

Event History Analysis With Roman Emperor's Data

Andy Grogan-Kaylor

2023-11-19

Table of contents

- 1 Introduction** **1**
- 2 Get Data and Data Wrangling** **2**
- 3 Cox Proportional Hazards Model** **3**
 - 3.1 `stset` the Data 3
 - 3.2 Formula for the Hazard 3
 - 3.3 Estimate the Cox Proportional Hazards Model 4
 - 3.4 Proportional Hazards Assumption 5
 - 3.5 Correcting For Violations of the Proportional Hazards Assumption 7
 - 3.6 Plot Curve Accounting for Covariates 8
- 4 References** **8**

1 Introduction

This example uses data on the ages at which Roman Emperors **began their reign**. Sources for this data are unclear, but it appears that the original source is <http://www.roman-emperors.org/> via <https://github.com/rfordatascience/tidytuesday/tree/master/data/2019/2019-08-13>.

2 Get Data and Data Wrangling

```
clear all

import delimited ///
"https://raw.githubusercontent.com/agrogan1/newstuff/master/categorical/survival-analysis-

list in 3 // list out an observation

generate birthyear = real(substr(birth, 1, 4)) // convert first 4 characters to real number

generate reignyear = real(substr(reign_start, 1, 4)) // convert first 4 characters to real number

generate agereign = reignyear - birthyear

drop if agereign <0 // drop negative ages at rise to power

histogram agereign

encode rise, generate(riseNUMERIC) // numeric version of rise to power

save emperors2.dta, replace
```

(encoding automatically selected: ISO-8859-1)
(16 vars, 68 obs)

```
+-----+
3. | index |      name |                                name_full |      birth |
|      3 | Caligula | GAIVS IVLIVS CAESAR AVGVSTVS GERMANICVS | 0012-08-31 |
+-----+
|      death | birth_~y | birth_~v |      rise | reign_st~t | reign_end |
| 0041-01-24 | Antitum | Italia | Birthright | 0037-03-18 | 0041-01-24 |
+-----+
|           cause | killer |           dynasty |           era |
| Assassination | Senate | Julio-Claudian | Principate |
+-----+
|                                           notes |
|      assassination may have only involved the Praetorian Guard |
+-----+
|                                           verif_who |
|                      Reddit user zonation |
+-----+
```

+-----+
(5 missing values generated)

(5 missing values generated)

(2 observations deleted)

(bin=7, start=4, width=10.714286)

file emperors2.dta saved

3 Cox Proportional Hazards Model

3.1 stset the Data

We need to `stset` the data so that Stata knows that this is survival data with special characteristics relevant to survival analysis. For those of you have used other commands that attach special characteristics to the data, this is similar to using `svyset` for complex survey data, `xtset` for panel data, or even to the `mi` suite of commands for multiple imputation.

The most commonly used syntax is something like `stset timevar, failure(failvar) id(id)`¹

There are many ways to specify `failvar`, we outline the most straightforward. Consult Stata help for your exact situation.

3.2 Formula for the Hazard

$h(t)$ the rate of occurrence.

$$h(t) = \lim_{\delta \rightarrow \infty} \frac{\text{probability of having an event before time } t + \delta}{\delta}$$

This definition per Johnson & Shih (2007).

$$h(t) = h_0(t)e^{\beta_1 x_1 + \beta_2 x_2 + \text{etc.}}$$

¹`failvar` is often something like `died`.

We don't directly estimate the hazard, but estimate the effect of covariates on the hazard.

3.3 Estimate the Cox Proportional Hazards Model

```
use emperors2.dta, clear

stset agereign // stset the data

sts graph, by(riseNUMERIC) // survival curve by cause of death

graph export survival.png, width(1000) replace

stcox ib5.riseNUMERIC // Cox model w BIRTHRIGHT (5) as reference cat.
```

Survival-time data settings

```
Failure event: (assumed to fail at time=agereign)
Observed time interval: (0, agereign]
Exit on or before: failure
```

```
-----
      66 total observations
       5 event time missing (agereign>=.)          PROBABLE ERROR
-----
      61 observations remaining, representing
      61 failures in single-record/single-failure data
    2,470 total analysis time at risk and under observation
                                     At risk from t =          0
                               Earliest observed entry t =      0
                                     Last observed exit t =      79
```

```
Failure _d: 1 (meaning all fail)
Analysis time _t: agereign
```

file

```
/Users/agrogan/Desktop/GitHub/newstuff/categorical/survival-analysis-and-event-hi
> story/emperors2/survival.png saved as PNG format
```

```
Failure _d: 1 (meaning all fail)
```

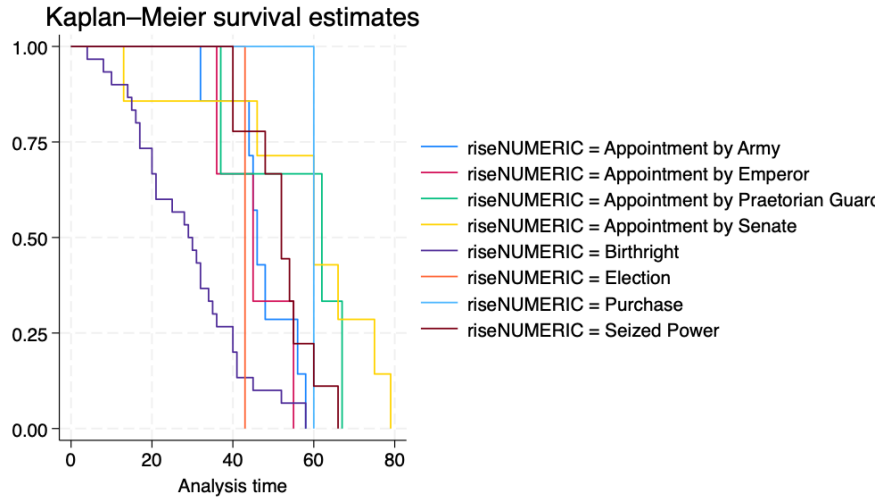



Figure 1: Survival Plot

```
stphplot, by(riseNUMERIC) scheme(michigan) legend(pos(6)) // graphical test of PH assumption
graph export ph.png, width(1000) replace
```

Test of proportional-hazards assumption

Time function: Analysis time

	rho	chi2	df	Prob>chi2
1.riseNUME~C	0.22588	2.76	1	0.0969
2.riseNUME~C	0.15586	1.37	1	0.2414
3.riseNUME~C	0.03204	0.05	1	0.8209
4.riseNUME~C	-0.04249	0.11	1	0.7352
5b.riseNUM~C	.	.	1	.
6.riseNUME~C	0.11688	0.78	1	0.3765
7.riseNUME~C	0.09609	0.47	1	0.4944
8.riseNUME~C	0.16179	1.47	1	0.2251
Global test		5.91	7	0.5504

Failure _d: 1 (meaning all fail)

```
Analysis time _t: agereign
```

```
file
```

```
/Users/agrogan/Desktop/GitHub/newstuff/categorical/survival-analysis-and-event-hi  
> story/emperors2/ph.png saved as PNG format
```

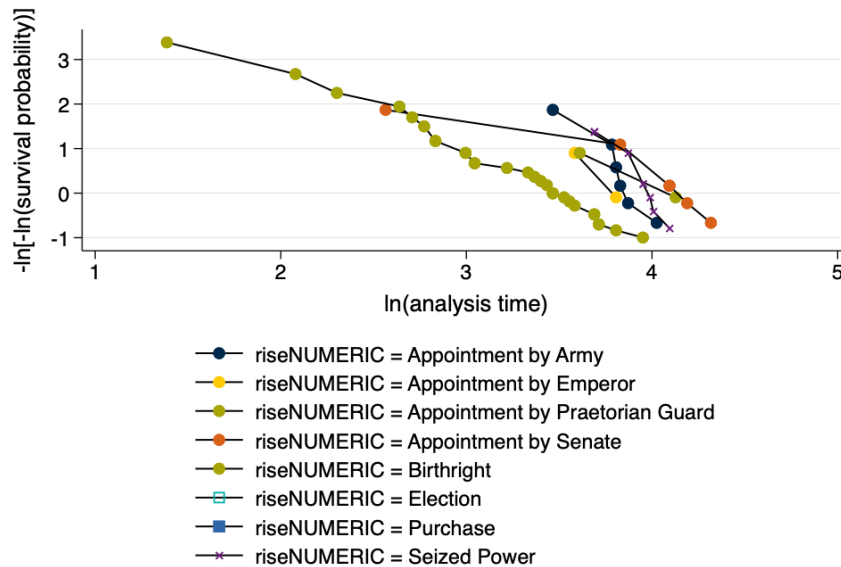


Figure 2: Graphical Assessment of Proportional Hazards Assumptions

3.5 Correcting For Violations of the Proportional Hazards Assumption

Had the proportional hazards assumption been violated, we could correct for this violation in one of two ways:

1. Estimating an interaction of the time variable (in this case `age`) with the variable violating the assumption.

e.g. `stcox age#ib5.riseNUMERIC`.

Note: In this relatively small sample this command fails to converge, perhaps because of sample size; or perhaps because there is no underlying violation of the proportional hazards assumption.

2. Using the `, strata(varname)` option to *stratify* on the variable violating the assumption.

Note that the command below provides results, but does not provide parameter estimates for the variable on which we are stratifying, `riseNUMERIC`.

```
stcox, strata(riseNUMERIC)
```

3.6 Plot Curve Accounting for Covariates

```
use emperors2.dta, clear

quietly: stset agereign // stset the data

quietly: stcox ib5.riseNUMERIC // Cox model

predict s, basesurv // predict baseline survival curve, accounting for covariates

summarize s

line s _t, sort c(J) // graph of survival curve

graph export survival2.png, width(1000) replace
```

(5 missing values generated)

Variable	Obs	Mean	Std. dev.	Min	Max
s	61	.3165069	.3181272	0	.9734189

```
file
```

```
/Users/agrogan/Desktop/GitHub/newstuff/categorical/survival-analysis-and-event-hi  
> story/emperors2/survival2.png saved as PNG format
```

4 References

Johnson, L. L., & Shih, J. H. (2007). CHAPTER 20 - An Introduction to Survival Analysis (J. I. Gallin & F. P. Ognibene, eds.). <https://doi.org/https://doi.org/10.1016/B978-012369440-9/50024-4>

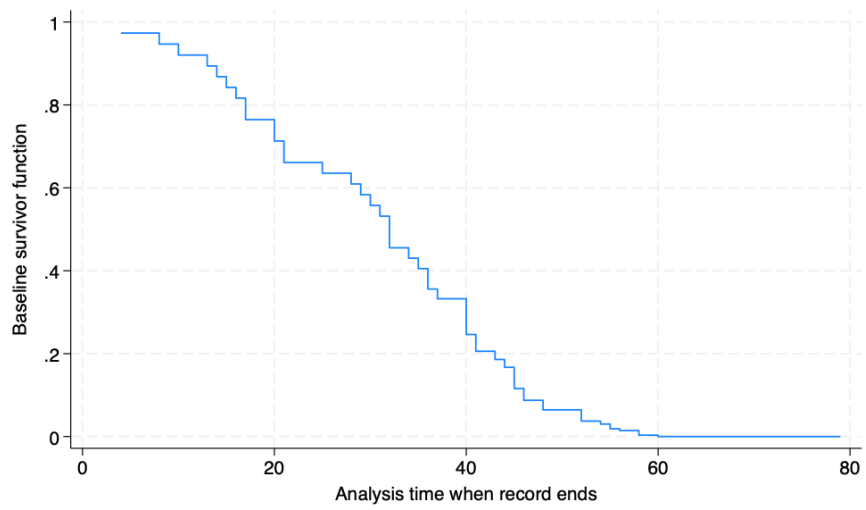


Figure 3: Survival Curve Accounting for Covariates