Open topic with navigation

Log-rank and Wilcoxon

Menu location: **Analysis_Survival_Log-rank and Wilcoxon**.

This function provides methods for comparing two or more survival curves where some of the observations may be censored and where the overall grouping may be stratified. The methods are nonparametric in that they do not make assumptions about the distributions of survival estimates.

In the absence of censorship (e.g. loss to follow up, alive at end of study) the methods presented here reduce to a Mann-Whitney (two sample Wilcoxon) test for two groups of survival times and a Kruskal-Wallis test for more than two groups of survival times. StatsDirect gives a comprehensive set of tests for the comparison of survival data that may be censored (Tarone and Ware, 1977; Kalbfleisch and Prentice, 1980; Cox and Oakes, 1984; Le, 1997).

The null hypothesis tested here is that the risk of death/event is the same in all groups.

Peto's log-rank test is generally the most appropriate method but the Prentice modified Wilcoxon test is more sensitive when the ratio of hazards is higher at early survival times than at late ones (Peto and Peto, 1972; Kalbfleisch and Prentice, 1980). The log-rank test is similar to the Mantel-Haenszel test and some authors refer to it as the Cox-Mantel test (Mantel and Haenszel, 1959; Cox, 1972).

Strata

An optional variable, strata, allows you to sub-classify the groups specified in the group identifier variable and to test the significance of this sub-classification (Armitage and Berry, 1994; Lawless, 1982; Kalbfleisch and Prentice, 1980).

Wilcoxon weights

StatsDirect gives you a choice of three different weighting methods for the generalised Wilcoxon test, these are Peto-Prentice, Gehan-Breslow and Tarone-Ware. The Peto-Prentice method is generally more robust than the others but the Gehan statistic is calculated routinely by many statistical software packages (Breslow, 1974; Tarone and Ware, 1977; Kalbfleisch and Prentice, 1980; Miller, 1981; Hosmer and Lemeshow 1999). You should seek statistical guidance if you plan to use any weighting method other than Peto-Prentice.

Hazard-ratios

An approximate confidence interval for the log hazard-ratio is calculated using the following estimate of standard error (SE):

$$\mathrm{SE} = \sqrt{\sum_{i=1}^{k} \frac{1}{e_{ij}}}$$

- where e_{ij} is the extent of exposure to risk of death (sometimes called expected deaths) for group i of k at the jth distinct observed time (sometimes called expected deaths) for group i of k (Armitage and Berry, 1994).

An exact conditional maximum likelihood estimate of the hazard ratio is optionally given. The exact estimate and its confidence interval (Fisher or mid-P) should be routinely used in preference to the above approximation. The exponents of Cox regression parameters are also exact estimators of the hazard ratio, but please note that they are not exact if Breslow's method has been used to correct for ties in the regression. Please consult with a statistician if you are considering using Cox regression.

Trend test

If you have more than two groups then StatsDirect will calculate a variant of the log-rank test for trend. If you choose not to enter group scores then they are allocated as 1,2,3 ... n in group order (Armitage and Berry, 1994; Lawless, 1982; Kalbfleisch and Prentice, 1980).

Technical validation

The general test statistic is calculated around a hypergeometric distribution of the number of events at distinct event times:

$$e_{ij} = \frac{n_{ij} d_j}{n_j}$$

$$\text{diagonal } (\widehat{\mathbf{V}}_j)_{ii} = \frac{n_{ij}(n_j - n_{ij})d_j(n_j - d_j)}{n_j^2(n_j - 1)}, \ i = 1, 2, \ldots k - 1$$

$$\text{off diagonal } (\widehat{\mathbf{V}}_j)_{ii} = \frac{n_{ij} n_{ij} d_j(n_j - d_j)}{n_j^2(n_j - 1)}, \ i, l = 1, 2, \ldots k - 1, k \neq l$$

$$\mathbf{U}_j = \sum_{i=1}^{k} w_j(d_{ij} - e_{ij})$$

$$\widehat{\mathbf{V}}_w = \mathbf{w}\widehat{\mathbf{V}}\mathbf{w}$$

$$\chi_{k-1}^2 = \mathbf{U}'\mathbf{V}_w^{-1}\mathbf{U}$$

$$\chi_{1 \text{ trend}}^2 = \frac{(\mathbf{c}'\mathbf{U})^2}{\mathbf{c}'\widehat{\mathbf{V}}_w\mathbf{c}}$$

- where the weight w_j for the log-rank test is equal to 1, and w_j for the generalised Wilcoxon test is n_i (Gehan-Breslow method); for the Tarone-Ware method w_j is the square root of n_i; and for the Peto-Prentice method w_j is the Kaplan-Meier survivor function multiplied by (n_i divided by n_i +1). e_{ij} is the expectation of death in group i at the jth distinct observed time where d_j events/deaths occurred. n_{ij} is the number at risk in group i just before the jth distinct observed time. The test statistic for equality of survival across the k groups (populations sampled) is approximately chi-square distributed on k-1 degrees of freedom. The test statistic for monotone trend is approximately chi-square distributed on 1 degree of freedom. **c** is a vector of scores that are either defined by the user or allocated as 1 to k.

Variance is estimated by the method that Peto (1977) refers to as "exact".

The stratified test statistic is expressed as (Kalbfleisch and Prentice, 1980):

$$\chi_{k-1}^2 = \left(\sum \mathbf{U}\right)'\left(\sum \widehat{\mathbf{V}}_w\right)^{-1}\left(\sum \mathbf{U}\right)$$

- where the statistics defined above are calculated within strata then summed across strata prior to the generalised inverse and transpose matrix operations.

Example
From [Armitage and Berry (1994, p. 479)](#).
Test workbook (Survival worksheet: Stage Group, Time, Censor).

The following data represent the survival in days since entry to the trial of patients with diffuse histiocytic lymphoma. Two different groups of patients, those with stage III and those with stage IV disease, are compared.
Stage 3: 6, 19, 32, 42, 42, 43*, 94, 126*, 169*, 207, 211*, 227*, 253, 255*, 270*, 310*, 316*, 335*, 346*

Stage 4: 4, 6, 10, 11, 11, 11, 13, 17, 20, 20, 21, 22, 24, 24, 29, 30, 30, 31, 33, 34, 35, 39, 40, 41*, 43*, 45, 46, 50, 56, 61*, 61*, 63, 68, 82, 85, 88, 89, 90, 93, 104, 110, 134, 137, 160*, 169, 171, 173, 175, 184, 201, 222, 235*, 247*, 260*, 284*, 290*, 291*, 302*, 304*, 341*, 345*

* = censored data (patient still alive or died from an unrelated cause)

To analyse these data in StatsDirect you must first prepare them in three workbook columns as shown below:

Stage group	Time	Censor
1	6	1
1	19	1
1	32	1
1	42	1

1	42	1
1	43	0
1	94	1
1	126	0
1	169	0
1	207	1
1	211	0
1	227	0
1	253	1
1	255	0
1	270	0
1	310	0
1	316	0
1	335	0
1	346	0
2	4	1
2	6	1
2	10	1
2	11	1
2	11	1
2	11	1
2	13	1
2	17	1
2	20	1
2	20	1
2	21	1
2	22	1
2	24	1
2	24	1
2	29	1
2	30	1
2	30	1
2	31	1
2	33	1
2	34	1
2	35	1
2	39	1
2	40	1
2	41	0
2	43	0
2	45	1
2	46	1
2	50	1
2	56	1
2	61	0
2	61	0
2	63	1
2	68	1
2	82	1
2	85	1
2	88	1
2	89	1
2	90	1
2	93	1
2	104	1
2	110	1
2	134	1

2	137	1
2	160	0
2	169	1
2	171	1
2	173	1
2	175	1
2	184	1
2	201	1
2	222	1
2	235	0
2	247	0
2	260	0
2	284	0
2	290	0
2	291	0
2	302	0
2	304	0
2	341	0
2	345	0

Alternatively, open the test workbook using the file open function of the file menu. Then select Log-rank and Wilcoxon from the Survival Analysis section of the analysis menu. Select the column marked "Stage group" when asked for the group identifier, select "Time" when asked for times and "Censor" for censorship. Click on the cancel button when asked about strata.

For this example:

Logrank and Wilcoxon tests

Log Rank (Peto):

For group 1 (Stage group = 1)
Observed deaths = 8
Extent of exposure to risk of death = 16.687031
Relative rate = 0.479414

For group 2 (Stage group = 2)
Observed deaths = 46
Extent of exposure to risk of death = 37.312969
Relative rate = 1.232815

test statistics:
-8.687031, 8.687031

variance-covariance matrix:
0.088912   -11.24706
-11.24706   11.24706

Chi-square for equivalence of death rates = 6.70971 P = 0.0096

Hazard Ratio, (approximate 95% confidence interval)
Group 1 vs. Group 2 = 0.388878, (0.218343 to 0.692607)

Conditional maximum likelihood estimates:

Hazard Ratio = 0.381485

Exact Fisher 95% confidence interval = 0.154582 to 0.822411

Exact Fisher one sided P = 0.0051, two sided P = 0.0104

Exact mid-P 95% confidence interval = 0.167398 to 0.783785

Exact mid-P one sided P = 0.0034, two sided P = 0.0068

Generalised Wilcoxon (Peto-Prentice):

test statistics:

-5.19836, 5.19836

variance-covariance matrix:

0.201506   -4.962627

-4.962627   4.962627

Chi-square for equivalence of death rates = 5.44529 P = 0.0196

Both log-rank and Wilcoxon tests demonstrated a statistically significant difference in survival experience between stage 3 and stage 4 patients in this study.

Stratified example

From Peto et al. (1977):

Group	Trial Time	Censorship	Stratum
1	8	1	1
1	8	1	2
2	13	1	1
2	18	1	1
2	23	1	1
1	52	1	1
1	63	1	1
1	63	1	1
2	70	1	2
2	70	1	2
2	180	1	2
2	195	1	2
2	210	1	2
1	220	1	2
1	365	0	2
2	632	1	2
2	700	1	2
1	852	0	2
2	1296	1	2
1	1296	0	2
1	1328	0	2
1	1460	0	2
1	1976	0	2
2	1990	0	2
2	2240	0	2

Censorship 1 = death event

Censorship 0 = lost to follow-up

Stratum 1 = renal impairment

Stratum 2 = no renal impairment

The table above shows you how to prepare data for a stratified log-rank test in StatsDirect. This example is worked through in the second of two classic papers by Richard Peto and colleagues (Peto et al., 1977, 1976). Please note that StatsDirect uses the more accurate variance formulae mentioned in the statistical notes section at the end of Peto et al. (1977).

P values