GSERM - 2019
Longitudinal Data Analysis

Exercise Two

Introduction and Data

The purpose of this exercise is to estimate and interpret parametric, Cox proportional hazards, and discrete-time models, and to compare the conclusions one might draw from them. The subject is leadership tenure; the data (available in CSV format) are drawn from the *Archigos* project, directed by Kristian Gleditsch, Hein Goemans, and Giacomo Chiozza (see here for details). They consist of data on 2,990 world leaders who served between 1875 and 2003, inclusive (NT = 15,244), and include time-varying information on how long the leader was in office – tenurestart and tenureend – and an indicator of when s/he left office (leftoffice).*

We're not going to get too complicated here; the relevant covariates are:

- age of the leader, in years,

- female, which means exactly what you think it does, and

- five indicator variables for region: LatinAmerica, Europe, Africa, MidEast, and Asia (with North America omitted as the implicit reference category).

Assignment

1. Begin by fitting one or more parametric model(s) of leader survival, using some combination of the seven covariates provided. Use your own judgement about (e.g.) if and how to specify interactions and other such things. Talk briefly about the results you obtain, being sure to include a discussion of the substantive importance of each covariate.

2. Reestimate the model in (1) using Cox's proportional hazards model. Again, use what you have learned in class and the readings to make informed, justifiable choices about what sort of model to estimate. Once again, briefly discuss your findings, this time with explicit attention to any differences between the parametric and Cox model results.

This assignment is due Tuesday, June 25, 2019 at 6:00 p.m. CSET. Exercises should be submitted *electronically*, in *PDF format*, by e-mailing them to Zorn (zorn@psu.edu). In addition to your responses to the items above, please include all code used to fit models, conduct diagnostics, generate plots, and so forth. This exercise is worth 100 points.

---

*For now, we'll put aside the important and fascinating question of the various means by which leaders can leave office, and treat all departures as equivalent to one another. Also, note that in this version of the data, leadership duration is measured in years. Those data were created by rounding day-level data, with some slightly-pernicious results. To create the relevant survival object in R , use:

```
Lead.S <- with(df, Surv(tenurestart,tenureend,leftoffice))
```

In Stata, use:

```
.  stset tenureend, id(obsid) failure(leftoffice)
```

Note that either or both of these will throw errors; you can ignore those errors for purposes of completing the assignment.