**BIOST 2049**

**Spring 2023**

**Research Article Dissection 4 – Logistic Regression**

**Due 3/15/23 by 11:59 via the Assignment Link posted to Canvas**

Please read the article “Predicting research use in a public health policy environment: results of a logistic regression analysis” by Zardo and Collie found on Canvas and answer the questions below.

This is an individual assignment. You can discuss the article with your classmates, but the answers you submit should be your own.

Please be brief but write in complete sentences with appropriate interpretations (a yes or no answer will not get full credit). A few sentences is all that is necessary to answer each question.

These questions also contain some review from Module 3 as it relates to logistic regression.

1. What is the goal of this paper?

Inform the design of interventions that increase research use in public health policy environments.

1. To assess the reliability of the survey used in this study, the authors used a split-half method. Is this similar to something we talked about in Module 3, Topic 9? Explain.

Split-half reliability involves correlating the results of one half of a survey with the other. This is like k-fold cross-validation, where model performance is evaluated across multiple folds of the dataset and consistent performance across folds increases the precision of our beliefs about the model performance, just as higher split-half reliability increases the reliability of the internal consistency or “measure performance”.

1. Briefly describe the statistical methods (not the survey testing methods).

They used single variable logistic regression and multiple variable logistic regression. Logistic regression models are a kind of generalized linear model.

1. What type of model building strategy did the authors use? Would you suggest a different method based off of what we learned in Module 3?

They first checked for significant covariance between research use and each of their 49 collected predictive factors. They then put significant predictors from this first stage analysis into a single model and tested for which factors contributed significantly to this single model.

1. On page 4 of the paper, in the results section, the authors say this “The model overall explained between 31.2% (Cox and Snell R square) and 42.4% (Nagelkerke R square) of the variance in research use.” Is this an appropriate interpretation? Why or why not.

Cox and Snell R square and Nagelkerke R square do not represent the proportion of variance explained in research use by the model. These are goodness of fit values. Thus, the authors’ interpretation is incorrect.

1. In table 2, interpret the beta coefficients and ORs for the variables 1. Skills for use of research and 2. Intention to use research in the next 12 months.

Skills for use of research: the beta coefficients indicate higher log odds of research use when medium or high-very high skills are present compared to when they are not. The odds ratios are the beta coefficients exponentiated. Values greater than 1 indicate higher odds when the predictors are present vs. when they are not

Intention to use research: the beta coefficient for yes implies a higher log odds of research use when intentions to use research in the next 12 months is present compared to when it is not. The odds ratio is the exponentiated beta coefficient. The odds ratio is greater than 1, indicating higher odds of research use when intention is present compared to when it is not.

1. Do you think there are analyses missing that should have been performed?

They should have used cross-validation to obtain confidence estimates in their model performance and to approximate the predictive performance of their models on new data.