

Simple Linear Regression

Test name: Simple Linear Regression

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

Rows in original dataset: 649

Rows used: 649

Outcome: G3

Predictor: G2

Estimated equation: $G3 = 0.121966 + (1.018490 \times G2)$

Pearson r: 0.918548

R-squared: 0.843730

Adjusted R-squared: 0.843489

F(1, 647): 3493.281566

Overall model p-value: 5.64240148959e-263

Slope p-value: 5.64240148958e-263

Slope 95% CI: [0.984653, 1.052328]

Residual standard error: 1.278096

Durbin-Watson: 1.846403

Interpretation: the slope is the expected outcome change for a one-unit increase in the predictor.

Decision rule: $p < 0.05$ and a slope confidence interval excluding zero indicate a statistically significant linear relationship.

Prediction intervals are wider than mean-response confidence intervals because they include individual outcome variation.

Charts generated: 8

Model summary

test_name	outcome	predictor	equation	complete_n	correlation	r_squared	adjusted_r_squared	f_statistic	f_p_value	rmse	mean	residual_standard	df_model	df_residual	aic	bic	decision
Linear Reg	G3	G3	G2 intercept + s	649	0.91855	0.84373	0.843493	493.2815	0.00000	1.27612	0.80802	1.27810	1.00000	647.00000	162.2786	171.1729	Significant linear

Regression coefficients

term	coefficient_b	standard_error	t_value	p_value	ci95_lower	ci95_upper	decision_alpha_0_05
Intercept	0.12197	0.20559	0.59324	0.55323	-0.28174	0.52568	Not statistically significant
Slope	1.01849	0.01723	59.10399	0.00000	0.98465	1.05233	Statistically significant

Regression ANOVA

source	sum_of_squares	df	mean_square	f_value	p_value
Regression	5706.37384	1	5706.37384	3493.28157	0.00000
Residual / Error	1056.89272	647	1.63353	nan	nan
Total	6763.26656	648	nan	nan	nan

Descriptive statistics

variable	n	mean	standard_deviation	minimum	maximum	skewness	kurtosis_excess
G3	649	11.90601	3.23066	0.00000	19.00000	-0.91291	2.71220
G2	649	11.57011	2.91364	0.00000	19.00000	-0.36028	1.66246

Pearson correlation

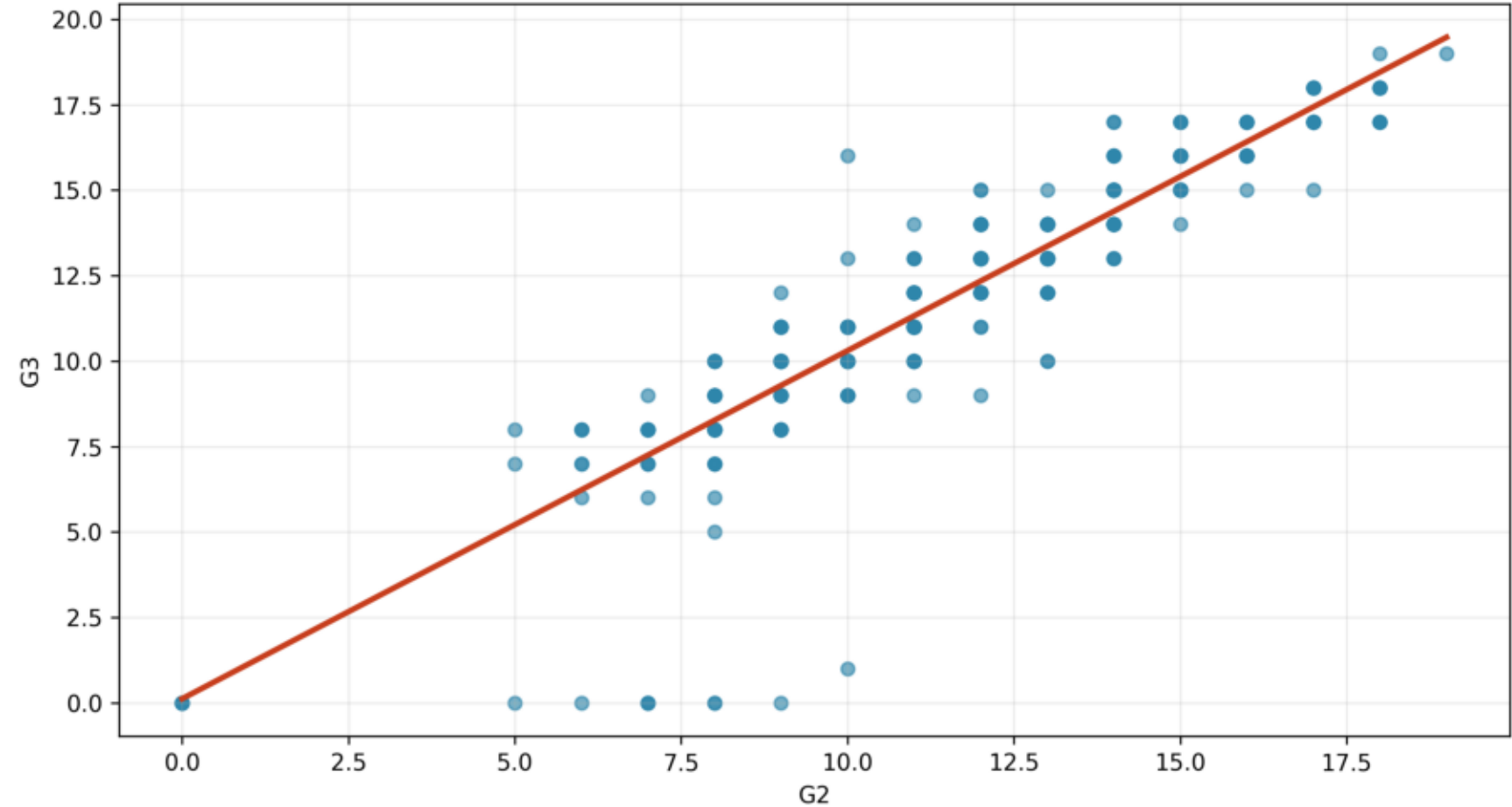
outcome	predictor	pearson_r	p_value	n
G3	G2	0.91855	0.00000	649

Diagnostic summary

diagnostic	statistic	p_value	guideline
Durbin-Watson	1.84640	nan	Approximately 2 suggests independent residuals
Jarque-Bera	10482.94671	0.00000	p >= 0.05 supports residual normality
Shapiro-Wilk	0.71085	0.00000	p >= 0.05 supports residual normality
Maximum absolute standardized residual	7.28908	nan	Values above 3 deserve review
Maximum absolute studentized residual	7.60227	nan	Values above 3 deserve review
Maximum leverage	0.02588	nan	Compare with 2p/n; p includes intercept
Maximum Cook's distance	0.08683	nan	Compare with 4/n
Residual skewness	-2.94142	nan	Values near zero indicate symmetry
Residual kurtosis	21.78965	nan	Normal-distribution kurtosis is approximately 3

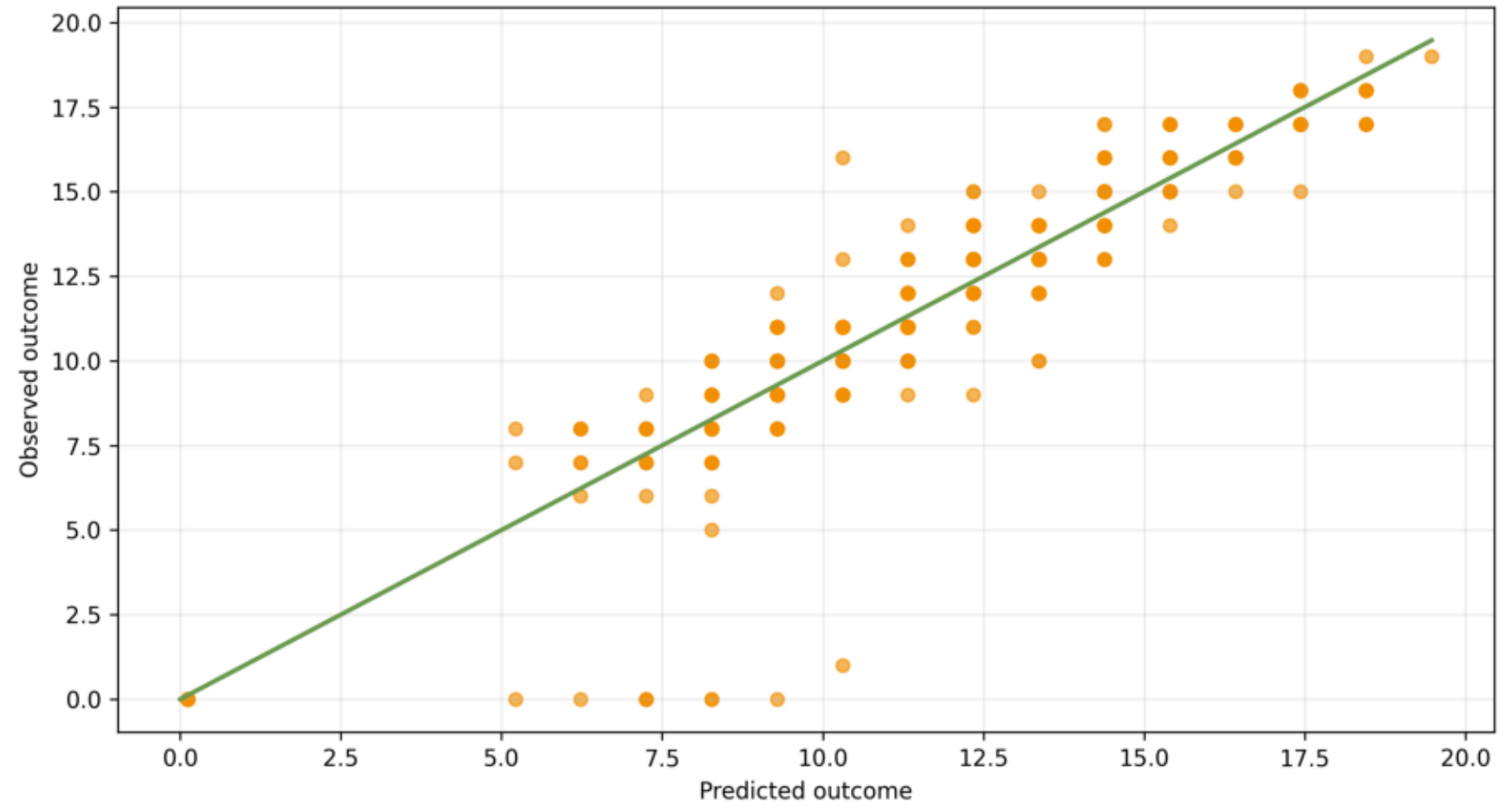
Simple linear regression fit

Observed G3 with fitted line from G2



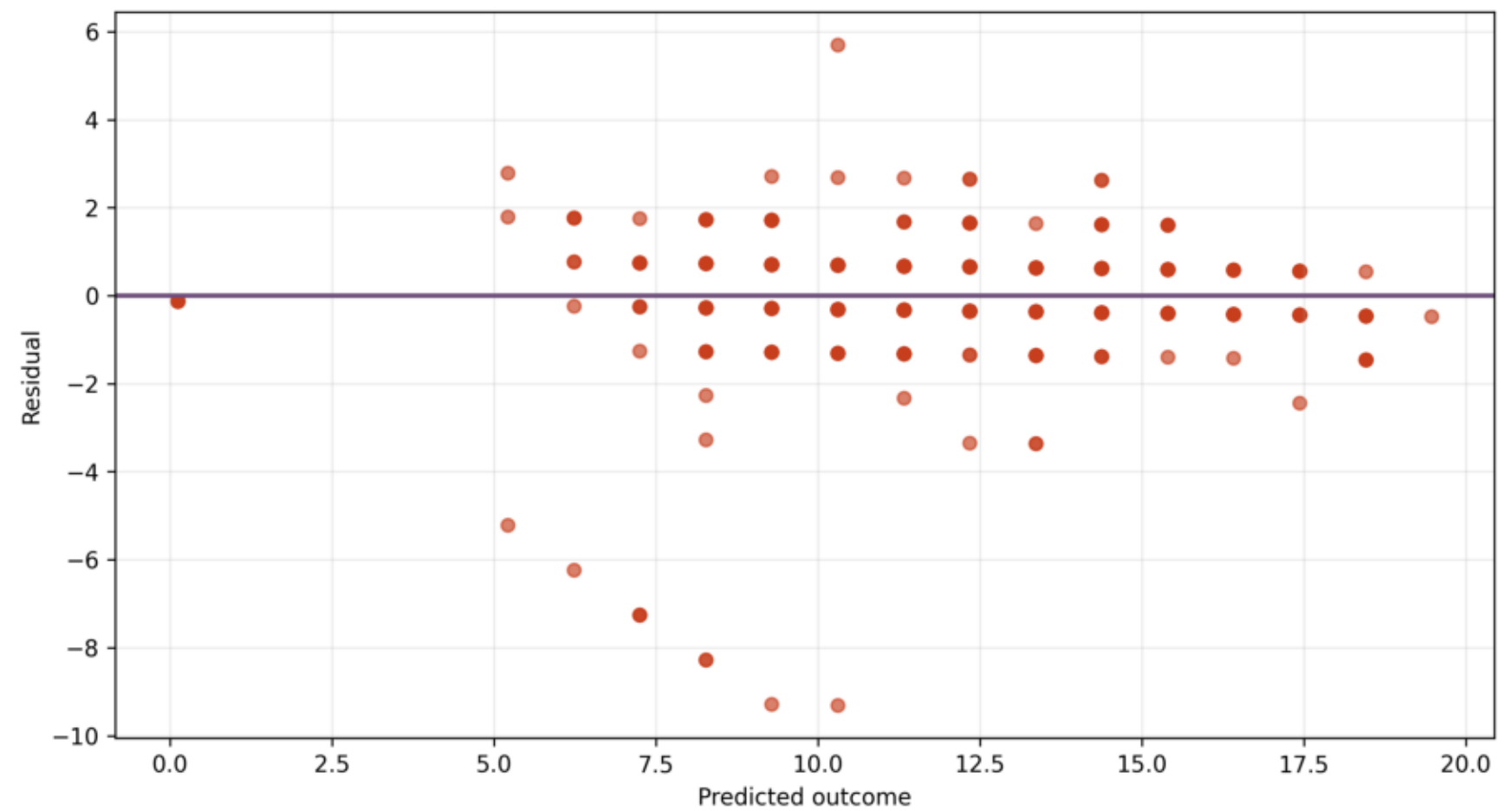
Observed versus predicted

Points closer to the diagonal indicate better fit



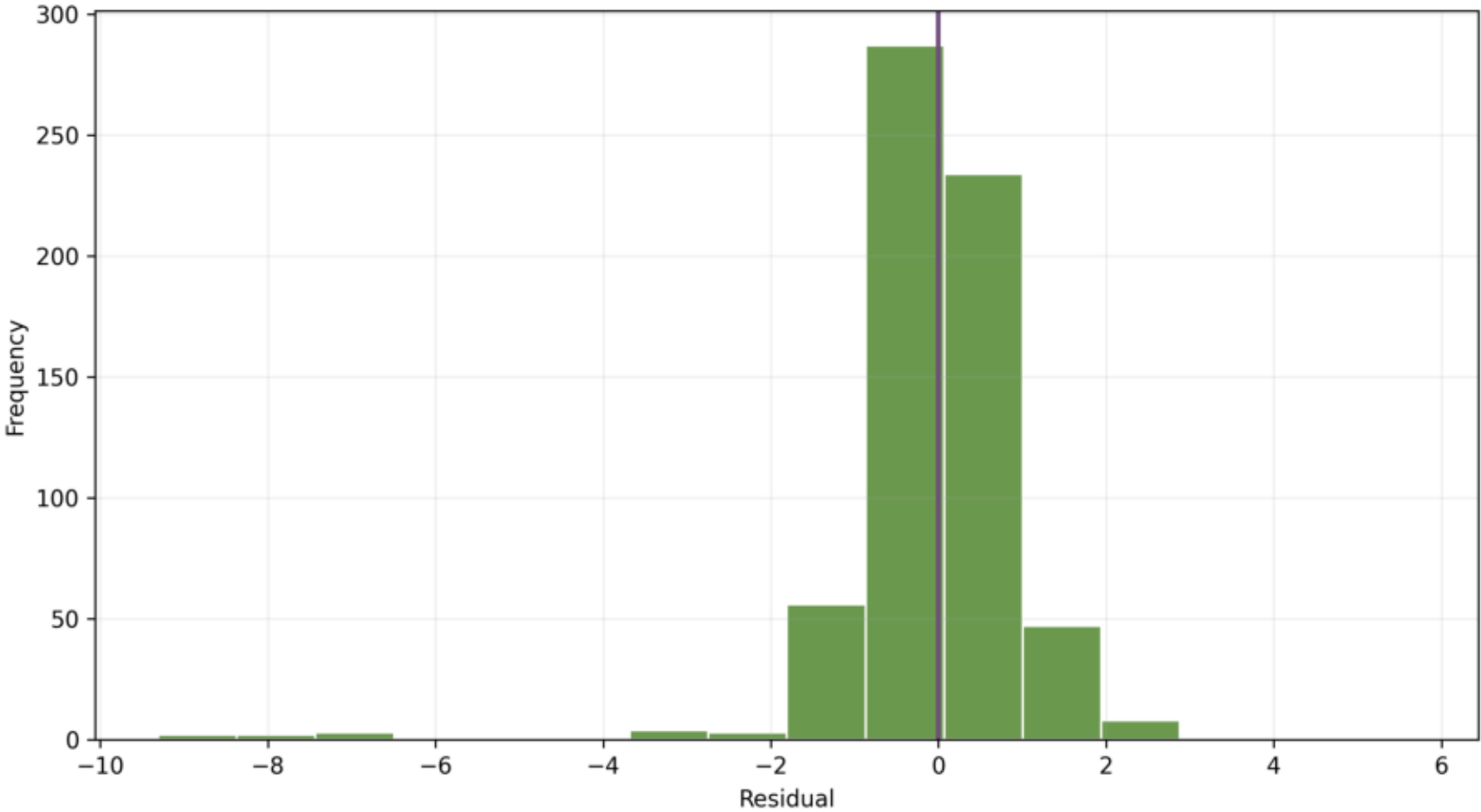
Residuals versus predicted values

Random scatter around zero supports linearity and constant variance



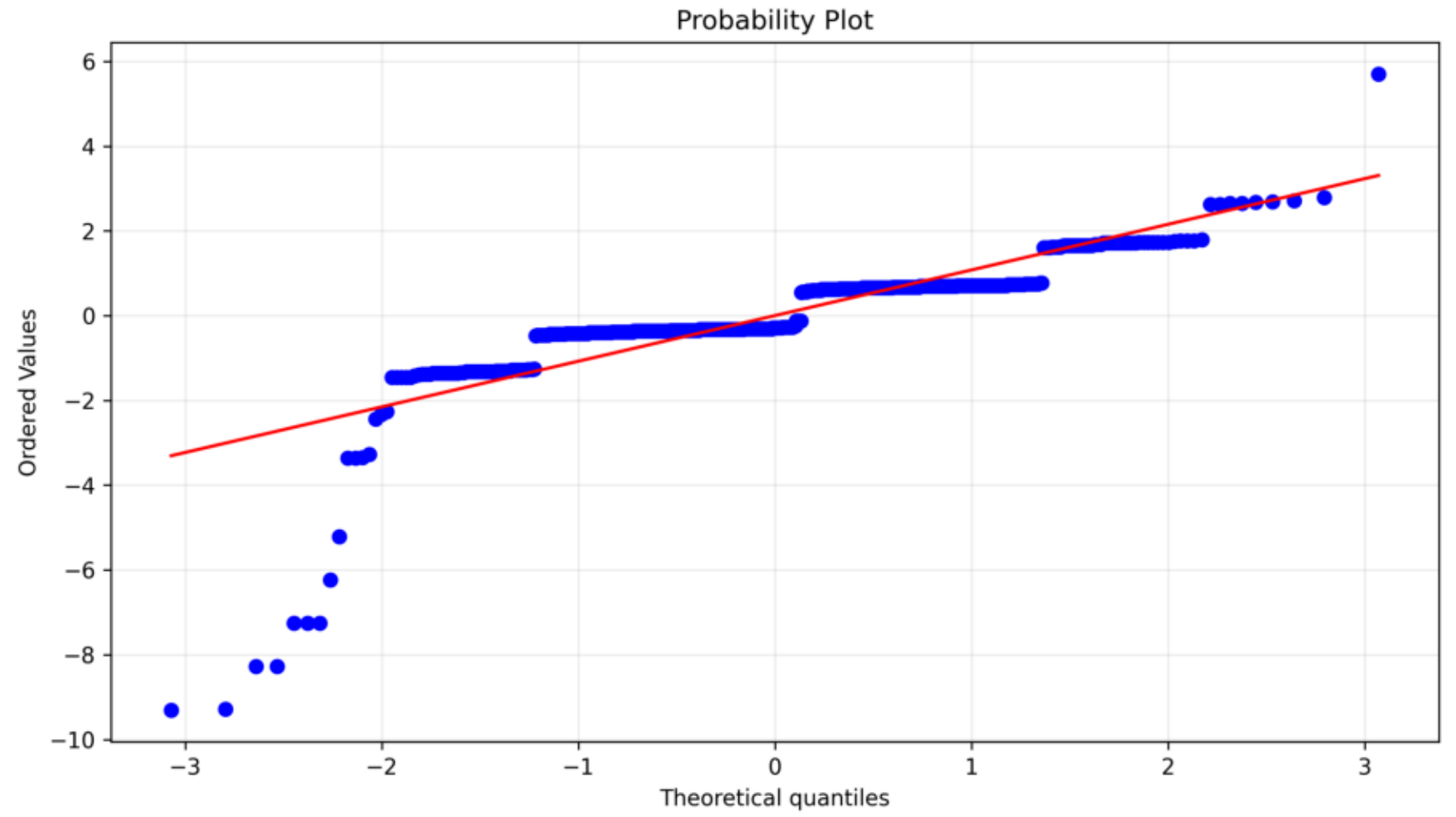
Residual distribution

A roughly symmetric distribution supports approximate normality



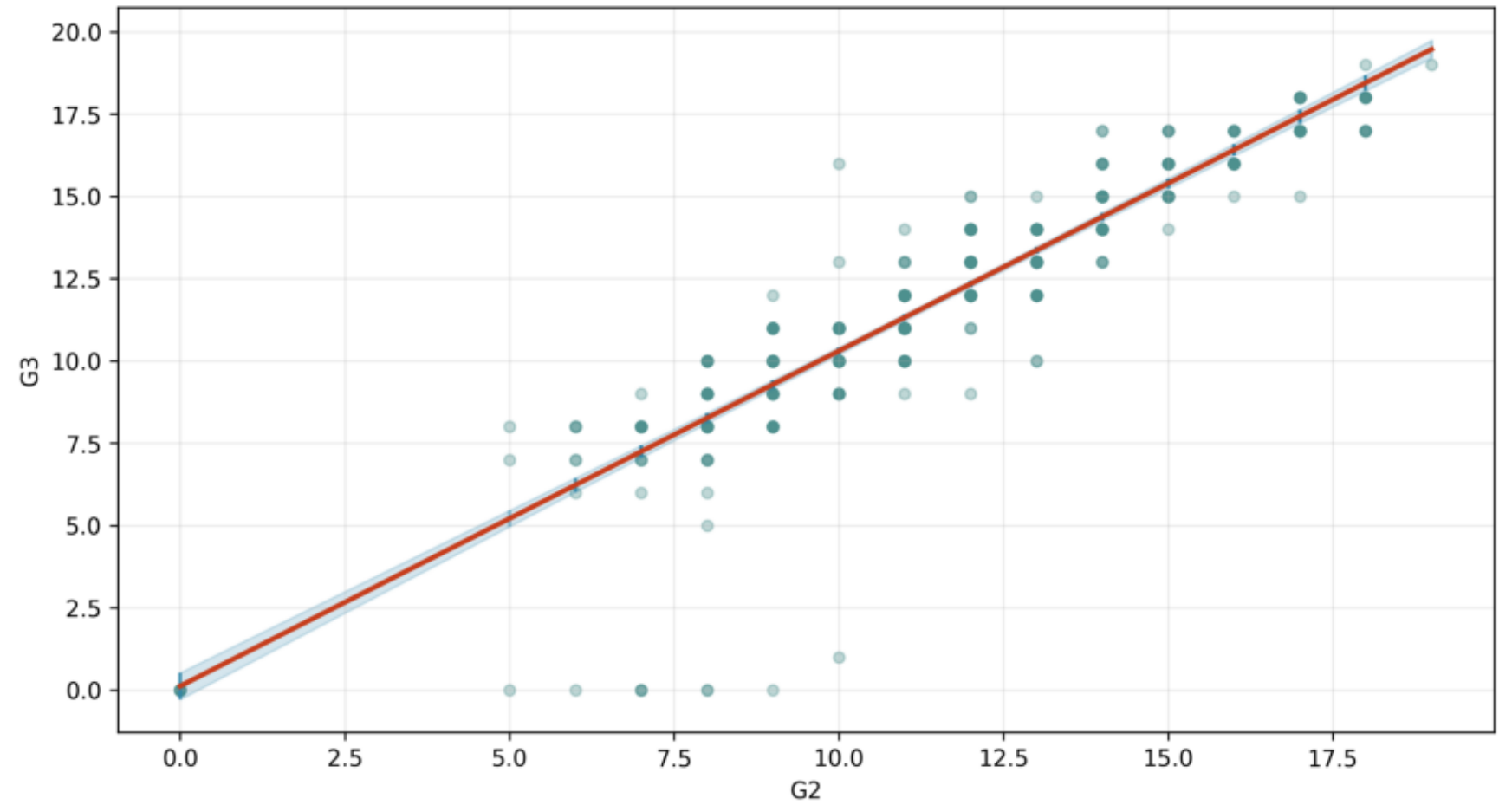
Normal Q-Q plot of residuals

Points close to the reference line support residual normality



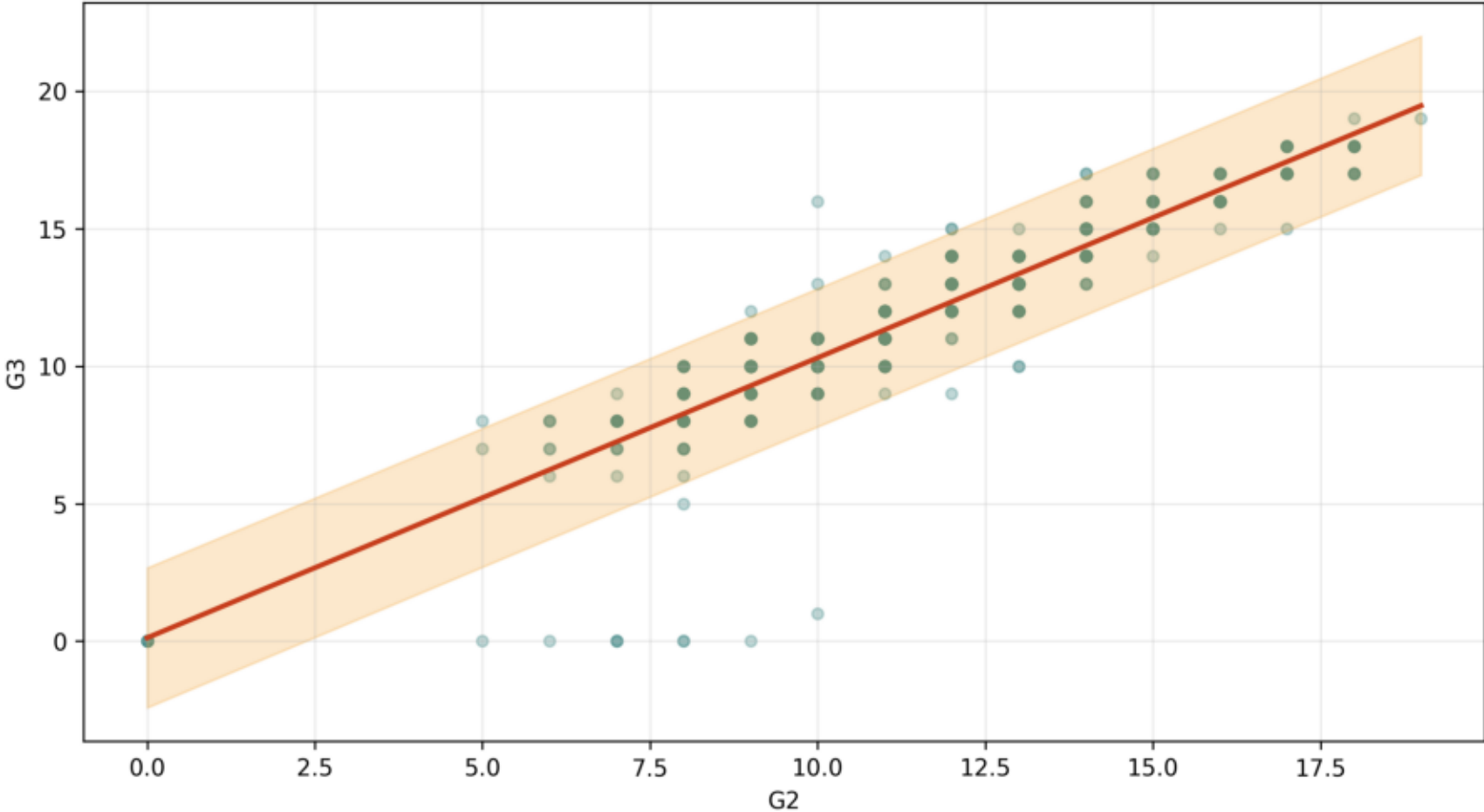
Fitted line with 95% mean confidence band

The band quantifies uncertainty in the expected outcome



Fitted line with 95% prediction band

Prediction intervals are wider because they include individual outcome variation



Influence diagnostic
Bubble size represents Cook's distance

