Survival Models in SAS

Part 6: PROC PHREG – Part 1

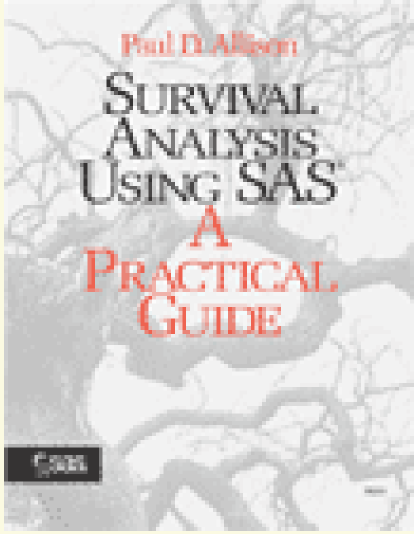
April 16, 2008

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: Introduction

The **proportional hazards regression model** was introduced by **David Cox** in a 1972 JRSS Series B paper.

This is one of the most cited papers in all of science.

The method is also called **Cox regression**.

Some properties of the Cox regression model:

1. A parametric assumption of the distribution of survival time is not necessary.
2. Time-dependent covariates are easily incorporated.
3. Stratified analysis is easily handled.
4. Adjustments for periods of time when the subject is not at risk can be made.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: The Proportional Hazards Model

The 1972 Cox paper proposed two innovations:

1. It introduced the **proportional hazards model** (even though the model can handle nonproportional hazards).

2. A new estimation method was derived, (maximum) **partial likelihood**.

Note: Some of the parametric models already introduced are also proportional hazards models, for example, the **Weibull** and **Gompertz** models.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: The Proportional Hazards Model

A basic model without time-varying covariates or nonproportional hazards is:

$$(1) \quad h_i(t) = \lambda_0(t)\exp\left(\beta_1 x_{i1} + \ldots + \beta_k x_{ik}\right)$$

$\lambda_0(t)$ is called the baseline hazard and is unspecified (except that $\lambda_0(t) \geq 0$).

Note that $\exp\left(\beta_1 x_{i1} + \ldots + \beta_k x_{ik}\right)$ guarantees that $h_i(t) \geq 0$.

Also, $h_i(t) = \lambda_0(t)$ whenever $x_{i1} = \ldots = x_{ik} = 0$.

Taking logs of both sides of (1) yields $\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$ where $\alpha(t) = \log \lambda_0(t)$.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: The Proportional Hazards Model

Special cases:

$\alpha(t) = \alpha t$ yields the **Gompertz** model.

$\alpha(t) = \alpha \log t$ yields the **Weibull** model.

For the **Cox model**, no assumptions are made for $\alpha(t)$.

Reason for the name **proportional hazards model**:

$$\frac{h_i(t)}{h_j(t)} = \exp\left(\beta_1(x_{i1} - x_{j1}) + \ldots + \beta_k(x_{ik} - x_{jk})\right)$$ **does not** depend on t.

Plots of the **hazard functions** over time for two subjects will be **parallel**.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood

Some properties of the **partial likelihood method**:

- The estimates of $\boldsymbol{\beta}$ do not depend on the baseline hazard $\lambda_0(t)$.
- The full likelihood function is factored into two parts: one part depends on both $\lambda_0(t)$ and $\boldsymbol{\beta}$ and the other part only depends on $\boldsymbol{\beta}$.
- The partial likelihood method ignores the first part and maximizes the second part.
- As a result the partial likelihood method is not fully efficient, but the loss in efficiency is small (Efron 1977) and it is still consistent and asymptotically normal.
- Partial likelihood estimates only depend on the **ranks** of the event times and not their actual values. Thus, any monotonic transformation of the event times does not affect the partial likelihood estimates.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

We’ll estimate a proportional hazards model using the **recidivism** data that was used for parametric models with **PROC LIFEREG**. The basic syntax for **PHREG** is the same as that for **LIFEREG**, except that a distribution is not specified.

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

```
                    The PHREG Procedure

                     Model Information

          Data Set                    SURVIVAL.RECID
          Dependent Variable          week
          Censoring Variable          arrest
          Censoring Value(s)          0
          Ties Handling               BRESLOW


          Number of Observations Read          432
          Number of Observations Used          432
```

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
432	114	318	73.61

Convergence Status

Convergence criterion (GCONV=1E-) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	1351.367	1318.241
AIC	1351.367	1332.241
SBC	1351.367	1351.395

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	33.1256	7	<.0001
Score	33.3828	7	<.0001
Wald	31.9875	7	<.0001

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

Analysis of Maximum Likelihood Estimates

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**

The partial likelihood method only uses the ranks of the event times in its calculations.

So there must be a way to deal with **tied events**. The default method for **PHREG** is the **Breslow** method. Three superior methods are discussed later.

Note that there is no intercept in the model. It is absorbed into the baseline hazard function $\alpha(t)$.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

The column labeled **Hazard Ratio** is e^{β}. It represents the relative change in the hazard function when the corresponding variable changes by one unit (controlling for all the other covariates).

So for a **dummy variable**, e^{β} represents the relative change in the hazard as the variable changes from 0 to 1.

For example, a hazard ratio of 0.685 for the dummy variable **`FIN`** means that the hazard of being arrested for those who received financial assistance is 69% of the hazard of those who did not receive financial assistance.

For quantitative covariates, a more useful calculation is $100(e^{\beta} - 1)$. This represents the percent change in the hazard as that covariate increase by one unit.

For example, a hazard ratio of 0.944 for **`AGE`** means that for each year increase in age, the hazard of being arrested decreases by 5.6%.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

The substantive conclusions from the Cox model are similar to those from the parametric model estimated by **LIFEREG**.

Namely, **`AGE`** and **`PRIO`** are highly significant and **`FIN`** is just significant at the 5% level.

Note that while the magnitudes and p-values for the two specifications are very similar, the signs are reversed. This is because **LIFEREG** estimates the model in *log-survival* time, while **PHREG** estimates a model in *log-hazard* format.

Note that only the parametric models that are also proportional hazard models (**exponential, Weibull, Gompertz**) can be interpreted in log-hazard format and compared to a **Cox** model. Distributions such as **gamma, log-logistic**, and **log-normal** do not produce proportional hazard models, and a comparison with a **Cox** model is not appropriate.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

The next example is a little more complicated and involves the famous **Stanford Heart Transplant Data** (Crowley and Hu, 1977).

The sample consists of 103 cardiac patients enrolled in the transplantation program between 1967 and 1974.

After enrollment in the program, patients waited varying lengths of time until a suitable donor heart was found.

Thirty patients died before receiving a transplant, while another four patients had still not received transplants at the termination date of April 1, 1974.

Patients were followed until death or until the termination date.

Of the 69 transplant recipients, only 24 were still alive at termination.

At the time of transplantation, all but four of the patients were tissue typed to determine the degree of similarity with the donor.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

The **input variables** are:

DOB	date of birth
DOA	date of acceptance into the program
DOT	date of transplant
DLS	date last seen (dead or censored)
DEAD	coded 1 if dead at DLS; otherwise coded as 0
SURG	coded 1 if patient had open-heart surgery prior to DOA; otherwise coded 0
M1	number of donor alleles with no match in recipient (1 through 4)
M2	1 if donor-recipient mismatch on HLA-A2 antigen; otherwise 0
M3	mismatch score

The variables DOT, M1, M2, and M3 are coded as missing for those patients who did not receive a transplant.
All four date measures are coded in the form *mm/dd/yy*.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

```
options yearcutoff = 1900;
data survival.stan;
      input dob mmddyy9. doa mmddyy9. dot mmddyy9. dls mmddyy9.
         id age dead dur surg trans wtime m1 m2 m3 reject;
      format dob doa dot mmddyy9.;
      surv1=dls-doa;
      surv2=dls-dot;
      wait=dot-doa;
      agetrans=(dot-dob)/365.25;
      ageaccpt=(doa-dob)/365.25;
      if wait = . then wait = 10000;
      agels=(dls-dob)/365.25;
cards;
01/10/37 11/15/67     .    01/03/68 1 30 1 50 0 0 . . . . .
Etc.
05/20/28 09/13/67     .    09/18/67 103 39 1 6 0 0 . . . . .
;
```

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

1st 10 observations in the Stanford Heart Transplant Data

dob	doa	dot	dls	dead	surg	m1	m2	m3
01/10/37	11/15/67	.	2924	1	0	.	.	.
03/02/16	01/02/68	.	2928	1	0	.	.	.
09/19/13	01/06/68	01/06/68	2942	1	0	2	0	1.11
12/23/27	03/28/68	05/02/68	3047	1	0	3	0	1.66
07/28/47	05/10/68	.	3069	1	0	.	.	.
11/08/13	06/13/68	.	3088	1	0	.	.	.
08/29/17	07/12/68	08/31/68	3789	1	0	4	0	1.32
03/27/23	08/01/68	.	3174	1	0	.	.	.
06/11/21	08/09/68	.	3227	1	0	.	.	.
02/09/26	08/11/68	08/22/68	3202	1	0	2	0	0.61

Several additional variables are also created:

```
surv1=dls-doa;   * days from acceptance until death;
surv2=dls-dot;   * days from transplant until death;
wait=dot-doa;    * days from acceptance until transplant;
agetrans=(dot-dob)/365.25;  * age at transplant;
ageaccpt=(doa-dob)/365.25;  * age at acceptance;
if wait = . then wait = 10000;
agels=(dls-dob)/365.25;     * age at death;
```

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

Question 1: Did transplantation decrease the hazard of death?

Approach: Cox regression of **SURV1** on transplant status (**TRANS**) controlling for **AGEACCPT** and **SURG**.

```
title "1st Cox Model for Stanford Heart Transplant Data";
proc phreg data=survival.stan;
      model surv1*dead(0) = trans surg ageaccpt;
run;
```

```
                Model Information
Data Set                      SURVIVAL.STAN
Dependent Variable            surv1
Censoring Variable            dead
Censoring Value(s)            0
Ties Handling                 BRESLOW
Number of Observations Read           103
Number of Observations Used           103
```

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
103	75	28	27.18

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	596.651	551.188
AIC	596.651	557.188
SBC	596.651	564.141

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	45.4629	3	<.0001
Score	52.0469	3	<.0001
Wald	46.6680	3	<.0001

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
trans	**1**	**-1.70813**	**0.27860**	**37.5902**	**<.0001**	**0.181**
surg	**1**	**-0.42130**	**0.37098**	**1.2896**	**0.2561**	**0.656**
ageaccpt	**1**	**0.05860**	**0.01505**	**15.1611**	**<.0001**	**1.060**

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

The results show significant effects of both transplant status and age of acceptance.

Each additional year of age at the time of acceptance into the program leads to a 6 percent increase in the hazard of death.

The hazard for those who received a transplant is only about 18 percent of the hazard of those who did not. Or equivalently (taking the reciprocal), those who did *not* receive a transplant are about 5 ½ times more likely to die at any given point in time.

However, the main reason why patients did not get transplants is that they died before a suitable donor could be found – thus the death rates are higher. The covariate is actually a *consequence* of the dependent variable: an early death prevents a patient from getting a transplant.

One solution is to treat transplant status as a *time-dependent covariate* (to be covered later).

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

Question 2: Of those patients who did receive a transplant, why did some survive longer than others?

Approach: Cox regression of **SURV2** on covariates M1, M2, M3, AGETRANS, WAIT and DOT for just the transplant patients.

```
title "Cox Model for just those receiving transplants";
proc phreg data=survival.stan;
   where trans=1;
   model surv2*dead(0)=surg m1 m2 m3 agetrans wait dot;
run;
```

Note that the origin has changed to the date of the transplant from the date of entry into the program.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Examples

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
surg	1	-0.77029	0.49718	2.4004	0.1213	0.463
m1	1	-0.24857	0.19437	1.6355	0.2009	0.780
m2	1	0.02958	0.44268	0.0045	0.9467	1.030
m3	**1**	**0.64407**	**0.34276**	**3.5309**	**0.0602**	**1.904**
agetrans	**1**	**0.04927**	**0.02282**	**4.6619**	**0.0308**	**1.050**
wait	1	-0.00197	0.00514	0.1469	0.7015	0.998
dot	1	-0.0001650	0.0002991	0.3044	0.5811	1.000

The two significant estimates imply that each additional year of age when the transplant takes place increases the hazard of dying by about 5 percent and that the hazard of dying almost doubles for those with a unit increase in the measure of tissue mismatch (m3).

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Mathematical and Computational Details

Recall the notation for survival models:

Given n independent observations ($i = 1,\ldots,n$) the data consists of three parts:

t_i = time of the event

δ_i = indicator variable equal to 1 if observation not censored and 0 if censored

$\mathbf{x}_i = [x_{i1} \ldots x_{ik}]$ = vector of covariate values

An ordinary likelihood is written $L = \prod_{i=1}^{n} L_i$ where L_i is the likelihood contribution for the ith observation.

The **partial likelihood** is a product of the likelihoods for all the *events* that are *observed*.

If there are J events, then $PL = \prod_{j=1}^{J} L_j$ where L_j is the likelihood contribution for the jth event.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Mathematical and Computational Details

We’ll see how the factors L_j are formed with an example.
The data is from **Collett** (1994) and consists of 45 breast cancer patients.

The variable **SURV** contains the survival time in months, beginning with the month of surgery.

Twenty-six of the women died (**DEAD** = 1) during the observation period. Thus, there are 26 terms in the partial likelihood.

The variable **X** has a value of 1 if the tumor had a positive marker for possible metastasis; otherwise it equals 0.

To avoid complications with tied data, the survival time for patient 8 is changed from 26 to 25.

The data listed on the next page are sorted by survival time to simplify the construction of the partial likelihood.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Mathematical and Computational Details

Breast cancer Dataset (Collett, 1994)

Obs	surv	dead	x
1	5	1	1
2	8	1	1
3	10	1	1
4	13	1	1
5	18	1	1
6	23	1	0
7	24	1	1
8	**25**	1	1
9	**26**	1	1
10	31	1	1
11	35	1	1
12	40	1	1
13	41	1	1
14	47	1	0
15	48	1	1
16	50	1	1
17	59	1	1
18	61	1	1
19	68	1	1
20	69	1	0
21	70	0	0
22	71	0	0
23	71	1	1
24	76	0	1
25	100	0	0
26	101	0	0
27	105	0	1
28	107	0	1
29	109	0	1
30	113	1	1
31	116	0	1
32	118	1	1
33	143	1	1
34	148	1	0
35	154	0	1
36	162	0	1
37	181	1	0
38	188	0	1
39	198	0	0
40	208	0	0
41	212	0	0
42	212	0	1
43	217	0	1
44	224	0	0
45	225	0	1

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Mathematical and Computational Details

The partial likelihood method is based on the **ordering** of the event times and the **risk set** at each event time.

The first event (death) occurs to patient 1 in month 5. To form the partial likelihood term L_1, we need the probability that patient 1 is the one (only one in this case since we don't have any ties in this dataset) to fail at $t = 5$ given that patients 1 through 45 are all at risk at $t = 5$.

This probability can be shown to be: $L_1 = \dfrac{h_1(5)}{h_1(5) + h_2(5) + \ldots + h_{45}(5)}$.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Mathematical and Computational Details

The second event (death) occurs to patient 2 in month 8. The risk set in this case is patients 2 through 45 (since patient 1 is gone now).

$$L_2 = \frac{h_2(8)}{h_2(8) + h_3(8) + \ldots + h_{45}(8)}.$$

We can continue this way for each successive event, dropping from the risk set those who have experienced the event prior to the current event time (death).

Any **censored observations** are also dropped from a risk set if they occur before the current event time. For example, the 21st death occurred to patient 22 in month 71. Patient 21 was censored at month 70, so her hazard does not appear in the denominator of L_{21}.

When a censored observation occurs at the same time as an event, the convention is to include the censored observation in the risk set for that event.

Thus, patient 23 who was censored in month 71 does show up in the denominator of L_{21}.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Mathematical and Computational Details

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

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

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 expression is not valid if there are tied events, but it is valid for ties between a single event and several censored observations.

As with maximum likelihood estimation, $\log PL = \sum_{i=1}^{n} \delta_i \left[\boldsymbol{\beta} \mathbf{x}_i - \log\left(\sum_{j=1}^{n} Y_{ij} e^{\boldsymbol{\beta} \mathbf{x}_j} \right) \right]$ is the function actually maximized.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Mathematical and Computational Details

Convergence problems can arise if there is a dummy explanatory variable X such that all observations having one of the values of X (0 or 1) occur in censored observations.

In such cases, an estimated value for the parameter of X may be reported when in actuality the parameter is approaching plus or minus infinity.

```
title "Cox Model for the Breast cancer Dataset (Collett, 1994)";
proc phreg data=survival.breast;
       model surv*dead(0) = x;
run;
```

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Mathematical and Computational Details

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	173.914	170.030
AIC	173.914	172.030
SBC	173.914	173.288

Testing Global Null Hypothesis: BETA=0

Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	**3.8843**	**1**	**0.0487**
Score	**3.5194**	**1**	**0.0607**
Wald	**3.2957**	**1**	**0.0695**

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	**Hazard Ratio**
x	1	0.90933	0.50089	3.2957	0.0695	**2.483**

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Mathematical and Computational Details

Note that the *p*-value for the **Wald** and **Score** tests are above the 5% cutoff while that for the **Log-Likelihood** test is below. Sample too small?

The estimated hazard ratio of 2.483 says that the hazard for death for those whose tumor has the positive marker was nearly 2 ½ times the hazard for those without the positive marker.

In a model with a single binary covariate (such as this one), an alternative to testing for different survival curves for the two groups is use **PROC LIFETEST**.

In this case, the *p*-value for the *log-rank test* is identical to that of **PHREG** above. That's because the two tests are equivalent for this special case.

Chapter 5: Estimating Cox Regression Models with PROC PHREG: Partial Likelihood: Mathematical and Computational Details

```
title "Log-rank Test for the Breast cancer Dataset";
proc lifetest data=survival.breast;
       time surv*dead(0);
       strata x;
run;
```

The LIFETEST Procedure

Summary of the Number of Censored and Uncensored Values

Stratum	x	Total	Failed	Censored	Percent Censored
1	0	13	5	8	61.54
2	1	32	21	11	34.38
Total		45	26	19	42.22

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	**3.5194**	**1**	**0.0607**
Wilcoxon	4.1766	1	0.0410
-2Log(LR)	4.3600	1	0.0368