Classification & Regression

Repeated Measures – Mixed Effects Models
Recap: Compare Two Groups

KM Estimates for Leukemia Data

- Treatment
- Control

Weeks in Remission

0 5 10 15 20 25 30 35
Recap: Word of Caution

• What we will talk about today operates under the assumption that the curves we are comparing are “proportional”
  – I.e. the distance between them remains the same except for a constant factor

• This is of course a strong assumption, but is the basis for much of the survival analysis methodology
  – There are a number of methods to account for cases where this is not true, we can address these outside of class if anyone uses this for their projects
Statistically Testing Differences

• When looking to compare two survival curves, the method is often dictated by the question asked
  – Comparison at fixed time points
    • Is 5 year survival rate after an individual’s first heart attack different between two groups
  – Comparison of whole curve
    • Is the probability of dying after admission different for two (or more) groups
Modeling Time To Event Data: Pt. 2

• Survival data are generally described and modelled in terms of two related probabilities, namely *survival* and *hazard*.
  – The survival probability $S(t)$ is the probability that an individual survives from the time origin to a specified future time $t$.
  – The hazard is usually denoted by $h(t)$ is the probability that an individual who is under observation at a time $t$ has an event at that time.
    • Put another way, it represents the instantaneous event rate for an individual who has already survived to time $t$.

• The hazard relates to the incident (current) event rate, while survival reflects the cumulative non-occurrence.
Hazard Function

\[ h(t) = \lim_{\Delta t \to \infty} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \]

- The hazard is **not a probability** because we are dividing the probability by a time interval. Rather it is a rate.
  - Since we are dealing with the condition of survival up to time t, sometimes the hazard function is referred to as **conditional failure rate**.
- The rate of occurrence of the event at duration t equals the density of events at t, divided by the probability of surviving to that duration without experiencing the event.
Hazard Function

- Significantly more complex to compute
  - Often using the Nelson–Aalen estimator
- Is not monotonically decreasing,
  - Hazard can change (increase or decrease) over time.
Some Common Hazard Function Shapes

- **Increasing hazard** function: often seen at retirement age.
- **Constant risk** of death from accident or rare disease (Poisson Process).
- **Bathtub** curve: often real life from birth to old age.
- **Decreasing** risk, e.g. after an operation.
Hazard Ratio

• The ratio of the number of events observed to that expected assuming the null hypothesis

\[
HR = \frac{\frac{Obs_a}{Exp_a}}{\frac{Obs_b}{Exp_b}}
\]

• The total expected and observed from each group
  - Easy to estimate from the log rank calculations
### Hazard Ratio

- \( HR = \frac{\text{Obs}_a}{\text{Exp}_a} \div \frac{\text{Obs}_b}{\text{Exp}_b} \)
- \( HR = \frac{4}{2.89} \div \frac{3}{4.110} \)
- \( = 1.896 \)

#### Table

<table>
<thead>
<tr>
<th>Time, Weeks</th>
<th>Number at Risk Group 1</th>
<th>Number at Risk Group 2</th>
<th>Total Number at Risk</th>
<th>Number of Relapses Group 1</th>
<th>Number of Relapses Group 2</th>
<th>Total Number of Relapses</th>
<th>Expected Number of Relapses in Group 1</th>
<th>Expected Number of Relapses in Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>8</td>
<td>8</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>8</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.467</td>
<td>0.533</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>7</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.462</td>
<td>0.538</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>7</td>
<td>12</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.417</td>
<td>0.583</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.444</td>
<td>0.556</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.600</td>
<td>0.400</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>3</td>
<td></td>
<td>2.890</td>
<td>4.110</td>
<td></td>
</tr>
</tbody>
</table>
Interpretation of HR

• Let’s pretend we have a hazard ratio of 2
  – This technically means that treatment will cause the patient to heal faster, but in a very specific sense.

• In the context of hazard ratio, “fast” means that a treated patient who has not yet healed by a certain time has twice the chance of being healed at the next point in time compared to someone in the control group.

Taken From Hazard Ratio in Clinical Trials Spotswood L. Spruance, et. al.
Building on What We Know

- Kaplan–Meier curves are great, but they have two distinct limitations.
  - Time to event data, focuses on only a single feature
    - Why is this an issue?
Building on What We Know

• Kaplan–Meier curves are great, but they have two distinct limitations.
  – Time to event data, focuses on only a single feature
    • It does not allow us to account for confounding variables!
      – We can of course account for this by creating multiple curves, but we reduce the information available to model/compare each group
Building on What We Know

• Kaplan–Meier curves are great, but they have two distinct limitations.
  – Time to event data, focuses on only a single feature
    • It does not allow us to account for confounding variables!
      – We can of course account for this by creating multiple curves, but we reduce the information available to model/compare each group
  – Give us an “average” view of the population.
    • Yet, we can also have covariate data at the individual level, either continuous or categorical, that we would like to use. For this, we turn to survival regression
Limitations

• Requires categorical “groups”
• These “groups” can not change over time
  – The groups are not shifting along the X-axis
Tying it All Together
Cox (proportional hazards) Regression

\[ h(t) = h_0(t) \times e^{x_1\beta_1 + x_2\beta_2 + \cdots} \]
Cox (proportional hazards) Regression

\[ h(t) = h_0(t) \times e^{x_1 \beta_1 + x_2 \beta_2 + \cdots} \]

Baseline Hazard

Covariate features
Cox (proportional hazards) Regression

\[
\frac{h(t)}{h_0(t)} = e^{x_1\beta_1 + x_2\beta_2 + \ldots}
\]

There is no intercept in the model:
“intercept” is really the unspecified baseline hazard, \(h_0(t)\)
Cox (proportional hazards) Regression

\[
\frac{h(t)}{h_0(t)} = e^{x_1 \beta_1 + x_2 \beta_2 + \ldots}
\]

\[
\ln \left( \frac{h(t)}{h_0(t)} \right) = \ln \left( e^{x_1 \beta_1 + x_2 \beta_2 + \ldots} \right) = \ln \left( e^{x_1 \beta_1 + x_2 \beta_2 + \ldots} \right) = \\
\ln \left( \frac{h(t)}{h_0(t)} \right) = x_1 \beta_1 + x_2 \beta_2 + \ldots
\]
Cox (proportional hazards) Regression

\[ \ln \left( \frac{h(t)}{h_0(t)} \right) = x_1 \beta_1 + x_2 \beta_2 + \cdots \]

- From here, the standard regression model
- Assumptions
  - Proportional hazards (this includes each of the covariates)
- Fitted with an optimization method
  - Similar to logistic regression, (instead of MLE we use partial likelihood)
- Handles confounding variables
  - Same as other regression (interpret each where all others are held constant)
What If Covariates Aren’t Proportional?

- Sometimes a covariate may not obey the proportional hazard assumption. In this case, we can allow a factor without estimating its effect to be adjusted using a stratified Cox model.
  - The math is a bit beyond this course, but it is implemented in python much like groups in Mixed effects models.
    - We include it within the model, but do not estimate coefficients
    - Must be nominal
Interpretation

• The estimated coefficients in the Cox proportional hazards regression model, represent the change in the expected log of the hazard ratio relative to a one unit change in $X_i$, holding all other predictors constant.
  – Just like logistic regression, it is useful to transform the coefficients

• It is highly useful to center your predictors
  – Think of the intercept...
Similar to Other Regression Outputs

```r
> summary(cox.ph.m)
Call:
coxph(formula = Surv(Time2Event, attrition) ~ CUSTOMER_AGE + MoB, data = wSurv.data)
n= 1957, number of events= 607

 coef exp(coef)   se(coef)     z     Pr(>|z|)    
CUSTOMER_AGE -0.003104  0.996900  0.004429 -0.701 0.4834  
MoB -0.276361  0.758539  0.018237 -15.154 <2e-16 ***

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

exp(coef) exp(-coef) lower .95    upper .95
CUSTOMER_AGE 0.996900  1.003048  0.98834  1.00561
MoB 0.758539  1.318053  0.7319  0.78611
```

Standard Error of Beta Coefficients

Z is wald statistics and is calculated by dividing $\beta$ with its standard error. It is assumed as asymptotically standard normal under the hypothesis that the corresponding $\beta$ is zero.

P Value corresponding to Z statistics. If P value is lower than 5% then Null Hypothesis of $\beta=0$ can be rejected for 95% confidence level.
Time Dependent Covariates!

• Violation of the proportional hazards assumption for a given covariate is equivalent to that covariate having a significant interaction with time.

  Ex.
  – A study investigating the relationship between cancer rate and smoking. It is possible (likely) individuals can change smoking habits throughout the study.

• Very useful in real-world settings
  – But extreme care should be taken on prediction outside of known ranges (extrapolation)

<table>
<thead>
<tr>
<th>id</th>
<th>start</th>
<th>stop</th>
<th>group</th>
<th>z</th>
<th>event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>False</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>False</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>True</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>False</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>12</td>
<td>1</td>
<td>1</td>
<td>True</td>
</tr>
</tbody>
</table>
Next Class

• Monday Quiz 2
  – Linear Regression
  – Logistic Regression
  – Mixed Effects Modes

• As with Quiz 1
  – No coding, No Math
  – Full class time for quiz
  – Look over assignment 3

• Review session Sunday Evening in iCeNSA 6pm-8pm