



BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Survival Analysis and Clinical Trials

Course Code: 23ST04DEC

Unit-II

Non-parametric Survival Models

Dr. T. Jai Sankar

Associate Professor and Head

Department of Statistics

Ms. I. Angel Agnes Mary

Guest Faculty

Department of Statistics

UNIT – II

Non-parametric Survival Models

Introduction to Life Tables

The life-table method is one of the oldest techniques for measuring mortality and describing the survival experience of a population. A **Life Table** is a presentation of summary of the mortality experiences of a community during a given period in the form of a table. The Life Table exhibits the number of persons living and dying at each age, on the basis of the experience of a *cohort*. It also gives the probability of dying and living separately. The Life Table tells the life history of a *cohort*.

Cohort is a group of individuals who born at the same time and experienced the same mortality conditions.

Types of Life Tables:

There are two types of Life tables are given:

- **Cohort or Generation Life Table**
- **Period Life Table**

The Cohort or Generation Life Table “summarizes the age specific mortality experience of a given birth cohort (a group of persons all born at the same time) for its life and thus extends over many calendar years.” On the other hand, the “Period Life Table summarizes the age specific mortality conditions pertaining to a given or other short time period.”

Uses of Life Table

- Actuaries compute rate of premiums for persons of different age groups using Life Table.
- It helps to assess the accuracy of census figures, death and birth registrations.
- It helps to evaluate the impact of family planning on population growth.
- It enables to assess the increase in the life span due to new scientific inventions, sophisticated medical treatments and improved living conditions.
- Estimates of migration can be made from Life Table.

Assumptions of Life Table:

A hypothetical cohort of life table usually comprises of 1,000 or 10,000 or 1,00,000 births.

1. The deaths are equally distributed throughout the year.
2. The cohort of people diminishes gradually by death only.
3. The cohort is closed to the in-migration and out-migration.
4. The death rate is related to a pre-determined age specific death rate.
5. The cohort of persons dies at a fixed age which does not change.
6. There is no change in death rates overtime.
7. The cohort of life tables are generally constructed separately for males and females.
- 8.

Construction of Life Table

Construction of Life Table begins with a *cohort* population. The following are the standard set of components of a Life Table:

- 1) Age (x)
- 2) Survivorship function
- 3) Number of deaths in the age interval ($x, x+1$)
- 4) Probability for a person surviving at the age x to die before $x+1$ years
- 5) Probability for a person aged x years to survive upto $x+1$ years
- 6) Number of persons lived in aggregate in the age interval ($x, x+1$).
- 7) Number of years lived by the cohort at and above the age x years.
- 8) Expectation of life.

These components are described below with their respective notations and formula required to compute each of them.

- 1) x : Age, in years
- 2) $l(x)$: Number of survivors at the exact age of x years.

For instance, $l(25)$ denotes the number of persons who survive to the moment of attaining age 25 years. Hence, $l(x)$ is a decreasing function of x .

$l(0)$ is called **radix** of the Life Table or *cohort*, which is conventionally taken as 1,00,000.

- 3) $d(x)$: Number of persons among $l(x)$ persons who die before reaching the age x years. *i.e.*, $d(x) = l(x) - l(x+1)$
- 4) $q(x)$: Probability for a person surviving at the age x will die before $x+1$ years. *i.e.*,
 $q(x) = d(x) / l(x)$

It is the proportion of persons dying between the ages of x and $x+1$ to the number of persons surviving at the age of x years, *i.e.*, at the beginning of the corresponding age interval.

5) $p(x)$: Probability for a person aged x years to survive up to $x+1$ years

$$p(x) = 1 - q(x), \text{ or equivalently, } p(x) = l(x+1) / l(x)$$

It is the proportion of persons surviving up to $(x + 1)$ years to the number of persons at the age of x years.

6) $L(x)$: Number of persons lived in aggregate in the age interval $(x, x+1)$

$$\text{i.e., } L(x) = \frac{l(x) + l(x+1)}{2} \text{ or equivalently } L(x) = l(x) - \frac{1}{2}d(x)$$

7) $T(x)$: Number of years lived by the **cohort** at and above the age x *i.e.*, $T(x) = L(x) + L(x+1) + L(x+2) + \dots$ or equivalently $T(x+1) = T(x) - L(x)$.

Total number of years lived by the **cohort** after x years of age.

8) $e^0(x)$: Expectation of life

$$e^0(x) = T(x) / l(x)$$

It gives the average number of years a person of age x years is expected to survive under the existing mortality conditions.

Example

A part of the Life Table of a population is given hereunder with incomplete information. Find that information from the given data and complete the Life Table.

Age (in years)	$l(x)$	$d(x)$	$p(x)$	$q(x)$	$L(x)$	$T(x)$	$e^0(x)$
72	4412	-	-	-	-	-	-
73	3724	-	-	-	-	-	-
74	3201	642	-	-	-	26567	-

Solution:

Values of the missing entries can be calculated from the given data applying the respective formulae as follows:

The number of persons who die before reaching age $x = 72$ and 73 years can be calculated as $d(72) = l(72) - l(73)$

$$= 4412 - 3724$$

$$\begin{aligned}
 &= 688 \\
 d(73) &= l(73) - l(74) \\
 &= 3724 - 3201 \\
 d(73) &= 523.
 \end{aligned}$$

Values of $q(x)$ are estimated as

$$\begin{aligned}
 q(72) &= \frac{d(72)}{l(72)} \\
 &= \frac{688}{4412} \\
 &= 0.1559 \\
 q(73) &= \frac{d(73)}{l(73)} \\
 &= \frac{523}{3724} \\
 q(73) &= 0.1404 \\
 q(74) &= \frac{d(74)}{l(74)} \\
 &= \frac{642}{3201} \\
 &= 0.2006.
 \end{aligned}$$

Values of $p(x)$ are estimated from the corresponding values of $q(x)$ as

$$\begin{aligned}
 p(72) &= 1 - q(72) \\
 &= 1 - 0.1559 = 0.8441 \\
 p(73) &= 1 - q(73) \\
 &= 1 - 0.1404 = 0.8596 \\
 p(74) &= 1 - q(74) \\
 &= 1 - 0.2006 = 0.7994.
 \end{aligned}$$

Values of $L(x)$ are estimated using its relationship with $l(x)$ and $d(x)$ as follows:

$$L(72) = \frac{l(72) + l(73)}{2}$$

$$= \frac{4412 + 3724}{2}$$

$$= 4,068$$

$$L(73) = \frac{l(73) + l(74)}{2}$$

$$= \frac{3724 + 3201}{2}$$

$$L(74) = l(74) - \frac{d(74)}{2}$$

$$= 3201 - \frac{642}{2}$$

$$= 2880.$$

The value of $T(x)$ is estimated for $x = 72$ and 73 from the given value of $T(74)$ and the estimated values of $L(72)$ and $L(73)$ as

$$p(73) = L(73) + T(74)$$

$$= 3463 + 26567 = 30,030.$$

$$p(72) = L(72) + T(73)$$

$$4068 + 30030 = 34,098.$$

The life expectancy of the **cohort** at the age $x = 72, 73$ and 74 years is estimated using the relationship

$$e^0(x) = \frac{T(x)}{l(x)}$$

as

$$e^0(72) = \frac{34098}{4412} = 7.73$$

$$e^0(73) = \frac{30030}{3724} = 8.06$$

$$e^0(74) = \frac{26567}{3201} = 8.30$$

The completed Life Table is as follows:

x	$l(x)$	$d(x)$	$p(x)$	$q(x)$	$L(x)$	$T(x)$	$e^0(x)$
72	4412	688	0.8441	0.1559	4,068	34,098	7.73
73	3724	523	0.8596	0.1404	3,463	30,030	8.06
74	3201	642	0.7994	0.2006	2,880	26,567	8.30

Failure (or hazard) rate

The failure rate is defined for non-repairable populations as the (instantaneous) rate of failure for the survivors to time t during the next instant of time. It is a rate per unit of time similar in meaning to reading a car speedometer at a particular instant and seeing 45 mph. The next instant the failure rate may change and the units that have already failed play no further role since only the survivor's count.

The failure rate (or hazard rate) is denoted by $h(t)$ and is calculated from

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{R(t)} = \text{the instantaneous (conditional) failure rate.}$$

The failure rate is sometimes called a "conditional failure rate" since the denominator $1 - F(t)$ (i.e., the population survivors) converts the expression into a conditional rate, given survival past time t .

Since $h(t)$ is also equal to the negative of the derivative of $\ln[R(t)]$, we have the useful identity:

$$F(t) = 1 - \exp \left[- \int_0^t h(t) dt \right]$$

If we let

$$H(t) = \int_0^t h(t) dt$$

be the Cumulative Hazard Function, we then have $F(t) = 1 - e^{-H(t)}$. Two other useful identities that follow from these formulas are:

$$h(t) = - \frac{d \ln R(t)}{dt}$$

$$H(t) = - \ln R(t) .$$

It is also sometimes useful to define an average failure rate over any interval (T_1, T_2) that "averages" the failure rate over that interval. This rate, denoted by $AFR(T_1, T_2)$, is a single number that can be used as a specification or target for the population failure rate over that interval. If T_1 is 0, it is dropped from the expression. Thus, for example, $AFR(40,000)$ would be the average failure rate for the population over the first 40,000 hours of operation.

The formulas for calculating AFR values are:

$$AFR(T_2 - T_1) = \frac{\int_{T_1}^{T_2} h(t) dt}{T_2 - T_1} = \frac{H(T_2) - H(T_1)}{T_2 - T_1} = \frac{\ln R(T_1) - \ln R(T_2)}{T_2 - T_1}$$

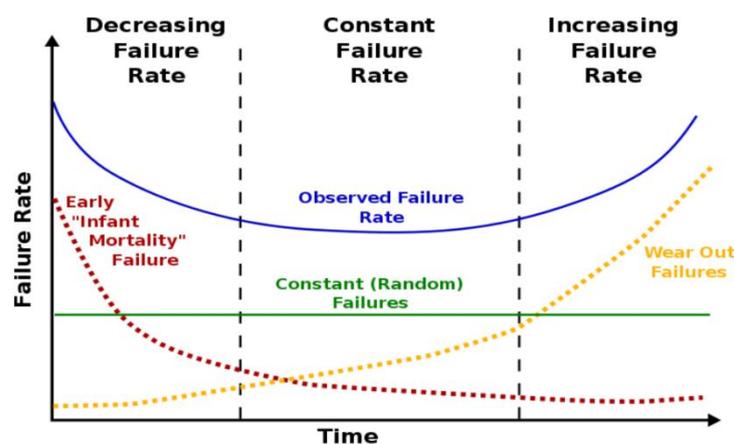
and

$$AFR(0, T) = AFR(T) = \frac{H(T)}{T} = \frac{-\ln R(T)}{T} .$$

IFR and DFR

- **Increasing Failure Rates** describe things which are more likely to fail with age, such as machines whose parts wear out.
- **Decreasing Failure Rates** describe things that are less likely to wear out with time: a business that has lasted two centuries is less likely to go bankrupt than one that has lasted two years.
- **Constant Failure Rates** describe things with exponential lifetimes.
- **Bathtub-Shaped Failure Rates** describe things that have relatively high failure rates when very young or very old, but flat rates in middle age (such as human beings and some machines).

Graphs of some of the main kinds of failure rate behavior



Mean Residual Life

Mean residual life is the expected remaining life.

$$\text{mrl}(x) = E(X - x \mid X > x)$$

It is the area under $S(x)$ to the right of x divided by $S(x)$.

$$\text{mrl}(x) = \frac{\int_x^\infty (t - x)f(t)dt}{S(x)} = \frac{\int_x^\infty S(t)dt}{S(x)}.$$

(Hint: integration by parts and $f(t)dt = -dS(t)$.)

Mean and Variance (Elementary Property)

The mean and variance of life can be expressed in terms of survival function.

The mean life is $\mu = \text{mrl}(0)$.

$$\mu = E(X) = \int_0^\infty tf(t)dt = \int_0^\infty S(t)dt.$$

The variance of X is

$$\text{var}(X) = 2 \int_0^{\infty} tS(t)dt - \left[\int_0^{\infty} S(t)dt \right]^2$$

Estimating the Survival Function

There are three methods. The first method is a parametric approach. This method assumes a parametric model (e.g., exponential distribution) of the data and we estimate the parameter first then form the estimator of the survival function. A second approach is to compute the EDF first and then converted it to an estimator of the survival function. The last approach is a powerful nonparametric method called the Kaplan-Meier estimator.

Parametric Approach:

Assume that we model the distribution as an exponential distribution with unknown parameter λ . An estimator of λ is

$$\hat{\lambda} = \frac{1}{\bar{T}_n} = \frac{n}{\sum_{i=1}^n T_i}$$

Then we estimate the survival function using

$$\hat{S}_1(t) = \hat{\lambda}e^{-\hat{\lambda}t} = \frac{e^{-\frac{t}{\bar{T}_n}}}{\bar{T}_n}, \quad t \geq 0.$$

EDF Approach:

Recall that the EDF $F(t)$ will be

$$\hat{F}(t) = \frac{1}{n} \sum_{i=1}^n I(T_i \leq t).$$

Then the survival function can be estimated by

$$\hat{S}_2(t) = 1 - \hat{F}(t) = \frac{1}{n} \sum_{i=1}^n I(T_i > t).$$

Kaplan-Meier estimator

When we estimate and graph survival curves using the Kaplan-Meier (KM) method. The estimated survival probabilities are computed using a product limit formula.

The data layout based on the Kaplan-Meier survival curves are derived is as follows: **(table)**

The first column in the table gives ordered survival times from smallest to largest. The second column gives frequency counts of failures at each distinct failure time. The third column gives frequency counts, denoted by q_f , of those persons censored in the time interval starting with failure time $t_{(f)}$ up to but not including the next failure time, denoted by $t_{(f+1)}$. The last column gives the risk set, which denotes the collection of individuals who have survived at least to time $t_{(f)}$.

Ordered failure times, $t_{(f)}$	# of failures m_f	# censored in $[t_{(f)}, t_{(f+1)})$, q_f	Risk set, $R(t_{(f)})$
$t_{(0)} = 0$	$m_0 = 0$	q_0	$R(t_{(0)})$
$t_{(1)}$	m_1	q_1	$R(t_{(1)})$
$t_{(2)}$	m_2	q_2	$R(t_{(2)})$
.	.	.	.
.	.	.	.
.	.	.	.
$t_{(k)}$	m_k	q_k	$R(t_{(k)})$

Example

The remission times (weeks) for two groups of leukemia patients are given below:

Group 1 - (n = 21): treatment	Group 2 - (n = 21): Placebo
6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

To compare the survival curves for the two treatments by calculating survival probabilities using Kaplan-Meier method.

Procedure

- To calculate number of events occur in two treatment group.
- To calculate number of censorings occur in two treatment group.
- To calculate number of persons at risk in two treatment group.
- To estimate probability of surviving.

$$P_i = \text{surviving past } t_{(f)} / \text{Total number of persons at risk}$$

- To calculate the estimated survivorship function,

$$\hat{S}(t) = \frac{\text{number of patients surviving longer than } t}{\text{total number of patients}}$$

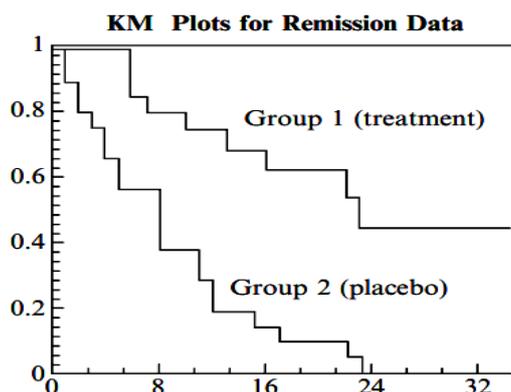
Calculation

Treatment Group

t_f	n_f	m_f	q_f	P_f	$S^{\wedge}(t)$
0	21	0	0	21/21	1
6	21	3	1	18/21	$1 \times 0.857 = 0.857$
7	17	1	1	16/17	$0.857 \times 0.941 = 0.807$
10	15	1	2	14/15	$0.807 \times 0.933 = 0.753$
13	12	1	0	11/12	$0.735 \times 0.917 = 0.690$
16	11	1	3	10/11	$0.690 \times 0.909 = 0.627$
22	7	1	0	6/7	$0.627 \times 0.857 = 0.537$
23	6	1	5	5/6	$0.537 \times 0.833 = 0.448$

Placebo Group

t_f	n_f	m_f	q_f	P_f	$S^{\wedge}(t)$
0	21	0	0	21/21	1
1	21	2	0	19/21	0.905
2	19	2	0	17/21	0.809
3	17	1	0	16/21	0.762
4	16	2	0	14/21	0.667
5	14	2	0	12/21	0.571
8	12	4	0	8/21	0.381
11	8	2	0	6/21	0.286
12	6	2	0	4/21	0.190
15	4	1	0	3/21	0.143
17	3	1	0	2/21	0.095
22	2	1	0	1/21	0.048
23	1	1	0	0/21	0



Conclusion

In an experiment comparing two groups of treatments the Kaplan-Meier curve for group 1 is consistently higher than the Kaplan-Meier curve for group 2. These figures indicate that group 1, which is the treatment group, has better survival prognosis than group 2, the placebo group. Moreover, as the number of weeks increases, the two curves appear to get farther apart, suggesting that the beneficial effects of the treatment over the placebo are greater the longer one stays in remission.

Example

The remission times (weeks) for two groups of leukemia patients are given below:

Group 1 - (n = 21): treatment	Group 2 - (n = 21): Placebo
6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

To compare the survival curves for the two treatments using Log-Rank test.

Procedure

- To calculate the observed counts (failures) occurred in two treatment group.
- To calculate the proportion in risk set occurred in two treatment group.
- To calculate the expected counts occurred in two treatment group.
- To estimate the variance is,

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

- To calculate the log-rank test statistic is,

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

Null hypothesis

H_0 : there is no significant difference between the two survival curves.

Calculation

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)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Var	0.49	0.49	0.25	0.48	0.47	0.65	0.24	0.87	0.23	0.45	0.42	0.19	0.20	0.17	0.18	0.30	0.20

$$(O_2 - E_2) = 10.26$$

$$\text{Var}(O_i - E_i) = 6.29$$

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

$$= ((10.26^2)/6.29)$$

$$= 16.79$$

Then we calculate the approximate formula is of the classic chi-square form that sums over each group being compared,

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

$$(O_1 - E_1) = -10.26 \text{ and } (O_2 - E_2) = 10.26$$

$$E_1 = 19.25 \text{ and } E_2 = 10.75$$

$$\chi^2 = (-10.26^2)/19.25 + (10.26^2)/10.75$$

$$= 5.46 + 9.77$$

$$= 15.23$$

The chi-square value obtained is 15.23, which is slightly smaller than the log-rank statistic of 16.79.

Table value at 1df = 3.84

Conclusion

For this test the decision rule is to Reject H_0 if $\chi^2 > 3.84$. We observe $\chi^2 = 15.23$, which exceeds the critical value of 3.84. Therefore, we reject H_0 . Hence, we conclude that the two survival curves are different.

Gehan's Test

The Gehan Generalized Wilcoxon Test is a distribution-free two-sample test and it is a generalization of the Wilcoxon test that samples right-censored observations.

In Gehan's generalized Wilcoxon test every observation x_i or x_i^+ in group 1 is compared with every observation y_j or y_j^+ in group 2 and a score U_{ij} is given to the result of every comparison. Calculate the test statistic is,

$$Z = \frac{W}{\sqrt{\text{Var}(W)}},$$

where

$$W = \sum_{i=1}^{n_1} U_i$$

$$U_i = R_{1i} + R_{2i}$$

$$\text{Var}(W) = \frac{n_1 n_2 \sum_{i=1}^{n_1+n_2} U_i^2}{(n_1 + n_2)(n_1 + n_2 - 1)}$$

H₀: The two treatments are equally effective.

Gehan test statistic $\sim \chi^2$ with 1df under H₀.

Example

Ten female patients with breast cancer are randomized to receive either CMF (cyclic administration of cyclophosphamide, methotrexate, and fluorouracil) or no treatment after a radical mastectomy. At the end of two years, the following times to relapse (or remission times) in months are recorded:

CMF (group 1): 23, 16+, 18+, 20+, 24+

Control (group 2): 15, 18, 19, 19, 20

To compare the treatments using Gehan wilcoxon test.

Procedure

- State the null and alternative hypothesis.
- To calculate Rank from left to right in each treatment group.
- To calculate Rank from right to left in each treatment group.
- To estimate W,

$$W = \sum_{i=1}^{n_1} U_i$$

- To estimate Variance of W,

$$\text{Var}(W) = \frac{n_1 n_2 \sum_{i=1}^{n_1+n_2} U_i^2}{(n_1 + n_2)(n_1 + n_2 - 1)}$$

- To calculate Gehan wilcoxon test statistic is

$$Z = \frac{W}{\sqrt{\text{Var}(W)}}$$

Calculation

Hypothesis

H_0 : The two treatments are equally effective i.e., $S_1 = S_2$.

H_1 : CMF more efficient than no treatment i.e., $S_1 > S_2$.

Calculating U_i for Gehan's Generalized Wilcoxon Test

Observations of Two Samples in Ascending Order	15	16+	18	18+	19	19	20	20+	23	24+
Computation of R_{1i}										
Rank from left to right, omitting Censored observations	1		2		3	4	5		6	
Assign next-higher rank to censored observations		2		3				6		7
Reduce the rank of tied observations to the lower rank for the value						3				
R_{1i}	1	2	2	3	3	3	5	6	6	7
Computation of R_{2i}										
Rank from right to left	10	9	8	7	6	5	4	3	2	1
Reduce the rank of tied observations to the lowest rank for the value					5					
Reduce the rank of censored observations to 1		1		1				1		1
R_{2i}	10	1	8	1	5	5	4	1	2	1
$U_i = R_{1i} - R_{2i}$	-9	1	-6	2	-2	-2	1	5	4	6
U_i^2	81	1	36	4	4	4	1	25	16	36

Thus, $W = 1 + 2 + 5 + 4 + 6 = 18$

$$\text{Var}(W) = \frac{5 \times 5 \times 208}{(5+5)(5+5-1)} = \frac{25 \times 208}{10 \times 9} = 57.78$$

$$Z = \frac{W}{\sqrt{\text{Var}(W)}} = \frac{18}{\sqrt{57.78}} = 2.368$$

$$Z\text{-table value} = Z_{0.05} = 1.96$$

Conclusion

For this test the decision rule is to Reject H_0 if $Z > 1.96$. We observe $Z = 2.368$, which exceeds the critical value of 1.96. Therefore, we reject H_0 . Hence, we conclude that the CMF more efficient than no treatment i.e., $S_1 > S_2$.

Cox-Mantel Test

Let $t_{(1)} < t_{(2)} < \dots < t_{(k)}$ be the distinct failure times in the two groups together and m_i be the number of failure times equal to t_i , or the multiplicity of t_i , so that

$$\sum_{i=1}^k m_{(i)} = r_1 + r_2$$

Further, let $R(t)$ be the set of people still exposed to risk of failure at time t , whose failure or censoring times are at least t . Here $R(t)$ is called the *risk set* at time t . Let n_{1t} and n_{2t} be the number of patients in $R(t)$ that belong to treatment groups 1 and 2, respectively. The total number of observations, failure or censored in $R(t_{(i)})$, is $r_{(i)} = n_{1t} + n_{2t}$.

$$U = r_2 - \sum_{i=1}^k m_{(i)} A_{(i)}$$

$$I = \sum_{i=1}^k \frac{m_{(i)}(r_{(i)} - m_{(i)})}{r_{(i)} - 1} A_{(i)}(1 - A_{(i)})$$

where $r_{(i)}$ is the number of observations, failure or censored, in $R(t_{(i)})$ and $A_{(i)}$ is the proportion of $r_{(i)}$ that belong to group 2. The two-sample test for Cox-Mantel is,

$$C = \frac{U}{\sqrt{I}}$$

Example

Ten female patients with breast cancer are randomized to receive either CMF (cyclic administration of cyclophosphamide, methotrexate, and fluorouracil) or no treatment after a radical mastectomy. At the end of two years, the following times to relapse (or remission times) in months are recorded:

CMF (group 1): 23, 16+, 18+, 20+, 24+

Control (group 2): 15, 18, 19, 19, 20

To perform the Cox-Mantel test.

Procedure

- State the null and alternative hypothesis.
- To calculate m_i , n_{1t} , n_{2t} , $r_{(i)}$ and $A_{(i)}$.
- To estimate U and I is,

$$U = r_2 - \sum_{i=1}^k m_{(i)} A_{(i)}$$

$$I = \sum_{i=1}^k \frac{m_{(i)}(r_{(i)} - m_{(i)})}{r_{(i)} - 1} A_{(i)}(1 - A_{(i)})$$

- To calculate Cox-Mantel test statistic is

$$C = \frac{U}{\sqrt{I}}$$

Calculation

Hypothesis

H_0 : The two treatments are equally effective i.e., $S_1 = S_2$.

H_1 : CMF more efficient than no treatment i.e., $S_1 > S_2$.

Computations of Cox--Mantel Test

t_i	m_i	n_{1t}	n_{2t}	$r_{(i)}$	$A_{(i)}$	$m_i * A_{(i)}$	I
15	1	5	5	10	0.5	0.5	0.25
18	1	4	4	8	0.5	0.5	0.25
19	2	3	3	6	0.5	1	0.4
20	1	3	1	4	0.25	0.25	0.1875
23	1	2	0	2	0	0	0
Total	6	17	13	30	1.75	2.25	1.0875

Where $r_1 = 1$ and $r_2 = 5$

Then, $U = 5 - 2.25$

$$= 2.75$$

$$I = 0.25 + 0.25 + 0.4 + 1.0875$$

$$= 1.0875$$

The Cox-Mantel test statistic is

$$C = \frac{U}{\sqrt{I}} = \frac{2.75}{\sqrt{1.0875}} = 2.637$$

$$Z\text{-table value} = Z_{0.05} = 1.96$$

Conclusion

For this test the decision rule is to Reject H_0 if $Z > 1.96$. We observe $Z = 2.637$, which exceeds the critical value of 1.96. Therefore, we reject H_0 . Hence, we conclude that the CMF more efficient than no treatment i.e., $S_1 > S_2$.

Z critical Table Value

Confidence ($1-\alpha$), 100%	Significance α	Critical Value $Z_{\alpha/2}$
90%	0.1	1.645
95%	0.05	1.96
98%	0.02	2.326
99%	0.01	2.576

Chi-Square Table

Df	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.01	0.02	0.051	0.103	0.211	4.605	5.991	7.378	9.21	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.86
5	0.412	0.554	0.831	1.145	1.61	9.236	11.07	12.833	15.086	16.75
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.69	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.18	2.733	3.49	13.362	15.507	17.535	20.09	21.955
9	1.735	2.088	2.7	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.94	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.92	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.3
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.66	5.629	6.571	7.79	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.39	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.26	9.591	10.851	12.443	28.412	31.41	34.17	37.566	39.997
21	8.034	8.897	10.283	11.591	13.24	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.26	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.98	45.559
25	10.52	11.524	13.12	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.16	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.29
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.42	76.154	79.49
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.54	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169