Hurdle Model

Test name: Hurdle Model

Dataset: D:\DATA ANALYSIS\H Regression Tests and Models\Hurdle Model\dataset.csv

Rows in original dataset: 649

Columns in original dataset: 33

Rows used in hurdle model: 649

Count outcome source: absences

Count outcome created: hurdle_count = round(absences) clipped at zero

Zero count observations: 244

Positive count observations: 405

Numeric predictors: G1, G2, G3, studytime, failures, age, Medu

Categorical predictors: school, sex, address

Hurdle formula: hurdle_positive ~ G1 + G2 + G3 + studytime + failures + age + Medu + C(school) + C(sex) + C(address)

Positive-count formula: hurdle_count ~ G1 + G2 + G3 + studytime + failures + age + Medu + C(school) + C(sex) + C(address)

Hurdle part: Binomial GLM with logit link for P(count > 0).

Positive-count part: Poisson GLM with log link fitted only on positive counts.

Expected count: predicted probability of positive count multiplied by predicted positive-count mean.

Positive-count Pearson dispersion: 3.0749

Positive-count dispersion differs from 1; consider robust or negative-binomial sensitivity reporting.

Interpretation: hurdle coefficients are odds-ratio effects; positive-count coefficients are rate-ratio effects.

Decision guide: p < 0.05 and confidence interval not crossing 1 indicates stronger evidence for a predictor effect.

Charts generated: 8

Hurdle logit coefficients

model	term	coefficient	standard_err	z_value	p_value	ci95_lower	ci95_upper	odds_ratio	odds_ratio_ci	odds_ratio_ci	decision_alpha_
Hurdle binary	Intercept	0.9936	1.3490	0.7365	0.4614	-1.6504	3.6376	2.7009	0.1920	37.9990	Not statistically
Hurdle binary	C(school)[T.M	-0.6276	0.2001	-3.1363	0.0017	-1.0198	-0.2354	0.5339	0.3607	0.7903	Statistically sig
Hurdle binary	C(sex)[T.M]	-0.2063	0.1789	-1.1531	0.2489	-0.5570	0.1444	0.8136	0.5729	1.1553	Not statistically
Hurdle binary	C(address)[T.	-0.0988	0.1977	-0.4998	0.6172	-0.4863	0.2887	0.9059	0.6149	1.3346	Not statistically
Hurdle binary	G1	-0.2425	0.0677	-3.5803	0.0003	-0.3753	-0.1098	0.7846	0.6871	0.8960	Statistically sig
Hurdle binary	G2	-0.1562	0.0878	-1.7795	0.0752	-0.3283	0.0158	0.8553	0.7201	1.0160	Not statistically
Hurdle binary	G3	0.2668	0.0742	3.5946	0.0003	0.1213	0.4123	1.3058	1.1290	1.5103	Statistically sig
Hurdle binary	studytime	-0.1657	0.1062	-1.5595	0.1189	-0.3739	0.0425	0.8473	0.6880	1.0435	Not statistically
Hurdle binary	failures	0.1250	0.1786	0.6999	0.4840	-0.2250	0.4750	1.1331	0.7985	1.6080	Not statistically
Hurdle binary	age	0.0841	0.0753	1.1176	0.2637	-0.0634	0.2316	1.0877	0.9386	1.2606	Not statistically
Hurdle binary	Medu	0.0806	0.0801	1.0067	0.3141	-0.0763	0.2376	1.0840	0.9265	1.2682	Not statistically

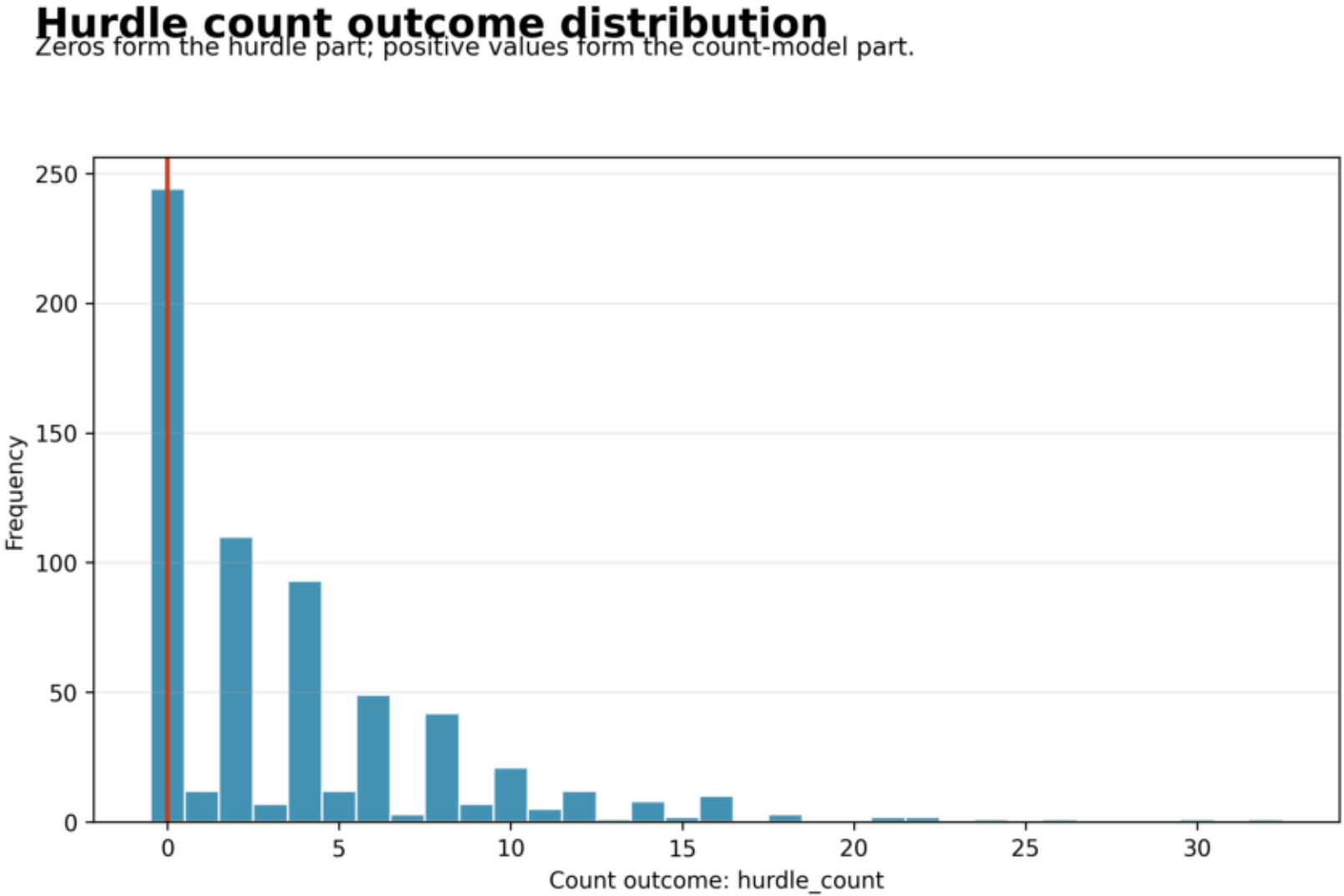
Positive-count coefficients

model	term	coefficient	standard_err	z_value	p_value	ci95_lower	ci95_upper	rate_ratio	rate_ratio_ci9	rate_ratio_ci9	decision_alpha_
Positive-coun	Intercept	0.8712	0.3262	2.6703	0.0076	0.2318	1.5106	2.3897	1.2608	4.5294	Statistically sig
Positive-coun	C(school)[T.M	-0.4265	0.0528	-8.0843	0.0000	-0.5299	-0.3231	0.6528	0.5887	0.7239	Statistically sig
Positive-coun	C(sex)[T.M]	-0.0064	0.0440	-0.1446	0.8850	-0.0925	0.0798	0.9937	0.9116	1.0831	Not statistically
Positive-coun	C(address)[T.	0.0831	0.0495	1.6783	0.0933	-0.0139	0.1802	1.0867	0.9861	1.1975	Not statistically
Positive-coun	G1	0.0205	0.0162	1.2636	0.2064	-0.0113	0.0524	1.0207	0.9888	1.0538	Not statistically
Positive-coun	G2	0.0016	0.0253	0.0635	0.9494	-0.0479	0.0511	1.0016	0.9532	1.0524	Not statistically
Positive-coun	G3	-0.0712	0.0243	-2.9316	0.0034	-0.1188	-0.0236	0.9313	0.8880	0.9767	Statistically sig
Positive-coun	studytime	-0.1022	0.0273	-3.7416	0.0002	-0.1557	-0.0486	0.9029	0.8558	0.9525	Statistically sig
Positive-coun	failures	-0.0045	0.0347	-0.1294	0.8970	-0.0726	0.0636	0.9955	0.9300	1.0657	Not statistically
Positive-coun	age	0.1050	0.0175	6.0055	0.0000	0.0708	0.1393	1.1108	1.0733	1.1495	Statistically sig
Positive-coun	Medu	-0.0141	0.0207	-0.6841	0.4939	-0.0547	0.0264	0.9860	0.9468	1.0267	Not statistically

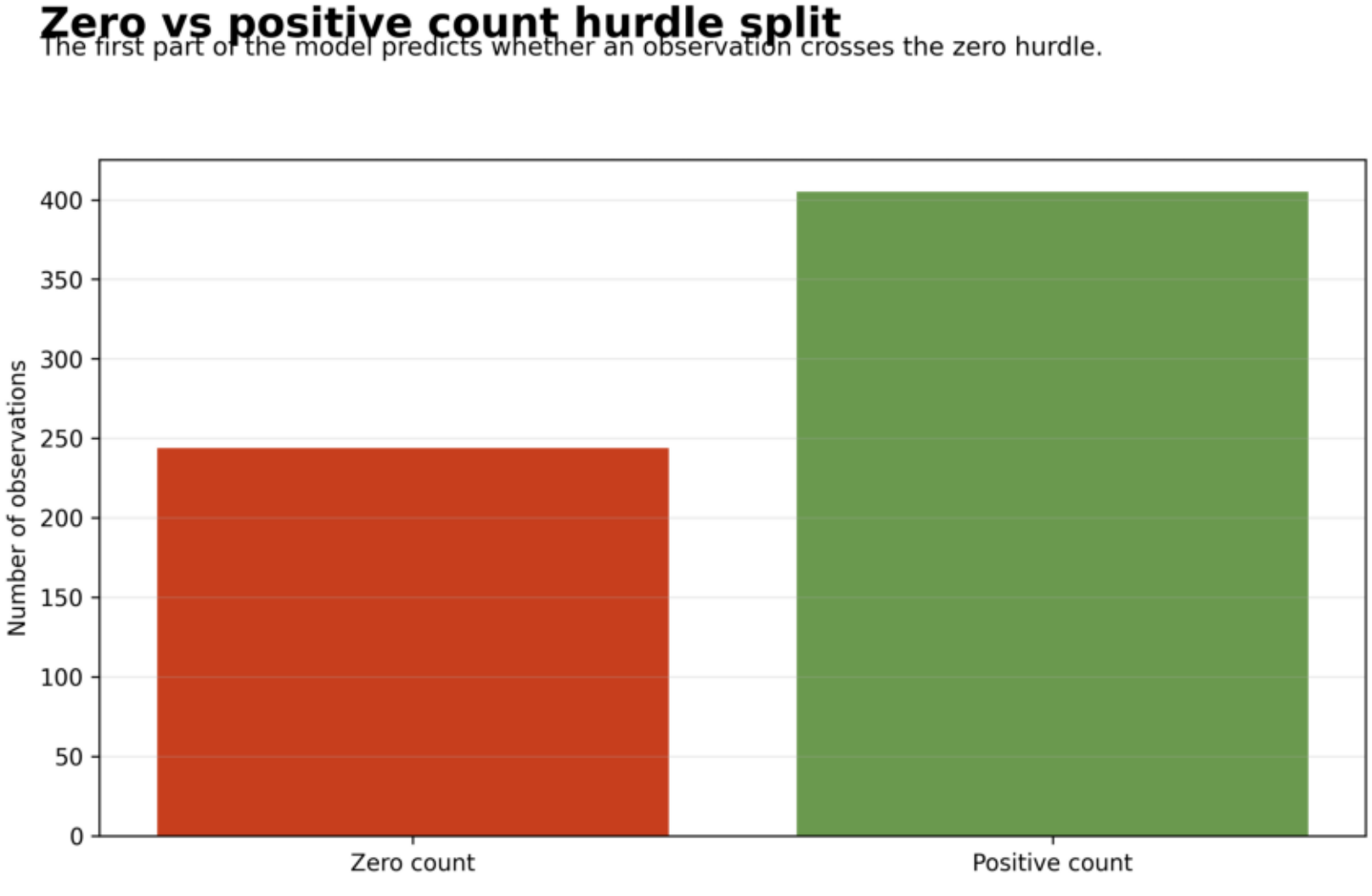
Hurdle model fit statistics

model_part	family	link	n	n_zero	n_positive	df_model	df_resid	log_likelihood	aic	pearson_disp	note
Hurdle binary	Binomial	logit	649.0000	244.0000	405.0000	10.0000	638.0000	-406.2263	834.4525		Models whethe
Positive coun	Poisson	log	405.0000	0.0000	405.0000	10.0000	394.0000	-1211.4880	2444.9760	3.0749	Fitted to positiv

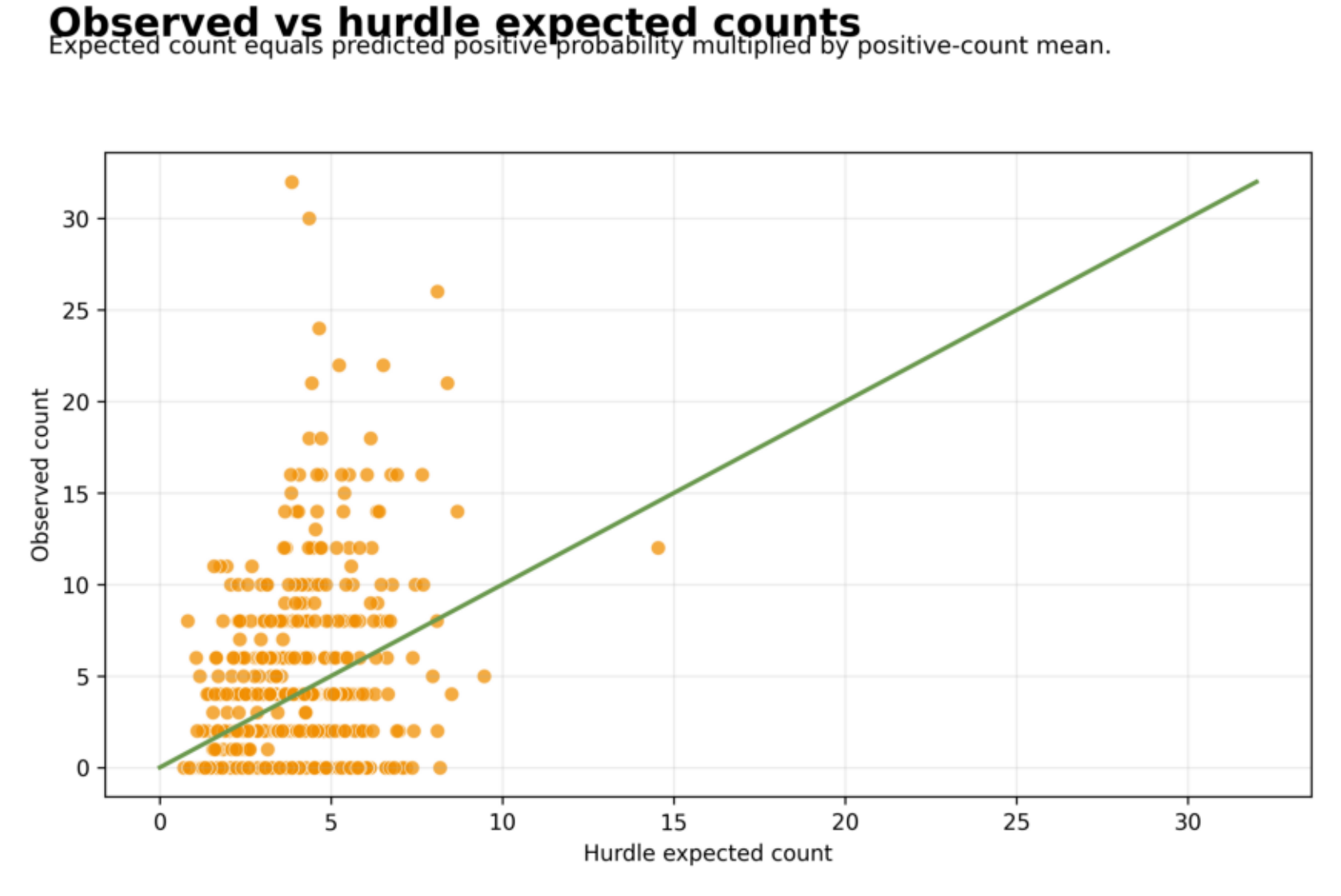
01_hurdle_count_outcome_distribution.png



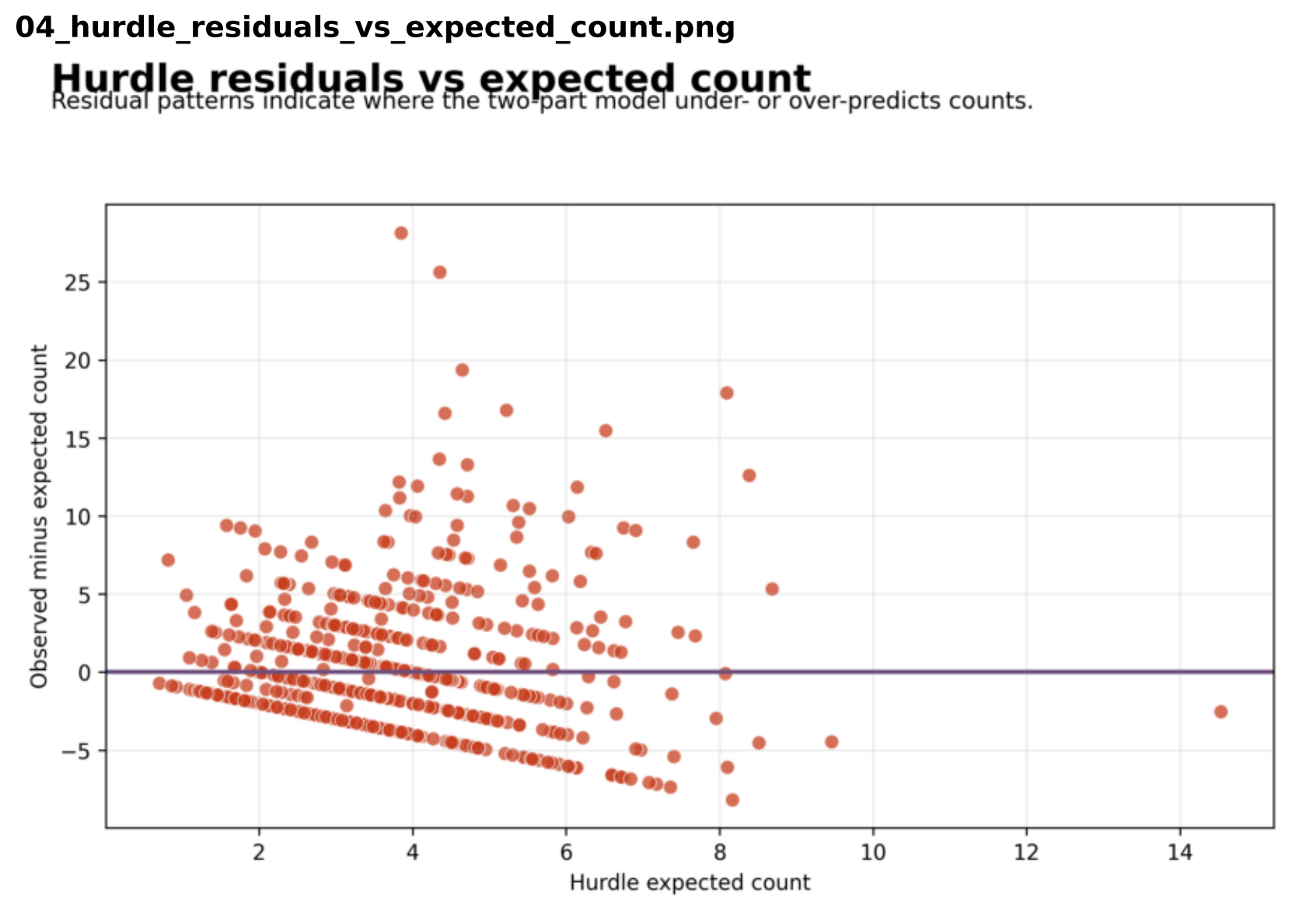
02_hurdle_zero_positive_split.png



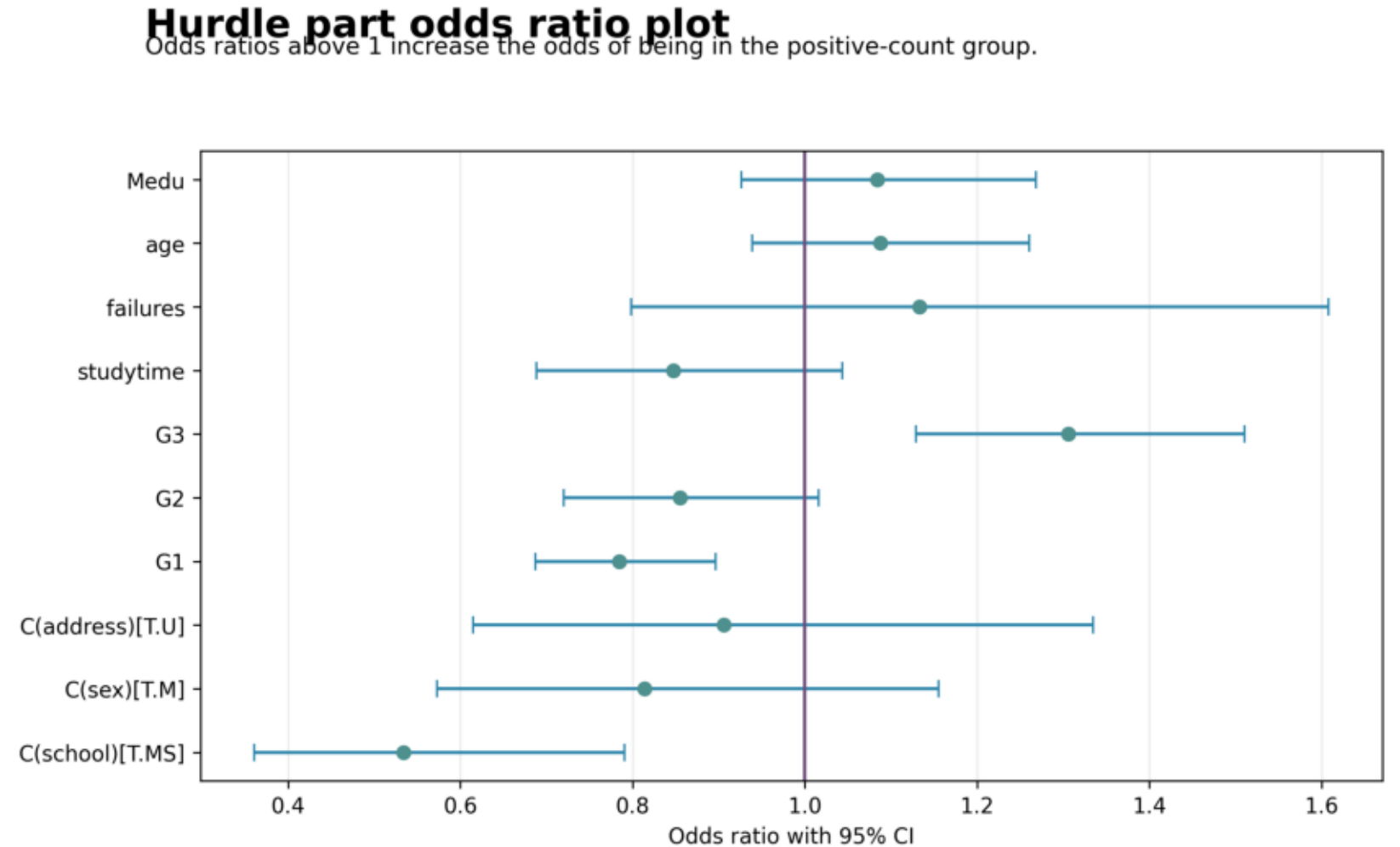
03_hurdle_observed_vs_expected_count.png



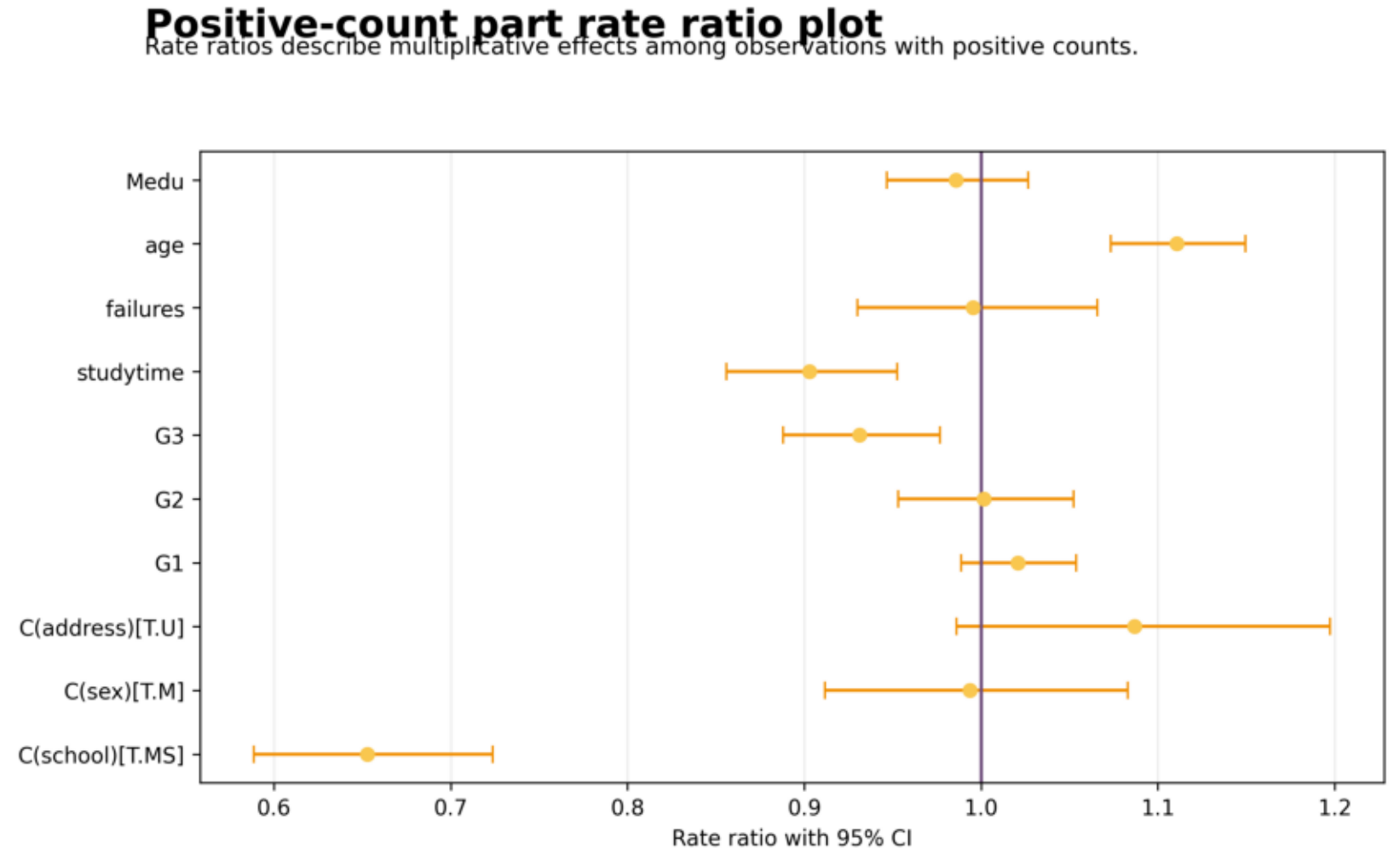
04_hurdle_residuals_vs_expected_count.png



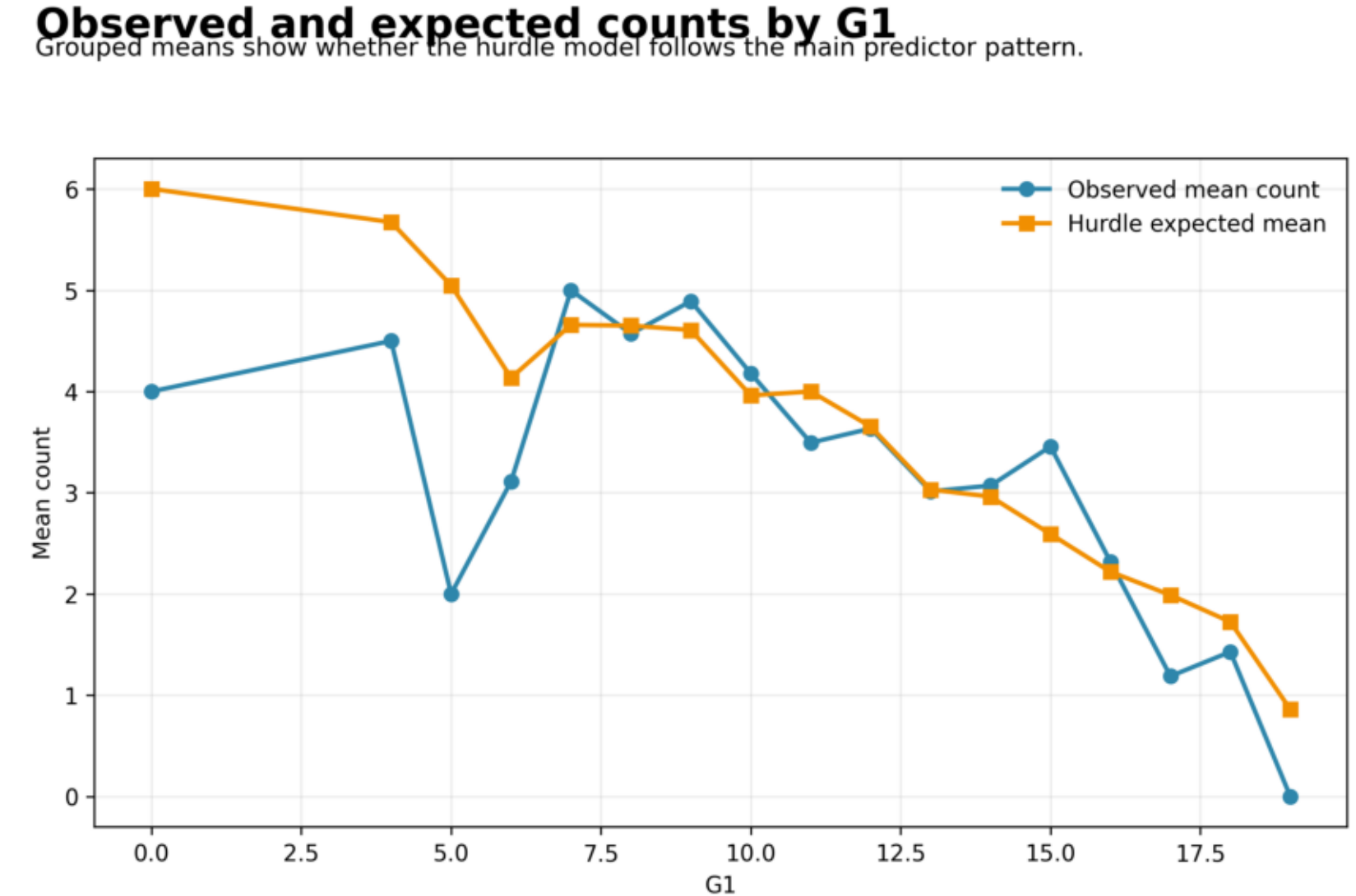
05_hurdle_logit_odds_ratio_plot.png



06_hurdle_positive_count_rate_ratio_plot.png



07_hurdle_observed_expected_by_main_predictor.png



08_hurdle_predicted_positive_probability.png