R instructions for the 11th seminar

Data set *Emamma.RData* presents data of 1000 female patients with breast cancer diagnosis treated in Masaryk Oncology Institute in Brno. The list of selected variables follows:

AGE age when diagnosis was determined;
TIME survival time in months;
Death the status indicator, (0=alive, 1=dead);
SIDE (left and right);
CHT chemotherapy (yes/no);
CHT_Ttype type of chemotherapy (no chemotherapy, CMF, FAC, other);
HT hormonal therapy (yes/no);
LR local relapse (yes/no);
MTS metastases (yes/no);
MP menopause (0 - premenopausal, 1 - postmenopausal);
HISTOL histology (1- ductal, 2 - lobular, 3 - modular, 4 - other);
STAGE stage of tumor disease (1, 2, 3, 4, higher values mean later stage)

R Instructions for the problem 1:

For the whole data set:

•build the Kaplan-Meier estimate of survival function.

```
!library(survival)
S<-Surv(Emamma$TIME,event=Emamma$Death)
SResults<-survfit(S∼1,conf.type="plain", type="kaplan-meier")
plot(SResults,conf.int=F,xlab="survival time", ylab="survival probability")
```

•Find out median, lower and upper quartile for the survival time.

```
SResults
```

•Create confidence intervals for survival function.
(It is based on formula: $lower = \hat{S}(t) - \sqrt{\hat{Var}(\hat{S}(t))} \cdot u_{1-\alpha/2}$; $upper = \hat{S}(t) + \sqrt{\hat{Var}(\hat{S}(t))} \cdot u_{1-\alpha/2}$)

```
plot(SResults,conf.int=T,xlab="survival time", ylab="survival probability")
```