stats

MDPI

Case Report

Bayesian Model Averaging and Regularized Regression as Methods for Data-Driven Model Exploration, with Practical Considerations

Hyemin Han

Educational Psychology Program, University of Alabama, Tuscaloosa, AL 35487, USA; hyemin.han@ua.edu; Tel.: +1-205-348-0746

Abstract: Methodological experts suggest that psychological and educational researchers should employ appropriate methods for data-driven model exploration, such as Bayesian Model Averaging and regularized regression, instead of conventional hypothesis-driven testing, if they want to explore the best prediction model. I intend to discuss practical considerations regarding data-driven methods for end-user researchers without sufficient expertise in quantitative methods. I tested three data-driven methods, i.e., Bayesian Model Averaging, LASSO as a form of regularized regression, and stepwise regression, with datasets in psychology and education. I compared their performance in terms of cross-validity indicating robustness against overfitting across different conditions. I employed functionalities widely available via *R* with default settings to provide information relevant to end users without advanced statistical knowledge. The results demonstrated that LASSO showed the best performance and Bayesian Model Averaging outperformed stepwise regression when there were many candidate predictors to explore. Based on these findings, I discussed appropriately using the data-driven model exploration methods across different situations from laypeople's perspectives.

Keywords: data-driven analysis; model exploration; variable selection; Bayesian Model Averaging; regularized regression; LASSO; stepwise regression; cross-validation; overfitting

check for updates

Citation: Han, H. Bayesian Model Averaging and Regularized Regression as Methods for Data-Driven Model Exploration, with Practical Considerations. *Stats* **2024**, *7*, 732–744. https://doi.org/10.3390/stats7030044

Academic Editor: Wei Zhu

Received: 24 June 2024
Revised: 11 July 2024
Accepted: 12 July 2024
Published: 18 July 2024

Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Researchers in psychology and educational research who use quantitative methods are often interested in exploring the best prediction or regression models with data-driven analysis methods [1]. Previously, many of them tended to test a priori hypotheses based on existing theories and literature via hypothesis-driven analysis [2,3]. However, recent advances in computer and data science allow researchers to gather large-scale datasets relevant to their research topics in a feasible manner. The open science movement enables them to access various large-size open datasets for secondary analysis [4]. In such a situation, the data-driven model exploration that researchers could not easily perform in the past has become a powerful research method to develop the most plausible explanation model based on data and to generate research questions and hypotheses for subsequent research projects [1,5,6].

Despite the potential benefit of the data-driven approach in quantitative research, researchers need help implementing such an approach in their research projects. One of the most fundamental challenges is that conventional hypothesis-driven analysis, such as testing one model based on statistical indicators, e.g., *p*-values, that most quantitative researchers are familiar with, is unsuitable for achieving the goal [7–9]. As mentioned, the intended use of conventional hypothesis testing is testing on a specific model rather than comparing the aptitude of multiple candidate models for model exploration [2,3]. Let us imagine an illustrative situation when one mistakenly employs the hypothesis-driven analysis method to identify the best regression model from a collected dataset. One may eventually test one specific model, the full model with all candidate predictors in the dataset.

Then, one may identify which predictor should be in the resultant model by examining whether the responsive p-value is lower than 0.05. In such a case, the resultant model is likely erroneous primarily due to inflated false positives and overestimating model parameters, e.g., coefficients [10,11]. Overfitting is another significant issue [5,8,12,13]. When researchers employ a full model, the model may include unnecessary candidate predictors not fundamentally associated with the dependent variable. In such a case, an apparent R^2 value might become higher. However, the full model may be overfitted to the data used for regression. Thus, it cannot accurately predict the dependent variable outside the used dataset. That results in suboptimal cross-validity. More specifically, in such a case, the estimated regression model might be statistically valid in the interval defined by the values and data in the employed dataset. However, outside the intervals, it may perform an extrapolation without statistically significant guarantee of the estimates. The abovementioned illustrative example suggests researchers should use analysis methods developed for data-driven model exploration rather than conventional hypothesis testing in such a case.

Furthermore, we should also consider the examination of the relative importance of predictors as a purpose for regression-based analysis in psychological and educational studies [14]. Researchers in psychology and education are frequently interested in the extent to which each predictor contributes to predicting the outcome variable of interest [14]. Although conventional indicators, such as p-values and R^2, have been employed to examine the relative importance of variables, they do not demonstrate the importance accurately [15]. I already explained previously the reason why p-values cannot be used for variable selection or evaluation [10,11]. In the case of R^2, due to the intercorrelation between predictors and multicollinearity, it is practically difficult to calculate the pure contribution of each predictor for evaluating its relative importance [14,16,17]. Hence, methods providing accurate information about the true relative importance of each variable after considering the intercorrelation and multicollinearity, such as dominance analysis and relative weight calculation [16,18], might be beneficial for quantitative studies in psychology and education [14]. Although their primary goal is not exploring best candidate models and predictors unlike data-driven exploration methods, their outcomes, relative importance, can offer some insights into which predictors should be seriously considered in explaining the association between predictors and the outcome variable in regression models of interest.

Hence, it would be informative to briefly overview and discuss methods to examine relative importance within the context of the current paper focusing on data-driven model exploration. Although data-driven model exploration methods pursue a different goal—identifying the best model and predictors—the data-driven methods can provide better information regarding the relative importance of variables compared with conventional hypothesis testing methods, particularly testing p-values [14,15]. For instance, it would be possible to determine that predictors which survived data-driven exploration may be more important in predicting an outcome variable than others [19]. I will discuss implications of data-driven methods within the context of examining relative importance in the discussion section after overviewing exploration methods and testing results.

In the following subsection, I will briefly overview several data-driven methods widely used in the field for background information.

1.1. Background

According to recent studies in psychology and education performing model exploration with data-driven analysis methods, the most accessible methods available in widely-used statistical software, such as *R*, are stepwise regression, Bayesian Model Averaging (*BMA*), and regularized regression [5,20–24]. Let us begin with overviewing the first method, stepwise regression. *R* implements the stepwise regression as a native statistical function, i.e., *step*. When a researcher conducts stepwise regression, there are different modes of model selection [25,26]. First, forward stepwise regression starts with a null

model and adds candidate predictors step-wisely until the process fulfills a criterion. Second, when one performs backward stepwise regression, one begins with a full model and removes one predictor per step until a criterion is satisfied. In the case of *step* in *R* [27], the default option employs both stepwise methods while examining the Akaike Information Criterion (AIC) as an indicator for each step.

Compared with the conventional approach, testing a full model and selecting predictors using *p*-values [25], stepwise regression is a better model exploration method as it employs a data-driven approach in its rationale. Furthermore, it is straightforward to use even without knowledge of advanced statistics, e.g., Bayesian statistics. Because stepwise regression suggests only one model at the end [27], from a psychological or educational researcher, who is often interested in clearly identifying factors contributing to predicting their dependent variable of interest, it is easy to understand and interpret the result. However, there are also limitations warranting our caution. Depending on which method (e.g., forward vs. backward vs. both) or criterion (e.g., AIC, *p*-values, etc.) is employed, stepwise regression may recommend different models across trials [28,29]. Given it selects only one model among all possible candidate models, model uncertainty becomes a significant concern. Because researchers can only collect sample data from a part of the reality—the population, in most cases, depending on which subset is chosen during the sampling and data collection processes—the model suggested by stepwise regression can significantly vary. Although it is possible to alleviate overfitting, as the final model does not include all predictors, some methodology researchers still criticize stepwise regression on the grounds that it can overestimate model parameters [30,31].

Second, *BMA* employs Bayesian statistics to average multiple candidate models with high posterior probabilities (see the Supplementary Materials for theoretical details) [9,32]. In *R*, the *BMA* package implements the functionality [33]. *BMA* utilizes Bayesian statistics to calculate the posterior probabilities indicating the likelihood of the inclusions of candidate predictors and models. It starts with assigning prior probabilities, the probabilities of the abovementioned predictor or model inclusions, before observing the data [2,3,34]. By default, each predictor and model has an equal chance of inclusion before data observation. The *BMA* procedure updates the prior distributions into the posterior distributions while observing collected data following the Bayes Theorem [34]. If the examined data suggests the inclusion of a specific predictor or model is more likely than the prior probability before observation, then the posterior probability becomes higher. The opposite can also occur if data does not support the prior probability. Once the procedure is completed, *BMA* calculates the mean coefficient of each predictor based on the posterior probabilities and the estimated coefficients in candidate models [33,35]. The posterior probability of a specific model works as a weight during the calculation process. If one model demonstrates a high posterior probability, the coefficients of predictors included in that model highly contribute to the averaged coefficients, as the model weight is high. Likewise, coefficients in a model with a low posterior probability contribute relatively less to the mean coefficients in the final result.

Compared with the previously introduced methods, *BMA* has several benefits. First, given it considers multiple candidate models, not one single model, with posterior probabilities, *BMA* is robust against model uncertainty [9,29,36]. It is less susceptible to uncertainty and overfitting because it does not choose one specific model. The previous studies reporting the superior cross-validity of *BMA* compared with conventional model exploration methods support such a point [20,22]. Moreover, since it implements Bayesian statistics, *BMA* can suggest which model is more plausible than others via posterior probabilities [2,3,34]. However, there are issues that researchers need to consider regarding this method. Unlike the conventional methods, *BMA* does not suggest one single model without clear indicators for inference, such as *p*-values of coefficients [9,20]. Although such a feature can provide methodological benefits as mentioned above, from the perspectives of researchers in psychology and education whose primary interest is not about methodological advances, the results from *BMA* might be difficult to interpret. This can be particularly

problematic if the researchers want to identify one specific model for explanation rather than prediction [37]. If they do not have background knowledge in Bayesian statistics, interpreting *BMA* results can be challenging.

Third, the regularized regression applies regularization processes to penalize unnecessary candidate predictor coefficients so that they shrink toward zero and minimize possible overfitting (see the Supplementary Materials for theoretical details) [38]. An *R* package, *glmnet*, is one of the most widely used packages implementing regularized regression [39,40]. Regularized regression attempts to penalize unnecessary coefficients by unfavoring models with higher coefficients. At the same time, it also considers the log-likelihood to prefer a model maintaining predictability [5,6,38]. Consequently, regularized regression suggests a model with reasonable predictability with its unnecessary coefficients approximated to zero [8,41]. Among various forms of regularized regression, the most frequently used form is the least absolute shrinkage and selection operator (LASSO), which considers the sum of the absolute values of coefficients [42]. LASSO attempts to minimize the sum value to penalize unnecessary coefficients [8].

The most distinctive benefit of LASSO and regularized regression is its robustness against overfitting, like the case of *BMA* [5]. Previous studies in psychology using LASSO (and similar regularized regression) have demonstrated that models selected via LASSO were more robust against overfitting in terms of better cross-validity compared with full models and models suggested by conventional methods [5,8,43]. The recommended models were more stringent accordingly. Despite these methodological advantages in model exploration, LASSO has several practical issues that should be considered by researchers in psychology and education. Like *BMA*, its result is more complicated to understand and interpret than the result of conventional regression when inference, not prediction, becomes a primary interest [7,8,37]. Regularized regression generates a model with regularized coefficients without other indicators for statistical inference, such as *p*-values. Even compared with *BMA*, which provides posterior probabilities [9,20], it might be less suitable for inference and explanation. Hence, psychological and educational researchers interested in an explanation model rather than prediction may experience practical difficulties while employing this method.

As overviewed, data-driven model exploration methods, including BMA and LASSO (and other forms of regularized regression), provide significant methodological benefits when researchers are interested in exploring the best model with data instead of testing a priori hypotheses [1,7]. I especially note that they are relatively robust against overfitting compared with conventional full-model testing [20,30,43]. Hence, when researchers can only collect sample, not population, data from a part of the whole group of interest, the data-driven methods should be standard approaches [9,44]. The methodological benefit becomes maximized when the ratio of observations to the number of candidate predictors in the data is small [9]. When researchers are interested in exploring data with multiple competing candidate predictors, particularly when the sample size is small, the data-driven methods would be the only valid means to achieve their goal.

1.2. Current Study

Despite the abovementioned methodological benefits of data-driven methods, several issues warrant further examination that I intend to conduct in this paper. First, most methodological papers (e.g., simulation papers) reporting the superior performance of the methods are authored by methodology experts (e.g., Wang et al. (2004) [45]). From the perspective of end-user researchers with insufficient statistical knowledge, it may be difficult to understand and test such benefits within the context of their research projects. Second, the previous papers have focused on introducing specific methodologies and reporting their performance from the methodological experts' perspectives. Hence, they could not provide sufficient concrete information about how their methods can improve end users' research outcomes. Also, the lack of practical guidelines about choosing a specific method given the purpose of a research project and the nature of collected data is

concerning. End-user researchers may be interested in such points explained and discussed from a practical point of view.

To address the abovementioned issues in applying data-driven methods, especially the insufficiency of practical information and guidelines for end-user researchers, I will examine the performance of the data-driven model exploration methods with data collected from studies in psychology and education, not simulated data. I will test stepwise regression, *BMA*, and LASSO as regularized regression with three datasets. I will employ representative R functionalities, i.e., *step*, *BMA*, and *glmnet*, with default settings for the tests to provide more practical insights to researchers who start with default functionalities without advanced statistical expertise. The performance of each method will be evaluated by cross-validation accuracy to examine the proneness towards overfitting. Moreover, I will compare the performance of the three methods while the ratio of observations to candidate predictors varies. Although previous research suggested that the novel data-driven exploration methods (e.g., *BMA* and LASSO) might perform better than conventional or stepwise regression when the abovementioned ratio is low, more concrete information about the point is still warranted. Hence, I will test the three methods while changing the ratio by altering the subsample size with one dataset with many candidate predictors.

With the evidence from the performance tests mentioned above, I will provide practical guidelines and considerations for employing data-driven model exploration methods, i.e., stepwise regression, *BMA*, and LASSO, to end-user researchers in psychology and education. I will discuss their strengths and limitations with evidence collected from tests with datasets in the fields. This paper will offer practical insights about utilizing the tools for psychological and educational researchers who want to perform model exploration instead of a priori hypothesis testing.

2. Materials and Methods

2.1. Test Datasets

I employed three datasets collected by studies in psychology to examine the performance of the three data-driven model exploration methods, i.e., stepwise regression, *BMA*, and LASSO. I acquired the datasets from open data repositories. I determined which datasets to employ based on the ratio of observations to the number of candidate predictors. Given that the previous literature recommends a minimum ratio ranging from 10 to 50 [46], I selected the three datasets with a ratio approximating those recommended numbers (90.42, 22.92, and 7.86, respectively).

The first dataset is from the field of positive psychology to examine which psychological factors predict one's sense of purpose contributing to flourishing (see Han (2024) [47] for the full information about this study). The original dataset and source codes for analyses in the study are available at https://doi.org/10.17605/OSF.IO/6VUK3 (accessed on 11 July 2024). The study employed six candidate predictors, i.e., empathic concern, perspective taking (see Davis (1983) [48] for further details about the Interpersonal Reactivity Index, the measure for these factors), moral identity internalization, moral identity symbolization (see Aquino and Reed (2002) [49] for further details about the Moral Identity Scale), post-conventional moral reasoning (see Choi et al. (2019) [50] for further details about the behavioral Defining Issues Test), and moral growth mindset (see Han et al. (2020) [51] for further details about the Moral Growth Mindset Measure), to predict purpose among adolescents (see Bronk et al. (2018) [52] for the Claremont Purpose Scale to assess the sense of purpose).

I employed the second dataset in moral and personality psychology exploring which character strengths significantly predict developed moral reasoning in post-conventional reasoning (see Han et al. (2022) [20] for the full information about this study). The original data and analysis source code files are openly available at https://doi.org/10.17605/OSF.IO/GCR9E (accessed on 11 July 2024). In this study, the Global Assessment of Character Strengths with 72 items was employed (see McGrath (2023) [53] for the details about this

assessment). In this case, the ratio of observations to the number of candidate predictors was 22.92 (1100 responses × 50%/24 candidate predictors).

Finally, the last dataset is from public health psychology (see Blackburn et al. (2022) [54] for details about the original study and Han (2022) [55] for the full descriptions about the dataset, including procedures for data collection and pre-processing). The original study's data and source code files are available at https://doi.org/10.17605/OSF.IO/Y4KGH. The study employed the COVID-19 vaccine intent measure (see Han (2022) [56] for the measurement).

For the full descriptions for the three datasets, refer to the Supplementary Materials.

2.2. Test Procedures

I used cross-validation to test the performance of the three exploration methods with the three datasets [57]. During the cross-validation process, I randomly separated the whole dataset into two subsets, one for training (model exploration; 50%) and one for cross-validation (50%) [58]. Once the subsets were created, I performed model exploration with the three methods. Because I intended to provide practical insights for end-user researchers with insufficient methodological expertise, I employed the default settings and functionalities without setting any special parameter during the exploration process. I conducted stepwise regression (default *R* function *step*), *BMA* (*bicreg* in *BMA* package), and LASSO (*cv.glmnet* in *glmnet* package) with the training dataset [27,35,39,40]. I extracted the estimated coefficients of candidate predictors from the regression results. All estimated coefficients from each method were stored in matrices for cross-validation.

Then, I performed cross-validation with the stored estimated coefficients from the training process [57]. I calculated the root mean squared error (RMSE) by comparing the estimated and actual dependent variable values [59]. I deemed an exploration method demonstrating the smallest RMSE value as one with the best cross-validation performance [57,60], which is the least susceptible to overfitting, and thus, the best exploration method in the present study. The estimated dependent variable values were calculated with the stored coefficients and predictor values from the cross-validation dataset. I repeated this cross-validation procedure 1000 times. For each time, randomly separating the training vs. cross-validation datasets was performed for repeated performance measures. The whole training and cross-validation test processes were performed for three datasets.

I conducted the statistical analysis of the performance data collected from the repeated cross-validation tests. I examined whether one method demonstrated a significantly smaller RMSE than another via the Bayesian *t*-test implemented in the *BayesFactor* package [61]. Consequently, three sets of Bayesian *t*-tests, i.e., *BMA* vs. LASSO, *BMA* vs. stepwise regression, and LASSO vs. stepwise regression, were performed for each dataset. I intentionally employed the Bayesian *t*-test instead of its conventional frequentist counterpart because Bayesian analysis is robust against inflated false positives when multiple comparisons are performed even without any correction [34,62]. I used Bayes Factors (BFs) indicating the extent to which the alternative hypothesis was more strongly supported by evidence than the null hypothesis for inference [63,64]. When 2log(BF) was 3 or higher [34,63,65], I concluded that there was a significant difference that was positively supported by data.

I also conducted additional cross-validation tests with the first dataset (including the purpose and moral psychology indicators) to examine how changes in the ratio of observations to the number of candidate predictors affect the performance of the methods more closely. For this, I analyzed the first dataset with four different sizes of subsets, i.e., n = 100, 200, 400, and 800. For these four conditions, the ratio values were 8.33 (100 responses × 50%/6 candidate predictors; only 50% of the subset was used for training), 16.67 (200 responses × 50%/6 candidate predictors), 33.33 (400 responses × 50%/6 candidate predictors), and 66.67 (800 responses × 50%/6 candidate predictors), respectively. The four different subset sizes were determined to make the ratio values range from below 10 to above 50 [46]. I applied the same procedures for the training and cross-validity check used for the whole dataset tests with this randomly selected subset. For example,

in the case of the condition of n = 100, 100 out of 1085 responses were randomly selected for each trial. Then, 50 were used for training and 50 were reserved for cross-validation. Finally, I calculated RMSE values with the cross-validation subset based on the coefficients estimated with the training subset. I repeated the abovementioned process 1000 times for each condition. Once all RMSEs were calculated, I performed Bayesian *t*-tests to compare RMSEs between the three model exploration methods for each n. I examined how the performance trend changes across different subset sizes, from 100 to 800, resulting in varied ratios of observations to the number of candidate predictors, from 8.33 to 66.67.

3. Results

3.1. First Dataset: Purpose and Moral Psychological Indicators

I reported the overall performance outcomes when the first dataset was examined in Table 1 (see CPS [full]) and Figure S1. For additional information, I also reported the effect sizes for this and subsequent tests in Cohen's d. LASSO demonstrated the best performance (the smallest RMSE) compared to all other methods. Interestingly, *BMA* was not superior to stepwise regression in this condition.

Table 1. Results from data exploration method performance analyses.

	BMA vs. LASSO		**BMA vs. Stepwise**		**LASSO vs. Stepwise**	
	2log(BF)	**Cohen's *d***	**2log(BF)**	**Cohen's *d***	**2log(BF)**	**Cohen's *d***
CPS (full)	1032.83	1.35	566.63	0.88	229.50	−0.52
GACS	16.67	0.15	380.59	−0.69	460.33	−0.77
Trust	29.76	0.19	94.39	−0.33	130.69	−0.38
CPS (n = 100)	37.96	−0.21	332.11	−0.64	46.07	−0.23
CPS (n = 200)	473.10	0.79	38.38	0.21	245.62	−0.54
CPS (n = 400)	721.52	1.04	61.59	0.27	426.50	−0.74
CPS (n = 800)	713.33	1.03	239.52	0.53	278.34	−0.58

3.2. Second Dataset: Character Strengths and Moral Reasoning

The results from the tests of the second dataset are presented in Table 1 (see GACS) and Figure S2. LASSO performed significantly better than *BMA* and stepwise regression. *BMA* also reported a significantly smaller RMSE than stepwise regression.

3.3. Third Dataset: Trust and COVID-19 Vaccine Intent

I demonstrated the results from the statistical analysis of the data exploration performance done with the third dataset in Table 1 (see Trust) and Figure S3. LASSO showed the best performance, *BMA* the second, and stepwise regression the third.

3.4. Performance Trends across Different Sample Sizes

I examined how the performance varied as the ratio of observation to the number of candidate predictors changed with the first dataset. The results from the performance analyses when n = 100, 200, 400, and 800 are presented in the four last rows in Table 1 and Figure S4. When $n \geq 200$, the ratio was 16.67 or larger; as when the whole dataset was analyzed, LASSO demonstrated the best performance, while the performance of *BMA* was inferior to that of stepwise regression. However, when the sample size and the ratio were smallest (n = 100), *BMA* outperformed stepwise regression. The superiority of *BMA* to stepwise regression was only observed when n = 100.

4. Discussion

In this study, I examined the performance of data-driven model exploration methods, i.e., stepwise regression, *BMA*, and LASSO, with cross-validation. I also investigated how

the ratio of observations to the number of predictors influenced the performance of each method. When the three datasets with different ratio values were tested, LASSO based on regularized regression demonstrated the best performance in all cases. *BMA* outperformed stepwise regression when the ratio values were low (the first dataset with n = 100, the second and third datasets). When the first dataset with the relatively high ratio value (90.42) was analyzed, *BMA* demonstrated a worse performance than stepwise regression except for when n = 100 (ratio = 8.33). Generally, these findings support my prediction that novel approaches for data-driven model exploration, i.e., *BMA* and regularized regression (LASSO in this study), would perform better than stepwise regression in terms of better cross-validity and robustness against overfitting. Among them, I found that LASSO performed best in all cases even when there were many observations (the first dataset). Interestingly, *BMA* did not perform better than stepwise regression when the ratio of observations to the number of candidate predictors was large.

The first noteworthy point is that LASSO, a form of regularized regression, demonstrated the best performance in all cases. This might be due to the nature and mechanism of *cv.glmnet* that I used in the present study for LASSO [6,66]. When one performs *cv.glmnet*, this function in *glmnet* package searches for the regression parameters that maximize cross-validation accuracy when the default options are used [39,40,67]. Consequently, the resultant coefficients become shrunken while unnecessarily large coefficients are penalized due to this behavioral tendency [5,6]. It might result in better cross-validation performance and robustness against overfitting compared with other methods that do not directly target optimizing cross-validity. *BMA* and stepwise regression are not interested in minimizing cross-validation RMSE. Such a difference in the primary purpose of the functionality might contribute to the superior performance of LASSO to others when cross-validity and overfitting become the main concerns. Hence, if a researcher's primary purpose is finding a prediction model that is most robust against overfitting and best explains and predicts phenomena outside of the collected data, LASSO may be the first candidate to be considered [37].

Despite the methodological benefit of regularized regression, as mentioned in the introduction, there is a significant practical limitation that end-users may need to consider. The method was most robust against overfitting, so it seems the most appropriate method for model exploration and predictor selection from the methodological perspective. However, it does not present any additional useful information for inference, such as p-values, that many psychological and educational researchers might be interested in [20,37]. Hence, it would be difficult to interpret the findings particularly when researchers are interested in their statistical implications (e.g., which predictor is statistically significant, etc.). One possible solution might be that researchers provide in-depth interpretations based on domain knowledge during discussion instead of merely relying on quantitative interpretations, particularly binary decision-making (e.g., $p < 0.05$) [68,69]. This might require additional labor during studies. However, given that many statisticians raised concerns about using simple thresholds for statistical inference, this will be a more fundamental solution to deal with overfitting and the quality of result interpretation during data-driven analysis [2,70–72].

BMA demonstrated good performance generally when the ratio of observations to the number of candidate predictors was low, particularly when it did not satisfy the most lenient cutoff, 10 [9,46,73]. In other cases, particularly when the number of candidate predictors was small (6, the first dataset), it did not demonstrate performance superior to stepwise regression. Interestingly, when the number of candidate predictors was high (24, the second dataset), *BMA* performed better than stepwise regression even if the ratio value, 22.92, was higher than the lenient cutoff. Given these results, *BMA* performs well when the collected data size is smaller than the number of candidate predictors. This effect increases as the number of candidate predictors increases. It is consistent with what the *BMA* developers argued: *BMA* is good at addressing model uncertainty when the abovementioned ratio is small [9,23].

Because *BMA* does not directly aim to optimize cross-validity like regularized regression, it might be plausible to see that its performance is better than regularized regression. Even if that is the case, *BMA* has a unique benefit compared with regularized regression for applied, not methodology, researchers. Because its results provide more information for inference, such as the posterior probabilities of candidate models and predictors, researchers interested in inference and explanation will find it useful [9,20]. Unlike LASSO and regularized regression, which do not usually provide such information, *BMA* suggests the selection of inclusion of which model or predictor is more likely to occur. Compared with conventional frequentist indicators, such as *p*-values, Bayesian indicators calculated by *BMA* are epistemologically superior because they directly indicate whether an alternative hypothesis, not a null hypothesis, is more strongly supported by evidence. Of course, the results from *BMA* are less straightforward to interpret than those from stepwise regression because it averages multiple models [20]. However, if researchers can understand and explain the Bayesian output indicators appropriately, *BMA* might be a good alternative when many candidate predictors are in a small dataset (e.g., when the ratio is low, <10; or when the number of candidate predictors is large, such as 24 in the second dataset).

Stepwise regression demonstrated acceptable performance only under limited circumstances when there were plenty of observations in the analyzed dataset. Its performance was superior to *BMA* only when the number of the candidate predictor was small. It never outperformed regularized regression. This result suggests that stepwise regression, which has been regarded as a traditional approach for model and variable selection, might not be optimal in many cases. If researchers intend to explore the best models and predictors with unspecific candidate predictors, either regularized regression or *BMA* should be considered. Selecting a single model from many candidates might seem attractive because the result is straightforward to understand. However, due to model uncertainty and the possibility of inflated false positives as discussed in the introduction, the model suggested by stepwise regression can be misleading [9,23,29,73]. Hence, instead of relying on this conventional method, perhaps including stepwise regression that has been widely used so far, researchers may start with the abovementioned methods appropriate for data exploration at the beginning. Then, once a potential model is identified, they can employ hypothesis-focused analysis to acquire more information for inference [74].

I will also consider how data-driven methods are related to efforts to examine the relative importance of candidate predictors. As I mentioned in the introduction, data-driven analysis and analysis of relative importance (e.g., dominance analysis and relative weight analysis) pursue different goals [19]. However, I assumed that results from data-driven analysis may inform psychological and educational researchers who are interested in investigating the extent to which predictors of interest contribute to predicting the outcome variable of interest. At the least, the outcomes of data-driven exploration, the suggested best model and candidate predictors, can indirectly inform such researchers. We can expect that variables that survived the exploration process are deemed to be more important than those that did not.

There are some caveats regarding the potential usefulness of data-driven methods in relative importance examination. Obviously, stepwise regression is likely to be least informative. It only suggests one candidate model and is susceptible to model uncertainty [9,29]. Also, it does not provide any additional information other than the suggested model. That said, researchers cannot get more direct information about the relative importance of each survived predictor. The same issue regarding the lack of direct importance indicators can be applied to regularized regression [75]. *BMA* might be the best method among three candidates for this purpose. In fact, Shou and Smithson [19] suggested that *BMA* can provide useful information about relative importance when it was compared with dominance analysis. Perhaps the posterior probabilities of candidate predictors, which are about the extent to which the predictors are supposed to be included in prediction models, can be used as more direct indicators for the relative importance of the predictors.

However, there are several issues that we need to consider while utilizing *BMA* for a method for relative importance analysis. Although *BMA* and dominance analysis behave similarly in most cases, as mentioned previously, they pursue two different goals at the first place, model averaging (or exploration) and testing relative importance [19], at the beginning. Thus, *BMA*'s relative superiority in examining relative importance should be considered a collateral benefit. Also, the consistency between *BMA* and dominance analysis becomes lower as multicollinearity increases [19]. If multicollinearity becomes a serious concern, then researchers may need to test the multicollinearity issue before conducting *BMA* when relative importance is also of their interest. Despite the issue when multicollinearity is severe, I would like to note that one positive aspect of *BMA* related to multicollinearity is that it can potentially (and perhaps partially) provide one way to address the issue [76]. Future studies should be done to investigate the benefits of *BMA* in relative importance analysis more accurately.

Although I could test the performance of different model exploration methods and discuss practical implications based on evidence in this study, several limitations warrant additional research. First, I did not test analysis methods addressing between-group differences, such as multilevel modeling, to examine data collected from different groups. Given the employed *R* functionalities widely available to end users only allowed testing within-group population-level predictors, I could not address data-driven methods for multilevel modeling (e.g., *explore.models* [7]). Researchers interested in examining cross-cultural or multi-group differences with advanced statistical knowledge may need to test data-driven methods considering between-group effects in future studies. Second, because I was primarily interested in practical considerations for end users likely to use the functionalities with default settings, I did not examine the performance when customized settings were applied. How data-driven functions behave when different settings are employed was out of the scope of this study. Since fine-tuning settings and parameters for optimization can significantly influence analysis outcomes, such a possibility needs to be tested and discussed in future research.

5. Concluding Remarks

In this paper, I tried to suggest several practical guidelines about how to employ data-driven model exploration methods by examining the widely available methods of stepwise regression, *BMA*, and LASSO. Researchers in psychology and education may refer to the suggestions to conduct more accurate data-driven analysis depending on the nature of their data to be examined. For additional information for practice, I suggest that future works should be done in the field. Perhaps, as I mentioned while discussing limitations, data collected from different groups may need to be further tested with data-driven methods to inform cross-cultural researchers. Furthermore, researchers may consider adjusting options for the *R* functionalities, and test and compare outcomes to examine how to maximize the performance of each method. With the findings from the proposed future studies, practical researchers will obtain more useful insights into how to utilize the data-driven model exploration methods in an optimal way based on their research purposes and situational factors.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/stats7030044/s1, Additional Information about Datasets; Additional Information about Data-driven Methods; Figure S1: Performance (RMSE) test with the first dataset; Figure S2: Performance (RMSE) test with the second dataset; Figure S3. Performance (RMSE) test with the third dataset; Figure S4: Performance (RMSE) trends when the ratio of observations to the number of candidate predictors varied. Reference [77] is cited in the Supplementary Materials.

Author Contributions: Conceptualization, H.H.; methodology, H.H.; software, H.H.; formal analysis, H.H.; writing—original draft preparation, H.H.; writing—review and editing, H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable given only secondary datasets openly available to the public were used in this study.

Informed Consent Statement: Not applicable given only secondary datasets openly available to the public were used in this study.

Data Availability Statement: All data and source files to run the tests conducted in this study are openly available to the public via the Open Science Framework repository: https://doi.org/10.17605/OSF.IO/ZSYTG (accessed on 11 July 2024).

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Jack, R.E.; Crivelli, C.; Wheatley, T. Data-Driven Methods to Diversify Knowledge of Human Psychology. *Trends Cogn. Sci.* **2018**, *22*, 1–5. [CrossRef] [PubMed]
2. Wagenmakers, E.-J. A Practical Solution to the Pervasive Problems of p Values. *Psychon. Bull. Rev.* **2007**, *14*, 779–804. [CrossRef] [PubMed]
3. Wagenmakers, E.-J.; Marsman, M.; Jamil, T.; Ly, A.; Verhagen, J.; Love, J.; Selker, R.; Gronau, Q.F.; Šmíra, M.; Epskamp, S.; et al. Bayesian Inference for Psychology. Part I: Theoretical Advantages and Practical Ramifications. *Psychon. Bull. Rev.* **2018**, *25*, 35–57. [CrossRef] [PubMed]
4. Weston, S.J.; Ritchie, S.J.; Rohrer, J.M.; Przybylski, A.K. Recommendations for Increasing the Transparency of Analysis of Preexisting Data Sets. *Adv. Methods Pract. Psychol. Sci.* **2019**, *2*, 214–227. [CrossRef] [PubMed]
5. McNeish, D.M. Using Lasso for Predictor Selection and to Assuage Overfitting: A Method Long Overlooked in Behavioral Sciences. *Multivar. Behav. Res.* **2015**, *50*, 471–484. [CrossRef] [PubMed]
6. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. [CrossRef]
7. Han, H. A Method to Explore the Best Mixed-Effects Model in a Data-Driven Manner with Multiprocessing: Applications in Public Health Research. *EJIHPE* **2024**, *14*, 1338–1350. [CrossRef] [PubMed]
8. Han, H.; Dawson, K.J. Applying Elastic-Net Regression to Identify the Best Models Predicting Changes in Civic Purpose during the Emerging Adulthood. *J. Adolesc.* **2021**, *93*, 20–27. [CrossRef] [PubMed]
9. Hoeting, J.A.; Madigan, D.; Raftery, A.E.; Volinsky, C.T. Bayesian Model Averaging: A Tutorial. *Stat. Sci.* **1999**, *14*, 382–401. [CrossRef]
10. Lu, M.; Zhou, J.; Naylor, C.; Kirkpatrick, B.D.; Haque, R.; Petri, W.A.; Ma, J.Z. Application of Penalized Linear Regression Methods to the Selection of Environmental Enteropathy Biomarkers. *Biomark. Res.* **2017**, *5*, 9. [CrossRef]
11. Feher, B.; Lettner, S.; Heinze, G.; Karg, F.; Ulm, C.; Gruber, R.; Kuchler, U. An Advanced Prediction Model for Postoperative Complications and Early Implant Failure. *Clin. Oral Implants Res.* **2020**, *31*, 928–935. [CrossRef] [PubMed]
12. Babyak, M.A. What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models. *Psychosom. Med.* **2004**, *66*, 411–421. [CrossRef] [PubMed]
13. Ng, A.Y. Preventing "Overfitting" of Cross-Validation Data. In Proceedings of the Machine Learning: Fourteenth International Conference (ICML 97), Nashville, TN, USA, 8–12 July 1997.
14. Johnson, J.W.; LeBreton, J.M. History and use of relative importance indices in organizational research. *Organ. Res. Methods* **2004**, *7*, 238–257. [CrossRef]
15. Kruskal, W.; Majors, R. Concepts of relative importance in recent scientific literature. *Am. Stat.* **1989**, *43*, 2–6. [CrossRef]
16. Budescu, D.V.; Azen, R. Beyond global measures of relative importance: Some insights from dominance analysis. *Organ. Res. Methods* **2004**, *7*, 341–350. [CrossRef]
17. Lipovetsky, S.; Conklin, W.M. Predictor relative importance and matching regression parameters. *J. Appl. Stat.* **2015**, *42*, 1017–1031. [CrossRef]
18. Johnson, J.W. A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivar. Behav. Res.* **2000**, *35*, 1–19. [CrossRef] [PubMed]
19. Shou, Y.; Smithson, M. Evaluating predictors of dispersion: A comparison of dominance analysis and Bayesian model averaging. *Psychometrika* **2015**, *80*, 236–256. [CrossRef] [PubMed]
20. Han, H.; Dawson, K.J.; Walker, D.I.; Nguyen, N.; Choi, Y.-J. Exploring the Association between Character Strengths and Moral Functioning. *Ethics Behav.* **2022**, *33*, 286–303. [CrossRef]
21. Galasso, V.; Pons, V.; Profeta, P.; Becher, M.; Brouard, S.; Foucault, M. Gender Differences in COVID-19 Attitudes and Behavior: Panel Evidence from Eight Countries. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 27285–27291. [CrossRef]
22. Han, H.; Dawson, K.J. Improved Model Exploration for the Relationship between Moral Foundations and Moral Judgment Development Using Bayesian Model Averaging. *J. Moral Educ.* **2022**, *51*, 204–218. [CrossRef]
23. Raftery, A.E.; Zheng, Y. Discussion: Performance of Bayesian Model Averaging. *J. Am. Stat. Assoc.* **2003**, *98*, 931–938. [CrossRef]
24. Brown, D.L. Faculty Ratings and Student Grades: A University-Wide Multiple Regression Analysis. *J. Educ. Psychol.* **1976**, *68*, 573–578. [CrossRef]

25. Henderson, D.A.; Denison, D.R. Stepwise Regression in Social and Psychological Research. *Psychol. Rep.* **1989**, *64*, 251–257. [CrossRef]
26. Ghani, I.M.M.; Ahmad, S. Stepwise Multiple Regression Method to Forecast Fish Landing. *Procedia-Soc. Behav. Sci.* **2010**, *8*, 549–554. [CrossRef]
27. DataCamp; Step: Choose a Model by AIC in a Stepwise Algorithm 2024. Available online: https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/step (accessed on 11 July 2024).
28. Clyde, M. Model Uncertainty and Health Effect Studies for Particulate Matter. *Environmetrics* **2000**, *11*, 745–763. [CrossRef]
29. George, E.I.; Clyde, M. Model Uncertainty. *Stat. Sci.* **2004**, *19*, 81–94. [CrossRef]
30. Hawkins, D.M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12. [CrossRef]
31. Kumar, S.; Attri, S.D.; Singh, K.K. Comparison of Lasso and Stepwise Regression Technique for Wheat Yield Prediction. *J. Agrometeorol.* **2021**, *21*, 188–192. [CrossRef]
32. Raftery, A.E.; Madigan, D.; Hoeting, J.A. Bayesian Model Averaging for Linear Regression Models. *J. Am. Stat. Assoc.* **1997**, *92*, 179–191. [CrossRef]
33. Raftery, A.E.; Hoeting, J.A.; Volinsky, C.T.; Painter, I.; Yeung, K.Y. Package "BMA". Available online: https://cran.r-project.org/web/packages/BMA/BMA.pdf (accessed on 11 July 2024).
34. Han, H. A Method to Adjust a Prior Distribution in Bayesian Second-Level fMRI Analysis. *PeerJ* **2021**, *9*, e10861. [CrossRef] [PubMed]
35. Raftery, A.E.; Painter, I.S.; Volinsky, C.T. BMA: An R Package for Bayesian Model Averaging. *Newsl. R Proj.* **2005**, *5*, 2–8.
36. Hinne, M.; Gronau, Q.F.; Van den Bergh, D.; Wagenmakers, E.-J. A Conceptual Introduction to Bayesian Model Averaging. *Adv. Methods Pract. Psychol. Sci.* **2020**, *3*, 200–215. [CrossRef]
37. Yarkoni, T.; Westfall, J. Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspect. Psychol. Sci.* **2017**, *53*, 174569161769339. [CrossRef] [PubMed]
38. Zou, H.; Hastie, T. Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [CrossRef]
39. Friedman, J.; Hastie, T.; Tibshirani, R.; Narasimhan, B.; Tay, K.; Simon, N.; Qian, J. Glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. Available online: https://cran.r-project.org/web/packages/glmnet/index.html (accessed on 11 July 2024).
40. Hastie, T.; Qian, J. Glmnet Vignette. Available online: https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html (accessed on 11 July 2024).
41. Kim, M.-H.; Banerjee, S.; Park, S.M.; Pathak, J. Improving Risk Prediction for Depression via Elastic Net Regression-Results from Korea National Health Insurance Services Data. In *AMIA Annual Symposium Proceedings*; AMIA Symposium: San Francisco, CA, USA, 2016; Volume 2016, pp. 1860–1869.
42. Finch, W.H.; Hernandez Finch, M.E. Regularization Methods for Fitting Linear Models with Small Sample Sizes: Fitting the Lasso Estimator Using R. *Pract. Assess. Res. Eval.* **2019**, *21*, 7. [CrossRef]
43. Doebler, P.; Doebler, A.; Buczak, P.; Groll, A. Interactions of Scores Derived from Two Groups of Variables: Alternating Lasso Regularization Avoids Overfitting and Finds Interpretable Scores. *Psychol. Methods* **2023**, *28*, 422–437. [CrossRef]
44. Fei, S.; Chen, Z.; Li, L.; Ma, Y.; Xiao, Y. Bayesian Model Averaging to Improve the Yield Prediction in Wheat Breeding Trials. *Agric. For. Meteorol.* **2023**, *328*, 109237. [CrossRef]
45. Wang, D.; Zhang, W.; Bakhai, A. Comparison of Bayesian Model Averaging and Stepwise Methods for Model Selection in Logistic Regression. *Stat. Med.* **2004**, *23*, 3451–3467. [CrossRef]
46. Heinze, G.; Dunkler, D. Five Myths about Variable Selection. *Transpl. Int.* **2017**, *30*, 6–10. [CrossRef]
47. Han, H. Exploring the Relationship between Purpose and Moral Psychological Indicators. *Ethics Behav.* **2022**, *34*, 28–39. [CrossRef]
48. Davis, M.H. Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach. *J. Personal. Soc. Psychol.* **1983**, *44*, 113–126. [CrossRef]
49. Aquino, K.; Reed, A. The Self-Importance of Moral Identity. *J. Personal. Soc. Psychol.* **2002**, *83*, 1423–1440. [CrossRef] [PubMed]
50. Choi, Y.-J.; Han, H.; Dawson, K.J.; Thoma, S.J.; Glenn, A.L. Measuring Moral Reasoning Using Moral Dilemmas: Evaluating Reliability, Validity, and Differential Item Functioning of the Behavioural Defining Issues Test (bDIT). *Eur. J. Dev. Psychol.* **2019**, *16*, 622–631. [CrossRef]
51. Han, H.; Dawson, K.J.; Choi, Y.R.; Choi, Y.-J.; Glenn, A.L. Development and Validation of the English Version of the Moral Growth Mindset Measure [Version 3; Peer Review: 4 Approved]. *F1000Research* **2020**, *9*, 256. [CrossRef] [PubMed]
52. Bronk, K.C.; Riches, B.R.; Mangan, S.A. Claremont Purpose Scale: A Measure That Assesses the Three Dimensions of Purpose among Adolescents. *Res. Hum. Dev.* **2018**, *15*, 101–117. [CrossRef]
53. McGrath, R.E. A Summary of Construct Validity Evidence for Two Measures of Character Strengths. *J. Personal. Assess.* **2023**, *105*, 302–313. [CrossRef] [PubMed]
54. Blackburn, A.M.; Vestergren, S.; the COVIDiSTRESS II Consortium. COVIDiSTRESS Diverse Dataset on Psychological and Behavioural Outcomes One Year into the COVID-19 Pandemic. *Sci. Data* **2022**, *9*, 331. [CrossRef] [PubMed]
55. Han, H. Trust in the Scientific Research Community Predicts Intent to Comply with COVID-19 Prevention Measures: An Analysis of a Large-Scale International Survey Dataset. *Epidemiol. Infect.* **2022**, *150*, e36. [CrossRef]
56. Han, H. Testing the Validity of the Modified Vaccine Attitude Question Battery across 22 Languages with a Large-Scale International Survey Dataset: Within the Context of COVID-19 Vaccination. *Hum. Vaccines Immunother.* **2022**, *18*, 2024066. [CrossRef]

57. De Rooij, M.; Weeda, W. Cross-Validation: A Method Every Psychologist Should Know. *Adv. Methods Pract. Psychol. Sci.* **2020**, *3*, 248–263. [CrossRef]
58. Bengio, Y.; Grandvalet, Y. No Unbiased Estimator of the Variance of K-Fold Cross-Validation. *Adv. Neural Inf. Process. Syst.* **2003**, *16*, 513–520.
59. Tuarob, S.; Tucker, C.S.; Kumara, S.; Giles, C.L.; Pincus, A.L.; Conroy, D.E.; Ram, N. How Are You Feeling?: A Personalized Methodology for Predicting Mental States from Temporally Observable Physical and Behavioral Information. *J. Biomed. Inform.* **2017**, *68*, 1–19. [CrossRef] [PubMed]
60. Lorenz, E.; Remund, J.; Müller, S.C.; Traunmüller, W.; Steinmaurer, G.; Pozo, D.; Ruiz-Arias, J.A.; Fanego, V.L.; Ramirez, L.; Romeo, M.G.; et al. Benchmarking of Different Approaches to Forecast Solar Irradiance. In Proceedings of the 24th European Photovoltaic Solar Energy Conference, Hamburg Germany, 21–25 September 2009; pp. 21–25.
61. Morey, R.D.; Rouder, J.N.; Jamil, T.; Urbanek, K.; Ly, A. Package 'BayesFactor. Available online: https://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf (accessed on 11 July 2024).
62. Berry, D.A.; Hochberg, Y. Bayesian Perspectives on Multiple Comparisons. *J. Stat. Plan. Inference* **1999**, *82*, 215–227. [CrossRef]
63. Wagenmakers, E.-J.; Love, J.; Marsman, M.; Jamil, T.; Ly, A.; Verhagen, J.; Selker, R.; Gronau, Q.F.; Dropmann, D.; Boutin, B.; et al. Bayesian Inference for Psychology. Part II: Example Applications with JASP. *Psychon. Bull. Rev.* **2018**, *25*, 58–76. [CrossRef] [PubMed]
64. Meskó, N.; Kowal, M.; Láng, A.; Kocsor, F.; Bandi, S.A.; Putz, A.; Sorokowski, P.; Frederick, D.A.; García, F.E.; Aguilar, L.A.; et al. Exploring Attitudes Toward "Sugar Relationships" across 87 Countries: A Global Perspective on Exchanges of Resources for Sex and Companionship. *Arch. Sex. Behav.* **2024**, *53*, 811–837. [CrossRef] [PubMed]
65. Kass, R.E.; Raftery, A.E. Bayes Factors. *J. Am. Stat. Assoc.* **1995**, *90*, 773–795. [CrossRef]
66. Ahmed, S.E.; Hossain, S.; Doksum, K.A. LASSO and Shrinkage Estimation in Weibull Censored Regression Models. *J. Stat. Plan. Inference* **2012**, *142*, 1273–1284. [CrossRef]
67. Scaliti, E.; Pullar, K.; Borghini, G.; Cavallo, A.; Panzeri, S.; Becchio, C. Kinematic Priming of Action Predictions. *Curr. Biol.* **2023**, *33*, 2717–2727.e6. [CrossRef]
68. Štěrba, Z.; Šašinka, Č.; Stachoň, Z.; Kubíček, P.; Tamm, S. Mixed Research Design in Cartography: A Combination of Qualitative and Quantitative Approaches. *Kartographische Nachrichten* **2014**, *64*, 262–269. [CrossRef]
69. Conn, V.S.; Chan, K.C.; Cooper, P.S. The Problem With *p*. *West. J. Nurs. Res.* **2014**, *36*, 291–293. [CrossRef] [PubMed]
70. Berger, J.O.; Sellke, T. Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence. *J. Am. Stat. Assoc.* **1987**, *82*, 112–122. [CrossRef]
71. Wasserstein, R.L.; Lazar, N.A. The ASA's Statement on p-Values: Context, Process and Purpose. *Am. Stat.* **2016**, *70*, 129–133. [CrossRef]
72. Cohen, J. The Earth Is Round ($p < 0.05$). *Am. Psychol.* **1994**, *49*, 997–1003. [CrossRef]
73. Raftery, A.E. Bayesian Model Selection in Social Research. *Sociol. Methodol.* **1995**, *25*, 111. [CrossRef]
74. Dreisbach, C.; Maki, K. A Comparison of Hypothesis-Driven and Data-Driven Research: A Case Study in Multimodal Data Science in Gut-Brain Axis Research. *CIN Comput. Inform. Nurs.* **2023**, *41*, 497–506. [CrossRef] [PubMed]
75. Mizumoto, A. Calculating the relative importance of multiple regression predictor variables using dominance analysis and random forests. *Lang. Learn.* **2023**, *73*, 161–196. [CrossRef]
76. Lee, Y.; Song, J. Robustness of model averaging methods for the violation of standard linear regression assumptions. *Commun. Stat. Appl. Methods* **2021**, *28*, 189–204. [CrossRef]
77. Fragoso, T.M.; Bertoli, W.; Louzada, F. Bayesian model averaging: A systematic review and conceptual classification. *Int. Stat. Rev.* **2018**, *86*, 1–28. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.