Improving Performance of Prediction Rule Ensembles with Relaxed and Adaptive Lasso

Hilbert, Anne

Universiteit
Leiden
The Netherlands

Improving Performance of Prediction Rule Ensembles with Relaxed and Adaptive Lasso

Anne Hilbert

Thesis advisor: Dr. Marjolein Fokkema, Institute of Psychology, Leiden University
Independent reader: Prof. Dr. Mark de Rooij, Institute of Psychology, Leiden University

Defended on January 10th , 2024

MASTER THESIS

STATISTICS AND DATA SCIENCE
UNIVERSITEIT LEIDEN

Contents

Abstract

Prediction models play a paramount role in various fields such as psychology and medicine, where the aim is to maximize predictive performance while ensuring high interpretability and stability. Prediction rule ensembles are a recent statistical learning method that address the black-box problem from common machine learning methods. First, an ensemble of trees is fitted, and by employing sparse regression, such as the lasso, only a subset of those trees is retained in the final ensemble, enhancing interpretability. However, the lasso suffers from drawbacks, considering that the optimal penalty parameter for variable selection may lead to an over-shrinkage of large coefficients. This study investigates if accuracy, sparsity, and stability of prediction rule ensembles can be improved by using the adaptive or relaxed lasso, or their combination. In the adaptive lasso, weight parameters are assigned to each coefficient in the penalty term, while in the relaxed lasso the lasso coefficients are debiased towards unpenalized values. In addition, in this study we compared if the results differ if the model selection was based on the lambda-1se or lambda-min criterion and between continuous and binary outcomes. For this, the models were evaluated on nine benchmark datasets using repeated 10-fold cross-validation. The results show that all lasso variations improve model sparsity significantly while maintaining high accuracy, but at the cost of stability. The relaxed and adaptive lasso select sparser models than the standard lasso while achieving good stability of variable selection, but at the cost of less stable predictions. The relaxed adaptive lasso yields the sparsest model, but is the most unstable. Regarding lambda criterion, for continuous outcomes the lambda-minimum criterion leads to highly unstable results and diminishes the effect of lasso approach used. For binary outcomes, the lambda-1se criterion only improves accuracy and sparsity, but not stability, while for continuous outcomes it improves all performance diagnostics.

Keywords: prediction rule ensembles, relaxed lasso, adaptive lasso, relaxed adaptive lasso, stability, sparsity, accuracy

Introduction

Common applications of statistics are prediction models, which are gaining increasing popularity as they play a pivotal role in many fields. Examples are in psychology, medicine, economy, and more. Goals of prediction models are selecting relevant predictors, and making accurate predictions without over-fitting the data. Furthermore, for practical applications, it is crucial that the results derived from prediction models remain interpretable and stable. Stability against chance fluctuations in the data is key as it ensures that the model can be used in different settings and the results are reliable and reproducible (Nogueira et al., 2018).

Examples of the application of prediction models are aptitude tests to predict outcomes such as a student's graduation or a job applicant's performance (Fokkema & Strobl, 2020). In the fields of clinical psychology and medicine, predictive models are employed to determine the likelihood of patient recovery or relapse (Fokkema & Strobl, 2020). Additionally, these models are used to predict potential diagnoses and guide the assessment of necessary medical screenings. Another example application is in credit scoring, predicting an applicant's credit-worthiness, or by companies to predict customer churning or improving recommendation systems. In all of these examples, model interpretability is crucial in order to know which factors to address to achieve the desired outcome.

Prediction Rule Ensembles

In machine learning, a useful tool to make predictions are decision trees, in which predictions are made based on rules in the format of "if [condition] then [outcome]" statements. In a single decision tree each observation is evaluated against a condition to make predictions. The interpretation is straightforward but the predictions may be inaccurate because the model is too simple. To increase accuracy machine learning methods such as random forest or bagging build many decision trees, forming an ensemble of trees. The predictions are made based on the aggregated results from all trees in the ensemble. This reduces the risk of over-fitting but makes the results more complex and difficult to interpret (Fokkema & Strobl, 2020). Tree ensembles are often called black-boxes because it is complex to understand how the model arrived at the prediction. To make data-driven decisions, interpretability is crucial. Prediction rule ensembles address the so-called black box problem by combining tree ensembling and sparse regression to only retain a subset of the trees from the initial ensemble. By fitting a more parsimonious model prediction rule ensemble strive to balance accuracy and interpretability. As the sparse regression method the lasso can be used to select which rules and/or linear terms stay in the final ensemble (J. H. Friedman & Popescu, 2008). However, the lasso suffers from some drawbacks. While it works well in model selection, it does not always select the most optimal model in terms of prediction accuracy (Dalalyan et al., 2017). In addition, the results of the lasso are unstable when multicollinearity is present (Zhao & Yu, 2006). To address these problems, variations of the lasso, such as the adaptive or relaxed lasso, have been proposed (Meinshausen, 2007; Zou, 2006). This study investigates if model accuracy, sparsity and stability of prediction rule ensembles can be improved by using these lasso variations.

RuleFit

An example of a prediction rule ensemble is the RuleFit algorithm by J. H. Friedman and Popescu (2008). In RuleFit, rules and linear terms are generated with gradient boosting, in more detail explained below. The rules and linear terms are also referred to as base learners. In the second step of the RuleFit algorithm, the lasso is used to make predictions and to select the final model by retaining only a subset of the initially fitted trees (J. H. Friedman & Popescu, 2008). In the lasso, all base learners are regressed on the response variable. If a rule/linear term applies to a given observation it is coded as one, otherwise as zero. Through the use of a penalty parameter, some base learners are dropped from the final ensemble, if their coefficient is shrunken to zero. Based on the regression coefficients, importance measures for each base learner and for each predictor can be estimated. This enhances the interpretation of the outcome as it can be identified which predictors and base learners are the most relevant in predicting the response variable (Fokkema & Strobl, 2020).

Rule Generation

To obtain the initial tree ensemble, gradient boosting can be employed. In RuleFit the rules in each iteration of gradient boosting are generated with the classification and regression tree (CART) algorithm (J. H. Friedman & Popescu, 2008). Alternatively, the conditional inference tree (ctree) algorithm, proposed by Hothorn et al. (2006), can be used. The advantage of ctree over CART is that ctree leads to more unbiased variable selection, is robust, and can be used for complex data (Hothorn et al., 2006). Unbiased in this context means that every predictor has the same likelihood of being selected if all predictors are uncorrelated to the response variable, regardless of measurement scale or missing values (Hothorn et al., 2006). The ctree algorithm ensures this by selecting the predictor variables based on a conditional inference test before generating the node splitting criteria (Hothorn et al., 2006).

In gradient boosting, the trees are fitted sequentially with each new tree learning from misclassifications of previous trees (Ayyadevara, 2018). The predictions are updated in each iteration, scaled by the learning rate. The learning rate $v \in [0, 1]$ regulates the influence from previous trees and is needed to avoid over-fitting (Natekin & Knoll, 2013). Research has shown a learning rate close to zero, for example $v = 0.01$, performs well in tree ensemble algorithms (J. H. Friedman, 2001). Furthermore, misclassified observations will have larger residuals/gradients, and thus exert more influence on the new predictions than the correctly classified observations. This allows new trees to learn from errors made by previous trees. The algorithm will stop when fitting a new tree would not improve the predictions significantly or when the maximum number of trees are reached. In addition, the maximum tree depths can be specified (Natekin & Knoll, 2013). Alternative methods to fit the trees in prediction rule ensembles are bagging or random forest (Fokkema & Strobl, 2020). An advantage is that no assumptions are made on the distribution of the response variable and they can thus be applied to many different types of data (Natekin & Knoll, 2013).

Rule Selection

After a forest of trees is fitted with gradient boosting, the next step in prediction rule ensembles is to select a small set of rules and linear terms to increase model sparsity and interpretability. For this, the Least Absolute Shrinkage and Selection Operator (LASSO) and three variations from the LASSO are compared in this study.

The Lasso

The lasso, first proposed by Tibshirani (1996), is a penalized regression method with a $L1$ penalty in the loss function. The size of the penalty is regulated by parameter λ. The lasso estimate can be written as follows:

$$\hat{\beta}^{\text{lasso}} = \arg\min_{\beta} \left(||y - \sum_{j=1}^{p} X_j \beta_j||^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right) \tag{1}$$

The larger λ, the more the coefficients will be shrunken toward zero and some coefficients will become exactly zero and thus be dropped from the model (Tibshirani, 1996). Other examples of variable selection methods are forward stepwise regression and best subset selection. However, these methods are computationally more expensive than the lasso, and are therefore not feasible when the number of predictors/base learners is large (Hastie et al., 2020). An advantage of the lasso is that in contrast to best subset selection and stepwise regression it is a convex optimization problem, meaning that there is only one local minimum, which is also the global minimum. Thus, the results are expected to be stable because the algorithm cannot get stuck in a local minimum (Hastie et al., 2020; Zou, 2006). In terms of variable selection the lasso is less aggressive and as a result leads to less parsimonious solutions especially when the signal-to-noise ratio (SNR) is high (Hastie et al., 2020). Another distinction is that in the lasso a shrinkage is also applied to the nonzero coefficients. A downside of this is that this can lead to an over-shrinkage of large coefficients (Dalalyan et al., 2017). Variations such as the relaxed or adaptive lasso have been proposed to counter these problems, as will be described in more detail below.

Relaxed Lasso

The relaxed lasso used in this study is a simplified version of the original relaxed lasso proposed by Meinshausen (2007). In the simplified relaxed lasso a multiplicative factor γ is introduced to control the strength of the regularization (Hastie et al., 2020). The new beta coefficients are calculated as a function of λ and γ:

$$\hat{\beta}^{\text{relaxed}}(\lambda, \gamma) = \gamma \hat{\beta}^{\text{lasso}}(\lambda) + (1 - \gamma)\hat{\beta}^{\text{ML}}(\lambda) \tag{2}$$

The $\hat{\beta}^{\text{ML}}$ value is estimated by fitting an unpenalized model to the predictors selected by the lasso. By multiplying the beta coefficients estimated with the lasso by $\gamma \in [0, 1]$, and adding them to the original coefficients from the OLS multiplied by $(1 - \gamma)$, the coefficients of predictors remaining in the model are debiased towards their OLS values. When $\gamma = 1$ the original lasso is obtained. In the ordinary lasso, a smaller lambda value that is suited for predictions may not be optimal for variable selection as redundant predictors will be selected. A larger lambda value suited for variable selection leads to an over-shrinkage of large coefficients. The factor $\gamma < 1$ in the relaxed lasso can mitigate this by reducing the penalty performed on predictors remaining in the model. Hastie et al. (2020) found

that the relaxed lasso achieves similar accuracy as the ordinary lasso when the signal-to-noise ratio (SNR) is low, but outperforms the ordinary lasso when the SNR is high. In addition, the relaxed lasso has a faster convergence rate than the original lasso and selects sparser models (Meinshausen, 2007).

Adaptive Lasso

In the adaptive lasso, a weight ω is imposed on the penalty of each predictor. By imposing a smaller weight on large coefficients, the penalty for important base learners is reduced (Zou, 2006), addressing the problem of over-shrinkage of large coefficients (Meinshausen & Bühlmann, 2006). The equation of the adaptive lasso can be written as follows:

$$\hat{\beta}^{\text{adaptive}} = \arg\min_{\beta} \left(||y - \sum_{j=1}^{p} X_j \beta_j||^2 + \lambda \sum_{j=1}^{p} \omega_j |\beta_j| \right) \tag{3}$$

The second term is the L_1 penalty. Vector ω_j contains the weights for each coefficient j and can be calculated as $\hat{\omega} = \frac{1}{|\hat{\beta}^{\text{ridge}}|}$. When the coefficient $\hat{\beta}$ of the predictor is larger, the weight will be smaller, reducing the shrinkage of lambda. The $\hat{\beta}$ parameter is estimated in a ridge regression in this study. Alternatively, it could be estimated from an OLS regression (Zou, 2006), but the ridge regression estimator is more stable. The ridge regression makes use of a $L2$ penalty term. In the $L2$ penalty, the squared coefficient β^2 is multiplied by λ, which is different from the $L1$ penalty in which the absolute value $|\beta|$ is multiplied by λ. Furthermore, research by Zou (2006) shows that the adaptive lasso meets the oracle properties; namely it can make as optimal predictions as if the underlying model was known. This is in contrast to the ordinary lasso which only consistently selects the true model when multicollinearity is low. In addition, they proved that the near-minimax optimal shrinkage property of the ordinary lasso also applies to the adaptive lasso. This means that the most optimal solution that can be obtained with nearly minimum risk will be selected (Zou, 2006).

Relaxed Adaptive Lasso

The relaxed adaptive lasso combines the relaxed and adaptive lasso. It is a very recent method and therefore not many research findings exist yet. Zhang et al. (2022) conducted a simulation study comparing the three lasso variations on linear models with continuous outcome variables. They found that the relaxed adaptive lasso can make more accurate predictions than the ordinary, relaxed, or adaptive lasso. The relaxed lasso and adaptive lasso tend to shrink too many coefficients to zero, while the relaxed adaptive lasso tended to select the right number of predictors. Moreover, the relaxed adaptive lasso has the same convergence rate as the relaxed lasso, namely $O_p(n^{-1})$. The adaptive lasso has a slower convergence rate, followed by the ordinary lasso (Zhang et al., 2022). However, no research findings exist on the use of relaxed adaptive lasso for non-continuous outcome variables or its application in rule ensembling methods. This study will address this research gap by comparing and evaluating prediction rule ensembles with rule selection based on the ordinary, relaxed, adaptive or relaxed adaptive lasso.

The following sections of this paper are structured as follows. In the Method section, the models and datasets are described and it is explained how the model metrics predictive accuracy, sparsity, and stability were measured. In

the Results section the findings for each of the mentioned metrics are discussed. The paper ends with a Discussion and Conclusion section.

Method

Fitted Models

All analyses were conducted in R version 4.3.1 (R Core Team, 2023). The models were fitted with function `pre` (Fokkema, 2020), which employs function `cv.glmnet` from package `glmnet` (J. Friedman et al., 2010). By default the standard lasso is used for rule selection. To obtain results for the relaxed lasso, the argument `relax` is set to 'TRUE'. The argument will be passed on to `cv.glmnet` which fits the relaxed lasso as described by Hastie et al. (2020). For the adaptive lasso, `ad.alpha` is set to 0, which specifies that the weights in the adaptive lasso are calculated based on coefficients from the ridge regression. For the parameter tuning the default settings of `pre` were used. Predictors were winsorized before being included as linear terms in estimation of the final model. For this, values below the 5th or above the 95th percentile were set to the 5th or 95th percentile respectively. This reduces the effect of outliers on the model. The decision trees were generated based on the ctree algorithm and fitted using gradient boosting with a learning rate of 0.01. Furthermore, 500 trees were built with a maximum tree depth of three. Each tree was fitted on a randomly drawn sub-sample consisting of 50% of the observations. The penalty parameter λ was selected based on k-fold cross-validation, and can be chosen either based on the lambda resulting in the lowest cross-validated error estimate, or the highest lambda value yielding a cross-validated error within 1 SE of the minimum (Fokkema & Strobl, 2020). The results for the predictions and importance measures based on the 1-SE versus the minimum criterion can be extracted from the same ensemble model, and were compared in this study. Both criteria are further referred to as the lambda-1se and lambda-min criterion.

Importance Measures

For each predictor and base learner importance measures were calculated indicating how much they contribute to the predictions. The importance measures of the base learners are estimated by multiplying the absolute value of their coefficients with their sample standard deviation. Equation 4 gives the formula of the importance measures of linear terms and Equation 5 of rules.

$$I_j = |\hat{b}_j| \times \text{sd}(l_j(x_j)) \tag{4}$$

$$I_k = |\hat{a}_k| \times \sqrt{s_k(1 - s_k)} \tag{5}$$

Here, s_k is the proportion of training observations to which rule k applies to. Accordingly, the standard deviation of rule k is estimated as $\sqrt{s_k(1 - s_k)}$. A larger regression coefficient indicates that the rule or linear term is more important and by multiplying it by its standard deviation it is accounted for its variance.

The importance measures can be further standardized with regard to the outcome variable, by dividing it by the standard deviation of the outcome variable y:

$$I'_j = |\hat{b}_j| \times \frac{sd(l_j(x_j))}{sd(y)} \tag{6}$$

$$I'_k = |\hat{a}_k| \times \frac{\sqrt{s_k(1-s_k)}}{sd(y)} \tag{7}$$

The importance measures of each predictor are calculated as follows:

$$J_j = I_j + \sum_{x_j \in r_k} \frac{I_k}{c_k} \tag{8}$$

Here, I_j is the importance measure of a linear term of predictor j, and $\sum_{x_j \in r_k} \frac{I_k}{c_k}$ is the sum of the importance measures of each rule that contains predictor j, divided by the number of conditions c_k in the rule. The sum gives the importance measure of the predictor (J. H. Friedman & Popescu, 2008). The higher the importance measure, the more relevant the predictor. If a predictor or base learner has been dropped from the model, its importance measure will be equal to zero.

Datasets

For this study eight benchmark datasets were used to implement and evaluate the four fitted models. Table 1 shows a summary of the datasets. Dataset three was used twice, with two different outcome variables. Thus there are nine datasets in total. A detailed description of all datasets can be found in Appendix A.

Table 1

Summary of Datasets

	Name	Outcome variable	Outcome type	p	N	Reference
1	High School Grades	grades	Continuous	30	649	Cortez, 2014
2	Sensation Seeking and Delinquency	delinquent behaviour	Continuous	25	1076	Roth and Herzberg, 2004
3	Personality and Drug Consumption	cannabis consumption	Continuous	12	1885	Fehrman et al., 2016
4	Personality and Drug Consumption	ecstasy consumption	Continuous	12	1885	Fehrman et al., 2016
5	Objectivity of Article	label (objective/subjective)	Binary	59	1000	Rizk and Awad, 2018
6	Breast Cancer Screening	benign/malignant tumour	Binary	30	569	Wolberg et al., 1995
7	Sleep Quality	sleep quality (poor/good)	Binary	17	546	Norbury and Evans, 2018
8	University Graduation	graduate/dropout	Binary	28	3630	Realinho et al., 2021
9	ADHD Screening	ADHD vs. control group	Binary	93	220	Trognon and Richard, 2022

Model Evaluations

The models were compared based on out-of-sample prediction accuracy, sparsity, and stability. For this, a repeated 10-fold cross validation with 10 repeats was conducted. The seed was set, ensuring that the same fold splitting was used for the same repeat in each model.

Predictive Accuracy

In each cross-validation the `pre` model was fitted on 90% of the data and predictions were made on the remaining 10%, which served as a validation set. This results in a vector containing the cross-validated predictions for each observation. In each of the 10 repeats, for continuous outcomes the Mean Squared Error (MSE) was calculated, and for binary outcomes the Squared Error Loss (SEL) based on the predicted probabilities, and the area under the ROC curve (AUC). To provide a measure of the effect size of predictive accuracy, the variance accounted for (VAF) was estimated by the coefficient of determination as follows:

$$R^2 = 1 - \frac{\text{MSE}}{\text{var}(y)} \tag{9}$$

In case of a binary outcome, the SEL was used instead of the MSE. The R^2 will be higher with increasing number of terms; to correct for this, the adjusted R^2 was estimated:

$$R^2_{\text{adjusted}} = 1 - \frac{(1 - R^2) \times (n - 1)}{n - \hat{p} - 1} \tag{10}$$

In this formula, n indicates the sample size, $\hat{p}$ the average number of terms selected across the 10 folds in the given repeat, and R^2 the coefficient of determination. The adjusted R^2 is independent of the measurement scale and can therefore be compared across datasets and provides an indication of the SNR.

It was also evaluated how stable the predictions are across the 10 repeats. For this, Euclidean distances between predictions were estimated between each pair of repeats. The average of the distances is reported in the Results and the distributions were plotted with boxplots. Given that this measure is dependent on the measurement scale of the outcome variable, it can only be compared within the same dataset. Additionally, the standard deviation of the MSE or SEL were calculated. The more variation, the less stable are the predictions.

Sparsity

The sparsity was estimated by the number of base learners and predictors selected in each fold of the cross-validation and for each repeat, resulting in 100 (folds × repeats) outcomes. The mean and standard deviation are reported in the Results section. We assume that a more parsimonious model has higher interpretability.

Stability

Model stability refers to robustness to small changes in the training data and is crucial in order to receive reproducible and generalizable results (Nogueira et al., 2018). In this study the stability of variable selection, importance measures, and predictions was evaluated.

Variable Selection. If a variable has been selected its importance measure will be > 0. In every fitted model a variable has either been selected (1) or not (0), resulting in a binary outcome for each fold. In order to estimate the stability of variable selection while correcting for the total number of variables, and the proportion of variables selected, a stability measure proposed by Nogueira et al. (2018) was calculated with following formula:

$$\hat{\phi}(Z) = 1 - \frac{\frac{1}{p}\sum_{j=1}^{p} s_j^2}{\frac{\overline{p}}{p}(1 - \frac{\overline{p}}{p})} \tag{11}$$

Here, $\overline{p}$ is the average proportion of predictors selected across the folds, p the total number of predictors, and s_j^2 the sample variance of selection of predictor j. The numerator is the average over the predictor's sample variance of selection. The denominator is the sample variance of the proportion of variables selected. If variable selection would be random, numerator and denominator would be equal and thus stability would be 0. The measure will be 1 only if selection of all predictors has a sample variance of 0, indicating that the same predictors have been selected in each fold and repeat. Nogueira et al. (2018) suggest following rules of thumb to interpret the stability:

- $\hat{\phi}(Z) > 0.75$: Excellent stability
- $0.4 < \hat{\phi}(Z) < 0.75$: Intermediate to good stability
- $\hat{\phi}(Z) < 0.4$: Poor stability

The stability between methods was compared with confidence intervals, calculated using a formula by Nogueira et al. (2018). The formula assumes that the stability measure follows a standard normal distribution.

Importance Measures. The stability of importance measures was evaluated with Euclidean distances. The standardized importance measures of the predictors were saved across each fold and repeat resulting in a dataframe containing a column for each predictor and a row with the importance measure for each fold. Euclidean distances between each row were calculated giving a measure of how much the importance measures differ across the folds. A larger distance indicates that the importance measures are less stable.

Model Comparison

To test whether the differences between the models are statistically significant, for each dataset linear mixed effects models were fitted, using the `lmer` function from the `lmerTest` package in R (Kuznetsova et al., 2017). The repeats are within-model effects. A full-factorial 2 (lambda criteria) × 2 (relaxed lasso) × 2 (adaptive lasso) design was used. The lambda-min criterion and standard lasso were taken as the reference categories. Furthermore, a three-way interaction between the lambda criterion, adaptive, and relaxed lasso was added to assess if the differences between lasso approaches differ depending on the lambda criterion.

Three linear mixed models were fitted. First, the difference in accuracy was tested with MSE/SEL as the outcome, and in a second model the adjusted R^2. For binary outcomes, an additional model was fitted, predicting the AUC. Third, to test the difference in sparsity the outcome variable was the number of predictors. The significance of the random effect was tested with a likelihood-ratio test. If the variance of the random-intercept variance was not significantly different than 0 (alpha = 0.05), we fitted a fixed-instead of a mixed-effects model. To test the differences in stability, the distances between predictions and between importance measures of predictors were regressed on method and criterion using a linear model. For the distances no random effect was needed as the distances were calculated between each pair of results and there is thus no dependency between observed values. To compare the stability of variable selection, confidence intervals were compared.

Results

To avoid repetition of similar results, we chose to report the results of datasets High School Grades and ADHD as typical examples of regression and classification problems. Results for all datasets are presented in the Appendix, visualizations in Appendix B and results of significance tests in Appendix C. The R-code and dataset files used in this study are available at https://github.com/Anne3478/thesis-pre-with-relaxed-adaptive-lasso.git.

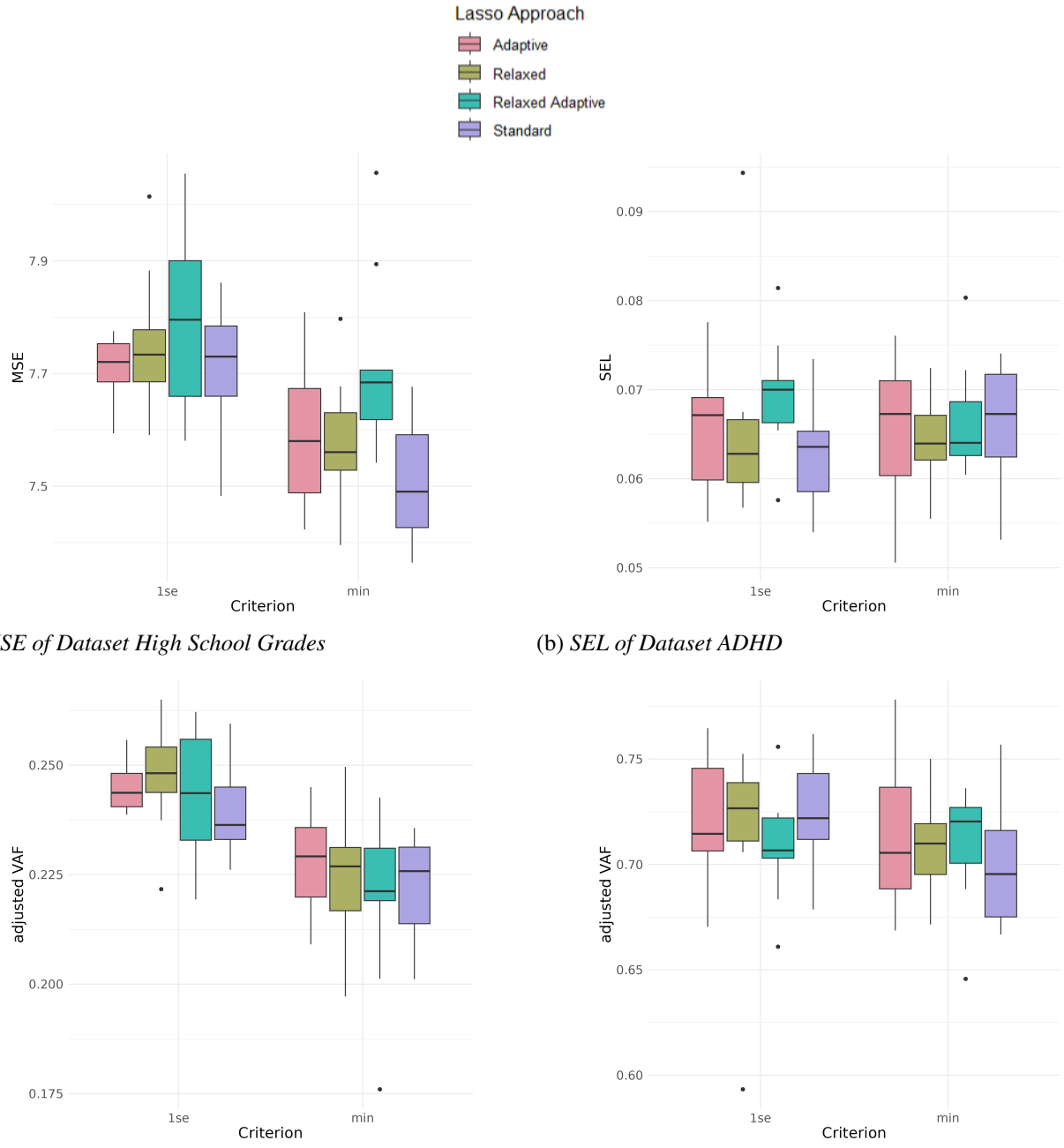
Predictive Accuracy

Table 2 presents the predictive accuracy for all datasets. In most regression and classification datasets, the standard lasso with the lambda-min criterion yields the best MSE/SEL value but the differences to the other lasso approaches are not large. Figure 1a shows that in dataset High School Grades the differences between the lasso approaches are not substantial and smaller than the differences between the two lambda criteria. This aligns with findings from other regression datasets. In two regression datasets, under the lambda-1se criterion, both the relaxed and standard lasso yield a lower MSE than the other approaches. In classification dataset ADHD, Figure 1b, the relaxed adaptive lasso has the highest SEL, with no significant differences observed among the remaining three lasso approaches. In the other classification datasets the pattern is similar. Furthermore, in both regression and classification datasets, the relaxed adaptive lasso has the largest variance compared to the other methods; suggesting that its predictions are less stable.

While the MSE/SEL is lower with increasing number of terms, the adjusted R^2 accounts for the number of base learners selected. As seen in Figure 1c, in dataset High School Grades the standard lasso has the lowest and the relaxed lasso the highest adjusted R^2, although the differences are not substantial. The same pattern is found in datasets Cannabis Consumption and Ecstasy consumption, while in dataset Delinquency no differences between lasso approaches are observed. In all regression datasets the adjusted R^2 is higher for the lambda-1se criterion than the lambda-min criterion. In dataset ADHD (Figure 1d), similar as to the other classification datasets, no considerable difference in adjusted R^2 between lasso approaches are found. Linear mixed models indicate that the main and interaction effects of relaxed and adaptive lasso on adjusted R^2 are not significant. In seven out of nine datasets, the lambda-1se criterion performed significantly better in terms of adjusted R^2 compared to the lambda-min criterion. Moreover, based on the significance tests and visualizations, the AUC seems to be unaffected by lambda criterion and lasso approach used.

Figure 1

Predictive Accuracy of Dataset High School Grades and ADHD



(a) *MSE of Dataset High School Grades*

(b) *SEL of Dataset ADHD*

(c) R^2_{adj} *of Dataset High School Grades*

(d) R^2_{adj} *of Dataset ADHD*

Table 2

Predictive Accuracy

		Lasso		Relaxed Lasso		Adaptive Lasso		Relaxed Adaptive	
Dateset		1se	min	1se	min	1se	min	1se	min
High School Grades	MSE	7.714	**7.503**	7.751	7.580	7.709	7.587	7.796	7.714
	SE	0.110	0.110	0.130	0.110	0.060	0.120	0.160	0.150
	R^2_{adj}	0.239	0.222	**0.248**	0.225	0.245	0.227	0.243	0.219
Delinquency	MSE	2.049	**2.033**	2.068	2.045	2.059	2.060	2.080	2.072
	SE	0.020	0.040	0.030	0.030	0.020	0.040	0.030	0.030
	R^2_{adj}	**0.100**	0.072	0.098	0.073	**0.100**	0.073	0.096	0.074
Cannabis Consumption	MSE	2.664	2.643	2.668	**2.640**	2.690	2.695	2.689	2.681
	SE	0.010	0.030	0.010	0.020	0.010	0.030	0.030	0.020
	R^2_{adj}	0.479	0.462	**0.481**	0.467	0.476	0.455	0.479	0.461
Ecstasy Consumption	MSE	2.007	**1.979**	2.006	1.984	2.027	1.999	2.044	2.008
	SE	0.010	0.010	0.020	0.010	0.010	0.010	0.030	0.020
	R^2_{adj}	0.250	0.246	**0.255**	0.248	0.246	0.243	0.243	0.242
Objectivity	SEL	0.130	**0.127**	0.132	0.130	0.132	0.128	0.135	0.131
	SE	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	R^2_{adj}	0.420	0.413	0.416	0.408	0.418	**0.423**	0.410	0.412
	AUC	0.810	**0.820**	0.810	**0.820**	0.810	**0.820**	0.810	0.810
Breast Cancer	SEL	0.031	**0.030**	0.033	0.032	0.032	0.031	0.038	0.033
	SE	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	R^2_{adj}	0.859	**0.862**	0.852	0.854	0.857	**0.862**	0.836	0.853
	AUC	**0.960**	**0.960**	0.950	0.950	**0.960**	**0.960**	0.950	0.950
Sleep Quality	SEL	0.216	**0.215**	0.218	0.218	0.217	0.217	0.221	0.221
	SE	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	R^2_{adj}	0.112	0.098	0.113	0.094	**0.114**	0.094	0.104	0.091
	AUC	0.660	0.660	0.660	0.660	0.650	0.650	0.650	0.650
University Graduation	SEL	0.072	**0.071**	0.073	**0.071**	0.073	0.072	0.073	0.072
	SE	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	R^2_{adj}	0.691	0.692	0.691	**0.693**	0.691	0.692	0.691	0.692
	AUC	0.910	0.910	0.910	0.910	0.910	0.910	0.910	0.910
ADHD	SEL	**0.063**	0.066	0.065	0.064	0.065	0.065	0.069	0.066
	SE	0.010	0.010	0.010	<0.001	0.010	0.010	0.010	0.010
	R^2_{adj}	**0.725**	0.702	0.716	0.708	0.722	0.716	0.709	0.710
	AUC	**0.920**	**0.920**	0.910	**0.920**	0.910	**0.920**	0.910	**0.920**

Note. The values denote the means over the ten repeats of the mean squared error (MSE), squared error loss (SEL), adjusted R^2, and area under the receive curve (AUC). The SE indicates the standard error of the MSE or SEL. Best values are in bold.

Sparsity

Table 3 and 4 show a summary of the number of predictors and number of base learners selected across folds and repeats. As expected, the lambda-1se criterion results in sparser models than the lambda-min criterion, both in terms of variable and base learner selection. As shown in Figure 2, in regression datasets, when the lambda-1se criterion is used, the relaxed lasso leads to the sparsest model in terms of variable selection and the relaxed adaptive lasso in terms of base learner selection. Interestingly, under the lambda-min criterion, the differences in sparsity between the lasso approaches is diminished and the variance is larger suggesting that when the base learner selection is less stable the differences between lasso approaches become less visible. In classification datasets, the relaxed adaptive lasso leads to the sparsest model, both in terms of variable and base learner selection. The standard lasso yields the least sparse model in all datasets.

Table 3

Number of Variables Selected

		Lasso		Relaxed Lasso		Adaptive Lasso		Relaxed Adaptive	
Dataset		1se	min	1se	min	1se	min	1se	min
High School Grades	*M*	14.25	21.13	**8.67**	20.08	14.69	20.58	10.70	19.61
	SD	2.11	1.89	3.09	2.77	2.10	1.89	3.14	2.75
Delinquency	*M*	12.08	20.32	**8.89**	18.71	12.54	19.07	10.05	17.86
	SD	1.86	2.30	2.20	3.15	2.49	2.19	2.53	2.64
Cannabis Consumption	*M*	10.04	11.92	**8.62**	11.75	11.01	11.95	10.33	11.93
	SD	1.24	0.31	1.25	0.66	1.03	0.26	1.36	0.26
Ecstasy Consumption	*M*	8.24	10.75	**6.15**	10.16	8.89	10.58	7.78	10.24
	SD	0.85	0.82	1.11	1.13	1.06	0.85	1.25	0.93
Objectivity	*M*	25.20	35.60	22.15	33.53	20.80	30.14	**14.79**	29.75
	SD	3.05	3.72	3.89	5.14	3.27	3.96	4.21	4.50
Breast Cancer	*M*	11.96	13.71	11.35	13.63	10.80	12.04	**8.69**	11.46
	SD	1.32	1.42	1.89	1.51	1.25	1.33	2.03	2.06
Sleep Quality	*M*	8.39	10.78	**6.32**	9.69	8.31	10.45	**6.80**	9.13
	SD	1.32	1.73	1.81	2.07	1.20	1.76	1.34	2.04
University Graduation	*M*	20.98	25.30	**18.92**	24.59	20.88	24.66	**19.35**	24.27
	SD	1.90	1.01	2.17	1.46	1.88	1.08	2.25	1.37
ADHD	*M*	15.92	21.38	15.08	21.37	12.05	16.81	**10.23**	16.27
	SD	2.45	3.27	4.23	3.16	2.35	2.70	4.02	3.24

Note. *M* indicates the mean number of predictors selected across folds and repeats, and SD its standard deviation. Latter can be interpreted as the standard error. Best values are in bold.

Table 4

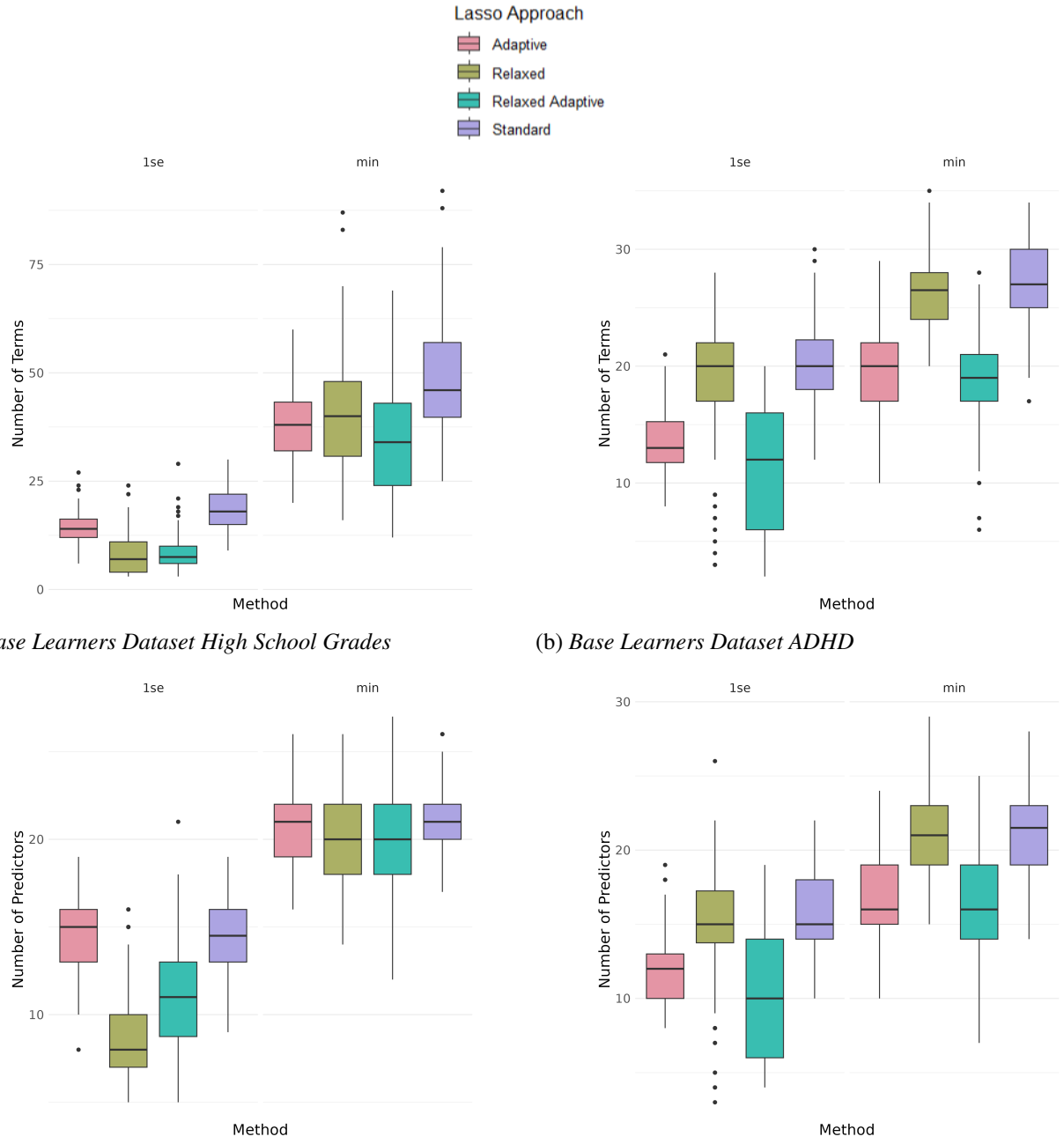
Number of Base Learners Selected

		Lasso		Relaxed Lasso		Adaptive Lasso		Relaxed Adaptive	
Dataset		1se	min	1se	min	1se	min	1se	min
High School Grades	*M*	18.71	48.85	**8.30**	40.94	14.25	38.47	**8.32**	34.35
	SD	4.85	12.02	5.13	13.57	3.79	9.09	4.20	12.30
Delinquency	*M*	18.65	58.27	11.06	50.97	13.12	43.78	**7.74**	37.28
	SD	3.76	14.57	4.55	19.81	4.35	11.23	3.69	14.56
Cannabis Consumption	*M*	43.75	115.11	31.56	100.26	34.86	104.54	**24.78**	94.67
	SD	8.45	28.02	11.43	33.10	12.15	26.10	12.67	30.92
Ecstasy Consumption	*M*	25.34	62.9	14.40	53.69	17.13	51.35	**10.46**	44.53
	SD	3.73	18.1	5.34	17.72	5.03	14.80	5.55	17.56
Objectivity	*M*	34.57	65.44	29.04	57.76	21.19	44.58	**11.71**	42.73
	SD	5.67	14.64	7.47	17.22	6.11	10.45	6.35	11.58
Breast Cancer	*M*	27.89	33.12	25.74	33.08	19.07	24.70	**10.66**	21.54
	SD	3.39	4.13	6.31	4.56	2.65	2.91	6.30	6.67
Sleep Quality	*M*	13.41	24.44	7.61	19.30	9.05	19.61	**5.15**	13.89
	SD	3.27	6.84	3.88	7.89	2.47	6.57	2.23	8.05
University Graduation	*M*	55.73	105.10	44.50	91.90	42.22	82.35	**33.19**	76.52
	SD	9.36	13.09	8.46	21.02	7.42	10.73	7.31	16.89
ADHD	*M*	20.15	26.88	18.72	26.48	13.74	19.94	**11.29**	19.12
	SD	3.51	3.66	5.53	3.20	2.90	3.48	5.19	3.77

Note. M indicates the mean number of base learners selected across folds and repeats, and SD its standard deviation. Latter can be interpreted as the standard error. Best values are in bold.

Figure 2

Sparsity: Number of Base Learners and Predictors Selected



(a) *Base Learners Dataset High School Grades*

(b) *Base Learners Dataset ADHD*

(c) *Predictors Dataset High School Grades*

(d) *Predictors Dataset ADHD*

Stability

Stability was assessed with three different measures; the stability of variable selection, the distances between predictions, and the distances between importance measures. Table 5 shows the results. In regression datasets, the stability on all three measures is higher for the lambda-1se than the lambda-min criterion. On the other hand, for classification datasets, no differences between the lambda-1se and lambda-min criterion are found.

Stability of Variable Selection

Figure 3 shows the stability of variable selection with confidence intervals. In dataset High School Grades, under the lambda 1se criterion, the standard lasso has a significantly higher stability, while the adaptive and relaxed lasso are more stable than the relaxed adaptive lasso. When the lambda-min criterion is used, the stability is significantly lower and the differences between lasso approaches are not present. The effect of lasso approach differs per dataset. As seen in Table 5, for datasets Delinquency and Ecstasy Consumption, both the relaxed and standard lasso yield the highest stability, and for dataset Cannabis Consumption only the relaxed lasso.

In classification datasets, no clear pattern of significant differences in stability of variable selection are found, and their seems to be an interaction effect between lasso approach and criterion. Under the lambda-1se criterion, in half of the classification datasets the standard and adaptive lasso seem to be slightly more stable, and in the remaining datasets, the relaxed and standard lasso. In dataset ADHD, as can be seen in Figure 3, under the lambda-1se criterion, the standard and adaptive lasso perform best and under the lambda-min criterion the adaptive and relaxed adaptive lasso.

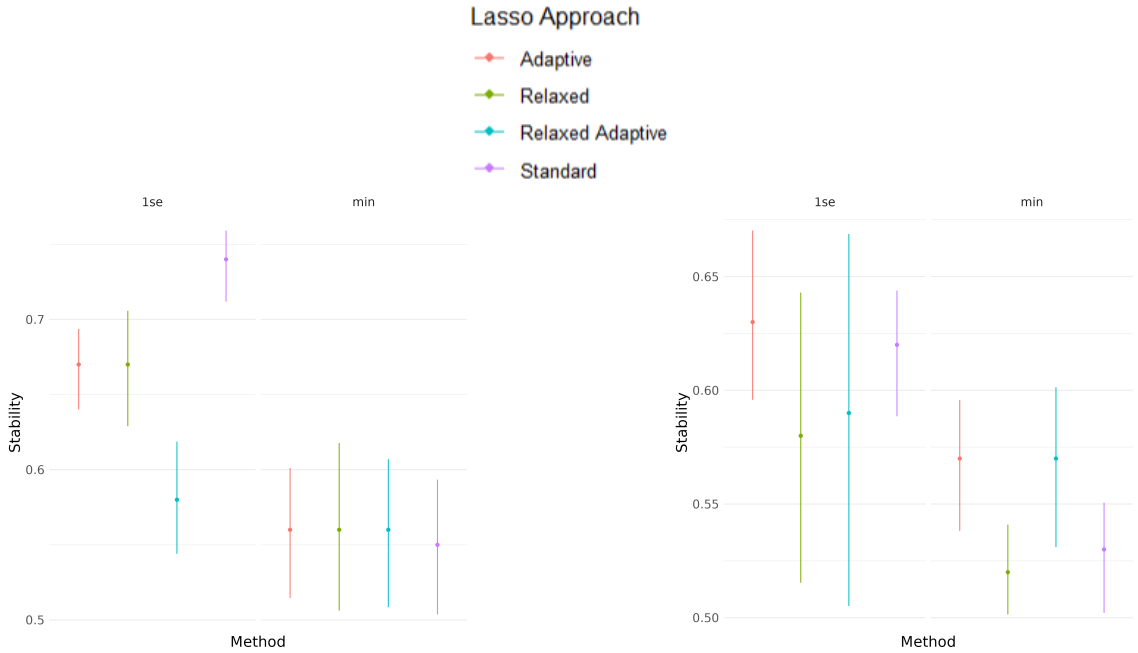
An interesting finding is that the results differ for regression and classification datasets. In all regression datasets, under the lambda-min criterion the stability is significantly lower with no difference between lasso approaches. In classification datasets, the difference between lambda criteria is not significant and the patterns between lasso approaches differ per lambda criterion, indicating an interaction effect. No clear conclusions can be drawn on which lasso approach performs best, as the results differ across datasets. There is a slight indication that for regression datasets the relaxed and standard lasso are the most stable and for classification datasets the adaptive and standard lasso, however the results do not reach significance.

Stability of Importance Measures

Table 5 shows that in all datasets the standard lasso with lambda-1se criterion yields the smallest distances. Similarly to the stability of variable selection, the importance measures are more stable for the lambda-1se than the lambda-min criterion. Interestingly, again this relationship is stronger for continuous than for binary outcomes. The results of the linear models show that the interaction effects between lasso approach and lambda criterion are significant in all datasets. This makes it harder to interpret the main effects. Inspecting the visualizations in Figure 4 and B8, under the lambda-1se criterion, the standard lasso seems to yield slightly more stable importance measures, followed by the adaptive and the relaxed lasso, and the most unstable is the relaxed adaptive lasso. In datasets Objectivity and Graduation, both the adaptive and the standard lasso are the most stable. No difference between the lasso approaches are observed under the lambda-min criterion for any dataset.

Figure 3

Stability of Variable Selection



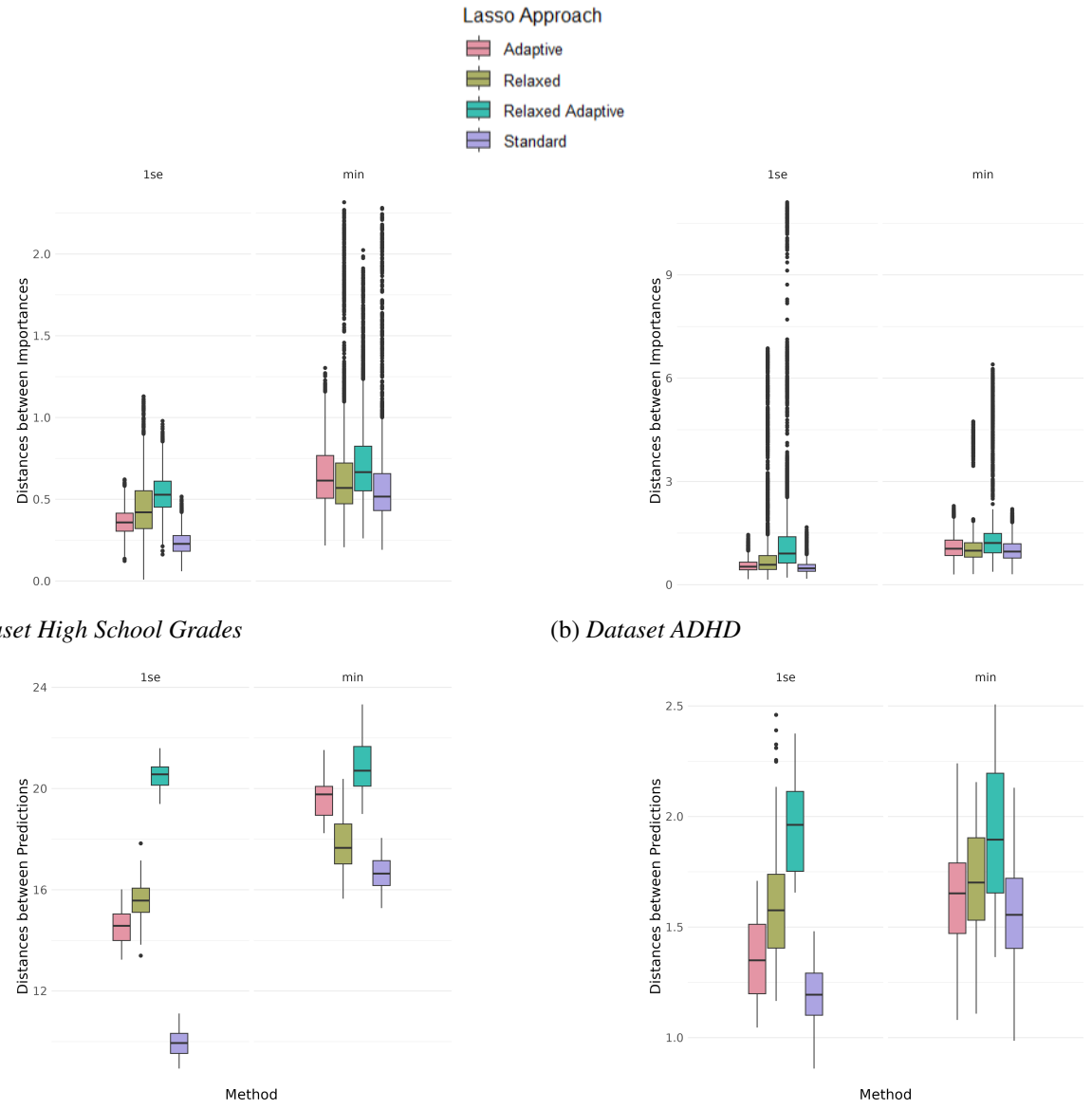
(a) *Dataset High School Grades* (b) *Dataset ADHD*

Stability of Predictions

Figure 4 displays the distribution of distances between predictions for dataset High School Grades and ADHD. A similar pattern is found in all datasets. The standard lasso yields the smallest distances, meaning the model is more stable. For the adaptive and relaxed lasso the distances are similar and smaller than for the relaxed adaptive lasso. The results of the linear models show that in all datasets, the interaction effects between lasso approach and lambda criterion are significant. Inspecting the visualizations in Figure B9, it seems that the differences between lasso approaches are less pronounced under the lambda-min criterion, and the effect of lambda criterion on stability is different per lasso approach. In all datasets, the differences between lambda criteria is the strongest for the standard lasso, showing significantly more stable results under the lambda-1se criterion. In three datasets, the relaxed adaptive lasso results in higher distances under the lambda-1se criterion, which is the opposite effect as for the other lasso approaches. Furthermore, looking at the distribution of the MSE/SEL in Figure B1, the variance seems to be highest for the relaxed adaptive lasso, which is in line with the finding that the predictions by the relaxed adaptive lasso are the most unstable.

Figure 4

Distribution of Distances between Importance Measures and Predictions



(a) *Dataset High School Grades*

(b) *Dataset ADHD*

(c) *Dataset High School Grades*

(d) *Dataset ADHD*

Table 5

Stability measures by model

		Lasso		Relaxed Lasso		Adaptive Lasso		Relaxed Adaptive	
Dateset		1se	min	1se	min	1se	min	1se	min
High School Grades	$\hat{\phi}(Z)$	**0.74**	0.55	0.67	0.56	0.67	0.56	0.58	0.56
	d(importances)	**0.23**	0.59	0.44	0.65	0.36	0.65	0.53	0.73
	d(predictions)	**9.95**	16.66	15.59	17.83	14.53	19.58	20.50	20.96
Delinquency	$\hat{\phi}(Z)$	**0.69**	0.35	0.66	0.39	0.60	0.44	0.55	0.47
	d(importances)	**0.11**	0.32	0.20	0.37	0.16	0.34	0.23	0.36
	d(predictions)	**5.70**	10.74	9.18	11.42	8.67	12.38	12.36	12.80
Cannabis Consumption	$\hat{\phi}(Z)$	0.37	0.01	**0.67**	0.04	0.19	0.00	0.30	0.01
	d(importances)	**0.15**	0.54	0.27	0.54	0.28	0.69	0.31	0.73
	d(predictions)	**9.34**	13.90	12.60	14.16	14.77	17.65	17.21	17.38
Ecstasy Consumption	$\hat{\phi}(Z)$	0.73	0.45	**0.75**	0.45	0.66	0.45	0.66	0.44
	d(importances)	**0.07**	0.25	0.17	0.28	0.13	0.28	0.18	0.31
	d(predictions)	**5.89**	9.24	10.38	10.06	10.26	11.38	14.21	12.18
Objectivity	$\hat{\phi}(Z)$	**0.65**	**0.65**	0.63	0.61	0.61	**0.64**	0.56	0.61
	d(importances)	**0.28**	0.62	0.56	0.64	0.33	0.54	0.53	0.58
	d(predictions)	**2.06**	2.95	2.91	3.19	2.70	3.01	3.77	3.27
Breast Cancer	$\hat{\phi}(Z)$	0.77	0.71	0.75	0.71	**0.80**	0.76	0.72	0.74
	d(importances)	**0.52**	0.87	1.19	0.88	0.58	0.81	1.61	0.93
	d(predictions)	**1.38**	1.70	1.74	1.76	1.57	1.76	2.62	1.97
Sleep Quality	$\hat{\phi}(Z)$	**0.64**	0.54	0.60	0.55	**0.64**	0.56	0.57	0.56
	d(importances)	**0.19**	0.34	0.30	0.39	0.26	0.46	0.39	0.50
	d(predictions)	**1.43**	1.95	2.09	2.22	2.09	2.46	2.93	2.78
Graduation	$\hat{\phi}(Z)$	0.62	0.63	0.66	0.58	0.57	**0.68**	0.57	0.60
	d(importances)	**0.45**	0.82	0.92	1.01	0.57	0.90	0.93	1.07
	d(predictions)	**2.21**	3.31	2.80	3.49	2.99	3.83	3.50	4.05
ADHD	$\hat{\phi}(Z)$	**0.62**	0.53	0.58	0.52	**0.63**	0.57	0.59	0.57
	d(importances)	**0.52**	1.01	1.03	1.07	0.57	1.08	1.47	1.50
	d(predictions)	**1.19**	1.58	1.65	1.71	1.35	1.63	1.97	1.93

Note. The $\hat{\phi}(Z)$ measure indicates the stability of variable selection and lies between 0 and 1, with a higher value indicating higher stability (Nogueira et al., 2018). The d(importances) and d(predictions) refer to the euclidean distances between the importance measures and predictions respectively. Larger distance indicates lower stability. Note that the distances are dependent on the measurement scale and therefore not comparable across different datasets. Best values for each measure are in bold.

Discussion

In this study, we evaluated if model accuracy, sparsity, and stability of prediction rule ensembles can be improved by using the adaptive, relaxed, or relaxed adaptive lasso instead of the standard lasso. The standard lasso faces some drawbacks, such as difficulties in finding a regularization parameter that yields both optimal variable selection and optimal predictive accuracy (Dalalyan et al., 2017). The relaxed and adaptive lasso address this problem by introducing additional parameters, that mitigate over-shrinkage of large coefficients (Meinshausen, 2007; Zou, 2006). However, there is limited research about comparing these variations of the lasso in the context of prediction rule ensembles.

The results suggest that sparsity of prediction rule ensembles can be improved by using the relaxed or adaptive lasso, but at the cost of stability. While the relaxed adaptive lasso selects the sparsest model, this is at the cost of significantly lower stability of predictions and slightly lower stability of variable selection and importance measures. The relaxed and adaptive lasso outperform the standard lasso in terms of sparsity while maintaining a high level of stability in variable selection, but make less stable predictions. In terms of accuracy the four lasso variations perform equally well. Furthermore, the lambda-1se criterion explains more variance than the lambda-min criterion when adjusting for the number of base learners selected, selects sparser models, and for regression datasets also improves stability. The differences between lasso approaches diminish under the lambda-min criterion in regression datasets as the results become highly unstable. In classification datasets, no significant differences in stability between the lambda criteria are found.

Predictive Accuracy

In terms of lasso approach, there are no significant differences in predictive accuracy found. In most regression datasets, the adjusted R^2 is highest for the relaxed lasso but the effect did not reach significance. Moreover, the lambda-1se criterion yields a significantly higher adjusted R^2 than the lambda-min criterion. The lambda-1se criterion is better at balancing sparsity and accuracy, while the lambda-min criterion is prone to select more noise variables which decreases the value of the adjusted R^2. Past research has shown that under low signal-to-noise ratio (SNR), the adaptive lasso and relaxed lasso perform equally well in terms of accuracy as the standard lasso, but yield higher predictive accuracy when the SNR is high (Meinshausen, 2007; Hastie et al., 2020; Zou, 2006). In this study, no difference in lasso approach is observed in relation to the SNR. These differences in findings might be explained by the fact that the previous studies compared variations of the lasso in linear penalized regression while the current study investigates the role in the context of prediction rule ensembles.

Sparsity

In terms of model sparsity, the results of this study are in line with previous research, showing that the relaxed, adaptive, and relaxed adaptive lasso select sparser models than the standard lasso (Huang et al., 2008; Meinshausen, 2007; Zhang et al., 2022). While the previous studies focused on linear models and assessed sparsity by the number of variables selected, this study additionally compared continuous and binary outcomes, and also evaluated the number of base learners selected. For both outcome types, the relaxed adaptive lasso is the sparsest model in terms

of base learner selection, followed by comparable results between the relaxed and adaptive lasso, and the least sparse the standard lasso. In most regression datasets, the relaxed lasso leads to the sparsest model in terms of variable selection and in classification datasets the relaxed adaptive lasso.

Furthermore, an interesting finding is that for continuous outcomes, the differences in sparsity between lasso approaches disappear when the lambda-min instead of the lambda-1se criterion is used. On the other hand, when the outcome is binary, no interaction between lasso approach and lambda criterion is observed. In most datasets, the variance of number of base learners selected is much higher for the lambda-min than the lambda-1se criterion, indicating lower stability. Nogueira et al. (2018) found that models that select more noise variables are less stable. In the context of prediction rule ensembles, this suggests that under the lambda-min criterion, more irrelevant base learners are selected, making the model less stable and reducing the effects of the lasso variations.

Stability

Previous research suggests that the standard lasso is unstable in variable selection when multicollinearity is high (Zhao & Yu, 2006). That is because in the presence of multicollinearity, the standard lasso may randomly select one of the correlated predictors (Dalalyan et al., 2017). The adaptive lasso addresses this by adding the weight parameter to adapt the shrinkage on each predictor individually. However, research comparing the stability between the different variations of the lasso is lacking.

The results of this study suggest that for regression datasets the relaxed and standard lasso seem slightly more stable in variable selection and in classification datasets the adaptive and standard lasso, but the pattern of results are not consistent across all datasets. In terms of importance measures and predictions the standard lasso seems the most stable. The least stable predictions are made by the relaxed adaptive lasso. This suggests that as more parameters are tuned in the model, the predictions become more unstable but this has only small effect on the stability of variable selection. Both the γ parameter in the relaxed lasso, and the weight ω in the adaptive lasso, are tuned based on the data. This might increase the risk of over-fitting as data dependency increases.

Furthermore, all regression datasets have significantly higher stability when the lambda-1se compared to the lambda-min criterion is used. Nogueira et al. (2018) found that when applying the ordinary lasso in logistic regression, a slightly larger lambda value is needed for achieving optimal stability compared to achieving optimal predictive accuracy. By sacrificing a small loss of accuracy, the stability increases as the risk of over-fitting is reduced. This is line with the findings for regression datasets, which show that the lambda-1se criterion yields significantly higher stability. More research is needed to explain why this finding could not be replicated in classification datasets. Nogueira et al. (2018) found the effect is stronger when more noise variables are present. However, in the current study no relationship between the SNR and stability is observed.

Strength and Limitations

The results of this study provide valuable insights into the use of the relaxed and adaptive lasso, or their combination, in prediction rule ensembles. This is the first study investigating model performance of these methods in the context of rule ensembles and the first to compare the lasso variations both with the lambda-min and lambda-1se criterion. Previous research on the lasso variations primarily focused on prediction accuracy and sparsity, lacking

evaluation of its stability. This study sheds light on the impact of variations of the lasso on stability of variable selection, importance measures and predictions.

An advantage of this study is that two outcome variables were compared, providing information on possible differences between linear and geneneralized linear models. Moreover, a strength is the use of repeated cross-validation, which makes the results less dependent on a single choice of folds. The datasets used differ in the number of participants and number of predictors, making the results more generalizable. Future research could extend this study by looking at different outcome types such as poisson, cox, or multinomial, and at high-dimensional data.

A limitation is that standard errors for predictive accuracy were only computed across repeats, but not computed across folds and repeats of the cross-validation. Thus, they quantify uncertainty due to different possible ways to separate a dataset into 10 equally-sized folds, but do not quantify uncertainty due to taking different subsamples of observations. Another limitation is that the results for stability of predictions are not standardized and therefore not comparable across datasets. While this study compared the lasso approaches in nine datasets, an additional analysis aggregating the results from all datasets is missing. In addition, future research could investigate which factors influenced the differences in sparsity and stability, and may help assessing how to further improve prediction rule ensembles. Past research suggests that SNR is an influencing factor but in the current study SNR did not systematically impact performance differences between lasso approaches.

Conclusion

Based on the results of this study, practical advice can be given on which lasso approach should be chosen for optimal performance of prediction rule ensembles. The choice of lasso approach should be dependent on if accuracy, sparsity, or stability is prioritized, or the trade-off between them. If accuracy is most important, the standard lasso with lambda-min criterion performs best, if sparsity is more relevant the relaxed adaptive lasso and for optimal stability the standard lasso with lambda-1se criterion. To achieve optimal trade-off between the three measures the relaxed or adaptive lasso with lambda-1se criterion are the best choices as both yield sparser models than the standard lasso, higher stability than the relaxed adaptive lasso, and maintain the same level of accuracy. The results of this study can be summarized as follows:

- Standard lasso with lambda-min criterion performs best in terms of prediction error, but is the least sparse.
- For optimal trade-off between accuracy and sparsity, lambda-1se criterion performs significantly better than lambda-min. No significant differences between lasso approaches.
- The relaxed adaptive lasso with lambda-1se criterion selects sparsest model.
- Stability of variable selection is equally high for all lasso approaches and higher for lambda-1se than lambda-min criterion.
- The standard lasso with lambda-1se criterion yields most stable importance measures and predictions.
- The least stable predictions are made by the relaxed adaptive lasso.

- In regression datasets, the lambda-1se criterion yields significantly more stable results than the lambda-min criterion.
- The effect of lambda-criterion is more pronounced in regression than in classification datasets.

Acknowledgements

We would like to acknowledge following sources of datasets in our study: The UCI Machine Learning Repository, Mendeley data, and studies by Roth and Herzberg (2004) and Trognon and Richard (2022), who made their data available.

References

Ayyadevara, V. K. (2018). Gradient boosting machine. In *Pro machine learning algorithms : A hands-on approach to implementing algorithms in python and R* (pp. 117–134). Apress. https://doi.org/10.1007/978-1-4842-3564-5_6

Cortez, P. (2014). Student Performance [Dataset]. *UCI Machine Learning Repository*. https://doi.org/https://doi.org/10.24432/C5TG7T

Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. https://doi.org/https://hdl.handle.net/1822/8024

Dalalyan, A. S., Hebiri, M., & Lederer, J. (2017). On the prediction performance of the Lasso. *Bernoulli*, *23*(1), 552–581. https://doi.org/10.3150/15-BEJ756

Fehrman, E., Egan, V., & Mirkes, E. (2016). Drug consumption (quantified) [Dataset]. *UCI Machine Learning Repository*. https://doi.org/https://doi.org/10.24432/C5TC7S

Fehrman, E., Muhammad, A. K., Mirkes, E. M., Egan, V., & Gorban, A. N. (2017). The five factor model of personality and evaluation of drug consumption risk [Dataset]. *Data Science: Innovative developments in data analysis and clustering*, 231–242.

Fokkema, M. (2020). Fitting prediction rule ensembles with R package pre. *Journal of Statistical Software*, *92*(12), 1–30. https://doi.org/10.18637/jss.v092.i12

Fokkema, M., & Strobl, C. (2020). Fitting prediction rule ensembles to psychological research data: An introduction and tutorial. *Psychological Methods*, *25*(5), 636. https://doi.org/https://psycnet.apa.org/doi/10.1037/met0000256

Friedman, J., Tibshirani, R., & Hastie, T. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/https://www.jstor.org/stable/2699986

Friedman, J. H., & Popescu, B. E. (2008). Predictive learning via rule ensembles. *The Annals of Applied Statistics*, *2*(3), 916–954. https://doi.org/10.1214/07-AOAS148

Hastie, T., Tibshirani, R., & Tibshirani, R. (2020). Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons. *Statistical Science*, *35*(4), 579–592. https://doi.org/10.1214/19-STS733

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674. https://doi.org/https://www.jstor.org/stable/27594202

Huang, J., Ma, S., & Zhang, C.-H. (2008). Adaptive lasso for sparse high-dimensional regression models. *Statistica Sinica*, 1603–1618. https://doi.org/https://www.jstor.org/stable/24308572

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Martins, M. V., Tolledo, D., Machado, J., Baptista, L. M., & Realinho, V. (2021). Early prediction of student's performance in higher education: A case study [Dataset]. *Trends and Applications in Information Systems and Technologies: Volume 1 9*, 166–175.

Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis*, *52*(1), 374–393. https://doi.org/https://doi.org/10.1016/j.csda.2006.12.019

Meinshausen, N., & Bühlmann, P. (2006). Variable selection and high-dimensional graphs with the lasso. *Annals of Statistics*, *34*, 1436–1462.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*, 21. https://doi.org/https://doi.org/10.3389/fnbot.2013.00021

Nogueira, S., Sechidis, K., & Brown, G. (2018). On the stability of feature selection algorithms. *The Journal of Machine Learning Research*, *18*(1), 6345–6398.

Norbury, R., & Evans, S. (2018). Time to think: Subjective sleep quality, trait anxiety and university start time [Dataset]. *Mendeley Data*. https://doi.org/10.17632/mxsjysrt8j.1

Norbury, R., & Evans, S. (2019). Time to think: Subjective sleep quality, trait anxiety and university start time. [Dataset]. *Psychiatry Research*, *271*, 214–219. https://doi.org/https://doi.org/10.1016/j.psychres.2018.11.054

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Realinho, V., Vieira Martins, M., Machado, J., & Baptista, L. (2021). Predict students' dropout and academic success [Dataset]. *UCI Machine Learning Repository*. https://doi.org/https://doi.org/10.24432/C5MC89

Rizk, Y., & Awad, M. (2018). Sports articles for objectivity analysis [Dataset]. *UCI Machine Learning Repository*. https://doi.org/https://doi.org/10.24432/C5801R

Roth, M., & Herzberg, P. Y. (2004). A validation and psychometric examination of the Arnett Inventory of Sensation Seeking (AISS) in German Adolescents [Dataset]. *European Journal of Psychological Assessment*, *20*(3), 205–214. https://doi.org/https://doi.org/10.1027/1015-5759.20.3.205

Street, W. N., Wolberg, W. H., & Mangasarian, O. L. (1993). Nuclear feature extraction for breast tumor diagnosis [Dataset]. *Biomedical Image Processing and Biomedical Visualization*, *1905*, 861–870. https://doi.org/https://doi.org/10.1117/12.148698

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Trognon, A., & Richard, M. (2022). Questionnaire-based computational screening of adult ADHD [Dataset]. *BMC psychiatry*, *22*(1), 401. https://doi.org/https://doi.org/10.1186/s12888-022-04048-1

Wolberg, W., Mangasarian, O., Street, N., & Street, W. (1995). Breast Cancer Wisconsin (Diagnostic) [Dataset]. *UCI Machine Learning Repository*. https://doi.org/https://doi.org/10.24432/C5DW2B

Zhang, R., Zhao, T., Lu, Y., & Xu, X. (2022). Relaxed adaptive lasso and its asymptotic results. *Symmetry*, *14*(7), 1422. https://doi.org/https://doi.org/10.3390/sym14071422

Zhao, P., & Yu, B. (2006). On model selection consistency of lasso. *The Journal of Machine Learning Research*, *7*, 2541–2563.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429. https://doi.org/https://www.jstor.org/stable/27639762

Appendix A
Dataset Descriptions

Dataset 1: High School Grades

In this dataset, the Portuguese language grades of secondary-school students are predicted based on demographics, such as gender and family educational background, social-emotional factors such as relationships and alcohol consumption, and school related factors such as number of school absences. The data was gathered 2005 and 2006 at two schools in Portugal using school reports and a questionnaire (Cortez & Silva, 2008).

Dataset 2: Youth Delinquency

This dataset comes from a study by Roth and Herzberg (2004) who investigated the validity of the Arnett Inventory of Sensation Seeking (AISS) scale and the relationship between sensation seeking and delinquency in adolescence. Delinquency was measured with 5 questions from the Youth Self-Report (YSR)- Subscale Delinquent Behaviour questionnaire. In the current study the sum of these questions is taken and predicted by the validated items from the sensation seeking questionnaire and demographics such as gender and age. After removing 160 observations due to missing values, the sample consists of 1076 German high school students.

Dataset 3: Drug Consumption

This dataset contains information about personality, demographics, and drug consumption for 18 different illegal drugs. The current study focuses on predicting cannabis and ecstasy consumption by the five personality traits, measured with the Revised NEO-Five Factor Inventory (NEO-FFI-R), and scores from the Barratt Impulsiveness Scale (BIS-11) and the Impulsiveness Sensation Seeking Scale (ImpSS). The data were collected in 2011 and 2012 with an online questionnaire. Participants are adults, mainly from English-native speaking countries (Fehrman et al., 2017).

Dataset 5: Objectivity of Article

This dataset contains data about the linguistic properties of 1000 sports articles (Rizk & Awad, 2018). Features were extracted using the Stanford Part-Of-Speech (POS) tagger software. Examples are grammatical properties such as the frequency of using adverbs, plural/singular nouns, and determiners, leading to a collection of 59 features. The outcome variable is to predict if the article is objective or subjective (Rizk & Awad, 2018).

Dataset 6: Breast Cancer

The Breast Cancer dataset comes from an oncological study by Street et al. (1993) who analyzed 569 images of cell nuclei. Using a computer vision diagnostic system they extracted ten features from the nuclei, such as the radius, smoothness, and symmetry. For each of these ten features they calculated the mean, maximum, and standard deviation. Thus, in total there are 30 predictor variables, three for each of the ten features. It is notable that high multicollinearity is expected for the measure of the same characteristic. The outcome variable is whether the cell nuclei come from a benign or malignant breast cancer tumour.

Dataset 7: Sleep Quality

Norbury and Evans (2019) investigated the relationship between sleep quality, mental health, and preferred university starting time. For this, 546 university students from two universities in England completed an online survey. The outcome variable is subjective sleep quality measured with the Pittsburgh Sleep Quality Index (PSQI) and dichotomized into poor/good sleep. Examples of predictors are trait anxiety, day-time dozing, sleep duration, preferred university starting time, chronotype (morning versus evening person) as well as tobacco, alcohol, and caffeine consumption (Norbury & Evans, 2019).

Dataset 8: University Graduation

This dataset comes from research by Martins et al. (2021) who studied methods to identify students at risk of academic failure in order to support these students at an early stage. The participants are undergraduate students at the Polytechnic Institute of Portalegre (IPP), Portugal. Data was collected from the academic year 2008/09 to 2018/19 and contains information about demographics, socio-economic factors such as student debts or parents' employment situation, and student's academic path such as admission grades and high school grades. The outcome is if the student graduated on time, graduated with a delay or did not graduate. In the current study, the outcome is dichotomized and only includes students who either graduated on time or not graduated. Based on this, 794 observations were excluded of students who graduated with a delay, leading to a final sample size of 3630 students.

Dataset 9: ADHD

The last dataset comes from an online study investigating the psychometric properties of an ADHD screening scale (Trognon & Richard, 2022). The sample are adults from the general French population of which 110 have been diagnosed with ADHD and the remaining 110 participants serve as a control group. Predictors are demographics such as age and gender, and items from three questionnaires; 43 items adapted from the DSM-5, 21 items from the Depression Anxiety Stress Scale (DASS), and 26 items of the the Scale of Adherence to the Values of the Ideal Democracy (AVDI). The last scale serves as a control scale and is not expected to be correlated to the diagnosis of ADHD (Trognon & Richard, 2022).

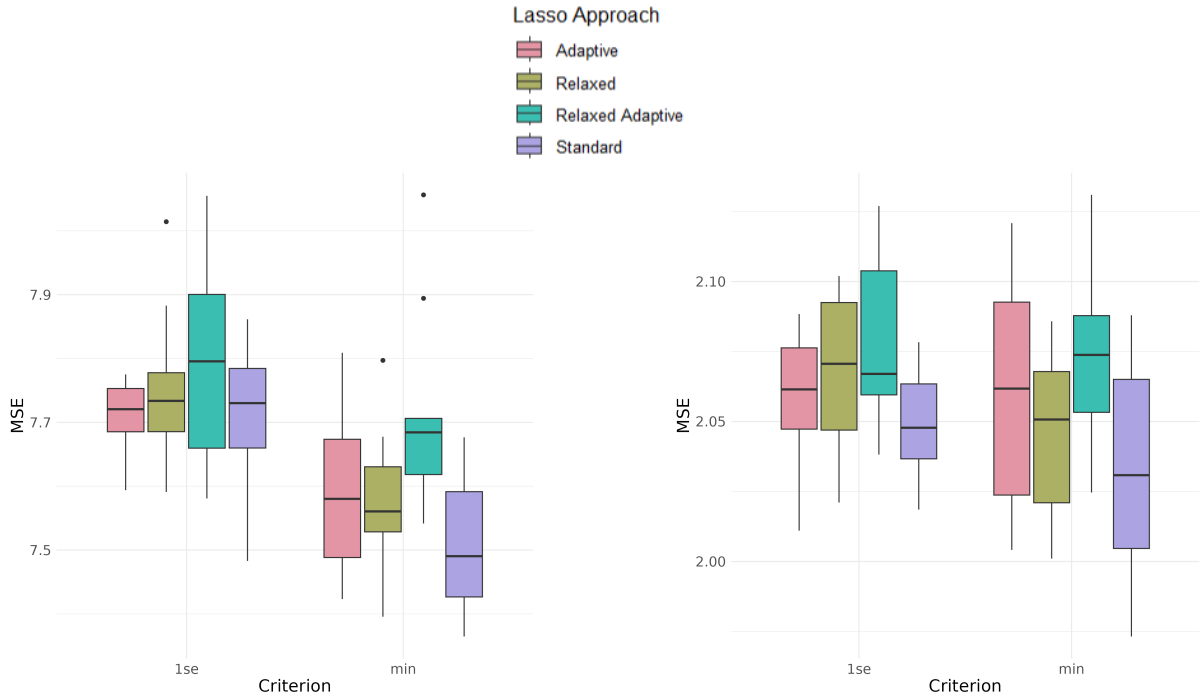
Appendix B
Visualizations

Accuracy

MSE/SEL

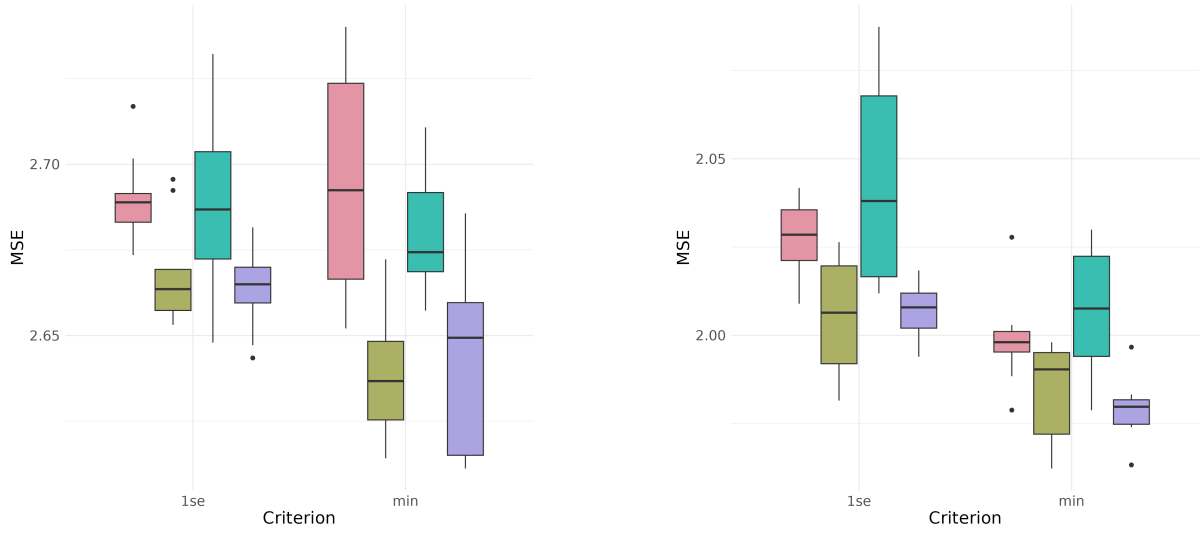
Figure B1

MSE by Dataset for Continuous Outcomes



(a) *Dataset 1: High School Grades*

(b) *Dataset 2: Delinquency*

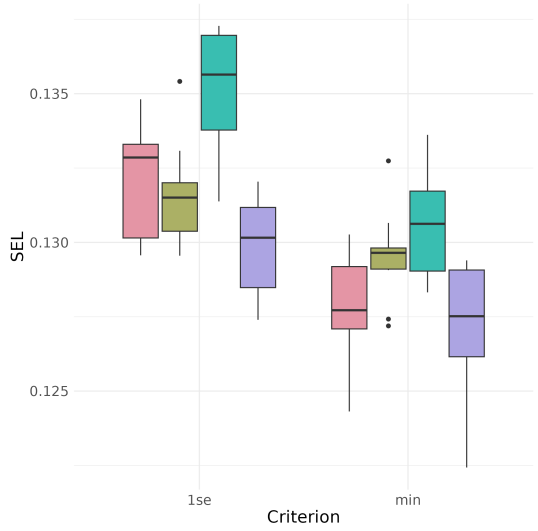


(c) *Dataset 3: Cannabis Consumption*

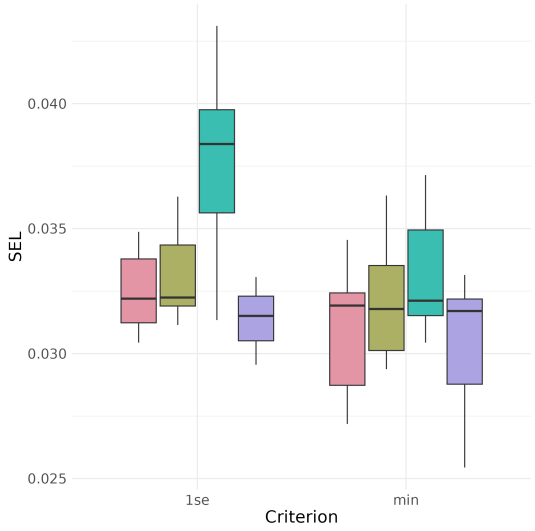
(d) *Dataset 4: Ecstasy Consumption*

Figure B2

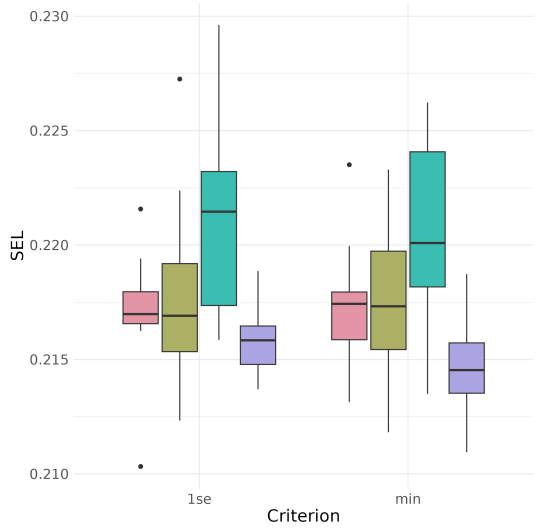
SEL by Dataset for Binary Outcomes



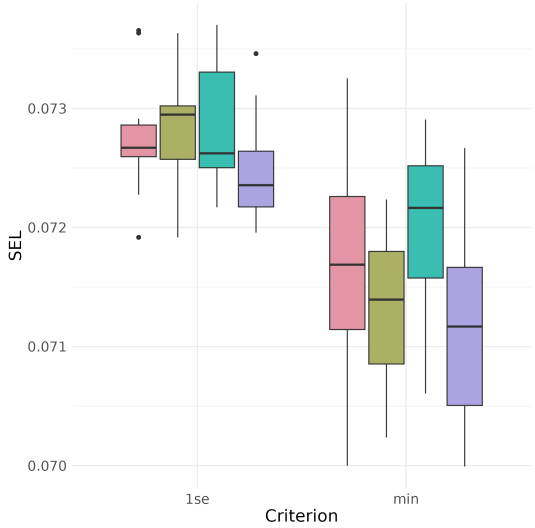
(a) *Dataset 5: Objectivity*



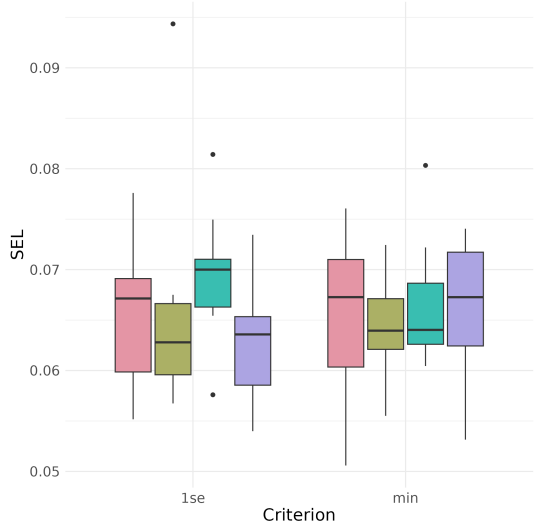
(b) *Dataset 6: Breast Cancer*



(c) *Dataset 7: Sleep Quality*



(d) *Dataset 8: Graduation*

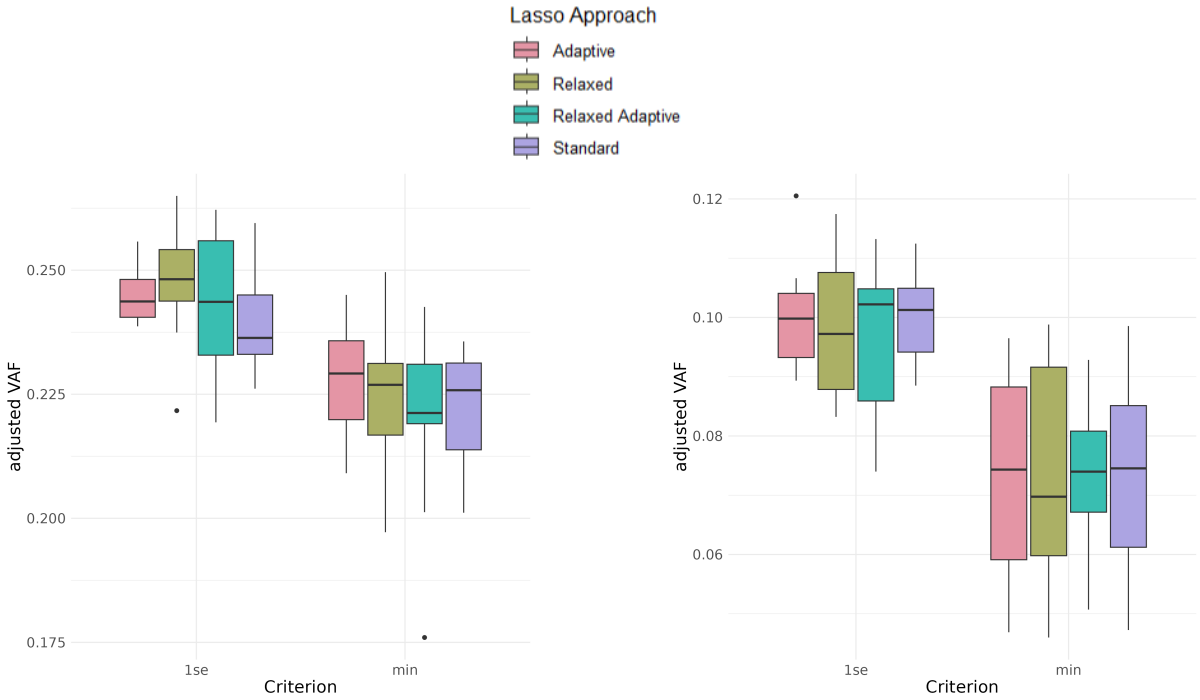


(e) *Dataset 9: ADHD*

Adjusted Variance Accounted For (VAF)

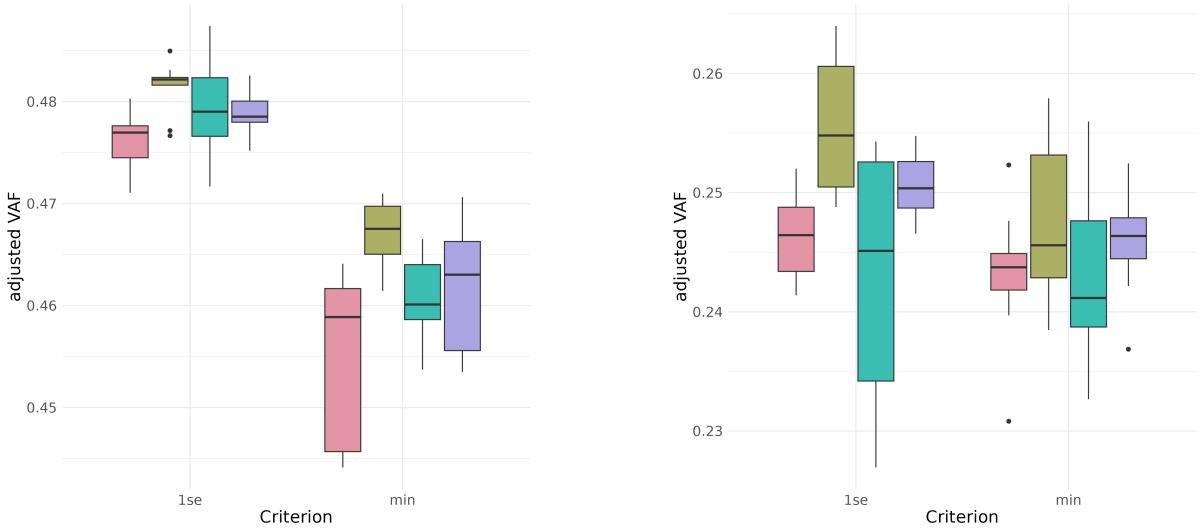
Figure B3

R^2_{adj} *by Dataset*



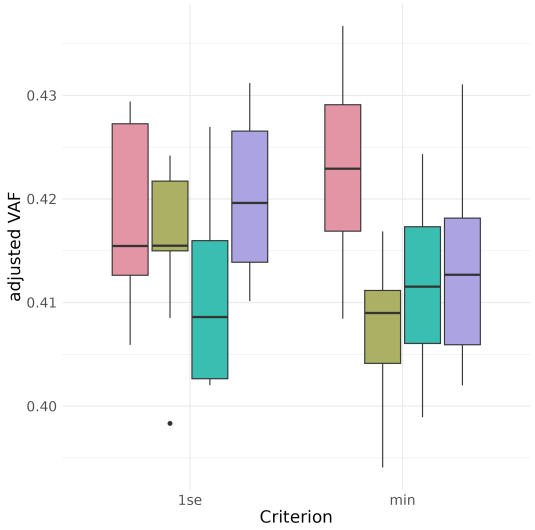
(a) *Dataset 1: High School Grades*

(b) *Dataset 2: Delinquency*

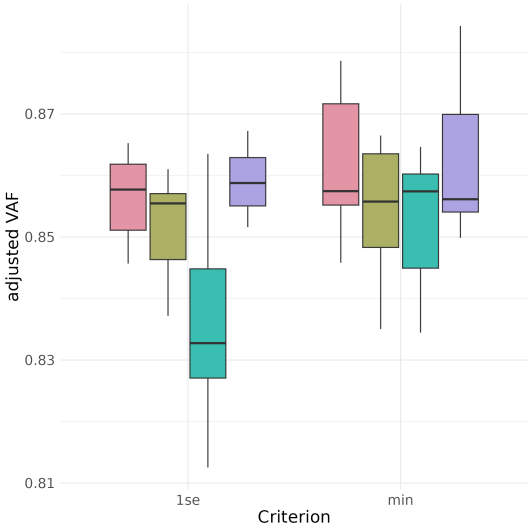


(c) *Dataset 3: Cannabis Consumption*

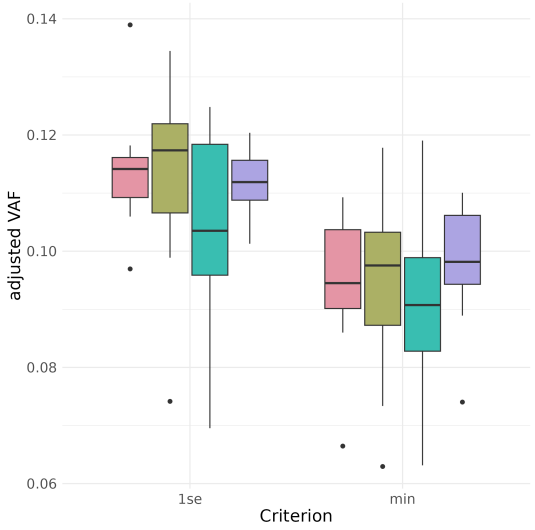
(d) *Dataset 4: Ecstasy Consumption*



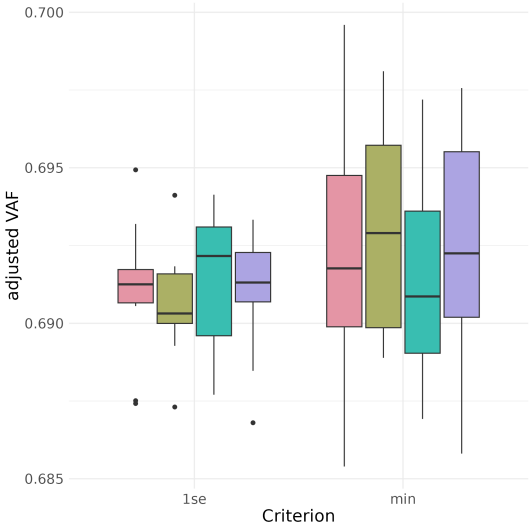
(e) *Dataset 5: Objectivity*



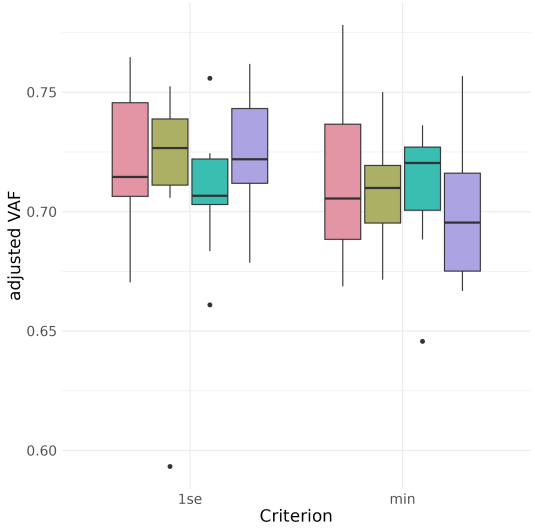
(f) *Dataset 6: Breast Cancer*



(g) *Dataset 7: Sleep Quality*



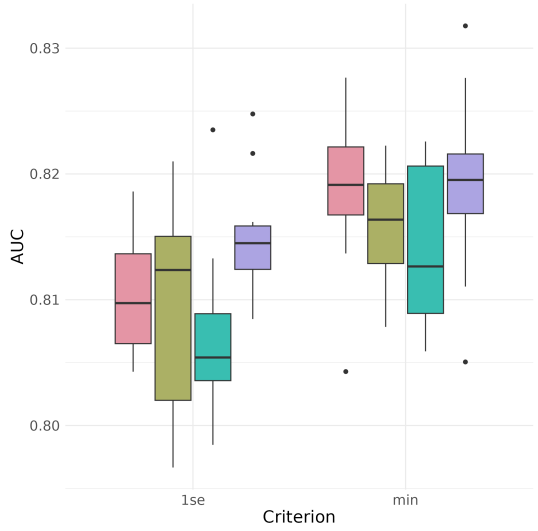
(h) *Dataset 8: Graduation*



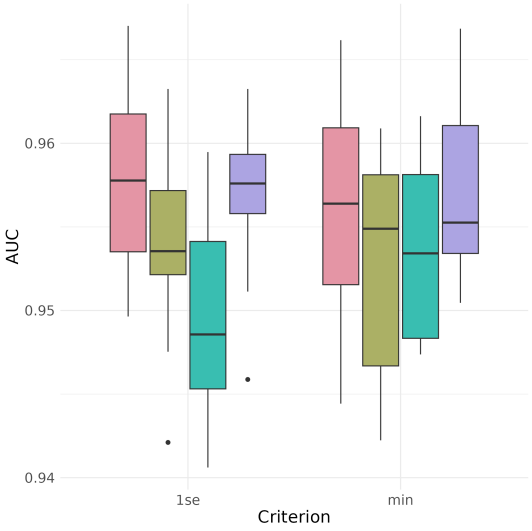
(i) *Dataset 9: ADHD*

Figure B4

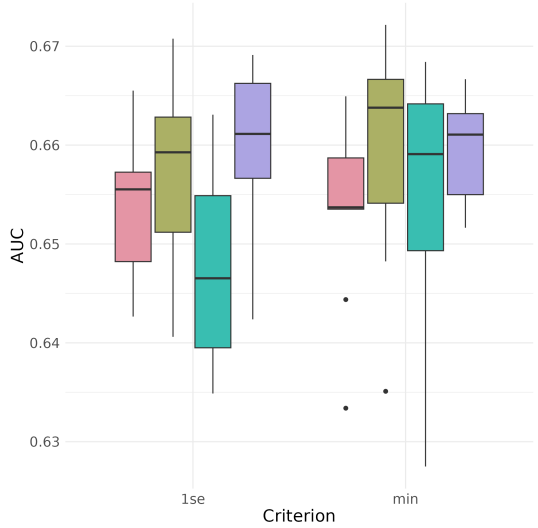
AUC by Dataset for Binary Outcomes



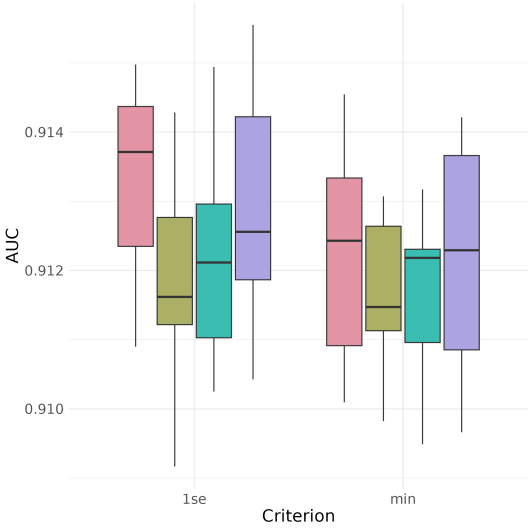
(a) *Dataset 5: Objectivity*



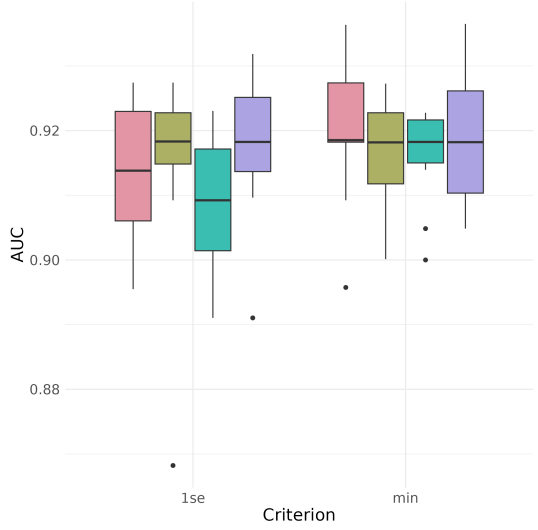
(b) *Dataset 6: Breast Cancer*



(c) *Dataset 7: Sleep Quality*



(d) *Dataset 8: Graduation*



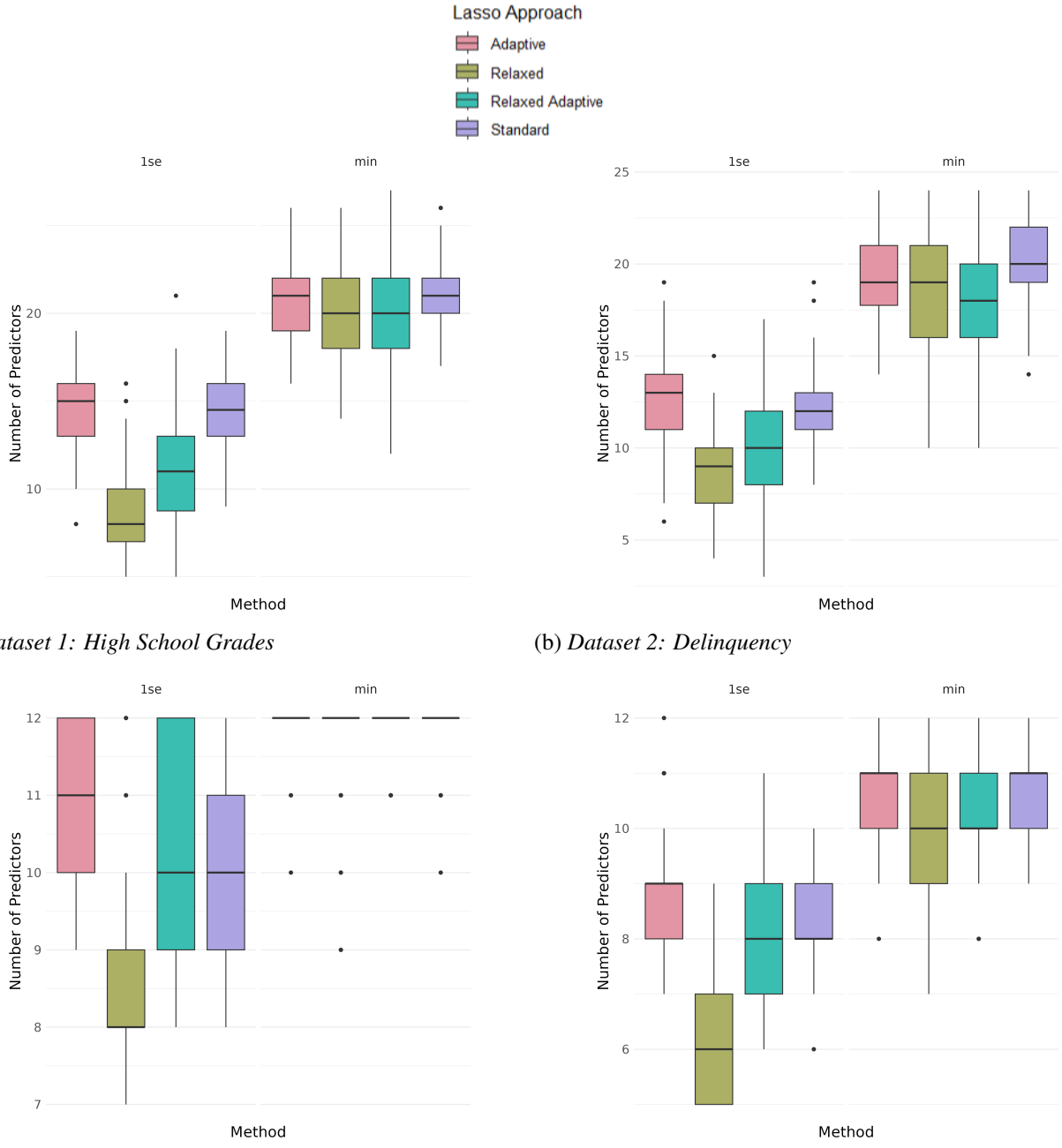
(e) *Dataset 9: ADHD*

Sparsity

Number of Predictors

Figure B5

Distribution of Number of Predictors by Dataset

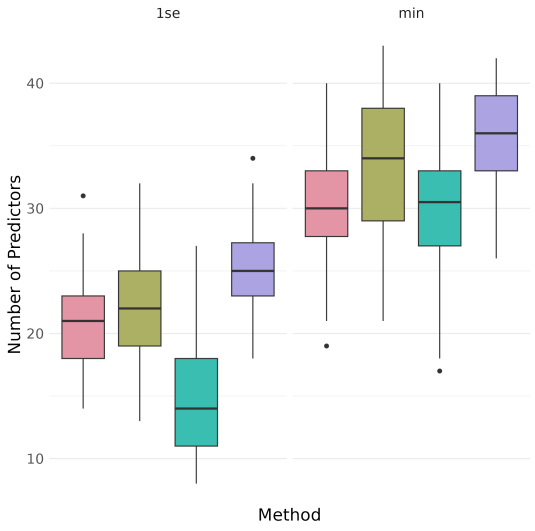


(a) *Dataset 1: High School Grades*

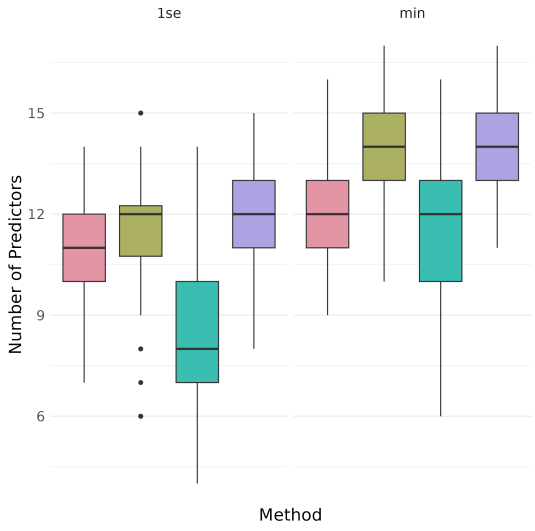
(b) *Dataset 2: Delinquency*

(c) *Dataset 3: Cannabis Consumption*

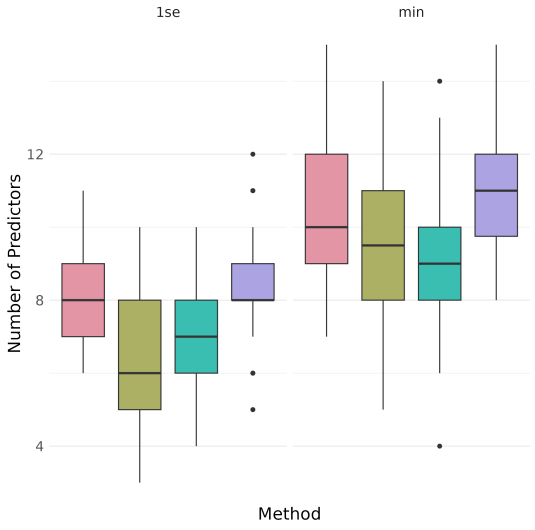
(d) *Dataset 4: Ecstasy Consumption*



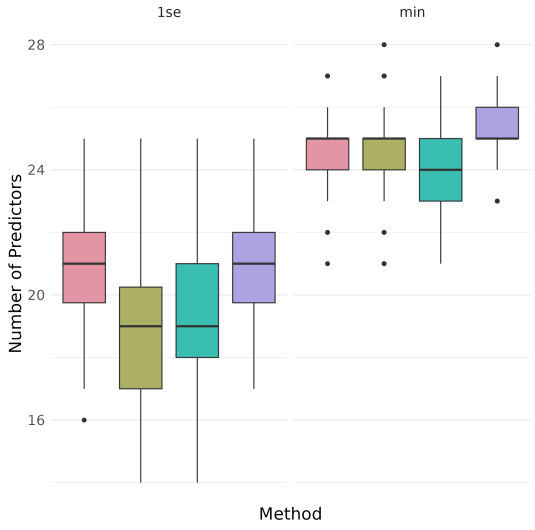
(e) *Dataset 5: Objectivity*



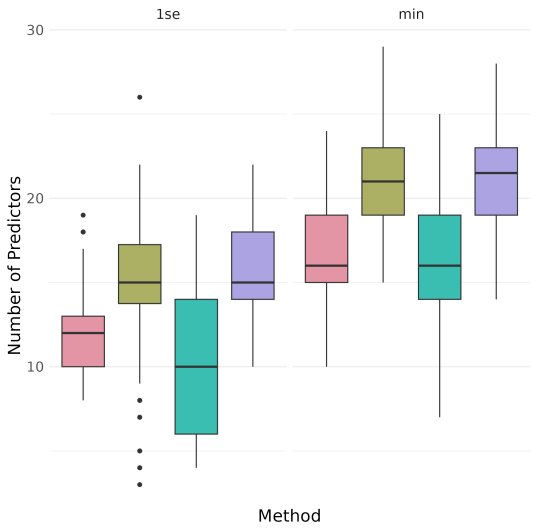
(f) *Dataset 6: Breast Cancer*



(g) *Dataset 7: Sleep Quality*



(h) *Dataset 8: Graduation*

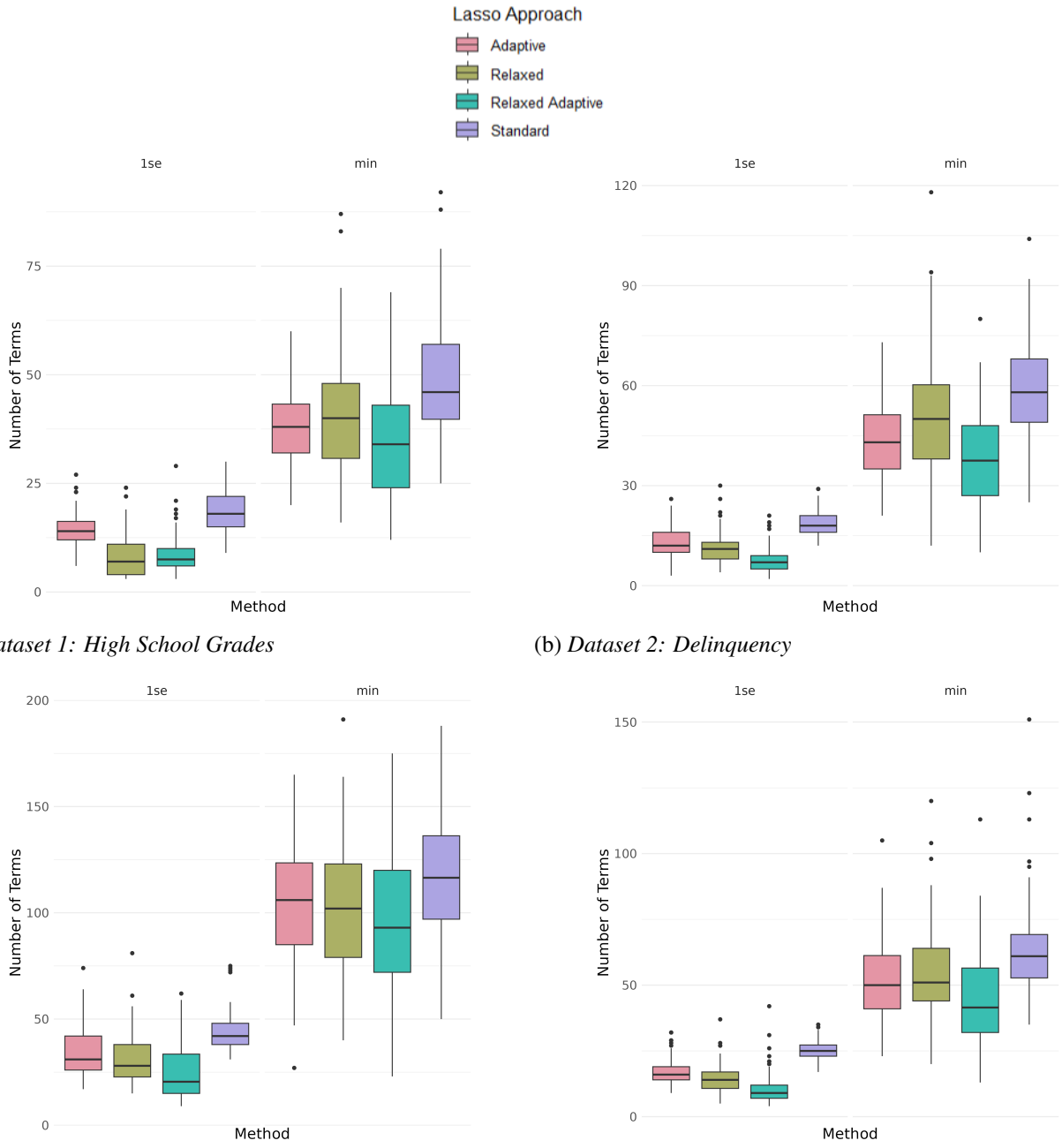


(i) *Dataset 9: ADHD*

Number of Base Learners

Figure B6

Distribution of number of base learners by Dataset

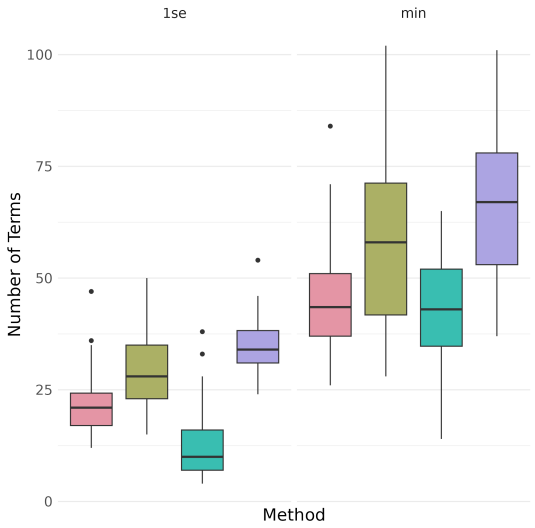


(a) *Dataset 1: High School Grades*

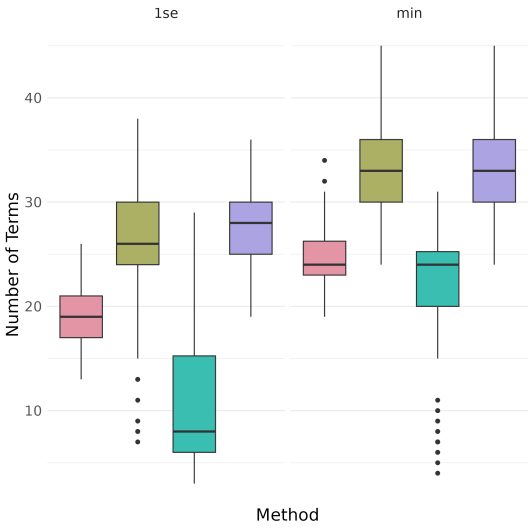
(b) *Dataset 2: Delinquency*

(c) *Dataset 3: Cannabis Consumption*

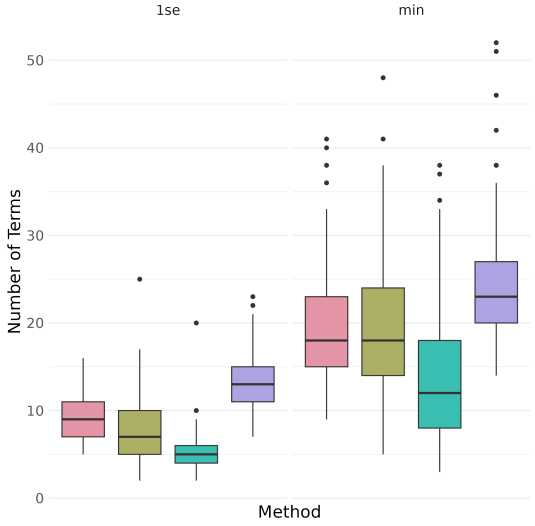
(d) *Dataset 4: Ecstasy Consumption*



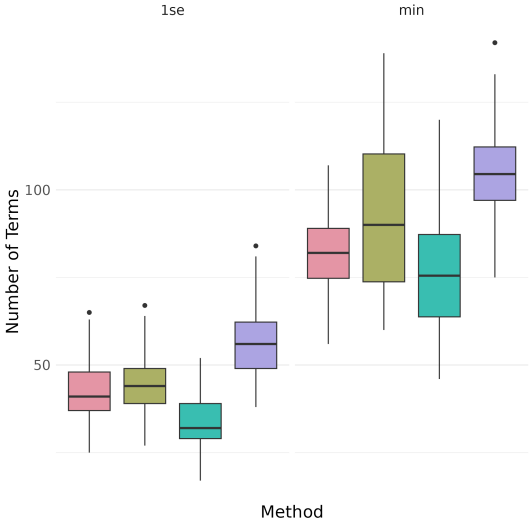
(e) *Dataset 5: Objectivity*



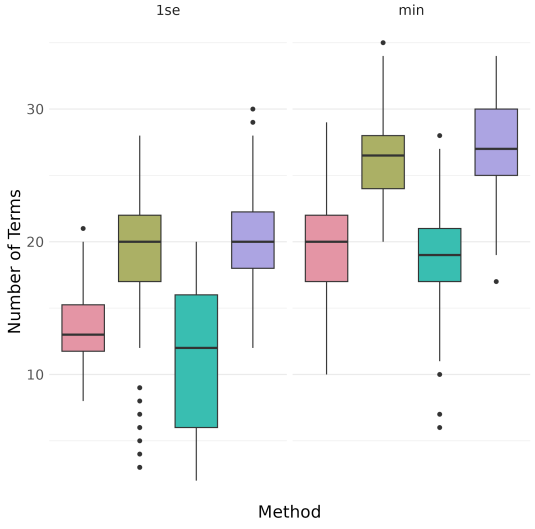
(f) *Dataset 6: Breast Cancer*



(g) *Dataset 7: Sleep Quality*



(h) *Dataset 8: Graduation*



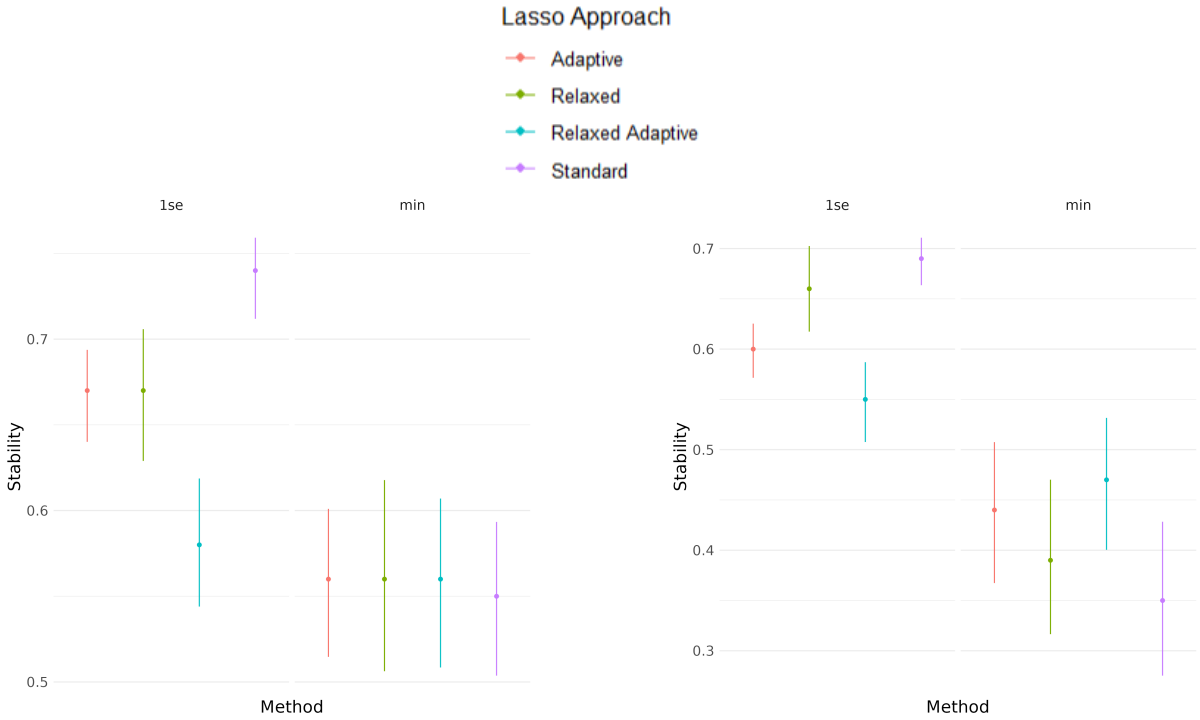
(i) *Dataset 9: ADHD*

Stability

Stability of Variable Selection

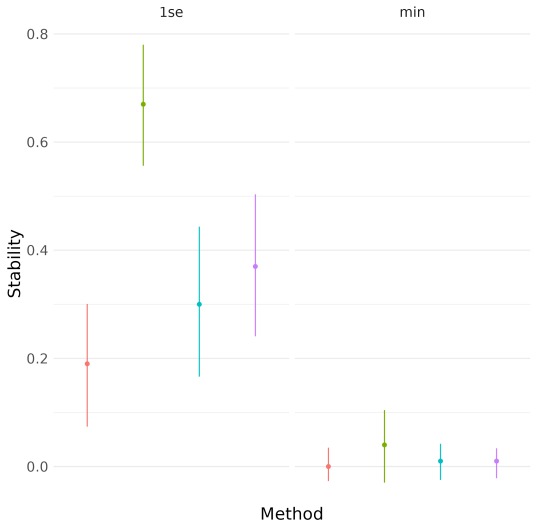
Figure B7

Stability of Variable Selection by Dataset

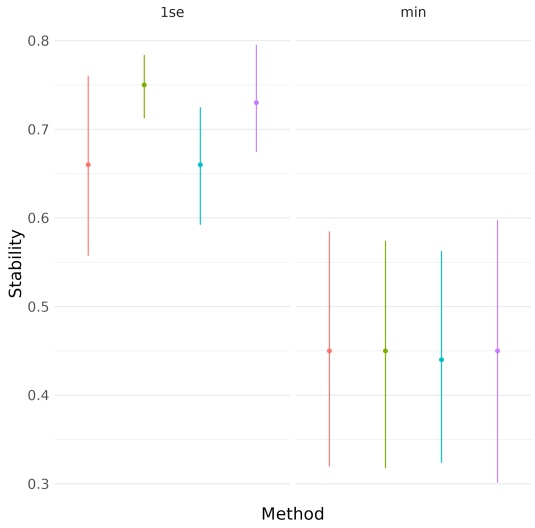


(a) *Dataset 1: High School Grades*

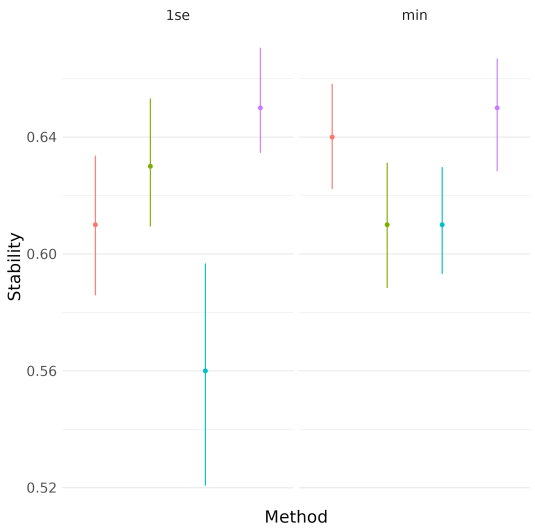
(b) *Dataset 2: Delinquency*



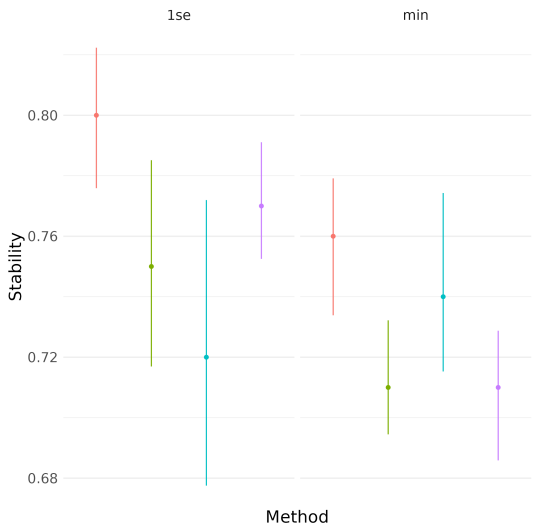
(c) *Dataset 3: Cannabis Consumption*



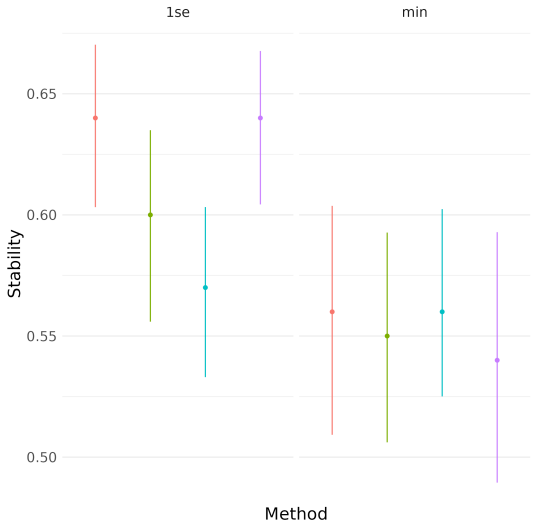
(d) *Dataset 4: Ecstasy Consumption*



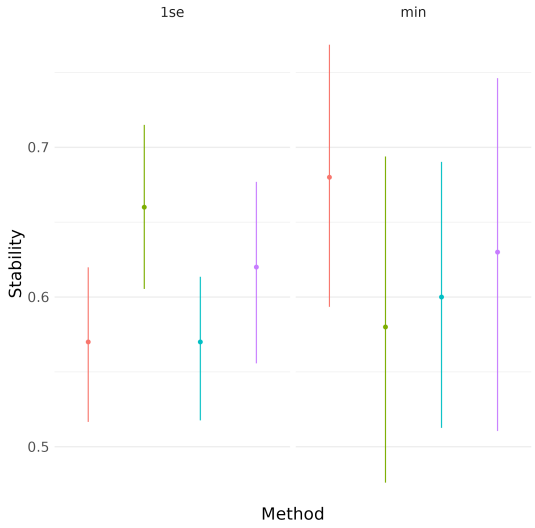
(e) *Dataset 5: Objectivity*



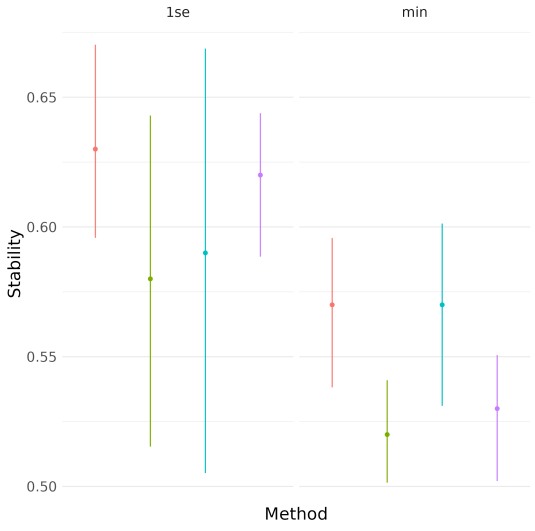
(f) *Dataset 6: Breast Cancer*



(g) *Dataset 7: Sleep Quality*



(h) *Dataset 8: Graduation*

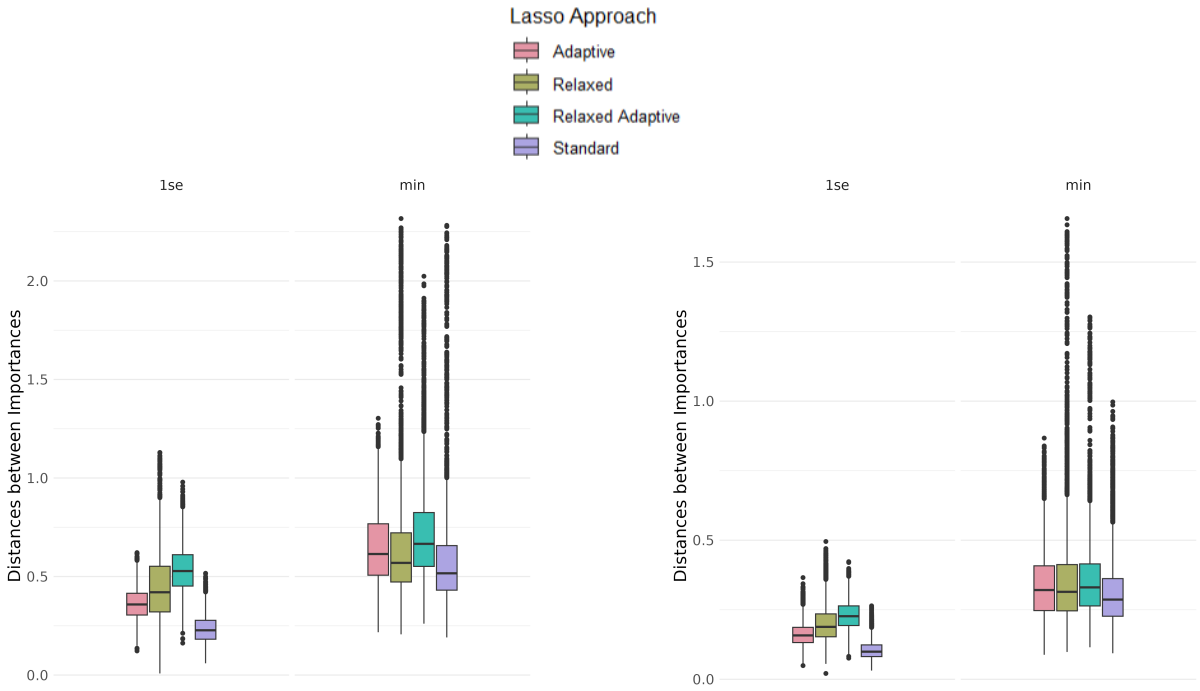


(i) *Dataset 9: ADHD*

Distances between Importance Measures

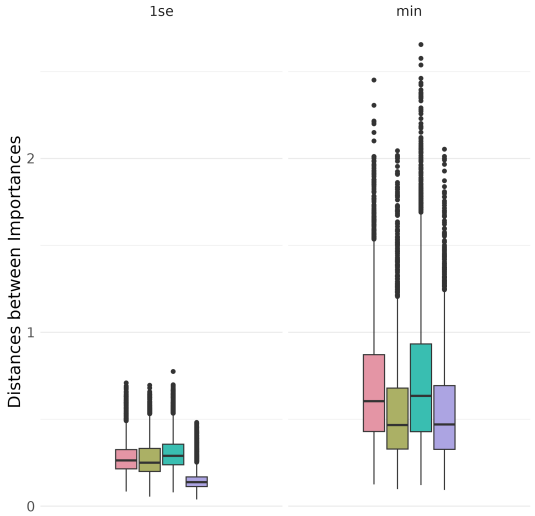
Figure B8

Distances between Importance Measures by Dataset

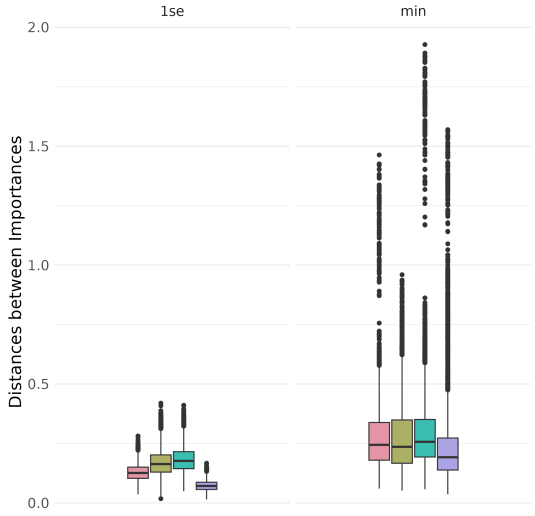


(a) *Dataset 1: High School Grades*

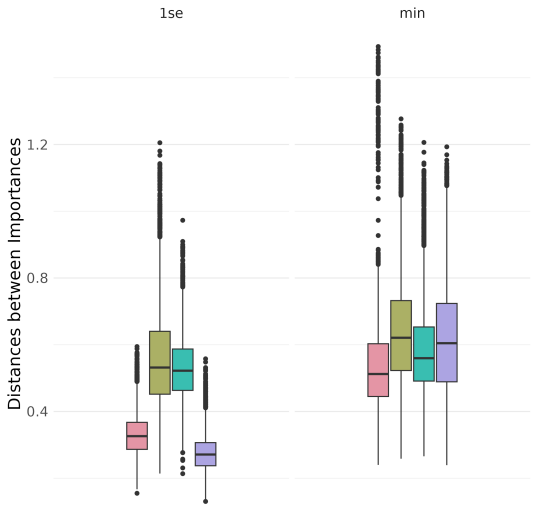
(b) *Dataset 2: Delinquency*



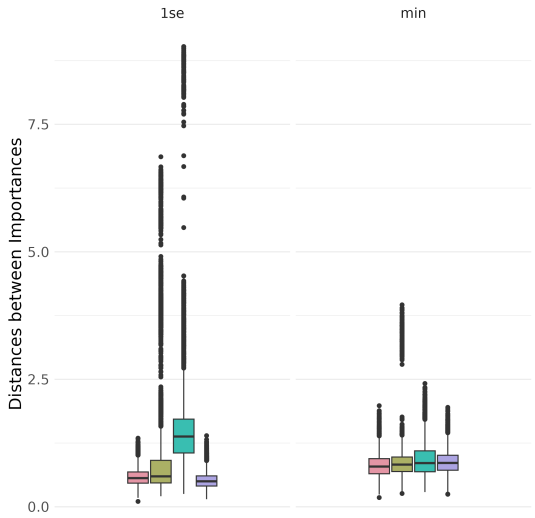
(c) *Dataset 3: Cannabis Consumption*



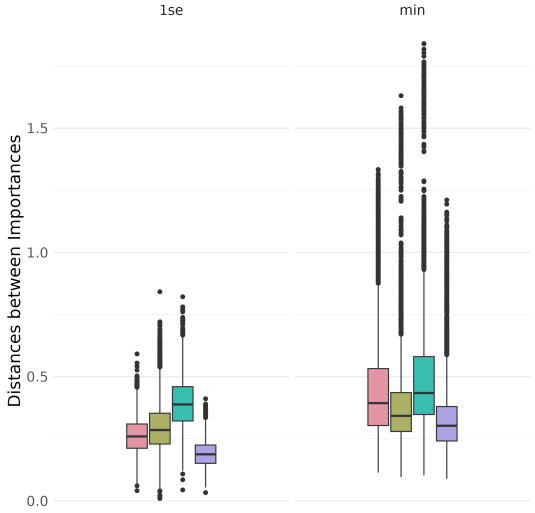
(d) *Dataset 4: Ecstasy Consumption*



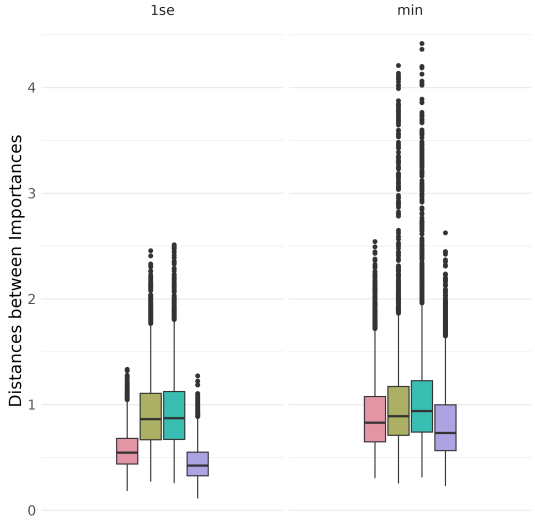
(e) *Dataset 5: Objectivity*



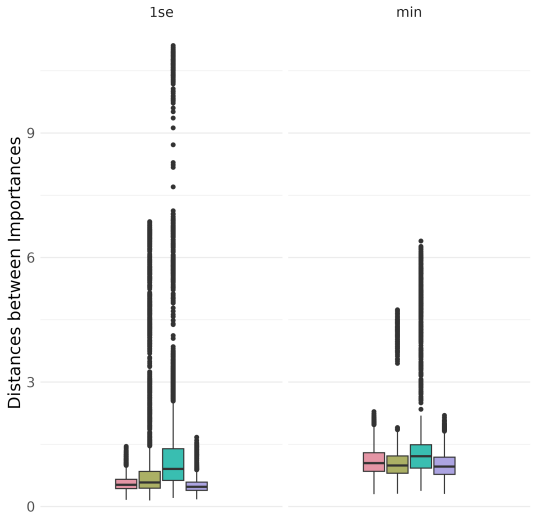
(f) *Dataset 6: Breast Cancer*



(g) *Dataset 7: Sleep Quality*



(h) *Dataset 8: Graduation*

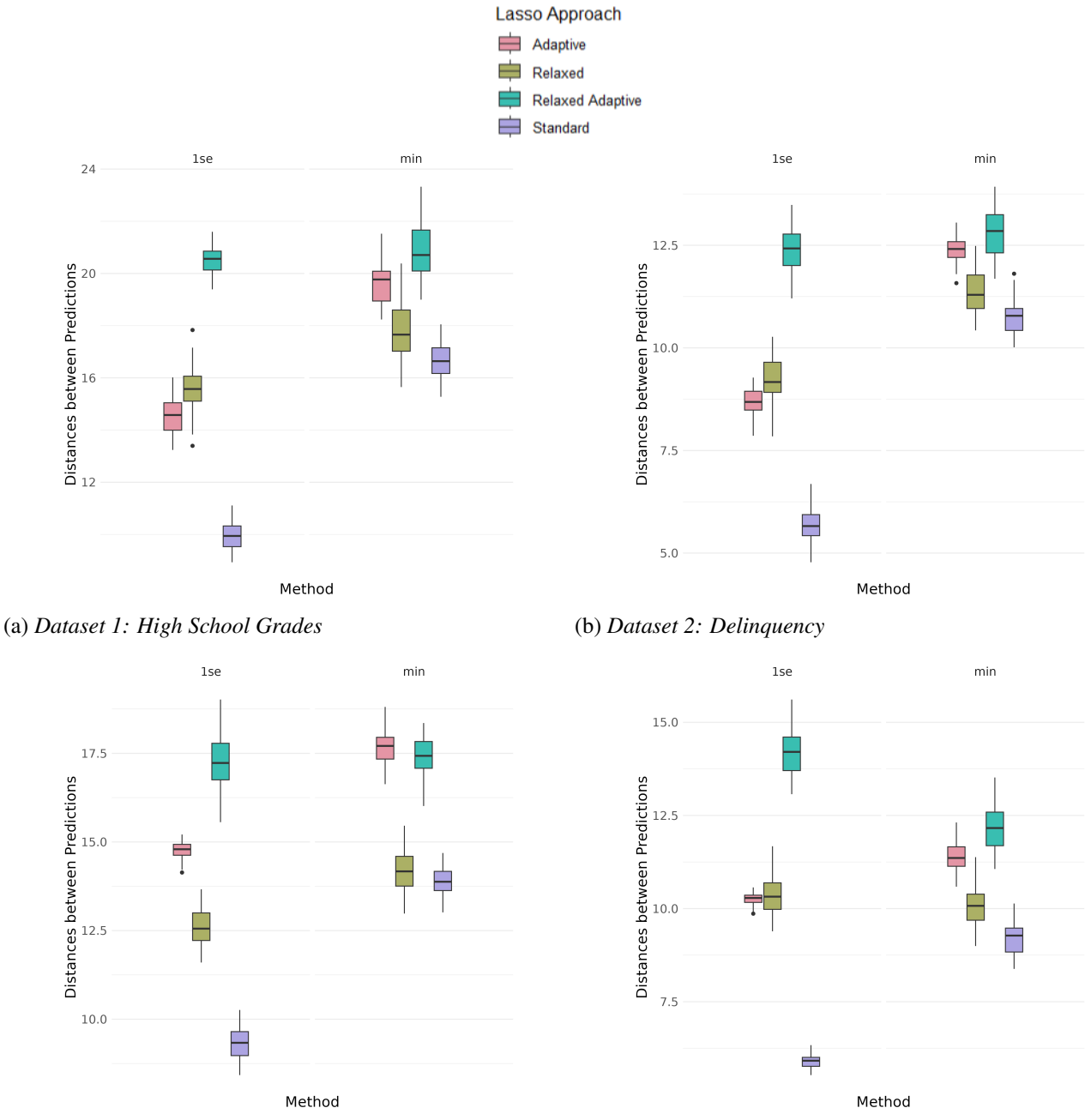


(i) *Dataset 9: ADHD*

Distances between Predictions

Figure B9

Distances between Predictions by Dataset

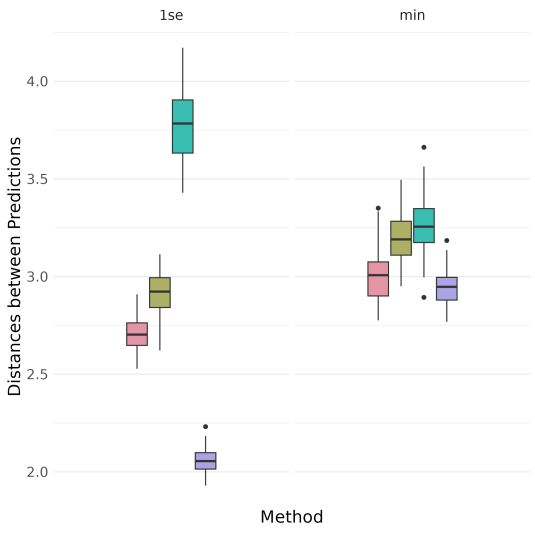


(a) *Dataset 1: High School Grades*

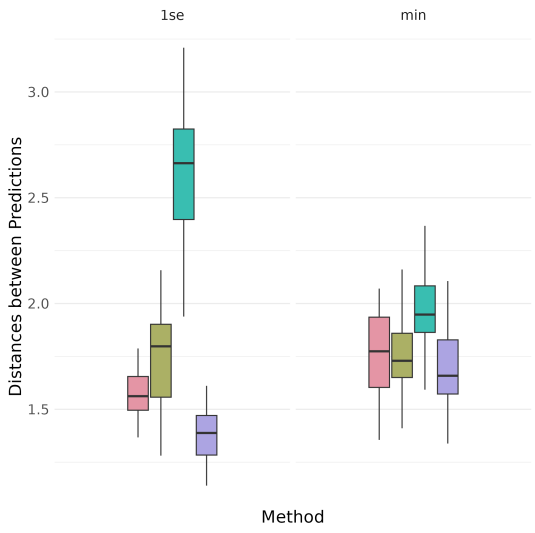
(b) *Dataset 2: Delinquency*

(c) *Dataset 3: Cannabis Consumption*

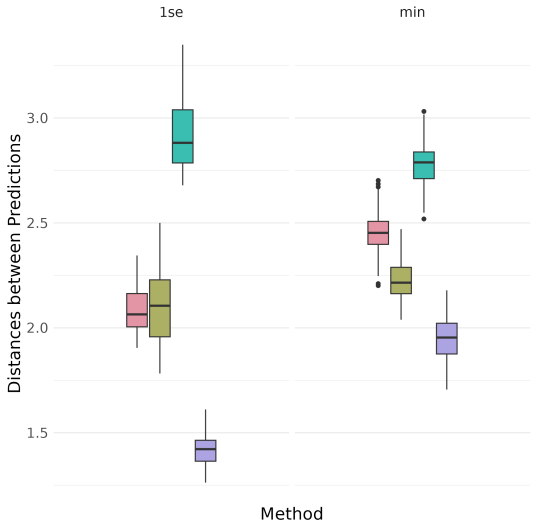
(d) *Dataset 4: Ecstasy Consumption*



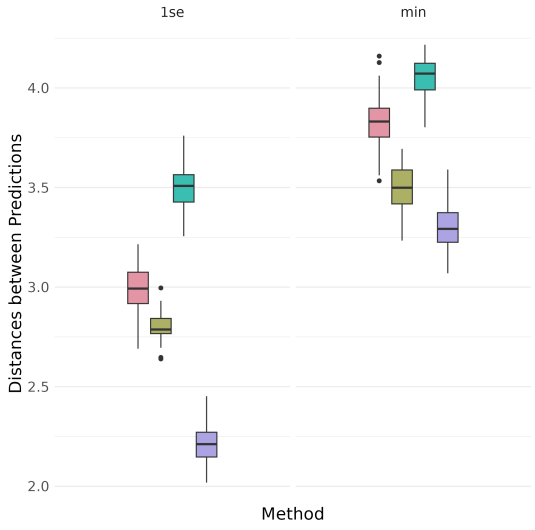
(e) *Dataset 5: Objectivity*



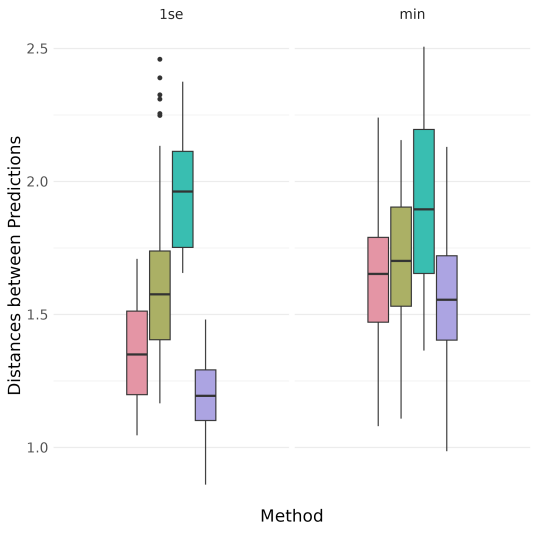
(f) *Dataset 6: Breast Cancer*



(g) *Dataset 7: Sleep Quality*



(h) *Dataset 8: Graduation*



(i) *Dataset 9: ADHD*

Appendix C

Statistical Significance Tests

This section displays the results of the linear mixed models and linear models. As a reference group, the standard lasso with the lambda-min criterion was used in all tests.

Dataset 1: High School Grades

Table C1

Linear Mixed Model of MSE by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	7.503	0.039	50.44	194.241	>0.001***
1SE	0.211	0.047	63.00	4.457	>0.001***
Relaxed	0.077	0.047	63.00	1.635	0.107
Adaptive	0.084	0.047	63.00	1.769	0.082.
1SE:Relaxed	−0.041	0.067	63.00	−0.613	0.542
1SE:Adaptive	−0.089	0.067	63.00	−1.335	0.187
Relaxed:Adaptive	0.050	0.067	63.00	0.745	0.459
1SE:Relaxed:Adaptive	0.001	0.095	63.00	0.012	0.991

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C2

Linear Mixed Model of adjusted VAF by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.223	0.004	47.767	54.646	>0.001***
1SE	0.016	0.005	63.000	3.334	0.001**
Relaxed	0.002	0.005	63.000	0.437	0.664
Adaptive	0.005	0.005	63.000	0.955	0.343
1SE:Relaxed	0.007	0.007	63.000	0.960	0.341
1SE:Adaptive	0.001	0.007	63.000	0.171	0.865
Relaxed:Adaptive	−0.010	0.007	63.000	−1.426	0.159
1SE:Relaxed:Adaptive	0.000	0.010	63.000	−0.039	0.969

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C3

Linear Model of Number of Predictors by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	21.13	0.252	83.962	>0.001***
1SE	−6.88	0.356	−19.331	>0.001***
Relaxed	−1.05	0.356	−2.950	0.003**
Adaptive	−0.55	0.356	−1.545	0.123
1SE:Relaxed	−4.53	0.503	−9.000	>0.001***
1SE:Adaptive	0.99	0.503	1.967	0.050*
Relaxed:Adaptive	0.08	0.503	0.159	0.874
1SE:Relaxed:Adaptive	1.51	0.712	2.121	0.034*

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C4

Linear Model of Distances between Importance Measures by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	0.590	0.003	201.239	>0.001***
1SE	−0.355	0.004	−85.674	>0.001***
Relaxed	0.060	0.004	14.386	>0.001***
Adaptive	0.060	0.004	14.504	>0.001***
1SE:Relaxed	0.150	0.006	25.616	>0.001***
1SE:Adaptive	0.067	0.006	11.368	>0.001***
Relaxed:Adaptive	0.018	0.006	3.141	0.002**
1SE:Relaxed:Adaptive	−0.056	0.008	−6.755	>0.001***

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C5

Linear model of Distances between Predictions by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	16.665	0.122	136.385	>0.001***
1SE	−6.715	0.173	−38.859	>0.001***
Relaxed	1.166	0.173	6.750	>0.001***
Adaptive	2.920	0.173	16.896	>0.001***
1SE:Relaxed	4.474	0.244	18.306	>0.001***
1SE:Adaptive	1.662	0.244	6.802	>0.001***
Relaxed:Adaptive	0.208	0.244	0.850	0.396
1SE:Relaxed:Adaptive	0.121	0.346	0.351	0.726

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Dataset 2: Delinquency

Table C6

Linear Mixed Model of MSE by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	2.033	0.010	25.41	207.214	>0.001***
1SE	0.016	0.010	63.00	1.644	0.105
Relaxed	0.012	0.010	63.00	1.275	0.207
Adaptive	0.027	0.010	63.00	2.801	0.007**
1SE:Relaxed	0.007	0.014	63.00	0.517	0.607
1SE:Adaptive	−0.017	0.014	63.00	−1.250	0.216
Relaxed:Adaptive	−0.001	0.014	63.00	−0.042	0.967
1SE:Relaxed:Adaptive	0.002	0.019	63.00	0.123	0.903

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C7

Linear Mixed Model of adjusted VAF by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.072	0.005	26.153	16.058	>0.001***
1SE	0.028	0.005	63.000	6.175	>0.001***
Relaxed	0.001	0.005	63.000	0.204	0.839
Adaptive	0.001	0.005	63.000	0.183	0.856
1SE:Relaxed	−0.003	0.006	63.000	−0.469	0.640
1SE:Adaptive	−0.001	0.006	63.000	−0.080	0.936
Relaxed:Adaptive	0.000	0.006	63.000	−0.058	0.954
1SE:Relaxed:Adaptive	−0.002	0.009	63.000	−0.253	0.801

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C8

Linear Model of Number of Predictors by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	20.32	0.245	83.012	>0.001***
1SE	−8.24	0.346	−23.803	>0.001***
Relaxed	−1.61	0.346	−4.651	>0.001***
Adaptive	−1.25	0.346	−3.611	>0.001***
1SE:Relaxed	−1.58	0.490	−3.227	0.001**
1SE:Adaptive	1.71	0.490	3.493	0.001**
Relaxed:Adaptive	0.40	0.490	0.817	0.414
1SE:Relaxed:Adaptive	0.30	0.692	0.433	0.665

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C9

Linear Model of Distances between Importance Measures by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	0.317	0.002	185.056	>0.001***
1SE	−0.211	0.002	−87.127	>0.001***
Relaxed	0.053	0.002	22.018	>0.001***
Adaptive	0.024	0.002	9.950	>0.001***
1SE:Relaxed	0.041	0.003	12.001	>0.001***
1SE:Adaptive	0.033	0.003	9.504	>0.001***
Relaxed:Adaptive	−0.030	0.003	−8.747	>0.001***
1SE:Relaxed:Adaptive	0.003	0.005	0.525	0.599

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C10

Linear Model of Distances between Predictions by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	10.742	0.072	148.199	>0.001***
1SE	−5.044	0.103	−49.209	>0.001***
Relaxed	0.675	0.103	6.586	>0.001***
Adaptive	1.641	0.103	16.010	>0.001***
1SE:Relaxed	2.811	0.145	19.389	>0.001***
1SE:Adaptive	1.334	0.145	9.200	>0.001***
Relaxed:Adaptive	−0.254	0.145	−1.756	0.080.
1SE:Relaxed:Adaptive	0.460	0.205	2.242	0.026*

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Dataset 3: Cannabis Consumption

Table C11

Linear Mixed Model of MSE by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	2.643	0.007	38.195	382.769	>0.001***
1SE	0.021	0.008	63.000	2.626	0.011*
Relaxed	−0.003	0.008	63.000	−0.397	0.693
Adaptive	0.052	0.008	63.000	6.591	>0.001***
1SE:Relaxed	0.007	0.011	63.000	0.662	0.511
1SE:Adaptive	−0.026	0.011	63.000	−2.329	0.023*
Relaxed:Adaptive	−0.011	0.011	63.000	−0.997	0.323
1SE:Relaxed:Adaptive	0.006	0.016	63.000	0.407	0.686

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C12

Linear Mixed Model of adjusted VAF by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.462	0.002	47.359	301.793	>0.001***
1SE	0.017	0.002	63.000	9.125	>0.001***
Relaxed	0.005	0.002	63.000	2.775	0.007**
Adaptive	−0.007	0.002	63.000	−3.942	>0.001***
1SE:Relaxed	−0.003	0.003	63.000	−0.963	0.339
1SE:Adaptive	0.005	0.003	63.000	1.819	0.074.
Relaxed:Adaptive	0.001	0.003	63.000	0.289	0.774
1SE:Relaxed:Adaptive	0.000	0.004	63.000	−0.117	0.907

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C13

Linear Model of Number of Predictors by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	11.92	0.091	130.590	>0.001***
1SE	−1.88	0.129	−14.564	>0.001***
Relaxed	−0.17	0.129	−1.317	0.188
Adaptive	0.03	0.129	0.232	0.816
1SE:Relaxed	−1.25	0.183	−6.847	>0.001***
1SE:Adaptive	0.94	0.183	5.149	>0.001***
Relaxed:Adaptive	0.15	0.183	0.822	0.412
1SE:Relaxed:Adaptive	0.59	0.258	2.285	0.023*

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C14

Linear Model of Distances between Importance Measures by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	0.541	0.003	155.439	>0.001***
1SE	−0.390	0.005	−79.171	>0.001***
Relaxed	−0.002	0.005	−0.455	0.649
Adaptive	0.145	0.005	29.500	>0.001***
1SE:Relaxed	0.125	0.007	17.979	>0.001***
1SE:Adaptive	−0.016	0.007	−2.362	0.018*
Relaxed:Adaptive	0.048	0.007	6.931	>0.001***
1SE:Relaxed:Adaptive	−0.146	0.010	−14.787	>0.001***

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C15

Linear Model of Distances between Predictions by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	13.899	0.082	170.509	>0.001***
1SE	−4.555	0.115	−39.516	>0.001***
Relaxed	0.261	0.115	2.264	0.024*
Adaptive	3.754	0.115	32.564	>0.001***
1SE:Relaxed	2.993	0.163	18.357	>0.001***
1SE:Adaptive	1.669	0.163	10.240	>0.001***
Relaxed:Adaptive	−0.530	0.163	−3.249	0.001**
1SE:Relaxed:Adaptive	−0.280	0.231	−1.216	0.225

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Dataset 4: Ecstasy Consumption

Table C16

Linear Mixed Model of MSE by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	1.979	0.005	41.323	392.486	>0.001***
1SE	0.028	0.006	63.000	4.788	>0.001***
Relaxed	0.005	0.006	63.000	0.859	0.394
Adaptive	0.020	0.006	63.000	3.331	0.001**
1SE:Relaxed	−0.006	0.008	63.000	−0.743	0.460
1SE:Adaptive	0.001	0.008	63.000	0.107	0.915
Relaxed:Adaptive	0.004	0.008	63.000	0.494	0.623
1SE:Relaxed:Adaptive	0.013	0.012	63.000	1.118	0.268

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C17

Linear Mixed Model of adjusted VAF by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.246	0.002	39.738	124.676	>0.001***
1SE	0.005	0.002	63.000	2.110	0.039*
Relaxed	0.002	0.002	63.000	0.831	0.409
Adaptive	−0.003	0.002	63.000	−1.164	0.249
1SE:Relaxed	0.003	0.003	63.000	0.913	0.365
1SE:Adaptive	−0.002	0.003	63.000	−0.515	0.608
Relaxed:Adaptive	−0.003	0.003	63.000	−0.787	0.434
1SE:Relaxed:Adaptive	−0.006	0.005	63.000	−1.226	0.225

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C18

Linear Model of Number of Predictors by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	10.75	0.101	105.990	>0.001***
1SE	−2.51	0.143	−17.499	>0.001***
Relaxed	−0.59	0.143	−4.113	>0.001***
Adaptive	−0.17	0.143	−1.185	0.236
1SE:Relaxed	−1.50	0.203	−7.395	>0.001***
1SE:Adaptive	0.82	0.203	4.042	>0.001***
Relaxed:Adaptive	0.25	0.203	1.232	0.218
1SE:Relaxed:Adaptive	0.73	0.287	2.545	0.011*

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C19

Linear Model of Distances between Importance Measures by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	0.254	0.002	125.298	>0.001***
1SE	−0.181	0.003	−63.262	>0.001***
Relaxed	0.024	0.003	8.542	>0.001***
Adaptive	0.030	0.003	10.522	>0.001***
1SE:Relaxed	0.072	0.004	17.724	>0.001***
1SE:Adaptive	0.026	0.004	6.405	>0.001***
Relaxed:Adaptive	0.004	0.004	0.994	0.320
1SE:Relaxed:Adaptive	−0.046	0.006	−8.090	>0.001***

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C20

Linear Model of Distances between Predictions by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	9.243	0.070	132.643	>0.001***
1SE	−3.349	0.099	−33.989	>0.001***
Relaxed	0.822	0.099	8.341	>0.001***
Adaptive	2.141	0.099	21.722	>0.001***
1SE:Relaxed	3.663	0.139	26.285	>0.001***
1SE:Adaptive	2.226	0.139	15.975	>0.001***
Relaxed:Adaptive	−0.023	0.139	−0.163	0.871
1SE:Relaxed:Adaptive	−0.514	0.197	−2.606	0.010**

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Dataset 5: Objectivity

Table C21

Linear Mixed Model of SEL by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.127	0.001	44.494	216.150	>0.001***
1SE	0.003	0.001	63.000	3.833	>0.001***
Relaxed	0.002	0.001	63.000	3.285	0.002**
Adaptive	0.001	0.001	63.000	0.841	0.404
1SE:Relaxed	−0.001	0.001	63.000	−0.594	0.555
1SE:Adaptive	0.002	0.001	63.000	1.639	0.106
Relaxed:Adaptive	0.000	0.001	63.000	0.486	0.628
1SE:Relaxed:Adaptive	0.001	0.001	63.000	0.646	0.521

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C22

Linear Mixed Model of AUC by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.819	0.002	56.816	403.049	>0.001***
1SE	−0.004	0.003	63.000	−1.646	0.105
Relaxed	−0.003	0.003	63.000	−1.279	0.206
Adaptive	−0.001	0.003	63.000	−0.233	0.816
1SE:Relaxed	−0.002	0.004	63.000	−0.516	0.608
1SE:Adaptive	−0.004	0.004	63.000	−1.125	0.265
Relaxed:Adaptive	−0.001	0.004	63.000	−0.323	0.748
1SE:Relaxed:Adaptive	0.003	0.005	63.000	0.655	0.515

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C23

Linear Mixed Model of adjusted VAF by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.413	0.003	39.696	156.398	>0.001***
1SE	0.007	0.003	63.000	2.251	0.028*
Relaxed	−0.006	0.003	63.000	−1.883	0.064.
Adaptive	0.010	0.003	63.000	3.348	0.001**
1SE:Relaxed	0.001	0.004	63.000	0.335	0.738
1SE:Adaptive	−0.012	0.004	63.000	−2.780	0.007**
Relaxed:Adaptive	−0.006	0.004	63.000	−1.318	0.192
1SE:Relaxed:Adaptive	0.002	0.006	63.000	0.338	0.736

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C24

Linear Model of Number of Predictors by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	35.60	0.402	88.640	>0.001***
1SE	−10.40	0.568	−18.310	>0.001***
Relaxed	−2.07	0.568	−3.644	>0.001***
Adaptive	−5.46	0.568	−9.613	>0.001***
1SE:Relaxed	−0.98	0.803	−1.220	0.223
1SE:Adaptive	1.06	0.803	1.320	0.187
Relaxed:Adaptive	1.68	0.803	2.092	0.037*
1SE:Relaxed:Adaptive	−4.64	1.136	−4.085	>0.001***

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C25

Linear Model of Distances between Importance Measures by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	0.617	0.002	333.186	>0.001***
1SE	−0.340	0.003	−129.872	>0.001***
Relaxed	0.023	0.003	8.606	>0.001***
Adaptive	−0.076	0.003	−28.938	>0.001***
1SE:Relaxed	0.260	0.004	70.188	>0.001***
1SE:Adaptive	0.129	0.004	34.932	>0.001***
Relaxed:Adaptive	0.021	0.004	5.806	>0.001***
1SE:Relaxed:Adaptive	−0.104	0.005	−19.962	>0.001***

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C26

Linear Model of Distances between Predictions by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	2.945	0.019	158.117	>0.001***
1SE	−0.886	0.026	−33.648	>0.001***
Relaxed	0.241	0.026	9.158	>0.001***
Adaptive	0.064	0.026	2.413	0.016*
1SE:Relaxed	0.614	0.037	16.471	>0.001***
1SE:Adaptive	0.581	0.037	15.590	>0.001***
Relaxed:Adaptive	0.017	0.037	0.463	0.644
1SE:Relaxed:Adaptive	0.198	0.053	3.767	>0.001***

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Dataset 6: Breast Cancer

Table C27

Linear Mixed Model of SEL by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.030	0.001	45.342	41.411	>0.001***
1SE	0.001	0.001	63.000	1.045	0.300
Relaxed	0.002	0.001	63.000	1.880	0.065.
Adaptive	0.000	0.001	63.000	0.465	0.644
1SE:Relaxed	0.000	0.001	63.000	0.093	0.926
1SE:Adaptive	0.001	0.001	63.000	0.544	0.588
Relaxed:Adaptive	0.001	0.001	63.000	0.527	0.600
1SE:Relaxed:Adaptive	0.003	0.002	63.000	1.601	0.114

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C28

Linear Mixed Model of AUC by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.957	0.002	53.061	497.831	>0.001***
1SE	0.000	0.002	63.000	−0.207	0.837
Relaxed	−0.004	0.002	63.000	−1.831	0.072.
Adaptive	−0.001	0.002	63.000	−0.337	0.737
1SE:Relaxed	0.002	0.003	63.000	0.533	0.596
1SE:Adaptive	0.002	0.003	63.000	0.566	0.574
Relaxed:Adaptive	0.002	0.003	63.000	0.451	0.653
1SE:Relaxed:Adaptive	−0.007	0.005	63.000	−1.549	0.126

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C29

Linear Mixed Model of adjusted VAF by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.862	0.003	43.834	263.191	>0.001***
1SE	−0.003	0.004	63.000	−0.718	0.475
Relaxed	−0.007	0.004	63.000	−1.930	0.058.
Adaptive	0.000	0.004	63.000	0.083	0.934
1SE:Relaxed	0.000	0.005	63.000	0.024	0.981
1SE:Adaptive	−0.003	0.005	63.000	−0.519	0.606
Relaxed:Adaptive	−0.002	0.005	63.000	−0.357	0.722
1SE:Relaxed:Adaptive	−0.011	0.008	63.000	−1.461	0.149

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C30

Linear Model of Number of Predictors by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	13.71	0.163	84.065	>0.001***
1SE	−1.75	0.231	−7.588	>0.001***
Relaxed	−0.08	0.231	−0.347	0.729
Adaptive	−1.67	0.231	−7.241	>0.001***
1SE:Relaxed	−0.53	0.326	−1.625	0.105
1SE:Adaptive	0.51	0.326	1.564	0.118
Relaxed:Adaptive	−0.50	0.326	−1.533	0.126
1SE:Relaxed:Adaptive	−1.00	0.461	−2.168	0.030*

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C31

Linear Model of Distances between Importance Measures by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	0.874	0.010	89.188	>0.001***
1SE	−0.351	0.014	−25.345	>0.001***
Relaxed	0.010	0.014	0.711	0.477
Adaptive	−0.061	0.014	−4.417	>0.001***
1SE:Relaxed	0.653	0.020	33.360	>0.001***
1SE:Adaptive	0.123	0.020	6.280	>0.001***
Relaxed:Adaptive	0.103	0.020	5.273	>0.001***
1SE:Relaxed:Adaptive	0.260	0.028	9.396	>0.001***

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C32

Linear Model of Distances between Predictions by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	1.700	0.030	57.613	>0.001***
1SE	−0.323	0.042	−7.737	>0.001***
Relaxed	0.055	0.042	1.311	0.191
Adaptive	0.060	0.042	1.431	0.153
1SE:Relaxed	0.308	0.059	5.216	>0.001***
1SE:Adaptive	0.135	0.059	2.289	0.023*
Relaxed:Adaptive	0.157	0.059	2.665	0.008**
1SE:Relaxed:Adaptive	0.524	0.083	6.273	>0.001***

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Dataset 7: Sleep Quality

Table C33

Linear Mixed Model of SEL by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.215	0.001	23.788	199.015	>0.001***
1SE	0.001	0.001	63.000	1.149	0.255
Relaxed	0.003	0.001	63.000	2.853	0.006**
Adaptive	0.003	0.001	63.000	2.756	0.008**
1SE:Relaxed	−0.001	0.001	63.000	−0.686	0.495
1SE:Adaptive	−0.002	0.001	63.000	−1.139	0.259
Relaxed:Adaptive	0.000	0.001	63.000	0.166	0.869
1SE:Relaxed:Adaptive	0.002	0.002	63.000	0.891	0.376

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C34

Linear Mixed Model of AUC by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.660	0.003	61.258	217.840	>0.001***
1SE	0.000	0.004	63.000	0.080	0.937
Relaxed	0.000	0.004	63.000	0.045	0.965
Adaptive	−0.006	0.004	63.000	−1.457	0.150
1SE:Relaxed	−0.003	0.006	63.000	−0.621	0.537
1SE:Adaptive	0.000	0.006	63.000	−0.061	0.951
Relaxed:Adaptive	0.001	0.006	63.000	0.107	0.915
1SE:Relaxed:Adaptive	−0.004	0.008	63.000	−0.474	0.637

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C35

Linear Mixed Model of adjusted VAF by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.098	0.004	21.908	21.986	>0.001***
1SE	0.014	0.004	63.000	3.358	0.001**
Relaxed	−0.003	0.004	63.000	−0.848	0.400
Adaptive	−0.004	0.004	63.000	−0.872	0.387
1SE:Relaxed	0.005	0.006	63.000	0.881	0.381
1SE:Adaptive	0.006	0.006	63.000	1.025	0.309
Relaxed:Adaptive	0.000	0.006	63.000	0.014	0.989
1SE:Relaxed:Adaptive	−0.012	0.008	63.000	−1.421	0.160

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C36

Linear Model of Number of Predictors by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	10.78	0.169	63.879	>0.001***
1SE	−2.39	0.239	−10.014	>0.001***
Relaxed	−1.09	0.239	−4.567	>0.001***
Adaptive	−0.33	0.239	−1.383	0.167
1SE:Relaxed	−0.98	0.338	−2.904	0.004**
1SE:Adaptive	0.25	0.338	0.741	0.459
Relaxed:Adaptive	−0.23	0.338	−0.681	0.496
1SE:Relaxed:Adaptive	0.79	0.477	1.655	0.098.

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C37

Linear Model of Distances between Importance Measures by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	0.341	0.002	147.806	>0.001***
1SE	−0.151	0.003	−46.254	>0.001***
Relaxed	0.051	0.003	15.610	>0.001***
Adaptive	0.120	0.003	36.873	>0.001***
1SE:Relaxed	0.058	0.005	12.514	>0.001***
1SE:Adaptive	−0.047	0.005	−10.286	>0.001***
Relaxed:Adaptive	−0.009	0.005	−1.864	0.062.
1SE:Relaxed:Adaptive	0.031	0.007	4.682	>0.001***

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C38

Linear Model of Distances between Predictions by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	1.954	0.018	107.498	>0.001***
1SE	−0.529	0.026	−20.578	>0.001***
Relaxed	0.265	0.026	10.316	>0.001***
Adaptive	0.502	0.026	19.543	>0.001***
1SE:Relaxed	0.404	0.036	11.099	>0.001***
1SE:Adaptive	0.159	0.036	4.385	>0.001***
Relaxed:Adaptive	0.060	0.036	1.637	0.103
1SE:Relaxed:Adaptive	0.119	0.051	2.317	0.021*

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Dataset 8: Graduation

Table C39

Linear Mixed Model of SEL by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.071	0.000	56.593	331.003	>0.001***
1SE	0.001	0.000	63.000	4.847	>0.001***
Relaxed	0.000	0.000	63.000	0.552	0.583
Adaptive	0.000	0.000	63.000	1.822	0.073.
1SE:Relaxed	0.000	0.000	63.000	0.430	0.669
1SE:Adaptive	0.000	0.000	63.000	−0.592	0.556
Relaxed:Adaptive	0.000	0.000	63.000	0.233	0.816
1SE:Relaxed:Adaptive	0.000	0.001	63.000	−0.570	0.571

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C40

Linear Mixed Model of AUC by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	0.912	0.000	1929.296	>0.001***
1SE	0.001	0.001	1.031	0.306
Relaxed	0.000	0.001	−0.681	0.498
Adaptive	0.000	0.001	0.236	0.814
1SE:Relaxed	−0.001	0.001	−0.732	0.466
1SE:Adaptive	0.000	0.001	0.407	0.686
Relaxed:Adaptive	0.000	0.001	−0.267	0.790
1SE:Relaxed:Adaptive	0.000	0.001	0.187	0.852

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C41

Linear Mixed Model of adjusted VAF by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.692	0.001	57.088	738.501	>0.001***
1SE	−0.001	0.001	63.000	−1.157	0.252
Relaxed	0.000	0.001	63.000	0.417	0.678
Adaptive	0.000	0.001	63.000	−0.134	0.894
1SE:Relaxed	−0.001	0.002	63.000	−0.517	0.607
1SE:Adaptive	0.000	0.002	63.000	0.109	0.914
Relaxed:Adaptive	−0.001	0.002	63.000	−0.607	0.546
1SE:Relaxed:Adaptive	0.002	0.002	63.000	0.743	0.460

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C42

Linear Model of Number of Predictors by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	25.30	0.170	148.918	>0.001***
1SE	−4.32	0.240	−17.980	>0.001***
Relaxed	−0.71	0.240	−2.955	0.003**
Adaptive	−0.64	0.240	−2.664	0.008**
1SE:Relaxed	−1.35	0.340	−3.973	>0.001***
1SE:Adaptive	0.54	0.340	1.589	0.112
Relaxed:Adaptive	0.32	0.340	0.942	0.347
1SE:Relaxed:Adaptive	0.21	0.481	0.437	0.662

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C43

Linear Model of Distances between Importance Measures by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	0.824	0.005	159.642	>0.001***
1SE	−0.373	0.007	−51.097	>0.001***
Relaxed	0.184	0.007	25.197	>0.001***
Adaptive	0.079	0.007	10.882	>0.001***
1SE:Relaxed	0.282	0.010	27.287	>0.001***
1SE:Adaptive	0.043	0.010	4.122	>0.001***
Relaxed:Adaptive	−0.021	0.010	−2.034	0.042*
1SE:Relaxed:Adaptive	−0.090	0.015	−6.172	>0.001***

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C44

Linear Model of Distances between Predictions by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	3.306	0.016	203.620	>0.001***
1SE	−1.094	0.023	−47.643	>0.001***
Relaxed	0.187	0.023	8.161	>0.001***
Adaptive	0.527	0.023	22.965	>0.001***
1SE:Relaxed	0.399	0.032	12.295	>0.001***
1SE:Adaptive	0.250	0.032	7.696	>0.001***
Relaxed:Adaptive	0.029	0.032	0.907	0.365
1SE:Relaxed:Adaptive	−0.102	0.046	−2.213	0.028*

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Dataset 9: ADHD

Table C45

Linear Mixed Model of SEL by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.066	0.002	17.561	28.296	>0.001***
1SE	−0.003	0.002	63.000	−1.633	0.107
Relaxed	−0.001	0.002	63.000	−0.693	0.491
Adaptive	−0.001	0.002	63.000	−0.494	0.623
1SE:Relaxed	0.004	0.003	63.000	1.504	0.138
1SE:Adaptive	0.004	0.003	63.000	1.372	0.175
Relaxed:Adaptive	0.003	0.003	63.000	1.099	0.276
1SE:Relaxed:Adaptive	−0.002	0.004	63.000	−0.428	0.670

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C46

Linear Mixed Model of AUC by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.919	0.004	27.76	254.639	>0.001***
1SE	−0.001	0.004	63.00	−0.360	0.720
Relaxed	−0.002	0.004	63.00	−0.491	0.625
Adaptive	0.001	0.004	63.00	0.385	0.701
1SE:Relaxed	−0.001	0.005	63.00	−0.183	0.855
1SE:Adaptive	−0.005	0.005	63.00	−0.894	0.375
Relaxed:Adaptive	−0.002	0.005	63.00	−0.445	0.658
1SE:Relaxed:Adaptive	0.000	0.007	63.00	0.001	0.999

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C47

Linear Mixed Model of adjusted VAF by Method

	Estimate	Std. Error	df	t-value	$P(>\|t\|)$
(Intercept)	0.702	0.010	17.371	69.989	>0.001***
1SE	0.024	0.008	63.000	2.905	0.005**
Relaxed	0.006	0.008	63.000	0.793	0.431
Adaptive	0.014	0.008	63.000	1.775	0.081.
1SE:Relaxed	−0.016	0.012	63.000	−1.409	0.164
1SE:Adaptive	−0.018	0.012	63.000	−1.523	0.133
Relaxed:Adaptive	−0.012	0.012	63.000	−1.085	0.282
1SE:Relaxed:Adaptive	0.009	0.016	63.000	0.526	0.601

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C48

Linear Model of Number of Predictors by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	21.38	0.324	65.968	>0.001***
1SE	−5.46	0.458	−11.912	>0.001***
Relaxed	−0.01	0.458	−0.022	0.983
Adaptive	−4.57	0.458	−9.971	>0.001***
1SE:Relaxed	−0.83	0.648	−1.280	0.201
1SE:Adaptive	0.70	0.648	1.080	0.281
Relaxed:Adaptive	−0.53	0.648	−0.818	0.414
1SE:Relaxed:Adaptive	−0.45	0.917	−0.491	0.624

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C49

Linear Model of Distances between Importance Measures by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	1.008	0.013	78.401	>0.001***
1SE	−0.493	0.018	−27.093	>0.001***
Relaxed	0.064	0.018	3.494	>0.001***
Adaptive	0.076	0.018	4.183	>0.001***
1SE:Relaxed	0.455	0.026	17.709	>0.001***
1SE:Adaptive	−0.025	0.026	−0.962	0.336
Relaxed:Adaptive	0.352	0.026	13.684	>0.001***
1SE:Relaxed:Adaptive	0.036	0.036	0.989	0.323

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

Table C50

Linear Model of Distances between Predictions by Method

	Estimate	Std. Error	t-value	$P(>\|t\|)$
(Intercept)	1.583	0.036	43.494	>0.001***
1SE	−0.395	0.051	−7.671	>0.001***
Relaxed	0.132	0.051	2.559	0.011*
Adaptive	0.044	0.051	0.857	0.392
1SE:Relaxed	0.330	0.073	4.535	>0.001***
1SE:Adaptive	0.115	0.073	1.580	0.115
Relaxed:Adaptive	0.172	0.073	2.361	0.019*
1SE:Relaxed:Adaptive	−0.015	0.103	−0.144	0.885

Note. p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001