# APTS - Survival Analysis
# Assessment

Ingrid Van Keilegom

August 2017

Consider the enclosed data set coming from the following book :

> Hosmer, D.W., Lemeshow, S. and May, S. (2008). *Applied Survival Analysis : Regression Modeling of Time to Event Data.* Second edition, Wiley, New York.

These data are coming from the *German Breast Cancer Study Group*, and the goal of the study is to analyse the effect of hormone therapy on the time to recurrence of breast cancer for women who have been treated for breast cancer in the past.

Consider the following subset of variables :

| Variable label | Variable type | Description |
|---|---|---|
| Time to recurrence | continuous | in days |
| Censoring indicator | categorical | 0 = censored |
| | | 1 = recurrence |
| Hormone therapy | categorical | 1 = yes, 2 = no |
| Age at diagnosis | continuous | in years |
| Menopausal status | categorical | 1 = postmenopausal |
| | | 2 = premenopausal |
| Number of progesterone receptors | continuous | 1 - 2380 |
| Number of estrogen receptors | continuous | 1 - 1144 |
| Tumor size | continuous | in mm |
| Number of nodes involved | continuous | 1 - 51 |
| Tumor grade | categorical | 1, 2, 3 |

**Part A**

We focus in Part A on the time to recurrence, the censoring indicator and the categorical variable 'hormone therapy'.

1. For the two levels of the variable 'hormone therapy', compute the survival distribution. Plot them on the same graph. What do the graphs suggest ?

2. For each level obtain an appropriate estimator and confidence interval for the 3 quartiles of the survival curves. Interpret the results.

3. Conduct a single test of differences between the survival curves. Justify your choice of test.

**Part B**

Consider now all variables in the data set.

1. We are interested in estimating the relative risk of survival for the two levels of the variable 'hormone therapy'. Assume that this relative risk, as well as any other relative risk, is constant over time, and independent of any other covariates.

    - Build an appropriate model for these data, using the model building procedures seen in class.
    - Find an estimator and confidence interval for the relative risk of hormone therapy under this model.

2. We now consider parametric regression models for the data.

    - Build an appropriate parametric model for the data, taking (for simplicity) the same variables as the ones selected for the previous model.
    - Give point and interval estimates of the coefficient of hormone therapy both in the AFT model and in the linear model representation.

Summarise your conclusions in a short report, which should be clear and well structured. Motivate well the procedures you used, and interpret the obtained results.