

About Dataset

Description

The Breast Cancer dataset contains patient records from a 1984-1989 trial conducted by the German Breast Cancer Study Group (GBSG) of 720 patients with node-positive breast cancer; it retains the 686 patients with complete data for the prognostic variables.

Format

A data set with 686 observations and 11 variables.

Pid: patient identifier

Age: age, years

Meno: menopausal status (0= premenopausal, 1= postmenopausal)

Size: tumor size, mm

Grade: tumor grade

Nodes: number of positive lymph nodes

Pgr: progesterone receptors (fmol/l)

Er: estrogen receptors (fmol/l)

Hormon: hormonal therapy, 0= no, 1= yes

Rfstime: recurrence-free survival time; days to first of recurrence, death or last follow-up

Status: 0= alive without recurrence, 1= recurrence or death

Details

These data sets are used in the paper by Royston and Altman. The Rotterdam data is used to create a fitted model, and the GBSG data is used for the validation of the model. The paper gives references for the data source.

References

Patrick Royston and Douglas Altman, External validation of a Cox prognostic model: principles and methods. BMC Medical Research Methodology 2013, 13:33