ARTICLE TEMPLATE

Taylor & Francis Rmarkdown template for authors (LaTeX-based Interact layout + Chicago author-date reference style)

Weihao Li[a], John Smith[b], Dominik Leutnant[c,†,‡]

[a]Taylor & Francis, 4 Park Square, Milton Park, Abingdon, UK; [b]Institut für Informatik, Albert-Ludwigs-Universität, Freiburg, Germany; [c]Muenster University of Applied Sciences - Institute for Infrastructure, Water, Resources, Environment, Correnstr. 25, 48149 Münster, Germany

ARTICLE HISTORY
Compiled April 20, 2022

ABSTRACT
This template is for authors who are preparing a manuscript for a Taylor & Francis journal using the LaTeX document preparation system and the 'interact} class file, which is available via selected journals' home pages on the Taylor & Francis website.

KEYWORDS
Sections; lists; figures; tables; mathematics; fonts; references; appendices

problem: residual plot diagnostics conventional test: too sensitive
background:

1. residual plot for model diagnostics

a. residual is widely used
b. what are the types of residual plots
c. comparison

2. conventional test: F, BP
3. visual test: lineup, theory

desc of experiment: 1. simulation setup 2. experimental design 3. result

comparison of conventional tests: 1. power (visual test vs. conventional test) (visual test most different one (everything test, any departure)) 2. investigate the difference (gap), give examples 3. conventional is too sensitive 4. make conventional less sensitive (vary alpha)

conclusion: 1. too sensitive, visual test is needed/preferable 2. visual test is infeasible in large scale (expensive) 3. future work (role of computer vision)

Title:

CONTACT Weihao Li. Email: weihao.li@monash.edu, John Smith. Email: john.smith@uni-freiburg.de, Dominik Leutnant. Email: leutnant@fh-muenster.de

1. Introduction

Diagnostics of the classical normal linear regression model conventionally involve evaluating the fitness of the proposed model, detecting the presence of influential observations and outliers, checking the validity of model assumptions and many more. Tools such as summary statistics, hypothesis testing, and data plots are essential for a systematic and detailed examination of the regression model (Mansfield and Conerly 1987).

Graphical summaries in which residuals are plotted against fitted values or other functions of the predictor variables that are approximately orthogonal to residuals are refereed to as standard residual plots. They are commonly used to identify patterns which are indicative of nonconstant error variance or nonlinearity (Cook and Weisberg 1982). Raw residuals and studentized residuals are the two most frequently used residuals in standard residual plots. The debt on which type of residuals should be used always present. While raw residuals are the most common output of computer regression software package, by applying a scaling factor, the ability of revealing nonconstant error variance in standard residual plots will often be enhanced by studentized residuals in small sample size (Gunst and Mason 2018).

As a two-dimensional representation of a model in a p-dimensional space, standard residual plots project data points onto the variable of the horizontal axis, which is a vector in p-dimensional space. Observations with the same projection will be treated as equivalent as they have the same position of the abscissa. Therefore, standard residual plots are often useful in revealing model inadequacies in the direction of the variable of the horizontal axis, but could be inadequate for detecting patterns in other directions, especially in those perpendicular to the variable of the horizontal axis. Hence, in practice, multiple standard residual plots with different horizontal axes will be examined.

Overlapping data points is a general issue in scatter plots not limited to standard residual plots, which often makes plots difficult to interpret because visual patterns are concealed. Thus, for relatively large sample size, Cleveland and Kleiner (1975) suggests the use of robust reference lines to give aids to eye in seeing patterns, which nowadays, are usually replaced with a spline or local polynomial regression line.

One of the useful supplements to standard residual plots is partial residual plot. While the standard residual plots indicate deviations of each predictor from linearity, the partial residual plots can be used to access the extent and the direction of the linearity. These plots can therefore be used to determine the importance of each predictor in the presence of all the others and assess the extent of nonlinearity in a given predictor. This is in addition to providing information on correct transformations and extreme data points.

References

Cleveland, William S, and Beat Kleiner. 1975. "A graphical technique for enhancing scatterplots with moving statistics." *Technometrics* 17 (4): 447–454.

Cook, R Dennis, and Sanford Weisberg. 1982. *Residuals and influence in regression*. New York: Chapman and Hall.

Gunst, Richard F, and Robert L Mason. 2018. *Regression analysis and its application: a data-oriented approach*. CRC Press.

Mansfield, Edward R, and Michael D Conerly. 1987. "Diagnostic value of residual and partial

residual plots." *The American Statistician* 41 (2): 107–116.