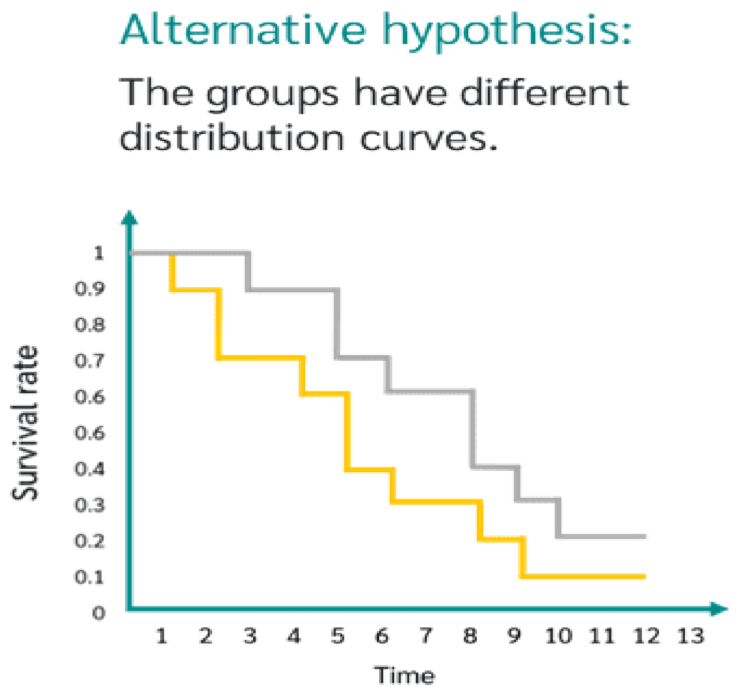
Cox Model

Rahul Telang

Kaplan Meier

- KM provides valuable information about survival
- It also allows us to compare survival wrt to a covariate



- But what if there are more than one covariate involved?

Parametric Models

- Now, we need a “regression” like model to account for covariates.
- Broadly regression works by fitting a linear line and estimates parameters by minimizing sum of square of error.
- When linear model is a not feasible, as in survival data, we need to make assumption on distribution of dependent variable (t - time for the event).
  - In linear regression, we implicitly assume that dependent variable follows normal distribution ($f(t) = \phi(.)$).
  - But normal distribution will not be correct choice for t.

Parametric model

- It is well established that, for timing data, the suitable distributions are Exponential, Weibull, Gamma etc.
- If it is Exponential distribution, then we know

Exponential pdf $f(t) = \lambda e^{-\lambda t}$

Exponential cdf $F(t) = 1 - e^{-\lambda t}$

Hazard $h(t) = \frac{f(t)}{1-F(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$

- Before we think about adding covariates, the first question if how does one estimate the parameter λ of the distribution?

Enter Log likelihood..

- In non-linear regression, the most common way to estimate λ is using likelihood principle.
- Take a detour of maximum likelihood......

Use of Cox Regression

- While KM and survival is widely used, due to its non-parametric nature, it has limitations.
- Timing data is observed in wide verity of situations. One needs a more general approach to handle those data.
- Cox regression is a workhorse for any timing data. For timing data, any data analysis is incomplete if it has not considered Cox regression.
- The following highlights where Cox model is used
  - Modeling Bankruptcy Prediction
  - Customer time to churn in the wireless telecommunications industry
  - Survival Analysis To Understand Customer Retention
  - Pricing life insurance premiums

Cox

- Cox model also relies on likelihood but avoids using density function for log likelihood...
- Cox's insight is that one can estimate parameters using partial likelihood . Thus, if the covariates X affect hazard like -

$$h(t, X) = h_0(t)e^{X\beta}$$

  Where $h_0(t)$ is the baseline hazard and (X) are linear predictors and (β) is the vector of parameters.

- Then in the linear form, one can write this as

$$\log\big(h(t, X)\big) = \log\big(h_0(t)\big) + X\beta$$

Cox Regression

- In Cox regression,

  $$\log(h(t/X)) = \log(h_0(t)) + X\beta.$$

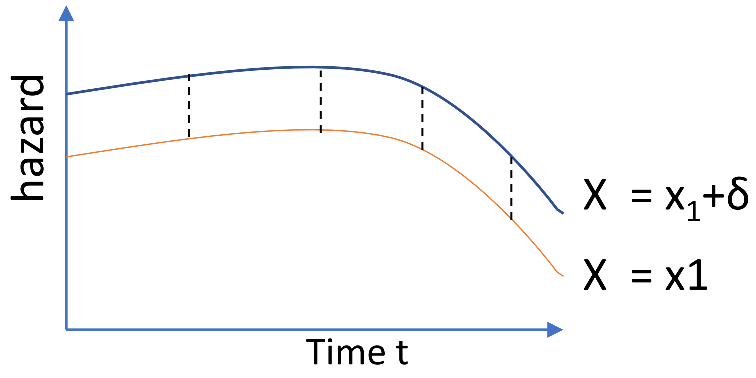
- Dependent variable is log of hazard. The dummy can be thought of as $\beta_0 = \log(h_0(t))$
  - But this is not a standard OLS model we are familiar with
  - Since we know that y in this context is not normally distributed. So, we can not estimate the model in a familiar method.
  - The most popular approach is to use log likelihood methods. Cox model uses partial log Likelihood method.

Cox regression

- Because Cox makes no assumption on distribution (hazard), $\beta_0 = \log(h_0(t))$ treated as a nuisance parameter.
  - This is the reason likelihood can be simplified
- The baseline hazard h_0 is not estimated. Instead, in Cox, one examines how covariates shift hazard functions relative to some baseline hazard.
- Cox requires some assumptions
  - Cox works on the order of duration (most time, second most time and so on..) than the absolute value of duration.
  - This means, there can not be any ties (one cannot have same duration in the data).
  - The data is used for estimation until it is dropped after censoring (like in Kaplan-Meier).

Cox Proportional hazard Model

- The key is that baseline hazard (h_o) can be separated from covariates X.
  - $h(t, X) = h_0(t)e^{X\beta}$
  - In this formulation, hazard h(.) is not a function of covariates X and covariates $e^{(X\beta)}$ does not involve *t*.
- Therefore, covariates shift the hazard $h_0(t)$ in a **proportionally.** The hazard proportionally goes up or down independent of time.
- Hence the assumption is that the individual predictors, ***X*, are time-invariant**.



Cox model

- In Cox, **the baseline hazard does not have to be specified**, and hence and can be separated from covariates X. In other parametric hazard functions, one must fully specify hazard function which makes estimation non-trivial.
  - Because Cox does not make any assumption about the shape of the underlying hazards, it gets us out from assuming a distribution for hazard.
- When we are more interested in studying how survival varies as a function of explanatory variables (the relative rates) rather than the shape of the underlying hazard function (the absolute rate), then Cox is widely used.

Interpreting the estimates

- The interpretation in Cox models follows regression framework. Recall we are modeling "time to event" for a subject.
- Hazard rate signals propensity of then even to occur (or fail).
- Recall, we have
  - $\log\big(h(t,X)\big) = \log\big(h_0(t)\big) + X\beta$
- Or,
  - $\log\big(h(t,X)\big) - \log\big(h_0(t)\big) = X\beta$
- Or $\log\left(\frac{h(t,X)}{h_0(t)}\right)$=Xβ
  - β indicates the impact of a unit increase in X on log of hazard rate.
- Notice that we do not assume a functional form for hazard. We cannot say how X affects y (time to event). We can only say how X affects hazard relative to the baseline hazard $h_0(t)$.
  - So, the impact of covariate is measured as hazard ratio.

Interpreting the estimates

- A unit change in X causes β unit change in the log of hazard ratio.
- We can exponentiate the coefficients, exp(β) to get the change in the hazard ratio.
- When
  - exp(β) > 1, hazard of time to event has gone up due a unit change in X. If exp(β) = 1.5, then hazard has increased by (1.5 - 1) = 50%.
  - exp(β) = 0, there is no change in hazard,
  - exp(β) < 1, hazard has gone down. If exp(β) = 0.8, then hazard has decreases by (1 -0.8) = 20%.

Example

- Take an example of customer churn - firm wants to understand how long users stay with them (tenure) and what factors affect churn. They can potentially design personalized strategies. Since this is a timing data (“when”), Cox is the natural choice
- We use the data on TV churn posted on canvas. The data is at customer level with attributes
  - demographics,
  - Tenure (in months)
  - phone/Internet connections,
  - streaming TV/movies,
  - type of contract they have (monthly, yearly, 2 year)
  - Bill (monthly, total)
  - Whether they churn.

Data

- Customer data

gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...
Female	0	Yes	No	1	No	No phone service	DSL	No	...
Male	0	No	No	34	Yes	No	DSL	Yes	...
Male	0	No	No	2	Yes	No	DSL	Yes	...
Male	0	No	No	45	No	No phone service	DSL	Yes	...
Female	0	No	No	2	Yes	No	Fiber optic	No	...

```
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   7043 non-null   object
 15  Contract          7043 non-null   object
 16  PaperlessBilling  7043 non-null   object
 17  PaymentMethod     7043 non-null   object
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   object
 20  Churn             7043 non-null   object
```

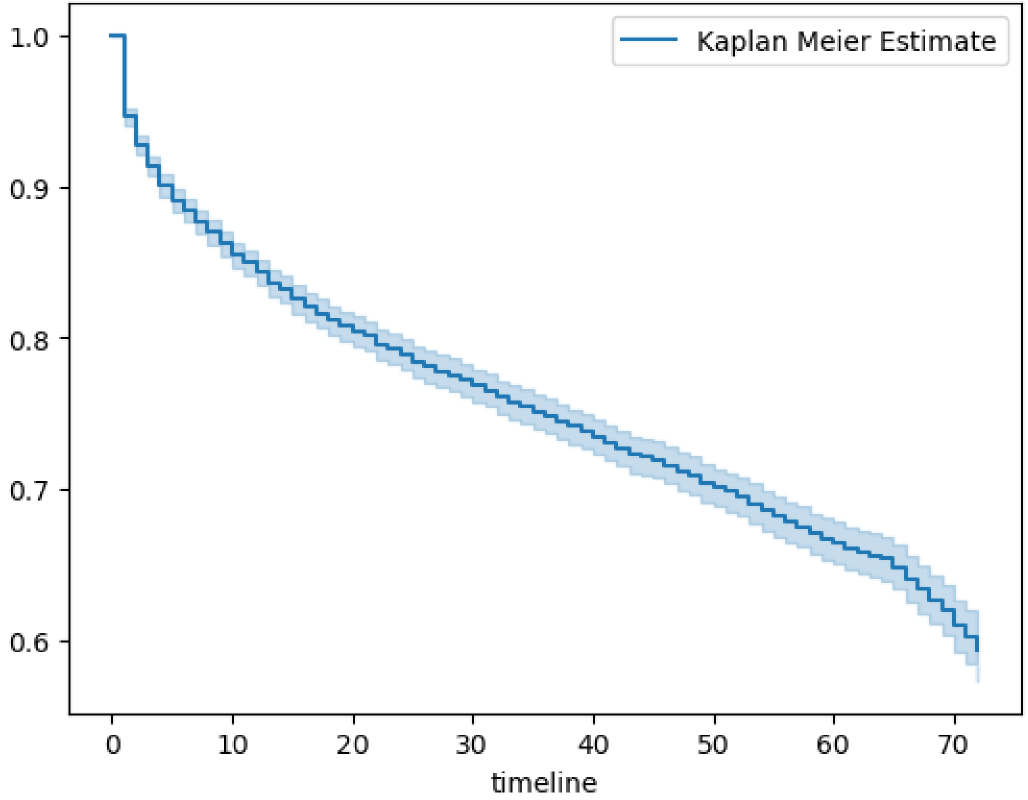
KM Curve

- Convert text data to numerical data.
- As before, we first plot KM curve for how long a user survives (tenure) by using kaplanmeier fitter

  - from lifelines import KaplanMeierFitter
  - durations = df['tenure’]
  - event_observed = df['Churn'] ## It has the churned (1) and censored is (0)
  - km = KaplanMeierFitter() ## instantiate the class to create an object

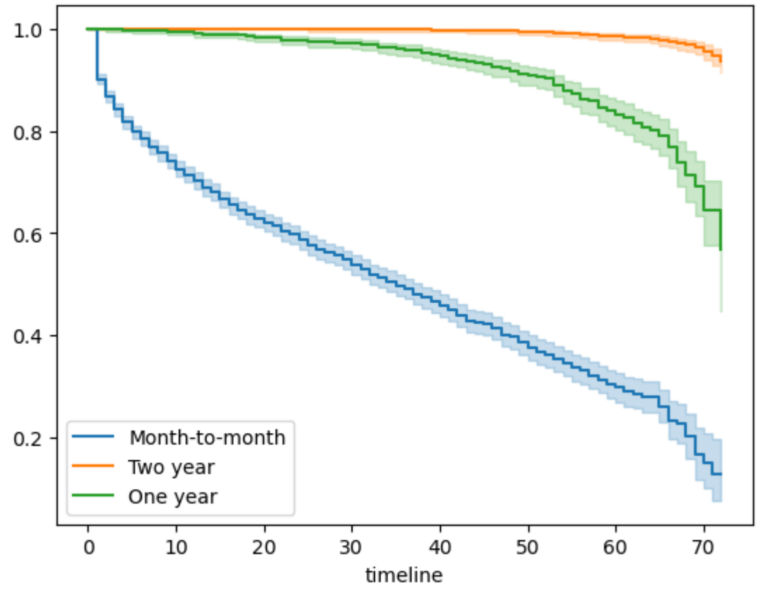
KM Curve

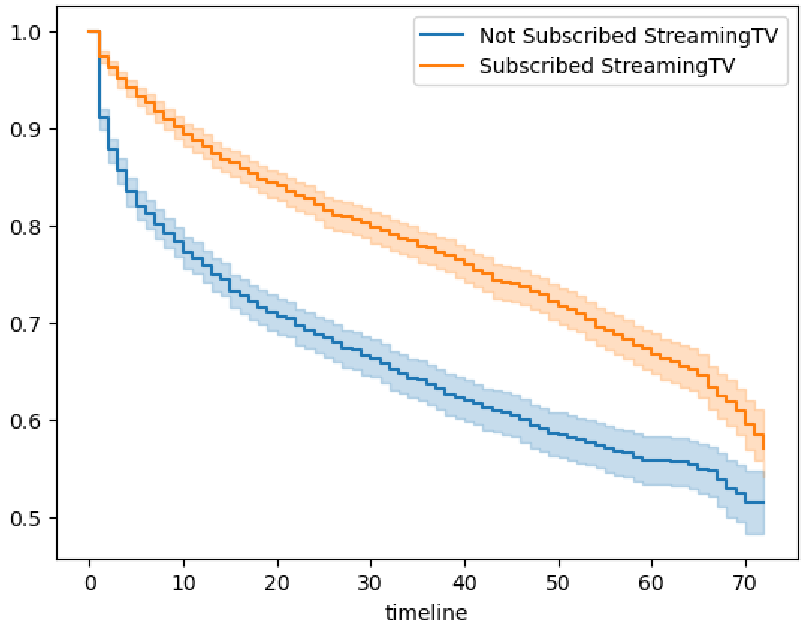
- Hazard is sharper in earlier timeline, but it becomes regular with time.



Compare different segments

- People sign up for different contracts (month to month, one year, two year). How does the survival curve look for them? Or people subscribe to streaming TV vs not.





Cox Model

- We estimate the impact of various covariates on hazard of churn and see how they impact the outcome.
- Given covariates X = [‘Mont_Charge’, ‘senior’, ‘gender’, ‘partner’, ‘dependent’, ‘phoneSerice’, ‘StreamingTV’],
- We want to estimate
  - $h(t, X) = h_0(t)e^{(X\beta)} = h_0(t)e^{(\beta_1 Month_{Charge} + \beta_2 senior + \ \dots)}$
- To do this in python, we import coxfitter library from lifelines and fit Cox model
  - from lifelines import CoxPHFitter
  - cph = CoxPHFitter()
  - cph.fit(data = “**tenure**", event_col=“**churn**")

Results

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	z	p	-log2(p)
MonthlyCharges	-0.01	0.99	0.00	-0.01	-0.01	-6.13	<0.005	30.06
SeniorCitizen	0.40	1.49	0.06	0.29	0.50	7.16	<0.005	40.15
gender_Male	-0.01	0.99	0.05	-0.10	0.08	-0.23	0.82	0.29
Partner_Yes	-0.81	0.45	0.05	-0.92	-0.70	-14.93	<0.005	164.98
Dependents_Yes	-0.36	0.70	0.07	-0.49	-0.22	-5.21	<0.005	22.37
PhoneService_Yes	0.69	2.00	0.10	0.49	0.90	6.65	<0.005	34.96
StreamingTV_No internet service	-2.10	0.12	0.13	-2.36	-1.84	-15.79	<0.005	184.09
StreamingTV_Yes	-0.19	0.83	0.06	-0.31	-0.07	-3.10	<0.005	9.03

model	lifelines.CoxPHFitter
duration col	'tenure'
event col	'Churn'
baseline estimation	breslow
number of observations	7043
number of events observed	1869
partial log-likelihood	-15182.39

Concordance	0.71
Partial AIC	30380.78
log-likelihood ratio test	941.30 on 8 df
-log2(p) of ll-ratio test	654.95

Python output

- Before interpreting the results, lets look at the numbers displayed in Cox output.
- Partial Log likelihood – Cox maximizes partial likelihood. A bigger value (less negative), suggests a better fit.
- Partial AIC - AIC stands for Akaike Information Criterion.
- AIC is used to compare between two model to test which one is a better model. Lower value of AIC is a better fit
- It is calculated as
    - AIC = 2k - 2 (partial Log Likelihood)
    - Where k is the number of parameters.

AIC

- AIC is trying to balance how many parameter to include. Lower value of AIC is a better fit. The following example illustrates this.

Log Likelihood - LL	Parameters (N)	AIC = -2LL+2K
-100.2	3	206.4
-98.6	5	207.2

- More parameters always leads to the lower value of likelihood (recall LL is negative). So, we penalize AIC by adding the cost of parameters (2k). In the example above, adding two more parameters lowers log likelihood but AIC goes up as we add extra parameters.
- In our example in earlier slide, AIC = (-2*(-15182.39)+2*8)= 30380.78

Likelihood ratio test

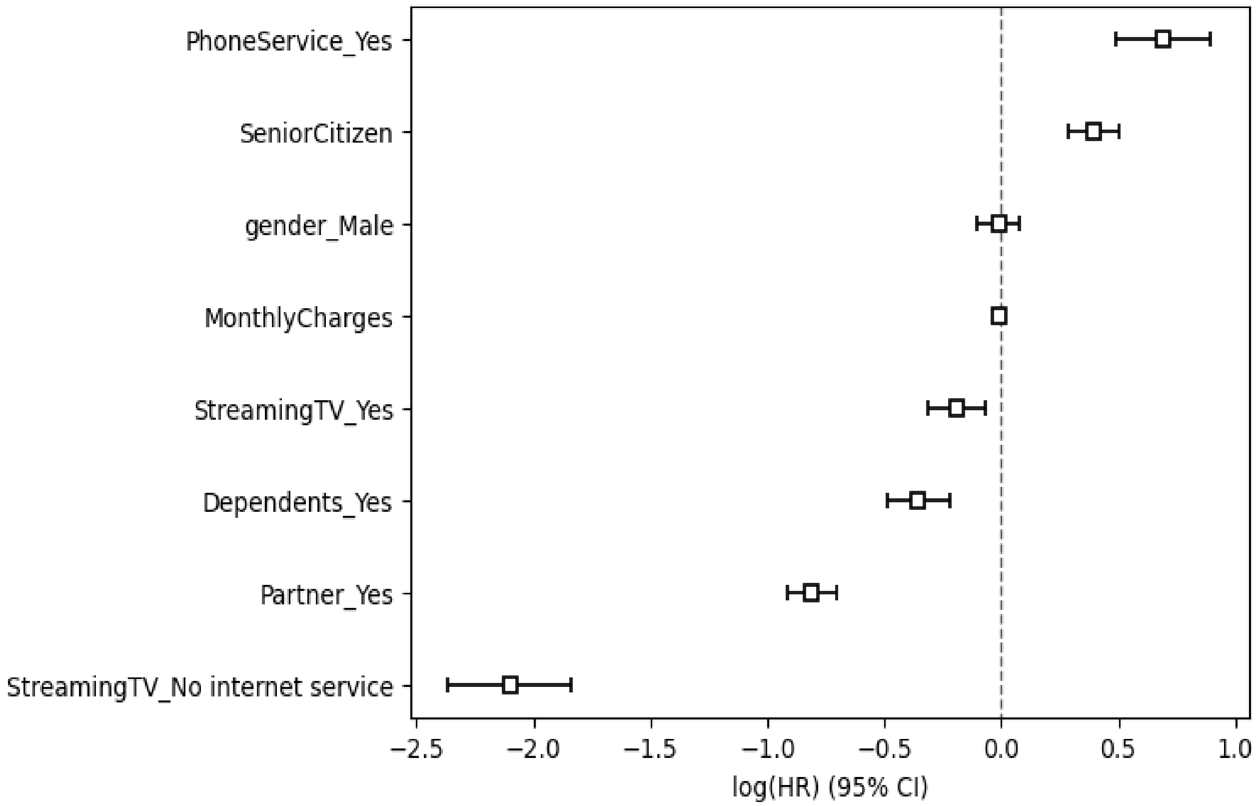
- The likelihood ratio (LR) test is a test of hypothesis in which two different maximum likelihood estimates of a parameter are compared in order to decide whether to reject a restriction on the parameter.
- In the previous example, we are testing whether 7 covariates jointly are significant.
  - log-likelihood ratio test 941.30 on 8 df
- It can be tested using Chi square statistics

Concordance

- Concordance – higher risk (hazard) should result in shorter survival time. Concordance ratio calculates that how many pairs match (high risk -> low survival time). It they do not then they are called discordant pairs. Higher value signals better fit.
- This statistics is a way to think about the “fit” of the model – how many times it correctly classifies the data - ROC curve for Cox models.
- It should be kept in mind that concordance only suggests how correctly the order of survival is predicted but not the absolute value of survival.
- A concordance value = 0.5 suggest that the model is no better than a random algorithm.
- Our model value of 0.71 is good.

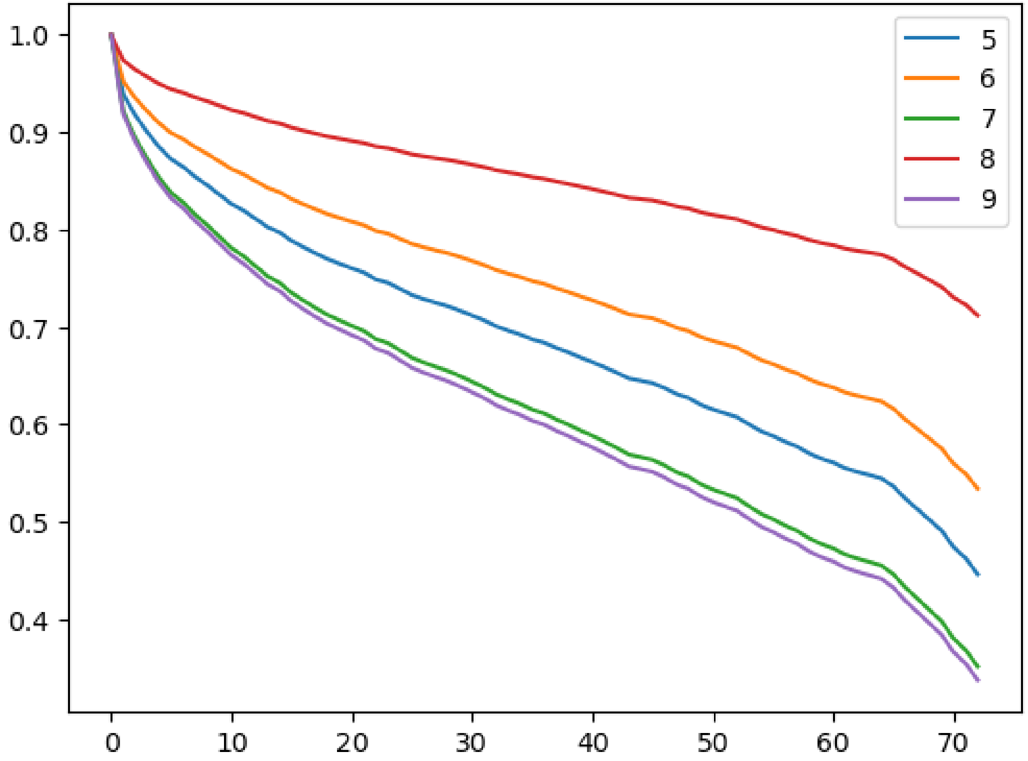
Interpreting covariates

- We can plot each covariate
  - cph.plot()
- Since Senior citizen is a discrete value, it suggests that when seniorcitizen = 1, it increases hazard by 49% (1.49-1)
- Monthly Charge is a continuous measure. The estimate suggests that a unit increase in Monthly Charge lowers the hazard by 1% (0.99-1)
- (StreamingTV_No internet service) is a discrete value and estimate suggest that it reduces hazard a lot (0.12-1) = 88%.



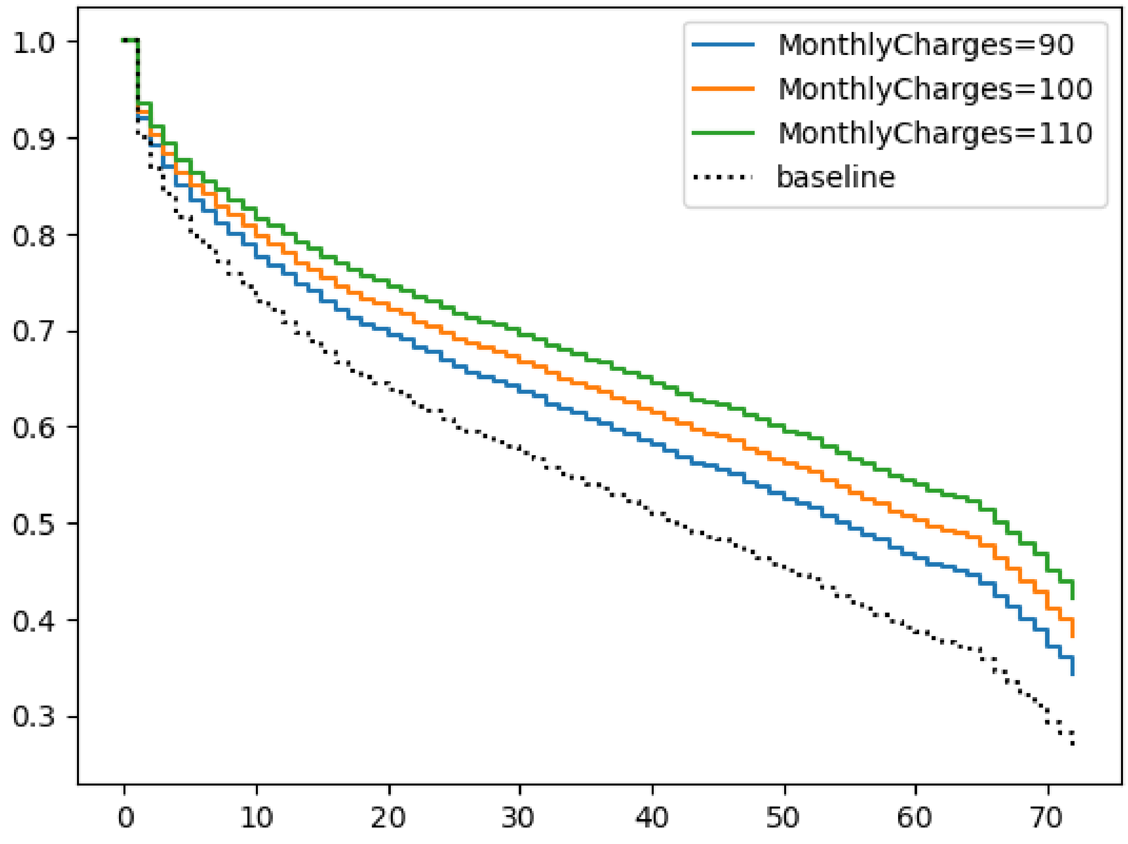
Survival function for individuals

- We can then plot the survival functions for individuals. User 8 has the highest survival rate and 9 the lowest.
- predict_survival_function(5:10).plot()



Marginal effect of change in one covariate

- Sometimes we are interested in understanding how change in one covariate affects survival while other covariates remain unchanged.
- If we change monthly charge, how does the survival change when other covariates remain constant?
- We can use command partial_effects_on_outcome
- cph.plot_partial_effects_on_outcome(covariates = 'MonthlyCharges', values = [90, 100, 110])

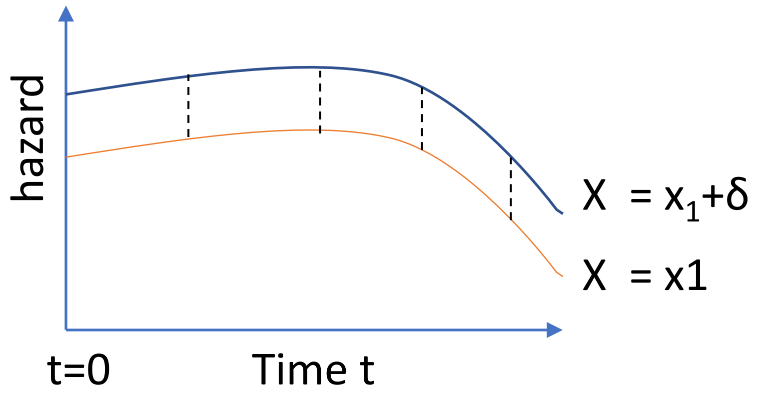


Assumptions

- A key assumption in Cox proportional hazard model is that effect of covariates is proportional.
- The proportional hazard assumption is that *all* observations, which have different survival time, have the same hazard function, separated by a unique scaling factor. So, the *shape* of the hazard function is the same for all observations, and only a scalar multiple changes per individual.
  - $h_i(t) = a_i h(t)$
    - $where\ a_i = e^{X\beta}$
- At the core of the assumption is that a_i is does not change with time. So, the hazard ratio is constant
  - $\frac{h_i(t)}{h_i(t)} = \frac{a_i}{a_j}$
- The hazard ratio remains same independent of time.

Proportional Hazard

- Proportionality means that the covariates shift the hazard ratio proportionally. If the two observations have different survival times the hazard ration will always remain same. They will differ only by a constant if covariates are different. Even if the covariate value changes, the ratio is assumed to remain same.



Data

- Python provides utility for testing proportionality assumption.
- We use data from time_varying.csv posted on canvas.

data

- We use a popular dataset which contains data from an experimental study of recidivism of 432 male prisoners, who were observed for a year after being released from prison. These are attributes
  - **week**: week of first arrest after release, or censoring time.
  - **arrest**: the event indicator, equal to 1 for those arrested during the period of the study and 0 for those who were not arrested.
  - **fin**: a factor, with levels “yes” if the individual received financial aid after release from prison, and “no” if he did not; financial aid was a randomly assigned factor manipulated by the researchers.
  - **age**: in years at the time of release.
  - **race**: a factor with levels “black” and “other”.
  - **wexp**: a factor with levels “yes” if the individual had full-time work experience prior to incarceration and “no” if he did not.
  - **mar**: a factor with levels “married” if the individual was married at the time of release and “not married” if he was not.
  - **paro**: a factor coded “yes” if the individual was released on parole and “no” if he was not.
  - **prio**: number of prior convictions.
  - **educ**: education, a categorical variable coded numerically, with codes 2 (grade 6 or less), 3 (grades 6 through 9), 4 (grades 10 and 11), 5 (grade 12), or 6 (some post-secondary).

Cox regression

- We estimate the following equation
  - $h\left(\frac{t}{X}\right) = h_0 \exp(\beta_{fin}\, fin + \beta_{age} age + \cdots \beta_{wexp}\, wexp + \cdots.)$

- h_0 is the baseline hazard.
- The impact of covariates is to shift the hazard up or down.

Results

- We first estimate Cox regression and use all covariates except education

model	lifelines.CoxPHFitter
duration col	'week'
event col	'arrest'
baseline estimation	breslow
number of observations	432
number of events observed	114
partial log-likelihood	-658.748
time fit was run	2023-01-31 23:13:53 UTC
model	untransformed variables

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
fin	-0.379	0.684	0.191	-0.755	-0.004	0.470	0.996	0.000	-1.983	0.047	4.398
age	-0.057	0.944	0.022	-0.101	-0.014	0.904	0.986	0.000	-2.611	0.009	6.791
race	0.314	1.369	0.308	-0.290	0.918	0.748	2.503	0.000	1.019	0.308	1.698
wexp	-0.150	0.861	0.212	-0.566	0.266	0.568	1.305	0.000	-0.706	0.480	1.058
mar	-0.434	0.648	0.382	-1.182	0.315	0.307	1.370	0.000	-1.136	0.256	1.965
paro	-0.085	0.919	0.196	-0.469	0.299	0.626	1.348	0.000	-0.434	0.665	0.589
prio	0.091	1.096	0.029	0.035	0.148	1.036	1.159	0.000	3.194	0.001	9.476

Concordance	0.640
Partial AIC	1331.495
log-likelihood ratio test	33.266 on 7 df
-log2(p) of ll-ratio test	15.370

Proportional model

- The Cox proportional model assumes that each covariate proportionally shifts the hazard function independent of time.
- Therefore, it is important to check is the proportionality assumption holds.

Proportionality Assumption

- We test the assumption of proportionality. LifeLines package provide functions to check this assumption. There are multiple ways this can be done. One of them is use “check_assumptions” function

cph.check_assumptions(dt_TV, p_value_threshold=0.05, show_plots = True)

- Km –Kaplan Meier
- The command also provides steps to perform to overcome violation of proportionality

null_distribution	chi squared
degrees_of_freedom	1
model	<lifelines.CoxPHFitter: fitted with 432 total ...
test_name	proportional_hazard_test

		test_statistic	p	-log2(p)
age	km	11.03	<0.005	10.12
	rank	11.45	<0.005	10.45
fin	km	0.02	0.89	0.17
	rank	0.02	0.90	0.15
mar	km	0.60	0.44	1.19
	rank	0.71	0.40	1.32
paro	km	0.12	0.73	0.45
	rank	0.13	0.71	0.49
prio	km	0.02	0.88	0.18
	rank	0.02	0.89	0.17
race	km	1.44	0.23	2.12
	rank	1.43	0.23	2.11
wexp	km	7.48	0.01	7.32
	rank	7.31	0.01	7.19

Proportionality test

- We can also use “proportional_hazard_test “

- from lifelines.statistics import proportional_hazard_test
- results = proportional_hazard_test(cph, dt_TV)

```
cph.check_assumptions(dt_TV, p_value_threshold=0.05,
show_plots =  True)
```

time_transform	rank
null_distribution	chi squared
degrees_of_freedom	1
model	<lifelines.CoxPHFitter: fitted with 432 total ...
test_name	proportional_hazard_test

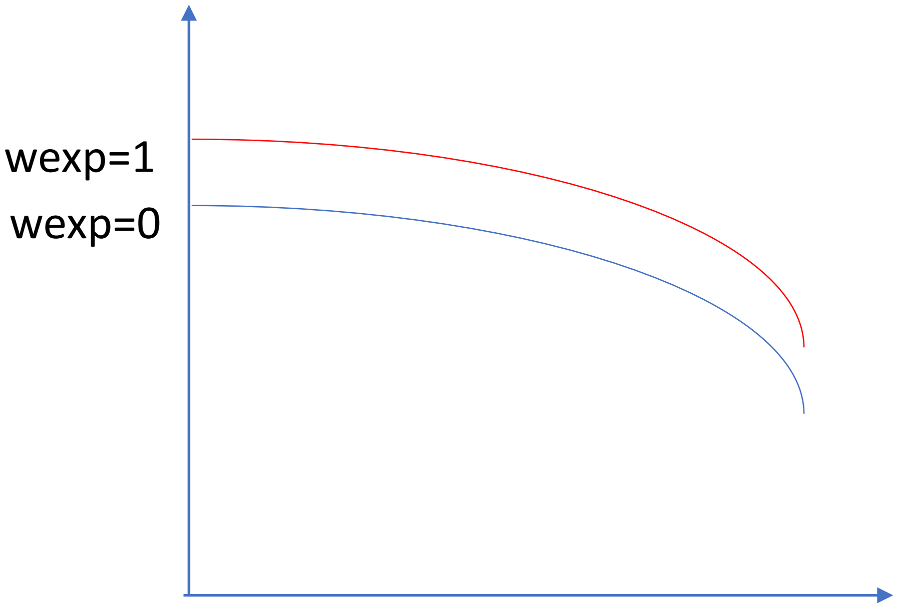
	test_statistic	p	-log2(p)
age	11.45	<0.005	10.45
fin	0.02	0.90	0.15
mar	0.71	0.40	1.32
paro	0.13	0.71	0.49
prio	0.02	0.89	0.17
race	1.43	0.23	2.11
wexp	7.31	0.01	7.19

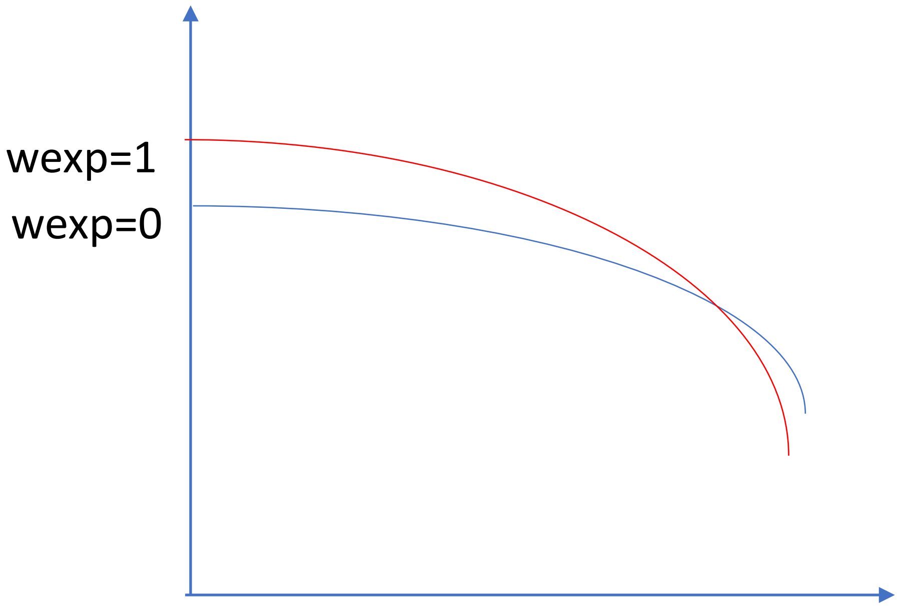
Results

- The results show that age and wexp do not fulfil the proportionality assumption.
- What is likely reason?
- Lets look at wexp first.
  - **wexp**: a dummy variable - “yes” if the individual had full-time work experience prior to incarceration and “no” if she did not.
- When using wexp as a covariate, we are assuming that that the shape of the hazard is same for all individuals but wexp shifts the hazard up or down.

Hazards may not be same

- The assumption is hazard shape is same for all individuals and wexp shifts the hazard.
- But what if the hazard is not the same? Suppose it is different for wexp=0 and wexp=1.





Strata

- Now using wexp as a covariate is violating the assumption.
- One possibility is to let the hazard be different between these subgroups (strata). One can then estimate different hazard for each strata.
  - Variable 'wexp' failed the non-proportional test: p-value is 0.0063. Advice: with so few unique values (only 2), you can include `strata=['wexp', ...]`
- Stratification means that we are allowing different hazards for that variable (akin to fixed effect model).
- When we stratify the model on wexp, we are estimating
  - $h\left(\frac{t}{X}\right) = h_{0\,wexp=0} \exp(\beta_{fin}\, fin + \beta_{age} age + \cdots)$
  - $h\left(\frac{t}{X}\right) = h_{0\,wexp=1} \exp\left(\beta_{fin}\, fin + \beta_{age} age + \cdots\right)$

Strata

- Note that now wexp is not a covariate but we are allowing two different baseline hazards for different wexp.
- We assume different hazard shape for strata but other covariates (fin, age....) remain the same.
- How does one measure the effect of covariates (β_{fin} , ...)?
- The covariate will affect each strata differently, so the model reports the effect of covariate averaged across strata. So β_{fin} is reporting the effect of fin on hazard after accounting for wexp=0 and wexp=1.

Strata

- cph.fit(dt_TV, 'week', 'arrest', strata=['wexp’])
- *Strata* is possible with dummies but complicated with continuous variables.

model	lifelines.CoxPHFitter
duration col	'week'
event col	'arrest'
strata	wexp
baseline estimation	breslow
number of observations	432
number of events observed	114
partial log-likelihood	-580.89
time fit was run	2023-02-01 01:22:36 UTC
model	wexp in strata

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%
fin	-0.38	0.68	0.19	-0.76	-0.01
age	**-0.06**	**0.94**	**0.02**	**-0.10**	**-0.01**
race	0.31	1.36	0.31	-0.30	0.91
mar	-0.45	0.64	0.38	-1.20	0.29
paro	-0.08	0.92	0.20	-0.47	0.30
prio	0.09	1.09	0.03	0.03	0.15

Concordance	0.61
Partial AIC	1173.77
log-likelihood ratio test	23.77 on 6 df
-log2(p) of ll-ratio test	10.77

Proportionality assumption

	test_statistic	p	-log2(p)
age	11.45	<0.005	10.45
fin	0.02	0.90	0.15
mar	0.71	0.40	1.32
paro	0.13	0.71	0.49
prio	0.02	0.89	0.17
race	1.43	0.23	2.11
wexp	7.31	0.01	7.19

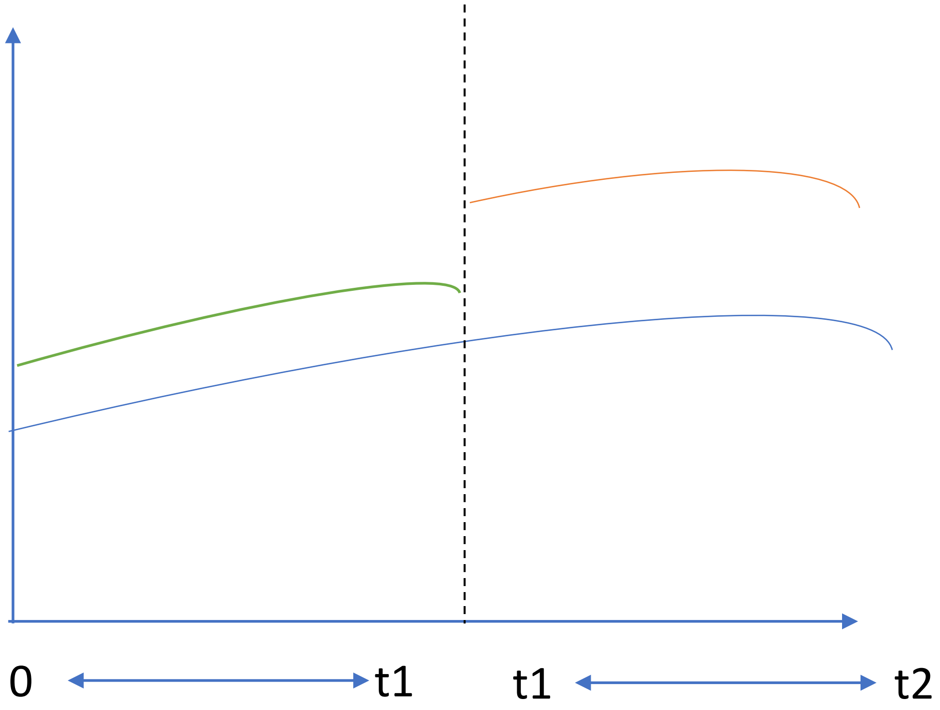
- Wexp is discrete and we could stratify. But age is continuous.
- Proportionality assumption is violated. See the recommendation when proportional assumption fails
  - *Advice 3: try adding an interaction term with your time variable*

Time varying covariates?

- Recall, in the cox model, covariates are not a function of time
  - $h(t, X) = h_0(t)e^{X\beta}$
- However, if covariate enter the hazard as
  - $h(t, X) = h_0(t)e^{X(t)\beta}$
- The covariate value changes with time. How does one accommodate it?
- The basic idea is to create a “panel”. If you are studying for period time [0,t] and covariate changes value at t1, and t2, then create a panel [0,t1], [t1,t2], [t2,t]. Within those segments, covariate is not changing.
- The proportionality assumption in that time-period can be satisfied.

Time varying covariate

- Covariate value changes at time t1. We allow covariate to remain same from 0 to t1 and t1 to t2 and so on.
- This satisfies the proportionality assumption and allows us to estimate the effect of time varying covariate.



Time varying covariates

To incorporate time varying covariates, we note the times when covariate value changes (lets say n times).
We split a subject from a single row into *n* new rows, and each new row lists the value of covariates during that time value of covariate remains the same.
As an example, we can write the data as

id	start	stop	var1	var2	event
1	0	4	0.1	1.4	False
1	4	8	0.1	1.2	False
1	8	10	0.1	1.5	True
2	0	12	0.5	1.6	False

Here id1 fails at time 10, but the covariate var2, changes value to 1.4, 1.2 and 1.5 at different times (0,4 and 8). So id 1's one row is split into 3 rows with event occurring (True or False) at time 10. But covariate value changes at time 4 and 8. In those periods, event will not occur.

We want to incorporate time varying age. We use python modules to do it more efficiently.

Time varying covariates

- First, we import to_episodic_format to split the data in the long form
  - from lifelines.utils import to_episodic_format
  - to_episodic_format(dt_TV, duration_col='week', event_col='arrest', time_gaps=1.)

	stop	start	arrest	age	fin	id	mar	paro	prio	race	wexp
0	1.0	0.0	0	27	0	0	0	1	3	1	0
1	2.0	1.0	0	27	0	0	0	1	3	1	0
2	3.0	2.0	0	27	0	0	0	1	3	1	0
3	4.0	3.0	0	27	0	0	0	1	3	1	0
4	5.0	4.0	0	27	0	0	0	1	3	1	0
5	6.0	5.0	0	27	0	0	0	1	3	1	0
6	7.0	6.0	0	27	0	0	0	1	3	1	0
7	8.0	7.0	0	27	0	0	0	1	3	1	0
8	9.0	8.0	0	27	0	0	0	1	3	1	0
9	10.0	9.0	0	27	0	0	0	1	3	1	0

Time varying covariate

- We interact age with time
- dt_TV_long['time*age'] = dt_TV_long['age'] * dt_TV_long['stop’]
- We then import CoxTimeVaryingFitter to accommodate the tine varying covariate we created
- ctv = CoxTimeVaryingFitter()

- ctv.fit (dt_TV_long,
  id_col='id',
  event_col='arrest',
  start_col='start',
  stop_col='stop',
  strata=['wexp']
)

This will estimate the model after accounting for time and age interaction.

Results

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
age	0.073	1.075	0.040	-0.005	0.151	0.995	1.163	0.000	1.830	0.067	3.893
fin	-0.386	0.680	0.191	-0.760	-0.011	0.468	0.989	0.000	-2.018	0.044	4.520
Mar	-0.397	0.672	0.382	-1.147	0.352	0.318	1.422	0.000	-1.039	0.299	1.743
paro	-0.098	0.907	0.196	-0.481	0.285	0.618	1.330	0.000	-0.501	0.616	0.698
prio	0.090	1.094	0.029	0.034	0.146	1.035	1.158	0.000	3.152	0.002	9.267
race	0.295	1.343	0.308	-0.310	0.899	0.733	2.458	0.000	0.955	0.340	1.558
time*age	-0.005	0.995	0.002	-0.008	-0.002	0.992	0.998	0.000	-3.337	0.001	10.203

- The interaction lowers the hazard for recidivism (rearrest) slightly (1-0.995)%.

Revisiting Cox model

- Advantage of Cox model is that we make no assumption on hazard function and only focus on the impact of covariates. Many times, we do not have clear idea of the shape of hazard. If we assume a wrong hazard, the impact of covariates is questionable.
- By avoiding assuming hazard, we do not have to estimate full likelihood model and can simplify estimation by partial likelihood.
- However, hazard tends to be a function to time which other parametric model allows. It has been shown that when hazard are known, a parametric model (like Weibull) outperforms Cox model.
  - One can use WeibullFitter(), ExponentialFitter(),… etc. if the distribution is known
- Parametric form allows for extrapolation (predict outside the range in the data)

Summary

- Cox is a popular model to estimate the effect of covariates on survival.
- It does not impose any restrictions on hazard but allows estimation in a robust and regression like format.
- It's a proportional hazard model; the impact of covariates is proportional
- Many assumptions like proportionality and time varying covariates need to be checked and accounted for in the model.