Nonparametric analysis: basic tools

Data description: The bladder cancer dataset contains recurrent event outcome information on 85 patients followed for the recurrence of bladder cancer tumor after transurethral surgical excision. The esposure of interest is the effect of the drug treatment fo thiotepa. Control variables are initial number and intial size of tumors. The status variable is 1 for last recurrence and 0 for everything else (including death for any reason (only for practical purposes)). Time to event is recorded in months. The main aim of this study is to evaluate if the treatment is protective for recurrence.

1. Working on the pooled sample (85 patients):

```
cd "working directory path"
use bladder.dta,clear
```

2. Describe data looking at frequency distribution of variables by treatment and interpret the results

```
. describe
. tabulate rx
. tab number rx,col row chi2
. sum size if rx==1,d
. sum size if rx==2,d
. ttest size,by(rx)
. tab event rx,col row chi2
```

3. Declare data to be survival-data in order to take into account the follow-up and interpret the new generated variables

```
. stset stop, fail(event) id(id)
. list _st _d _t _t0 if _n<=5
```

4. Describe survival data.

```
. stdes
. stsum
. stci
```

5. Estimate and plot the survival function by the Kaplan-Meier method with Greenwood confidence bands.

```
. sts list
```

6. Estimate and plot the cumulative hazard function by the Aalen-Nelson with confidence bands.

```
. sts list,na
```

7. Provide an estimate for the cumulative hazard using the relationship between survival and hazard (S(t)=exp(-cumhaz(t))) and represent it.

```
. sts generate km=s
. gen na_km=-log(km)
. sort stop
. list(stop na_km)
. scatter na_km stop,connect(J) ytit("Cumulative hazard") xtit("Time")
```

8. Compare the results obtained in (e) and (f).

```
twoway (scatter na stop, connect(J)) (scatter na_km stop, connect(J)), ///
ytit("Cumulative hazard") xtit("Time") ///
legend (lab(1 "Nelson-Aalen") lab(2 "Nelson-Aalen by K-M"))
```

9. Estimate and plot the survival function by treatment.

```
. sort rx
. sts list,by(rx)
. sts graph,by(rx)
```

10. Compare the survival function at 19, 27, 41 months for the two treatment groups.

```
. sts list, by(rx) compare at(19 27 41)
```

11. Represent the survival function by treatment showing also the number of patients at risk and censoring.

```
. sts graph, by(rx) lost atrisk ylab(0.50 1)
```

12. Estimate and plot the cumulative hazard function by treatment.

```
. sort rx
. sts list,by(rx) na
. sts graph,by(rx) na
```

13. Represent the hazard function (smoothed estimate) by treatment.

```
. sts graph,by(rx) hazard
```

14. Test the difference between the survival curves by treatment, by using i)log-rank test, ii)Wilcoxon-Breslow-Gehan test, iii) Tarone-Ware test, iv)Peto-Peto-Prentice test.

```
. sts test rx
. sts test rx,wilcoxon
. sts test rx,tware
. sts test rx,peto
```

15. Estimate and plot the survival function by size of tumour

```
. tab size
. egen size_cat=cut(size),at(1,2,3,4,10)
. tab size_cat size
. sort size_cat
. sts list,by(size_cat)
. sts graph,by(size_cat)
. sts test size_cat
```

16. Estimate and plot the survival function by number of tumours

```
. tab number
. egen number_cat=cut(number),at(1,2,3,4,10)
. tab number_cat number
. sort number_cat
. sts list,by(number_cat)
. sts graph,by(number_cat)
. sts test number_cat
```

17. Estimate and plot the survival function by number of tumours and treatment

```
. tab number_cat rx
. sts graph,by(number_cat rx)
```

18. After re-loading data, estimate and plot the survival function by the Life-Table method on data grouped by year.

```
. use bladder.dta,clear
. ltable stop event,intervals(12)
. ltable stop event,intervals(12) graph
```

19. Estimate and plot the survival function by treatment by the Life-Table method on data grouped by year.

```
. ltable stop event,intervals(12) graph by(rx)
```

20. Estimate the mortality rate by the Life-Table method on data grouped by year.

```
. ltable stop event,intervals(12) haz
```

21. Load data bladder2.dta. It differs for previous data because information is given for both time at second last recurrence and time at last recurrence. The main aim of this study is to evaluate if the treatment is protective for recurrence + conditional on having already suffered from a recurrence.

    Working on the pooled sample (85 patients):

```
cd "working directory path"
use bladder2.dta,clear
```

22. Declare data to be survival-data in order to take into account the follow-up and interpret the new generated variables.

```
. stset stop, fail(event) entry(start) id(id)
. list _st _d _t _t0 if _n<=5
```

23. Describe survival data.

```
. stdes
. stsum
```

24. Estimate and plot the survival function by the Kaplan-Meier method by treatment.

```
. sort rx
. sts list,by(rx)
. sts graph,by(rx)
```

25. Test the difference between the survival curves by treatment, by using log-rank test.

```
. sts test rx
```