

Principal Component Regression

Test name: Principal Component Regression

Dataset: D:/DATA ANALYSIS/H Regression Tests and Models/Principal Component Regression/dataset.csv

Rows in original dataset: 649

Columns in original dataset: 33

Rows used in PCR model: 649

Outcome variable: G3

Predictor columns after dummy coding: 12

Selected principal components: 8

Component selection rule: lowest cross-validated RMSE among available components

Explained variance retained: 0.8655

Training R-squared: 0.8122

Training adjusted R-squared: 0.8099

Training RMSE: 1.3989

Training MAE: 0.9058

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

Categorical predictors: school, sex, address, famsize

Formula: $G3 \sim \text{principal components of } G1 + G2 + \text{studytime} + \text{failures} + \text{absences} + \text{age} + \text{Medu} + \text{Fedu} + \text{schoolMS} + \text{sexM} + \text{addressU} + \text{famsizeLE3}$

Interpretation: PCA creates orthogonal components from standardized predictors; linear regression is fitted to those selected components.

Reporting guide: state selected components, retained variance, model fit, reconstructed coefficients, and residual diagnostics.

Charts generated: 10

PCR model fit

test_name	outcome	n	predictor_columns_after_dummies	selected_components
Principal Component Regression	G3	649	12	8
selection_rule variance_rule_suggested_components				
lowest cross-validated RMSE among available components				8
explained_variance_selected_components	r_squared_training	adjusted_r_squared_training	rmse_training	mae_training
0.8654798	0.8122032	0.8098557	1.398943	0.9057747
sse	mse	intercept_original_scale		
1270.12	1.984562	-3.020182		

Component selection

components	cv_rmse	cv_mae	cv_r_squared	cumulative_explained_variance
1	2.065414	1.5197495	0.5906427	0.2427676
2	1.883615	1.3439357	0.6595350	0.3759206
3	1.814780	1.3121234	0.6839641	0.4815545
4	1.788135	1.3049951	0.6931763	0.5751919
5	1.760295	1.2552064	0.7026562	0.6567901
6	1.655255	1.1513348	0.7370835	0.7304653
7	1.426850	0.9277601	0.8046358	0.8027810
8	1.419374	0.9231348	0.8066776	0.8654798

Component variance

component	component_number	explained_variance_ratio	cumulative_explained_variance	selected_in_final_model
PC1	1	0.24276755	0.2427676	TRUE
PC2	2	0.13315305	0.3759206	TRUE
PC3	3	0.10563392	0.4815545	TRUE
PC4	4	0.09363736	0.5751919	TRUE
PC5	5	0.08159825	0.6567901	TRUE
PC6	6	0.07367520	0.7304653	TRUE
PC7	7	0.07231564	0.8027810	TRUE
PC8	8	0.06269879	0.8654798	TRUE

Original-scale coefficients

term	original_scale_coefficient	standardized_predictor_coefficient
Intercept	-3.020181563	NA
G1	0.494354416	1.35713394
G2	0.484716961	1.41229008
studytime	-0.069431819	-0.05759436
failures	-0.707365129	-0.41963386
absences	-0.005445754	-0.02527243
age	0.262101509	0.31927571
Medu	0.015300050	0.01735870
Fedu	-0.028178784	-0.03099472
schoolMS	-0.573306728	-0.27333899
sexM	-0.275694155	-0.13569315
addressU	-0.095369013	-0.04388335

interpretation

Predicted outcome when all original predictors are zero

Approximate PCR coefficient reconstructed from selected principal components

Approximate PCR coefficient reconstructed from selected principal components

Approximate PCR coefficient reconstructed from selected principal components

Approximate PCR coefficient reconstructed from selected principal components

Approximate PCR coefficient reconstructed from selected principal components

Approximate PCR coefficient reconstructed from selected principal components

Approximate PCR coefficient reconstructed from selected principal components

Approximate PCR coefficient reconstructed from selected principal components

Approximate PCR coefficient reconstructed from selected principal components

Approximate PCR coefficient reconstructed from selected principal components

Approximate PCR coefficient reconstructed from selected principal components



















