# Does Parental Education Level Affect Student Performance?

Bellevue University

Hasnaa Elidrissi

11/16/2024

# Statistical Question & Hypothesis

Does parental education level affect student performance (final grade)?

Null Hypothesis (H0): Parental education levels do not significantly affect the student's final grade.

Alternative Hypothesis (H1): Higher parental education levels positively affect the student's final grade.

# Variables

| Field Name | Description | Type | Possible values |
|---|---|---|---|
| sex | student's sex | binary | 'F' - female or 'M' - male |
| age | student's age | numeric | from 15 to 22 |
| famsize | family size | binary | 'LE3' - less or equal to 3 or 'GT3' - greater than 3 |
| Pstatus | parent's cohabitation status | binary | 'T' - living together or 'A' - apart |
| Medu | mother's education | numeric | 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education) |
| fedu | mother's education | numeric | 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education) |
| Fjob | father's job | nominal | 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other' |
| guardian | student's guardian | nominal | 'mother', 'father' or 'other' |
| studytime | weekly study time | numeric | 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours) |
| failures | number of past class failures | numeric | n if 1<=n<3, else 4 |
| schoolsup | extra educational support | binary | yes or no |
| famsup | family educational support | binary | yes or no |
| G3 | final grade | numeric | from 0 to 20, output target |

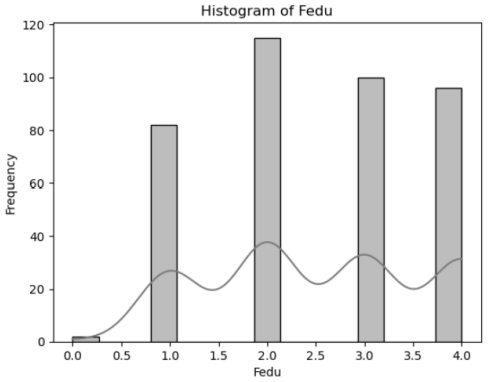
# Descriptive Analysis of Variables

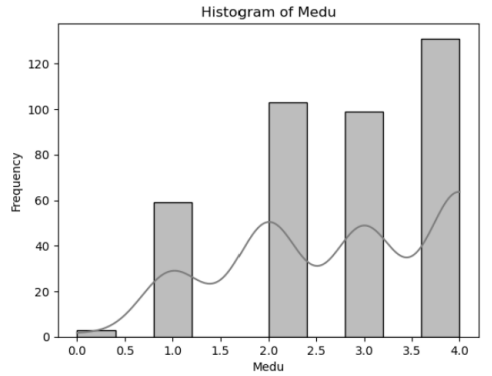
- **Parental Education Levels**: Both Medu and Fedu have means close to 3, indicating that many parents have at least completed secondary education. This could suggest a potential for a positive influence on student performance if higher parental education is correlated with better support for education at home.
- **Study Time**: Most students spend a moderate amount of time studying (2-5 hours per week). It would be interesting to see if those who study more (higher studytime values) perform better (G3).
- **Failures**: The low mean and low percentiles for failures show that most students in the dataset have not failed classes. However, we might expect those with higher failure counts to have lower grades, which could serve as an indicator of academic struggles.
- **Student Performance (G3)**: The average grade of 10.4 and a median of 11, with a wide range, shows that there's a diversity in academic performance among students. The distribution could reveal whether students' grades are clustered around the mean or spread out more widely, which we can further explore through histograms and scatter plots.

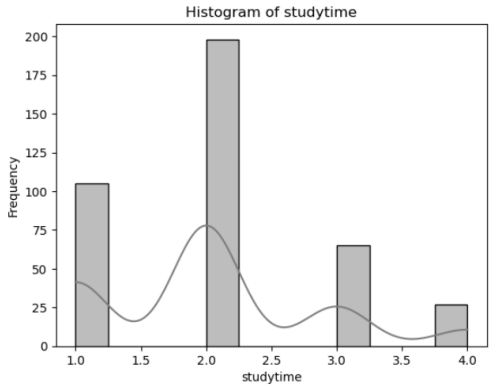
| | age | Medu | Fedu | studytime | failures | G3 |
|---|---|---|---|---|---|---|
| **count** | 395.000000 | 395.000000 | 395.000000 | 395.000000 | 395.000000 | 395.000000 |
| **mean** | 16.696203 | 2.749367 | 2.521519 | 2.035443 | 0.334177 | 10.415190 |
| **std** | 1.276043 | 1.094735 | 1.088201 | 0.839240 | 0.743651 | 4.581443 |
| **min** | 15.000000 | 0.000000 | 0.000000 | 1.000000 | 0.000000 | 0.000000 |
| **25%** | 16.000000 | 2.000000 | 2.000000 | 1.000000 | 0.000000 | 8.000000 |
| **50%** | 17.000000 | 3.000000 | 2.000000 | 2.000000 | 0.000000 | 11.000000 |
| **75%** | 18.000000 | 4.000000 | 3.000000 | 2.000000 | 0.000000 | 14.000000 |
| **max** | 22.000000 | 4.000000 | 4.000000 | 4.000000 | 3.000000 | 20.000000 |

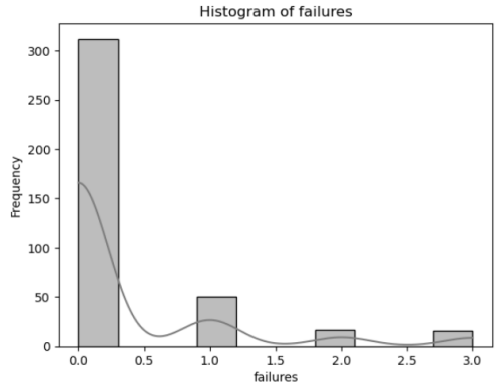
# Distributions

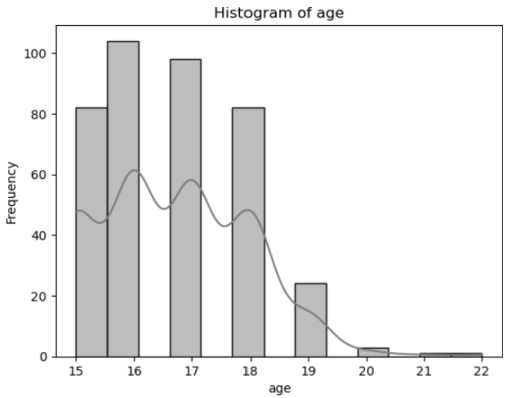
- Parent Education Levels (Medu and Fedu): Most parents have completed secondary education, with some reaching higher education. Low parental education levels (0) are outliers and might impact student performance.
- Study Time and Failures: The skewed distribution of failures and the variation in studytime suggest potential areas to investigate further. Higher failures might correlate with lower grades, and different study times could influence student performance.
- Student Performance (G3): The distribution of G3 appears close to a normal distribution, with most students clustered around the middle of the grading scale. The few very low and high grades could represent students with unique challenges or strengths.

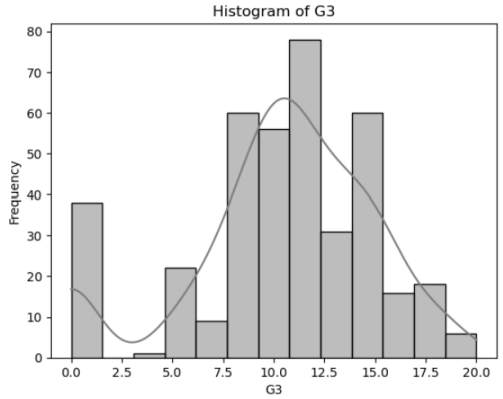


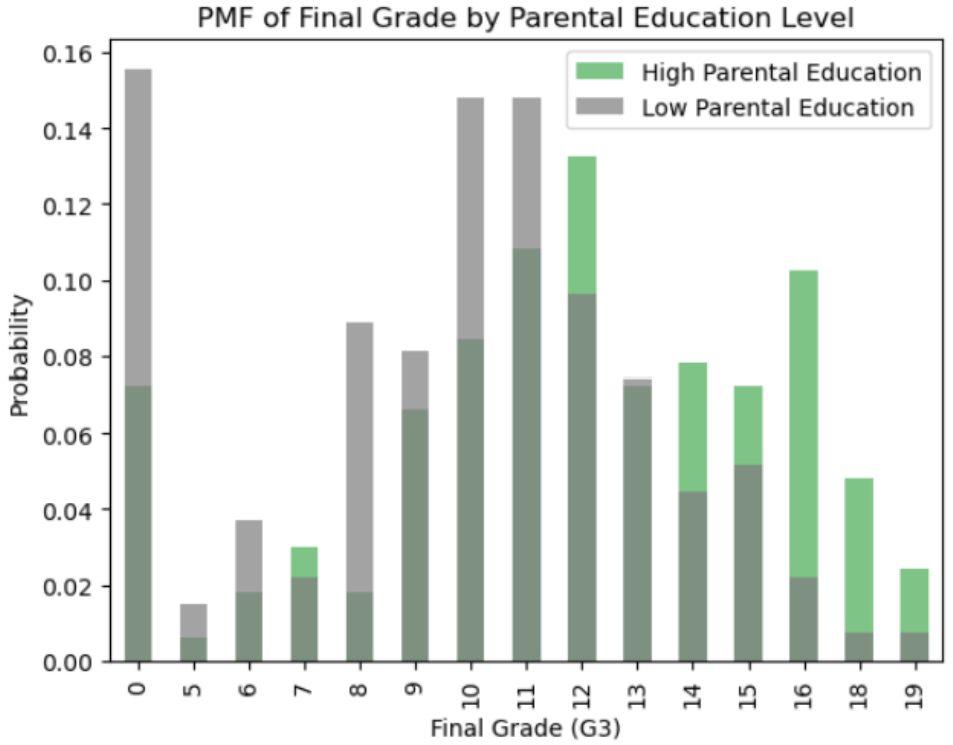



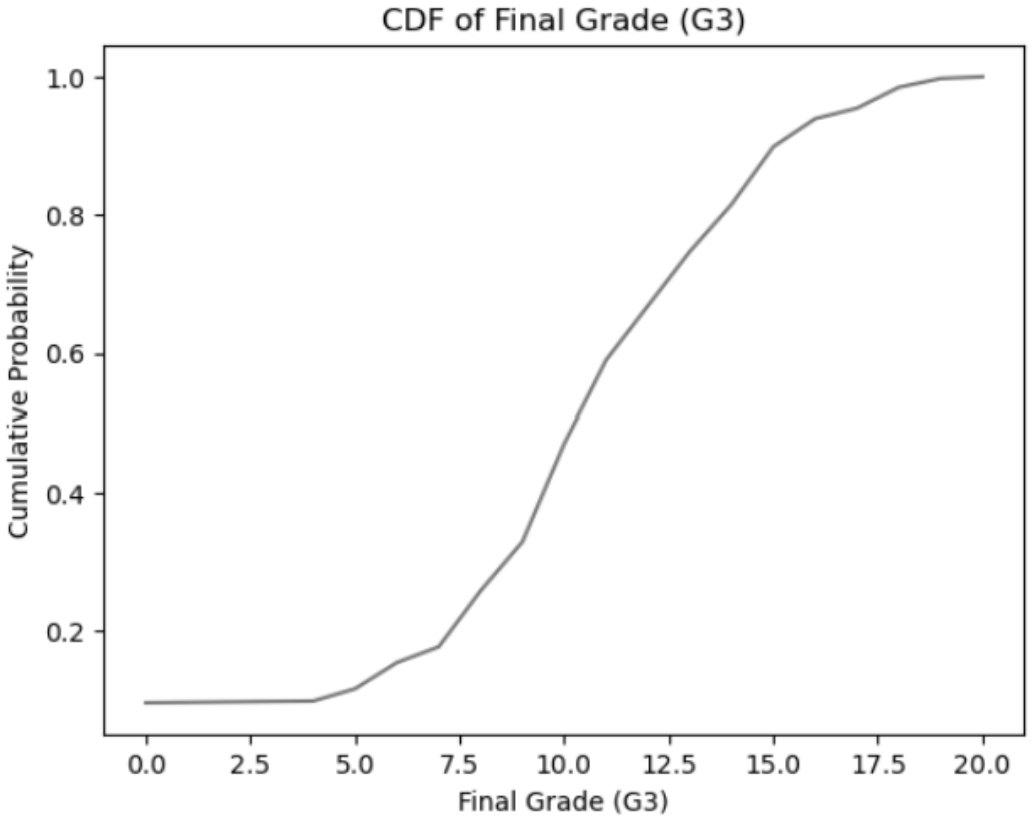
# PMF (Probability Mass Function) Comparison

- **Scenarios**: Students with parents who have higher education levels (e.g., Medu and Fedu ≥ 3) and those with lower levels (e.g., Medu and Fedu < 3).
- **Insights and Implications:**
  - Potential Positive Influence of Parental Education: The chart suggests that higher parental education levels are associated with better academic performance, as students from these backgrounds are less likely to have low grades and more likely to score in the higher range.
  - Educational Support Needs: The data may imply a need for additional support for students from families with lower parental education levels to help bridge the performance gap.
  - Further Analysis: To strengthen these insights, a statistical test (such as a t-test or regression analysis) could confirm the significance of parental education's effect on grades. Additionally, considering other variables like studytime and failures may reveal more about factors influencing student performance.



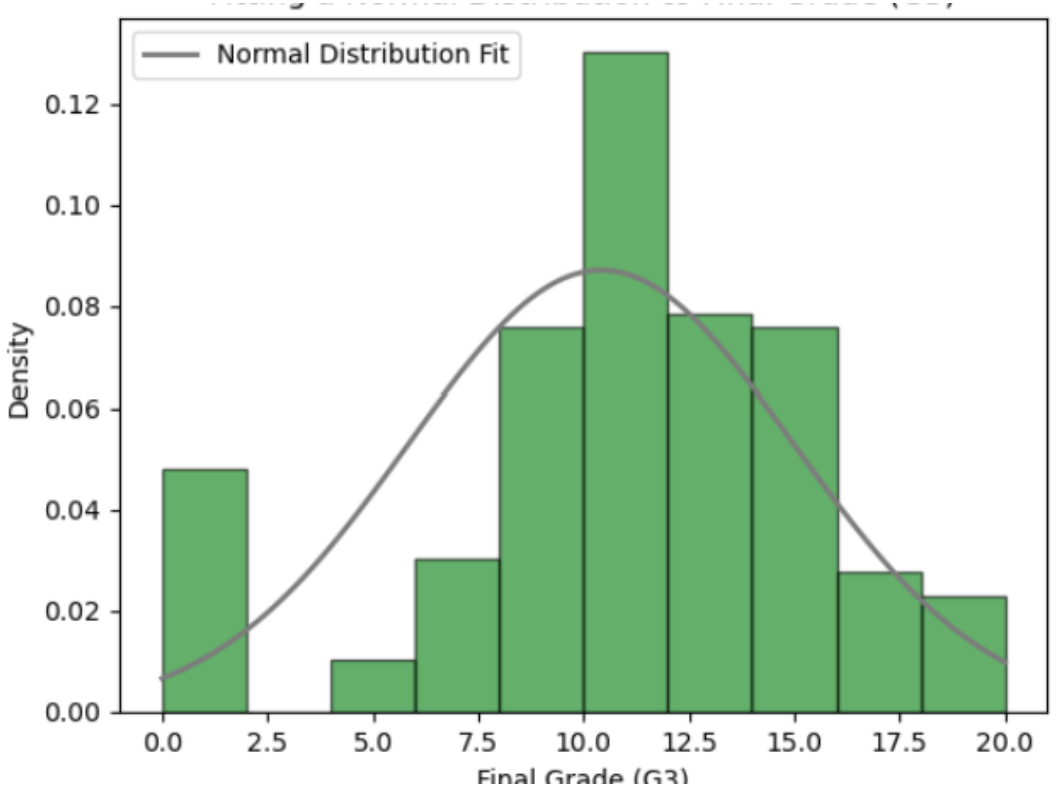
# CDF (Cumulative Distribution Function) of G3

- **Implications and InsightsGrade**
  - Distribution: The CDF shows that most students are performing at an average level, with fewer at the extremes. This could mean that there is limited academic disparity among the majority of students, with only few outliers.
  - Impact on Hypothesis: Since the majority of students fall in the middle range, it may be challenging to attribute high or low parental education as the sole factor influencing performance.
  - Further analysis, such as a regression, might be necessary to examine subtle variations.
  - Potential Interventions: If we aim to improve performance, focusing on moving students from the lower-middle range (grades 8–10) up to higher grades could be a strategic goal, as the CDF shows this as the area with the largest group of students.



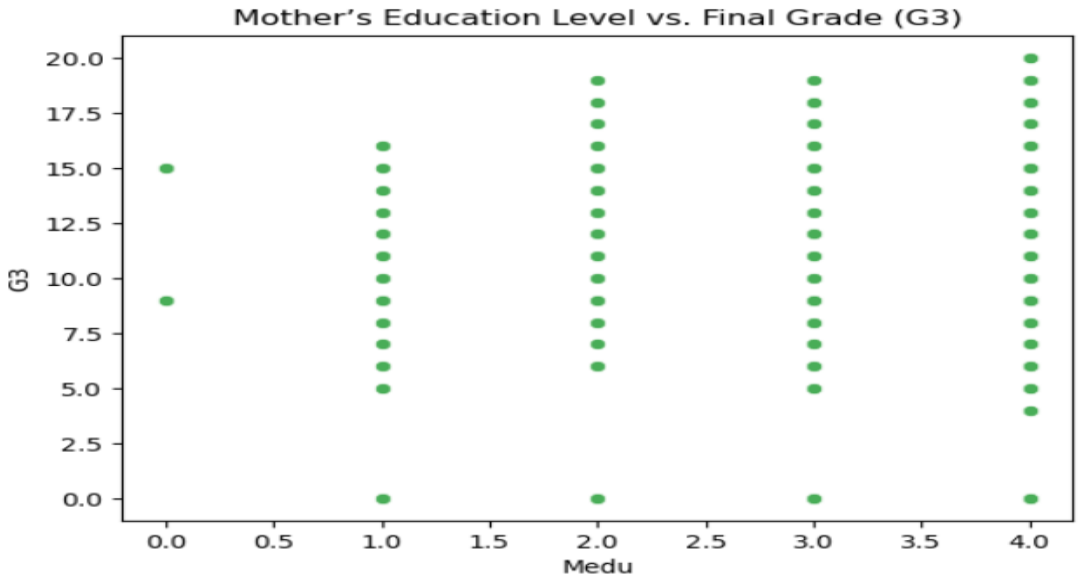
# Analytical distribution

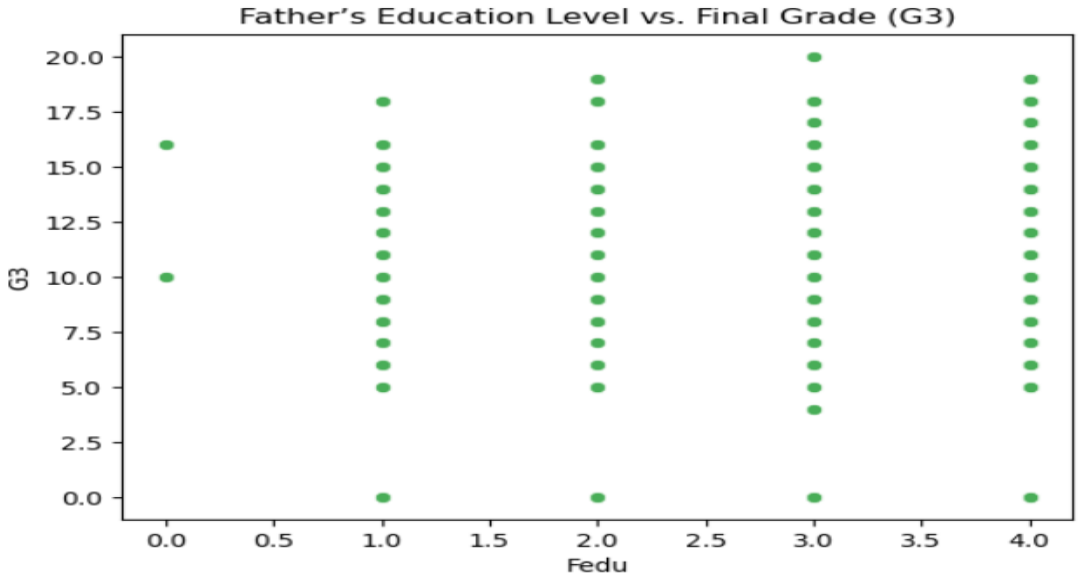
- **Suitability of the Normal Distribution:** Although the normal distribution is not a perfect fit, it provides a reasonable approximation for G3. This allows us to use techniques that assume normality, such as z-scores, to interpret grades relative to the mean.
- **Student Performance:** The fit suggests that most students perform near the mean (grades 10–12), with fewer students at very low or very high scores. This could imply that student performance is generally consistent, with a few outliers who may need additional academic support.
- **Potential for Further Investigation:** Given the slight deviations at the extremes, it may be useful to investigate whether certain factors (such as parental education, study time, or past failures) contribute to students being outliers in their performance.



# Scatter Plot Analysis

- **Weak Correlation:** Based on these scatter plots, there doesn't appear to be a strong linear correlation between parental education levels (Medu and Fedu) and student final grades (G3). The data points are widely dispersed with no clear pattern of incerease or decrease in G3 with higher parental education levels.
- **Need for Further Analysis:** Since the scatter plots don't show a clear relationship, further analysis (such as multiple regression) can help determine if there is any indirect influence when other variables, like study time and failures, are controlled.
- **Potential Moderating Variables:** Other factors, such as socio-economic status or home environment, might moderate the relationship between parental education and student performance, making it more complex than a direct linear relationship.



# Hypothesis Testing

- **P-value Analysis:** The p-value of 0.00034 is significantly less than the common significance level of 0.05. This indicates that there is strong evidence against the null hypothesis.
- **Decision:** Since the p-value is well below 0.05, we reject the null hypothesis.
- **Conclusion:** There is a statistically significant difference in final grades (G3) between students with high and low parental education levels. This suggests that parental education level may have an impact on student performance, with students from higher-educated backgrounds potentially achieving better grades.

# Regression Analysis

**Interpretation**

- Mother's Education (Medu) appears to have a positive and statistically significant effect on G3, implying that higher materal education may be associated with better student performance.
- Father's Education (Fedu) does not show a significant effect on G3, indicating that in this model, father's education level does not significantly contribute to predicting final grades.
- The low R-squared value suggests that other factors, not included in this model, play a more substantial role in determining student performance. Variables such as study habits, socioeconomic status, or school resources could potentially have more influence.

**Conclusion**

- The regression analysis indicates a small but significant relationship between mother's education level and student performance. However, the model explains only a small portion of the variance in grades, suggesting that parental education alone is not a strong predictor of academic success. This result aligns with the t-test findings, which showed a difference in G3 based on parental education but highlights that other factors are likely at play in detrmining student performance.

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | G3 | R-squared: | 0.048 |
| Model: | OLS | Adj. R-squared: | 0.043 |
| Method: | Least Squares | F-statistic: | 9.802 |
| Date: | Mon, 11 Nov 2024 | Prob (F-statistic): | 7.01e-05 |
| Time: | 10:42:50 | Log-Likelihood: | -1151.5 |
| No. Observations: | 395 | AIC: | 2309. |
| Df Residuals: | 392 | BIC: | 2321. |
| Df Model: | 2 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | 7.8205 | 0.648 | 12.073 | 0.000 | 6.547 | 9.094 |
| Medu | 0.8359 | 0.264 | 3.168 | 0.002 | 0.317 | 1.355 |
| Fedu | 0.1176 | 0.265 | 0.443 | 0.658 | -0.404 | 0.639 |

| | | | |
|---|---|---|---|
| Omnibus: | 34.382 | Durbin-Watson: | 1.976 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 40.826 |
| Skew: | -0.765 | Prob(JB): | 1.36e-09 |
| Kurtosis: | 3.375 | Cond. No. | 12.1 |