BIOST P8110: Applied Regression II
Lecture 12 - PROC PHREG (Part I)

Qixuan Chen
Department of Biostatistics
Columbia University

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

1. Fit Cox model
2. Estimate hazards ratios and 95% CI
3. Perform hypothesis tests
4. Do model selection
5. Estimate survival functions for specific covariate values

NAME:
Framingham Data (Levy 1999)

SIZE:
4699 observations

SOURCE:
The Framingham Heart Study has collected cardiovascular risk factor data and long-term follow-up on almost 5000 residents of the town of Framingham, Massachusetts. Example used in chapter 7.5 of Dupont (2008).

DESCRIPTIVE ABSTRACT:
In this study, we will investigate the effect of baseline diastolic blood pressure on coronary heart disease adjusted for other risk factors (gender, age, and BMI).

LIST OF VARIABLES:

Variables	Name	Description
1	id	subject id
2	followup	followup days
3	chdfate	CHD status (1 = CHD, 0 = censored)
4	sex	gender (1 = male, 2 = female)
5	dbp	diastolic blood pressure
6	age	age
7	bmi	body mass index
8	dbp_c	nominal scale variable of diastolic blood pressure (1 – dbp ≤ 60, 2 – 60 < bdp ≤ 70, 3 – 70 < bdp ≤ 80, 4 – 80 < bdp ≤ 90, 5 – 90 < bdp ≤ 100, 6 – 100 < bdp ≤ 110, 7 – dbp > 110)

1. Read in the data set

```
/*input the data set*/
data framingham;
      infile C:\Framingham.csv' delimiter = ',' MISSOVER DSD;
      input id followup chdfate sex dbp age bmi;
run;

/*create a nominal variable 'dbp_c'*/
data framingham; set framingham;
      if dbp <= 60 then dbp_c = 1;
      else if dbp <= 70 then dbp_c = 2;
      else if dbp <= 80 then dbp_c = 3;
      else if dbp <= 90 then dbp_c = 4;
      else if dbp <= 100 then dbp_c = 5;
      else if dbp <= 110 then dbp_c = 6;
      else if dbp > 110 then dbp_c = 7;
run;
```

2. Fit a Cox model for "dbp_c"

2.1 SAS Syntax

```
proc phreg data=framingham;
      class dbp_c(ref=first)/param=ref;
      model followup*chdfate(0) = dbp_c /risklimits covb ties=EFRON;
      title "Model 1: DBP_c";
run;
```

1) The PROC PHREG statement invokes the procedure to fit a Cox proportional hazards model.
2) The CLASS statement names the categorical variables to be used in the analysis.
    a. The CLASS statement must precede the MODEL statement.
    b. You can specify various *options* for each variable by enclosing them in parentheses after the variable name. For example, the first ordered level of the "dbp_c" is used as the reference cell. You can also use "(ref=last)" for the last ordered level of a categorical variable as the reference cell.
    c. You can also specify global *options* for the CLASS statement by placing them after a slash (/).
    d. The parameterization method for the categorical variable has many options, which can be specified using "param=". Here we use the reference cell method.

3) The MODEL statement identifies the variables to be used as the failure time variables, the censoring variable, and the explanatory effects, including covariates and interactions.
    a. The failure time variable is "followup", and the censoring variable is "chdfate". The censoring variable must be numeric. The failure time variables must contain nonnegative values. Any observation with a negative failure time is excluded from the analysis, as is any observation with a missing value for any of the variables listed in the MODEL statement.
    b. A list of censoring values (separated by blanks or commas if there is more than one) is enclosed in parentheses. If the censoring variable takes on one of these values, the corresponding failure time is considered to be censored.

c. The variables following the equal sign are the explanatory variables (sometimes called independent variables or covariates) for the model.
d. Many options available in the MODEL statement, which can be specified after a slash (/). The three options we use here are:
    i. RISKLIMITS: produces confidence intervals for hazard ratios of main effects not involved in interactions. The confidence coefficient can be specified with the ALPHA= option.
    ii. COVB: displays the estimated covariance matrix of the parameter estimates.
    iii. TIES=: specifies how to handle ties in the failure time. There are four methods available. They are "BRESLOW", "DISCRETE", "EFRON", and "EXACT".
        - Both the EXACT and DISCRETE methods produce exact results based on true partial likelihood estimates, but they have different assumptions for the ties. Both methods can take a considerable amount of computer resources, especially the EXACT method.
        - If the exact methods are too time-consuming, use the EFRON approximation, at least for model exploration.
        - If ties are not extensive, the EFRON and BRESLOW methods provide satisfactory approximations.
        - If there are no ties, all four methods result in the same likelihood and yield identical estimates. The default, TIES=BRESLOW, is the most efficient method when there are no ties.
        - For more information for the different methods for ties, see pages 144-153 (Allison 2010).

2.2 SAS Output

Model 1: DBP_c

The PHREG Procedure

Model Information

Data Set	WORK.FRAMINGHAM
Dependent Variable	followup
Censoring Variable	chdfate
Censoring Value(s)	0
Ties Handling	EFRON

Number of Observations Read	4699
Number of Observations Used	4699

Class Level Information

Class	Value	Design Variables					
dbp_c	1	0	0	0	0	0	0
	2	1	0	0	0	0	0
	3	0	1	0	0	0	0
	4	0	0	1	0	0	0
	5	0	0	0	1	0	0
	6	0	0	0	0	1	0
	7	0	0	0	0	0	1

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
4699	1473	3226	68.65

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	23669.640	23447.801
AIC	23669.640	23459.801
SBC	23669.640	23491.571

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	221.8389	6	<.0001
Score	259.7158	6	<.0001
Wald	237.1973	6	<.0001

Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
dbp_c	6	237.1973	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Rati Confidence Limits	Label
dbp_c	2	1	0.67741	0.24709	7.5162	0.0061	1.969	1.213 3.195	dbp_c 2
dbp_c	3	1	0.93917	0.24073	15.2211	<.0001	2.558	1.596 4.100	dbp_c 3
dbp_c	4	1	1.11714	0.24092	21.5009	<.0001	3.056	1.906 4.901	dbp_c 4
dbp_c	5	1	1.51230	0.24313	38.6892	<.0001	4.537	2.817 7.307	dbp_c 5
dbp_c	6	1	1.83926	0.25442	52.2605	<.0001	6.292	3.821 10.360	dbp_c 6
dbp_c	7	1	2.24777	0.27124	68.6765	<.0001	9.467	5.563 16.109	dbp_c 7

Estimated Covariance Matrix

Parameter		dbp_c2	dbp_c3	dbp_c4	dbp_c5	dbp_c6	dbp_c7
dbp_c2	dbp_c 2	0.0610521584	0.0555588292	0.0555598984	0.0555619846	0.0555643335	0.0555667270
dbp_c3	dbp_c 3	0.0555588292	0.0579488305	0.0555648606	0.0555701688	0.0555763746	0.0555821291
dbp_c4	dbp_c 4	0.0555598984	0.0555648606	0.0580445402	0.0555776862	0.0555877318	0.0555973630
dbp_c5	dbp_c 5	0.0555619846	0.0555701688	0.0555776862	0.0591136568	0.0556105026	0.0556288823
dbp_c6	dbp_c 6	0.0555643335	0.0555763746	0.0555877318	0.0556105026	0.0647307787	0.0556705266
dbp_c7	dbp_c 7	0.0555667270	0.0555821291	0.0555973630	0.0556288823	0.0556705266	0.0735692271

2.3 Answer the following questions using the above SAS outputs

a. Write down Model 1.

b. Is diastolic blood pressure (DBP) an important cardiovascular risk factor?

c. What is the estimate and 95% CI of the hazards ratio between people with DBP > 110 and people with DBP ≤ 60?

d. What is the estimate and 95% CI of the hazards ratio between people with DBP > 110 and people with 60 < DBP ≤ 70?

e. Is the hazard of cardiovascular disease significantly different between people with DBP > 110 and people with 60 < DBP ≤ 70?

2.4 All pairs of hazard ratios can be estimated by using the HAZARDRATIO statement:

```
proc phreg data=framingham;
      class dbp_c(ref=first)/param=ref;
      model followup*chdfate(0) = dbp_c /risklimits covb ties=EFRON;
      hazardratio 'all pairs' dbp_c/ diff=all;
      title "Model 1: Test all pairs";
run;
```

SAS output:

Model 1: Test all pairs

all pairs: Hazard Ratios for dbp_c

Description	Point Estimate	95% Wald Confidence Limits	
dbp_c 1 vs 2	0.508	0.313	0.824
dbp_c 1 vs 3	0.391	0.244	0.627
dbp_c 1 vs 4	0.327	0.204	0.525
dbp_c 1 vs 5	0.220	0.137	0.355
dbp_c 1 vs 6	0.159	0.097	0.262
dbp_c 1 vs 7	0.106	0.062	0.180
dbp_c 2 vs 3	0.770	0.647	0.916
dbp_c 2 vs 4	0.644	0.541	0.767
dbp_c 2 vs 5	0.434	0.360	0.523
dbp_c 2 vs 6	0.313	0.247	0.397
dbp_c 2 vs 7	0.208	0.154	0.281

dbp_c 3 vs 4	0.837	0.730	0.960
dbp_c 3 vs 5	0.564	0.485	0.656
dbp_c 3 vs 6	0.407	0.329	0.502
dbp_c 3 vs 7	0.270	0.204	0.357
dbp_c 4 vs 5	0.674	0.579	0.784
dbp_c 4 vs 6	0.486	0.393	0.600
dbp_c 4 vs 7	0.323	0.244	0.427
dbp_c 5 vs 6	0.721	0.579	0.899
dbp_c 5 vs 7	0.479	0.360	0.639
dbp_c 6 vs 7	0.665	0.482	0.917

3. Add confounders

3.1 SAS Syntax

```
/* Cox regression model for DBP and the potential confounders */
proc phreg data=framingham;
      class dbp_c(ref=first) sex(ref=last) /param=ref;
      model followup*chdfate(0) = dbp_c sex age BMI /ties=EFRON;
      title "Model 2: DBP+sex+age+BMI";
run;
```

3.2 SAS Output

Model 2: DBP+sex+age+BMI

The PHREG Procedure

Model Information

Data Set	WORK.FRAMINGHAM
Dependent Variable	followup
Censoring Variable	chdfate
Censoring Value(s)	0
Ties Handling	EFRON

Number of Observations Read	4699
Number of Observations Used	4690

Class Level Information

Class	Value	Design Variables					
dbp_c	1	0	0	0	0	0	0
	2	1	0	0	0	0	0
	3	0	1	0	0	0	0
	4	0	0	1	0	0	0
	5	0	0	0	1	0	0
	6	0	0	0	0	1	0
	7	0	0	0	0	0	1
sex	1	1					
	2	0					

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
4690	1472	3218	68.61

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	23650.536	22991.385
AIC	23650.536	23009.385
SBC	23650.536	23057.034

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	659.1516	9	<.0001
Score	688.3623	9	<.0001
Wald	642.0029	9	<.0001

Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
dbp_c	6	84.8182	<.0001
sex	1	175.6803	<.0001
age	1	259.6367	<.0001
bmi	1	38.0537	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
dbp_c 2	1	0.53527	0.24725	4.6867	0.0304	1.708	dbp_c 2
dbp_c 3	1	0.66959	0.24132	7.6990	0.0055	1.953	dbp_c 3
dbp_c 4	1	0.68244	0.24239	7.9269	0.0049	1.979	dbp_c 4
dbp_c 5	1	0.91521	0.24577	13.8669	0.0002	2.497	dbp_c 5
dbp_c 6	1	1.28118	0.25708	24.8363	<.0001	3.601	dbp_c 6
dbp_c 7	1	1.54674	0.27583	31.4438	<.0001	4.696	dbp_c 7
sex 1	1	0.70377	0.05310	175.6803	<.0001	2.021	sex 1
age	1	0.05310	0.00330	259.6367	<.0001	1.055	
bmi	1	0.04045	0.00656	38.0537	<.0001	1.041	

3.3 Answer the following questions using the above SAS outputs

a. Write down Model 2.

b. Is Model 2 significantly better than the null model?

c. Is Model 2 significantly better than Model 1?

4. Test interactions

4.1 SAS Syntax

```
/* Test if BMI is an effect modifier for the relation between DBP and CHD */
proc phreg data=framingham;
      class dbp_c(ref=first) sex(ref=last) /param=ref;
      model followup*chdfate(0) = dbp_c sex age BMI dbp_c*BMI/risklimits covb
                                                              ties=EFRON;
      title "Model 3: DBP+sex+age+BMI+DBP*BMI";
run;
```

4.2 SAS Output

Model 3: DBP+sex+age+BMI+DBP*sex

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	23650.536	22972.359
AIC	23650.536	23002.359
SBC	23650.536	23081.775

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	678.1772	15	<.0001
Score	707.1921	15	<.0001
Wald	645.8151	15	<.0001

Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
dbp_c	6	26.6389	0.0002
sex	1	171.2208	<.0001
age	1	250.5607	<.0001
bmi	1	16.5512	<.0001
bmi*dbp_c	6	20.3223	0.0024

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
dbp_c	2	1	6.02135	1.96079	9.4303	0.0021	.	.	.	dbp_c 2
dbp_c	3	1	7.06858	1.90429	13.7784	0.0002	.	.	.	dbp_c 3
dbp_c	4	1	6.81148	1.90498	12.7851	0.0003	.	.	.	dbp_c 4
dbp_c	5	1	7.48475	1.90971	15.3611	<.0001	.	.	.	dbp_c 5
dbp_c	6	1	6.62717	1.96161	11.4138	0.0007	.	.	.	dbp_c 6
dbp_c	7	1	9.02014	2.01090	20.1208	<.0001	.	.	.	dbp_c 7
sex	1	1	0.69725	0.05329	171.2208	<.0001	2.008	1.809	2.229	sex 1
age		1	0.05240	0.00331	250.5607	<.0001	1.054	1.047	1.061	
bmi		1	0.29815	0.07329	16.5512	<.0001	.	.	.	
bmi*dbp_c	2	1	-0.22737	0.07697	8.7268	0.0031	.	.	.	dbp_c 2 * bmi
bmi*dbp_c	3	1	-0.26437	0.07455	12.5750	0.0004	.	.	.	dbp_c 3 * bmi
bmi*dbp_c	4	1	-0.25392	0.07443	11.6383	0.0006	.	.	.	dbp_c 4 * bmi
bmi*dbp_c	5	1	-0.26975	0.07449	13.1139	0.0003	.	.	.	dbp_c 5 * bmi
bmi*dbp_c	6	1	-0.22679	0.07595	8.9178	0.0028	.	.	.	dbp_c 6 * bmi
bmi*dbp_c	7	1	-0.29962	0.07741	14.9817	0.0001	.	.	.	dbp_c 7 * bmi

The PHREG Procedure

Estimated Covariance Matrix

Parameter		dbp_c2	dbp_c3	dbp_c4	dbp_c5	dbp_c6
dbp_c2	dbp_c 2	3.844678348	3.501381214	3.501450548	3.501598569	3.501457512
dbp_c3	dbp_c 3	3.501381214	3.626324752	3.502030043	3.502452217	3.502094586
dbp_c4	dbp_c 4	3.501450548	3.502030043	3.628942110	3.504117415	3.503447214
dbp_c5	dbp_c 5	3.501598569	3.502452217	3.504117415	3.646978126	3.504924552
dbp_c6	dbp_c 6	3.501457512	3.502094586	3.503447214	3.504924552	3.847913468
dbp_c7	dbp_c 7	3.501960794	3.502943414	3.504892484	3.506762864	3.505793211
sex1	sex 1	0.000312879	0.000113058	-0.000502647	-0.001373419	-0.002149061
age		-0.000009533	-0.000054246	-0.000134819	-0.000193717	-0.000134653
bmi		0.136004780	0.136028277	0.136072745	0.136114966	0.136085505
dbp_c2bmi	dbp_c 2 * bmi	-0.149707733	-0.135995901	-0.135998522	-0.136004648	-0.135999962
dbp_c3bmi	dbp_c 3 * bmi	-0.135995467	-0.140827150	-0.136018076	-0.136033337	-0.136021356
dbp_c4bmi	dbp_c 4 * bmi	-0.135997768	-0.136017487	-0.140639759	-0.136088729	-0.136065594
dbp_c5bmi	dbp_c 5 * bmi	-0.136000534	-0.136027625	-0.136080856	-0.141071763	-0.136109121
dbp_c6bmi	dbp_c 6 * bmi	-0.135991221	-0.136010853	-0.136053968	-0.136102701	-0.147618794
dbp_c7bmi	dbp_c 7 * bmi	-0.136003275	-0.136032923	-0.136092699	-0.136151066	-0.136127559

Estimated Covariance Matrix

Parameter		dbp_c7	sex1	age	bmi	dbp_c2bmi
dbp_c2	dbp_c 2	3.501960794	0.000312879	-0.000009533	0.136004780	-0.149707733
dbp_c3	dbp_c 3	3.502943414	0.000113058	-0.000054246	0.136028277	-0.135995901
dbp_c4	dbp_c 4	3.504892484	-0.000502647	-0.000134819	0.136072745	-0.135998522
dbp_c5	dbp_c 5	3.506762864	-0.001373419	-0.000193717	0.136114966	-0.136004648
dbp_c6	dbp_c 6	3.505793211	-0.002149061	-0.000134653	0.136085505	-0.135999962
dbp_c7	dbp_c 7	4.043725281	-0.001546461	-0.000225183	0.136147849	-0.136019280
sex1	sex 1	-0.001546461	0.002839343	0.000022177	-0.000017151	-0.000011221
age		-0.000225183	0.000022177	0.000010960	-0.000005510	0.000000319
bmi		0.136147849	-0.000017151	-0.000005510	0.005370772	-0.005367683
dbp_c2bmi	dbp_c 2 * bmi	-0.136019280	-0.000011221	0.000000319	-0.005367683	0.005923914
dbp_c3bmi	dbp_c 3 * bmi	-0.136052265	-0.000004348	0.000001869	-0.005368467	0.005367316
dbp_c4bmi	dbp_c 4 * bmi	-0.136117289	0.000014764	0.000004591	-0.005369955	0.005367402
dbp_c5bmi	dbp_c 5 * bmi	-0.136168631	0.000045412	0.000006101	-0.005371060	0.005367521
dbp_c6bmi	dbp_c 6 * bmi	-0.136131375	0.000081603	0.000003971	-0.005369849	0.005367183
dbp_c7bmi	dbp_c 7 * bmi	-0.154098167	0.000058835	0.000006604	-0.005371514	0.005367633

Estimated Covariance Matrix

Parameter		dbp_c3bmi	dbp_c4bmi	dbp_c5bmi	dbp_c6bmi	dbp_c7bmi
dbp_c2	dbp_c 2	-0.135995467	-0.135997768	-0.136000534	-0.135991221	-0.136003275
dbp_c3	dbp_c 3	-0.140827150	-0.136017487	-0.136027625	-0.136010853	-0.136032923
dbp_c4	dbp_c 4	-0.136018076	-0.140639759	-0.136080856	-0.136053968	-0.136092699
dbp_c5	dbp_c 5	-0.136033337	-0.136088729	-0.141071763	-0.136102701	-0.136151066
dbp_c6	dbp_c 6	-0.136021356	-0.136065594	-0.136109121	-0.147618794	-0.136127559
dbp_c7	dbp_c 7	-0.136052265	-0.136117289	-0.136168631	-0.136131375	-0.154098167
sex1	sex 1	-0.000004348	0.000014764	0.000045412	0.000081603	0.000058835
age		0.000001869	0.000004591	0.000006101	0.000003971	0.000006604
bmi		-0.005368467	-0.005369955	-0.005371060	-0.005369849	-0.005371514
dbp_c2bmi	dbp_c 2 * bmi	0.005367316	0.005367402	0.005367521	0.005367183	0.005367633
dbp_c3bmi	dbp_c 3 * bmi	0.005557798	0.005368065	0.005368430	0.005367843	0.005368630
dbp_c4bmi	dbp_c 4 * bmi	0.005368065	0.005539753	0.005370198	0.005369245	0.005370619
dbp_c5bmi	dbp_c 5 * bmi	0.005368430	0.005370198	0.005548727	0.005370710	0.005372249
dbp_c6bmi	dbp_c 6 * bmi	0.005367843	0.005369245	0.005370710	0.005767713	0.005371364
dbp_c7bmi	dbp_c 7 * bmi	0.005368630	0.005370619	0.005372249	0.005371364	0.005992256

4.3 Answer the following questions using the above SAS outputs

a. Write down Model 3.

b. Is Model 3 significantly better than Model 2?

c. Is BMI a significant effect modifier for the relation between DBP and CHD?

d. What is the estimate and 95% CI of the hazards ratio between resident A (DBP > 110 and BMI = 30) and resident B (DBP ≤ 60 and BMI = 30)?

4.4 Question (d) can also be done using HAZRDRATIO statement:

```
proc phreg data=framingham;
      class dbp_c(ref=first) sex(ref=last) /param=ref;
      model followup*chdfate(0) = dbp_c sex age BMI dbp_c*BMI/alpha=0.05
                                                    risklimits covb ties=EFRON;
      hazardratio 'all pairs diff at BMI=30' dbp_c / at (BMI=30) diff=ALL;
      title "Model 3: DBP+sex+age+BMI+DBP*sex (Q 4.3.d)";
run;
```

SAS Output:

Model 3: DBP+sex+age+BMI+DBP*sex (Q 4.3.d)

The PHREG Procedure

all pairs diff at BMI=30: Hazard Ratios for dbp_c

Description	Point Estimate	95% Wald Confidence Limits	
dbp_c 1 vs 2 At bmi=30	2.225	0.939	5.272
dbp_c 1 vs 3 At bmi=30	2.369	1.034	5.425
dbp_c 1 vs 4 At bmi=30	2.239	0.983	5.099
dbp_c 1 vs 5 At bmi=30	1.836	0.806	4.184
dbp_c 1 vs 6 At bmi=30	1.193	0.517	2.752
dbp_c 1 vs 7 At bmi=30	0.969	0.412	2.282
dbp_c 2 vs 3 At bmi=30	1.065	0.767	1.478
dbp_c 2 vs 4 At bmi=30	1.006	0.735	1.377
dbp_c 2 vs 5 At bmi=30	0.825	0.602	1.131
dbp_c 2 vs 6 At bmi=30	0.536	0.380	0.758
dbp_c 2 vs 7 At bmi=30	0.436	0.294	0.645
dbp_c 3 vs 4 At bmi=30	0.945	0.772	1.157
dbp_c 3 vs 5 At bmi=30	0.775	0.632	0.951
dbp_c 3 vs 6 At bmi=30	0.504	0.393	0.646
dbp_c 3 vs 7 At bmi=30	0.409	0.300	0.559
dbp_c 4 vs 5 At bmi=30	0.820	0.685	0.982
dbp_c 4 vs 6 At bmi=30	0.533	0.424	0.670
dbp_c 4 vs 7 At bmi=30	0.433	0.322	0.582
dbp_c 5 vs 6 At bmi=30	0.650	0.516	0.818
dbp_c 5 vs 7 At bmi=30	0.528	0.392	0.710
dbp_c 6 vs 7 At bmi=30	0.812	0.585	1.128

5. Estimate conditional survival functions

5.1 SAS Syntax

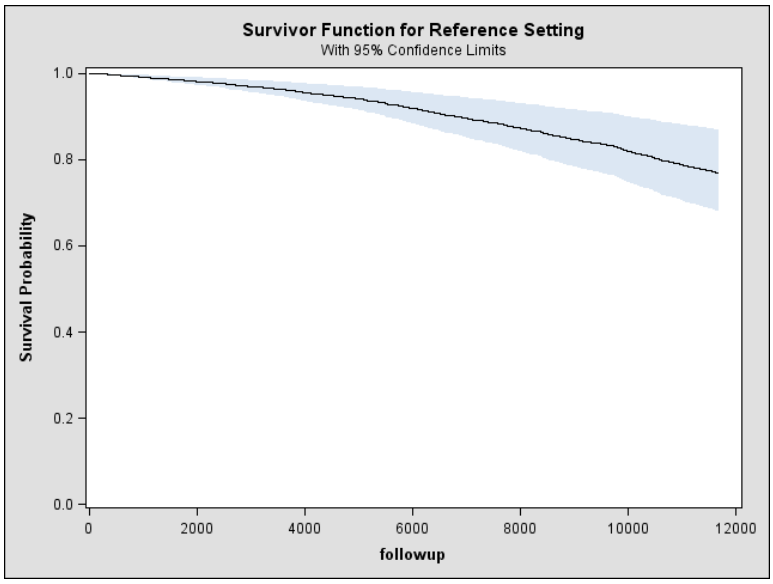
```
/*Estimate the survival function for a person with the continuous covariates
taking the mean value and the categorical variables taking the reference level*/

ods graphics on;
proc phreg data=framingham plots(cl)=s;
      class dbp_c(ref=first) sex(ref=last) /param=ref;
      model followup*chdfate(0) = dbp_c sex age BMI dbp_c*BMI/ties=EFRON;
      baseline out=a survival=s lower=lcl upper=ucl;
      title "Estimate S(t) for covariates at the mean values or reference cell";
run;
ods graphics off;
```

1) When there are no time-dependent covariates, the Cox model can be written as $S_i(t) = [S_0(t)]^{\exp(\beta x_i)}$, where $S_i(t)$ is the probability that individual i with covariate values x_i survive to time t. $S_0(t)$ is called the baseline survival function, and can be estimated by a nonparametric maximum likelihood method.

2) The BASELINE statement and the PLOTS option can be used to get $S_i(t)$ with average values for continuous covariates and reference cell for categorical variables.
   a. The OUT=A option outputs the survival estimates in a data set called A.
   b. The SURVIVAL=S stores the survival probabilities in a variable named S.
   c. The UPPER and LOWER options store the upper and lower 95% confidence limits in variables UCL and LCL, respectively.

5.2 SAS Output

Reference Set of Covariates for Plotting			
age	bmi	dbp_c	sex
46.030064	25.631706	1.000000	2.000000



The first 10 lines of output from data A:

Estimate S(t) for covariates at the mean values or reference cell

Obs	dbp_c	sex	age	bmi	followup	s	lcl	ucl
1	1	2	46.0301	25.6317	0	1.00000	.	.
2	1	2	46.0301	25.6317	18	0.99991	0.99974	1.00000
3	1	2	46.0301	25.6317	35	0.99983	0.99958	1.00000
4	1	2	46.0301	25.6317	109	0.99974	0.99943	1.00000
5	1	2	46.0301	25.6317	147	0.99966	0.99928	1.00000
6	1	2	46.0301	25.6317	169	0.99957	0.99914	1.00000
7	1	2	46.0301	25.6317	199	0.99948	0.99900	0.99996
8	1	2	46.0301	25.6317	201	0.99940	0.99887	0.99992
9	1	2	46.0301	25.6317	209	0.99931	0.99874	0.99989
10	1	2	46.0301	25.6317	265	0.99922	0.99860	0.99985

5.3 Get predictions about survival time for particular sets of covariate values:

```
/* Estimating survival curves*/
data one;
input id sex dbp_c age bmi;
datalines;
1 1 1 50 25
2 2 2 60 30
;
run;

ods graphics on;
proc phreg data=framingham plots(cl overlay)=survival;
      class dbp_c(ref=first) sex(ref=last) /param=ref;
      model followup*chdfate(0) = dbp_c sex age BMI dbp_c*BMI/ties=EFRON;
      baseline covariates=one out=pred survival=_all_ /rowid=id;
      title "Estimate S(t) for the two subjects in the data 'one'";
run;
ods graphics off;

proc print data=pred;
run;
```

1) The PLOTS (OVERLAY) option is used to request the curves to be overlaid.
2) In the BASELINE statement,
    a. The COVARIATES names the SAS data set that contains the sets of covariate values for which the quantities of interest are estimated. All variables in the COVARIATES= data set are copied to the OUT= data set. Thus, any variable in the COVARIATES= data set can be used to identify the covariate sets in the OUT= data set.
    b. Specifying SURVIVAL=_ALL_ is equivalent to specifying SURVIVAL=, STDERR=, LOWER=, and UPPER=.
    c. ROWID names a variable in the COVARIATES= data set for identifying the baseline function curves in the plots