COMPETING RISKS

Dr. Aric LaBarr
Institute for Advanced Analytics
MSA Class of 2014

INTRODUCTION THROUGH EXAMPLES

Medical Example

- Cancer researcher finds a medicine that cures cancer.
- Run a medical study where you follow 100 patients for 5 years after giving them cancer cure to see how many die.
- In year 4, 7 of these patients travel together to Iceland and die in a volcano accident.
- The other 93 patients made it to the end of five years without passing away.

Medical Example

- Cancer researcher finds a medicine that cures cancer.
- Run a medical study where you follow 100 patients for 5 years after giving them cancer cure to see how many die.
- In year 4, 7 of these patients travel together to Iceland and die in a volcano accident.
- The other 93 patients made it to the end of five years without passing away.

WHAT IS THE MORTALITY RATE?

Medical Example

- Cancer researcher finds a medicine that cures cancer.
- Run a medical study where you follow 100 patients for 5 years after giving them cancer cure to see how many die.
- In year 4, 7 of these patients travel together to Iceland and die in a volcano accident.
- The other 93 patients made it to the end of five years without passing away.

WHAT IS THE MORTALITY RATE?

DOES 7% FEEL RIGHT?

Customer Example

- Observe customers over the past year to try and analyze **voluntary** churn.
- Of the 1000 customers in the data set, 240 left voluntarily, while 60 left involuntarily.

WHAT IS THE CUSTOMER CHURN RATE?

Customer Example

- Observe customers over the past year to try and analyze **voluntary** churn.
- Of the 1000 customers in the data set, 240 left voluntarily, while 60 left involuntarily.

WHAT IS THE CUSTOMER CHURN RATE?

DOES 30% FEEL RIGHT?

Other Forms of Right Censoring

- The most common form of right censoring is where there is a specific cut-off date (Type I censoring).
- Other forms include the following:
  - Censor after a certain number of events occur (Type II censoring).
  - Censoring occurs due to other reason outside of investigator controls (random censoring).

?

COMPETING RISKS

Competing Risks

- Medical Example:
  - The 7 patients who did not die due to cancer for 4 years of the study should be included as **censored** observations.

- Customer Example:
  - The 60 customers that left involuntarily should be considered **censored** observations.
  - If a customer is forced to leave at day 100, then they didn’t volunteer to leave on days 1-99!

Competing Risks

- All of the previous examples in the class have dealt with the notion of events being similar in nature.
- *Competing risks* distinguishes different kinds of events and treats them differently in the analysis.
- In competing risks, the occurrence of one type of event excludes an observation from another type of event.
  - If a person dies due to a volcano accident, they can't die due to cancer.
  - If a customer is forced to leave, then they can't voluntarily leave.

Type-Specific Hazard Rates

- When there are multiple event types, the hazard function contains two variables – *T* and *J*.
- The type specific hazard function is as follows:

$$h_{i,j}(t) = \lim_{\Delta t \to 0} \frac{P(t \le T_i < t + \Delta t, J_i = j \mid T_i \ge t)}{\Delta t}$$

$$h_i(t) = \sum_j h_{i,j}(t)$$

- The interpretation stays the same, just type specific.

Type-Specific Hazard Rates

- When there are multiple event types, the hazard function contains two variables – *T* and *J*.
- The type specific hazard function is as follows:

$$h_{i,j}(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T_i < t + \Delta t, J_i = j | T_i \geq t)}{\Delta t}$$

$$h_i(t) = \sum_j h_{i,j}(t)$$

- The interpretation stays the same, just type specific.

Type-Specific Hazard Probabilities

- Hazard probabilities are the same as the rates in how they change.
- When there are multiple event types, the hazard function contains two variables – *T* and *J*.
- The type specific hazard function is as follows:

$$h_{i,j}(t) = P(t \leq T_i < t + 1, J_i = j | T_i \geq t)$$

$$h_i(t) = \sum_j h_{i,j}(t)$$

- The interpretation stays the same, just type specific.

Other Forms of Right Censoring

- The most common form of right censoring is where there is a specific cut-off date (Type I censoring).
- Other forms include the following:
  - Censor after a certain number of events occur (Type II censoring).
  - Censoring occurs due to other reason outside of investigator controls (random censoring).

POTENTIAL PROBLEM!!!

Competing Risks Censoring

- In modeling a specific type of risk in competing risk models, the other risks are considered censored at the time of event.
- These censored observations must be *uninformative* to be considered unbiased.
  - Those who are at high risk for one event are no more or less likely to experience other kinds of events (conditional on the covariates).
  - If someone dies due to heart disease at age 50 it gives us no information about his risk of dying of a car accident at that age.

?

ESTIMATION WITHOUT COVARIATES

Graphical Differences

- Are the different types of events similar in how they fail?
- Events with similar failure rates (or survival rates) could potentially be modeled and grouped together.
- Already seen how to test for similarities.

Graphical Differences

```
data const;
     set Survival.Leaders;
     event=(lost=1);
     type=1;
run;

data nat;
     set Survival.Leaders;
     event=(lost=2);
     type=2;
run;

data noncon;
     set Survival.Leaders;
     event=(lost=3);
     type=3;
run;

data Survival.Leaders2;
     set const nat noncon;
run;

proc lifetest data=Survival.Leaders2 plots=lls;
     time years*event(0);
     strata type / diff=all;
run;
```

ESTIMATION WITH COVARIATES

Modeling Type-Specific Hazards

- Type-Specific hazards can be modeled with both proportional hazard models …

$$\log h_{i,j}(t) = \alpha_j(t) + \beta_{1,j}x_{i,1} + \cdots + \beta_{k,j}x_{i,k}$$

- … and accelerated failure time (AFT) models :

$$\log T_{i,j} = \beta_0 + \beta_{1,j}x_{i,1} + \cdots + \beta_{k,j}x_{i,k} + \sigma e_i$$

Likelihood Estimation

- There are two ways to estimate the parameters through likelihoods:
  1. Maximize the entire likelihood (or partial likelihoods)
  2. Maximize pieces of the likelihood function separately

- SAS requires separate models for the estimation.
  - Allows for different variables in each portion as well as different distributions (AFT models).

Time in Power for World Leaders Data

- Compiled by Bienen and van de Walle in 1991.
- Primary leaders of all countries between 1960 and 1987.
- Number of years the leader was in power and the manner they lost power:
  - Still in power (0)
  - Constitutional means (1)
  - Death from natural causes (2)
  - Non-constitutional means (3)

Variables to Predict Year

- Manner – how the leader reached power (0: constitutional, 1: non-constitutional)
- Start – year of entry to power
- Military – background of leader (1: military, 0: civilian)
- Age – age at time of entry
- Conflict – level of ethnic conflict (1: medium/high, 0:low)
- LogInc – log of GNP per capita
- Growth – avg. annual growth rate of GNP
- Pop – population in millions
- Land – land area in 1000 km^2
- Literacy – literacy rate (unknown year)
- Region – 0: Middle East, 1: Africa, 2: Asia, 3: Latin America

Comparing Covariates in Cox Models

- In Cox regression models, there is no underlying distribution to worry about matching across types of risk.
- When working with Cox regression models, competing risks can assessed by looking at the covariates that are significant in the different types of event models.
- Use log-likelihood test to test if $\beta_j = \beta$ for all j.

Comparing Covariates in Cox Models

```
proc phreg data=Survival.Leaders;
     class region;
     model years*lost(0) = manner start military
                           age conflict loginc growth
                           pop land literacy region;
run;

proc phreg data=Survival.Leaders;
     class region;
     model years*lost(0,2,3) = manner start military
                               age conflict loginc growth
                               pop land literacy region;
run;

proc phreg data=Survival.Leaders;
     class region;
     model years*lost(0,1,3) = manner start military
                               age conflict loginc growth
                               pop land literacy region;
run;

proc phreg data=Survival.Leaders;
     class region;
     model years*lost(0,1,2) = manner start military
                               age conflict loginc growth
                               pop land literacy region;
run;
```

Comparing Covariates in Cox Models

```
data LRT;
        All = 3455.69;
        Const = 1482.715;
        Natural = 225.583;
        NonConst = 1593.741;

        Sum = Const + Natural + NonConst;
        Diff = All - Sum;

        P_value = 1 - probchi(Diff,26);
run;

proc print data=LRT;
run;
```

AFT Models with Competing Risks

- Accelerated Failure Time models have a similar structure to Cox regression models when dealing with competing risks.
- With AFT Models, distributions need to be evaluated for all types of failure!

AFT Models with Competing Risks

```
data Survival.Leaders3;
     set Survival.Leaders;
     lower = years;
     upper = years;
     if years = 0 then do;
          lower = .;
          upper = 1;
     end;
     if lost in (0,1,2) then upper = .;
run;

proc lifereg data=Survival.Leaders3;
     class region;
     model (lower, upper) = manner start military
                            age conflict loginc
                            literacy region / dist=gamma;
run;
```

CONDITIONAL PROCESSES

Independent Events?

- The competing risks approach presumes that each event type has its own hazard that governs **both** the occurrence and timing of events of that type.
- They are assumed to be independent processes acting in parallel with each other.
- Example:
    - Death due to natural causes vs. forcible removal from power.

Independent Events?

- In a business setting, this independence assumption rarely seems reasonable.
- Example:
  - Consider the event to be buying a personal computer.
  - Two types:
    - Mac
    - PC
  - These aren’t two independent processes where we see what happens first.
- One process governs when you will buy a computer, while another process determines choice of computer.

Two-Stage Modeling

- One process governs when you will buy a computer, while another process determines choice of computer.
  1. Survival Analysis on buying a computer.
  2. Logistic regression on which type of computer to buy.