Assignment 1: Survival Analysis

Faizan Khalid Mohsin

January 23, 2019

Question 1

Question 1 d:

```
age = rep(22:29, 2)
death = c(433, 412, 337, 331, 287, 242, 215, 192,
      24, 36, 66, 102, 138, 171, 185, 200)
personyears = c(91444, 86835, 75892, 63241, 52023, 42123, 36915, 32215,
                8556, 12708, 23203, 35415, 46207, 55675, 60470, 64770)
z = c(rep(0, 8), rep(1, 8))

model2 = glm(death ~ z + factor(age) + offset(log(personyears)),
             family = poisson(link="log") )

summary(model2)
```

```
##
## Call:
## glm(formula = death ~ z + factor(age) + offset(log(personyears)),
##     family = poisson(link = "log"))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.44325  -0.14770   0.00831   0.30377   1.05482
##
## Coefficients:
##                 Estimate Std. Error  z value Pr(>|z|)
## (Intercept)    -5.348768   0.046822 -114.235  < 2e-16 ***
## z              -0.602158   0.041853  -14.388  < 2e-16 ***
## factor(age)23   0.004702   0.066494    0.071  0.94362
## factor(age)24  -0.044167   0.068449   -0.645  0.51876
## factor(age)25   0.097314   0.067515    1.441  0.14948
## factor(age)26   0.145055   0.068388    2.121  0.03392 *
## factor(age)27   0.179327   0.069648    2.575  0.01003 *
## factor(age)28   0.183552   0.070734    2.595  0.00946 **
## factor(age)29   0.197412   0.071658    2.755  0.00587 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 226.5601  on 15  degrees of freedom
## Residual deviance:   2.2334  on  7  degrees of freedom
## AIC: 131.32
##
## Number of Fisher Scoring iterations: 3
```

The coefficient of the marital status is -0.602. Therefore, its exponential is 0.5477151. Which is the rate ratio between being single and married. Hence, the mortality rate of people who are single is twice that of those who are single (after adjusting for age). Or in other words … . The hazard rate.

Question 1 e:

The below code produces the expected number of events in age/marital status.

We look at expected number of events in each age/marital status per 100,000 person years.

Need to exponentiate the coefficient estimates and we get the following:

```
coeff = model2$coefficients

alpha = coeff[1]
beta = coeff[2]

coef_vect_z0 = c(alpha, coeff[-(1:2)] + alpha )
coef_vect_z1 = c(alpha, coeff[-(1:2)] + alpha ) + beta

expected_events = data.frame(Age = 22:29,
                  Single = round(exp(coef_vect_z0)*100000, 0),
                  Married = round(exp(coef_vect_z1)*100000, 0)
                  )

rownames(expected_events) = NULL
kable(expected_events, caption = "Expected number of events for each age/marital status group.")
```

Table 1: Expected number of events for each age/marital status group.

Age	Single	Married
22	475	260
23	478	262
24	455	249
25	524	287
26	550	301
27	569	311
28	571	313
29	579	317

Question 1 f:

```
model3 = glm(death ~ z + factor(age) + z:factor(age) +
               offset(log(personyears)), family = poisson(link="log") )
summary(model3)
```

```
##
## Call:
## glm(formula = death ~ z + factor(age) + z:factor(age) + offset(log(personyears)),
##     family = poisson(link = "log"))
##
```

```
## Deviance Residuals:
##  [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##
## Coefficients:
##                  Estimate Std. Error  z value Pr(>|z|)
## (Intercept)     -5.352744   0.048057 -111.383  < 2e-16 ***
## z               -0.523590   0.209705   -2.497  0.01253 *
## factor(age)23    0.002003   0.068823    0.029  0.97679
## factor(age)24   -0.064239   0.072642   -0.884  0.37652
## factor(age)25    0.100155   0.073011    1.372  0.17013
## factor(age)26    0.152785   0.076117    2.007  0.04472 *
## factor(age)27    0.193333   0.080260    2.409  0.01600 *
## factor(age)28    0.207009   0.083430    2.481  0.01309 *
## factor(age)29    0.230052   0.086705    2.653  0.00797 **
## z:factor(age)23  0.007864   0.272362    0.029  0.97697
## z:factor(age)24  0.078191   0.249189    0.314  0.75369
## z:factor(age)25 -0.073738   0.238330   -0.309  0.75702
## z:factor(age)26 -0.090084   0.233895   -0.385  0.70013
## z:factor(age)27 -0.102622   0.232285   -0.442  0.65864
## z:factor(age)28 -0.120222   0.232449   -0.517  0.60502
## z:factor(age)29 -0.133998   0.232776   -0.576  0.56485
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 2.2656e+02  on 15  degrees of freedom
## Residual deviance: 1.4699e-13  on  0  degrees of freedom
## AIC: 143.08
##
## Number of Fisher Scoring iterations: 3
```

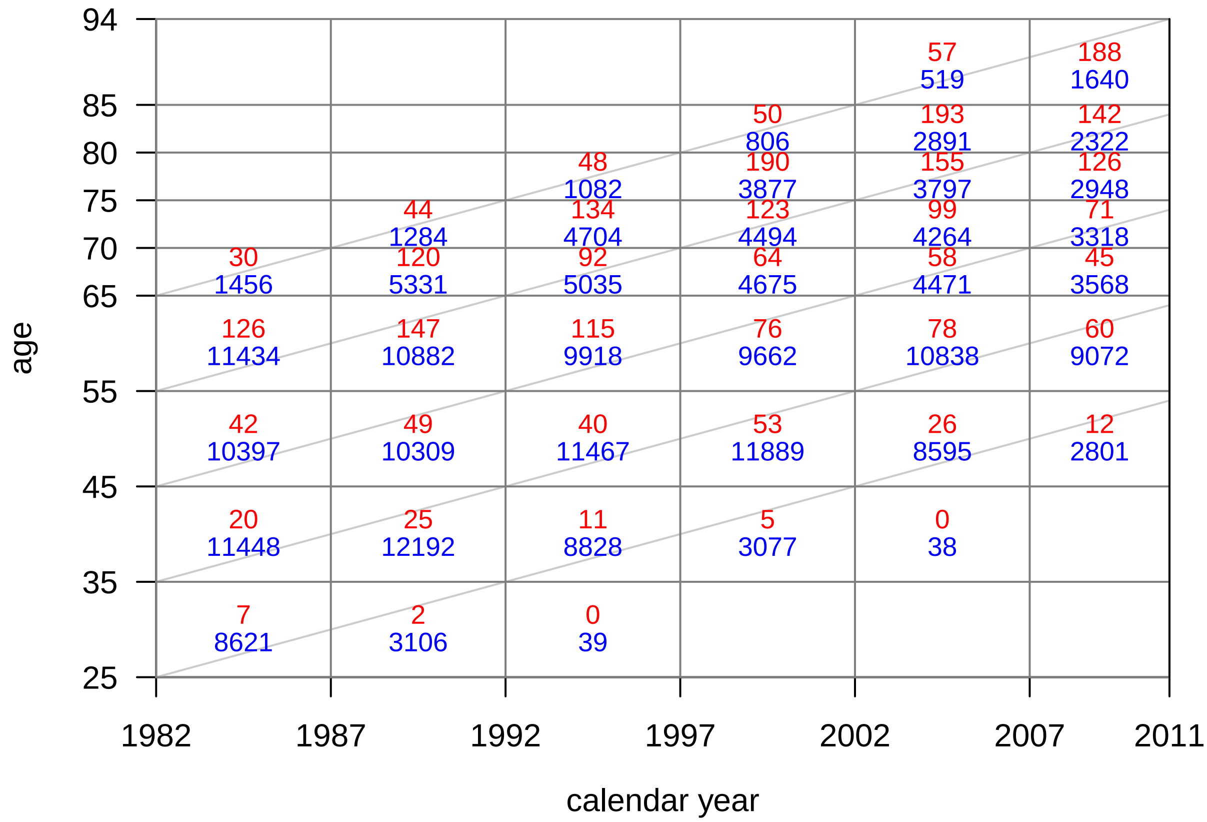
Question 4

Question 4 a:

```r
rm(list=ls())

data=read.csv("http://individual.utoronto.ca/osaarela/finrisk82.csv", sep = ";")

###### Function to transform the data.

dataset = data
dataset$events <- as.numeric(dataset$events)
dataset$followupyears <- as.numeric(dataset$followupyears)
dataset$year <- as.numeric(dataset$year)
ncol(dataset)
nrow(dataset)
by(as.numeric(dataset$events), dataset$endpoint, sum)
by(as.numeric(dataset$followupyears), dataset$endpoint, sum)

# Plot Lexis diagram:
```

```r
str(dataset)
table(dataset$endpoint)
frdeaths <- dataset[dataset$endpoint == 'DEATH',]
chddeaths <- dataset[dataset$endpoint == 'CHD2',!(names(dataset) %in% c('endpoint','followupyears'))]
names(chddeaths)[names(chddeaths) == 'events'] <- 'chd'
frdeaths <- merge(frdeaths, chddeaths)
frdeaths <- frdeaths[,!(names(frdeaths) %in% 'endpoint')]
table(frdeaths$events >= frdeaths$chd)

frdeaths$yearmid <- 1984.5 * (frdeaths$year >= 1982 & frdeaths$year < 1987) +
  1989.5 * (frdeaths$year >= 1987 & frdeaths$year < 1992) +
  1994.5 * (frdeaths$year >= 1992 & frdeaths$year < 1997) +
  1999.5 * (frdeaths$year >= 1997 & frdeaths$year < 2002) +
  2004.5 * (frdeaths$year >= 2002 & frdeaths$year < 2007) +
  2009 * (frdeaths$year >= 2007 & frdeaths$year < 2011)
frdeaths$agemid <- 30 * (frdeaths$agegr == '<35') +
  40 * (frdeaths$agegr == '35-44') +
  50 * (frdeaths$agegr == '45-54') +
  60 * (frdeaths$agegr == '55-64') +
  67.5 * (frdeaths$agegr == '65-69') +
  72.5 * (frdeaths$agegr == '70-74') +
  77.5 * (frdeaths$agegr == '75-79') +
  82.5 * (frdeaths$agegr == '80-84') +
  89.0 * (frdeaths$agegr == '>85')
fragg <- aggregate(frdeaths[,c('events','followupyears')], by=list(frdeaths$yearmid, frdeaths$agemid), 

# postscript(file.path(outpath, 'frlexis.eps'), width=6, height=6, paper='special', horizontal=FALSE)
op <- par(mar=c(4,4,0,0), mgp=c(2,1,0))
minyr <- 1982
maxyr <- 2011
minage <- 25
maxage <- 94

plot(NULL, NULL, type='n', xlim=c(minyr, maxyr), ylim=c(minage, maxage), axes=FALSE,
     xlab='calendar year', ylab='age', main='')
axis(1, at=c(seq(minyr, maxyr, by=5), 2011), las=1, pos=minage)
axis(2, at=c(25,35,45,55,65,70,75,80,85,94), las=1, pos=minyr)

ygrid <- c(seq(minyr, 2007, by=5), 2011)
agrid <- c(25,35,45,55,65)
segments(rep(minyr, length(agrid)), agrid, pmin(minyr + (maxage - agrid), maxyr), pmin(agrid + (maxyr -

ygrid <- c(seq(minyr, 2007, by=5), 2011)
agrid <- c(25,35,45,55,65,70,75,80,85,94)
segments(rep(minyr, length(agrid)), agrid, rep(maxyr, length(agrid)), agrid, col='gray50')
segments(ygrid, rep(minage, length(ygrid)), ygrid, rep(maxage, length(ygrid)), col='gray50')

lines(c(maxyr, maxyr), c(minage, maxage))

ygrid <- seq(1840, 1950, by=10)
agrid <- seq(20, 90, by=5)
for(i in 1:nrow(fragg)) {
  text(fragg[i,'Group.1'], fragg[i,'Group.2'], fragg[i,'events'], pos=3, offset=0.15, cex=0.8, col='red
```

```r
  text(fragg[i,'Group.1'], fragg[i,'Group.2'], round(fragg[i,'followupyears']), pos=1, offset=0.15, cex=
}
```




```r
par(op)
# dev.off()

# Analysis variables:

frdeaths$ageg <- 1 * (frdeaths$agegr == '<35') +
  2 * (frdeaths$agegr == '35-44') +
  3 * (frdeaths$agegr == '45-54') +
  4 * (frdeaths$agegr == '55-64') +
  5 * (frdeaths$agegr == '65-69') +
  6 * (frdeaths$agegr == '70-74') +
  7 * (frdeaths$agegr == '75-79') +
  8 * (frdeaths$agegr == '80-84') +
  9 * (frdeaths$agegr == '>85')
frdeaths$yearg <- frdeaths$year - min(frdeaths$year) + 1
frdeaths$sexg <- 0 * (frdeaths$sex == 'men') +
  1 * (frdeaths$sex == 'women')
frdeaths$area <- 0 * (frdeaths$rua == 'FIN-EASa') +
  1 * (frdeaths$rua == 'FIN-WESa')
nyears <- length(unique(frdeaths$yearg))
nagegroups <- length(unique(frdeaths$ageg))

# Poisson regression for total mortality:
```

```r
modelfit <- glm(events ~ as.factor(yearg) + as.factor(ageg) + sexg + area, offset=log(followupyears), da
summary(modelfit)
```

```r
#summary(as.factor(frdeaths$area))

# Poisson regression for CHD mortality:
model_chd <- glm(chd ~ as.factor(yearg) + as.factor(ageg) + sexg + area, offset=log(followupyears), data
summary(model_chd)
```

```
##
## Call:
## glm(formula = chd ~ as.factor(yearg) + as.factor(ageg) + sexg +
##     area, family = poisson(link = "log"), data = frdeaths, offset = log(followupyears))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4236  -0.8387  -0.2500   0.3324   2.5996
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -22.671183 644.686237  -0.035 0.971947
## as.factor(yearg)2   -0.088995   0.365978  -0.243 0.807874
## as.factor(yearg)3   -0.411659   0.385413  -1.068 0.285476
## as.factor(yearg)4   -0.522625   0.385984  -1.354 0.175734
## as.factor(yearg)5   -0.191774   0.350204  -0.548 0.583963
## as.factor(yearg)6   -0.564356   0.373972  -1.509 0.131277
## as.factor(yearg)7   -0.133568   0.337142  -0.396 0.691975
## as.factor(yearg)8   -0.003897   0.324351  -0.012 0.990415
## as.factor(yearg)9   -0.296691   0.333802  -0.889 0.374098
## as.factor(yearg)10  -0.265291   0.327366  -0.810 0.417721
## as.factor(yearg)11  -0.615961   0.342148  -1.800 0.071817 .
## as.factor(yearg)12  -0.666657   0.340316  -1.959 0.050120 .
## as.factor(yearg)13  -0.641137   0.332519  -1.928 0.053840 .
## as.factor(yearg)14  -0.755372   0.333069  -2.268 0.023334 *
## as.factor(yearg)15  -0.732211   0.327430  -2.236 0.025336 *
## as.factor(yearg)16  -0.971682   0.336107  -2.891 0.003840 **
## as.factor(yearg)17  -0.795724   0.324579  -2.452 0.014224 *
## as.factor(yearg)18  -0.790958   0.320629  -2.467 0.013629 *
## as.factor(yearg)19  -1.052493   0.327903  -3.210 0.001328 **
## as.factor(yearg)20  -1.269514   0.334395  -3.796 0.000147 ***
## as.factor(yearg)21  -0.925976   0.317807  -2.914 0.003572 **
## as.factor(yearg)22  -1.162812   0.324709  -3.581 0.000342 ***
## as.factor(yearg)23  -1.214519   0.323213  -3.758 0.000172 ***
## as.factor(yearg)24  -1.177274   0.318902  -3.692 0.000223 ***
## as.factor(yearg)25  -1.463688   0.327034  -4.476 7.62e-06 ***
## as.factor(yearg)26  -1.427742   0.324009  -4.406 1.05e-05 ***
## as.factor(yearg)27  -1.748577   0.336079  -5.203 1.96e-07 ***
## as.factor(yearg)28  -1.774635   0.335458  -5.290 1.22e-07 ***
## as.factor(yearg)29  -1.376675   0.319735  -4.306 1.66e-05 ***
## as.factor(ageg)2    14.360501 644.686394   0.022 0.982228
## as.factor(ageg)3    16.728064 644.686216   0.026 0.979299
## as.factor(ageg)4    17.910959 644.686206   0.028 0.977836
## as.factor(ageg)5    18.635925 644.686211   0.029 0.976939
## as.factor(ageg)6    19.379127 644.686213   0.030 0.976019
```

```
## as.factor(ageg)7    20.040912 644.686214   0.031 0.975201
## as.factor(ageg)8    20.649429 644.686218   0.032 0.974448
## as.factor(ageg)9    21.513103 644.686224   0.033 0.973380
## sexg                -0.948166   0.070821 -13.388  < 2e-16 ***
## area                -0.160443   0.070951  -2.261 0.023740 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 2121.50  on 719  degrees of freedom
## Residual deviance:  606.49  on 681  degrees of freedom
## AIC: 1623.9
##
## Number of Fisher Scoring iterations: 16
```

```r
# Poisson regression for CHD mortality:
# frdeaths$nonchd = frdeaths$events - frdeaths$chd

model_nonchd <- glm(events - chd ~ as.factor(yearg) + as.factor(ageg) + sexg + area, offset=log(followu
summary(model_chd)
```

```
##
## Call:
## glm(formula = chd ~ as.factor(yearg) + as.factor(ageg) + sexg +
##     area, family = poisson(link = "log"), data = frdeaths, offset = log(followupyears))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4236  -0.8387  -0.2500   0.3324   2.5996
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -22.671183 644.686237  -0.035 0.971947
## as.factor(yearg)2    -0.088995   0.365978  -0.243 0.807874
## as.factor(yearg)3    -0.411659   0.385413  -1.068 0.285476
## as.factor(yearg)4    -0.522625   0.385984  -1.354 0.175734
## as.factor(yearg)5    -0.191774   0.350204  -0.548 0.583963
## as.factor(yearg)6    -0.564356   0.373972  -1.509 0.131277
## as.factor(yearg)7    -0.133568   0.337142  -0.396 0.691975
## as.factor(yearg)8    -0.003897   0.324351  -0.012 0.990415
## as.factor(yearg)9    -0.296691   0.333802  -0.889 0.374098
## as.factor(yearg)10   -0.265291   0.327366  -0.810 0.417721
## as.factor(yearg)11   -0.615961   0.342148  -1.800 0.071817 .
## as.factor(yearg)12   -0.666657   0.340316  -1.959 0.050120 .
## as.factor(yearg)13   -0.641137   0.332519  -1.928 0.053840 .
## as.factor(yearg)14   -0.755372   0.333069  -2.268 0.023334 *
## as.factor(yearg)15   -0.732211   0.327430  -2.236 0.025336 *
## as.factor(yearg)16   -0.971682   0.336107  -2.891 0.003840 **
## as.factor(yearg)17   -0.795724   0.324579  -2.452 0.014224 *
## as.factor(yearg)18   -0.790958   0.320629  -2.467 0.013629 *
## as.factor(yearg)19   -1.052493   0.327903  -3.210 0.001328 **
## as.factor(yearg)20   -1.269514   0.334395  -3.796 0.000147 ***
## as.factor(yearg)21   -0.925976   0.317807  -2.914 0.003572 **
## as.factor(yearg)22   -1.162812   0.324709  -3.581 0.000342 ***
```

```
## as.factor(yearg)23  -1.214519   0.323213  -3.758 0.000172 ***
## as.factor(yearg)24  -1.177274   0.318902  -3.692 0.000223 ***
## as.factor(yearg)25  -1.463688   0.327034  -4.476 7.62e-06 ***
## as.factor(yearg)26  -1.427742   0.324009  -4.406 1.05e-05 ***
## as.factor(yearg)27  -1.748577   0.336079  -5.203 1.96e-07 ***
## as.factor(yearg)28  -1.774635   0.335458  -5.290 1.22e-07 ***
## as.factor(yearg)29  -1.376675   0.319735  -4.306 1.66e-05 ***
## as.factor(ageg)2    14.360501 644.686394   0.022 0.982228
## as.factor(ageg)3    16.728064 644.686216   0.026 0.979299
## as.factor(ageg)4    17.910959 644.686206   0.028 0.977836
## as.factor(ageg)5    18.635925 644.686211   0.029 0.976939
## as.factor(ageg)6    19.379127 644.686213   0.030 0.976019
## as.factor(ageg)7    20.040912 644.686214   0.031 0.975201
## as.factor(ageg)8    20.649429 644.686218   0.032 0.974448
## as.factor(ageg)9    21.513103 644.686224   0.033 0.973380
## sexg                -0.948166   0.070821 -13.388  < 2e-16 ***
## area                -0.160443   0.070951  -2.261 0.023740 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 2121.50  on 719  degrees of freedom
## Residual deviance:  606.49  on 681  degrees of freedom
## AIC: 1623.9
##
## Number of Fisher Scoring iterations: 16
```