NEW TEST STATISTICS TO ASSESS THE GOODNESS-OF-FIT OF LOGISTIC REGRESSION MODELS

JASSIM NASSIR HUSSAIN

UNIVERSITI SAINS MALAYSIA

2013

NEW TEST STATISTICS TO ASSESS THE GOODNESS-OF-FIT OF LOGISTIC REGRESSION MODELS

by

JASSIM NASSIR HUSSAIN

Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

April 2013

ACKNOWLEDGEMENTS

Throughout this endeavor, my PhD study in the School of Mathematical Sciences, Universiti Sains Malaysia (USM), could never have been fulfilled without the full support from my supervisor, Institute of Postgraduate Studies (IPS), the School's office, my friends and my family, who helped make the most productive, meaningful and memorable time of my life. I would like to take this opportunity to thank all of them for their support, inspiring guidance, helpful discussion, beautiful friendship and generous love.

First and foremost, I would like to express my sincere appreciation to my supervisor, Prof. Dr. Low Heng Chin, for everything she has done to make my doctoral study the most rewarding experience forever. I owe special thanks to Assoc. Prof. Dr. Rasimah Aripin, who thoroughly reviewed the SAS program. It was hard work and I really appreciate it. Also, special thanks go to Prof. Dr. Quah Soon Hoe for his help, support and valuable advices.

I would also like to thank all USM staff and members especially the members of the Institute of Postgraduate Studies (IPS) for their warm-hearted help, financial support through the USM fellowship, and helpful advices during my doctoral study.

My appreciation and thanks would also go to the School of Mathematical Sciences committee for giving me the chance to be a tutor and for the financial support through the research grant scheme. And here my appreciation also goes to all the staff that I worked with for their infinite patience and generous help during my teaching experience, which will be the wealth of my whole life.

On behalf of all the graduate students in the department a special thanks to our wonderful School staff, technicians in the Computer Lab in the School and library staff. I cannot thank you enough for all you do above and beyond the job description to help the students and faculty in our department.

This acknowledgement would not be complete without thanking all the graduate students in the department. Thank you all for your love, support and helping us to make Penang a home away from home.

Last but not the least; I would like to give my special thanks to my beloved wife and best friend, my son, Atheer. Their support throughout this process is endless and I would not be where I am today without their support and love. I am forever grateful. I would also thank my dear parents, brothers and sisters for their support during my overseas study. Their love is always the motivation to pursue my goals.

TABLE OF CONTENTS

.

CHAPTER SEVEN: CONCLUSIONS

.

LIST OF TABLES

.

LIST OF FIGURES

Page

LIST OF SYMBOLS

Symbol	Description	page
$g(\mu)$	Link function in generalized linear model	1
$\pi(\mathbf{x}_i)$	Probability of success $\pi(\mathbf{x}_i) = \pi_i = Pr(Y = 1\|\mathbf{x_i})$	3
$logit\big(\pi(\mathbf{x}_i)\big)$	The logistic transformation $= \ln\{\pi(\mathbf{x}_i)/\big(1 - \pi(\mathbf{x}_i)\big)\}$	3
Y	The response (dependent) variable y_i , $i = 1, 2, \dots, n$	3
$\mathbf{x_i}$	Vector of covariates where $\mathbf{x}_i = (x_{i1}, \cdots, x_{iP})$ and x_{ij} is i^{th} observation in j^{th} characteristic, $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, P$	3
n	The number of observations (elements) in the sample	4
P	The number of covariates (characteristics) in the model	4
ln or log	The natural logarithm	4
$\boldsymbol{\beta}$	$\beta_0, \beta_1, \dots, \beta_j, \dots, \beta_P$ are the unknown parameters in the model	4
$\hat{\pi}_i = \hat{\pi}(\mathbf{x}_i)$	The estimated of the probability success π_i	8
$\hat{r}_i$	The estimated residuals $\hat{r}_i = y_i - \hat{\pi}(\mathbf{x}_i)$	8
F_0	Completely specified probability distribution function	12
x_i	The i^{th} observation in the sample of univariate variable X	12
$\hat{x}_i$	The estimated values of the i^{th} observation in univariate variable	12
$V(x_i)$	The variance of univariate variable X	12
G	Number of categories of the variable	13
X^2	Criterion of goodness-of-fit test statistics	14
n_g	The observed number of the observations included into the g^{th} category	14
m_g	The expected number of the observations included into the g^{th} category	14

K	Number of groups (groups) $k = 1, 2, \dots, K$	14
R_k	The groups, $R_1, \cdots, R_K$	14
p_k	The probability of assigning the observation in k^{th} group	14
θ	The unknown parameter of the model	15
e_g	The "expected" frequency equal to the sampling expectation, $e_g = E(N_g)$ in Armitage (1966) study	16
N_g	Observed number of failures in the g^{th} category in Armitage (1966) study	16
p_{kg}	The observed proportion in the $(k, g)^{th}$ cell in Armitage (1966) study	16
t_g	The total observations in g^{th} category in Armitage (1966) study	16
d	The boundary of the group, $(-\infty,\ d_1), (d_1,\ d_2), \cdots, (d_{K-1},\ \infty)$	17
λ	The root of the determinant equation $\left\|\tilde{\boldsymbol{I}} - (1 - \boldsymbol{\lambda})\hat{\boldsymbol{I}}\right\| = 0$, $\lambda_1, \dots, \lambda_P$	17
$\tilde{\boldsymbol{I}}$	The information matrix per observation for θ when it is estimated from the observed data	17
$\hat{\boldsymbol{I}}$	The information matrix per observation for θ when it is estimated from the original data	17
$\hat{C}$	Hosmer and Lemeshow GOF test	19
g_k	Covariate patterns with in the k^{th} group.	20
O_k	The number of observed response variables of the k^{th} group	20
m_g	Number of observations in the g^{th} covariate pattern	20
E_k	The number of expected response variables of the k^{th} group	20
$\bar{\pi}_k$	The average of estimated probabilities for k^{th} group	20
$\hat{\pi}_g$	The estimated probability for the g^{th} covariate pattern	20
n_k	The number of observations in the k^{th} group	20
$\hat{H}$	The second GOF test of Hosmer and Lemeshow	20

s_g	The score of the response variable of g^{th} category	22
Z_i	The observed score	23
μ_i	The mean score	23
$\hat{\mu}_i$	The predicted mean score	23
I_{ik}	The indicator variable $I_{ik} = 1$ if $\hat{\mu}_i$ is in group *k*, and *0* otherwise	23
γ_k	The unknown parameter of the grouping effect	23
H_0	The null hypothesis	23
w_i	The sampling weighting due to selection probabilities	24
L	The likelihood function	24
C_k	Fagerland *et al*. (2008) GOF test statistic	28
$\hat{T}'$	The vector of deciles of risk of the weighted residual	30
$\hat{\boldsymbol{V}}(\hat{\boldsymbol{T}})$	The estimated variance–covariance matrix of the weighted residual	30
$\hat{M}'$	The vector of deciles of the estimated means of the residuals	31
$I^{(k)}$	The indicator function of belonging i^{th} element of **X** to k^{th} group	32
$(\boldsymbol{V}^{-})$	Generalized inverse of the matrix $\boldsymbol{V}$	33
$\tilde{K}_K$	Zhang (1999) GOF test statistic	34
$\tilde{Q}_K$	The deviations between the observed group probabilities obtained from the nonparametric maximum likelihood and the maximum semiparametric likelihood estimator	34
$F(x)$ or $G(x)$	The cumulative distribution function of X	35
$f(x)\ or\ g(x)$	The probability density functions (*pdf*) corresponding to $F(x)$ and $G(x)$	35
$\hat{q}_k$	The observed group probabilities obtained from the nonparametric maximum likelihood estimator	36
$\tilde{q}_k$	The observed group probabilities obtained from the maximum semiparametric likelihood estimator	36

L_k	The general chi-square statistic of Deng et al. (2009)	36
T_{EL}	Evans and Li (2005) GOF test statistic	37
$X^2_{P^*}$	Xie et al. (2008) chi-square GOF test statistic	38
T^*	Tsiastis (1980) and Xie et al. (2008) Score GOF test statistic	39
X^{2*}, D^{2*}	Pulkstenis and Robinson (2002) GOF test statistics	42
SD_λ	The family of power divergence statistics of GOF test	43
G^2	The log likelihood ratio statistic	44
T^2	The Freeman-Tukey statistic	44
χ^2	Pearson's chi-square statistic	44
NM^2	The Neyman modified χ^2 statistic	44
GM^2	The modified log likelihood ratio statistic	44
a_λ	Some kind of distance basically compares the ratio of observed frequencies to expected frequencies raised to the power λ	45
λ	A real parameter $\lambda \in R$	45
v_g	The variance of number of successes in g^{th} covariate category	46
y_g	Dependent variable in the g^{th} covariate category defined as $y_g = (1 - 2\hat{\pi}_g)/v_g$	46
B	The correction factor for the variance	46
Z	Osius and Rojek (1992) GOF test statistic	46
$\boldsymbol{\eta_i}$ or $\boldsymbol{\eta}(\mathbf{x}_i)$	The linear combination of the covariates in LRM	47
ST	Stukel (1988) GOF test statistic	47
$l(X)$	The maximum log-likelihood estimation from the assumed model	47
$l(X, D)$	The maximum log-likelihood estimation from the generalized LRM	47
$P\hat{R}_1$	Royston (1992) monotone GOF test statistic	48

$P\hat{R}_2$	Royston (1992) quadratic GOF test statistic	48
$\xi_{ik}^{(j)}$	The j^{th} element of a P-dimensional linear predictor	50
U	Lin (2010) GOF test statistic based on the unweighted sum of residual squares	51
T_h	Lin and Chen (2011) GOF test	52
R_h	The vector of nonparametric smoothing residuals	52
$\widetilde{D}$	The Bartlett-type (transformed deviance) statistic proposed by Taneichi, *et al*.(2011)	54
$\boldsymbol{c_k}$	The vector of initial seeds for k^{th} group	60
S_j	The minimum distance by which the seeds of groups must be separated	60
$\tilde{\boldsymbol{c}}_{\boldsymbol{k}}$	The vector of means for k^{th} group	62
D_i	The distance function between the i^{th} observation covariates and the group centroid $\boldsymbol{c}_k$	62
$d(\mathbf{x}_i, \boldsymbol{c}_k)^2$	The Euclidean distance measurement between the observation $\mathbf{x}_i$ and the group centre $\boldsymbol{c}_k$	62
$I(\mathbf{x}_i, R_k)$	The association function of observation $\mathbf{x}_i$ with the group k	63
S_{il}	Gower's general similarity measure between two elements i and l	65
d_{il}	Gower's dissimilarity measure between two elements i and l	65
S_{ilj}	The similarity score denotes the contribution provided by the j^{th} covariate	65
W_{ilj}	The weight of the j^{th} covariate	65
r_j	The range of the observations for the j^{th} covariate	65
R^2	The measure of the dissimilarity or separation of groups	72
I_{igk}	The indicator variable of determining the location of the observed and the estimated response variable y_i and $\hat{\pi}_i$ in g^{th} category of k^{th} group	75
w_k	The weight of k^{th} group	76

n_k	The number of elements in k^{th} group	76
y_{igk}	The i^{th} element of the g^{th} category of the observed response variable in k^{th}group	76
O_{gk}	The observed frequency in the g^{th} category and in k^{th} group	76
$\hat{\pi}_{igk}$	The i^{th} estimated probability in the g^{th} category of the k^{th} group	79
E_{gk}	The weighted expected frequency for g^{th} category in k^{th} group	79
X_c^2	The first developed GOF test (grouped chi-square test statistic)	80
D_c^2	The second developed GOF test (grouped deviance test statistic)	80
$o_p(x_n)$	The random variable of smaller order than x_n for large n, the subscript p indicates that the sequence has a probabilistic rather than deterministic behavior.	81
$O_p(x_n)$	A random variable of same order of x_n	81
$\mathbf{V}$	Idempotent and a symmetric matrix	82
$\boldsymbol{\phi}$	An eigenvector of $\mathbf{V}$	82
$R(\mathbf{V})$	The rank of $\mathbf{V}$	83

.

LIST OF ABBREVIATIONS

Abbreviation	Description	Page
GLM	Generalized Linear Model	1
GOF	Goodness-Of-Fit	1
LRM	Logistic Regression Model	1
Logit	Logistic link function (Log of the odds ratio)	2
MLE	Maximum Likelihood Estimation	4
MLRM	Mixed Logistic Regression Model	4
CSM	Group - Specific Model	4
PAM	Population - Averaged Model	4
RELEM	Random Effect Logistic Regression Model	4
GEE	Generalized Estimation Equation	4
LR	Likelihood Ratio	6
AIC	Akaike Information Criterion	6
BIC	Bayesian Information Criterion	6
OLS	Ordinary Least Square	6
Pdf	Probability Distribution Function	14
ESS	Error sum of squares	63
CC	Compactness Criterion (or Cluster Criterion)	72
MESS	Mean of error sum of squares	72
TSS	Total Sum Squares	72
SAS	Statistical Analysis System (Software)	103
PC	Personal Computer	103

LIST OF PUBLICATIONS

1. Hussain, J.N. Heng, C. L. and Abbas, F.M., (2008) An overview of evaluation criteria in logistic regression model. *Mathematics Journal: Special edition Part II,* 431-437.

2. Hussain, J.N. Heng, C. L. & Abbas, F.M., (2008) Sensitivity analysis for survival regression models, *Mathematics Journal, Special edition Part II*, 525-533.

3. Hussain, J.N., (2008) Sensitivity analysis to select the most influential risk factors in a logistic regression model, *International Journal of Quality, Statistics and Reliability*, 4, 1-10.

4. Hussain, J.N. Heng, C. L. and Abbas, F.M., (2007) Modeling and assessing the important factors of water quality. Presented In: the *2nd Regional Conference (ECOMOD 2007*), School of Biological Sciences, USM, Penang, 28-30 August.

5. Hussain, J.N. and Heng, C. L., (2008) Chi-Square Goodness-of-fit tests for the Logistic Regression Model: Which one is best? Presented in: *the 2nd International Conference on Science and Technology: Application in Industry and Education (ICSTIE 2008)*. Universiti Teknologi Mara, Penang, 12-13 December.

6. Hussain, J.N. and Heng, C. L., (2009) An alternative method to Construct goodness-of-fit Test for multinomial logistic regression model**.** Presented in: *the 5th Asian Mathematical Conference (AMC2009)*, Kuala Lumpur, 22 – 26 June.

.

STATISTIK UJIAN BARU UNTUK MENILAI KEBAGUSAN PENYUAIAN MODEL REGRESI LOGISTIK

ABSTRAK

Model regresi logistik binari (binomial) merupakan satu daripada model linear teritlak (GLMs). Ia digunakan apabila pembolehubah bersandar adalah dikotomi dan pembolehubah tidak bersandar terdiri daripada lain-lain jenis. Model regresi lojistik digunakan dalam pelbagai bidang termasuk biomedikal dan sains kemasyarakatan. Menilai kebagusan penyuaian (GOF) dianggap sebagai langkah penting selepas menyuaikan model bagi menunjukkan kecukupan Model regresi lojistik untuk menyuaikan cerapan. Ujian GOF ditakrif sebagai penilaian ketepatan anggaran hasil dengan data cerapan. Terdapat dua teknik yang boleh digunakan untuk membangunkan statistik ujian GOF khi kuasa dua. Teknik pertama adalah berasaskan kepada cerapan individu. Teknik ini tidak digemari di dalam model regresi lojistik kerana beberapa sebab sepert taburan yang diperolehi dan nilai-nilai *P* adalah salah. Manakala teknik kedua adalah berasaskan kepada strategi kelompok dan teknik ini kerap digunakan.

Beberapa stategi kelompok, di mana pengujian statistik GOF diasaskan, telah dicadangkan supaya menilai GOF bagi model regresi lojistik. Kesemua strategi dan pengujian statistik GOF mempunyai batasan. Sehubungan itu, cadangan strategi kelompok yang baru dan pembangunan pengujian statistik GOF baru yang berasaskan strategi ini bertujuan untuk menilai kecukupan model regresi lojistik didorong oleh batasan strategi kelompok dan pengujian statistik GOF sedia ada. Seterusnya objektif utama tesis ini adalah cadangan strategi kelompok baru dan

pembangunan dua ujian statistik GOF X_c^2 dan D_c^2 berdasarkan untuk menilai kecukupan model regresi lojistik. Ujian statistik GOF yang dibangunkan direka bagi memperolehi kebagusan penyuaian model regresi lojistik yang menyeluruh supaya dapat digunakan ke atas sebarang set data.

Kajian ini melaksanakan lapan simulasi yang mewakili lapan situasi yang berbeza berserta model regresi lojistik yang berbeza dan menganalisis tiga set data klinikal. Objektif utama melaksanakan kajian simulasi ini adalah untuk mengkaji prestasi strategi kelompok yang dicadangkan serta membangunkan ujian statistik GOF X_c^2 dan D_c^2; dan membandingkan prestasi tersebut dengan ujian statistik GOF semasa.

Semua penemuan dari kajian simulasi dan penganalisisan set data klinikal menunjukkan bahawa strategi kumpulan yang dicadang mempunyai kemampaan untuk mengasingkan elemen-elemen set data. Ujian statistik GOF yang baru mempunyai taburan khi kuasa dua. Ia mempunyai kuasa yang cukup untuk mengesan sisihan dari model regresi lojistik yang benar. Ujian statistik GOF yang baru mempunyai kuasa pengesanan yang tinggi berbanding dengan ujian statistik sediada bagi faktor yang berbeza. Secara umumnya, penemuan kajian menunjukkan bahawa cadangan strategi kluster dan ujian statistik GOF yang berasaskan strategi ini mempunyai potensi untuk digunakan sebagai cadangan strategi pengasingan yang dicadangkan dan ujian statistik GOF untuk menilai kecukupan model regresi lojistik.

.

NEW TEST STATISTICS TO ASSESS THE GOODNESS-OF-FIT OF LOGISTIC REGRESSION MODELS

ABSTRACT

The binary (or binomial) logistic regression model (LRM) is one of the generalised linear models (GLMs). It is used when the dependent variable is dichotomous and the independent variables are of any type. LRM are popular in many applications and in different disciplines including biomedical and social sciences. Assessing the goodness-of-fit (GOF) is considered to be the important step after fitting the model to show the adequacy of the LRM in fitting the observations. The GOF test is defined as an evaluation of how well the estimated outcomes agree with the observed data. Two techniques may be used to construct the GOF test statistics of chi-square type. The first technique is based on ungrouped observations. This technique is not preferred in the LRM for many reasons including that the obtained distribution and $p-\text{values}$ are incorrect. The second technique is the preferred technique where it is based on grouping the observations.

Many strategies of grouping and GOF test statistics based on these strategies of grouping have been proposed to assess the GOF for the LRMs. All these strategies and GOF test statistics have their own limitations. Hence, proposing a new strategy of grouping and developing new GOF test statistics based on the proposed strategy of grouping to assess the adequacy of the fitted LRM are motivated by the limitations of currently available strategies of grouping and GOF test statistics. Consequently, the main objectives of this thesis are to propose a new strategy of grouping based on partitional group analysis and to develop two GOF test statistics, X_c^2 and D_c^2 based on

the new strategy of grouping to assess the adequacy of the LRMs. The developed GOF test statistics are designed to enable us to determine the overall GOF of the LRM to any data set.

Eight simulation studies representing different data settings with different LRMs are implemented and three clinical datasets are analysed. The main objectives of conducting these simulation studies are to examine the performance of the proposed strategy of grouping and the developed GOF test statistics X_c^2 and D_c^2; and to compare their performance with the existing GOF test statistics.

All the results from analysing these simulation studies and clinical datasets show that the proposed strategy of grouping has adequate efficiency to partition the elements of the dataset. The new GOF test statistics have a chi-square distribution. They have adequate power of detection for the departure from the true LRMs. The new GOF test statistics have a high power of detection compared to the existing test statistics for different factors. In general, these results show that the proposed strategy of grouping and GOF test statistics based on it have a potential use in practice as a recommended strategy of partitioning and as GOF test statistics to assess the adequacy of the LRMs.

CHAPTER ONE

INTRODUCTION

1.1 Introduction

Generalised linear models (GLMs) are extended from ordinary regression models to model the relationship between covariates and several types of nonnormal response variables (i.e., continuous, dichotomous, counts). Binary (or binomial) logistic regression model is one of these models. It is used when the dependent variable is a dichotomy and the independent variables are of any type. Corresponding with the increase in application of logistic regression model (LRM), there has been an increase in statistical research on this model. These researches aim to evaluate the modelling process of these models which involves many activities such as assessing the overall goodness-of-fit (GOF), choosing the relevant distribution of error, selecting variables to be included in the systematic component and specifying the link function $g(\mu)$ to be used. Therefore, one area of current research is the development of new methods to assess the overall GOF of this model, because assessing overall GOF for the LRM is considered as the principle activity in the modelling process. On the other hand, GOF refers to the adequacy of the fitted model; this may include detection of when important covariates are omitted, when the link function is not appropriate or when the functional form of modelling covariates is not correct. Consequently, assessing the overall GOF for the LRM is widely studied and many strategies of grouping and test statistics based on them have been proposed.

.

1.2 Background of the Study

GLMs represent a unified statistical framework aiming to investigate the effect of a set of explanatory variables on the (conditional) mean of the response variable (Muggeo and Ferrara, 2008). Nowadays, GLMs are part of the standard empirical research and they are commonly employed in several disciplines. McCullagh and Nelder (1989) described these models in great detail and indicated that the term '*generalised linear model*' is due to Nelder and Wedderburn (1972) who described how a collection of seemingly dissimilar statistical techniques could be unified. GLM is a linear model for a transformed mean of a response variable that has distribution in the natural exponential family. Three components compose a generalised linear model: a *random component* which identifies the response variable Y and its probability distribution; a *systematic component* which specifies the explanatory variables used in a linear predictor function; and a *link function* specifies the function of $E(Y)$ that the model associates to the systematic component (Agresti, 2002).

GLMs consist of a large family of models. LRM is one of them, to model the relationship between dichotomous response variable and any type of covariates. LRM is GLM with binomial random component, logit link function and any type of covariates. Therefore, there is no guarantee that a certain LRM fits the data well in practice. Consequently, overall GOF of the resulting model should be examined after the coefficients in LRM have been estimated. Thus, assessing GOF for LRM is considered as a main activity in the modelling process. The following subsections give an overview about the LRM and the overall GOF test.

.

1.2.1 Logistic regression model

The logistic (logit) link function is widely used in GLMs when the response variable is not numerical, but categorical, e.g. a binary response variable as alive or dead, success or failure to give a binary LRM, or a more than two categories response variable (nominal or ordinal) e.g. disease status (nil, first stage, second stage, and high stage of disease) to give a multinomial LRM. The logistic link function or logit of probability of success $\pi(\mathbf{x}_i) = \pi_i = Pr(Y = 1|\mathbf{x_i})$ is $logit(\pi(\mathbf{x}_i)) = ln\{\pi(\mathbf{x}_i) / (1 - \pi(\mathbf{x}_i))\}$ and it has been defined as log odds ratio of success. The models with this type of link function are called LRMs (Agresti, 2002). Berkson (1944) stated that the LRM was discovered in the beginning of the twentieth century to describe the population growth, and it was called the "logistic" function. Many authors considered LRM as the most important model for categorical response data such as (Hosmer and Lemeshow (2000) and Agresti (2002). It is used increasingly in a wide variety of applications. Early uses were in biomedical studies; however, the past 20 years have also seen much use in social science research and marketing. Recently, logistic regression has become a popular tool in business applications as well. Another area of increasing application is genetics.

LRMs involves the association between covariates and binomial response variable which are found in different disciplines including biomedical research, ecology, health policy, and biology (Hosmer and Lemeshow, 2000 and Dreisitl *et al*., 2005). The logistic transformation or logit is preferred from other transformations, because it helps to transform any value of $\pi(\mathbf{x}_i)$ in the range (0 to 1) to corresponding values of $logit(\pi(\mathbf{x}_i))$ in the range $(-\infty \text{ to } +\infty)$. The LRM for the dependence of

$\pi(\mathbf{x}_i) = Pr(Y = 1|\mathbf{x}_i)$ on the vector $\mathbf{x}_i = (x_{i1}, \cdots, x_{iP})$ of P covariates, where x_{ij} is i^{th} observation in j^{th} characteristic, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, P$, is (Agresti, 2002):

$$logit\, \pi(\mathbf{x}_i) = ln\left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}\right) = \boldsymbol{\beta}'\mathbf{x}_i \qquad (1.1)$$

where $\pi(\mathbf{x}_i)$ is the corresponding probability of i^{th} response variable $Y = y_i$; $i = 1, 2 \ldots, n$ and β_j; $j = 0, 1, \cdots, P$ are the unknown parameters. The mechanism of maximum likelihood estimation (MLE) and model fitting for the LRMs are special cases of the GLMs fitting (Cohen *et al.*, 2003).

Another kind of LRM is the mixed logistic regression model (MLRM) to accommodate the correlated and over dispersed data by adding random effects to the linear predictor. Its application is useful in various disciplines, such as the analysis of grouped data including longitudinal data or repeated measures (Hosmer and Lemeshow, 2000 and Feddag and Mesbah, 2006). MLRM is popular for grouped binary data as well as for the independent binary data. It is consisting of two main types of models to associate grouped binary data with any type of grouped covariates. The first model is the cluster-specific model (CSM) and the second model is the population-averaged model (PAM). The CSM includes cluster effects which are useful for assessing the effects of cluster-varying covariates. Alternatively, PAM does not include cluster effects and they are most useful for assessing the effects of cluster level covariates (Hosmer and Lemeshow, 2000). Typically, the random effect logistic regression model (RELRM) is used as representative of a CSM and the logistic generalised estimation equation (GEE) method is used as a representative of a PAM

to estimate the parameters. Estimated parameters and estimated standard error can be used to calculate the odds ratios and to construct GOF tests for these models (Hosmer and Lemeshow, 2000 and Neter *et al*., 1996). Even though the evaluation of the overall GOF for these models is widely studied, it remains far less studied than in the case of the ordinary LRMs. The following subsection discusses the GOF test in the LRMs.

1.2.2 GOF test for the LRM

GOF test is considered as the main component of any modelling process. It is defined as an evaluation of how well the model predicted outcomes agree with the observed data (Pulkstenis and Robinson, 2002). Although, GOF test refers to adequacy of the model, it is widely referred to as *lack-of-fit,* because it measures how far is the model from the data (Archer *et al.*, 2007). The most important factor which causes the lack-of-fit for the LRM is the problems with the linear component such as omitting the higher order terms of covariates; deleting the important covariates related to the response variables from the model and deleting the influential observations. Outliers can also lead to a poor fitting (Collett, 2003). Basically, GOF tests intend to detect the presence of the lack-of-fit of the model to the observed data without indicating the nature of the problem (Pregibon, 1981). Consequently, the model is said to fit poorly if the model's residual variation is large and systematic (Hosmer *et al.*, 1997). This is the case when the predicted values produced by the LRMs do not accurately reflect the observed values in the modelling process.

In general, the modeling process of the LRM can be divided into five activities. Consequently, the GOF test statistics can be classified based on these activities into the following groups (Hussain *et al.*, 2008):

1) the overall GOF tests which are used to test the overall fitting of the model, for example, the deviance, Pearson chi-square test, Hosmer and Lemeshow tests.

2) the power of association tests which are used to measure the power of association in LRMs such as Pseudo-R^2 indices, which include the likelihood index**,** Cox and Snell index and Nagelkerke index.

3) individual parameter tests which are used to test the goodness of estimation of the parameters of the model, such as likelihood ratio (LR) test for single predictor, Wald score test, and Partial deviance.

4) the link function tests which are used to test the suitability of chosen link function, e.g., Box-Tidwell transformation test and Logit step test.

5) the model comparison tests which are used to compare the nested or not nested models, such as the maximum likelihood ratio test for nested model, meanwhile Akaike information criterion (AIC) and Bayesian information criterion (BIC) for not nested models. Table 1.1 presents a comparison between the modelling test statistics in the LRMs and the modelling test statistics in ordinary least square (OLS) regression models (Hussain *et al*., 2008):

.

Table 1.1 comparisons of modeling test statistics in both LRM and OLS regression models

Test statistics in modeling activities	LRMs	OLS models
Over all GOF tests	a- The likelihood ratio (LR) b- H-L test of overall fitting c- Score tests	*F*-test
Power of association	Pseudo-R^2 such as: a- Likelihood index b- Cox and Snell index c- Nagelkerke index	Coefficient of determination (R^2)
Coefficients evaluation	a- LR test for single predictor b- Wald score test c- Partial deviance	*t*-test
Link function	a- Box-Tidwell transformation test b- Logit step test	*t*-test
Comparison of fitted models	a- M L test for nested model b- AIC, BIC	Information criteria. (AIC, BIC)

In the setting of LRMs, GOF test statistics of the first group in Table 1.1 can be constructed by using one of two techniques based on the chi-square approach. The first technique has been used in the OLS regression model, which is based on ungrouped observations. In this technique, the process of fitting the model to the data is considered as a way of replacing a set of data values y by a set of fitted data $\hat{y}$ from the model involving a small number of parameters P. This technique is not preferred in the LRMs for many reasons, such as that the obtained distribution and the calculated p-values are incorrect because they are based on parameters coming from ungrouped data and the estimation of the variance is imprecise especially in the categorical data (Kuss, 2002 and Collett, 2003).

Therefore, the alternative technique which is used in LRMs is based on grouping the observations. In this technique, the process of fitting the model is considered as a comparison between the observed and the fitted numbers of data in each group. This

technique is preferred to construct the GOF tests in the LRMs (Hosmer and Lemeshow, 2000). According to this technique, there are four groups of strategies of grouping have been proposed by the researchers. The first group involves strategies based on grouping the estimated probabilities $\hat{\pi}_i = \hat{\pi}(\mathbf{x}_i)$. The second group involves strategies based on grouping the estimated residuals $\hat{r}_i = y_i - \hat{\pi}(\mathbf{x}_i)$ from the fitted model. The third group involves strategies based on grouping the range of the covariates using the categories of the categorical covariates. The fourth group involves strategies based on combined grouping for the range of the response variable and the range of the covariates using their categories.

Consequently, many strategies of grouping have been proposed based on these ideas (see Section 2.3). There are other types of strategies that are used to construct the GOF based on grouping or without grouping the observations and approaches other than chi-square approach (see Section 2.4). Hence, numerous strategies of grouping and GOF test statistics have been proposed based on these strategies of grouping. This assists to investigate that there is none with the acceptance as a reliable strategy of grouping or GOF test statistic just as in the OLS regression models (McCulloch, 2000).

1.3 Problem identification and the importance of the study

LRMs have popularity in many applications and disciplines especially when the categorical response variables are present. Assessing the GOF is considered as the first and the most important step, after fitting the LRMs, to show the adequacy of the

model to fit the observations. GOF test statistics based on the chi-square approach and the strategies of grouping are widely used as a measure of how far the fitted values using the LRMs deviate from the observed values in each of K distinct groups.

The problem of this study is derived from the fact that there are many strategies of grouping, and GOF test statistics based on these strategies of grouping have been proposed to assess the adequacy of the LRMs, but none are considered as the best strategy of grouping or the best GOF test statistic. Consequently, the importance of this study is derived from the limitations of current strategies for grouping and the existing GOF test statistics used to assess LRMs (see Sections 2.2 and 2.3). In other words, the need for this study has been motivated by these limitations of currently available strategies of grouping and the GOF test statistics that were built based on these strategies.

1.4 Objectives of the Study

This study aims to:

1) propose a new strategy of grouping based on partitional cluster analysis methods to group the elements of the data set using the range of the covariates and the range of the response variable.
2) develop two GOF test statistics of chi-square type based on the proposed strategy of grouping.
3) extend the proposed strategy of grouping and the developed GOF to assess the fit of the mixed LRM.

4) investigate the performance of the new strategy of grouping and the new GOF test statistics by conducting simulation studies, analysing clinical data sets and comparing the performance of the proposed strategy of grouping and the proposed GOF test statistics with the performance of existing strategies of grouping and existing GOF test statistics.

1.5 Organisation of thesis

This thesis consists of seven chapters. The first chapter is devoted to give a background about LRMs, GOF test, the problem of the study, objectives of the study, and the importance of the study and the organisation of this thesis. Chapter two consists of the overview of literature in the field of GOF construction. Chapter three discusses the proposed strategy of grouping, the developed GOF test statistics to assess the LRMs; the large sample distribution of the new GOF test statistics; extension of the proposed strategy of grouping and the developed GOF tests to assess the fit of the mixed LRM, and the criteria for evaluating the performance of GOF test statistics. Chapter four is devoted to describing the simulation studies. Chapter five evaluates the performance of the proposed strategy of grouping and the new GOF tests by discussing the results of conducting the simulation studies, whereas chapter six is devoted to investigate the performance of the proposed strategy of grouping and the new GOF tests through analysing clinical data sets. The conclusions and the future works are the subjects of the last chapter.

.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The overall GOF test occupies the first step in statistical modelling process. The GOF test shows whether the predicted values are accurate representative of the observed values, or whether the model fitted the observed data accurately. The GOF tests in LRMs may be classified into five groups according to modelling activities (see section 1.2.2). The first group includes the overall GOF tests of chi-square type. Two approaches have been used to construct GOF tests of chi-square type. The first approach is to assess the fitting of the observed sample data to the expected distribution. In this approach, the process of fitting the model to the data set is considered as a way of replacing a set of data values y by a set of fitted data $\hat{y}$ from a model involving a small number of parameters, P. This approach is not preferred in the LRMs for many reasons. The first reason is the estimation of the variance of categorical covariates is computationally complex and sometimes imprecise. The second reason is the GOF test statistic is completely independent of the observations and contains no information about the model fit when the number of the categories of the variable is large as in the case of presence the continuous covariates. The second approach is more complicated than the first approach which includes assessing the fitting of the models to the observed data set. In this approach, the process of fitting the model to data set is considered as a way of comparison between the observed and the expected counts in each group of the data set which is grouped into distinct groups and then employing the chi-square approach. This approach is preferred in

constructing the GOF test in the LRMs. This chapter gives an overview of the important previous works on constructing the GOF tests for these two approaches.

2.2 Constructing GOF tests for univariate variable

The chi-square GOF tests occupied a central position in statistical theory, and it is difficult to imagine another GOF test which would has the same generality of application. In the first approach, GOF tests of chi-square type are widely used as a measure of how far the observed sample data deviates from the expected distribution. The idea of constructing the GOF test of chi-square type originated from Karl Pearson in 1900 as cited in Pearson (1936) and Pearson (1922). Karl Pearson proposed chi-square test to assess the GOF in many cases. His first contribution was a GOF test of a chi-square type to assess how far the sample distribution of univariate $X = x_1, \cdots, x_n$ deviates from a completely specified probability distribution function F_0. In this GOF test, the process of fitting the model to the data is considered as a way of replacing a set of observation values x_i by a set of estimated values $\hat{x}_i$ from the model. Here, the model is represented by the probability distribution function of the population F_0. The measure of his GOF is known as the Pearson chi-square given by:

$$X^2 = \sum_{i=1}^{n} \frac{(x_i - \hat{x}_i)^2}{V(x_i)} \tag{2.1}$$

where x_i is the i^{th} observation in the sample of univariate variable, $i = 1,2, \ldots, n$, n is the number of the observations in the sample, $\hat{x}_i$ is the estimated value of the i^{th} observation in univariate variable and $V(x_i)$ is the variance of univariate variable X.

This GOF test has many disadvantages. The estimation of the variance $V(x_i)$ is computationally complex and sometimes imprecise especially in the categorical data set. There are different estimation methods (moments, maximum likelihood and least squares) which lead to different estimation of the parameters of the model, or the assumed distribution, whichever one is appropriate. This GOF test statistic is also completely independent of the observations and contains no information about the distribution fit in the situation when each observation has its individual frequency (Collett, 2003). Furthermore, GOF test statistic calculated by this method almost does not have the chi-square distribution, because the replacement of a parameter by an estimator usually increases the variability of the results (Kuss, 2002).

Karl Pearson second contribution was GOF test of a chi-square type to assess the fitting of the sample distribution of variable X which falls into G categories as shown in Table 2.1 to a specific population distribution. The purpose of this method is to construct a GOF test of chi-square type to examine the null hypothesis that the distribution of the sample $x_1, \cdots, x_G$ is equal to a completely specified distribution F_0.

Table 2.1 One way cross-classification of observations according to the categories

Categories No.	1	2	…	G
n_g	n_1	n_2	…	n_G
m_g	m_1	m_2	…	m_G

No.= number of elements in each category

He introduced the now classical chi-square test:

$$X^2 = \sum_{g=1}^{G} \frac{\left(n_g - m_g\right)^2}{m_g} \tag{2.2}$$

where n_g is defined as the observed number of the observations included in the g^{th} category, $g = 1, 2, \ldots, G$; G is the number of categories of the variable, m_g is the expected number of observations in g^{th} category calculated according to the hypothetical probability distribution function (pdf) of the sample. The X^2 is considered as a GOF criterion to assess the agreement between the observed pdf of the sample and completely specified pdf of the population. The distribution of X^2 is approximately chi-square with $G - 1$ degrees of freedom for large sample size n. The validity of this distribution, however, relies on the assumption of large n_g and the test shows unsatisfactory behavior with sparse data (Kuss, 2002). McCullagh and Nelder (1989) have shown that $X^2 = n$ in the extreme case when every individual observation has its own covariate category ($n_g \equiv 1$), where the sample size n is not a sensible measure of GOF.

The objective of the previous chi-square tests is to examine the null hypothesis stating that the frequency distribution (F) of certain events observed in a sample is consistent with a particular theoretical distribution (F_0). Fisher (1924) mentioned that it is rare to test this hypothesis, and the more common situation is to test the null hypothesis H_0 that F is a member of a certain parametric family F_θ. He established the classical Pearson-Fisher chi-square test in the case of the continuous variable. He proposed partitioning the range of the variable into K groups which is defined as $R_1, \cdots, R_K$ and he defined $p_1, \cdots, p_K$, as the probabilities of assigning the observation in each group which are functions of the true parameter of the model θ and the true

value of θ is almost known. He formed a chi-square type statistic under the condition $K > P$, as follows:

$$X^2 = \sum_{k=1}^{K} \frac{\left(O_k - n\hat{p}_k(\theta)\right)^2}{n\hat{p}_k(\theta)\left(1 - \hat{p}_k(\theta)\right)} \tag{2.3}$$

where P is the number of the variables, O_k is the observed number of observations falling into k^{th} group, $n\hat{p}_k(\theta)$ is the expected number of the observations in the k^{th} group and θ is the unknown parameters of the model estimated by using minimum chi-square method of estimation. This X^2 has asymptotically a chi-square distribution with $K - P - 1$ degrees of freedom for large sample size n. This test statistic also does not differ much from the previous test statistic. Consequently all or some of the disadvantages of the previous formula may be found in this GOF of test. In addition, Fisher's results are valid only if θ is estimated by the minimum chi-square estimator.

The previous works focused often on constructing the GOF test statistics for the numerical variables with normal distribution. Meanwhile, Armitage (1966) proposed a method of obtaining an asymptotically valid chi-square test with $G - 1$ degrees of freedom for the categorical data. He supposed that the data was divided into G categories, denoted as $g = 1, 2, \cdots, G$ and the secondary subdivision for each category is divided into K groups, denoted as $k = 1, 2, \cdots, K$. Armitage (1966) described his notations as in the following table:

.

Table 2.2 The notations used in Armitage (1966) study

Groups	Categories $1 \ldots . g \ldots . G$	Summation
1 ⋮ k ⋮ K	Failures = N_{kg} Successes $= t_{kg} - N_{kg}$ Total cases = t_{kg}	n_k $t_k - n_k$ t_k
Summation	Failures= N_g Successes= $t_g - N_g$ Total cases = t_g	n t - n t

The observed proportion in the $(k, g)^{th}$ cell is $p_{kg} = N_{kg}/t_{kg}$; where N_{kg} is the number of the failure observations in the k^{th} group of g^{th}category and t_{kg} is the total number of cases in the same group and category. The proposed test is sensitive to systematic discrepancies between N_g and $E(N_g)$ for particular categories. These discrepancies may be tested by using the proposed chi-square type GOF test statistic given by:

$$X^2 = \sum_g (N_g - e_g)^2 \left\{ \left(\frac{1}{e_g} \right) + \left(\frac{1}{t_g - e_g} \right) \right\} \qquad (2.4)$$

where $N_g = \sum_k N_{kg}$ is the observed number of failures in the g^{th} category, $t_g = \sum_k t_{kg}$ is the total observations in g^{th} category, $e_g = E(N_g) = \sum_k e_{kg}$ is the "expected" frequency which is equal to the sampling expectation and $e_{kg} = E(N_{kg}) = n_k t_{kg}/t_k$, n_k and t_k are the failures and the total observations in k^{th} group respectively. Armitage (1966) stated that this GOF test statistic did not follow, even asymptotically, the chi-square distribution with $G - 1$ degrees of freedom.

.

Several authors have also shown their interest in the general problem of testing the GOF of univariate variable that come from a specific parametric distribution family. Dahiya (1971), Dahiya and Gurland (1973), Moore (1971, 1977), Tate and Hye (1973), Spruill (1976) and Fellegi (1980) had proposed other formulas of GOF test statistics using a chi-square approach. All of these formulas are based on the original approach of chi-square proposed by Pearson but with different strategies of construction.

The distribution of this X^2 when the continuous variables are involved is examined by many authors such as Neyman and Pearson (1931), Chernoff and Lehmann (1954) and Watson (1957). They showed that if the groups probabilities $p_1,\ p_2, \dots, p_k$ are prescribed, the groups are chosen, the normal distribution was used in fitting continuous variables with pdf as $f(x;\ \theta_1, \theta_2, \cdots, \theta_k)$ and the range of x is partitioned into K groups $(-\infty,\ d_1), (d_1,\ d_2), \cdots,\ (d_{K-1},\ \infty)$, then the asymptotic chi-square distribution with $K - P - 1$ degrees of freedom of X^2 estimated as in Eq. (2.3) does not hold. They showed that under the condition of $K > P$, the statistic X^2 is asymptotically distributed as:

$$\chi^2_{K-P-1} + \lambda_1 x_1^2 + \cdots + \lambda_P x_P^2 \tag{2.5}$$

where $\lambda_1, \dots, \lambda_P$ are the roots of the determinant equation $\left|\tilde{\boldsymbol{I}} - (1 - \boldsymbol{\lambda})\hat{\boldsymbol{I}}\right| = 0$. Here $\tilde{\boldsymbol{I}}$ is the information matrix per observation for θ when it is estimated from the observed data, $\hat{\boldsymbol{I}}$ is the information matrix per observation for θ when it is estimated from the original data, all of them in the interval (0, l) and $x_1, \dots, x_P$ are standard normal variables, independent of each other. They have also remarked that the efficiency of

the GOF test statistic is affected by the replacement process because the replacement process of parameters by estimators increases the variability of the results. Several authors such as Moore and Spruill (1975) and Broffitt and Randles (1977) have also done related works on this topic, and they showed that the distribution of chi-square statistic estimated by Eq. (2.3) is bounded by the known chi-square distributions that is $\chi^2_{K-P-1} \leq X^2 \leq \chi^2_{K-1}$.

2.3 Constructing GOF tests for models

Overall the GOF test statistics can be constructed by using one of two techniques based on the chi-square approach to assess the adequacy of the models. The first technique has been used in the OLS regression models, which is based on the ungrouped observations. This technique is not preferred in the LRMs for many reasons; therefore, the alternative technique in LRMs is based on grouping the observations (see Section 1.2.2 for more details). Since 1980 much of the interest has been changed to test the GOF of the models arising in different disciplines. The reason for this new interest is that in many situations, the GLMs or specifically the LRMs have become a widely used and accepted method of analysis of binary outcome variables. This popularity comes from the availability of easily used software and the ease of interpretation of the results of the fitted model (Hosmer *et al.*, 1997).

Corresponding with this increase in applications of these models has been an increase in statistical research on the development of new methods to assess the adequacy of the fitted model. Several authors have proposed the strategies of

grouping based on either partitioning the range of the estimated probabilities of response variable or the range of the covariates or both of them or the residuals to construct GOF tests. The following subsections present the previous works according to the type of the strategy of grouping.

2.3.1 GOF tests based on partitioning the range of the estimated response variable ($\hat{\pi}_i$)

The first work which is considered as a base for this kind of works is by the Hosmer and Lemeshow (1980). They proposed a strategy of grouping to partition the range of the estimated response variable. This strategy of grouping is called *deciles of risks*. It is based on collapsing the columns of $2 \times J$ table into fixed number of groups $K = 10$ of equal sizes. The rows of the table correspond to the two values of the response variable $y = 1$ or 0 and the columns correspond to the J possible covariate patterns. According to this strategy of grouping, the n observations of the estimated probability from fitting the LRM were ranked in ascending order. Then, the first group would consist of $n/10$ observations with the smallest estimated probabilities $\hat{\pi}_i$. The second group would consist of $n/10$ observations with the next smallest estimated probabilities, and so on.

After all groups are formed, a chi-square type GOF test statistic is constructed based on the comparison between the observed and the expected number of observations in each group such as:

$$\hat{C} = \sum_{k=1}^{K} \frac{(O_k - n\bar{\pi}_k)^2}{n\bar{\pi}_k(1-\bar{\pi}_k)} \tag{2.6}$$

where $O_k = \sum_{i=1}^{n_k} y_i$ is the observed number of events/successes in the k^{th} group, g_k is the covariate patterns with m_g observations in the g^{th} covariate pattern and $\bar{\pi}_k = \sum_{g=1}^{g_k} m_g \hat{\pi}_g / n_k$ is the average of estimated probabilities for k^{th} group, $\hat{\pi}_g$ denotes the predicted probability for the g^{th} covariate pattern, n_k is the number of observations in the k^{th} group.

The second strategy of grouping proposed by Hosmer and Lemeshow in 1989 is called *fixed cut-off points,* which differs from the previous one where the range of the ranked estimated response variable $\hat{\pi}_i$ is partitioned into K groups according to prespecified *fixed cut-off points*. This strategy of grouping gives fixed groups with approximately equal sample sizes. A GOF test statistic $\hat{H}$ is constructed based on comparing the observed frequency O_k with the estimated frequency E_k in the group k as defined in the previous statistic $\hat{C}$ in Eq. (2.6). These GOF test statistics $\hat{C}$ and $\hat{H}$ are widely used for the following reasons:

1) these strategies of grouping are clear and easy to implement.
2) these GOF test statistics based on these strategies are naturally attractive and easy to compute;
3) these GOF test statistics have good properties based on the simulation studies;
4) the $\hat{C}$ GOF test statistic is widely available in computer packages.
5) lack of a better GOF test statistic also contributes to their popularity.

However, in spite of these good properties of these GOF test statistics $\hat{C}$ and $\widehat{H}$, they have the following disadvantages:

1) Kuss (2002) pointed out the fact that these strategies of grouping collect all observations with low values in a single group and high values in another group. Therefore, it might be possible that the first groups have low expected frequencies for the success events and the last groups have low expected frequencies for fault events, both facts questioning the validity of the chi-square distribution for these GOF test statistics;
2) the distribution of these GOF test statistics was derived based on the simulation study. Thus it may differ if the setting of the simulation study is different;
3) these GOF test statistics have low power to detect specific types of lack of fit (such as nonlinearity in an explanatory variable);
4) the scale of partitioning in the strategy of grouping based on cut-off point technique is chosen subjectively. Thus the values of $\widehat{H}$ GOF test statistic are affected by this choice.
5) when the number of groups is 5 or less, these GOF test statistics become not sensitive to the fitting of the model; they will almost indicate that the model fits the data well, because the estimated variance of $\hat{\beta}$ may become unreliable since there are few degrees of freedom for the estimate (Archer *et al.*, 2007).

These disadvantages of Hosmer and Lemeshow (1980, 1989) GOF tests motivated researchers to improve these strategies and their GOF test statistics or adapt these GOF tests to assess other kind of LRMs. Lipstiz *et al.* (1996) extended the Hosmer and Lemeshow (1980) GOF test statistic to assess the fitting of the ordinal

LRM. They proposed a strategy of grouping based on assigning a score to the response variable categories to construct the proposed GOF test statistic. They considered the multinomial response variable, since the response variable is categorical, where the i^{th} individual's response ($i = 1, 2, \ldots, n$) falls into one of G possible categories ($g = 1, \ldots, G$). They defined the indicator random variable $Y_{ig} = 1$ if the i^{th} individual has response g and equal 0 otherwise, with $\sum_{g=1}^{G} Y_{ig} = 1$. They proposed grouping the Y_{ig}'s together to form the response vector $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iG})'$ and defined the probability of response g as $\pi_{ig} = E(Y_{ig})$, with $\sum_{g=1}^{G} \pi_{ig} = 1$. Consequently, the random vector $\boldsymbol{Y_i}$ has a multinomial distribution with probability vector $\boldsymbol{\pi_i} = E(\boldsymbol{Y}_i) = (\pi_{i1}, \ldots, \pi_{iG})'$. Finally, each individual is assumed to have a $P \times 1$ covariate vector $\mathbf{x}_i = (\mathrm{x}_{i1}, \ldots, \mathrm{x}_{iP})'$. There are many choices of link function relating the elements of $\boldsymbol{\pi}_i$ to the covariates. The general form is :

$$L_{ig} = L_{ig}(\boldsymbol{\pi}_i) = \alpha_g + \mathbf{x_i'}\boldsymbol{\beta}_1 \tag{2.7}$$

where L_{ig} is the link function include the 'cumulative' logit, the 'continuation ratio' logit and the 'adjacent categories' logit, α_g and $\boldsymbol{\beta}_1$ are the unknown parameters of the model and whatever the choice of link function, for all these models $\boldsymbol{\pi}_i = \boldsymbol{\pi_i}(\boldsymbol{\beta})$, where $\boldsymbol{\beta}' = (\boldsymbol{\alpha}', \boldsymbol{\beta}_1')$. Lipstiz *et al.* (1996) suggested the following steps to construct the proposed GOF test statistic appropriate for any of the ordinal regression models:

1) choose one of the following methods of scoring to assign the score s_g to the response variable of category *g*:
   a) using equally spaced (integer) scores such as $s_g = g,\ g = 1, 2, \ldots, G$ or
   b) assume $s_1 = 1,\ s_2 = \cdots = s_G = 0$, where $\boldsymbol{s} = (s_1, \cdots, s_G)'$.

2) three types of score for the i^{th} response variable where the observed score is $Z_i = \sum_{g=1}^{G} s_g Y_{ig}$. The mean score is $\mu_i = \mu_i(\boldsymbol{\beta}) = E(Z_i) = \sum_{g=1}^{G} s_g \pi_{ig}$, and the predicted mean score is $\hat{\mu}_i = \sum_{g=1}^{G} s_g \hat{\pi}_{ig}$.

3) follow Hosmer and Lemeshow (1980) in using the percentiles to partition the predicted scores $\hat{\mu}_i$ into $K = 10$ groups. The groups are formed according to the tenths of the predicted scores of equal size, where the first group contains $n/10$ the smallest predicted scores and the last group has $n/10$ largest predicted scores.

4) define the indicator variable $I_{ik} = 1$ if $\hat{\mu}_i$ is in group k, and 0 otherwise according to the partitioning strategy above, where $k = 1, \ldots, K-1$.

5) consider that the grouping process will add a random effect to the model, then, the alternative model:

$$L_{ig} = L_{ig}(\boldsymbol{\pi}_i) = \alpha_g + \mathbf{x_i'}\boldsymbol{\beta_1} + \sum_{k=1}^{K-1} \gamma_k I_{ik} \tag{2.8}$$

is considered to assess the GOF of model in Eq. (2.7), where γ_k is the unknown parameter of the grouping effect. The $\mathbf{x_i'}\boldsymbol{\beta_1}$ is the covariates effect and $\sum_{k=1}^{K-1} \gamma_k I_{ik}$ is the grouping effect.

6) propose the likelihood ratio, Wald or score statistic to assess the GOF of the model in Eq. (2.7). One of these GOF tests is used to examine the hypothesis that the grouping effects are not significant $H_0: \gamma_1 = \gamma_2 = \cdots = \gamma_{K-1} = 0$. If this hypothesis is accepted, then the model in Eq. (2.7) is correctly specified, otherwise the model is incorrectly specified.

All or some properties of the Hosmer and Lemeshow (1980) GOF test statistic may be found in Lipstiz *et al.* (1996) GOF test statistic. Furthermore, the GOF test in this setting do not test the GOF of the full model, but test only the significance of the effect of the grouping process.

The above works proposed GOF tests to examine the adequacy of the LRM for the random sample. Alternatively, Graubard *et al.* (1997) proposed GOF test to examine the adequacy of LRM when it was fitted for a multistage, a stratified or a cluster samples, with different sampling weight w_i due to selection probabilities. They proposed grouping strategy for establishing deciles of risk for the estimated probabilities $\hat{\pi}(\mathbf{x}_i)$ from fitting the LRM to these samples. The maximum likelihood estimation for $\hat{\pi}(\mathbf{x}_i)$ would not be valid for fitting the LRM for these kinds of sampling. Instead, Graubard *et al.* (1997) used pseudo-maximum likelihood method to estimate the parameters $\boldsymbol{\beta}$*'s* from the LRM in the case of the complex sampling which is given by:

$$L = \prod_{i=1}^{n} \pi_i^{w_i y_i} (1 - \pi_i)^{w_i (1 - y_i)} \tag{2.9}$$

where w_i are the sample weights and L is the likelihood function. The weighted estimates $\widehat{\boldsymbol{\beta}}'s$ are the values of $\boldsymbol{\beta}'s$ maximise the function as in Eq. (2.9). They substitute these $\widehat{\boldsymbol{\beta}}'s$ into the model as in Eq. (1.1) to obtain the weighted estimated probabilities $\hat{\pi}_i$'s. Then these weighted estimated probabilities $\hat{\pi}_i$ are divided into weighted deciles, which have a weighted of one-tenth of the n observations in the data set in each group.

Graubard *et al.* (1997) defined the number of weighted outcomes in the k^{th} decile as $O_k = \sum_{i=1}^{n_k} w_{ki} y_{ki}$ and the number of weighted expected outcomes as $e_k = \sum_{i}^{n_k} w_{ki} \hat{\pi}_{ki}$, where, w_{ki} is the sample weight for the i^{th} observation in the k^{th} decile of risk. Finally, they suggested constructing a Wald form test statistic as:

$$W_d = (\boldsymbol{O} - \boldsymbol{E})' \boldsymbol{V_d^{-1}} (\boldsymbol{O} - \boldsymbol{E}) \tag{2.10}$$