# Exercise 1: Expected residual life and estimation of survival functions

(13.3.2019)

**Expected residual life and survival function**

The expected residual life at time $t$ uniquely determines a continuous survival function with finite mean.

**Survival analysis using R**

- Tools available in the package *survival*, *KMsurv* and *eha*
- Check if this package has been installed (type installed.packages() at the console and press enter)
- If the above-mentioned packages are not installed then type install.packages("KMsurv") at the console to install *KMsurv* provided you have internet connection. If you have a zip file of the package in your PC then go to Packages Menu and select Install Package(s) from local zip files... and then select the zip file.
- To use the functions available in the package *survival*, load the package by typing *library(survival)*
- Functions which we will be using during the course:
    - *Surv* - create a survival object
    - *lifetab* - Life-table estimates (available in the package *KMsurv*)
    - *survfit* - Kaplan-Meier estimates
    - *survdiff* - log-rank test
    - *weibreg* - parametric survival regression model available in *eha*
    - *coxph* - Cox proportional hazards model
    - *rexp, rweibull, rgamma*: - generate random sample from exponential, Weibull and gamma distribution

**Estimation of survival function**

Load the data set (Veterans administration lung cancer trial, cf. Kalbfleisch and Prentice, 2002) from the R survival package:

```
library(survival)
?veteran # for explanation of the variables in the data set
data(veteran) # load the data
str(veteran) # show records of the data
```

1. Plot a histogram of the survival times corresponding to uncensored observations (veteran$status == 1). (See help(hist))
   Note: Try

```
hist(veteran$time[veteran$status == 1], nclass=30, main='', xlab='Survival time (days)',
    col='gray90')
```

by changing values of nclass and also without nclass.

2. Create an output file where the histogram is stored. Remember to first defint outpath where the file should be stored.

```
pdf(file.path(outpath, 'survtimes.pdf'))
op <- par(mar=c(4,4,0,0), mgp=c(2,1,0))
hist(veteran$time[veteran$status == 1], nclass=30, main='', xlab='Survival time (days)',
    col='gray90')
par(op)
dev.off()
```

3.    a. Use the *survifit* routine in R to calculate the Kaplan-Meier estimate of the overall survival in the data. In the survival routines of R, the response variable needs to be specified as a *survival* object. If the observed failure time variable is *time* and failure indicator variable is *status*, the response variable is created as

```
Surv(time, status)
```

Applying the plot command to the output object from the *survfit* routine, you can draw the estimate and its confidence limits. Experiment with different confidence levels (e.g. 80% and 95%). You can also practice with the plot command options (e.g. *xlab*, *ylab*).

   b. Plot the Kaplan-Meier estimates of the survival functions separately for the two treatment groups (standard vs. test). Does there appear to be a difference between the two groups in survival? Irrespective of the treatment group, compare the survival in groups defined by the histological type of tumor (variable *celltype*). You may also like to explore the effect on survival of the other covariates in the data.

   c. Compare the two treatments by the log-rank test. You can find this in the *survdiff* routine. Compare then the effect of celltype on survival.

4. Data matrix cervix contains grouped survival data for two cohorts of women, diagnosed with stage I or stage II cervix cancer. Use the *lifetab* routine in R library *KMsurv* to create life tables for both groups.