Wisconsin Illinois SAS® Users
Spring Conference

Woman's Club
Milwaukee, *Wisconsin*

June 26, 2013

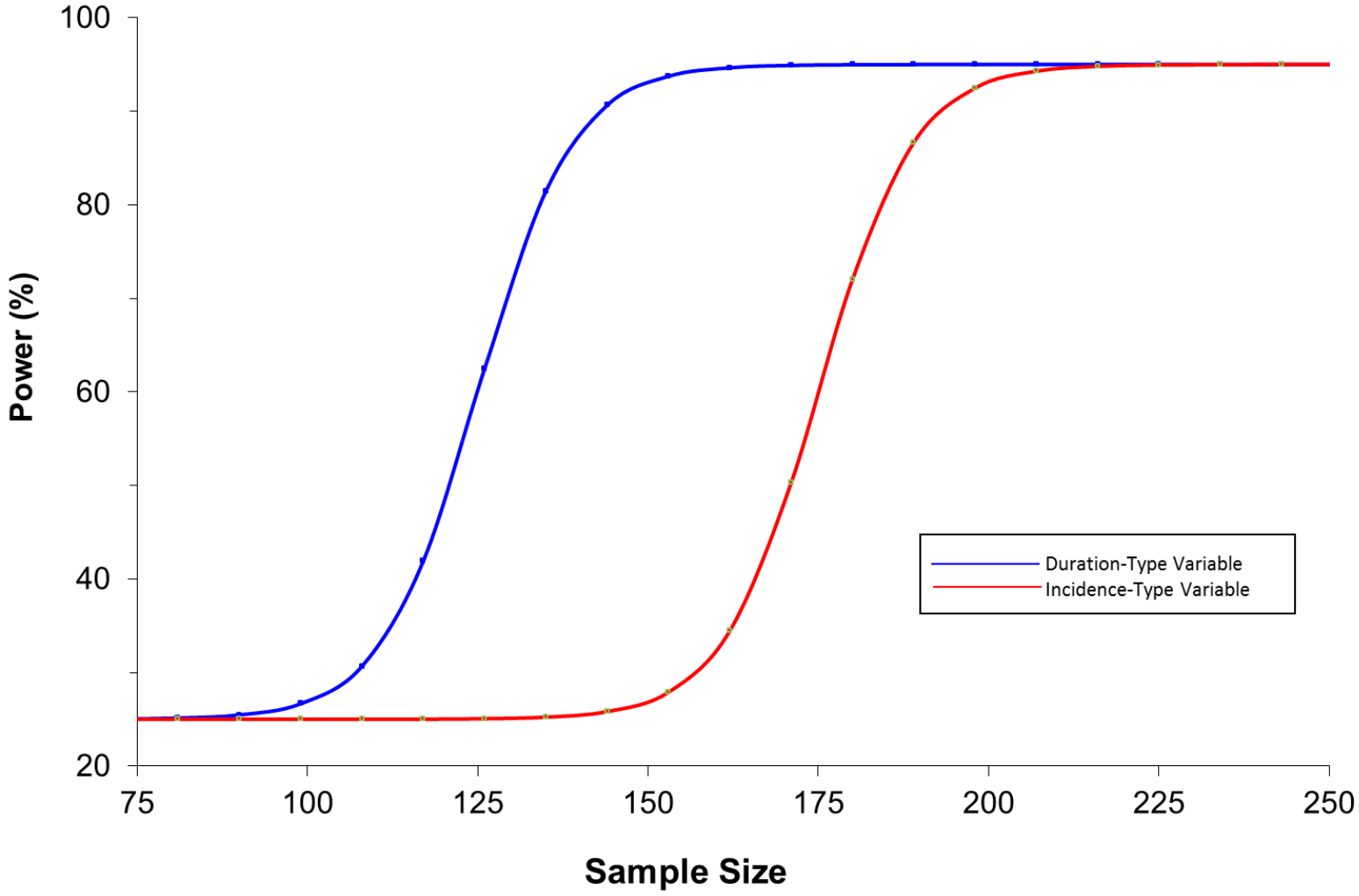
An Introduction to the Analysis of
Failure-Time Data in Industrial Settings

Michael G. Wilson

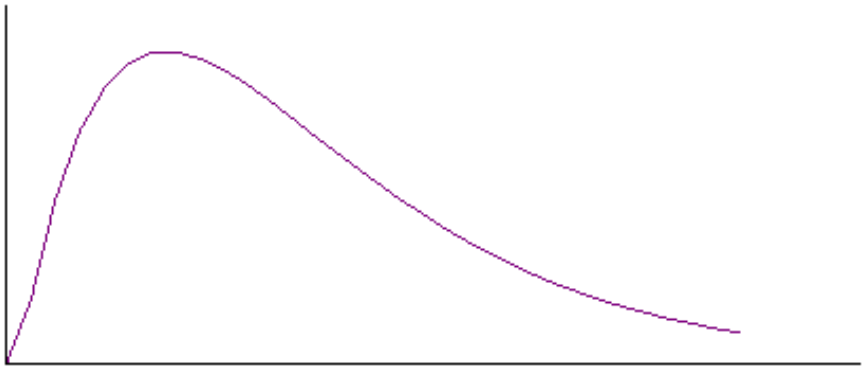
The Nature of Duration Data

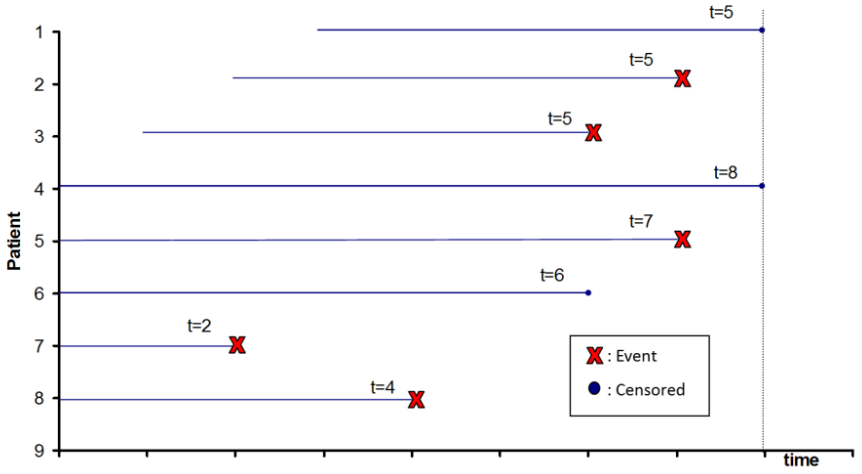
- Is structured as the time to some well-defined endpoint event
- Economic Duration data measure the time to some event.
  - Unemployment benefits stop
  - Mutual Fund fails
  - Recession ends
  - Length of a Union Strike
- Medical Duration Data is often called Survival Data
  - Time to Death (Survival)
  - Time to an Adverse Event
  - Time to Tumor Remission

Advantage of Duration Data

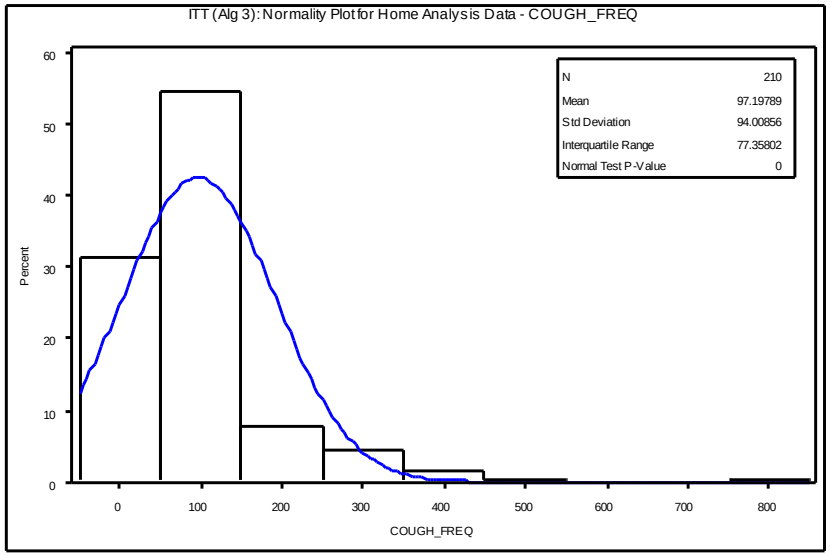


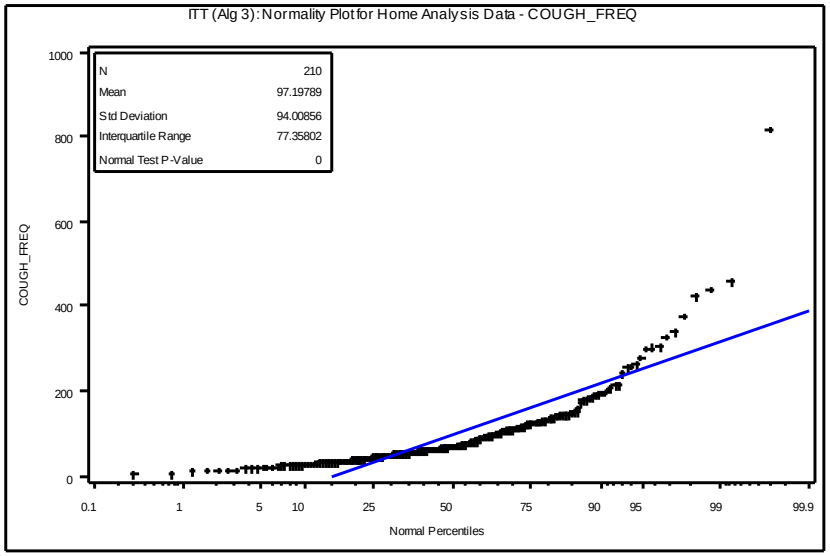
Disadvantage of Duration Data is that it is difficult to analyze

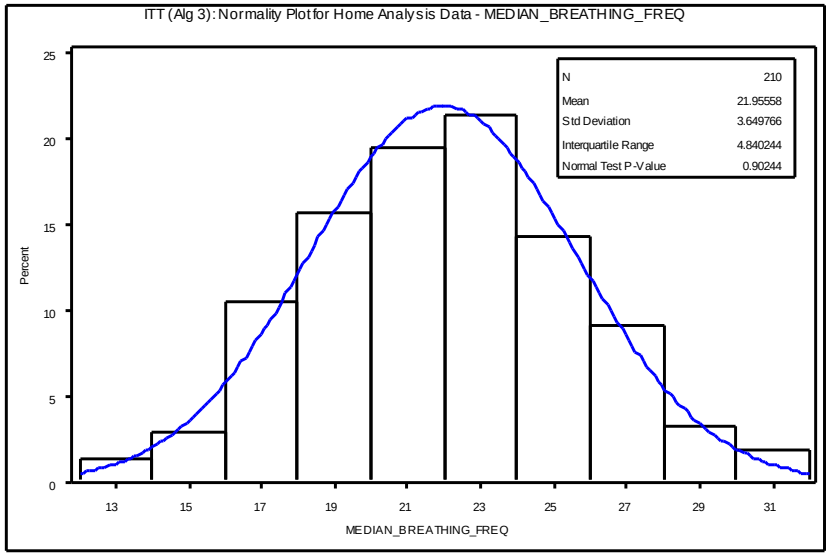


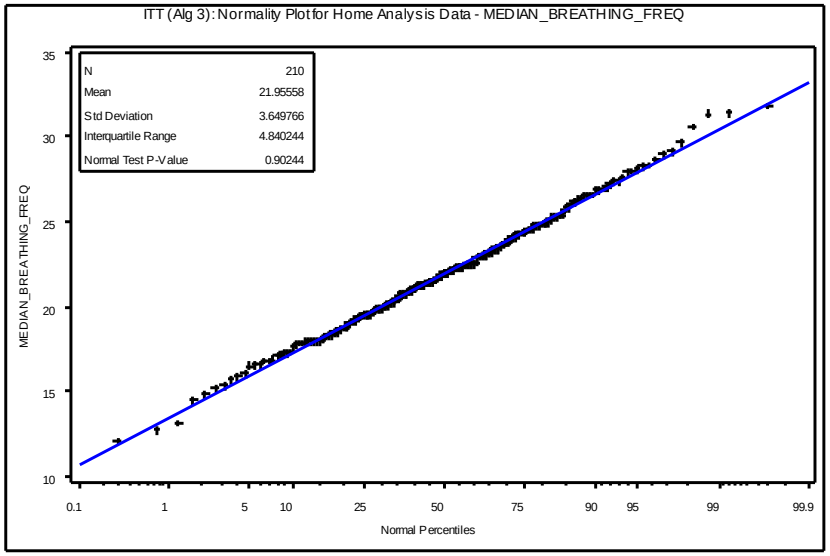


QQ Plots Show Non-normality





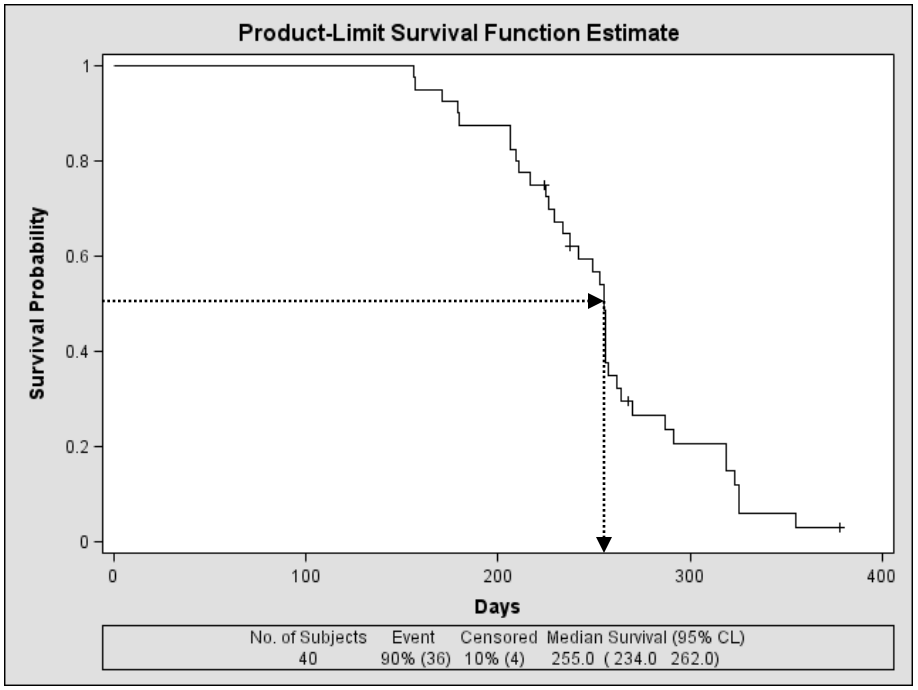




Outline of Selected Analysis Methods for Duration Curves

- The Product-Limit Method
- Accelerated Failure Time Models
- Proportional Hazards Regression

The Survival (Duration) Distribution



- S(t) is the probability that an event time is greater than or equal to t, where t can be any non-negative number.
- $S(t) = P[\ T \geq t\ ] = 1 - P[\ T < t\ ]$
  $= 1 - F(t)$
- $S(0) = 1.0$
- $S(\infty) = 0.0$
- $S(255) = 0.5$

The Probability Density Function

- The probability density function (pdf) is the first derivative of the cdf, F(y)
  - Slope of the cdf and the negative slope of Survival
  - d/dt [F(y)] = d/dt [1- S(y)] = -d/dt[S(y)]
- Probability of failure over a very small time interval

$$f(y) = \lim_{\Delta y \to 0} P(\text{a subject fails between y and y } + \Delta\text{y}) / \Delta\text{y}$$

$$f(y) = \lim_{\Delta y \to 0} P(y < Y < y + \Delta y) / \Delta y$$

- The probability density function is an unconditional probability

Hazard Function

- The hazard function is a conditional density
- The probability that a subject fails over the next instant conditioned that the subject has survived up to the beginning of the interval.

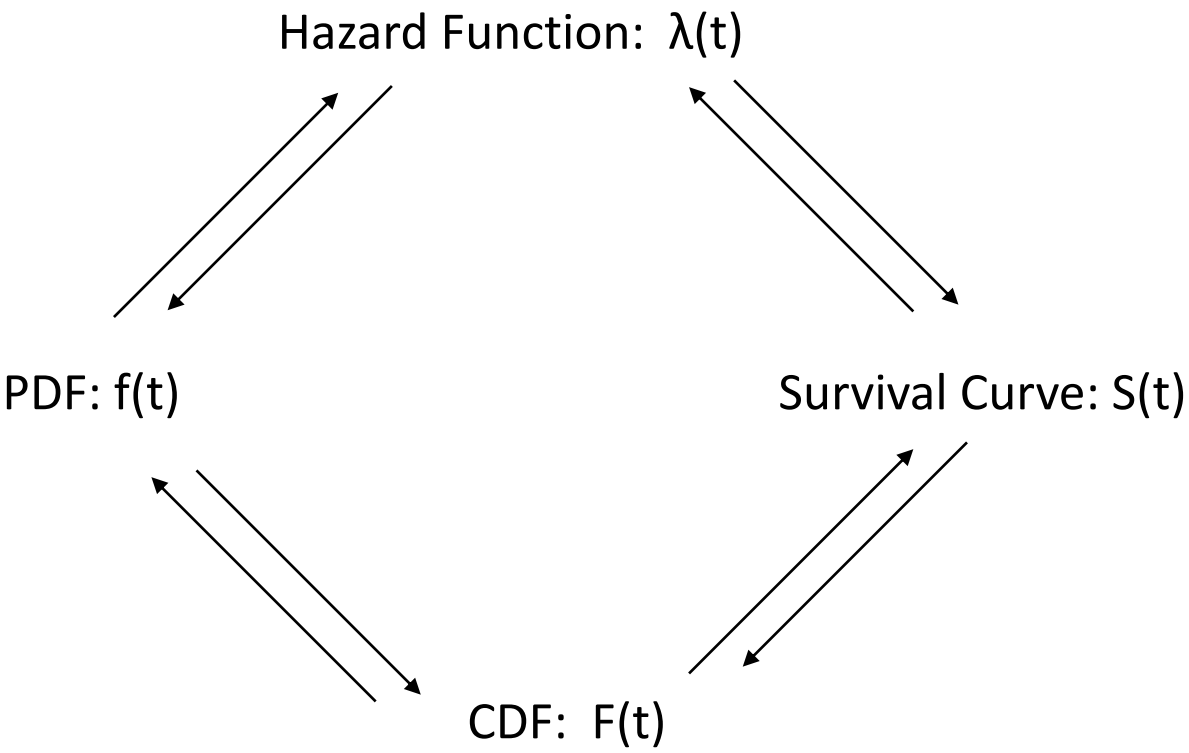
$$h(y) = \lim_{\Delta y \to 0} P(\text{a subject fails between y and } \Delta y \mid \text{the subject survives to y}) / \Delta y.$$

$$= \lim_{\Delta y \to 0} P(y < Y \leq y + \Delta y | Y > y) / \Delta y$$

$$= \frac{f(y)}{1 - F(y)}$$

$$= \frac{f(y)}{S(y)}$$

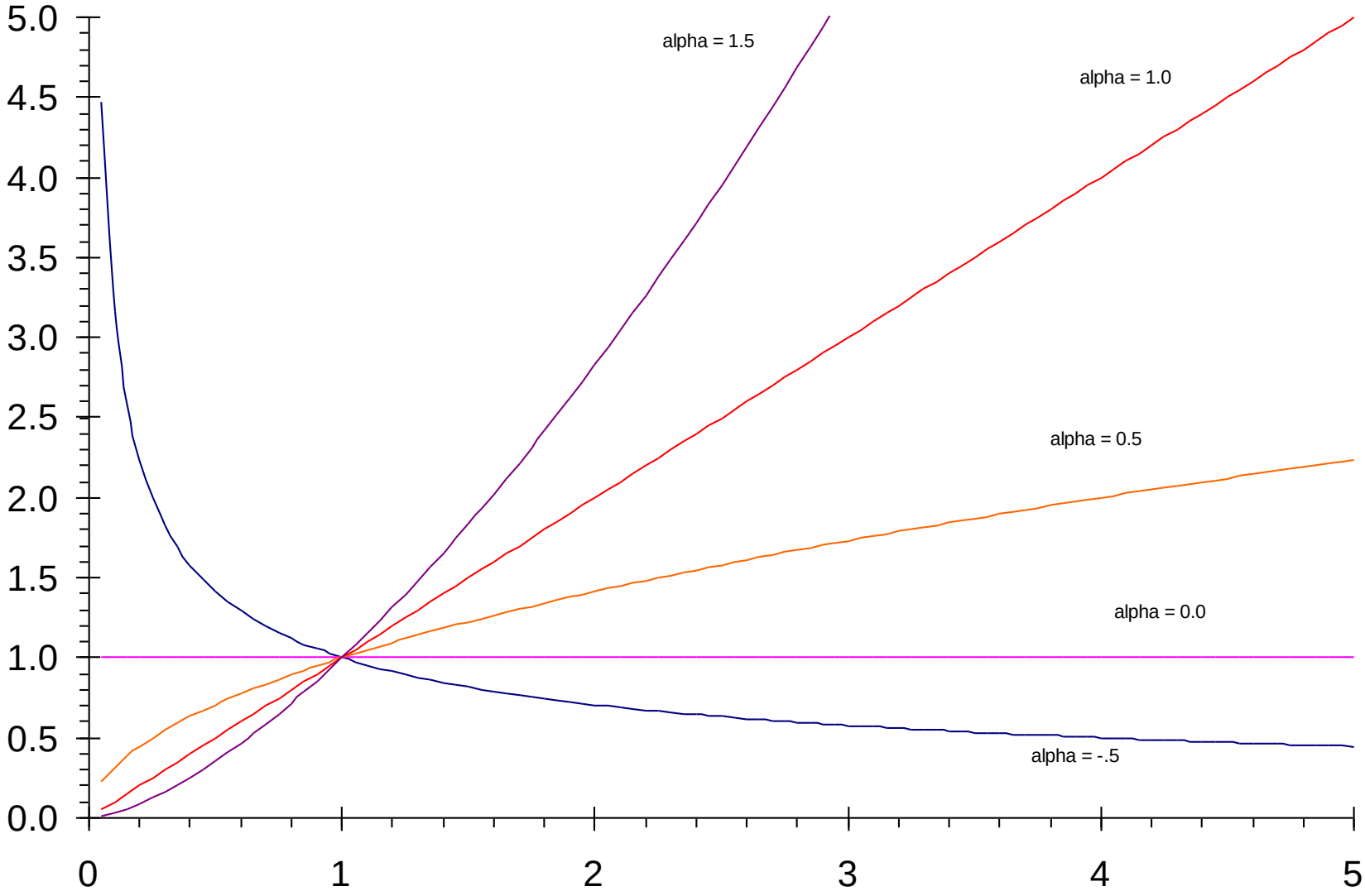
The Hazard Function is Central to Describing the Probability Distributions of Duration Data



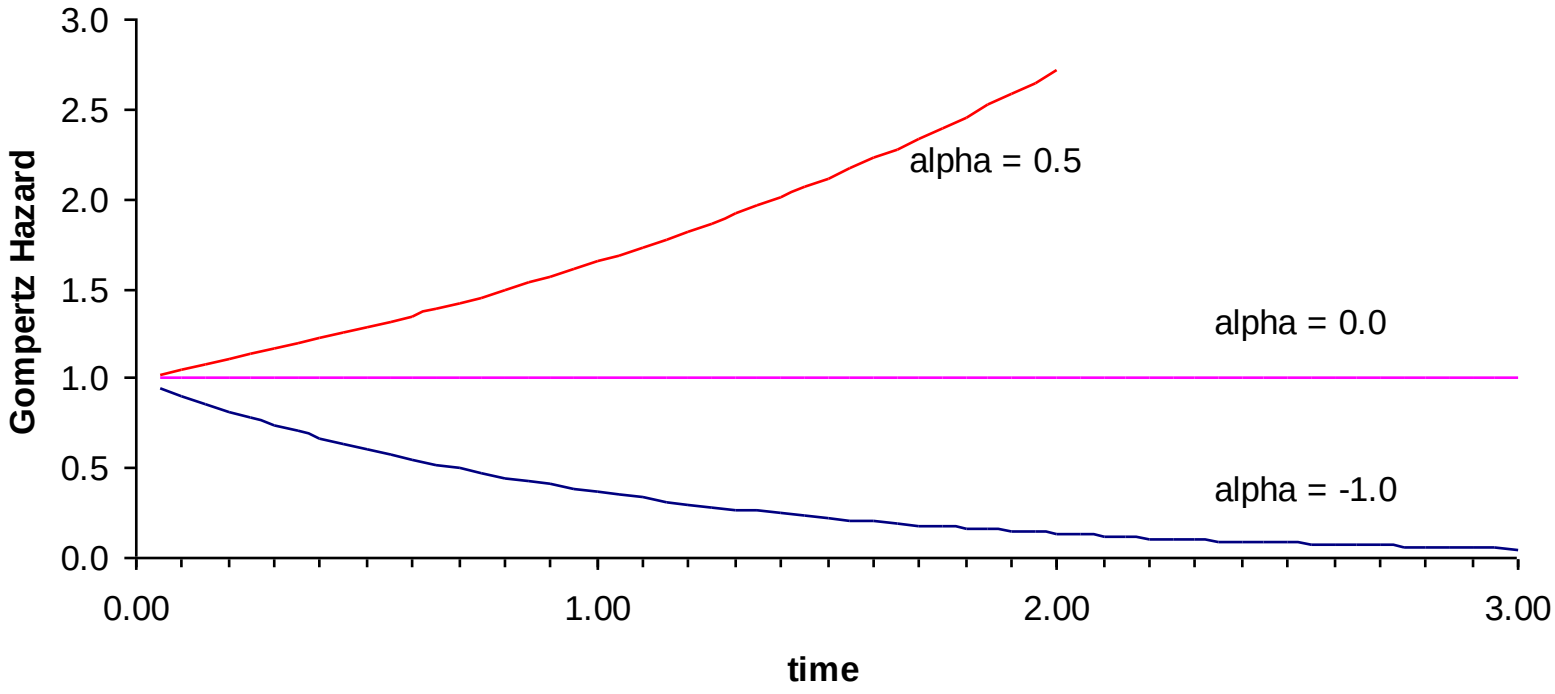
Selected Simple Hazard Families

Hazard	Form of Log Hazard	Log Hazard	Family
$h(t) = \lambda$	Constant	$\log h(t) = \mu$	Exponential
$h(t) = \lambda\gamma^t$	Linear	$\log h(t) = \mu + \alpha t$	Gompertz
$h(t) = \lambda t^\alpha$	Log Linear	$\log h(t) = \mu + \alpha \log(t)$	Weibull

Weibull Family of Hazard Densities



Gompertz Family of Hazard Densities



- Unlike the Weibull, the Gompertz intercept is lambda, which in some situations is more realistic than zero or infinity
- Gompertz is not readily available in SAS

Proportional Hazard Functions

- The Exponential, Weibull and Gompertz models are all members of a general class known as proportional hazards models.
- Two hazard functions $h_1(y)$ and $h_2(y)$ are said to be proportional iff:

$$h_1(y) = \lambda\, h_2(y) \text{ , for all } y > 0,$$

where λ is constant.

- It can be shown, $S_1(y) = [S_2(y)]^{\lambda}$

Product-Limit Estimates

- Formally the Product Limit nonparametric estimation of the survival function at time y is given by:

$$S(y) = \prod_{y_{(k)} < y} \left(1 - \frac{d_k}{n_k}\right).$$

- Multiply together a series of conditional probabilities
- Nonparametric

Computing the Product Limit Estimates

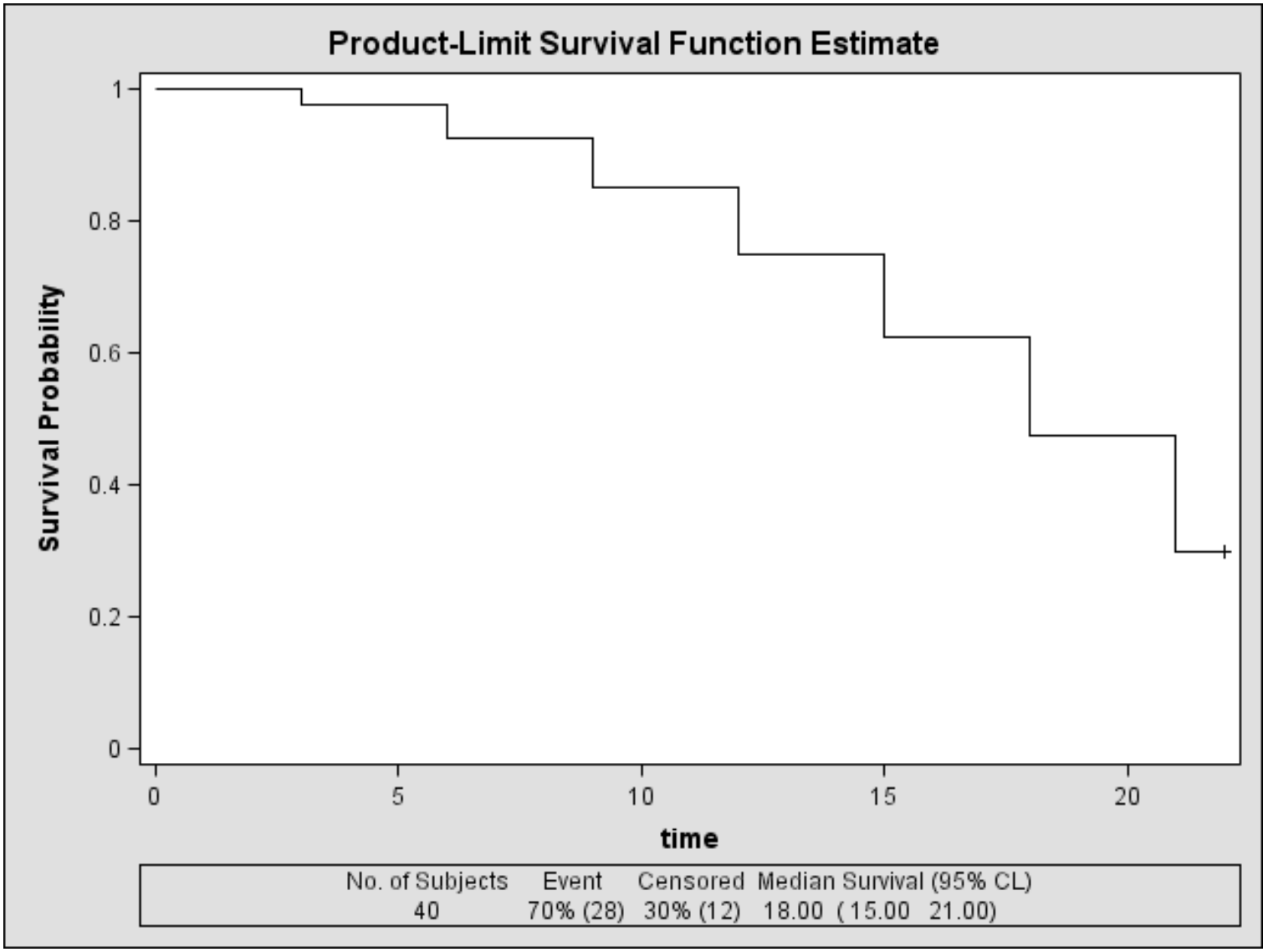
- In SAS, only a PROC LIFETEST and TIME statements are required:

```
proc lifetest;
     time months;
     run;
```

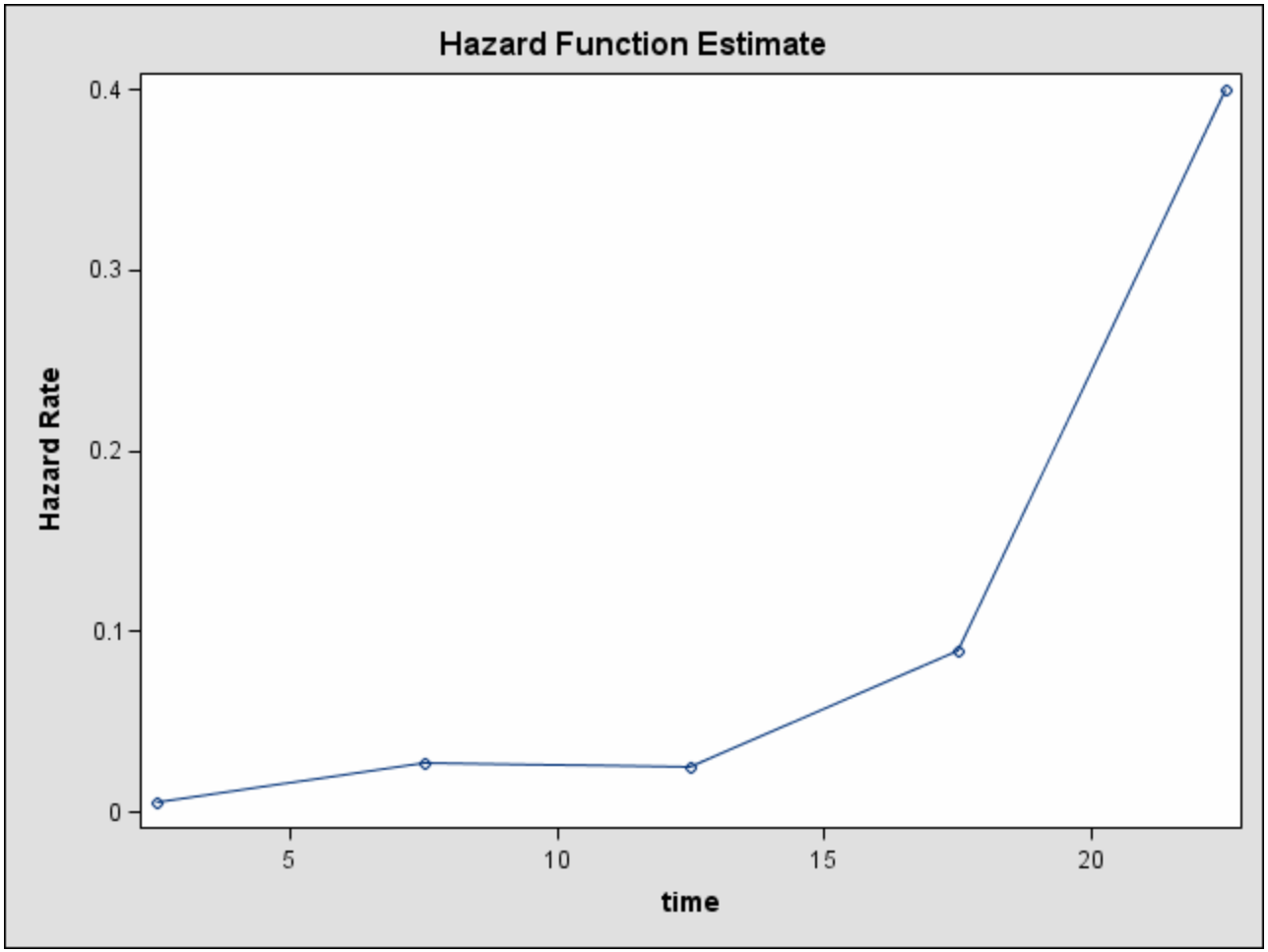
Product-Limit Large Sample Confidence Intervals

$$\left[\hat{S(y)}\right] \pm z(\alpha/2)\sqrt{\text{var}\left[\hat{S(y)}\right]}$$

Empirical Survival Distribution is a Step Function of PL Estimates



Empirical Hazard Function for Mutual Fund Example



Use the LT option to get hazard plots.

Experimental Cancer Treatment Study

- Consider the results of a small randomized cancer trial.
- 40 cases have been randomly assigned to two treatment groups
  - Experimental Drug + Standard
  - Standard Therapy Only
- The response is the time from randomization to death.
- Interested in whether the survival distributions differ between the two treatments.
- The data set Exposed contains four variables:

DAYS	Follow-up time from randomization to death
STATUS	Event indicator with value 1 for death time and value 0 for censored time
TREATMENT	Treatment indicator, 1 if experimental, 0 if standard
SEX	M and F

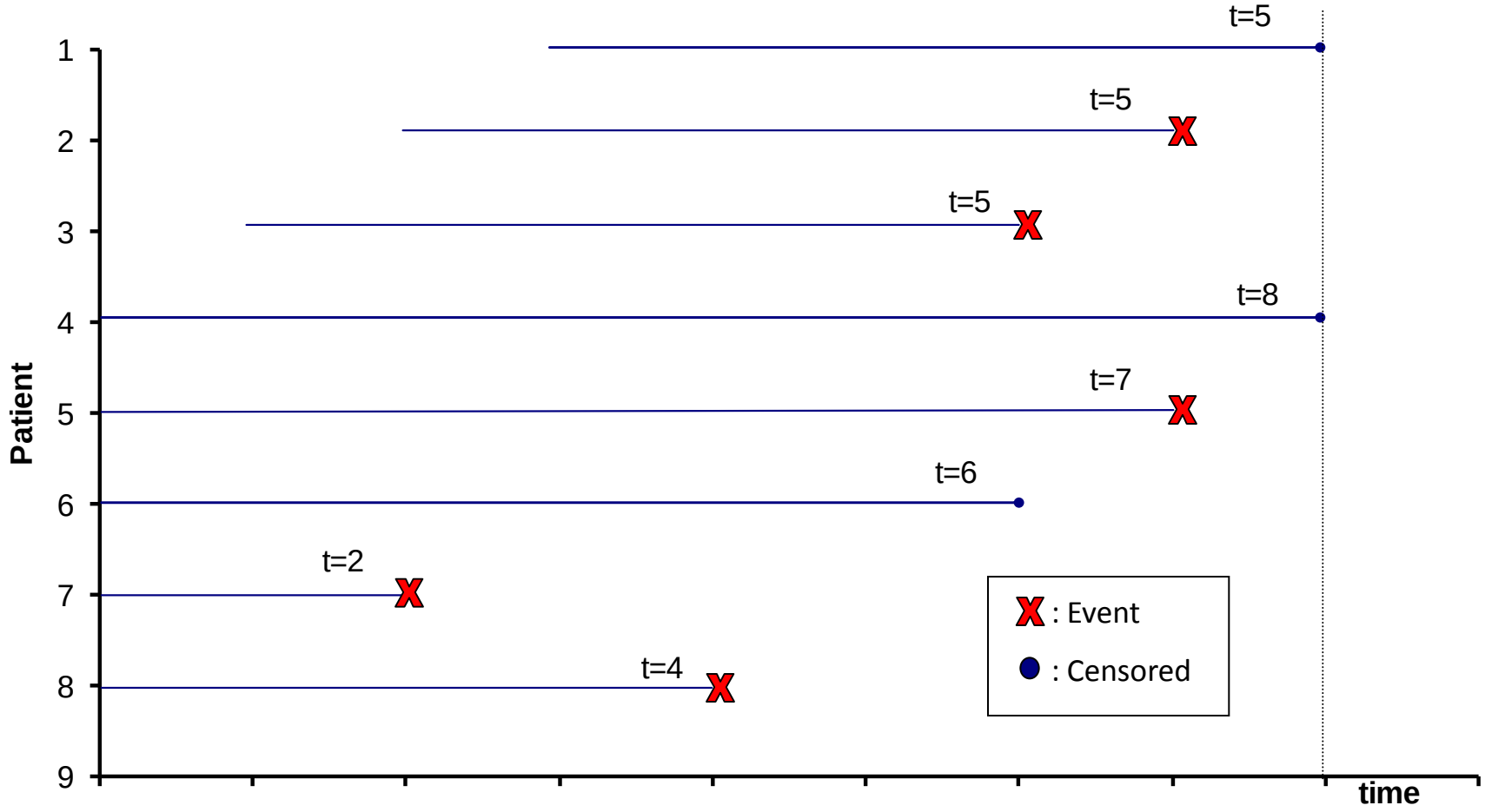
Experimental Cancer Treatment Results

Experimental		Standard	
Days	Sex	Days	Sex
179	F	237*	F
256	F	156	F
262	M	270	M
256	F	257	M
255	M	242	M
224*	F	157	F
225	F	249	M
287	M	180	F
319	M	226	F
264	M	268*	M
378*	M	291	M
355	M	323	M
319	M	253	M
256	M	206	F
171	F	206	F
325	M	237	M
325	M	211	F
217	F	229	F
255	F	234	F
256	F	209	F
Mean = 269.2		Mean = 232.5	

Types of Censoring

Directional	Right	Study stopped before the event occurs
	Left	Event occurred before the study started
	Interval	Event occurs between visits, a < T < b
Fixed	Type I	Under the control of the investigator When the study conclusion is scheduled; Fixed censoring time
	Type II	Fixed number of events to occur 50 patients out of 100 have an event
	Single	One censoring time
	Multiple	If two or more censoring times e.g. 3 and 5 years
Random		Lost to Follow-up; competing risk These occur before the fixed censoring
	Non-informative	Those censored at time, t, should be representative of all subjects with the same covariates, still alive at t None of the covariates predict censoring
	Informative	Censoring mechanism and Event related

Censoring Pattern in Calendar Time



Non-Informative Censoring

- All standard Survival Analysis Methods assume that censoring is random and non-informative
  - Those censored at time, t, should be representative of all subjects with the same covariates, still alive at t (Cox and Oakes, 1984)
  - In most situations adding entry time as a covariate in the model will help to adjust for the dependency
- The non-informative assumption cannot be checked by any statistical test
  - A crude sensitivity test is offered by assuming best and worst case scenarios (Allison, chapter 8)
  - Otherwise,
    - Do everything you can to reduce drop-out censoring
    - Measure covariates likely to affect censoring
    - Cautiously interpret tests and confidence intervals
- There are no SAS procedures for Informative Censoring

PL Estimates with Non-Informative Right Censoring

- Censored data are paired Time to Occurrence
  - In this example, the time variable is DAYS and
  - the censoring variable is STATUS
  - With value 0 indicating censored observations
- Data for the time to occurrence of a pre-defined event obtained from n subjects of a clinical trial can be arranged as $(y_1,c_1),\ldots,(y_n,c_n)$
  - where y_i is the observed time for subject i, and,

$$c_i = \begin{Bmatrix} 1, \text{ if } y_i \text{ is the survival time,} \\ 0, \text{ if } y_i \text{ is censored.} \end{Bmatrix}$$

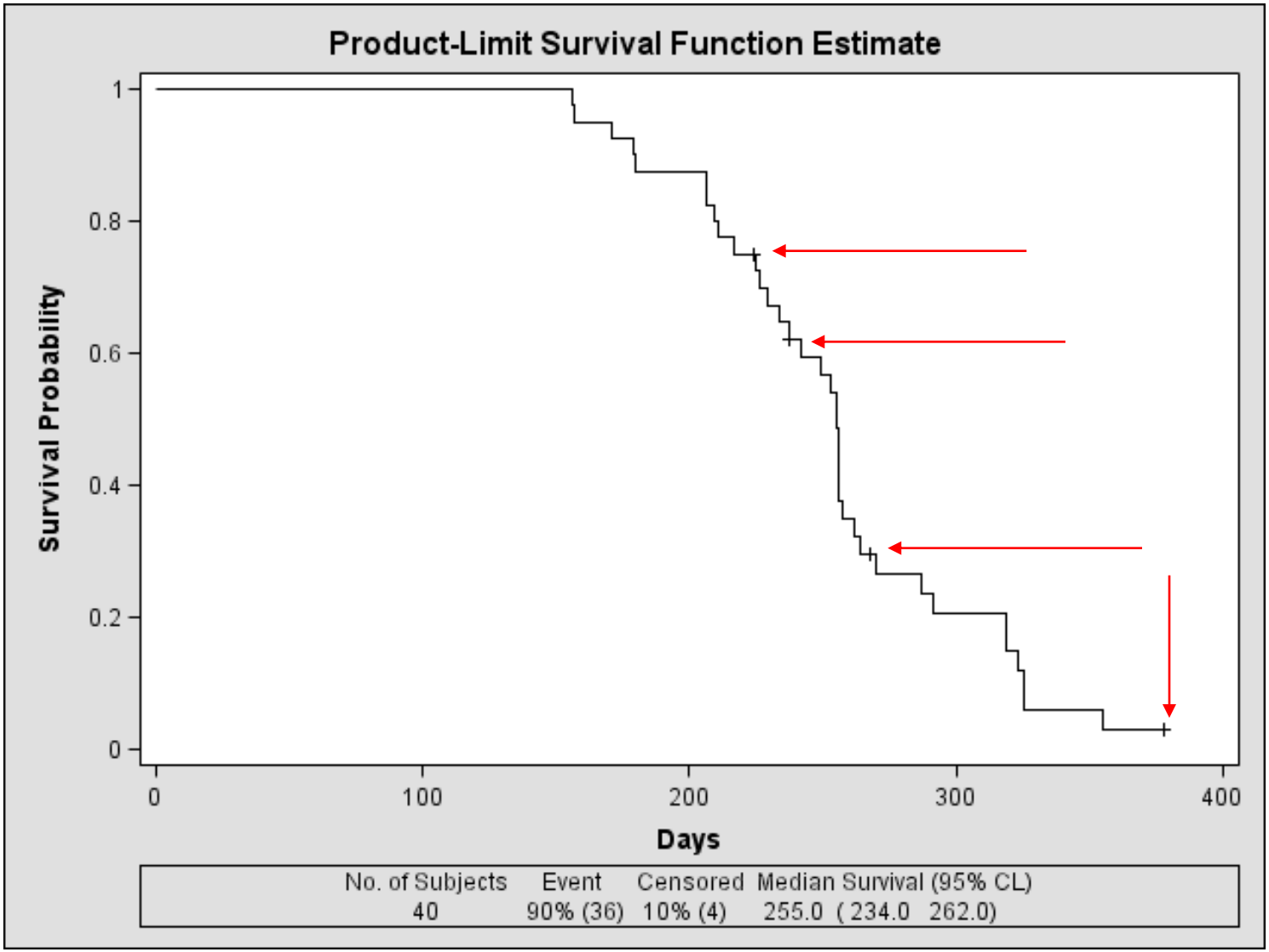
- The following statements compute the product-limit estimate for the censored data:

```
proc lifetest;
     time days*status(0);
     run;
```

Product-Limit Survival Estimates

Days	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.000	1.0000	0	0	0	40
156.000	0.9750	0.0250	0.0247	1	39
157.000	0.9500	0.0500	0.0345	2	38
171.000	0.9250	0.0750	0.0416	3	37
179.000	0.9000	0.1000	0.0474	4	36
180.000	0.8750	0.1250	0.0523	5	35
206.000	0.8250	0.1750	0.0601	7T	33
209.000	0.8000	0.2000	0.0632	8	32
211.000	0.7750	0.2250	0.0660	9	31
217.000	0.7500	0.2500	0.0685	10	30
224.000*				10	29
225.000	0.7241	0.2759	0.0708	11	28
. . .	. . .	. . .	. . .	.	.
237.000	0.6207	0.3793	0.0773	15	24
237.000*				15	23
242.000	0.5937	0.4063	0.0785	16	22
249.000	0.5667	0.4333	0.0795	17	21
253.000	0.5397	0.4603	0.0801	18	20
255.000	0.4858	0.5142	0.0807	20T	18
256.000	0.3778	0.6222	0.0788	24TTT	14
257.000	0.3508	0.6492	0.0776	25	13
262.000	0.3238	0.6762	0.0762	26	12
264.000	0.2969	0.7031	0.0745	27	11
268.000*				27	10
270.000	0.2672	0.7328	0.0727	28	9
287.000	0.2375	0.7625	0.0704	29	8
319.000	0.1484	0.8516	0.0599	32T	5
323.000	0.1187	0.8813	0.0548	33	4
325.000	0.0594	0.9406	0.0404	35T	2
355.000	0.0297	0.9703	0.0291	36	1
378.000*	0.0297			36	0

Survival Function for Experimental Cancer Treatment Data



Comparison between Survival Functions

- In the Experiment Cancer Treatment Study, we're interested in whether the survival distributions differ between the two treatments.
- The statistical hypotheses to determine whether the experimental therapy can improve survival probability compared to standard treatment are typically expressed as:
- Null H_o: $S_1 = S_2$ versus
- Alternative H_a: $S_1 \neq S_2$

Comparison between Survival Functions

- One commonly used method for comparing two survival functions is the log-rank test.
- The log-rank test is known to be the most powerful test for the alternative hypothesis that hazard function are proportional.
- Null H_o: $S_1(y) = S_2(y)$ versus
- Alternative H_a: $S_1(y) = [S_2(y)]^{\lambda}$
- The log-rank test for comparing two independent samples is the same as the Mantel-Haenszel statistic for combining results from different strata.
- Under the assumption of hypergeometric distribution, the conditional expected number of patients with occurrence of events for d_{1k} at the event time $y_{(k)}$ for the test drug is given by:

$$e_{1k} = \frac{n_{1k} d_{1k}}{n_k}$$

Comparing Survival Curves

- In the Experiment Cancer Treatment Study, we're interested in whether the survival distributions differ between the two treatments

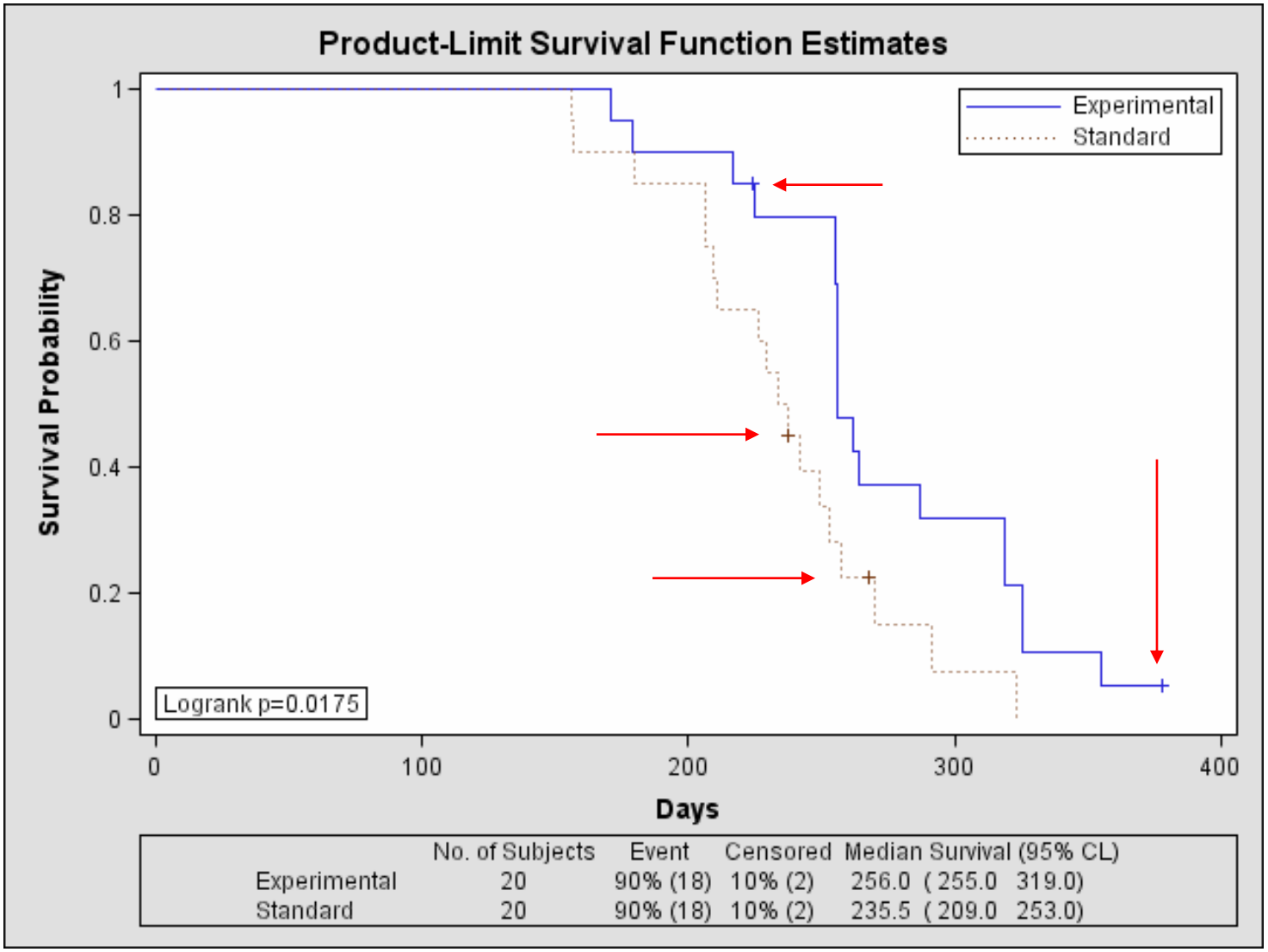
```
proc lifetest;
     time Days*Status(0);
     strata Treatment;
     run;
```

Comparison between Survival Functions

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	5.6485	1	0.0175
Wilcoxon	5.0312	1	0.0249
-2Log(LR)	0.1983	1	0.6561

PL Estimates for Experimental Cancer Study



Comparison between Survival Functions

- The log-rank statistic is the same as the Mantel-Haenszel:

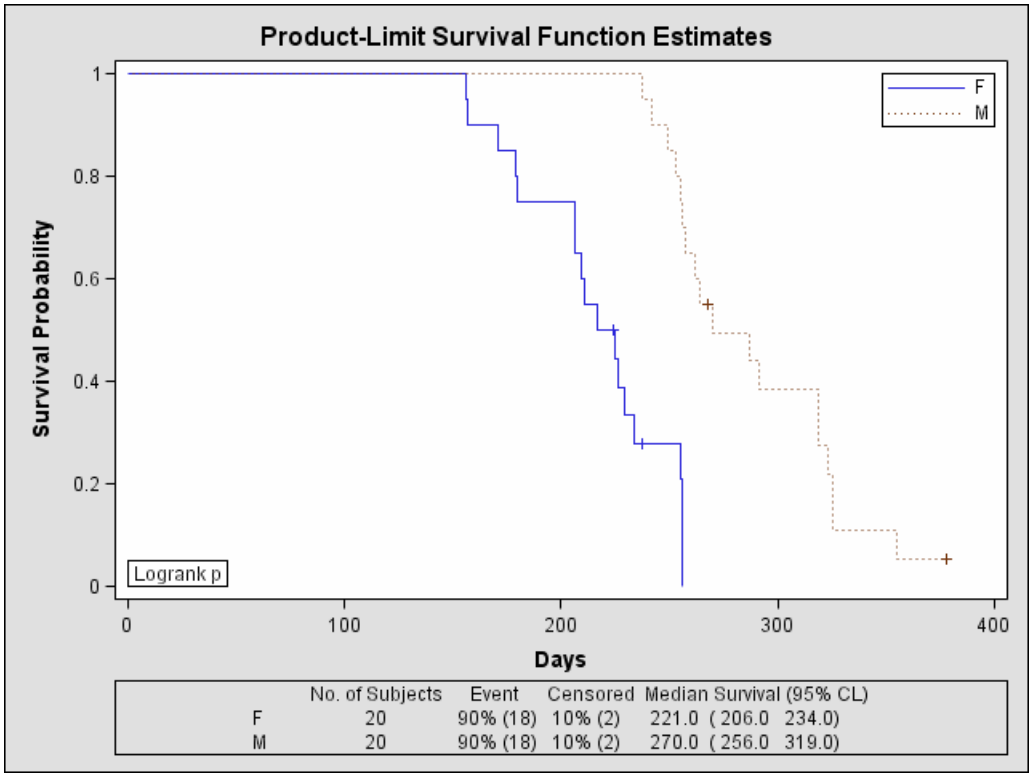
$$X_{LR} = \frac{\sum_{k=1}^{K} (d_{1k} - e_{1k})^2}{\sum_{k=1}^{K} V_{1k}}$$

- Under the null hypothesis of equal survival functions, the log-rank statistic approximately follows a central chi-square distribution with 1 degree of freedom when sample size is moderate.
- Hence, the null hypothesis is rejected at the αth significance level if:

$$X_{LR} > X^2(\alpha, 1)$$

Stratified Log-Rank Tests

- We also plotted the survival curves by gender
- A separation in survival curves was observed



- So we wanted to test the effect of treatment stratified for gender

Tests for Treatment Effect Adjusting for Gender

- SAS Code

```
proc lifetest;
     time Days*Status(0);
     strata sex / group = Treatment;
     run;
```

- SAS Output

Stratified Test of Equality over Group

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	7.2466	1	0.0071
Wilcoxon	5.9179	1	0.0150

Outline of Selected Analysis Methods for Duration Curves

- The Product-Limit Method
- Accelerated Failure Time Models
- Proportional Hazards Regression

Accelerated Failure Time Models

- Models
  - The estimates of parametric regression
    - Weibull
    - Exponential
    - Log-Normal
  - for censored survival data
  - Uses the method of Maximum likelihood
- $S_i(t) = S_j(\Phi_{ij}t)$, for all t
- Advantages
  - Accommodates left and interval censoring
  - Test shape of the hazard function
  - More efficient (smaller standard errors)
- Parametric methods for estimation of survival time can be found in
  - Lee (1992) and
  - Marubini and Valsecchi (1995)

Recidivism Dataset

- N = 432 inmates from Maryland Penitentiary
- Censoring Time = 1 year
- Dates of arrest recorded
- Does parole lengthen the time to next arrest?

Fitting AFT Models

- SAS Code

```
proc lifereg data = recid02;
     class educ01;
     model week*arrest(0)=fin age race wexp mar paro prio
                          educ01 / dist = weibull covb;
                          run;
```

  - Class statement creates indicator variables for polytomous covariates
  - Rich array of survival distributions including the
    - generalized gamma distribution

Recidivism Model with Education as a CLASS Variable

Analysis of Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	4.4680	0.5171	3.4544	5.4816	74.65	<.0001
fin		1	0.2690	0.1379	-0.0012	0.5392	3.81	0.0510
age		1	0.0392	0.0159	0.0079	0.0705	6.04	0.0140
race		1	-0.2524	0.2229	-0.6893	0.1845	1.28	0.2575
wexp		1	0.0773	0.1522	-0.2209	0.3755	0.26	0.6114
mar		1	0.3013	0.2732	-0.2342	0.8368	1.22	0.2701
paro		1	0.0658	0.1396	-0.2078	0.3394	0.22	0.6374
prio		1	-0.0585	0.0213	-0.1004	-0.0167	7.53	0.0061
educ01	3	1	-0.5116	0.3090	-1.1172	0.0941	2.74	0.0978
educ01	4	1	-0.3536	0.3243	-0.9892	0.2819	1.19	0.2755
educ01	5	0	0.0000					

Outline of Selected Analysis Methods for Duration Curves

- The Product-Limit Method
- Accelerated Failure Time Models
- Proportional Hazards Regression

Primary Biliary Cirrhosis Dataset

- In their 1991 book, Fleming and Harrington published a survival analysis dataset from the Mayo clinic involving primary biliary cirrhosis. The Mayo PBC Model published by Dickson (1989) has been extremely important in liver disease research
- The data consist of 418 patients among which 161 had died
- A subset of variables are:

TIME	Follow-up time, in years
STATUS	Event indicator with value 1 for death time and value 0 for censored time
AGE	Age in years from birth to study registration
ALBUMIN	Serum albumin level in gm/dl
BILIRUBIN	Serum bilirubin level in mg/dl
EDEMA	Edema presence, ordinal 0.0, 0.5, 1.0
PROTIME	Prothrombin time in seconds

Proportional Hazards Regression
Model Specification

- $\boldsymbol{h}(t;\boldsymbol{x}) = \boldsymbol{h_o}(t)\exp\{\beta_1 x_1 + \ldots + \beta_k x_k\}$
- where:
  - $\boldsymbol{h}(t;\boldsymbol{x})$ is the hazard function at time t for a subject with covariates values $x_1,\ldots,x_k$
  - $\boldsymbol{h_o}(t)$ is the baseline hazard function, that is, when all covariates equal zero
  - exp is the exponential function: $\exp(x)=e^x$
  - x_i is the i^{th} covariate in the model and
  - β_i is the regression coefficient for the i^{th} covariate x_i

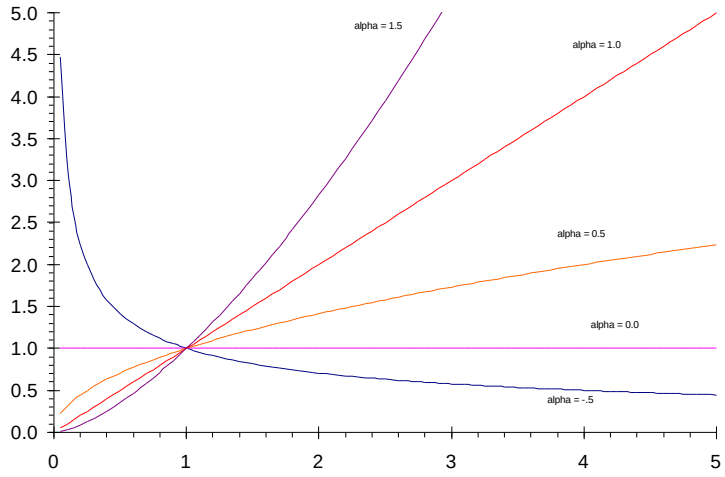
Proportional Hazards Regression
Remarks

$$\boldsymbol{h}(t;\boldsymbol{x}) = \boldsymbol{h_o}(t)\exp\{\beta_1 x_1 + \ldots + \beta_k x_k\}$$

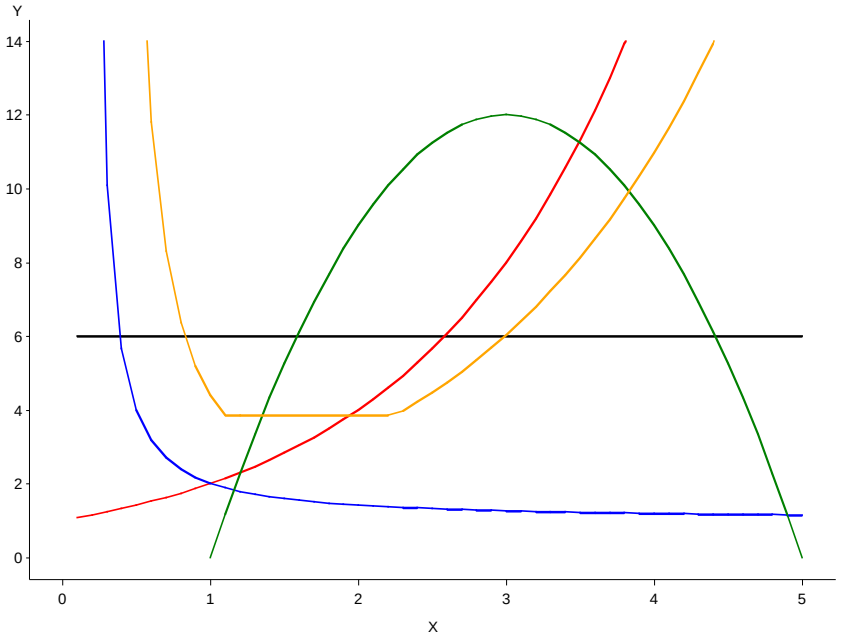
- Proportional hazards regression predicts the hazard function, not 'Y' as in ordinary regression
- Baseline hazard, $\boldsymbol{h_o}(t)$, can take any functional form with a positive value
- The exponential function of the covariates will be positive
- The model is multiplicative. There is no intercept; it is absorbed into the baseline function
- The log-transform model is additive.
  - $\ln[\boldsymbol{h}(t;\boldsymbol{x})] = \ln[\boldsymbol{h_o}(t)] + \{\beta_1 x_1 + \ldots + \beta_k x_k\}$

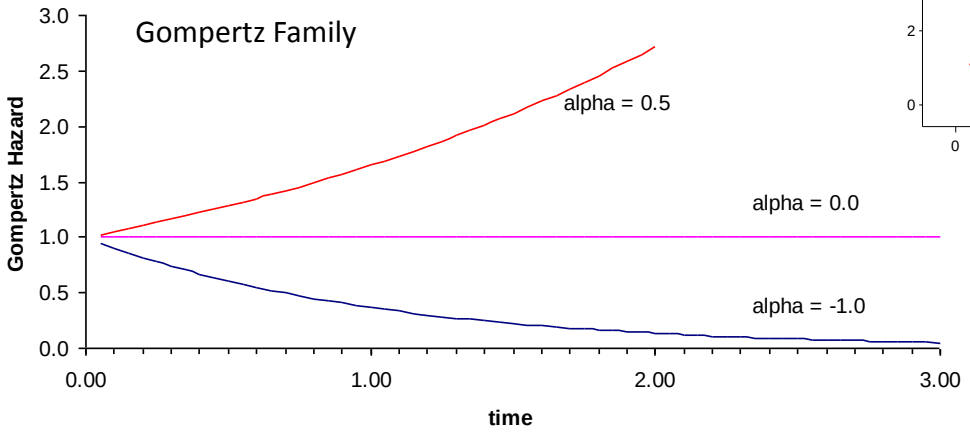
Examples of Baseline Hazards $h_o(t)$

Weibull Family



Unspecified Shapes





Partial Likelihood

$$PL = \prod_{i \in events} \frac{h(t_i)}{\sum_{j \in R_i} h(t_j)} = \prod_{i \in events} \frac{h_0(t_i)\exp\{\beta x_1 + \ldots + \beta x_k\}}{\sum_{j \in R_i} h_0(t_j)\exp\{\beta x_1 + \ldots + \beta x_k\}}$$

- Where
  - t_i is the i^{th} death time,
  - x_i are the covariates,
  - R_i is the risk set at time t_i (*i.e.* still alive and uncensored)

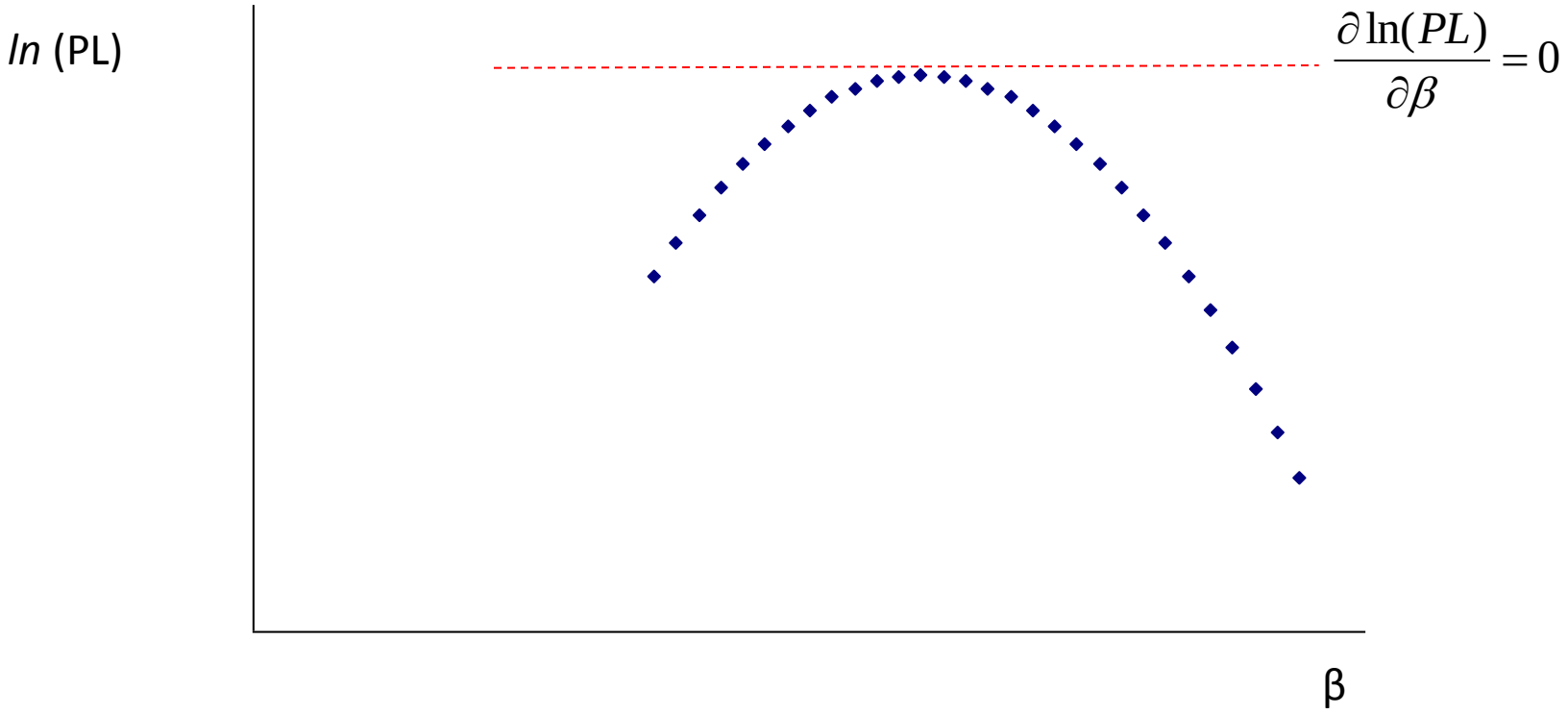
- The numerator is the hazard of death for the subject who died at time t_i
- The denominator is the sum of hazards of death for all subjects in the risk set at time t_i
- The ratio reflects the likelihood of death for subject *i*

Partial Likelihood Estimation of the PH Model

$$PL = \prod_{i \in events} \frac{h(t_i)}{\sum_{j \in R_i} h(t_j)} = \prod_{i \in events} \frac{h_0(t_i)\exp\{\beta x_1 + \ldots + \beta x_k\}}{\sum_{j \in R_i} h_0(t_j)\exp\{\beta x_1 + \ldots + \beta x_k\}}$$

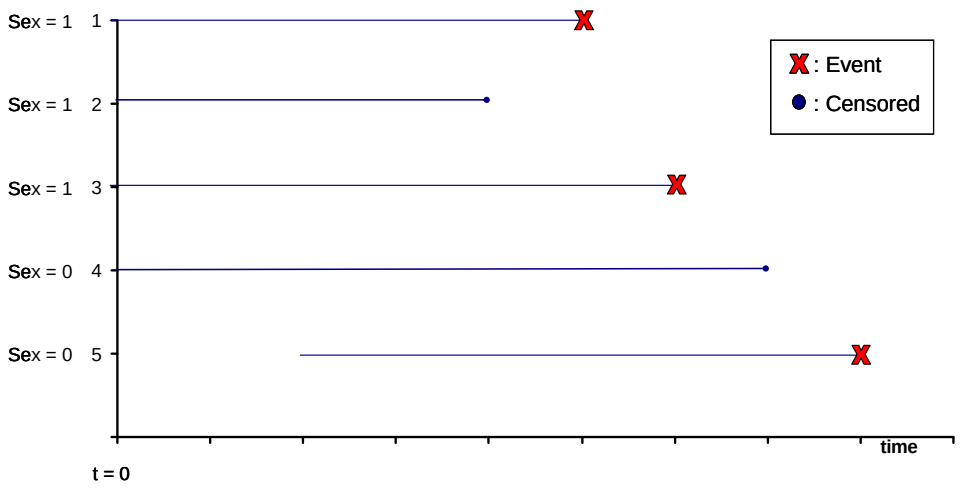
- The likelihood function can be factored into two parts: the dependent on $\boldsymbol{h_o}(t)$ and the part dependent on the betas, $\beta_1, \ldots, \beta_k$.
- Partial likelihood uses only the betas, $\beta_1, \ldots, \beta_k$, ignoring the baseline hazard function, $\boldsymbol{h_o}(t)$.
- Some information is lost, so the standard errors are slightly increased
- Because the form of the baseline hazard function, $\boldsymbol{h_o}(t)$, is not specified some robustness in the betas, $\beta_1, \ldots, \beta_k$, is gained
- The estimates of the betas, $\beta_1, \ldots, \beta_k$, are still, asymptotically unbiased (consistent) and asymptotically normally distributed

Method of Partial Likelihood Estimation

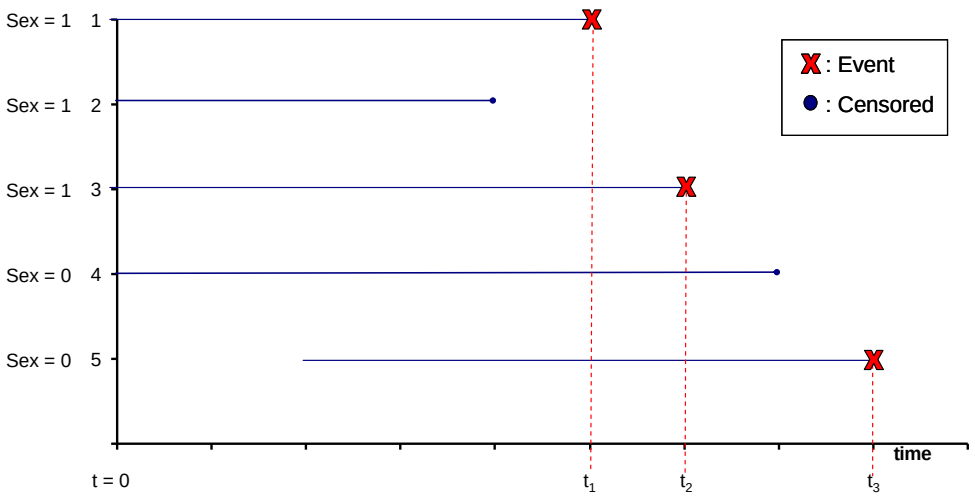


Simple PL Example
One Covariate, Five Observations

$$PL = \prod_{i \in events} \frac{h(t_i)}{\sum_{j \in R_i} h(t_j)} = \prod_{i \in events} \frac{h_0(t_i)\exp\{\beta x_i\}}{\sum_{j \in R_i} h_0(t_j)\exp\{\beta x_j\}}$$



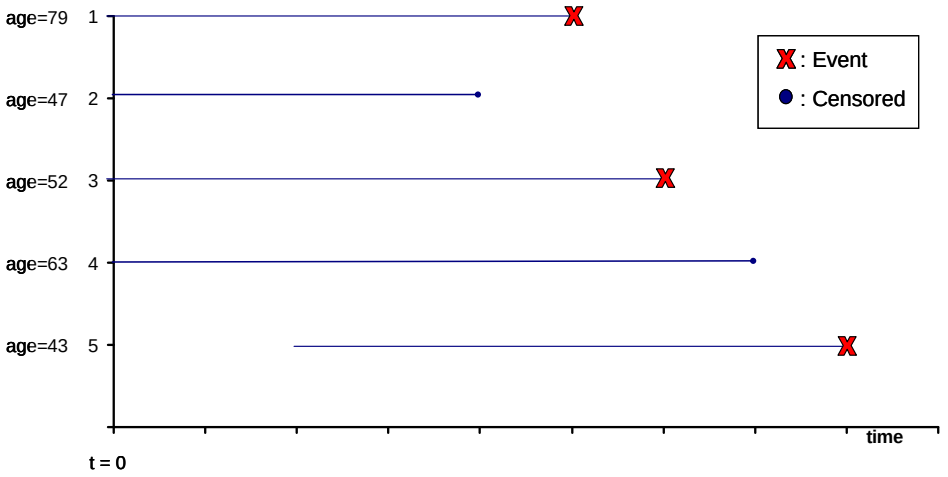
Simple PL Example
One Dichotomous Covariate, Five Observations



$$PL = \frac{e^{1\beta}}{e^{1\beta} + e^{1\beta} + e^{0\beta} + e^{0\beta}} \times \frac{e^{1\beta}}{e^{1\beta} + e^{0\beta} + e^{0\beta}} \times \frac{e^{0\beta}}{e^{0\beta}}$$

- This example shows a special case.
- If β=0, PL = 1/4 x 1/3 x 1
- PL increases as β increases
- The value of β that maximizes the PL is positive infinity

Simple PL Example
One Continuous Covariate, Five Observations



$$PL = \frac{e^{79\beta}}{e^{79\beta} + e^{52\beta} + e^{63\beta} + e^{43\beta}} \times \frac{e^{52\beta}}{e^{52\beta} + e^{63\beta} + e^{43\beta}} \times \frac{e^{43\beta}}{e^{43\beta}}$$

- This example shows a second special case.
- Death occurs in order, so PL increases as β increases
- The value of β that maximizes the PL is positive infinity
- The term for the last death equals 1 and will not contribute to the PL.

Time Scale for Partial Likelihood

- PL only uses the ranks of time values
- Actual time is not used
- Years, Months and Weeks do not matter
- Log(Week) vs. Week does not matter

- Later we will see that for some time-dependent covariates (time*covariate interactions), the choice of time scale does matters

Methods for Tied Event Times in Partial Likelihood

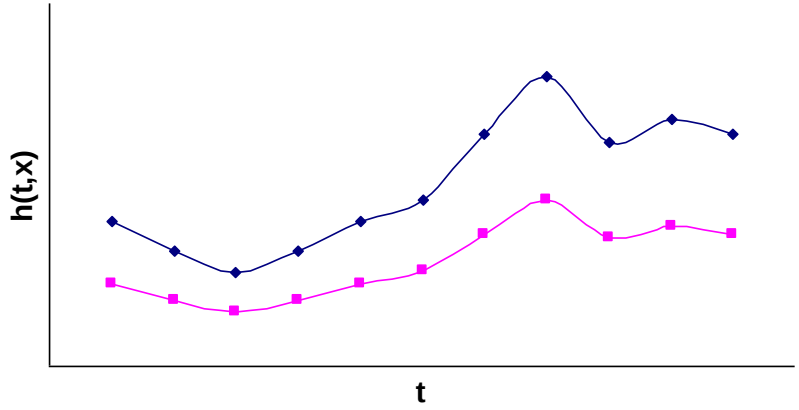
- Exact
  - Best
  - Computationally intensive
- Discrete
  - Exact
  - Proportional Odds not Proportional Hazards
- Efron
  - Best of approximations
  - Splus default
- Breslow
  - Worst approximation
  - SAS Default
  - Only available method in SPSS
- If no ties methods give identical results

Possible Remedial Measures for Issues in PH Modeling

	Potential implication	Suggested remedy
1	Outliers and Influential Observations	PH Influence Statistics
2	Interaction among Covariates	Martingale Plots
3	Incorrect Functional Form	Cumulative Residual Plots
4	Covariate excluded wrongly	Refit model with covariate included
5	Covariate modeled incorrectly (e.g. nonlinear effect)	Fit non-linear term (e.g. squared term)
6	Quasi-complete separation	Firths Penalized Likelihood Maximum
7	Correlated Responses	Quasi-likelihood estimation of the Sandwich Variance
8	Non-PH	?

Wilson (2008).

When Hazards are Proportional



$$h_1(t) = \gamma \cdot h_2(t)$$

- Hazard functions for two different levels of a covariate are proportional for all values of t
- For example, if men have twice the risk of myocardial infarction compared to women at age 40, the risk is twice at age 60
- The baseline function for age can have any form
- The estimates of the betas are consistent (i.e. asymptotically unbiased) and asymptotically normally distributed.

How do you know?

Hazard Ratio from PH Regression

- The hazard ratio for two individuals with a different covariate, like gender

$$\text{HR (Gender)} = \frac{\boldsymbol{h}(t;x = 1)}{\boldsymbol{h}(t;x = 0)} = \frac{\boldsymbol{h_o}(t)\exp\{\beta(x_1=1)\}}{\boldsymbol{h_o}(t)\exp\{\beta(x_1=0)\}}$$

$$= \exp\{\beta(x_1=1 - x_1=0)\} = e^{\beta}$$

- Notice how the baseline hazard function, $\boldsymbol{h_o}(t)$, cancels out because it is the same at all time points, t.

Coding Categorical Variables with Missing Values

- Get into the habit of coding for missing values routinely

```
if edema = . then do;
    MildED = .;
    SevED  = .;
    end;
 else if edema = 0.0 then do;
    MildED = 0;
    SevED  = 0;
    end;
 else if edema = 0.5 then do;
    MildED = 1;
    SevED  = 0;
    end;
 else if edema = 1.0 then do;
    MildED = 0;
    SevED  = 1;
    end;
```

- Include the Indicator variables in the model statement

```
proc phreg data = liver02;
     model Time*Status(0) = sex age MildED SevED;
     run;
```

The PHREG Procedure

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	1740.013	1505.341
AIC	1740.013	1517.341
SBC	1740.013	1535.829

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	234.6717	6	<.0001
Score	329.2670	6	<.0001
Wald	240.7266	6	<.0001

Introduction to Survival Analysis

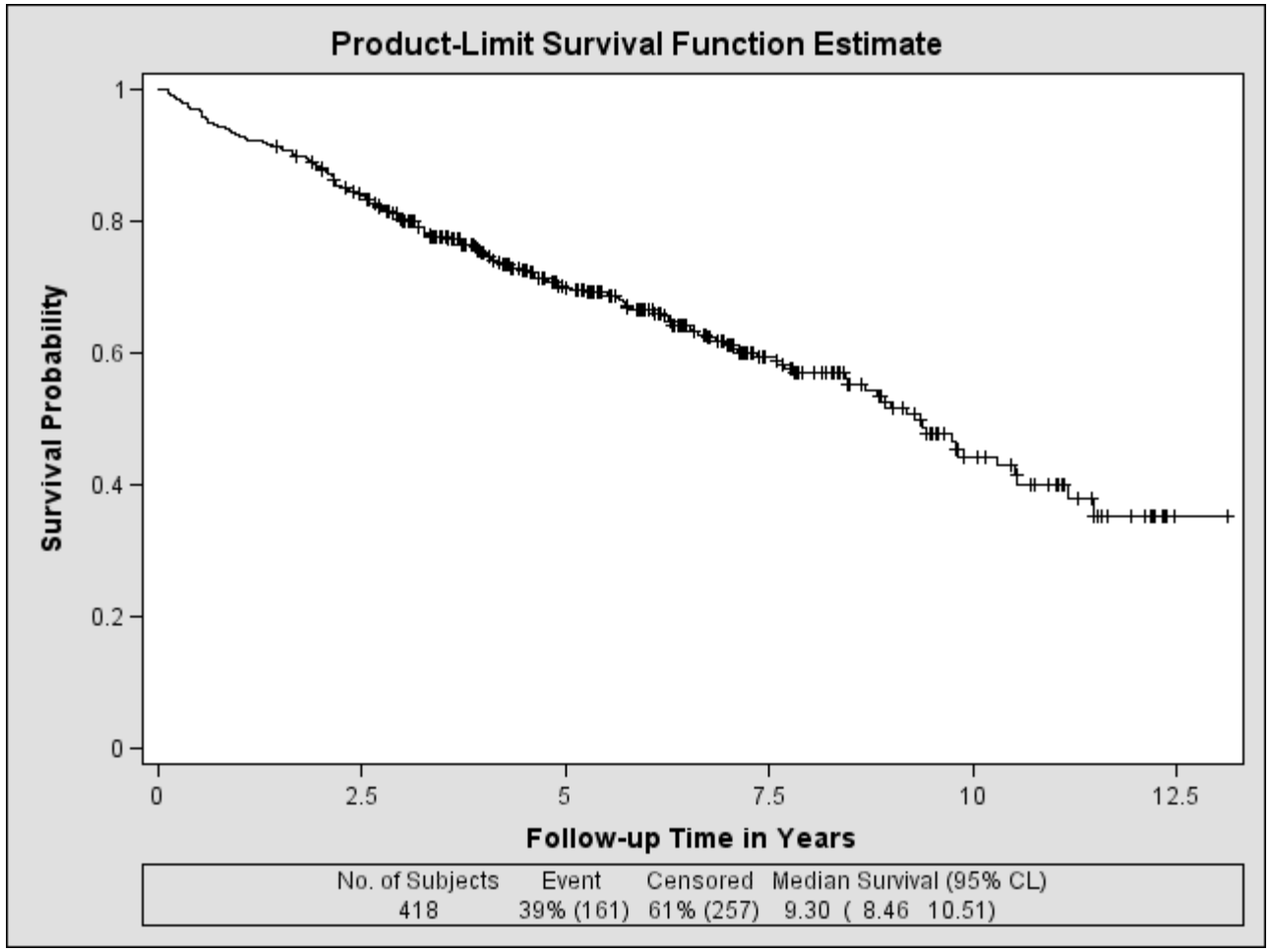
The PHREG Procedure
Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
logBili	1	0.86715	0.08291	109.3945	<.0001	2.380
logProtime	1	2.35341	0.77093	9.3189	0.0023	10.521
logAlbumin	1	-2.55101	0.64818	15.4893	<.0001	0.078
Age	1	0.04019	0.00768	27.3524	<.0001	1.041
MildED	1	0.27028	0.22495	1.4437	0.2295	1.310
SevED	1	0.97348	0.28943	11.3126	0.0008	2.647

Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
test1	4.7467	1	0.0294

Survivor Function for PBC Dataset



Polytomous Parameters Estimates

- The coefficient for MildED reflects the difference between Mild Edema and the reference group (Edema = none)
  - $\beta < 0$: lower death risk for Mild than No Edema
  - $\beta > 0$: higher death risk for Mild than No Edema
- The coefficient for SevED reflects the difference between Severe Edema and the reference group (Edema = none)
  - $\beta < 0$: lower death risk for Severe than No Edema
  - $\beta > 0$: higher death risk for Severe than No Edema
- Additionally, a comparison between Mild vs. Severe might be interesting
  - Mild vs. Severe is the difference between (Mild vs. None) and (Severe vs. None)
  - Use the Test statement
  - Could get all pairwise comparisons

Measure of Explained Variation

- In ordinary regression, R^2 = proportion of variation explained by the model.
- For likelihood-based procedures, we use the Generalized R^2
  - = $1 - \exp\{-(2/n)\,(LL(\beta) - LL(0))\}$,
    - where LL = Log Likelihood
  - = $1 - \exp\{-X^2/n\}$,
    - where X^2 = LR test statistic
- For PH model, perfect fit implies $LL(\beta) = 0$.
- So a maximum possible Generalized $R^2 = 1 - \exp\{-(2/n)\,LL(0)\}$
- If n is large the Generalized R^2 is approximately 1
- For PBC data,
  - Generalized $R^2 = 1 - \exp\{-234.6/418\} = 0.43$
  - Maximum possible Generalized $R^2 = 1 - \exp\{-1740.0/418\} = 0.98$

Liver Cirrhosis Treatment Study

- Randomized clinical trial at Mayo
- Survival of patients with liver cirrhosis (NEJM 306:319-326)
- Compare a new treatment, D-penicillamine with placebo
- Data collected at randomization
  - Presence/absence of ascites
  - Prothrombin time in seconds -10

Proportional Hazards Model Results

- $h(t) = h_0(t) \exp\{ -0.135 X_{TRT} + 1.737 X_A + 0.346 X_P \}$
  - X_{TRT}: 1 = D-penicillamine, 0 = placebo
  - X_A: 1 = ascites, 0 = no ascites
  - X_P: Prothrombin time – 10 (Continuous variable, units in seconds)
- $h_0(t)$ is the event rate at time t in the placebo arm for subjects without ascites with a prothrombin time of 10 seconds
- Hazard rate of death two years post randomization for a subject on this trial who received the new treatment, had ascites at randomization and a prothrombin time of 10 seconds compared to a similar subject who received placebo?
- $RR = \exp\{ -0.135 \} = 0.87$

Hazard Rate Calculations

$$\lambda(t) = \lambda_0(t)\exp\{\text{-0.135 } X_{TRT} + 1.737\ X_A + 0.346\ X_P\}$$

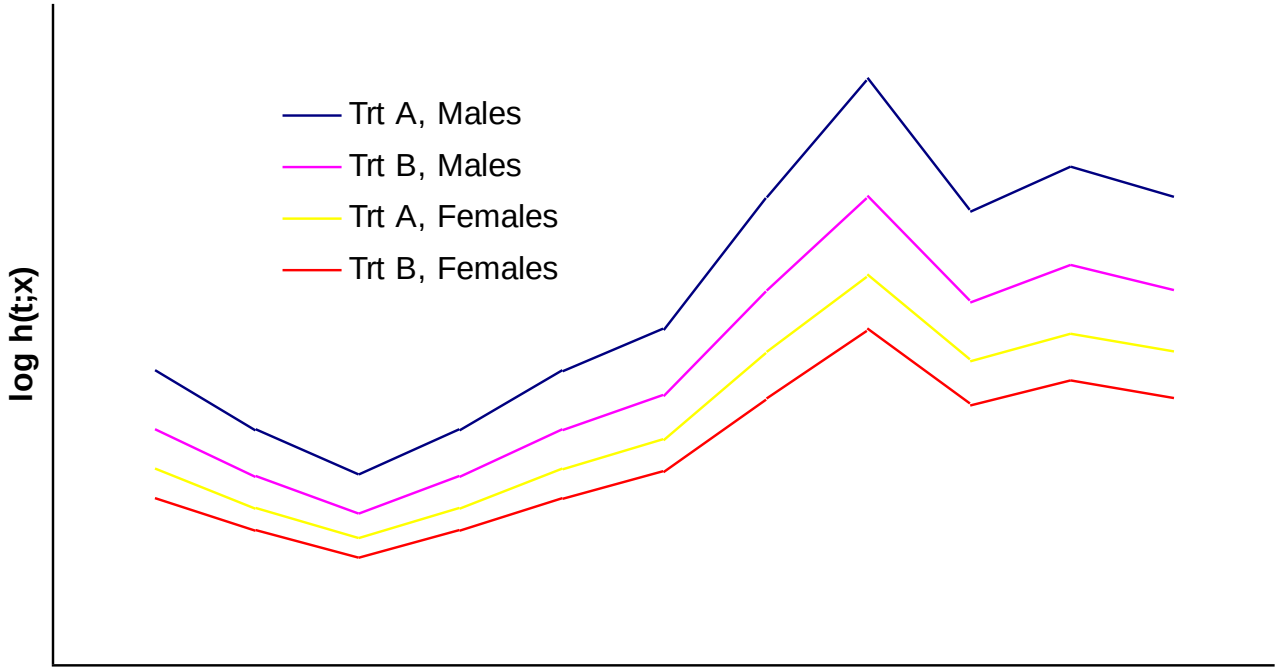
$$\frac{\lambda_{person1}(t)}{\lambda_{person2}(t)} = \frac{\lambda_0(t)\exp\{\text{-0.135 *1+1.737*1+0.346 *0}\}}{\lambda_0(t)\exp\{\text{-0.135 *0+1.737 *1+0.346 *0}\}} =$$

$$\frac{e^{-0.135} * e^{1.737} * e^{0}}{e^{0} * e^{1.737} * e^{0}} =$$

$\exp\{-0.135\} = 0.87$ is the relative rate of death for subjects who received treatment compared to those who received placebo

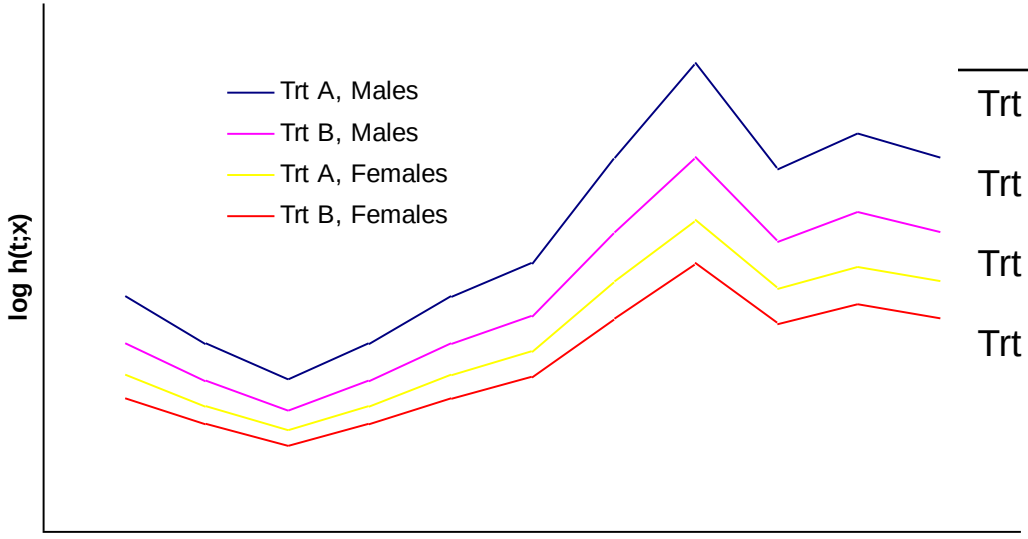
Dichotomous Covariate Interactions

- $\boldsymbol{h}(t;\boldsymbol{x}) = \boldsymbol{h_o}(t)\exp\{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2\}$



Dichotomous Covariate Interactions

- $\boldsymbol{h}(t;\boldsymbol{x}) = \boldsymbol{h_o}(t)\exp\{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2\}$



	x1	x2	h(t;x)
Trt A, Males	0	1	h0(t) e{β2}
Trt B, Males	1	1	h0(t) e{β1+β2+ β3}
Trt A, Females	0	0	h0(t)
Trt B, Females	1	0	h0(t) e{β1}

Modified from Marubini and Valsecchi

Dichotomous Covariate Interactions

- Because the baseline hazard function $h_0(t)$ is not estimated the previous plot is hypothetical and cannot be created
- Therefore, we have two options to present interactions:
- First, present separate hazard ratios by level of covariate

For example, Let $\beta_1 = -0.6$ (trt) $\beta_2 = 1.4$ (gender) $\beta_3 = -0.4$ (intx)

Males: HR (Trt B vs. Trt A) $= \exp(\beta_1 + \beta_3) = \exp(-0.6 - 0.4)$

$= \exp(-1) = 0.4$

Females: HR (Trt B vs. Trt A) $= \exp(\beta_1) = \exp(-0.6)$

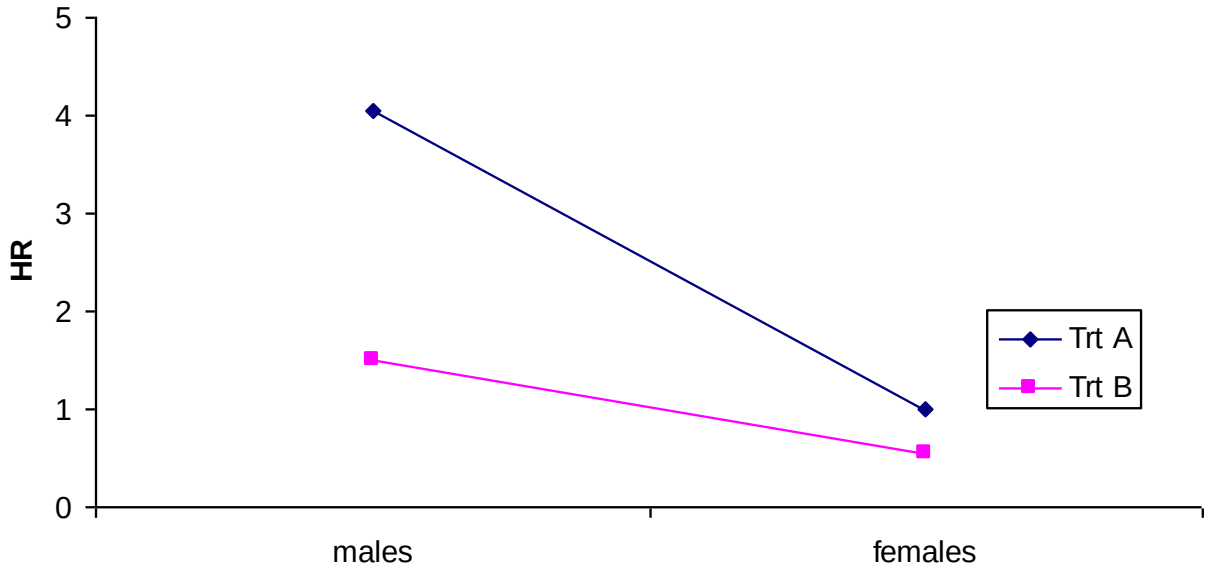
$= \exp(-0.6) = 0.6$

Trt B is better than A, but larger effect in males

Dichotomous Covariate Interactions

- Secondly, plot the hazard ratios in comparison to the reference group
- The reference group for this example is Females on Treatment A

	Trt	HR	h(t;x)
Males	A	4.0	e{β2}
	B	1.5	e{β1+β2+ β3}
Females	A	1.0	e{1}
	B	0.6	e{β1}



Continuous Covariate Interactions

- Continuous-by-Categorical Interaction
  - Plot the HR by the continuous covariate with separate line for the levels of the categorical covariate
- Continuous-by-Continuous Interaction
  - Plot the HR by one of the continuous covariates, with separate lines for selected values of the other covariate

Checks on the Proportional Hazards Assumption

- Graphical
  - Log Cumulative Hazard
  - Schoenfeld Partial Residuals
  - Standardized Score Process
- Analytical
  - Standardized Score Process
  - Schoenfeld Partial Residuals
  - Covariate by Time Interaction
    - Continuous
    - Categorical

Log Cumulative Hazard

- If the hazards are proportional: $h_1(t) = \gamma \cdot h_2(t)$
  it can be shown that:

$$\log[-\log S_1(t)] = log\gamma + \log[-\log S_2(t)]$$

- We recognize –log(S(t)) as the cumulative hazard function, H(t)
- So this relationship implies that the Log Cumulative Hazard for the two groups differs by a constant.
- Plots of the log cumulative hazard would be parallel
- Departures from parallelism would be consistent with violations of the PH assumption.

Synthetic, Censored Time-to-Event Datasets

- Clinical Trial Format for Acute Coronary Syndrome
- A subset of variables are:

TIME	Follow-up time, in years
STATUS	Event indicator with value 1 for death time and value 0 for censored time
TREATMENT	Dichotomous Treatment Variable
AGE	Age in years continuous
TROPININ	Diagnostic Cardiac Tropinin-T (cTnT; in µg/L)

Failure times generated from Weibull hazard:

$h(t) = \lambda\gamma(\lambda t)^{\gamma-1}$

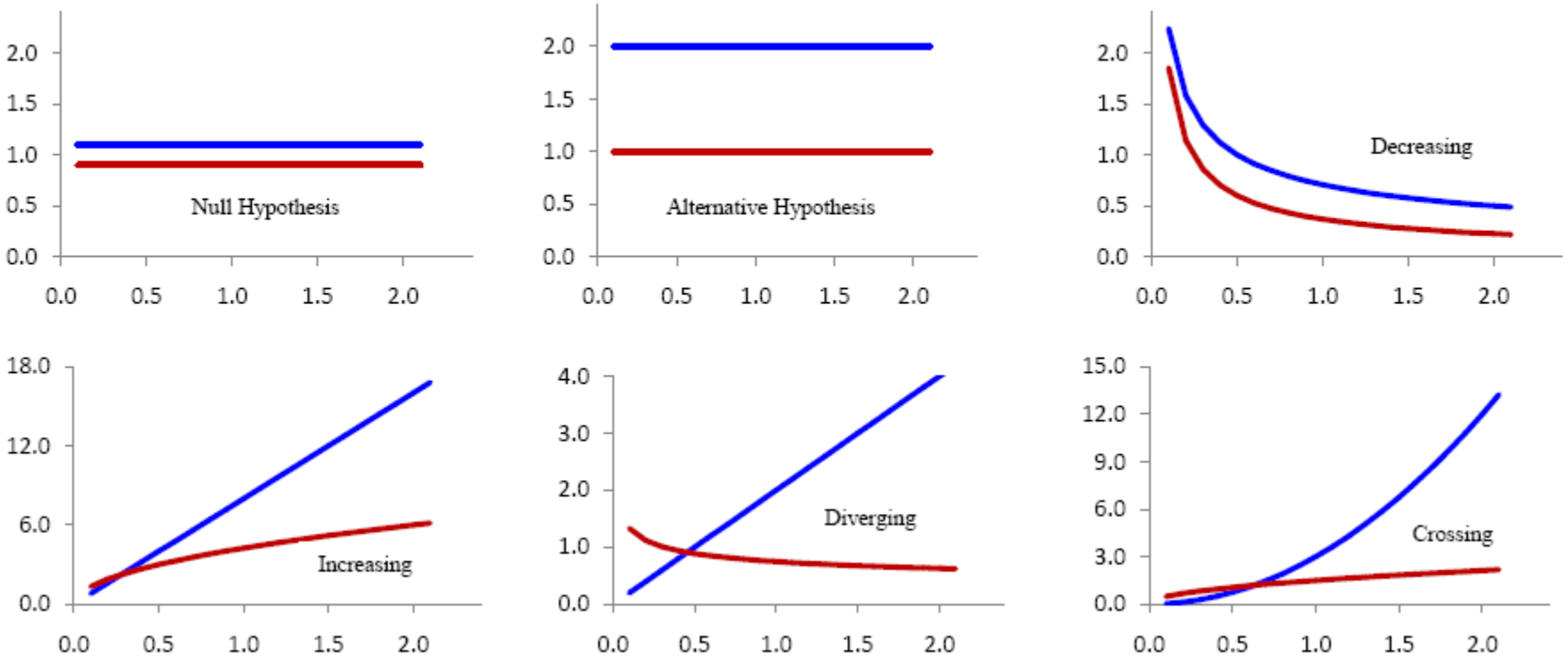
Melanson 2007

Weibull Parameter Values for Six Hazard Patterns

		Control Group		Investigational Group	
Case	**Hazards Pattern**	**Shape (λ)**	**Scale (γ)**	**Shape (λ)**	**Scale (γ)**
1	Constant (1)	2.00	2.00	2.00	2.00
2	Constant (2)	1.00	1.00	1.00	2.00
3	Decreasing	0.30	2.00	0.50	2.00
4	Increasing	1.50	2.00	2.00	2.00
5	Diverging	0.75	1.00	2.00	1.00
6	Crossing	1.50	1.00	3.00	1.00

Adapted from Ng'andu 1997

Six Patterns of Hazard Functions by Time



The control groups are shown in red while investigational groups are shown in blue.

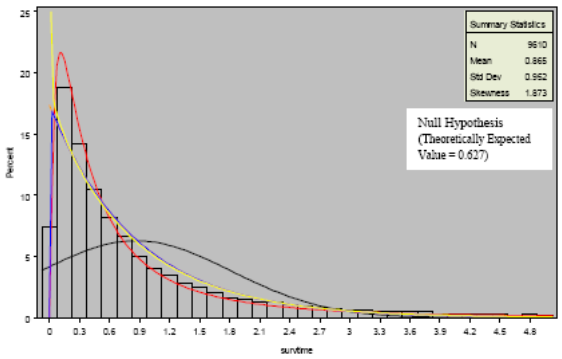
Ordinarily Generated Graphs for PH

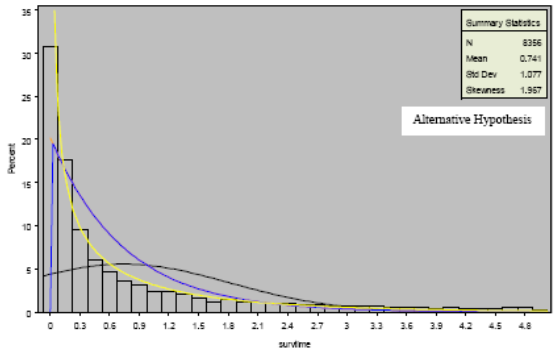
- Histograms with distributions
- Survival Curves
- Hazard Curves
- Cumulative Hazard Curves

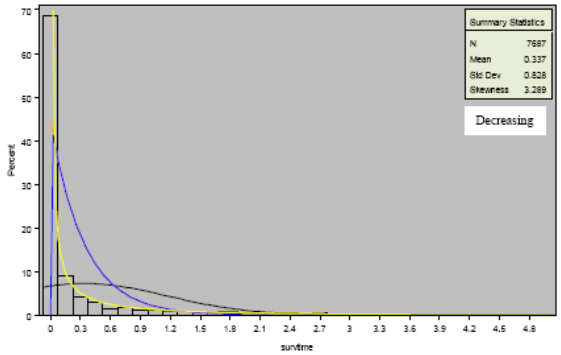
Generate Histograms

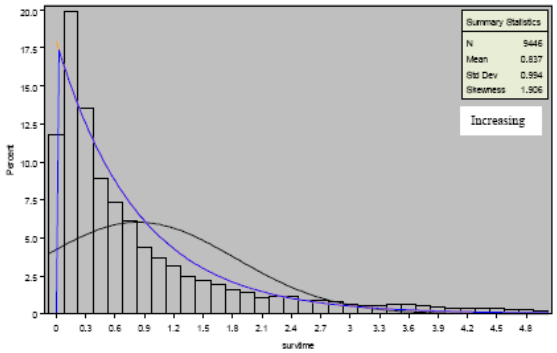
```
%let pctlist = 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 ;

footnote5 'Normal = Black, LogNormal = Red, Exponential = Orange,
           Weibull = Blue, Gamma = Yellow';
ods trace on;
ods output GoodnessOfFit = gof01 FitQuantiles = fq01
           ParameterEstimates = pe01;
proc univariate data=plotdata pctldef=5;
     var survtime;
     histogram /  normal      (l=3  color=black percents = &pctlist)
                  lognormal   (theta = 0   zeta  = est sigma = est l=1
                               color=red percents = &pctlist)
                  exponential (theta = est sigma = est l=4
                               color=orange percents = &pctlist)
                  weibull     (theta = est sigma = est l=2
                               color=blue   percents = &pctlist)
                  gamma       (theta = est sigma = est l=8
                               color=yellow percents = &pctlist)
                  cframe  = ligr
                  vaxis   = axis1
                  name    = 'MyHist';
      inset n mean(5.3) std='Std Dev'(5.3) skewness(5.3)
             / pos = ne  header = 'Summary Statistics' cfill = ywh;
      axis1 label=(a=90 r=0);
      run;
 footnote5 ' ';
```

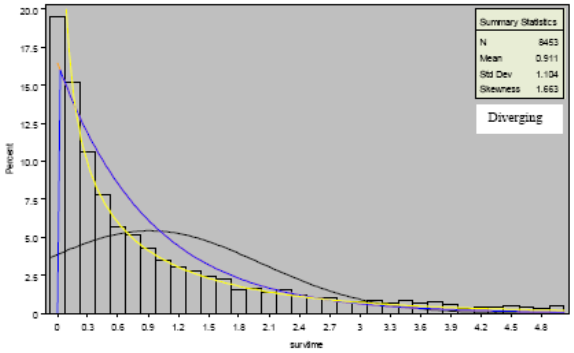
Histograms of Time-to-Event Data

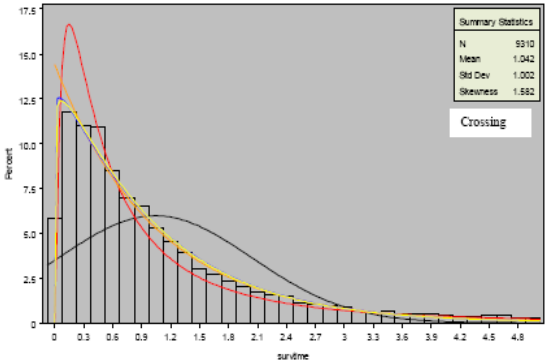










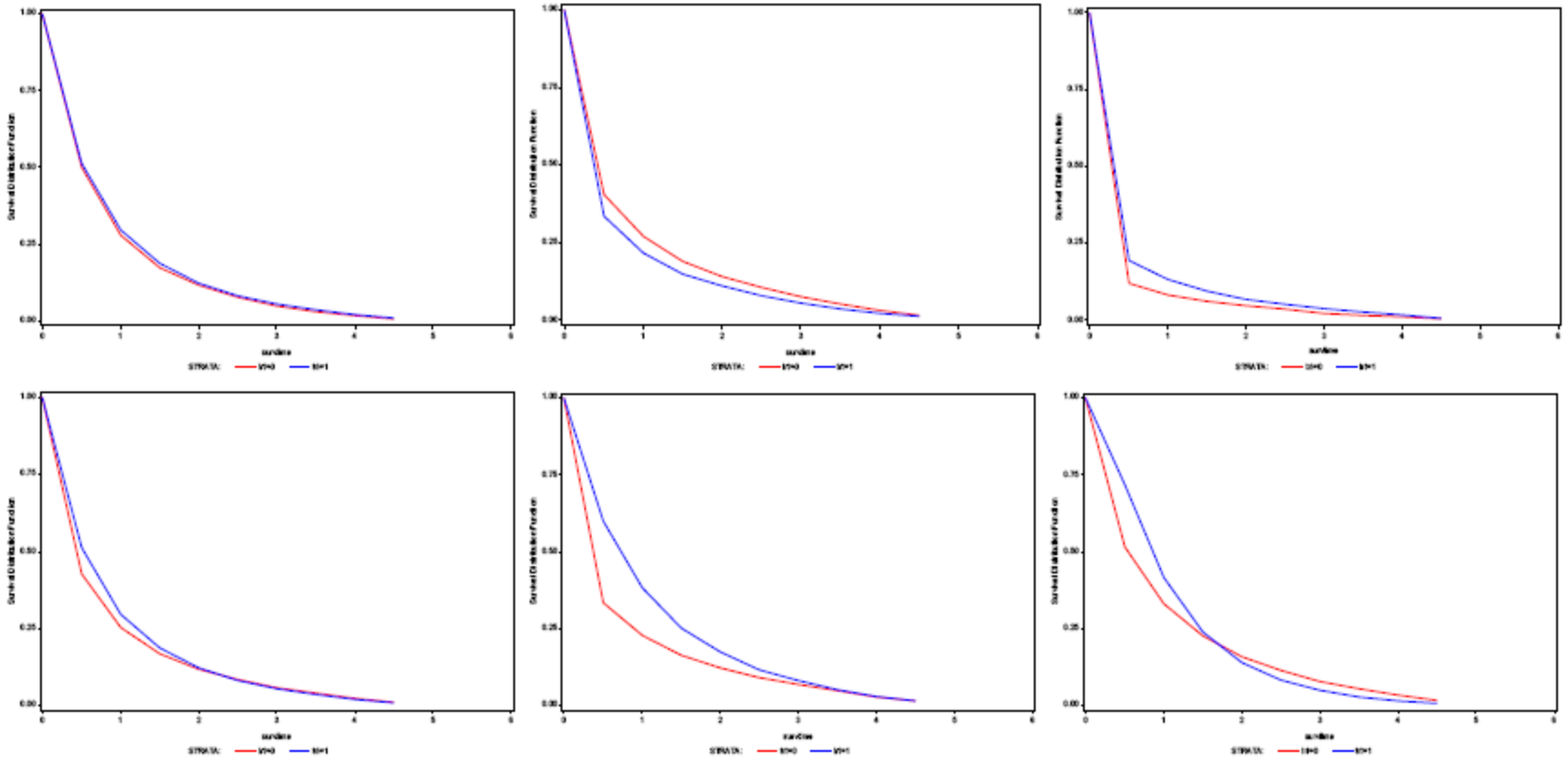


Normal = Black; Lognormal = Red; Exponential = Orange; Weibull = Blue; Gamma= Yellow

Generate Survival Curves

```
symbol1 value=none interpol=join color = red;
symbol2 value=none interpol=join color = blue;
proc lifetest data = plotdata method = life plots = (s) graphics
              intervals = 1 to 9 by 2;
              strata trt;
              time survtime;
              run;
```

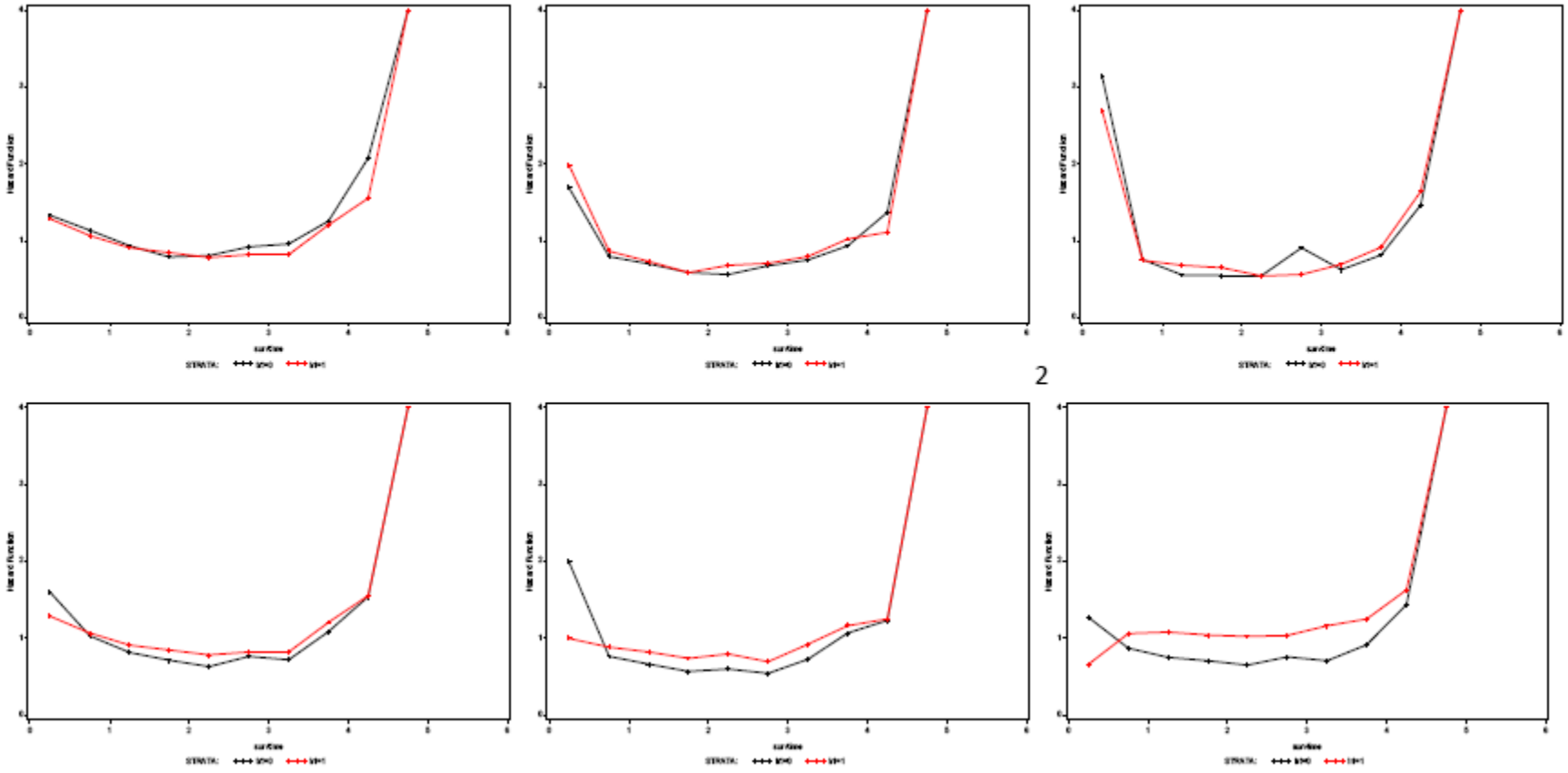
Six Patterns of Survival Functions by Time



Generate Hazards

```
proc lifetest data = plotdata method = life plots = (h) graphics
              intervals = 1 to 9 by 2;
              strata trt;
              time survtime;
              run;
```

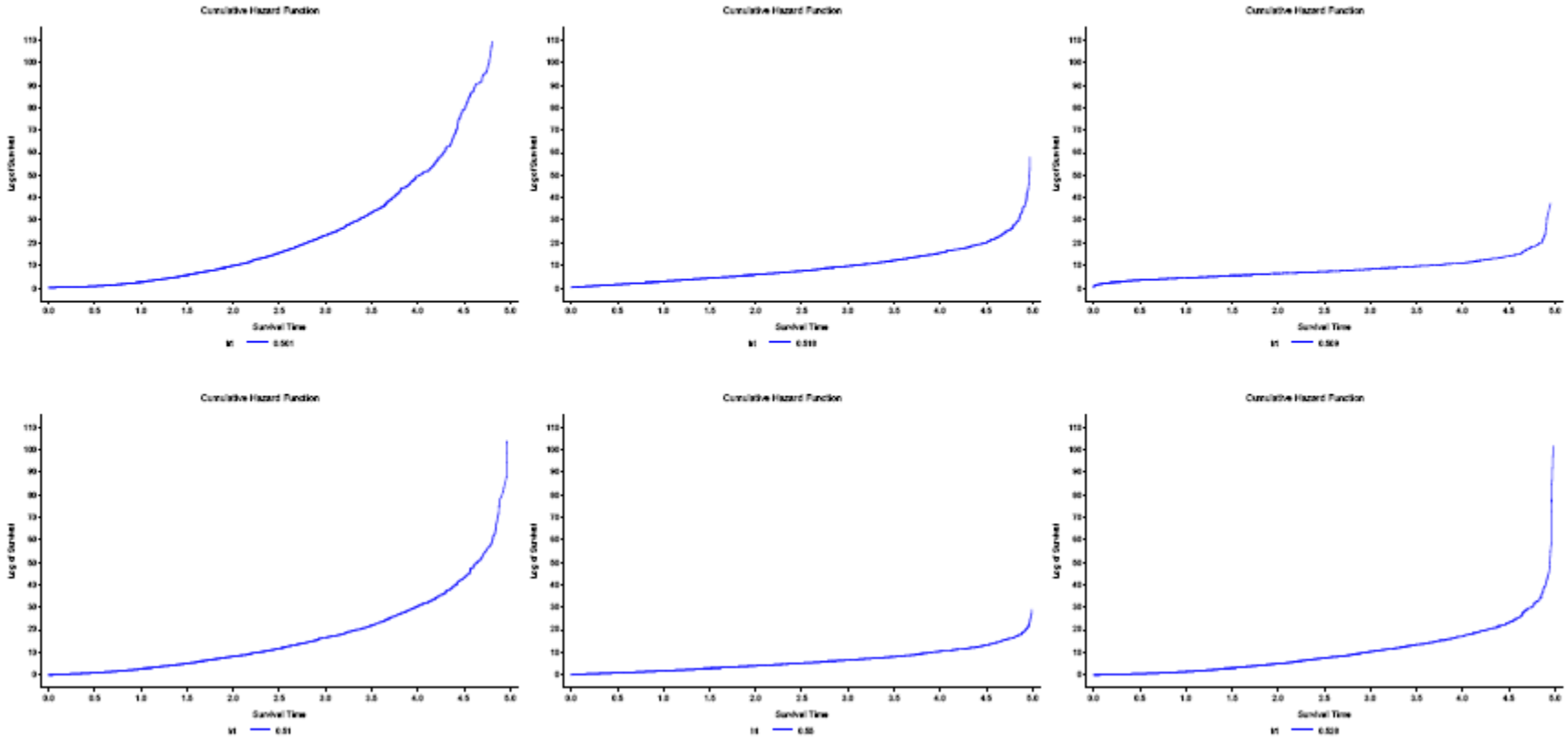
Six Patterns of Hazard Functions by Time



Cumulative Hazard Function

```
proc phreg data = plotdata;
     model survtime*event(0) = x1 x2 x3 trt  / ties = efron;
     baseline out = out03 survival = s logsurv = ls loglogs = lls; run;

data out04;
     set out03;
     ls = -ls;
     logsurvtime = log(survtime); run;

goptions reset=all ftext="arial" htext=1.5 hsize=11 in vsize=8 in
          device=emf rotate=landscape gsfname=PlotOut4 gsfmode=replace;
filename PlotOut4 "%nrbquote(&projlib)sasout\&saspgmn._OGSA04.emf";
symbol1 value=none interpol=join color = blue w=2 ;
axis1 minor=none offset=(2,2) label=(a=90 h=1.6 'Log of Survival')
      order=0 to 110 by 10 value=(h=1.5);
axis2 minor=none offset=(2,2) label=(h=1.6 'Survival Time')
      order=0 to 5 by .5 value=(h=1.5);

title4 h=1.8 'Cumulative Hazard Function';
proc gplot data = out04;
     plot ls*survtime = trt / vaxis = axis1 haxis = axis2 nofr;
     run;
```

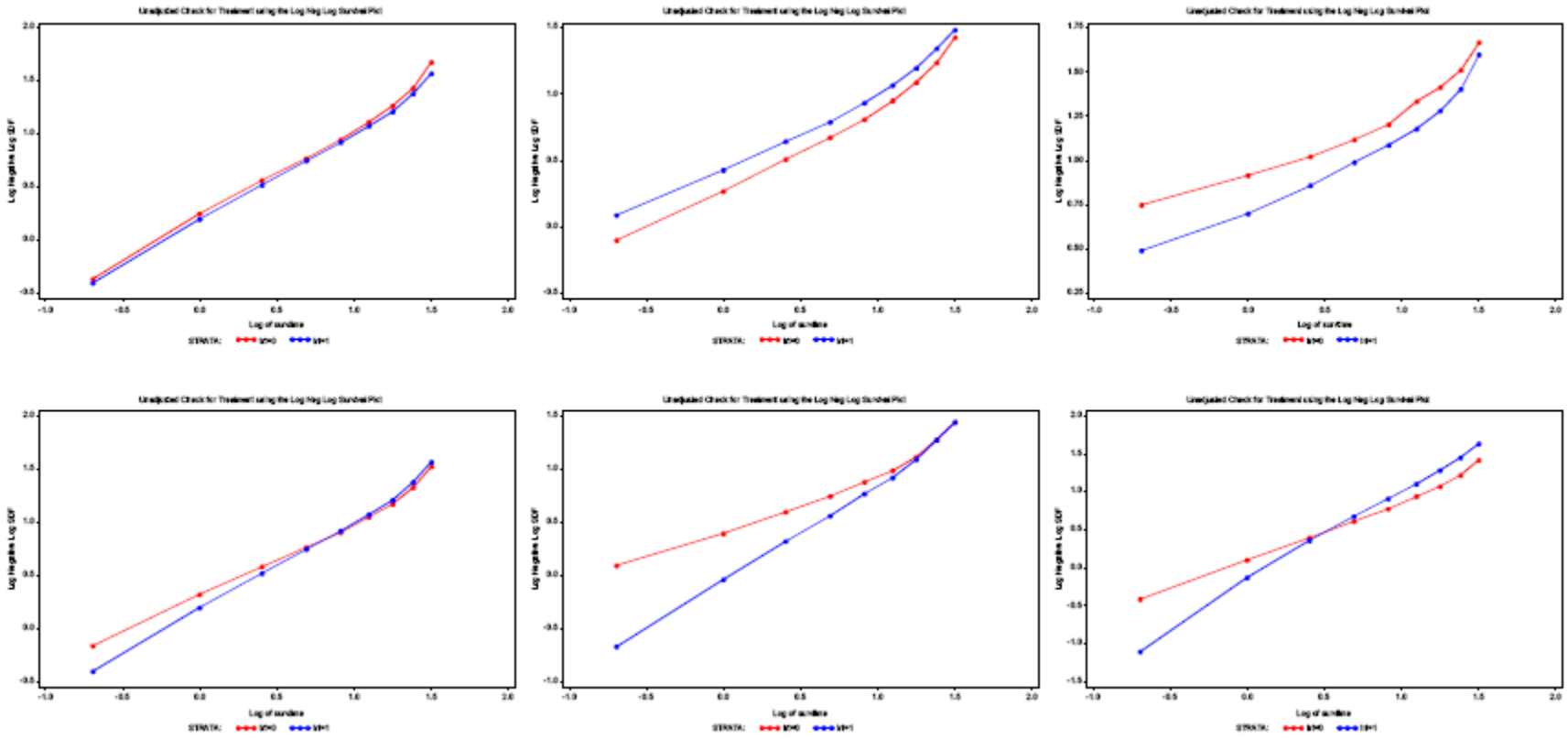
Six Patterns of the Cumulative Hazard Functions by Time



Unadjusted Log Cumulative Hazard

```
goptions reset=all ftext="arial" htext=1.5 hsize=11 in vsize=8 in
          device=emf rotate=landscape gsfname=PlotOut5 gsfmode=replace;

filename PlotOut5 "%nrbquote(&projlib)sasout\&saspgmn._GCPH01.emf";

symbol1 value=dot w=2 h=1 l=1 interpol=join color = red;
symbol2 value=dot w=2 h=1 l=1 interpol=join color = blue;

title4 'Unadjusted Check for Treatment using the Log Neg Log Survival Plot';
proc lifetest data = plotdata method = life plots = (lls) graphics
              intervals = 0 to 5 by 0.5;
              strata trt;
              time survtime;
              run;
```

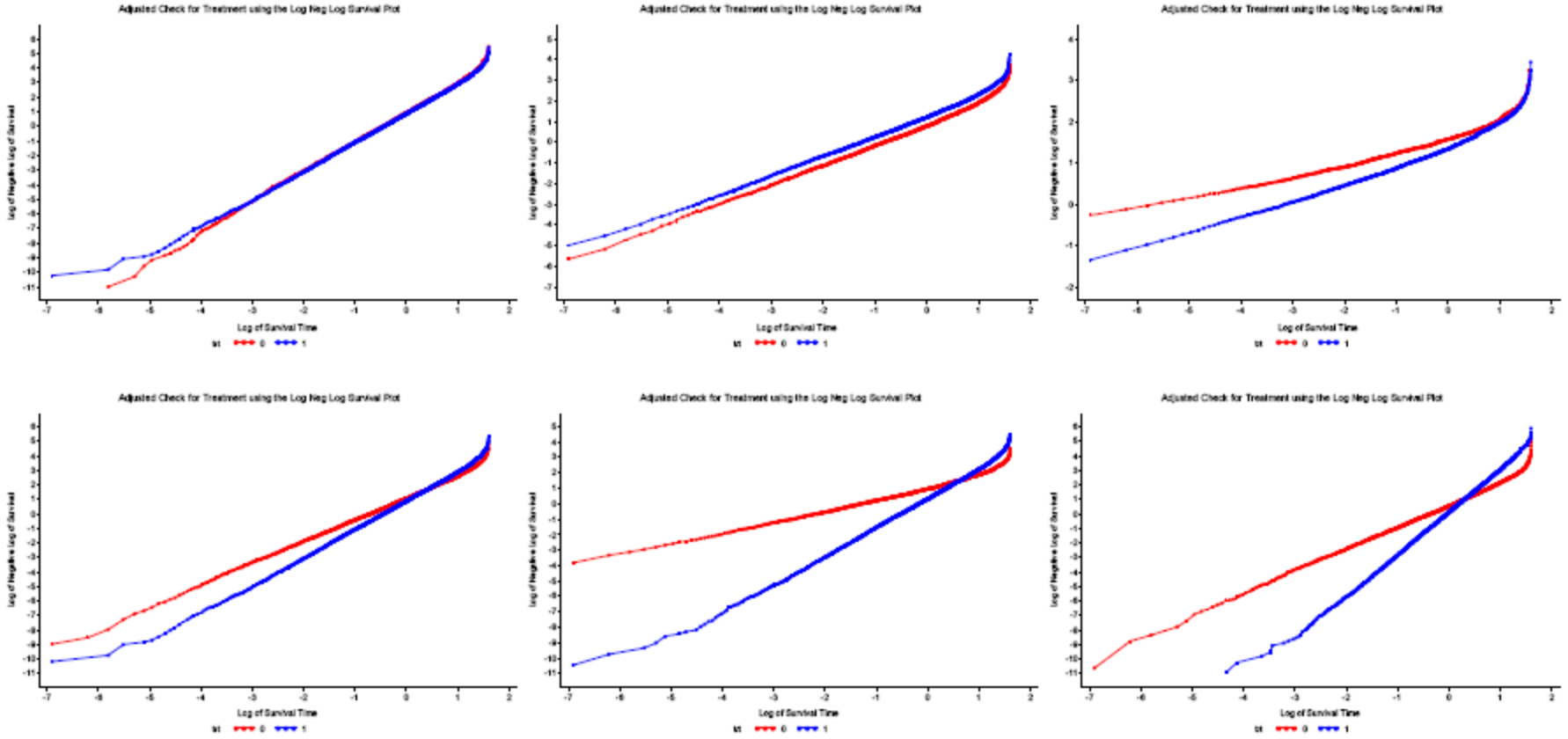
Six Patterns of the Unadjusted Log Cumulative Hazard Functions by Log Time



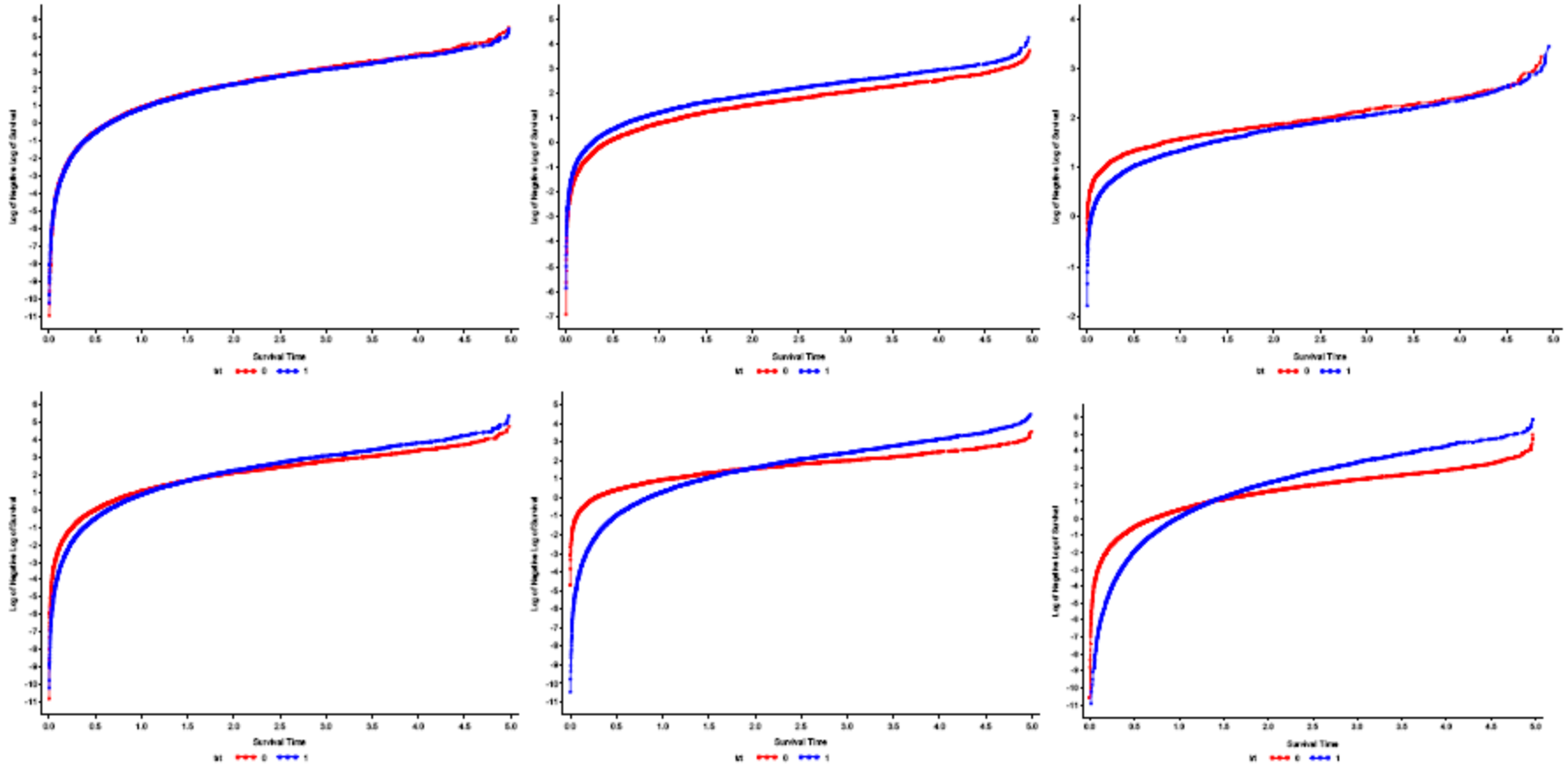
Adjusted Log Cumulative Hazard

```
title4 'Adjusted Check for Treatment using the Log Neg Log Survival Plot';
proc phreg data = plotdata;
     model survtime*event(0) = x1 x2 x3 / ties = efron;
     strata trt;
     baseline out = out01
              survival = s
              logsurv = ls
              loglogs = lls;
              run;

data out02;
     set out01;
     label logsurv = 'Negative Log of Survival';
     if ls = . then logsurv = .;
     else logsurv = -ls;
     if survtime = 0 then logsurvtime = .;
     else logsurvtime = log(survtime);
     run;
```

Six Patterns of the Adjusted Log Cumulative Hazard Functions by Log Time



Six Patterns of the Adjusted Log Cumulative Hazard Functions by Time



Schoenfeld Partial Residuals

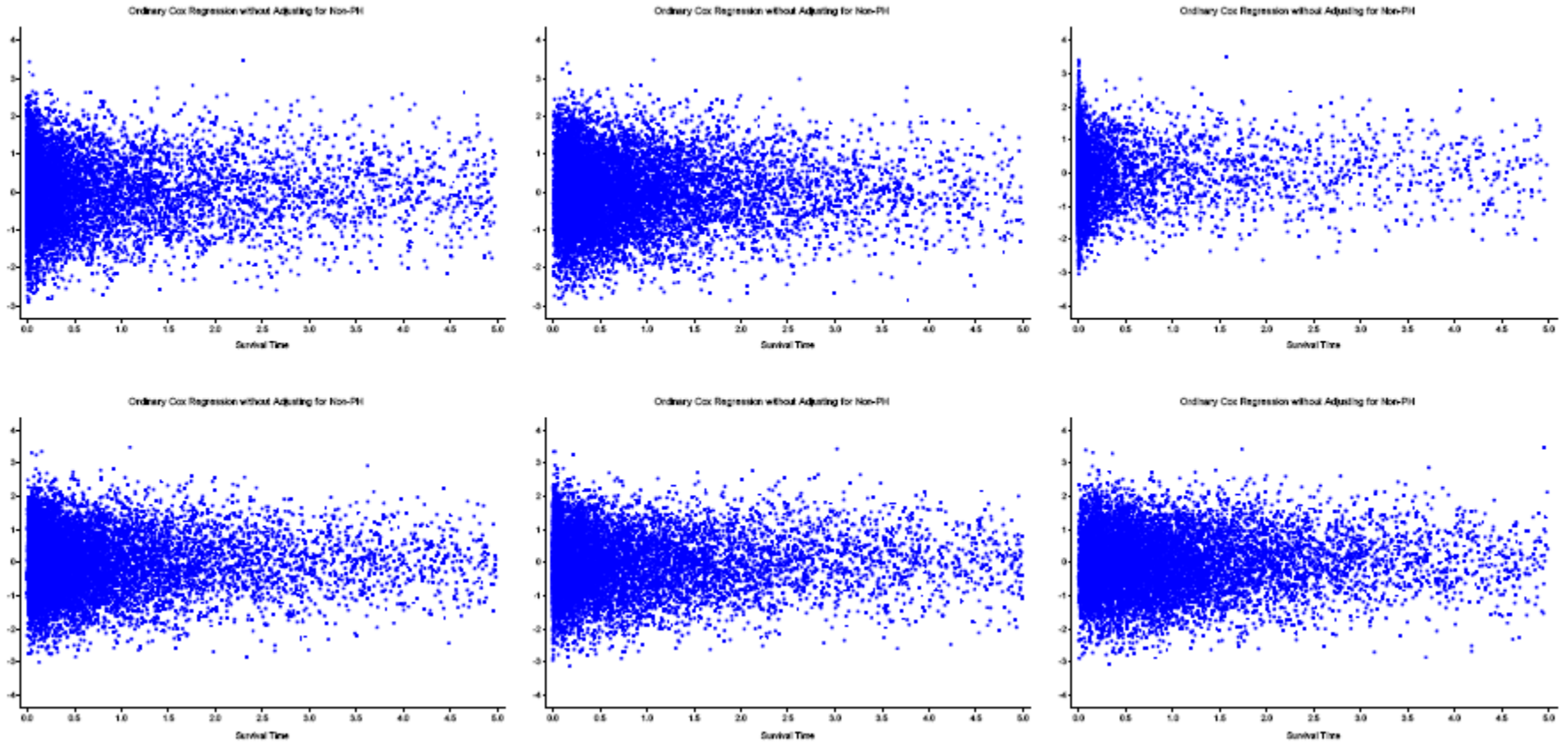
$$\hat{r}_{ik} = x_{ik} - \hat{\bar{x}}_{w_i k}$$

- For each covariate in a PH regression, a Shoenfeld Partial Residual can be calculated for each case that was not censored.
- The residual is the difference between covariate value for an individual and a weight mean of covariate values for all individuals.
- Under proportional hazards, a plot of this residuals against time should be approximately flat.
- Note:
  - These residuals are time independent and
  - Baseline Hazard Function independent

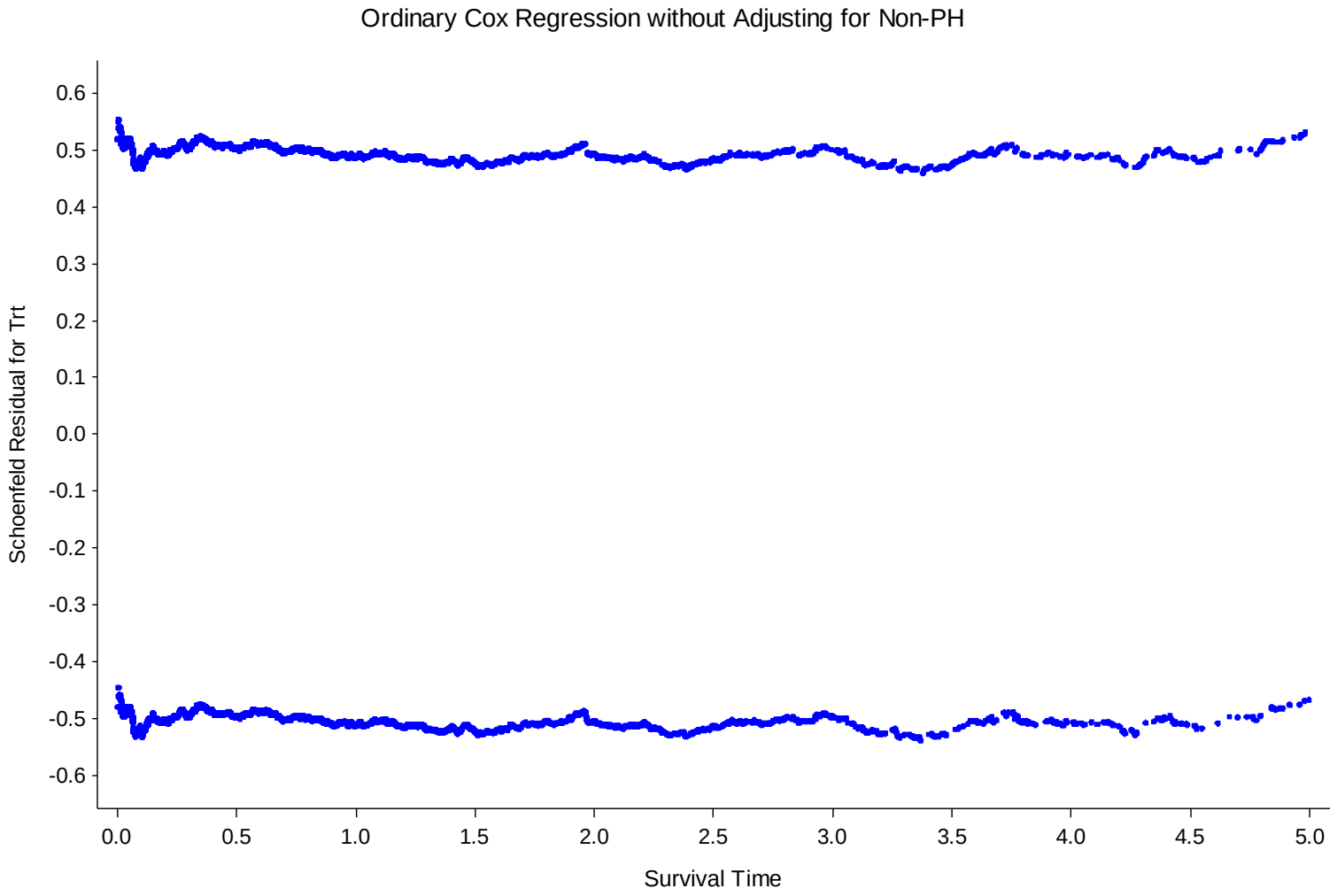
Schoenfeld Partial Residuals

```
title4 'Ordinary Cox Regression without Adjusting for Non-PH';
proc phreg data = ads01;
     model survtime*event(0) = x1 x2 x3 trt / ties = efron;
     output   out = out05
              ressch = schx1 schx2 schx3 schtrt;
              run;
```

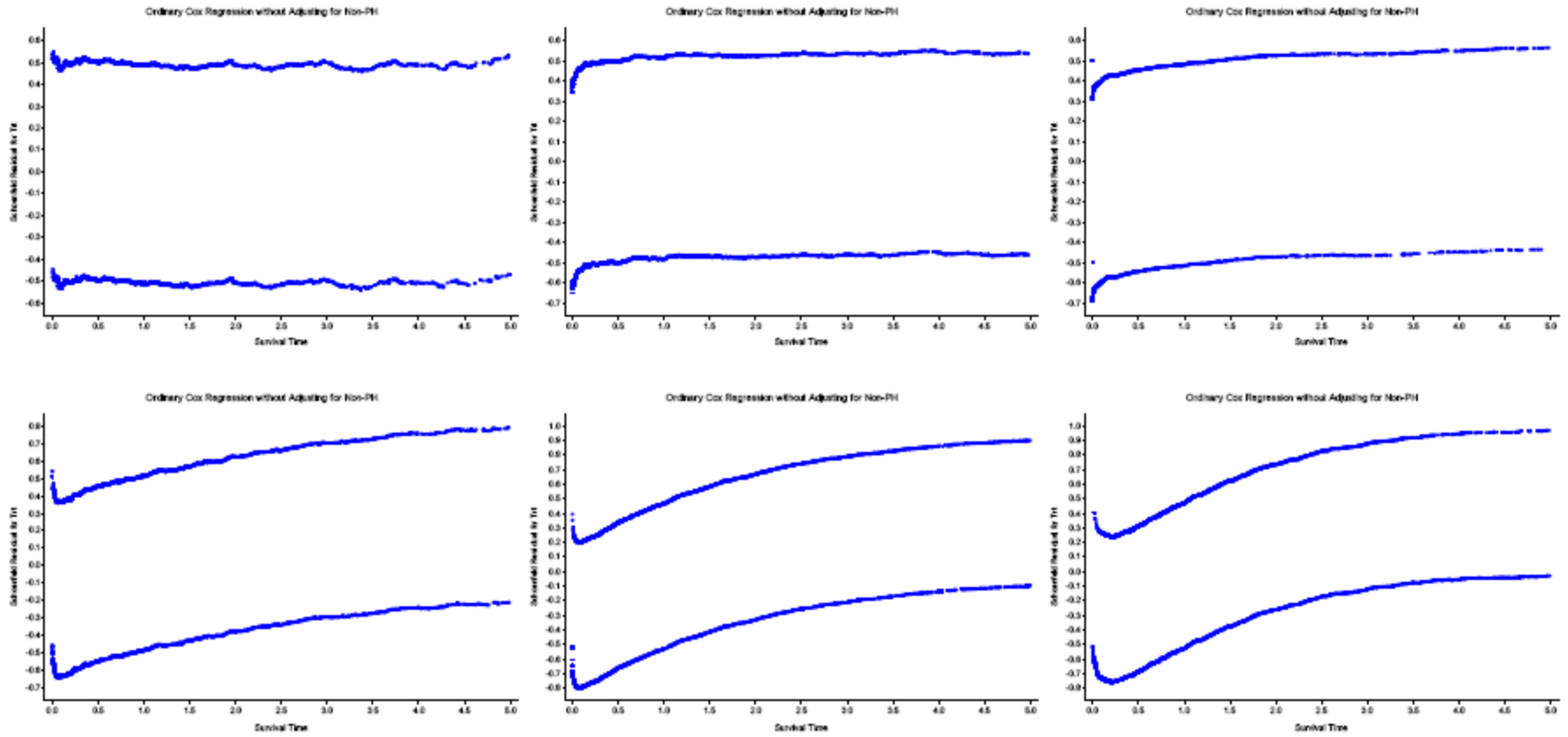
Schoenfeld Residuals for a Continuous Variable known to follow the Proportional Hazards Assumption by Survival Time



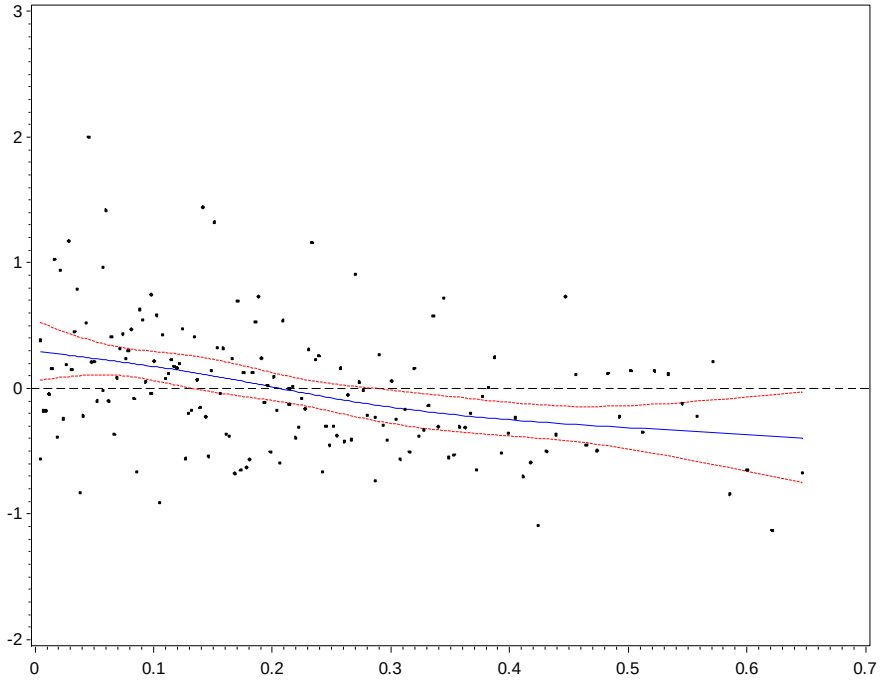
Schoenfeld Residuals for a Categorical Variable under the Null Hypothesis by Survival Time



Schoenfeld Residuals for a Categorical Variable under different Proportional Hazards Assumptions by Survival Time



Consider Adding a LOESS smoothing line through the Schoenfeld Residuals




```
ods output ScoreResults = ScoreResults01;
proc loess data = phregout;
     model resid = covariate;
     score; run;
```

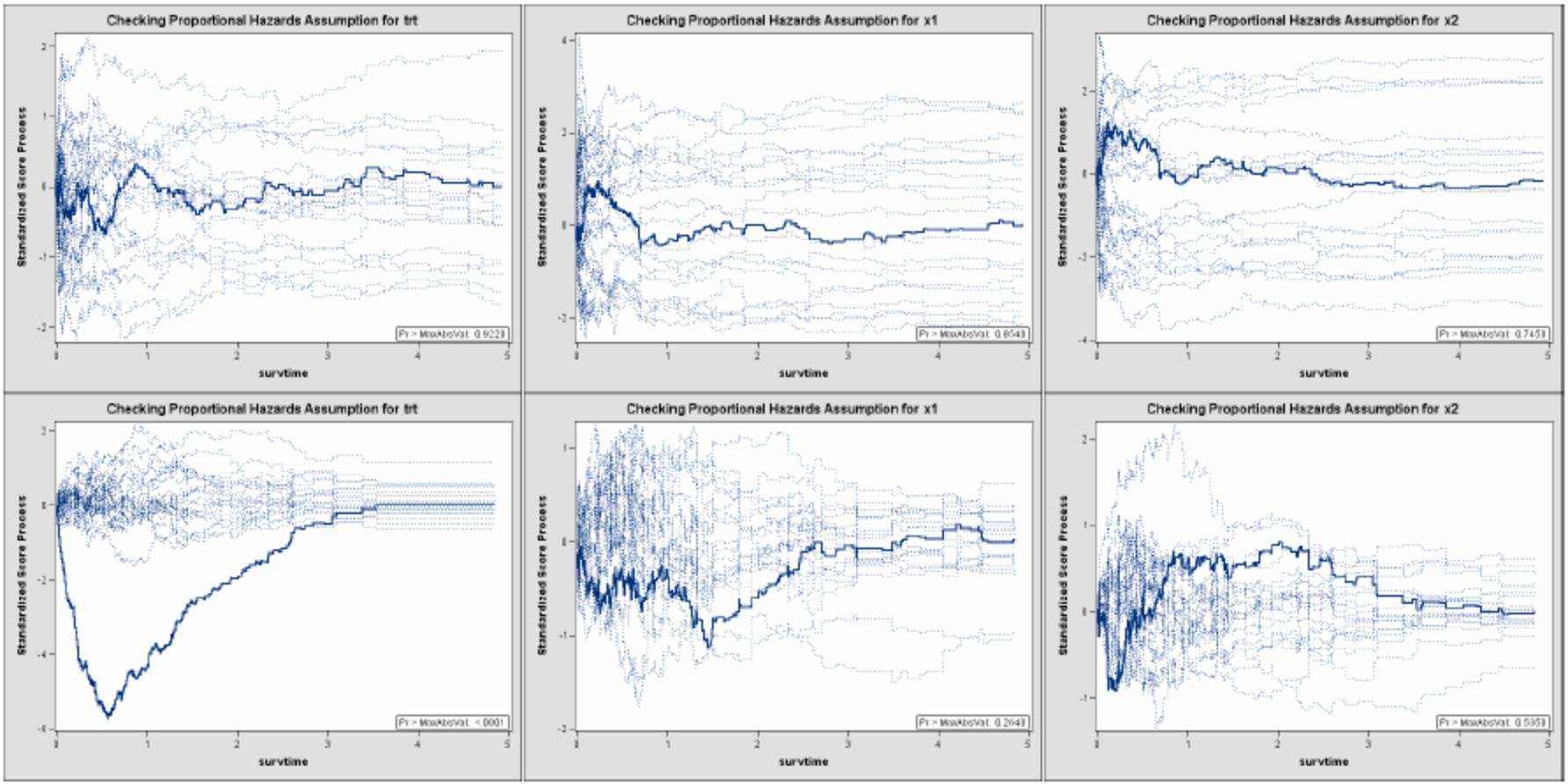
Standardized Score Process

- Martingale Residual Transform
  - Martingale is a special sequence of random variables where the conditional expected value of the next observation, given all the past observations, is equal to the last observation
- Tied down Brownian Process
  - Start and end at zero
- Paths (or process) under the null hypothesis is simulated
- Atypical observed paths are evidence of violations of proportional hazards

Standardized Score Process

```
title4 h=1.8 'Standardized Score Process Plots to Assess the PH assumption';
title5 h=1.8 'Kolmogorov-type Supremum Test using the Assess statement';
proc phreg data = plotdata02;
     model survtime*event(0) = trt x1 x2 x3
           / ties = efron;
     assess var=(trt ) ph / resample seed = 46163  ;
run;
title4 ' ';
title5 ' ';
ods graphics off;
ods html close;
```

Standardized Score Processes for three covariates by Survival Time



Standardized Score Process

- PH statistic is sensitive to alternatives for which covariates have a monotonically increasing or decreasing effect over time
- Lin (1993) showed this PH statistic is consistent against non-proportional hazards

Supremum Test for Proportionals Hazards Assumption

Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
trt	5.3397	1000	46163	<.0001
x1	1.2089	1000	46163	0.1530
x2	1.2693	1000	46163	0.1750

Therneau 1990

Supremum Test for Proportional Hazards Assumption

Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsValue
fin	0.5423	1000	974156000	0.972
age	1.8135	1000	974156000	0.441
race	0.9424	1000	974156000	0.749
wexp	1.3007	1000	974156000	0.641
mar	0.9368	1000	974156000	0.781
paro	0.5385	1000	974156000	0.975
prio	0.6172	1000	974156000	0.943

Schoenfeld Partial Residuals Test

$$z = \rho\sqrt{(n_u - 2)/(1-\rho^2)}$$

- Harrell (1986) developed simple test
- Based on Fisher's *z*-transform of Pearson's correlation between
  - The partial residuals and
  - Rank order of time

- This test statistic tends to be positive if the ratio of the hazards for high values of the covariate increases over time
- Otherwise, it is negative if the hazard ratio decreases over time

- Easily calculated in a data step, and not available in the PHREG procedure

Covariate by Time Interaction

- Cox (1972) proposed adding a time-dependent interaction variable to the model and test its significance.
- The Partial Likelihood Function has the same form with and woithout these time-dependent variables

$$h(t; X(t)) = h_0(t)\, exp\left\{\sum_{i=1}^{p} \beta_i X_i + \sum_{i=1}^{p} \gamma_i X_i g_i(t)\right\}$$

  - where, $g_i(t)$ is some non-zero function of time
  - The hazard ratio is constant for all t only when $\gamma_i=0$.

Covariate by Time Interaction

- To test the null hypothesis that γ=0 compute the likelihood ratio test

$$-2\ln\left[\frac{L\left(\widehat{\beta},\mathbf{0}\right)}{L\left(\widehat{\beta},\widehat{\gamma}\right)}\right] \sim X^2$$

- The creation of this interaction with time is complex data manipulation because that value changes.
- The PHREG procedure is exceptional for creating these variables because it provides a rich subset of DATA step operators and functions

Covariate by Time Interaction

```
*** nph06s01 ***;
proc phreg data = ads01;
     model survtime*event(0) = x1 x2 x3 trt trttime
     / ties = efron;
     trttime = trt*logsurvtime;
     run;
```

Covariate by Time Interaction for Crossing Hazards (Case 6)

Output Appendix 2.
Table 8.

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
x1	1	-0.80131	0.01273	3962.2248	<.0001	0.449
x2	1	1.61956	0.01625	9935.5621	<.0001	5.051
x3	1	0.01907	0.01099	3.0130	0.0826	1.019
trt	1	0.11559	0.02313	24.9682	<.0001	1.123
x1time	1	-0.00327	0.00244	1.7974	0.1800	0.997

Output Appendix 3.
Table 9.

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
x1	1	-0.90746	0.01236	5388.3594	<.0001	0.404
x2	1	1.80893	0.01735	10868.2407	<.0001	6.104
x3	1	0.08778	0.01018	74.3177	<.0001	1.092
trt	1	-0.54525	0.02627	430.7541	<.0001	0.580
trttime	1	0.60965	0.01208	2549.0946	<.0001	1.840

Categorical Interaction

- Interaction effect might not be linear, so categorize the time-dependent variable:

```
*** nph06s01 ***;
proc rank data = ads01 out = ads04 group = 5;
     var survtime;
     ranks rsurvtime;
     run;

title4 'Test of Interaction between trt and time using Dummy Variables';
proc phreg data = ads04;
     model survtime*event(0) = x1 x2 x3
           trttime0 trttime1 trttime2
           trttime3 trttime4
           / ties = efron;
     trttime0 = trt * rsurvtime_0;
     trttime1 = trt * rsurvtime_1;
     trttime2 = trt * rsurvtime_2;
     trttime3 = trt * rsurvtime_3;
     trttime4 = trt * rsurvtime_4;
     run;
```

- Number of intervals should be
  - subject-matter based
  - Relatively equal number of events and censored observations
  - to be keep standard errors similar

Categorical Interaction

Output Appendix 4.
Table 11.

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
x1	1	-0.85988	0.01307	4327.4703	<.0001	0.423
x2	1	1.72533	0.01952	7811.5800	<.0001	5.614
x3	1	0.02681	0.01075	6.2132	0.0127	1.027
trttime0	1	-0.21911	0.06389	11.7617	0.0006	0.803
trttime1	1	-0.11454	0.04248	7.2682	0.0070	0.892
trttime2	1	-0.03552	0.03583	0.9824	0.3216	0.965
trttime3	1	0.10308	0.03430	9.0299	0.0027	1.109
trttime4	1	0.58883	0.04436	176.1789	<.0001	1.802

Comparisons of PH Test Performance

- Ng'andu (1997) showed that the test statistics
  - The Score Process test
  - The Schoenfeld Partial Residuals test
  - Covariate by time interaction test
- Are practically equally powerful
- The Test for Continuous Interaction with time has the advantage of its simplicity

Modeling Non-Proportionality

- Superior Estimates of the time-dependent covariate
- Improved Precision for the other covariates in the model

Output Appendix 3.
Table 9.

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
x1	1	-0.90746	0.01236	5388.3594	<.0001	0.404
x2	1	1.80893	0.01735	10868.2407	<.0001	6.104
x3	1	0.08778	0.01018	74.3177	<.0001	1.092
trt	1	-0.54525	0.02627	430.7541	<.0001	0.580
trttime	1	0.60965	0.01208	2549.0946	<.0001	1.840

Analysis of Maximum Likelihood Estimates

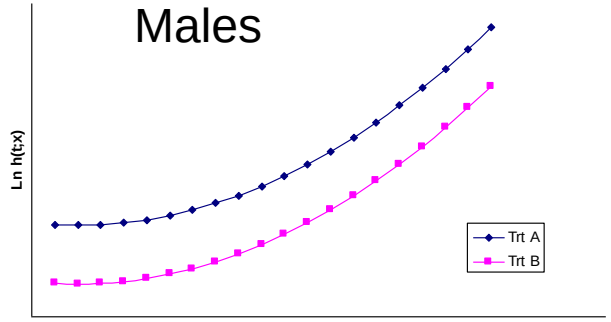
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
x1	1	-0.75099	0.01258	3563.9049	<.0001	0.472
x2	1	1.53452	0.01705	8100.7074	<.0001	4.639
x3	1	-0.01535	0.01178	1.6977	0.1926	0.985
trt	1	-0.00532	0.02256	0.0556	0.8136	0.995

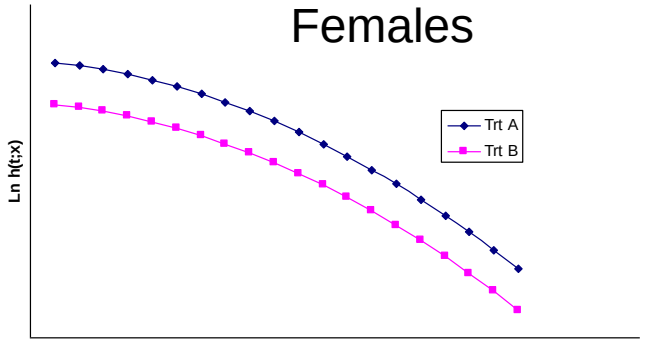
Stratification

- Stratification allows the baseline hazard function to vary across strata
- When the covariate which is non-proportional
  - Is categorical
  - Not of direct interest
  - Too difficult to model (functional form)
- Limited method
  - Cannot include a variable as a covariate and as a stratification
  - Stratified PH models are used when the stratification variables are known to affect the outcome but the estimates of the effects are considered to be of secondary importance (Hosmer and Lemeshow 1999)

Stratified Models

- $\boldsymbol{h_i}\,(t;\boldsymbol{x}) = \boldsymbol{h_{oi}}(t)\exp\{\beta_1 x_1 + \ldots + \beta_k x_k\}$
- Creates a different baseline hazard for each stratum, but the same covariate estimates





- Male and females have a non-PH relationship
- The treatment effect is PH within both males and females
- The hazard ratio is the same in both groups

Stratified Models

- SAS Code

```
proc phreg data = recid;
     model week*arrest(0) = fin / ties = exact;
     strata agecat;
     run;
```

- In the strata statement you can have more than one covariate but it must be categorical or categorized like age above
- Or use the flexibility of the strata statement

```
strata age (10 to 30 by 10);
```

- A separate strata is created from each combination
- There are four strata with 2 bi-level variables

```
strata sex fin;
```

- If you over-stratify, you lose power and information

Stratified Models with Interaction

- In the regular stratified model, there are different hazard functions in each stratum, but the same hazard ratios for the covariates
- You might suspect that a covariate effect differs over a strata
- If so, add an interaction term for strata and covariate to the stratified model
- SAS Code

```
proc phreg data = recid;
     model week*arrest(0) = fin fin*race;
     strata race;
     run;
```

- This model is equivalent to fitting separate models for each level of race with “fin” as the only covariate
- Test the Interaction coefficient if determine if “fin” is different for race.

Stratification vs. Time*Covariate Interactions

- Time x Covariate Interaction
  - Must choose form of time dependency, e.g. x*t vs. x*log(t)
  - The parameter estimate, $\beta_{2,}$ is easily interpretable for clients
- Stratification
  - Takes less computational time
  - Models any non-PH relationship; no need to choose form
  - No inference is possible for the stratification variable
    - Only makes sense for “nuisance variables”
  - You cannot stratify by a variable and also include it as a covariate
  - Can be useful in modeling clustered data. Use a different strata for each cluster

Principles in Survival Analysis

- Plot your data
- Plot your survivor function
- Identify outliers
- Check for Missing Values
- Check for Informative Censoring
- Examine proportional hazards assumption
- Evaluate the functional form of covariates
- Cautiously interpret tests and confidence intervals