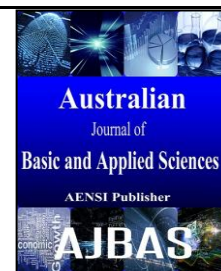




ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

The Effective Model of Linear Regressions for Colorectal Cancer Stages in General Hospital: A Case Study in Kuala Lumpur.

¹Muhammad Ammar Bin Shafi, ²Mohd Saifullah Bin Rusiman, ³Siti Noor Asyikin Binti Mohd Razali, ⁴Mahathir Bin Mohamad

¹Faculty of Science, Technology and Human Development, UTHM, 86400 Parit Raja, Batu Pahat, Johor. Malaysia

²Faculty of Science, Technology and Human Development, UTHM, 86400 Parit Raja, Batu Pahat, Johor. Malaysia

³Faculty of Science, Technology and Human Development, UTHM, 86400 Parit Raja, Batu Pahat, Johor. Malaysia

⁴Faculty of Science, Technology and Human Development, UTHM, 86400 Parit Raja, Batu Pahat, Johor. Malaysia

ARTICLE INFO

Article history:

Received 12 February 2015

Accepted 1 March 2015

Available online 28 March 2015

Keywords:

Fuzzy Regression, Multiple Linear Regression, Colorectal Cancer.

ABSTRACT

Regression analysis has become popularity among several sources of researcher and standard tolls in analyzing data. This structure was represented two commonly statistical models such as multiple linear regression and extended fuzzy correlation and regression analysis (Ni, 2005). Colorectal cancer was applied and case in Malaysia. The quality of life in CRC patients to detect the early CRC stage is still very poor, mainly ad-hoc and not implemented as a national wide programme. This study aims to determine the best model to measuring the mortality rate of patients by CRC stages at hospital using mean square error compared. Secondary data of 180 patients have colorectal cancer and received treatment in hospital recorded by nurses and doctors. Based on the results of regression, extended fuzzy correlation and regression analysis (Ni, 2005) is the best model to measuring the mortality rate of patients who have colorectal cancer after received treatment in hospital.

© 2015 AENSI Publisher All rights reserved.

To Cite This Article: Muhammad Ammar Bin Shafi, Mohd Saifullah Bin Rusiman, Siti Noor Asyikin Binti Mohd Razali, Mahathir Bin Mohamad., The Effective Model of Linear Regressions for Colorectal Cancer Stages in General Hospital: A Case Study in Kuala Lumpur. *Aust. J. Basic & Appl. Sci.*, 9(8): 7-10, 2015

INTRODUCTION

Regression analysis has become one of the standard tools in analysing the data. Its popularity comes from several sources. The mathematical equation gained from its analysis could explain relationship between the dependent and independent variables. It provides much explanatory power, especially due to its multivariate nature. It is widely available in computer packages and easy to interpret. It has been widely used in applied sciences, economic, engineering, computer, social sciences and other fields (Agresti, A., 1996).

Nonlinear modeling is of interest to many researchers in modeling statistics, rather than linear modeling. The functional form obtained should be approximately near to the real data. If the functional form is far away from the real data, its mean the estimation will be inconsistency, bias and so on. However, other difficulties may arise with a non-linearity approach (Rousseeuw, et. al, 2003). Nowadays, there are many modeling resulting from the regression analysis such as multiple regression, quadratic regression, cubic regression, logit model, probit model, exponential model, growth model, neural network regression and fuzzy regression.

Colorectal cancer is cancer of the colon and rectum. The colon and rectum are two parts in human body to play an important role to digest food and past waste. Colorectal cancer is one of the commonest diseases malignancies in the world (Malaysian Oncological Society, 2007). According (WHO,2012), Colorectal cancer is the fourth leading died of the cancer person. This cancer is rising in many countries especially in the Asian Region in Malaysia.

Colorectal cancer is the third leading cause of cancer deaths in Malaysia. Lung cancer is the highest rate cause of cancer deaths by 17.93. Followed by breast cancers is 15.83 and the third one is colorectal cancers by 13.10 rates of cancer deaths (WHO data, 2010). Data from the Ministry of Health of Malaysia out of statement that increasing in colorectal cancer admission rates 8.1% in 1987 to 11.9% in 1995 (MOH, 1995). Recent studies have shown an increasing colorectal cancer (CRC) in Asian population (Medical J Malaysia, 2008). According to the Second Paper of the National Cancer Registry, 2003, colorectal cancer got for 14.2% of male cancer and 10.1% of female cancer in Malaysia. It can be conclude commonest cancer among men and otherwise third most common among women

Corresponding Author: Muhammad Ammar Bin Shafi, Faculty of Science, Technology and Human Development, UTHM, 86400 Parit Raja, Batu Pahat, Johor. Malaysia,
Tel: 0195182167; E-mail: ammar26121991@gmail.com.

respectively. CRC including four stages an early stage until final stage. The stage I it is means cancer is confined to the inner lining of the colon or rectum, stage II means cancer spreads through the wall of the colon or rectum, stage III means cancer spreads to nearby lymph nodes and lastly stage IV means cancer spreads to distant parts of the body, such as the liver or lungs (MOS, 2007)

Method And Analysis:

The study was a linear model. The sample consisted of 180 patients in range 21 until 90 years old who are having symptoms and colon cancer by any of four stages who are receive the treatment at general hospital around Kuala Lumpur. There are including male and female of patients and in various ethnics. The data are from hospital in Kuala Lumpur. Furthermore, the analysis used multiple linear regression and extended fuzzy correlation and regression Analysis (Ni, 2005). The study has been discussed with mean square error, root mean square error and fitted of model.

Data Analysis:

The statistical software SPSS 20, excel 2010 and matlab 2008 was used to analyse the data. Exploratory data analysis used to explore the behaviour of the data. Some exploration on the demographic characteristic applied. For categorical variables, data presented in frequency and percentage distribution while for continuous variable,

descriptive statistics measured by mean or average, standard deviation and correlation of variables were applied.

Results:

The majority's demographic of patients after receiving treatment most colorectal cancer of patients in General Hospital Kuala Lumpur was Malay and male and most of them were in average age of 61 years old and cancer colon for most patients in colon side.

Multiple Linear Regression:

Firstly, assumptions of multiple linear regression should be fulfill before using the data to analyzed. This study applied three assumptions which are constant variance, normality and multicollinearity. All the assumptions were satisfied and when these assumptions are met the results may be trustworthy.

Multiple linear regression is one common model used by statistician especially in medical health. This model has been used to study and analyse twenty five predictor variables and further detecting colorectal cancer by stage. After analyse done, just only eleven predictor variables were significance for colorectal cancer. All variables were significance are age at diagnosis, icd10 site, TNM staging, family history, Crohn's disease, history of cancer, gastric, ovarian, Intestinal obstruction, anemia and abdominal. There are significant variables had been measured by p – value < 0.05 .

Table 1: Summary of parameter estimate multiple linear regression.

independent variables	Beta (β)	Sig. Value
(Constant)	76.056	*0.000
Gender	1.977	0.286
Age at Diagnosis (years)	-0.421	*0.000
Ethnic Group	2.231	0.114
ICD 10 Site	3.459	*0.027
TNM Staging	0.961	*0.040
Family History	-16.738	*0.000
Diabetes Mellitus	0.832	0.679
Crohn's Disease	5.035	*0.012
Ulcerative colitis	1.508	0.432
polyp	-1.577	0.396
History of cancer(s)	5.557	*0.007
Endometrial	-0.869	0.648
Gastric	-6.517	*0.001
Small bowel	-3.33	0.087
Hepatobiliary	-1.047	0.589
Urinary tract	2.597	0.176
Ovarian	12.865	*0.000
Other cancer	3.248	0.084
Intestinal Obstruction	-4.35	*0.022
Colorectal	1.227	0.511
weight_loss	3.063	0.1
Diarrhoe	0.75	0.695
Anemia	-7.943	*0.003
blood_stool	-2.158	0.233
abdominal	-3.994	*0.038

*Significant at 0.05

All the significance variables influence the colorectal cancer effects. The estimated multiple linear regression model for symptoms and factors of

colorectal cancer model is as follow:

$$\hat{Y} = 76.056 + -0.421 \text{ age} + 3.459 \text{ icd10} + 0.961 \text{ TNM Staging} - 16.738 \text{ family history} + 5.035$$

Crohn's disease + 5.557 history of cancer – 6.517 gastric + 12.865 ovarian – 4.350 intestinal obstruction – 7.943 anemia – 3.994 abdominal

Other than that, ANOVA analysis had been done. The results shown the mean square error term is 129.558 and RMSE is 11.3826. The P-value for

the F test statistic is less than 0.05, providing strong evidence against the null hypothesis. The squared multiple correlation $R^2 = SSR/SST = 21396.590/41348.550 = 0.517$, indicating that 51.7% of the variability in the tumour size variable is explained by the twenty five variables.

Table 2: ANOVA for multiple linear regression.

Source	Sum of Squares	df	Mean Square	F Value	P Value
Regression	21396.59	25	855.864	6.606	0
Residual	19951.96	154	129.558		
Total	41348.55	179			
RMSE			11.3826		

Extended By Benchmarking Models Under Fuzziness (Chung, 2012):

This model has been approached by William Chung on 2012. Extended benchmarking models under fuzzy environment can be developed by means of fuzzy linear regression among variables. A

normalization of coefficient is applied to this model to get the better results. This model has been analysis using Matlab 2008.

The new model of normalization under fuzziness has been carrying out. This model will measure the parameter for each variable.

Table 3: summary of mean square error for model.

MSE	110.8442
RMSE	10.5283

Table 3 shows there are twenty five independent variables has been analysed and a new coefficient produced and mean square error for the benchmarking under fuzziness model colorectal cancer was 110.8442 respectively. Otherwise, RMSE value was 10.5293. The estimated benchmarking under fuzziness model for colorectal cancer of patients is as follow:

$$\hat{Y} = 64.2147 + 1.9772 \text{ gender} - 0.4213 \text{ age} + 2.2313 \text{ ethnic} + 3.4595 \text{ icd10} + 0.9615 \text{ TNM Staging} + 16.7380 \text{ family history} - 0.8324 \text{ diabetes mellitus} - 5.0349 \text{ Crohn's disease} - 1.5080 \text{ ulcerative colitis} + 1.5765 \text{ polyp} - 5.5572 \text{ history of cancer} + 0.8686 \text{ endometrial} + 6.5167 \text{ gastric} + 3.3303 \text{ small bowel} + 1.0469 \text{ hepatobiliary} - 2.5967 \text{ urinary tract} - 12.8651 \text{ ovarian} - 3.2476 \text{ other cancer} + 4.3497 \text{ intestinal obstruction} - 1.2271 \text{ colorectal} - 3.0630 \text{ weight loss} - 0.7497 \text{ diarrhoea} + 7.9433 \text{ blood stool} + 2.1583 \text{ anemia} + 3.9941 \text{ abdominal}$$

Summary Of Results:

In summary, the present study proved Extended Fuzzy Correlation and Regression Analysis (Ni, 2005) model is the best model to measure the predicting staging of colorectal cancer faced by patients in General Hospital around Kuala Lumpur. Furthermore, analyses using two methods which are mean square error and root mean square error to measure comparing between Multiple Linear Regression and Extended Fuzzy Correlation and Regression Analysis (Ni, 2005) model and the results all two methods are in Extended Fuzzy Correlation and Regression Analysis (Ni, 2005) side. It can be concluded, Extended Fuzzy Correlation and Regression Analysis (Ni, 2005) is better model than multiple linear regression and it clearly shows that this data is well fitted with Extended Fuzzy

Correlation and Regression Analysis (Ni, 2005) model.

Limitations Of The Study:

This study will only focus on 180 patients who have a Colorectal Risk Cancer symptom at General Hospital in Kuala Lumpur for the period year 2012. The analysis cannot be replicated to another hospital or clinic due to methods and ways of colorectal cancer patients stated. The data was private by the general hospital in Kuala Lumpur.

Conclusions:

This study was aim to describe the staging of colorectal cancer of the patients disease by length of colorectal cancer mm and to reduce the mortality of patients died. It was found that most patients are in stage 2 of colorectal cancer in General Hospital Kuala Lumpur. Furthermore, most colorectal cancer of patients in General Hospital Kuala Lumpur was Malay and male and most of them were in average age of 61 years old and cancer colon for most patients in colon side or area.

There are 25 determinants of factors and causes of colorectal cancer which are gender, ethnic, age, icd10, TNM staging, family history of colon cancer, Diabetes Mellitus, Crohn's Disease, Ulcerative colitis, polyp, History of cancer(s), Endometrial, Gastric, Small bowel, Hepatobiliary, Urinary tract, Ovarian, Other cancer, Intestinal Obstruction, Colorectal, weight loss, diarrhoea, anaemia, blood stool, abdominal pain. In order to investigate predicting staging of colorectal cancer of significant determinants, this study was performed four applied of linear regression. From these four models, the significant determinants are listed as age, icd10, TNM staging, family history of colon cancer, Crohn's

Disease, History of cancer(s), Gastric, Ovarian, Intestinal Obstruction, anaemia, abdominal pain.

Moreover, this study also interested to know which model fit the data well based on two methods of measuring which are mean square error and root mean square error. Based on two comparing methods, Extended Fuzzy Correlation and Regression Analysis (Ni, 2005) is to found out to be well fitted model in predicting staging of colorectal cancer patients in General Hospital around Kuala Lumpur.

ACKNOWLEDGEMENT

Praise is to Allah S.W.T for the blessing and giving me the strength and health to complete this project paper. First and foremost, I would like to thanks to my kindness supervisor, Dr. Mohd Saifullah Bin Rusiman for his invaluable advice, comments and constant guidance during the progress of my project paper. I would like to express my special thanks to grant vot E020 that have gave to me through my supervisor is graduate research assistant. With this grant i can be supported and be encouraged to do this paper. No such valuable words than "Thank You" to appreciate the eagerness and kindness. More than words could express on paper or could be spoken in words. Hopefully this research project would be guidance in future. Thank You.

REFERENCES

- Agresti, A., 1996. An Introduction to Categorical Data Analysis. *New York : John Wiley & Sons, Inc.*
- Aliev, R.A., B. Fazlollahi and R. Vahidov, 2002. Genetic algorithms-based fuzzy regression analysis. *Journal of Soft Computing*, 6: 470-475. Springer-Verlag.
- Benne, D.H., J.D. Hardcastle, 1996. Early Diagnosis and Screening. Colorectal Cancer. Ed Williams NS. *Churchill Livingstone*; 21-37.
- Bezdek, J.C., 1974. Cluster validity with fuzzy set. *Journal Cybernetic*, 3: 58-72.
- Bezdek, J.C., 1981. Pattern recognition with fuzzy objective function algorithms. *USA: Kluwer Academic Publishers*
- Bottomley, A., 2002. The cancer patient and quality of life. *The Oncologist*, 7: 120-5.
- Cancer Research UK. Cancer Stats Key Facts, 2010. <http://www.cancercenter.com/colon-cancer/quality-life.cfm> (25 Oct. 2011)
- Center, M.M., A. Jemal, E. Ward, 2009. International trends in colorectal cancer incidence rates. *Cancer Epidemiol Biomarkers*, 19: 1688-94.
- Dubois, D.J. and P. Henry, 1980. Fuzzy set and systems : Theory and applications. *Academic Press Inc.*
- Dunn, J., 1974. A fuzzy relative of the ISODATA process and its use in detecting compact, well separated cluster. *Cybernetics*, 3(3): 32-57.
- Hyodo, I., Suzuki H, Takahashi K, *et al.* 2010. Present status and perspective of colorectal cancer in Asia: Colorectal Cancer Working Group Report in 30th Asia-Pacific Cancer Conference. *Jpn J ClinOncol*, 40: 38-43.
- Jemal, A., R. Siegel, E. Ward, 2008. Cancer Statistics. *C.A: A Cancer J Clin*, 58: 71-96.
- Kong, C.K., A.C. Roslani, C.W. Law, 2010. Impact of socioeconomic Class on Colorectal Cancer Patient Outcomes in Kuala Lumpur and Kuching, Malaysia. *APJCP*, 11: 969-74.
- Lazim, A., Z. Nadia, 2012. Matrix Driven Multivariate Fuzzy Linear Regression Model In Car Sales. *Journal of Applied Sciences*.
- Natrah, M.S., W.P. SharifaEzat, M.A. Syed, A.M. Mohd Rizal, S. Saperi, 2012. Quality of Life in Malaysian Colorectal Cancer Patients : A Preliminary Result. *APJCP*, 13: 1-6.
- National Cancer Registry, 2006. Ministry of Health Malaysia. 2006. Malaysian Cancer Statistics – Data And Figure Peninsular Malaysia.
- The First Annual Report of the National Cancer Patient Registry-Colorectal Cancer 2007-2008. 2010.
- Nwoye, E., L.C. Khor, S.S. Dlay, W.L. Woo, 2008. Spectral And Statistical Features In Fuzzy Neural Expert Machine For Colorectal Adenomas And Adenocarcinoma Classification.
- Pedryoz, W., 2005. Knowledge-based clustering: From data to information granules. *John Wiley & Sons*.
- Shchneider, J., N. Bitterlich, G. Schulze, 2005. Improved Sensitivity In The Diagnosis Of Gastro-Intestinal Tumors By Fuzzy Logic-Based Tumor Marker Profiles Including The Tumor M2-PK. *Anticancer Research*, 25: 1507-1516.
- Mahmoud Taheri, S., 2003. Trends In Fuzzy Statistics. *Austrian Journal Of Statistics*, 32(3): 239-257.
- Tanaka, H., 1987. Fuzzy Data Analysis By Possibilities Linear Models. *Fuzzy Set Syst*, 24: 363-375.
- Zadeh, L.A., 1965. Fuzzy sets. *Inform. Control*, 8: 338-358