



Group Variable Selection Methods with Quantile Regression: A Simulation Study

Hussein A. Hashem 

Department of Mathematics, College of Science, University of Duhok, Kurdistan Region, Iraq.

Article information

Article history:

Received :May 11,2024

Revised: October 5,2024

Accepted: December 16, 2024

Available online June 1,2025

Keywords:

Variable Selection, Group Variable Selection, Quantile Regression, Group Lasso, Regularization.

Abstract

In many cases, covariates have a grouping structure that can be used in the analysis to identify important groups and the significant members of those groups. This paper reviews some group variable selection methods that utilize quantile regression. The study compares seven previously proposed group variable selection methods, namely the group Lasso estimate, the quantile group Lasso (median group Lasso) estimate, the quantile group adaptive Lasso estimate, the sparse group Lasso estimate, the group scad estimate, the group mcp estimate, and the group gel estimate through a simulation study. The simulation study helps determine which methods perform best in all linear regression scenarios.

Correspondence:

Hussein A. Hashem

hussein.hashem@uod.ac

DOI: [10.33899/ijjoss.2025.187759](https://doi.org/10.33899/ijjoss.2025.187759) , ©Authors, 2025, College of Computer Science and Mathematics, University of Mosul.
This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Variable selection is a crucial task for analyzing high-dimensional data in various research fields such as biology, signal processing, and collaborative filtering. For instance, microarray experiments measure thousands of variables (genes, proteins) simultaneously. However, the data sets produced by these experiments are typically large in terms of the number of predictors (X) but small in terms of the number of biological samples (n). This problem is commonly known as the “large p and small n problem” and poses significant challenges to conventional statistical techniques, especially in regression analysis.

With the advancement of computer and data collection technologies, the size of databases has continued to increase. In response to this, various statistical methodologies have been developed over the past few decades to address the challenges posed by these large amounts of data. One of the major challenges is parameter estimation, model and variable selection. There have been several regression methods proposed for fitting multiple regression models, particularly in cases where the least-squares method cannot be used.

In 1996, Tibshirani [1] introduced a statistical method called Lasso (Least Absolute Shrinkage and Selection Operator), which aims to minimize the residual sum of squares while subject to a constraint on the L_1 norm. This approach leads to some coefficients being estimated as exactly zero, which helps to perform variable selection and estimation simultaneously. Since then, many extensions of the Lasso have been developed such as adaptive Lasso [2], Smoothly Clipped Absolute Deviation (SCAD)[3], and so on.

Quantile regression, which was first introduced by Koenker and Bassett in 1978[4], is a statistical technique that can be used to estimate different quantiles (e.g. the median) of a conditional distribution. It enables us to compare how predictor variables

affect different quantiles of the response variable. This provides valuable insights into how the relationship between variables changes across the distribution of the response variable.

Several methods have been proposed to perform variable selection in high-dimensional data with outliers by combining regularized and robust regression methods. One such method is the Huber Lasso method, proposed by Rosset and Zhu in 2007[5], which combines Huber's criterion loss with a Lasso penalty. Another method, proposed by Wang et al. in 2007[6], is the LAD-adaptive Lasso method, which combines the idea of Least Absolute Deviation (LAD) and L_1 -norm refers to the same concept. LAD is a more user-friendly term often used in statistics, while L_1 -norm is the more mathematical term used in various fields like linear algebra and machine learning. Both terms describe the sum of the absolute values of the differences between a set of data points and a central point (often the median). Additionally, Lambert-Lacroix and Zwald 2011[7] developed a method called Huber's Criterion with an adaptive Lasso, which combines Huber's loss function and adaptive Lasso penalty.

Fujisawa and Eguchi [8] proposed the gamma divergence for regression, which measures the difference between two conditional probability density functions. Arnold and Tibshirani [9] implemented the dual algorithm available in the R package genLasso. Taddy[10] introduced the gamma Lasso (GL) algorithm, which is a more computationally efficient, multi-convex relaxation of best variable selection. Yi and Huang [11] developed Semismooth Newton Coordinate Descent (SNCD), an algorithm that provides better efficiency and scalability for computing the solution paths of penalized quantile regression. Qin et al. [12] proposed the Maximum Tangent Likelihood Estimation (MTE) method. Christidis et al. [13] introduced the Split Regularized Regression (SRR) method, which is a more computationally efficient, multi-convex relaxation of best-split selection. Finally, Zhu et al. [14] proposed Whitening Lasso (WLasso), which removes correlations by applying a whitening transformation to the data before using the generalized Lasso criterion designed by Tibshirani and Taylor [15].

When the grouping structure is unknown and needs to be estimated, a group penalty can be applied. In biological studies, genetic data often comes with background scientific information. For instance, genes that share the same biological pathway are often found in a neighborhood, forming a group.

Several penalty methods have been proposed to consider the grouping structure. The Group Lasso, which uses the coefficients norm within a group, was first proposed by Bakin [16] and later extended by Yuan and Lin [17]. Huang et al. [18] then introduced group SCAD and group Minimax Concave Penalty (MCP) to select important groups for covariates with grouping structures. In the context of quantile regression models, Ciuperca [19] proposed an adaptive group Lasso with an adaptive Lasso penalty and established the sparsity and asymptotic normality of their methods. Kato [20] investigated the Group Lasso penalty for high-dimensional sparse quantile regression models and achieved a non-asymptotic error bound for estimation error. For the classification problem, Hashem et al. [21] explored the Group Lasso penalty approach.

Cai et al. [22] conducted a study on sparse group Lasso for high-dimensional double sparse linear regression. In this type of regression, the parameter of interest exhibits both element-wise and group-wise sparsity simultaneously. This problem is a significant example of a simultaneously structured model, which is a widely studied topic in the fields of statistics and machine learning. Huang et al. [23] examined various coding strategies and reference categories, and they concluded that the selection outcomes of lasso models heavily rely on these choices. This creates practical challenges when the lasso is employed with real-world data.

Moreover, McDonald [24] proposed a new R package for computing sparse Group Lasso, while Li et al. [25] introduced an adaptive sparse Group Lasso penalty for Logistic regression, which is used for cancer data diagnosis.

In the following section, we will provide an overview of various methods for selecting group variables in linear regression.

2. Methods

We will explain the regression regularization methods using the standard model of multiple linear regression. Let the data $(x_1, y_1), \dots, (x_n, y_n)$, and the design matrix denoted by $\mathbf{X} = (x_1^T, \dots, x_n^T)^T$, the general linear model is usually written as $y = \mathbf{X}\beta + u$

Here are the regression coefficients the random errors, x_i the regressors for observation $i, i = 1, \dots, n$ and $y = (y_1, \dots, y_n)^T$. The ordinary least squares (OLS) method estimates by minimizing the residual squared error, i.e. $\hat{\beta}_{OLS} = \min_{\beta} \{(y - \mathbf{X}\beta)^T(y - \mathbf{X}\beta)\}$. In general, OLS typically produces estimators that have low biases but high variances. To improve the accuracy of predictions, it is often necessary to slightly increase the bias to reduce the variance. We need to refer to it as a solution for specific problems in the model. For example, Ridge Regression in a linear model can be used for multiple regression models that suffer from multicollinearity problems.:

- Ridge regression introduces a bias-variance trade-off.
- shrinking coefficients reduce variance (better generalization) but introduce a slight bias.
- The λ parameter controls the strength of the penalty and the balance between bias and variance.

2.1 Lasso Regression

The Least Absolute Shrinkage and Selection Operator (LASSO), introduced by Tibshirani in 1996[1], is a widely utilized method for estimating regression coefficients and conducting variable selection in high-dimensional data settings. LASSO employs a regularization technique by imposing an L_1 –penalty on the regression coefficients, inducing shrinkage towards zero and promoting sparsity in the model. This method proves particularly beneficial when the number of predictor variables (p) significantly exceeds the number of samples (n).

Typically, the intercept (β_0) is exempt from the penalty, and its handling involves centring the input and response variables before model fitting. The primary objective of LASSO is to minimize the residual sum of squares while constraining the sum of absolute coefficient values to be less than a constant. The LASSO estimate ($\hat{\beta}$) comprises the coefficients that minimize this objective function.

2.2 Group Lasso Methods

In some real-world applications involving data analysis, it is common to have predictors that can be grouped naturally. In such cases, selecting groups of variables is of interest. Genetic data, for instance, can be grouped such that a group of genes corresponds to the same biological pathway. To accommodate this kind of situation, the group Lasso method was introduced by Yuan and Lin in 2006 [17]. This method is ideal for shrinking entire groups of predictors to 0 or estimating the regression coefficients for the entire group. The regression coefficients of groups will either all be 0 or all be nonzero.

For the group Lasso method, assume the predictor variables can be naturally grouped into k groups for $k = 1, \dots, K$, where each group consists of p_k predictor variables such that $\sum_{k=1}^K p_k = p$. Within each group k , there are j predictors for $j = 1, \dots, p_k$. The predictor variables should be standardized so that each x_{ij} has mean 0 and variance 1 for $j = 1, \dots, p$. The criterion to be minimized is:

$$\frac{1}{2} \sum_{i=1}^n (y_i - \sum_{k=1}^K x_{ik} \beta_k)^2 + n\lambda \sum_{k=1}^K \|\beta_k\|_2$$

where $\lambda \geq 0$ is a tuning parameter, y_i is the i th response, x_{ik} is a $1 \times p_k$ vector of predictors in the k th group for the i th observation, and β_k is a $p_k \times 1$ vector of regression coefficients for group k . As for the criterion above, for each group of predictors, minimize the sum of the squared distances, while simultaneously shrinking unimportant groups with the Lasso penalty (the L_2 the norm in this case). The tuning parameter λ controls the rate of shrinkage and can be chosen using cross-validation. In particular, Yuan and Lin [17] use a shrinkage parameter based on an approximate C_p -type criterion. The Lasso method is a popular technique for selecting predictors while estimating their values simultaneously. However, it is not suitable for data with outliers or high multicollinearity. The group Lasso, which uses the Least Square Estimate (LSE), is particularly vulnerable to outliers and may not perform well. The shooting algorithm is used to compute the group Lasso. Although the shooting algorithm was originally proposed for the Lasso method, it was later adapted for the group Lasso by Yuan and Lin in 2006 [17].

2.3 Group Descent Algorithms(*grpreg*)

A statistical method called "grouped penalties" is useful when dealing with models that have a large number of predictors. However, this method is often limited to linear regression models or models in which the members of a group are orthogonal to each other. To solve this problem, Breheny and Huang [26] combined the ideas of coordinate descent optimization and local approximation of penalty functions to create a new algorithm that can be used for fitting models with grouped penalties. This algorithm is both stable and fast, even when there are many more variables than there are samples. Although the algorithm was initially developed for models with grouped penalties, it can be applied to other penalized regression problems in which the penalties are complicated. The R package developed by Breheny and Huang [26] contains all the necessary group-related methods, except for ElasticNet, which is available separately.

2.4 Quantile Regression

The Ordinary least squares (OLS) regression estimates the mean response based on predictor variables. However, an alternative approach known as least absolute deviation (LAD) regression estimates the conditional median function. LAD regression is particularly advantageous in scenarios with response outliers and heavy-tailed errors, as it offers greater robustness.

In 1978, Koenker and Bassett [4] introduced quantile regression (QR) as an extension of LAD regression. QR estimates the conditional quantile function of the response, thereby providing comprehensive insights into the conditional distribution of the response variable. QR inherits the desirable properties of LAD regression while offering a more informative model overall.

Here's a brief review of quantile regression models. Given the data $(x_1, y_1), \dots, (x_n, y_n)$, unlike the mean regression model which models the conditional mean $E(y|X) = X\beta$.

Koenker and Bassett [15] proposed the linear quantile regression model for the θ th quantile ($0 < \theta < 1$) as

$$y_i = x_i^T \beta + u_i, \quad i = 1, \dots, n$$

Where $\beta = (\beta_1, \dots, \beta_p)^T \in R^p$ and u_i 's are independent with their θ th quantiles equal to zero.

Quantile regression offers a flexible and comprehensive approach to modelling the relationship between response variables and predictors by varying the quantile parameter θ . Notably, when θ equals 0.5, quantile regression reduces to the least absolute deviation regression or median regression, renowned for its robustness to outliers. This method estimates the conditional quantiles of a response variable and is widely acknowledged for its robustness to outliers, making it a preferred choice in such scenarios. The Least Absolute Deviation (LAD) regression is essentially the same as median regression because both LAD regression and median regression aim to minimize the absolute deviations between the predicted values and the actual values in the data. LAD regression minimizes the sum of the absolute values of the residuals (differences between predicted and actual values). Median regression aims to find the line (or hyperplane in higher dimensions) that minimizes the absolute deviations of the data points from a central point - the median.

A significant advantage of quantile regression is a powerful tool when the assumptions of least squares regression are not met or when you need a more detailed understanding of the relationship between variables across different parts of the conditional distribution. However, its interpretation and computational aspects require careful consideration. In practice, the coefficients can be consistently estimated by solving a minimization problem, providing reliable parameter estimates across various quantiles of interest. $\min_{\beta} \sum_{i=1}^n \rho_{\theta}(y_i - x_i^T \beta)$

where $\rho(\cdot)$ is an outlier-resistant loss function called the objective function

$$\rho_{\theta}(t) = \begin{cases} \theta t & \text{if } t \geq 0 \\ -(1 - \theta)t & \text{if } t < 0 \end{cases}, \text{ where } 0 < \theta < 1.$$

The inaugural application of regularization in quantile regression occurred in 2004, spearheaded by Koenker. In this pioneering work, the LASSO penalty was introduced to address random effects within a mixed-effect quantile regression framework. The objective was to induce shrinkage of the random effects towards zero, leveraging the regularization properties of the LASSO method. This innovative approach marked a significant advancement in the field, offering a novel means of addressing model complexity and improving estimation precision in mixed-effect quantile regression models.

3. Simulation Study

In this section, we compare group variable selection methods in low-dimensional settings with sparse and non-sparse coefficients ($p = 50, n = 100$) and high-dimensional settings with sparse coefficients ($p = 100, n = 50$). For the sparse settings, we use a classical simulation setting, e.g. Yu et al. [27] and Li et al. [28] where $y = \beta_0 + x\beta + u$, with $\beta_0 = 0$ and we create a group structure by simulating 10 groups, each consisting of 10 covariates. The 100 variables are assumed to follow a multivariate normal distribution $N(0; \Sigma)$, with Σ having a diagonal block structure. Each block corresponds to one group and is defined by the matrix $r^{|i-k|}$, $i = 1, \dots, 10$, $k = 1, \dots, 10$. For the correlation r , we experiment both with $r = 0.95$ (well-defined group structure) and $r = 0.5$. For the β values we consider three cases:

- (1) The values for the first three groups are given by $\beta_j = (0.5, 1, 1.5, 2, 2.5, 2, 2, 2, 2, 2), (2, 2, 1, 1, 1, 1, 3, 3, 3, 3), (1, 1, 1, 2, 2, 2, 3, 3, 3, 3)$, and they are set to zero for all other groups, which corresponds to the sparse case with group structures in the predictors.
- (2) $\beta_j = (1, 2, 3, 4, 5, 0.1, 0.2, 0.3, 0.4, 0.5)$, and they are set to zero for all other groups, which corresponds to the very sparse case with group structures in the predictors.
- (3) $\beta_j = 0.1$ for all j , which corresponds to a dense case.

For the error ϵ , we will examine the following distributions, which are skewed due to the presence of outliers, to assess the robustness of the compared methods:

- normal: $N(0; 1)$
- Laplace distribution with location 0 and scale 1: $Laplace(0, 1)$
- A t distribution with 3 degrees of freedom: t_3
- Gamma distribution: $G(3, 1)$
- A mixture of two normal distributions: $0.1N(0, 100) + 0.9N(0, 1)$
- A mixture of two Laplace distributions: $0.1Laplace(0, 1) + 0.9Laplace(0, 2)$
- Chi-square distributions: $\chi_{(3)}^2$

We compare the group variable selection methods described in the previous section, namely:

- "*grp.lasso*": group Lasso penalty (Yuan and Lin,[29])
- "*qgrp.lasso*": quantile group Lasso (median group Lasso) (see Sherwood et al., [22]).
- "*qgrad.lasso*": quantile group adaptive Lasso (see Sherwood et al., [22]).
- "*sparse.grp.lasso*": sparse group Lasso penalty (group Lasso + Lasso), extra parameters *tau* (see Xiong et al., [27], Huling and Chien, [30]).
- "*grp.scad*": group smoothly clipped absolute deviation, extra parameters *gamma* (see Xiong et al., [29], Huling and Chien, [30]).
- "*grp.mcp*": group minimax concave penalty, extra parameters *gamma* (see Xiong et al., [29], Huling and Chien, [30]).
- "*grp.gel*": group exponential Lasso (Breheny, [31])

For the *grp.lasso* and *sparse.grp.lasso* methods we use the R package *oem*, for the *grp.scad*, *grp.mcp* and *grp.gel* methods we use the R package *grpreg* for *qgrp.lasso* and *qgrad.lasso* we use the R package *rqPen* [32]

3.1 Simulation 1: low-dimensional with sparse coefficients (Case 1)

In this section, we are analyzing data that has low-dimension and sparse coefficients. The dataset we are working with has 50 variables and 100 observations. We present the simulation results in Figure 1, Table 1.A, and Table 1.B, where we examine the cases of low correlation ($r = 0.5$) and high correlation ($r = 0.95$) among the predictors. Figure 1 displays the median model error over 500 iterations. The mean error produces similar results, with the model error computed by $(\hat{\beta} - \beta)^T S_x (\hat{\beta} - \beta)$, where $\hat{\beta}$ are the estimated parameters and S_x the sample covariance.

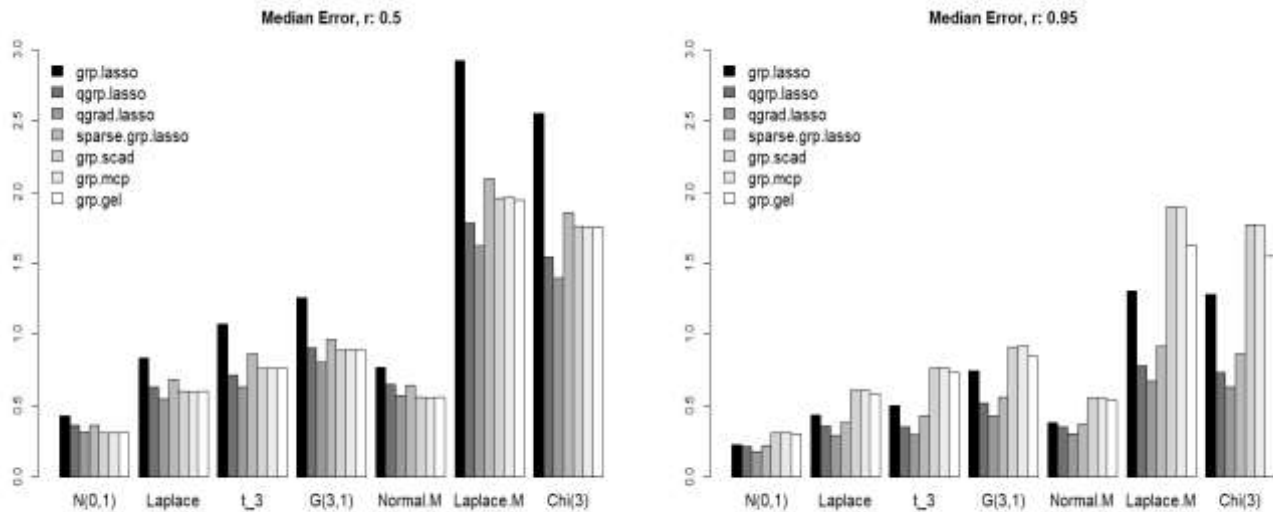


Figure 1: Comparison of group variable selection methods under different error distributions. The median model error over 500 replications for Simulation 1 when $p = 50$ and $n = 100$.

Table1.A: Average Median Model Error over 500 replications for the case: $p = 50, n = 100, r = 0.5$, and β values as in Simulation 1.

	lasso	qgrp.lasso	qgrad.lasso	sparse.lasso	scad	mcp	gel
N(0,1)	0.427	0.365	0.312	0.364	0.305	0.306	0.306
Laplace	0.834	0.628	0.548	0.685	0.598	0.597	0.602
t_3	1.071	0.715	0.633	0.862	0.763	0.762	0.761
G(3,1)	1.257	0.906	0.804	0.964	0.885	0.891	0.887
Normal. M	0.769	0.650	0.569	0.645	0.557	0.555	0.556
Laplace. M	2.922	1.783	1.621	2.096	1.958	1.965	1.943
Chi(3)	2.551	1.546	1.394	1.856	1.757	1.754	1.755

Table1.B: Average Median Model Error over 500 replications for the case: $p = 50, n = 100, r = 0.95$, and β values as in Simulation 1.

	lasso	qgrp.lasso	qgrad.lasso	sparse.lasso	scad	mcp	gel
N(0,1)	0.226	0.212	0.177	0.217	0.309	0.308	0.301
Laplace	0.435	0.355	0.292	0.382	0.608	0.608	0.584
t_3	0.502	0.351	0.293	0.429	0.763	0.765	0.730
G(3,1)	0.746	0.520	0.427	0.561	0.911	0.916	0.847
Normal. M	0.381	0.348	0.301	0.369	0.553	0.554	0.534
Laplace. M	1.304	0.779	0.674	0.916	1.896	1.892	1.630
Chi(3)	1.279	0.731	0.631	0.863	1.768	1.763	1.549

Our results indicate that the grp. scad, grp.mcp, and grp.gel methods do not perform well. However, the qgrad.lasso method outperforms all other methods when predictors are highly correlated, for most error distributions.

For most LASSO problems, the standard lasso function is the recommended choice due to its efficiency and simplicity. Use glmnet if you need the flexibility of L_1/L_2 regularization or are working with classification problems. Consider sparse.lasso only for very large and sparse datasets where memory limitations become a concern. Avoid qgrad.lasso unless you have a specific reason to use the QGD algorithm for research or experimentation. The best choice depends on the specific characteristics of your data and the computational resources available. If you're unsure, start with the standard lasso function and explore alternatives like sparse.lasso if efficiency becomes a bottleneck with large datasets.

3.2 Simulation 2: high-dimensional with sparse coefficients (Case 1)

We are examining a scenario that is similar to simulation 3.1, but with a different sample size and multiple predictors. Specifically, we are dealing with a high-dimensional simulation where the coefficients are sparse and p equals 100, while n equals 50. The median model error across multiple replications is reported in Figure 2, Table 2.A, and Table 2.B. The model error is calculated in the same way as in Figure 1.

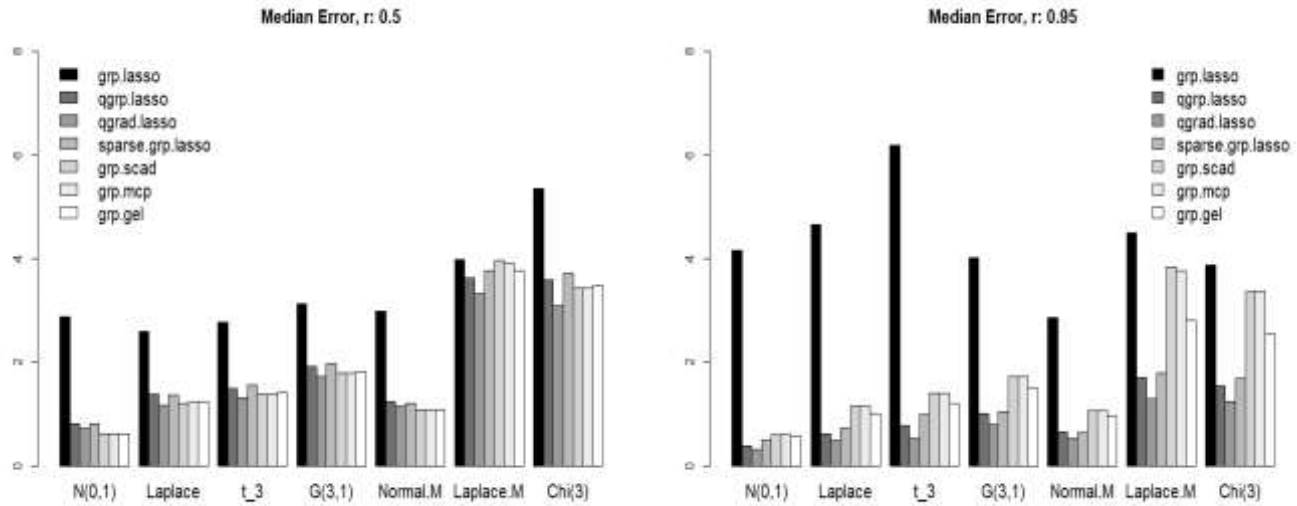
Figure 2: Comparison of group variable selection methods under different error distributions. The median model error over 500 replications for Simulation 2 when $p = 100$ and $n = 50$.

Table2.A: Average Median Model Error over 500 replications for the case: $p = 100$, $n = 50$, $r = 0.5$, and β values as in Simulation 1.

	lasso	qgrp.lasso	qgrad.lasso	sparse.lasso	scad	mcp	gel
N(0,1)	2.879	0.817	0.725	0.810	0.612	0.612	0.618
Laplace	2.588	1.380	1.167	1.357	1.206	1.213	1.231
t_3	2.773	1.493	1.304	1.562	1.379	1.385	1.413
G(3,1)	3.119	1.920	1.722	1.961	1.794	1.796	1.805
Normal. M	2.992	1.241	1.144	1.209	1.079	1.079	1.080
Laplace. M	3.987	3.627	3.325	3.775	3.968	3.913	3.764
Chi(3)	5.352	3.586	3.093	3.716	3.434	3.432	3.473

Table2.B: Average Median Model Error over 500 replications for the case: $p = 100$, $n = 50$, $r = 0.95$, and β values as in Simulation 1.

	lasso	qgrp.lasso	qgrad.lasso	sparse.lasso	scad	mcp	gel
N(0,1)	4.161	0.379	0.312	0.486	0.591	0.591	0.555
Laplace	4.665	0.620	0.484	0.730	1.152	1.152	1.006
t_3	6.179	0.774	0.533	0.990	1.398	1.397	1.203
G(3,1)	4.021	1.005	0.795	1.040	1.716	1.720	1.481
Normal. M	2.853	0.652	0.527	0.636	1.069	1.069	0.958
Laplace. M	4.493	1.708	1.293	1.784	3.844	3.769	2.810
Chi(3)	3.869	1.543	1.233	1.677	3.364	3.377	2.560

The results of the study show that grp.lasso method does not perform well when the predictors are highly correlated. On the other hand, the qgrad.lasso method outperforms all other methods as departures from normality increase.

3.3 Simulation 3: low- dimensional with very sparse coefficients (Case 2)

To examine how well group variable selection methods perform in Simulation 1, we created a fresh simulation scenario. In this new setup, we have a very sparse problem similar to Case 2 where most of the coefficients are equal to zero. Figure 3 depicts the median model error across repeated trials, with the same method of calculating model error as seen in Figure 1.

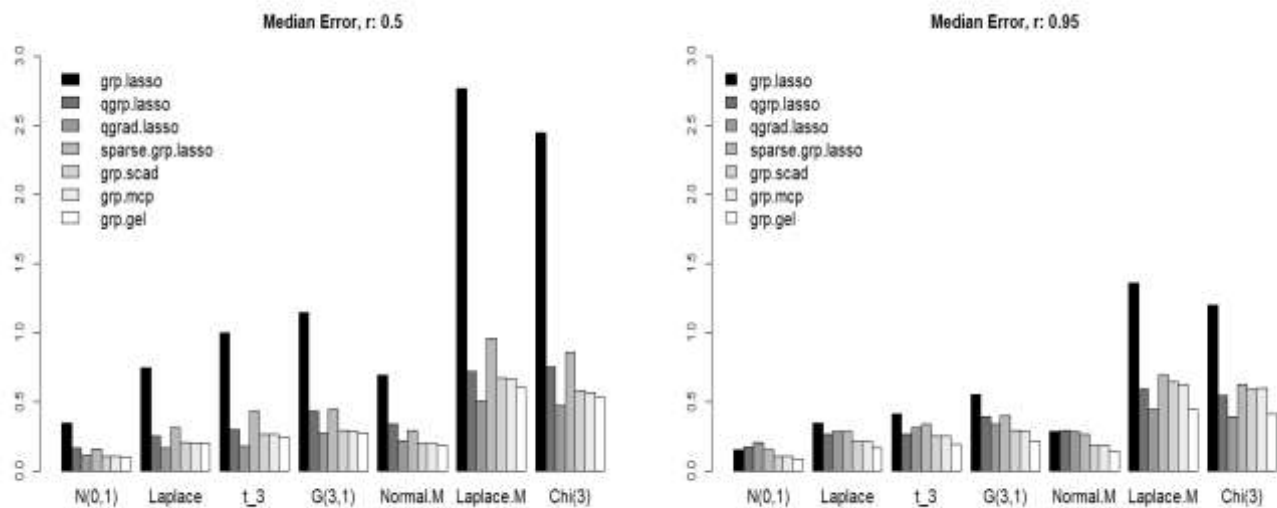
Figure 3: Comparison of group variable selection methods under different error distributions. The median model error over 500 replications for Simulation 3 when $p = 50$ and $n = 100$.ss

Table3.A: Average Median Model Error over 500 replications for the case: $p = 50, n = 100, r = 0.5$, and β values as in Simulation 3.

	lasso	qgrp.lasso	qgrad.lasso	sparse.lasso	scad	mcp	gel
N(0,1)	0.346	0.166	0.107	0.157	0.104	0.103	0.100
Laplace	0.747	0.251	0.161	0.313	0.203	0.199	0.198
t_3	1.002	0.301	0.179	0.429	0.263	0.263	0.246
G(3,1)	1.144	0.434	0.271	0.443	0.293	0.285	0.272
Normal. M	0.691	0.336	0.218	0.292	0.196	0.195	0.187
Laplace. M	2.766	0.715	0.502	0.956	0.670	0.664	0.604
Chi(3)	2.447	0.754	0.472	0.862	0.581	0.564	0.532

Table3.B: Average Median Model Error over 500 replications for the case: $p = 50, n = 100, r = 0.95$, and β values as in Simulation 3.

	lasso	qgrp.lasso	qgrad.lasso	sparse.lasso	scad	mcp	gel
N(0,1)	0.150	0.170	0.207	0.153	0.101	0.101	0.083
Laplace	0.343	0.267	0.281	0.283	0.213	0.209	0.161
t_3	0.413	0.261	0.308	0.336	0.250	0.249	0.193
G(3,1)	0.550	0.391	0.337	0.398	0.289	0.287	0.211
Normal. M	0.282	0.293	0.283	0.264	0.183	0.184	0.146
Laplace. M	1.358	0.591	0.447	0.693	0.642	0.627	0.445
Chi(3)	1.200	0.544	0.392	0.625	0.590	0.597	0.412

Based on the data presented in Figure 3, Table 3A, and Table 3B, our simulation study concludes that the group exponential Lasso (grp.gel) is the most effective method as non-normality increases. This is especially true when the predictors are strongly correlated.

3.4 Simulation 4: high-dimensional with very sparse coefficients (Case 2)

We are exploring a scenario similar to simulation 3.3 but with a larger number of predictors and a different sample size. Specifically, we are examining a high-dimensional simulation with sparse coefficients, where there are 100 predictors and 50 observations. Figure 4 displays the median model error across 500 replications. The model error is calculated in the same way as in Figure 3.

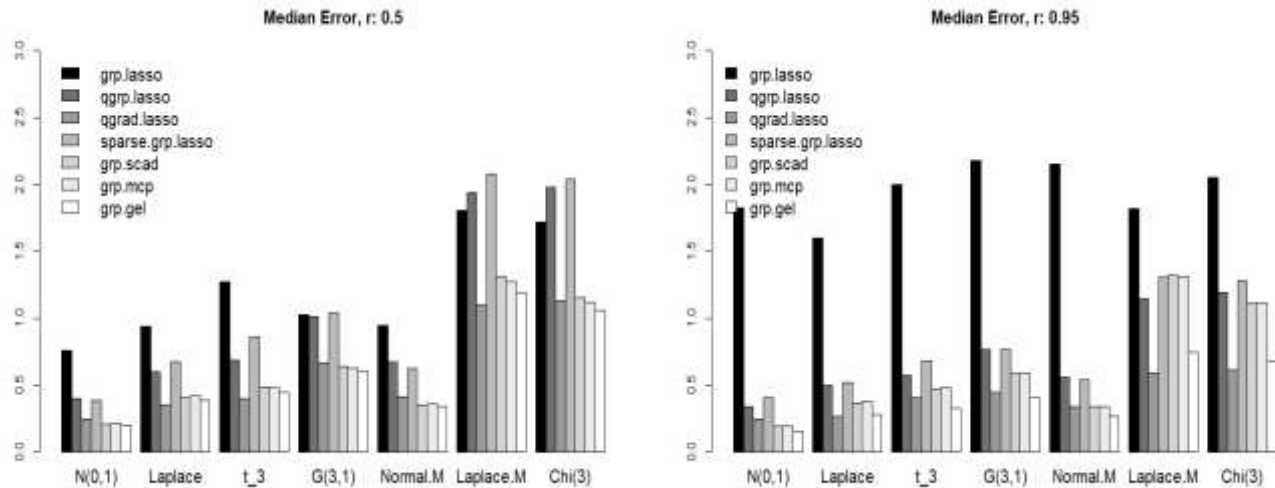


Figure 4: Comparison of group variable selection methods under different error distributions. The median model error over 500 replications for Simulation 4 when $p = 100$ and $n = 50$.

Table4.A: Average Median Model Error over 500 replications for the case: $p = 100, n = 50, r = 0.5$, and β values as in Simulation 3.

	lasso	qgrp.lasso	qgrad.lasso	sparse.lasso	scad	mcp	gel
N(0,1)	0.755	0.397	0.243	0.383	0.204	0.209	0.194
Laplace	0.935	0.597	0.354	0.671	0.406	0.420	0.383
t_3	1.271	0.683	0.396	0.855	0.482	0.480	0.443
G(3,1)	1.024	1.014	0.668	1.038	0.640	0.626	0.606
Normal. M	0.945	0.672	0.414	0.626	0.352	0.357	0.339
Laplace. M	1.804	1.941	1.099	2.079	1.313	1.274	1.190
Chi(3)	1.722	1.980	1.133	2.041	1.151	1.122	1.059

Table4.B: Average Median Model Error over 500 replications for the case: $p = 100, n = 50, r = 0.95$, and β values as in Simulation 3.

	lasso	qgrp.lasso	qgrad.lasso	sparse.lasso	scad	mcp	gel
N(0,1)	1.825	0.336	0.242	0.403	0.189	0.189	0.148
Laplace	1.602	0.496	0.267	0.518	0.367	0.375	0.276
t_3	2.002	0.568	0.406	0.675	0.469	0.476	0.326
G(3,1)	2.183	0.767	0.443	0.763	0.585	0.584	0.402
Normal. M	2.154	0.559	0.341	0.540	0.339	0.339	0.262
Laplace. M	1.819	1.142	0.593	1.306	1.316	1.314	0.745
Chi(3)	2.055	1.193	0.613	1.282	1.112	1.110	0.680

Based on the findings presented in Figure 5, Table 4A, and Table 4B, our simulation study concludes that the grp.gel and qgrad.lasso methods outperform all other methods as the degree of deviation from normality increases. This is especially noticeable when the predictors are strongly correlated.

3.5 Simulation 5: low- dimensional with non-sparse coefficients (Case 3)

To examine how well group variable selection methods perform in non-sparse settings, we conducted a new simulation that closely resembled case 3. This simulation involved a non-sparse situation, and we analyzed the median model error over 500 replications for the scenarios where the number of variables (p) is 50 and the number of observations (n) is 100. The results of this analysis are presented in Figure 5.

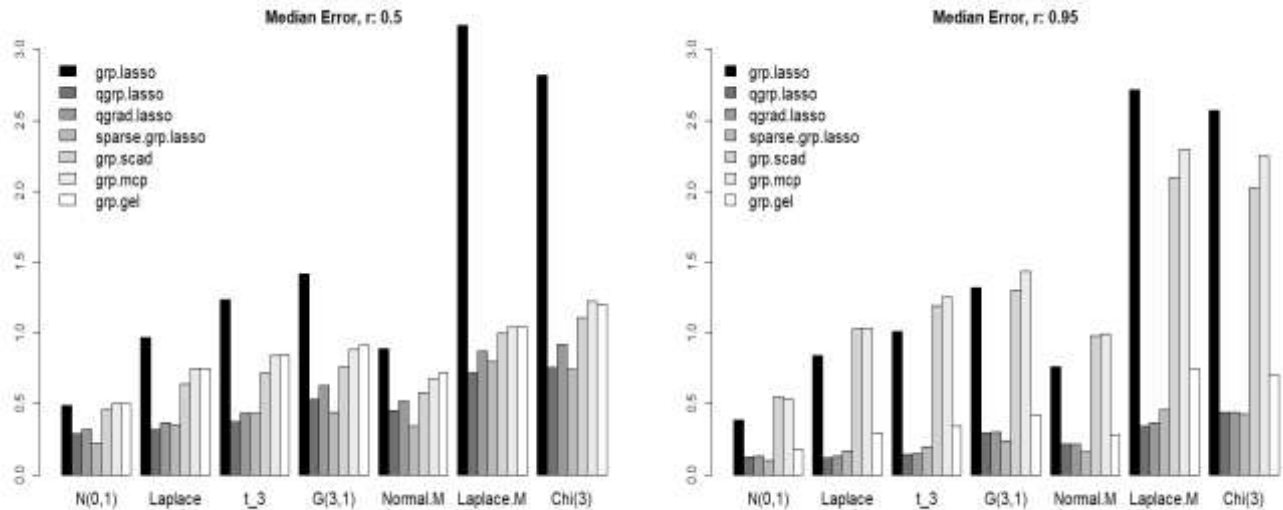
Figure 5: Comparison of group variable selection methods under different error distributions. The median model error is over 500 replications for Simulation 5 when $p = 50$ and $n = 100$.

Table5.A: Average Median Model Error over 500 replications for the case: $p = 50, n = 100, r = 0.5$, and β values as in Simulation 5.

	lasso	qgrp.lasso	qgrad.lasso	sparse.lasso	scad	mcp	gel
N(0,1)	0.491	0.290	0.316	0.225	0.459	0.503	0.499
Laplace	0.969	0.316	0.365	0.350	0.639	0.750	0.750
t_3	1.232	0.372	0.432	0.432	0.718	0.841	0.844
G(3,1)	1.420	0.531	0.628	0.438	0.763	0.883	0.913
Normal. M	0.887	0.450	0.520	0.344	0.580	0.678	0.717
Laplace. M	3.170	0.717	0.872	0.800	0.998	1.040	1.040
Chi(3)	2.819	0.758	0.919	0.743	1.108	1.226	1.200

Table5.B: Average Median Model Error over 500 replications for the case: $p = 50, n = 100, r = 0.95$, and β values as in Simulation 3.

	lasso	qgrp.lasso	qgrad.lasso	sparse.lasso	scad	mcp	gel
N(0,1)	0.389	0.128	0.133	0.098	0.550	0.533	0.181
Laplace	0.842	0.122	0.132	0.170	1.030	1.036	0.293
t_3	1.013	0.145	0.150	0.197	1.196	1.254	0.345
G(3,1)	1.321	0.296	0.302	0.239	1.305	1.434	0.422
Normal. M	0.764	0.214	0.217	0.159	0.980	0.995	0.280
Laplace. M	2.714	0.344	0.368	0.460	2.098	2.296	0.748
Chi(3)	2.572	0.438	0.440	0.427	2.021	2.247	0.699

Based on the findings shown in Figure 5 and Tables 5.A and 5.B, our simulation study confirms that both the qgrp.lasso and sparse.lasso methods perform better than all other methods as the extent of non-normality increases. This is especially apparent when the predictors are highly correlated.

3.6 Simulation 6: high-dimensional with non-sparse coefficients (Case 3)

To examine the effectiveness of group variable selection methods in Simulation 2, we established a fresh simulation. This simulation is similar to case 3 in that it involves a non-sparse situation. Figure 6 displays the median model error from 500 replications for the scenarios where $p = 100$ and $n = 50$.

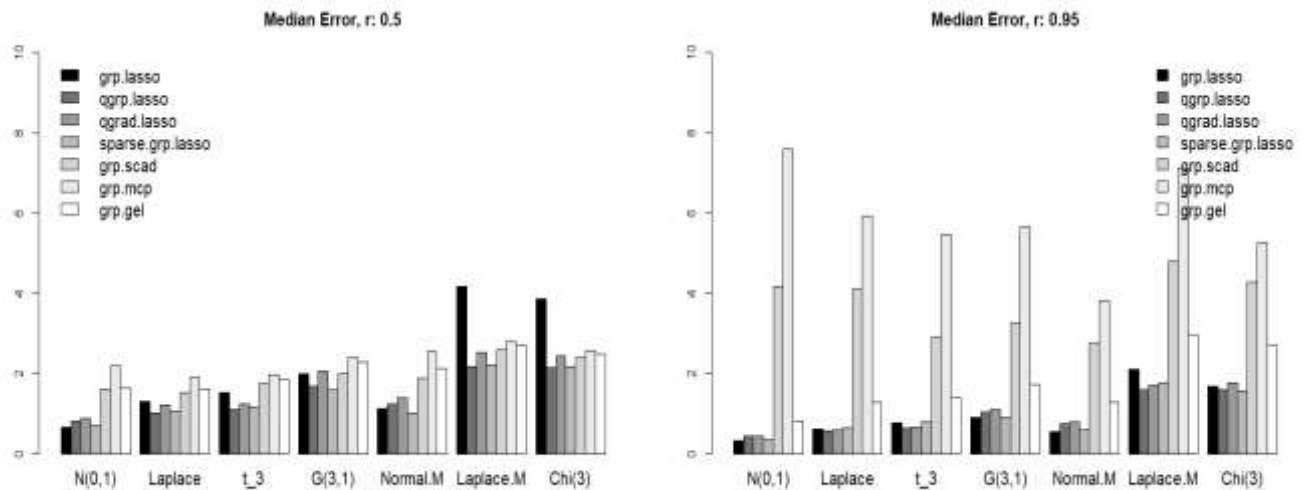


Figure 6: Comparison of group variable selection methods under different error distributions. The median model error is over 500 replications for Simulation 6 when $p = 100$ and $n = 50$.

Table 6.A: Average Median Model Error over 500 replications for the case: $p = 50, n = 100, r = 0.5$, and β values as in Simulation 5.

	lasso	qgrp.lasso	qgrad.lasso	sparse.lasso	scad	mcp	gel
N(0,1)	0.650	0.824	0.871	0.713	1.623	2.194	1.634
Laplace	1.297	1.015	1.217	1.065	1.517	1.905	1.615
t_3	1.527	1.103	1.233	1.178	1.751	1.973	1.865
G(3,1)	2.004	1.682	2.055	1.594	1.993	2.383	2.289
Normal. M	1.116	1.240	1.398	1.019	1.878	2.556	2.130
Laplace. M	4.182	2.169	2.526	2.223	2.592	2.794	2.715
Chi(3)	3.857	2.151	2.449	2.164	2.391	2.567	2.479

Table6.B: Average Median Model Error over 500 replications for the case: $p = 50, n = 100, r = 0.95$, and β values as in Simulation 3.

	lasso	qgrp.lasso	qgrad.lasso	sparse.lasso	scad	mcp	gel
N(0,1)	0.327	0.439	0.439	0.351	4.175	7.607	0.819
Laplace	0.603	0.570	0.600	0.632	4.104	5.937	1.283
t_3	0.781	0.625	0.660	0.786	2.910	5.455	1.402
G(3,1)	0.894	1.034	1.114	0.912	3.254	5.672	1.731
Normal. M	0.548	0.750	0.799	0.590	2.740	3.810	1.272
Laplace. M	2.109	1.586	1.711	1.761	4.797	7.143	2.969
Chi(3)	1.671	1.588	1.766	1.568	4.277	5.260	2.713

According to the results presented in Figure 6, Table 5A and Table 5B, our simulation study has confirmed that the qgrp.lass method outperforms all other methods as the degree of departure from normality increases. Moreover, the results also indicate that grp.mcp is the worst performing method, particularly when the predictors are highly correlated and there is a significant deviation from normality.

4. Conclusions

In statistics, many methods rely on the assumption of normality. However, these methods may not be suitable for data that deviate significantly from normality, such as when outliers are present. Recently, group variable selection methods have been developed, such as the group Lasso method and the quantile group Lasso (median group Lasso). These methods are particularly useful in high-dimensional settings, where the number of predictors (p) is greater than the sample size (n). In a simulation study, we found that the quantile group adaptive Lasso (qgrad.lasso) and the group exponential Lasso (grp.gel) methods outperformed other group methods, especially in cases where there was a large departure from normality.

Acknowledgment

The author is very grateful to the University of Duhok, College of Science for their provided facilities, which helped improve this work's quality.

References

1. Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* 58, 267–288.
2. Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association* 101, 1418–1429.
3. Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association* 96(456), 1348–1360.
4. Koenker, R. and G. W. Bassett (1978). Regression quantiles. *Econometrica* 46, 33–50.
5. Rosset, S. and Zhu, J. (2007). Piecewise linear regularized solution paths. *The Annals of Statistics* 35 (3), 1012–1030.
6. Wang, H., Li, G., and Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the LAD-Lasso. *Journal of Business & Economic Statistics* 25, 347 - 355.
7. Lambert-Lacroix, S. and Zwald, L. (2011). Robust regression through the Huber's criterion and adaptive lasso penalty. *Electronic Journal of Statistics* 5, 1015–1053.
8. Fujisawa, H. and Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination, *Journal of Multivariate Analysis*, 99(9), 2053–2081.
9. Arnold, T. B., and Tibshirani, R. J. (2014). Efficient implementations of the generalized Lasso dual-path algorithm. *Journal of Computational and Graphical Statistics*, 25(1):1–27, 2016.
10. Taddy, M. (2017). One-step estimator paths for concave regularization, *Journal of Computational and Graphical Statistics* pp. 1–12.
11. Yi, C. Huang, J. (2016). Semismooth Newton Coordinate Descent Algorithm for Elastic-Net Penalized Huber Loss Regression and Quantile Regression. *Journal of Computational and Graphical Statistics* 3. 547–557.
12. Qin, Y. Li, S. and Yu, Y. (2017) Penalized Maximum Tangent Likelihood Estimation and Robust Variable Selection, <https://arxiv.org/pdf/1708.05439.pdf>.
13. Christidis, A.-A., Lakshmanan, L., Smucler, E., and Zamar, R. (2020). Split regularized regression. *Technometrics* 62.3, pp. 330–338.
14. Zhu, W., L'evy-Leduc, C., and Tern'es, N. (2021). A variable selection approach for highly correlated predictors in high-dimensional genomic data. *Bioinformatics*, 37(16), 2238– 2244.
15. Tibshirani, R. J., and Taylor, J. (2011). The solution path of the generalized Lasso. *Ann.Stat.*, 39(3), 1335–1371.
16. Bakin S (1999). Adaptive regression and model selection in data mining problems (Ph.D. thesis), The Australian National University.
17. Yuan M and Lin Y (2006). Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, B (Statistical Methodology), 68, 49–67.
18. Huang J, Breheny P, and Ma S (2012). A Selective Review of Group Selection in High-Dimensional Models, *Statistical Science*, 27, 481–499.
19. Ciuperca, G. (2019). Adaptive group LASSO selection in quantile models, *Statistical Papers*, 60, 173–197.
20. Kato, K. (2011). Group Lasso for high dimensional sparse quantile regression models, *arXiv:1103.1458 v2 [stat.ME]*.
21. Hashem H., Vinciotti V., Alhamzawi, R, and Yu, K. (2016). Quantile regression with group lasso for classification, *Advances in Data Analysis and Classification*, 10, 375–390.
22. Cai, T. T., Zhang, A. R., and Zhou, Y. (2022). Sparse group lasso: Optimal sample complexity, convergence rate, and statistical inference. *IEEE Transactions on Information Theory*, 68, 5975–6002.
23. Huang, Y., Tibbe, T., Tang, A., & Montoya, A. (2023). Lasso and Group Lasso with Categorical Predictors: Impact of Coding Strategy on Variable Selection and Prediction. *Journal of Behavioral Data Science*, 3(2), 15-42.
24. McDonald, D. J. (2022). sparsegl: An R Package for Estimating Sparse Group Lasso. <https://arxiv.org/pdf/2208.02942.pdf>.
25. Li, J. , Liang, K. and Song,X. (2022). Logistic regression with adaptive sparse group Lasso penalty and its application in acute leukemia diagnosis. *Computers in Biology and Medicine*, Volume 141.
26. Breheny P and Huang J. (2015) Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and Computing*, 25: 173-187.
27. Yu, K., C. Cathy, C. Reed, and D. Dunson (2013). Bayesian variable selection in quantile regression. *Statistics and Its Interface* 6, 261–274
28. Li, Q., R. Xi, and N. Lin (2010). Bayesian regularized quantile regression. *Bayesian Analysis* 5, 1–24
29. Xiong S, Dai B, Huling J, Qian PZ. (2016). Orthogonalizing EM: A design-based least squares algorithm. *Technometrics*; 58(3): 285-93.

30. Huling, J.D.; Chien, P. (2018). Fast Penalized Regression and Cross-Validation for Tall Data with the OEM Package. J. Stat. Softw.
31. Breheny P. (2015). The group exponential lasso for bi-level variable selection. Biometrics, 71: 731-740.
32. Sherwood, B. ,Li, S. and Maidman, A. (2016). rqPen: Penalized Quantile Regression. R package version 1.4.

طرق اختيار مجموعة المتغيرات مع الانحدار الكمي: دراسة محاكاة

حسين عبد الرحمن هاشم

قسم الرياضيات، كلية العلوم ، جامعة دهوك، دهوك، العراق

الخلاصة: في العديد من الحالات، تحتوي المتغيرات المشتركة على بنية تجميع يمكن استخدامها في التحليل لتحديد المجموعات المهمة والمتغيرات المهمة في تلك المجموعات. يستعرض هذا البحث بعض طرق اختيار متغيرات المجموعة التي تستخدم الانحدار الكمي. تقارن الدراسة سبع طرق اختيار متغيرات المجموعة المقترحة مسبقاً، وهي تقدير مجموعة لاسو (group Lasso)، وتقدير مجموعة كمية لاسو (quantile group Lasso) (المجموعة المتوسطة)، وتقدير لاسو التكيفي للمجموعة الكمية (the quantile group adaptive Lasso)، وتقدير لاسو للمجموعة المتناثرة (the sparse group Lasso)، وتقدير scad للمجموعة، وتقدير mcp للمجموعة. ، وتقدير gel للمجموعة من خلال دراسة المحاكاة. تساعد دراسة المحاكاة في تحديد الأساليب التي تحقق أفضل أداء في جميع سيناريوهات الانحدار الخطي.

الكلمات المفتاحية: اختيار متغير ؛ اختيار متغير المجموعة الانحدار الكمي ، لاسو المجموعة ، التنظيم.