

Univariate Analysis - Student Performance Dataset

Raju Ahmed

2025-12-27

1. Introduction

In this document, I present my univariate analysis for the Student Performance dataset covering: **Binary Variables** (sex, paid, activities, higher, internet), **Ordinal Variables** (Medu, famrel), **Numeric Variable** (age), and one **Bivariate Analysis** (Medu vs G3).

2. Data Loading

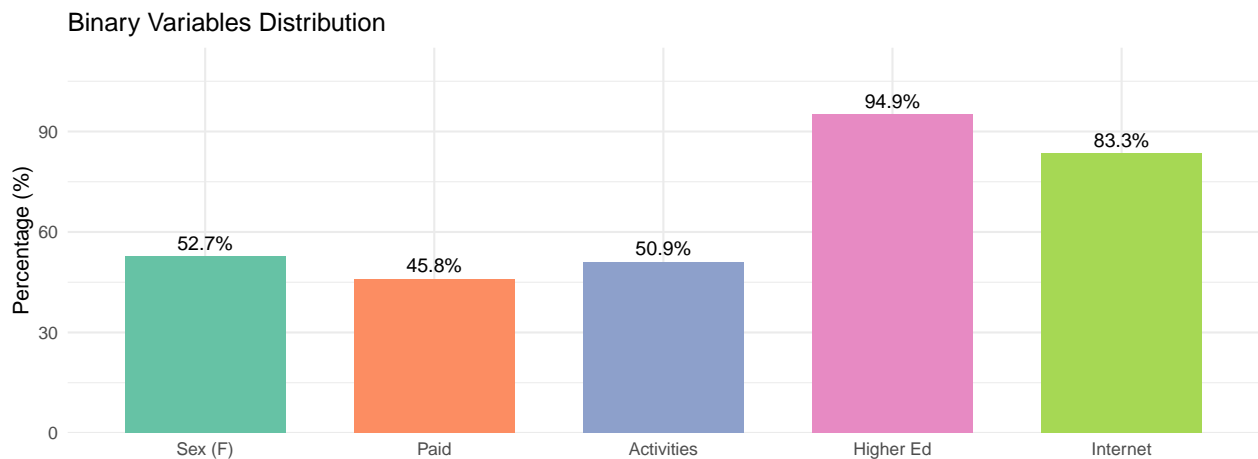
```
data <- read_csv("student-mat-selected.csv")
cat("Dataset:", nrow(data), "rows x", ncol(data), "columns\n")
```

```
## Dataset: 395 rows x 13 columns
```

3. Binary Variables

Table 1: Binary Variables Summary

Variable	Yes/F (n)	Yes/F (%)	No/M (n)	Mode
sex (F/M)	208	52.7	187	F
paid	181	45.8	214	no
activities	201	50.9	194	yes
higher	375	94.9	20	yes
internet	329	83.3	66	yes

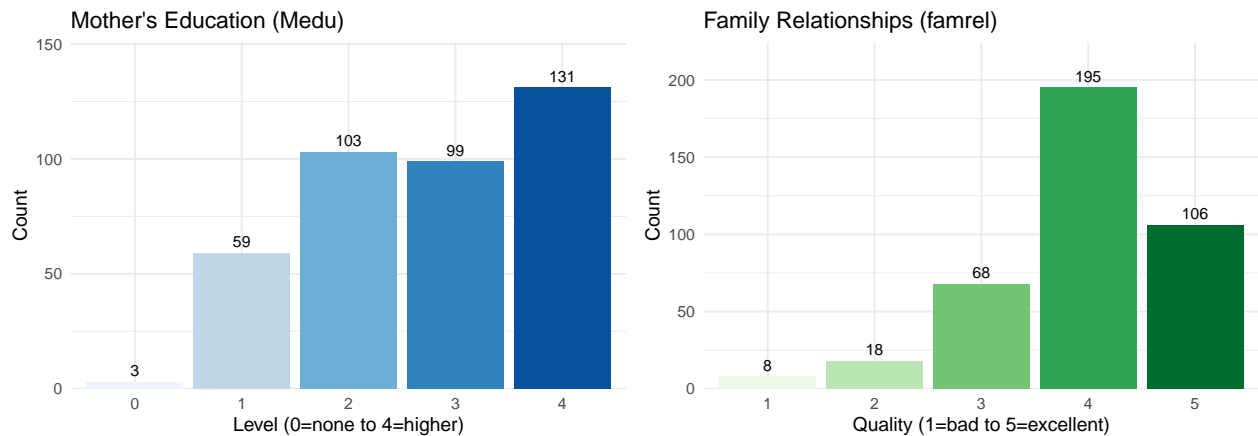


Interpretation: Sex is balanced (52.7% female). 45.8% take paid classes, 50.9% do activities, 94.9% want higher education, 83.3% have internet.

4. Ordinal Variables

Table 2: Ordinal Variables Summary

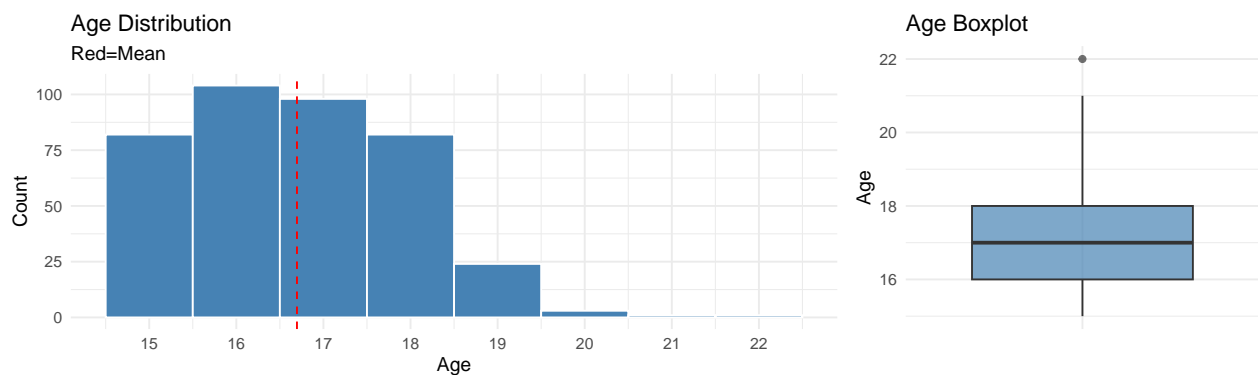
Variable	Mode	Median	Distribution
Medu (0-4)	4	3	3>59, 103>99>131 (skewed to higher)
famrel (1-5)	4	4	4>195, 5>86 (mostly good)



Interpretation: Medu mode=4 (higher ed), median=3. Most mothers have secondary or higher education. Famrel mode=4 (good), median=4. Most students report good family relationships.

5. Numeric Variable: Age

Mean: 16.7 | Median: 17 | Mode: 16 | SD: 1.28 | Range: 15 - 22 | IQR: 2



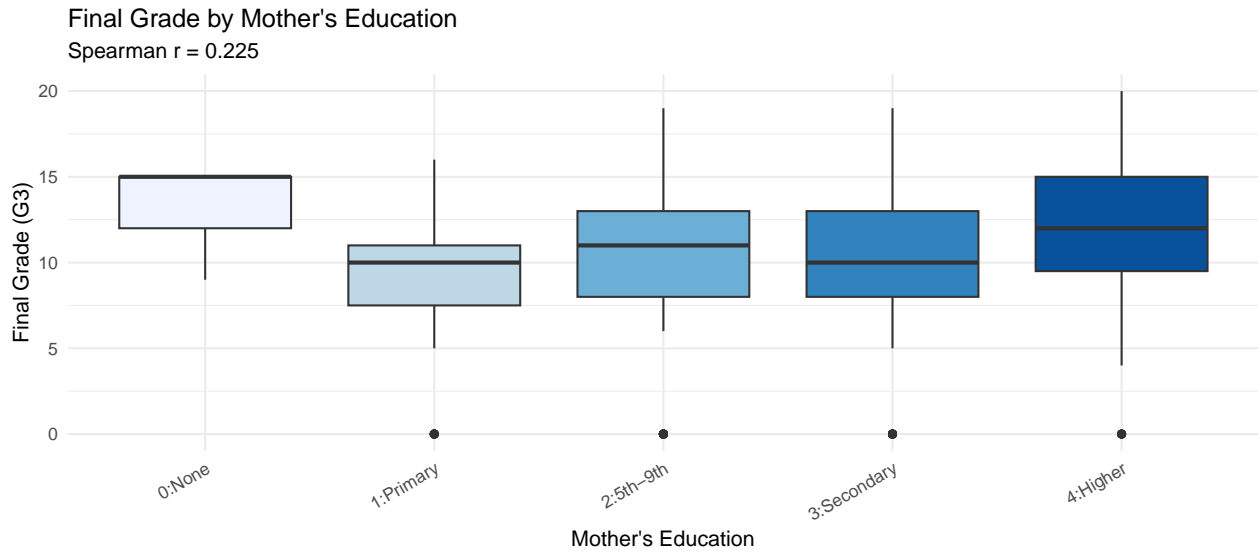
Interpretation: Ages 15-22, mean=16.7, median=17, mode=16. Slightly right-skewed with some older students.

6. Bivariate: Mother's Education vs Final Grade

Table 3: G3 by Medu Level

Medu	n	Mean_G3	SD
0	3	13.00	3.46
1	59	8.68	4.36
2	103	9.73	4.64
3	99	10.30	4.62
4	131	11.76	4.27

Spearman $r = 0.225$



Interpretation: Positive correlation ($r=0.225$). Medu=0 has high mean (13.0) but only $n=3$ - small sample artifact. Excluding Medu=0, clear trend: grades increase from 8.68 (level 1) to 11.76 (level 4). Mother's education is a meaningful predictor of student performance.