
Análise do Desempenho Escolar em Matemática no Ensino Secundário

Diana Francisco, Lucas Almeida, Rita Pereira

PL3, 202100637

PL3, 202100067

PL3, 202200170

Resumo: Este trabalho analisa o desempenho escolar dos alunos do ensino secundário, no âmbito da disciplina de matemática, de duas escolas portuguesas. Foram escolhidas e estudadas as variáveis: escola; idade; horas de estudo semanal; aulas extras pagas; número de faltas e nota final. A análise incluiu os pontos: estatística descritiva; testes de ajustamento; testes de independência e regressão linear. Conclui-se que as aulas extras pagas influenciam ligeiramente as horas de estudo, e que não existe uma correlação significativa entre o número de faltas e as notas finais.

Palavras-chave: desempenho escolar; análise estatística; ajustamento; independência; regressão-linear.

1. Introdução

O trabalho teve como objetivo a análise do desempenho dos alunos do ensino secundário, em matemática, através dos dados recolhidos de duas escolhas portuguesas. A base de dados foi obtida do repositório “UCI” e através da mesma, foram selecionadas variáveis convenientes às etapas da análise que seria desenvolvida neste trabalho. A análise visa identificar padrões e relações entre as variáveis escolhidas, contribuindo para a compreensão dos fatores que influenciam o desempenho escolar.

2. Base de Dados

A base de dados utilizada foi retirada do repositório “UCI”: <https://archive.ics.uci.edu/dataset/320/student+performance>, e a mesma contém informação sobre um total de 395 alunos, de duas escolhas portuguesas: Gabriel Pereira e Mousinho da Silveira.

As variáveis escolhidas para análise foram:

- Escola: qualitativa nominal (GP, MS)
- Idade: quantitativa contínua (15-22 anos)
- Horas de estudo semanal: qualitativa ordinal (1-4 níveis)
- Aulas extras pagas: qualitativa nominal (sim, não)
- Número de faltas: quantitativa discreta (0-75 faltas)
- Nota final: quantitativa contínua (0-20 valores)

Esta base de dados é composta por 395 alunos, sendo 349 da escola Gabriel Pereira e 46 alunos da escola Mousinho da Silveira.

2.1. Análise Descritiva

A maioria dos alunos (88%) frequenta a escola Gabriel Pereira (GP), enquanto apenas 12% estão na escola Mousinho da Silveira (MS). Essa disparidade é visível no **Gráfico 1**, que mostra a predominância da escola GP.

Tabela 1 – Frequências absolutas e relativas da variável “escola”

Escola	Frequências absolutas (ni)	Frequências relativas (fi)
GP	349	0.8835
MS	46	0.1165

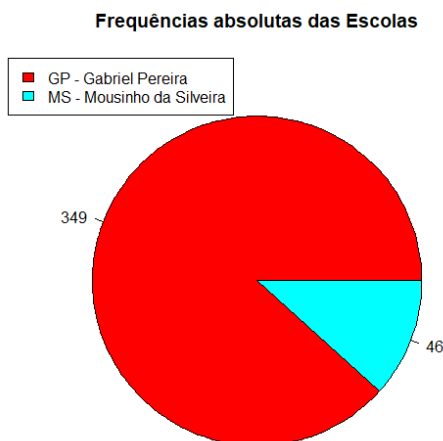


Gráfico 1 – Distribuição das escolas (GP, MS)

A variável foi categorizada em quatro níveis:

- Menos de 2 horas (27% dos alunos).
- Entre 2 e 5 horas (50%).
- Entre 5 e 10 horas (16%).
- Mais de 10 horas (7%).

O **Gráfico 2** ilustra que a maioria dos alunos estuda entre 2 e 5 horas por semana, seguido por aqueles que estudam menos de 2 horas. Poucos alunos dedicam mais de 10 horas semanais ao estudo.

Tabela 2 – Frequências absolutas e relativas da variável “*estudo semanal*”

Estudo semanal	Frequências absolutas (ni)	Frequências relativas (fi)	Frequências absolutas acumuladas (NI)	Frequências relativas acumuladas (FI)
<2h	105	0.2658	105	0.2658
entre 2 e 5h	198	0.5013	303	0.7671
entre 5 e 10h	65	0.1646	368	0.9316
>10h	27	0.0684	395	1

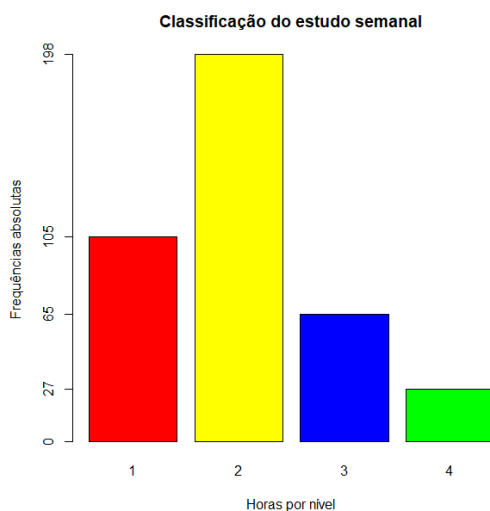


Gráfico 2 – Distribuição das horas de estudo semanal (4 níveis)

Cerca de 54% dos alunos não pagam por aulas extras, enquanto 46% optam por esse recurso. O **Gráfico 3** mostra uma distribuição quase equilibrada entre as duas categorias.

Tabela 3 – Frequências absolutas e relativas da variável “*aulas extras pagas*”

Aulas extras pagas	Frequências absolutas (ni)	Frequências relativas (fi)
não	214	0.5418
sim	181	0.4582

Frequências absolutas das aulas extra pagas

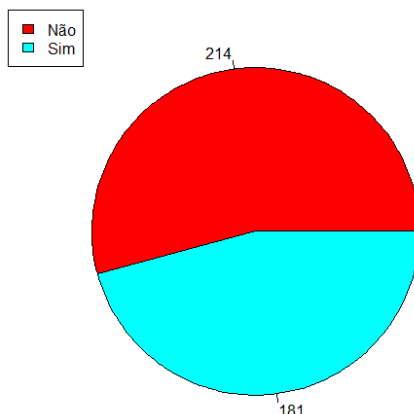


Gráfico 3 – Distribuição das aulas extras pagas (sim, não)

A idade dos alunos varia entre 15 e 22 anos, com a maior concentração nas idades de 16 anos (26%) e 17 anos (25%). O **Gráfico 4** destaca essa distribuição, mostrando que a maioria dos alunos está na faixa etária prevista para o ensino secundário.

Tabela 4 – Frequências absolutas e relativas da variável “idade”

Idade	Frequências absolutas (ni)	Frequências relativas (fi)	Frequências absolutas acumuladas (NI)	Frequências relativas acumuladas (FI)
15	82	0.2076	82	0.2076
16	104	0.2633	186	0.4709
17	98	0.2481	284	0.7190
18	82	0.2076	366	0.9266
19	24	0.0608	390	0.9873
20	3	0.0076	393	0.9949
21	1	0.0025	394	0.9975
22	1	0.0025	395	1

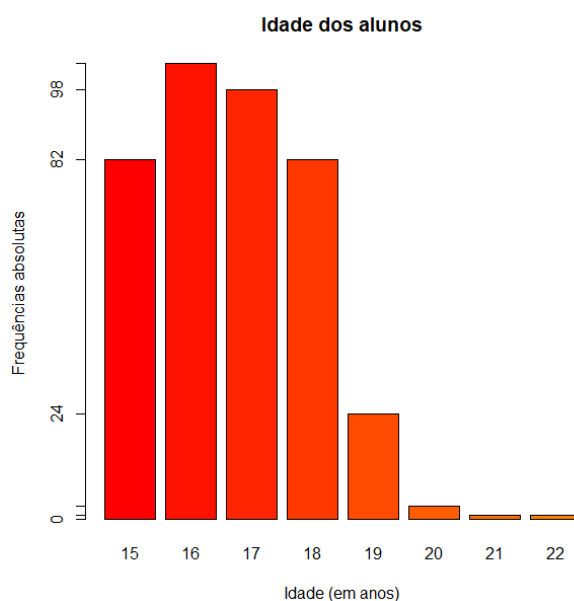


Gráfico 4 – Distribuição da idade (15-22 anos)

A maioria dos alunos (78%) tem menos de 8,33 faltas, enquanto apenas 3% ultrapassam 33,3 faltas. O **Gráfico 5** revela uma distribuição assimétrica, com a maior parte dos alunos a apresentar um número baixo de faltas.

Tabela 5 – Frequências absolutas e relativas da variável “número de faltas”

NºFaltas	Frequências absolutas (ni)	Frequências relativas (fi)	Frequências absolutas acumuladas (NI)	Frequências relativas acumuladas (FI)
[0;8.33[309	0.7823	309	0.7823
[8.33;16.7[60	0.1519	369	0.9342
[16.7;25[17	0.0430	386	0.9772
[25;33.3[4	0.0101	390	0.9873
[33.3;41.7[2	0.0051	392	0.9924
[41.7;50[0	0.0000	392	0.9924
[50;58.3[2	0.0051	394	0.9975
[58.3;66.7[0	0.0000	394	0.9975
[66.7;75]	1	0.0025	395	1

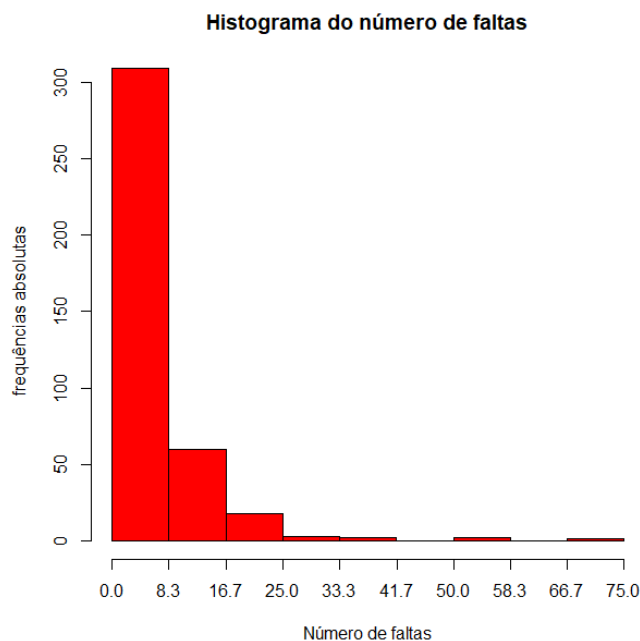


Gráfico 5 – Distribuição do número de faltas (0-75 faltas)

As notas finais variam entre o 0 e 20 valores, com a maior concentração (23%) na classe [7,67; 9,56[. O **Gráfico 6** mostra uma distribuição onde os alunos agrupam-se em diferentes níveis de desempenho.

Tabela 6 – Frequências absolutas e relativas da variável “*nota final*”

Nota final	Frequências absolutas (ni)	Frequências relativas (fi)	Frequências absolutas acumuladas (NI)	Frequências relativas acumuladas (FI)
[2; 3.89[8	0.0203	8	0.0203
[3.89;5.78[20	0.0506	28	0.0709
[5.78; 7.67[45	0.1139	73	0.1848
[7.67;9.56[89	0.2253	162	0.4101
[9.56;11.4[58	0.1468	220	0.5570
[11.4;13.3[78	0.1975	298	0.7544
[13.3;15.2[54	0.1367	352	0.8911
[15.2; 17.1[31	0.0785	383	0.9696
[17.1;19]	12	0.0304	395	1

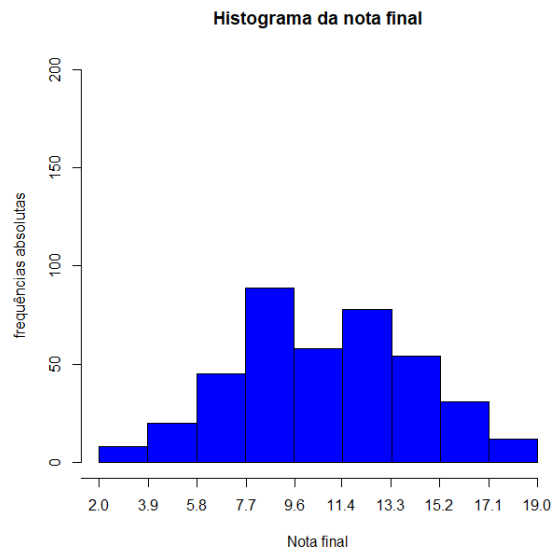


Gráfico 6 – Distribuição da nota final (0-20 valores)

2.2. Testes de Ajustamento

Foram realizados testes de ajustamento para verificar a distribuição de origem das variáveis escolhidas:

- Número de faltas: testar para uma distribuição Binomial negativa.
 - H_0 : o número de faltas segue uma distribuição binomial negativa.
 - H_1 : o número de faltas não segue uma distribuição binomial negativa.
 - Tipo de teste: teste de ajustamento do qui-quadrado.
 - Decisão pelo p-value: aceitar H_0 , com um p-value de 0.123 para um nível de significância de 5%.
 - Decisão pela região crítica: aceitar H_0 , com uma estatística de teste de 4.1833 para uma região crítica $[5.9915; +\infty[$.
- Nota final: testar para uma distribuição Normal.
 - H_0 : a nota final segue uma distribuição normal.
 - H_1 : a nota final não segue uma distribuição normal.
 - Tipo de teste: teste Lilliefors.

- Decisão pelo p-value: rejeitar H_0 , com um p-value de 0.0017 para um nível de significância de 5%.
- Idade: testar para uma distribuição Qui-Quadrado.
 - H_0 : a idade segue uma distribuição qui-quadrado.
 - H_1 : a idade não segue uma distribuição qui-quadrado.
 - Tipo de teste: teste de ajustamento do qui-quadrado.
 - Decisão pelo p-value: rejeitar H_0 , com um p-value de 0 para um nível de significância de 5%.
 - Decisão pela região crítica: rejeitar H_0 , com uma estatística de teste de 378.13 para uma região crítica $[5.9915, +\infty[$.

2.3. Teste de Independência e Regressão Linear Simples

Foi realizado um teste de independência entre as variáveis: aulas extras pagas e horas de estudo semanal.

Tabela 7 – Tabela de contingência entre a variável “*aulas extras pagas*” e a variável “*horas de estudo semanal*”

Aulas extras pagas	<2h	2 a 5h	5h a 10h	>10h
não	73	99	31	11
sim	32	99	34	16

As hipóteses são:

- H_0 : as variáveis são independentes.
- H_1 : as variáveis são dependentes.

Ao fazer um teste do qui-quadrado obtivemos um p-value de 0.0023, ou seja, com base no p-value a decisão é de rejeitar H_0 , para um nível de significância de 5%. Pela região crítica a decisão de rejeitar H_0 mantêm-se, tendo uma estatística de teste de 14.418 e uma região crítica $[7.8147; +\infty[$. As medidas de associação foram as seguintes:

- Coeficiente de contingência: 0.1876.
- V de Cramér: 0.1910.

Na regressão linear simples foi feita com a variável da nota final e número de faltas.

- Variável dependente: nota final.
- Variável independente: número de faltas.
- Correlação de Pearson: -0.012, ou seja, fraca.

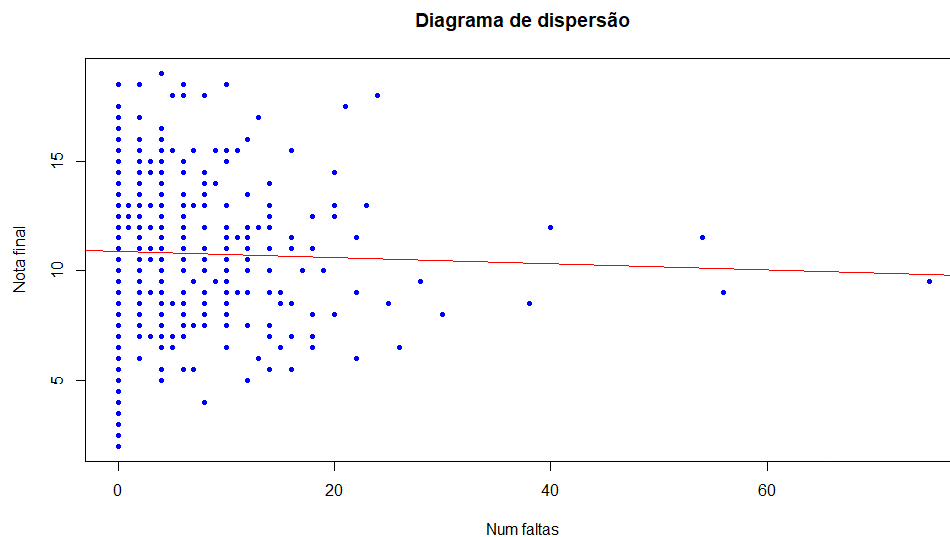


Gráfico 7 – Diagrama de dispersão com a reta de regressão linear.

Podemos concluir que não há uma relação linear significativa entre faltas e notas.

3. IA

A utilização de IA neste trabalho foi limitada ao RStudio. Ao surgir dificuldades com o código do script foi necessário recorrer à ferramenta de IA: “*chatgpt*” em casos onde os slides de apoio teórico não foram o suficiente para resolver determinados erros. Foi também usada com a intenção de apoiar na escrita de comentários relevantes para o melhor entendimento do código.

4. Conclusões

Por fim, ao ter realizado este trabalho foi possível chegar às seguintes conclusões:

- A nota final e a idade não seguem as distribuições sugeridas, enquanto o número de faltas segue uma distribuição binomial negativa.
- As aulas extras pagas influenciam fracamente as horas de estudo semanal.
- Não existe uma correlação significativa entre o número de faltas e as notas finais.

Referências Bibliográficas

- OpenAI. ChatGPT. Disponível em <https://www.openai.com/chatgpt>.

-
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
 - UCI Machine Learning Repository. Student Performance Data Set. Disponível em <https://archive.ics.uci.edu/dataset/320/student+performance>.