

Regression Assumptions

Test name: Regression Assumptions

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

Rows in original dataset: 649

Rows used in the model: 649

Outcome: G3

Predictors: G1, G2, studytime, failures, absences, age

Formula: $G3 \sim G1 + G2 + studytime + failures + absences + age$

R-squared: 0.850632

Adjusted R-squared: 0.849236

RMSE: 1.254409

Overall F: 609.352742

Overall p-value: 3.25413e-261

Durbin-Watson: 1.860956

Maximum VIF: 4.154857

Maximum Cook's distance: 0.176952

Cases flagged by at least one influence rule: 58

Assumption decisions:

Ramsey RESET F test: Potential violation

Shapiro-Wilk test: Potential violation

Jarque-Bera test: Potential violation

Breusch-Pagan LM test: Potential violation

White LM test: Potential violation

Durbin-Watson statistic: Supported

Correlation of fitted values with absolute residuals: Potential violation

Maximum VIF: Supported

Scaled-predictor condition number: Supported

Charts generated: 10

Interpretation rule: use formal tests together with residual and influence plots rather than relying on one statistic alone.

Model summary

test_name	outcome	predictors	n_complete	number_of	r_squared	adjusted_r	rmse	f_value	f_p_value	aic	bic	durbin_wat	condition_n
Regression	G3	G1, G2, stu	649	6	0.85063	0.84924	1.2544	609.35	3.2541e-26	2143	2174.3	1.861	4.2502

Assumption tests

assumption	diagnostic	statistic	p_value	guideline	decision
Linearity / functional form	Ramsey RESET F test	10.774	0.0010852	p >= 0.05 supports the sp	Potential violation
Residual normality	Shapiro-Wilk test	0.7547	6.2445e-30	p >= 0.05 supports norma	Potential violation
Residual normality	Jarque-Bera test	10153	0	p >= 0.05 supports norma	Potential violation
Homoscedasticity	Breusch-Pagan LM test	33.054	1.0237e-05	p >= 0.05 supports const	Potential violation
Homoscedasticity	White LM test	75.547	1.7529e-06	p >= 0.05 supports const	Potential violation
Independence of errors	Durbin-Watson statistic	1.861		Values near 2 support inde	Supported
Residual variance pattern	Correlation of fitted values	-0.20964	7.0257e-08	Small non-significant corre	Potential violation
Multicollinearity	Maximum VIF	4.1549		VIF below 5 is generally ac	Supported
Multicollinearity	Scaled-predictor condition	4.2502		Values below about 30 are	Supported

Multicollinearity diagnostics

predictor	vif	tolerance	vif_decision
G1	4.1549	0.24068	Acceptable
G2	4.0914	0.24442	Acceptable
studytime	1.0881	0.91907	Acceptable
failures	1.3061	0.76564	Acceptable
absences	1.0486	0.95363	Acceptable
age	1.1549	0.8659	Acceptable

Influence summary

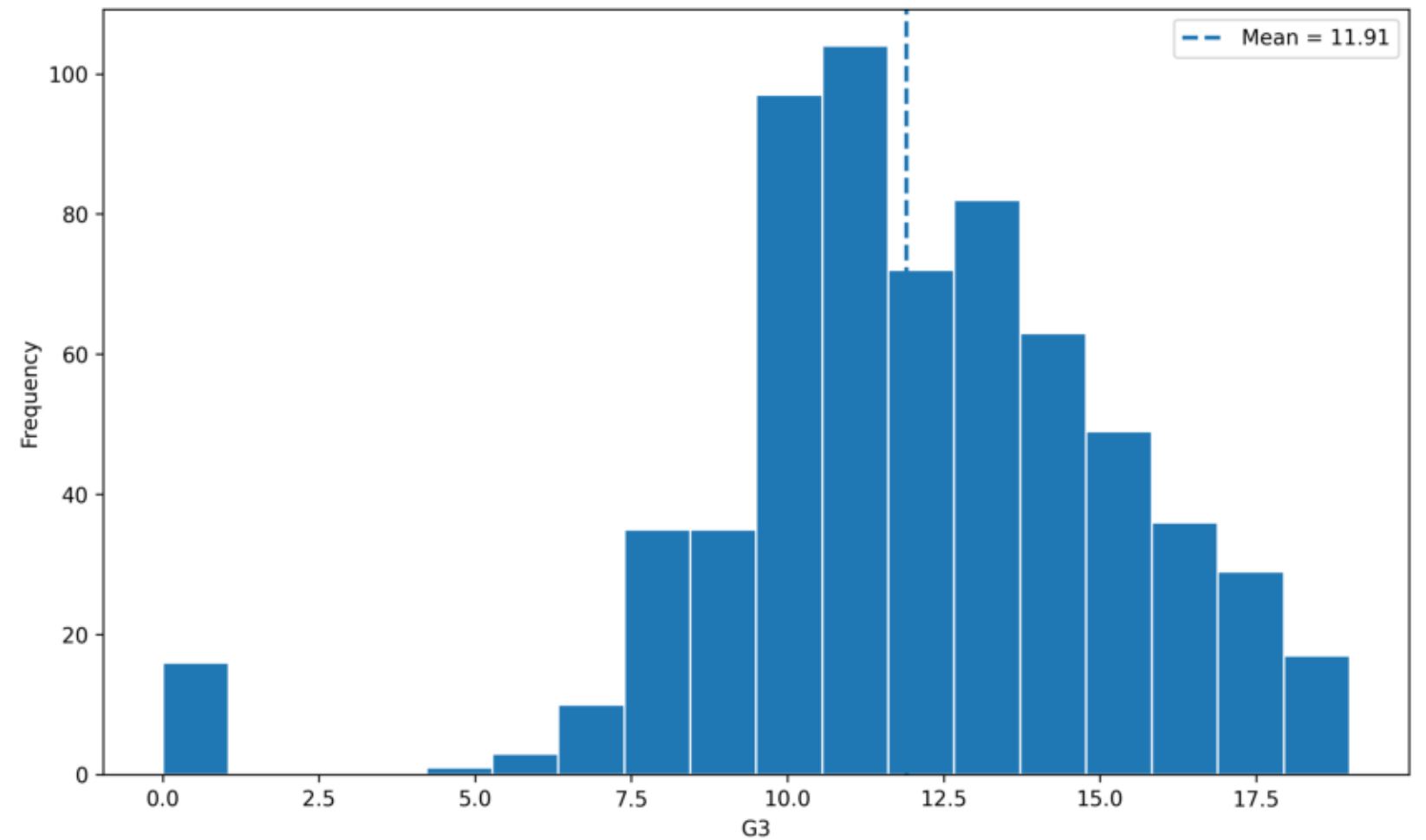
n	predictors	studentized	leverage_th	cooks_dist	dffits_thres	residual_out	high_lever	high_cooks	high_dffits	any_influen	maximum_	maximum_	maximum_ab
649	6	3	0.021572	0.0061633	0.20771	10	45	26	26	58	0.17695	0.10331	7.5907

Coefficients

term	coefficient	standard_error	t_value	p_value	ci95_lower	ci95_upper	decision_alpha_0_05
const	-0.54531	0.76264	-0.71504	0.47484	-2.0429	0.95225	Not statistically significant
G1	0.14217	0.036589	3.8856	0.00011267	0.070321	0.21402	Statistically significant
G2	0.8834	0.03421	25.823	2.2065e-101	0.81622	0.95058	Statistically significant
studytime	0.094579	0.061966	1.5263	0.12743	-0.027102	0.21626	Not statistically significant
failures	-0.23316	0.094932	-2.456	0.014312	-0.41957	-0.046742	Statistically significant
absences	0.022679	0.010874	2.0857	0.037398	0.0013272	0.044031	Statistically significant
age	0.023641	0.043473	0.54381	0.58676	-0.061726	0.10901	Not statistically significant

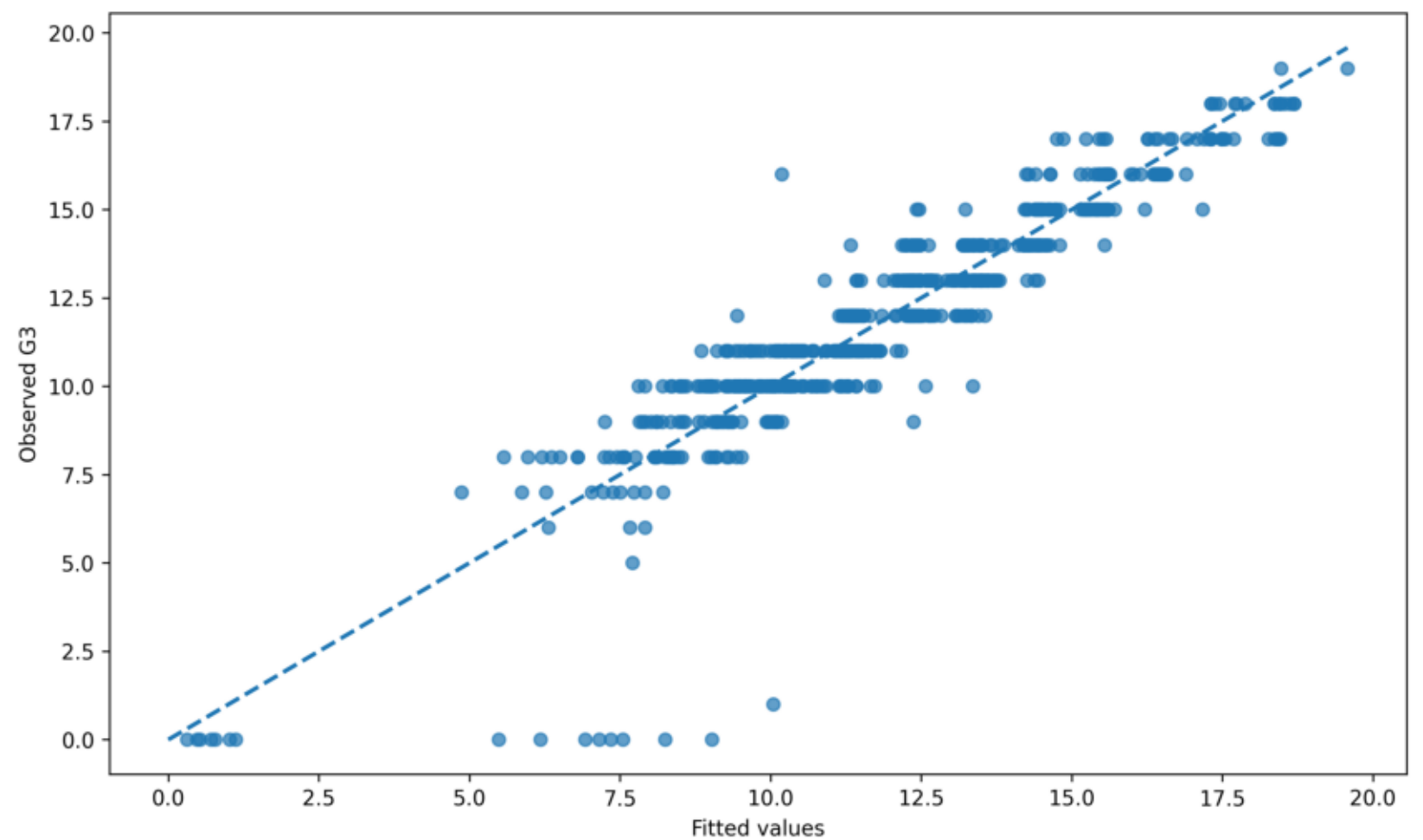
Outcome distribution

Inspect the dependent variable before evaluating residual assumptions.



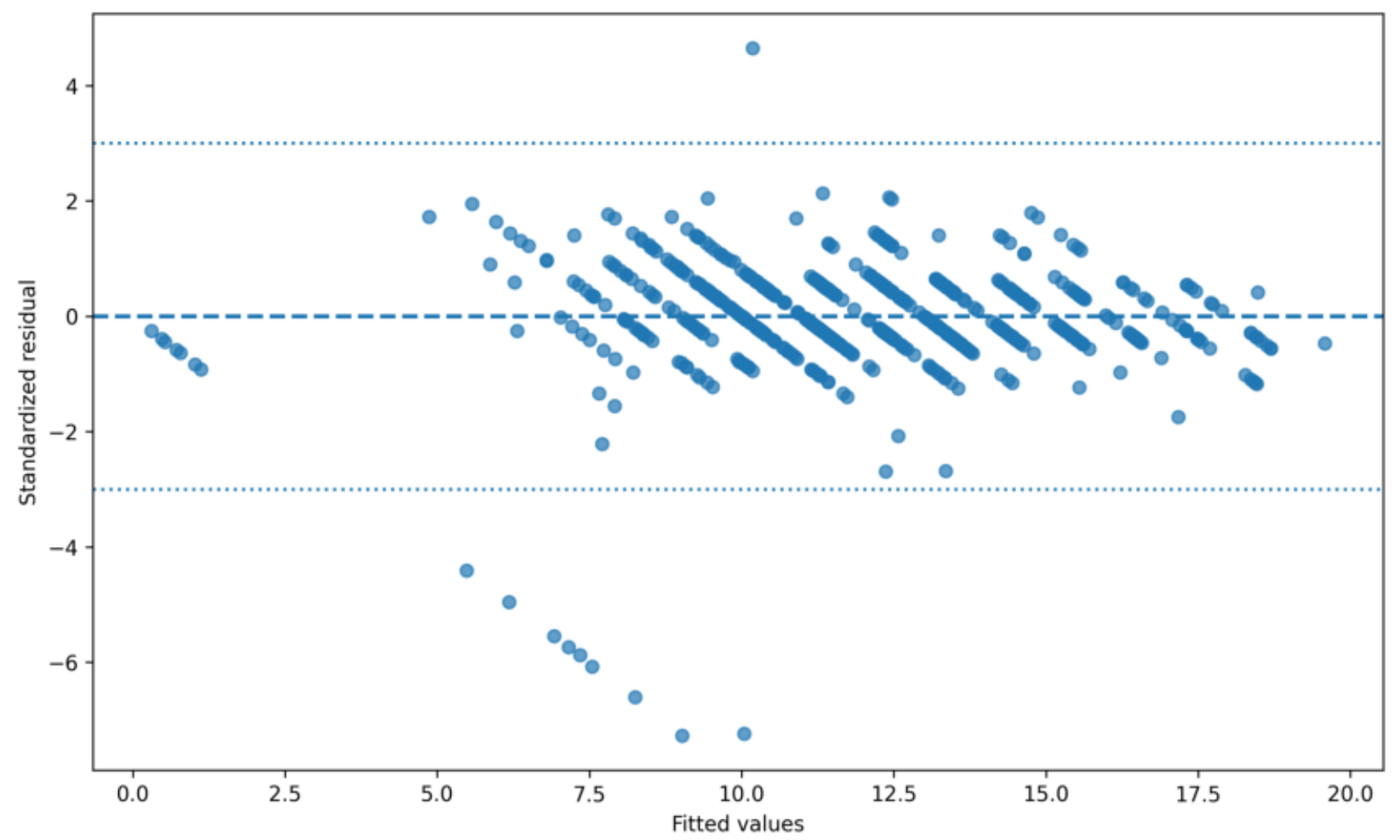
Observed versus fitted values

Points close to the diagonal indicate stronger model agreement.



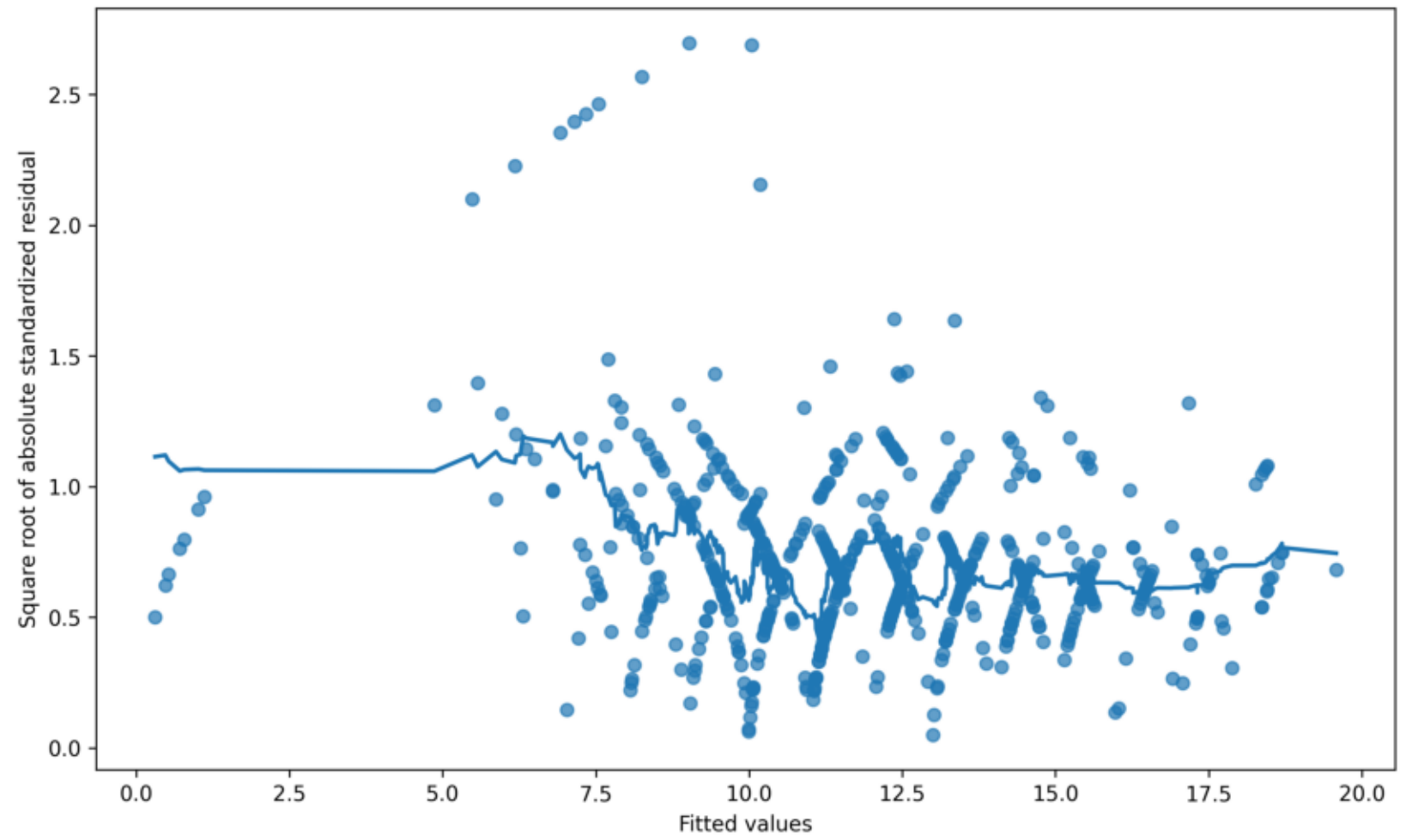
Residuals versus fitted values

A pattern-free band around zero supports linearity and constant variance.



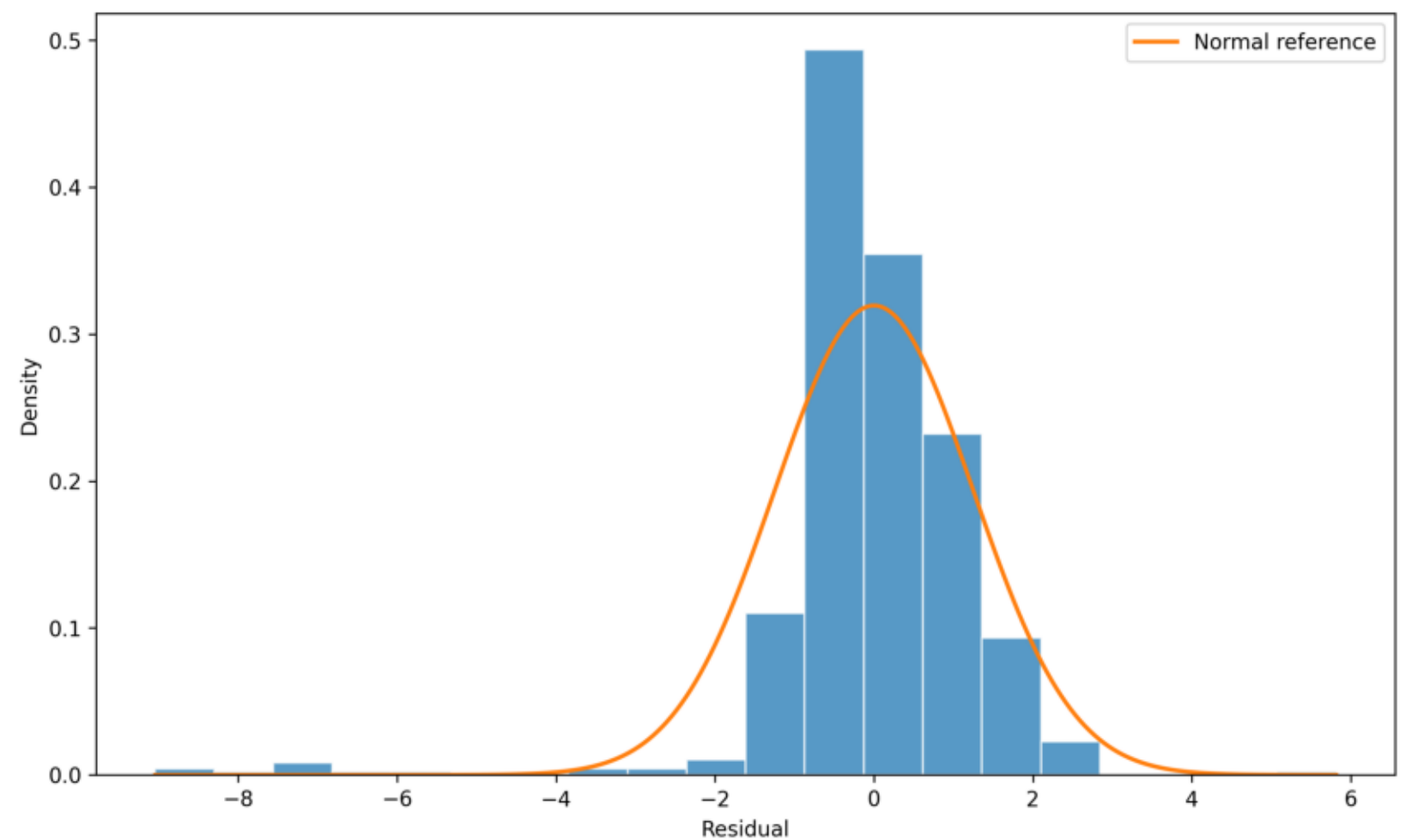
Scale-location plot

A roughly horizontal spread supports homoscedasticity.



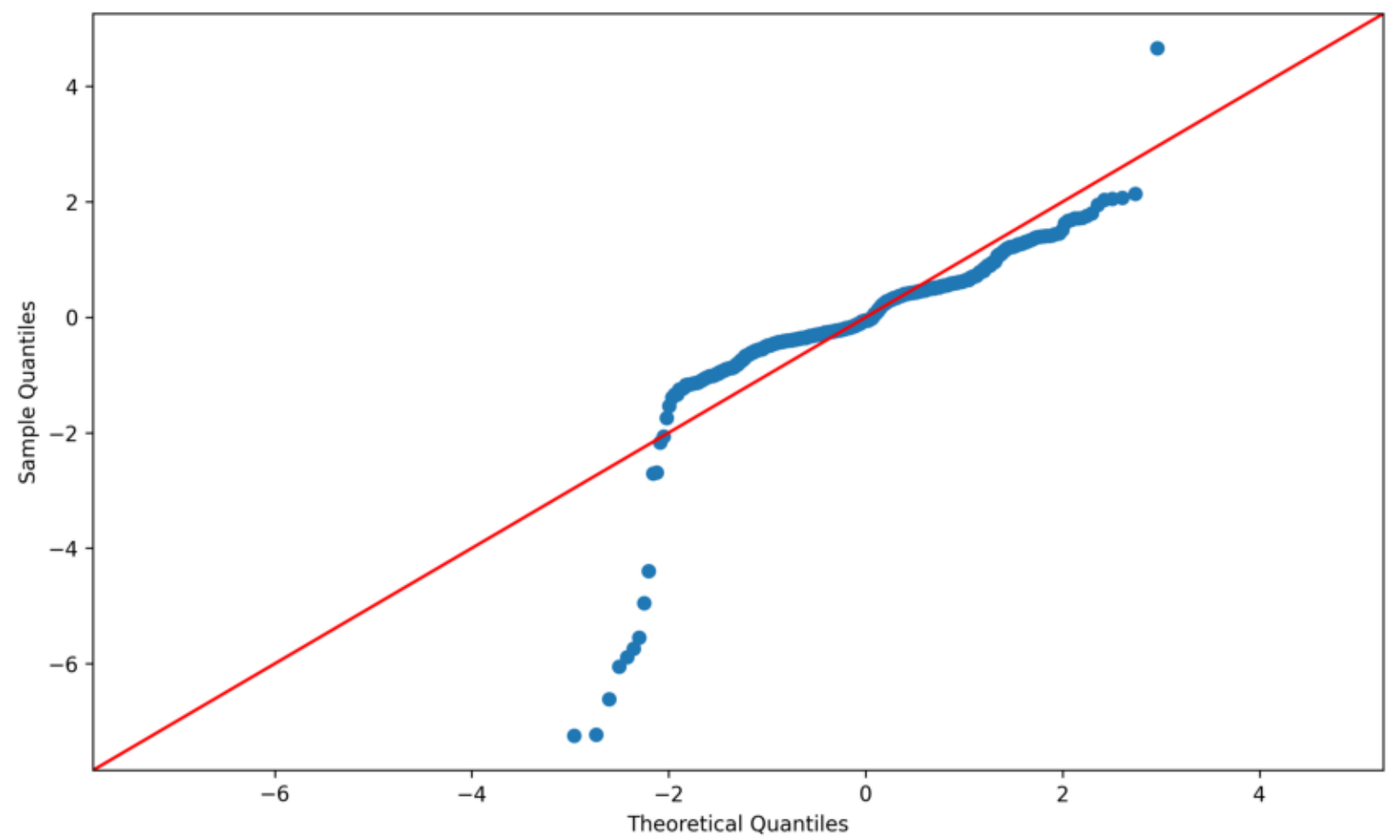
Residual distribution

Approximate symmetry and agreement with the reference curve support normality.



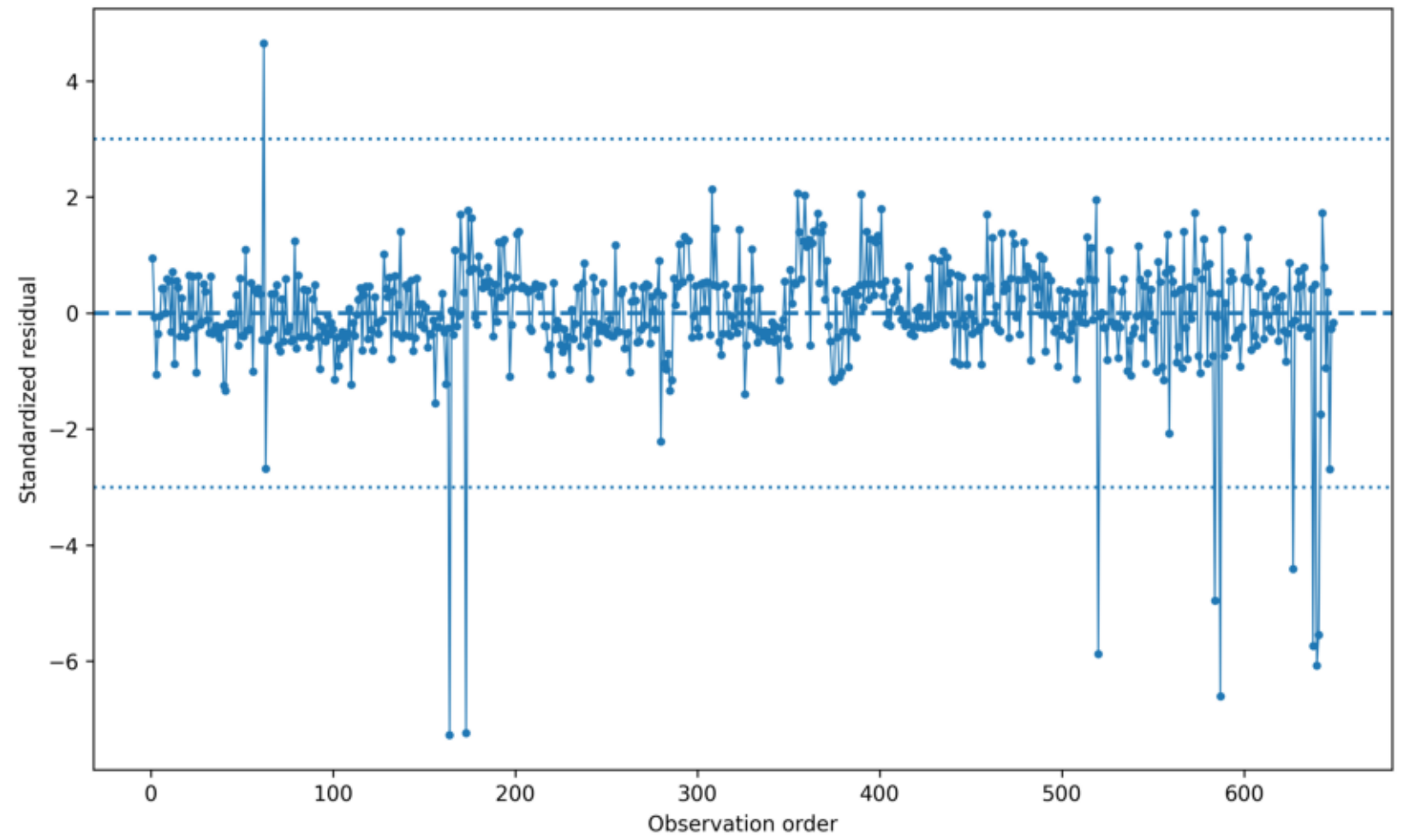
Normal Q-Q plot of residuals

Points close to the reference line support approximate residual normality.



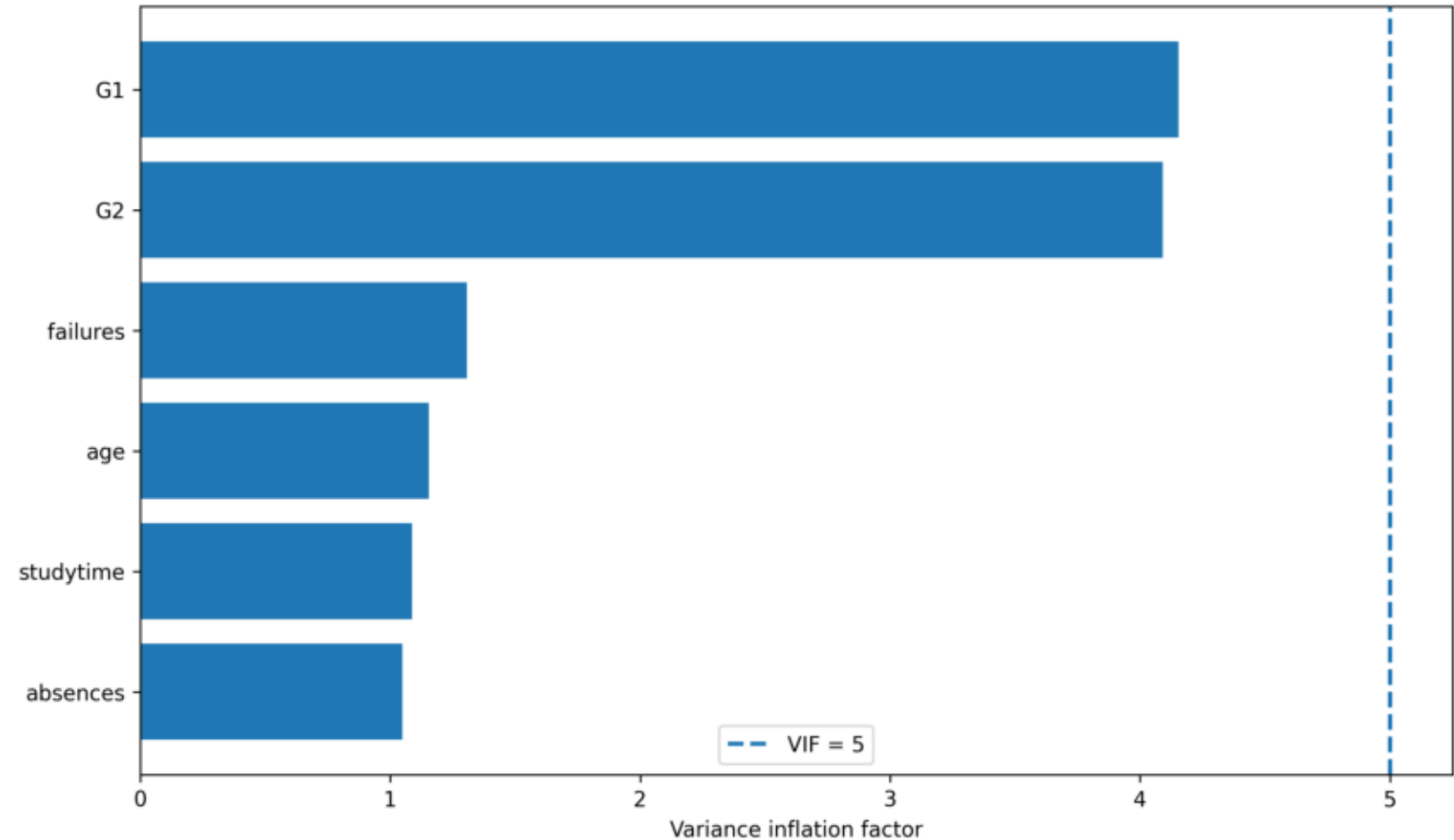
Residual sequence

Long runs or cycles can signal dependence or omitted structure.



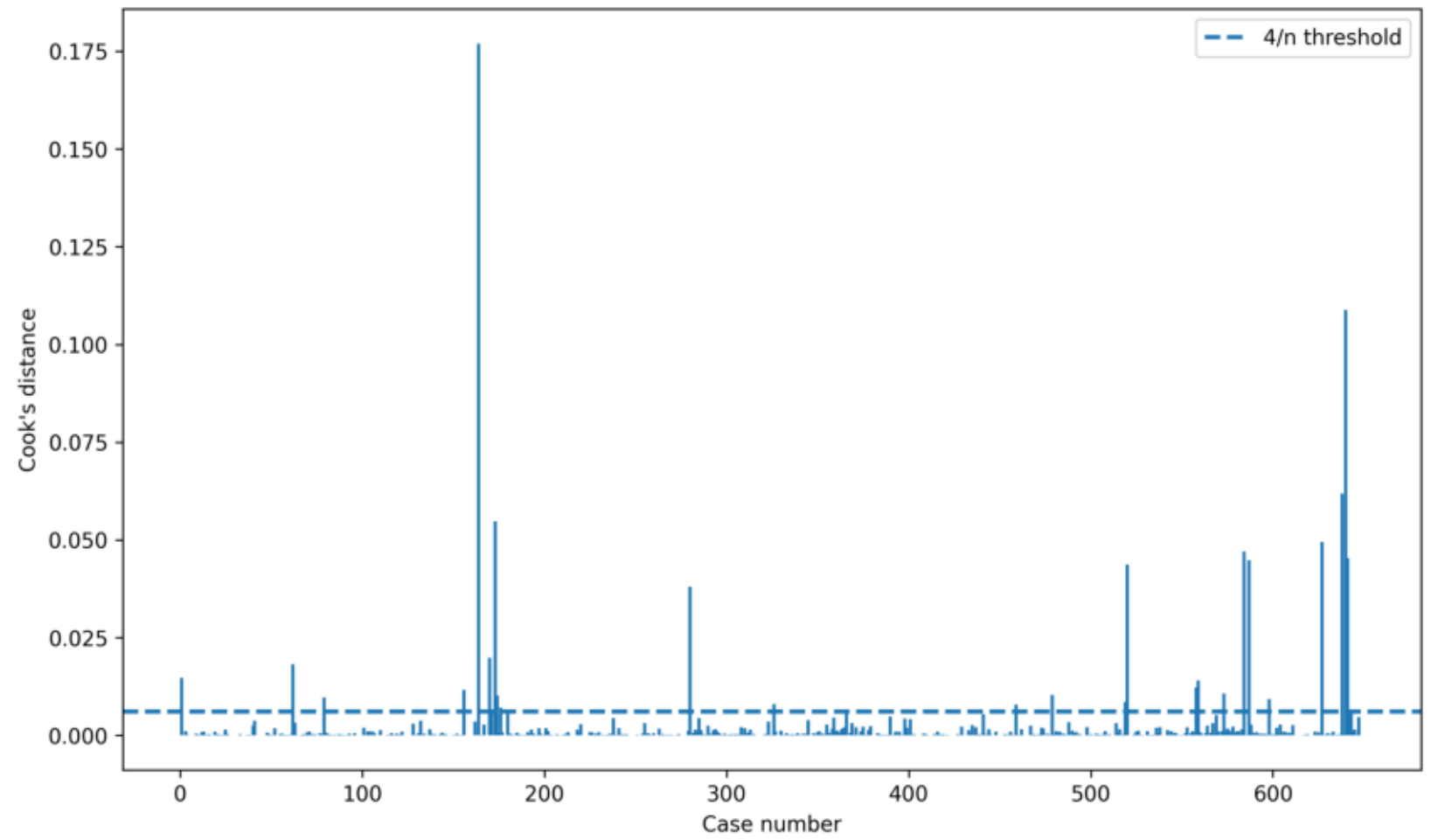
Multicollinearity diagnostics

VIF values below 5 are generally acceptable for routine regression work.



Cook's distance

Cases above the reference threshold deserve influence review.



Leverage and studentized residuals

Larger points have greater Cook's distance and may be influential.

