

Weighted Least Squares Regression

Test name: Weighted Least Squares Regression

Dataset: D:/DATA ANALYSIS/H Regression Tests and Models/Weighted Least Squares Regression/dataset.csv

Rows in original dataset: 649

Columns in original dataset: 33

Rows used in WLS model: 649

Outcome variable: G3

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

Categorical predictors: school, sex, address

Formula: $G3 \sim G1 + G2 + \text{studytime} + \text{failures} + \text{absences} + \text{age} + \text{school} + \text{sex} + \text{address}$

OLS baseline model fitted first.

Weights created as inverse estimated residual variance from OLS residuals.

Mean scaled WLS weight: 1

Minimum scaled WLS weight: 0.5477

Maximum scaled WLS weight: 1.5136

Manual Breusch–Pagan style $p < 0.05$ suggests heteroscedasticity, so WLS is useful to consider.

OLS adjusted R-squared: 0.8501

WLS adjusted R-squared: 0.862

Interpretation: WLS coefficients estimate the expected outcome after giving more influence to observations with lower estimated error variance.

Decision guide: compare OLS and WLS coefficients, residual plots, and heteroscedasticity diagnostics before reporting the weighted model.

Reporting guide: state outcome, predictors, weight-construction method, coefficients, 95% CIs, p-values, fit statistics, and diagnostic conclusion.

Charts generated: 7

Weighted Least Squares coefficients

model	term	coefficient	standard_error
Weighted Least Squares	(Intercept)	-0.36602869	0.71775561
Weighted Least Squares	G1	0.14187344	0.03560749
Weighted Least Squares	G2	0.85608955	0.03404758
Weighted Least Squares	studytime	0.05059617	0.05814358
Weighted Least Squares	failures	-0.27662126	0.09680590
Weighted Least Squares	absences	0.01505165	0.01044721
Weighted Least Squares	age	0.04361666	0.04082367
Weighted Least Squares	schoolMS	-0.15184518	0.10992918
Weighted Least Squares	sexM	-0.16796462	0.09623508
Weighted Least Squares	addressU	0.07592353	0.10754779

t_value	p_value	ci95_lower	ci95_upper
-0.5099628	6.102536e-01	-1.775473466	1.04341608
3.9843706	7.545174e-05	0.071951600	0.21179527
25.1439214	1.619461e-97	0.789230896	0.92294821
0.8701937	3.845213e-01	-0.063579405	0.16477175
-2.8574835	4.409112e-03	-0.466717391	-0.08652512
1.4407333	1.501498e-01	-0.005463371	0.03556667
1.0684160	2.857364e-01	-0.036548098	0.12378141
-1.3813001	1.676695e-01	-0.367711277	0.06402092
-1.7453575	8.140332e-02	-0.356939855	0.02101061
0.7059515	4.804756e-01	-0.135266281	0.28711334

decision_alpha_0_05

Not statistically significant

Statistically significant

Statistically significant

Not statistically significant

Statistically significant

Not statistically significant

Not statistically significant

OLS vs WLS fit statistics

	model	n	df_residual	r_squared	adj_r_squared	
Ordinary Least Squares	649		639	0.8522223	0.8501409	
Weighted Least Squares	649		639	0.8639355	0.8620191	
	sigma	aic	bic	rmse	weighted_rmse	f_statistic
	1.250640	2144.009	2193.239	1.240967		NA
	409.4514					
	1.157174	2054.904	2104.134	1.242420	1.148224	450.8113
	f_p_value					
	1.440344e-258					
	5.255230e-270					

WLS diagnostics

	diagnostic statistic
Manual Breusch-Pagan style test using fitted values	
	19.224437
	OLS residual SD 1.241925
	WLS residual SD 1.243375
p_value	
1.162162e-05	
NA	
NA	
	interpretation
p < 0.05 suggests heteroscedasticity and supports WLS consideration.	
	Baseline unweighted residual spread.
	Residual spread after inverse-variance weighting.













