```
R version 3.0.3 (2014-03-06) -- "Warm Puppy"
> fox = read.table("http://socserv.mcmaster.ca/jfox/Courses/soc761/Rossi.txt", header =
> dim(fox)
[1] 432  62
> fox[1:10,1:12]
   week arrest fin age race wexp mar paro prio educ emp1 emp2
1    20      1   0  27    1    0   0    1    3    3    0    0
2    17      1   0  18    1    0   0    1    8    4    0    0
3    25      1   0  19    0    1   0    1   13    3    0    0
4    52      0   1  23    1    1   1    1    1    5    0    0
5    52      0   0  19    0    1   0    1    3    3    0    0
6    52      0   0  24    1    1   0    0    2    4    0    0
7    23      1   0  25    1    1   1    1    0    4    1    1
8    52      0   1  21    1    1   0    1    4    3    0    0
9    52      0   0  22    1    0   0    0    6    3    0    0
10   52      0   0  20    1    1   0    0    0    5    0    1
> library(survival)
Loading required package: splines
> survdiff(Surv(week, arrest) ~ fin, data=fox) # log-rank test
Call:
survdiff(formula = Surv(week, arrest) ~ fin, data = fox)

        N Observed Expected (O-E)^2/E (O-E)^2/V
fin=0 216       66     55.6      1.96      3.84
fin=1 216       48     58.4      1.86      3.84

 Chisq= 3.8  on 1 degrees of freedom, p= 0.0501
> #regression version of 2-group comparison
> cox2samp <- coxph(Surv(week, arrest)    ~ fin , data=fox)
> summary(cox2samp)
Call:
coxph(formula = Surv(week, arrest) ~ fin, data = fox)
  n= 432, number of events= 114
       coef exp(coef) se(coef)      z Pr(>|z|)
fin -0.3691    0.6914   0.1897 -1.945   0.0517 .

    exp(coef) exp(-coef) lower .95 upper .95
fin    0.6914      1.446    0.4767     1.003

Concordance= 0.546  (se = 0.024 )
Rsquare= 0.009   (max possible= 0.956 )
Likelihood ratio test= 3.84  on 1 df,   p=0.05013
Wald test            = 3.78  on 1 df,   p=0.05174
Score (logrank) test = 3.83  on 1 df,   p=0.05042

> mod.allison <- coxph(Surv(week, arrest) ~ fin + age + race + wexp + mar +
                                     paro + prio, data=fox) #full cox model
> summary(mod.allison)
Call:
coxph(formula = Surv(week, arrest) ~ fin + age + race + wexp +
    mar + paro + prio, data = fox)
  n= 432, number of events= 114
         coef exp(coef) se(coef)      z Pr(>|z|)
fin  -0.37942   0.68426  0.19138 -1.983  0.04742 *
age  -0.05744   0.94418  0.02200 -2.611  0.00903 **
race  0.31390   1.36875  0.30799  1.019  0.30812
wexp -0.14980   0.86088  0.21222 -0.706  0.48029
mar  -0.43370   0.64810  0.38187 -1.136  0.25606
paro -0.08487   0.91863  0.19576 -0.434  0.66461
```

- The results of the Cox regression for time to first arrest are as follows:

Covariate	b_j	e^{b_j}	SE(b_j)	z_j	p_j
`fin`	-0.379	0.684	0.191	-1.983	$.047$
`age`	-0.057	0.944	0.022	-2.611	$.009$
`race`	0.314	1.369	0.308	1.019	$.310$
`wexp`	-0.150	0.861	0.212	-0.706	$.480$
`mar`	-0.434	0.648	0.382	-1.136	$.260$
`paro`	-0.085	0.919	0.196	-0.434	$.660$
`prio`	0.091	1.096	0.029	3.195	$.001$

  where:
  - b_j is the maximum partial-likelihood estimate of β_j in the Cox model.
  - e^{b_j}, the exponentiated coefficient, gives the effect of x_j in the multiplicative form of the model — more about this shortly.
  - SE(b_j) is the standard error of b_j, that is the square-root of the corresponding diagonal entry of the estimated asymptotic coefficient-covariance matrix.

  - $z_j = b_j/\text{SE}(b_j)$ is the Wald statistic for testing the null hypothesis H_0: $\beta_j = 0$; under this null hypothesis, z_j follows an asymptotic standard-normal distribution.
  - p_j is the two-sided p-value for the null hypothesis H_0: $\beta_j = 0$.
    - Thus, the coefficients for `age` and `prio` are highly statistically significant, while that for `fin` is marginally so.

- The estimated coefficients b_j of the Cox model give the linear, additive effects of the covariates on the log-hazard scale.
  - Although the *signs* of the coefficients are interpretable (e.g., other covariates held constant, getting financial aid decreases the hazard of rearrest, while an additional prior incarceration increases the hazard), the *magnitudes* of the coefficients are not so easily interpreted.

- It is more straightforward to interpret the exponentiated coefficients, which appear in the multiplicative form of the model,

$$\widehat{h}_i(t) = \widehat{h}_0(t) \times e^{b_1 x_{i1}} \times e^{b_2 x_{i2}} \times \cdots \times e^{b_k x_{ik}}$$

  - Thus, increasing x_j by 1, holding the other x's constant, multiplies the estimated hazard by e^{bj}.
  - For example, for the dummy-regressor `fin`, $e^{b_1} = e^{-0.379} = 0.684$, and so we estimate that providing financial aid *reduces* the hazard of rearrest — other covariates held constant — by a factor of 0.684 — that is, by $100(1 - 0.684) = 31.6$ percent.
  - Similarly, an additional prior conviction *increases* the estimated hazard of rearrest by a factor of $e^{b_7} = e^{0.091} = 1.096$ or $100(1.096 - 1) = 9.6$ percent.

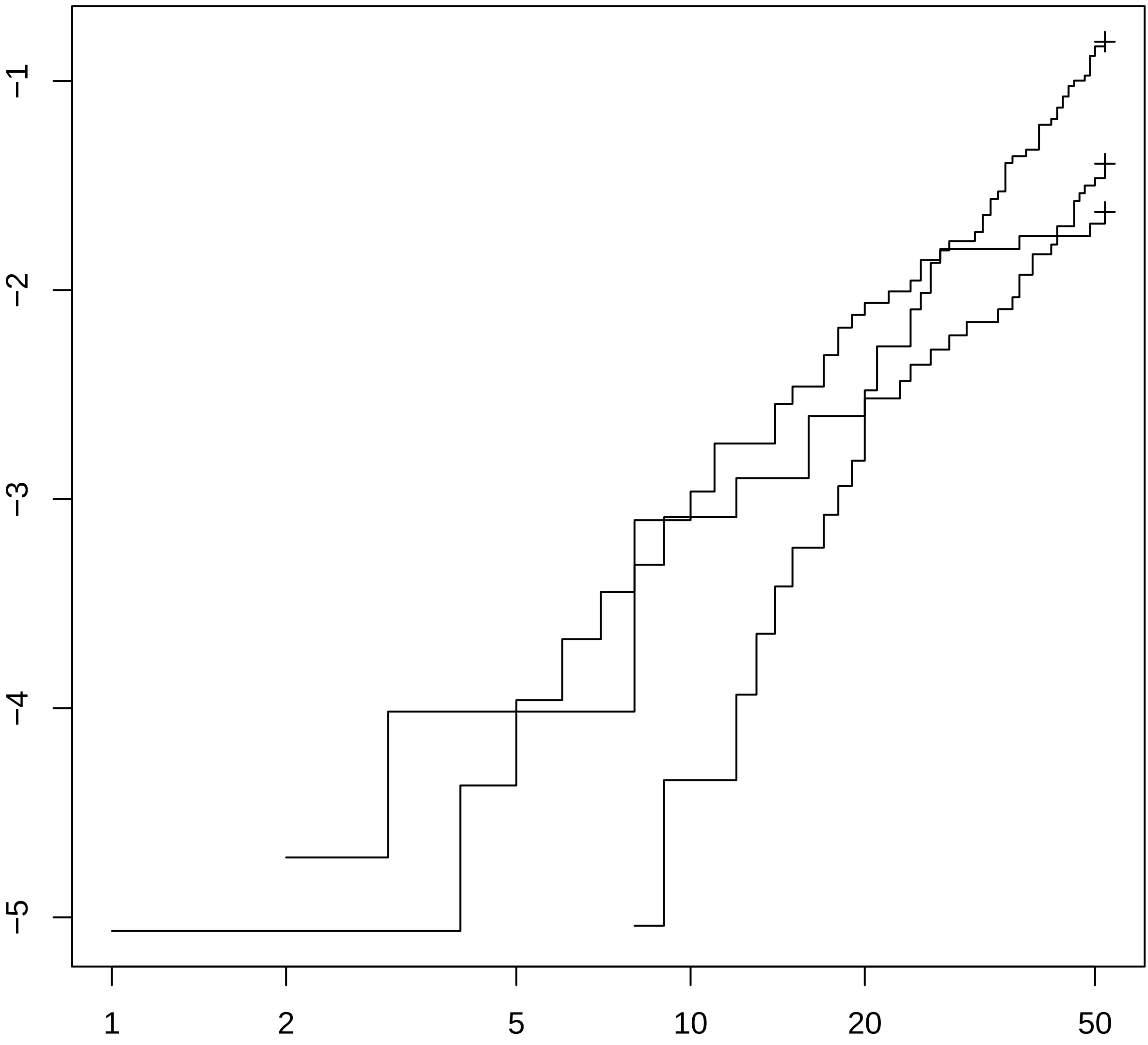
```
R version 3.0.3 (2014-03-06) -- "Warm Puppy"
> fox = read.table("http://socserv.mcmaster.ca/jfox/Courses/soc761/Rossi.txt", header =
> dim(fox)
[1] 432  62
> fox[1:10,1:12]
   week arrest fin age race wexp mar paro prio educ emp1 emp2
1    20      1   0  27    1    0   0    1    3    3    0    0
2    17      1   0  18    1    0   0    1    8    4    0    0
3    25      1   0  19    0    1   0    1   13    3    0    0
4    52      0   1  23    1    1   1    1    1    5    0    0
5    52      0   0  19    0    1   0    1    3    3    0    0
6    52      0   0  24    1    1   0    0    2    4    0    0
7    23      1   0  25    1    1   1    1    0    4    1    1
8    52      0   1  21    1    1   0    1    4    3    0    0
9    52      0   0  22    1    0   0    0    6    3    0    0
10   52      0   0  20    1    1   0    0    0    5    0    1
> library(survival)
Loading required package: splines
> survdiff(Surv(week, arrest) ~ fin, data=fox) # log-rank test
Call:
survdiff(formula = Surv(week, arrest) ~ fin, data = fox)

        N Observed Expected (O-E)^2/E (O-E)^2/V
fin=0 216       66     55.6      1.96      3.84
fin=1 216       48     58.4      1.86      3.84

 Chisq= 3.8  on 1 degrees of freedom, p= 0.0501
> #regression version of 2-group comparison
> cox2samp <- coxph(Surv(week, arrest)    ~ fin , data=fox)
> summary(cox2samp)
Call:
coxph(formula = Surv(week, arrest) ~ fin, data = fox)
  n= 432, number of events= 114
       coef exp(coef) se(coef)      z Pr(>|z|)
fin -0.3691    0.6914   0.1897 -1.945   0.0517 .

    exp(coef) exp(-coef) lower .95 upper .95
fin    0.6914      1.446    0.4767     1.003

Concordance= 0.546  (se = 0.024 )
Rsquare= 0.009   (max possible= 0.956 )
Likelihood ratio test= 3.84  on 1 df,   p=0.05013
Wald test            = 3.78  on 1 df,   p=0.05174
Score (logrank) test = 3.83  on 1 df,   p=0.05042

> mod.allison <- coxph(Surv(week, arrest) ~ fin + age + race + wexp + mar +
                                     paro + prio, data=fox) #full cox model
> summary(mod.allison)
Call:
coxph(formula = Surv(week, arrest) ~ fin + age + race + wexp +
    mar + paro + prio, data = fox)
  n= 432, number of events= 114
         coef exp(coef) se(coef)      z Pr(>|z|)
fin  -0.37942   0.68426  0.19138 -1.983  0.04742 *
age  -0.05744   0.94418  0.02200 -2.611  0.00903 **
race  0.31390   1.36875  0.30799  1.019  0.30812
wexp -0.14980   0.86088  0.21222 -0.706  0.48029
mar  -0.43370   0.64810  0.38187 -1.136  0.25606
paro -0.08487   0.91863  0.19576 -0.434  0.66461
```

```
prio  0.09150   1.09581  0.02865  3.194  0.00140 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     exp(coef) exp(-coef) lower .95 upper .95
fin     0.6843     1.4614    0.4702    0.9957
age     0.9442     1.0591    0.9043    0.9858
race    1.3688     0.7306    0.7484    2.5032
wexp    0.8609     1.1616    0.5679    1.3049
mar     0.6481     1.5430    0.3066    1.3699
paro    0.9186     1.0886    0.6259    1.3482
prio    1.0958     0.9126    1.0360    1.1591

Concordance= 0.64  (se = 0.027 )
Rsquare= 0.074   (max possible= 0.956 )
Likelihood ratio test= 33.27  on 7 df,   p=2.362e-05
Wald test            = 32.11  on 7 df,   p=3.871e-05
Score (logrank) test = 33.53  on 7 df,   p=2.11e-05

> # fin coeff changed just enough to be signif
> cox.zph(mod.allison)
            rho    chisq        p
fin     0.00646  0.00502 0.943519
age    -0.26455 11.27897 0.000784
race   -0.11224  1.41652 0.233977
wexp    0.22976  7.14021 0.007537
mar     0.07295  0.68627 0.407435
paro   -0.03618  0.15496 0.693841
prio   -0.01366  0.02304 0.879353
GLOBAL       NA 17.65862 0.013609
> # bad, global rejects, components age and wexp contributors, prop hazards assumption

> # a couple of sideshows
> attach(fox)
> fivenum(age)
[1] 17 20 23 27 44
> # Grouping variable to stratify by age
> age.group <- cut(fox$age, c(min(fox$age), 21, 26, max(fox$age)))
> table(age.group)
age.group
(17,21] (21,26] (26,44]
    159     155     112
> agefit = survfit(Surv(week, arrest) ~ age.group, data = fox)
> plot(agefit, fun = 'cloglog')
> # not so nonparallel, but test shows it strong
> plot(agefit, fun = 'cloglog')
> # Estimate of baseline hazard
> plot(basehaz(mod.allison), pch=23, bg='yellow')
> # you need to pick off and replot (like in RQ) to get axes as typical
> # Survival curve at "mean values" of covariates (as in sAS)
> plot(survfit(mod.allison), lty=c(1,2,2), col=c('red', 'green', 'green'), lwd=c(2,1,1)
> # Xmas theme

> # Schoenfeld residuals with LOWESS smoother line
> fox.schoenfeld <- residuals(mod.allison, type='schoenfeld')
> sort.time <- sort(fox$week[fox$arrest >0])
>

> # Schoenfeld residuals with LOWESS smoother line
> fox.schoenfeld <- residuals(mod.allison, type='schoenfeld')
```

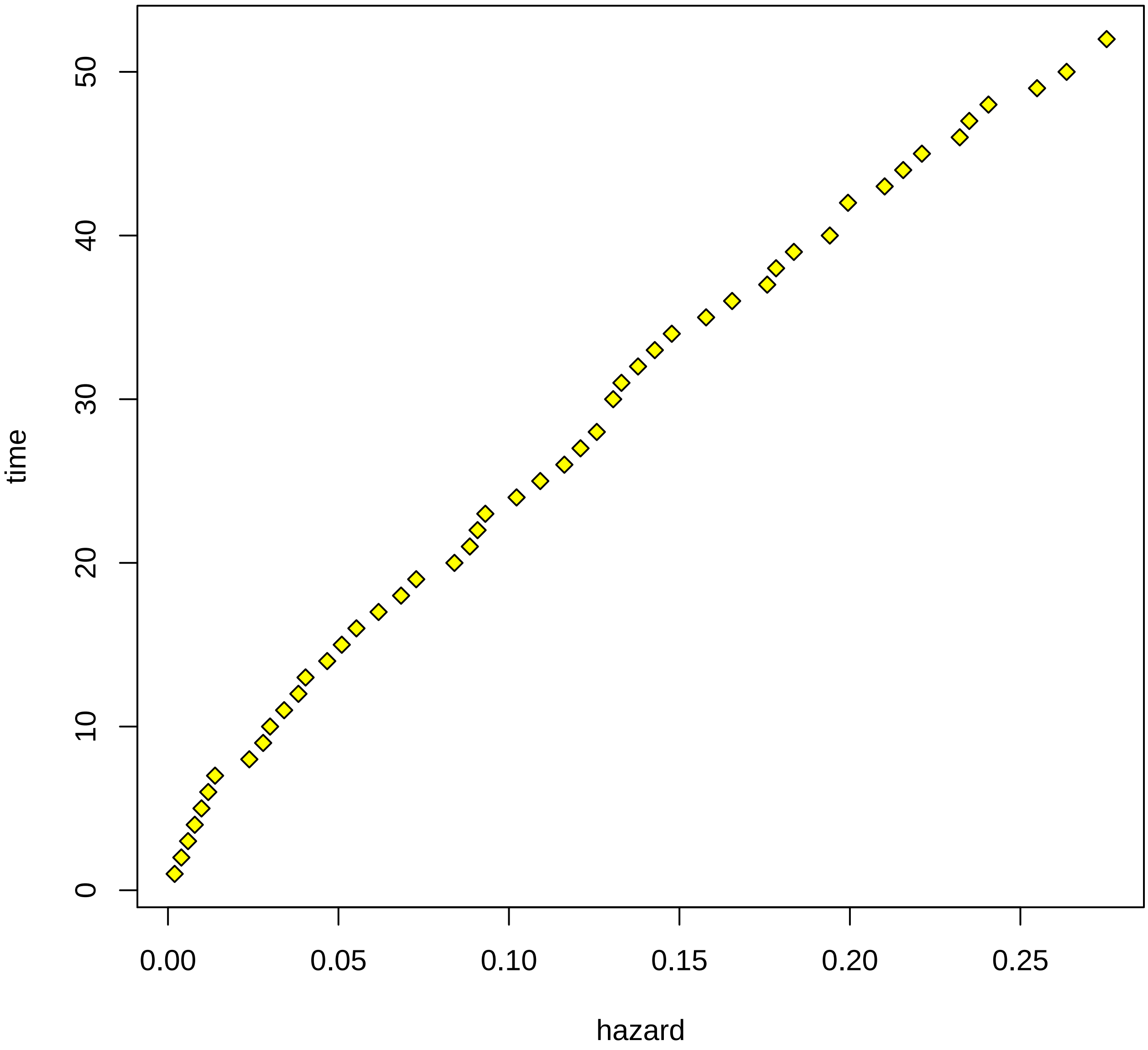
```
> sort.time <- sort(fox$week[fox$arrest >0])
>
> for (var in c('age', 'fin', 'prio')) {
+   plot(sort.time, fox.schoenfeld[,var], ylab=paste('Schoenfeld residuals for ', var),
+   lines(lowess(sort.time, fox.schoenfeld[,var]), col='red', lwd=2)
+ }
> plot(cox.zph(mod.allison))
> # i.e. you can build the cox.zph plots yourself

> ### how to work the time-dependence on age, could stratify on age-group
> mod.allisonST <- coxph(Surv(week, arrest)
   ~ fin + strata(age.group) + race + wexp + mar + paro + prio, data=fox) #stratify on
> summary(mod.allisonST)
Call:
coxph(formula = Surv(week, arrest) ~ fin + strata(age.group) +
    race + wexp + mar + paro + prio, data = fox)

  n= 426, number of events= 111
   (6 observations deleted due to missingness)

         coef exp(coef) se(coef)      z Pr(>|z|)
fin  -0.39446   0.67405  0.19327 -2.041 0.041251 *
race  0.28713   1.33260  0.30971  0.927 0.353874
wexp -0.15410   0.85719  0.21947 -0.702 0.482598
mar  -0.39777   0.67182  0.38397 -1.036 0.300241
paro -0.06869   0.93362  0.19774 -0.347 0.728302
prio  0.09534   1.10003  0.02861  3.332 0.000861 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     exp(coef) exp(-coef) lower .95 upper .95
fin     0.6740     1.4836    0.4615    0.9845
race    1.3326     0.7504    0.7262    2.4453
wexp    0.8572     1.1666    0.5575    1.3179
mar     0.6718     1.4885    0.3165    1.4259
paro    0.9336     1.0711    0.6337    1.3756
prio    1.1000     0.9091    1.0400    1.1635

Concordance= 0.605  (se = 0.047 )
Rsquare= 0.044   (max possible= 0.92 )
Likelihood ratio test= 18.95  on 6 df,   p=0.004243
Wald test            = 21.25  on 6 df,   p=0.001652
Score (logrank) test = 21.52  on 6 df,   p=0.00148

> cox.zph(mod.allisonST)
            rho    chisq       p
fin    -0.01685  0.03254 0.85686
race   -0.11049  1.35555 0.24431
wexp    0.22528  7.09299 0.00774
mar     0.06555  0.52464 0.46887
paro   -0.02561  0.07323 0.78669
prio    0.00485  0.00277 0.95800
GLOBAL       NA 11.23084 0.08150
> # that makes global nonsig, wexp is still a bit of concern, but many would declare vi
> # and fin is significant for this model !!
```

```
> ########## week 9 topic, time dependence  ##########
> # use the unfold function from John Fox to make a long form version for time dependen
      (lag 1 week) of "emp" measure
> source("http://homes.chass.utoronto.ca/~bfox/ICPSR-R-course/unfold.R") # or use from
```

cloglog for age stratification



baseline (flipped)



survival at ave covariates