

**PREDICTIVE ACCURACY OF FITTED LOGISTIC REGRESSION
MODEL USING RANKED SET SAMPLES**

by

Kevin Carl Peña Santos

A thesis submitted in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE (STATISTICS)

School of Statistics, University of the Philippines
Diliman, Quezon City

March 2011

ABSTRACT

Separation of likelihood and rare events are two interrelated problems in fitting the logistic regression model. We propose to resolve this problem by drawing the sample using ranked set sampling. An extensive simulation study was conducted to assess the performance of a logistic regression model fitted from ranked set samples and compared to those estimates using simple random samples. RSS performs best in small populations regardless of the distribution of the binary response variable in the population. As the sample and population sizes increase, the predictive ability under RSS also improves but it stabilizes to become comparable to SRS. Furthermore, RSS mitigates the problem of separation of likelihood especially when the population size is relatively large. For small population, it is advisable to obtain at least 5% sampling rate to avoid the said problem. RSS also performed better than SRS with 1% sampling rate. Hence, RSS provides high predictive accuracy and keeps costs at low levels.

The predictive accuracy of the fitted model using RSS is robust to omitted variable problem and model inadequacy as long as there is an available concomitant variable whose correlation with the binary response variable is strong enough.

Keywords: logistic regression model, ranked set sampling, rare events, separation of likelihood