

Univariate Analysis - Student Performance Dataset

Kezia Fernandes

2025-12-29

Contents

1	Introduction	1
2	Continuous Variable: G3 (Final Grade)	1
3	Numeric Discrete Variable: Absences	2
4	Numeric Discrete Variable: Failures	2
5	Ordinal Variables: Study Time and Travel Time	3

1 Introduction

This dataset originates from the secondary education domain and focuses on analyzing factors associated with student academic performance in Mathematics at two Portuguese secondary schools. The data capture multiple dimensions of a student's profile, combining academic outcomes, demographic characteristics, and socio-educational factors. Information was collected through a combination of school records (such as grades and absences) and student questionnaires, providing both objective and self-reported measures relevant to educational performance.

For this analysis, a subset of 13 variables was selected to reflect key aspects influencing student achievement while maintaining analytical clarity. These variables include demographic attributes (sex, age), family and background indicators (mother's education level, quality of family relationships), school-related factors (study time, travel time, past failures, absences), support and engagement variables (paid classes, extracurricular activities, internet access), educational aspirations (desire for higher education), and the final Mathematics grade (G3) as the outcome variable.

The dataset contains a mix of binary nominal variables (e.g., sex, internet access), ordinal categorical variables (e.g., study time, travel time, family relationship quality), and numeric discrete variables (e.g., age, failures, absences). The final grade (G3), measured on a scale from 0 to 20, is treated as a continuous numeric variable. This structure makes the dataset well suited for univariate statistical analysis, allowing for an initial exploration of distributions, central tendencies, and variability across different types of educational and socio-demographic factors.

2 Continuous Variable: G3 (Final Grade)

2.1 Descriptive Statistics

Total = 395 | Mean = 10.42 | Median = 11 | Mode = 10 | SD = 4.58 | Variance = 20.99 | CV = 0.44

Five-Number Summary:

Min = 0 | Q1 = 8 | Median = 11 | Q3 = 14 | Max = 20 | IQR = 6

Shape:

Skewness = -0.73

2.2 Visualizations

2.3 Interpretation

The mean final grade is 10.42 with median of 11. The SD of 4.58 shows considerable variability (CV = 0.44). The skewness of -0.73 indicates a left-skewed distribution with more high-performing students. Grades span 0 to 20, with 50% scoring between 8 and 14 (IQR = 6).

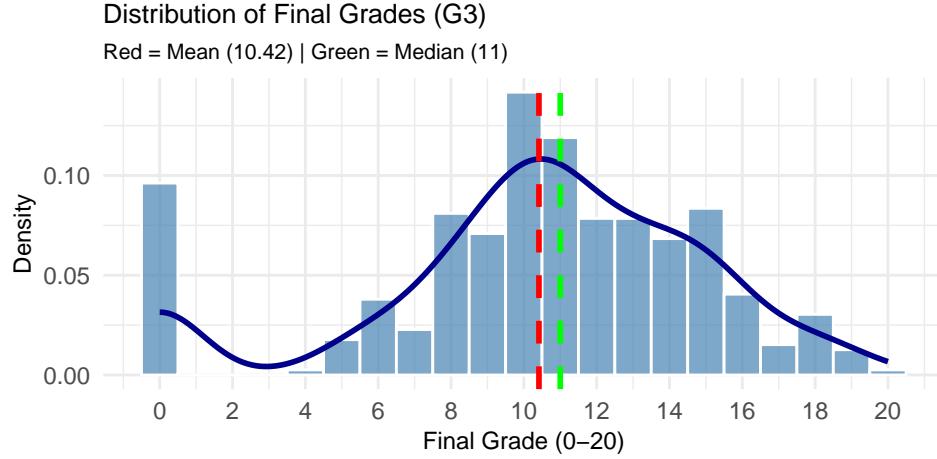


Figure 1: Distribution of Final Grades showing left skewness

3 Numeric Discrete Variable: Absences

3.1 Descriptive Statistics

$N = 395$ | Mean = 5.71 | Median = 4 | Mode = 0 | SD = 8 | Variance = 64.05 | CV = 1.402

Five-Number Summary:

Min = 0 | Q1 = 0 | Median = 4 | Q3 = 8 | Max = 75 | IQR = 8

Shape:

Skewness = 3.658 | Zero absences: 29.1 %

3.2 Visualizations

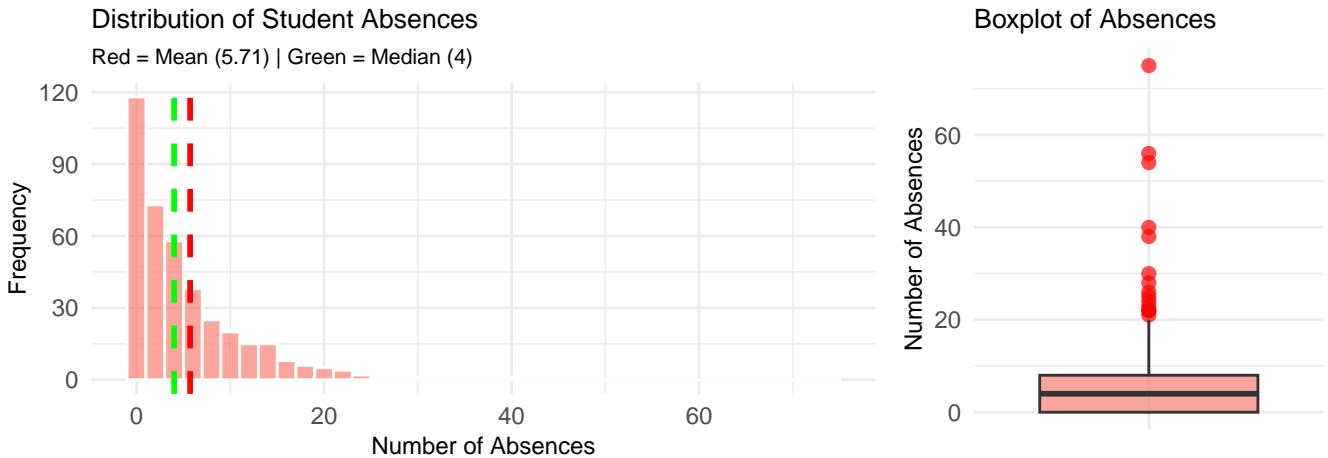


Figure 2: Distribution and outliers in student absences

3.3 Interpretation

Students average 5.71 absences with median of 4, but mode is 0 (29.1% had perfect attendance). The SD of 8.0 is notably large with CV = 1.402, indicating extremely high variability. The skewness of 3.658 shows an extremely right-skewed distribution. Absences range from 0 to 75, with 50% having 0-8 absences (IQR = 8).

4 Numeric Discrete Variable: Failures

4.1 Descriptive Statistics

$N = 395$ | Mean = 0.33 | Median = 0 | Mode = 0 | SD = 0.74 | Variance = 0.55 | Range = 0 - 3

4.2 Frequency Distribution

Table 1: Frequency Distribution of Past Class Failures

Failures	Count	Percentage	Cumulative
0	312	78.99	78.99
1	50	12.66	91.65
2	17	4.30	95.95
3	16	4.05	100.00

4.3 Visualizations

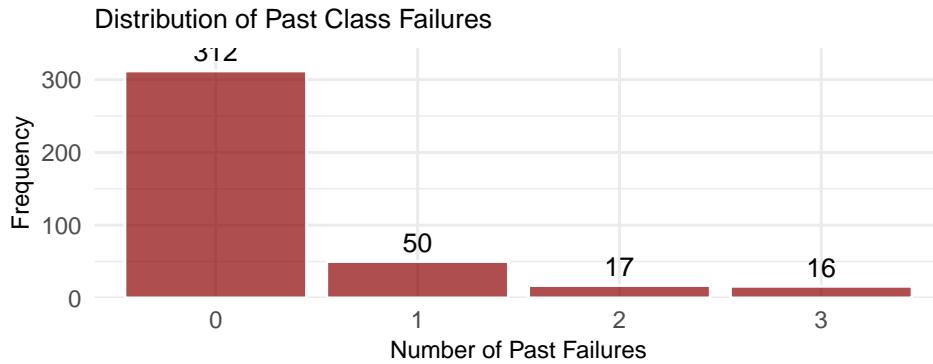


Figure 3: Distribution showing most students never failed

4.4 Interpretation

Mean is 0.33 failures with median and mode of 0. An impressive 78.99% have never failed a class. Only 8.35% have failed 2+ classes, representing a small at-risk group. The SD of 0.74 indicates limited variability.

5 Ordinal Variables: Study Time and Travel Time

Study time categories: 1 = <2 hours/week, 2 = 2-5 hours, 3 = 5-10 hours, 4 = >10 hours.

Travel time categories: 1 = <15 min, 2 = 15-30 min, 3 = 30-60 min, 4 = >60 min.

5.1 Frequency Distributions

Table 1: Study Time Distribution

Category	Count	Percentage	Cumulative
1	105	26.58	26.58
2	198	50.13	76.71
3	65	16.46	93.16
4	27	6.84	100.00

Table 2: Travel Time Distribution

Category	Count	Percentage	Cumulative
1	257	65.06	65.06
2	107	27.09	92.15
3	23	5.82	97.97
4	8	2.03	100.00

Study Time - Central Tendency:

Mode = 2 | Median = 2

Travel Time - Central Tendency:

Mode = 1 | Median = 1

5.2 Visualizations

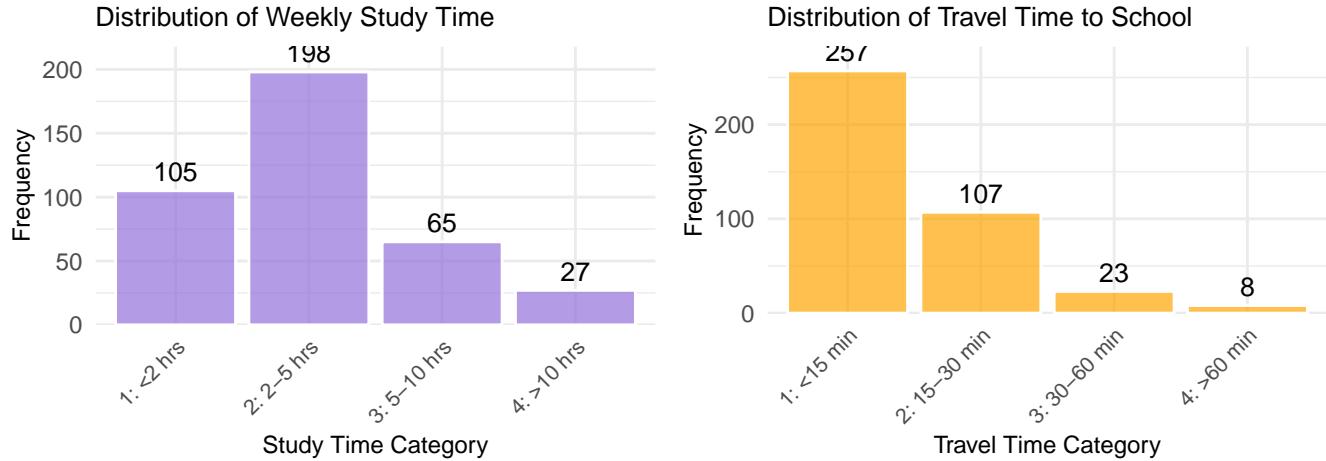


Figure 4: Study time and travel time distributions

5.3 Interpretation

Study Time: Most common is category 2 (2-5 hours/week) with 50.13% of students. Over 26% study <2 hours weekly (potentially insufficient). Only 6.84% study >10 hours. Median of 2 confirms typical student studies 2-5 hours weekly.

Travel Time: Majority (65.06%) live very close to school (<15 min commute). Mode and median are both 1. Additional 27.09% have 15-30 min commutes (92.15% total within 30 min). Only 7.85% face commutes >30 min. Short travel times likely minimize fatigue.