BIOST P8110: Applied Regression II
Lecture Note 9 - PROC PHREG (Part II)

Qixuan Chen

Department of Biostatistics

Columbia University

Spring 2017

This lecture's big ideas - how to use PROC PHREG in SAS 9.2-9.4 to:

1. Fit Cox model with time-dependent covariates
2. Fit Cox model with nonproportional hazards
   a. Testing the proportionality assumption
   b. Interactions with time as time-dependent covariates
   c. Nonproportionality via stratification

Section I: Time-dependent covariates

NAME:
STAN: Standford Heart Transplant Patients (Crowley and Hu; 1977)

SIZE:
103 observations

SOURCE:
The sample consisted of 103 cardiac patients who were enrolled in the transplantation program between 1967 and 1974. After enrollment, patients waited varying lengths of time until a suitable donor heart was found. Patients were followed until death or until the termination date of April 1, 1974. Of the 69 transplant recipients, only 24 were still alive at termination.

LIST OF VARIABLES:

Variables	Name	Description
1	DOB	is the date of birth
2	DOA	is the date of acceptance into the program
3	DOT	is the date of transplant
4	DLS	is the date last seen (death date or censoring date)
5	DEAD	status at last seen (1=dead; 0=otherwise)
6	SURG	had open-heart surgery before DOA (1=yes; 0=no)

1. Read in the data set and create variables needed for analysis

```
/*Read in the data set*/
libname lecture "C:\9_PROC_PHREG";
data stan; set lecture.stan;
run;

/*Create the variables needed for analysis*/
data stan; set stan;
surv1 = dls - doa;
ageaccpt = (doa - dob)/365.25;
wait = dot - doa;
if dot = . then trans = 0;
else trans = 1;
keep surv1 dead wait surg ageaccpt trans;
run;
```

2. Fit a Cox model for "trans" as a time-invariant covariate

```
/*Fit a cox model treat "trans" as a time-invariant variable*/
proc phreg data=stan;
model surv1*dead(0) = trans surg ageaccpt / ties=efron;
title "Cox model 1: with 'trans' as a time-invariant variable";
run;
```

```
                Cox model 1: with 'trans' as a time-invariant variable

                              The PHREG Procedure

                      Analysis of Maximum Likelihood Estimates

                    Parameter      Standard                                Hazard
    Parameter    DF    Estimate       Error    Chi-Square    Pr > ChiSq     Ratio

    trans         1    -1.44903     0.26363      30.2101        <.0001       0.235
    surg          1    -0.00280     0.39767       0.0000        0.9944       0.997
    ageaccpt      1    -0.01203     0.00317      14.3969        0.0001       0.988
```

3. Fit a Cox model for "trans" as a time-dependent covariate

1) Time-dependent covariates

➢ Time-dependent covariates are those that may change in value over the course of observation.

➢ We can modify a Cox model to include time-dependent covariates. For example, if we treat "trans" a time-dependent covariate, we have

$$h_i(t) = h_0(t)\exp\{\beta_1 surg_i + \beta_2 ageaccpt_i + \beta_3 trans_i(t)\}.$$

This says that the hazard at time t depends on the value of SURG and AGEACCPT as well as the value of TRANS at time t.

2) Why the finding in the Cox model 1 is misleading?

- In model 1, we attempted to determine whether a transplant raised or lowered the risk of death by examining the effect of a time-invariant covariate TRANS that was equal to 1 if the patient ever had a transplant and 0 otherwise. The reason that the finding was misleading is that who died quickly after acceptance into the program had less time available to get transplants.

3) How to handle the time-dependent covariate in PROC PHREG?

- SAS syntax

```
/*Fit a cox model treat "trans" as a time-dependent variable*/
proc phreg data=stan;
model surv1*dead(0) = plant surg ageaccpt / ties=efron;
if wait >= surv1 or wait=. then plant = 0; else plant = 1;
title "Cox model 2: with 'trans' as a time-dependent variable";
run;
```

```
              Cox model 2: with 'trans' as a time-dependent variable

                             The PHREG Procedure

                      Analysis of Maximum Likelihood Estimates

                      Parameter      Standard                                  Hazard
     Parameter    DF     Estimate        Error    Chi-Square    Pr > ChiSq      Ratio

     plant         1      0.06426      0.30319        0.0449        0.8321       1.066
     surg          1     -0.47084      0.37834        1.5488        0.2133       0.624
     ageaccpt      1     -0.00848      0.00292        8.4180        0.0037       0.992
```

- The IF statement defines the new time-varying covariate PLANT. Note that programming statements must follow the MODEL statement. Unlike an IF statement in the DATA step, which only operates on a single case at a time, this IF statement compares waiting times for patients who were at risk of a death with survival times for patients who experienced events. Thus, SURV1 in this statement is typically not the patient's own survival time, but the survival time of other patients who died.
- Model 1 shows that the hazard for those who received a transplant is lower than that for those who did not. While Model 2 shows transplantation has no effect on the hazard of death.

4) For more examples of time-dependent covariates, see page 153-172 (Allison 2010).

Section II: Cox models with nonproportional hazards

NAME:
RECID: Arrest times for released prisoners (Rossi, Berk, and Lenihan; 1980)

SIZE:
432 observations

SOURCE:
The RECID dataset contains information about 432 inmates who were released from Maryland state prisons in the early 1970s. The aim of this study was to determine the efficacy of financial aid to released inmates as a means of reducing recidivisms. Half of the inmates were randomly assigned to receive financial aid. They were followed for 1 year after their release and were interviewed monthly during that period.

LIST OF VARIABLES:

Variables	Name	Description
1	WEEK	is the week of first arrest
2	ARREST	has a value of 1 if arrested; 0 otherwise
3	FIN	has a value of 1 if received financial aid; 0 otherwise
4	AGE	is the age in years at the time of release
5	RACE	has a value of 1 if black; 0 otherwise
6	WEXP	has a value of 1 if had full-time work experience before incarceration; 0 otherwise
7	MAR	has a value of 1 if was married at the time of release; 0 otherwise
8	PARO	has a value of 1 if release on parole; 0 otherwise
9	PRIO	is the number of convictions before the current incarceration

---

1. **Background**

   ➢ Whenever you introduce time-dependent covariates into a Cox model, it is no longer accurate to call it a proportional hazards (PH) model. – Because the time-dependent covariate will change at different rates for different individuals, so the ratios of their hazards cannot remain constant.
   ➢ But suppose you don't have any time-dependent covariates. How do you know whether your data satisfy the PH assumption, and what happens if the assumption is violated?
   ➢ The PH assumption assumes that the effect of each variable is the same at all points in time. If the effect of a covariate varies with time, the PH assumption is violated for that variable.
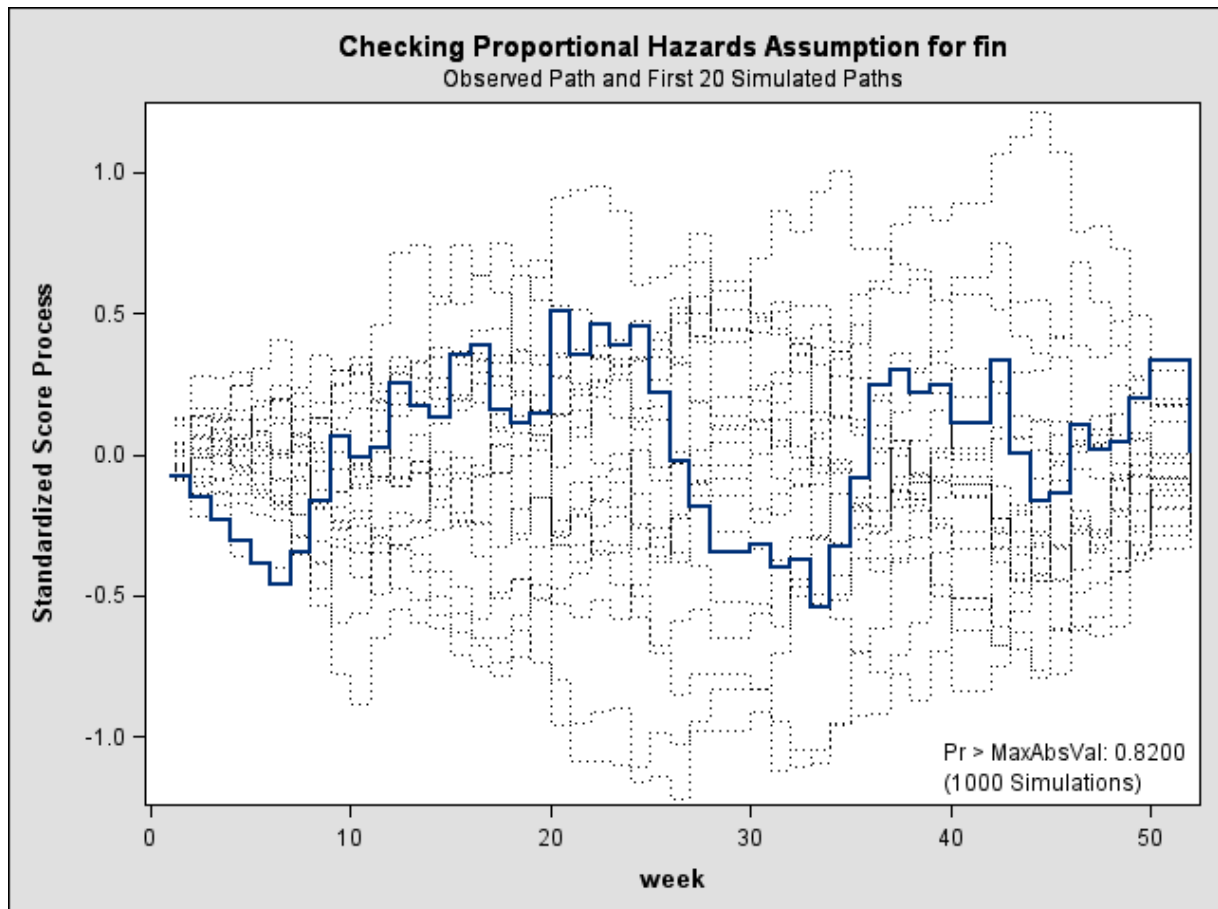
➢ If we estimate a PH model when the assumption is violated for some variable, then the coefficient that we estimate for that covariate is a sort of average effect over the range of times observed in the data.
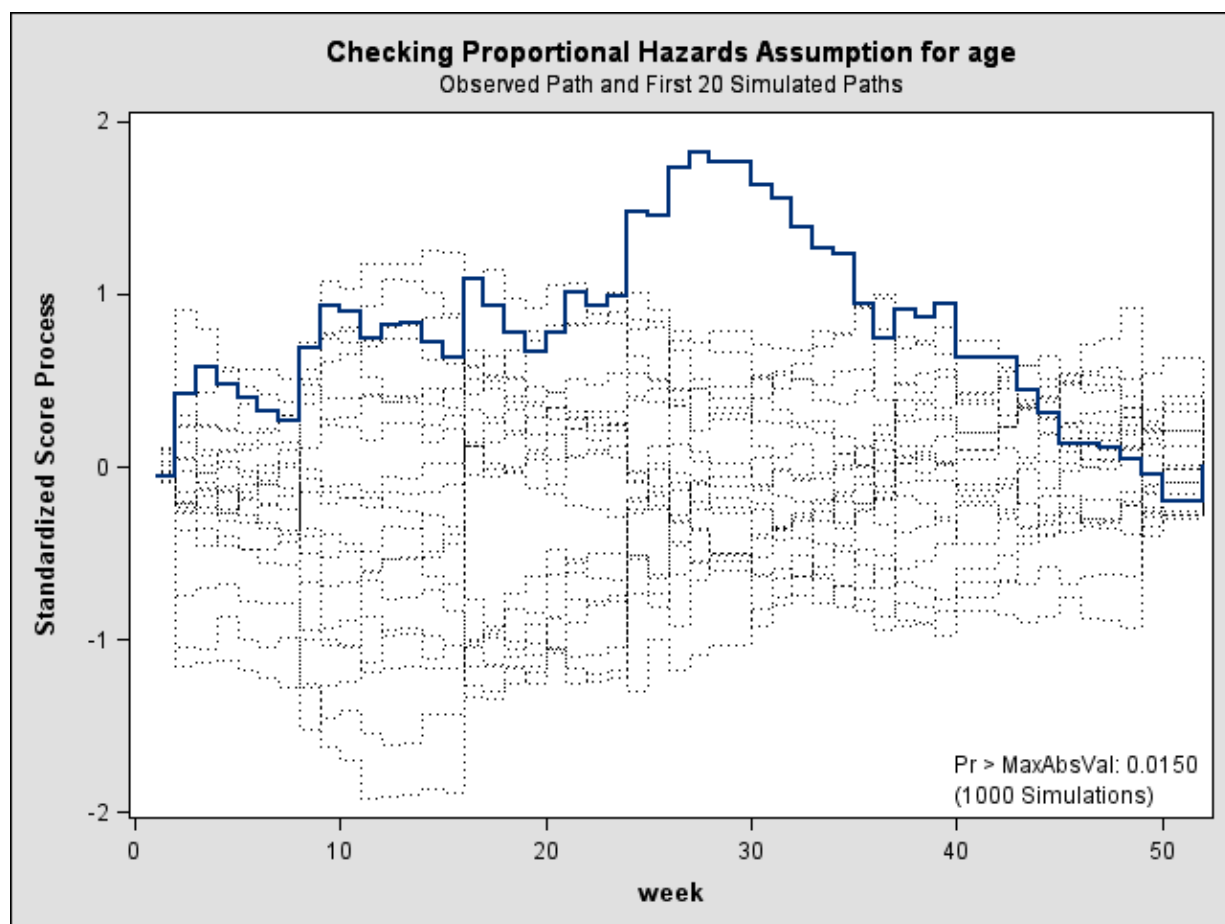
2. Testing the PH assumption with the ASSESS statement

1) SAS syntax

```
/*Read in the data set*/
data recid; set lecture.recid;
run;

/*Test the PH assumption*/
ods graphics on;
proc phreg data=recid;
model week*arrest(0) = fin age race wexp mar paro prio / ties = efron;
assess PH / resample;
run;
ods graphics off;
```

2) SAS outputs

Checking Proportional Hazards Assumption for age
Observed Path and First 20 Simulated Paths

Pr > MaxAbsVal: 0.0150
(1000 Simulations)

Supremum Test for Proportionals Hazards Assumption				
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
fin	0.5408	1000	437159001	0.8200
age	1.8192	1000	437159001	0.0150
race	0.9435	1000	437159001	0.2250
wexp	1.3008	1000	437159001	0.0880
mar	0.9349	1000	437159001	0.2380
paro	0.5383	1000	437159001	0.8280
prio	0.6104	1000	437159001	0.7480

3) Interpretations

➢ In PROC PHREG, martingale residuals are used to test for nonproportionality, and has been incorporated into the ASSESS statement.

➢ For each covariate, the ASSESS statement produces a graphical display of the empirical score process, which is based on the martingale residuals. The solid line is the observed empirical score process. The dashed lines are empirical score processes based on 20 random simulations that embody the PH assumption. If the observed process deviates markedly from the simulated processes, it is evidence against the PH assumption.

➢ In the lower right corner of graphical output, we get a more quantitative assessment in the form of p-value. For age, among 1000 simulated paths, only 1.5 percent of them have extreme points that exceeded the most extreme point of the observed path. The p-value was produced by the RESAMPLE option.

➢ Note that if you only want to have the table without graphics, simply not include the ODS GRAPHICS statements.

➢ The ASSESS statement cannot be used when there are time-dependent covariates.

3. Two methods to extend the Cox model to allow for nonproportional hazards

1) Interactions with time as time-dependent covariates

➢ To represent the interaction between a covariate and time in a Cox model, we can write
$$h_i(t) = h_0(t)\exp\{\beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i2}t\} \text{ or } h_i(t) = h_0(t)\exp\{\beta_1 X_{i1} + (\beta_2 + \beta_3 t)X_{i2}\}$$

➢ If β_3 is positive, the effect of X_2 increases linearly with time. β_2 can be interpreted as the effect of X_2 at time 0.

➢ The interaction term can be defined as a time-dependent covariate. The SAS syntax is

```
/*Add time and age interaction term for nonproportionality */
proc phreg data=recid;
model week*arrest(0) = fin age race wexp mar paro prio ageweek/ ties = efron;
ageweek = age*week;
title "Cox model with age*week interaction";
run;
```

```
                Cox model with age*week interaction

                       The PHREG Procedure

                Analysis of Maximum Likelihood Estimates
```

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
fin	1	-0.37823	0.19129	3.9096	0.0480	0.685
age	1	0.03692	0.03917	0.8883	0.3459	1.038
race	1	0.32290	0.30804	1.0989	0.2945	1.381
wexp	1	-0.12224	0.21285	0.3298	0.5658	0.885
mar	1	-0.41162	0.38212	1.1604	0.2814	0.663
paro	1	-0.09293	0.19583	0.2252	0.6351	0.911
prio	1	0.09354	0.02869	10.6294	0.0011	1.098
ageweek	1	-0.00369	0.00146	6.4241	0.0113	0.996

- ➤ The interaction term is significant, which confirms what we found using the ASSESS statement.
- ➤ For any suspected covariate, simply add to the model a time-dependent covariate representing the interaction of the original covariate and time. If the interaction covariate does not have a significant coefficient, then we may conclude that the PH assumption is not violated for that variable.

2) **Nonproportionality via stratification**

- ➤ Another approach to nonproportionality is stratification, a technique that is most useful when the covariate that interacts with time is both categorical and not of direct interest.
- ➤ EXAMPLE:

NAME:
MYEL: Myelomatosis patients (Peto et al. 1977)

SIZE:
25 observations

SOURCE:
The MYEL data set contains survival times for 25 patients diagnosed with myelomatosis. The patients were randomly assigned to two drug treatments.

LIST OF VARIABLES:

Variables	Name	Description
1	DUR	the time in days from the point of randomization to either death or censoring
2	STATUS	has a value of 1 if death; 0 censoring
3	TREAT	a value of 1 or 2 for two treatments
4	RENAL	has a value of 1 if renal function was normal or 0 for impaired functioning at the time of randomization

---

- ➤ SAS syntax:

```
data myel; set lecture.myel;
run;

/*Test for intercation term*/
proc phreg data=myel;
model dur*status(0) = treat renal renaltime/ ties = exact;
renaltime = renal*dur;
title "Interaction model";
run;

/*Stratify by renal*/
proc phreg data=myel;
model dur*status(0) = treat/ ties = exact;
strata renal;
title "stratification";
run;
```

```
                        Interaction model

                Analysis of Maximum Likelihood Estimates

                   Parameter      Standard                              Hazard
      Parameter    DF   Estimate      Error    Chi-Square   Pr > ChiSq    Ratio

      treat         1    1.45713     0.65974      4.8781       0.0272      4.294
      renal         1  -23.35615        3299      0.0001       0.9944      0.000
      renaltime     1    3.05883   412.43077      0.0001       0.9941     21.303


                          stratification

                Analysis of Maximum Likelihood Estimates

                   Parameter      Standard                              Hazard
      Parameter    DF   Estimate      Error    Chi-Square   Pr > ChiSq    Ratio

      treat         1    1.46398     0.65965      4.9254       0.0265      4.323
```

➢ The statistical models:

Model 1 for impaired: $h_i(t) = h_{01}(t)\exp\{\beta_1 treat_i\}$

Model 2 for not impaired: $h_i(t) = h_{00}(t)\exp\{\beta_1 treat_i\}$

Or combine the two models as $h_i(t) = h_{renal}(t)\exp\{\beta_1 treat_i\}$

➢ The stratification has been implemented in PROC PHREG using STRATA statement.

3) Compare interaction and stratification methods

➢ The interaction method requires choosing a particular form for the interaction, but stratification allows for any change in the effect of a covariate over time.

➢ Stratification takes less computing time. This can be important in working with large sample.

➢ No estimates are obtained for the effect of the stratifying variable. As a result, stratification only makes sense for nuisance variables whose effects have little or no interest.

➢ If the form of the interaction with time is correctly specified, the explicit interaction method should yield more efficient estimates of the coefficients of the other covariates.

➢ It is a trade-off between robustness and efficiency.