



LOGRANK-TEST

Comparing Survival Functions

دکتر رحیم زاده، آمار زیستی

Log rank test

- To compare two survival curves produced from two groups treatment and control we use the rather curiously named log rank test.
- **The assumptions used in this test are:**
- That the survival times are ordinal or continuous.
- That the risk of an event in one group relative to the other does not change with time. (Parallel Curve)
- Thus, if treatment reduces the risk of death in patients, then this risk reduction does not change with time (the so called proportional hazards assumption).

Log-Rank Test

- The null hypothesis is that there is no difference in survival between the two groups or that there is no difference between the populations in the probability of death at any point.
- The log rank test is a non-parametric test and makes no assumptions about the survival distributions.
- The log rank test compares the observed number of events in each group to what would be expected if the null hypothesis were true (i.e., if the survival curves were identical).
- H_0 : The two survival curves are identical (or $S_{1t} = S_{2t}$) versus
- H_1 : The two survival curves are not identical (or $S_{1t} \neq S_{2t}$, at any time t) ($\alpha=0.05$).

Log-Rank Test (Approximate)

- The log rank statistic is approximately distributed as a chi-square test statistic. There are several forms of the test statistic, and they vary in terms of how they are computed. We use the following:

$$\chi^2 = \sum \frac{(\sum O_{jt} - \sum E_{jt})^2}{\sum E_{jt}}$$

- where $\sum O_{jt}$ represents the sum of the **observed number of events** in the j^{th} group over time (e.g., $j=1,2$) and $\sum E_{jt}$ represents the sum of the **expected number of events** in the j^{th} group over time.
- The sums of the observed and expected numbers of events are computed for each event time and summed for each comparison group.
- The log rank statistic has degrees of freedom equal to $k-1$, where k represents the number of comparison groups. If $k=2$, so the test statistic has 1 degree of freedom.

دکتر رحیم زاده، آمار زیستی

Log-Rank Test

- To compute the test statistic, we need the observed and expected number of events at each event time. The observed number of events are from the sample and the expected number of events are computed assuming that the null hypothesis is true (i.e., that the survival curves are identical).
- To generate the expected numbers of events we organize the data into a life table with rows representing each event time, regardless of the group in which the event occurred. We also keep track of group assignment. We then estimate the proportion of events that occur at each time (O_t/N_t) using data from both groups combined under the assumption of no difference in survival (i.e., assuming the null hypothesis is true).
- We multiply these estimates by the number of participants at risk at that time in each of the comparison groups (N_{1t} and N_{2t} for groups 1 and 2 respectively).

Log-Rank Test

- Specifically, we compute for each event time t , the number at risk in each group, N_{jt} (e.g., where j indicates the group, $j=1, 2$) and the number of events (deaths), O_{jt} , in each group. The table below contains the information needed to conduct the log rank test to compare the survival curves above. Group 1 receive either the brief intervention focused on abstinence from alcohol, and group 2 standard prenatal care.

Example

An investigator wishes to evaluate the efficacy of a brief intervention to • prevent alcohol consumption in pregnancy. Pregnant women with a history of heavy alcohol consumption are recruited into the study and randomized to receive either the brief intervention focused on abstinence from alcohol or standard prenatal care. The outcome of interest is relapse to drinking. Women are recruited into the study at approximately 18 weeks gestation and followed through the course of pregnancy to delivery (approximately 39 weeks gestation). The data are shown below and indicate whether women relapse to drinking and if so, the time of their first drink measured in the number of weeks from randomization. For women who do not relapse, we record the number of weeks from randomization that they are alcohol free.

Standard Prenatal Care			Brief Intervention	
Relapse	No Relapse		Relapse	No Relapse
19	20		16	21
6	19		21	15
5	17		7	18
4	14			18
				5

To compute the test statistic, we organize the data according to event (relapse) times and determine the numbers of women at risk in each treatment group and the number who relapse at each observed relapse time.

Time, Weeks	Number at Risk - Group 1 N_{1t}	Number at Risk - Group 2 N_{2t}	Number of Relapses - Group 1 O_{1t}	Number of Relapses - Group 2 O_{2t}
4	8	8	1	0
5	7	8	1	0
6	6	7	1	0
7	5	7	0	1
16	4	5	0	1
19	3	2	1	0
21	0	2	0	1

Log Rank Test

EXAMPLE

$$O_1 - E_1 = -10.26$$

$$O_2 - E_2 = 10.26$$

Expected cell counts:

$$e_{1f} = \left(\frac{n_{1f}}{n_{1f} + n_{2f}} \right) \times (m_{1f} + m_{2f})$$

$\uparrow$ $\uparrow$
 Proportion # of failures over
 in risk set both groups

$$e_{2f} = \left(\frac{n_{2f}}{n_{1f} + n_{2f}} \right) \times (m_{1f} + m_{2f})$$

Two groups:

$O_2 - E_2$ = summed observed minus expected score for group 2

$$\text{Log-rank statistic} = \frac{(O_2 - E_2)^2}{\text{Var}(O_2 - E_2)}$$

$$\text{Var}(O_i - E_i) = \sum_j \frac{n_{1f} n_{2f} (m_{1f} + m_{2f}) (n_{1f} + n_{2f} - m_{1f} - m_{2f})}{(n_{1f} + n_{2f})^2 (n_{1f} + n_{2f} - 1)}$$

$i = 1, 2$

Log-Rank Test

time	# nf1	#nf2	# mf1	# mf2	E F g1	E F g2	o1-e1	o2-e2	n1f*nf2	m1f+mf2	(n1f+nf2-m1f-m2f)	numariator	deminator	num/dem	
4	8	8	1	0	0.5	0.5	0.5	-0.5	64	1	15	960	3840	0.25	
5	7	8	1	0	0.466667	0.533333	0.533333	-0.53333	56	1	14	784	3150	0.248889	
6	6	7	1	0	0.461538	0.538462	0.538462	-0.53846	42	1	12	504	2028	0.248521	
7	5	7	0	1	0.416667	0.583333	-0.41667	0.416667	35	1	11	385	1584	0.243056	
16	4	5	0	1	0.444444	0.555556	-0.44444	0.444444	20	1	8	160	648	0.246914	
19	3	2	1	0	0.6	0.4	0.4	-0.4	6	1	4	24	100	0.24	
21	0	2	0	1	0	1	0	0	0	1	1	0	4	0	
			4	3	2.889316	4.110684	1.110684	-1.11068							1.477379
							0.835005	0.835005	0.835005	0.835005					

Log-Rank Test (approximation)

We now compute the test statistic:

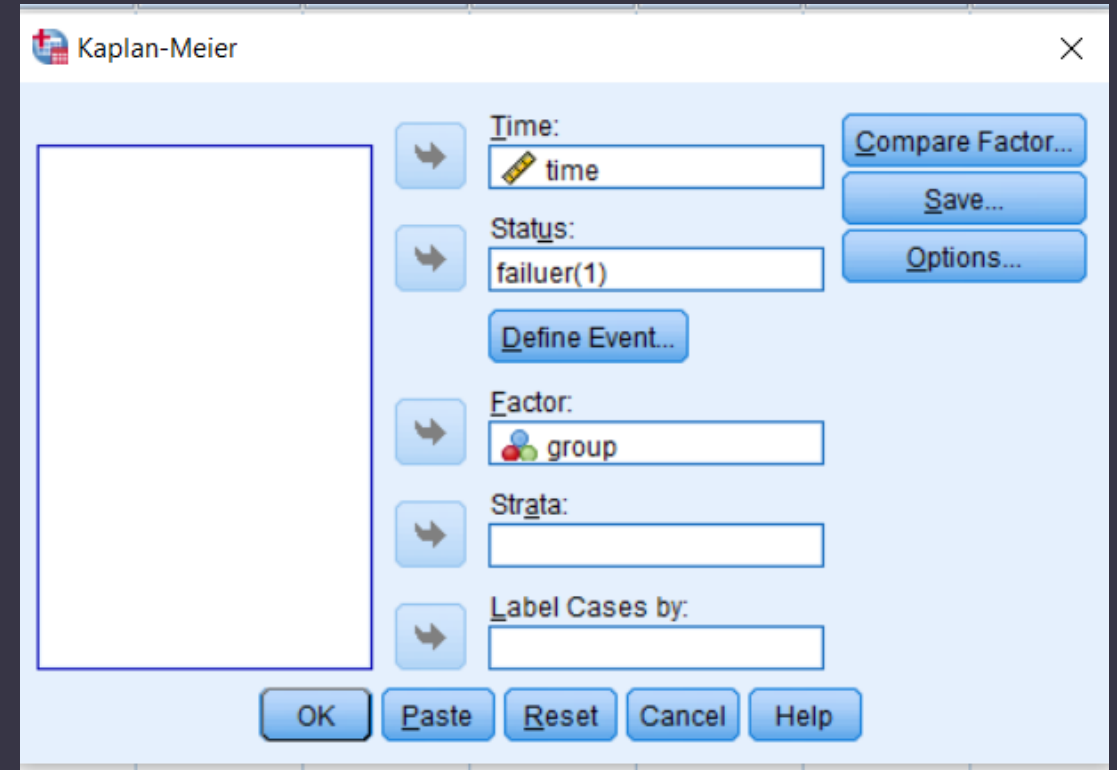
$$\chi^2 = \sum \frac{(\sum O_{jt} - \sum E_{jt})^2}{\sum E_{jt}} = \frac{(4 - 2.890)^2}{2.890} + \frac{(3 - 4.110)^2}{4.110} = 0.426 + 0.300 = 0.726$$

Conclusion. Do not reject H_0 because $0.726 < 3.84$. We do not have statistically significant evidence at $\alpha=0.05$, to show that the time to relapse is different between groups.

Notice: This approximate formula involving only observed and expected values without variance or covariance calculations can also be used when there are more than two groups being compared. The chi-square value obtained is 0.726, which is slightly smaller than the log-rank statistic of 0.835.

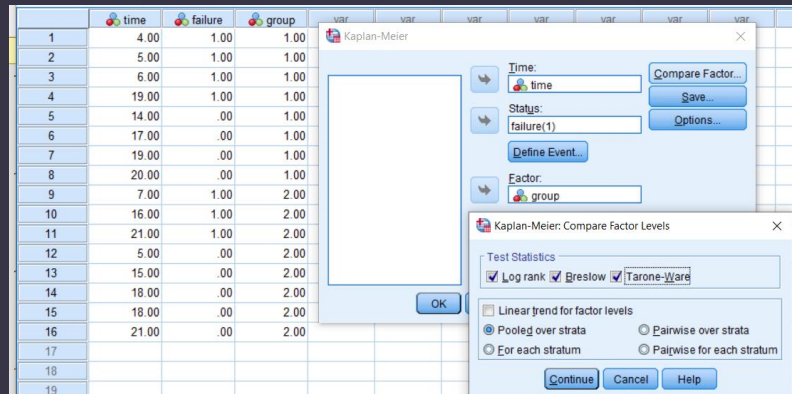
Comparing Several Group in SPSS

- In SPSS there are three test to comparison Survival in several groups
- Log-Rank
- Breslow
- Taron-Ware
- **Note:** All of this are existed in STATA



Comparing Survival Function in SPSS

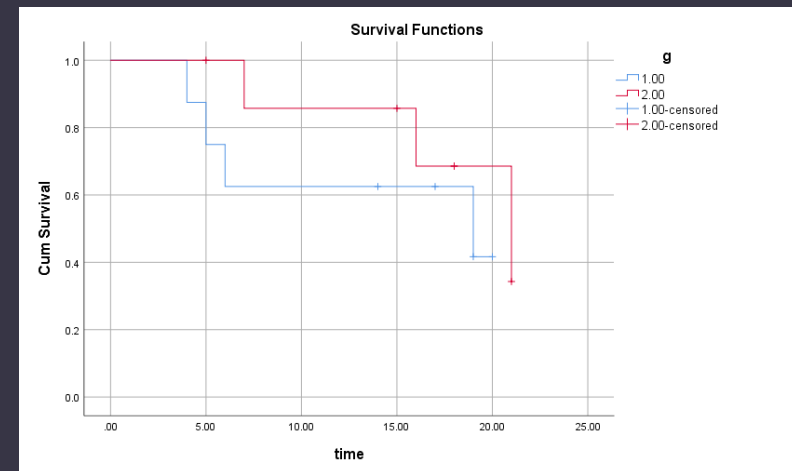
Analysis/ Survival/ Kaplan-Meier •



Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.835	1	.361
Breslow (Generalized Wilcoxon)	1.148	1	.284
Tarone-Ware	.983	1	.321

Test of equality of survival distributions for the different levels of group.



دکتر رحیم زاده، آمار زیستی

sts test STATA Command

- There are other versions of the log rank statistic as well as other tests to compare survival functions between independent groups. For example, a popular test is the modified Wilcoxon test which is sensitive to larger differences in hazards earlier as opposed to later in follow-up (in STATA).
- logrank, cox, wilcoxon, tware, peto, and fh(p q) specify the test of equality desired.
- logrank is the default, unless the data are pweighted, in which case cox is the default and is the only possibility.
- trend specifies that a test for trend of the survivor function across three or more ordered groups be performed.
- strata(varlist) requests that a stratified test be performed.
- detail modifies strata(); it requests that, in addition to the overall stratified test, the tests for the individual strata be reported. detail is not allowed with cox.

sts test command in STATA

Stata/MP 17.0 - C:\Users\M.R\Desktop\cancer.dta

File Edit Data Graphics Statistics User Window Help

History

Filter commands here

Command _rc

31 hausman 198

32 lrtest 198

33 stcox age i:drug

34 lrtest 198

35 stcox

36 lrtest 198

37 sts graph, by(drug)

38 stphplot, by(drug)

39 sum age

40 gen age_c=1 198

41 gen age_c=1 if age<55 198

42 gen age_c=1 if age<55 198

43 gen age_c=1 if age<55

44 replace age_c=2 if age>55

45 replace age_c=2 if age>55

46 sum age_c

47 tab age_c

48 drop age_c

49 gen age_c=1 if age<=57

50 drop age_c

51 gen age_c=1 if age<=57

52 drop age_c

53 gen age_c=1 if age<57

54 replace age_c=2 if age>=57

55 tab age_c

56 sts test drug, logrank strata(age_c) 198

57 sts test drug, logrank strata(age_c)

58 sts test drug, logrank

59 stmc drug, by(age_c)

60 sts test drug, logrank strata(age_c)

Command

drug	Observed events	Expected events
1	11	4.07
2	1	4.08
3	3	6.85
Total	15	15.00

chi2(2) = 18.28
Pr>chi2 = 0.0001

-> age_c = 2

drug	Observed events	Expected events
1	8	3.35
2	5	4.32
3	3	8.34
Total	16	16.00

chi2(2) = 13.96
Pr>chi2 = 0.0009

-> Total

drug	Observed events	Expected events*
1	19	7.42
2	6	8.40
3	6	15.19
Total	31	31.00

sts test - Test equality of survivor functions

Main if/in Options

Variables: drug Survival settings...

Perform test:

☒ Log-rank ☐ Cox ☐ Wilcoxon

☐ Tarone-Ware ☐ Peto-Peto-Prentice

☐ Fleming-Harrington $S(\text{time}-1)^p [(1-S(\text{time}-1))^q]$:

p q

☐ Test trend of the survivor function across three or more ordered groups

Test type

☒ Unstratified ☐ Stratified on variables:

☒ Display overall test results ☐ Display individual test results

OK Cancel Submit

Variables

Filter variables here

Name	Label
studytim	Months to death or end...
died	1 if patient died
drug	Drug type (1=placebo)
age	Patient's age at start of ...
_st	1 if record is to be used...
_d	1 if failure; 0 if censored
_t	Analysis time when reco...
_t0	Analysis time when reco...
age_c	

Properties

Variables

Name	Label
studytim	Months to death or end...
died	1 if patient died
drug	Drug type (1=placebo)
age	Patient's age at start of ...
_st	1 if record is to be used...
_d	1 if failure; 0 if censored
_t	Analysis time when reco...
_t0	Analysis time when reco...
age_c	

Data

Frame	Label
default	
Filename	cancer.dta
Label	Patient Survival in Dr...
Notes	
Variables	9
Observations	48
Size	816

C:\Users\M.R\Desktop

Type here to search

89°F Sunny 12:39 AM 7/20/2023

The Wilcoxon (Breslow–Gehan) test

- sts test, wilcoxon performs the generalized Wilcoxon test of Breslow (1970) and Gehan (1965).
- This test is appropriate when hazard functions are thought to vary in ways other than proportionally and when censoring patterns are similar across groups.
- The Wilcoxon test statistic is constructed by weighting the contribution of each failure time to the overall test statistic by the number of subjects at risk.
- Thus, it gives heavier weights to earlier failure times when the number at risk is higher.
- As a result, this test is susceptible to differences in the censoring pattern of the groups.

The Tarone–Ware test

- t_{ts} test, t_{ware} performs a test suggested by Tarone and Ware (1977), with weights equal to the square root of the number of subjects in the risk pool at time t_i .
- Like Wilcoxon's test, this test is appropriate when hazard functions are thought to vary in ways other than proportionally and when censoring patterns are similar across groups.
- The test statistic is constructed by weighting the contribution of each failure time to the overall test statistic by the square root of the number of subjects at risk.
- Thus, like the Wilcoxon test, it gives heavier weights, although not as large, to earlier failure times.
- Although less susceptible to the failure and censoring pattern in the data than Wilcoxon's test, this could remain a problem if large differences in these patterns exist between groups.

The Peto–Peto–Prentice test

- sts test, peto performs an alternative to the Wilcoxon test proposed by Peto and Peto (1972) and Prentice (1978).
- The test uses as the weight function an estimate of the overall survivor function, which is similar to that obtained using the Kaplan–Meier estimator.
- This test is appropriate when hazard functions are thought to vary in ways other than proportionally, but unlike the Wilcoxon–Breslow–Gehan test, it is not affected by differences in censoring patterns across groups.

The generalized Fleming–Harrington tests

- `sts test, fh(p q)` performs the Harrington and Fleming (1982) class of test statistics.
- The weight function at each distinct failure time, t , is the product of the Kaplan–Meier survivor estimate at time $t - 1$ raised to the p power and $1 -$ the Kaplan–Meier survivor estimate at time $t - 1$ raised to the q power.
- Thus, when specifying the Fleming and Harrington option, you must specify two nonnegative arguments, p and q .
- When $p > q$, the test gives more weights to earlier failures than to later ones.
- When $p < q$, the opposite is true, and more weight is given to later failures than to earlier ones.
- When p and q are both zero, the weight is 1 at all failure times and the test reduces to the log-rank test.

The “Cox” test

- The term Cox test is our own, and this test is a variation on the log-rank test using Cox regression. One way of thinking about the log-rank test is as a Cox proportional hazards model on indicator variables for each of the groups. The log-rank test is a test that the coefficients are zero or, if you prefer, that the hazard ratios are one. The log-rank test is, in fact, a score test of that hypothesis performed on a slightly different (partial) likelihood function that handles ties more accurately. Many researchers think that a (less precise) score test on the precise likelihood function is preferable to a (more precise) likelihood-ratio test on the approximate likelihood function used in Cox regression estimation.

Why The “Cox” test?

- Our purpose in offering sts test, cox is not to promote its use instead of the log-rank test but to provide a test for researchers with sample-weighted data. If you have sample weights (if you specified pweights when you stset the data), you cannot run the log-rank or Wilcoxon tests. The Cox regression model, however, has been generalized to sample-weighted data, and Stata's stcox can fit models with such data. In sample-weighted data, the likelihood-ratio statistic is no longer appropriate, but the Wald test based on the robust estimator of variance is.

Log-Rank test Advantages

- A primary advantage of the log-rank test is the ease with which it can be explained to non statisticians, because the test statistic is the difference between the observed and expected number of failures within groups.

Log-Rank test Limitations

- First, It does not provide an effect estimate or the related confidence interval to compare the survival in different patient groups .
- Second, it does not permit the adjustment of confounders in etiological research or predictors in prognostic research .
- Finally, the KM method requires data categorization, so calculation of the incremental increase (or decrease) in the hazard of a given event associated with one unit (or any number of units) increase in the candidate risk factor is not possible

Weight of Different test

The Log-Rank test gives equal weight to failures at each time and is optimal when the hazards are proportional. Sometimes one might have reason to give more weight to differences at earlier or later times. The table below shows some alternatives:

Test	Weight
Mantel-Haenszel or log-rank	1
Wilcoxon-Breslow-Gehan	n
Tarone-Ware	$\sqrt{n}$
Peto-Peto-Prentice	$\tilde{S}(t_j)$
Fleming-Harrington	$\hat{S}(t_{j-1})^p(1 - \hat{S}(t_{j-1}))^q$

Weighting the test statistic for two groups

Test statistic:

$$\frac{\left(\sum_f w(t_{(f)}) (m_{if} - e_{if}) \right)^2}{\text{var} \left(\sum_j w(t_{(f)}) (m_{if} - e_{if}) \right)}$$

$i = 1, 2$

$f = f\text{th failure time}$

$w(t_{(f)}) = \text{weight at } f\text{th failure time}$

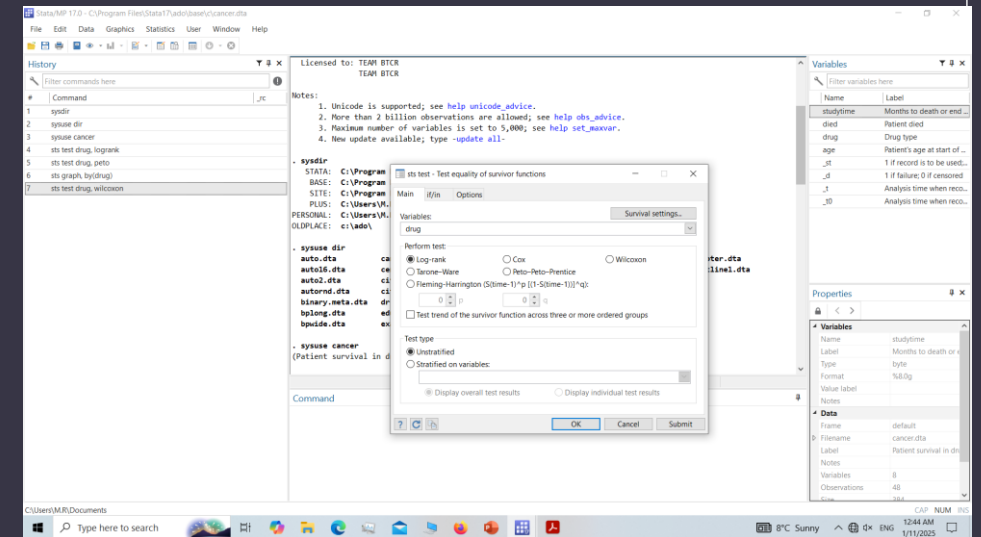
Comparison Survival Analysis with STATA

- sts test group, logrank •
- sts test group, cox •
- sts test group, Wilcoxon •
- sts test group, tware •
- sts test group, peto •
- sts test group, fh(1 1) •

Equality of survivor functions
Log-rank test

group	Observed events	Expected events
1	4	2.89
2	3	4.11
Total	7	7.00

chi2(1) = 0.84
Pr>chi2 = 0.3608



Example

- A small clinical trial is run to compare two combination treatments in patients with advanced **gastric cancer**. Twenty participants with stage IV gastric cancer who consent to participate in the trial are randomly assigned to receive chemotherapy before surgery or chemotherapy after surgery. The primary outcome is death and participants are followed for up to 48 months (4 years) following enrollment into the trial. The experiences of participants in each arm of the trial are shown below.

Chemotherapy Before Surgery		Chemotherapy After Surgery	
Month of Death	Month of Last Contact	Month of Death	Month of Last Contact
8	8	33	48
12	32	28	48
26	20	41	25
14	40		37
21			48
27			25
			43

Example

time	failure	group	time	failure	group
8	0	1	25	0	2
8	1	1	25	0	2
12	1	1	28	1	2
14	1	1	33	1	2
20	0	1	37	0	2
21	1	1	41	1	2
26	1	1	43	0	2
27	1	1	48	0	2
32	0	1	48	0	2
40	0	1	48	0	2

Log-Rank Test (approximation)

We can now compute the test statistic:

$$\chi^2 = \sum \frac{(\sum O_{jt} - \sum E_{jt})^2}{\sum E_{jt}} = \frac{(6 - 2.620)^2}{2.620} + \frac{(3 - 6.380)^2}{6.380} = 4.360 + 1.791 = 6.151$$

The test statistic is approximately distributed as chi-square with 1 degree of freedom. Thus, the critical value for the test can be found in the table of Critical Values of the χ^2 Distribution.

For this test the decision rule is to Reject H_0 if $\chi^2 > 3.84$. We observe $\chi^2 = 6.151$, which exceeds the critical value of 3.84. Therefore, we reject H_0 . We have significant evidence, $\alpha=0.05$, to show that the two survival curves are different.

time	# nf1	#nf2	# mf1	# mf2	E F g1	E F g2	o1-e1	o2-e2	n1f*nf2	m1f+mf2	(n1f+nf2- m1f-m2f)	numariator	deminator	num/dem
8	10	10	1	0	0.5000	0.5000	0.5000	-0.5000	100	1	19	1900	7600	0.2500
12	8	10	1	0	0.4444	0.5556	0.5556	-0.5556	80	1	17	1360	5508	0.2469
14	7	10	1	0	0.4118	0.5882	0.5882	-0.5882	70	1	16	1120	4624	0.2422
21	5	10	1	0	0.3333	0.6667	0.6667	-0.6667	50	1	14	700	3150	0.2222
26	4	8	1	0	0.3333	0.6667	0.6667	-0.6667	32	1	11	352	1584	0.2222
27	3	8	1	0	0.2727	0.7273	0.7273	-0.7273	24	1	10	240	1210	0.1983
28	2	8	0	1	0.2000	0.8000	-0.2000	0.2000	16	1	9	144	900	0.1600
33	1	7	0	1	0.1250	0.8750	-0.1250	0.1250	7	1	7	49	448	0.1094
41	0	5	0	1	0.0000	1.0000	0.0000	0.0000	0	1	4	0	100	0.0000
			6	3	2.6206	6.3794	3.3794	-3.3794						1.6513
					6.921									
					4	6.9214	6.916	6.9160						

Logrank test Formoula

Expected cell counts:

$$e_{1f} = \left(\frac{n_{1f}}{n_{1f} + n_{2f}} \right) \times (m_{1f} + m_{2f})$$

$\uparrow$ $\uparrow$
 Proportion # of failures over
 in risk set both groups

$$e_{2f} = \left(\frac{n_{2f}}{n_{1f} + n_{2f}} \right) \times (m_{1f} + m_{2f})$$

$$\begin{aligned} & \text{Var}(O_i - E_i) \\ &= \sum_j \frac{n_{1f} n_{2f} (m_{1f} + m_{2f}) (n_{1f} + n_{2f} - m_{1f} - m_{2f})}{(n_{1f} + n_{2f})^2 (n_{1f} + n_{2f} - 1)} \\ & i = 1, 2 \end{aligned}$$

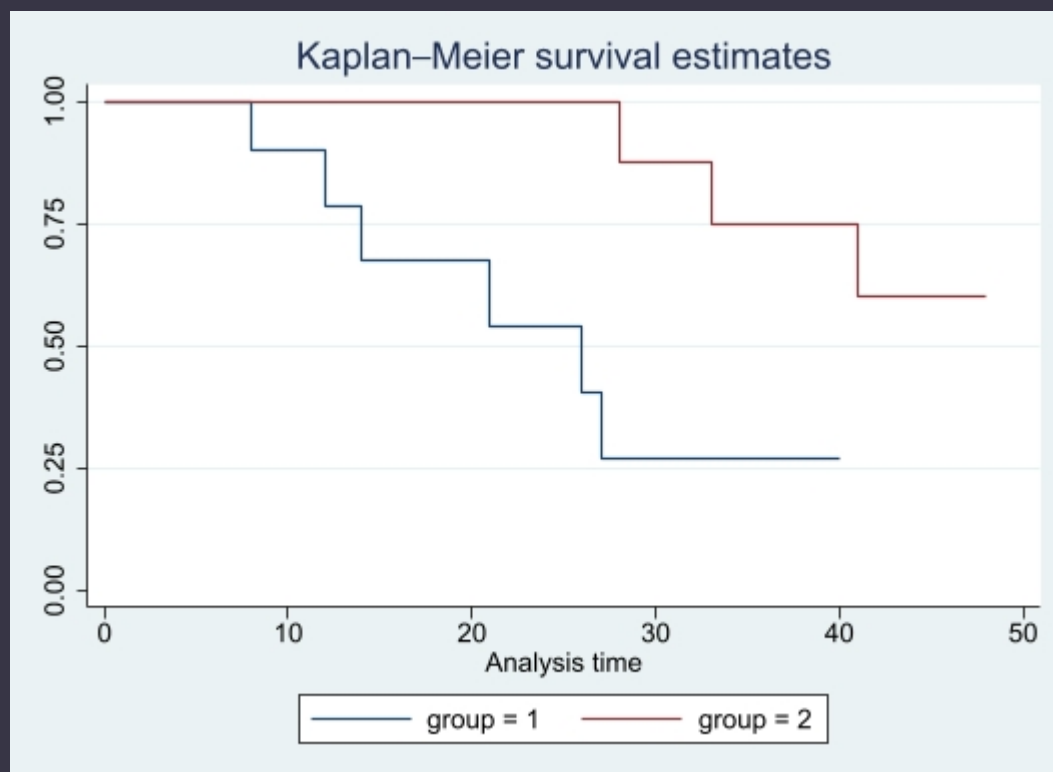
Two groups:

$O_2 - E_2$ = summed observed minus expected score for group 2

$$\text{Log-rank statistic} = \frac{(O_2 - E_2)^2}{\text{Var}(O_2 - E_2)}$$

STATA Log-Rank test

`sts graph, by(group)`



`stset time, failure(failure)`

Equality of survivor functions
Log-rank test

group	Observed events	Expected events
1	6	2.62
2	3	6.38
Total	9	9.00

$\text{chi2}(1) = 6.92$

$\text{Pr}>\text{chi2} = 0.0085$

STATA Command

sts list, by(group)

Kaplan-Meier survivor function
By variable: group

Time	At risk	Fail	<u>Net</u> <u>lost</u>	Survivor function	Std. error	[95% conf. int.]	
group=1							
8	10	1	1	0.9000	0.0949	0.4730	0.9853
12	8	1	0	0.7875	0.1340	0.3809	0.9426
14	7	1	0	0.6750	0.1551	0.2906	0.8825
20	6	0	1	0.6750	0.1551	0.2906	0.8825
21	5	1	0	0.5400	0.1731	0.1812	0.8007
26	4	1	0	0.4050	0.1747	0.0999	0.7014
27	3	1	0	0.2700	0.1604	0.0414	0.5838
32	2	0	1	0.2700	0.1604	0.0414	0.5838
40	1	0	1	0.2700	0.1604	0.0414	0.5838
group=2							
25	10	0	2	1.0000	.	.	.
28	8	1	0	0.8750	0.1169	0.3870	0.9814
33	7	1	0	0.7500	0.1531	0.3148	0.9309
37	6	0	1	0.7500	0.1531	0.3148	0.9309
41	5	1	0	0.6000	0.1817	0.1955	0.8523
43	4	0	1	0.6000	0.1817	0.1955	0.8523
48	3	0	3	0.6000	0.1817	0.1955	0.8523

STATA Command

```
sts list, by( group ) compare
```

Kaplan-Meier survivor function

By variable: group

group		1	2
Time			
	8	0.9000	1.0000
	13	0.7875	1.0000
	18	0.6750	1.0000
	23	0.5400	1.0000
	28	0.2700	0.8750
	33	0.2700	0.7500
	38	0.2700	0.7500
	43	.	0.6000
	48	.	0.6000

STATA Command

```
sts list, cumhaz by( group ) compare
```

Nelson-Aalen cumulative hazard function

By variable: group

group		1	2
Time	8	0.1000	0.0000
	13	0.2250	0.0000
	18	0.3679	0.0000
	23	0.5679	0.0000
	28	1.1512	0.1250
	33	1.1512	0.2679
	38	1.1512	0.2679
	43	.	0.4679
	48	.	0.4679

Choosing a Test

- Results of different weightings usually lead to similar conclusions
- The best choice is test with most power
- Power depends on how the null is violated
- There may be a clinical reason to choose a particular weighting
- Choice of weighting should be a priori
- decision on which statistical test to use rather than fish for a desired p-value. Fishing for a desired result may lead to bias.

Remission data

EXAMPLE

Expanded Table (Remission Data)

f	$t(f)$	# failures		# in risk set		# expected		Observed-expected	
		m_{1f}	m_{2f}	n_{1f}	n_{2f}	e_{1f}	e_{2f}	$m_{1f}-e_{1f}$	$m_{2f}-e_{2f}$
1	1	0	2	21	21	$(21/42) \times 2$	$(21/42) \times 2$	-1.00	1.00
2	2	0	2	21	19	$(21/40) \times 2$	$(19/40) \times 2$	-1.05	1.05
3	3	0	1	21	17	$(21/38) \times 1$	$(17/38) \times 1$	-0.55	0.55
4	4	0	2	21	16	$(21/37) \times 2$	$(16/37) \times 2$	-1.14	1.14
5	5	0	2	21	14	$(21/35) \times 2$	$(14/35) \times 2$	-1.20	1.20
6	6	3	0	21	12	$(21/33) \times 3$	$(12/33) \times 3$	1.09	-1.09
7	7	1	0	17	12	$(17/29) \times 1$	$(12/29) \times 1$	0.41	-0.41
8	8	0	4	16	12	$(16/28) \times 4$	$(12/28) \times 4$	-2.29	2.29
9	10	1	0	15	8	$(15/23) \times 1$	$(8/23) \times 1$	0.35	-0.35
10	11	0	2	13	8	$(13/21) \times 2$	$(8/21) \times 2$	-1.24	1.24
11	12	0	2	12	6	$(12/18) \times 2$	$(6/18) \times 2$	-1.33	1.33
12	13	1	0	12	4	$(12/16) \times 1$	$(4/16) \times 1$	0.25	-0.25
13	15	0	1	11	4	$(11/15) \times 1$	$(4/15) \times 1$	-0.73	0.73
14	16	1	0	11	3	$(11/14) \times 1$	$(3/14) \times 1$	0.21	-0.21
15	17	0	1	10	3	$(10/13) \times 1$	$(3/13) \times 1$	-0.77	0.77
16	22	1	1	7	2	$(7/9) \times 2$	$(2/9) \times 2$	-0.56	0.56
17	23	1	1	6	1	$(6/7) \times 2$	$(1/7) \times 2$	-0.71	0.71
Totals		9	(21)			19.26	(10.74)	-10.26	(-10.26)

EXAMPLE

$$O_1 - E_1 = -10.26$$

$$O_2 - E_2 = 10.26$$

Approximate formula:

$$X^2 \approx \sum_i^{\# \text{ of groups}} \frac{(O_i - E_i)^2}{E_i}$$

EXAMPLE

$$X^2 = \frac{(-10.26)^2}{19.26} + \frac{(10.26)^2}{10.74}$$

$$= 15.276$$

Log-rank statistic = 16.793

Remission data

EXAMPLE

Using Stata: Remission Data

Group	Events observed	Events expected
1	9	19.25
2	21	10.75
Total	30	30.00

Log rank = $\text{chi2}(2) = 16.79$
P-Value = $\text{Pr} > \text{chi2} = 0.000$

EXAMPLE

$$O_2 - E_2 = 10.26$$

$$\text{Var}(O_2 - E_2) = 6.2685$$

$$\begin{aligned}\text{Log - rank statistic} &= \frac{(O_2 - E_2)^2}{\widehat{\text{Var}}(O_2 - E_2)} \\ &= \frac{(10.26)^2}{6.2685} = 16.793\end{aligned}$$

Log-Rank Test with more than 2 Groups

Raw DATA

Calculation Expected

g1	g2	g3
2	1	3
5 ⁺	3 ⁺	7
8	6	8
10 ⁺	9	13 ⁺
14 ⁺	11	16 ⁺

time	ng1(O1)	ng2(O2)	ng3(O3)	fg1	fg2	fg3	E1	E2	E3
1	5	5	5	0	1	0	0.333	0.333	0.333
2	5	4	5	1	0	0	0.357	0.286	0.357
3	4	4	5	0	0	1	0.308	0.308	0.385
6	3	3	4	0	1	1	0.600	0.600	0.800
8	3	2	3	1	0	1	0.750	0.500	0.750
9	2	2	2	0	1	0	0.333	0.333	0.333
11	1	1	2	0	1	0	0.250	0.250	0.500
				2	4	3	2.932	2.610	3.458

Log-Rank Test (approximation)

Approximate formula:

$$\chi^2 \approx \sum_i^{\text{\# of groups}} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \sum_{i=1}^g \frac{(O_i - E_i)^2}{E_i} = 1.07$$

Because three groups are being compared here, $G = 3$ and the degrees of freedom for the log-rank test is thus $G-1$, or 2.

The log-rank statistic is computed to be 1.07, which has a P-value=0.414 . Thus, the conclusion from the log-rank test is that there is not significant difference among the three survival curves three groups.

Stratified log rank test

- The stratified log rank test provides a method of testing the equivalence of survival curves controlling for the stratified variable.
- With this test the summed observed minus expected scores O-E are calculated within strata of each group and then summed across strata. The only difference between the unstratified and stratified approaches is that for the unstratified approach, the observed minus expected number of events for each failure time is summed over all failure times for each group (i).
- With the stratified approach, the observed minus expected number of events is summed over all failure times for each group within each stratum and then summed over all strata.
- Either way, the null distribution is chi-square with $G - 1$ degrees of freedom, where G represents the number of groups being compared (not the number of strata).

Stratified log rank test

The stratified approach can also be applied to any of the weighted variations of the log rank test (e.g., Wilcoxon).

A limitation of the stratified approach is the reduced sample size within each stratum.

This is particularly problematic with the remission dataset, which has a small sample size to begin with.

Log rank unstratified

$$O_i - E_i = \sum_j (m_{ij} - e_{ij})$$

i = group #, f = f th failure time

Log rank stratified

$$O_i - E_i = \sum_s \sum_f (m_{ifs} - e_{ifs})$$

i = group #, f = j th failure time,
 s = stratum #

Stratified or unstratified (G groups)
Under H_0 :

log rank statistic $\sim \chi^2$ with
 $G - 1$ df

Stratified log rank test (gastric cancer data)

time	failure	group	Strata	time	failure	group	Strata
8	0	1	1	25	0	2	1
8	1	1	1	25	0	2	1
12	1	1	2	28	1	2	2
14	1	1	2	33	1	2	2
20	0	1	1	37	0	2	1
21	1	1	1	41	1	2	1
26	1	1	2	43	0	2	2
27	1	1	2	48	0	2	2
32	0	1	1	48	0	2	2
40	0	1	2	48	0	2	2

Stratified log rank test

Strata1

time	failure	group	Strata
8	0	1	1
8	1	1	1
20	0	1	1
21	1	1	1
32	0	1	1
25	0	2	1
25	0	2	1
37	0	2	1
41	1	2	1

Strata2

time	failure	group	Strata
12	1	1	2
14	1	1	2
26	1	1	2
27	1	1	2
40	0	1	2
28	1	2	2
33	1	2	2
43	0	2	2
48	0	2	2
48	0	2	2
48	0	2	2

Stratified log rank test

Strata1

time	# nf1	#nf2	# mf1	# mf2	E F g1	E F g2	o1-e1	o2-e2
8	5	4	1	0	0.56	0.44	0.44	-0.44
21	2	4	1	0	0.33	0.67	0.67	-0.67
41	0	1	0	1	0	1	0	0
			2	1	0.89	2.11		

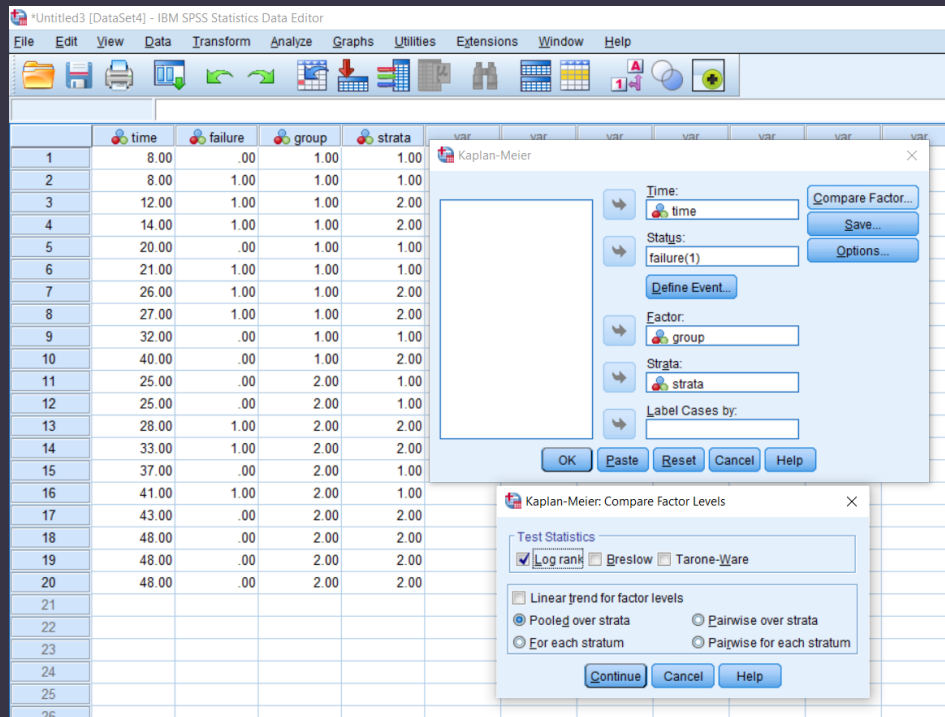
strata2

time	# nf1	#nf2	# mf1	# mf2	E F g1	E F g2	o1-e1	o2-e2
12	5	6	1	0	0.45	0.55	0.55	-0.55
14	4	6	1	0	0.4	0.6	0.6	-0.6
26	3	6	1	0	0.33	0.67	0.67	-0.67
27	2	6	1	0	0.25	0.75	0.75	-0.75
28	1	6	0	1	0.143	0.857	-0.14	0.143
33	1	5	0	1	0.167	0.833	-0.17	0.167
			4	2	1.74	4.26		

Expected in group1 are sum of the
 $0.89+1.74=2.63$ and

Expected in group2 are sum of the
 $2.11+4.26=6.37$

Stratified log rank test



Overall Comparisons^a

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	6.949	1	.008

Test of equality of survival distributions for the different levels of group.

a. Adjusted for strata.

Stratified log rank test in STATA

`sts test group, logrank strata(strata) •`

Equality of survivor functions

Stratified log-rank test

Strata: strata

group	Observed events	Expected events*
1	6	2.64
2	3	6.36
Total	9	9.00

* aggregated over stratum-specific results

`chi2(1) = 6.95`

`Pr>chi2 = 0.0084`

Homework

Practice1: Calculate Log-Rank test for this data, and check the answer with the SPSS and STATA •

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

- Practice 2: calculate the Variance and test statistics for the example in page 41.
- : Read the pages 67-823Practice