**Introduction:**

Prostate cancer is the second most common and fifth most aggressive neoplasm among men worldwide resulting in a large burden of both morbidity and mortality (Adam Barsouk et al.2020). Many prostate cancers are detected based on elevated plasmatic levels of prostate-specific antigen (psa > 4 ng/mL) which is normally expressed by prostate tissue(Prashanth Rawla et al. (2019). Despite the fluctuations in the reputation of PSA for its role in the management of prostate cancer, more studies are in demand to validate its significance along with other clinical parameters. Hence in the current study a detailed analysis was done on the data from 97 prostate cancer patients with sufficient clinical information used as dependent variables. The prostate cancer study was reported by Hastie et al., 2001 and the datasets were adopted from their reports to investigate the role of psa. Multiple regression model combined with descriptive statistical terms were used to determine the relation of psa with other independent variables. The results from the reconstructed regression model by eliminating the few outliers suggested that psa levels are associated with the cancer volume and also as predictive marker for prostate cancer.

**Data Sources**

The prostate cancer dataset was obtained from the previously published report by Hastie

et al., 2001. The dataset represents 97 prostate cancer patients who were about to undergo radical prostectomies. Each entry of the dataset has been given a unique ID and 8 other variables such as psa, cancerv, weight, age, hyperplasia, seminal, capsular and score.

**Statistical Analysis**

The dataset was retrieved in .csv format. The data was analyzed using the statistical

software R version 4.0.3 (2020-10-10). The aim of the study is to find the relationship

between response variable i.e. psa and explanatory variables such as cancerv, weight, age, hyperplasia, seminal, capsular and score using linear regression model. The data was found to be with no missing values and ready to use for the analysis.

Interactive shiny app is developed to visualize the linear relationship and distribution of independent variables.

**Shiny app:**

1. Created UI (user interface) to give access to users for selecting the independent variables.
2. App has dropdown box to select independent variables (cancerv, weight, age, hyperplasia, seminal, capsular and score).
3. Server returns the linear relationship and distribution of explanatory variables of dotplot for independent variable with response variable(psa = Prostate-Specific Antigen).
4. The app returns a dotplot of distribution of each type of explanatory variables that selected.

**Conclusion:** From the analysis we can see that the PSA (Prostate-Specific Antigen) levels are linearly dependent on cancer volume, age, seminal vesicle invasion, Gleason score and can help to predict the prostate-specific antigen levels that in turn helps to predict the prostate cancer

**References:**

1) Adam Barsouk, Epidemiology, Staging and Management of Prostate Cancer,

PMID: 32698438.

2) Prashanth Rawla, Epidemiology of Prostate Cancer, World J Oncol.

2019;10(2):63-89

3) Trevor Hastie, Robert Tibshirani, Jerome Friedman (1900-1993), The

Elements of Statistical Learning., 2nd ed., Springer.