

Màster universitari en Estadística i Investigació Operativa



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH



UNIVERSITAT DE
BARCELONA

Lifetime Data Analysis, Course 2021/22

Exercises Topics 3 and 4

Delivery of exercises: November 28, 2021.

Exercise 1 (1.5 points)

The following R output provides some information on the survival times (measured in days) in a randomized clinical trial that compared two treatments for a certain cancer:

```
> survfit(Surv(stime, status) ~ treat, cancer)
```

	n	events	median	0.95LCL	0.95UCL
treat=Treatment 1	13	7	638	268	NA
treat=Treatment 2	13	5	NA	475	NA

```
> summary(survfit(Surv(stime, status) ~ treat, cancer))
```

Treatment=1

time	n.risk	n.event	survival
59	13	1	
115	12	1	
156	11	1	
268	10	1	
329	9	1	
431	8	1	
638	5	1	

Treatment=2

time	n.risk	n.event	survival
353	13	1	
365	12	1	
464	9	1	
475	8	1	
563	7	1	

- Why does the `survfit` function not provide an estimation of the median under Treatment 2?
- Why does the `survfit` function not provide an upper limit of the 95% confidence interval of the median under Treatment 1?
- Estimate the values of the survival functions at 365 days using both the Kaplan-Meier and the Nelson-Aalen estimator.
- Estimate the values of the hazard and the cumulative hazard functions at 365 days.

Exercise 2 (3.5 points)

The data frame `kidney` of the `KMsurv` package (<https://CRAN.R-project.org/package=KMsurv>) contains the times (in months) until a renal infection of 119 dialysis patients. They were divided into two study groups according to the placement of the catheter: either surgically or percutaneously.

- Draw the (estimated) survival functions in both study groups and interpret the resulting graph.

- (b) Estimate the median time until renal infection in both groups. Comment on the results.
- (c) Use the function `bshazard` of the `bshazard` package (<https://CRAN.R-project.org/package=bshazard>) to draw smoothed estimates of the hazard functions in both study groups. Interpret the graph and compare it with the one of the survival functions.
- (d) Which are the (estimated) probabilities of not suffering any renal infection after 6, 12, 18, 24, and 30 months in both study groups? Give also the corresponding 95% confidence intervals.
- (e) Which are the lower and upper limits of the linear EP confidence bands with level 95% in both study groups after 6, 12, 18, 24, and 30 months in both study groups.
- (f) Comment on the differences between confidence bands and the confidence intervals.

Note: In R, the confidence bands can be computed, e.g., with the function `km.ci` of the `km.ci` package (<https://cran.r-project.org/web/packages/km.ci/index.html>).

Exercise 3 (2.5 points)

The file `Elderly.txt` contains part of the data of a study on elderly people in Barcelona that was carried out between 1986 and 1994¹. Following, we will use these data to estimate the survival function of men older than 65 years. In particular:

- (a) Explain why the survival times are left-truncated.
- (b) Draw the number of people at risk (of dying) as a function of age.
- (c) Draw the conditional survival functions for men aged 70 and 85 years, respectively. Which are the estimated probabilities of surviving 90 years in both cases?
- (d) Which are the corresponding estimated probabilities of surviving 90 years, if left truncation was ignored? Comment on the results.

Exercise 4 (2.5 points)

The data frame `turnover` in the R workspace `Assign2Exer4.RData` is part of

“... a real dataset shared from Edward Babushkin’s blog used to predict an Employee’s risk of quitting ...”²

It includes the following variables

```
> Hmisc::Label(turnover)
```

```
label(stag)      <- 'Time in company until turnover or end of study [months]'
label(event)     <- 'Turnover indicator'
label(gender)    <- 'Gender'
label(headgend)  <- 'Gender of the supervisor'
```

- (a) Draw the survival functions of time until turnover for both women and men. What do you observe?
- (b) Draw the survival functions separately for both genders of the supervisor. What do you observe?
- (c) Test the hypothesis that time to turnover does not depend on the employee’s gender using any test of the Fleming Harrington family of tests. What do you conclude?
- (d) Test the hypothesis that time to turnover does not depend on the employee’s gender using a stratified test. What do you conclude?
- (e) Comment on the differences between both tests.

¹Lamarca, R., J. Alonso, G. Gómez G, A. Muñoz (1998): Left-truncated data with age as time scale: an alternative for survival analysis in the elderly population. *J Gerontol A Biol Sci Med Sci.*, 53(5), M337–M343

²<https://www.kaggle.com/davinwijaya/employee-turnover>