# Practical session 2:
# Estimation of survival and hazard functions

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

   ```
   librarsturvival)
   ?veteran # for explanations for the variables in the data set
   data(veteran) # load the data
   str(veteran) # show records of the data
   ```

   (a) Use the *s*urvfit routine in R to calculate the Kaplan-Meier estimate of 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 *t*ime and the failure indicator variabIe is *s*tatus, the response variable is created as

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

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

   (b) Plot the Kaplan-Meier estimates of the survival functions separater 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 then survival in groups defined by the histological type of tumor (variable celltype). You can also explore the effect survival of the other covariates in the data.

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

2. 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 library *K*Msurv to create life tables for both cohorts. You can load the required library by command *l*ibrary(KMsurv).

3. Read oldmort data from the library *eha*. Compare the survival probability of men and women graphically as well as using the log-rank test. Note: Function surv_test from the library(coin) is needed since survdiff does not allow left truncated data.

   Also, plot cumulative hazard function by birthplace separately for men and women, and interprete the results.

4. See `https://cran.r-project.org/web/views/Survival.html` for R functions available for survival analysis.