Part 6
Lecture 1 Survival Analysis - Parametric

1

Medical Imaging
UNIVERSITY OF TORONTO

MiDATA

Who I am...

Pascal Tyrrell, PhD

Associate Professor

Department of Medical Imaging, Faculty of Medicine

Institute of Medical Science, Faculty of Medicine

Department of Statistical Sciences, Faculty of Arts and Science

ONE OF 2 PROTECTIVE CREAMS WAS APPLIED TO THE ARM OF EACH PERSON IN A STUDY. A PAINFUL IRRITANT WAS ADDED TO THE SAME ARM

CREAM A   HAD TIME UNTIL A PERSON EXPERIENCED PAIN ON 17 PEOPLE

  N=19   HAD TIME LAST RECORDED ON 2 PEOPLE STILL WITH NO PAIN

CREAM B   HAD TIME UNTIL A PERSON EXPERIENCED PAIN ON 13 PEOPLE

  N=19   HAD TIME LAST RECORDED ON 6 PEOPLE STILL WITH NO PAIN

COMPARING TIME TO DEATH

Drug A   Time to cure for 17 people. Last seen alive on 2 people

Drug B   Time to cure for 13 people. Last seen alive on 6 people

COMPARING TIME TO EVENT

-Headaches...

```
OPTIONS PS=65 LS=100 NODATE NONUMBER ;
DATA HEADACHE ;
INPUT MINUTES GROUP CENSOR @@ ;   DATALINES  ;
11 1 0    12 1 0    19 1 0    19 1 0    19 1 0    19 1 0    21 1 0
20 1 0    21 1 0    21 1 0    20 1 0    21 1 0    20 1 0    21 1 0
25 1 0    27 1 0    30 1 0    14 2 0    16 2 0    16 2 0    21 2 0
21 2 0    23 2 0    23 2 0    23 2 0    23 2 0    23 2 0    24 2 0
24 2 0    30 2 0    21 1 1    24 1 1    25 2 1    26 2 1    32 2 1
30 2 1    32 2 1  20 2 1
; RUN ;


PROC SORT DATA = HEADACHE ; BY CENSOR GROUP ; RUN ;
PROC MEANS DATA = HEADACHE N MEAN STDDEV CLM ;
VAR MINUTES ; RUN ;
PROC MEANS DATA = HEADACHE N MEAN STDDEV CLM ;
BY GROUP ; VAR MINUTES ; RUN ;
PROC MEANS DATA = HEADACHE N MEAN STDDEV CLM ;
BY CENSOR GROUP ; VAR MINUTES ; RUN ;
```

MiDATA

The MEANS Procedure
Analysis Variable : MINUTES

N	Mean	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
38	22.03	4.89	20.42	23.63

GROUP	N	Mean	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
1	19	**20.58**	4.34	18.49	22.67
2	19	**23.47**	5.09	21.02	25.93

CALCULATING MEANS BY CENSOR AND GROUP VARIABLES

The MEANS Procedure   Analysis Variable : MINUTES

CENSOR	GROUP	N	Mean	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
0	1	17	20.35	4.51	18.03	22.67
	2	13	21.62	4.21	19.07	24.16
1	1	2	22.50	2.12	3.44	41.56
	2	6	27.50	4.72	22.54	32.46

Medical Imaging
UNIVERSITY OF TORONTO

MiDATA

The LIFEREG Procedure

```
PROC LIFEREG DATA = HEADACHE ;  CLASS GROUP ;
MODEL  MINUTES  *  CENSOR( 1 )  =  GROUP  ;
RUN    ;
```

The SAS System

The LIFEREG Procedure

Model Information	
Data Set	WORK.HEADACHE
Dependent Variable	Log(MINUTES)
Censoring Variable	CENSOR
Censoring Value(s)	1
Number of Observations	38
Noncensored Values	30
Right Censored Values	8
Left Censored Values	0
Interval Censored Values	0
Number of Parameters	3
Name of Distribution	Weibull
Log Likelihood	-9.37930239

Number of Observations Read	38
Number of Observations Used	38

Class Level Information		
Name	Levels	Values
GROUP	2	1 2

DATASET WITH ALL DATA HAS 38 SUBJECTS

Type III Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
GROUP	1	6.0540	0.0139

Analysis of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**95% Confidence Limits**		**Chi-Square**	**Pr > ChiSq**
Intercept		1	3.3091	0.0589	3.1938	3.4245	3161.70	<.0001
GROUP	1	1	-0.1933	0.0786	-0.3473	-0.0393	6.05	0.0139
GROUP	2	0	0.0000	.	.	.	.	.
Scale		1	0.2122	0.0304	0.1603	0.2809		
Weibull Shape		1	4.7128	0.6742	3.5604	6.2381		

DATASET WITH NO CENSORED DATA HAS 30 SUBJECTS

```
DATA  N30  ;     SET HEADACHE  ;
IF CENSOR = 0  ;   RUN  ;
PROC  LIFEREG  DATA  =  N30    ;
MODEL  MINUTES  =  GROUP   ;  RUN ;
```

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
GROUP	1	0.3239	0.5693

Analysis of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.0619	0.1047	2.8568	3.2671	855.79	<.0001
GROUP	1	0.0388	0.0683	-0.0949	0.1726	0.32	0.5693
Scale	1	0.1850	0.0252	0.1417	0.2415		
Weibull Shape	1	5.4052	0.7350	4.1406	7.0560		

Recall that the two means were similar: 20.35 and 21.62 minutes.

Medical Imaging
UNIVERSITY OF TORONTO

MiDATA

DATASET WITH ONLY CENSORED DATA HAS 8 SUBJECTS

```
DATA   N8   ;    SET  HEADACHE;    IF  CENSOR = 1  ;     RUN  ;
PROC   LIFEREG     DATA = N8 ;    MODEL   MINUTES = GROUP ;   RUN ;
```

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
GROUP	1	8.2856	0.0040

Analysis of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	2.8775	0.1579	2.5681	3.1870	332.20	<.0001
GROUP	1	0.2533	0.0880	0.0808	0.4258	8.29	0.0040
Scale	1	0.1072	0.0321	0.0596	0.1928		
Weibull Shape	1	9.3262	2.7920	5.1865	16.7699		

Recall that the two means were similar: 22.50 and 27.50 minutes.

MiDATA

THE VALUE OF CENSORED DATA

Suppose the researcher thought the whole idea of using CENSORED observations was too complicated.

They thought the idea of using survival times in which the event did not even occur was strange.

However, if they chose to use only the 30 uncensored times the p value was 0.5693. They would have missed the significant finding (p = 0. 0139).

A little math....

❑ ***Survivorship function:*** probability of surviving past a given time (event-free past t); probability, must be positive and between 0 & 1.

$$S(t) = PROB(T > t)$$

❑ ***Probability density function:*** Unconditional probability that the event will occur at EXACT time (between t and t + delta); must be positive and between 0 & 1

$$f(t) = \lim_{\Delta t \to 0} \frac{\Pr\{t \le T \le t + \Delta t\}}{\Delta t}$$

❑ ***Hazard function:*** Instantaneous risk that an event will occur at a given time, GIVEN no event up until that time (i.e., conditional); because the hazard function is NOT a probability it can be greater than 1; but must be positive

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr\{t \le T \le t + \Delta t / T \ge t\}}{\Delta t}$$

General Parametric Model – PROC LIFEREG

Accelerated Failure Time (AFT) models – models that can be linearized by taking logs

$$\log T_i = \beta_o + \beta_{1xi1} + \ldots + \beta_{pxip} + \sigma\varepsilon$$

where T_i is the failure time for i subjects (modelled as a function of p covariates)

Available Distributions:

Distribution of T	Distribution of error term
Weibull	Extreme value (2 parameters)
Exponential	Extreme value (1 parameter)
Gamma	Log gamma
Log-logistic	Logistic
Log-normal	Normal

Medical Imaging
UNIVERSITY OF TORONTO

MiDATA

Advantages of Parametric Models:

❑ Allow testing hypotheses about the shape of the hazard function

❑ Use the method of full maximum likelihood estimation

❑ More efficient estimate are provided if the shape of the survival distribution is known

❑ Regression coefficients can be clinically meaningful, and can be related to those from a PH model

❑ Can provide estimates of survival time

❑ Residuals can be computed (i.e., differences between observed and predicted survival times)

❑ Can control for confounding variables