Lecture Four: Comparing Multiple Samples: Non-parametric tests

1. Graphics

- Estimate survivor curve by, for example, Splus "survfit()"
- Visual examination: plot the estimated survivor curve (use Splus "plot.survfit()")
- Example (plot): The study of prostatic cancer patients (Table 1.4)

2. Characterization of Differential Survival

- Survivor functions for two groups, $S_1(t)$, $S_2(t)$
- Statistical Formulation

Null hypothesis	H_0: $S_1(t) = S_2(t)$
Alternative hypothesis	H_a: $S_1(t) \neq S_2(t)$

- Specific Alternative Hypothesis
  - Example 1

    H_a: $S_1(t) < S_2(t)$

  - Example 2

    H_a: $S_1(t) > S_2(t)$, $t < t_0$

  - Example 3

    H_a: $S_1(t) > S_2(t)$, $t < t_0$
    $S_1(t) < S_2(t)$, $t \geq t_0$

- Sources contributing to the difference exhibited in data
  - Difference due to sample/data variation (chance)
  - Difference due to treatment

- Quantifying difference due to treatments
  - Eliminating random (sample) variation?
  - "separating" treatment effects from random variation?
    - What is the likelihood/chance of observing such a difference exhibited in the data if treatment effects are absent under the null hypothesis?

- The smaller the likelihood/chance, the stronger the evidence that the treatment effects are present (the null hypothesis is incorrect)
- This chance or likelihood is often quantified by p-value (probability)

3. Components/Reasoning of hypothesis testing

- Null hypothesis, e.g. H_0: $S_1(t) = S_2(t)$
- Sample(s) of data
- Testing statistics (e.g. Chi-square statistic) that are sensitive to certain departure from the null hypothesis
- A test statistic is subject to sampling variation, and is a random variable
- The test statistic follows a sampling distribution under the null hypothesis; it must follow a different distribution if the null hypothesis is not true.
- Upon observing the value of the test statistic based on sample, compare this value with the reference sampling distribution under the null hypothesis
  - If the null hypothesis is true, the observed value of the statistic would not likely to be extreme (little evidence against the null)
  - If extreme,
    - More likely (power), the value of the test statistic came from a distribution different from that under the null, hence evidence against the null
    - Although very unlikely, the null hypothesis may still be true, this unlikeliness is measured by the p-value or type I error.

4. Nonparametric Tests: log-rank Tests

- Intuition: an illustration using two group comparison

If there is no treatment effects on survival (the null hypothesis), the survivorship would be the same for the two groups besides random variation. Suppose we observe an event, this event could have occurred to any individual with equal chance regardless of his/her group membership. From data analytic standpoint, the two groups of data would blend well—if we order the data, there would be no group segregation—a building block for many non-parametric techniques.

- For each time interval $(t_{(j-1)}, t_{(j)}]$, in which there is only one distinct failure time (allow ties), we have a 2 by 2 table

Group	# of deaths at $t_{(j)}$	# of surviving beyond $t_{(j)}$	# at risk just before $t_{(j)}$
I	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
II	d_{2j}	$n_{2j} - d_{2j}$	n_{2j}
Total	d_j	$n_j - d_j$	n_j

- n : the size of risk sets
- d: the number of failures
- subscript: treatment groups

- Analysis of a single 2 by 2 table

  - The null hypothesis H_0: $S_1(t) = S_2(t)$, implies that failure probabilities $q_{01} = q_{02}$
    - If an event is to occur, every individual at risk, regardless of his/her being in treatment I or treatment II group, has the equal chance being the "victim"
    - Therefore, the event coming from I is of n_{1j} / n_j; coming from II is of the chance n_{2j} / n_j
    - Given d_j events in this time interval, we expect

      $e_{1j} = d_j * n_{1j}/n_j$
      $e_{2j} = d_j * n_{2j}/n_j$

      events from I and II, respectively.

  - Discrepancy between the observed failure d_{1j} and the expected number of failure e_{1j} in I would be an evidence against H_0
  - To test the significance of this discrepancy within the time window under consideration, $(t_{(j-1)}, t_{(j)}]$, 2 by 2 table analysis would be appropriate (snapshot)
  - Below is a review of the 2 by 2 table
    - $d_{1j}|d_j$ or equivalently $d_{2j}|d_j$ (why?) provides information about the difference in failure rate between the two groups
    - Under the null, the discrepancy $d_{1j} - e_{1j}$ would be small
    - we compare $d_{1j} - e_{1j}$ with the distribution of $d_{1j}|d_j$ to determine if the discrepancy is significant

      ✓ $d_{1j}|d_j$—follows hypergeometric distribution (*d_{1j} "deaths" without replacement from $n_j = n_j - d_j + d_j$*) if we assume the fixed marginal

      ✓ Mean: $E(d_{1j}|d_j) = e_{1j}$

      ✓ Variance

$$v_{1j} = \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$$

  - We have a sequence of 2 by 2 tables over time, one for each time interval
  - How to connect this sequence of snapshots together?

- Log-rank test: summarizing a sequence of 2 by 2 tables with equal weight

$$U_L = \sum_{j=1}^{r} (d_{1j} - e_{1j})$$

- Given d_j , n_{1j} and n_{2j} *(using conditional likelihood arguments- Kalbfleisch & Prentice)*

$$V_L \equiv \text{var}(U_L) = \sum_{j=1}^{r} v_{1j}$$

- Mantel-Haenszel/log-rank statistic (**by central limit theorem**)

$$\frac{U_L^2}{V_L} \sim \chi^2(1)$$

when the null hypothesis is true.

- Example 2.12: Prognosis for women with breast cancer (Table 1.2, p7)

Output (see Table 2.8 for calculation by hand):

Stratum 1: GROUP = 0 (Negative staining)

Product-Limit Survival Estimates

SURVT	Survival	Failure	Standard Error	Number Failed	Number Left
0.000	1.0000	0	0	0	13
23.000	0.9231	0.0769	0.0739	1	12
47.000	0.8462	0.1538	0.1001	2	11
69.000	0.7692	0.2308	0.1169	3	10
148.000	0.6410	0.3590	0.1522	4	5
181.000	0.5128	0.4872	0.1673	5	4

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	.	181.000	.
50	.	148.000	.
25	148.000	47.000	.

Stratum 2: GROUP = 1(positive staining)

Product-Limit Survival Estimates

SURVT	Survival	Failure	Standard Error	Number Failed	Number Left
0.000	1.0000	0	0	0	32
5.000	0.9688	0.0313	0.0308	1	31
8.000	0.9375	0.0625	0.0428	2	30
10.000	0.9063	0.0938	0.0515	3	29
13.000	0.8750	0.1250	0.0585	4	28
18.000	0.8438	0.1563	0.0642	5	27
24.000	0.8125	0.1875	0.0690	6	26
26.000	0.7500	0.2500	0.0765	8	24
31.000	0.7188	0.2813	0.0795	9	23
35.000	0.6875	0.3125	0.0819	10	22
40.000	0.6563	0.3438	0.0840	11	21
41.000	0.6250	0.3750	0.0856	12	20
48.000	0.5938	0.4063	0.0868	13	19
50.000	0.5625	0.4375	0.0877	14	18
59.000	0.5313	0.4688	0.0882	15	17
61.000	0.5000	0.5000	0.0884	16	16
68.000	0.4688	0.5313	0.0882	17	15
71.000	0.4375	0.5625	0.0877	18	14
113.000	0.3938	0.6063	0.0892	19	9
118.000	0.3445	0.6555	0.0906	20	7
143.000	0.2953	0.7047	0.0900	21	6

Summary Statistics for Time Variable SURVT

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	.	113.000	.
50	64.500	40.000	143.000
25	28.500	18.000	50.000

Summary of the Number of Censored and Uncensored Values

Stratum	GROUP	Total	Failed	Censored	Percent Censored
1	0	13	5	8	61.54
2	1	32	21	11	34.38
Total		45	26	19	42.22

Testing Homogeneity of Survival Curves for SURVT over Strata

Rank Statistics

GROUP	Log-Rank	Wilcoxon
0	-4.5651	-159.00
1	4.5651	159.00

Covariance Matrix for the Log-Rank Statistics

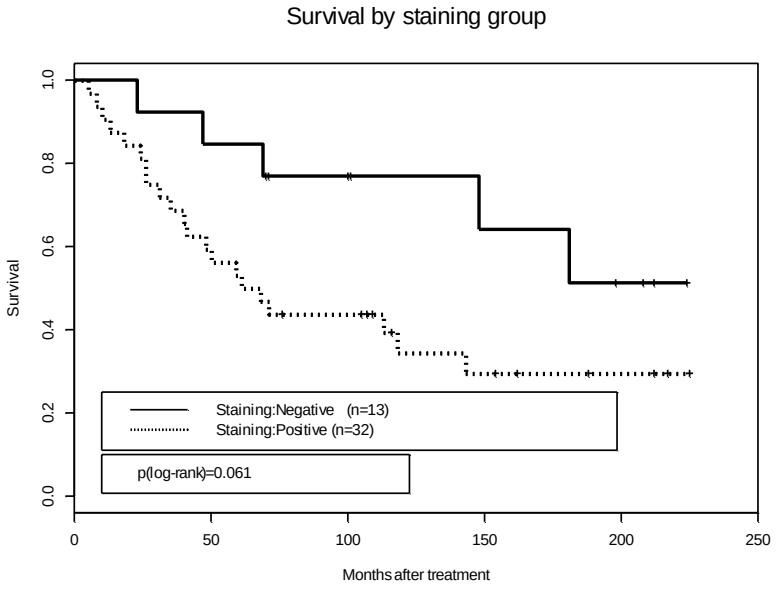
GROUP	0	1
0	5.92900	-5.92900
1	-5.92900	5.92900

Covariance Matrix for the Wilcoxon Statistics

GROUP	0	1
0	6048.14	-6048.14
1	-6048.14	6048.14

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	3.5150	1	0.0608
Wilcoxon	4.1800	1	0.0409
-2Log(LR)	4.3563	1	0.0369



- **Conclusion**: The discrepancy between the observed failure time and expected failure time under the null hypothesis is marginal; there is some evidence that the prognosis of a breast cancer patient is dependent on the result of the staining procedure.

SAS code:

```
Options ls = 80;
libname fu '../sdata';
data fu.hpa;
infile '../data/hpa.dat';
input survt censor group;
filename gsasfile 'hpa.gsf';
goptions gaccess=gsasfile ROTATE=LANDSCAPE gsfmode=replace device=ps;
 proc lifetest plots=(s) ;
        time survt*censor(0);
        strata  group;
run;
```

Splus function for generating the plot above:

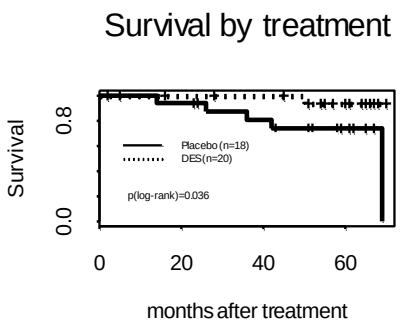
```
 hpa.s<-function(){
     tmpdf <- importData("../sdata/hpa.sas7bdat ")
    wmf.graph("hpa.wmf")
    stain.km <- survfit(Surv(survt, censor)~group, data=tmpdf)
    plot(stain.km,xlab="months after treatment", lty=1:2,
       ylab="Survival", xlim=c(0, 250), ylim=c(0,1),mark.time=T,  lwd=2)
     title("Survival by staining group")
     legend(10,0.35, paste(c("Staining:Negative  ","Staining:Positive"),
                c("(n=13)","(n=32)")), lty=1:2)
     legend(10,0.2,"p(log-rank)=0.061")
    dev.off()
 }
```

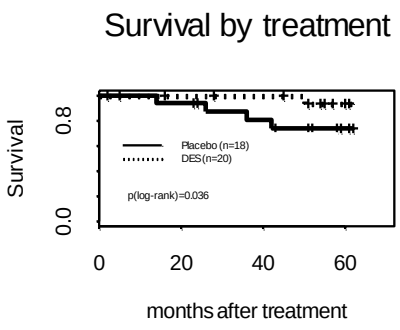
Assignment three:

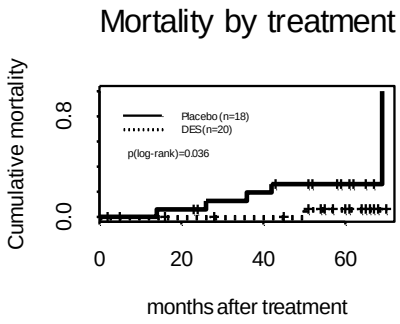
For the recurrence of bladder cancer study (Table B.2, page 493), use Kaplan-Meier estimator to investigate survival and its association with treatments (Placebo and thiotepa). In this homework, you shall

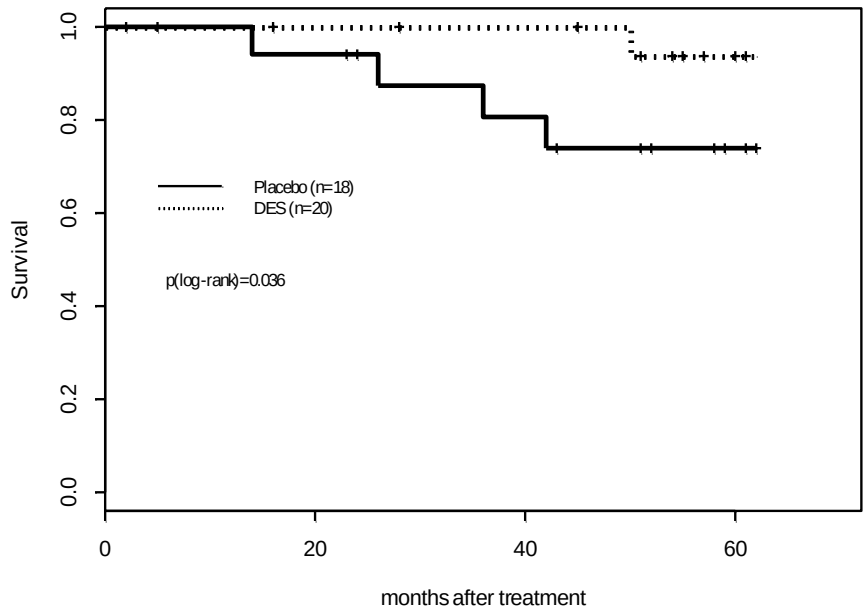
- Conduct log-rank tests to test the significance of the difference, if any, between two treatment groups.
- Interpret the results.
- Generate the Kaplan-Meier plot with appropriate legends.

Note: Generate a permanent SAS (or other software) dataset for the whole dataset (more assignments will be based on this data). In this assignment, you ignore all other covariates but treatment.









No. at Risk Placebo	18	16	12	6
No. at Risk DES	20	18	17	10