

H1 GLM Summary — MATH-500, Qwen2.5-1.5B

GLM (correct ~ C(problem) + step_std + aha), cluster-robust SEs by problem

Variant	N	Share Ah	Accuracy	AME	Coef	p-value
formal	123014	0.0243	0.3011	0.0214	0.1633	0.2973
gpt	123014	0.0217	0.3011	-0.0858	-0.7049	0.0000
words	123014	0.0120	0.3011	-0.0485	-0.3863	0.0017

Group accuracies (aha=1 vs aha=0):

Variant	acc(aha=1)	acc(aha=0)	Δ (pp)	n1	n0
Formal	0.0174	0.3081	-29.07	2984	120030
Gpt	0.1635	0.3042	-14.07	2673	120341
Words	0.2590	0.3016	-4.26	1479	121535