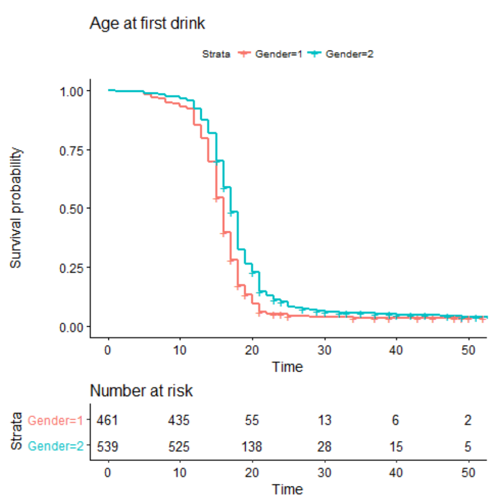
Non-parametric methods: comparing groups

Shannon Pileggi

STAT 417

OUTLINE

Comparing survival experiences of independent populations



Which group appears to have the larger median survival time?

1. `1 = males`
2. `2 = females`

Briefly comment on the survival experiences by gender.

Comparing survival experiences of independent populations

The previous figure reveals an observed difference in the estimated survival curves among males and females based on the samples of ages. Why can't we stop here?

Inference procedures to compare survival experiences

To compare survival curves over the range of time t among several independent populations, we can conduct formal tests.:

1.

2.

Age at first drink, by-hand example

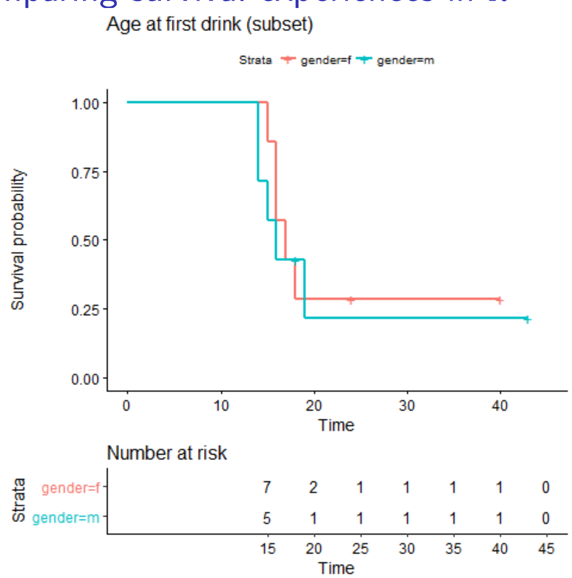
Males	43+	15	19	14	18+	16	14
Females	18	15	17	16	40+	24+	16

R Code

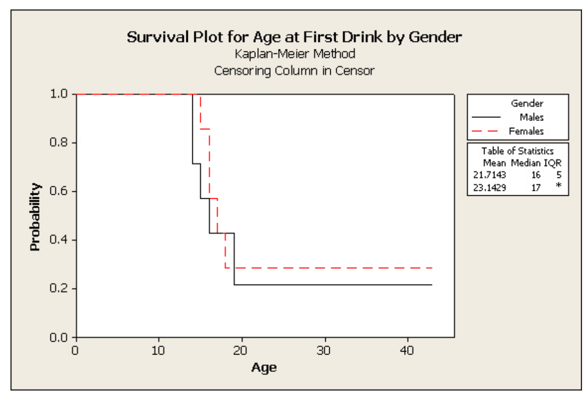
```
drinksub <- data.frame(time   = c(43, 15, 19, 14, 18, 16, 14,
                                  18, 15, 17, 16, 40, 24, 16),
                       censor = c( 0,  1,  1, 1,   0,  1,  1,
                                   1,  1,  1, 1,   0,  0,  1),
                       gender = c("m","m","m","m","m","m","m",
                                  "f","f","f","f","f","f","f"))

KM_obj <- survfit(Surv(time, censor) ~ gender, data = drinksub)

library(survminer)
ggsurvplot(KM_obj, data = drinksub,
           risk.table = TRUE,
           title = "Age at first drink (subset)")
```

R Code

Comparing survival experiences in R



Comparing survival experiences in Minitab



Log-rank test

The log-rank test can be used to compare survival experiences for **two** populations. The general null and alternative hypotheses are:

H_0:

H_a:

H_0 and H_a for age at first drink

In the age at first drink example, let `1 = males` and `2 = females`. Then the null and alternative hypotheses are:

H_0:

H_a:

Log-rank test details

The approach is to compare the total number of *observed* events to the total number of *expected* events in Group 1 under the assumption that $S_1(t) = S_2(t)$, where:

- m
- n_{1i}
- n_{2i}
- $n_i = n_{1i} + n_{2i}$
- d_{1i}, d_{2i}, and d_i

Log-rank test details

It is typically assumed that the number of event occurrences that will occur in time interval i, for Group 1 follows a *hypergeometric* probability distribution.

- Let d_i/n_i be the overall proportion of individuals at time $t_{(i)}$ who experience the event. Then the expected number of event occurrences in Group 1 at time $t_{(i)}$, denoted E_{1i}, is given by:
- Also, based on the assumption that the number of event occurrences at time $t_{(i)}$ for Group 1 follows a hypergeometric distribution, the variance for the number of event occurrences for Group 1 at time $t_{(i)}$, denoted V_{1i}, is given by:

Age at first drink, by-hand example

Males	43+	15	19	14	18+	16	14
Females	18	15	17	16	40+	24+	16

Write out the ordered event times:

Order	1	2	3	4	5	6	7
Males							
Females							

Age at first drink, by-hand example

i	$t_{(i)}$	Interval	n_i	d_i	d_i/n_i	n_{1i}	d_{1i}	n_{2i}	E_{1i}	V_{1i}
0		$[0, 14)$								
1		$[14, 15)$								
2		$[15, 16)$								
3		$[16, 17)$								
4		$[17, 18)$								
5		$[18, 19)$								
6		$[19, 43)$								

Example calculations

E_{11}:

V_{11}:

Log-rank test details

We compare the total observed and total expected counts over *all* the complete event times with the statistic:

$$\frac{observed - expected}{sd} =$$

- $\sum_{i=1}^{m} d_{1i}$ is the sum of the observed event occurrences
- $\sum_{i=1}^{m} E_{1i}$ is the sum of the expected event occurrences
- $\sum_{i=1}^{m} V_{1i}$ is the variance of the total number of event occurrences over the m complete event times

Log-rank test details

- This quantity is essentially a z-score, i.e:
- By the Central Limit Theorem, when each sample size n_i is "reasonably large," the above term follows approximately a standard normal distribution.

Log-rank test details

- It is more common to see the square of the statistic reported in statistical software.
- The square of the above statistic is called the **log-rank test statistic** (for two groups) given by:
- This test statistic follows a **chi-square distribution** with 1 degree of freedom.

Discussion

If the observed and expected number of events are far apart, then

- χ^2_L will be ("big", "small"),
- the corresponding p-value will be ("big", "small"),
- leading to (evidence / no evidence) to reject H_0.

Age at first drink, by-hand example

Calculate the value of the log-rank test statistic:

Minitab output

```
Minitab Output

...
Test Statistics

Method     Chi-Square  DF  P-Value
Log-Rank     0.131114   1    0.717
Wilcoxon     0.425355   1    0.514

Minitab Output
```

Conclusion?

Wilcoxon test

- The log-rank test statistic can be considered a special case of the following statistic:

  where the w_i's are "weights" all equal to _______.
- The **Wilcoxon test statistic**, denoted X_W^2, is also a special case of above statistic with weights $w_i = n_i$, i.e:

$$X_W^2 = \frac{\left[\sum_{i=1}^{m} n_i(d_{1i} - E_{1i})\right]^2}{\sum_{i=1}^{m} n_i^2 V_{1i}}$$

  where n_i is the number of subjects at risk just prior to time $t_{(i)}$
- The Wilcoxon test statistic, X^2, also follows a χ^2-distribution with 1 degree of freedom.

Wilcoxon test statistic

Verify that the value of the Wilcoxon test statistics is 0.4253.

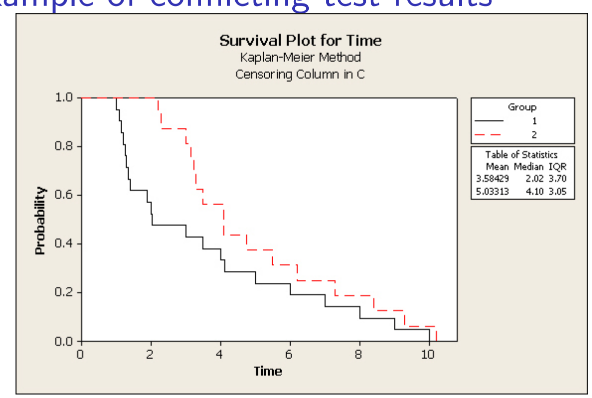
Comparison of log-rank and Wilcoxon tests

1.

2.

3.

Example of conflicting test results



Minitab Output

```
Method     Chi-Square  DF  P-Value
Log-Rank      2.13966   1    0.144
Wilcoxon      4.27216   1    0.039
```

Minitab Output

Log-rank test results in R

R Code

```
survdiff(Surv(time, censor) ~ gender, data = drinksub)
```

R Code

R Output

```
Call:
survdiff(formula = Surv(time, censor) ~ gender, data = drinksub)

         N Observed Expected (O-E)^2/E (O-E)^2/V
gender=f 7        5     5.54    0.0523     0.131
gender=m 7        5     4.46    0.0649     0.131

 Chisq= 0.1  on 1 degrees of freedom, p= 0.717
```

R Output

Getting started (2 groups)

Log-rank (2 groups)

Software output (2 groups)

> 2 groups

Comparing survival experiences of > 2 populations

- The log-rank and Wilcoxon tests can easily be extended to $k > 2$ populations.
- The null and alternative hypotheses are now:

  H_0:

  H_a:

- Test statistic computations involve variance/covariance matrix calculations, so no mathematical details will be presented here. The distribution of the test statistic is:

Comparing survival experiences of > 2 populations

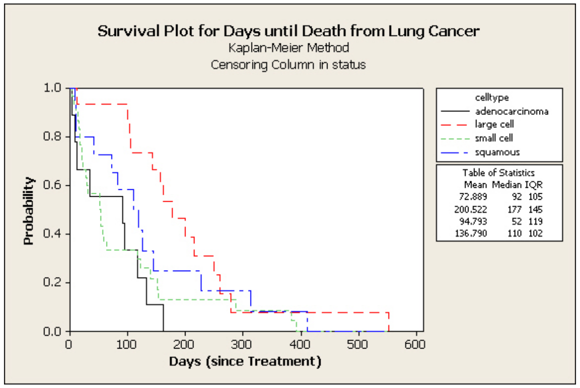
- Test statistic computations involve variance/covariance matrix calculations, so no mathematical details will be presented here. The distribution of the test statistic is:
- The extension of the log-rank test to more than two groups can be implemented in Minitab and R, while the Wilcoxon test can only be implemented in Minitab.

Lung cancer example (VALCSG)

- There are two primary classifications of lung cancer based on the size and appearance of the malignant cells under a microscope: *small cell* and *non-small cell*.
- **Small cell** lung cancer: aggressive and spreads quickly through the body
- Non-small cell lung cancer spreads more slowly and can be further sub-classified depending on origin and how the cells spread:
  - **Squamous** cell carcinoma: Cancer that begins in squamous cells, which are thin, flat cells that look like fish scales.
  - **Large cell** carcinoma: Cancer that may begin in several types of large cells.
  - **Adenocarcinoma**: Cancer that begins in the cells that line the alveoli and make substances such as mucus

The Veterans Administration Lung Cancer Study Group (VALCSG) investigated the effects of two treatments (a standard treatment and test treatment) on the survival of patients.

Lung cancer example (VALCSG)



Lung cancer example (VALCSG)

Minitab Output

```
...
Test Statistics

Method     Chi-Square  DF  P-Value
Log-Rank       9.6408   3    0.022
Wilcoxon      11.5123   3    0.009
```

Minitab Output

H_0:

H_a:

Lung cancer example (VALCSG)

- State conclusion based on log-rank test at $\alpha = 0.01$.

- State conclusion based on Wilcoxon test at $\alpha = 0.01$.

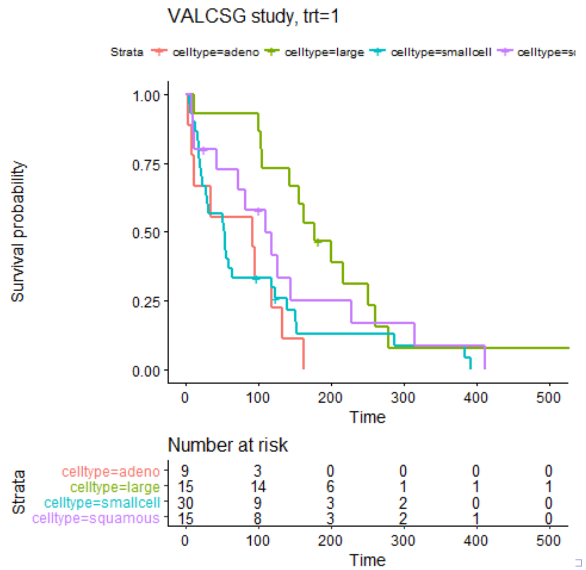
Lung cancer example (VALCSG) in R

R Code

```
# focus on treatment group 1 only
veteran1 <- veteran[veteran$trt == 1, ]

KM_obj <- survfit(Surv(time, status) ~ celltype,
                  data = veteran1)

ggsurvplot(KM_obj, data = veteran1,
           risk.table = TRUE,
           title = "VALCSG study, trt=1")

survdiff(Surv(time, status) ~ celltype, data = veteran1)
```

R Code

Lung cancer example (VALCSG) in R



Lung cancer example (VALCSG) in R

R Output

```
Call:
survdiff(formula = Surv(time, status) ~ celltype, data = veteran1)

                    N Observed Expected (O-E)^2/E (O-E)^2/V
celltype=adeno      9        9     5.17     2.832     3.194
celltype=large     15       14    23.17     3.626     6.217
celltype=smallcell 30       28    21.14     2.225     3.503
celltype=squamous  15       13    14.52     0.159     0.211

 Chisq= 9.6  on 3 degrees of freedom, p= 0.0219
```

R Output