GLM Section

Nathalie Blasco

2025-11-19

Results

Generalized Linear Model (GLM)

A generalized linear model (GLM model) using a Gaussian link was fit including week, site, treatment, age, and sex as predictors (Table 2).

Table 2: GLM Model Summary (Gaussian Link)

term	estimate	std.error	statistic	p.value
(Intercept)	42.059	3.160	13.310	0.000
week	0.235	0.089	2.635	0.009
treat	0.148	0.626	0.236	0.813
age	0.018	0.042	0.440	0.660
sex	-2.066	1.059	-1.950	0.052
site	-0.135	0.190	-0.711	0.478

The coefficient for week was statistically significant, indicating an overall linear trend in TWSTRS score over time. However, because the GLM assumes independence of observations, the repeated measures within individuals violate this assumption. As a result, the standard errors and p-values may be underestimated, and inference should be interpreted with caution. This motivates the subsequent use of correlation-aware models such as GEE and GLMM.

GLM Model Diagnostics

Diagnostics were assessed for the GLM model.

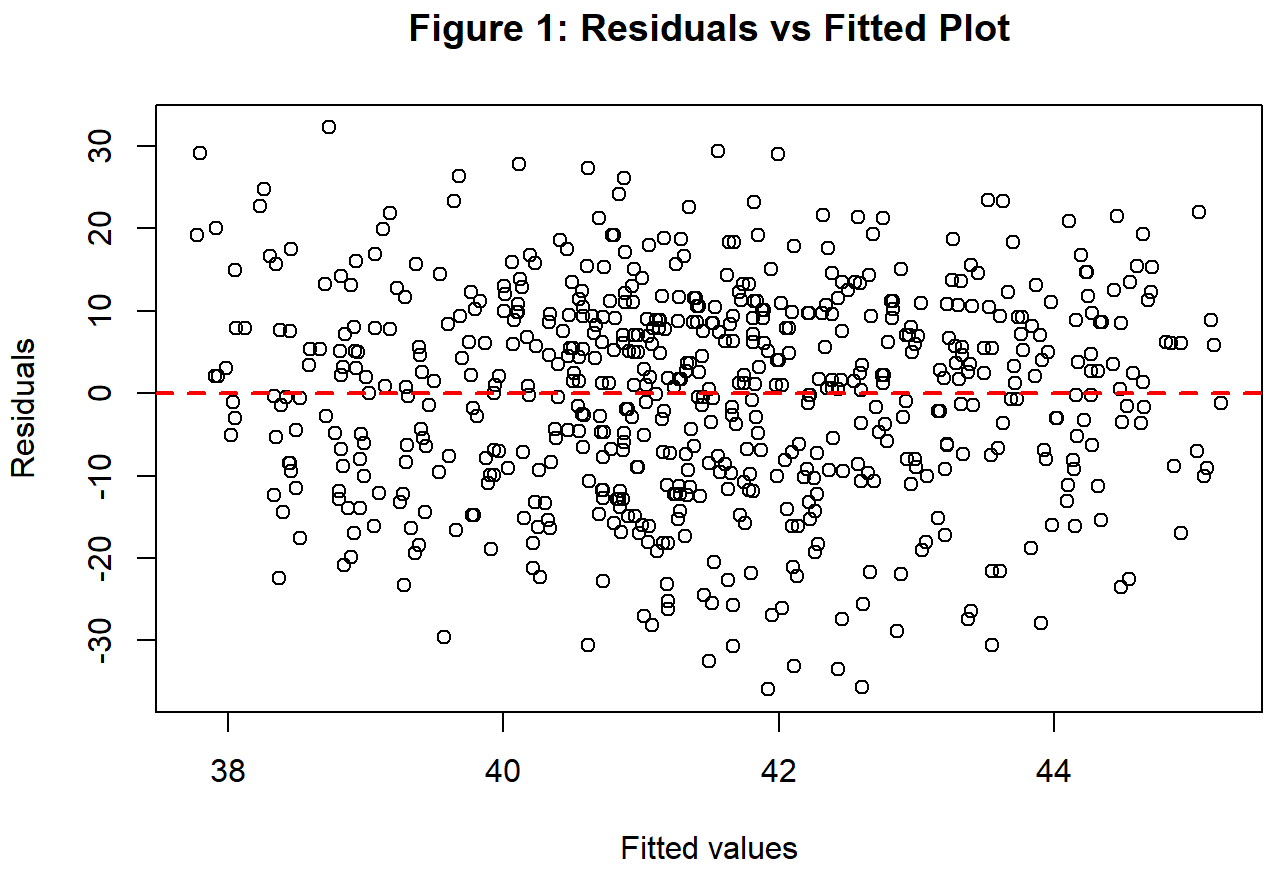


Figure 1 shows the residuals versus the fitted values. The absence of strong patterns or fanning suggests that the linearity and homoscedasticity assumptions are reasonably met.

Figure 2: Q-Q Plot

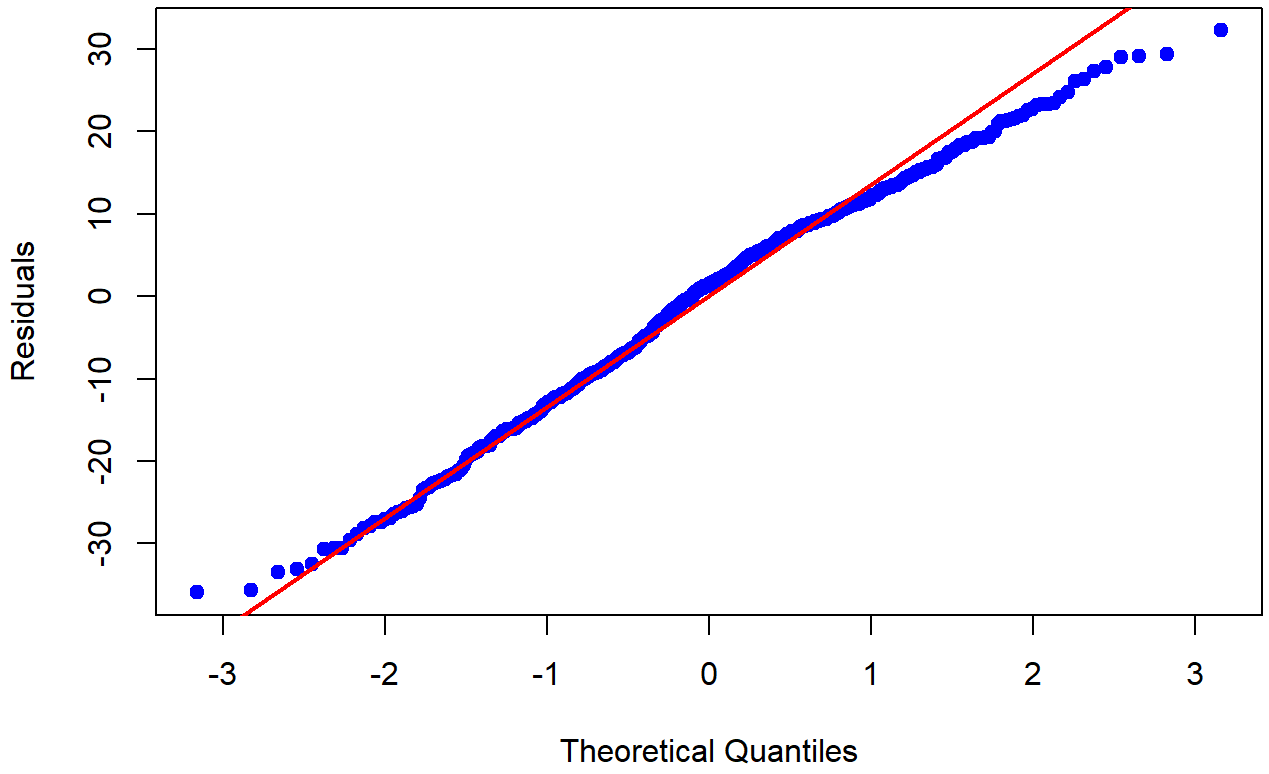
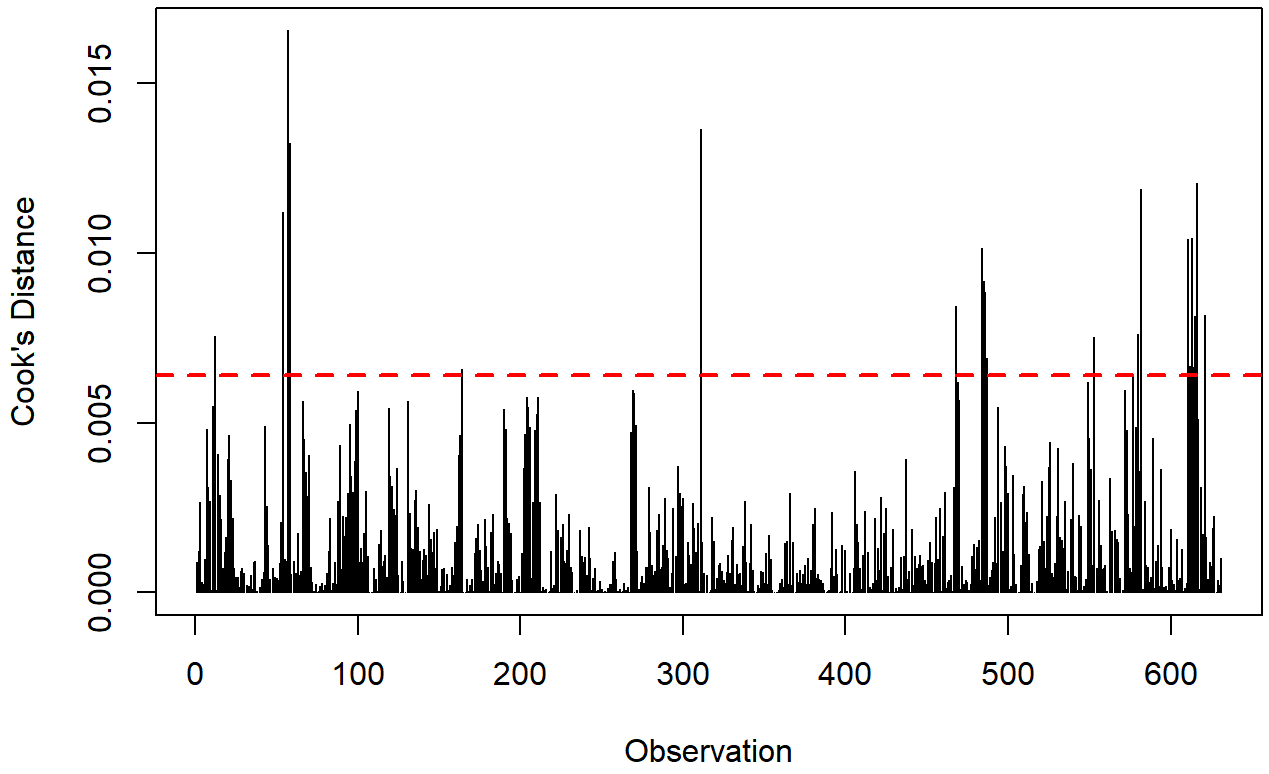


Figure 2 presents a QQ plot of the residuals, which largely follow the 45 degree reference line with mild deviations in the tails. This indicates that the normality assumption is approximately satisfied.

Figure 3: Cook's Distance Plot



Finally, figure 3 displays Cook's distance for all observations. The conventional threshold, 4/(n-k-1), where n is the number of observations and k is the number of predictors, was used to assess the influence of the observations. Several observations exceeded the threshold and were considered potentially influential and warranted further investigation. Therefore, a second GLM was fit, excluding those observations.

Table 3: GLM Model Summary (Excluding Influential Observations)

term	estimate	std.error	statistic	p.value
(Intercept)	41.350	2.978	13.884	0.000
week	0.280	0.085	3.306	0.001
treat	0.566	0.587	0.964	0.335
age	0.052	0.040	1.296	0.195
sex	-2.956	1.002	-2.950	0.003
site	-0.342	0.182	-1.881	0.060

Excluding influential observations resulted in modest shifts in the parameter estimates and slightly improved precision. Notably, the effects of sex and week became stronger and statistically significant. The direction of all effects remained consistent with the original model, suggesting that the influential observations primarily affected precision rather than the overall patterns of association.

Overall, while the GLM provides a reasonable descriptive summary, its assumptions are not appropriate for longitudinal data with correlated repeated measures. The following GEE and GLMM analyses address this limitation by explicitly modeling within-subject correlation and random effects.