**Data Description**

**Data Source**

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

<https://archive.ics.uci.edu/ml/citation_policy.html>

**Summary of Data**

The Breast Cancer Wisconsin (Diagnostic) Dataset is a publicly available dataset created by Dr. William H. Wolberg from the University of Wisconsin and was donated to the UCI Machine Learning Repository. It contains measurements of the characteristics of breast cancer tumors with no missing values. The 569 observations were gathered from a digitized image of a fine needle aspirate (FNA) of breast masses of female patients diagnosed with breast cancer. There are 12 variables in this dataset 2 of which are nominal and categorical and the remaining are quantitative. For the Hotelling T2 test this dataset will be divided into two groups malignant patients and benign patients. The description of each variable is given in the below Table 1.0.

**Variables Description**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Type** | **Unit of Measurement** | **Description** |
| **id** | Nominal | N/A | This is a unique identifier assigned to each patient. |
| **diagnosis** | Categorical | N/A | This is a qualitative variable that identifies the type of the tumor whether it is malignant (M) or benign (B). |
| **radius\_mean** | Quantitative | Millimeters | This feature measures the mean of distances from the center to points on the perimeter of the tumor. It represents the average size of the tumor. |
| **texture\_mean** | Quantitative | N/A | This feature measures the mean of gray-scale values in the image of the tumor. It represents the variation in the pixel intensities in the image. It is measured on a scale from 0 to 100. |
| **perimeter\_mean** | Quantitative | Millimeters | This feature measures the mean perimeter of the tumor, which is the length of its boundary. |
| **area\_mean** | Quantitative | Millimeters Squared | This feature measures the area of the tumor, which is the total number of pixels inside the boundary. It is measured in square mm. |
| **smoothness\_mean** | Quantitative | N/A | This feature measures the mean local variation in radius lengths of the tumor. It represents how much the radius of the tumor changes at different points along its boundary. It is measured on a scale from 0 to 1. |
| **compactness\_mean** | Quantitative | N/A | This feature measures the ratio of the perimeter squared to the area of the tumor, minus 1.0. It represents how tightly the tumor is packed together. It is measured on a scale from 0 to 1. |
| **concavity\_mean** | Quantitative | N/A | This feature measures the mean severity of concave portions of the contour of the tumor. It represents the amount of concavity in the boundary of the tumor. It is measured on a scale from 0 to 1. |
| **concavepoints\_mean** | Quantitative | N/A | This feature measures the number of concave portions of the contour of the tumor. It represents the number of inwardly curved sections in the boundary of the tumor. It is measured on a scale from 0 to 1. |
| **symmetry\_mean** | Quantitative | N/A | This feature measures how symmetric the tumor is. It represents how similar the left and right halves of the tumor are. It is measured on a scale from 0 to 1. |
| **fractaldimension\_mean** | Quantitative | N/A | This feature measures the mean "coastline approximation" of the tumor. It represents how much the boundary of the tumor is convoluted. It is measured on a scale from 0 to 1. |

Table 1.0 – Variable Description