

Statistical Analysis Report: Student Performance

Name: Osama

Roll no : DS-034/2025

Course: Statistics & Probability for Data Science (CT-581)

Title: Impact of Demographic and Academic Factors on Student Performance

Introduction

This report analyzes a real-world dataset titled "**Students Performance in Exams**", which includes information on students' scores in math, reading, and writing, alongside demographic details such as gender, parental education level, and test preparation course completion.

The objective is to apply statistical methods to understand:

- Whether the test preparation course improves math scores,
- If there's any association between gender and course participation,
- Whether parental education impacts performance,
- And whether reading scores can predict math scores.

These insights can help in designing educational interventions tailored to student backgrounds.

Descriptive Statistics

We begin with an overview of the student performance data, focusing on math, reading, and writing scores.

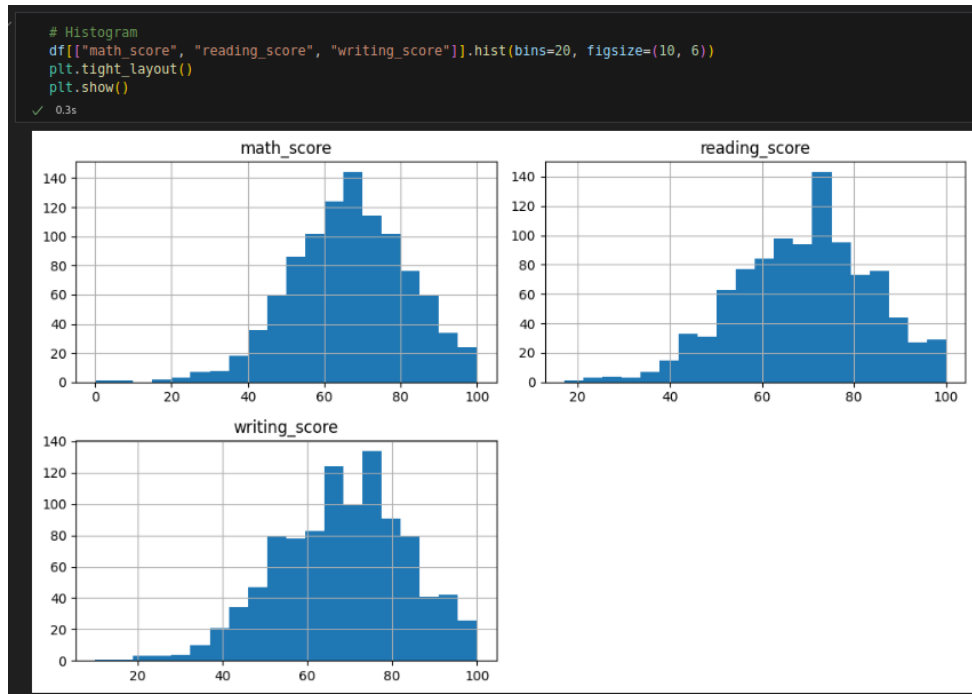
Summary Statistics

Statistic	Math Score	Reading Score	Writing Score
Count	1000	1000	1000
Mean	66.09	69.17	68.05
Std	15.16	14.60	15.20
Min	0.00	17.00	10.00
25%	57.00	59.00	57.75
Median (50%)	66.00	70.00	69.00
75%	77.00	79.00	79.00
Max	100.00	100.00	100.00

Visualizations

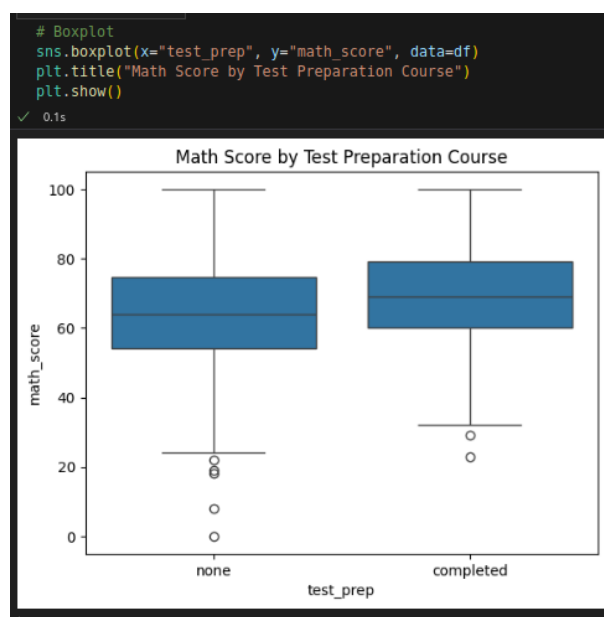
Histograms

- Distributions are approximately normal for all three subjects.
- Majority of scores cluster around the mean (60–80 range).



Boxplot (Math Score vs. Test Prep Course)

- Median math score is higher for students who **completed** the test preparation course.
- Students with no prep show more low-score outliers.



Correlation Heatmap



	Math Score	Reading Score	Writing Score
Math Score	1.00	0.82	0.80
Reading	0.82	1.00	0.89
Writing	0.80	0.89	1.00

Insight: All scores are **strongly positively correlated**, especially between reading and writing.

Methodology

We applied the following statistical techniques using Python:

1. **T-test (Hypothesis Testing):** Compare math scores between students who completed vs. did not complete the test preparation course.
2. **ANOVA:** Examine the effect of parental education level on math scores.
3. **Simple Linear Regression:** Predict math scores using reading scores.
4. **Chi-Square Test:** Determine the association between parental education level and test preparation course completion.

Statistical Tests and Results

1- T-Test: Effect of Test Preparation on Math Score

Objective: To test whether there is a significant difference in math scores between students who completed the test preparation course and those who did not.

Hypotheses:

- **H₀ (Null):** There is no significant difference in math scores between students who completed and did not complete test preparation.
- **H₁ (Alt):** There is a significant difference in math scores between the two groups.

```
completed = df[df["test_prep"] == "completed"]["math_score"]
none = df[df["test_prep"] == "none"]["math_score"]

t_stat, p_val = stats.ttest_ind(completed, none)
print("T-test p-value:", p_val)
```

✓ 0.0s

T-test p-value: 1.5359134607147415e-08

Result:

- **p-value** = 1.5359134607147415e-08

Conclusion: Since $p < 0.05$, we reject H_0 . Test preparation has a statistically significant effect on math scores.

2- Chi-Square Test: Association between Gender and Test Preparation

Objective: To determine whether gender is associated with completing the test preparation course.

Hypotheses:

- **H₀ (Null):** Gender and test preparation completion are independent.
- **H₁ (Alt):** Gender and test preparation completion are associated.

```
contingency = pd.crosstab(df["gender"], df["test_prep"])
chi2, p, dof, expected = stats.chi2_contingency(contingency)

print("Chi-square p-value:", p)
```

✓ 0.0s

Chi-square p-value: 0.9008273880804724

Result:

- **p-value** = 0.9008273880084724

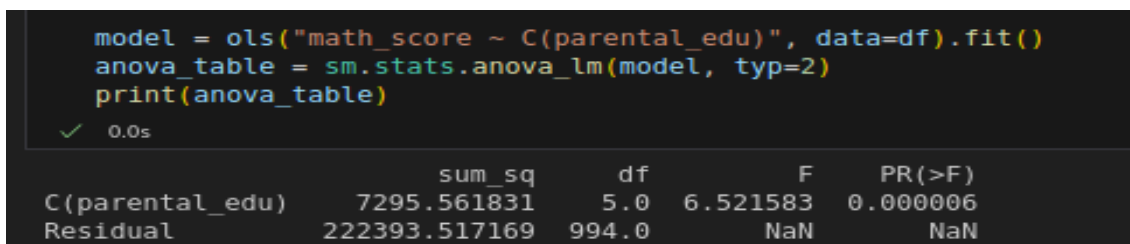
Conclusion: Since $p > 0.05$, we fail to reject H_0 . There is no significant association between gender and completing test preparation.

3- ANOVA: Effect of Parental Education on Math Score

Objective: To test whether parental education level significantly affects students' math scores.

Hypotheses:

- **H_0 (Null):** All parental education groups have equal mean math scores.
- **H_1 (Alt):** At least one group has a different mean.



```
model = ols("math_score ~ C(parental_edu)", data=df).fit()
anova_table = sm.stats.anova_lm(model, typ=2)
print(anova_table)
```

✓ 0.0s

	sum_sq	df	F	PR(>F)
C(parental_edu)	7295.561831	5.0	6.521583	0.000006
Residual	222393.517169	994.0	NaN	NaN

Result:

- **p-value** = 0.000006

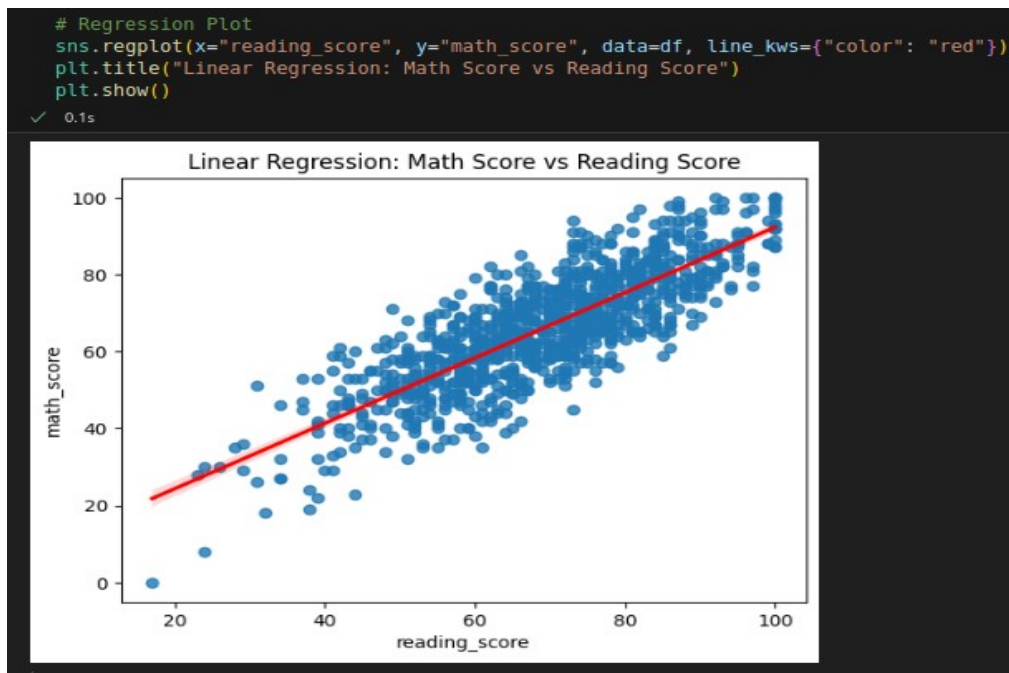
Conclusion: Since $p < 0.05$, we reject H_0 . Parental education level significantly affects math performance.

4- Simple Linear Regression: Math Score vs Reading Score

Objective: To determine whether reading score predicts math score.

Hypotheses:

- **H_0 (Null):** There is no linear relationship between reading and math scores.
- **H_1 (Alt):** There is a linear relationship between reading and math scores.



Result:

- A strong positive trend observed in the regression line and scatter plot.

Conclusion: There is a statistically significant positive linear relationship. Reading score is a strong predictor of math score.

Conclusion

This analysis reveals several insights:

- Completing a test preparation course significantly improves math performance.
- Parental education level influences math scores.
- Gender has no significant impact on test preparation course completion.
- Reading score is a reliable predictor of math score.

These findings can inform policies targeting academic improvement strategies.

Software Used

Python

- Libraries: pandas, scipy.stats, statsmodels, matplotlib, seaborn

Dataset source:

<https://www.kaggle.com/datasets/spscientist/students-performance-in-exams>