

Negative Binomial Regression

Test name: Negative Binomial Regression

Dataset: D:\DATA ANALYSIS\H Regression Tests and Models\Negative Binomial Regression\dataset.csv

Rows in original dataset: 649

Columns in original dataset: 33

Rows used in negative binomial model: 649

Count outcome created: nb_count = round(G3) clipped at zero

Numeric predictors: G1, G2, studytime, failures, absences, age

Categorical predictors: school, sex, address

Formula: nb_count ~ G1 + G2 + studytime + failures + absences + age + C(school) + C(sex) + C(address)

Family/link: Negative Binomial count model with log link.

Poisson Pearson dispersion estimate: 0.2431

Negative binomial alpha: 0.00000001

Poisson AIC: 3009.7990

Negative binomial AIC: 3009.7992

Poisson dispersion is below 0.75, so overdispersion is not evident in this fitted model.

Interpretation: coefficients are log-count effects; exp(coefficient) is the incidence rate ratio.

Decision guide: $p < 0.05$ and IRR CI not crossing 1 indicates stronger evidence for a predictor effect.

Reporting guide: state count outcome, link, predictors, alpha/dispersion, IRRs, 95% CIs, and model comparison.

Charts generated: 8

Negative Binomial coefficients

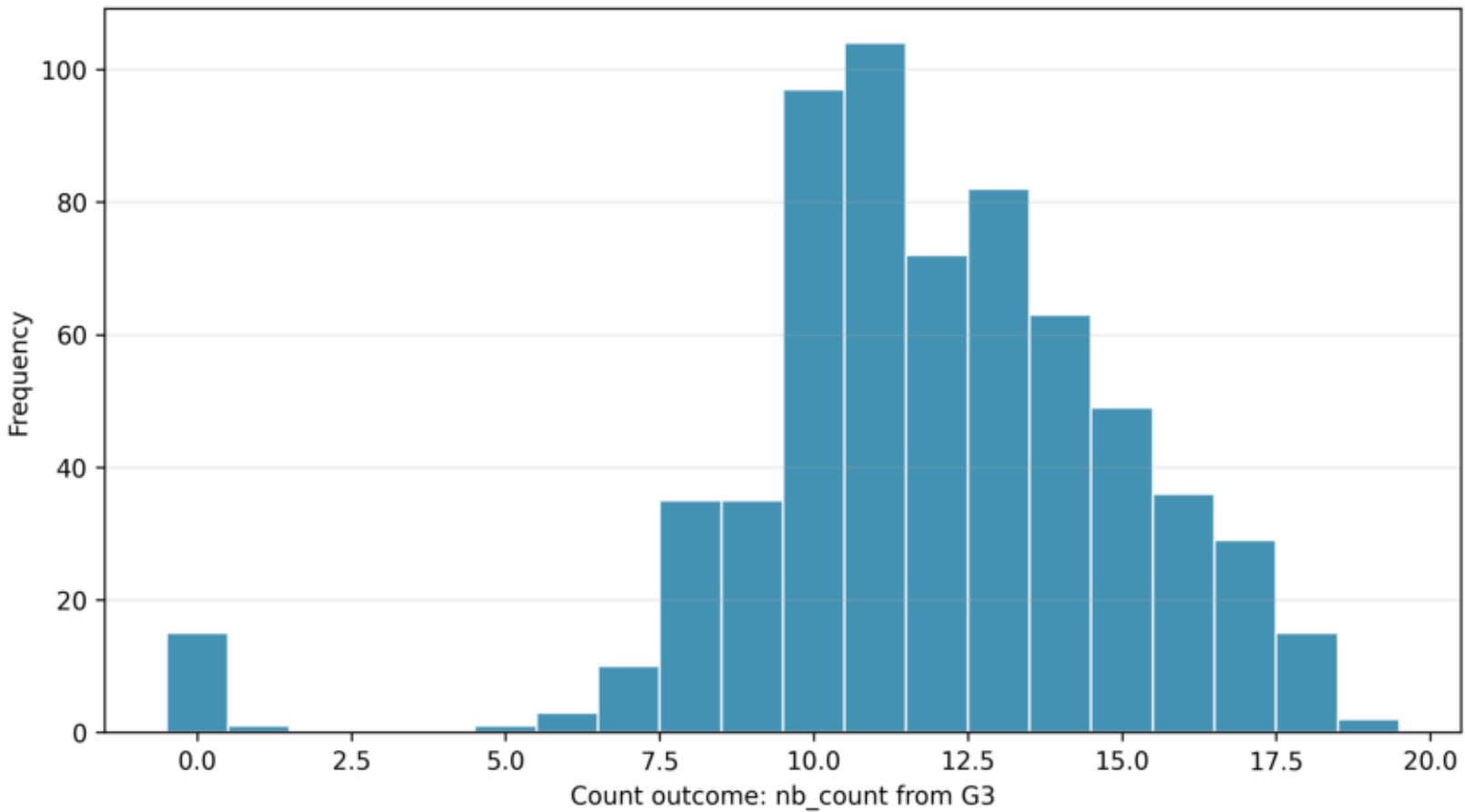
model	term	log_count_co	standard_err	z_value	p_value	ci95_lower_lo	ci95_upper_lo	incidence_rat	irr_ci95_lowe	irr_ci95_uppe	decision_alpha
GLM Negative	Intercept	1.5753	0.1798	8.7616	0.0000	1.2229	1.9276	4.8320	3.3969	6.8733	Statistically sig
GLM Negative	C(school)[T.M	-0.0334	0.0277	-1.2051	0.2282	-0.0877	0.0209	0.9672	0.9161	1.0211	Not statistically
GLM Negative	C(sex)[T.M]	-0.0168	0.0241	-0.6951	0.4870	-0.0641	0.0305	0.9834	0.9379	1.0310	Not statistically
GLM Negative	C(address)[T	0.0084	0.0271	0.3097	0.7568	-0.0447	0.0615	1.0084	0.9563	1.0634	Not statistically
GLM Negative	G1	-0.0022	0.0091	-0.2410	0.8095	-0.0199	0.0156	0.9978	0.9803	1.0157	Not statistically
GLM Negative	G2	0.0864	0.0088	9.7846	0.0000	0.0691	0.1037	1.0902	1.0715	1.1092	Statistically sig
GLM Negative	studytime	0.0027	0.0145	0.1867	0.8519	-0.0258	0.0312	1.0027	0.9745	1.0317	Not statistically
GLM Negative	failures	-0.0406	0.0255	-1.5934	0.1111	-0.0905	0.0093	0.9602	0.9135	1.0094	Not statistically
GLM Negative	absences	0.0027	0.0026	1.0164	0.3095	-0.0025	0.0078	1.0027	0.9975	1.0078	Not statistically
GLM Negative	age	-0.0059	0.0103	-0.5768	0.5641	-0.0261	0.0142	0.9941	0.9742	1.0143	Not statistically

Poisson vs Negative Binomial fit statistics

model	family	link	n	df_model	df_resid	log_likelihod	aic	bic	pearson_chi	poisson_disp	negative_binom	note
Standard Po	Poisson	log	649.0000	9.0000	639.0000	-1494.8995	3009.7990	-3890.7360	155.3659	0.24313905		Negative Bino
GLM Negativ	Negative Bin	log	649.0000	9.0000	639.0000	-1494.8996	3009.7992	-3890.7361	155.3658		1e-08	Negative Bino

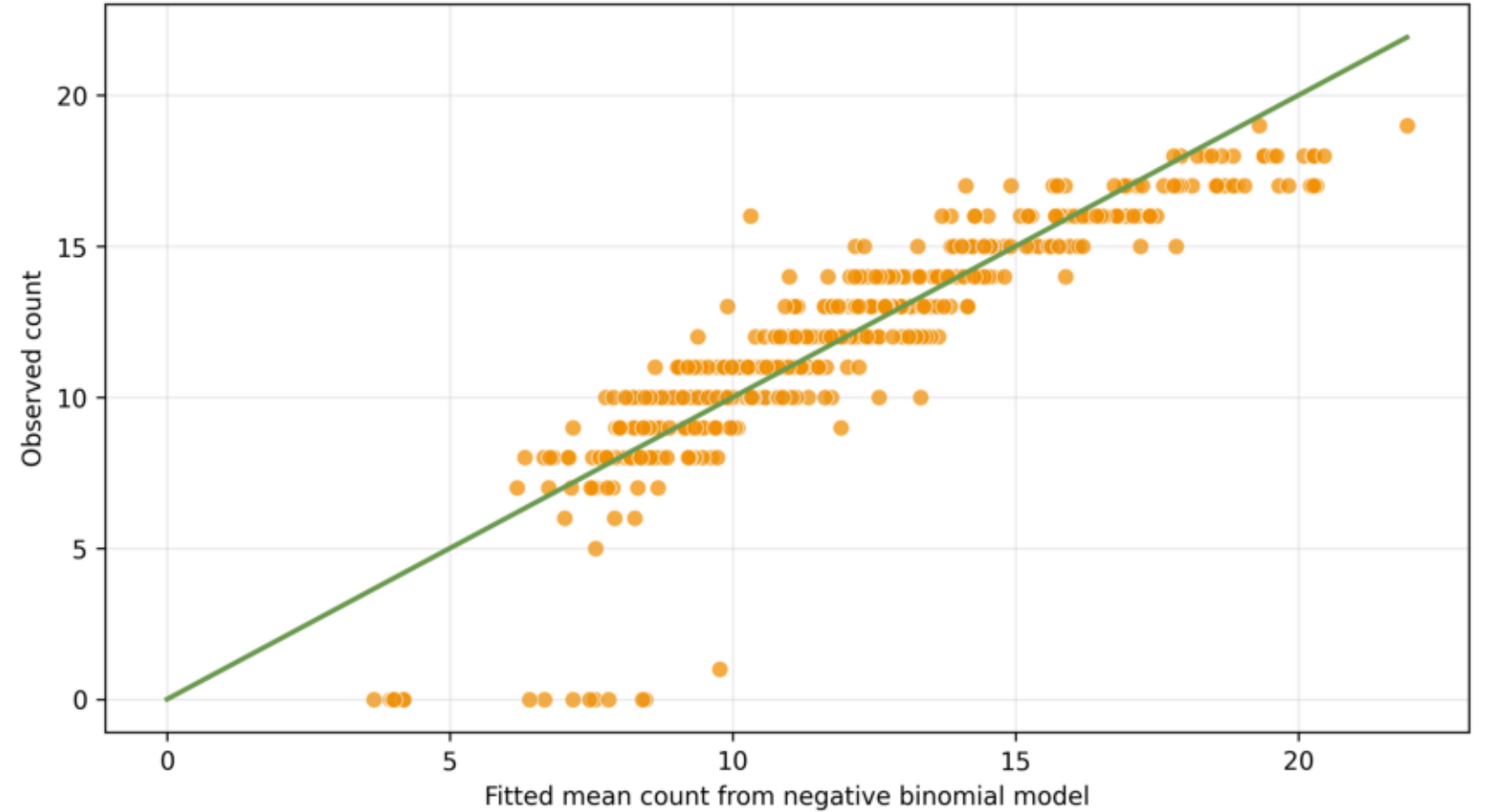
Negative binomial count outcome distribution

The rounded non-negative outcome is modeled as count data.



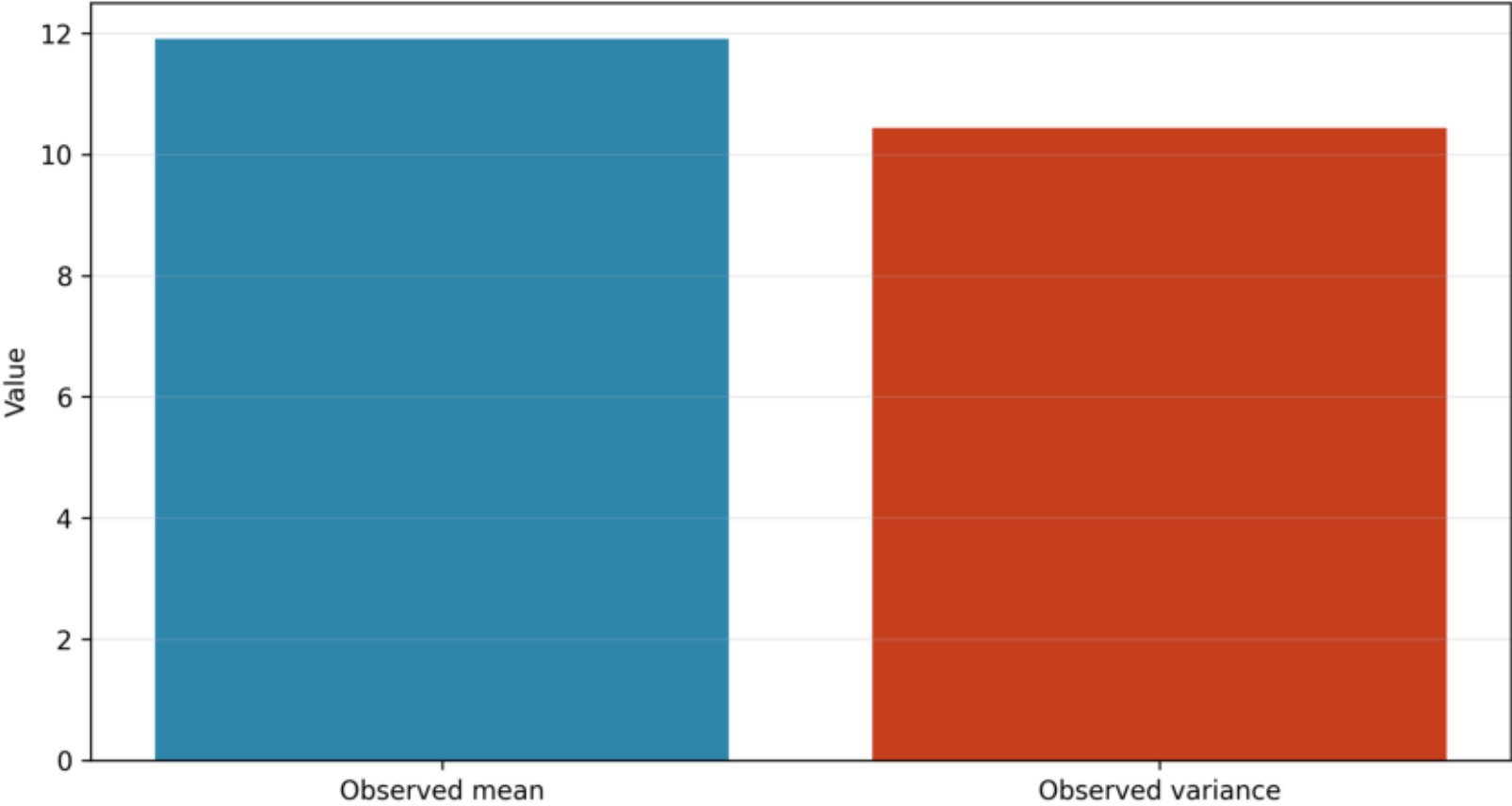
Observed vs fitted count values

Points near the diagonal indicate stronger agreement between observed and fitted counts.



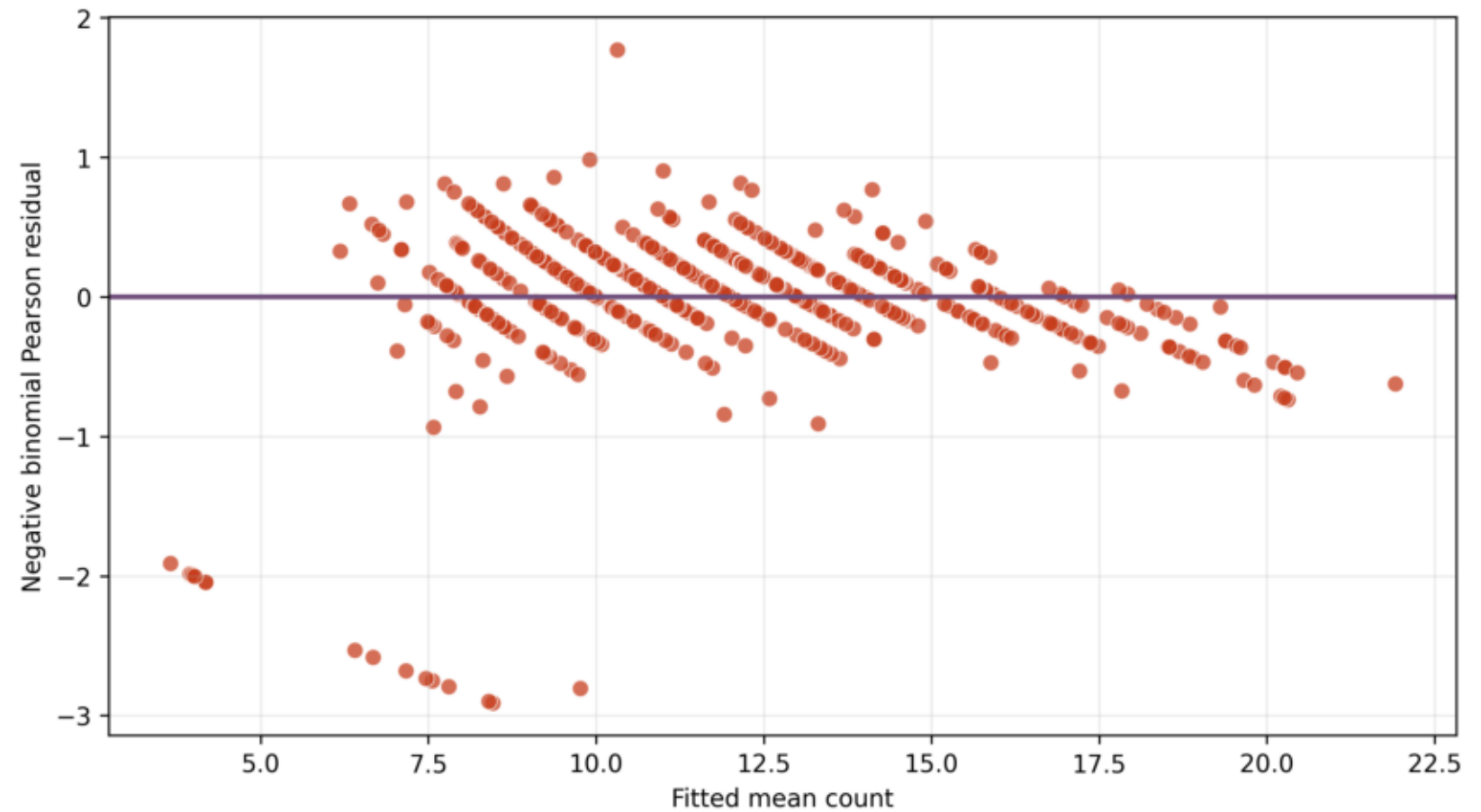
Count mean and variance check

Negative binomial regression is useful when count variance exceeds the Poisson mean assumption.



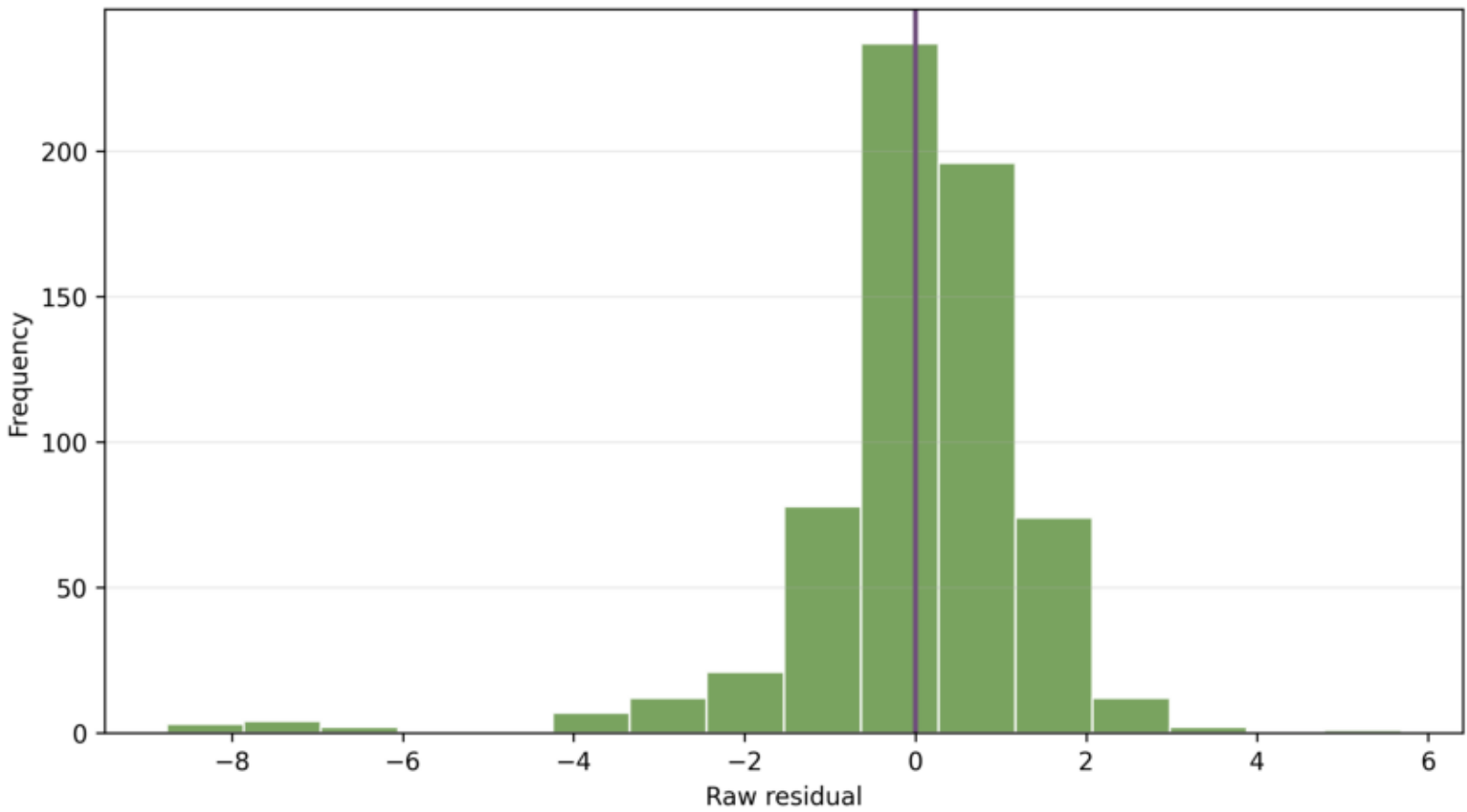
Pearson residuals vs fitted values

Large structure in residuals can indicate remaining lack of fit.



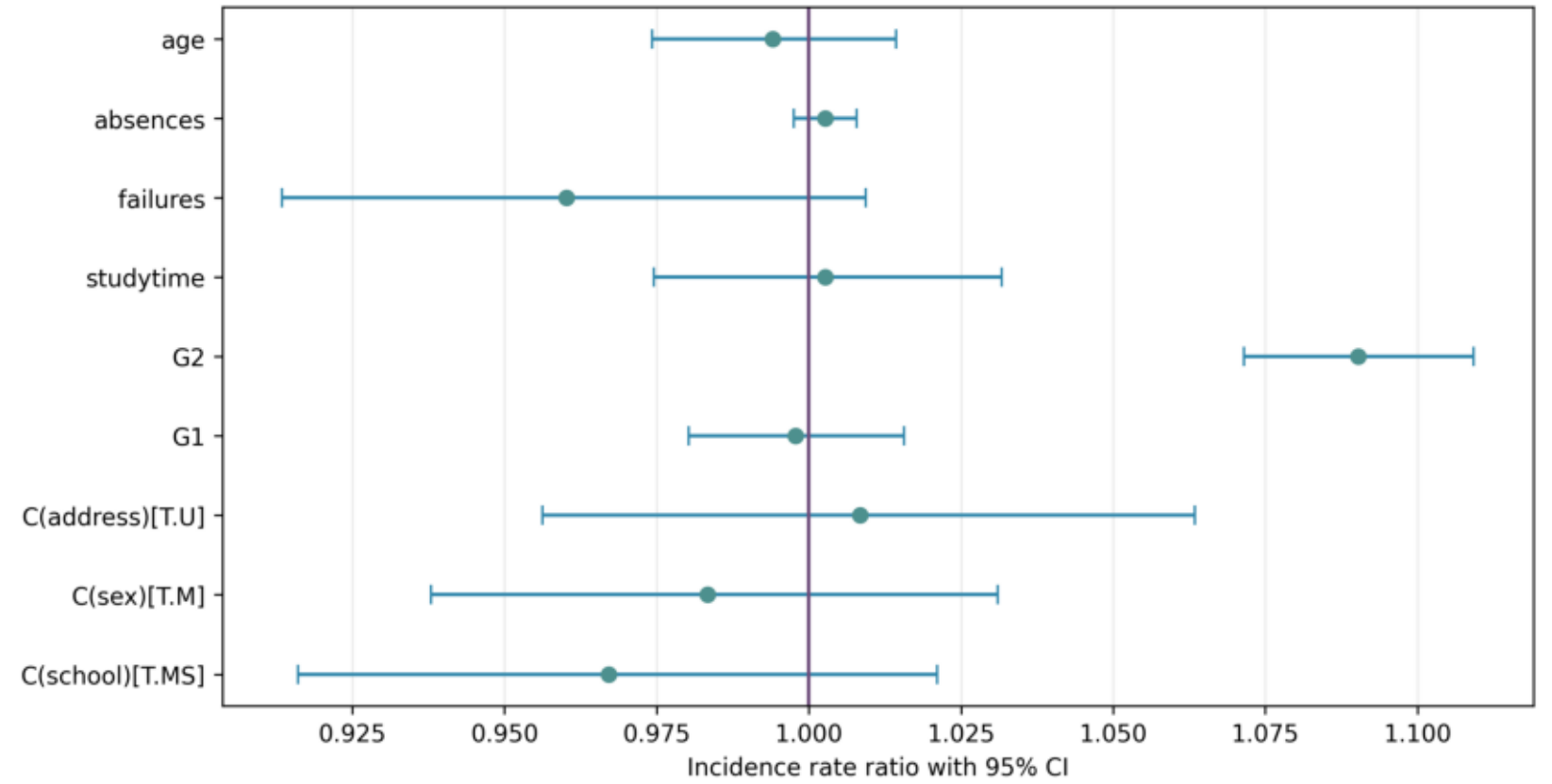
Negative binomial residual distribution

Residuals summarize observed count minus fitted mean count.



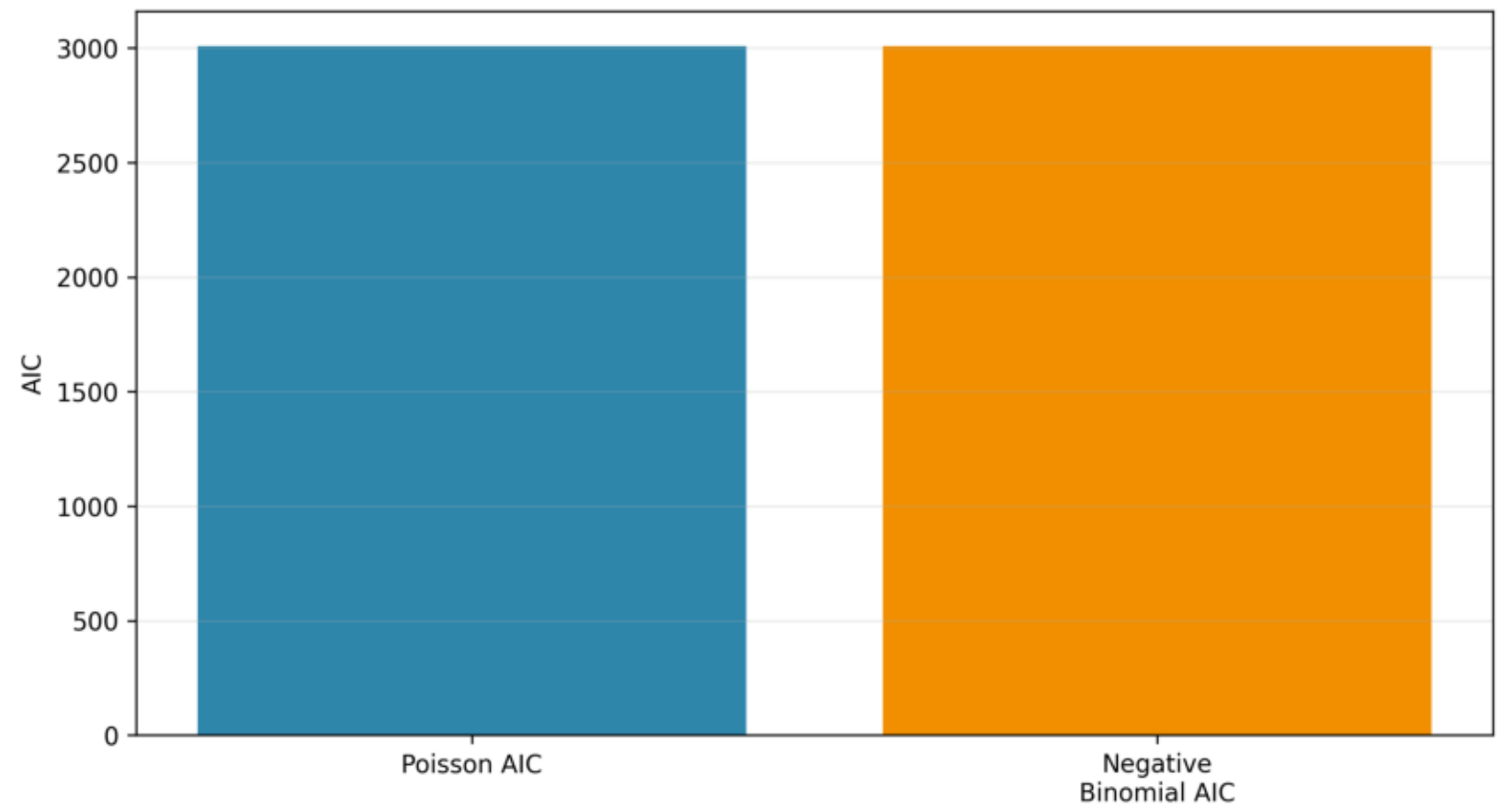
Negative binomial incidence rate ratio plot

Intervals not crossing 1 indicate stronger predictor evidence.



Poisson vs negative binomial model AIC

Lower AIC suggests better relative model fit when likelihoods are comparable.



Observed and fitted count means by G1

This chart compares grouped observed means against count-model fitted means.

