Chapter 9
Pairing Regression Model and R Programming

ABSTRACT

This chapter provides an overview of the popularity of linear regression in medical research. In particular, the implementation of R programming in executing regression analysis was given emphasis using RStudio as the integrated development environment. Importing dataset and consequent exploratory data analysis were also presented in this chapter. In addition, the use of the most popular visualization package in R was also demonstrated. At the very least, this chapter encapsulates the whole process of regression analysis from data ingestion, visualization, model building, and model testing. Software and packages used in this chapter were all free and open-source, which ensures that the implemented codes will be reproduceable.

INTRODUCTION

Regression analysis is an essential statistical method for the evaluation of medical data. It facilitates the recognition and categorization of associations among and between numerous factors. It also supports the discovery of prognostically significant risk factors and the estimation of risk scores for a particular prognosis (Schneider et al., 2010). Depending on the purpose or goal of the research study, the regression model can serve as a 'classification' or as a 'prediction' technique. This book chapter provides medical practitioners with the basic skills needed to critically read and interpret research output

using linear regression methods. This Chapter explains the application of statistical models to a wide variety of medical investigative studies and clinical trials. Regression methods are used to appropriately answer the key design questions posed and take due account of any effects of potentially influencing co-variables. It begins with a revision of basic statistical concepts, followed by implementation of these methods in R Programming. The various methods of modelling are covered in a non-technical manner so that the principles can be easily applied in everyday practice. Data drawn from published studies are used to exemplify statistical concepts throughout the discussion. This book chapter is designed for clinicians, public health and environmental health professionals, para-medical research professionals, laboratory-based researchers and students (Tai & Machin, 2013).

In this chapter we will discuss simple linear regression and how we can easily calculate this technique using R programming. A regression analysis allows you to predict a variable (dependent variable, Y) based on one or more variables (independent variable, X). Depending on whether there are one or more independent variables, a distinction is made between simple and multiple linear regression analysis (*Linear Regression - DATAtab*, n.d.).

For simplicity, let us consider an X variable and a Y variable. The independent variable is on the x-axis and the dependent variable is on the y-axis. The dataset that we will use here is a sample dataset of laboratory confirmed SARS -CoV2 patients with several variables which included if the patient is being admitted to an intensive care unit or not (ICU), their running hospitalization in days (HOS) and the level of their lactate dehydrogenase (LDH). The relationship between the HOS and LDH is expressed in a form of a scatter plot with each patient ICU status being tagged. The R code to generate the plot is given as well.

Basic scatter plot

```
df <- read.csv("validation_set.csv")[,-1]
ggplot(df, aes(y=HOS, x=LDH, color=ICU)) + geom_point()+
 labs(title = "Relationship between the Number of
Hospitalization and LDH", subtitle ="among ICU and Non-ICU
SARS-CoV2 Positive Patients")+
 labs(y="Number of Hospitalization (Days)")+
 labs(x="Level of LDH")+
 labs(color = "ICU Patient or Not?")
```

7 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/pairing-regression-model-and-r-programming/314293

Related Content

The Impact and Implication of Artificial Intelligence on Thematic Healthcare and Quality of Life
Bongs Lainjoand Hanan Tmouche (2023). *International Journal of Applied Research on Public Health Management (pp. 1-17).*
www.irma-international.org/article/impact-implication-artificial-intelligence-thematic/318140

Intentional Use of Digital Technology in Graduate Epidemiology Education
Charlotte Baker (2021). *Research Anthology on Public Health Services, Policies, and Education (pp. 244-256).*
www.irma-international.org/chapter/intentional-use-of-digital-technology-in-graduate-epidemiology-education/281979

Digital Transformation Challenges for the Development of Quality Electronic Medical Records in Greece
Dimitrios G. Katehakisand Angelina Kouroubali (2022). *Quality of Healthcare in the Aftermath of the COVID-19 Pandemic (pp. 112-134).*
www.irma-international.org/chapter/digital-transformation-challenges-for-the-development-of-quality-electronic-medical-records-in-greece/292424

Profile: Plague
(2022). *Historical and Epidemiological Analyses on the Impact of Infectious Disease on Society (pp. 215-235).*
www.irma-international.org/chapter/profile/303302

Seeking Solutions to the Challenges in Health Services in Developing Nations: Primary Health Care Revisited
Jessica McCormack, Patrick Rawstorneand Mohamud Sheikh (2015). *Transforming Public Health in Developing Nations (pp. 1-51).*
www.irma-international.org/chapter/seeking-solutions-to-the-challenges-in-health-services-in-developing-nations/133674