Short Statistical Note: Student Academic Performance (UCI)

Innokentii Batushanskii

October 2023

1 Purpose

This note summarizes a small empirical exercise demonstrating standard statistical inference tools on an open tabular dataset. The goal is to report interpretable estimates with uncertainty (confidence intervals) and to cross-check selected results using resampling methods.

2 Data

The analysis uses the mathematics sample of the UCI Student Performance dataset ($N = 395$). The outcome is the final grade $G3$ on a 0–20 scale. The project repository contains the raw data in `data/raw/student-mat.csv` and code in `notebooks/01_paper.ipynb`.

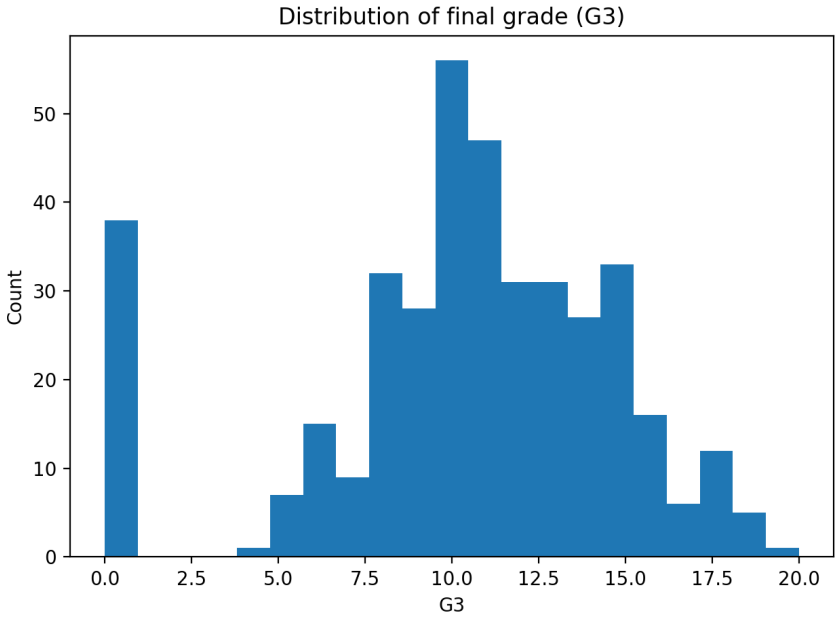

Figure 1: Distribution of final grades ($G3$).

3 Methods

3.1 Group comparisons

Two binary group comparisons were evaluated using Welch two-sample t-tests. Effect sizes are reported using Cohen's d. For robustness, each comparison was validated using (i) a permutation test for the mean difference and (ii) a nonparametric bootstrap confidence interval for the mean difference.

3.2 One-way ANOVA

Differences in mean $G3$ across the categorical variable `reason` were evaluated using one-way ANOVA. Inference was validated by a permutation ANOVA, and effect size is reported using η^2.

3.3 Linear regression

Two OLS specifications were estimated and inference was based on HC3 heteroskedasticity-robust standard errors:

$$G3_i = X_i'\beta + \varepsilon_i \quad \text{(Model A: excludes prior grades)} \tag{1}$$

$$G3_i = X_i'\beta + \gamma_1 G1_i + \gamma_2 G2_i + \varepsilon_i \quad \text{(Model B: includes prior grades).} \tag{2}$$

Selected coefficients from Model A were additionally evaluated using a pairs bootstrap (2,000 resamples).

4 Results

4.1 Welch t-tests with resampling checks

Table 1 reports mean differences (group A minus group B), Welch 95% CIs, Cohen's d, bootstrap CIs, and permutation p-values.

Table 1: Binary group comparisons for $G3$ (Welch t-test with resampling checks).

Comparison (A–B)	$\bar{G3}_A$	$\bar{G3}_B$	Diff.	Welch 95% CI	d	p_{perm}
F – M	9.966	10.914	-0.948	[-1.851, -0.045]	-0.208	0.041
No internet – Yes	9.409	10.617	-1.208	[-2.413, -0.002]	-0.265	0.052

Notes: "Diff." is the mean difference in grade points. Welch p-values were 0.0396 (sex) and 0.0496 (internet). Bootstrap 95% CIs for the mean differences were sex: [-1.833, -0.060] and internet: [-2.422, -0.018].

4.2 ANOVA: school choice motivation

Mean $G3$ varies across `reason` categories (course, home, other, reputation), but the omnibus evidence is limited:

$$F(3, 391) = 2.080, \quad p = 0.102, \quad p_{\text{perm}} = 0.103, \quad \eta^2 = 0.0157.$$

This indicates a small share of variance explained by `reason` in this sample.

4.3 Regression: Model A (excludes prior grades)

Model A explains 21.9% of grade variation ($R^2 = 0.219$, adj. $R^2 = 0.168$). Table 2 reports selected coefficients with HC3 95% CIs.

4.4 Regression: Model B (includes prior grades)

Including prior grades substantially increases explanatory power ($R^2 = 0.842$, adj. $R^2 = 0.831$). The coefficients on prior grades are:

$$\widehat{\gamma}_1 = 0.178 \quad (G1), \qquad \widehat{\gamma}_2 = 0.956 \quad (G2),$$

both statistically precise in this sample. Several covariate coefficients from Model A shrink toward zero once $G1$ and $G2$ are included.

Table 2: Selected coefficients from Model A (OLS with HC3 robust 95% CIs).

Term	Estimate	95% CI (HC3)
Male (vs female)	1.343	[0.392, 2.295]
Failures = 1 (vs 0)	-2.605	[-4.074, -1.136]
Failures = 2 (vs 0)	-4.411	[-6.843, -1.978]
Failures = 3 (vs 0)	-4.936	[-7.332, -2.540]
Family support = yes	-1.005	[-1.937, -0.073]
Weekend alcohol = 4 (vs 1)	-1.354	[-2.641, -0.067]

4.5 Bootstrap check for selected Model A coefficients

Table 3 compares HC3 confidence intervals to pairs-bootstrap percentile intervals (2,000 resamples) for three selected Model A coefficients.

Table 3: Model A coefficient intervals: HC3 vs pairs bootstrap.

Term	HC3 95% CI	Bootstrap 95% CI
Male (vs female)	[0.392, 2.295]	[0.430, 2.281]
Failures = 1 (vs 0)	[-4.074, -1.136]	[-4.030, -1.191]
Study time = 3 (vs 1)	[-0.083, 2.839]	[-0.040, 2.806]

5 Summary

Across methods, prior academic failures show a large negative association with final grades in Model A. The binary group comparisons (sex, internet access) yield modest mean differences with nontrivial uncertainty. Conditioning on prior grades (Model B) changes the fitted relationships substantially and yields much higher explanatory power, consistent with strong academic persistence in this dataset.

Reproducibility: All figures and tables referenced here are generated by `notebooks/01_paper.ipynb` and stored under `reports/figures/` and `reports/tables/`.