Using Censored Data to Predict the Time Required to Complete a Bachelor's Degree

Ronald D. Thompson, Ph.D.
Associate Institutional Research Analyst

Office of Institutional Research & Planning
University of Missouri System
721 Lewis Hall
Columbia, Missouri 65211.

ABSTRACT

ThompsonRD: Using Censored Data to Predict the Time Required to Complete a Bachelor's Degree
Mid-America Association for Institutional Research (MidAIR)
31st Annual Conference
November 9-11, 2011
MidAIRTalk_20111111_Splash

MidAIR 2011 Conference Proposal

Title: Using Censored Data to Predict the Time Required to Complete a Bachelor's Degree

Author: Ronald D. Thompson, Ph.D.
Associate Institutional Research Analyst
Office of Institutional Research & Planning
University of Missouri System

Abstract: The time required for a first-time-college undergraduate student admitted to a Midwestern research university to complete a bachelor's degree is predicted using censored completions data.

Detailed Description:

Time-to-completion statistics are usually computed using student cohort groups which entered a university many years earlier. For example, to compute today (in Fall Semester 2011) the average time-to-completion of first-time-college (FTC) undergraduate students, one approach is to utilize the times-to-completion of the entering Fall Semester 2001 cohort (say), since ten years of completions data are available for this cohort—the tacit assumption being, of the students in this Fall Semester 2001 cohort who will graduate, only a very small, negligible percentage will require longer than ten years to do so.

The concern is, if our primary interest is in predicting the time-to-completion of a student belonging to our current new Fall Semester 2011 entering cohort, a statistic based on the yet,

ificant amount of unknown, or incomplete, data.

hether completions data corresponding to more recent FTC undergraduate

Outline

- Reporting Mean Time to Receive First-Bachelor's Degree
- Censored Data
- Weibull Model
- Estimating Mean Time using a Weibull model and censored data in the case of a FTC undergraduate cohort
- Next Steps
- Q & A

Time-to-First-Bachelor's Degree (in Yrs), by FTC Cohort

--- CALYEAR=1999 --------------------------

The FREQ Procedure

Time_to_Degree	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2.73	28	1.36	28	1.36
2.93	8	0.39	36	1.75
3.31	75	3.64	111	5.39
3.73	874	42.41	985	47.79
3.93	78	3.78	1063	51.58
4.31	405	19.65	1468	71.23
4.73	388	18.83	1856	90.05
4.93	27	1.31	1883	91.36
5.31	63	3.06	1946	94.42
5.73	38	1.84	1984	96.26
5.93	7	0.34	1991	96.60
6.31	20	0.97	2011	97.57
6.73	12	0.58	2023	98.16
7.31	4	0.19	2027	98.35
7.73	4	0.19	2031	98.54
7.93	3	0.15	2034	98.69
8.31	5	0.24	2039	98.93
8.73	4	0.19	2043	99.13
8.93	2	0.10	2045	99.22
9.31	3	0.15	2048	99.37
9.73	4	0.19	2052	99.56
10.73	2	0.10	2054	99.66
10.93	1	0.05	2055	99.71
11.31	4	0.19	2059	99.90
11.73	2	0.10	2061	100.00

Frequency Missing = 785

Source: UM EMSAS Fall Enrollment & Completions Tbls ThompsonRD MidAIR 11/11/2011

--- CALYEAR=2000 --------------------------

The FREQ Procedure

Time_to_Degree	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.93	1	0.04	1	0.04
2.31	3	0.13	4	0.18
2.73	31	1.38	35	1.56
2.93	6	0.27	41	1.83
3.31	105	4.68	146	6.51
3.73	977	43.56	1123	50.07
3.93	89	3.97	1212	54.03
4.31	424	18.90	1636	72.94
4.73	371	16.54	2007	89.48
4.93	29	1.29	2036	90.77
5.31	77	3.43	2113	94.20
5.73	48	2.14	2161	96.34
5.93	5	0.22	2166	96.57
6.31	18	0.80	2184	97.37
6.73	18	0.80	2202	98.17
6.93	2	0.09	2204	98.26
7.31	5	0.22	2209	98.48
7.73	4	0.18	2213	98.66
7.93	4	0.18	2217	98.84
8.31	6	0.27	2223	99.11
8.73	6	0.27	2229	99.38
9.31	3	0.13	2232	99.51
9.73	4	0.18	2236	99.69
9.93	3	0.13	2239	99.82
10.73	4	0.18	2243	100.00

Frequency Missing = 768

Source: UM EMSAS Fall Enrollment & Completions Tbls ThompsonRD MidAIR 11/11/2011

University of Missouri System

PERF. MEASURE: Mean time out table of Times-to-Degree for FTC-1998 Col

MAJOR CONCERN: What to report?

*** Sketch a graph of completion times w.r.t. implied cut-off NOW! ***

ISSUES:

1. Don’t know how many remaining students will eventually graduate, and how many won’t ever.

2. For remaining students who will eventually graduate, don’t know when.

3. Any mean time that’s reported must necessarily underestimate.

4. How to choose the cut-off? (Presumably want this performance measure to reflect the current univ and current students as much as possible.)

University of Missouri System

COHORT: FTC, full-time, degree-seeking, undergrad, recent HS grad from a MO high school, satisfy UM-required HS core courses and either ACTCOM >= 24 or “120 Rule”. (Non-Resident Aliens excluded.)

METRIC: Elapsed Time to First-Bachelor’s Degree (in Yrs)

	FTC Cohort		
	1998	**1999**	**2000**
Cohort Size	2,974	2,846	3,011
Completions Through SP2011 (STRM='3927')			
N (No. of Grads)	2,197	2,061	2,243
Mean	4.3	4.3	4.2
Median	4.3	3.9	3.7
Max	12.3	11.7	10.7
Elapsed Time = 10 years			
N (No. of Grads)	2,188	2,052	2,239
Mean	4.3	4.2	4.2
Median			
Max	9.7	9.7	9.9
Elapsed Time = 8 years			
N (No. of Grads)	2,174	2,034	2,217
Mean	4.2	4.2	4.2
Median	4.1	3.9	3.7
Max	7.9	7.9	7.9

ThompsonRD:

Freq Dj-

Freq D

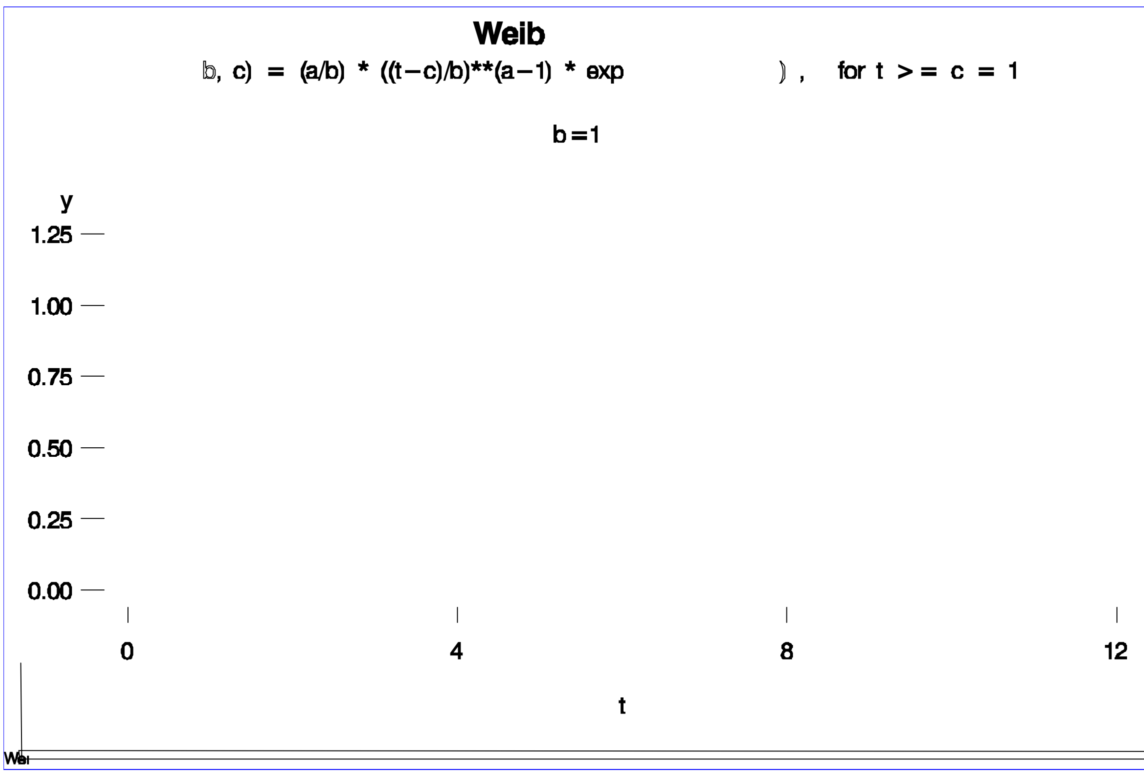
Time – to – Fi ohort

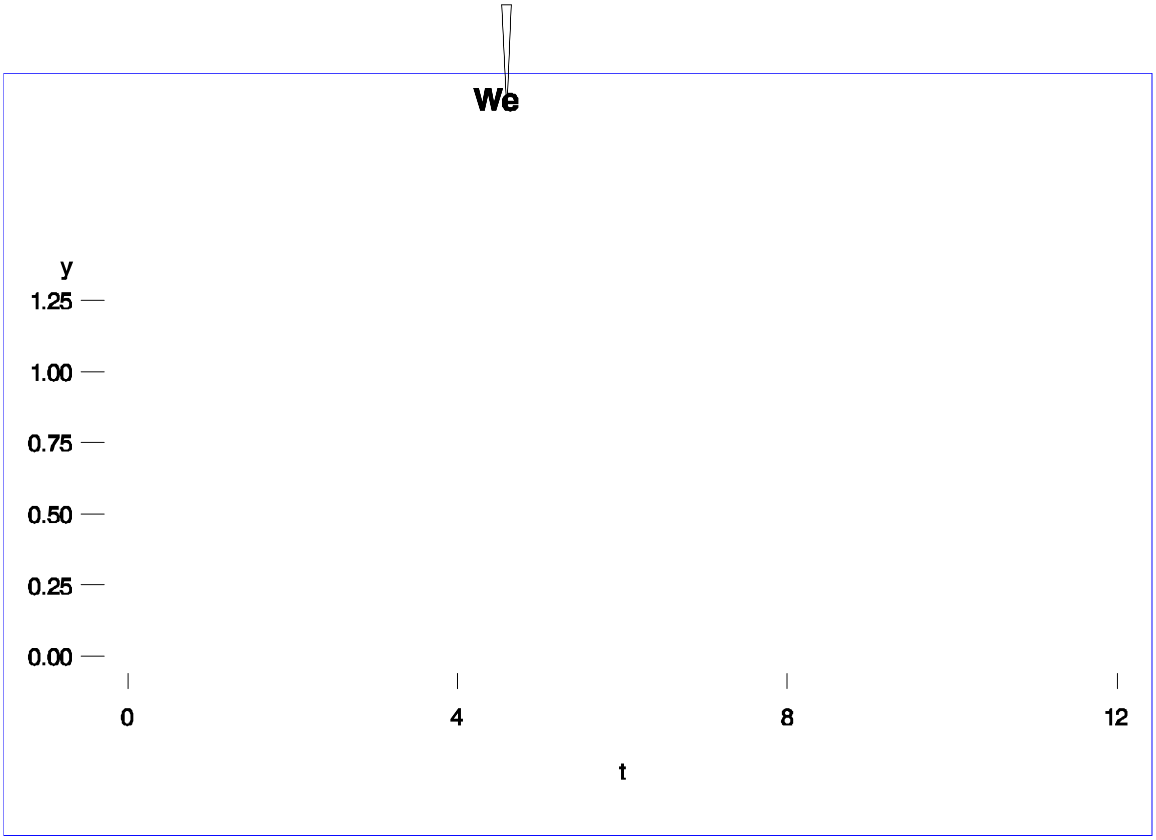
FTC, FT, DS, U/G, Recent HS Grad from a MO HS,
Satis

Source: UM EMSAS Fall Enrollment & Completions Tbls

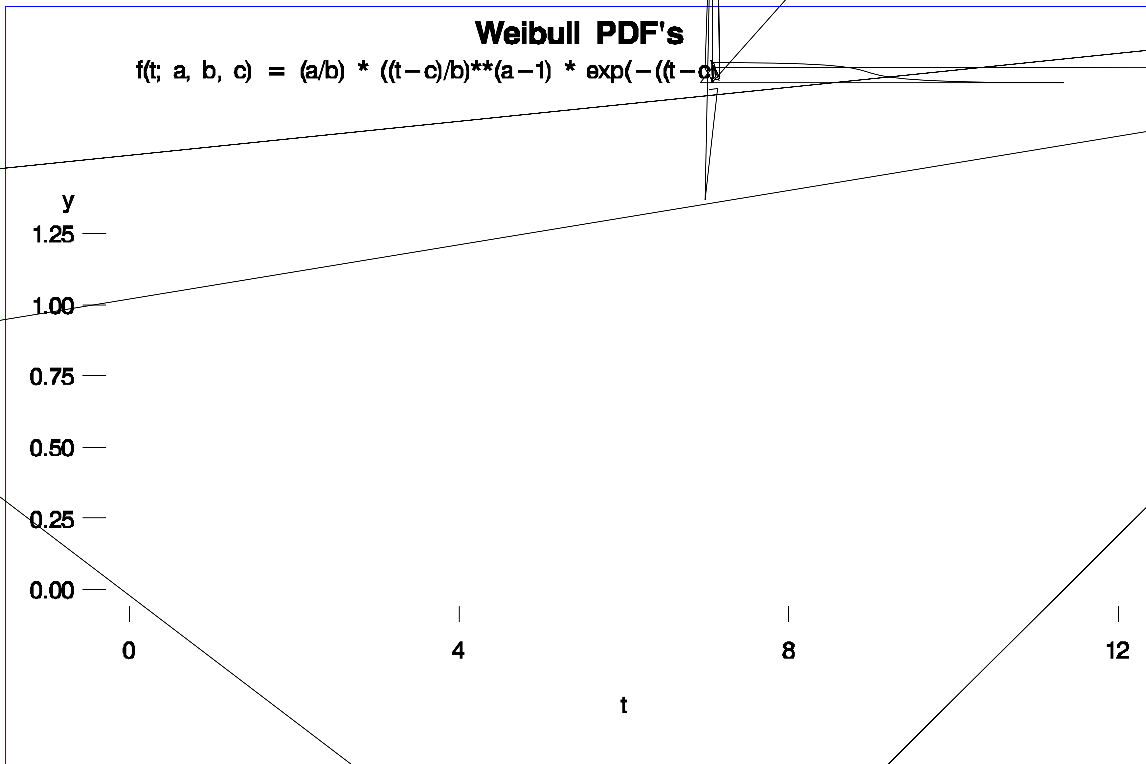
ThompsonRD MidAIR

Weibull Distribution





We
y
1.25
1.00
0.75
0.50
0.25
0.00
0
4
8
12
t


Weibull PDF's
f(t; a, b, c) = (a/b) * ((t−c)/b)**(a−1) * exp(−((t−c)
y
1.25
1.00
0.75
0.50
0.25
0.00
0
4
8
12
t

Weibull MLE's for FTC-1998 Cohort (N = 2,974)

1. Critical Decisions:

   a. Choose cut-off of 5 years (i.e., end of SS of 5th year, SS2003).

      1,983 graduated

      991 NOT graduated

      - 254 enrolled at FS census of next FS (FS2003)
      - 737 NOT enrolled

   b. Assume 254 will eventually graduate, and
      Assume of the 737 will

2. Likelihood Function

ὒ | ,ᛠ |– 2, Ὼὤὸὤ
پ,ٹ ٺ ٻ

θ ῏Ω ὸ ‖ ,ᛠ ,– 2 · 1 ῏Ο ὸ | ,ᛠ ,– 2
ڛ ٺ پ صزة ٻ ٺ پ

3. Log-Likelihood Function

ὒᶻ | ,ᛠ |– 2, Ὼὤὸὤ
ٹ

ὑ ῏Ω · ὰὲ ῏Ω ὼ ‖ ,ᛠ ,– 2 254 · 5 2 / ᛠ
ش ٺ پ

ML Estimato

4. Solution

| 4.11

↑ 2.43

– 2.00

Next Steps

1. Get true start times.
   Here, assumed start of FS (= 3rd week in Aug = 7.75/12) for everyone.

2. Investigate completion times.
   Here, assumed ends of FS (= mid Dec = 11.5/12), SP (= mid May = 4.5/12), or SS (= end July = 7.0/12).

3. Investigate Weibull location parameter –.
   Here, assumed – .

4. Obtain Posterior Bayes Estimates of Weibull parameters | and ↑ (and –), and then ‘, the mean time to graduation.
   Here, obtained Maximum Likelihood Estimates (MLE's) of | , ↑ , and

   In

   - of 1st FS after cut-off (end of 5th yr), will in fact graduate.

   - NOT registered at FS Census of 1st FS after cut-off will graduate.

6. Investigate cut-off.
   Here, used elapsed time of 5 yrs (end of SS of 5th yr).

7. Investigate reproducibility.
   Here, worked with FTC-1998 Cohort.

8. Entertain research questions.
   For Example: Do "today's" students at your institution take longer to graduate, on average, than past students? If so, then what are the practical implications, if any, to your institution (e.g., on policy, funding, curricula, and facilities including classrooms, laboratories, and housing)?

Summary

- Described in general terms the challenges of reporting mean time to receive a first-bachelor's degree.
- Motivated the use of censored data when estimating this mean time.
- Described the 3-parameter Weibull model and motivated its use here.
- Described the process of obtaining a Maximum Likelihood Estimator of this mean time, using the Weibull model and censored data drawn on a cohort of FTC 1998 undergraduate students. Discussed this estimate relative to the actual realized mean time for this cohort.
- Listed and discussed the next logical steps anticipated for this on-going res