PROMPT**

I ran a linear regression on my dataset:

OLS Regression Results
==
Dep. Variable: mmse_score R-squared: 0.085
Model: OLS Adj. R-squared: 0.084
Method: Least Squares F-statistic: 118.4
Date: Tue, 18 Aug 2026 Prob (F-statistic): 1.99e-26
Time: 09:44:33 Log-Likelihood: -3408.9
No. Observations: 1273 AIC: 6822.
Df Residuals: 1271 BIC: 6832.
Df Model: 1

Covariance Type: nonrobust
coef std err t P>|t| [0.025 0.975]
const 23.1136 0.285 81.114 0.000 22.555 23.673
edu_clean 1.6618 0.153 10.881 0.000 1.362 1.961
Omnibus: 398.011 Durbin-Watson: 2.012
Prob(Omnibus): 0.000 Jarque-Bera (JB): 1394.628
Skew: -1.507 Prob(JB): 1.45e-303
Kurtosis: 7.148 Cond. No. 6.78
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
R-squared: 0.08521884342638919
Saved.
Outcome variable: MMSE score, continuous
Predictive variable: edu_clean (highest level of education attained), ordinal

My preliminary interpretation:
Individuals with higher levels of education attained are associated with modestly higher MMSE scores.
Educational attainment demonstrates a statistically significant, positive, but modest strength of association with late-life cognitive performance, predicting a 1.66 point increase in MMSE performance for every 1 level of education acquired. This relationship explains 8.5% of the total variance in MMSE scores ($R^2 = 0.085$, $p<0.001$).
Critique my interpretation. Don't rewrite it. Specifically:
Where am I overclaiming?
Where am I underclaiming?
What would a critical reviewer challenge?

** FEEDBACK **

Clear Distinction Between Significance and Magnitude: You avoided the common mistake of confusing a small p-value with a massive effect. By noting that $p < 0.001$ coexists with

explaining only 8.5% of the variance, you accurately communicated that the finding is statistically real but leaves 91.5% of variance unexplained.

Complete Core Reporting: You successfully integrated all essential descriptive statistics—the unstandardized slope ($\beta = 1.66$), the coefficient of determination ($R^2 = 0.085$), and the significance level ($p < 0.001$)—into a clear narrative.

Where You Are Overclaiming

* **Equal-Interval Linearity:** Stating a "1.66 point increase... for every 1 level" treats your ordinal education variable as a continuous interval scale. Because this specific output ran `edu_clean` as a single numeric variable ($Df = 1$), the model *forced* a constant linear slope. Claiming that each educational step yields the exact same 1.66-point gain overclaims linearity across ordinal categories without empirical proof.
* **Construct Scope:** Framing the predictor broadly as "education acquired" overstates what an ordinal variable captures compared to continuous years of formal schooling.

Where You Are Underclaiming

* **Clinical Impact:** Labeling the effect as purely "modest" downplays its real-world clinical significance. On a 30-point MMSE scale—where 2 to 3 points often separate normal cognition from mild cognitive impairment—a 1.66-point shift *per single tier* (and a ~3.32-point difference across the full range) is clinically substantial.
* **Precision and Certainty:** You omitted the 95% confidence interval ($[1.362, 1.961]$). Including this range strengthens your finding by showing that even the lower bound represents a clinically non-trivial effect (>1.36 points).

What a Critical Reviewer Would Challenge

* **Treating Ordinal Data as Continuous:** Reviewers will point out that `edu_clean` was entered as a single continuous term ($Df = 1$) rather than a factor with dummy variables (`C(edu_clean)`). They will demand evidence that the step from Tier 1 to 2 equals the step from Tier 2 to 3, or request categorical factor encoding to show discrete category comparisons.
* **Omitted Variable Bias:** This simple regression lacks essential controls for age and sex. Because older age correlates with both lower MMSE scores and lower historical educational attainment (cohort effects), the unadjusted coefficient likely conflates age-related cognitive decline with educational background.
* **Violation of Normality (Ceiling Effect):** The extreme residual skewness (-1.507) and kurtosis (7.148), along with highly significant Jarque-Bera tests ($p < 0.001$), highlight a severe ceiling effect common to MMSE scores. Reviewers will question standard error accuracy under non-robust OLS and request heteroscedasticity-robust standard errors (`HC3`) or models designed for bounded data.

* **Causal Implication vs. Cognitive Reserve:** Phrasing like "per 1 level... acquired" risks implying a direct protective intervention. Reviewers will emphasize that educational level often serves as a proxy for early-life intelligence, parental socioeconomic status, or baseline cognitive reserve rather than a direct causal mechanism.

**

AI confirmed that the format of my reporting, correctly stating the significance and direction of magnitude for my test.

AI raised that fitting an ordinal variable to a linear regression model without categorical dummy coding results in the variable being treated as having consistent jumps between levels. As a result, claiming that each 1 level increase in education attainment results in a 1.66 point increase in MMSE scores an overclaim. I'll revise this coding, incorporating dummy categories into my model.

Educational attainment demonstrates a statistically significant, positive association with late-life cognitive performance. This relationship explains 8.5% of the total variance in MMSE scores ($R^2 = 0.085$, $p<0.001$).