



SEMINAR SERIES 2018-2019

When: Thursday, October 4, 2018
Time: 3:30 – 4:30 pm
Refreshments at 3:15 pm
Where: Sidney Smith Hall Rm TBA
Speaker: Heping Zhang, Yale University
Host: Dehan Kong

Residuals and Diagnostics for Generalized Linear Models

Ordinal outcomes are common in scientific research and everyday practice, and we often rely on regression models to make inference. A long-standing problem with such regression analyses is the lack of effective diagnostic tools for validating model assumptions.

The difficulty arises from the fact that an ordinal variable has discrete values that are labeled with, but not, numerical values. The values merely represent ordered categories. In this paper, we propose a surrogate approach to defining residuals for an ordinal outcome Y . The idea is to define a continuous variable S as a "surrogate" of Y and then obtain residuals based on S . For the general class of cumulative link regression models, we study the residual's theoretical and graphical properties. We show that the residual has null properties similar to those of the common residuals for continuous outcomes. Our numerical studies demonstrate that the residual has power to detect misspecification with respect to 1) mean structures; 2) link functions; 3) heteroscedasticity; 4) proportionality; and 5) mixed populations.

The proposed residual also enables us to develop numeric measures for goodness-of-fit using classical distance notions. Our results suggest that compared to a previously defined residual, our residual can reveal deeper insights into model diagnostics. We stress that this work focuses on residual analysis, rather than hypothesis testing. The latter has limited utility as it only provides a single p-value, whereas our residual can reveal what components of the model are misspecified and advise how to make improvements.

This is a joint work with Dungang Liu, University of Cincinnati Lindner College of Business.