UNIVERSITI MALAYSIA TERENGGANU

PROSIDING SIMPOSIUM KEBANGSAAN SAINS MATEMATIK KE 16

JILID 2 : MATEMATIK TULEN, STATISTIK DAN MATEMATIK PENDIDIKAN

"Sains Matematik Memacu Keunggulan Pemikiran"

SKSM 16

MIMOS

PERSATUAN SAINS MATEMATIK MALAYSIA
PERSAMA

3 - 5 Jun 2008
Hotel Renaissance, Kota Bharu

Dasar Penerbitan UMT

Perpustakaan Negara Malaysia Data Pengkatalogan-dalam-Penerbitan

Simposium Kebangsaan Sains Matematik (ke-16:2008:Kota Bharu)
Sains matematik memacu keunggulan pemikiran: prosiding Simposium Kebangsaan Sains Matematik ke 16, 3-5 Jun 2008, Kota Bharu. Jilid 2, Matematik tulen, statistic dan matematik pendidikan/pengarang Adem Kilciman... [et.al].
ISBN 978-983-2888-92-5
1. Mathematics—congresses. 2. Science-- congress.
I. Kilicman, Adem. II. Universiti Malaysia Terengganu. III. Judul.
510

HAKSANGKAL

Prosiding ini disediakan sebagai himpunan kertas kerja yang dibentangkan di Simposium Kebangsaan Sains Matematik ke 16 yang dianjurkan oleh Jabatan Matematik, Fakulti Sains dan Teknologi, Universiti Malaysia Terengganu dan Persatuan Sains Matematik Malaysia (PERSAMA). Jabatan Matematik UMT dan PERSAMA tidak bertanggungjawab ke atas ketepatan dan kesempurnaan artikel dalam prosiding ini. Hakcipta bagi setiap sumbangan artikel adalah pada pengarang masing-masing. Keputusan dan interpretasi yang dinyatakan adalah hasil nukilan pengarang sendiri dan tidak semestinya mencerminkan pendapat Jabatan Matematik UMT dan PERSAMA.

DISCLAIMER

This proceeding was prepared as a collection of papers presented at the 16th National Symposium of Mathematical Sciences organized by Department of Mathematics, Faculty of Science and Technology, Universiti Malaysia Terengganu and Malaysian Mathematical Society (PERSAMA). Department of Mathematics UMT and PERSAMA do not hold any responsibility for the accuracy and completeness of the articles in the proceeding. Authors retain the copyright of their individual contributions. The results and interpretations expressed are those of the authors alone and do not necessarily state or reflect those of Department of Mathematics UMT or PERSAMA.

KANDUNGAN

MATEMATIK TULEN

SIMULATION ON GROUP DELETED GENERALIZED POTENTIALS FOR DIAGNOSTICS OF CENSORED SURVIVAL REGRESSION

[1]Abdul Kudus, [2]Noor Akma Ibrahim, [3]Isa Daud

[1]Department of Statistics, Bandung Islamic University, Indonesia
Institut Penyelidikan Matematik (INSPEM), Universiti Putra Malaysia, Malaysia
akudus69@yahoo.com

[2]Institut Penyelidikan Matematik (INSPEM), Universiti Putra Malaysia, Malaysia
Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, Malaysia
nakma@putra.upm.edu.my

[3]Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, Malaysia
drisa@science.upm.edu.my

Abstract :

The identification of multiple high leverage points (HLP) can not be successfully carried out by using a method developed for handling single HLP. The situation getting worse when handling Weibull distributed censored response data. We proposed method for identifying multiple HLPs. The proposed method is developed based on the similar problem for handling multiple HLPs in multiple linear regression setting. Simulation study is conducted for assessing its performance across a wide range of scenarios. We found that the proposed method had good capability to identify multiple HLP. As a general rule, the proposed method perform better in higher covariate space dimension, larger number of outlying regressor variables, larger outlying leverage distance, larger magnitude of unusualness in response-space, larger number of multiple points clouds and higher percentage of censoring.

Introduction

The statistical analysis of survival time data has become a topic of considerable interest to statisticians and workers in areas such as engineering, medicine and biological sciences. Such data may show structural complexity, but it is the presence of censoring which sets its analysis apart from traditional statistical techniques.

The Cox proportional hazards (PH) model is usually applied to analyze the relationship between survival time and explanatory variables. Kudus and Ibrahim (2005) extended the model for competing risks survival data. Whereas, Kudus *et al.* (2005) proposed the regression tree method based on Cox proportional hazards for analyzing competing risks survival data. The regression tree is further extended by using proportional hazards model for subdistribution and applied on breast cancer study (Ibrahim *et al.*, 2008).

Weibull regression model is a special case of Cox proportional hazards. Kalbfleisch and Prentice (2002) showed that Weibull regression model is not only having form of proportional hazards but also log-linear model. The problem arose when the model is fitted to data contained unusual observations, such as high leverage points (HLP). This problem can cause invalidity of the use of standard inferential procedure. It is thus important for the data analyst to be able to identify such observations.

If the dataset contain more than one HLP, which is likely to be the case in most data sets, the problem of identifying such observations become more difficult. There is evidence (Rousseeuw and Leroy, 1987) that the single HLP detection techniques have been proved to be ineffective to detect potential observations in the presence of multiple HLP in linear regression problem. Wisnowski *et al.* (2001) stated that many standard least-squares regression diagnostics quantities and plots have been shown to fail in the presence of multiple outliers (including multiple HLP), particularly if the observations are clustered in an outlying cloud. Imon (1996) introduced generalized potentials (p_{ii}*) and using it for identifying multiple HLP in linear regression setting.

Since there is no existing method for identifying multiple HLP in Weibull regression model, then it is thus important to develop a method for such problem. The proposed method must have a closed relation to the existing method for linear regression problem due to the advantage of log-linear form of Weibull regression model. Hence, it would be a generalization from the existing method. We assesed its performance through a series of Monte Carlo simulation in various multiple HLP configurations likely to be encountered in practice.

In Section 2 we review Weibull regression modeling through proportional hazards or log-linear models and their parameter estimations by means of maximum likelihood method. Section 3 presented single and multiple HLP identifications for Weibull regression model. Extensive Monte Carlo simulation to evaluate multiple HLP identification method is designed and conducted in Section 4 and last section presents a conclusion of the paper.

The Weibull Regression Modeling

The PH model

Consider survival time $T > 0$, sometime censored, and suppose that a vector of basic covariates $\underline{x}' = (x_1, x_2, \ldots)$ is available on each individual, their measurements having been taken at or before time 0. Aspect of $\underline{x}$ are expected to be predictive of subsequent survival time. The principal problem is that of modeling and determining the relationship between T and covariates x.

Recall that Weibull distribution with scale parameter λ and shape parameter γ has hazard function

$$h(t) = \gamma\lambda t^{\gamma-1} \tag{1}$$

To include the covariate vector $\underline{x}_i$ of the i^{th} individual, the hazard for a given $\underline{x}_i$ can be expressed as

$$\begin{aligned} h(t \mid \underline{x}_i) &= h_0(t)\cdot \exp\left(\sum \beta_j x_{ji}\right) \\ &= \gamma\lambda \exp\left(\sum \beta_j x_{ji}\right) t^{\gamma-1} \end{aligned} \tag{2}$$

Now, Weibull distribution has scale parameter $\tilde{\lambda} = \lambda \exp\left(\sum \beta_j x_{ji}\right)$ and shape parameter γ. The survivor function turns out to be

$$S(t \mid \underline{x}_i) = \exp\left[-\exp\left(\sum \beta_j x_{ji}\right)\lambda t^{\gamma}\right] \tag{3}$$

and probability density function is

$$f(t \mid \underline{x}_i) = \lambda\gamma \exp\left(\sum \beta_j x_{ji}\right) t^{\gamma-1} \exp\left\{-\exp\left(\sum \beta_j x_{ji}\right)\lambda t^{\gamma}\right\} \tag{4}$$

Given covariate $\underline{x}_i$, let t_i^0, $i = 1, \ldots, n$, be the true failure times of a sample of size n, assumed to be independent identically distributed with a Weibull distribution with hazard function (2). Assuming that the observations are subject to arbitrary right censoring, the period of follow-up for the ith individual is limited to a value c_i. Then, the observed failure time of the ith individual is given by $t_i = \min(t_i^0, c_i)$. Define δ_i such that $\delta_i = 0$ if $t_i^0 \geq c_i$ (a censored observation) and $\delta_i = 1$ if $t_i^0 < c_i$ (an observed failure of some kind).

The Weibull proportional hazards model is fitted by constructing the likelihood function of the n observations

$$L(\lambda,\gamma,\underline{\beta}) = \prod_{i=1}^{n} \{f(t_i)\}^{\delta_i} \{S(t_i)\}^{1-\delta_i} \tag{5}$$

where $f(.)$ is (4) and $S(.)$ is (3). It is easier to maximize logarithm of this function instead of likelihood function itself for obtaining estimate of the unknown parameters λ, γ and $\underline{\beta}$.

Log-linear model

Consider a log-linear model for the random variable T_i associated with the survival time of the i^{th} individual in a survival study, according to which

$$\ln T_i = \beta_0^* + \sum \beta_j^* x_{ji} + \sigma\varepsilon_i \tag{6}$$

In this model β_1^*, β_2^*, …, β_k^* are the unknown coefficients of the value of k explanatory variables X_1, X_2, …, X_k, and β_0^*, σ are two further parameters, known as the intercept and scale parameter, respectively. The quantity ε_i is a random variable used to model departure of the values of ln T_i from the linear part of the model. Suppose that ε follows standard extreme value distribution with probability density function given by

$$f(\varepsilon) = \exp(\varepsilon - e^{\varepsilon}\}, \text{ for } -\infty < \varepsilon < \infty \tag{7}$$

Let $\xi = e^{\varepsilon}$, then the probability density function of ξ is $f(\xi) = e^{-\xi}$ which is exponential distribution with unit mean.

Now consider the survivor function of T_i

$$\begin{aligned}
S(t \mid \underline{x}_i) &= P(T_i \geq t) \\
&= P(\ln T_i \geq \ln t) \\
&= P\left(\beta_0^* + \sum \beta_j^* x_{ji} + \sigma\varepsilon_i \geq \ln t\right) \\
&= P\left(\varepsilon_i \geq \frac{\ln t - \beta_0^* - \sum \beta_j^* x_{ji}}{\sigma}\right) \\
&= P\left(e^{\varepsilon_i} \geq \exp\left\{\frac{\ln t - \beta_0^* - \sum \beta_j^* x_{ji}}{\sigma}\right\}\right)
\end{aligned} \tag{8}$$

Since $\xi = e^{\varepsilon}$ has a unit exponential distribution, so $P(e^{\varepsilon} \geq \xi) = e^{-\xi}$. It then follows that

$$S(t \mid \underline{x}_i) = \exp\left[-\exp\left\{\frac{\ln t - \beta_0^* - \sum \beta_j^* x_{ji}}{\sigma}\right\}\right] \tag{9}$$

There is a direct correspondence between equation (3) and equation (9), in the sense that

$$\beta_j = -\frac{\beta_j^*}{\sigma}, \; \lambda = \exp\left(-\frac{\beta_0^*}{\sigma}\right) \text{ and } \gamma = \frac{1}{\sigma} \tag{10}$$

With log-linear form of Weibull regression model, the likelihood function has a simple form. Since probability density function of ε is (7), then density function of Y_i = ln T_i is given by

$$f(y \mid \underline{x}_i) = \frac{1}{\sigma}\exp\left(z_i - e^{z_i}\right) \tag{11}$$

and its survivor function is

$$S(y \mid \underline{x}_i) = \exp\left(-e^{z_i}\right) \tag{12}$$

where

$$z_i = \frac{y - \beta_0^* - \sum \beta_j^* x_{ji}}{\sigma} \tag{14}$$

The likelihood function based on $y_1, y_2, \ldots, y_n$ which are logarithm of the n observed survival times $t_1, t_2, \ldots, t_n$ is then

$$L\left(\underline{\beta}^*, \beta_0^*, \sigma\right) = \prod_{i=1}^{n} \{f(y_i)\}^{\delta_i}\{S(y_i)\}^{1-\delta_i} \tag{15}$$

The Identification of HLP for Weibull Regression Model

Single HLP Identification

HLP is observation which is isolated in the covariate space (i.e., far removed from the main body of points in the X space). They can be thought of as outliers in the covariate space (Chatterjee and Hadi, 1986). HLP need not be influential, and influential observations are not necessarily HLP.

There are several ways to understand the characteristics of leverage in the linear regression model $yi = \alpha_0 + \Sigma\alpha_j x_{ji} + \varepsilon_i$, one of them is $w_{ij} = \partial\hat{y}_i / \partial y_j$ which measure the amount of leverage of the response value y_j on the predicted value $\hat{y}_i$. w_{ii} most directly reflects the influence of y_i on the fit. The generalization of leverage from linear regression to more general models can be based on this viewpoint.

Wei *et al.* (1998) showed that generalized leverage derived by the above approach is

$$W\left(\underline{\hat{\theta}}\right) = \left\{(D_\theta)\left(-\ddot{l}_{\theta\theta}\right)^{-1}\left(-\ddot{l}_{\theta y}\right)\right\}\Big|_{\theta=\hat{\theta}} \tag{16}$$

where

$$D_\theta = \frac{\partial E(\underline{y})}{\partial \underline{\theta}^T},\ \ddot{l}_{\theta\theta} = \frac{\partial^2 l(\underline{\theta})}{\partial\underline{\theta}\partial\underline{\theta}^T} \text{ and } \ddot{l}_{\theta y} = \frac{\partial^2 l(\underline{\theta})}{\partial\underline{\theta}\partial\underline{y}^T}$$

For log-linear specification of Weibull regression model (6)

$E(\underline{y}) = \beta_0^* \underline{1} + X\underline{\beta}^*$, $\underline{\theta} = \left(\beta_0^*, \underline{\beta}^*, \sigma\right)$ and $l(\underline{\theta})$ is logarithm of likelihood function (15).

If we only concern with regression coefficient $\underline{\psi} = \left(\beta_0^*, \underline{\beta}^*\right)$, then the generalized leverage is

$$W\left(\underline{\hat{\psi}}\right) = \left\{X\left(X^T \Lambda X\right)^{-1} X^T \Lambda\right\}\Big|_{\psi=\hat{\psi}} \tag{17}$$

where $\Lambda = diag\{\exp(z_i)\}$, $i = 1, \ldots, n$ and z_i is (14).

Next, let $\tilde{X} = \Lambda^{\frac{1}{2}} X$, it is then

$$W(\hat{\beta}) = \tilde{X}\left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{X}^T \tag{18}$$

Thus the *generalized leverage* of the i^{th} observation is given by

$$w_{ii} = \tilde{x}_i^T \left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{x}_i, i = 1,2,\ldots,n \tag{19}$$

In linear regression setting, points with w_{ii} greater than $2(k+1)/n$ (twice the average value) are generally regarded as HLP (Hoaglin and Welsch, 1978). Since, the situation is different for log-linear model, then we use 0.2 as a calibration point which stated by Huber (1981) as risky to dangerous points.

Another statistic for HLP identification is called *potential*. Potential is leverage of the i^{th} point which is based on a fit to the data with the i^{th} case deleted, namely

$$p_{ii} = \tilde{x}_i^T \left(\tilde{X}_{(i)}^T \tilde{X}_{(i)}\right)^{-1} \tilde{x}_i, i = 1,2,\ldots,n \tag{20}$$

where $\tilde{X}_{(i)}$ is matrix $\tilde{X}$ with i^{th} case deleted. Simple relationship between generalized leverage (19) and potential (20) is

$$p_{ii} = \frac{w_{ii}}{1 - w_{ii}} \tag{21}$$

with cut-off point: Median(p_{ii}) + 3 MAD(p_{ii}), where MAD(p_{ii}) = Median{| p_{ii} – Median(p_{ii})|}/0.6745.

Multiple HLPs Identification

Multiple HLPs identification method for Weibull regression model will be proposed in this section. It is motivated by the anticipation that the single case deleted measure discussed in the previous section may be ineffective for the identification of multiple HLPs because of masking and/or swamping effects. Let R be a set of cases 'remaining' in the analysis and D be a set of cases 'deleted'. Hence R contains $(n–d)$ cases after $d < (n–k)$ cases in D are deleted. We assume that the last d rows of $\tilde{X}$ is D set. *Group deleted leverage* based on group deleted cases D is

$$w_{ii(R)} = \tilde{x}_i^T \left(\tilde{X}_R^T \tilde{X}_R\right)^{-1} \tilde{x}_i, i = 1,2,\ldots,n \tag{22}$$

with cut-off point: Median($w_{ii(R)}$) + 3 MAD($w_{ii(R)}$). By adopting generalized potential proposed by Imon (1996), then its corresponding *group deleted potential* turns out to be

$$p_{ii}^* = \begin{cases} \dfrac{w_{ii(R)}}{1 - w_{ii(R)}}; & \text{for } i \in R \\ w_{ii(R)}; & \text{for } i \in D \end{cases} \tag{23}$$

with cut-off point: $\text{Median}(p_{ii}^*) + 3\text{MAD}(p_{ii}^*)$

Simulation on Identification of Multiple HLP for Weibull Regression Model

Scenario

We use Monte Carlo simulation to examine the performance of the multiple HLPs detection procedures across a wide range of scenarios. The simulation generated 90% of clean observations and plant HLPs at location specified by the scenario and factor settings. The regressor variables levels for clean observations are generated from a multivariate normal distribution with a mean of μ_x=7.5 and standard deviation of σ_x=4.0. The choice of these parameters does not affect the result of the simulations. The survival time response for the i^{th} clean observation is generated from $\text{Weibull}\left(\lambda \exp\left(\sum \beta_j x_{ji}\right), \gamma\right)$, where $\lambda = 1$, $\beta_j = -3$ and $\gamma = 3/2$ which correspond to $\beta_0^* = 0$, $\beta_j^* = 2$ and $\sigma = 2/3$ in log-linear form of Weibull regression. For the planted HLPs, the j^{th} regressor variable value for the i^{th} observations is $x_{ji} = \bar{x}_{j,clean} + 4\delta_L + \varepsilon_{ji}^*$ where $\bar{x}_{j,clean}$ is the average of the clean value for the j^{th} regressor, δ_L is the magnitude of the outlying shift distance in X-space and ε_{ij}^* is a random variate from a $U(0,0.25)$. We use the ε_{ij}^* term to separate multiple observations in a cloud. If the i^{th} observation is both HLP and regression outlier, then response value t_i is $t_i = t_i^* + \delta_R \sigma_t$, where t_i^* is generated from $\text{Weibull}\left(\lambda \exp\left(\underline{x}_i^{*\prime} \underline{\alpha}\right), \gamma\right)$, δ_R is the magnitude of the outlying distance off the regression plane in standard deviation units, σ_t.

Beside all the above factors, we also consider the level of censoring percentage (p_c). The percentage of censoring is defined by letting C which follow Uniform(0,a_c), where a_c will be chosen such that it results in an overall probability of censoring p_c = P(T > $C|\underline{x}$).

Our simulation studies aim to characterize the effects of specific factors on two primary measures of performance: detection capability and false alarm rate. The false alarm rate is the probability that a clean observation is swamped and the complement of detection probability is the masking probability. The factors considered are:

- Magnitude of unusualness in X-space, δ_L (3 and 5)
- Number of clouds (1 and 2)
- Magnitude of unusualness in response-space, δ_R (5)
- Percentage of censoring, p_c = (0%, 5% and 10%)
- Dimension of data ((n,# x variables) = (40,2) and (60,6))
- Proportion of x variables with extreme values (all k variables, 1 out of k variables and 3 out of 6 variables)

Each procedure's performance is evaluated on its ability to detect the planted HLPs and avoid false alarms. Both are reported for 500 replications.

Result

Table 1 and 2 showed the simulation results for one cloud of HLPs with 3 and 5 unit magnitude of usualness in X-space, respectively. Multiple HLPs detection based on $w_{ii(R)}$ and p_{ii}^* showed better result. The same result is also obtained from simulation with δ_L = 5. We also conduct simulation for 2 clouds of HLPs and 2 clouds of HLPs located at different response-space magnitude. The results still showed better performance for $w_{ii(R)}$ and p_{ii}^*.

δ_L	p_c	(n,k)	x variables	(1)*	(2)	(3)	(4)
3σ_x	0%	40,2	2	0.035(0.001)	0.417(0.098)	0.663(0.071)	0.639(0.099)
(1 cloud)			1	0.020(0.001)	0.354(0.098)	0.483(0.077)	0.453(0.103)
		60,6	6	0.019(0.007)	0.172(0.101)	0.708(0.032)	0.667(0.079)
			3	0.025(0.006)	0.227(0.095)	0.568(0.035)	0.523(0.082)
			1	0.020(0.006)	0.198(0.095)	0.314(0.042)	0.271(0.091)
	5%	40,2	2	0.030(0.001)	0.444(0.094)	0.686(0.070)	0.665(0.095)
			1	0.017(0.001)	0.377(0.097)	0.512(0.076)	0.493(0.102)

		60,6	6	0.014(0.005)	0.110(0.076)	0.805(0.024)	0.783(0.059)
			3	0.032(0.005)	0.212(0.089)	0.600(0.032)	0.563(0.077)
			1	0.018(0.006)	0.197(0.093)	0.338(0.041)	0.283(0.088)
	10%	40,2	2	0.028(0.001)	0.452(0.092)	0.701(0.069)	0.677(0.095)
			1	0.016(0.001)	0.377(0.098)	0.515(0.075)	0.490(0.102)
		60,6	6	0.014(0.004)	0.099(0.071)	0.826(0.022)	0.805(0.054)
			3	0.029(0.005)	0.196(0.083)	0.645(0.029)	0.610(0.071)
			1	0.018(0.006)	0.197(0.091)	0.336(0.040)	0.278(0.087)
Average probabilities				0.022(0.004)	0.268(0.092)	0.580(0.049)	0.547(0.086)

Note: * Four detection methods compared: (1) generalized leverage, (2) potential, (3) group deleted leverage, (4) group deleted potential

Table 1. Scenario with detection and false alarm probabilities (in parenthesis) for 1 cloud multiple HLPs with $3\sigma_x$ of magnitude of usualness in X-space

δ_L	p_c	(n,k)	x variables	(1)*	(2)	(3)	(4)
$5\sigma_x$	0%	40,2	2	0.086(0.001)	0.523(0.093)	0.861(0.068)	0.849(0.095)
(1 cloud)			1	0.071(0.001)	0.497(0.090)	0.757(0.069)	0.740(0.096)
		60,6	6	0.030(0.007)	0.200(0.100)	0.867(0.029)	0.850(0.075)
			3	0.049(0.005)	0.274(0.094)	0.787(0.030)	0.760(0.076)
			1	0.045(0.006)	0.278(0.091)	0.550(0.034)	0.506(0.082)
	5%	40,2	2	0.078(0.001)	0.577(0.089)	0.871(0.065)	0.863(0.090)
			1	0.067(0.001)	0.527(0.089)	0.749(0.069)	0.731(0.095)
		60,6	6	0.014(0.003)	0.091(0.058)	0.937(0.018)	0.926(0.045)
			3	0.041(0.004)	0.230(0.074)	0.847(0.024)	0.821(0.061)
			1	0.038(0.005)	0.273(0.087)	0.586(0.032)	0.545(0.079)
	10%	40,2	2	0.068(0.001)	0.617(0.088)	0.886(0.062)	0.878(0.086)
			1	0.063(0.001)	0.546(0.086)	0.763(0.068)	0.751(0.093)
		60,6	6	0.016(0.003)	0.088(0.057)	0.939(0.018)	0.928(0.044)
			3	0.040(0.004)	0.219(0.068)	0.854(0.022)	0.835(0.057)
			1	0.034(0.005)	0.269(0.083)	0.618(0.031)	0.572(0.076)
Average probabilities				0.049(0.003)	0.347(0.083)	0.791(0.043)	0.770(0.077)

Note: * Four detection methods compared: (1) generalized leverage, (2) potential, (3) group deleted leverage, (4) group deleted potential

Table 2. Scenario with detection and false alarm probabilities (in parenthesis) for 1 cloud multiple HLPs with $5\sigma_x$ of magnitude of usualness in X-space

Conclusion

The group deleted procedure ($w_{ii(R)}$ and p^*_{ii}) have the better detection probability and have false alarm rate about the nominal 5% level. As a general rule, the proposed method perform better in higher covariate space dimension, larger number of outlying regressor variables, larger outlying leverage distance, larger magnitude of unusualness in response-space, larger number of multiple points clouds and higher percentage of censoring.

References

Chatterjee, S. and Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression. *Stat. Sci.* Vol. 1(3), pp: 379-416.

Hoaglin, D. C. and Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *Amer. Statist.* 32. 17-22.

Huber, P. (1981). *Robust Statistics*. New York: Wiley.

Ibrahim, Kudus, A., Daud, I. and Abu Bakar, M. R. (2008). Decision tree fro competing risks survival probability in breast cancer study. *Int. J. of Biomed. Sci.* Vol. 3(1).

Imon, A. H. M. R. (1996). Subsample methods in regression residual prediction and diagnostics. *PhD thesis*, School of Mathematics and Statistics, University of Birmingham, UK.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data 2nd Ed.* Wiley Interscience. New Jersey.

Kudus, A. and Ibrahim, N. A. (2005). Competing Risks Cox Proportional Hazard Modeling Using SAS. *SAS User Malaysia Forum 2005.* http://www.sas.com/offices/asiapacific/malaysia/events/sum2005docs/19.pdf

Kudus, A., Ibrahim, N. A., Abu Bakar, M. R. and Daud, I. (2005). Regression Trees for Competing Risks Survival Data. *CD Proceedings of International Conference on Applied Mathematics. Bandung*, August 22-26.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, Chapman & Hall, London.

Rousseeuw, P. J., and Leroy, A. (1987). *Robust regression and outlier detection.* Wiley, New York.

Wei, B-C., Hu, Y-Q and Fung, W-K. (1998). Generalized leverage and its application. *Scand. J. Stat.* Vol. 25, pp:25-37.

Wisnowski, J. W., Montgomery, D. C. and Simpson, J. R. (2001). A comparative analysis of multiple outlier detection procedures in the linear regression model. *Comput. Statist. Data Anal.* 36:351-382