

Categorical Data: Comparing SAS® and R for Logistic and Poisson Regression

Lillian Chow, San Francisco State University, San Francisco, California

ABSTRACT

Categorical data is some of the most common and available data we encounter on a day to day basis. Data that can be divided into groups or levels, for example race, gender, customer satisfaction and party affiliation, can all be considered categorical data. Though these data are extremely common, analyzing and interpreting categorical data can often be complex and elusive. This paper seeks to briefly summarize the different techniques used to test, interpret, and analyze categorical data using Logistic and Poisson regression, as well as explore methods of coding these techniques using SAS® and comparing it with similar code in R software. To illustrate the different techniques used for model checking and selection, interpretation of the coefficients, and identifying and resolving issues, I will use SAS/IML® software, focusing on the LOGISTIC and GENMOD procedures and corresponding options as well as PROC IML to call various R functions within SAS®.

INTRODUCTION

LOGISTIC REGRESSION

INTERPRETATION OF MODEL

Odds Ratio/ Logit Interpretation

Confidence Intervals

TESTING/ INFERENCE OF PARAMETERS

Wald Test/ Confidence Interval

Likelihood-Ratio

MODEL CHECKING AND SELECTION

Goodness of Fit (deviance)

Checking Model Residuals

Comparing AIC

IDENTIFYING AND RESOLVING ISSUES

POISSON REGRESSION

INTERPRETATION OF MODEL

Odds Ratio

Confidence Intervals

TESTING COEFFICIENT SIGNIFICANCE

Wald Test/ Confidence Interval

Likelihood Ratio Test

MODEL CHECKING AND SELECTION

Goodness of Fit

Checking Model Residuals

Comparing AIC

IDENTIFYING AND RESOLVING ISSUES

CONCLUSION