

a1_p3

R Markdown

A1 Problem 3a

The dataset contains data about breast cancer from the University of Wisconsin Hospitals.

The data is made available on July, 1992.

The data is from collected from 8 different groups from January 1998 - November 1991.

The dataset contains total 699 observations/instances from the 8 groups, and each group has different number of observations with group 1 having the most number of observations.

The dataset contains 10 variables/attributes and it is same throughout all 8 groups.

Variables/attributes from 2 to 10 has the value between 1-10, and some have same units and others having different and unique unit such as “Sample code number”.

Additionally, the dataset has a “Class” attribute with only 2 values, Benign & Malignant.

Finally, the dataset has total 16 missing attribute values.