Nonparametric Regression For Comparing The Mean Response of Breast Cancer Disease For Treatment With Application

By

Abeer Salem Abd-Rabo Omar

Assistant Lecturer, Department of Statistics
Faculty of Administrative Sciences, University of Aden

Supervised By

Prof.
Sahar Adel Raafat

Professor and Chairman, Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Benha University

Prof.
Farid Naguib Guirguis

Associate Professor of Statistics, Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Benha University

Abstract

There are two main sets of statistical models which are parametric and nonparametric regression models. The parametric regression models can be used when the explanatory variables are continuous or discrete. For the nonparametric regression models, it is used only when the explanatory variables are continuous and probability distribution are unknown.

The nonparametric regression has a great importance in many studies, where the nature of the relationship between the explanatory variables and the dependent variable is unknown. In other words, the probability distribution of these variables is unknown.

Therefore, the problem of response of breast cancer treatment is one of studies which have the nature of the relationship between the variables under study and response to treatment unknown.

Breast cancer is one of the most common types of cancers in Yemen, and the nature of the response to treatment is unknown. So, the study aimed to estimate the mean response.

The problem under study:

Given the prevalence of breast cancer in Arab societies in general, and Yemen in particular. For the seriousness of the disease which become a threat to a lot of women in Yemen which caused the casualties, material ,and due to the state mental experienced by from the disease or not?, and when there will be a response to treatment. The study aimed to estimate mean response to treatment.

Objectives of the study:

The aim of this study is to estimate the mean response of treatment by using nonparametric regression functions which are Nadaraya – Watson and KNN method. The researcher used visual basic software for application.

Hypotheses:

The study assumed:

1- The presence of significant age effects in influencing the response.

2- The presence of significant effects of the size of tumor in influencing the response.

3- Having significant effects to the amount of the drug in the treatment effect on the response.

4- Having significant effects for body surface area in influencing the response.

The limits of the study:

The study included patients with breast cancer in center Hope cancer in republic hospital in Aden since the establishment of the center in 2007 and until 2011.

The plan of the research:

The study included six chapters as follows:

Chapter One: The problem of breast cancer in Yemen and Litratre review:

This chapter introduces the problem of the spread of breast cancer in Yemen, its causes, symptoms and ways of treatment. As well as, the previous studies which used the nonparametric methods of estimating the regression function.

Chapter Two: Regression methods and cancer problem:

This chapter presents different regression methods used in the census to examine the relationship between the variable response (the dependent variable) and the explanatory variables. In addition to the advantages and disadvantages of each method, statistical properties of the estimators resulting from each method, and the reason of using the nonparametric way in the estimation of regression function which expresses the average response for cancer of various factors in flouncing the response.

Chapter Three: the methods used the nonparametric estimation:

This chapter addressed different methods used in the estimating the nonparametric regression function such as, kernel method, nearest neighborhood method, symmetric neighborhood, and spline regression method. We are concentrating on the kernel method by Ebanchincove formula which is considered one of the most common and efficient kernel function and most efficient. The different formulas used to estimate the kernel function such as Nadaraya – Watson, Gasser – Muler and Priestly-Chao. We concentrate on Nadaraya-Watson formula because it is the used method in this study.

This formula can be used only if the used model in the study is a random design model, where all the explanatory variables are a random variables change form one individual to another. In other words, explanatory variable is seen as a random variable change with the change of the samples, as the explanatory variable used in our study. Where we find that each of the age and size of the tumor, the amount of the property and body surface area change from patient to another.

Chapter Four: The problem of choosing smoothing parameter in the estimation of nonparametric regression function:

This chapter includes the different ways used to select the smoothing parameter and when to use each way.

Chapter Five: Application:

In this chapter, we apply the kernel function for Ebanchicnove and Nadarya – Watson formula to estimate mean response of breast cancer disease of the factors in the response. In addition to finding out which factor has the biggest impact in the response.

Chapter Six: Conclusion and recommendations:

We summarize in this chapter our findings of the study and the recommendations with discussion of the results that we reached by statistical methods. In addition to the results of the studies and medical research were consistent with the finding of medicine.