Survival Models in SAS

Part 7: PROC PHREG – Part 2

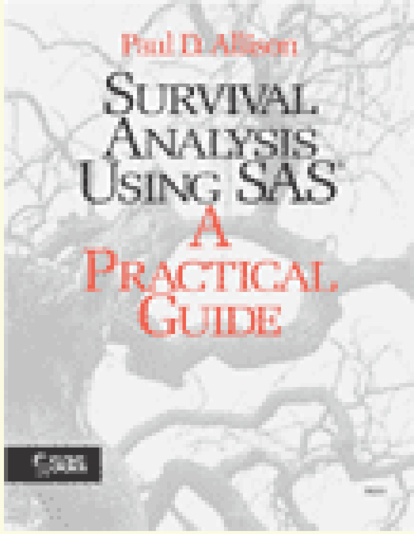
May 21, 2008

Charlie Hallahan

Chapter 5: Estimating Cox Regression Models with PROC PHREG

These talks are based on the book “**Survival Analysis Using the SAS System: A Practical Guide**” (1995) by Paul Allison.

The book is part of the SAS Books-by-Users series and can be found at http://www.sas.com/apps/pubscat/bookdetails.jsp?catid=1&pc=55233



Chapter 5: Estimating Cox Regression Models with PROC PHREG

This series of talks will cover

Chapter 1: Introduction

Chapter 2: Basic Concepts of Survival Analysis

Chapter 3: Estimating and Comparing Survival Curves with PROC LIFETEST

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG

Chapter 5: Estimating Cox Regression Models with PROC PHREG

Chapter 6: Competing Risks

Chapter 5: Estimating Cox Regression Models with PROC PHREG

Topics in Chapter 5:

Introduction
The Proportional Hazards Model
Partial Likelihood
Tied Data
Time-Dependent Covariates
Cox Models with Nonproportional Hazards
Interactions with Time as Time-Dependent Covariates
Nonproportionality via Stratification
Left Truncation and Late Entry into the Risk Set
Estimating Survivor Functions
Residuals and Influence Statistics
Testing Linear Hypotheses with the TEST Statement
Conclusion

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data

Recall the expression for the **partial likelihood** for data with **fixed covariates** from a **proportional hazards** model is:

$$PL = \prod_{i=1}^{n} \left[\frac{e^{\boldsymbol{\beta}\mathbf{x}_i}}{\sum_{j=1}^{n} Y_{ij} e^{\boldsymbol{\beta}\mathbf{x}_j}} \right]^{\delta_i} = \prod_{j=1}^{J} L_j$$

where $Y_{ij} = 1$ if $t_j \geq t_i$ and $Y_{ij} = 0$ if $t_j < t_i$. Note that even though this product is taken over all patients, the censored observations i are essentially ignored since $\delta_i = 0$.

This formula for PL is invalid if there are tied data, which is quite common.

A common technique with tied data is *Breslow's approximation* (the default in **SAS**).

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data

However, *Breslow's method* is appropriate only when there are few ties.

In general, a large proportion of the subjects are usually tied.

PHREG provides two other methods with the options **TIES = EXACT** or **TIES = DISCRETE**.

Both methods are actually exact methods. The difference is the assumption made about the ties occured.

EXACT assumes that there is a true but unknown ordering for the tied event times (i.e., *time* is continuous) while **DISCRETE** assumes that the events really occured at exactly the same time.

With the recidivism data, there was only one arrest during weeks 1 through 7. Thus, the calculations for L_1 through L_7 are straightforward.

However, in week 8, there five arrests. So, the calculation of L_8 must be altered.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data

week	Frequency	Percent	Frequency	Percent
1	1	0.23	1	0.23
2	1	0.23	2	0.46
3	1	0.23	3	0.69
4	1	0.23	4	0.93
5	1	0.23	5	1.16
6	1	0.23	6	1.39
7	1	0.23	7	1.62
8	5	1.16	12	2.78
9	2	0.46	14	3.24
10	1	0.23	15	3.47
11	2	0.46	17	3.94
12	2	0.46	19	4.40
13	1	0.23	20	4.63
14	3	0.69	23	5.32
15	2	0.46	25	5.79
16	2	0.46	27	6.25
17	3	0.69	30	6.94
18	3	0.69	33	7.64
19	2	0.46	35	8.10
20	5	1.16	40	9.26
21	2	0.46	42	9.72
22	1	0.23	43	9.95
23	1	0.23	44	10.19
24	4	0.93	48	11.11
25	3	0.69	51	11.81
26	3	0.69	54	12.50

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data

week	Frequency	Percent	Frequency	Percent
27	2	0.46	56	12.96
28	2	0.46	58	13.43
30	2	0.46	60	13.89
31	1	0.23	61	14.12
32	2	0.46	63	14.58
33	2	0.46	65	15.05
34	2	0.46	67	15.51
35	4	0.93	71	16.44
36	3	0.69	74	17.13
37	4	0.93	78	18.06
38	1	0.23	79	18.29
39	2	0.46	81	18.75
40	4	0.93	85	19.68
42	2	0.46	87	20.14
43	4	0.93	91	21.06
44	2	0.46	93	21.53
45	2	0.46	95	21.99
46	4	0.93	99	22.92
47	1	0.23	100	23.15
48	2	0.46	102	23.61
49	5	1.16	107	24.77
50	3	0.69	110	25.46
52	**322**	**74.54**	**432**	**100.00**

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Exact Method

The **EXACT** method assumes that the events recorded as simultaneous **really occurred at distinct times** (if we were able to measure more precisely).

Without knowing the correct order, we consider **all possible orderings**.

With five ties at time = 8, there are 5! = 120 **possible orderings**.

Denote the possibilities as A_i, $i = 1, \ldots , 120$.

We need the probability of the ***union*** of these events, i.e., $\Pr\{A_1 \text{ or } A_2 \text{ or } \ldots A_{120}\}$.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Exact Method

Thus, $L_8 = \sum_{i=1}^{120} \Pr\left(A_i\right).$

Each of these 120 probabilities $\Pr\left(A_i\right)$ is just a **standard partial likelihood**.

Suppose we (arbitrarily) label the five arrests at time 8 as 8, 9, 10, 11, 12.

Suppose A_1 denotes the ordering {8, 9, 10, 11, 12}.

Then $\Pr\left(A_1\right) = \left(\frac{e^{\beta x_8}}{e^{\beta x_8} + e^{\beta x_9} + \ldots + e^{\beta x_{432}}}\right)\left(\frac{e^{\beta x_9}}{e^{\beta x_9} + e^{\beta x_{10}} + \ldots + e^{\beta x_{432}}}\right)\cdots\left(\frac{e^{\beta x_{12}}}{e^{\beta x_{12}} + e^{\beta x_{13}} + \ldots + e^{\beta x_{432}}}\right)$

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Exact Method

If A_2 denotes the ordering {9, 8, 10, 11, 12}.

$$\text{Then } \Pr(A_2) = \left(\frac{e^{\beta x_9}}{e^{\beta x_8} + e^{\beta x_9} + \ldots + e^{\beta x_{120}}}\right)\left(\frac{e^{\beta x_8}}{e^{\beta x_8} + e^{\beta x_{10}} + \ldots + e^{\beta x_{120}}}\right)\cdots\left(\frac{e^{\beta x_{12}}}{e^{\beta x_{12}} + e^{\beta x_{13}} + \ldots + e^{\beta x_{120}}}\right)$$

Continue this way for the other 118 possible orederings.

Then L_8 is just the sum of these 120 probabilities.

The situation is simpler for week 9 because there are only two arrests (yielding 2! = 2 possible orderings. For L_9 we would have

$$L_9 = \left(\frac{e^{\beta x_{13}}}{e^{\beta x_{13}} + e^{\beta x_{14}} + \ldots + e^{\beta x_{432}}}\right)\left(\frac{e^{\beta x_{14}}}{e^{\beta x_{14}} + e^{\beta x_{15}} + \ldots + e^{\beta x_{432}}}\right) + \left(\frac{e^{\beta x_{14}}}{e^{\beta x_{13}} + e^{\beta x_{14}} + \ldots + e^{\beta x_{432}}}\right)\left(\frac{e^{\beta x_{13}}}{e^{\beta x_{13}} + e^{\beta x_{15}} + \ldots + e^{\beta x_{432}}}\right)$$

For week 10 (with only one arrest), we're back to the standard partial likelihood formula:

$$L_{10} = \left(\frac{e^{\beta x_{15}}}{e^{\beta x_{15}} + e^{\beta x_{16}} + \ldots + e^{\beta x_{432}}}\right)$$

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Exact Method

Note that if 10 events occur at the same time, then there are over 3 million possible orderings to evaluate.

Needless to say, the number of possible calculations needed to implement the exact method when there are many ties renders the method impractical.

However, statisticians have developed equivalent formulations involving the evaluation of a definite integral. This makes the method plausible.

The most popular approximation to the **EXACT** method is due to **Breslow** (1974). This is the default in **PHREG (TIES=BRESLOW).**

Efron (1977) proposed an alternative approximation (**TIES=EFRON**).

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Exact Method

The default **BRESLOW** method.

```
proc phreg data=survival.recid;
   model week*arrest(0)=fin age race wexp mar paro prio;
run;
```

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
fin	1	-0.37902	0.19136	3.9228	0.0476	0.685
age	1	-0.05724	0.02198	6.7798	0.0092	0.944
race	1	0.31415	0.30802	1.0402	0.3078	1.369
wexp	1	-0.15113	0.21212	0.5076	0.4762	0.860
mar	1	-0.43280	0.38180	1.2850	0.2570	0.649
paro	1	-0.08497	0.19575	0.1884	0.6642	0.919
prio	1	0.09114	0.02863	10.1331	0.0015	1.095

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Exact Method

The **EXACT** method.

```
proc phreg data=survival.recid;
   model week*arrest(0)=fin age race wexp mar paro prio
               / ties=exact;
run;
```

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
fin	1	-0.37942	0.19138	3.9305	0.0474	0.684
age	1	-0.05743	0.02200	6.8152	0.0090	0.944
race	1	0.31393	0.30800	1.0389	0.3081	1.369
wexp	1	-0.14981	0.21223	0.4983	0.4803	0.861
mar	1	-0.43372	0.38187	1.2900	0.2560	0.648
paro	1	-0.08486	0.19576	0.1879	0.6646	0.919
prio	1	0.09152	0.02865	10.2067	0.0014	1.096

In this example, the Breslow approximation does a good job as its parameter estimates agree with those of the exact method out to at least two decimal places.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Exact Method

The **EFRON** method.

```
proc phreg data=survival.recid;
   model week*arrest(0)=fin age race wexp mar paro prio
                 / ties=efron;
run;
```

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
fin	1	-0.37942	0.19138	3.9304	0.0474	0.684
age	1	-0.05743	0.02200	6.8152	0.0090	0.944
race	1	0.31392	0.30799	1.0389	0.3081	1.369
wexp	1	-0.14981	0.21223	0.4983	0.4803	0.861
mar	1	-0.43372	0.38187	1.2900	0.2560	0.648
paro	1	-0.08486	0.19576	0.1879	0.6646	0.919
prio	1	0.09152	0.02865	10.2067	0.0014	1.096

The Efron approximation improves on the Breslow by agreeing with the exact method's parameter estimates out to four decimal places.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Exact Method

It has been shown that the **Breslow** method deteriorates as the number of ties at a particular point in time becomes a large proportion of the number of cases at risk.

For the recidivism data, the number of ties of survival times at any given time never exceeds 2 percent – thus the **Breslow** method works fine in this example.

The author reports on a simulation example (no code given, so can't be replicated) where the number of ties as a percent of those at risk ranges from 19% to 32%.

As a result, the **Breslow** approximation is quite different from the exact methods. The **Efron** approximation is acceptable for qualitative conclusions as to which covariates are significant, but the parameter estimates themselves are not that accurate.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Discrete Method

The **DISCRETE** option in **PHREG** is also an exact method.

It is based on a fundamentally different model than the **EXACT** method (which assumes that time is continuous and that ties only result for the imprecise measurement of time).

Events are assumed to occur at discrete times (for example, a change in the political party of the president can only occur every four years at a minimum; a default on a monthly mortgage payment can only occur on a monthly basis).

Cox introduced the discrete method in his 1972 paper and involves a form of his partial likelihood method.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Discrete Method

Assume that time t can only take on integer values.

Let P_{it} be the conditional probability that subject i has an event at time t, given that an event has not already occurred for that subject.

P_{it} is sometimes called a **discrete - time hazard**.

The model for P_{it} is then a logit-type regression: $\log\left(\frac{P_{it}}{1-P_{it}}\right)=\alpha_t+\beta_1 x_{i1}+\ldots+\beta_k x_{ik}$.

α_t represents a set of constants that vary with time (playing the same role as $\alpha(t)$ in the continuous model).

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Discrete Method

This model is sometimes called a **proportional odds model**.

Defining the odds that individual i has an event at time t (given that i has not already had the event) as $O_{it} = \dfrac{P_{it}}{1-P_{it}}$, then $O_{it} = \exp\left(\alpha_t + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}\right)$.

Thus, $\dfrac{O_{it}}{O_{jt}} = \dfrac{\exp\left(\alpha_t + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}\right)}{\exp\left(\alpha_t + \beta_1 x_{j1} + \ldots + \beta_k x_{jk}\right)} = \exp\left(\boldsymbol{\beta}\left(\mathbf{x}_i - \mathbf{x}_j\right)\right)$ does not depend on t.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Discrete Method

How to estimate this model?

Chapter 7 shows that standard maximum likelihood methods can be used by treating the model as a standard logit model.

To use the partial likelihood method of estimation, the α_ts are treated as nuisance parameters and only the βs are estimated.

Assuming there are J unique event times, the **partial likelihood function** will contain J terms:

$$PL = \prod_{j=1}^{J} L_j$$ where L_j is the partial likelihood for the jth event.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Discrete Method

The **job duration dataset** consists of 100 simulated job durations, measured from the year of entry into the job until the year that the employee quit.

Durations after the fifth year are censored.

If the employee was fired before the fifth year, the duration is censored at the end of the last full year in which the employee was working.

There are only possible survival times: 1, 2, 3, 4, or 5 years.

A simple life table for these data:

Duration	Number Quit	Number Censored	Number At Risk	Quit/ At Risk
1	22	7	100	0.22
2	18	3	71	0.25
3	16	4	50	0.32
4	8	1	30	0.27
5	4	17	21	0.19

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Discrete Method

Three covariates were measured at the beginning of the job: years of schooling (ED), salary in thousands of dollars (SALARY), and the prestige of the occupation (PRESTIGE) measured on a scale from 1 to 100.

The original dataset has a variable `event` coded as 1 if the worker quit, 0 if they were censored at week 5, and 2 if they were fired before week 5. A new variable, `event2`, was created with a value of 1 if they quit and 0 if they were censored.

First 10 observations for the job duration dataset:

Obs	dur	event	ed	prestige	salary	event2
1	1	1	7	3	19	1
2	4	1	14	62	17	1
3	5	0	16	70	18	0
4	2	1	12	43	135	1
5	3	1	9	18	12	1
6	1	1	11	31	12	1
7	1	1	13	26	6	1
8	1	1	10	1	4	1
9	2	1	12	28	17	1
10	3	1	7	25	4	1

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Discrete Method

Results from using the three methods of treating ties discussed so far:

```
title "Cox model for job duration data";
proc phreg data=survival.jobdur;
        model dur*event2(0) = ed prestige salary / ties = breslow;
run;
```

Substitute `ties = efron` and `ties = exact` for the other two methods.

Ties Handling: BRESLOW (took 0.03 seconds)

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
ed	1	0.11644	0.05918	3.8719	0.0491	1.123
prestige	1	-0.06427	0.00959	44.9329	<.0001	0.938
salary	1	-0.01495	0.00792	3.5648	0.0590	0.985

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Discrete Method

Ties Handling: EFRON (took 0.04 seconds)

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
ed	1	0.14403	0.05954	5.8513	0.0156	1.155
prestige	1	-0.07980	0.00996	64.1916	<.0001	0.923
salary	1	-0.02015	0.00830	5.9013	0.0151	0.980

Ties Handling: EXACT (took 0.03 seconds)

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
ed	1	0.16433	0.06380	6.6341	0.0100	1.179
prestige	1	-0.09202	0.01240	55.1094	<.0001	0.912
salary	1	-0.02254	0.00884	6.5048	0.0108	0.978

The parameter estimates for the Breslow method are about one-third smaller than the Exact method, while the p-values are substantially higher. Efron's method is in closer agreement with the Exact method, but there is still an appreciable loss of accuracy in estimating the magnitudes of the coefficients..

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Discrete Method

For the job duration data, there are only five terms in the partial likelihood function, but each of these five terms is colossal.

At time 1, there were 22 people with events out of 100 who were at risk.

To get L_1, we ask the question: given that 22 events occurred, what is the probability that they occurred to these particular 22 people rather to some different set of 22 people coming from the 100 people at risk?

How many ways are there of selecting 22 people from among a set of 100?

Answer: 7.3321×10^{21} !!!! Call this number Q and let $q = 1$ to Q.

Assume $q = 1$ corresponds to the 22 people who actually experienced the event at time 1.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Discrete Method

For a given q, let Ψ_q be the product of the odds for all the individuals is that set.

If the individuals who actually experienced events are labeled $i = 1$ to 22, then

$$\Psi_1 = \prod_{i=1}^{22} O_{i1} \quad \text{and} \quad L_1 = \frac{\Psi_1}{\Psi_1 + \Psi_2 + \ldots + \Psi_Q} \quad .$$

This "*simple*" expression has over a trillion terms being summed in the denominator.

Fortunately, there is a recursive algorithm making the calculation feasible.

Is the discrete-time assumption reasonable for the job duration data? In practice, someone can quit their job at any time (i.e., treat time as continuous).

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – The Discrete Method

```
title "Cox model for job duration data";
proc phreg data=survival.jobdur;
        model dur*event2(0) = ed prestige salary / ties = discrete;
run;
```

Ties Handling: DISCRETE   (took 0.09 seconds)

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
ed	1	0.21938	0.08480	6.6929	0.0097	1.245
prestige	1	-0.12047	0.01776	46.0222	<.0001	0.886
salary	1	-0.02611	0.01020	6.5560	0.0105	0.974

Compared with the **EXACT** method, the chi-square statistics are similar. However, the coefficients for **ED** and **PRESTIGE** are about one-third larger for the **DISCRETE** method.

This is largely due to the fact that that two different models are being estimated, a **hazard model** (EXACT) and a **logit model** (DISCRETE). The logit coefficients are generally larger. For the logit model, $100(e^{\beta}-1)$ gives the percent change in the ***odds*** resulting from a one-unit increase in the covariate. Thus, a one-year increase in schooling increases the odds of quitting a job by $100(e^{0.219}-1)$ =24%.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – Comparison of Methods

It has been shown that if ties result from grouping continuous time data into intervals, the logit model converges to the proportional hazards model as the interval length gets smaller (**Thompson 1977**).

When there are no ties, the partial likelihoods for all four methods (the two exact methods and the two approximations) reduce to the same formula, although **PHREG** is still slightly faster with the **Breslow** method.

Using simulated datasets ranging in size from 100 to 1200, Allison shows that the estimation time hardly changes for the Breslow and Efron methods, while the EXACT method increase by a factor of 50 and the DISCRETE method by a factor of 9.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – Comparison of Methods

Allison summarizes the handling of ties with six points:

1. When there are no ties, all four options in **PHREG** give identical results.

2. When there are few ties, it makes little difference which method is used. Since computing times are comparable, you might as well use one of the exact methods.

3. When the number of ties is large, relative to the number at risk, the approximate methods tend to yield estimates that are biased towards 0.

4. Both the **EXACT** and **DISCRETE** methods produce exact results (i.e., true partial likelihood estimates), but the **EXACT** method assumes that ties arise from grouping continuous, untied data, while the **DISCRETE** method assumes that evens really occur at the same time. The choice should be based on substantive grounds, although qualitative results will usually be similar.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Tied Data – Comparison of Methods

5. Both of the exact methods need a substantial amount of computer time for large data sets containing many ties. This is especially true for the **EXACT** method where doubling the sample size increases computing time by at least a factor of 5.

6. If the exact methods are too time-consuming, use the **Efron** approximation, at least for model exploration. It’s nearly always better than the **Breslow** methods, with virtually no increase in computer time.