

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 model: 649

Outcome: G3

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

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

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

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_cases	number_of_predictors
Regression Assumptions		G3 G1, G2, studytime, failures, absences, age			649	6
r_squared	adjusted_r_squared	rmse	f_value	f_p_value	aic	bic durbin_watson
0.8506322	0.8492363	1.254409	609.3527	3.254128e-261	2144.955	2180.758 1.860956
condition_number_scaled_predictors						
		4.25023				

Assumption tests

assumption	diagnostic	statistic	p_value
Linearity / functional form	Ramsey RESET F test	6.0224690	2.563202e-03
Residual normality	Shapiro-Wilk test	0.7546974	6.244476e-30
Residual normality	Jarque-Bera test	10310.3963997	0.000000e+00
Homoscedasticity	Breusch-Pagan LM test	33.0541920	1.023705e-05
Independence of errors	Durbin-Watson statistic	1.8609563	NA
Residual variance pattern	Correlation of fitted values with absolute residuals	-0.2096445	7.025654e-08
Multicollinearity	Maximum VIF	4.1548567	NA
Multicollinearity	Scaled-predictor condition number	4.2502299	NA
guideline		decision	
p >= 0.05 supports the specified linear form		Potential violation	
p >= 0.05 supports normal residuals		Potential violation	
p >= 0.05 supports normal residuals		Potential violation	
p >= 0.05 supports constant residual variance		Potential violation	
Values near 2 support independence		Supported	
Small non-significant correlation is desirable		Potential violation	
VIF below 5 is generally acceptable		Supported	
Values below about 30 are generally acceptable		Supported	

Multicollinearity diagnostics

predictor	vif	tolerance	vif_decision
G1	4.154857	0.2406822	Acceptable
G2	4.091388	0.2444158	Acceptable
studytime	1.088056	0.9190701	Acceptable
failures	1.306096	0.7656406	Acceptable
absences	1.048621	0.9536337	Acceptable
age	1.154871	0.8658977	Acceptable

Influence summary

n predictors studentized_residual_threshold leverage_threshold cooks_distance_threshold dffits_threshold					
649	6	3	0.02157165	0.006163328	0.2077096
residual_outlier_count high_leverage_count high_cooks_distance_count high_dffits_count any_influence_flag_count					
	10	45	26	26	58
maximum_cooks_distance maximum_leverage maximum_absolute_studentized_residual					
	0.1769525	0.1033085	7.590715		

Coefficients

	term	coefficient	standard_error	t_value	p_value	ci95_lower	ci95_upper
	(Intercept)	-0.54531480	0.76263548	-0.7150399	4.748443e-01	-2.042876127	0.95224653
	G1	0.14216855	0.03658857	3.8856004	1.126739e-04	0.070320825	0.21401627
	G2	0.88339876	0.03420986	25.8229266	2.206497e-101	0.816222016	0.95057550
	studytime	0.09457944	0.06196635	1.5263031	1.274269e-01	-0.027101783	0.21626066
	failures	-0.23315622	0.09493193	-2.4560358	1.431172e-02	-0.419570816	-0.04674162
	absences	0.02267927	0.01087356	2.0857261	3.739772e-02	0.001327229	0.04403132
	age	0.02364114	0.04347324	0.5438091	5.867616e-01	-0.061725777	0.10900806
	decision_alpha_0_05						
Not statistically significant							
Statistically significant							
Statistically significant							
Not statistically significant							
Statistically significant							
Statistically significant							
Not statistically significant							



















