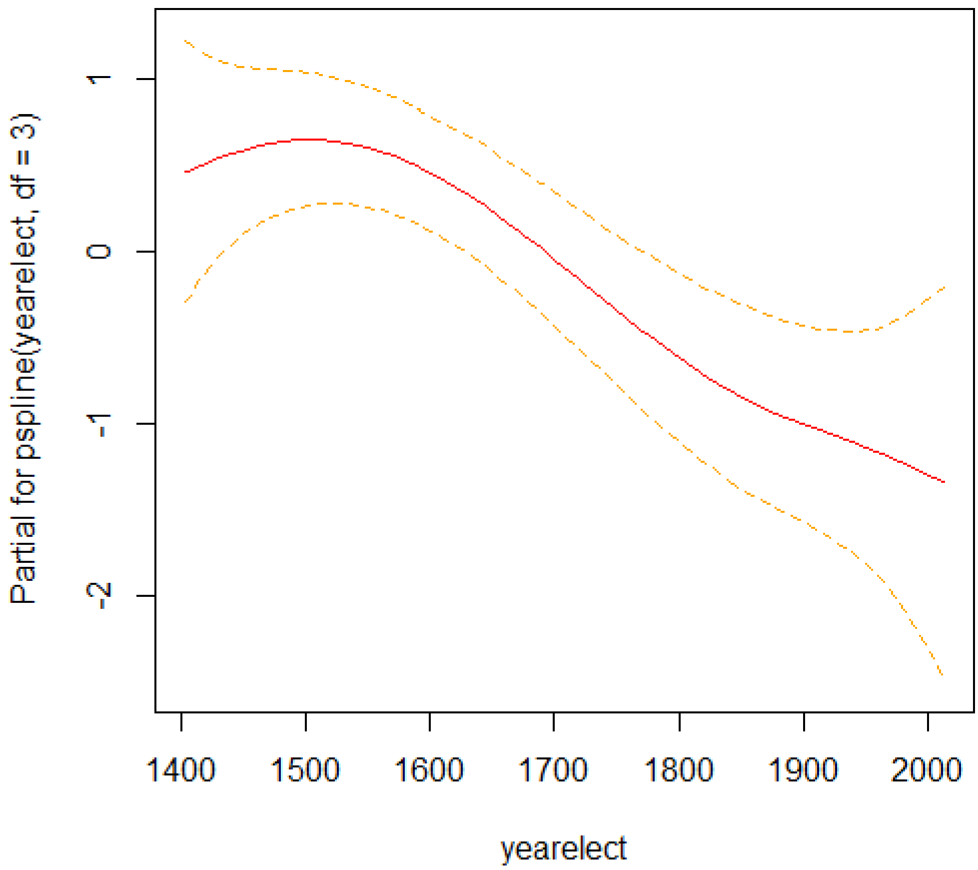
This curve looks more linear, at least starting from approximately year 1500. Thus the use of year in the model looks OK.

Use of psplines:

```
fit.pope1.spl=coxph(surv.pope ~pspline(ageelect,df=3) + yearelect)

termplot(fit.pope1.spl,se=T,terms=1)


fit.pope2.spl=coxph(surv.pope ~ageelect+pspline(yearelect,df=3))

termplot(fit.pope2.spl,se=T,terms=2)
```



Partial for pspline(ageelect, df = 3)
-2
-1
0
1
40
50
60
70
80
ageelect



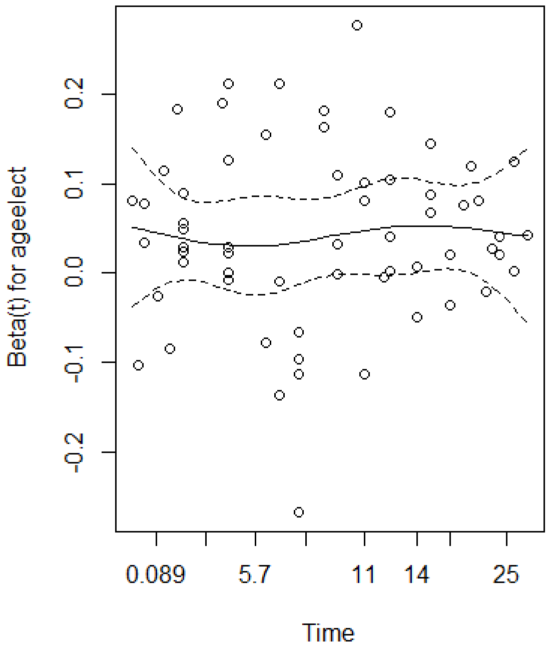
We chose df=3. The plots are then very close to the lowess smooths from the martingale residual plots. The conclusions are hence much the same.

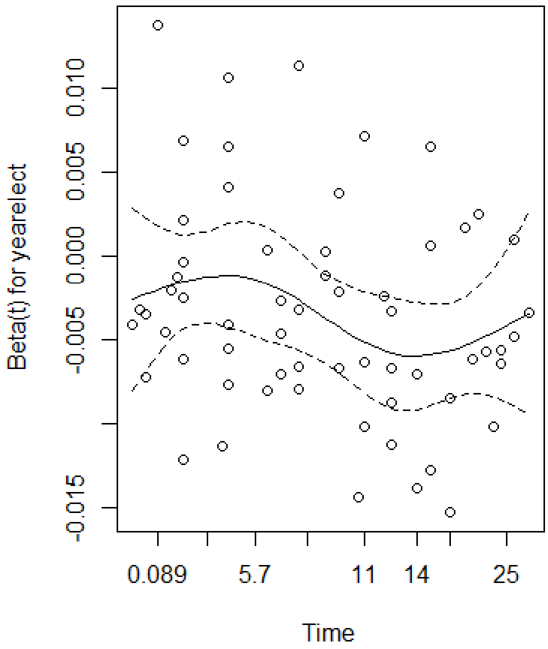
(f)

Schoenfeld residuals and their use are described in Slides 12.

```
cox.zph(fit.pope)

            chisq df     p

ageelect  0.00634  1 0.937

yearelect 3.45029  1 0.063

GLOBAL    3.74955  2 0.153
```

This output indicates that the beta-coefficient of ageelect does not change with t, while the one of yearelect might have a tendency to change with t. We now check this further by plotting. The beta looks rather constant for the ageelect, while there is a more strange behavior of beta for yearelect.



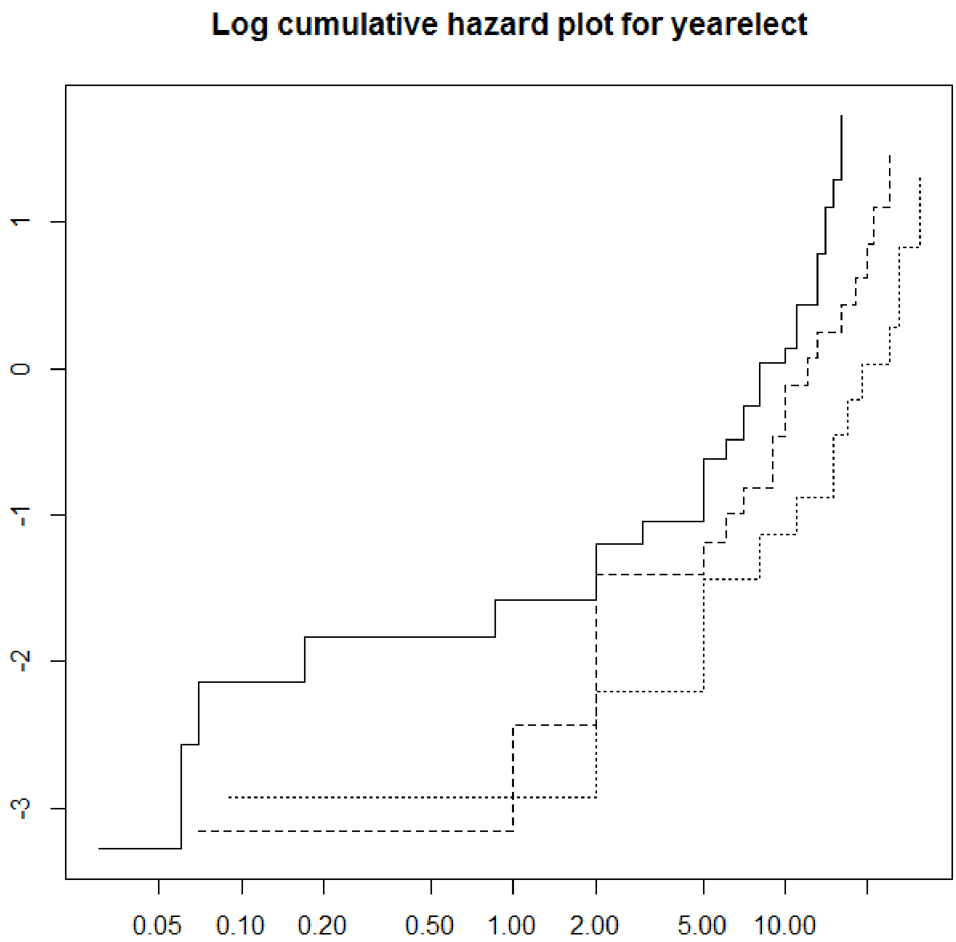


(g)

No running of R is asked for here. We are in the situation considered on page 23/40 of Slides 12. Thus we fit a Cox-model where `ageelect` is the covariate, while `yearelect` is stratified, for example using the `century` variable introduced in Problem 1c:

```
fit3=survfit(coxph(surv.pope ~ ageelect + strata(century)))
```

```
plot(fit3,fun="cloglog",main="Log cumulative hazard plot for
yearelect",lty=1:3)
```

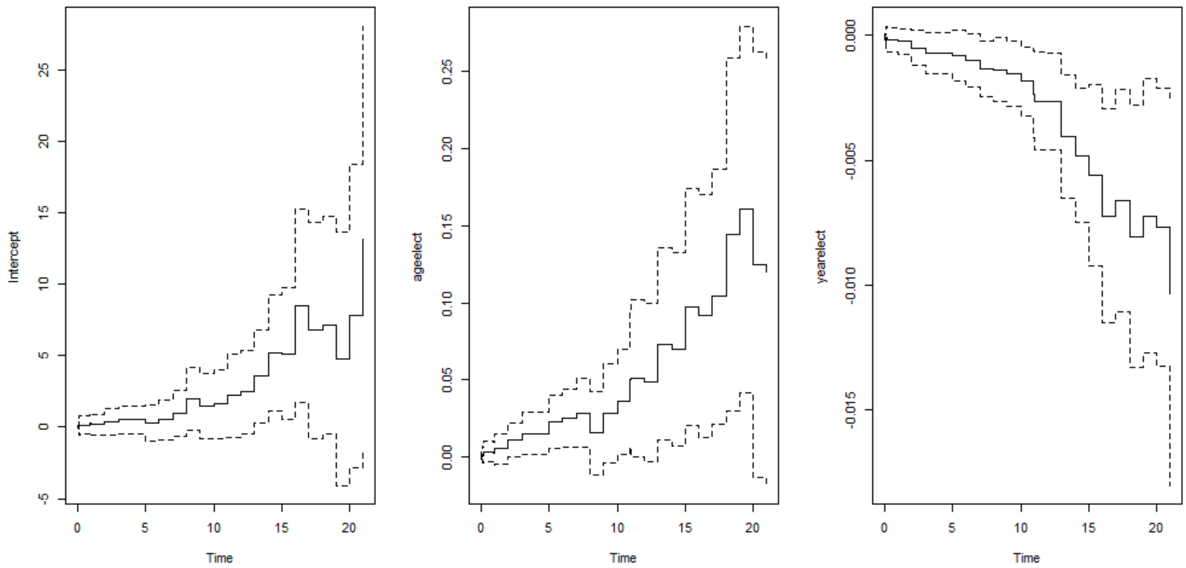


We have to check whether the curves can be viewed as parallell (i.e., having a constant vertical distance). This is perhaps questionable here, confirming the possible problem with the variable as revealed by the Schoenfeld residuals.

PROBLEM 3

Aalen's additive model

```
fit.pope.aalen=aareg(surv.pope ~ ageelect + yearelect)
```



The tendency of these plots (looking at the "derivatives") is that things are fairly constant up to time 10, but then the derivatives increase in absolute value. This effect may also be seen in the Nelson-Aalen plots in (c).

```
> print(fit.pope.aalen)

Call:

aareg(formula = surv.pope ~ ageelect + yearelect)


  n= 63

    27 out of 30 unique event times used


                slope      coef se(coef)     z        p

Intercept  0.505000  9.28e-02 3.55e-02  2.62 0.008920

ageelect   0.013100  1.30e-03 4.51e-04  2.88 0.004040

yearelect -0.000575 -8.53e-05 2.29e-05 -3.73 0.000193


Chisq=14.67 on 2 df, p=0.000651; test weights=Aalen
```

PROBLEM 4

First, restrict the dataset to popes with ageelect > 50:

```
data.popes.50 <- subset(popes.new, ageelect>50)
```

time.50 = data.popes.50$Survival

delta.50 = 1-data.popes.50$Censored

ageelect.50 = data.popes.50$Age.Election

yearelect.50 = data.popes.50$Year.Elected

Then the left truncation Surv object becomes

```
surv.pope.left = Surv(ageelect.50,ageelect.50+time.50,delta.50==1)
```

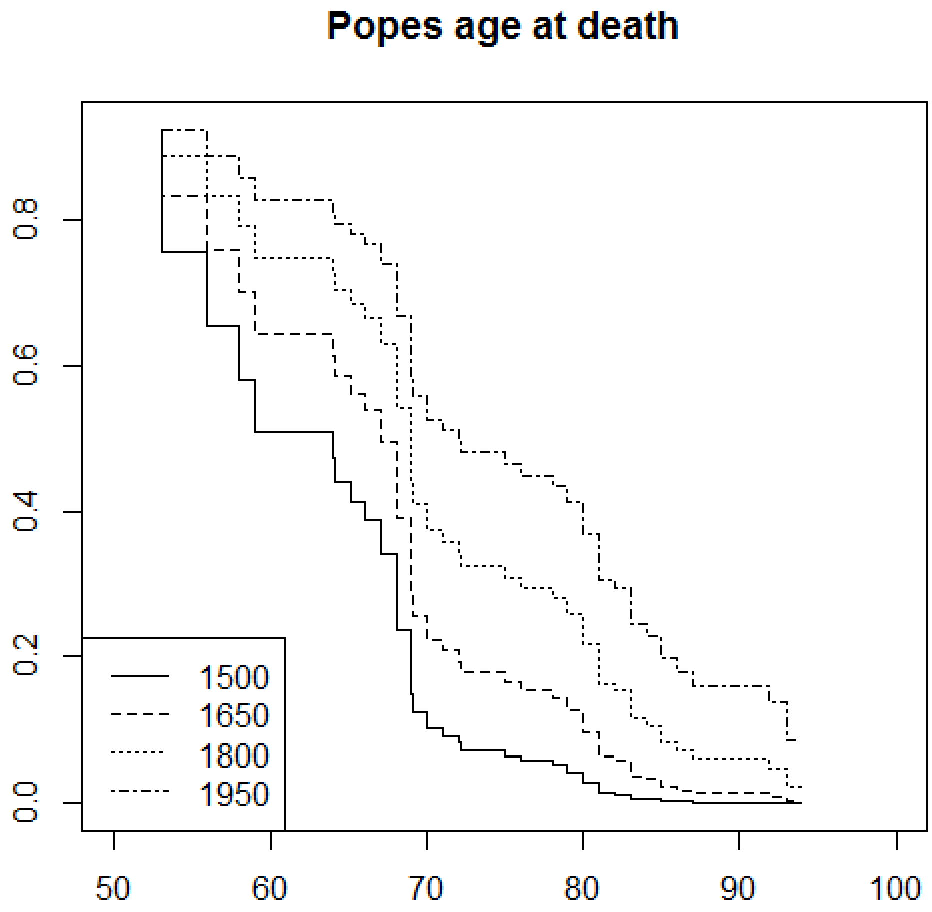
A Cox-regression with "yearelect" as covariate, is then performed by

```
fit.pope.left=coxph(surv.pope.left ~)

summary(fit.pope.left)


new.cov.left=data.frame(yearelect.50=c(1500,1650,1800,1950))
```

Kaplan-Meier plots:

These are plots for the conditional distributions of age at death for popes conditional on starting their pontification at age 50 or higher.

```
surv.new.left=survfit(fit.pope.left,newdata=new.cov.left)

plot(surv.new.left,mark.time=F,lty=1:4, xlim=c(50,100),main="Popes
age at death")

legend("bottomleft", c("1500", "1650", "1800", "1950"), lty=1:4)
```


Popes age at death
0.0
0.2
0.4
0.6
0.8
50
60
70
80
90
100
1500
1650
1800
1950