REPORT WRITING

PST107

BODY

In this walkthrough report I am going to discuss the statical methods that were used in the dataset. The dataset that is used in this report is about Breast cancer. This dataset of breast cancer was obtained from the 2017 November update of the SEER program of the NCI, which provides information on population-based cancer statistics. The dataset involved female patients with infiltrating duct and lobular carcinoma breast cancer (SEER primary cites recode NOS histology codes 8522/3) diagnosed in 2006-2010. Patients with unknown tumour size, examined regional LNs, positive regional LNs, and patients whose survival months were less than 1 month were excluded; thus, 4024 patients were ultimately included(Kaggle.com).

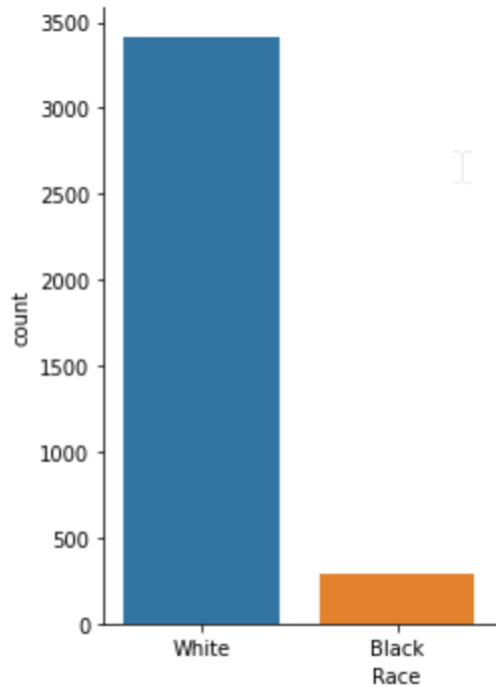
The dataset here is anyalyzed using analysis like univariate analysis ,bivariate analysis, distribution analysis, mean median,standard analayis etc.We are using collab research by google to show the statical analysis. All the data are showed in numbers and in a bar graph.

Introduction

Lets understand about the statical analysis on a large number of dataset. We are doing analysis on a large number of random samples. According to the law of large numbers, the mean of a population gets closer to its average proportional to the number of samples. But what makes this law of large number of important is the fact that after doing a random experiment a large number of times, the outcome of the experiment becomes stable.

Using statistical inferences, we can infer conclusions based on a sample taken from the population.There are many concepts and terms that we need to be aware of like population,sample,parameter and statistic.

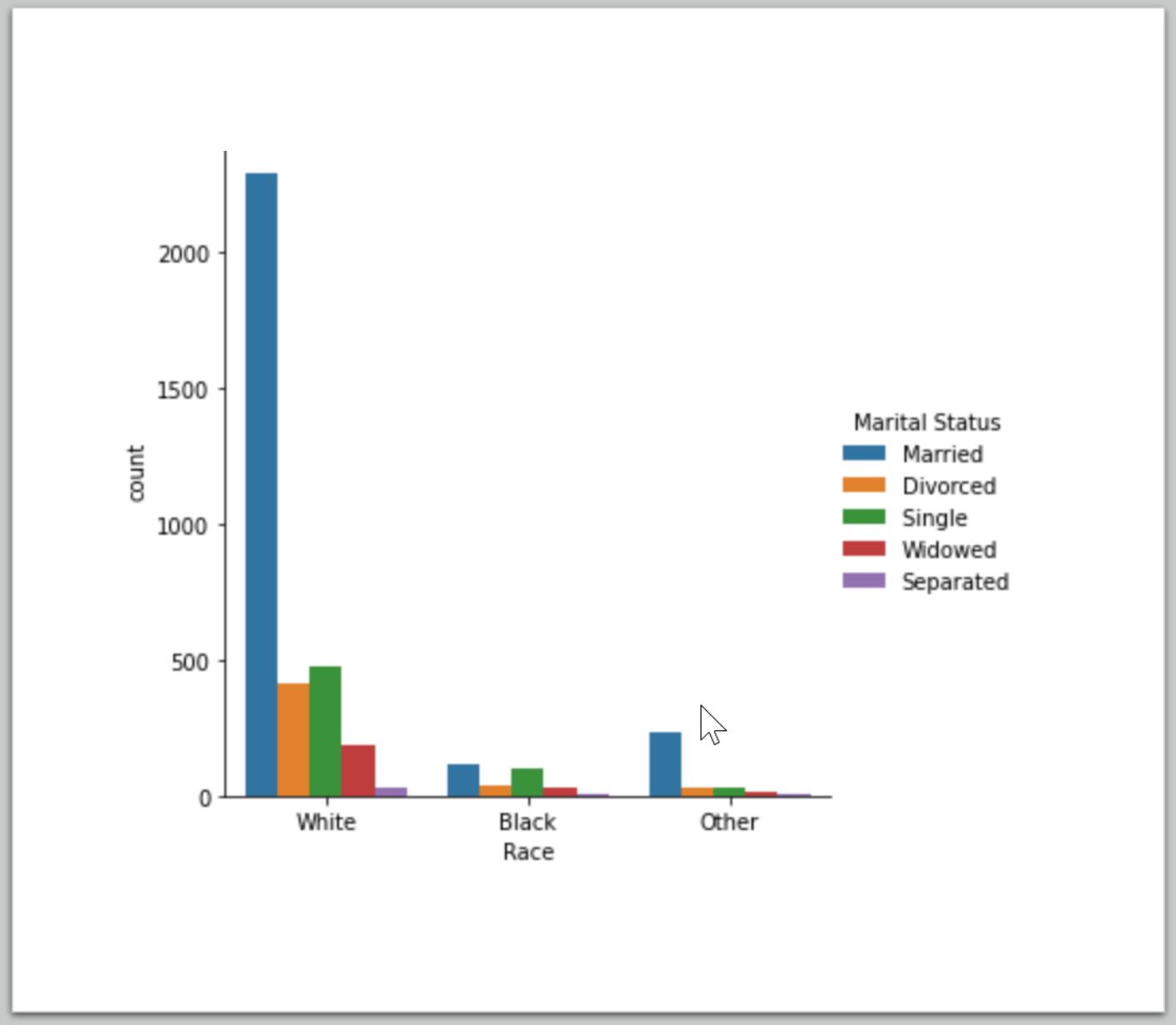
Lets talk about univariate analysis, Univariate summarize only one variable at a time.For examples:when you want to find the average household electrical bill value or you might want to study about a group od diabetic patients to find their weights and so on.



Here is a bar graph of an univariate analysis on the dataset based on the findings.It shows the ratio of white and black cloured skin based on their race.

Similarly, we have Bivariate analysis.The primary purpose of bivariate data is to compare the two sets of data to find a relationship between the two variables. If one variable influences the change in another variable ,then you have an independent and dependent variable.

In the figure there is a bar graph that shows Married women either white or black race has highest rate of breast cancer than other



We can also calculate mean,median and standard deviation from the dataset.

Conclusion

Stastical analysis can help companies and other scientist to analyze their data and costs,as well as recognize spendings trends.After identifying the information we can have insights about the data.