Cox Proportional Hazards Model

Model Introduction and Applications

Overview

- Survival Analysis Basics
- Cox Proportional Hazards Model
  - Hazard Ratio
  - Key Assumptions
  - Limitations
- Proportionality Testing with Schoenfeld Residuals
- Application 1: Application of Stratified Cox Model
- Application 2: Laser Coagulation on Retinopathy
- Application 3: Recidivism of Male Prisoners

Survival Analysis Basics

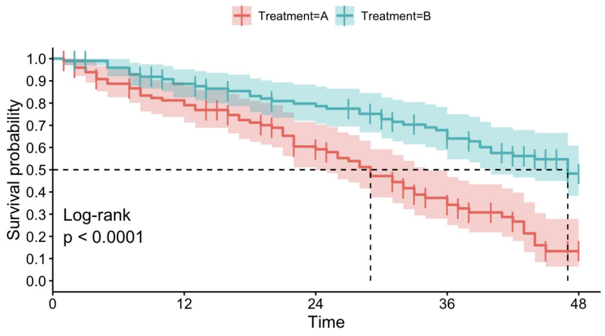
- An area of statistics that focuses on studying factors impacting “survival”
- “Survival” measures the time elapsed for an event of interest to occur
- Event types
  - Relapse, Progression, Death
- Use cases
  - Medicine - patient survival time analysis
  - Sociology - event-history analysis
  - Engineering - failure time analysis

Survival Data Characteristics

- Usually one event occurring per subject, e.g. death
- Usually positively skewed
- Prone to censoring, i.e. an event of interest does not occur for an individual during the timeframe of data collection

Methods in Survival Analysis

- Kaplan-Meier plot: non-parametric statistics used to visualize survival curves
- Log-rank test: hypothesis test to compare survival curves between groups
- Cox proportional hazards regression: regression model to study the effect of variables on survival



Cox Proportional Hazards Model

A regression model that performs multivariate analysis to investigate the connection between time and predictor variables.

$$h(t; x_i) = h_0(t)exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p) = h_0(t)exp(\beta x_i)$$

t - survival time

h(t) - hazard function, event rate at time t conditional on survival

$h_0(t)$ - baseline hazard, an arbitrary non-negative function of time

Hazard Ratio

The ratio between actual hazard rate and baseline hazard rate

$$HR = h(t)/h_0(t) = exp(\beta x_i)$$

Measures the effect of a predictor variable to event probability

HR > 1: Increase in hazard

HR = 1: No effect

HR < 1 : Decrease in hazard

Hazard Ratio

For two groups i and j, their hazard ratio is calculated as:

$$h(t;x_i)/h(t;x_j) = h_0(t)exp(\beta x_i)/h_0(t)exp(\beta x_j) = exp(\beta x_i)/exp(\beta x_j)$$

- The hazard ratio remains constant over time
- The hazard curves for different groups should be proportional and cannot cross

Key Assumptions

- Proportionality of hazards
  - $h(t)$ is proportional to $h_0(t)$, i.e. the size of the effect of the covariates are constant over time
- The covariates contribute linearly to the natural log of the hazard ratio

$$ln[h(t)/h_0(t)] = \beta x_i$$

Limitations

- The assumption of time-independence of the hazard ratio may be violated, as impacts of the covariates likely vary over time
  - Check via Schoenfeld Residuals
- The assumption of linearity may fail
  - Check via Martingale Residuals

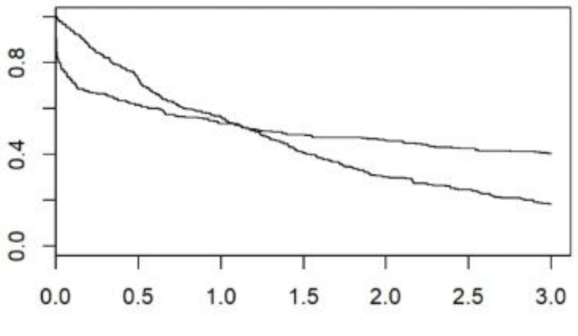
Proportionality Testing

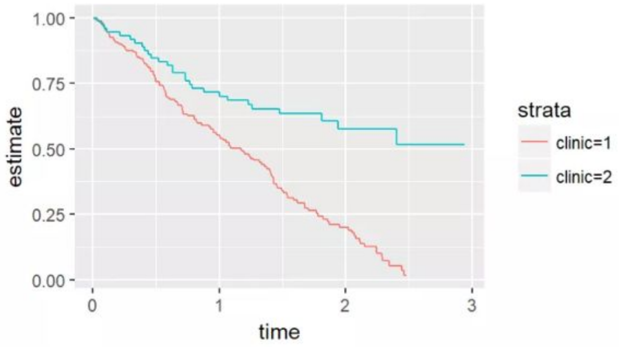
A simple way to test for proportionality is to look at the Kaplan-Meier curves.

The survival curve of two individual should never intersect.

(Consider y=2x and y=3x)

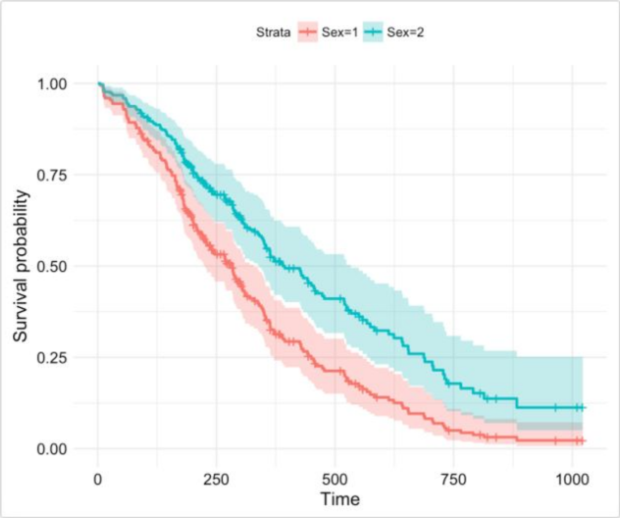
Some problematic survival curves:





Proportionality Testing

Sometimes it's hard to tell from the survival curves if the ratio is proportional. The graph method is good for testing non-proportionality.



be 100% sure the hazard

rati

en male and female is

cons

Proportionality Testing: Schoenfeld Residuals

To test proportionality accurately, we need to look at **Schoenfeld Residuals**

Schoenfeld Residuals = Observed covariate - Expected covariate

$$E[s_{t,j}] + \hat{\beta}_j = \beta_j(t)$$

Instead of a single residual for each individual, Schoenfeld residuals are computed with one per occurrence per covariate.

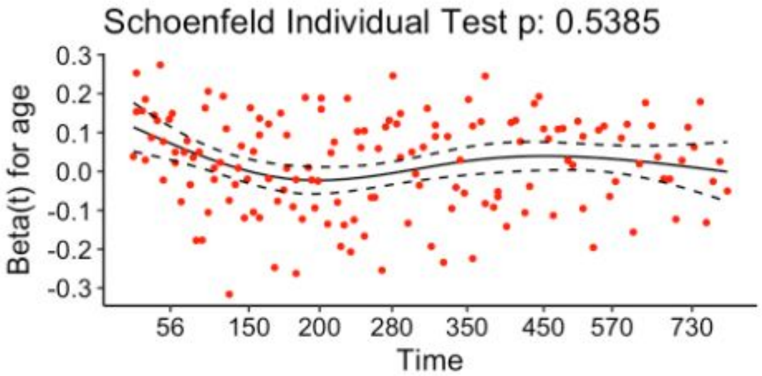
The proportional hazard assumption implies that: $\hat{\beta}_j = \beta_j(t)$

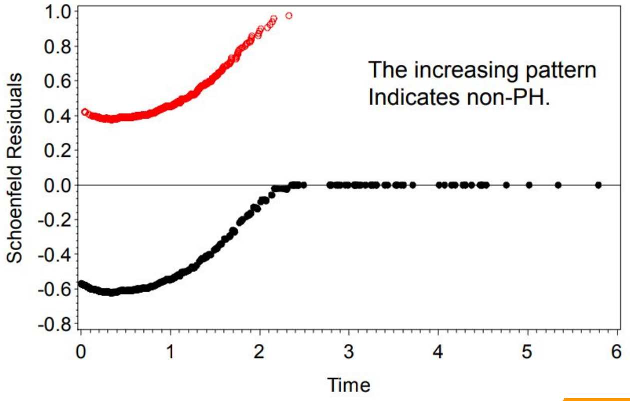
Hence: $E[s_{t,j}] = 0.$

Proportionality Testing: Schoenfeld Residuals

Assuming proportionality of the hazards, the Schoenfeld residuals are independent of time and should be centered around 0.

Thus, a plot suggesting a non-random pattern against time is evidence of non-proportionality.





Proportionality Testing: Schoenfeld Residuals

In R, we can conduct the following test to a fitted cox model to verify the proportion assumption numerically.

```
test.ph <- cox.zph(res.cox)
test.ph
```

```
              rho  chisq     p
age       -0.0483  0.378 0.538
sex        0.1265  2.349 0.125
wt.loss    0.0126  0.024 0.877
GLOBAL         NA  2.846 0.416
```

H0: Sum of Schoenfeld Residuals=0

Ha: Sum of Schoenfeld Residuals≠0

Remember, we want the Schoenfeld residual to be 0. So the optimal result would be to not reject the null. That is, the p-value should not be significant.

Conclusion: When the p-value is greater than 0.05. The proportionality assumption holds!

Application 1

Stratified Cox Model

Data Overview

	BIRTH_DATE_off	DEATH_DATE_off	sex	race	LUNG	PANCREAS	THYROID	COLORECTAL	MELANOMA	LUNG.DT	PANCREAS_DT	THYROID_DT	COLORECTAL_DT	MELANOMA_DT
1	1959/12/2	NULL	Male	Black/African-American	1	0	0	0	0	2013/9/1	#N/A	#N/A	#N/A	#N/A
2	1955/11/18	2016/4/20	Male	Black/African-American	1	0	0	0	0	2016/4/2	#N/A	#N/A	#N/A	#N/A
3	1946/1/14	2015/5/25	Female	Black/African-American	1	0	0	0	0	2015/4/11	#N/A	#N/A	#N/A	#N/A
4	1932/7/28	NULL	Female	Black/African-American	0	0	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A
5	1943/4/17	NULL	Male	White	0	0	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A
6	1950/8/20	NULL	Female	Black/African-American	0	0	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A
7	1946/11/3	NULL	Female	Black/African-American	0	0	1	0	0	#N/A	#N/A	2012/11/29	#N/A	#N/A
8	1938/6/10	NULL	Female	Black/African-American	0	0	1	0	0	#N/A	#N/A	2012/6/24	#N/A	#N/A
9	1947/2/27	NULL	Female	Black/African-American	0	0	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A
10	1957/10/18	NULL	Male	Black/African-American	1	0	0	0	0	2012/11/29	#N/A	#N/A	#N/A	#N/A
11	1954/2/12	NULL	Female	Black/African-American	0	0	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A
12	1941/8/9	NULL	Male	Black/African-American	0	0	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A
13	1940/2/4	NULL	Male	Black/African-American	1	0	0	0	0	2013/5/21	#N/A	#N/A	#N/A	#N/A
14	1971/8/11	NULL	Female	Black/African-American	0	0	0	1	0	#N/A	#N/A	#N/A	2014/1/1	#N/A
15	1948/3/22	2016/1/7	Female	Black/African-American	0	0	0	1	0	#N/A	#N/A	#N/A	2014/9/30	#N/A

This is a dataset that contains 3934 patients with one of the five cancers: lung, pancreas, thyroid, colorectal, and melanoma.

Data Overview

Use **Birth date** to extract **age** at discovery date

Use **Death date** and **Discovery date** to create **survival time**

	BIRTH_DATE_off	DEATH_DATE_off	sex	race	LUNG	PANCREAS	THYROID	COLORECTAL	MELANOMA	LUNG_DT	PANCREAS_DT	THYROID_DT	COLORECTAL_DT	MELANOMA_DT
1	1959/12/2	NULL	Male	Black/African-American	1	0	0	0	0	2013/9/1	#N/A	#N/A	#N/A	#N/A
2	1955/11/18	2016/4/20	Male	Black/African-American	1	0	0	0	0	2016/4/2	#N/A	#N/A	#N/A	#N/A
3	1946/1/14	2015/5/25	Female	Black/African-American	1	0	0	0	0	2015/4/11	#N/A	#N/A	#N/A	#N/A
4	1932/7/28	NULL	Female	Black/African-American	0	0	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A
5	1943/4/17	NULL	Male	White	0	0	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A
6	1950/8/20	NULL	Female	Black/African-American	0	0	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A
7	1946/11/3	NULL	Female	Black/African-American	0	0	1	0	0	#N/A	#N/A	2012/11/29	#N/A	#N/A
8	1938/6/10	NULL	Female	Black/African-American	0	0	1	0	0	#N/A	#N/A	2012/6/24	#N/A	#N/A
9	1947/2/27	NULL	Female	Black/African-American	0	0	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A
10	1957/10/18	NULL	Male	Black/African-American	1	0	0	0	0	2012/11/29	#N/A	#N/A	#N/A	#N/A
11	1954/2/12	NULL	Female	Black/African-American	0	0	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A
12	1941/8/9	NULL	Male	Black/African-American	0	0	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A
13	1940/2/4	NULL	Male	Black/African-American	1	0	0	0	0	2013/5/21	#N/A	#N/A	#N/A	#N/A
14	1971/8/11	NULL	Female	Black/African-American	0	0	0	1	0	#N/A	#N/A	#N/A	2014/1/1	#N/A
15	1948/3/22	2016/1/7	Female	Black/African-American	0	0	0	1	0	#N/A	#N/A	#N/A	2014/9/30	#N/A

Use **Death Date** to create **survival status:**
Null=0 (censored) non-null=1 (dead)

The observation is assumed to end on 2016-9-15.

Model Results

	time	status	age	sex	race
1	1110	0	53	Male	Black/African-American
2	18	1	60	Male	Black/African-American
3	44	1	69	Female	Black/African-American
10	1386	0	55	Male	Black/African-American
13	1213	0	73	Male	Black/African-American
16	4	1	71	Female	Black/African-American
23	65	1	60	Female	Black/African-American
26	610	0	61	Male	Black/African-American

```
Call:
coxph(formula = Surv(time, status) ~ age + sex + race, data = LUNG)

  n= 1408, number of events= 183

                                                coef  exp(coef)   se(coef)      z Pr(>|z|)
age                                        8.985e-03  1.009e+00  6.703e-03  1.340    0.180
sexMale                                    1.704e-01  1.186e+00  1.491e-01  1.143    0.253
raceBlack/African-American                 7.749e-01  2.170e+00  1.521e-01  5.095 3.48e-07 ***
raceAsian/Mideast Indian                   2.814e-01  1.325e+00  4.598e-01  0.612    0.541
raceNative Hawaiian/Other Pacific Islander -1.384e+01  9.715e-07  1.398e+03 -0.010    0.992
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                                           exp(coef) exp(-coef) lower .95 upper .95
age                                        1.009e+00  9.911e-01    0.9959     1.022
sexMale                                    1.186e+00  8.434e-01    0.8853     1.588
raceBlack/African-American                 2.170e+00  4.607e-01    1.6110     2.924
raceAsian/Mideast Indian                   1.325e+00  7.547e-01    0.5381     3.263
raceNative Hawaiian/Other Pacific Islander 9.715e-07  1.029e+06    0.0000       Inf

Concordance= 0.625  (se = 0.021 )
Likelihood ratio test= 30.45  on 5 df,   p=1e-05
Wald test            = 29.71  on 5 df,   p=2e-05
Score (logrank) test = 31.84  on 5 df,   p=6e-06
```

Base race is White

```
res.cox <- coxph(Surv(time, status) ~ age + sex + race, data =  LUNG)
```

PH Assumption Testing

```
#Test for PH assumption
test.ph <- cox.zph(res.cox)
test.ph
```

```
        chisq df      p
age       6.9  1 0.0086
sex       1.4  1 0.2363
race      5.8  3 0.1219
GLOBAL   13.4  5 0.0201
```

Both age and global p-value is less than 0.05, indicating the PH assumption is violated for age and the whole model

Stratification

The idea of this method is to split the analysis time into several intervals and Cox proportional model is stratified for these time intervals. The effect of fixed baseline covariates becomes stronger or weaker over time, which can be explored via stratification by time.

Essentially, this will let the baseline hazard function differ between subsamples.

Eg. if variable sex violates the PH assumption, use stratification to fit a stratified Cox model with different baseline hazard functions for each sex.

Eg. If variable age violates the PH assumption, use stratification to fit a stratified Cox model with different baseline hazard functions for different time intervals based on the pattern on the Schoenfeld residuals.

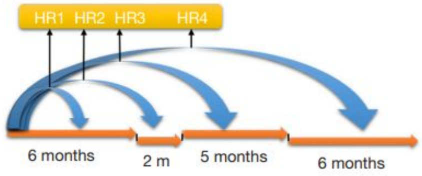
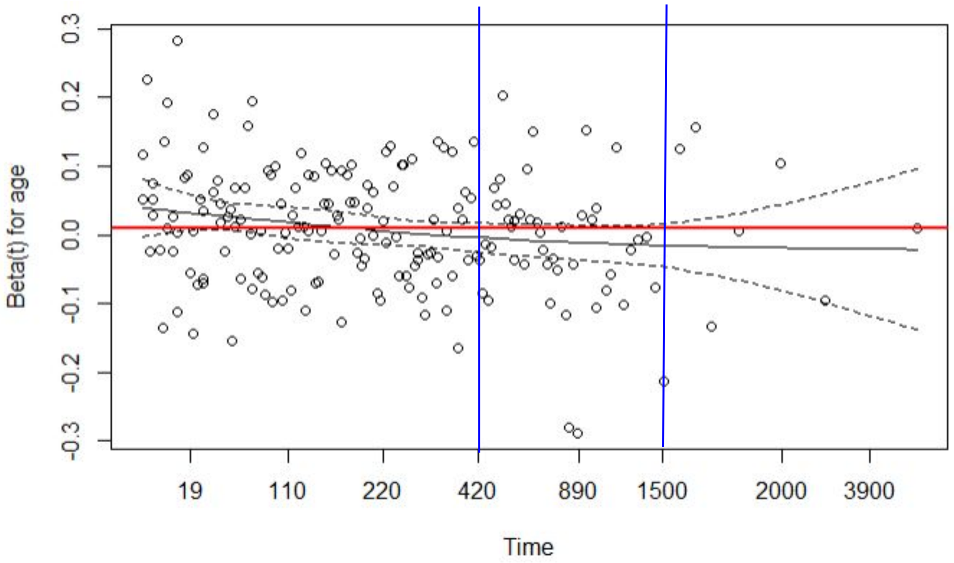


Figure 2 Time stratified effect of fixed baseline covariate on survival. Note that the effects of baseline covariate for different time windows are different, resulting in a series of hazard ratios.

Applying Stratification to Age



Plotting the schoenfeld residual of age against time, we can observe that from 0-420, 420-1500, the residuals are pretty even.
Therefore, dividing time into three intervals for age and run cox model

Stratified Data

```r
#Splitting the data into three time intervals
lung2 <- survSplit(Surv(time, status) ~ age + sex + race, data = LUNG,
                   cut = c(420,1500), episode = "time_group")
```

	time	status	age	sex	race
1	1110	0	53	Male	Black/African-American
2	18	1	60	Male	Black/African-American
3	44	1	69	Female	Black/African-American
10	1386	0	55	Male	Black/African-American
13	1213	0	73	Male	Black/African-American
16	4	1	71	Female	Black/African-American
23	65	1	60	Female	Black/African-American
26	610	0	61	Male	Black/African-American
27	278	1	80	Male	Black/African-American
28	801	0	64	Female	Black/African-American
30	689	0	51	Female	Black/African-American
32	1024	1	57	Female	Black/African-American
39	289	0	89	Male	White
40	1007	0	84	Female	Black/African-American
43	1095	0	64	Male	Black/African-American
49	718	0	56	Female	Black/African-American
50	853	0	70	Male	Black/African-American
57	615	0	74	Female	Black/African-American
66	707	0	81	Female	Black/African-American
67	529	0	68	Female	Black/African-American

	age	sex	race	tstart	time	status	time_group
1	53	Male	Black/African-American	0	420	0	1
2	53	Male	Black/African-American	420	1110	0	2
3	60	Male	Black/African-American	0	18	1	1
4	69	Female	Black/African-American	0	44	1	1
5	55	Male	Black/African-American	0	420	0	1
6	55	Male	Black/African-American	420	1386	0	2
7	73	Male	Black/African-American	0	420	0	1
8	73	Male	Black/African-American	420	1213	0	2
9	71	Female	Black/African-American	0	4	1	1
10	60	Female	Black/African-American	0	65	1	1
11	61	Male	Black/African-American	0	420	0	1
12	61	Male	Black/African-American	420	610	0	2
13	80	Male	Black/African-American	0	278	1	1
14	64	Female	Black/African-American	0	420	0	1
15	64	Female	Black/African-American	420	801	0	2
16	51	Female	Black/African-American	0	420	0	1
17	51	Female	Black/African-American	420	689	0	2
18	57	Female	Black/African-American	0	420	0	1
19	57	Female	Black/African-American	420	1024	1	2
20	89	Male	White	0	289	0	1
21	84	Female	Black/African-American	0	420	0	1
22	84	Female	Black/African-American	420	1007	0	2
23	64	Male	Black/African-American	0	420	0	1
24	64	Male	Black/African-American	420	1095	0	2

Stratified Model Results

```r
#fitting a new cox model using the time split
fit_tv_coef <- coxph(Surv(tstart, time, status) ~ age:strata(time_group) + sex + race,
                     data = lung2)
```

```
Call:
coxph(formula = Surv(tstart, time, status) ~ age:strata(time_group) + 
    sex + race, data = lung2)

  n= 2748, number of events= 183 

                                               coef  exp(coef)   se(coef)      z Pr(>|z|)    
sexMale                                   1.743e-01  1.190e+00  1.492e-01  1.168   0.2426    
raceBlack/African-American                7.804e-01  2.182e+00  1.522e-01  5.126 2.96e-07 ***
raceAsian/Mideast Indian                  2.914e-01  1.338e+00  4.598e-01  0.634   0.5262    
raceNative Hawaiian/Other Pacific Islander -1.384e+01  9.722e-07  1.399e+03 -0.010   0.9921    
age:strata(time_group)time_group=1        1.761e-02  1.018e+00  8.067e-03  2.183   0.0291 *  
age:strata(time_group)time_group=2       -1.145e-02  9.886e-01  1.252e-02 -0.915   0.3604    
age:strata(time_group)time_group=3       -8.405e-03  9.916e-01  3.185e-02 -0.264   0.7918    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                                          exp(coef) exp(-coef) lower .95 upper .95
sexMale                                   1.190e+00  8.401e-01    0.8886     1.595
raceBlack/African-American                2.182e+00  4.582e-01    1.6192     2.941
raceAsian/Mideast Indian                  1.338e+00  7.472e-01    0.5435     3.296
raceNative Hawaiian/Other Pacific Islander 9.722e-07  1.029e+06    0.0000       Inf
age:strata(time_group)time_group=1        1.018e+00  9.825e-01    1.0018     1.034
age:strata(time_group)time_group=2        9.886e-01  1.012e+00    0.9646     1.013
age:strata(time_group)time_group=3        9.916e-01  1.008e+00    0.9316     1.055

Concordance= 0.626  (se = 0.021 )
Likelihood ratio test= 34.43  on 7 df,   p=1e-05
Wald test            = 33.5  on 7 df,   p=2e-05
Score (logrank) test = 35.62  on 7 df,   p=9e-06
```

```r
test.ph <- cox.zph(fit_tv_coef)
test.ph
#Now the PH assumption holds.
```

```
                       chisq df    p
sex                     1.48  1 0.22
race                    5.00  3 0.17
age:strata(time_group)  3.97  3 0.26
GLOBAL                 10.67  7 0.15
```

Application 2

Laser Coagulation on Retinopathy

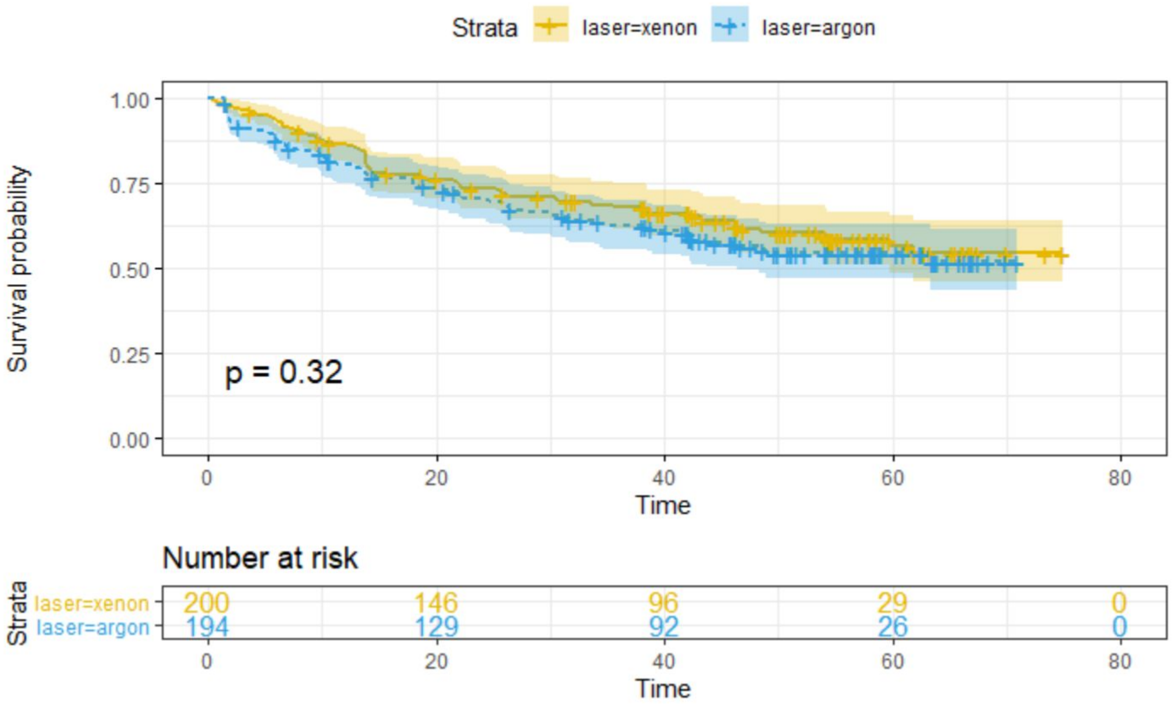
Data Overview

The “retinopathy” dataset in R contains 394 observations of 197 patients. These patients were selected in a random sample of patients with “high-risk” diabetic retinopathy. The first six rows of the dataset are shown below.

```{r}
head(retinopathy)
```

	id	laser	eye	age	type	trt	futime	status	risk
	<int>	<fctr>	<fctr>	<int>	<fctr>	<int>	<dbl>	<int>	<int>
1	5	argon	left	28	adult	1	46.23	0	9
2	5	argon	left	28	adult	0	46.23	0	9
3	14	argon	right	12	juvenile	1	42.50	0	8
4	14	argon	right	12	juvenile	0	31.30	1	6
5	16	xenon	right	9	juvenile	1	42.27	0	11
6	16	xenon	right	9	juvenile	0	42.27	0	11

Survival Probability Curves



Model Results

```
> #Fit cox model
> res.cox <- coxph(Surv(futime, status) ~ laser + eye + age + type
 + trt + risk, data =  retinopathy)
> summary(res.cox)
Call:
coxph(formula = Surv(futime, status) ~ laser + eye + age + type +
    trt + risk, data = retinopathy)

  n= 394, number of events= 155

                coef exp(coef)  se(coef)      z Pr(>|z|)
laserargon  0.165921  1.180480  0.162138  1.023   0.3062
eyeleft     0.217142  1.242521  0.167562  1.296   0.1950
age         0.010508  1.010563  0.009829  1.069   0.2851
typeadult  -0.157209  0.854526  0.293514 -0.536   0.5922
trt        -0.786751  0.455322  0.169254 -4.648 3.35e-06 ***
risk        0.141444  1.151936  0.055991  2.526   0.0115 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

           exp(coef) exp(-coef) lower .95 upper .95
laserargon    1.1805     0.8471    0.8591    1.6221
eyeleft       1.2425     0.8048    0.8947    1.7256
age           1.0106     0.9895    0.9913    1.0302
typeadult     0.8545     1.1702    0.4807    1.5190
trt           0.4553     2.1962    0.3268    0.6344
risk          1.1519     0.8681    1.0322    1.2855

Concordance= 0.629  (se = 0.023 )
Likelihood ratio test= 33.44  on 6 df,   p=9e-06
Wald test            = 31.93  on 6 df,   p=2e-05
Score (logrank) test = 33.09  on 6 df,   p=1e-05
```

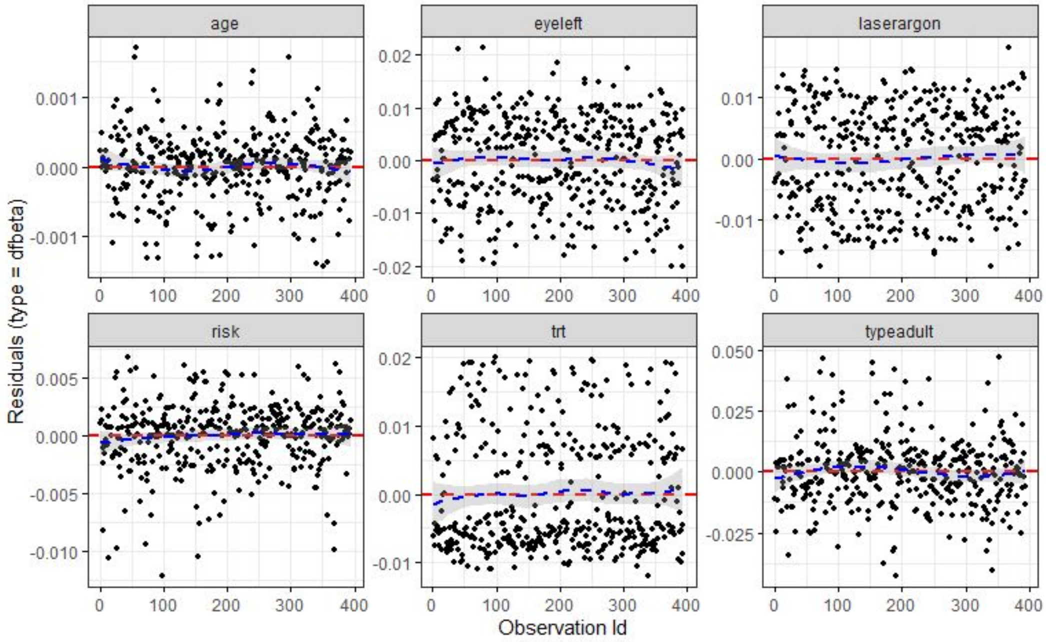
The results show that treatment and risk are statistically significant, with the p-values being less than alpha of .05.

To determine the effect on hazard of a coefficient, it will have to be exponentiated.

```
> exp(res.cox$coefficients["trt"])
      trt
0.4553217
> exp(res.cox$coefficients["risk"])
    risk
1.151936
```

Testing for Influential Outliers

```
> ggcoxdiagnostics(res.cox, type = "dfbeta", linear.predictions = FALSE,
 ggtheme = theme_bw())
```



We can test for influential outliers by visualizing the dfbeta values.

```
> res.cox$coefficients
 laserargon      eyeleft          age
  typeadult          trt         risk
 0.16592132  0.21714226  0.01050751
-0.15720861 -0.78675112  0.14144435
```

Testing PH Assumption

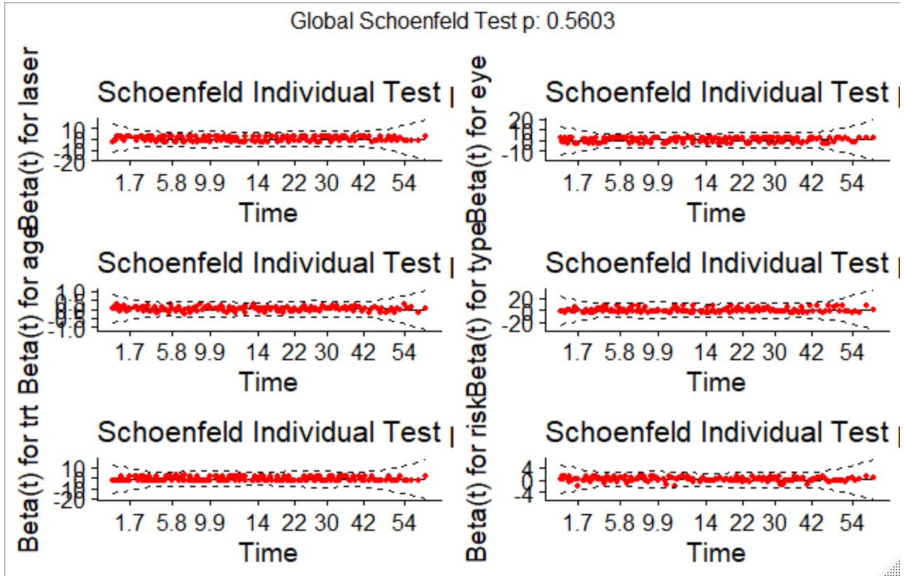
```
> #Test for PH assumption
> test.ph <- cox.zph(res.cox)
> test.ph
        chisq df    p
laser   0.672  1 0.41
eye     0.848  1 0.36
age     0.370  1 0.54
type    0.704  1 0.40
trt     0.604  1 0.44
risk    1.571  1 0.21
GLOBAL  4.872  6 0.56
```

The assumptions of the model need to be checked. First is the PH assumption.

P-values are all above alpha level of .05, indicating they are all not statistically significant, and so all variables pass the PH assumption tests. The global test is also not statistically significant.

Testing PH Assumption, cont'd

```
> #Visualizing the Schoenfeld residuals:
> ggcoxzph(test.ph)
```

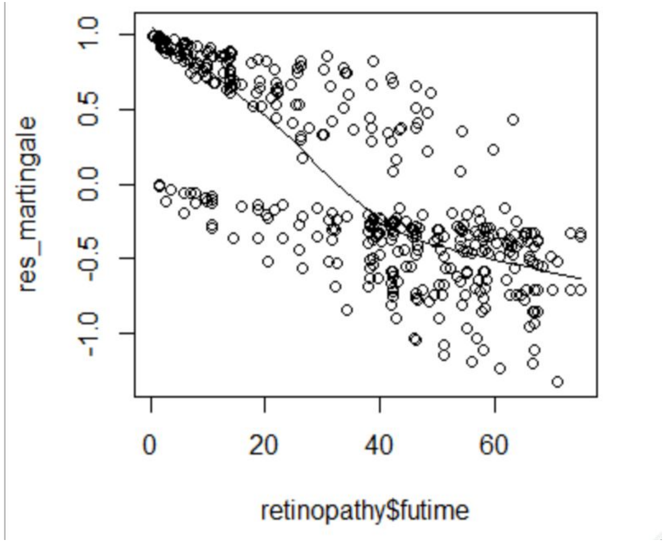


We can also test the PH assumption by examining the Schoenfeld residuals.

They appear to be distributed pretty evenly around zero, indicating no pattern with time. As such, this also appears to support the assumption of proportional hazards.

Testing Non-Linearity

```
> #Test for non-linearity
> res_martingale<-residuals(res.cox, type="martingale")
> scatter.smooth(retinopathy$futime,res_martingale)
```



The residuals appear to be fairly linear.

Application 3

Recidivism of Male Prisoners

Data Overview

The dataset contains data from an experimental study of recidivism of 432 male prisoners, who were observed for a year after being released from prison. There are 9 variables in the dataset.

```
##    week arrest fin age race wexp mar paro prio
## 1    20      1   0  27    1    0   0    1    3
## 2    17      1   0  18    1    0   0    1    8
## 3    25      1   0  19    0    1   0    1   13
## 4    52      0   1  23    1    1   1    1    1
## 5    52      0   0  19    0    1   0    1    3
## 6    52      0   0  24    1    1   0    0    2
```

Model I Summary

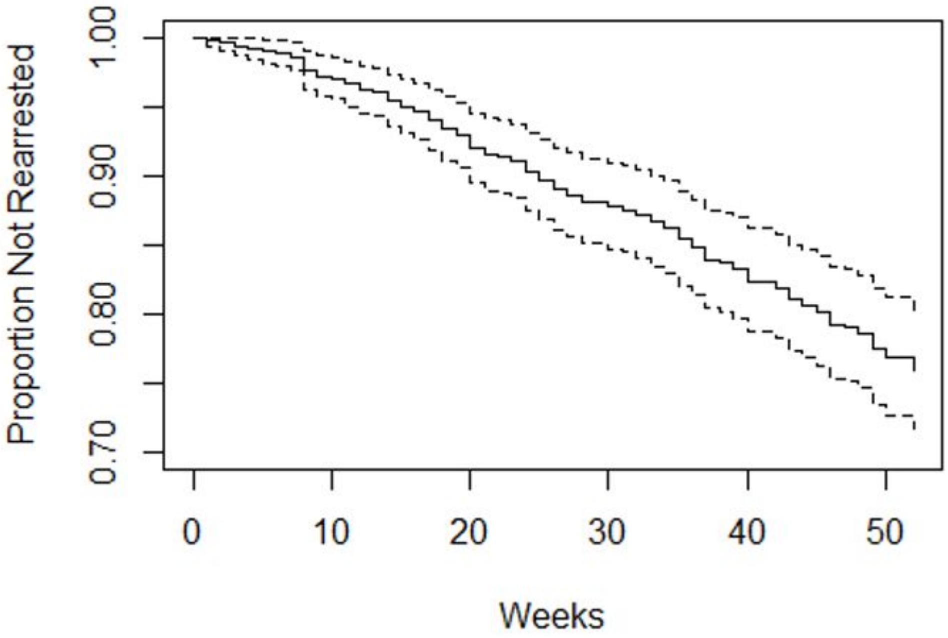
```
Call:
coxph(formula = Surv(week, arrest) ~ fin + age + race + wexp +
    mar + paro + prio, data = dat)

  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
```

```
Likelihood ratio test= 33.27  on 7 df,   p=2e-05
Wald test            = 32.11  on 7 df,   p=4e-05
Score (logrank) test = 33.53  on 7 df,   p=2e-05
```

The exponentiated coefficients in the second column of the first panel are interpretable as multiplicative effects on the hazard.

The likelihood-ratio, Wald statistic, and chi-square statistics are asymptotically equivalent tests of the omnibus null hypothesis that all of the β’s are zero.

Model I Survival Curves



Model II Summary

```
Call:
coxph(formula = Surv(week, arrest) ~ fin + prio + mar, data = dat)

  n= 432, number of events= 114 

         coef exp(coef) se(coef)      z Pr(>|z|)    
fin  -0.41386   0.66109  0.19011 -2.177 0.029485 *  
prio  0.10340   1.10893  0.02669  3.873 0.000107 ***
mar  -0.72885   0.48246  0.36688 -1.987 0.046966 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     exp(coef) exp(-coef) lower .95 upper .95
fin     0.6611     1.5126    0.4555    0.9596
prio    1.1089     0.9018    1.0524    1.1685
mar     0.4825     2.0727    0.2351    0.9903

Likelihood ratio test= 21.35  on 3 df,   p=9e-05
Wald test            = 22.78  on 3 df,   p=4e-05
Score (logrank) test = 23.25  on 3 df,   p=4e-05
```

By eliminating some covariates, we now have a new model.

The coefficients are all statistically significant.

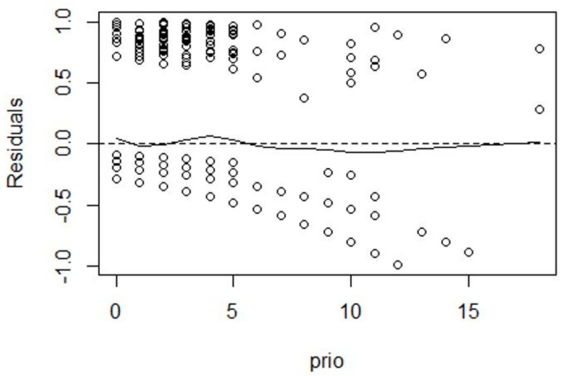
Model II Assumption Testing

Testing the PH assumption:

```
cox.zph(mod.3)

##           chisq df    p
## fin      0.0796  1 0.78
## prio     0.4558  1 0.50
## mar      0.9536  1 0.33
## GLOBAL   1.4357  3 0.70
```

There is strong evidence of proportional hazards for fin, prio, mar and global tests.

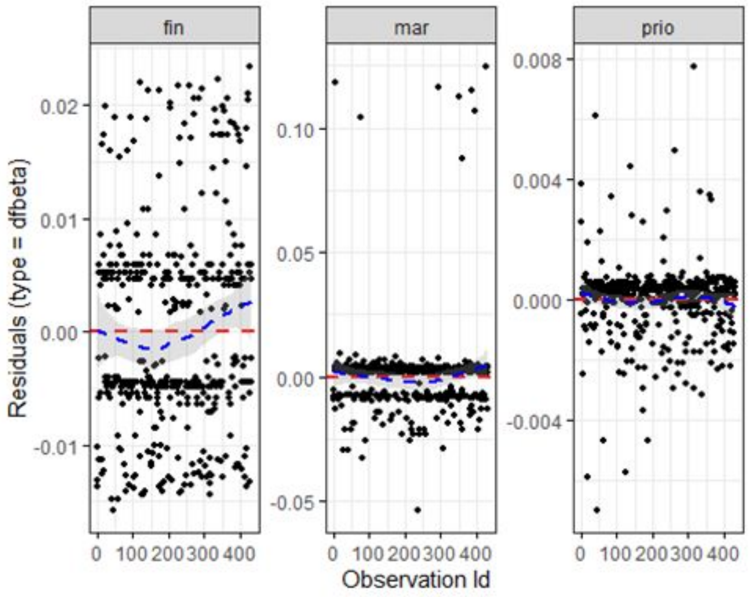
Checking for nonlinearity:



Nonlinearity is slight here.

Model II Assumption Testing, cont'd

Checking for influential outliers:



	coef	exp(coef)	se(coef)
fin	-0.41386	0.66109	0.19011
prio	0.10340	1.10893	0.02669
mar	-0.72885	0.48246	0.36688

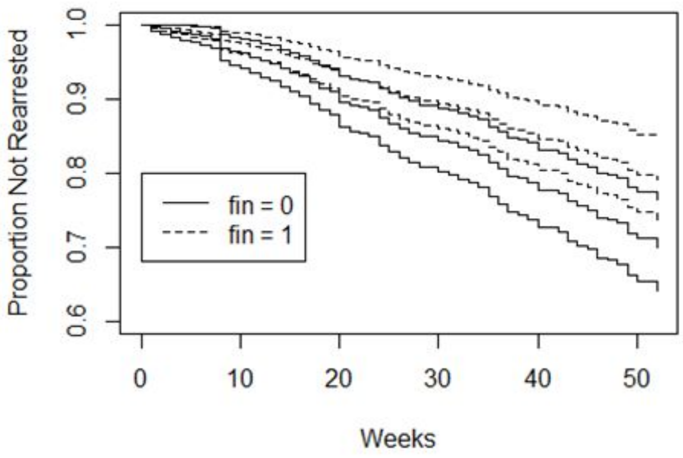
None of the observations is terribly influential individually

Prediction I: Est Survival Functions Based on Financial Aid Status

New Data 1:

```
dat.fin

##    fin        mar      prio
## 1    0 0.1226852 2.983796
## 2    1 0.1226852 2.983796
```

```
plot(survfit(mod.3, newdata=dat.fin), conf.int=T,lty=c(1,2), ylim=c(.6,
 1),xlab='Weeks', ylab='Proportion Not Rearrested')
```



Higher estimated "survival" for those receiving financial aid.

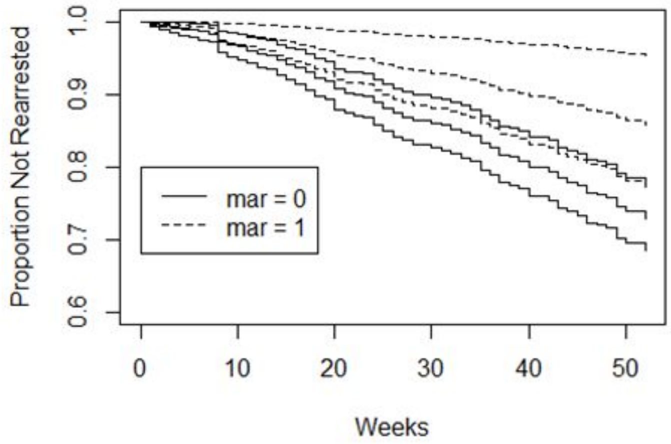
But the two confidence envelopes overlap substantially.

Prediction II: Est Survival Functions Based on Marital Status

New Data 2:

```
dat.fin2

##    mar fin      prio
## 1    0 0.5 2.983796
## 2    1 0.5 2.983796
```

```
plot(survfit(mod.3, newdata=dat.fin2), conf.int=T,lty=c(1,2), ylim=c(.6,
 1),xlab='Weeks', ylab='Proportion Not Rearrested')
```



Higher estimated "survival" for those who are married.

At later weeks, two confidence envelopes almost do not overlap any more.

References

- https://www.mrc-bsu.cam.ac.uk/wp-content/uploads/Survlecture2015.pdf
- http://www.sthda.com/english/wiki/cox-proportional-hazards-model
- http://www.sthda.com/english/wiki/cox-model-assumptions
- https://web.njit.edu/~wguo/Math%20659_2010/Chapters%205%20and%206%20%20from%20%5BSurvival%20Analysis_A%20Self%20Learning%20Text_Book%5D.pdf
- https://yosemite.epa.gov/sab/sabproduct.nsf/140FAAE2B008F03885257788005D1C86/$File/Gradient_+Comments_082010.pdf
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5107407/
- https://www.slideshare.net/ChandanReddy4/machine-learning-for-survival-analysis
- https://cran.r-project.org/web/packages/survival/survival.pdf
- https://cran.r-project.org/web/packages/survminer/survminer.pdf
- https://rpubs.com/kaz_yos/kleinbaum-ph-assumption
- https://myweb.uiowa.edu/pbreheny/7210/f15/notes/11-10.pdf