

Probit Regression

Test name: Probit Regression

Dataset: D:/DATA ANALYSIS/H Regression Tests and Models/Probit Regression/dataset.csv

Rows in original dataset: 649

Columns in original dataset: 33

Rows used in probit model: 649

Binary outcome created: probit_pass = 1 when G3 >= 10 otherwise 0

Observed event rate: 0.8459

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

Categorical predictors: school, sex, address

Formula: probit_pass ~ G1 + G2 + studytime + failures + absences + age + school + sex + address

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

Log likelihood: -89.2387

AIC: 198.4774

McFadden pseudo R squared: 0.68

Classification accuracy at 0.50 cutoff: 0.9384

Interpretation: coefficients are latent normal z-score effects; positive coefficients increase predicted probability.

Decision guide: $p < 0.05$ indicates statistical evidence for a predictor effect.

Reporting guide: state binary outcome cutoff, probit link, coefficients, p-values, model fit and classification results.

Charts generated: 8

Probit Regression coefficients

term	probit_coefficient	standard_error	z_value	p_value	ci95_lower	ci95_upper
(Intercept)	-12.8142088	2.12413161	-6.0326812	1.612614e-09	-16.97743023	-8.65098733
G1	0.3310649	0.07652429	4.3262726	1.516537e-05	0.18108009	0.48104980
G2	0.8003018	0.11339395	7.0577115	1.692670e-12	0.57805372	1.02254982
studytime	0.0894782	0.14708919	0.6083261	5.429712e-01	-0.19881132	0.37776772
failures	-0.1858637	0.13809304	-1.3459312	1.783247e-01	-0.45652112	0.08479365
absences	-0.0244432	0.02313968	-1.0563328	2.908162e-01	-0.06979613	0.02090973
age	0.2091438	0.09481825	2.2057337	2.740265e-02	0.02330345	0.39498415
schoolMS	-0.5888183	0.23730020	-2.4813226	1.308958e-02	-1.05391819	-0.12371851
sexM	-0.3122860	0.22635997	-1.3795989	1.677102e-01	-0.75594335	0.13137142
addressU	-0.1123543	0.22519916	-0.4989109	6.178422e-01	-0.55373656	0.32902793

decision_alpha_0_05

Statistically significant

Statistically significant

Statistically significant

Not statistically significant

Not statistically significant

Not statistically significant

Statistically significant

Statistically significant

Not statistically significant

Not statistically significant

interpretation

Positive coefficients increase the latent probit score and therefore increase predicted probability.

Positive coefficients increase the latent probit score and therefore increase predicted probability.

Positive coefficients increase the latent probit score and therefore increase predicted probability.

Positive coefficients increase the latent probit score and therefore increase predicted probability.

Positive coefficients increase the latent probit score and therefore increase predicted probability.

Positive coefficients increase the latent probit score and therefore increase predicted probability.

Positive coefficients increase the latent probit score and therefore increase predicted probability.

Probit Regression model fit statistics

model	family	link	n	df_residual	log_likelihood	null_log_likelihood	aic	mcFadden_pseudo_r2
Probit Regression	Binomial	Probit	649	639	-89.23871	-278.8928	198.4774	0.680025
classification_accuracy_cutoff_0_50 mean_predicted_probability observed_event_rate log_loss								
			0.9383667		0.8450701	0.8459168	0.1375019	

Probit Regression confusion matrix

actual_class	predicted_0	predicted_1
Actual 0	77	23
Actual 1	17	532















