

# 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:  $G3 = 0.121966 + (1.01849 \times G2)$

Pearson r: 0.918548

R-squared: 0.84373

Adjusted R-squared: 0.843489

F statistic: 3493.281566

Overall model p-value: 5.642401e-263

Slope p-value: 5.642401e-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 relation

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                | n_complete_cases | correlation_r | r_squared |
|--------------------------|-------------|-----------------------------------------------|-----------------------------|-----------|-------------------------|------------------|---------------|-----------|
| Simple Linear Regression | G3          | G2                                            | G3 = intercept + slope * G2 |           |                         | 649              | 0.918548      | 0.8437304 |
| adjusted_r_squared       | f_statistic | f_p_value                                     | rmse                        | mae       | residual_standard_error | df_model         | df_residual   |           |
| 0.8434889                | 3493.282    | 5.642401e-263                                 | 1.276125                    | 0.8080228 | 1.278096                | 1                | 647           |           |
| aic                      | bic         | decision                                      |                             |           |                         |                  |               |           |
| 2164.271                 | 2177.697    | Statistically significant linear relationship |                             |           |                         |                  |               |           |

## Regression coefficients

|  | term                          | coefficient_b | standard_error | t_value    | p_value       | ci95_lower | ci95_upper |
|--|-------------------------------|---------------|----------------|------------|---------------|------------|------------|
|  | Intercept                     | 0.1219661     | 0.20559337     | 0.5932396  | 5.532281e-01  | -0.2817447 | 0.5256769  |
|  | Slope                         | 1.0184903     | 0.01723218     | 59.1039894 | 5.642401e-263 | 0.9846526  | 1.0523281  |
|  | decision_alpha_0_05           |               |                |            |               |            |            |
|  | Not statistically significant |               |                |            |               |            |            |
|  | Statistically significant     |               |                |            |               |            |            |

# Regression ANOVA

| source           | sum_of_squares | df  | mean_square | f_value  | p_value       |
|------------------|----------------|-----|-------------|----------|---------------|
| Regression       | 5706.374       | 1   | 5706.373839 | 3493.282 | 5.642401e-263 |
| Residual / Error | 1056.893       | 647 | 1.633528    | NA       | NA            |
| Total            | 6763.267       | 648 | NA          | NA       | NA            |

# Descriptive statistics

| variable | n   | mean     | standard_deviation | minimum | maximum | skewness   | kurtosis_excess |
|----------|-----|----------|--------------------|---------|---------|------------|-----------------|
| G3       | 649 | 11.90601 | 3.230656           | 0       | 19      | -0.9086938 | 2.664626        |
| G2       | 649 | 11.57011 | 2.913639           | 0       | 19      | -0.3586190 | 1.626164        |

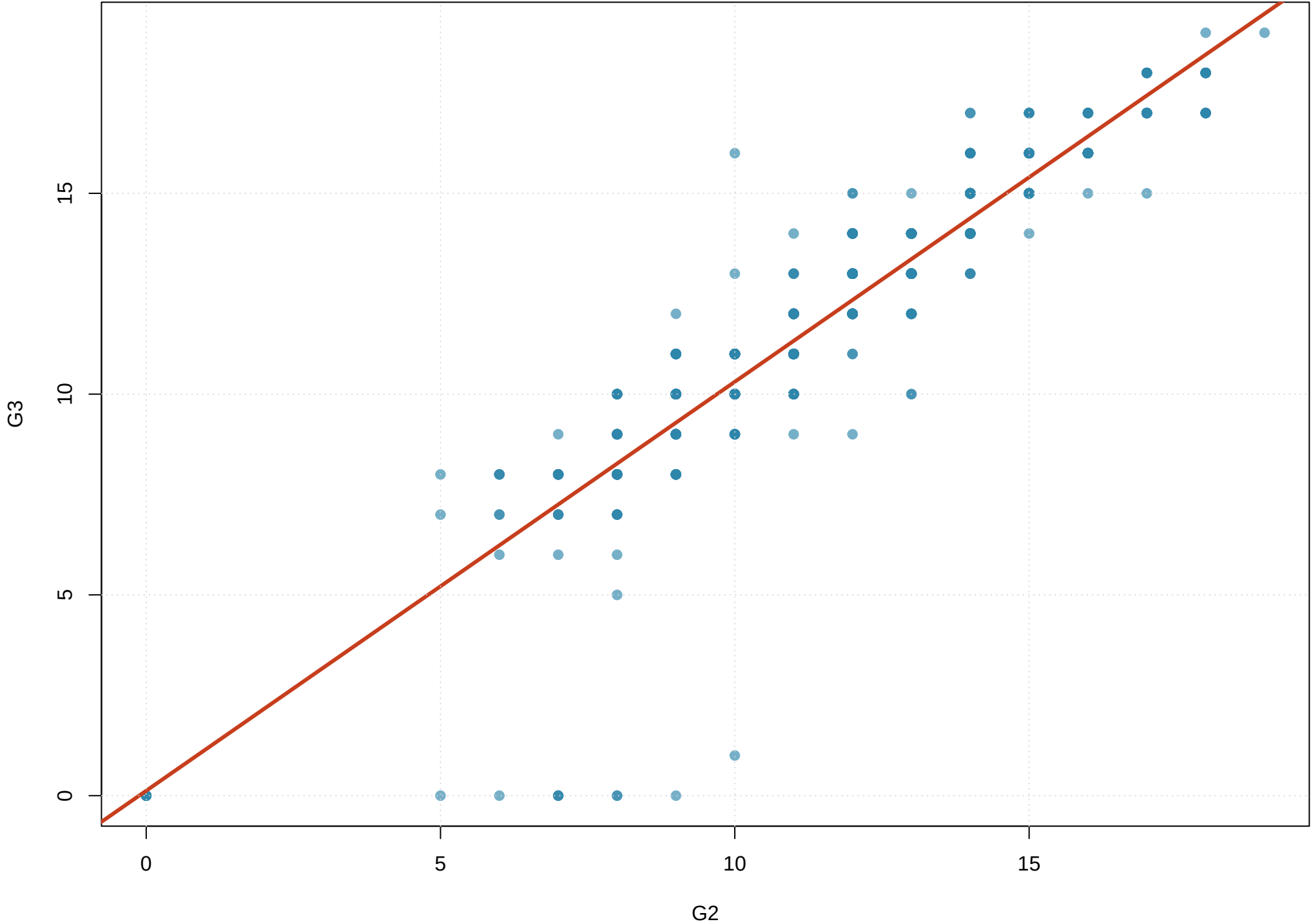
# Pearson correlation

| outcome | predictor | pearson_r | p_value       | n   |
|---------|-----------|-----------|---------------|-----|
| G3      | G2        | 0.918548  | 5.642401e-263 | 649 |

# Diagnostic summary

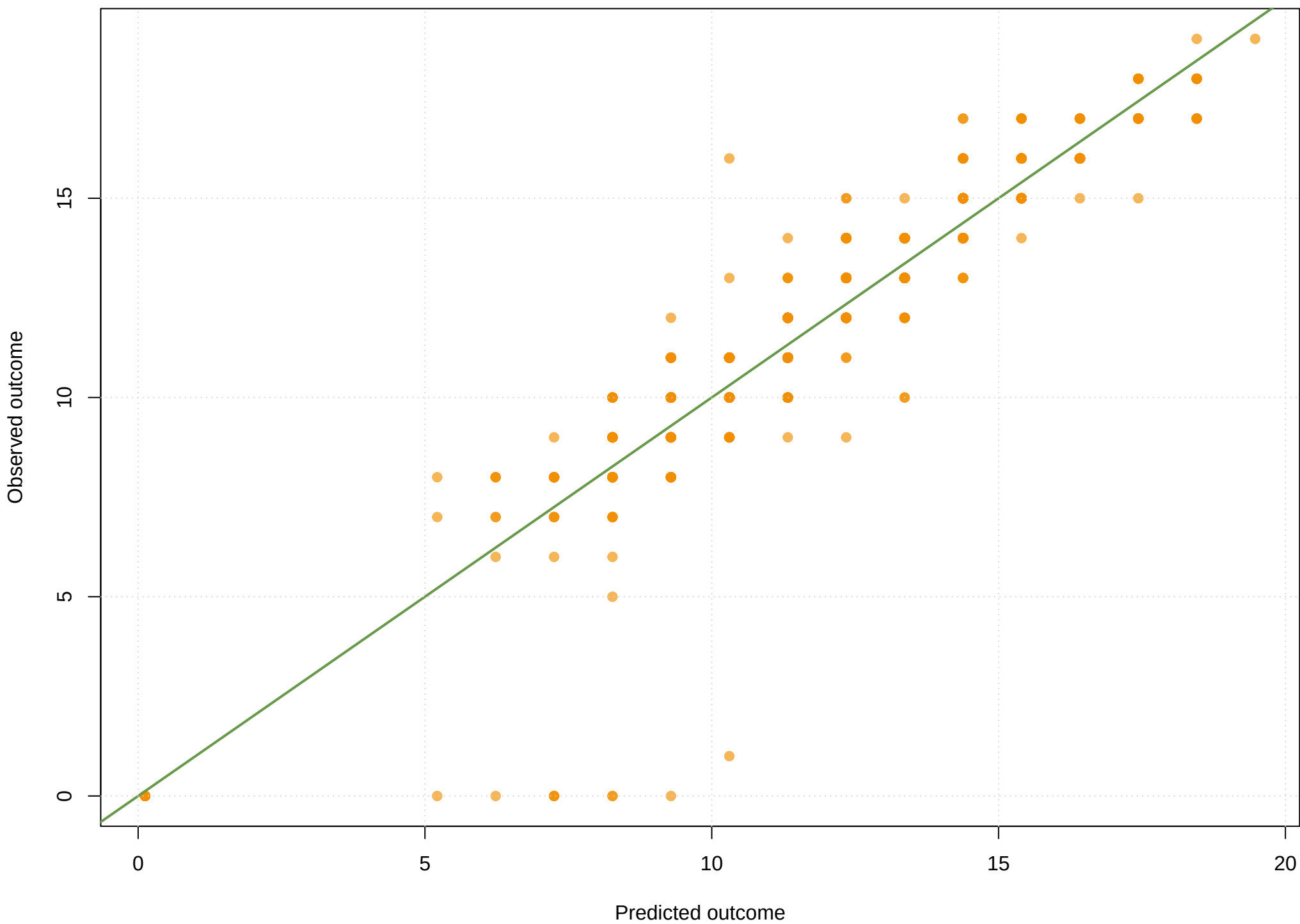
| diagnostic                             | statistic   | p_value      | guideline                                       |
|----------------------------------------|-------------|--------------|-------------------------------------------------|
| Durbin-Watson                          | 1.84640330  | NA           | Approximately 2 suggests independent residuals  |
| Shapiro-Wilk                           | 0.71084663  | 6.059646e-32 | p >= 0.05 supports residual normality           |
| Maximum absolute standardized residual | 7.28907875  | NA           | Values above 3 deserve review                   |
| Maximum absolute studentized residual  | 7.60227285  | NA           | Values above 3 deserve review                   |
| Maximum leverage                       | 0.02587567  | NA           | Compare with 2p/n; p includes intercept         |
| Maximum Cook's distance                | 0.08683126  | NA           | Compare with 4/n                                |
| Residual skewness                      | -2.93462909 | NA           | Values near zero indicate symmetry              |
| Residual kurtosis                      | 21.72255809 | NA           | Normal-distribution kurtosis is approximately 3 |

Simple linear regression fit

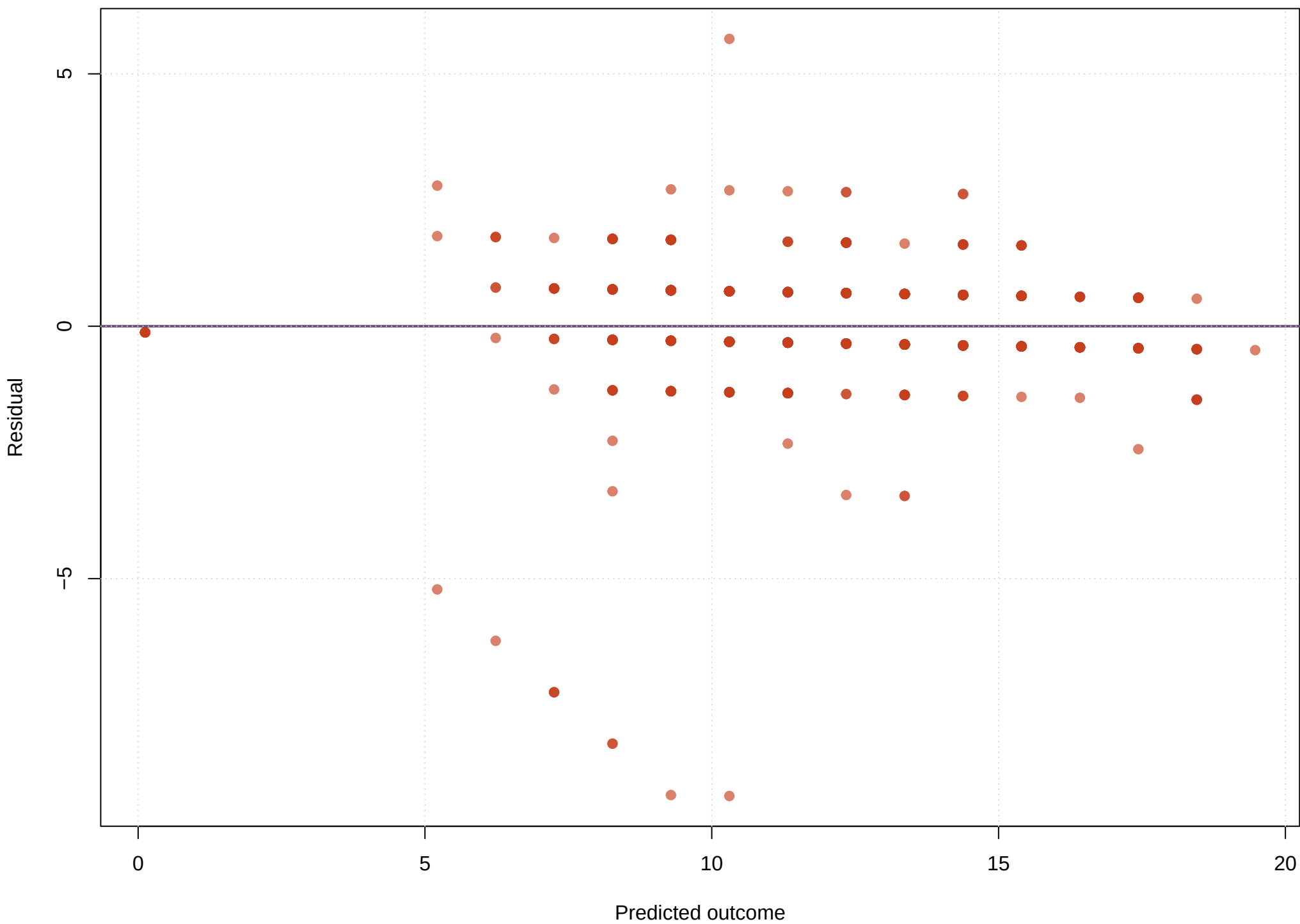




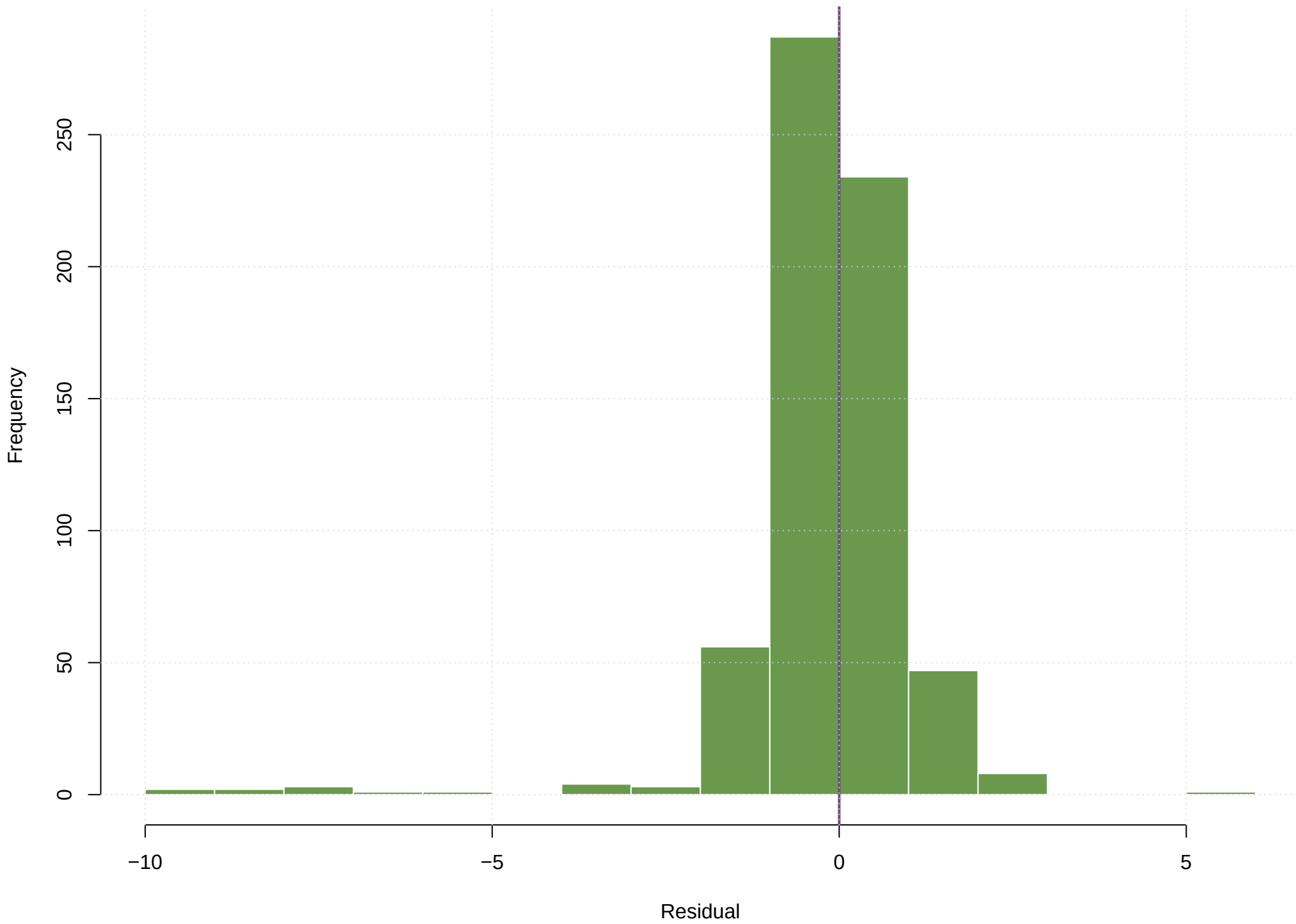
Observed versus predicted



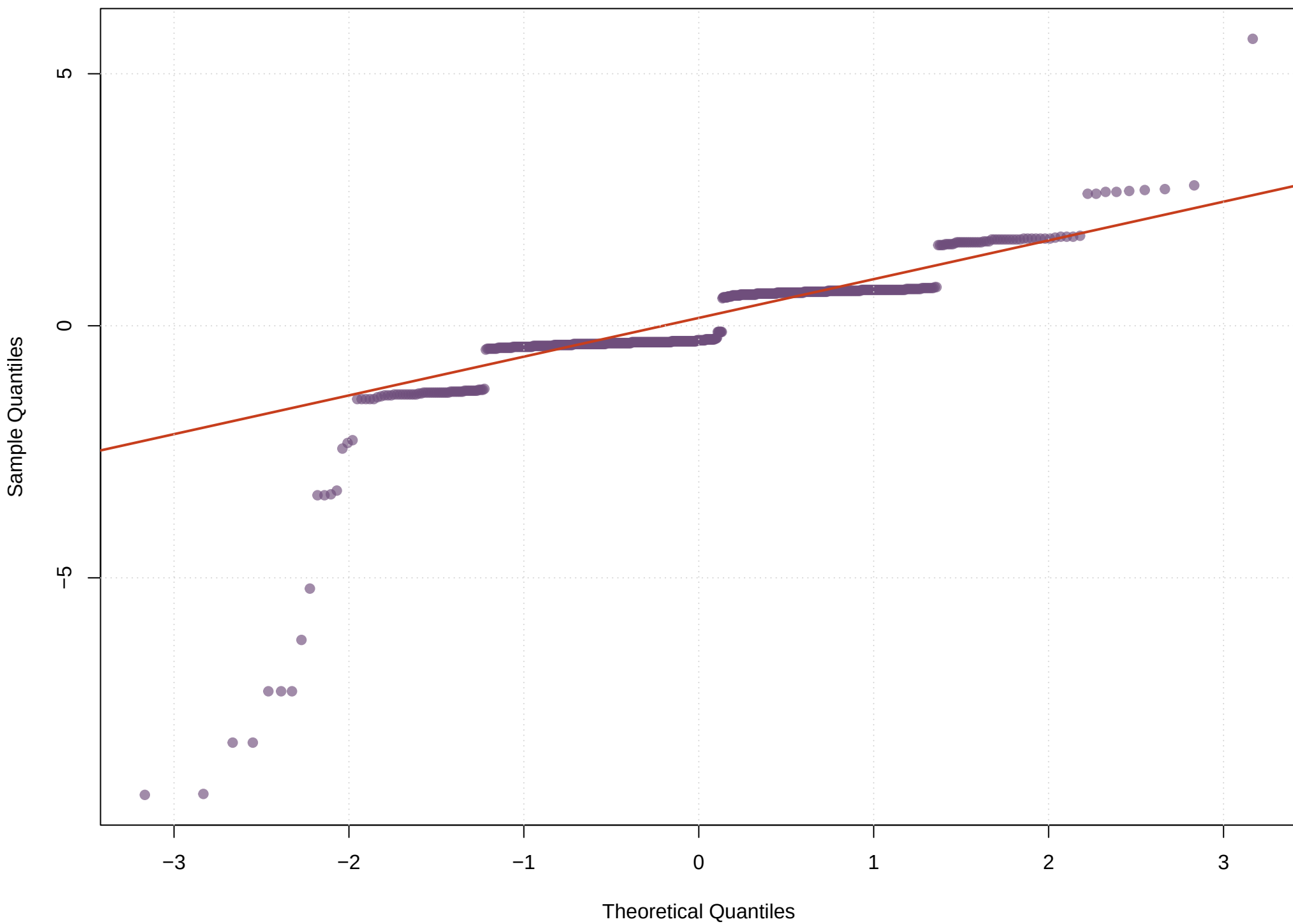
Residuals versus predicted values



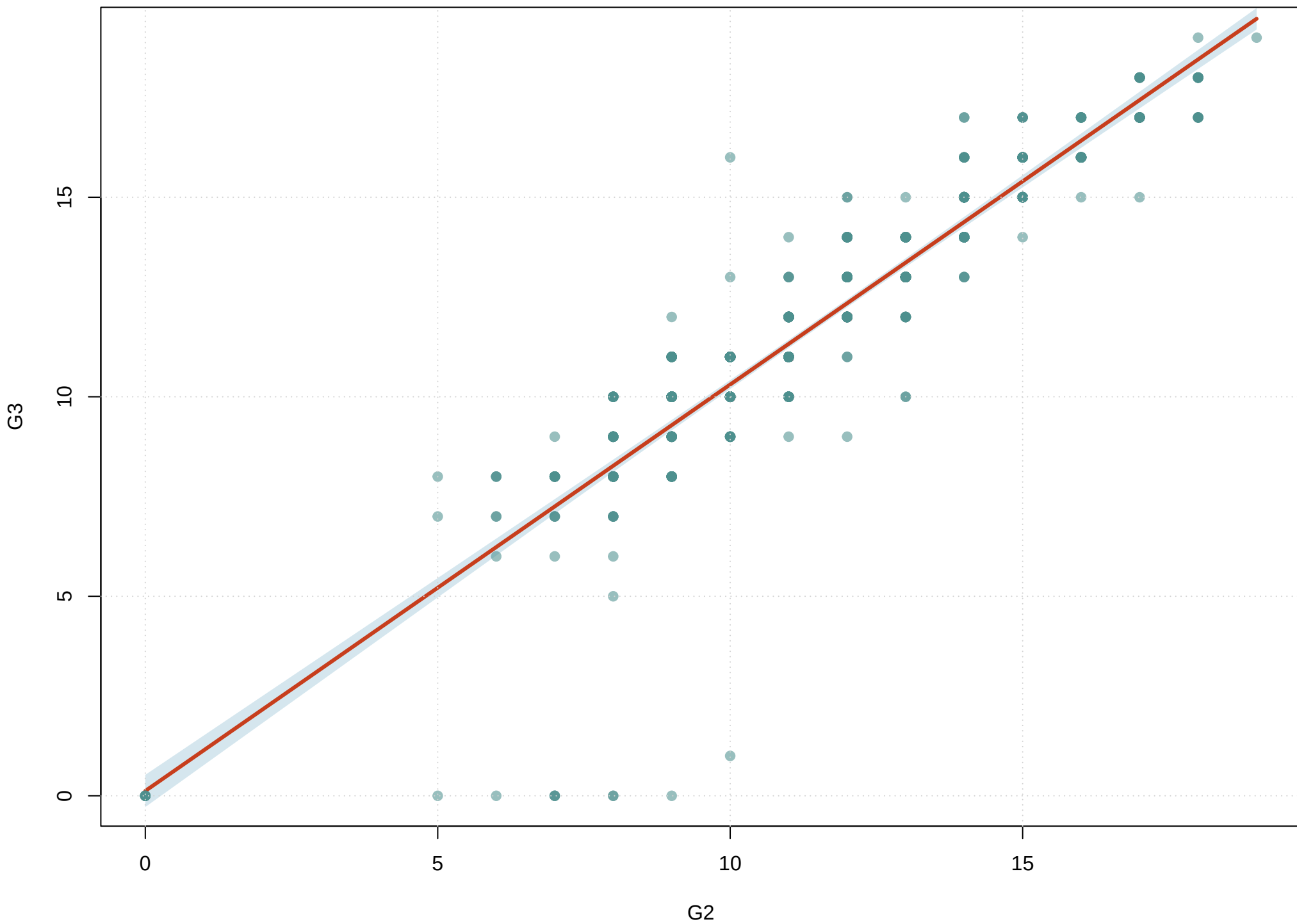
**Residual distribution**



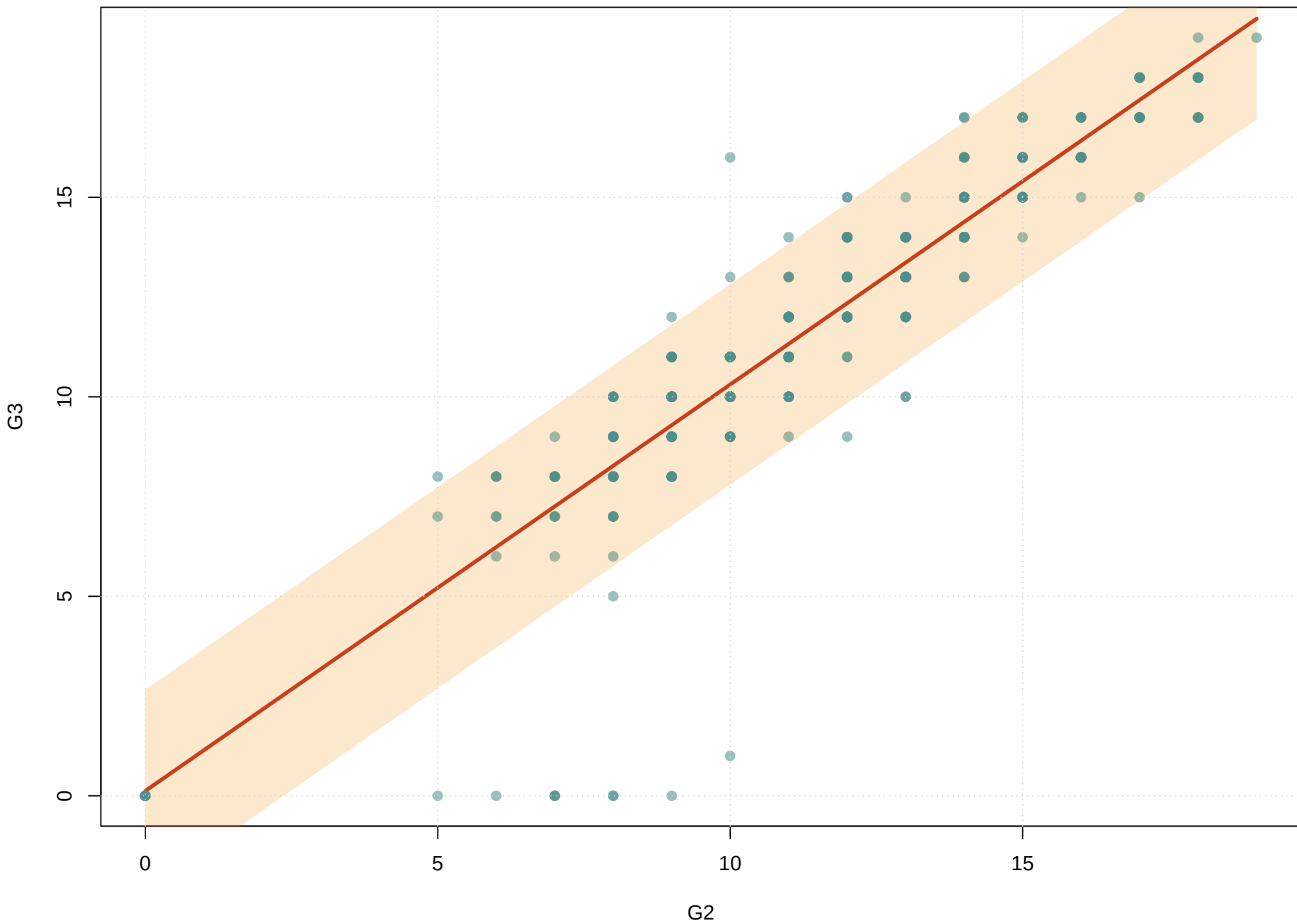
Normal Q-Q plot of residuals



Fitted line with 95% mean confidence band



Fitted line with 95% prediction band



Influence diagnostic

