INDUSTRIAL TRAINING REPORT
AT
UNIVERSITI TUN HUSSEIN ONN MALAYSIA
(UTHM) BATU PAHAT, JOHOR
BY
SITI AFIQAH BINTI MUHAMAD JAMIL
(2012220758)

INDUSTRIAL TRAINING REPORT
AT
UNIVERSITI TUN HUSSEIN ONN MALAYSIA (UTHM)
86400, PARIT RAJA, BATU PAHAT, JOHOR

PREPARED BY:

SITI AFIQAH BINTI MUHAMAD JAMIL

(2012220758)

REPORT SUBMITTED TO:

FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES

UNIVERSITI TEKNOLOGI MARA KELANTAN

AS PART OF REQUIREMENT

FOR

BACHELOR OF SCIENCE (HONS) (STATISTICS)

CS221

JANUARY 2016

MODELLING COUNT DATA:

AN APPLICATION TO A BREAST CANCER DATA IN MALAYSIA

ACKNOWLEDGEMENT

Bismillahirrahmanirrahim. Thanks to Allah SWT, for his wisdom and blessing giving me the opportunity to complete this research project which is about analysis of modelling counts data: an application to a breast cancer data. This research project was analysed based on the count data of breast cancer patients yearly in Malaysia.

First and foremost, I would like to acknowledge YM Tengku Mardhiah Tengku Jalal from Universiti Teknologi Mara (UiTM) and Dr. Mohd Asrul Affendi Bin Abdullah from Universiti Tun Hussein Onn Malaysia (UTHM) who had supervised and guided me a lot by giving me continuous commitment through this research project, excellent comment, and constructive criticism all the way to the end of this research project. I would like to extend my appreciation to the coordinator of the industrial training, Madam Siti Nurhafizah Binti Mohd Shafie for her cooperation.

Deepest thanks to parents and family members for their moral and financial support from the beginning till the end of this research project. Besides, thanks to all friends for their encouragement, suggestion, and support in completing this research project. Finally, I also would like to extend thanks to all of the people who contribute either direct or indirect in order to complete this research project.

ABSTRACT

Cancer is a disease of cells and virtually affects everyone as it is one of the most common factors that influenced the deaths of people in Malaysia. Breast cancer patient who were diagnosed, varies among institution in Malaysia. A study was conduct to review the deaths of breast cancer based on the breast cancer data in Malaysia. The breast cancer data contains 62 numbers of observations according to the duration of data collected from years 1980 until 2010. The data was collected yearly in Malaysia by the Institute for Health Metrics and Evaluation (IMHE) on the systematic analysis of collecting the data. There were three variables which comprised the number of deaths of breast cancer, the number of cases of breast cancer and the age group. The dependent variable was the number of deaths of breast cancer, while the independent variable includes the number of cases of breast cancer and the age group which comprised women aged 15-49 and women aged 50-79. The first objective of the study was to determine the trend pattern of the number of breast cancer deaths among women aged 15-49 and women aged 50-79. The second objective was to identify the best count data model that fit the number of breast cancer deaths among women aged 15-49 and women aged 50-79. The last objective was to identify factors that affecting the deaths of breast cancer from the cases of breast cancer and the women aged 15-49 and women aged 50-79. The breast cancer data was analysed using the EXCEL, SPSS software and SAS procedure. The finding indicates that: (1) the trend pattern of women aged 15-49 was lower compared to the pattern of women aged 50-79. (2) The best fit model to be used when modelling counts data was the Negative Binomial regression analysis model. (3) The factor that show high influences on analysing the number of deaths of breast cancer was number of cases.

Keywords: *Breast Cancer, Time-plot, Negative Binomial regression, SAS and SPSS.*