Correlation Assumptions Report

X variable: G2
Y variable: G3

Correlation results:

method	x_variable	y_variable	n	correlation	p_value	ci95_lower_fisher_z	ci95_upper_fisher_z	assumption_note
Pearson correlation	G2	G3	649	0.9185480	5.642401e−263	0.9056062	0.9297808	Use when relationship is linear, variables are approximately continuous and normal, and influential outliers are not driving the result.
Spearman rank correlation	G2	G3	649	0.9444512	7.066377e−315	NA	NA	Use as a robust alternative when monotonic relationship is present but Pearson assumptions are doubtful.

Assumption decision summary:

assumption	status	Recommended reporting choice	Decision	evidence
Complete paired cases	Pass			n = 649
Approximate normality for Pearson	Flag			See normality table and Q−Q plots
Linearity / monotonic agreement	Pass			|Pearson − Spearman| = 0.0259
Constant residual spread	Flag			Breusch−Pagan p = 7.195e−05
No influential bivariate outliers	Flag			Mahalanobis flags = 19 ; Cook flags = 20
		Pearson can be reported with caution, but Spearman should also be reported because at least one assumption needs attention.		

Normality diagnostics:

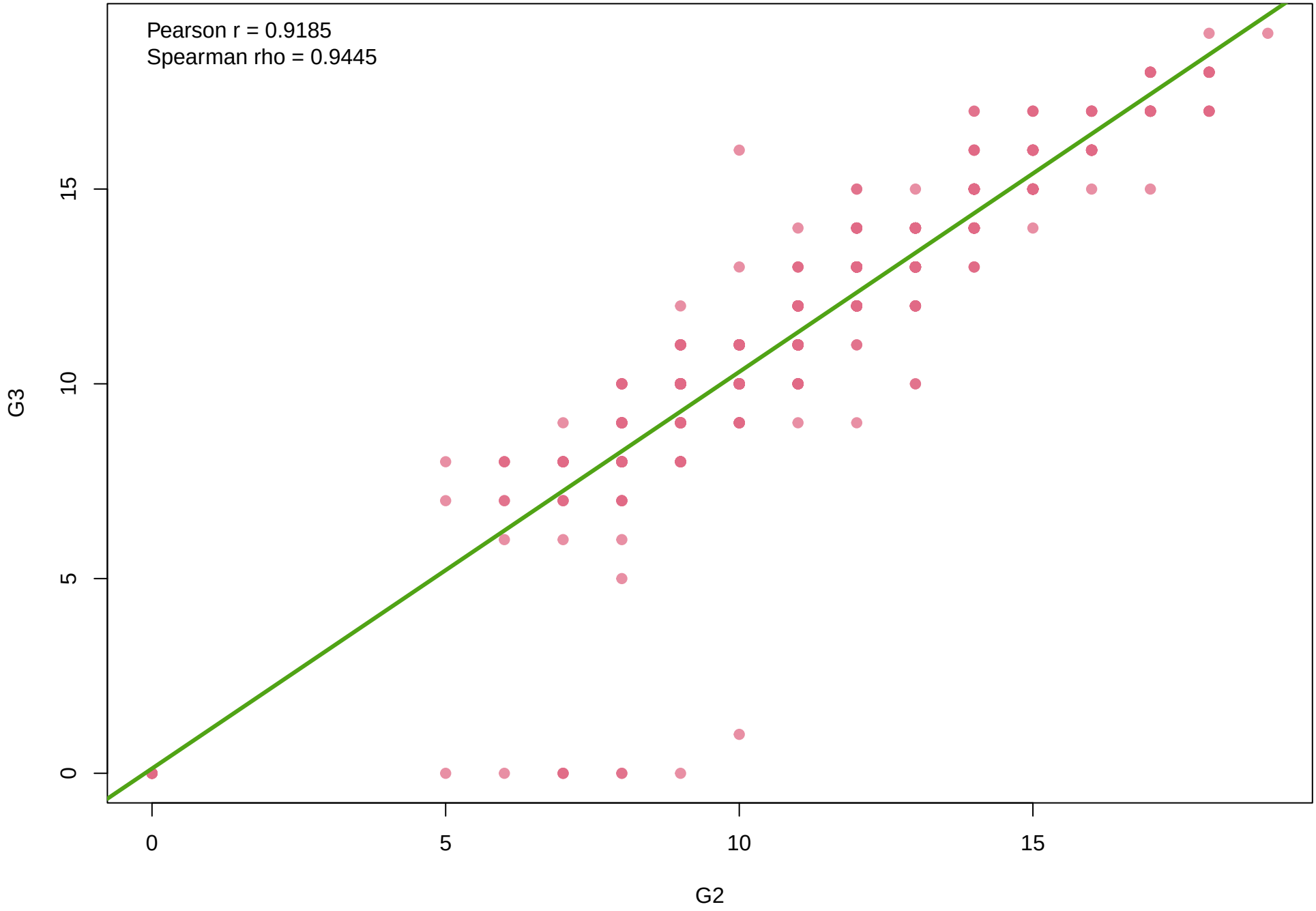
variable	n	mean	standard_deviation	minimum	maximum	skewness	excess_kurtosis	normality_method	normality_statistic	normality_p_value	normality_decision_alpha_0_05
G2	649	11.57011	2.913639	0	19	−0.3602826	1.662465	Shapiro−Wilk normality test	0.9616671	5.583292e−12	Possible non−normality
G3	649	11.90601	3.230656	0	19	−0.9129094	2.712204	Shapiro−Wilk normality test	0.9259809	2.415986e−17	Possible non−normality

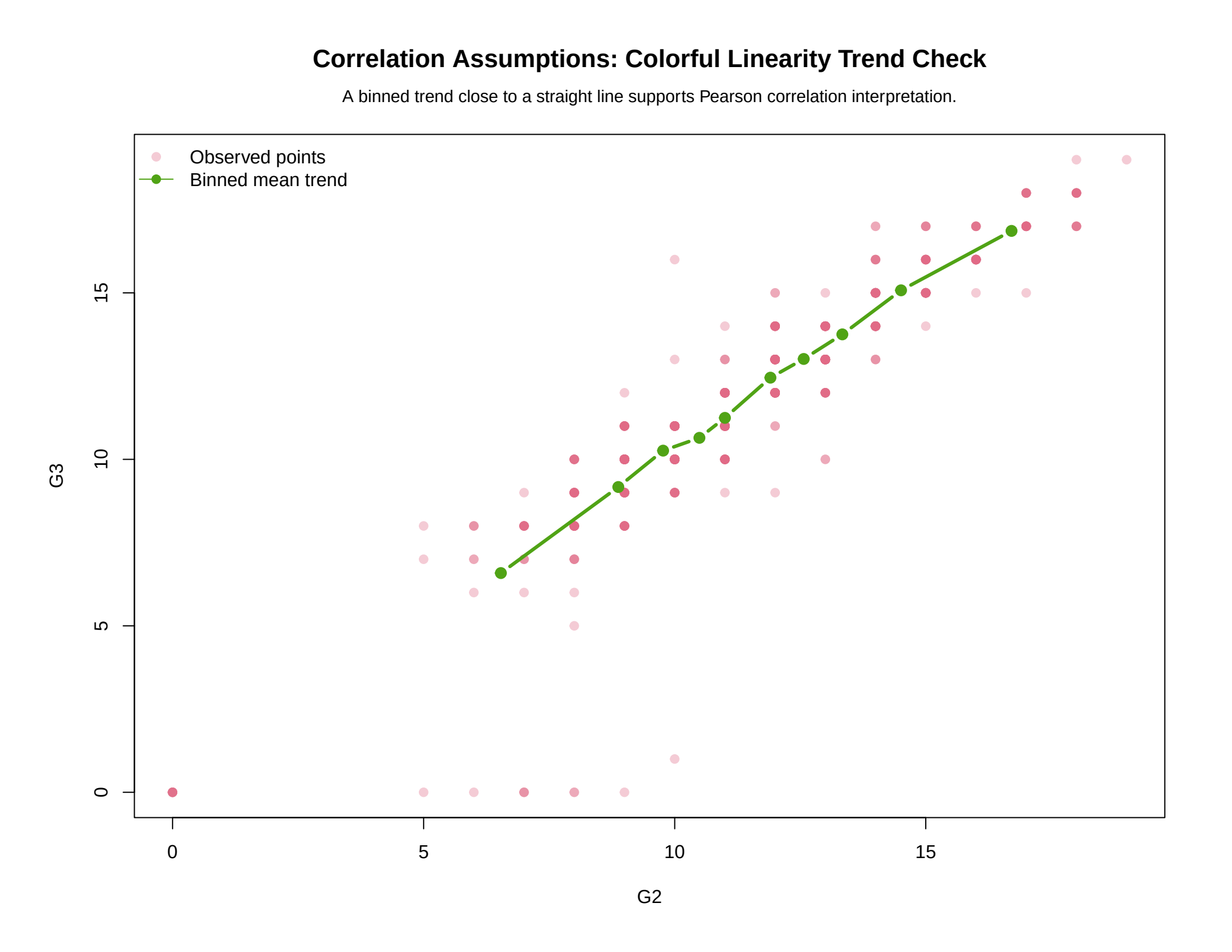
Linearity, residual, and outlier diagnostics:

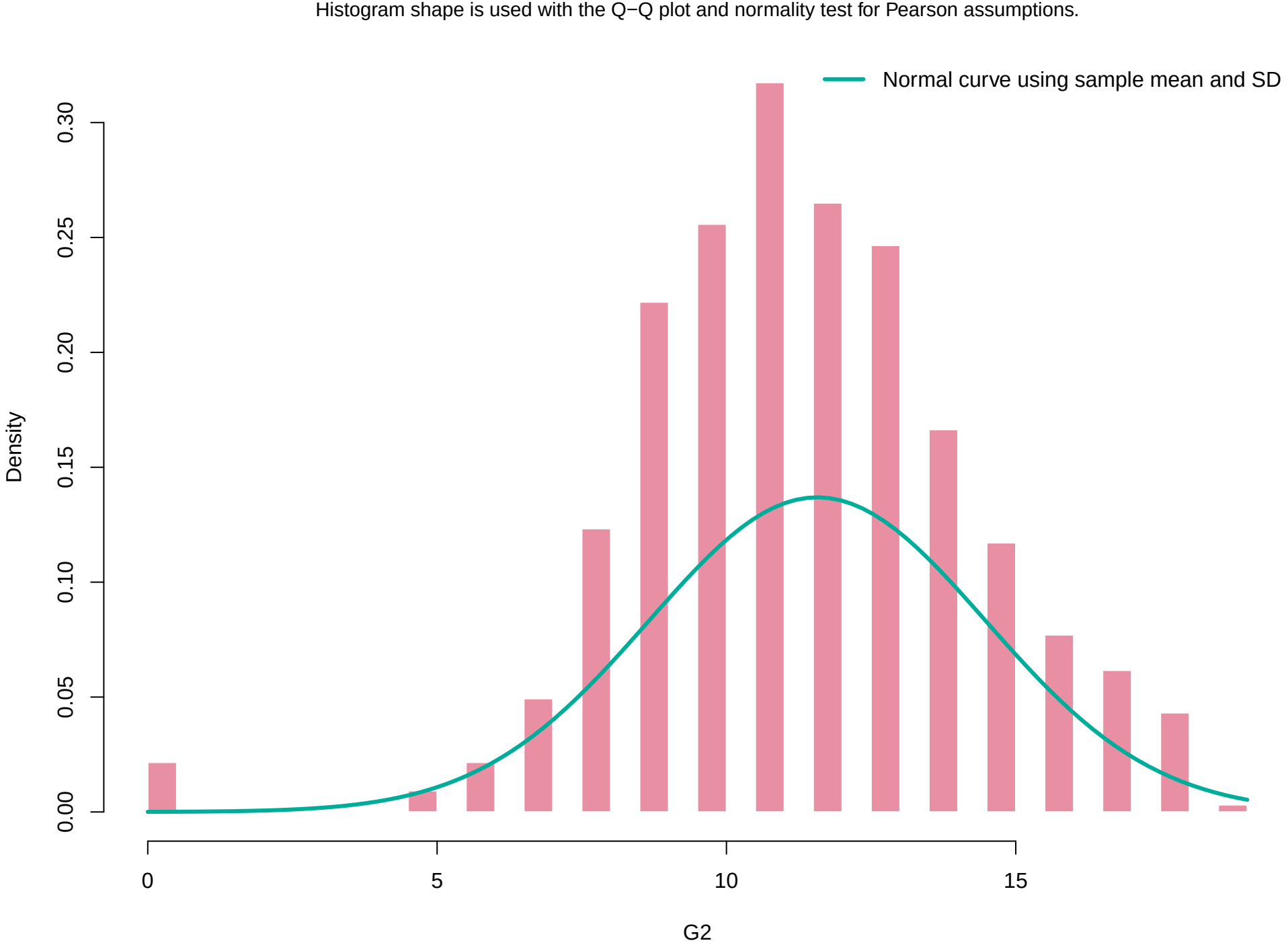
n_complete_pairs	pearson_r	pearson_p_value	spearman_rho	spearman_p_value	absolute_difference_pearson_spearman	regression_intercept	regression_slope	rmse	breusch_pagan_statistic	breusch_pagan_p_value	homoscedasticity_decision_alpha_0_05	mahalanobis_critical_chi_square_975_df2	mahalanobis_outlier_count_975	cooks_distance_threshold_4_over_n	cooks_distance_flag_count
649	0.918548	5.642401e−263	0.9444512	7.066377e−315	0.02590318	0.1219661	1.01849	1.276125	15.75877	7.195376e−05	Possible heteroscedasticity	7.377759	19	0.006163328	20

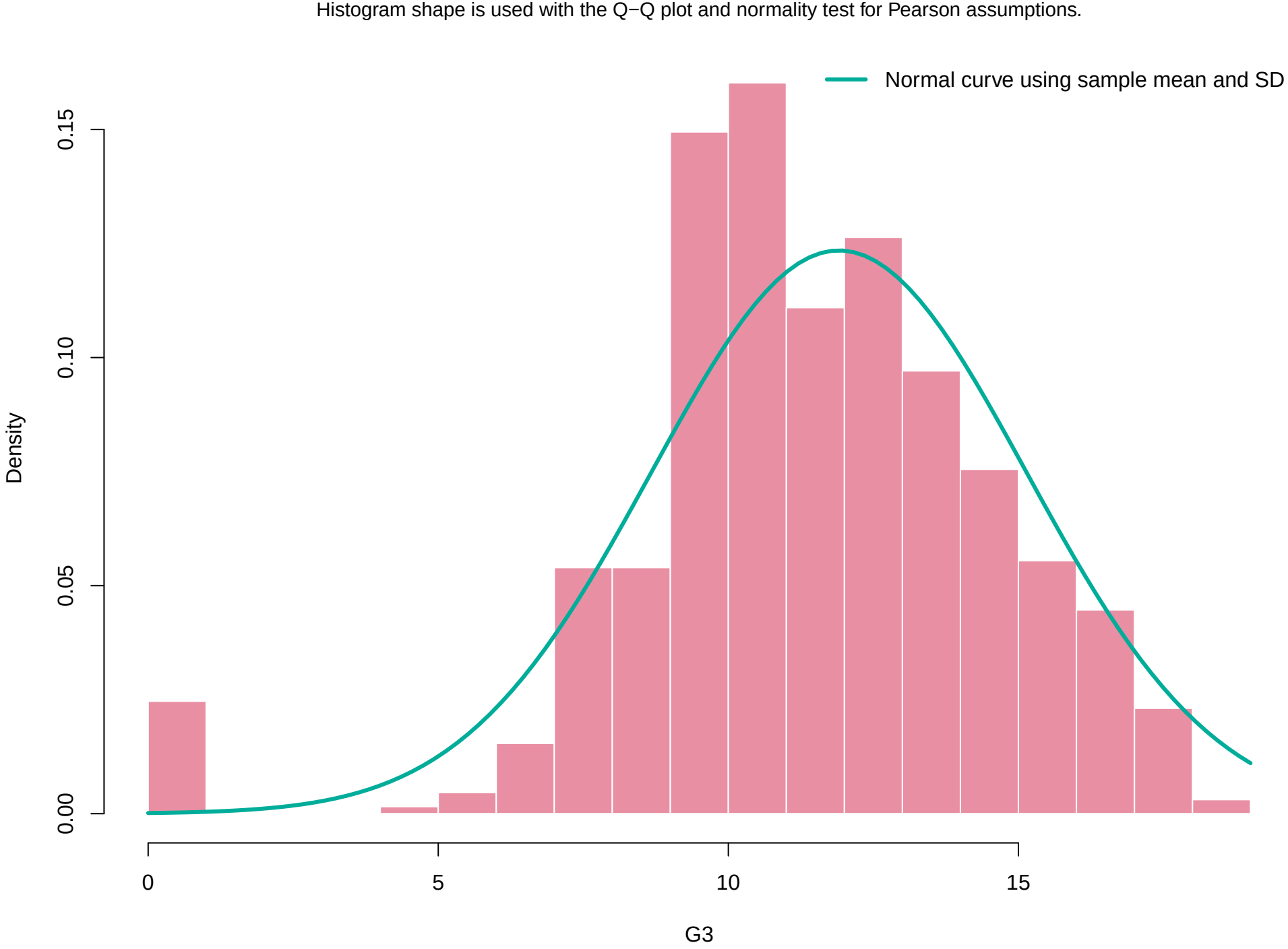
Correlation Assumptions: Scatterplot with Linear Fit

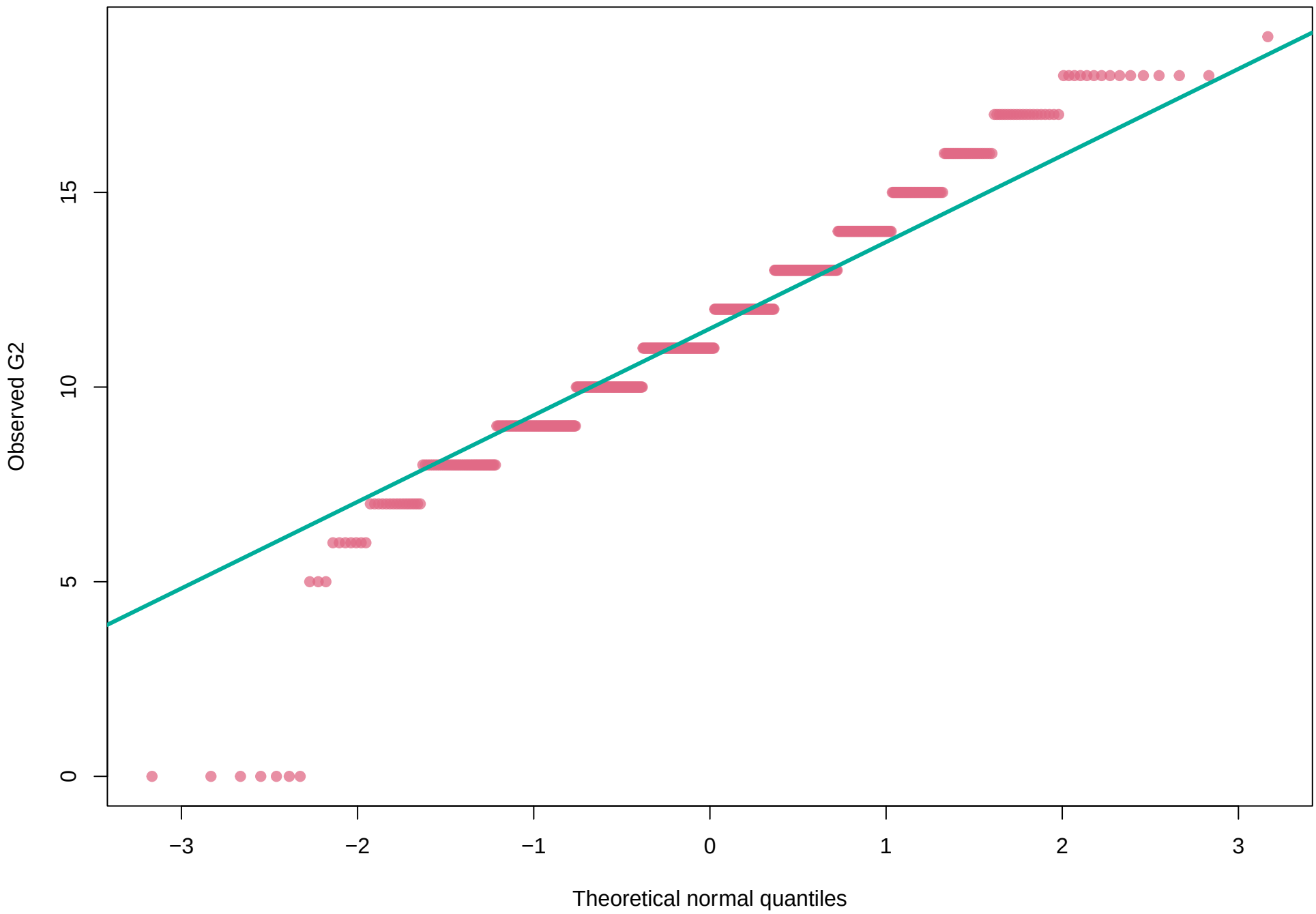
The main Pearson assumption is a reasonably linear relationship without extreme influential points.

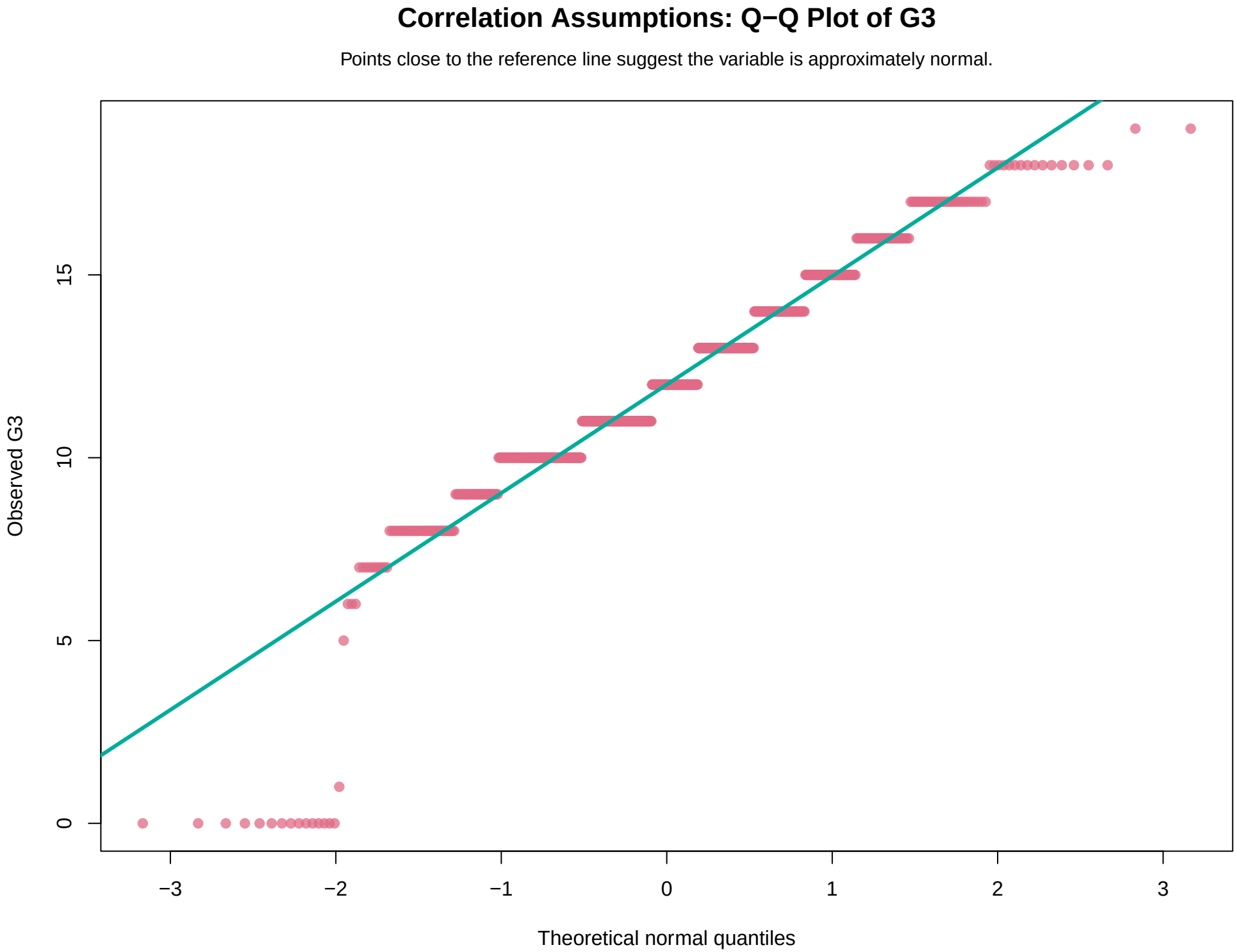


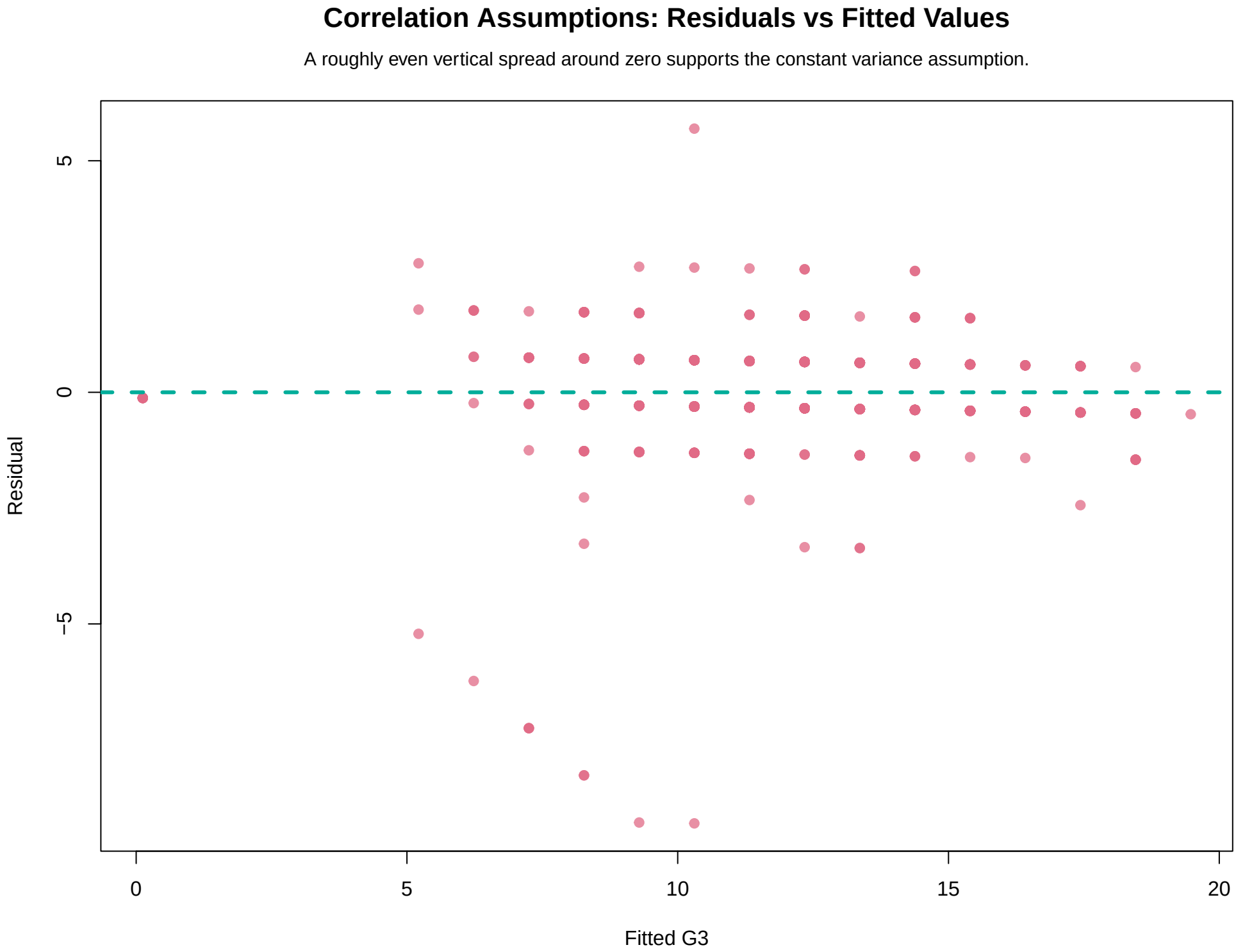

Correlation Assumptions: Colorful Linearity Trend Check
A binned trend close to a straight line supports Pearson correlation interpretation.
Observed points
Binned mean trend
G3
G2
0
5
10
15

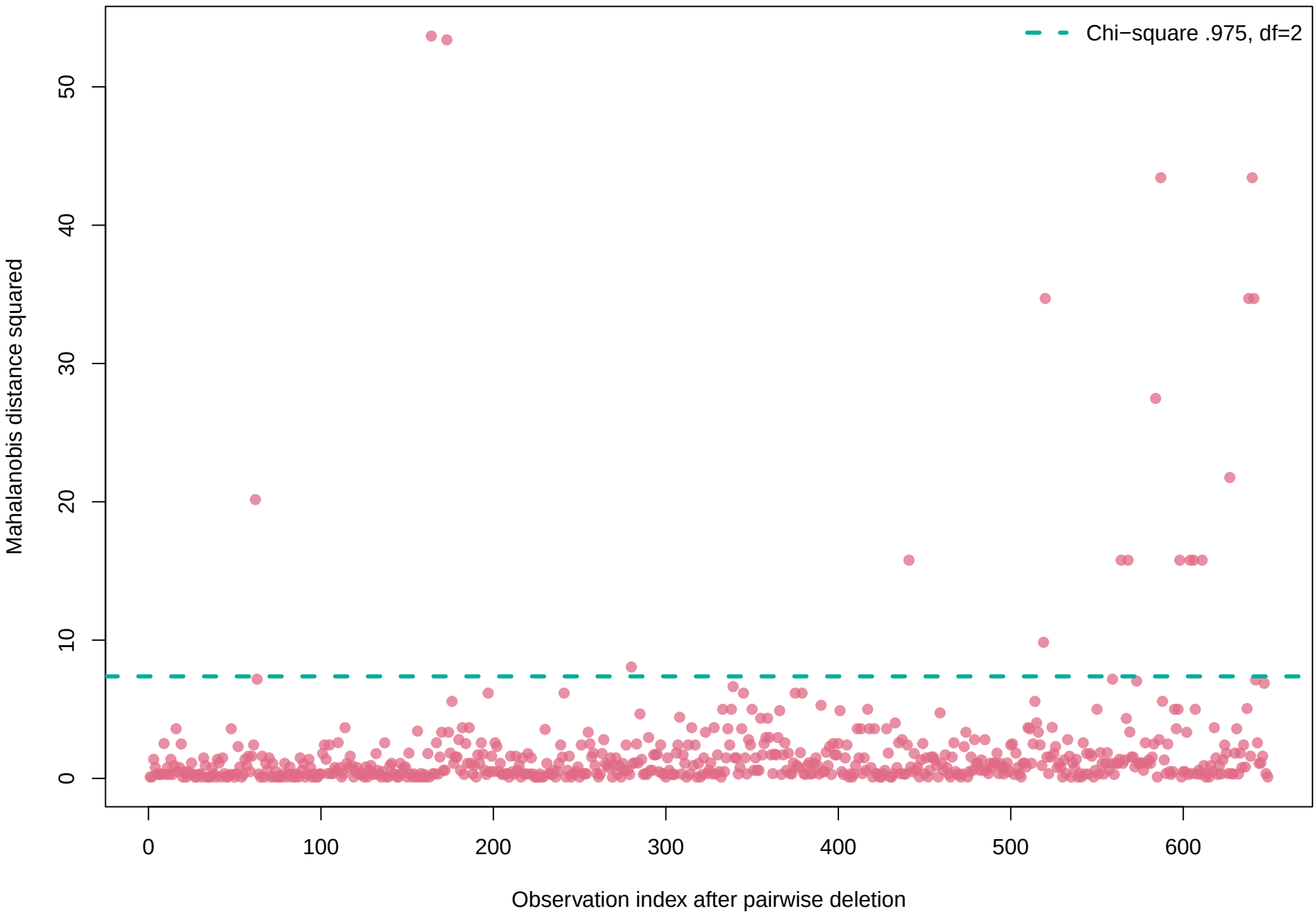

Correlation Assumptions: Histogram of G2
Histogram shape is used with the Q−Q plot and normality test for Pearson assumptions.
Normal curve using sample mean and SD
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
5
10
15
G2


Correlation Assumptions: Histogram of G3
Histogram shape is used with the Q−Q plot and normality test for Pearson assumptions.
Normal curve using sample mean and SD
Density
0.00
0.05
0.10
0.15
0
5
10
15
G3


Correlation Assumptions: Q−Q Plot of G2
Points close to the reference line suggest the variable is approximately normal.
Observed G2
0
5
10
15
−3
−2
−1
0
1
2
3
Theoretical normal quantiles


Correlation Assumptions: Q−Q Plot of G3
Points close to the reference line suggest the variable is approximately normal.
Observed G3
0
5
10
15
−3
−2
−1
0
1
2
3
Theoretical normal quantiles


Correlation Assumptions: Residuals vs Fitted Values
A roughly even vertical spread around zero supports the constant variance assumption.
Residual
5
0
−5
0
5
10
15
20
Fitted G3


Correlation Assumptions: Colorful Bivariate Outlier Check
Large Mahalanobis distances or high Cook's distance can show points that drive the correlation.
Chi−square .975, df=2
Mahalanobis distance squared
0
10
20
30
40
50
0
100
200
300
400
500
600
Observation index after pairwise deletion

Correlation Assumptions: Colorful Numeric Correlation Matrix

The selected pair is G2 with G3 ; nearby variables show context.