Statistics in Evidence Based Medicine II

Lecture 4: Survival Analysis

Rizwana Rehman, PhD
Regional Statistician
Southeast Epilepsy Centers of Excellence
Durham VA Medical Center, Durham NC
Rizwana.Rehman@va.gov
(919)286-0411 ext: 5024

Audio Information: Dial 1-855-767-1051
Conference ID 61304911

Epilepsy Centers of Excellence
SOUTHEAST

Text Books

- **Main:** **Statistics at Square One 12th edition (2010)**

  M J Campbell & T D V Swinscow

  http://www.phsource.us/PH/EPI/Biostats/

- **Secondary:** **Basic and Clinical Biostatistics (2004)**

  Beth Dawson, Robert G. Trapp

  http://www.accessmedicine.com/resourceTOC.aspx?resourceID=62

- For more information, program materials, and to complete evaluation for CME credit visit

  www.epilepsy.va.gov/Statistics

Audio Information: Dial 1-855-767-1051
Conference ID 61304911

Overview

- Vocabulary
- Kaplan-Meier Curve
- Log Rank Test
- Assumptions
- Reading and Reporting the Results
- Appendix- Survival Analysis with Openstat

What is Survival Analysis

- Survival analysis is a collection of statistical techniques for data analysis where the outcome variable is time until an event occurs.
- Survival data could come from clinical trials, epidemiological observational studies or lab experiments on animals.

Examples

- Time from birth to death
- Time from treatment of lung cancer to death
- The time from treatment until eradication of infection among patients who are treated with an antibiotic
- Disease free cohort/time until heart disease

Censoring

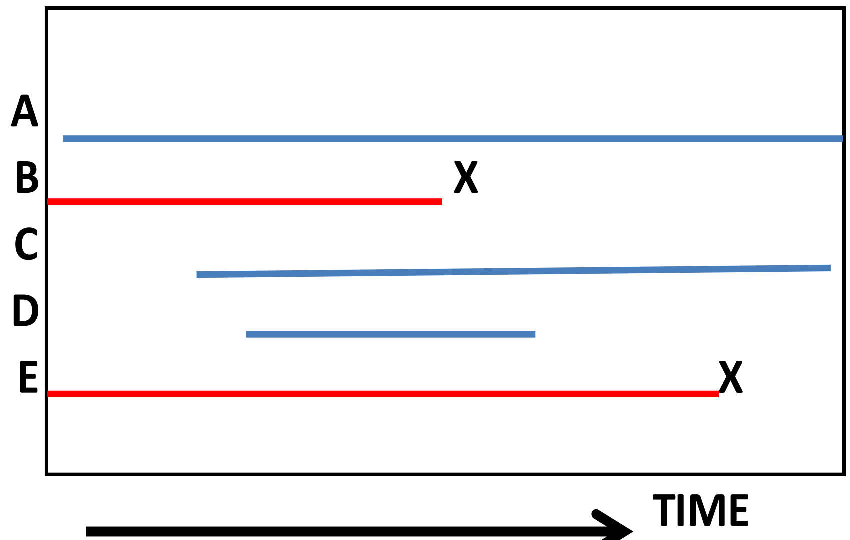
- Data are subject to censoring when the study ends before the event occurs.
  - A person does not experience the event during study.
  - A person is lost to follow up.
  - A person withdraws from the study.
  - A record is lost after a certain time.

Types of Censored Data

- Right Censored
  - True survival time is equal to or greater than observed survival time.
- Left Censored
  - True survival time is less than or equal to the observed survival time.
- Interval Censored
  - True survival time is within a known time interval.

Analysis of Survival Data

Standard techniques are not suitable for survival analysis.



X = Event Occus

Vocabulary of Survival Analysis

- Survival Time $T (T \geq 0)$- time individual has survived over some follow-up period.
  - A specific value of T is denoted by *t*
- Event
  - Death, disease incidence, relapse from remission, recovery
  - The start and end of event must be clearly identified

Vocabulary of Survival Analysis

- Survival Function $S(t)$
  - Probability a person has survived at least to time t $S(t)=P[T>t]$
  - The graph of $S(t)$ against t is called a survival curve.
- Hazard function $h(t)$
  - Hazard rate is simply the death rate; chance that a person will die today given that person was alive at the beginning of the day

Goals of Survival Analysis

- To estimate and interpret survivor and or hazard functions
- To compare survivor or hazard functions
- To assess the relationship of explanatory variables to survival time

Kaplan-Meier Curve

- Probability of surviving k or more periods (cumulative proportion surviving) entering the study is a product of the k observed survival rates for each period.

  $$S(k)= p_1 \times p_2 \times p_3 \times \text{---------} \times p_k$$

- The proportion surviving period i having survived up to period i is given by

  $$p_i=(r_i-d_i)/r_i$$

Rules

- The survival times including the censored values should be ordered in increasing duration.
- If a censored time has the same value as an uncensored time, then the uncensored should precede the censored.
- At each event (i) work out the number alive r_i before the event.
- Censoring does not alter the cumulative surviving.

Kaplan-Meier Example

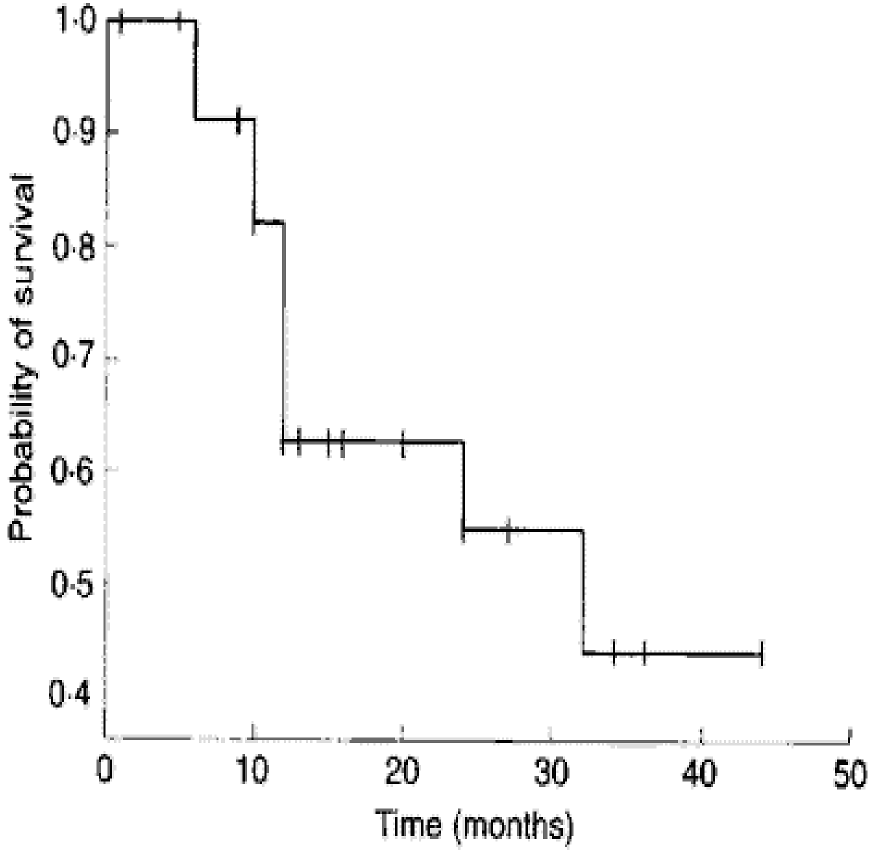
Table 12.1 Survival in 49 patients with Dukes' C colorectal cancer randomly assigned to either γ linoleic acid or control treatment	
Treatment	Survival time (months)
γ linoleic acid (n=25)	1+, 5+, 6, 6, 9+, 10, 10, 10+, 12, 12, 12, 12, 12+, 13+, 15+, 16+, 20+, 24, 24+, 27+, 32, 34+, 36+, 36+, 44+
Control (n=24)	3+, 6, 6, 6, 6, 8, 8, 12, 12, 12+, 15+, 16+, 18+, 18+, 20, 22+, 24, 28+, 28+, 28+, 30, 30+, 33+, 42

Statistics at Square one

Kaplan-Meier Example

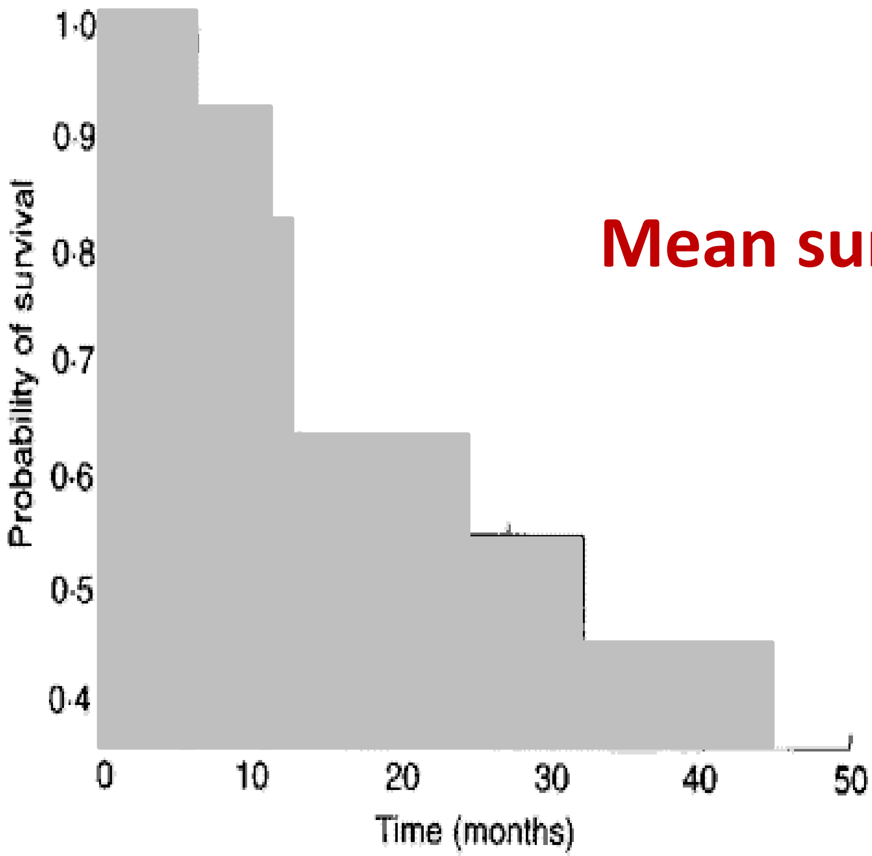
Case (i)	Survival time (months) (t_i)	Number alive (r_i)	Deaths (d_i)	Proportion surviving (ri-di)/ri	Cumulative Proportion S(t)
	0	0	0	-	1
1	1+	25	0	1	1
2	5+	24	0	1	1
3	6	23	2	0.9130	0.9130
4	6				
5	9+	21	0	1	0.9130
6	10	20	2	0.90	0.8217
7	10				
8	10+				
9	12	17	4	0.7647	0.6284
10	12				
11	12				
12	12				
13	12+				
14	13+	12	0	1	0.6284
15	15+	11	0	1	0.6284
16	16+	10	0	1	0.6284
17	20+	9	0	1	0.6284
18	24	8	1	0.875	0.5498
19	24+				
20	27+	6	0	1	0.5498
21	32	5	1	0.80	0.4399
22	34+				
23	36+				
24	36+				
25	44+				

Kaplan-Meier Survival Estimate Curve



Statistics at Square one

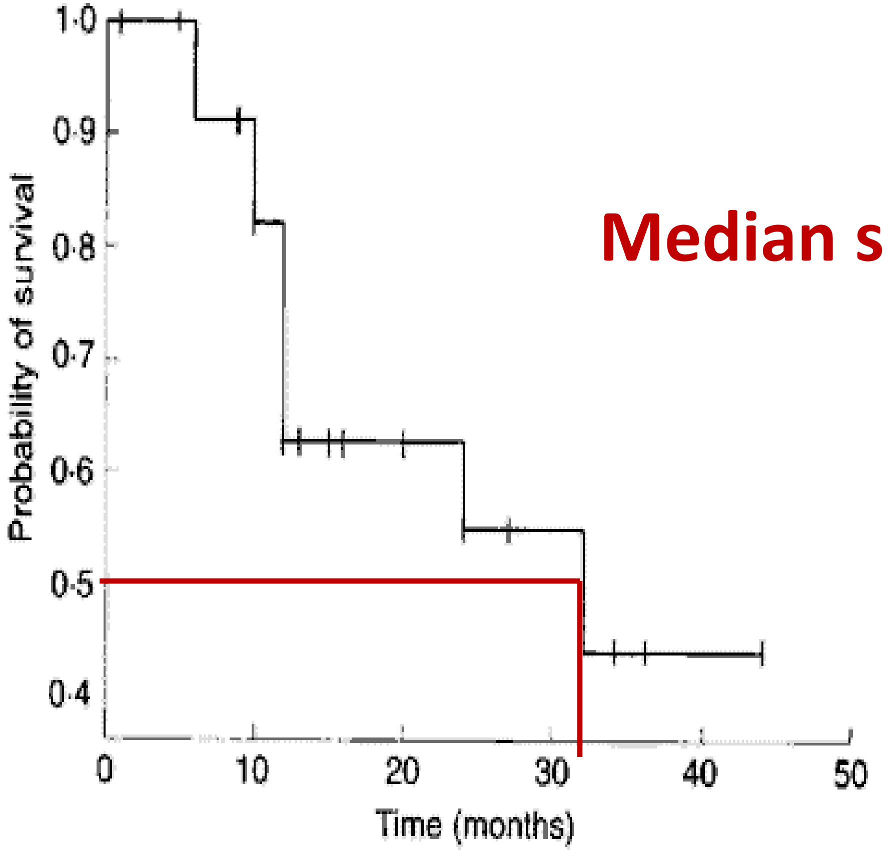
Estimation of Mean time from Kaplan-Meier Curve (Area under the curve)



Mean survival time≈29 months

Statistics at Square one

Estimation of Median time from Kaplan-Meier Curve (Time at which S(t)=0.50)



Median survival time≈32 months

Statistics at Square one

Log Rank Test to Compare Two Survival Functions

- H_0: Two groups (treatments) don't differ with respect to population survival functions.
- H_A: There is a difference between two survival functions.
- With O_A and O_B as actual number of events and E_A and E_B as expected number of events

$$X^2 = \frac{(O_A - E_A)^2}{E_A} + \frac{(O_B - E_B)^2}{E_B}$$

Computing E_A AND E_B

- Order the data for two groups combined.
- Consider each event starting at time t=0.
- At each event at time t_i consider the total alive (r_i) and total number still alive in group A (r_{Ai}) up to that point.

  $$E_{Ai}=(r_{Ai}/r_i)\times d_i$$

  $$E_A=\text{Sum}\ (E_{Ai})$$

- $E_B=n-E_A$ n=total number of events

Dukes' C Colorectal Cancer Example

Table 12.1 Survival in 49 patients with Dukes' C colorectal cancer randomly assigned to either γ linoleic acid or control treatment	
Treatment	Survival time (months)
γ linoleic acid (n=25)	1+, 5+, 6, 6, 9+, 10, 10, 10+, 12, 12, 12, 12, 12+, 13+, 15+, 16+, 20+, 24, 24+, 27+, 32, 34+, 36+, 36+, 44+
Control (n=24)	3+, 6, 6, 6, 6, 8, 8, 12, 12, 12+, 15+, 16+, 18+, 18+, 20, 22+, 24, 28+, 28+, 28+, 30, 30+, 33+, 42

Statistics at Square one

Calculations

Table 12.3 Calculation of log rank statistics for 49 patients randomly assigned to receive γ linoleic acid (A) or control (B)

Survival time (months) t_i	Group	Total at risk r	Number of events d_i	Total at risk in group A r_{Ai}	Expected number of events E_{Ai}
0		49			
1+	A	49	0	25	0
3+	B	48	0	24	0
5+	A	47	0	24	0
6	A	46	6	23	3.0
6	A				
6	B				
6	B				
6	B				
6	B				
8	B	40	2	21	1.05
8	B				
9+	A	38	0	21	0
10	A	37	2	20	1.0811
10	A				
10+	A				
12	A	34	6	17	3.0
12	A				
12	A				
12	A				
12	B				
12	B				
12+	A				

Calculations

$$X^2 = \frac{(O_A - E_A)^2}{E_A} + \frac{(O_B - E_B)^2}{E_B}$$

$$X^2 = \frac{(10 - 11.37)^2}{11.37} + \frac{(12 - 10.63)^2}{10.63} = 0.34$$

$$p = 0.44$$

Little evidence that this result would not have arisen by chance

Hazard Ratio

Hazard Ratio is a measure of the relative survival experience in the two groups.

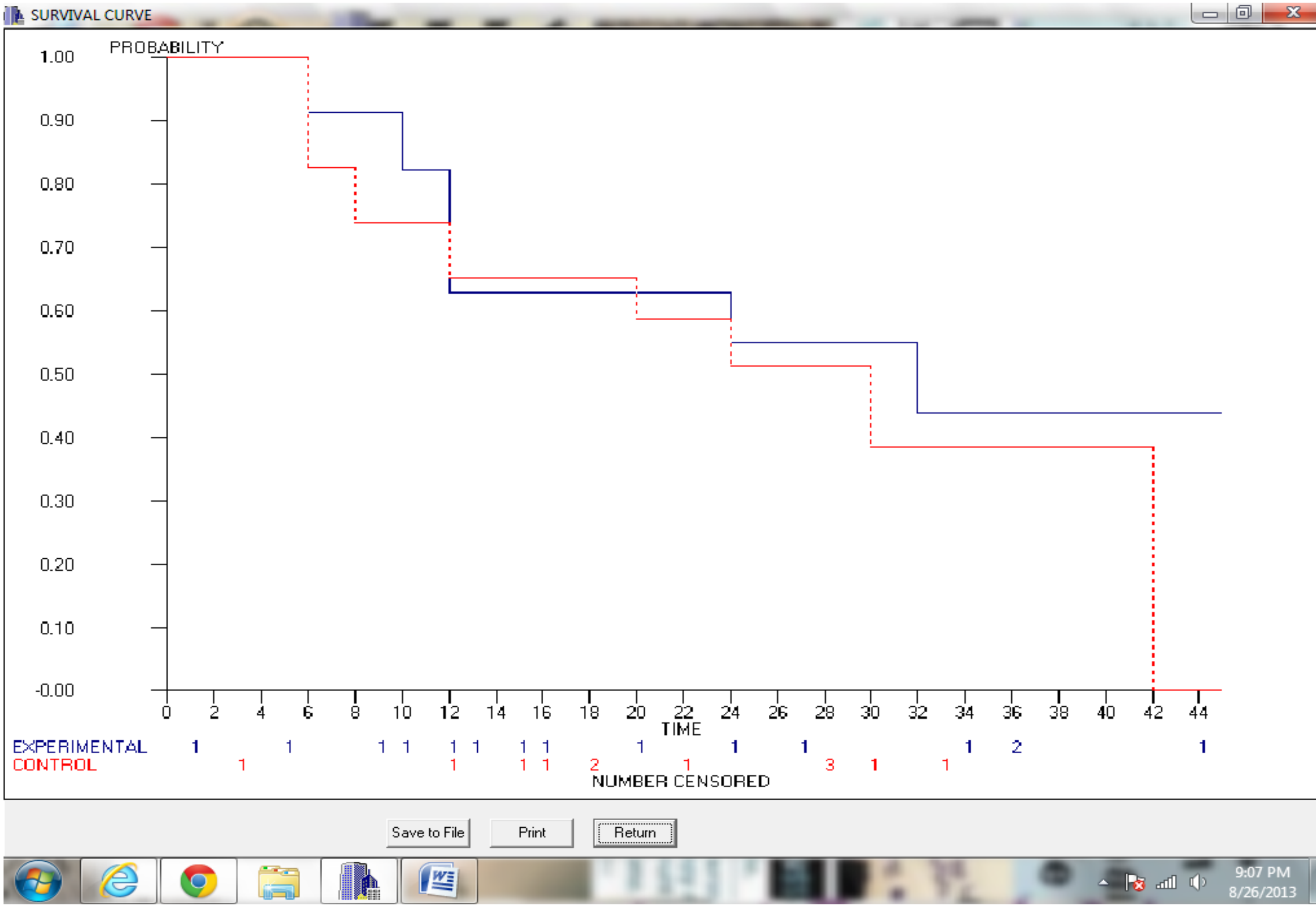
$$HR = \frac{O_A / E_A}{O_B / E_B} = \frac{10 / 11.37}{12 / 10.63} = 0.779$$

Survival Analysis in Openstat

22	0	2	9	0.0000	0.5870
24	2	1	8	0.8750	0.5498
24	0	2	7	0.0000	0.5136
24	0	1	7	0.0000	0.5498
27	0	1	6	0.0000	0.5498
28	0	2	7	0.0000	0.5136
28	0	2	6	0.0000	0.5136
28	0	2	5	0.0000	0.5136
30	1	2	4	0.7500	0.3852
30	0	2	3	0.0000	0.3852
32	1	1	5	0.8000	0.4399
33	0	2	2	0.0000	0.3852
34	0	1	4	0.0000	0.4399
36	0	1	3	0.0000	0.4399
36	0	1	2	0.0000	0.4399
42	1	2	1	0.0000	0.0000
44	0	1	1	0.0000	0.4399

Total Expected Events for Experimental Group = 11.375
Observed Events for Experimental Group = 10.000
Total Expected Events for Control Group = 10.625
Observed Events for Control Group = 12.000
Chisquare = 0.344 with probability = 0.442
Risk = 0.778, Log Risk = -0.250, Std.Err. Log Risk = 0.427
95 Percent Confidence interval for Log Risk = (-1.087,0.586)
95 Percent Confidence interval for Risk = (0.337,1.796)

Kaplan-Meier Curves in Openstat



Assumptions

- Uninformative censoring
- Length of follow up
- Completeness of follow up
- Cohort effect on survival
- Between center differences
- Risk of an event in one group relative to the other does not change with time.

More Advanced Methods

- An extension of log rank test is cox regression.
- Cox regression allows for patients related factors that could potentially affect the survival time of a patient.

Reading and Reporting

- Always report confidence intervals of measures.
  - The log rank test should be presented as

X^2(log rank)=0.34, d.f=1, p=0.44, estimated

relative risk=0.779, 95% confidence interval 0.34 to 1.80

- Don't read too much into the right hand part of a Kaplan-Meier Plot.
- Openstat is a free soft ware that can be used for survival analysis.

Helpful Resources

- Survival Analysis Part I: Basic Concepts and First Analysis

  http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2394262/pdf/89-6601118a.pdf

- Statistics Review: Survival Analysis

  http://www.biomedcentral.com/content/pdf/cc2955.pdf

- A comprehensive book
  - Survival Analysis A self Learning Text

    David G. Kleinbaum & Mitchel Klein

Thank you!

Questions/Comments

Rizwana.Rehman@va.gov

(919) 286-0411 ext: 5024

For more information, program materials, and to complete evaluation for CME credit visit

www.epilepsy.va.gov/Statistics

APPENDIX
Survival Analysis with OpenStat

Dukes' C Colorectal Cancer Example

Table 12.1 Survival in 49 patients with Dukes' C colorectal cancer randomly assigned to either γ linoleic acid or control treatment

Treatment	Survival time (months)
γ linoleic acid (n=25)	1+, 5+, 6, 6, 9+, 10, 10, 10+, 12, 12, 12, 12, 12+, 13+, 15+, 16+, 20+, 24, 24+, 27+, 32, 34+, 36+, 36+, 44+
Control (n=24)	3+, 6, 6, 6, 6, 8, 8, 12, 12, 12+, 15+, 16+, 18+, 18+, 20, 22+, 24, 28+, 28+, 28+, 30, 30+, 33+, 42

Statistics at Square one

Download & Tutorial

- Free download

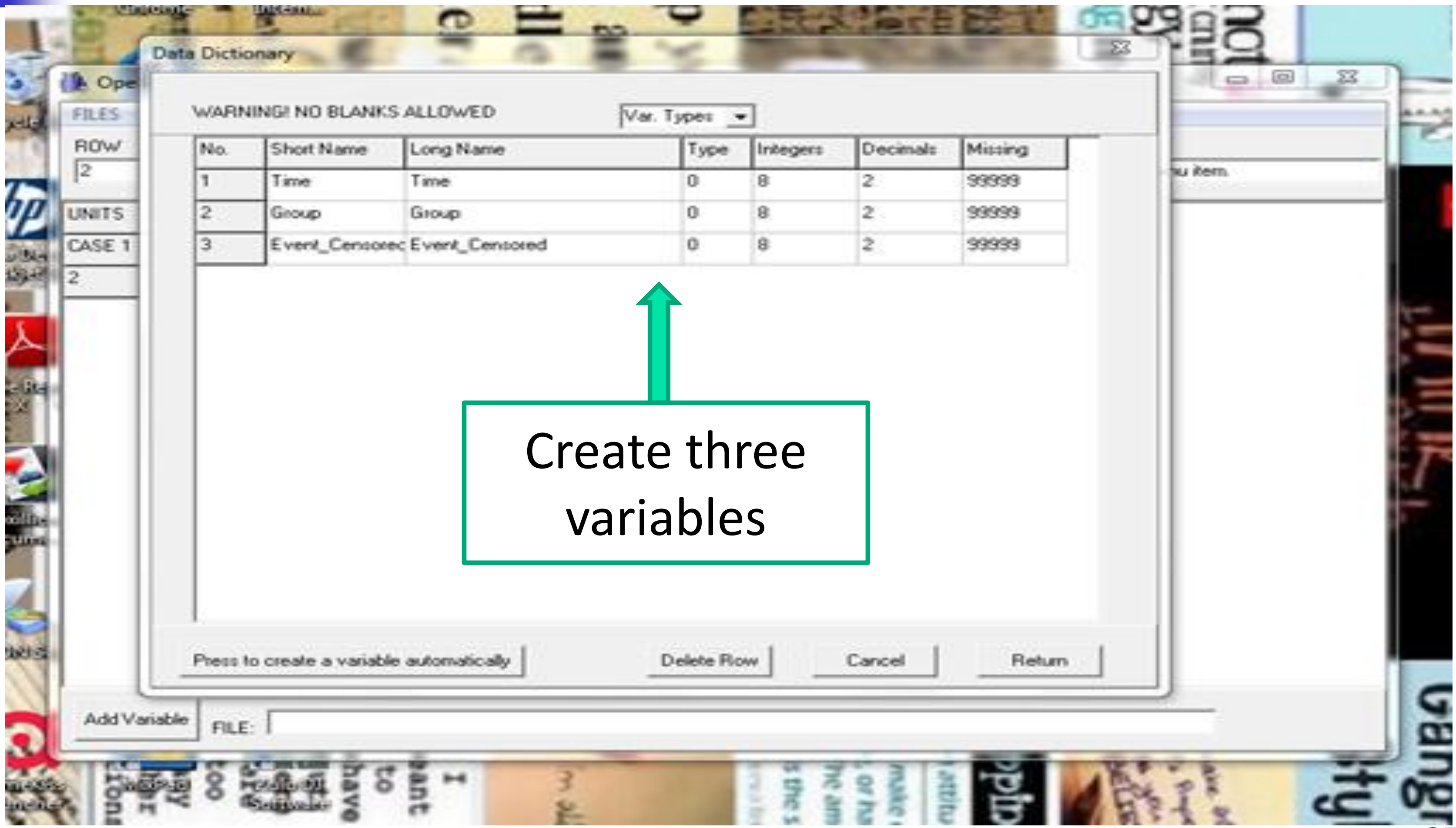
http://www.statprograms4u.com/OpenStatMain.htm

http://openstat.en.softonic.com/

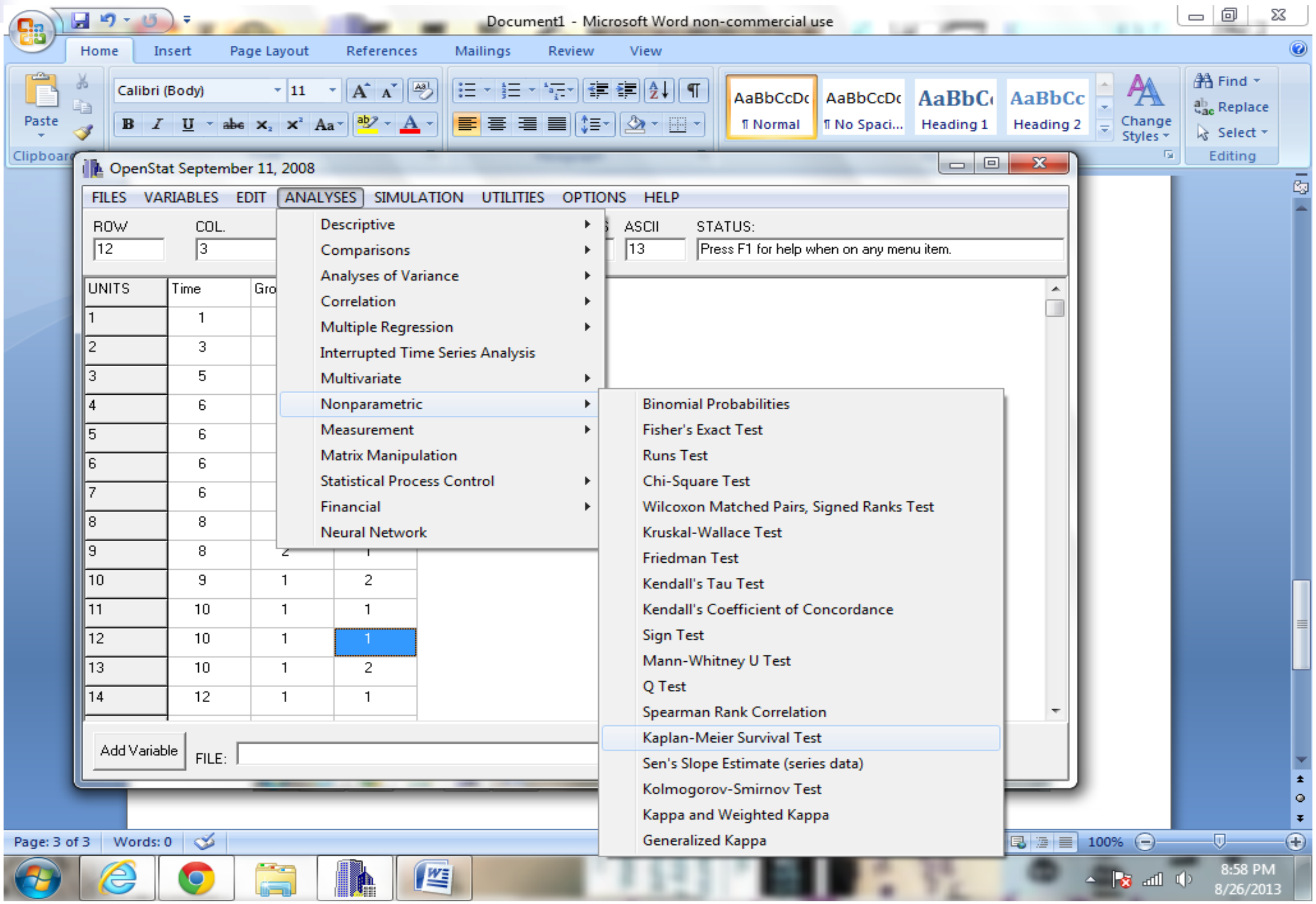
- Tutorial A good resource

http://www.google.com/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=2&ved=0CDcQFjAB&url=http%3A%2F%2Fwww.statprograms4u.com%2FOpenStatReferenceDoc.doc&ei=cpscUr--KMmxsASWrIDYCA&usg=AFQjCNEc964udnzMZ6CnAXsMzNG9K9MD1Q&sig2=PgJo_FkycjMIR4zGaXuI_A

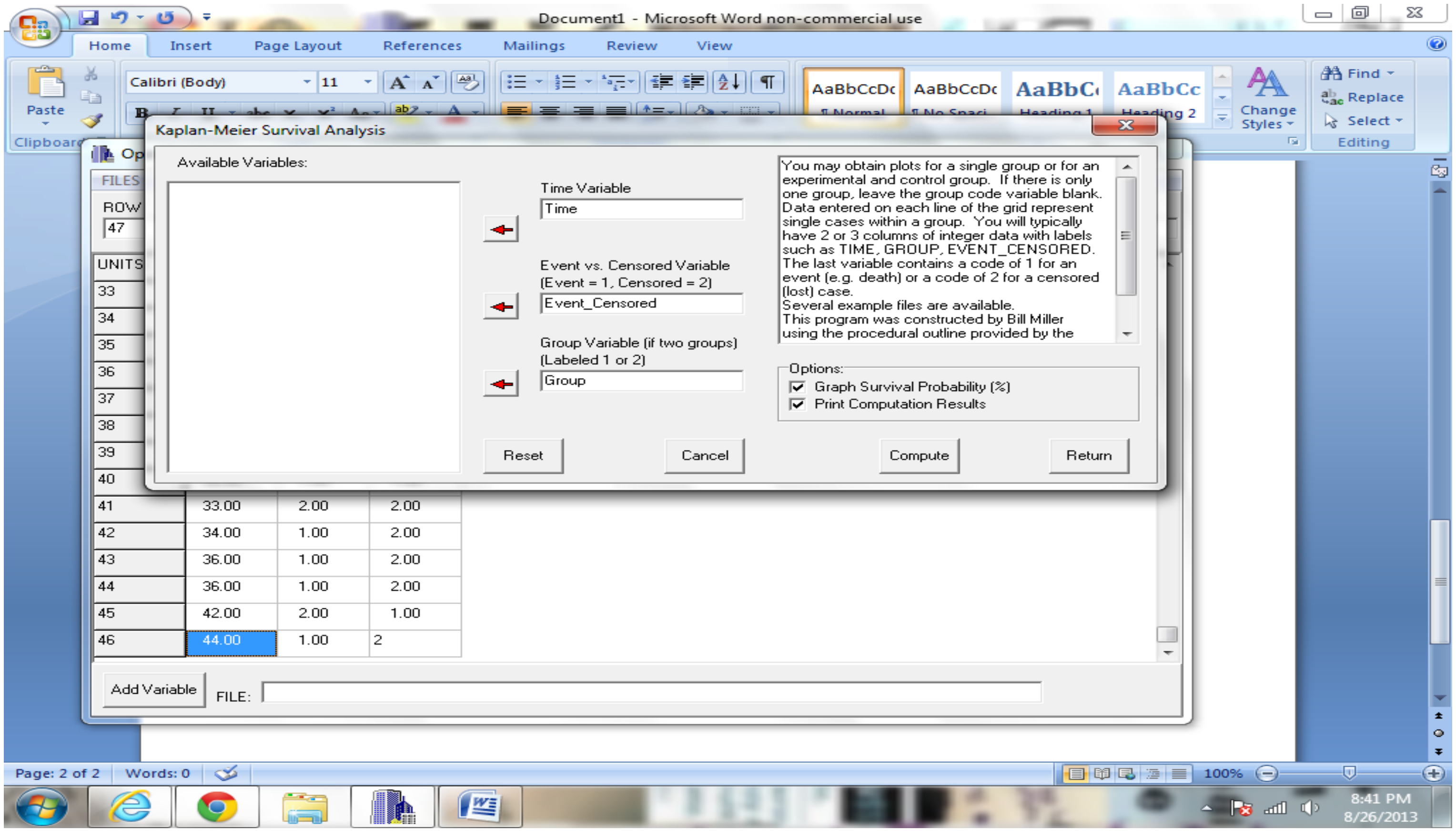
Entering the data



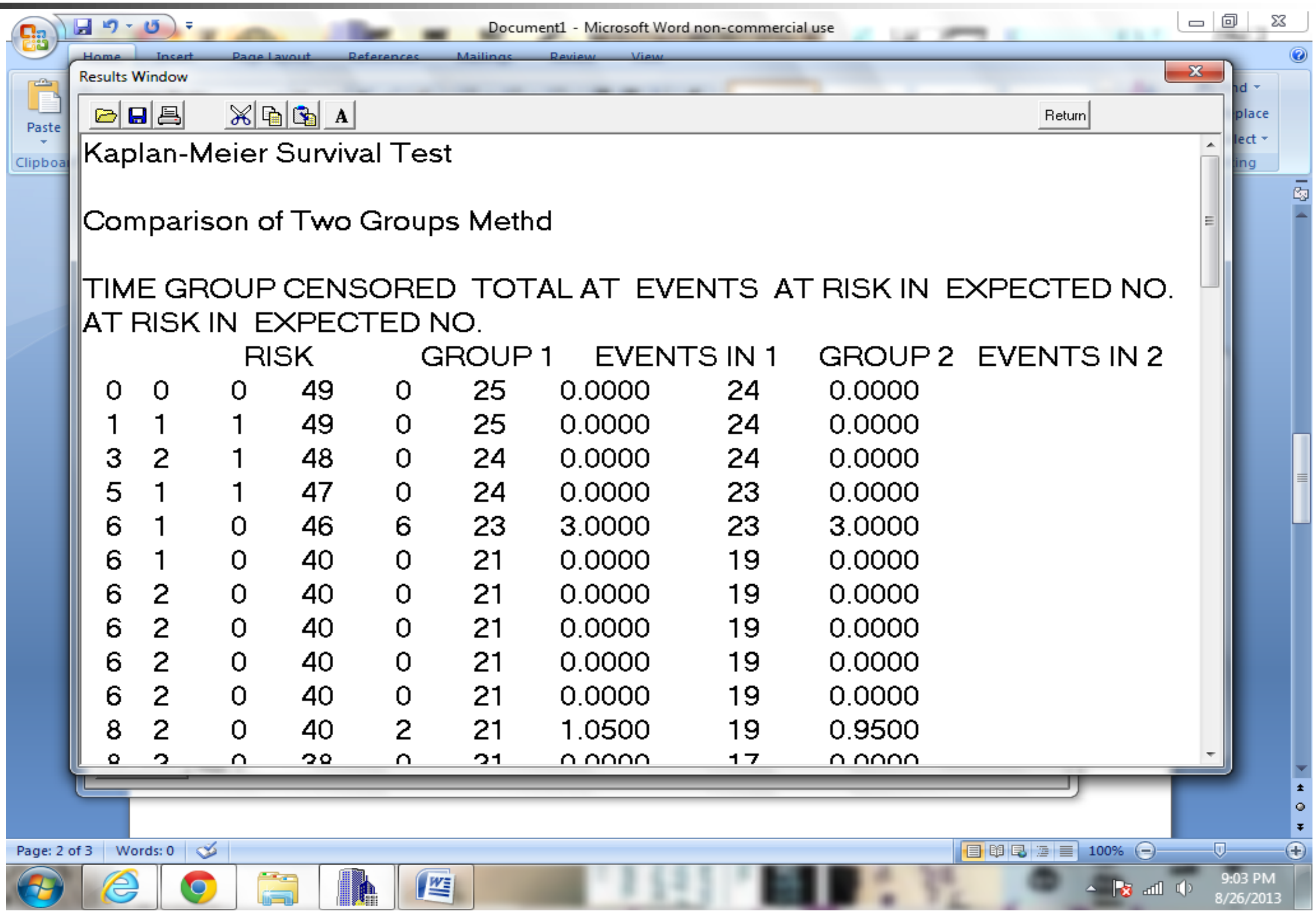
Choosing Kaplan-Meier



Selection of Variables on Dialogue Box



Results



Kaplan-Meier Survival Test

Comparison of Two Groups Methd

TIME GROUP CENSORED TOTAL AT EVENTS AT RISK IN EXPECTED NO. AT RISK IN EXPECTED NO.
RISK GROUP 1 EVENTS IN 1 GROUP 2 EVENTS IN 2

TIME	GROUP	CENSORED	TOTAL AT RISK	EVENTS	AT RISK IN GROUP 1	EXPECTED NO. EVENTS IN 1	AT RISK IN GROUP 2	EXPECTED NO. EVENTS IN 2
0	0	0	49	0	25	0.0000	24	0.0000
1	1	1	49	0	25	0.0000	24	0.0000
3	2	1	48	0	24	0.0000	24	0.0000
5	1	1	47	0	24	0.0000	23	0.0000
6	1	0	46	6	23	3.0000	23	3.0000
6	1	0	40	0	21	0.0000	19	0.0000
6	2	0	40	0	21	0.0000	19	0.0000
6	2	0	40	0	21	0.0000	19	0.0000
6	2	0	40	0	21	0.0000	19	0.0000
6	2	0	40	0	21	0.0000	19	0.0000
8	2	0	40	2	21	1.0500	19	0.9500
8	2	0	38	0	21	0.0000	17	0.0000

Important Results

22	0	2	9	0.0000	0.5870
24	2	1	8	0.8750	0.5498
24	0	2	7	0.0000	0.5136
24	0	1	7	0.0000	0.5498
27	0	1	6	0.0000	0.5498
28	0	2	7	0.0000	0.5136
28	0	2	6	0.0000	0.5136
28	0	2	5	0.0000	0.5136
30	1	2	4	0.7500	0.3852
30	0	2	3	0.0000	0.3852
32	1	1	5	0.8000	0.4399
33	0	2	2	0.0000	0.3852
34	0	1	4	0.0000	0.4399
36	0	1	3	0.0000	0.4399
36	0	1	2	0.0000	0.4399
42	1	2	1	0.0000	0.0000
44	0	1	1	0.0000	0.4399

Total Expected Events for Experimental Group = 11.375
Observed Events for Experimental Group = 10.000
Total Expected Events for Control Group = 10.625
Observed Events for Control Group = 12.000
Chisquare = 0.344 with probability = 0.442
Risk = 0.778, Log Risk = -0.250, Std.Err. Log Risk = 0.427
95 Percent Confidence interval for Log Risk = (-1.087,0.586)
95 Percent Confidence interval for Risk = (0.337,1.796)

Kaplan-Meier Curves