# Cox Regression Models

**Matthew Austin**

Institute for Advanced Analytics
MSA Class of 2019

# Proportional hazards:
## Overview

# The proportional hazards model

- An alternative to modeling failure time is to model the hazard
- In a **proportional hazards** (PH) model, we model the log (hazard) directly:

$$\log \lambda_i(t) = \log \lambda_0(t) + \mathbf{x}_i^\top \boldsymbol{\beta}$$

- And the hazard is

$$\lambda_i(t) = \lambda_0(t) e^{\eta_i}$$

- So here, the predictors shift the hazard rather than directly rescaling time
- The idea here is that there's some baseline hazard $\lambda_0(t)$ (which follows some distribution) that is doubled/tripled/halved/etc. based on an individual's factors $\mathbf{x}_i$

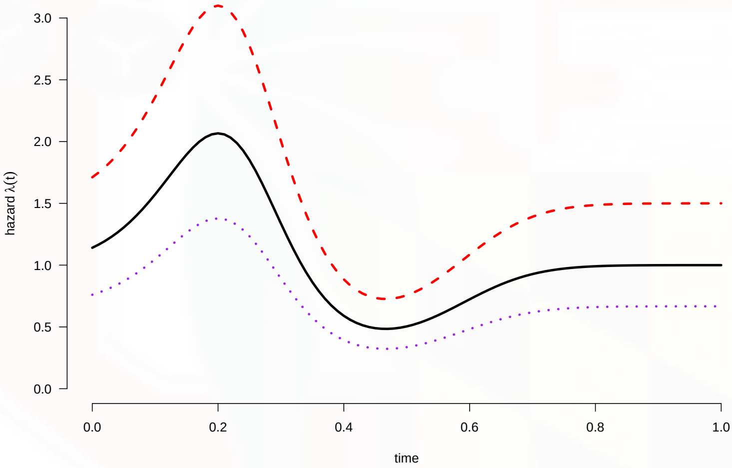
# Proportional hazards?

- If we look at two different individuals $\{\mathbf{x}_1, \mathbf{x}_2\}$, then
  - $\lambda_1(t) = \lambda_0(t)e^{\mathbf{x}_1^\top \boldsymbol{\beta}}$
  - $\lambda_2(t) = \lambda_0(t)e^{\mathbf{x}_2^\top \boldsymbol{\beta}}$
- The **hazard ratio** of these two people is:

$$\frac{\lambda_1(t)}{\lambda_2(t)} = e^{(\mathbf{x}_1 - \mathbf{x}_2)^\top \boldsymbol{\beta}}$$

- Baseline hazard cancels out!
- Since the hazard ratio doesn't depend on $t$, it is constant at all times—**proportional hazards**
- Basically, we're looking at the effect of some initial conditions/factors/predictors on the hazard

# Ohhh, proportional hazards.



# Interpreting estimates

- If we change $x_j$ by $c_j$:

$$\frac{\lambda_1(t)}{\lambda_2(t)} = \frac{e^{(x_j+c_j)\beta_j}}{e^{x_j\beta_j}} = e^{c_j\beta_j}$$

- $e^{c_j\beta_j}$ is the predicted hazard ratio for $x_j + c_j$ compared to $x_j$
    - $\beta_j = 0 \iff e^{\beta_j} = 1$ means $x_j$ has no effect on the hazard
    - $\beta_j > 0 \iff e^{\beta_j} > 1$ means $x_j$ is associated with an *increased* risk of failure (higher hazard)
    - $\beta_j < 0 \iff e^{\beta_j} < 1$ means $x_j$ is associated with a *decreased* risk of failure (lower hazard)

# Cox PH models: R syntax

```
fit <- coxph(Surv(week, arrest == 1) ~ fin + age + race +
             wexp + mar + paro + prio, data = recid)
```

| | coef | exp(coef) | se(coef) | z | Pr(>\|z\|) |
|---|---|---|---|---|---|
| fin | -0.38 | 0.68 | 0.19 | -1.98 | 0.05 |
| age | -0.06 | 0.94 | 0.02 | -2.61 | 0.01 |
| race | 0.31 | 1.37 | 0.31 | 1.02 | 0.31 |
| wexp | -0.15 | 0.86 | 0.21 | -0.71 | 0.48 |
| mar | -0.43 | 0.65 | 0.38 | -1.14 | 0.26 |
| paro | -0.08 | 0.92 | 0.20 | -0.43 | 0.66 |
| prio | 0.09 | 1.10 | 0.03 | 3.19 | 0.00 |

- For men receiving financial aid compared to those who didn't, the hazard ratio is $\beta_{\texttt{fin}} = e^{-0.38} = 0.68$
- The hazard is $\beta_{\texttt{prio}} = e^{0.09} = 1.1$ times higher for every additional prior conviction

# Cox PH models: SAS syntax

```
proc phreg data=survival.recid;
  model week*arrest(0) = fin age race wexp mar prio paro
                         / risklimits=pl ties=efron;
run;
```

# Proportional hazards:
## PH vs. AFT

# Differences between model types

- PH model: predictors have a multiplicative effect on the **hazard**

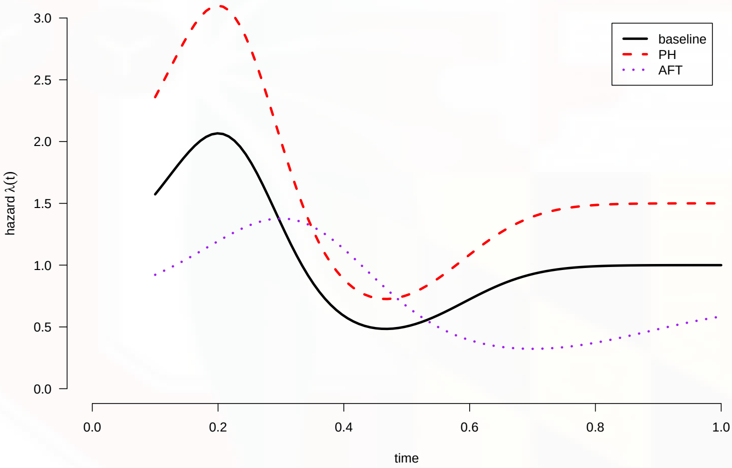
$$\lambda_i(t) = \lambda_0(t)e^{\eta_i}$$
$$\Longrightarrow S_i(t) = S_0(t)^{\exp\{\eta_i\}}$$

- AFT model: predictors have a multiplicative effect on **failure time**

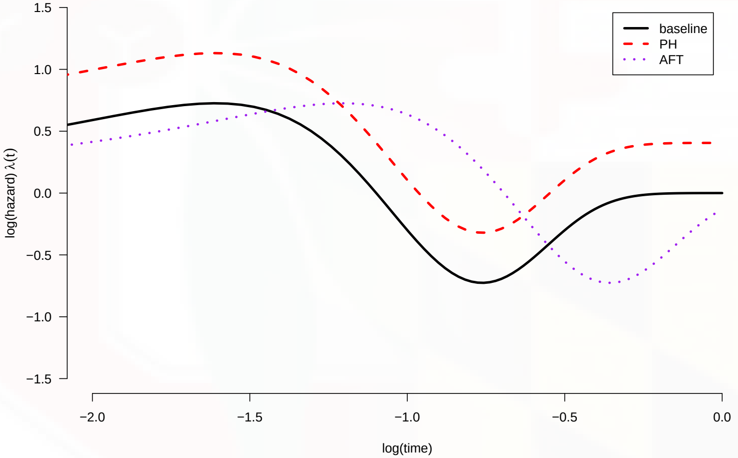
$$T_i = T_0 e^{\eta_i}$$
$$\Longrightarrow \lambda_i(t) = \lambda_0(te^{-\eta_i})e^{-\eta_i}$$
$$\Longrightarrow S_i(t) = S_0(te^{-\eta_i})$$

- (In both models, a good rule of thumb for degrees of freedom is $df < \frac{\#\text{events}}{\{10,15,20\}}$)

# PH vs. AFT: hazard vs. time



# PH vs. AFT as linear models: log (hazard) vs. log (time)



# The Weibull distribution

- A model is either a PH or AFT model; there is no way both assumptions can be satisfied simultaneously...
- Except... (*drumroll*)... for a Weibull distribution
- Fitting a Weibull AFT model automatically gives you everything you need to compute estimates for the corresponding Weibull PH model:

$$\boldsymbol{\beta}_{PH} = -\gamma \boldsymbol{\beta}_{AFT}$$

- (Explanation on course page if you care about why)
- Note: Although this relationship is true, your answers using this formula may not exactly match the `R/SAS` output from fitting both models—we will now discuss why

# The Cox regression model:
## Estimation

# Why use a different model?

- In building a model, it's often the case that we're interested in the effect of various predictors on the response rather than the actual response itself
- In survival data, since we don't always observe $T$, we might prefer to look at the relative event times instead of the actual differences in event times

# Semiparametric models

- In PH and AFT models, estimation depends on some distributional assumption about the either the baseline time or baseline hazard
- But in PH models, the likelihood can be split into a piece that has the baseline hazard and a piece that only depends on the predictors
- For this reason, it is possible to treat a PH model as a **semiparametric** model, where:
  - The piece depending on the predictors is parametric (e.g., $\mathbf{x}_i^\top \boldsymbol{\beta}$)
  - The piece depending on $\lambda_0(t)$ is nonparametric (we make no assumptions about its form/distribution)

# The Cox regression model

$$\log \lambda_i(t) = \log \lambda_0(t) + \mathbf{x}_i^\top \boldsymbol{\beta}$$
$$\Longrightarrow \lambda_i(t) = \lambda_0(t) e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$$

- Using the semiparametric model we've just introduced, we can basically ignore ever estimating anything about $\lambda_0(t)$ and just focus on the second part—this is the motivation for the **Cox regression model**
- In the Cox model, the baseline hazard $\lambda_0(t)$ is thus left completely unspecified
- This is still a PH model; the only difference now is that for estimation purposes, we don't need to care about the baseline hazard at all

# Partial likelihood

- Estimates are obtained by maximizing the "partial" likelihood—"partial" in the sense that we're only maximizing the part that depends on the predictors
- The partial likelihood is based on the **ranks** of failure times, which don't depend on the baseline hazard
- The idea is that since the PH model gives us hazard *ratios*, the actual distance between failure times isn't all that important, so ignoring it doesn't really matter when estimating the HR

# Comparative risk

- What the Cox model is estimating is essentially a subject's **relative** likelihood of failure at `time = t` compared to everyone else in the risk set at `time = t`
  - In other words: conditional on a failure happening at `time = t`, how likely was it to happen to subject $i$ out of everyone remaining at that time?
- By maximizing the partial likelihood, we're trying to get the $\hat{\boldsymbol{\beta}}$ that predicts the hazard was high for observations at their actual event time
- Any estimation/inference (coefficients, hazard ratios, etc.) is still valid, but contrary to the AFT, Cox models don't make any absolute predictions of time or risk

# Tied event times

- Since Cox regression is based on ranks, ties can be problematic
- There are a few common methods used to break ties and construct an appropriate partial likelihood for tied data: Efron, Breslow, exact
- If there are few/no ties, any of these methods will work just as well as another, but it's always safe to use Efron at the very least (default in `R`, but needs to be specified in `SAS`)
- For the case where time is discrete and events can actually occur at the exact same time, another option is the *discrete Cox model*
  - This is just a conditional logistic regression model
  - `ties = exact` in `coxph()`; `ties = discrete` (I think) in `proc phreg`

# Cox model assumptions

- Even though the Cox model gets us out of having to worry about any distributional assumptions, there are still model assumptions that we *do* need to check:
  - Linear relationship between $\mathbf{x}$ and $\log \lambda(t)$
  - **Proportional hazards**: no interactions involving time
- We'll talk about how to check these assumptions later

# The Cox regression model:
## Prediction

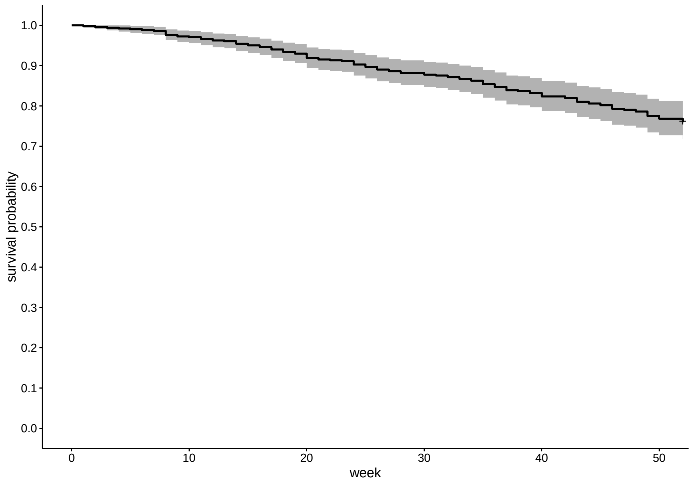
# Estimating survival curves

- In Cox regression models, each predictor is centered because #mathreasons, but the actual predictions are still relative to some unspecified baseline hazard rather than an absolute risk score
- However, once we've obtained $\hat{\boldsymbol{\beta}}$ from the partial likelihood, we can plug it into the full likelihood and nonparametrically estimate the remaining piece for $\lambda_0(t)$
- With $\hat{\boldsymbol{\beta}}$ and $\hat{\lambda}_0(t)$, we can estimate survival curves using the previously stated relationship:

$$\hat{S}_i(t) = \hat{S}_0(t)^{\exp\{\hat{\eta}_i\}}$$

# Survival curves: R syntax

```
ggsurvplot(survfit(fit), data = recid)
```



# But who is that curve for?

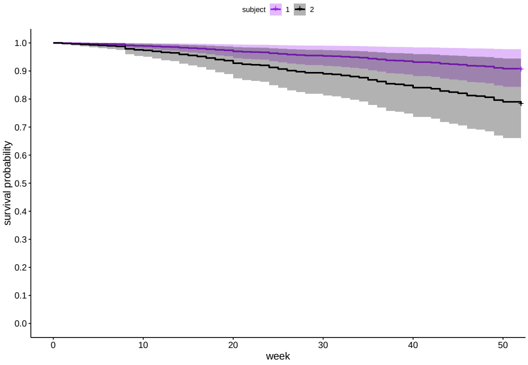
- The default survival curve represents an observation with all $x_j = 0$, but since the predictors are centered in Cox regression, $x_j = 0$ represents $\bar{x}_j$
- ...So who is this hypothetical person?

```
##   fin   age  race  wexp   mar  paro  prio
##  0.50 24.60  0.88  0.57  0.12  0.62  2.98
```

- From `?survfit.coxph`: "Serious thought has been given to removing the default value for `newdata`, which is to use a single 'pseudo' subject with covariate values equal to the means of the data set, since the resulting curve(s) almost never make sense."
  - (con't:) "It remains due to an unwarranted attachment to the option shown by some users *and by other packages*." $\longrightarrow$ shots fired at SAS

# Survival curves: R syntax

```
newdata <- data.frame(fin = c(1, 0), age = 30, race = 0,
                      wexp = c(1, 0), mar = 0, paro = 0,
                      prio = c(0, 4))
ggsurvplot(survfit(fit, newdata), data = newdata)
```



# Survival curves: SAS syntax

```
proc phreg data=survival.recid plots=survival;
  ...
run;

data ref;
  input fin age race wexp mar paro prio;
  1 30 0 1 0 0 0
  0 30 0 0 0 0 4
  ;
run;

proc phreg data=survival.recid plots(overlay)=survival;
  ...
  baseline covariates=ref out=refs;
run;
```