

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 >= 10 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 + school + sex + address

Family/link: binomial(link = 'logit').

ROC AUC: 0.9768

Accuracy at cutoff 0.5 : 0.9353

Sensitivity/recall: 0.9672

Specificity: 0.76

Interpretation: coefficients are log-odds effects; exp(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
Logistic Regression	(Intercept)	-25.30061097	4.26107399	-5.9376136	2.892006e-09	-33.65216252
Logistic Regression	G1	0.61743488	0.13966498	4.4208280	9.832340e-06	0.34369654
Logistic Regression	G2	1.53655782	0.22541183	6.8166690	9.317558e-12	1.09475876
Logistic Regression	studytime	0.17190772	0.27146011	0.6332706	5.265569e-01	-0.36014433
Logistic Regression	failures	-0.35140655	0.25181571	-1.3954910	1.628678e-01	-0.84495626
Logistic Regression	absences	-0.04330284	0.04429576	-0.9775842	3.282800e-01	-0.13012093
Logistic Regression	age	0.43973369	0.18153469	2.4223122	1.542209e-02	0.08393224
Logistic Regression	schoolMS	-0.95105542	0.43518929	-2.1853833	2.886075e-02	-1.80401075
Logistic Regression	sexM	-0.63262749	0.42256589	-1.4971097	1.343647e-01	-1.46084142
Logistic Regression	addressU	-0.05951009	0.41182528	-0.1445033	8.851031e-01	-0.86667281
ci95_upper_logit odds_ratio_exp_b odds_ratio_ci95_lower odds_ratio_ci95_upper		decision_alpha_0_05				
-16.94905942	1.028216e-11	9.357623e-14	4.356292e-08	Statistically significant		
0.89117322	1.854166e+00	1.410151e+00	2.437988e+00	Statistically significant		
1.97835688	4.648561e+00	2.988462e+00	7.230852e+00	Statistically significant		
0.70395976	1.187568e+00	6.975756e-01	2.021742e+00	Not statistically significant		
0.14214317	7.036976e-01	4.295761e-01	1.152742e+00	Not statistically significant		
0.04351526	9.576213e-01	8.779892e-01	1.044476e+00	Not statistically significant		
0.79553514	1.552294e+00	1.087555e+00	2.215626e+00	Statistically significant		
-0.09810008	3.863331e-01	1.646372e-01	9.065582e-01	Statistically significant		
0.19558643	5.311943e-01	2.320409e-01	1.216024e+00	Not statistically significant		
0.74765262	9.422260e-01	4.203478e-01	2.112036e+00	Not statistically significant		

Logistic regression fit statistics

model	family	link	n	events_1	non_events_0	event_rate	log_likelihood	null_deviance
Logistic Regression	binomial	logit	649	549	100	0.8459168	-87.76155	557.7855
residual_deviance	aic	mcfadden_pseudo_r2	roc_auc	accuracy	sensitivity	recall	specificity	
175.5231	195.5231	0.6853215	0.9767942	0.9352851	0.9672131		0.76	

Confusion matrix

	actual	predicted_0	predicted_1
Actual 0		76	24
Actual 1		18	531













