Student’s Name: Yuxiong Xie

Instructor’s Name: Esther Namubiru

Class and Section: EAP 506 SECTION P03

Due Date: 11/17/2015

**Article Analysis**

Academic articles are the essential of academic writing; they express authors’ ideas and researches. In statistical science, academic writing has a specific style that can be differentiated from other disciplines. As a graduate student of statistics, I had to read lots of professional articles and found some special features and unique types used in academic writing, such as structure, citation style and usage of tables and graphs. Thus, in order to learn these features of academic article in statistical field, I focused on the analysis of an article called “Does data splitting improve prediction?”, which was wrote by Julian J. Faraway (2014) and published in the journal of “Statistics and Computing”.

In this article, Faraway (2014) examined the differences between the full data strategy and the data splitting strategy for prediction and how data splitting influence predictions. He surveyed and divided data into two parts, and discussed four methods – Model validation, Scoring, Decomposition of performance and Estimators – to evaluate data splitting’s effectiveness. However, there are lots of preview researches regarding data prediction with the approach of integrating the model selection and parameter estimation, but lacking of studies that overcomes the drawbacks such as uncertainty in predictions, less data to select models and estimate the parameters. So the purpose of this article was to fill these gaps and use data splitting methods to improve prediction. He had a driving question; it was, “Will the gain from avoiding data re-use compensate for the loss in model selection and parameter estimation performance?” (p.2). Faraway (2014) also highlighted many researchers in his article. In particular, he connected his research to other studies on data splitting, such as Stone (1974), Dawid (1984), Meng and Xie (2013). Finally, his study is relevant to future studies on “attempting recommendations for other types of response or data of a more structured form” (Faraway, 2014, p.11).

According to Swales and Feak (2012, p.331), there are three moves defined by them in introduction and abstract. The first move is identified when the author introduces and reviews the previous research regarding data splitting methods (Swales and Feak, 2012, p.331). For example, Faraway (2014) described, “Stone (1974) provides an early history of data splitting. Dawid (1984) discusses the related but different issue of updating models as new data arrive sequentially” (p.2). The second move is noted when the author points out a gap in the current research (Swales and Feak, 2012, p.331). In this article, Faraway (2014) evaluated some methods to judge predictions in previous studies and found out there was lacking of studies on addressing uncertainty in predictions, so he referred to an extend research by starting that, “Now suppose we are required to express the uncertainty in our predictions, perhaps in the form of a prediction interval or more generally as a predictive distribution” (p.1). Lastly, this article explained the purpose of the research which is an example of the third move. For instance, Faraway (2014) used a question to indicate the aims, which was described as that, “Will the gain from avoiding data re-use compensate for the loss in model selection and parameter estimation performance? An answer to this question is the purpose of this paper.” (p.2). Here I will mention more important features that this article has.

The article is organized by following introduction, methods, simulations and conclusion pattern which is little different with IMRD pattern. The same between these two patterns is that article has included introduction and methods, but the difference is that this article replaced results and discussion parts by simulations and conclusion parts. In the introduction, Faraway (2014) described the information about how important it is to judge predictions, the solution to improve prediction by using data splitting methods and the purpose of this article. In next part, he discussed the usage of data splitting in a variety of fields and evaluated the effectiveness in the methods, such as Model validation, Scoring, Decomposition of performance and Estimators. In addition, the extra part of simulations is to analyze data splitting’s value and application (Faraway, 2014, p.2). The author expressed the importance and reason of using data splitting in four simulations, which make audiences to understand the value of this methods and support the idea – “data splitting can be a simple solution to solve uncertain predictions”. Finally, conclusion part contains the result and discussion about the investigation. What else, the article also contains titles and subtitles, which help audience to easily understand the contents which the author wants to explain. For example, in the third part, Faraway (2014) showed the title as “Simulations”, and the subtitles divided into four parts that describes the data splitting in different aspects, such as Box-Cox, Variable Selection, Outliers and Binary Response.

Moreover, this article presented the researches of literature review. Faraway (2014) described the four methods based on the theories in previous studies; while he analyzed data splitting in four simulations with the existing data. What is more, when the author presented new topics and ideas, he combined lots of graphs and tables to express the information. For instance, Faraway (2014) showed figures about “scoring performance for two predictive distributions compared for observed A and B” to explain the methods of Scoring (p.3). Further, the citation style which the author used is APA style. Faraway (2014) used the contributions of other authors by paraphrasing and citing with their names, years and pages, which makes the article clean and concise without many directly quotation marks. For instance, Faraway (2014) cited the idea of Meng and Xie (2013) by paraphrasing, like that “… should improve the model but Meng and Xie (2013) show that this is not necessarily ture…” (p.2). In the APA style, the most important feature is that it should include the year which the article published and the pages which the authors cited in order to make audience to know the details about the cited article.

In conclusion, analyzing this article is very helpful for my academic writing. As this article is the typical example of academic writing in Statistical Science, it indicates many features which I can follow, such as following IMRD patterns, using graphs and tables, APA citation style. Furthermore, the article analysis also helps me understand how to choose academic papers and broad my academic views in the field. Therefore, it will be easier for me to write my future papers and to do investigations.

**Reference**

Swales, J. M. & Feak, C. B. (2012). *Academic writing for graduate students, 3rd edition: essential skills and tasks*. Michigan ELT: University of Michigan Press.

Faraway, J. J. (2014). Does data splitting improve prediction?. *Statistics and Computing, 24,* 1-12. doi: 10.1007/s11222-014-9522-9