Regression in R

Regression in R Report
==========

Folder: D:/DATA ANALYSIS/H Regression Tests and Models/Regression in R
Dataset: D:/DATA ANALYSIS/H Regression Tests and Models/Regression in R/dataset.csv
Rows used: 649
Outcome: G3
Simple model: G3 ~ G2
Multiple model: G3 ~ G1 + G2 + studytime + failures + absences + age + Medu + Fedu

Simple regression results:
R−squared = 0.843730
Adjusted R−squared = 0.843489
F statistic = 3493.281566
Model p−value = 5.642401e−263
RMSE = 1.276125

Multiple regression results:
R−squared = 0.850777
Adjusted R−squared = 0.848912
F statistic = 456.111097
Model p−value = 1.454898e−258
RMSE = 1.247020
Strongest standardized predictor = G2 (Beta = 0.797983)

Multiple−regression coefficients:

model	term	coefficient_b	standard_error
Multiple linear regression	Intercept	−0.50115470	0.77395037

Model comparison

model	outcome	predictors	observations	number_of_predictors	r_squared	adjusted_r_squared	f_statistic	f_p_value	aic	bic	rmse	mae	residual_standard_error	durbin_watson	decision_alpha_0_05
Simple linear regression	G3	G2	649	1	0.8437304	0.8434889	3493.2816	5.642401e−263	2164.271	2177.697	1.276125	0.8080228	1.278096	1.846403	Model is statistically significant
Multiple linear regression	G3	G1, G2, studytime, failures, absences, age, Medu, Fedu	649	8	0.8507772	0.8489119	456.1111	1.454898e−258	2148.325	2193.079	1.247020	0.7800189	1.255758	1.861535	Model is statistically significant

Regression coefficients

model	term	coefficient_b	standard_error	t_value	p_value	ci95_lower	ci95_upper	standardized_beta	absolute_standardized_beta	decision_alpha_0_05
Simple linear regression	Intercept	0.12196613	0.20559337	0.5932396	5.532281e−01	−0.281744681	0.52567693	NA	NA	Not statistically significant
Simple linear regression	G2	1.01849034	0.01723218	59.1039894	5.642401e−263	0.984652599	1.05232809	0.918548004	0.918548004	Statistically significant
Multiple linear regression	Intercept	−0.50115470	0.77395037	−0.6475282	5.175223e−01	−2.020943673	1.01863427	NA	NA	Not statistically significant
Multiple linear regression	G1	0.14339666	0.03667185	3.9102656	1.020384e−04	0.071384979	0.21540835	0.121851980	0.121851980	Statistically significant
Multiple linear regression	G2	0.88480738	0.03436950	25.7439712	7.355488e−101	0.817316763	0.95229799	0.797983068	0.797983068	Statistically significant
Multiple linear regression	studytime	0.09663180	0.06209078	1.5562987	1.201314e−01	−0.025294474	0.21855808	0.024811369	0.024811369	Statistically significant
Multiple linear regression	failures	−0.23536120	0.09523444	−2.4713874	1.371819e−02	−0.422370939	−0.04835147	−0.043218630	0.043218630	Statistically significant
Multiple linear regression	absences	0.02276203	0.01091928	2.0845722	3.750404e−02	0.001320082	0.04420398	0.032697099	0.032697099	Statistically significant
Multiple linear regression	age	0.02268549	0.04366788	0.5195005	6.035914e−01	−0.063064155	0.10843513	0.008553694	0.008553694	Not statistically significant
Multiple linear regression	Medu	−0.04495115	0.05793762	−0.7758543	4.381216e−01	−0.158721956	0.06881965	−0.015786086	0.015786086	Not statistically significant
Multiple linear regression	Fedu	0.02202534	0.05931553	0.3713250	7.105183e−01	−0.094451233	0.13850192	0.007498896	0.007498896	Not statistically significant

Regression ANOVA

model	source	sum_of_squares	df	mean_square	f_value	p_value
Simple linear regression	Regression	5706.374	1	5706.373839	3493.2816	5.642401e−263
Simple linear regression	Residual	1056.893	647	1.633528	NA	NA
Simple linear regression	Total	6763.267	648	NA	NA	NA
Multiple linear regression	Regression	5754.033	8	719.254122	456.1111	1.454898e−258
Multiple linear regression	Residual	1009.234	640	1.576927	NA	NA
Multiple linear regression	Total	6763.267	648	NA	NA	NA

Multiple−regression VIF

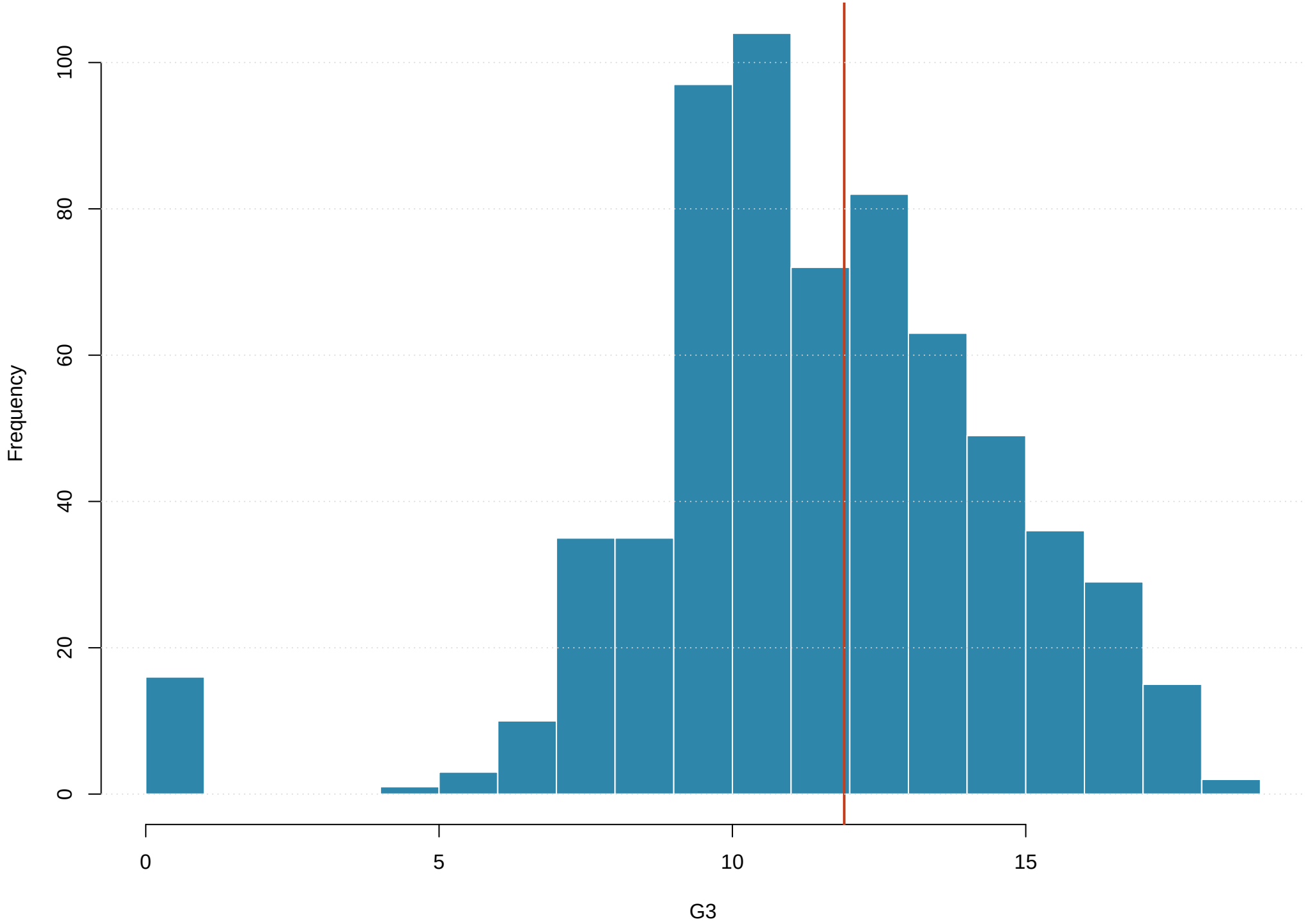
predictor	vif	tolerance	diagnostic
G1	4.164832	0.2401057	Acceptable
G2	4.120795	0.2426716	Acceptable
studytime	1.090085	0.9173597	Acceptable
failures	1.311611	0.7624211	Acceptable
absences	1.055187	0.9476995	Acceptable
age	1.162734	0.8600421	Acceptable
Medu	1.775552	0.5632051	Acceptable
Fedu	1.749166	0.5717012	Acceptable

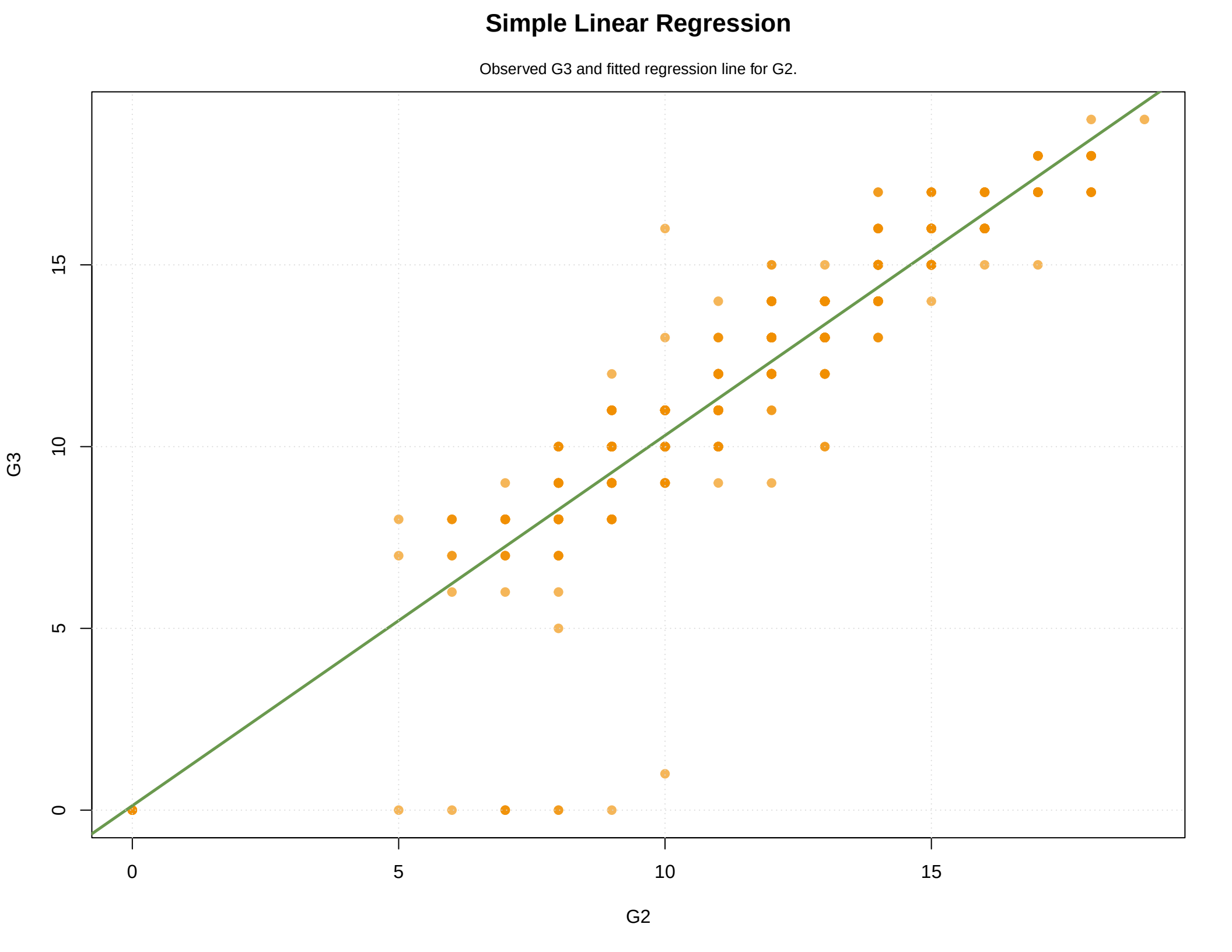
Diagnostic tests

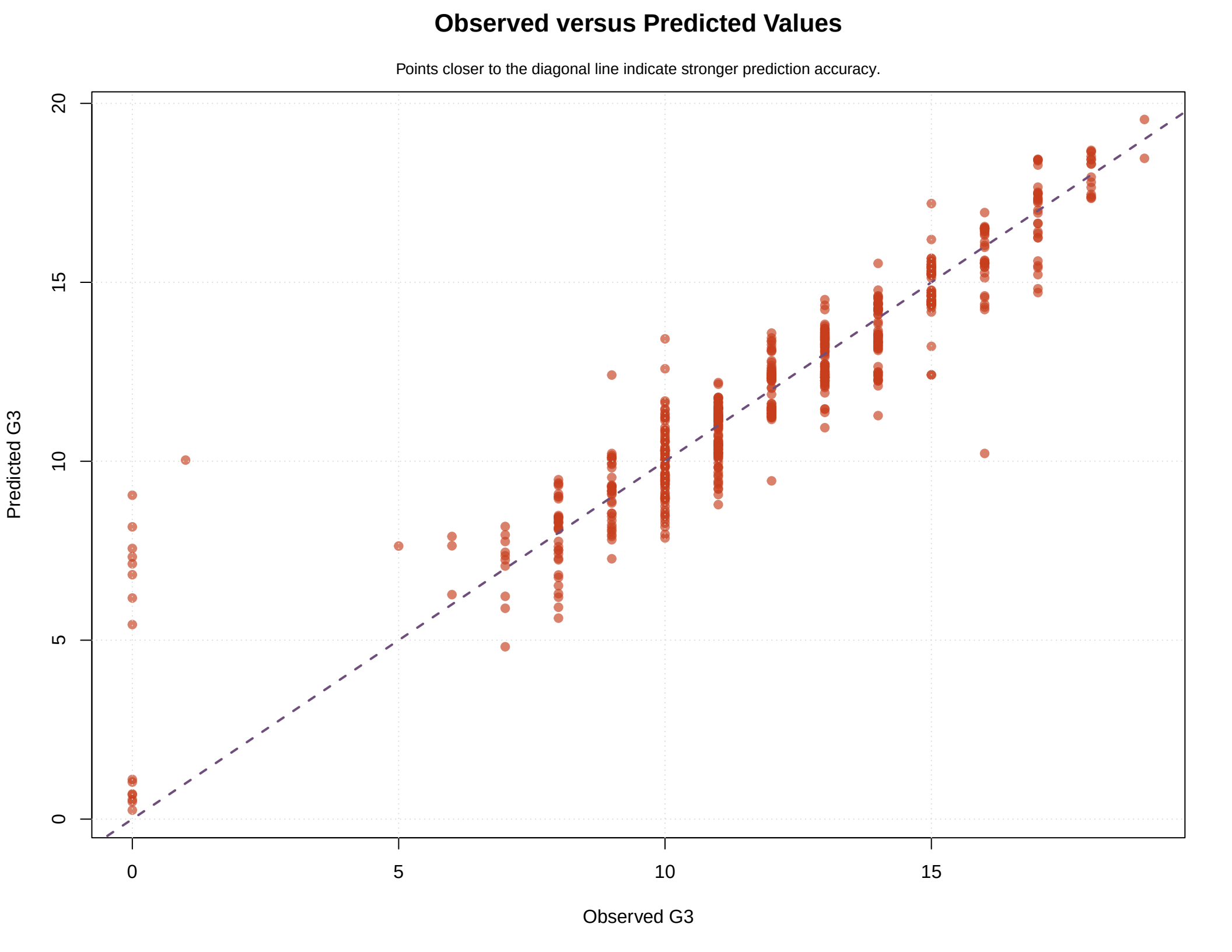
model	diagnostic	statistic	p_value	interpretation
Simple linear regression	Shapiro–Wilk residual normality	7.108466e−01	6.059646e−32	Potential non−normality
Simple linear regression	Jarque–Bera residual normality	1.041057e+04	0.000000e+00	Potential non−normality
Simple linear regression	Breusch–Pagan heteroscedasticity	1.575877e+01	7.195376e−05	Potential heteroscedasticity
Simple linear regression	Durbin–Watson residual independence	1.846403e+00	NA	Values near 2 indicate approximate independence
Multiple linear regression	Shapiro–Wilk residual normality	7.567934e−01	7.918469e−30	Potential non−normality
Multiple linear regression	Jarque–Bera residual normality	9.923742e+03	0.000000e+00	Potential non−normality
Multiple linear regression	Breusch–Pagan heteroscedasticity	3.572281e+01	1.973500e−05	Potential heteroscedasticity
Multiple linear regression	Durbin–Watson residual independence	1.861535e+00	NA	Values near 2 indicate approximate independence

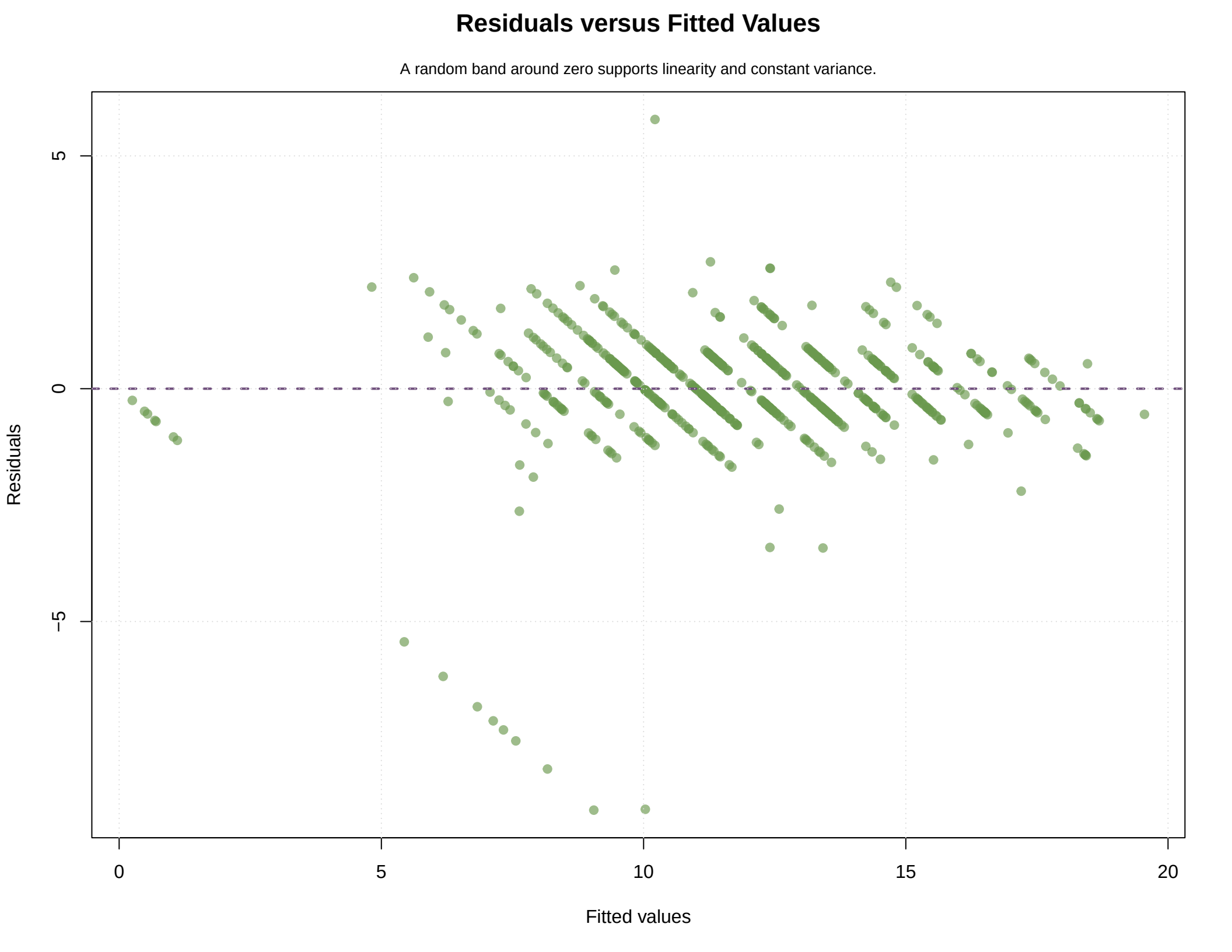
Outcome Distribution

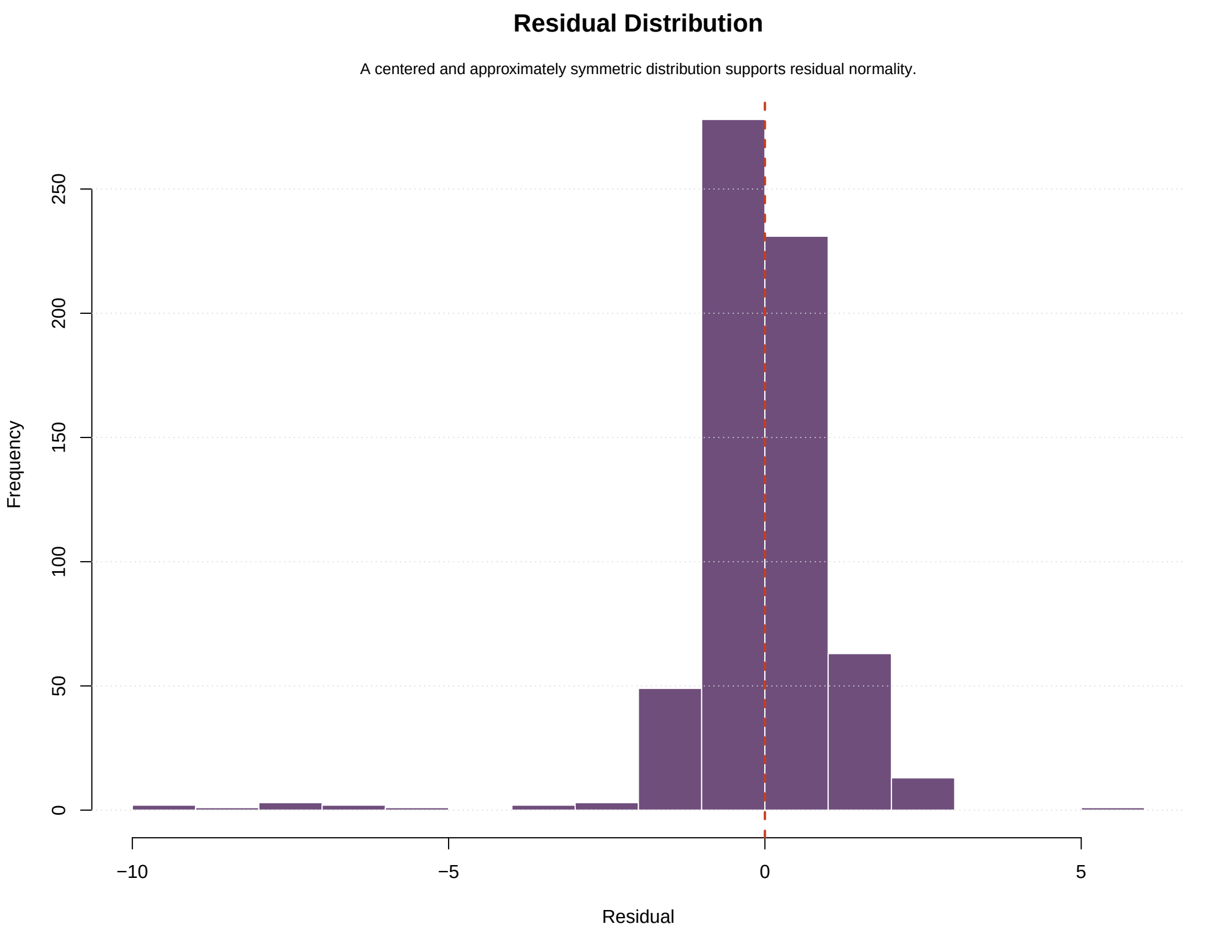
Distribution of the dependent variable used in the regression analyses.

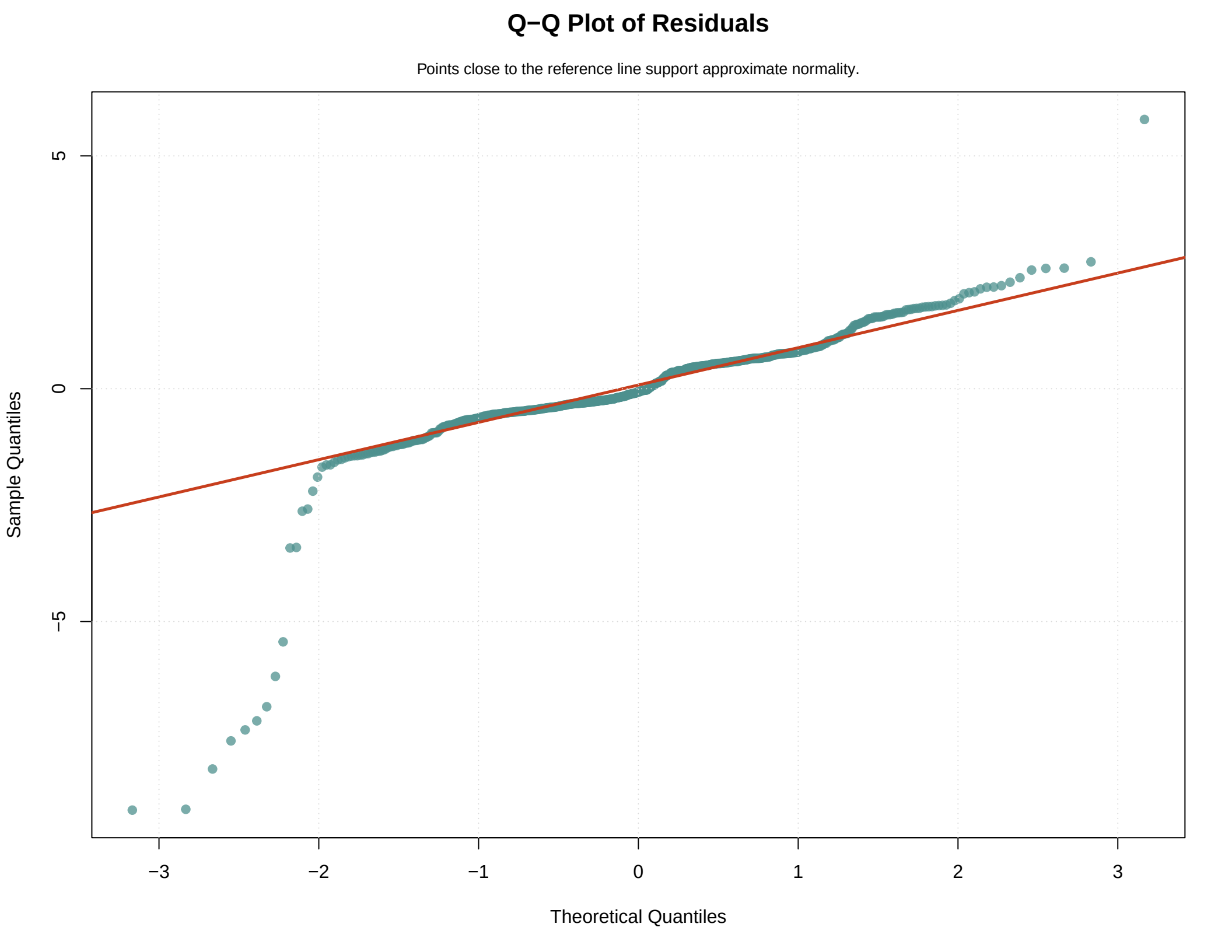


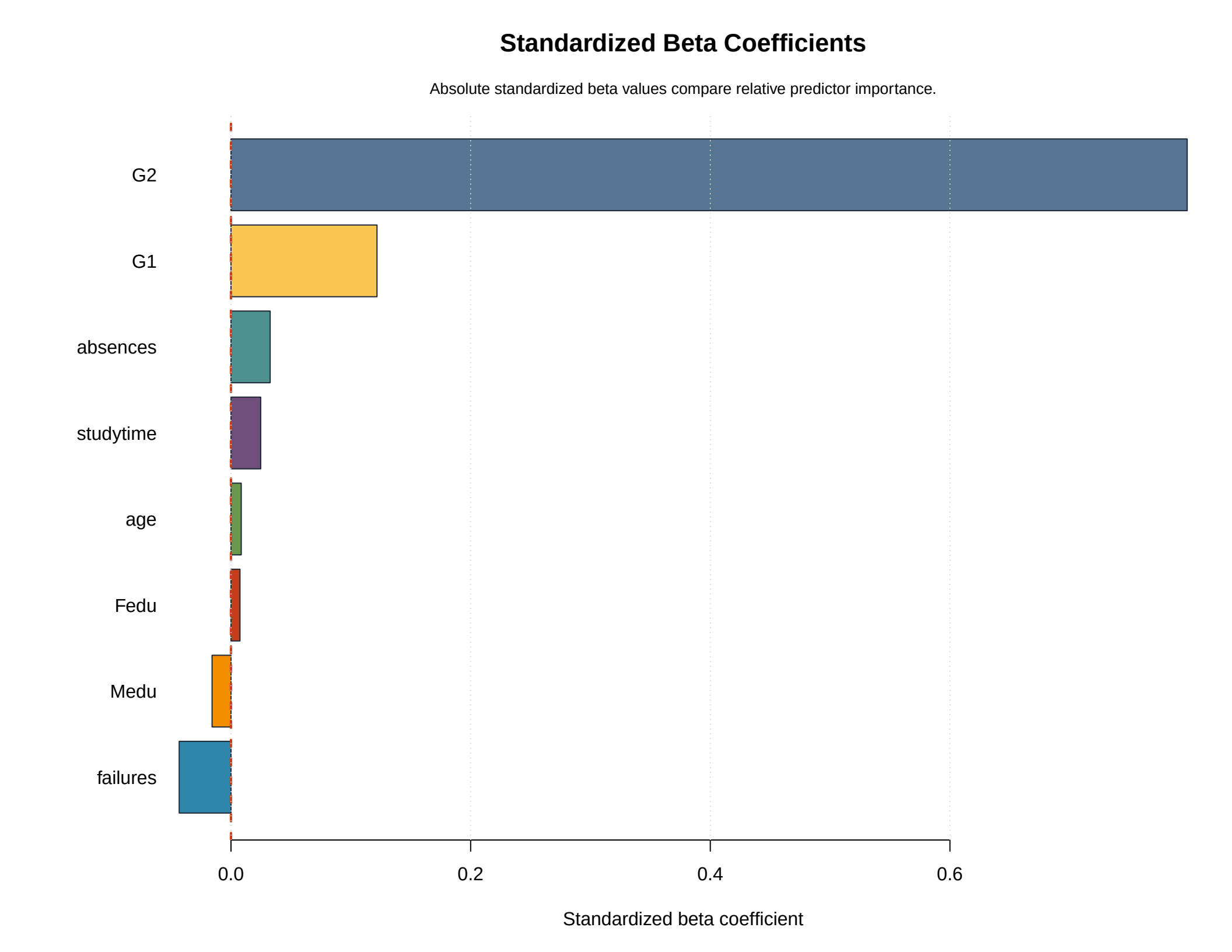

Simple Linear Regression
Observed G3 and fitted regression line for G2.
G3
G2
0
5
10
15
0
5
10
15

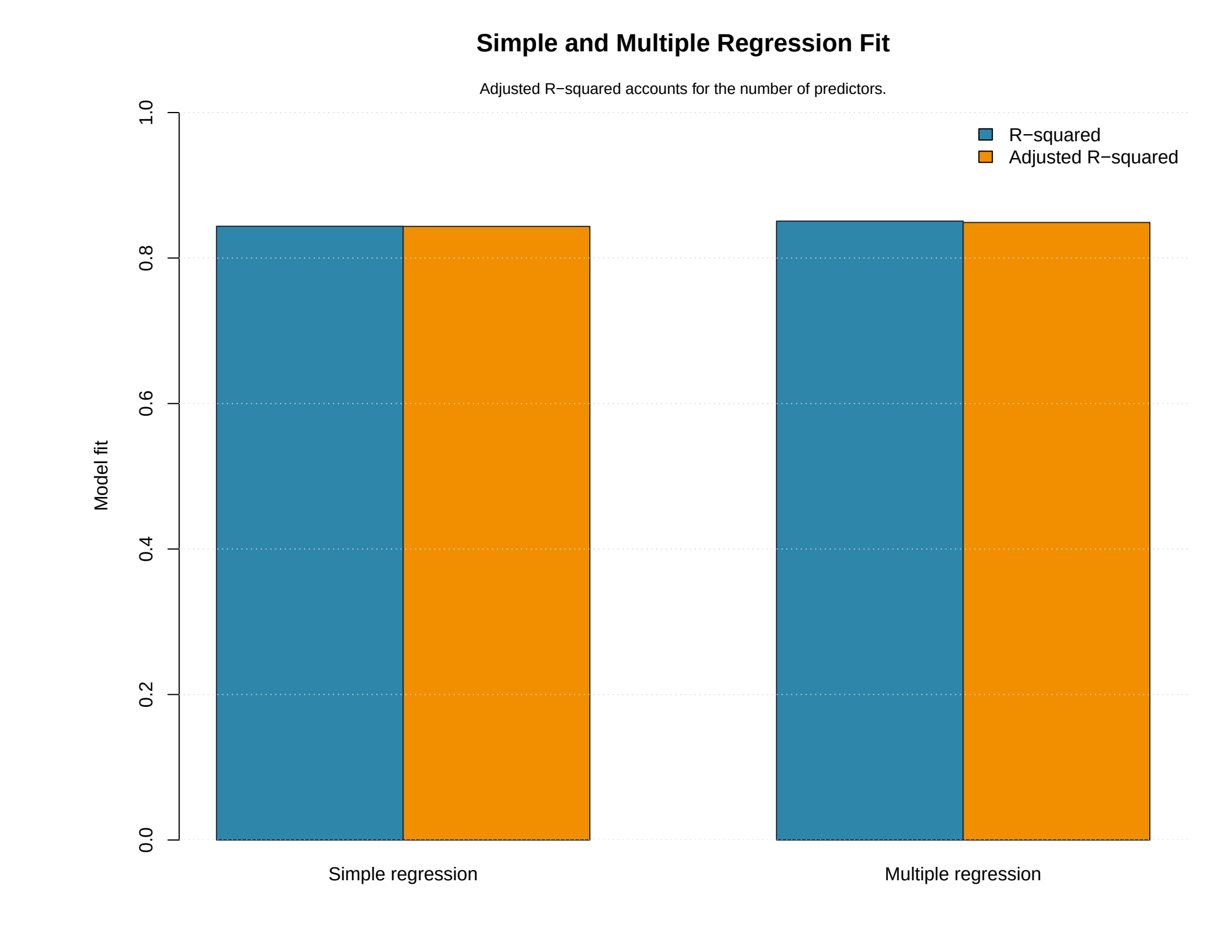

Observed versus Predicted Values
Points closer to the diagonal line indicate stronger prediction accuracy.
Predicted G3
Observed G3
0
5
10
15
20
0
5
10
15

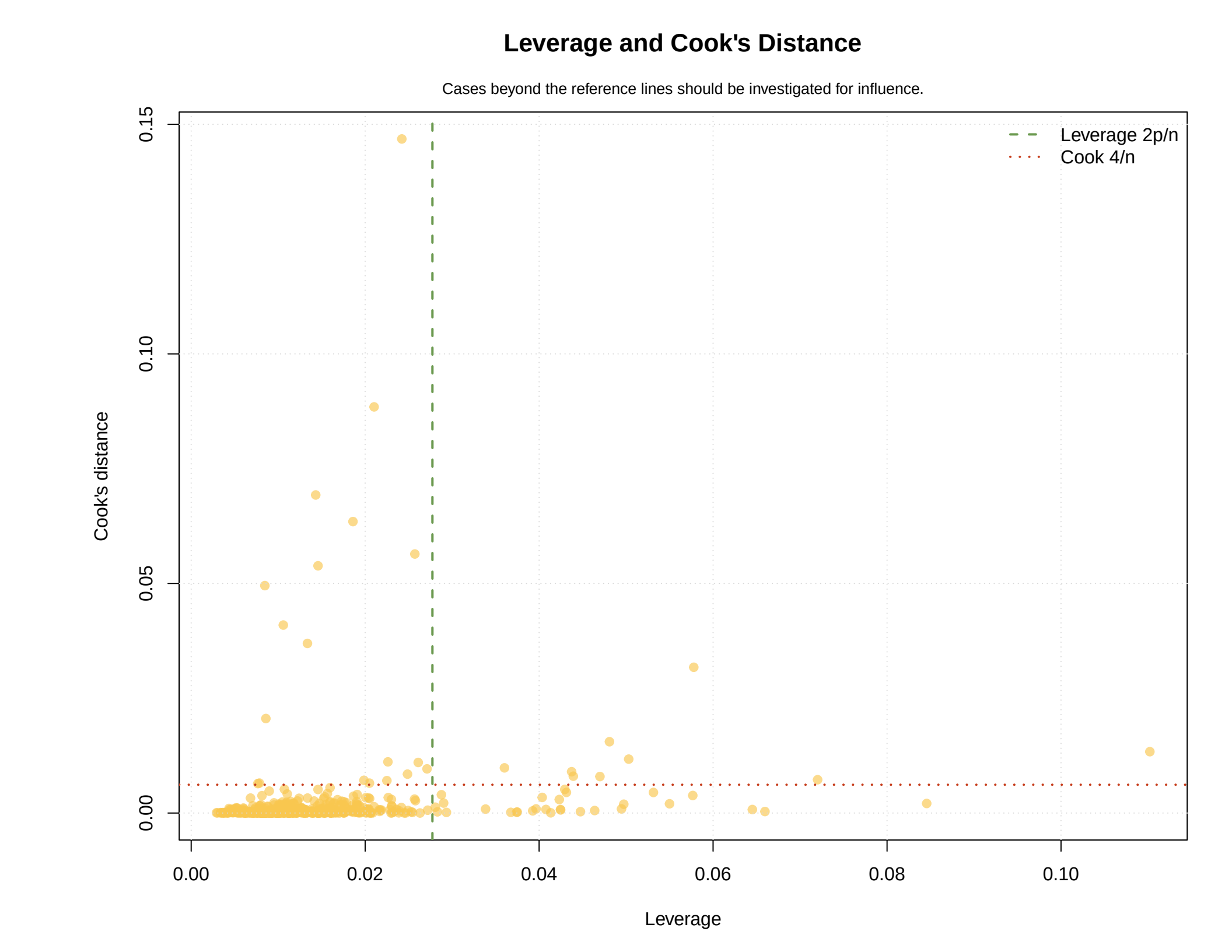

Residuals versus Fitted Values
A random band around zero supports linearity and constant variance.
Residuals
Fitted values
5
0
−5
0
5
10
15
20

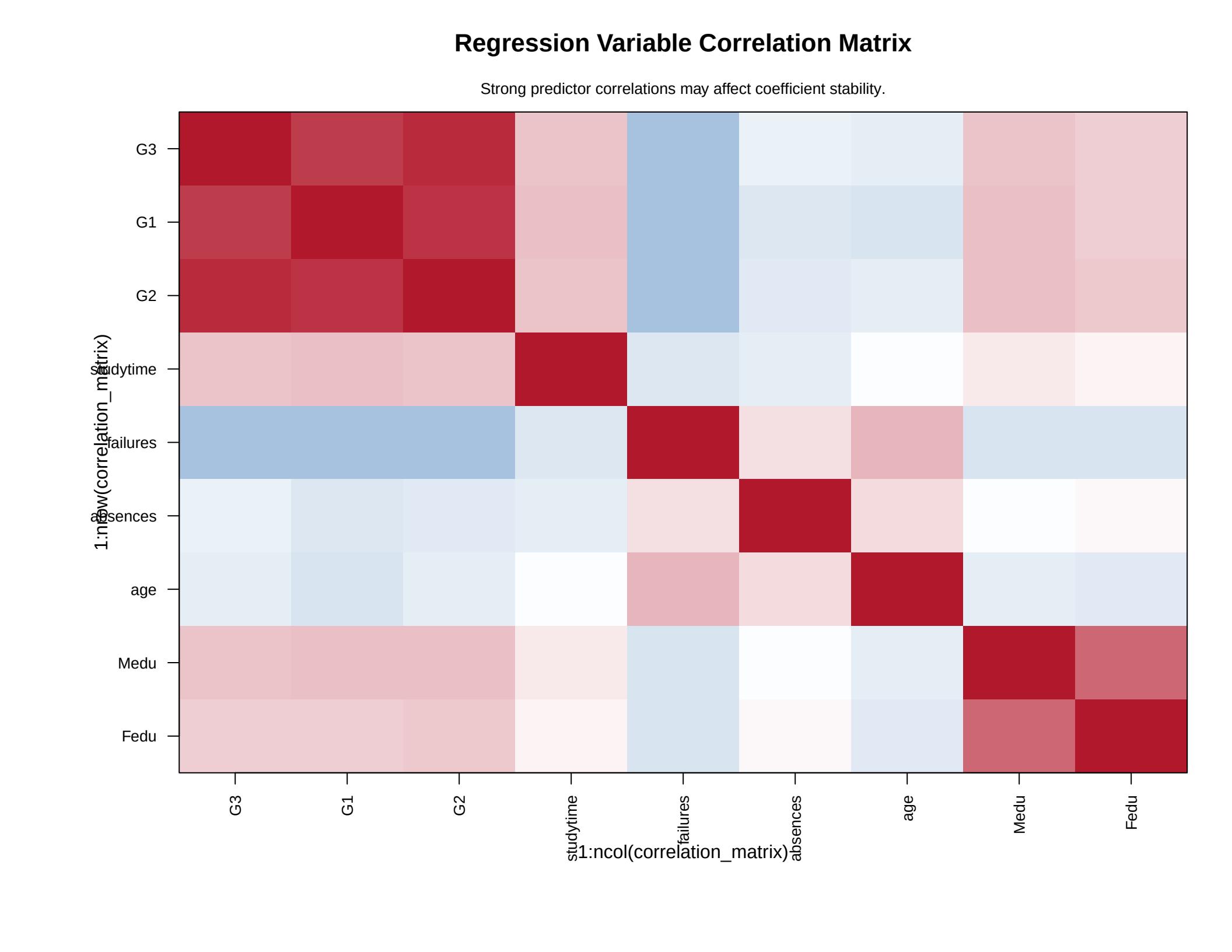

Residual Distribution
A centered and approximately symmetric distribution supports residual normality.
Frequency
0
50
100
150
200
250
−10
−5
0
5
Residual


Q−Q Plot of Residuals
Points close to the reference line support approximate normality.
Sample Quantiles
Theoretical Quantiles
5
0
−5
−3
−2
−1
0
1
2
3


Standardized Beta Coefficients
Absolute standardized beta values compare relative predictor importance.
G2
G1
absences
studytime
age
Fedu
Medu
failures
0.0
0.2
0.4
0.6
Standardized beta coefficient


Simple and Multiple Regression Fit
Adjusted R−squared accounts for the number of predictors.
R−squared
Adjusted R−squared
Model fit
0.0
0.2
0.4
0.6
0.8
1.0
Simple regression
Multiple regression


Leverage and Cook's Distance
Cases beyond the reference lines should be investigated for influence.
Leverage 2p/n
Cook 4/n
Cook's distance
0.00
0.05
0.10
0.15
0.00
0.02
0.04
0.06
0.08
0.10
Leverage


Regression Variable Correlation Matrix
Strong predictor correlations may affect coefficient stability.
1:nrow(correlation_matrix)
G3
G1
G2
studytime
failures
absences
age
Medu
Fedu
1:ncol(correlation_matrix)
G3
G1
G2
studytime
failures
absences
age
Medu
Fedu