# COX REGRESSION MODEL

Dr. Aric LaBarr

Institute for Advanced Analytics

# MORE ON PMLE

# Semiparametric Models

- In AFT and PH models, estimation depends on some distributional assumption around either the failure time or the baseline hazard.
- However, in PH models, Cox noticed that the likelihood can be split into two pieces:
  - 1st piece: depends on $h_0(t)$ and the parameters
    - Treat as non-parametric (no assumptions about form or distribution)
  - 2nd piece: **only** depends on the parameters
    - Treat as parametric (know the form)
- This is why it is called a **semiparametric** model.

# Cox Regression Model

- Using the semiparametric model approach, we can basically ignore ever estimating anything about the baseline hazard $h_0(t)$ – the **Cox regression model**.

- Basically, Cox disregarded the first piece of the likelihood and maximized the second piece – still a PH model.

# Partial Likelihood Estimation

- This is the more important piece of the work done by Sir David Cox in his original article.
- Estimates are obtained by maximizing the **partial likelihood** – only one piece that depends on the predictors, not the entire thing.
  - Done based on ranks of failure times – don’t depend on baseline hazard.
  - All we care about is ratios between hazards.

# Too Much Info on PMLE

- Since estimation for Cox regression uses ranks, ties can be problematic.
- Common methods to construct an appropriate partial likelihood for breaking ties: Efron (R default), Breslow (SAS default), exact
- If there are a new/no ties any of these would work just fine.
- Safe to go with Efron because it does better for higher numbers of ties.