

Binary Logistic Regression

Test name: Binary Logistic Regression

Dataset: D:\DATA ANALYSIS\H Regression Tests and Models\Binary Logistic Regression\dataset.csv

Rows in original dataset: 649

Columns in original dataset: 33

Rows used in binary logistic model: 649

Binary outcome created: blr_pass = 1 when G3 >= 10.000; otherwise 0

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

Categorical predictors: school, sex, address

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

Family/link: Binomial with logit link.

Model fitting method: Logit maximum likelihood

Likelihood-ratio p-value: 0.000000

McFadden pseudo R2: 0.6853

Nagelkerke R2: 0.7720

AUC ROC: 0.9768

Classification accuracy at cutoff 0.50: 0.9353

Hosmer-Lemeshow p-value: 0.034923

The likelihood-ratio test supports the fitted predictor set.

AUC indicates useful discrimination.

Hosmer-Lemeshow p-value is below 0.05, so calibration/lack-of-fit should be reviewed.

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

Decision guide: p < 0.05 and odds-ratio CI not crossing 1 indicates stronger evidence for a predictor effect.

Reporting guide: state the binary outcome, link, predictors, likelihood-ratio test, odds ratios, 95% CIs, AUC, classification cutoff, and calibration result.

Charts generated: 8

Binary logistic coefficients and odds ratios

model	term	logit_coef	standard_error	wald_z_value	p_value	ci95_lower_log	ci95_upper_log	odds_ratio_exp	odds_ratio_ci95_lower	odds_ratio_ci95_upper	decision_alpha_05
Logit maximum likelihood	Intercept	-25.3006	4.2614	-5.9372	0.0000	-33.6528	-16.9484	0.0000	0.0000	0.0000	Statistically significant
Logit maximum likelihood	C(school)[T.MS]	-0.9511	0.4352	-2.1853	0.0289	-1.8040	-0.0981	0.3863	0.1646	0.9066	Statistically significant
Logit maximum likelihood	C(sex)[T.M]	-0.6326	0.4226	-1.4970	0.1344	-1.4609	0.1956	0.5312	0.2320	1.2161	Not statistically significant
Logit maximum likelihood	C(address)[T.U]	-0.0595	0.4118	-0.1445	0.8851	-0.8667	0.7477	0.9422	0.4203	2.1121	Not statistically significant
Logit maximum likelihood	G1	0.6174	0.1397	4.4206	0.0000	0.3437	0.8912	1.8542	1.4101	2.4380	Statistically significant
Logit maximum likelihood	G2	1.5366	0.2254	6.8161	0.0000	1.0947	1.9784	4.6486	2.9884	7.2311	Statistically significant
Logit maximum likelihood	studytime	0.1719	0.2715	0.6332	0.5266	-0.3602	0.7040	1.1876	0.6976	2.0218	Not statistically significant
Logit maximum likelihood	failures	-0.3514	0.2518	-1.3954	0.1629	-0.8450	0.1422	0.7037	0.4296	1.1528	Not statistically significant
Logit maximum likelihood	absences	-0.0433	0.0443	-0.9775	0.3283	-0.1301	0.0435	0.9576	0.8780	1.0445	Not statistically significant
Logit maximum likelihood	age	0.4397	0.1815	2.4222	0.0154	0.0839	0.7956	1.5523	1.0875	2.2157	Statistically significant

Fit statistics

model	family	link	n	events_o	non_events_o	df_model	df_resid	log_likeli	log_likeli	likelihood	likelihood	likelihood	aic	bic	mcfadde	cox_sne	nagelker	auc_roc	note
Logit ma	Binomia	logit	649.000	549.000	100.000	9.0000	639.000	-87.7616	-278.892	382.262	9.0000	0.0000	195.523	240.277	0.6853	0.4451	0.7720	0.9768	Odds ratios

Classification metrics

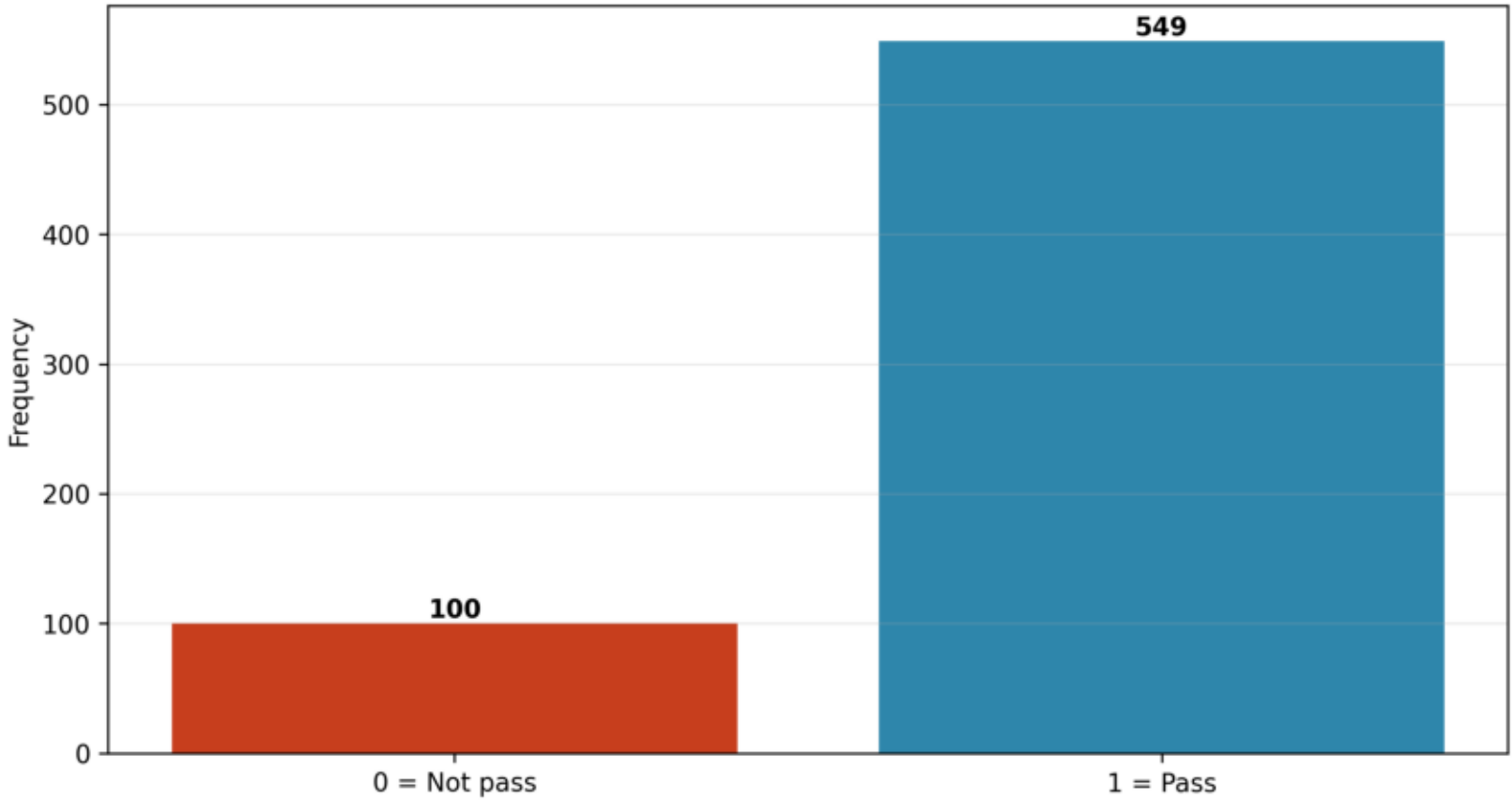
metric	value
True positive	531.0000
False positive	24.0000
True negative	76.0000
False negative	18.0000
Accuracy	0.9353
Sensitivity / recall	0.9672
Specificity	0.7600
Precision / PPV	0.9568
Negative predictive value	0.8085
F1 score	0.9620
Cutoff	0.5000

Hosmer-Lemeshow summary

hosmer_lemeshow_chi_square	df	p_value	groups
16.5690	8.0000	0.0349	10.0000

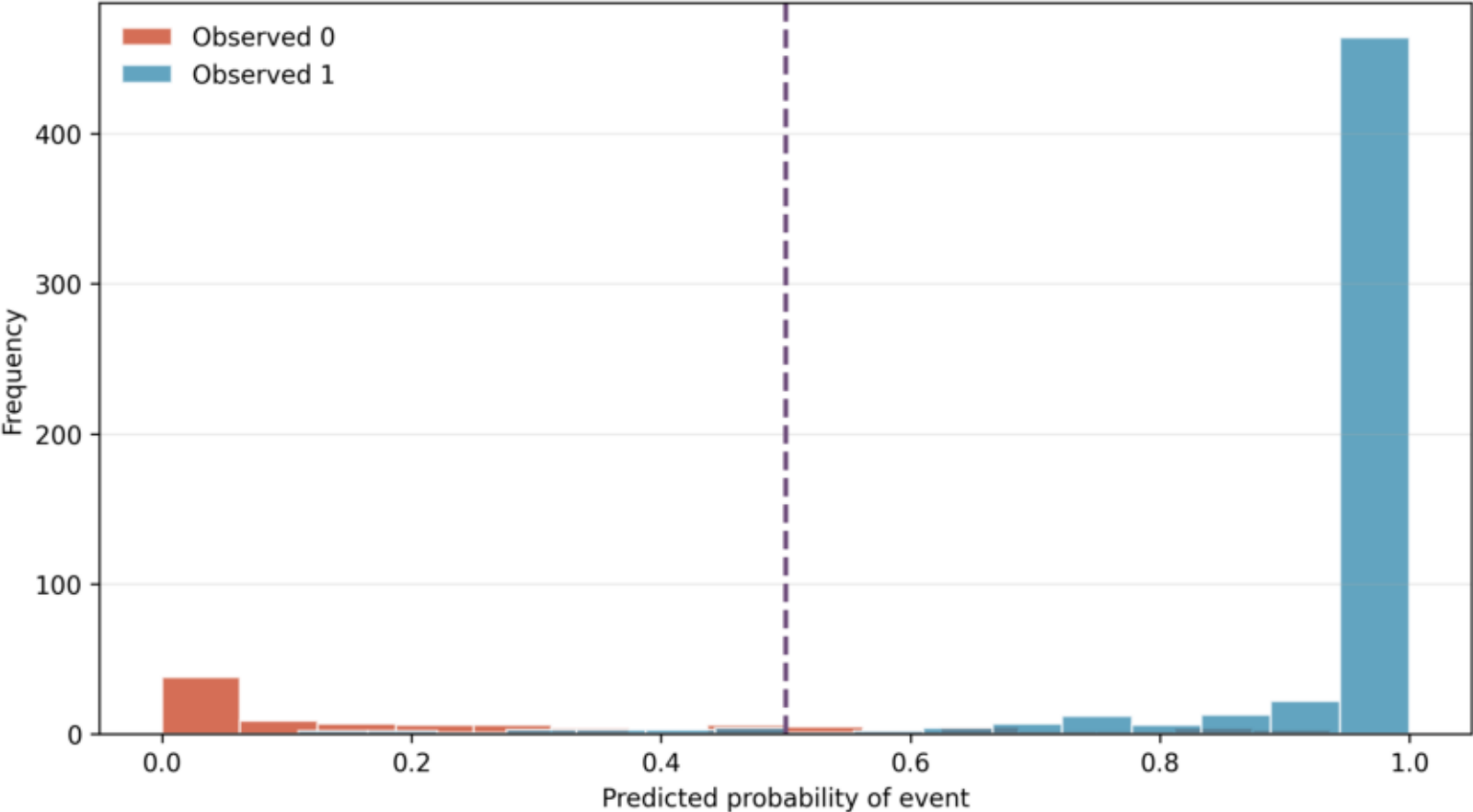
Binary outcome distribution

The target event is blr_pass=1 when G3 is at or above the selected pass threshold.



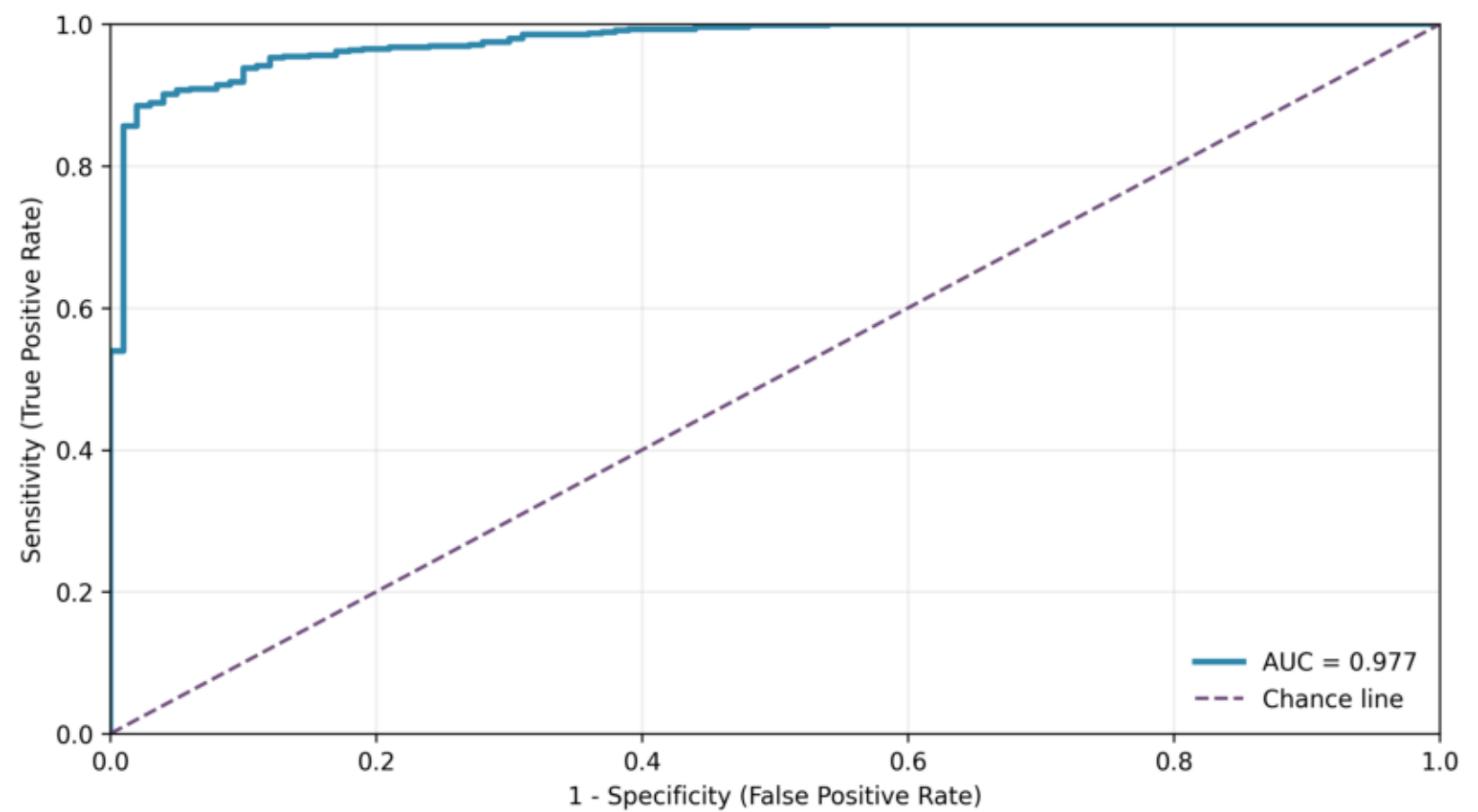
Predicted probability distribution by observed class

Better separation is shown when observed 1 cases concentrate toward higher probabilities.



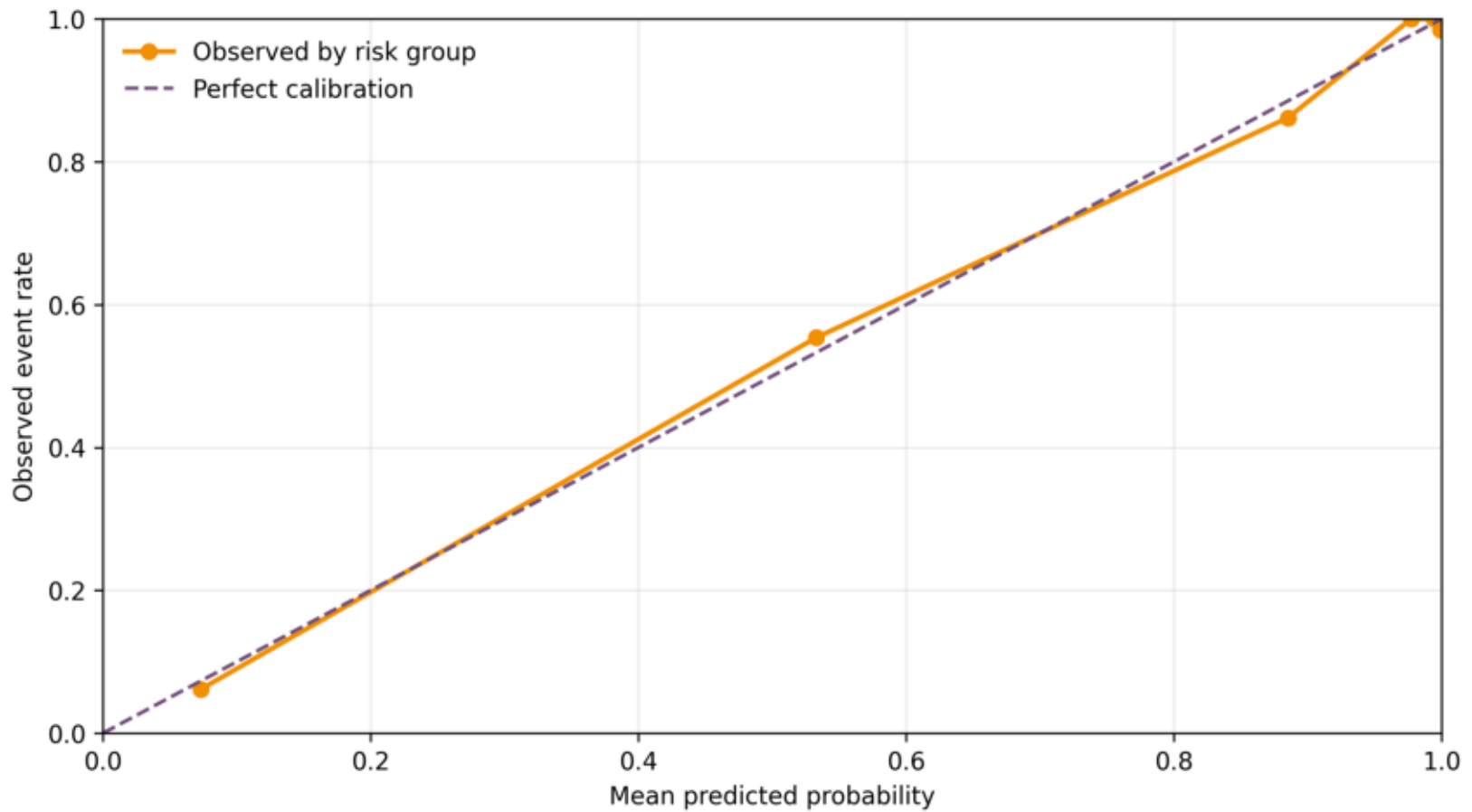
ROC curve for binary logistic regression

The ROC curve summarizes discrimination across all probability cutoffs.



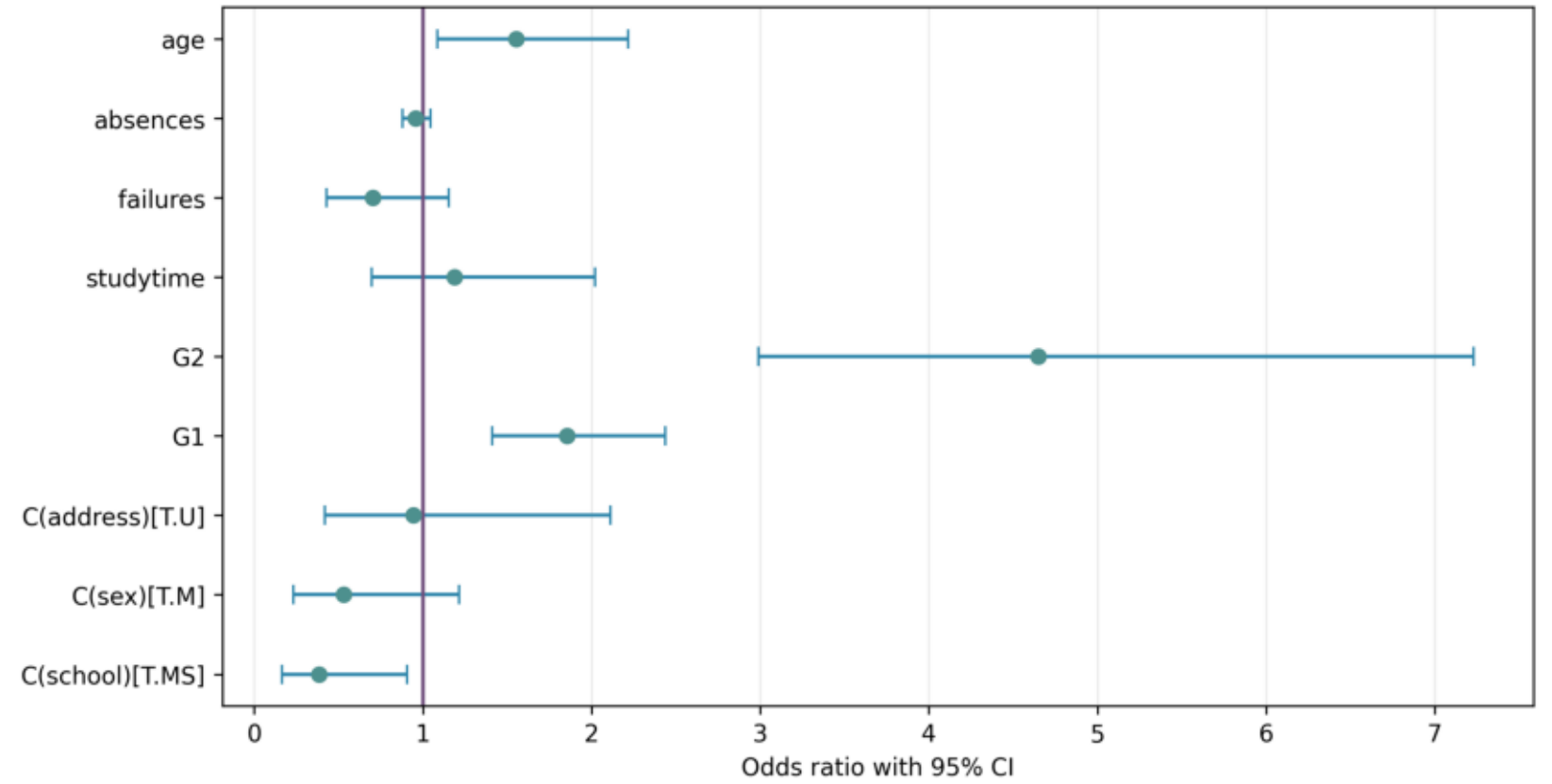
Calibration plot by predicted-risk groups

Points close to the diagonal show better agreement between predicted and observed event rates.



Binary logistic odds ratio plot

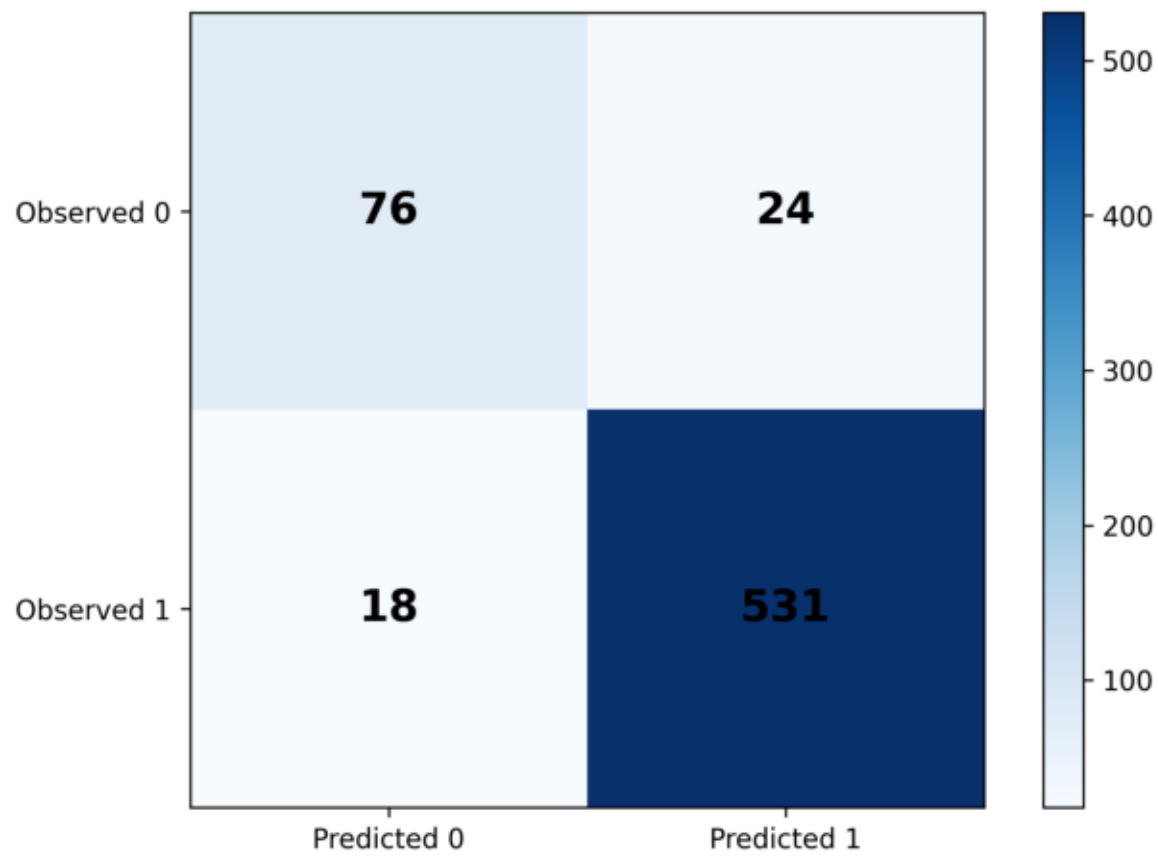
Intervals not crossing 1 indicate stronger evidence for a predictor effect on event odds.



06_binary_logistic_confusion_matrix.png

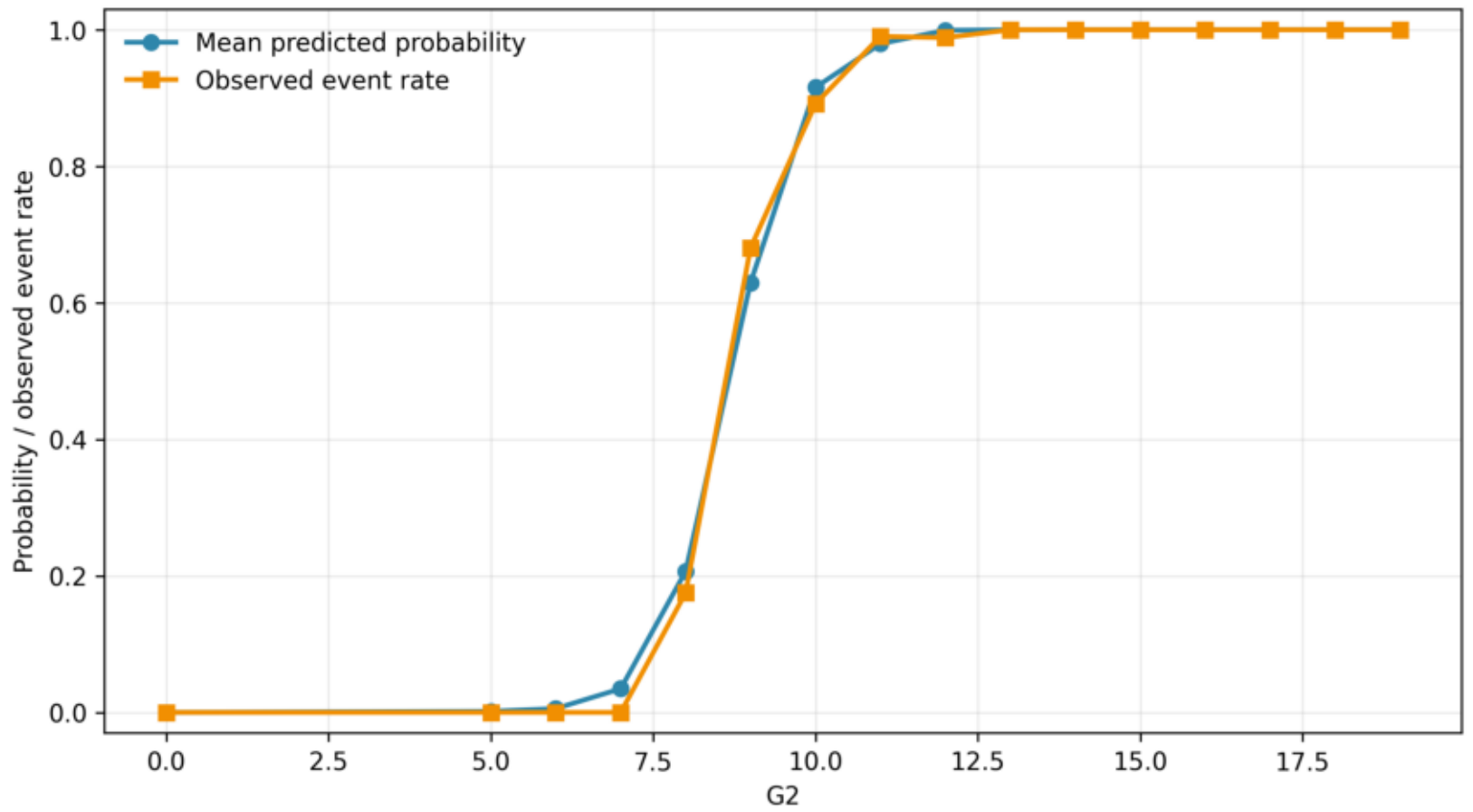
Classification confusion matrix

Classes use the probability cutoff of 0.50.



Predicted probability by G2

This chart compares grouped observed outcomes with model-predicted probabilities.



Pearson residuals vs predicted probability

Large residual patterns can indicate influential observations or model lack of fit.

