

**FUNDAÇÃO GETULIO VARGAS - FGV**

**TRABALHO I – STUDENT PERFORMANCE**

Ana Paula Pudo  
Davi Khoury  
Fábio Monteiro  
Lucas Sena Alves  
Marcos Soares

**MBA Business Analytics e Big Data | Inferência  
Estatística**

Prof. João Rafael Dias Pinto

Brasília  
2021

## SUMÁRIO

RELAÇÃO DE ILUSTRAÇÕES .....	3
TABELAS .....	3
GRÁFICOS .....	3
1. Introdução .....	4
1.1 Objetivo Geral .....	4
1.2 Objetivos Específicos .....	4
2. Metodologia .....	5
2.1 Problema .....	6
3. Descrição do conjunto de dados .....	6
Tabela 1: Análise quantitativa e qualitativa das variáveis .....	7
4. Rendimento geral dos alunos .....	7
5. Análise exploratória univariada dos dados: panorama geral dos alunos .....	8
5.1 Distribuição por escola e gênero .....	8
5.2 Distribuição por localização .....	8
5.3 Distribuição por situação parental e família .....	9
5.4 Distribuição por desempenho, estrutura e apoio .....	11
5.5 Distribuição por aspectos pessoais .....	12
6. Testes de hipótese e análise bivariada em relação ao rendimento escolar .....	13
6.1 Regressão linear - desempenho de alunos por idade .....	18
6.2 Diagrama sobre o desempenho de alunos por localização (urbano ou rural) .....	18
6.3 Diagrama sobre o desempenho por grau de escolaridade da mãe .....	19
6.4 Diagrama sobre o desempenho por grau de escolaridade do pai .....	20
6.5 Diagrama sobre o desempenho por grau de profissão da mãe .....	20
6.6 Diagrama sobre o desempenho por reprovações .....	21
6.7 Diagrama sobre o desempenho por reforço escolar .....	22
7. Diferenças de rendimento entre as duas escolas (disciplina matemática) .....	22
8. Resultados e conclusões .....	23
9. Referências Bibliográficas .....	24

## **RELAÇÃO DE ILUSTRAÇÕES**

### **TABELAS**

Tabela 1: Análise quantitativa e qualitativa das variáveis

### **GRÁFICOS**

Gráfico 1: Aluno x Rendimento Escolar  
Gráfico 2: Aluno x Desempenho Escolar  
Gráfico 3: Aluno x Distribuição de notas dos Estudantes  
Gráfico 4: Análise univariada de Alunos x Escola  
Gráfico 5: Análise univariada de Alunos x Sexo  
Gráfico 6: Análise univariada de Alunos x Endereço  
Gráfico 7: Análise univariada Locomoção Casa x Escola  
Gráfico 8: Análise univariada de Alunos por motivo de escolha da escola  
Gráfico 9: Análise univariada - Relacionamento Familiar  
Gráfico 10: Análise univariada Tamanho da Família  
Gráfico 11: Análise univariada Situação Conjugal  
Gráfico 12: Análise univariada Guarda Parental  
Gráfico 13: Análise univariada Nível Escolar dos Pais  
Gráfico 14: Análise univariada Nível Escolar das Mães  
Gráfico 15: Análise univariada Profissão das Mães  
Gráfico 16: Análise univariada Profissão dos Pais  
Gráfico 17: Análise univariada Reprovações  
Gráfico 18: Análise univariada Tempo de Estudo  
Gráfico 19: Análise univariada - Acesso à internet em casa  
Gráfico 20: Análise univariada - Suporte Familiar  
Gráfico 21: Análise univariada - Reforço Escolar  
Gráfico 22: Análise univariada - Desejam o Curso Superior  
Gráfico 23: Análise univariada - Ativ. Extracurriculares  
Gráfico 24: Análise univariada - Frequentaram o Maternal  
Gráfico 25: Análise univariada - Alunos que estão em um relacionamento  
Gráfico 26: Análise univariada - Tempo Livre na Semana  
Gráfico 27: Análise univariada – Saídas com amigos  
Gráfico 28: Análise univariada - Consumo Álcool Semana  
Gráfico 29: Análise univariada - Consumo Álcool Final de Semana  
Gráfico 30: Análise univariada - Situação de Saúde  
Gráfico 31: Distribuição de Rendimento Por Sexo  
Gráfico 32: Regressão linear simples de idade e desempenho  
Gráfico 33: Análise bivariada de tipo de endereço e desempenho  
Gráfico 34: Análise bivariada por moradia rural e urbana versus desempenho  
Gráfico 35: Análise bivariada desempenho escolar e escolaridade da mãe  
Gráfico 36: Análise bivariada desempenho escolar e escolaridade do pai

Gráfico 37: Análise bivariada desempenho escolar e profissão da mãe

Gráfico 38: Análise bivariada desempenho escolar e repetências

Gráfico 39: Análise bivariada desempenho escolar e reforço escolar

Gráfico 40: Análise bivariada desempenho entre escolas

## **1. Introdução**

O nível de educação de uma determinada sociedade é um dos elementos mais importantes para definir sua capacidade de realizar um desenvolvimento econômico sustentável. Segundo Abrão (2018), existem diversas formas de se medir o impacto da Educação na economia. Um país mais educado gasta menos com saúde pública, tem níveis de segurança mais elevados, já que apresenta criminalidade mais baixa. Maior escolaridade faz com que, direta ou indiretamente, uma economia funcione com mais eficiência.

Um exemplo prático disso pode ser percebido em Portugal. Em 2006, o país registrou o pior PIB per capita dos membros da zona do euro, segundo dados da consultoria Eurostat. Uma das explicações para este desempenho está nos seus níveis de educação básica. No mesmo ano, a taxa de desistência prematura nas escolas em Portugal foi de aproximadamente 40% para a faixa etária entre 18 - 24 anos, enquanto a média na UE foi de 15%.

Tal fenômeno acontece, entre outros fatores, pelo desempenho insuficiente em disciplinas básicas, como língua portuguesa e matemática, que fornecem bases importantes para outras disciplinas.

Para verificar quais variáveis podem impactar o desempenho de estudantes na realidade portuguesa, este estudo irá analisar uma base de dados do ano letivo 2005/2006 de estudantes do ensino secundário, proveniente de um questionário dado aos alunos e que também considerou reports técnicos sobre o desempenho em duas disciplinas: matemática e língua portuguesa. Nesta análise, será considerada apenas as notas da disciplina de matemática.

### **1.1 Objetivo Geral**

Analisar os aspectos socioeconômicos dos estudantes e verificar quais variáveis podem impactar no desempenho escolar, para que as autoridades possam definir políticas públicas que melhorem a performance acadêmica nos próximos anos.

### **1.2 Objetivos Específicos**

- Analisar e categorizar as informações dos estudantes
- Avaliar o rendimento geral dos alunos
- Identificar o panorama geral dos estudantes e realidade socioeconômica a que eles se inserem
- Identificar quais fatores podem interferir no desempenho escolar

- Propor sugestões para que os fatores de interferência possam ser contornados

## **2. Metodologia**

O estudo deste documento foi produzido a partir dados levantados pelo pesquisador Paulo Cortez, coordenador do grupo de pesquisa em Sistemas de Informação e Tecnologias da Universidade do Minho, localizada na cidade de Braga, em Portugal. A base de dados foi disponibilizada e baixada no site da Universidade da Califórnia em Irvine (UCI). As informações são referentes ao ano letivo 2005/2006.

A base de dados deste estudo é composta por 33 variáveis, entre notas dos alunos, características demográficas, socioeconômicas e acadêmicas, matriculados em duas escolas de Portugal: Gabriel Pereira e Mousinho da Silveira. Neste estudo, o recorte feito considerou apenas a disciplina de matemática.

A partir das análises realizadas, foi possível traçar um panorama geral das variáveis que influenciam no desempenho acadêmico dos alunos da matéria de matemática. As técnicas nos testes de hipóteses utilizadas foram Test t, Anova e Regressão Linear, além de técnicas de classificação das variáveis e de análise exploratória dos dados.

Como parâmetros adotados nas averiguações estatísticas demonstrados neste estudo, foram adotados como ponto de corte o driver de P-value como limiar no valor de 5% (P-value estatisticamente significativo  $\leq 0,05$ ). O percentual do intervalo de confiança ficou estabelecido em 95%.

As análises que serão descritas, tiveram como variável alvo as médias escolares dos alunos das escolas onde foram realizados os levantamentos e aquisições dos dados objeto deste estudo. A partir do alvo foram realizados os cruzamentos, nas análises com os demais elementos disponibilizados na base, na identificação de elementos estatisticamente significantes na extração de informações.

Todos os dados foram trabalhados no RStudio, onde também realizamos a plotagem de gráficos e tabelas. Para calcular todas as variáveis e realizar as análises necessárias, utilizamos códigos de programação da linguagem R.

As análises realizadas no documento foram divididas em seis partes. Após uma descrição e classificação inicial dos dados, será detalhado o desempenho e panorama gerais dos estudantes e as variáveis que podem interferir no desempenho, através de uma análise univariada dos dados. Na sequência, serão explorados os testes de hipóteses e as análises bivariadas relacionadas ao rendimento escolar. Serão analisadas também se as escolas possuem diferenças de desempenho e a sumarização dos resultados.

## 2.1 Problema

Considerando os aspectos socioeconômicos dos alunos e o desempenho acadêmico deles ao longo do ano letivo 2005/2006, este estudo se propõe a avaliar a correção entre a realidade dos estudantes e as notas obtidas no período. Quais fatores interferem no desempenho acadêmico? Quais são irrelevantes para boas notas? O que poderia ser feito para melhorar a performance dos alunos na disciplina avaliada?

## 3. Descrição do conjunto de dados

A base de dados analisada conta com 33 variáveis, classificadas da seguinte forma:

Item	Variáveis	Descrição das Variáveis
1	school - student's school (binary: "GP" - Gabriel Pereira or "MS" - Mousinho da Silveira)	Qualitativa Nominal
2	sex - student's sex (binary: "F" - female or "M" - male)	Qualitativa Nominal
3	age - student's age (numeric: from 15 to 22)	Quantitativa Contínua
4	address - student's home address type (binary: "U" - urban or "R" - rural)	Qualitativa Nominal
5	famsize - family size (binary: "LE3" - less or equal to 3 or "GT3" - greater than 3)	Qualitativa Nominal
6	Pstatus - parent's cohabitation status (binary: "T" - living together or "A" - apart)	Qualitativa Nominal
7	Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)	Qualitativa Ordinal
8	Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)	Qualitativa Ordinal
9	Mjob - mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")	Qualitativa Nominal
10	Fjob - father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")	Qualitativa Nominal
11	reason - reason to choose this school (nominal: close to "home", school "reputation", "course" preference or "other")	Qualitativa Nominal
12	guardian - student's guardian (nominal: "mother", "father" or "other")	Qualitativa Nominal
13	travelttime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)	Qualitativa Ordinal
14	studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)	Qualitativa Ordinal
15	failures - number of past class failures (numeric: n if 1<=n<3, else 4)	Qualitativa Ordinal
16	schoolsup - extra educational support (binary: yes or no)	Qualitativa Nominal
17	famsup - family educational support (binary: yes or no)	Qualitativa Nominal

18	paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)	Qualitativa Nominal
19	activities - extra-curricular activities (binary: yes or no)	Qualitativa Nominal
20	nursery - attended nursery school (binary: yes or no)	Qualitativa Nominal
21	higher - wants to take higher education (binary: yes or no)	Qualitativa Nominal
22	internet - Internet access at home (binary: yes or no)	Qualitativa Nominal
23	romantic - with a romantic relationship (binary: yes or no)	Qualitativa Nominal
24	famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)	Qualitativa Nominal
25	freetime - free time after school (numeric: from 1 - very low to 5 - very high)	Qualitativa Nominal
26	goout - going out with friends (numeric: from 1 - very low to 5 - very high)	Qualitativa Nominal
27	Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)	Qualitativa Nominal
28	Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)	Qualitativa Nominal
29	health - current health status (numeric: from 1 - very bad to 5 - very good)	Qualitativa Nominal
30	absences - number of school absences (numeric: from 0 to 93)	Quantitativa contínua
31	G1 - first period grade (numeric: from 0 to 20)	Quantitativa Contínua
32	G2 - second period grade (numeric: from 0 to 20)	Quantitativa Contínua
33	G3 - final grade (numeric: from 0 to 20, output target)	Quantitativa Contínua

Tabela 1: Análise quantitativa e qualitativa das variáveis

#### 4. Rendimento geral dos alunos

Pelas informações da base de dados, a média de pontuação das notas verificada foi de 10,42, conforme apresentado nos gráficos 2 e 3. A nota mínima para aprovação, na escala adotada em Portugal, são 10 pontos, enquanto a nota máxima é 20 pontos, por período avaliado. Percebe-se que 32,91% dos alunos obtiveram nota suficiente dentro da escala apresentada nos padrões de Portugal.

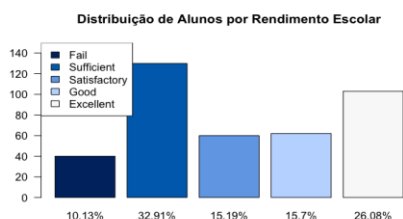


Gráfico 1: Aluno x Rendimento Escolar

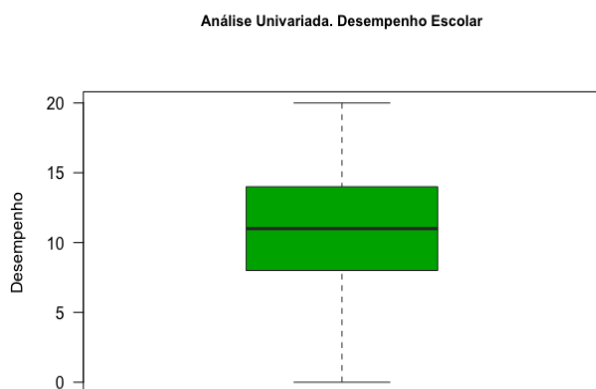


Gráfico 2: Aluno x Desempenho Escolar

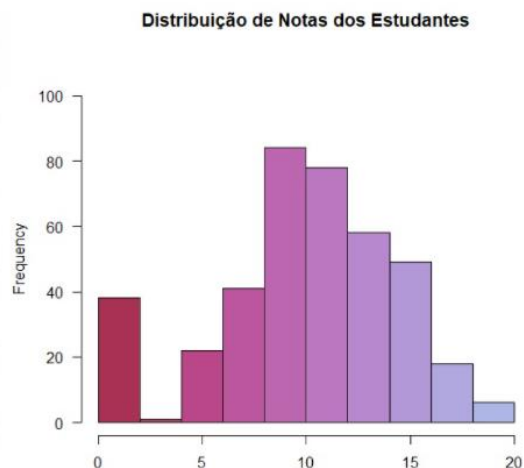


Gráfico 3: Distribuição de notas dos Estudantes

## 5. Análise exploratória univariada dos dados: panorama geral dos alunos

A Partir das análises realizadas, foi possível traçar um panorama geral dos alunos das escolas Gabriel Pereira e Mousinho da Silveira.

### 5.1 Distribuição por escola e gênero

O maior número de alunos estuda na escola Gabriel Pereira, que possui uma oferta maior de cursos. No quesito gênero, há um equilíbrio no percentual entre homens e mulheres.

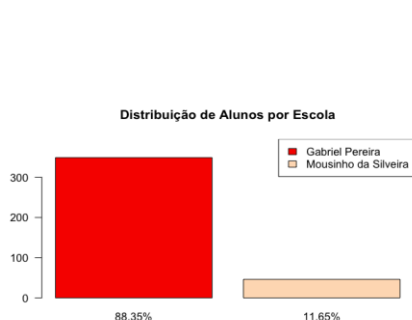


Gráfico 4: Análise univariada de Alunos x Escola

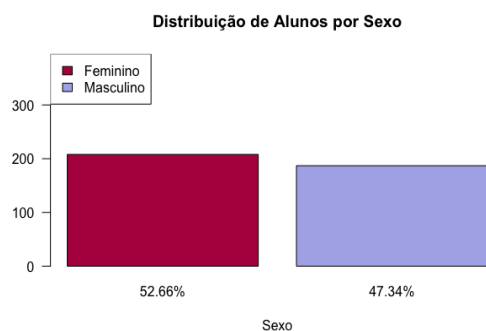


Gráfico 5: Análise univariada de Alunos x Sexo

### 5.2 Distribuição por localização

Aproximadamente 78% dos alunos residem na área urbana, sendo que a grande maioria reside próximo a escola, com tempo de deslocamento de apenas 15 minutos. Por viverem em área urbana, os alunos devem contar com acesso à internet e outras facilidades para adquirir



conhecimento e informação. A distância, aliás, foi a segunda razão mais citada para a escolha da matrícula em uma das escolas.

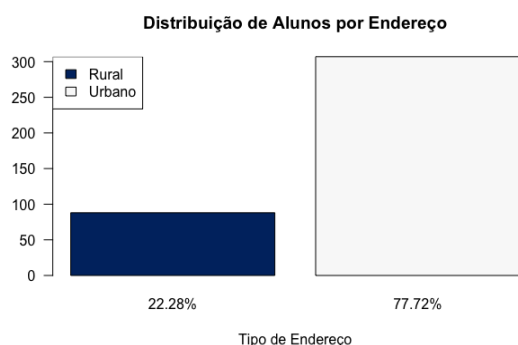


Gráfico 6: Análise univariada de Alunos x Endereço

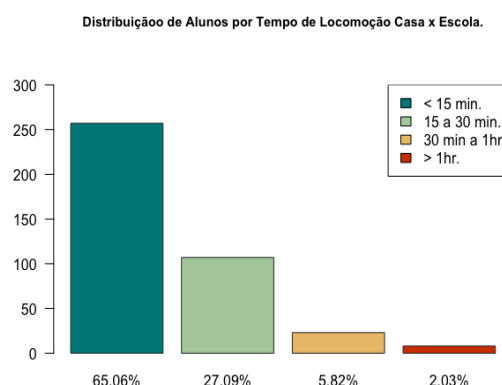


Gráfico 7: Análise univariada Locomoção Casa x Escola

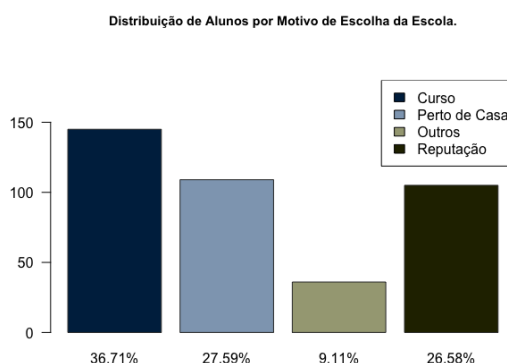


Gráfico 8: Análise univariada de Alunos por motivo de escolha da escola

### 5.3 Distribuição por situação parental e família

No cenário familiar, metade dos alunos apresentaram um bom relacionamento familiar. Em 89% dos casos, os alunos vivem em lares com pais casados. No restante, a maior parte dos filhos vive com as mães (69% dos casos). Majoritariamente, os estudantes vivem em famílias com até três membros (71%).

Avaliando o nível de escolaridade dos pais, cerca de 33% das mães possuem curso superior, contra os 24% dos pais. Tanto pais quanto mães trabalham majoritariamente em áreas não identificadas, fora dos segmentos de educação, saúde, serviços e do lar.

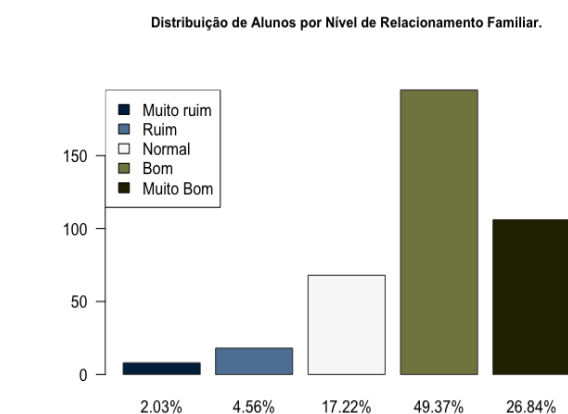


Gráfico 9: Análise univariada - Relacionamento Familiar

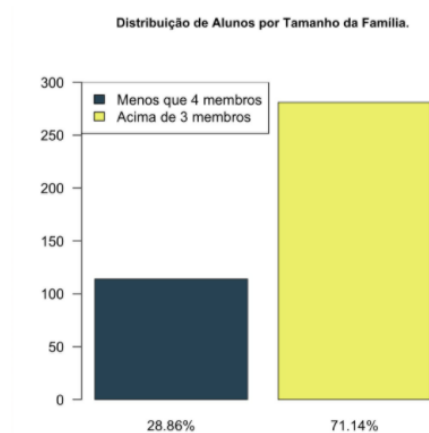


Gráfico 10: Análise univariada Tamanho da Família

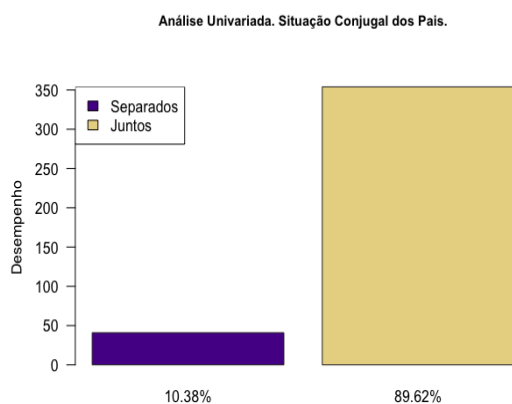


Gráfico 11: Análise univariada Situação Conjugal

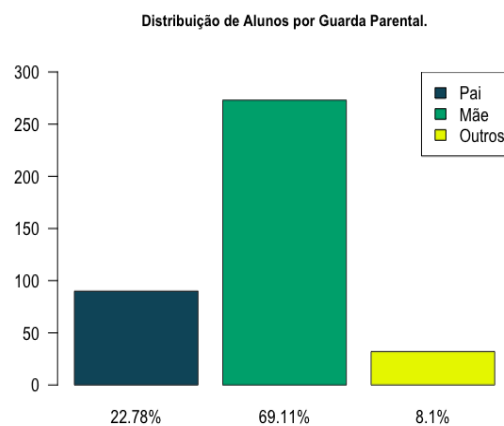


Gráfico 12: Análise univariada Guarda Parental

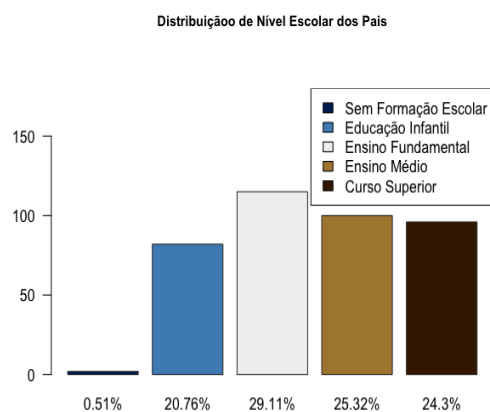


Gráfico 13: Análise univariada Nível Escolar dos Pais

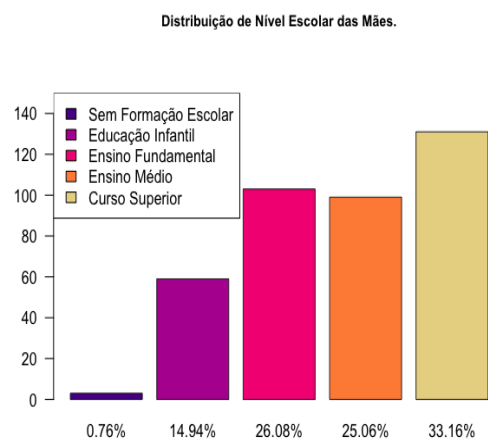


Gráfico 14: Análise univariada Nível Escolar das Mães

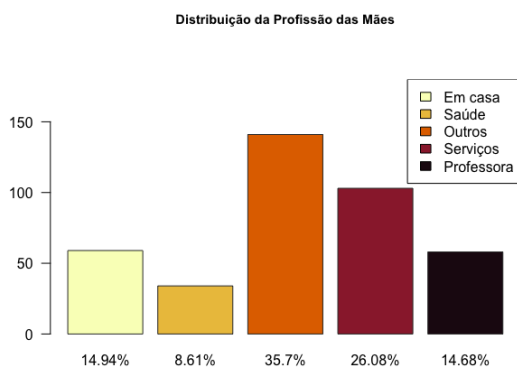


Gráfico 15: Análise univariada Profissão das Mães

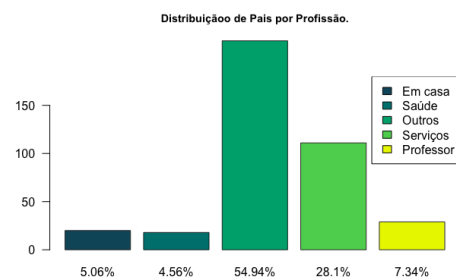


Gráfico 16: Análise univariada Profissão dos Pais

## 5.4 Distribuição por desempenho, estrutura e apoio

Os alunos matriculados nas duas escolas secundárias apresentaram um baixo nível de reprovações, com 79% deles conseguindo índices ao menos satisfatórios de desempenho. Entre as variáveis que ajudam a explicar esse resultado estão o tempo de estudo semanal, o suporte familiar, as atividades extracurriculares e as pretensões de conseguir uma vaga no ensino superior.

O apoio familiar na educação desde a primeira infância e estrutura para os estudos também são fatores que ajudam nos índices de desempenho positivo.

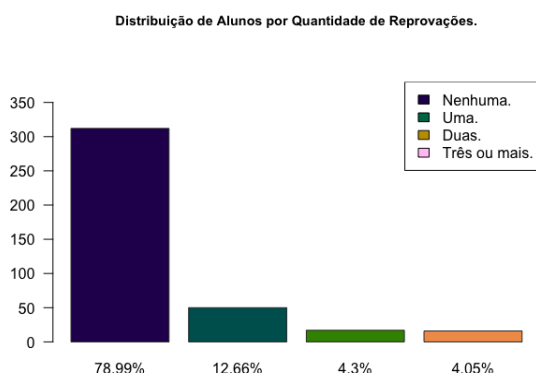


Gráfico 17: Análise univariada Reprovações

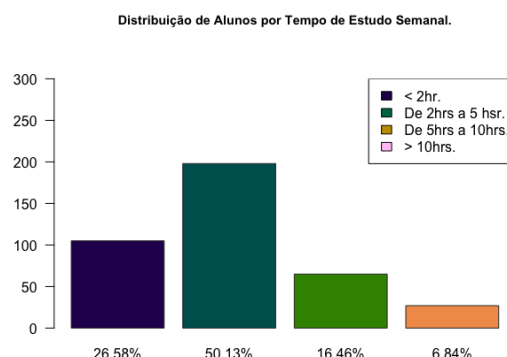


Gráfico 18: Análise univariada Tempo de Estudo Semanal

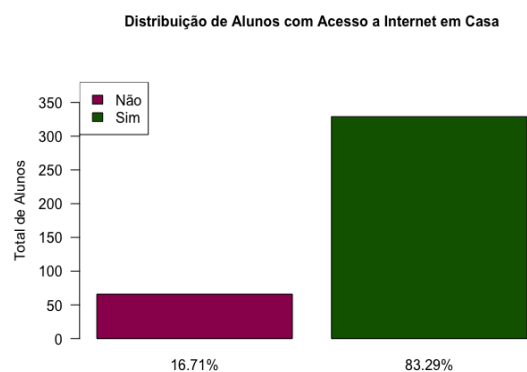


Gráfico 19: Análise univariada - Acesso à internet em

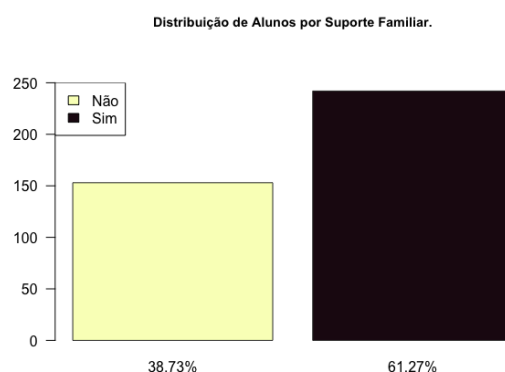


Gráfico 20: Análise univariada - Suporte Familiar

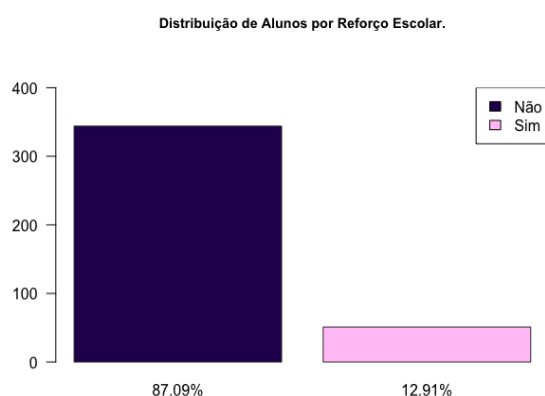


Gráfico 21: Análise univariada - Reforço Escolar

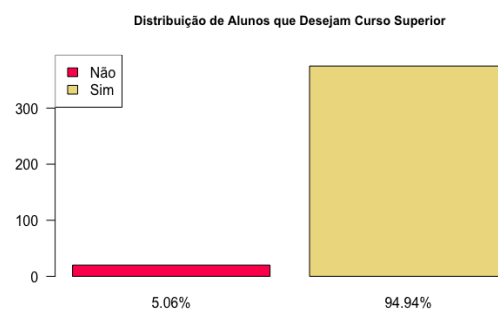


Gráfico 22: Análise univariada - Desejam o Curso Superior

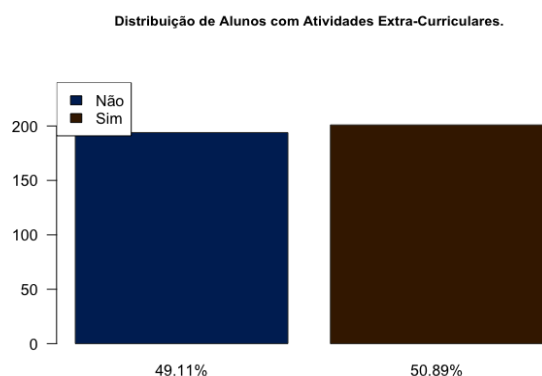


Gráfico 23: Análise univariada - Ativ. Extracurriculares

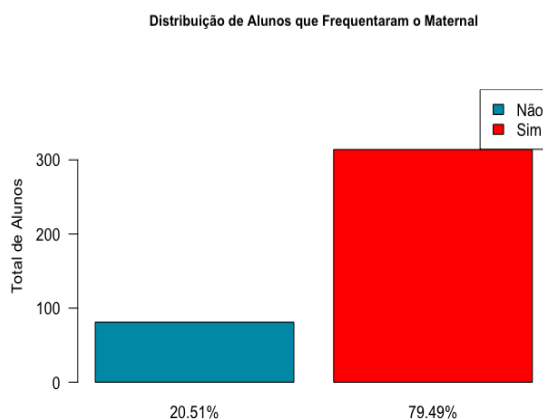


Gráfico 24: Análise univariada - Frequentaram o Maternal

## 5.5 Distribuição por aspectos pessoais

Os estudantes declararam majoritariamente não estarem em um relacionamento. Em seu tempo livre, os alunos gostam de sair, porém não foram identificadas altas taxas de consumo de álcool tanto durante a semana quanto no final de semana. A saúde geral mostra-se em bons níveis.

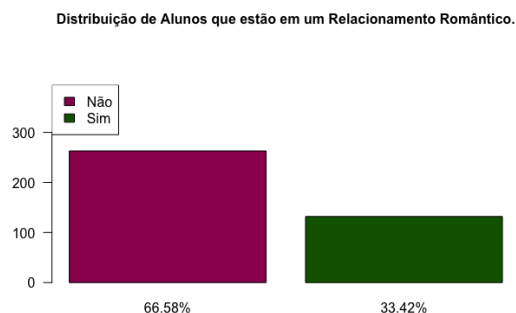


Gráfico 25: Análise univariada - Alunos que estão em um relacionamento

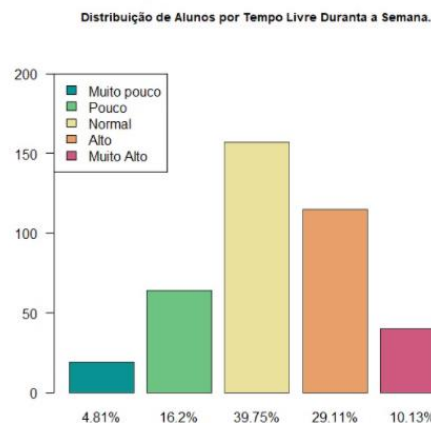


Gráfico 26: Análise univariada - Tempo Livre na Semana

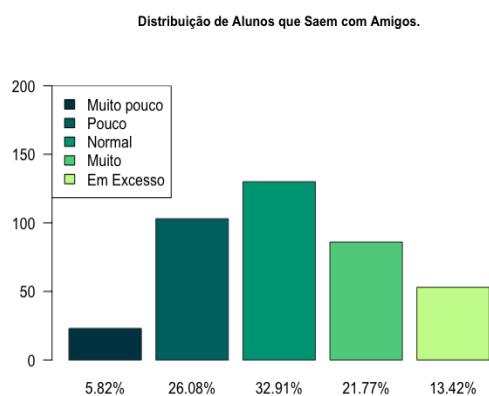


Gráfico 27: Análise univariada – Saídas com amigos

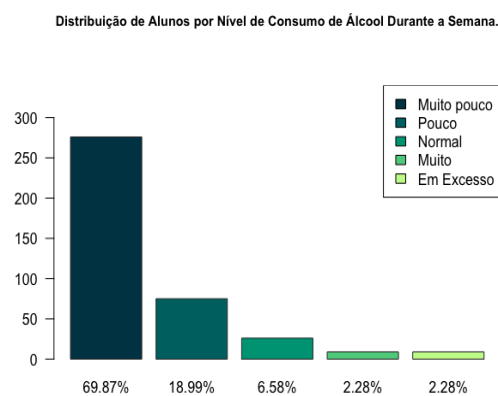


Gráfico 28: Análise univariada - Consumo Álcool Semana

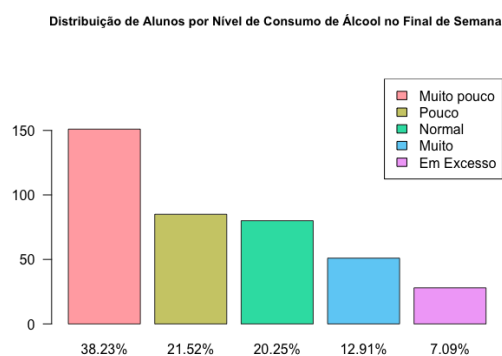


Gráfico 29: Análise univariada - Consumo Álcool Final de Semana

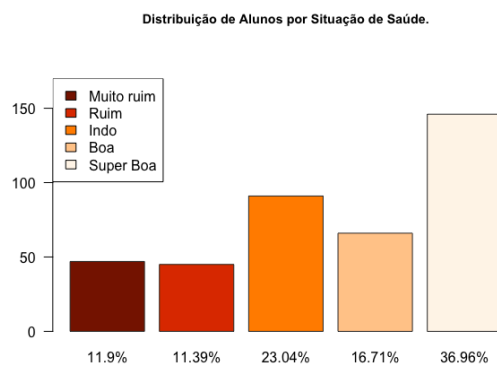


Gráfico 30: Análise univariada - Situação de Saúde

## 6. Testes de hipótese e análise bivariada em relação ao rendimento escolar

Considerando os dados levantados, temos informações que podem ser cruzadas para verificar quais elementos podem interferir, direta ou indiretamente, no desempenho estudantil dos alunos. Por isso, foram executados todos os testes de hipóteses em todas as variáveis. Os resultados estão seguir:

Variáveis			Teste de Variância		Teste T		ANOVA		Regressão Linear		Hipóteses dos testes		
	Descrição	Tipo	Não é igual	Driver	Alternativa	Driver	ANOVA	Nelch ANOV	Driver	Indicador	Hipótese Alternativa - Ha	Hipótes Nula - H0	Conclusão
Variáveis	Descrição	Type	var.equal	P-valor	Alternative	P-Valor	P-valor	P-valor	P-valor	R2	Ha	H0	Conclusão
sex	sex - student's sex (binary: "F" - female or "M" - male)	fct	VERDADEIRO	0,699	less	0,01993	---	---	---	---	Ha: O gênero não interfere no desempenho do aluno	H0: O gênero interfere no desempenho do aluno	P-valor menor que 5% rejeita a hipótese nula. O gênero não interfere no desempenho
age	age - student's age (numeric: from 15 to 22)	int	---	---	---	---	---	---	0,001	0,026	Ha: Existe relação estatisticamente significante entre a idade dos alunos e as médias escolares.	H0: Não há relação estatisticamente significante entre a idade dos alunos e as médias escolares.	P- Valor < que 5% . Identificamos uma relação linear inversa entre as variáveis de modo que a cada ano de idade, há uma redução de 0,5 decimos da nota final
address	address - student's home address type (binary: "U" - urban or "R" - rural)	fct	VERDADEIRO	0,988	less	0,0178	---	---	---	---	Ha: Alunos que moram na zona Rural possuem média escolar menor do que a média escolar de alunos que moram na zona Urbana.	H0: Alunos que moram na zona Rural possuem média escolar igual ou superior a média escolar de alunos que moram na zona Urbana.	P-valor < 5% temos evidencia para rejeitar H0 em favor de Ha. Temos probabilidade estatística de que alunos que moram em Zona Rural tem médias menores dos que moram em Zona Urbana.
famsize	famsize - family size (binary: "LE3" - less or equal to 3 or "GT3" - greater than 3)	dbl	VERDADEIRO	0,188	tow.sided	0,1062	---	---	---	---	Ha: A média escolar de alunos com família de até 3 membros difere da média escolar de alunos com família acima de 3 membros.	H0: a média escolar de alunos com família de até 3 membros não difere da média escolar de alunos com família cima de 3 membros.	P-valor > 5% falhamos em rejeitar H0. O tamanho da família não interfere no desempenho escolar.
Pstatus	Pstatus - parent's cohabitation status (binary: "T" - living together or "A" - apart)	fct	VERDADEIRO	0,593	tow.sided	0,2501	---	---	---	---	Ha: A situação conjugal dos pais interfere no desempenho escolar dos alunos.	H0: A situação conjugal dos pais não interfere no desempenho escolar dos alunos.	P-valor > que 5%, falhamos em rejeitar H0. A situação conjugal dos pais não interfere no desempenho dos alunos.
Medu	Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)	fct	---	---	---	---	9,24E-05	0,005	---	---	Ha: Ao menos uma das médias escolares dos alunos diferem com relação ao grau de escolaridade das mães.	H0: As médias escolares dos alunos não diferem com relação ao grau de escolaridade das mães.	Temos evidencia para rejeitar H0 em favor de Ha. Os resultados mostram que ao menos uma das médias se difere com relação ao grau de escolaridade das mães. Desta forma concluímos que a o grau de escolaridade das mães influencia no desempenho escolar dos alunos.
Fedu	Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)	fct	---	---	---	---	0,022	0,51	---	---	Ha: Ao menos uma das médias escolares dos alunos diferem com relação ao grau de escolaridade dos pais.	H0: As médias escolares dos alunos não diferem com relação ao grau de escolaridade dos pais	Temos evidencia para rejeitar H0 em favor de Ha. Considerando variâncias iguais, os resultados mostram que ao menos uma das médias se difere com relação ao grau de escolaridade dos pais. Desta forma concluímos que o grau de escolaridade dos pais influencia no desempenho escolar dos alunos.
Mjob	Mjob - mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")	fct	---	---	---	---	0,0051	0,006	---	---	Ha: Ao menos uma das médias escolares se diferem conforme a categoria da profissão das mães.	H0: As médias escolares não se diferem conforme a categoria da profissão das mães.	Temos evidencia para rejeitar H0 em favor de Ha. Os resultados mostram que ao menos uma das médias se difere com relação a categoria da profissão das mães. Desta forma concluímos que a categoria de profissão das mães influencia no desempenho escolar dos alunos.

<b>Fjob</b>	Fjob - father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")	fct	---	---	---	---	0,268	0,277	---	---	Ha: Ao menos uma das médias escolares se diferem conforme a categoria da profissão dos pais.	H0: As médias escolares não se diferem conforme a categoria da profissão dos pais.	Falhamos em rejeitar H0. Os resultados mostram que as médias escolares não diferem conforme a categoria da profissão dos pais.
<b>reason</b>	reason - reason to choose this school (nominal: close to "home", school "reputation", "course" preference or "other")	fct	---	---	---	---	0,102	0,092	---	---	Ha: As médias não se diferem conforme o motivo de escolha do colégio.	H0: Ao menos uma das médias se diferem conforme o motivo de escolha do colégio.	Falhamos em rejeitar H0. Os resultados mostram que as médias escolares não diferem com relação ao motivo da escolha da escola.
<b>guardian</b>	guardian - student's guardian (nominal: "mother", "father" or "other")	fct	---	---	---	---	0,205	0,226	---	---	Ha: Ao menos uma das médias escolares se difere em função do tipo de guarda parental.	H0: As médias escolares não se diferem em função do tipo de guarda parental.	Falhamos em rejeitar H0. Os resultados mostram que as médias escolares não diferem com relação ao tipo de guarda parental dos alunos. Desta forma concluímos que o tipo de guarda parental não interfere no desempenho escolar dos alunos.
<b>traveltime</b>	traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)	int	---	---	---	---	0,019	0,172	---	---	Ha: Ao menos uma das médias escolares se difere em função do tempo de locomoção até a escola.	H0: As médias escolares não se diferem em função do tempo de locomoção até a escola.	Falhamos em rejeitar H0. Considerando variâncias diferentes, os testes mostram que todas as médias se diferem em função do tempo de transporte de casa até o colégio. Desta forma concluímos que o tempo de casa até o colégio não interfere no desempenho dos alunos.
<b>studytime</b>	studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)	int	---	---	---	---	0,052	0,196	---	---	Ha: Ao menos uma das médias escolares se difere em função do tempo de estudo do aluno durante a semana	H0: As médias escolares se diferem em função do tempo de estudo do aluno durante a semana	Falhamos em rejeitar H0. Os estudos mostram que o tempo de estudo do aluno durante a semana não interfere no seu desempenho escolar.
<b>failures</b>	failures - number of past class failures (numeric: n if 1<=n<3, else 4)	int	---	---	---	---	<<<<<< 5%	<<<<<< 5%	---	---	Ha: Ao menos uma das médias se diferem em função de reprovações anteriores na mesma disciplina.	H0: As médias não se diferem em função de reprovações anteriores na mesma disciplina.	Temos evidencia para rejeitar H0 em favor de Ha. Os testes comprovam que o desempenho dos alunos se difere em função de reprovações anteriores na mesma disciplina. Desta forma podemos concluir que reprovações na disciplina tem relação com o desempenho do aluno.
<b>schoolsup</b>	schoolsup - extra educational support (binary: yes or no)	fct	FALSO	3,51E-05	greater	0,009871	---	---	---	---	Ha: A média escolar de alunos que não possuem reforço escolar é maior do que a média escolar de alunos que possuem.	H0: A média escolar de alunos que não possuem reforço escolar é menor ou igual a média escolar de alunos que possuem.	P-valor < 5% temos evidencia para rejeitar H0 em favor de Ha. A média escolar de alunos que não possuem reforço escolar é superior a dos alunos que possuem.
<b>famsup</b>	famsup - family educational support (binary: yes or no)	fct	VERDADEIRO	0,798	tow.side	0,437	---	---	---	---	Ha: A média de alunos que possuem suporte educacional familiar é diferente da média de alunos que não possuem.	H0: A média de alunos que possuem suporte educacional familiar é igual a média de alunos que não possuem.	P-valor > que 5%, falhamos em rejeitar H0. O suporte educacional familiar não interfere no desempenho escolar.

[illegible]



goout	goout - going out with friends (numeric: from 1 - very low to 5 - very high)	int	---	---	---	---	---	---	---	---	---	---	---
Dalc	Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)	fct	FALSO	0,001362	two.side	0,2195	---	---	---	---	Ha: A média escolar de alunos que consomem álcool acima da média é diferente a média escolar de alunos que não o fazem.	H0: A média escolar de alunos que consomem álcool acima da média é igual a média escolar de alunos que não o fazem.	P-valor > 5% falhamos em rejeitar H0. O consumo semanal de álcool (Dalc + Walc), acima da média estabelecida [média de (Dalc + Walc)/2]] não tem relação no desempenho escolar.
Walc	Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)	fct											
health	health - current health status (numeric: from 1 - very bad to 5 - very good)	int	---	---	---	---	---	---	---	---	---	---	---
absences	absences - number of school absences (numeric: from 0 to 93)	int	---	---	---	---	---	---	0,497	0,0011	Ha: Existe relação estatisticamente significante entre as faltas dos alunos e as médias escolares.	H0: Não há relação estatisticamente significante entre as faltas dos alunos e as médias escolares.	P-valor > que 5%, Falhamos em rejeitar H0. Não apresenta um padrão de lineariedade entre as faltas e o desempenho escolar

Foi verificado que as variáveis idade, área de residência, grau de escolaridade de mãe e pai, profissão da mãe, desempenho escolar e reforço escolar refletiram no desempenho escolar durante o ano letivo.

A seguir, serão apresentadas análises bivariadas detalhadas dessas variáveis, com a apresentação das hipóteses nula ( $H_0$ ) e alternativa ( $H_a$ ) e os respectivos testes de hipótese.

## 6.1 Regressão linear - desempenho de alunos por idade

No teste de regressão linear simples por idade, temos o seguinte cenário:

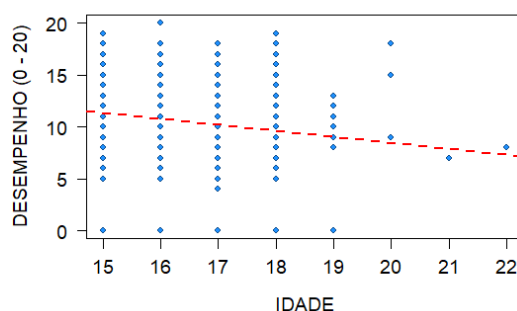


Gráfico 32: Regressão linear simples de idade e desempenho

P-valor: 0.001

**CONCLUSÃO:** Com o P-value é menor que 5%, identificamos uma relação linear inversa entre o desempenho dos alunos e sua idade, de modo que, a cada ano de idade dos alunos, há uma redução de 0.5 décimos da nota final.

## 6.2 Diagrama sobre o desempenho de alunos por localização (urbano ou rural)

$H_0$ : Alunos que moram na zona Rural possuem média escolar igual ou superior a média escolar de alunos que moram na zona Urbana.

$H_a$ : Alunos que moram na zona Rural possuem média escolar menor do que a média escolar de alunos que moram na zona Urbana.

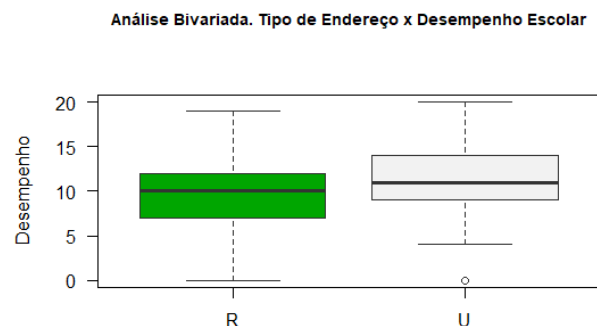


Gráfico 33: Análise bivariada de tipo de endereço e desempenho

Teste de Variância: 0,988. As variâncias não são diferentes  
T.test: 0,0178

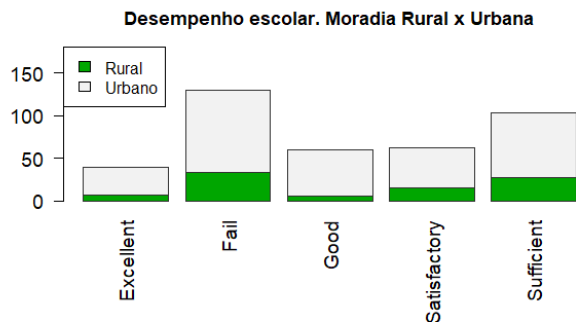


Gráfico 34: Análise bivariada por moradia rural e urbana versus desempenho

**CONCLUSÃO:** P-valor < 5% temos evidência para rejeitar H0 em favor de Ha. Temos probabilidade estatística de que alunos que moram em Zona Rural tem médias menores dos que moram em Zona Urbana.

### 6.3 Diagrama sobre o desempenho por grau de escolaridade da mãe

H0: As médias escolares dos alunos não diferem com relação ao grau de escolaridade das mães.

Ha: Ao menos uma das médias escolares dos alunos diferem com relação ao grau de escolaridade das mães.

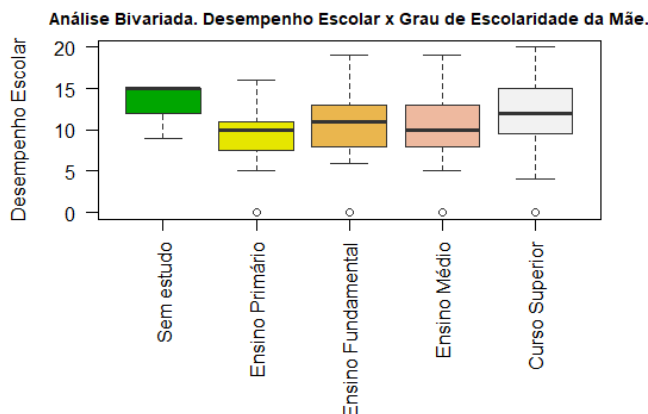


Gráfico 35: Análise bivariada desempenho escolar e escolaridade da mãe

Teste ANOVA: Variâncias iguais

P-valor: menor que 5%

Teste WELCH ANOVA: Variâncias diferentes

P-valor: menor que 5%

**CONCLUSÃO:** Temos evidência para rejeitar H0 em favor de Ha. Os resultados mostram que ao menos uma das médias se difere com relação ao grau de escolaridade das mães. Desta forma

concluimos que o grau de escolaridade das mães influencia no desempenho escolar dos alunos.

#### 6.4 Diagrama sobre o desempenho por grau de escolaridade do pai

H0: As médias escolares dos alunos não diferem com relação ao grau de escolaridade dos pais.

Ha: Ao menos uma das médias escolares dos alunos diferem com relação ao grau de escolaridade dos pais.

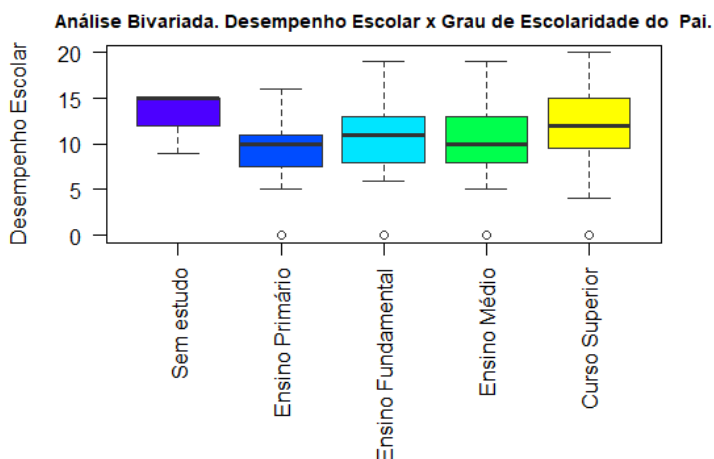


Gráfico 36: Análise bivariada desempenho escolar e escolaridade do pai

Teste ANOVA: Variâncias iguais

P-valor: menor que 5%

Teste WELCH ANOVA: Variâncias diferentes

P-valor: menor que 5%

**CONCLUSÃO:** Considerando variâncias iguais, os resultados mostram que ao menos uma das médias se difere com relação ao grau de escolaridade dos pais. Desta forma, concluimos que o grau de escolaridade dos pais influencia no desempenho escolar dos alunos.

#### 6.5 Diagrama sobre o desempenho por grau de profissão da mãe

H0: As médias escolares não se diferem conforme a categoria da profissão das mães.

Ha: Ao menos uma das médias escolares se diferem conforme a categoria da profissão das mães.

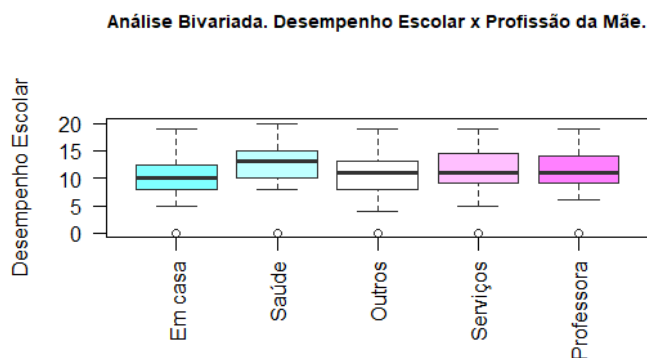


Gráfico 37: Análise bivariada desempenho escolar e profissão da mãe

Teste ANOVA: Variâncias iguais

P-valor: menor que 5%

Teste WELCH ANOVA: Variâncias diferentes

P-valor: menor que 5%

**CONCLUSÃO:** Considerando variâncias iguais, os resultados mostram que ao menos uma das médias se difere com relação ao grau de escolaridade dos pais. Desta forma, concluímos que o grau de escolaridade dos pais influencia no desempenho escolar dos alunos.

## 6.6 Diagrama sobre o desempenho por reprovações

H0: As médias não se diferem em função de reprovações anteriores na mesma disciplina.

Ha: Ao menos uma das médias se diferem em função de reprovações anteriores na mesma disciplina.

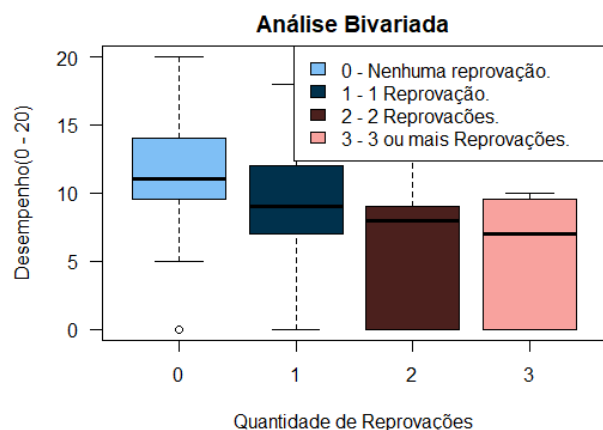


Gráfico 38: Análise bivariada desempenho escolar e repetências

Teste ANOVA: Variâncias iguais

P-valor: menor que 5%

Teste WELCH ANOVA: Variâncias diferentes

P-valor: menor que 5%

**CONCLUSÃO:** Os testes comprovam que o desempenho dos alunos se difere em função de reprovações anteriores na mesma disciplina. Desta forma, podemos concluir que reprovações na disciplina tem relação com o desempenho do aluno.

## 6.7 Diagrama sobre o desempenho por reforço escolar

H0: A média escolar de alunos que não possuem reforço escolar é menor ou igual a média escolar de alunos que possuem.

Ha: A média escolar de alunos que não possuem reforço escolar é maior do que a média escolar de alunos que possuem.

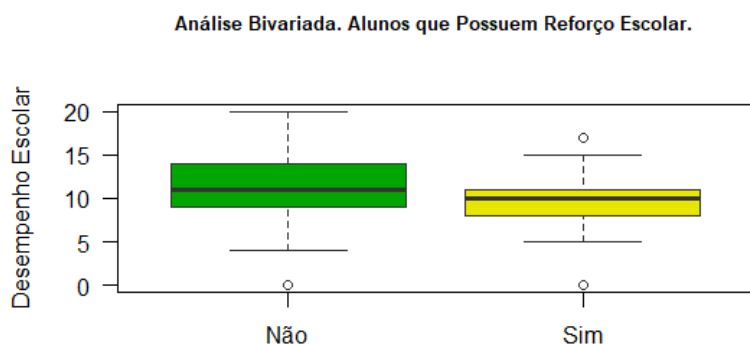


Gráfico 39: Análise bivariada desempenho escolar e reforço escolar

Teste de Variância: 0,00003512. As variâncias são diferentes.

T.test: 0,009

**CONCLUSÃO:** Como o P-valor é menor que 5%, rejeitamos a hipótese nula em favor de Ha. Portanto, concluímos que alunos que não possuem reforço escolar possuem média escolar superior a alunos que fazem o reforço.

## 7. Diferenças de rendimento entre as duas escolas (disciplina matemática)

As hipóteses colocadas para a comparação de desempenho entre as escolas foram as seguintes:

Ho: Não existe diferença de desempenho entre as escolas Gabriel Pereira e a Mousinho Silveira

Ha: Existe diferença de desempenho a escolas Gabriel Pereira e a Mousinho Silveira

A partir de uma análise bivariada, observa-se que a linha das médias está muito próxima, não havendo diferenças de desempenho entre as escolas. Entretanto, faz-se necessário a validação desta informação a partir do teste de comparação de médias. O teste executado foi o teste.t.

Como o p-valor = 0,372 é maior que 5% rejeitamos a hipótese nula. Portanto, concluímos que não existe diferença de desempenho entre as escolas Gabriel Pereira e a Mousinho da Silveira.

Gráfico X: Análise Bivariada Desempenho x Escolas

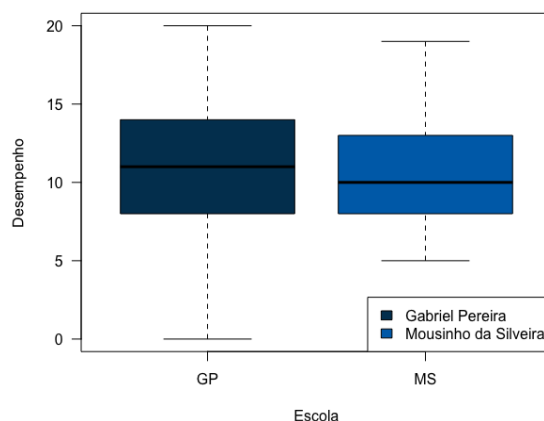


Gráfico 40: Análise bivariada desempenho entre escolas

## 8. Resultados e conclusões

Portugal possuía, à época do levantamento de desempenho de estudantes, uma população estimada de 10,5 milhões de habitantes. Portanto, a amostragem analisada neste estudo junto aos alunos das escolas Gabriel Pereira e Mousinho da Silveira é insuficiente para a tomada de decisões de políticas públicas para a melhoria do desempenho dos estudantes secundaristas portugueses, mas traz boas pistas das razões que interferiram nas notas abaixo das esperadas na comparação com as demais nações da União Europeia.

A vivência em zona rural ou urbana é um fator que tem impacto direto no desempenho. Isso deve ocorrer, entre outros fatores, pela infraestrutura precária em tais lugares. Investimentos em expansão de sinal de internet em alta velocidade e reforço na rede elétrica são dois fatores que podem ajudar a reduzir as diferenças em condições de estudo entre estudantes dessas duas áreas.

É notável que o grau de escolaridade dos pais interfere diretamente no desempenho escolar dos filhos. Pais e mães que estudam mais tempo replicam essa cultura aos filhos, criando um ciclo virtuoso de educação. Importante que autoridades continuem incentivando, abrindo sempre mais oportunidades de qualificação de ensino em instituições públicas e privadas.

Nas escolas analisadas, os estudantes que fazem reforço escolar possuem notas menores, o que requer uma investigação aprofundada sobre o assunto. É possível que determinados alunos necessitem de metodologias de ensino diferenciadas, que possam cativar e estimular a busca pelo conhecimento de maneiras menos ortodoxas que o ensino tradicional. O desempenho de repetentes tende a ser melhor, de acordo com os dados, o que é um efeito esperado.

## 9. Referências Bibliográficas

ABRÃO, Ana Carla. Como a educação influencia o desenvolvimento econômico do brasil? Todos pela educação. Disponível em: <https://todospelaeducacao.org.br/noticias/como-a-educacao-influencia-o-desenvolvimento-economico-do-brasil/> Acesso em 01/08/2021

P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUTURE BUSINESS TECHNOLOGY Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7.

PÚBLICO. Portugal registou o pior PIB per capita da zona euro em 2006. Disponível em: <https://www.publico.pt/2007/12/18/jornal/portugal-registou-o-pior-pib-per-capita-da-zona-euro-em-2006-241907> Acesso em 01/08/2021