SEQUENCE SURVIVORS: ANALYSIS OF GRADUATION OR TRANSFER IN THREE YEARS OR LESS

By: Leila Jamoosian
   Terra Morris
   Terrence Willett

Spring 2018
THE EVOLUTION OF INTELLECTUAL FREEDOM

I'm going to research whatever I want!

I'm going to research whatever my professor wants!

I'm going to research whatever my tenure committee wants!

I'm going to research whatever my grant committee wants!

I'm going to research whatever I want!

Before grad school

Grad student

Assistant professor

Tenured professor

Emeritus professor

"Research in Peace!"
<table>
<thead>
<tr>
<th>Developmental Stage</th>
<th>Number Alive</th>
<th>Mortality Factor(s)</th>
<th>Number Dying</th>
<th>Percent Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egg</td>
<td>200</td>
<td>predation</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>Larva</td>
<td>100</td>
<td>parasitization</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>Pupa</td>
<td>10</td>
<td>freezing</td>
<td>6</td>
<td>60</td>
</tr>
<tr>
<td>Adult</td>
<td>4</td>
<td>(2 females and 2 males if sex ratio is 1:1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Monarch Butterfly Life Cycle Diagram](image-url)
Table 1. Life table for the total population: United States, 2003

<table>
<thead>
<tr>
<th>Age</th>
<th>$q_x$</th>
<th>$l_x$</th>
<th>$d_x$</th>
<th>$L_x$</th>
<th>$T_x$</th>
<th>$\delta_x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>0.006865</td>
<td>100,000</td>
<td>687</td>
<td>99,394</td>
<td>7,743,016</td>
<td>77.4</td>
</tr>
<tr>
<td>1-2</td>
<td>0.000469</td>
<td>99,313</td>
<td>47</td>
<td>99,290</td>
<td>7,643,622</td>
<td>77.0</td>
</tr>
<tr>
<td>2-3</td>
<td>0.000337</td>
<td>99,267</td>
<td>33</td>
<td>99,250</td>
<td>7,544,332</td>
<td>76.0</td>
</tr>
<tr>
<td>3-4</td>
<td>0.000254</td>
<td>99,233</td>
<td>25</td>
<td>99,221</td>
<td>7,445,082</td>
<td>75.0</td>
</tr>
<tr>
<td>4-5</td>
<td>0.000194</td>
<td>99,208</td>
<td>19</td>
<td>99,199</td>
<td>7,345,861</td>
<td>74.0</td>
</tr>
<tr>
<td>5-6</td>
<td>0.000177</td>
<td>99,189</td>
<td>18</td>
<td>99,180</td>
<td>7,246,663</td>
<td>73.1</td>
</tr>
<tr>
<td>6-7</td>
<td>0.000160</td>
<td>99,171</td>
<td>16</td>
<td>99,163</td>
<td>7,147,482</td>
<td>72.1</td>
</tr>
<tr>
<td>7-8</td>
<td>0.000147</td>
<td>99,156</td>
<td>15</td>
<td>99,148</td>
<td>7,048,319</td>
<td>71.1</td>
</tr>
<tr>
<td>8-9</td>
<td>0.000132</td>
<td>99,141</td>
<td>13</td>
<td>99,134</td>
<td>6,949,171</td>
<td>70.1</td>
</tr>
<tr>
<td>9-10</td>
<td>0.000117</td>
<td>99,128</td>
<td>12</td>
<td>99,122</td>
<td>6,850,036</td>
<td>69.1</td>
</tr>
<tr>
<td>10-11</td>
<td>0.000109</td>
<td>99,116</td>
<td>11</td>
<td>99,111</td>
<td>6,750,914</td>
<td>68.1</td>
</tr>
<tr>
<td>11-12</td>
<td>0.000118</td>
<td>99,105</td>
<td>12</td>
<td>99,100</td>
<td>6,651,803</td>
<td>67.1</td>
</tr>
<tr>
<td>12-13</td>
<td>0.000157</td>
<td>99,094</td>
<td>16</td>
<td>99,086</td>
<td>6,552,704</td>
<td>66.1</td>
</tr>
<tr>
<td>13-14</td>
<td>0.000233</td>
<td>99,078</td>
<td>23</td>
<td>99,067</td>
<td>6,453,618</td>
<td>65.1</td>
</tr>
<tr>
<td>60-61</td>
<td>0.009827</td>
<td>87,774</td>
<td>863</td>
<td>87,343</td>
<td>1,945,136</td>
<td>22.2</td>
</tr>
<tr>
<td>61-62</td>
<td>0.010831</td>
<td>86,911</td>
<td>941</td>
<td>86,441</td>
<td>1,857,793</td>
<td>21.4</td>
</tr>
<tr>
<td>62-63</td>
<td>0.011672</td>
<td>85,970</td>
<td>1021</td>
<td>85,460</td>
<td>1,771,352</td>
<td>20.6</td>
</tr>
<tr>
<td>63-64</td>
<td>0.012891</td>
<td>84,949</td>
<td>1095</td>
<td>84,402</td>
<td>1,685,892</td>
<td>19.8</td>
</tr>
<tr>
<td>64-65</td>
<td>0.013908</td>
<td>83,854</td>
<td>1166</td>
<td>83,271</td>
<td>1,601,490</td>
<td>19.1</td>
</tr>
<tr>
<td>65-66</td>
<td>0.015003</td>
<td>82,688</td>
<td>1241</td>
<td>82,068</td>
<td>1,518,219</td>
<td>18.4</td>
</tr>
<tr>
<td>66-67</td>
<td>0.016267</td>
<td>81,448</td>
<td>1325</td>
<td>80,785</td>
<td>1,436,151</td>
<td>17.6</td>
</tr>
</tbody>
</table>
WHY DO WE WANT STUDENTS TO NOT “SURVIVE” COLLEGE MORE QUICKLY

- Traditional survival analysis tracks individuals until they perish
- Applied to a college setting, perishing is completing a degree or transfer
- Past analyses have given many years to allow students to show a completion (e.g. 8 years)
- With new incentives, completion in 3 years is critical
WHY SURVIVAL NOT LINEAR REGRESSION?

- Time is not normally distributed
- No approach to deal with Censors
To estimate time that students graduate or transfer in three years or less (time-to-event=1/ incidence rate) for a group of individuals.

- We can compare time to graduate or transfer in three years between two of more groups.

- Or assess the relation of co-variables to the three years graduation event such as number of credit, full time status, number of withdraws, placement type, successfully pass at least transfer level Math or English courses, …
Definitions:

- **Time-to-event**: Number of days from beginning the program until graduating

- **Outcome**: Completing AA/AS degree or transfer in three years or less

- **Censoring**: Students who drop or those who do not graduate in three years or less (takes longer to graduate or fail)

  - **Note 1**: Students who completed other goals such as certificate of achievement or skill certificate were not counted as success in this analysis

  - **Note 2**: First timers who got their first degree at Cabrillo (they may come back for second degree or more), we consider only first degree

  - **Note 3**: Analysis included students who were at Cabrillo at least one semester and took at least 12 credits
Statistics:

- **Total:** 20,727 students
- **Total 12>=credits:** 17,309 students
- **Achieved Degree (AA/AS) between 2007-17:** 2,629 Students (13%)
- **Achieved Degree (AA/AS) & Transfer-BA/MA (Transfer:12>= credits):** 1,311 Students (49.9%)
- **Calendar (eight different year cohorts):**
  - MIN-MAX Start at Cabrillo: 2007-2014
  - MIN-MAX Degree or Transfer: 2010-2017
    - 2007-2010
    - 2008-2011
    - ...
    - 2014-2017
- **Transfer who achieved BA or MA (Transfer:12>= credits) NOT Degree (AA/AS) between 2007-17:** 2,389 Students (13.80%)
- **Transfer (Transfer:12>= credits) (2-4, 0-4) and/or achieved a Degree (2-4,2-0) (AA/AS) Less than 3 Years in Calendar:** 1,328 Students (7.67%)
- **Censoring:** Students who drop or those who did not achieve degree in Calendar
Fig. 1.1 The vertical dashed lines indicate the start, end of accrual, and end of follow-up. The `x`'s denote events (complete at least one Math transfer level course in 3 semesters) and the open circles denote censoring events.
**SNAPSHOT OF DATA**

Note: ID.CC are not student ids
Statistics:

- **Case of Censoring:**
  - Dropping school temporary or permanently
  - Completing after the cut off day (1200 days)
  - Lost data about students (going to another college, data loss, etc.)
  - Being anonymous transfer students, no matching data at NSC (National Student Clearinghouse)

- Why “right censoring" method? Because we do not know when they leave the college, but we know how many days they were at Cabrillo college.
Statistics:

- **65% (N=13,425)** of students who had “Education Goal” of AA/AS degree (N=1,606, 8%) and of Transfer (N=11,819, 57%)
Calendar of interest (2007-2017)- eight years cohorts

Time (Number of days)- Random Variable

Probability of students not completing between $P(900 < T < 1200) = ?$

$T_i$, persistence, is not normally distributed

\[
f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}
\]
Density function of Time

Time

Density
NSC LIMITATION (NATIONAL STUDENT CLEANING HOUSE)

- Not all students trackable
  - Student’s agreement required for sharing data
  - Fuzzy match problem
- Not all schools participate in NSC (although >90% of U.S. schools do)
COUNT OF STUDENTS ACHIEVING AA/AS DEGREES BETWEEN 2007-17 2,629
COUNT OF STUDENTS ACHIEVING AA/AS DEGREES/TRANSFER LESS THAN THREE YEARS IN CALENDAR 1,328
\( S(t \leq 1200) = \text{Probability of not completing a degree more than three years (or 1200 days)} = 1 - F(t) \)

\( F(t) \) is cumulative distribution of \( f(t) \) which we like to know more about it more than \( f(t) \)

“Very Important”

**Risk rate (Hazard):** Risk of completion between those at the risk of completion.

If a group of students persists (survives) to time \( t \), the hazard will be the rate of students of surviving to the next instant.

\[
 h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t / T \geq t)}{\Delta t}
\]

\[ h(t) = \frac{f(t)}{S(t)} = -\frac{d \ln S(t)}{dt} \]
Kaplan-Meier Estimation and Curve

The easiest way to calculate survival (not completion rate) is Kaplan-Meier estimate. You can calculate it by hand, but not able to calculate covariate effects.

\[ S(999) = P(T > 999) = 1 - P(T < 999) = 87.5 \]

87.5% is the probability of the time from starting at Cabrillo College to complete (Degree or Transfer) is greater than 999 days.
What software run survival analyses? All of them!
SPSS
STATA
SAS
R

R sounds like the most fun. How can I learn more?

R4IR

R codes for this presentation:
https://github.com/LJamoosian/Survival-Analysis.git
Using Log Rank Test

\[ H_0 : S_1 = S_2 \]
\[ S_{\downarrow 1} = \frac{1}{10250} \approx 1 \]
\[ S_{\downarrow 2} = (1 - \frac{87}{10249}) \] * \[ \frac{1}{10250} \]
$S(T) = P(t \leq T) = 1 - P(t > T)$
Using Stratified Log Rank Test

\[ H_0 : S_1 = S_2 = S_3 = S_4 = S_5 \]
Survival Curves of Placement Level in Math

Math Placement Level: -1, -2, -3, -4, 0

$p < 0.0001$

Cabrillo Students Earning a Degree or Transferring within 1,000 days by Initial Math Placement Level

<table>
<thead>
<tr>
<th>Math Placement Level</th>
<th>Percent Completed in 1,000 Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>23%</td>
</tr>
<tr>
<td>-1</td>
<td>15%</td>
</tr>
<tr>
<td>-2</td>
<td>7%</td>
</tr>
<tr>
<td>-3</td>
<td>5%</td>
</tr>
<tr>
<td>-4</td>
<td>3%</td>
</tr>
</tbody>
</table>

Time (Number of Days)
“MATH IS A BARRIER. 92% OF STUDENTS NEAR THE TRANSFER GATE STILL NEED TO COMPLETE TRANSFER-LEVEL MATH REQUIREMENTS IN ORDER TO PROCEED ALONG THE PATH TO UNIVERSITY.”
WWW.RPGGROUP.ORG/THROUGH-THE-GATE

One Year Transfer Math Completion for 2014-2015 First Time Scorecard Cohort

Completion Non-Completion

Female 23% 77%
Male 31% 69%

* Pacific Islander
* Native American
* African American
Hispanic 14% 86%
Filipino 22% 79%
White 29% 71%
Asian 33% 67%

67%

* Rate suppressed due to low count
The education path for students at Cabrillo College is often too long, confusing, and costly.

Of first-time degree seeking students who entered Cabrillo in 2010, only 2 in 10 had transferred by 2016 and only 1 in 10 transferred with a degree.
Survival Curves of Placement Level in English

English Placement Level + -1 + -2 + -3 + 0

1 - Probability (Completion)

p < 0.0001

Time (Number of Days)
Survival Curves of Placement Level in Math

Placement Type + Others - Placement Test

p < 0.0001

Time (Number of Days)
Survival Curves of Graduated & Transferred Students to Higher Education Level

Graduation and Transfer Type: 0-0, 0-4, 2-0, 2-4

- 0-0 = no 2 or 4 yr degree
- 0-4 = 4 yr degree only
- 2-0 = 2 yr degree only
- 2-4 = 2 and 4 yr degree

p < 0.0001
Survival Curves of Full Time Status

Full Time Category: Mostly Full-Time/More than 75% + Mostly Part-Time/Less than 35% + Mix Full-Time/35%-75%

1 - Probability (Completion)

Time (Number of Days)

p < 0.0001
Survival Curves of Number of Times that Students Withdraw

Number of Withdraw: More Than 4 + None - One - Three + Two

1 - Probability (Completion)

Time (Number of Days)

p < 0.0001
KP Curve of Underrepresented Minorities

URM + No - Yes

$p < 0.0001$
WHAT OPTIONS WE HAVE TO SEE RELATION OF SURVIVAL PROBABILITY AND ALL COVARIATES?

- **Parametric method**: AFT model (Accelerated Failure Time)
- **Non-Parametric method**: Cox regression model
COX REGRESSION ANALYSIS
Why non-parametric model instead of parametric?

- No assumption about underlying distribution of survival time
- More efficient when no suitable theoretical distributions are known

COX MODEL ASSUMPTION
PROPORTIONAL HAZARD RATIO

$$\lambda(t) = \lambda_0(t) \cdot \exp(\beta_1 X_1 + \ldots + \beta_p X_p)$$
my.surv = Surv(db1$TIME, db1$STATUS)
coxph.fit = coxph(my.surv ~ as.factor(db1$mathlev) + as.factor(db1$englev) +
                 db1$`# COMP CRED` + as.factor(db1$Withdraw) + db1$`Financial Aid` +
                 as.factor(db1$GENDER) + as.factor(db1$FullTimeCat) + db1$URM.,
                 method = "efron") # other options "breslow" and "exact" and "efron"
summary(coxph.fit)

> summary(coxph.fit)
Call:
coxph(formula = my.surv ~ as.factor(db1$mathlev) + as.factor(db1$englev) +
       db1$`# COMP CRED` + as.factor(db1$Withdraw) + db1$`Financial Aid` +
       as.factor(db1$GENDER) + as.factor(db1$FullTimeCat) + db1$URM.,
       method = "efron")

n = 10218, number of events = 1392
(36 observations deleted due to missingness)

             coef  exp(coef)    se(coef)      z     Pr(>|z|)  
as.factor(db1$mathlev)1  0.371772  1.450303  0.069800  5.326  1.00e-07 ***
as.factor(db1$englev)1  0.119170  1.126561  0.061815  1.928  0.0539 .
db1$`# COMP CRED`    -0.009071  0.990970  0.010169 -8.486  < 2e-16 ***
as.factor(db1$Withdraw)None  2.221134  9.217779  0.101191 21.950  < 2e-16 ***
as.factor(db1$Withdraw)One  1.826218  6.210357  0.107606 16.971  < 2e-16 ***
as.factor(db1$Withdraw)Three  1.090732  2.976451  0.131803  8.275  < 2e-16 ***
as.factor(db1$Withdraw)Two  1.430103  4.179131  0.116956 12.228  < 2e-16 ***
db1$`Financial Aid`1   -0.120431  0.886538  0.059276 -2.032  0.0422 *
as.factor(db1$GENDER)M  -0.375020  0.687276  0.055145 -6.801  1.04e-11 ***
as.factor(db1$FullTimeCat)LowFullTime -2.468293  0.084729  0.083172 -29.677  < 2e-16 ***
as.factor(db1$FullTimeCat)MidFullTime -1.247531  0.287213  0.064772 -19.260  < 2e-16 ***
db1$URM.1               -0.316693  0.728554  0.058053 -5.455  4.89e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
<table>
<thead>
<tr>
<th>Variable</th>
<th>Exp(coef)</th>
<th>Exp(-coef)</th>
<th>Lower .95</th>
<th>Upper .95</th>
</tr>
</thead>
<tbody>
<tr>
<td>as.factor(db1$mathlev)1</td>
<td>1.45030</td>
<td>0.6895</td>
<td>1.26487</td>
<td>1.66292</td>
</tr>
<tr>
<td>as.factor(db1$englev)1</td>
<td>1.12656</td>
<td>0.8877</td>
<td>0.99802</td>
<td>1.27166</td>
</tr>
<tr>
<td>db1$<code># COMP CRED</code></td>
<td>0.99097</td>
<td>1.0091</td>
<td>0.98890</td>
<td>0.99305</td>
</tr>
<tr>
<td>as.factor(db1$Withdraw)None</td>
<td>9.21778</td>
<td>0.1085</td>
<td>7.55949</td>
<td>11.23983</td>
</tr>
<tr>
<td>as.factor(db1$Withdraw)One</td>
<td>6.21036</td>
<td>0.1610</td>
<td>5.02947</td>
<td>7.66851</td>
</tr>
<tr>
<td>as.factor(db1$Withdraw)Three</td>
<td>2.97645</td>
<td>0.3360</td>
<td>2.29883</td>
<td>3.85380</td>
</tr>
<tr>
<td>as.factor(db1$Withdraw)Two</td>
<td>4.17913</td>
<td>0.2393</td>
<td>3.32302</td>
<td>5.25581</td>
</tr>
<tr>
<td>db1$<code>Financial Aid</code>1</td>
<td>0.88654</td>
<td>1.1280</td>
<td>0.78930</td>
<td>0.99576</td>
</tr>
<tr>
<td>as.factor(db1$GENDER)M</td>
<td>0.68728</td>
<td>1.4550</td>
<td>0.61687</td>
<td>0.76572</td>
</tr>
<tr>
<td>as.factor(db1$FullTimeCat)LowFullTime</td>
<td>0.08473</td>
<td>11.8023</td>
<td>0.07198</td>
<td>0.09973</td>
</tr>
<tr>
<td>as.factor(db1$FullTimeCat)MidFullTime</td>
<td>0.28721</td>
<td>3.4817</td>
<td>0.25297</td>
<td>0.32609</td>
</tr>
<tr>
<td>db1$URM.1</td>
<td>0.72855</td>
<td>1.3726</td>
<td>0.65020</td>
<td>0.81635</td>
</tr>
</tbody>
</table>

Concordance = 0.804  (se = 0.008)
Rsquare = 0.167    (max possible = 0.915)
Likelihood ratio test = 1867  on 12 df,  p=0
Wald test         = 1740  on 12 df,  p=0
Score (logrank) test = 2129  on 12 df,  p=0
What recommendations should we give?
Congratulations!
You Survived!

Primary Contact:
Leila Jamoosian
lejamoos@cabrillo.edu