

# Lasso Regression

Test name: Lasso Regression

Dataset: D:\DATA ANALYSIS\H Regression Tests and Models\Lasso Regression\dataset.csv

Rows in original dataset: 649

Columns in original dataset: 33

Rows used in Lasso model: 649

Outcome variable: G3

Numeric predictors: G1, G2, studytime, failures, absences, age, Medu, Fedu

Categorical predictors: school, sex, address, famsize

Encoded predictor columns: 12

Cross-validation folds: 5

Selected alpha/lambda: 0.016927

Selected non-zero predictors: 10

Held-out test RMSE: 1.1156

Held-out test MAE: 0.7253

Held-out test R-squared: 0.8765

Lasso selected at least one predictor.

Interpretation: Lasso uses L1 regularization; coefficients equal to zero are removed from the selected model.

Decision guide: smaller test RMSE and larger test R-squared indicate better predictive performance.

Reporting guide: state outcome, predictors, scaling, selected alpha/lambda, retained predictors, RMSE, MAE, R-squared, and coefficient signs.

Charts generated: 7

# Lasso selected coefficients

| term        | standardized_lasso_coefficient | absolute_standardized_coefficient | selected_by_lasso | rank_by_absolute_coefficient |
|-------------|--------------------------------|-----------------------------------|-------------------|------------------------------|
| G2          | 2.4765                         | 2.4765                            | Selected          | 1.0000                       |
| G1          | 0.4670                         | 0.4670                            | Selected          | 2.0000                       |
| school_MS   | -0.1176                        | 0.1176                            | Selected          | 3.0000                       |
| Medu        | -0.0867                        | 0.0867                            | Selected          | 4.0000                       |
| absences    | 0.0834                         | 0.0834                            | Selected          | 5.0000                       |
| failures    | -0.0766                        | 0.0766                            | Selected          | 6.0000                       |
| sex_M       | -0.0640                        | 0.0640                            | Selected          | 7.0000                       |
| address_U   | 0.0532                         | 0.0532                            | Selected          | 8.0000                       |
| famsize_LE3 | 0.0485                         | 0.0485                            | Selected          | 9.0000                       |
| studytime   | 0.0412                         | 0.0412                            | Selected          | 10.0000                      |
| Fedu        | 0.0000                         | 0.0000                            | Shrunk to zero    | 11.0000                      |
| age         | -0.0000                        | 0.0000                            | Shrunk to zero    | 12.0000                      |

# Model performance metrics

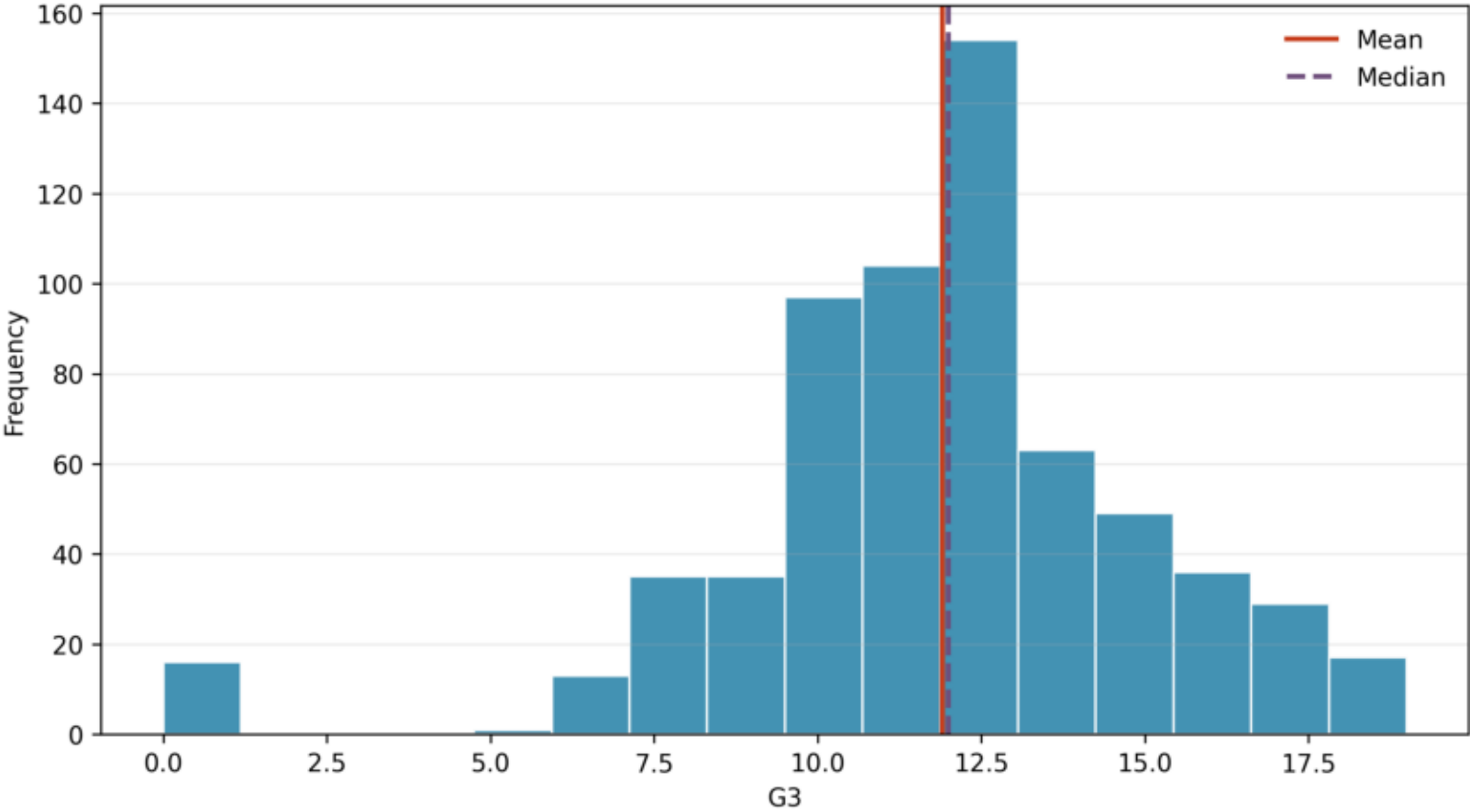
| model            | split | n        | rmse   | mae    | r_squared | selected_predictors | alpha_lambda |
|------------------|-------|----------|--------|--------|-----------|---------------------|--------------|
| Lasso Regression | Train | 486.0000 | 1.2852 | 0.8074 | 0.8425    | 10.0000             | 0.0169       |
| Lasso Regression | Test  | 163.0000 | 1.1156 | 0.7253 | 0.8765    | 10.0000             | 0.0169       |
| Ridge comparison | Train | 486.0000 | 1.2843 | 0.8128 | 0.8427    | 12.0000             | 1.1237       |
| Ridge comparison | Test  | 163.0000 | 1.1277 | 0.7337 | 0.8738    | 12.0000             | 1.1237       |
| OLS comparison   | Train | 486.0000 | 1.2842 | 0.8135 | 0.8427    | 12.0000             |              |
| OLS comparison   | Test  | 163.0000 | 1.1278 | 0.7344 | 0.8738    | 12.0000             |              |

# Cross-validation path

| alpha_lambda | mean_cv_mse | std_cv_mse | se_cv_mse | selected_alpha |
|--------------|-------------|------------|-----------|----------------|
| 15.8489      | 10.5067     | 2.0613     | 0.9219    |                |
| 14.3319      | 10.5067     | 2.0613     | 0.9219    |                |
| 12.9600      | 10.5067     | 2.0613     | 0.9219    |                |
| 11.7195      | 10.5067     | 2.0613     | 0.9219    |                |
| 10.5977      | 10.5067     | 2.0613     | 0.9219    |                |
| 9.5832       | 10.5067     | 2.0613     | 0.9219    |                |
| 8.6659       | 10.5067     | 2.0613     | 0.9219    |                |
| 7.8364       | 10.5067     | 2.0613     | 0.9219    |                |
| 7.0863       | 10.5067     | 2.0613     | 0.9219    |                |
| 6.4080       | 10.5067     | 2.0613     | 0.9219    |                |
| 5.7946       | 10.5067     | 2.0613     | 0.9219    |                |
| 5.2400       | 10.5067     | 2.0613     | 0.9219    |                |

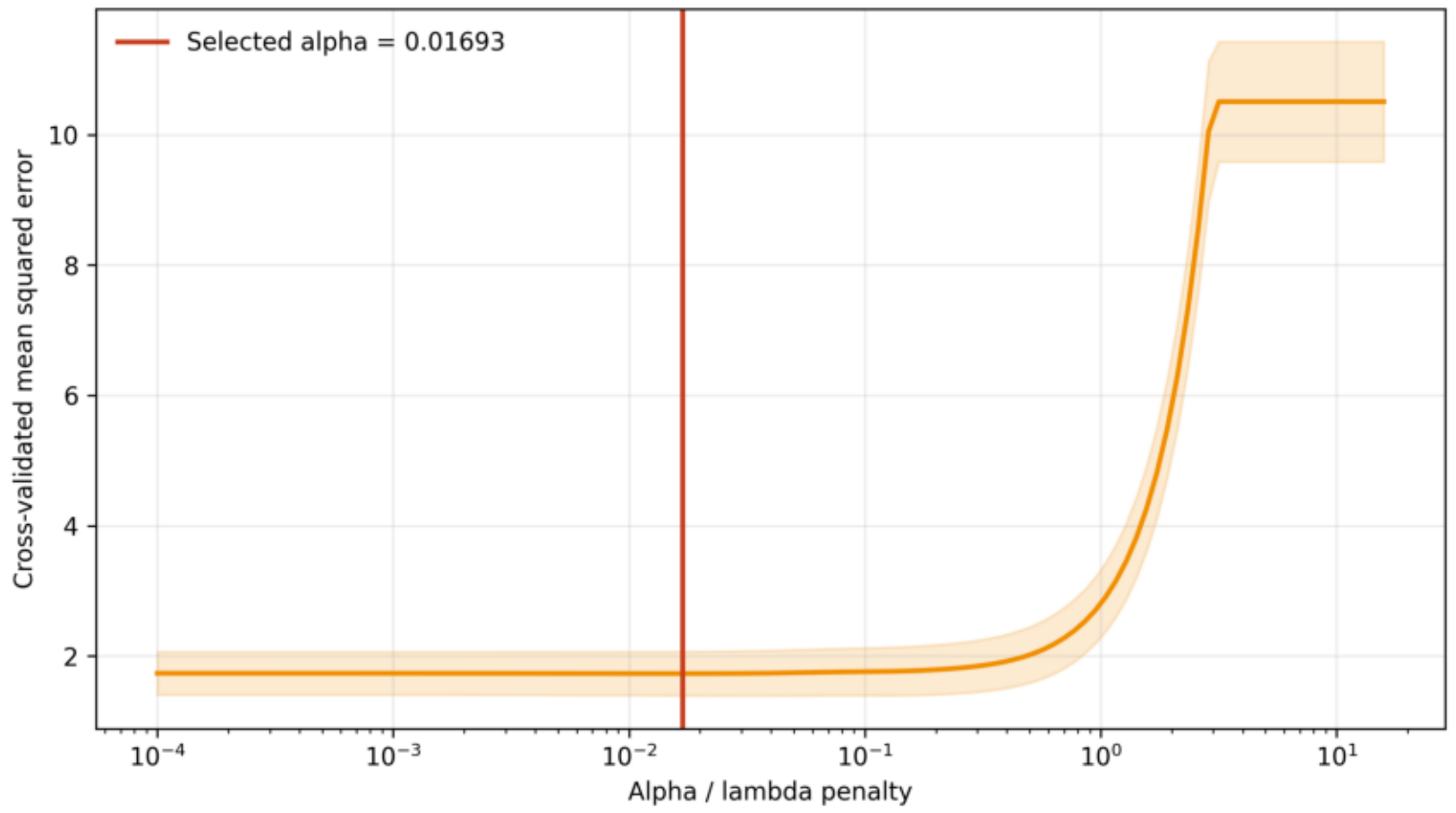
# Outcome distribution for Lasso Regression

The default continuous outcome is the final grade variable G3 when available.



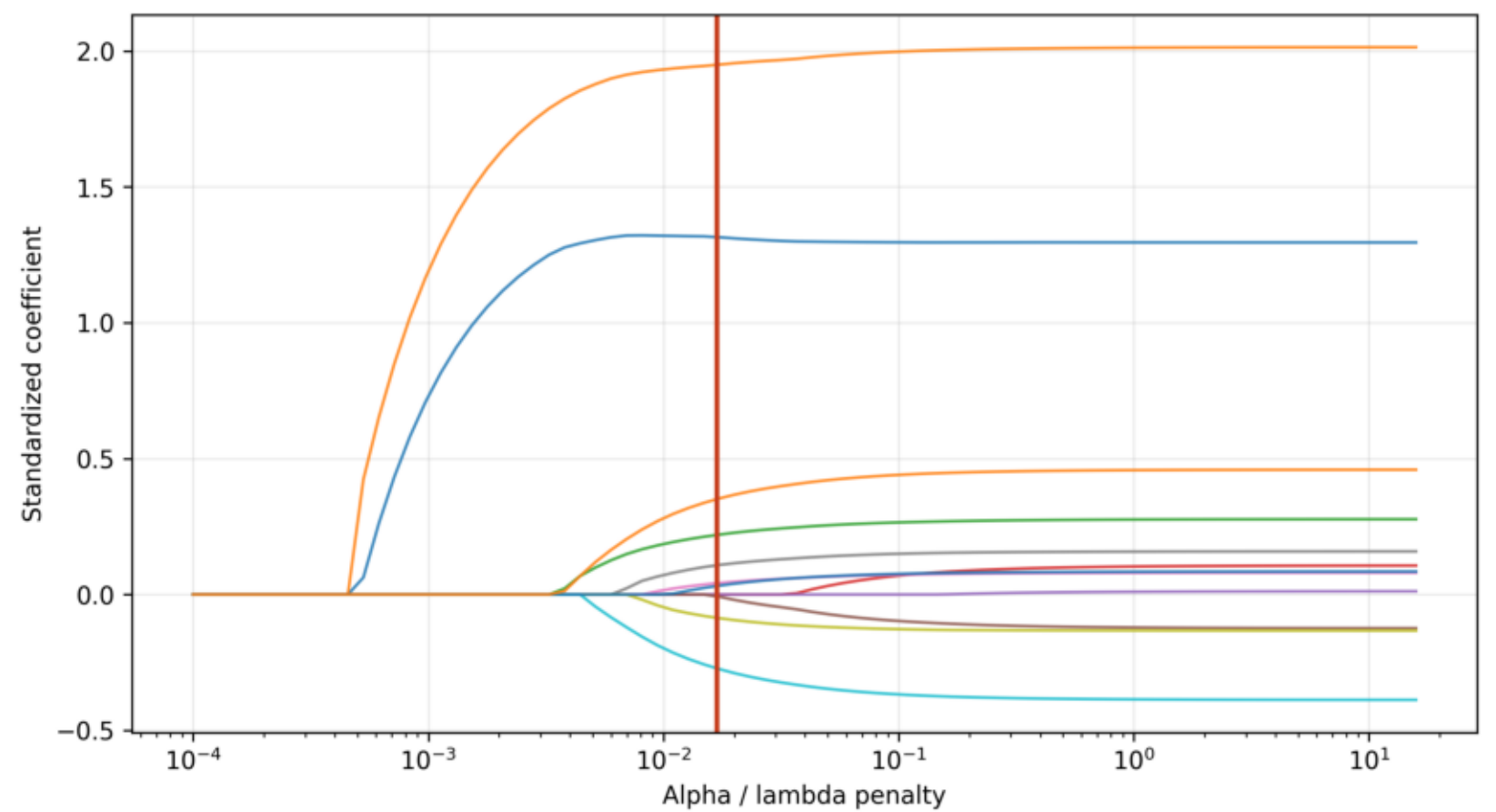
# Cross-validation curve for Lasso alpha

The selected alpha minimizes cross-validated prediction error.



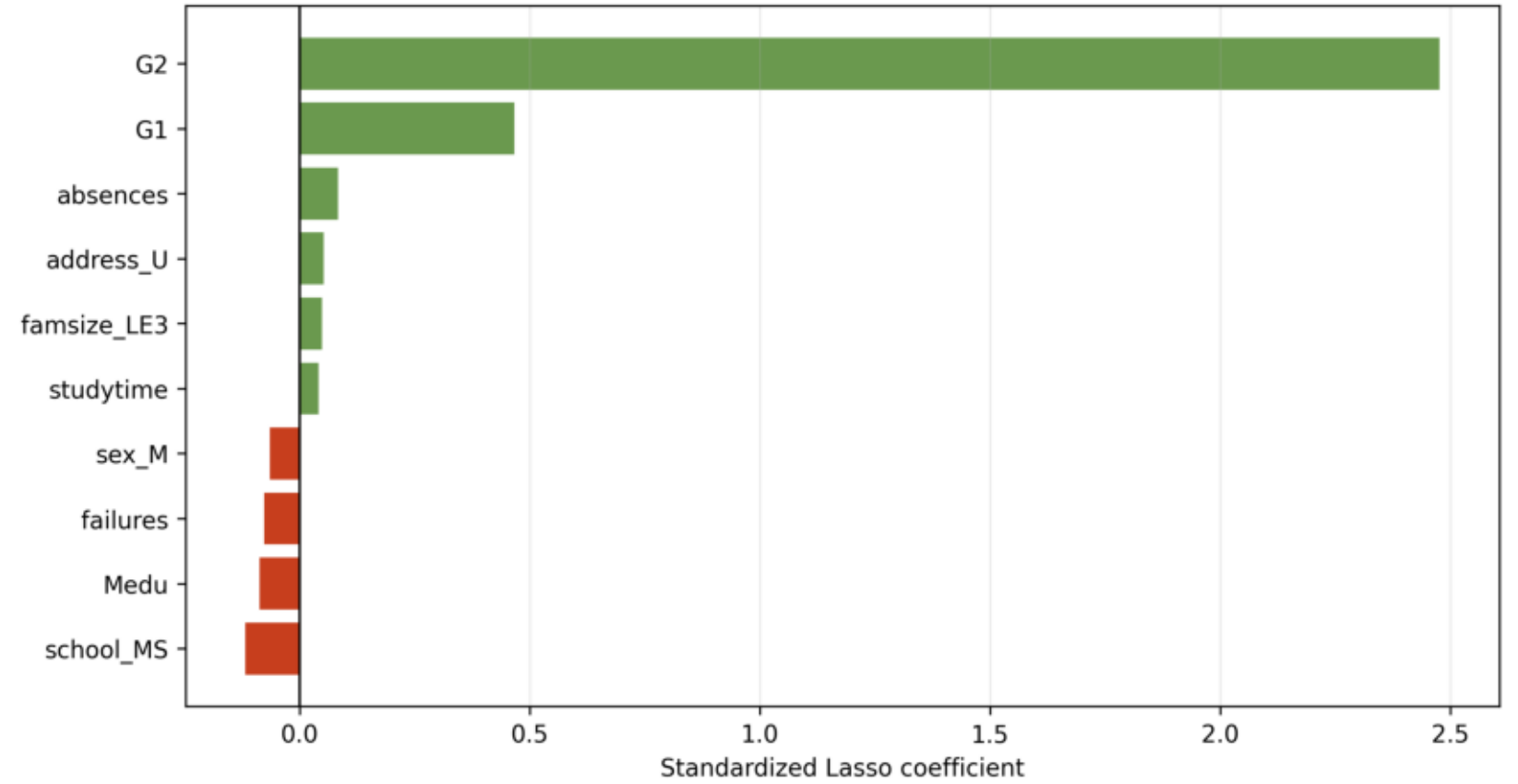
# Lasso coefficient path

As the L1 penalty increases, weak coefficients are shrunk toward zero.



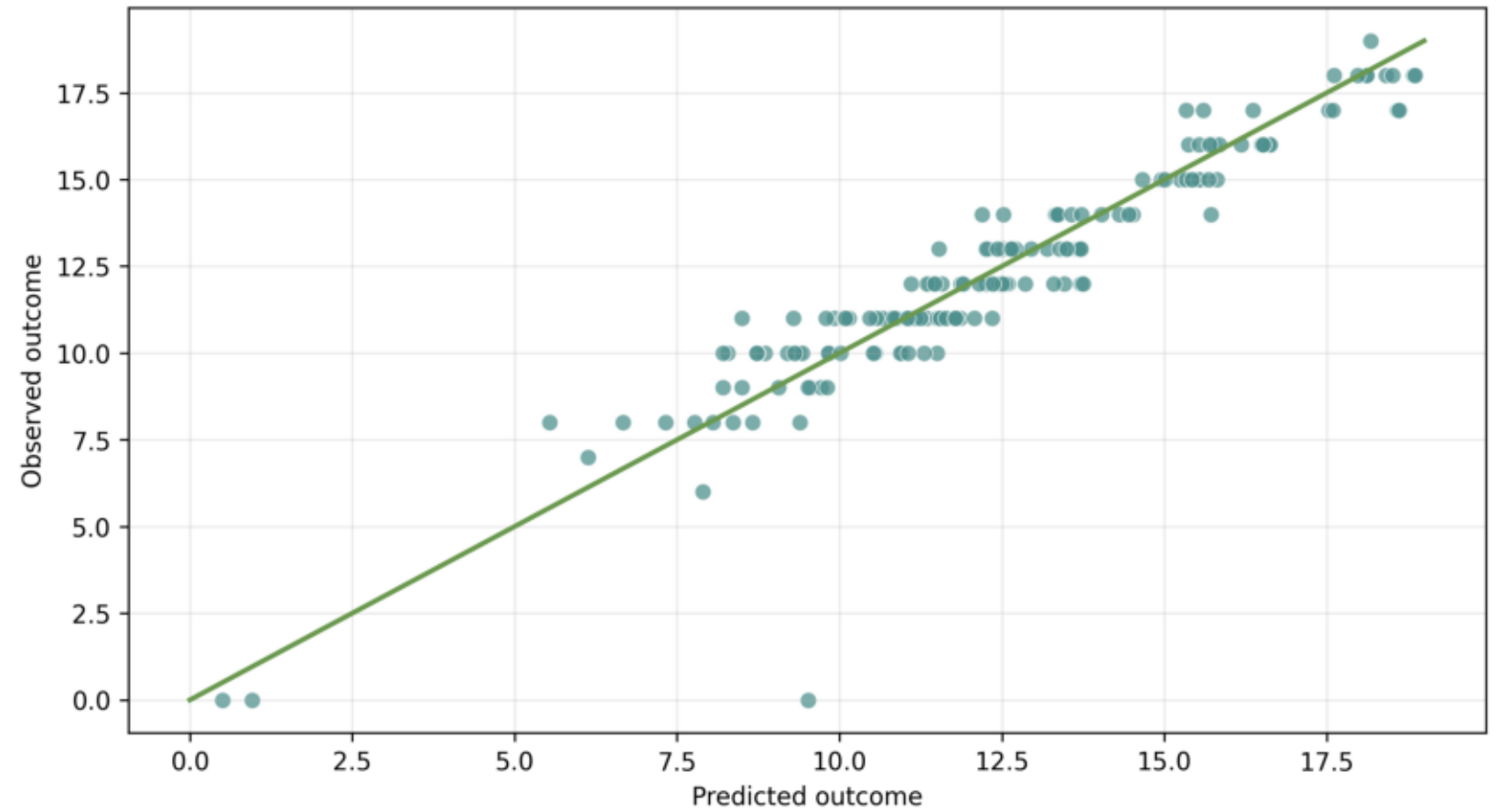
# Predictors selected by Lasso

Non-zero coefficients are the predictors retained by L1 regularization.



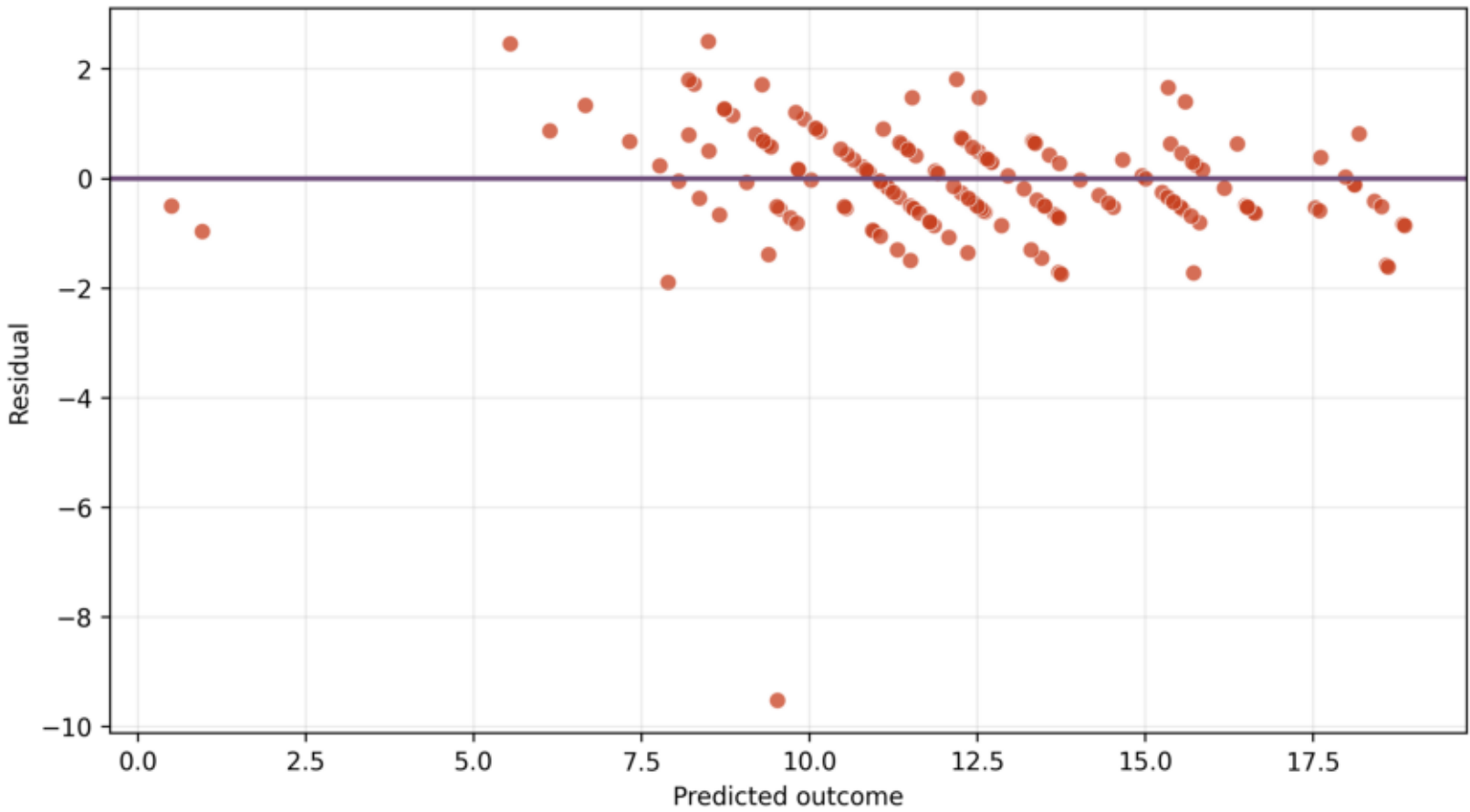
# Observed vs predicted values

Points near the diagonal show stronger prediction accuracy on the held-out test data.



# Residuals vs predicted values

Residuals should be centered around zero without a strong curved pattern.



# Prediction-error comparison

Lower test RMSE indicates better out-of-sample prediction accuracy.

