

Logistic Regression

Test name: Logistic Regression

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

Rows in original dataset: 649

Columns in original dataset: 33

Rows used in logistic model: 649

Binary outcome created: `logit_pass` = 1 when `G3` $\geq$ 10.00, otherwise 0

Events coded 1: 549

Non-events coded 0: 100

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

Categorical predictors: `school`, `sex`, `address`

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

Family/link: Binomial with logit link.

Fit method: Maximum likelihood GLM Binomial fit

ROC AUC: 0.9768

Accuracy at cutoff 0.50: 0.9353

Sensitivity/recall: 0.9672

Specificity: 0.7600

Interpretation: coefficients are log-odds effects; $\exp(\text{coefficient})$ is the estimated odds ratio.

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

Reporting guide: state binary outcome, cutoff, link, predictors, odds ratios, 95% CIs, AUC, accuracy, sensitivity and specificity.

Charts generated: 7

Logistic regression odds ratios

model	term	logit_coefficient	standard_error	z_value	p_value	ci95_lower_logit	ci95_upper_logit	odds_ratio_exact	odds_ratio_confidence	odds_ratio_confidence	decision_alpha
Logistic Regr	Intercept	-25.3006	4.2614	-5.9372	0.0000	-33.6528	-16.9484	0.0000	0.0000	0.0000	Statistically significant
Logistic Regr	C(school)[T.M]	-0.9511	0.4352	-2.1853	0.0289	-1.8040	-0.0981	0.3863	0.1646	0.9066	Statistically significant
Logistic Regr	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
Logistic Regr	C(address)[T.M]	-0.0595	0.4118	-0.1445	0.8851	-0.8667	0.7477	0.9422	0.4203	2.1121	Not statistically significant
Logistic Regr	G1	0.6174	0.1397	4.4206	0.0000	0.3437	0.8912	1.8542	1.4101	2.4380	Statistically significant
Logistic Regr	G2	1.5366	0.2254	6.8161	0.0000	1.0947	1.9784	4.6486	2.9884	7.2311	Statistically significant
Logistic Regr	studytime	0.1719	0.2715	0.6332	0.5266	-0.3602	0.7040	1.1876	0.6976	2.0218	Not statistically significant
Logistic Regr	failures	-0.3514	0.2518	-1.3954	0.1629	-0.8450	0.1422	0.7037	0.4296	1.1528	Not statistically significant
Logistic Regr	absences	-0.0433	0.0443	-0.9775	0.3283	-0.1301	0.0435	0.9576	0.8780	1.0445	Not statistically significant
Logistic Regr	age	0.4397	0.1815	2.4222	0.0154	0.0839	0.7956	1.5523	1.0875	2.2157	Statistically significant

Logistic regression fit statistics

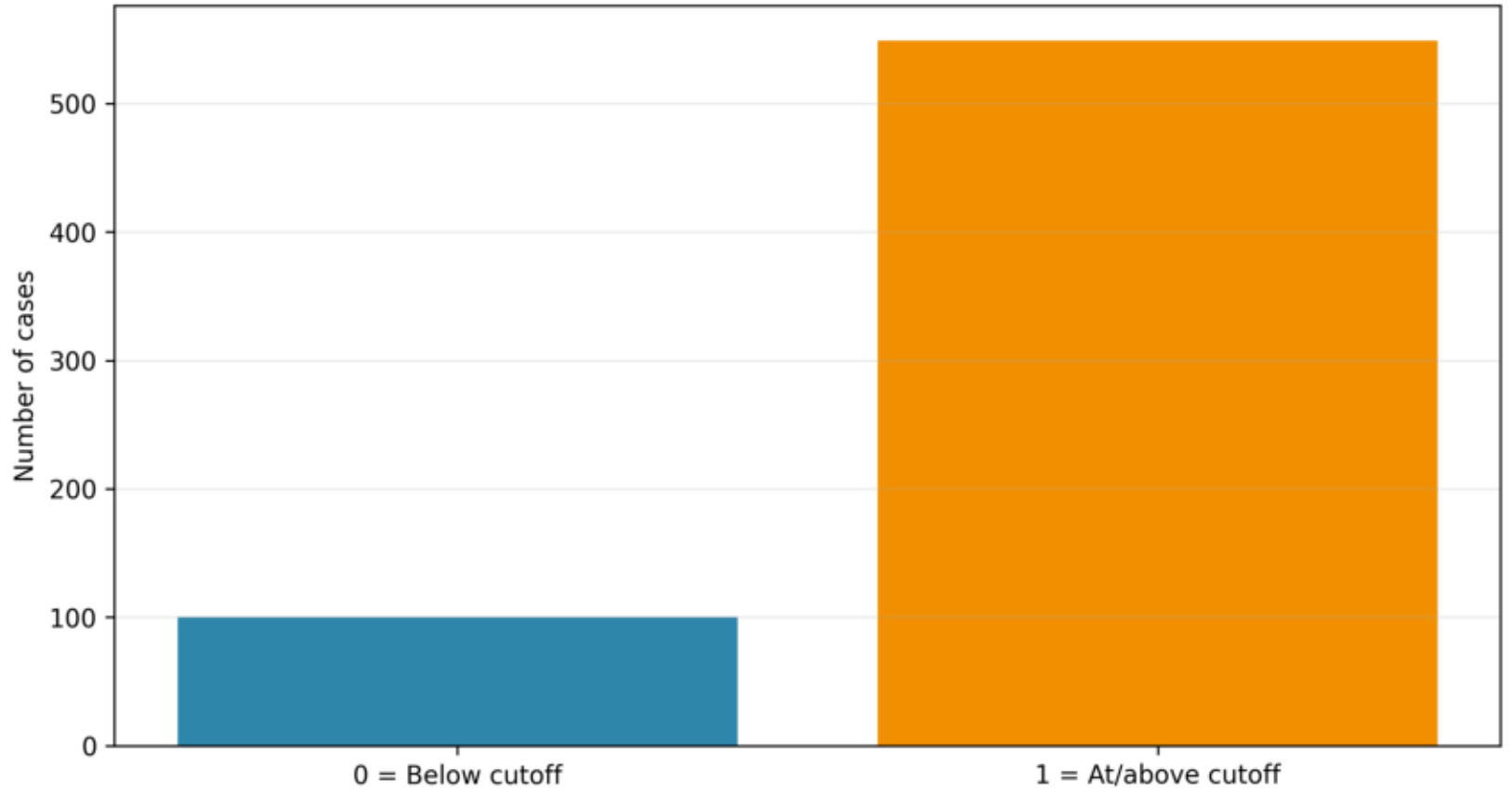
model	family	link	fit_not	n	events	non_ev	event_	log_lik	null_lo	mcfad	aic	bic	roc_auc	cutoff	accuracy	precision	sensitivity	specificity	f1_score	true_n	false_p	false_r	true_pos
Logisti	Binom	logit	Maxim	649.00	549.00	100.00	0.8455	-87.76	-278.8	0.6853	195.52	-3962.	0.9768	0.5000	0.9353	0.9568	0.9672	0.7600	0.9620	76.000	24.000	18.000	531.000

Confusion matrix

actual	predicted_0	predicted_1
Actual 0	76.0000	24.0000
Actual 1	18.0000	531.0000

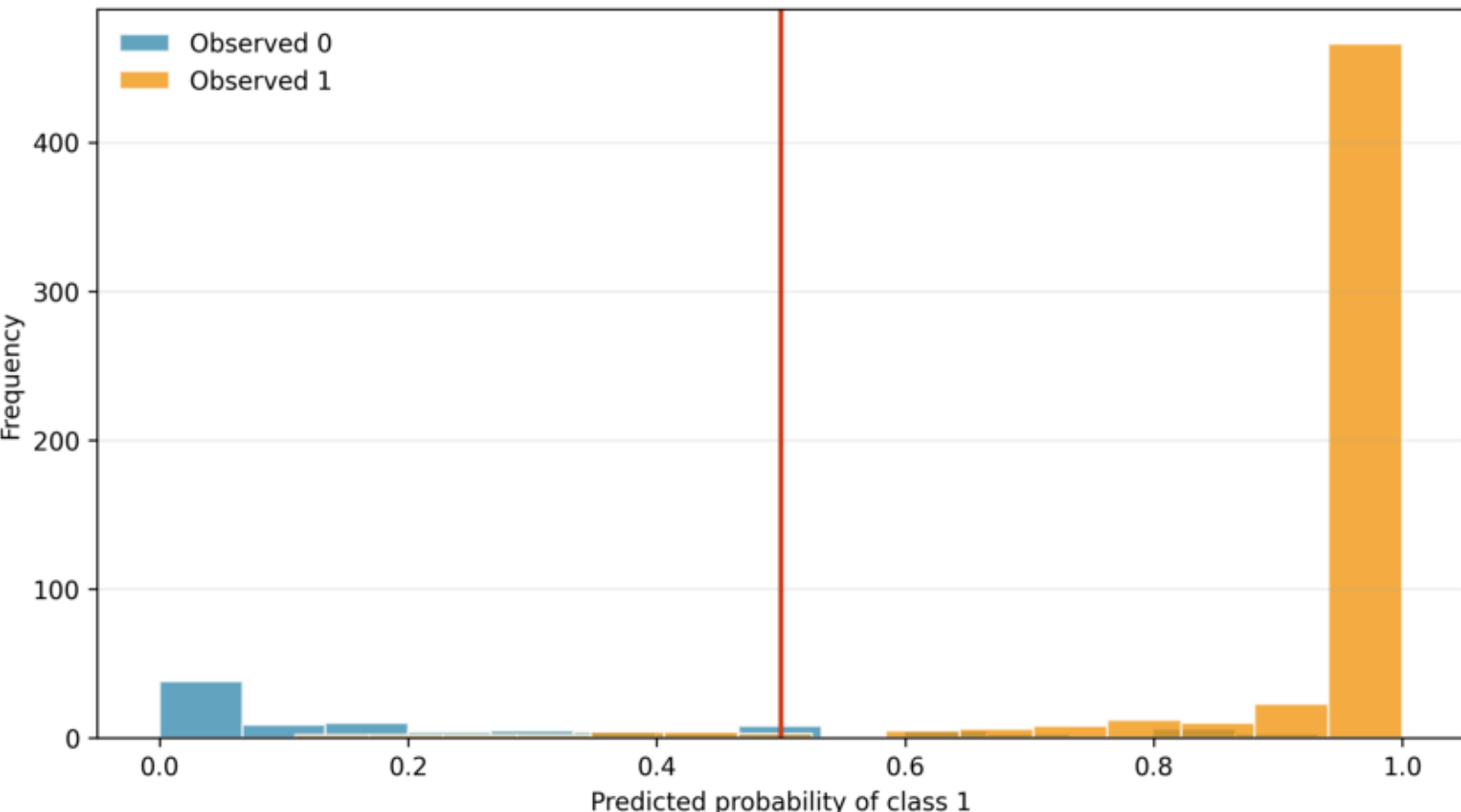
Binary outcome distribution

Outcome: logit_pass created from G3 using cutoff 10.00.



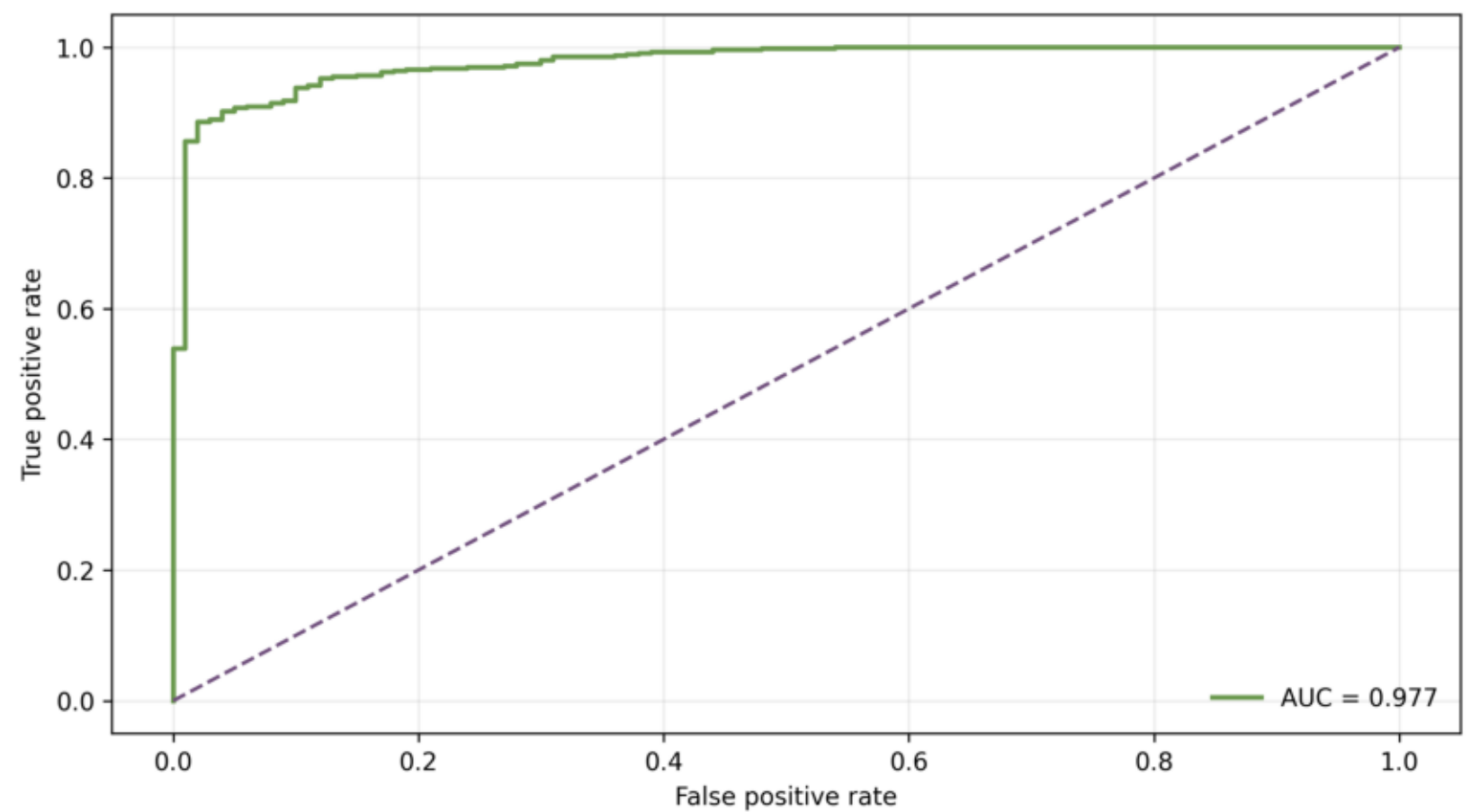
Predicted probability distribution

Good separation shows low probabilities for class 0 and high probabilities for class 1.



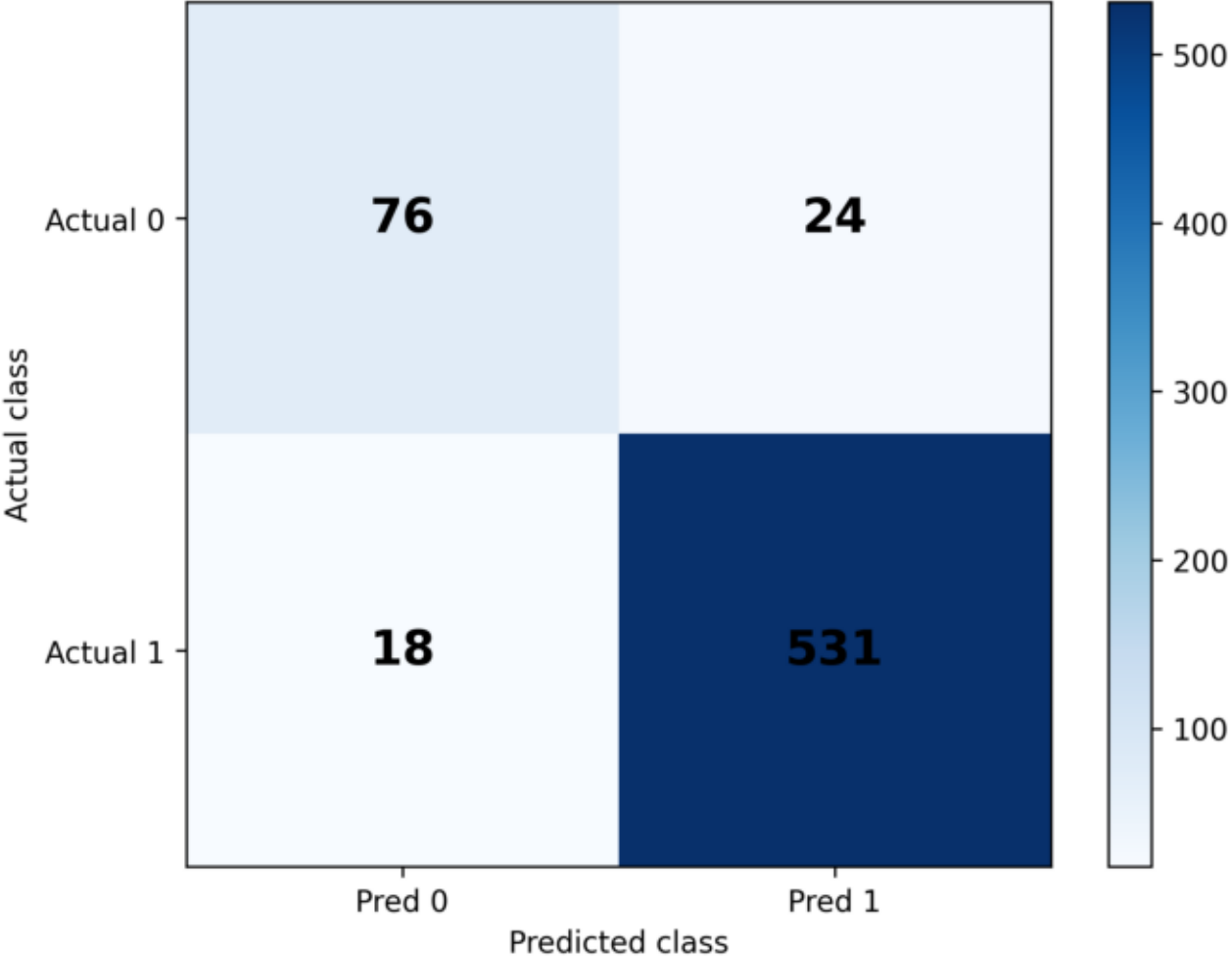
ROC curve

The ROC curve summarizes classification performance across probability thresholds.



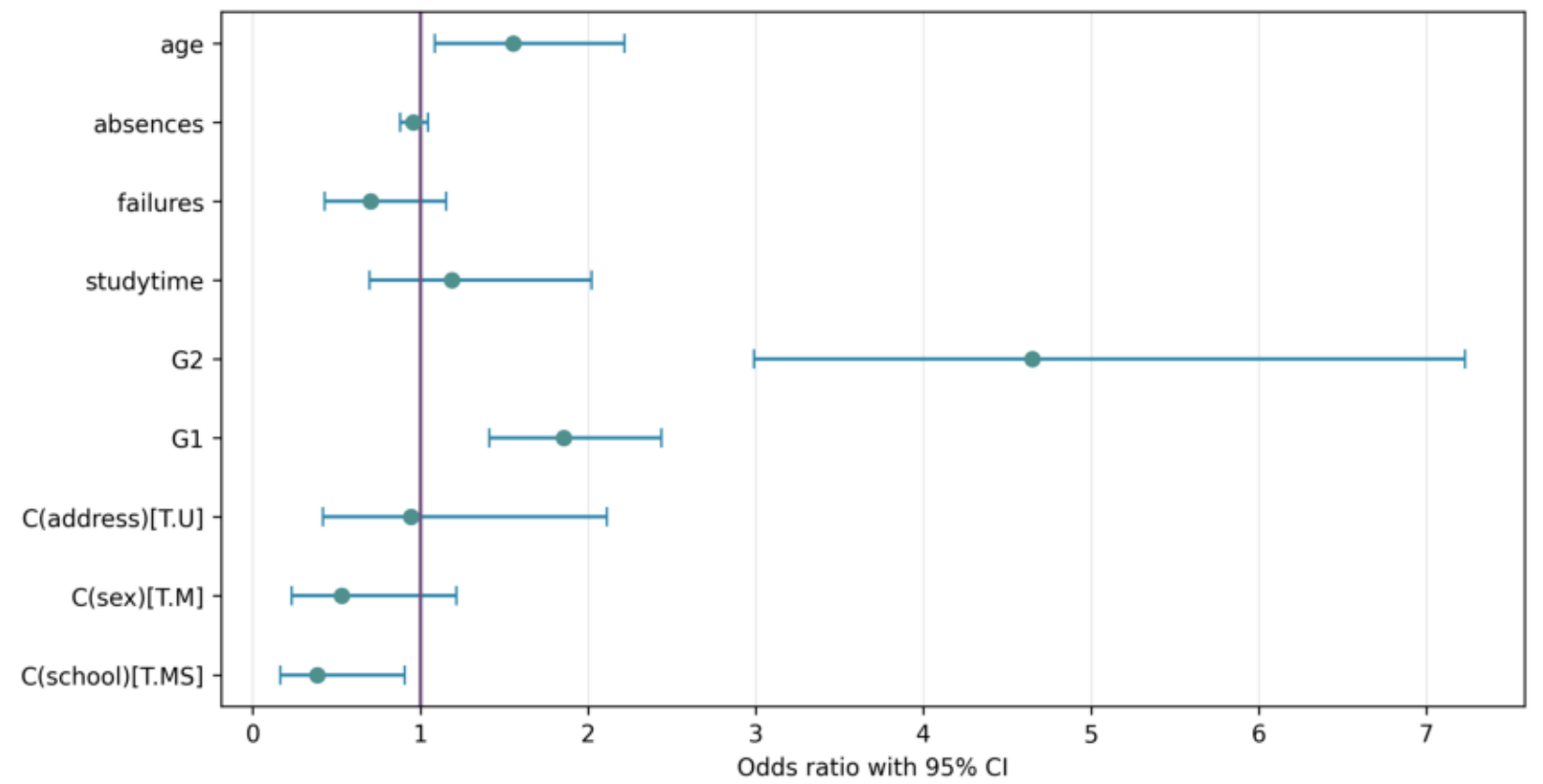
Confusion matrix

Classification uses predicted probability cutoff 0.50.



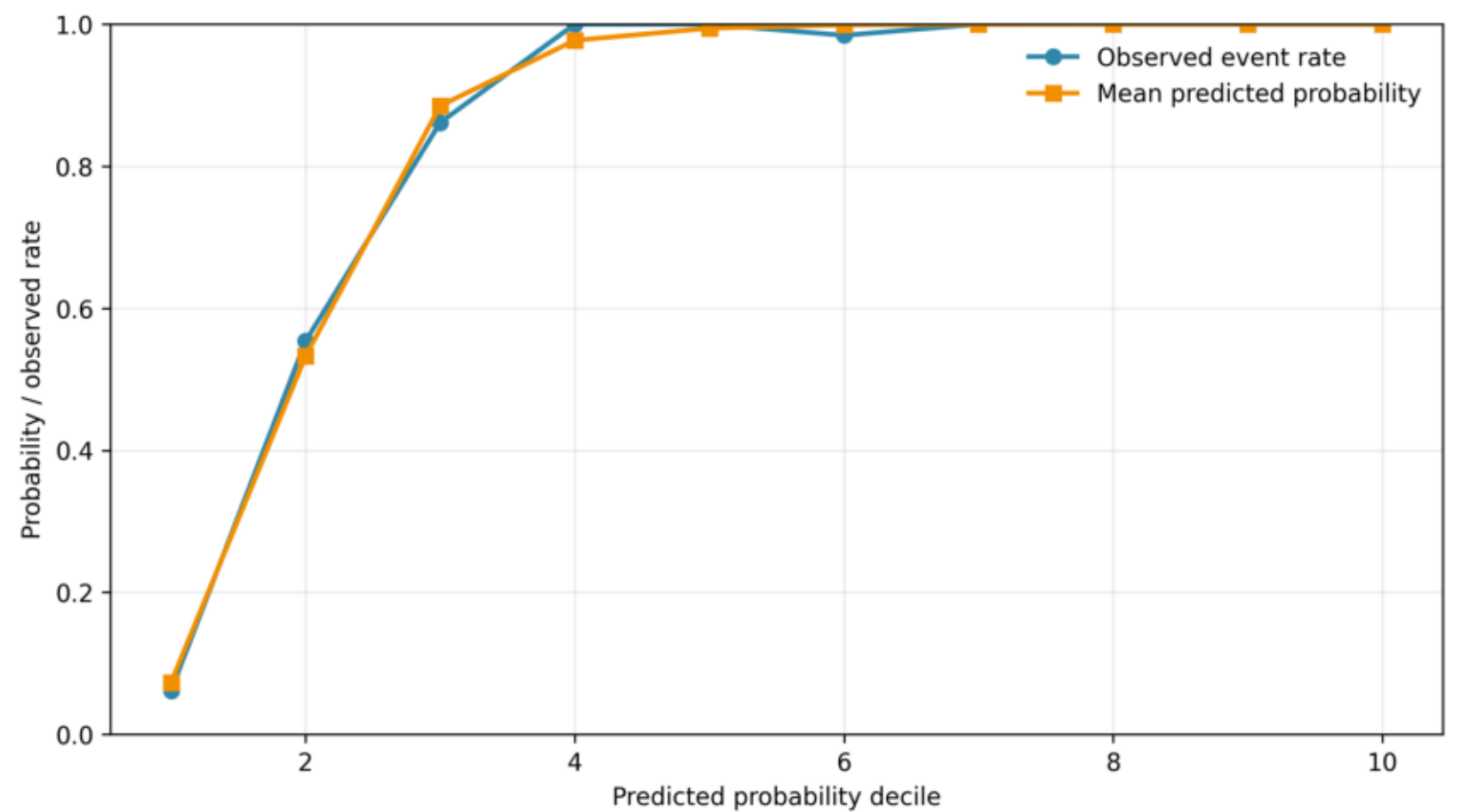
Logistic regression odds ratio plot

Odds ratios above 1 increase odds of class 1; below 1 decrease odds.



Calibration by predicted probability decile

Observed event rates should roughly follow the mean predicted probabilities.



Observed and predicted probability by G1

This chart compares actual event rate with fitted probability by the main predictor.

