1. Open the WHAS500 data set in the software program of your choice

Obs	admitdate	disdate	fdate	id	age	gender	hr	sysbp	diasbp	bmi	cvd	afb	sho	chf	av3	miord	mitype	year	los	dstat	lenfol	fstat	time_yrs
1	01/13/1997	01/18/1997	12/31/2002	1	83	0	89	152	78	25.5405	1	1	0	0	0	1	0	1	5	0	2178	0	5.96304
2	01/19/1997	01/24/1997	12/31/2002	2	49	0	84	120	60	24.0240	1	0	0	0	0	0	1	1	5	0	2172	0	5.94661
3	02/17/1997	02/27/1997	12/11/1997	4	70	0	65	123	76	26.6319	1	0	0	1	0	0	1	1	10	0	297	1	0.81314
4	03/01/1997	03/07/1997	12/31/2002	5	70	0	63	135	85	24.4126	1	0	0	0	0	0	1	1	6	0	2131	0	5.83436
5	03/11/1997	03/12/1997	03/12/1997	6	70	0	76	83	54	23.2424	1	0	0	0	1	0	0	1	1	1	1	1	0.00274

It is always a good idea to peek at the first few rows of a dataset to orient yourself at the start.

a. Calculate and graph on the same graph a Kaplan-Meier curve for the three cohorts associated with year. Does it appear as if these survival curves differ? If so, do they appear to violate the assumption of proportional hazards?

The LIFETEST Procedure

Summary of the Number of Censored and Uncensored Values

Stratum	year	Total	Failed	Censored	Percent Censored
1	1	160	78	82	51.25
2	2	188	77	111	59.04
3	3	152	60	92	60.53
Total		500	215	285	57.00

The survival curves are reasonable, with the possible exception of year 3. The latter half of that curve (corresponing to about 1.5 to 2 years appears to decline a bit too sharply

a. Calculate and graph on the same graph a Kaplan-Meier curve for the three cohorts associated with year. Does it appear as if these survival curves differ? If so, do they appear to violate the assumption of proportional hazards?

The LIFETEST Procedure

Testing Homogeneity of Survival Curves for time_yrs over Strata

Rank Statistics		
year	Log-Rank	Wilcoxon
1	-13.357	-4507.0
2	-1.616	-746.0
3	14.972	5253.0

Covariance Matrix for the Log-Rank Statistics			
year	1	2	3
1	44.5218	-29.5636	-14.9582
2	-29.5636	46.6305	-17.0669
3	-14.9582	-17.0669	32.0251

Covariance Matrix for the Wilcoxon Statistics			
year	1	2	3
1	6629211	-3954367	-2674844
2	-3954367	7022111	-3067744
3	-2674844	-3067744	5742588

The survival curves are reasonable, with the possible exception of year 3. The latter half of that curve (corresponing to about 1.5 to 2 years appears to decline a bit too sharply

a. Calculate and graph on the same graph a Kaplan-Meier curve for the three cohorts associated with year. Does it appear as if these survival curves differ? If so, do they appear to violate the assumption of proportional hazards?

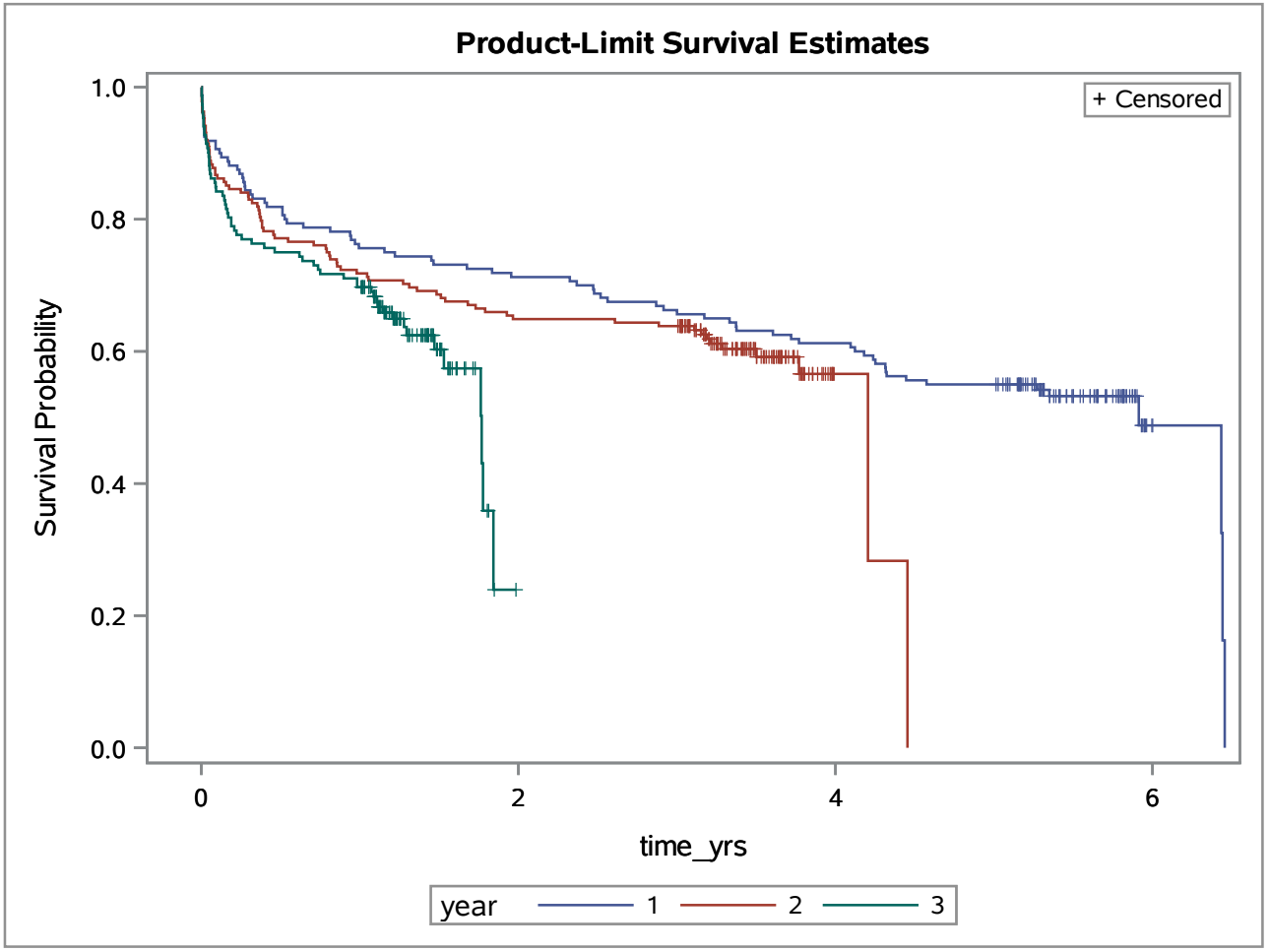
The LIFETEST Procedure

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	8.0786	2	0.0176
Wilcoxon	5.5936	2	0.0610
-2Log(LR)	38.6984	2	<.0001

The survival curves are reasonable, with the possible exception of year 3. The latter half of that curve (corresponing to about 1.5 to 2 years appears to decline a bit too sharply

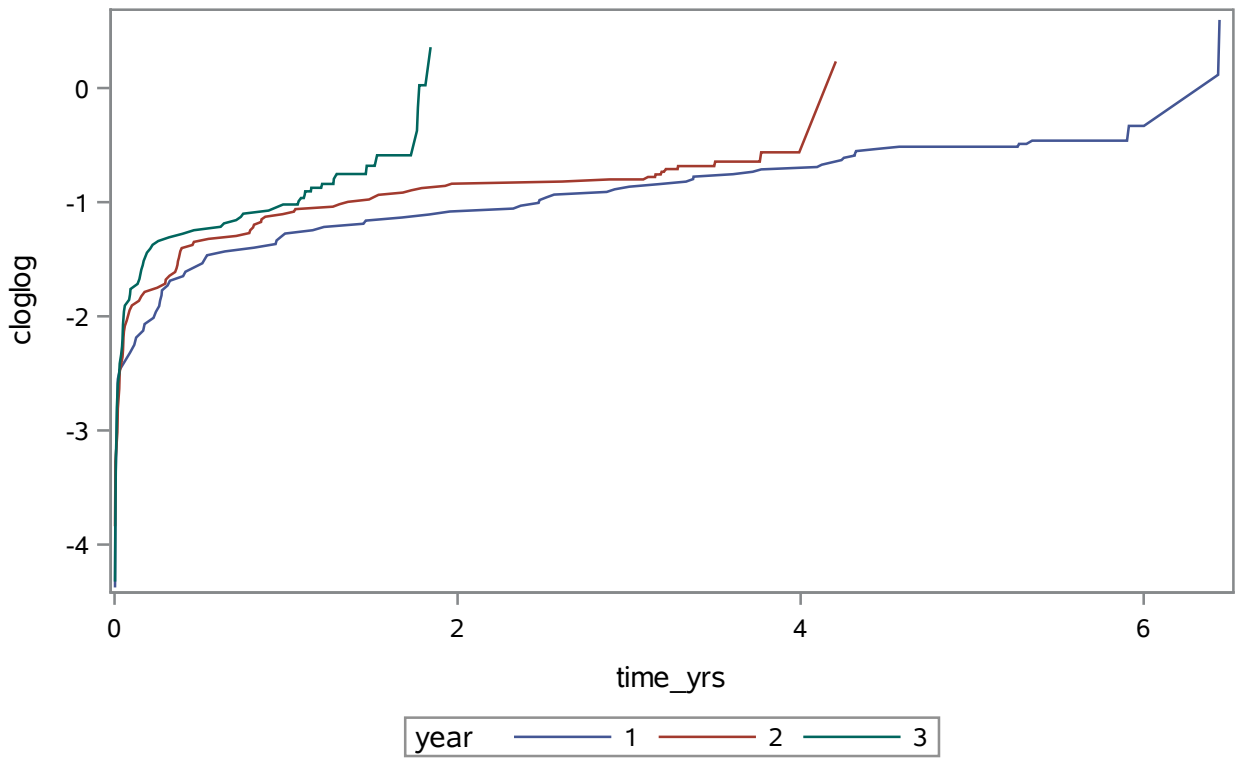
a. Calculate and graph on the same graph a Kaplan-Meier curve for the three cohorts associated with year. Does it appear as if these survival curves differ? If so, do they appear to violate the assumption of proportional hazards?

The LIFETEST Procedure



The survival curves are reasonable, with the possible exception of year 3. The latter half of that curve (corresponing to about 1.5 to 2 years appears to decline a bit too sharply

Introduction to survival analysis. Exercises 06, SAS
b. Calculate and interpret the complementary log-log plots.



These plots show the same pattern. The plots are reasonably parallel with the possible exception of year 3. proc print data=km_by_year(obs=10)

c. Calculate, plot, and interpret the Schoenfeld residuals from a Cox regression model with year as the only independent variable.

The PHREG Procedure

Model Information	
Data Set	WORK.TIME_RECODE
Dependent Variable	time_yrs
Censoring Variable	fstat
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	500
Number of Observations Used	500

Summary of the Number of Event and Censored Values			
Total	**Event**	**Censored**	**Percent Censored**
500	215	285	57.00

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

Model Fit Statistics		
Criterion	**Without Covariates**	**With Covariates**
-2 LOG L	2455.158	2447.653
AIC	2455.158	2449.653
SBC	2455.158	2453.023

The Schoenfeld residuals ...

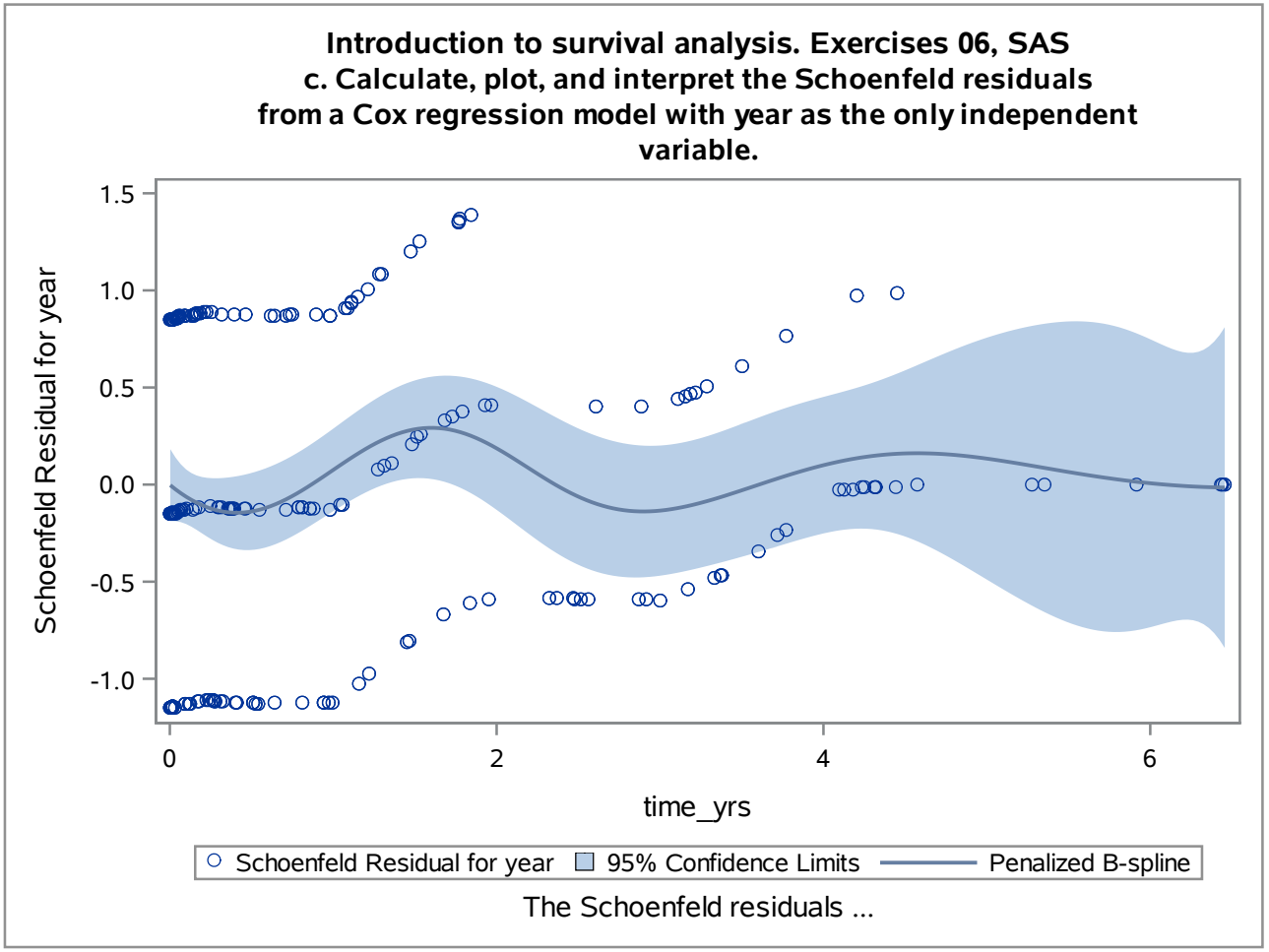
c. Calculate, plot, and interpret the Schoenfeld residuals from a Cox regression model with year as the only independent variable.

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	7.5055	1	0.0062
Score	7.5159	1	0.0061
Wald	7.4424	1	0.0064

Analysis of Maximum Likelihood Estimates						
Parameter	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
year	1	0.26531	0.09725	7.4424	0.0064	1.304

The Schoenfeld residuals ...


Introduction to survival analysis. Exercises 06, SAS
c. Calculate, plot, and interpret the Schoenfeld residuals from a Cox regression model with year as the only independent variable.
Schoenfeld Residual for year
1.5
1.0
0.5
0.0
-0.5
-1.0
0
2
4
6
time_yrs
Schoenfeld Residual for year
95% Confidence Limits
Penalized B-spline
The Schoenfeld residuals ...

d. Fit a Cox regression model with gender as an independent variable and include year as a strata. Create estimated survival plots for each strata comparing males to females.

The PHREG Procedure

Model Information	
Data Set	WORK.AUGMENT
Dependent Variable	time_yrs
Censoring Variable	fstat
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	3500
Number of Observations Used	500

Summary of the Number of Event and Censored Values

Stratum	year	Total	Event	Censored	Percent Censored
1	1	160	78	82	51.25
2	2	188	77	111	59.04
3	3	152	60	92	60.53
Total		500	215	285	57.00

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

The survival plots ...

d. Fit a Cox regression model with gender as an independent variable and include year as a strata. Create estimated survival plots for each strata comparing males to females.

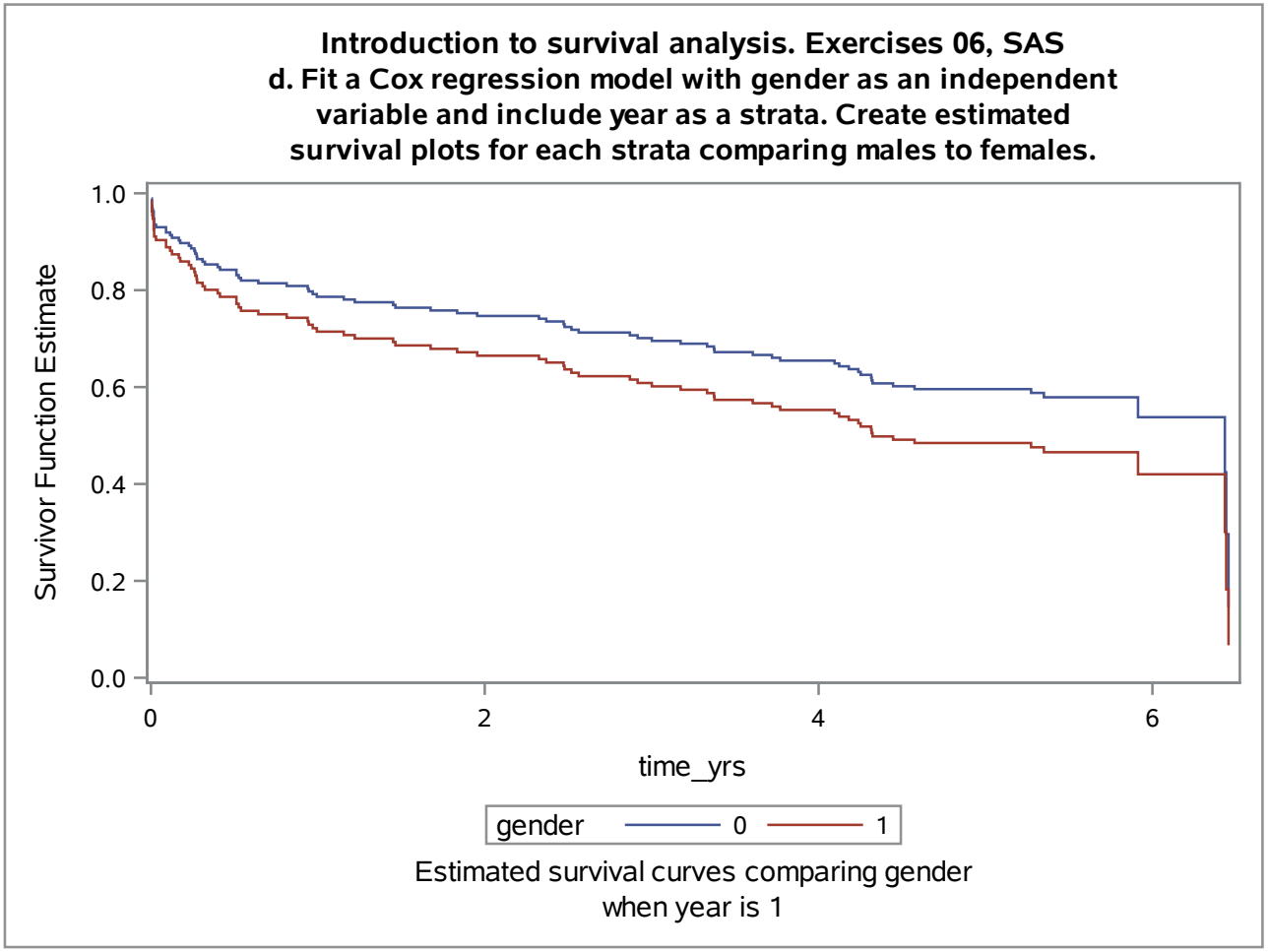
The PHREG Procedure

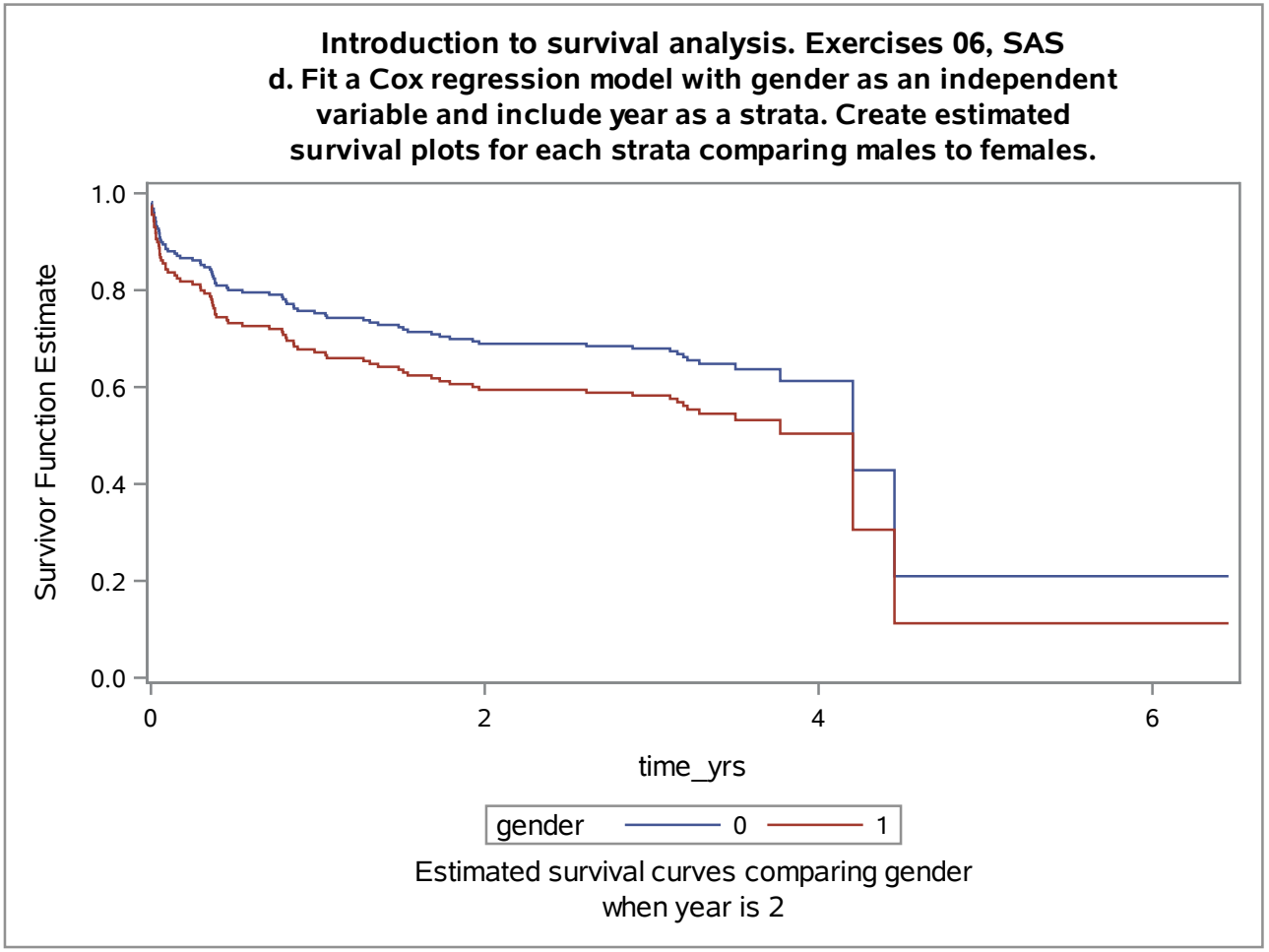
Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	2005.787	2000.032
AIC	2005.787	2002.032
SBC	2005.787	2005.403

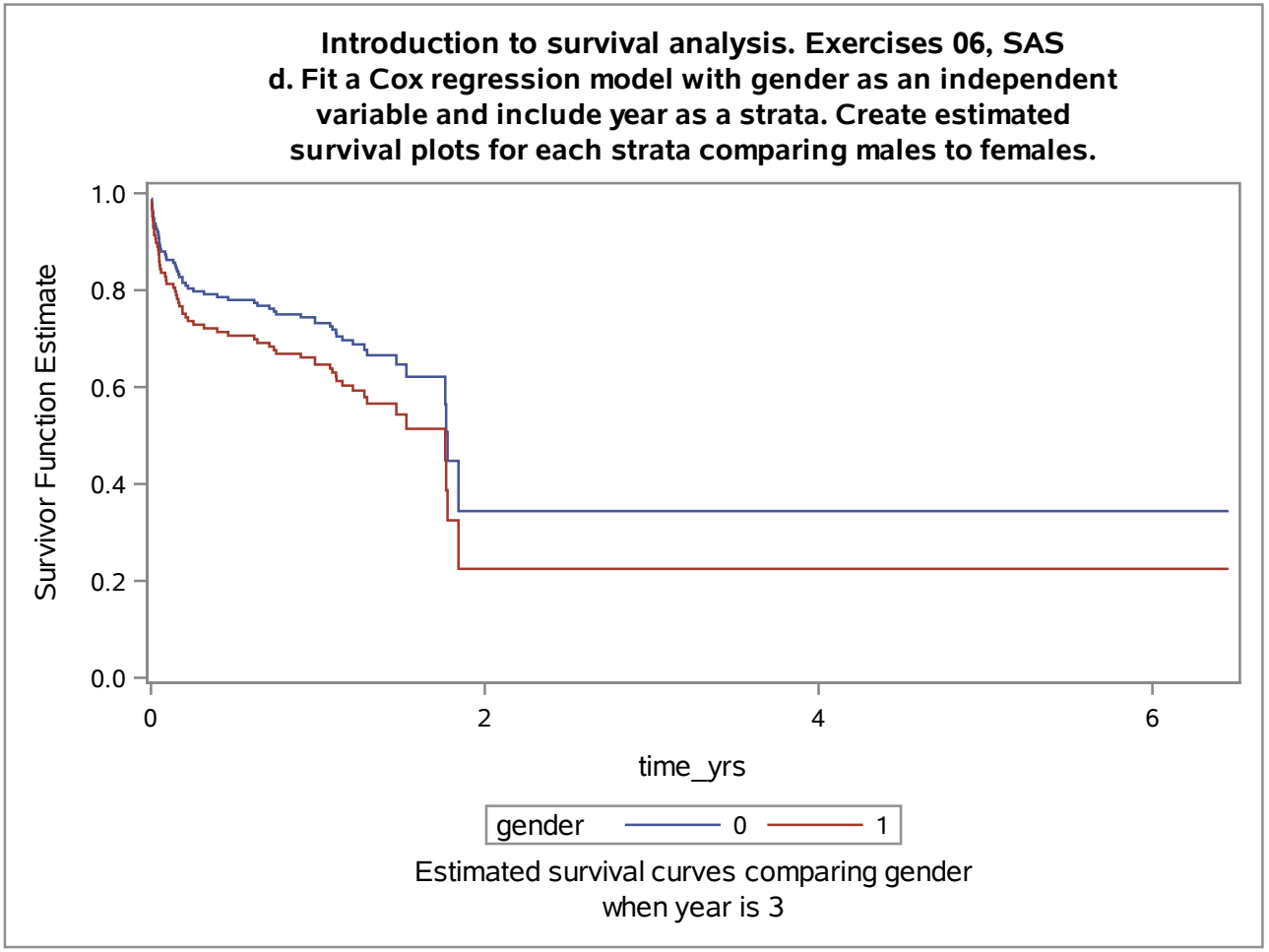
Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	5.7552	1	0.0164
Score	5.8755	1	0.0154
Wald	5.8222	1	0.0158

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
gender	1	0.33566	0.13911	5.8222	0.0158	1.399

The survival plots ...


Introduction to survival analysis. Exercises 06, SAS
d. Fit a Cox regression model with gender as an independent variable and include year as a strata. Create estimated survival plots for each strata comparing males to females.
Survivor Function Estimate
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
time_yrs
gender
0
1
Estimated survival curves comparing gender when year is 1


Introduction to survival analysis. Exercises 06, SAS
d. Fit a Cox regression model with gender as an independent variable and include year as a strata. Create estimated survival plots for each strata comparing males to females.
Survivor Function Estimate
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
time_yrs
gender
0
1
Estimated survival curves comparing gender when year is 2


Introduction to survival analysis. Exercises 06, SAS
d. Fit a Cox regression model with gender as an independent variable and include year as a strata. Create estimated survival plots for each strata comparing males to females.
Survivor Function Estimate
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
time_yrs
gender
0
1
Estimated survival curves comparing gender when year is 3