

Does Alcohol Affects Grades?

Semar Augusto da Cunha Mello Martins

1. Introdução:

O álcool é uma das drogas recreacionais mais comumente usadas pelo mundo, isso acontece devido à sua legalidade e à facilidade de acesso a ela. Contudo, já foi descoberto que o abuso de álcool pode causar vício, o que pode causar múltiplos problemas na vida das pessoas, como falta de motivação para estudar, perda do emprego, entre outros.

Nesse trabalho vamos avaliar se o consumo de álcool tem relação com a nota dos alunos. É importante notar que boa parte dos sujeitos avaliados na base de dados possui menos de 18 anos, por causa disso, a taxa de alcoólatras no estudo será inferior à média mundial.

2. Metodologia:

3. Base de Dados:

A base de dados escolhida foi “Student Alcohol Consumption”, disponibilizada inicialmente pelo UCI - Machine Learning Repository mas atualmente só foi possível encontrá-la no Kaggle.

A base corresponde a uma pesquisa com as notas de alunos do ensino médio com notas somente das matérias matemática e português. Além das notas dos alunos, existem vários atributos relacionados à vida de cada aluno.

Features:

- Escola do aluno
- Sexo
- Idade
- Endereço
- Tamanho da família
- Se os pais são separados ou não
- Educação da mãe
- Educação do pai
- Trabalho da mãe
- Trabalho do pai
- Motivo pela escolha da escola
- Quem é o guardião do estudante
- Tempo que o aluno demora para chegar à escola
- Tempo de estudo por semana
- Número de matérias repetidas pelo aluno
- Se a escola dá suporte extra ao aluno
- Se a família suporta o estudo do aluno
- Se o aluno frequenta aulas particulares
- Atividades extra-curriculares
- Se o aluno apresenta interesse em fazer faculdade
- Possui internet?
- Tem relações românticas?

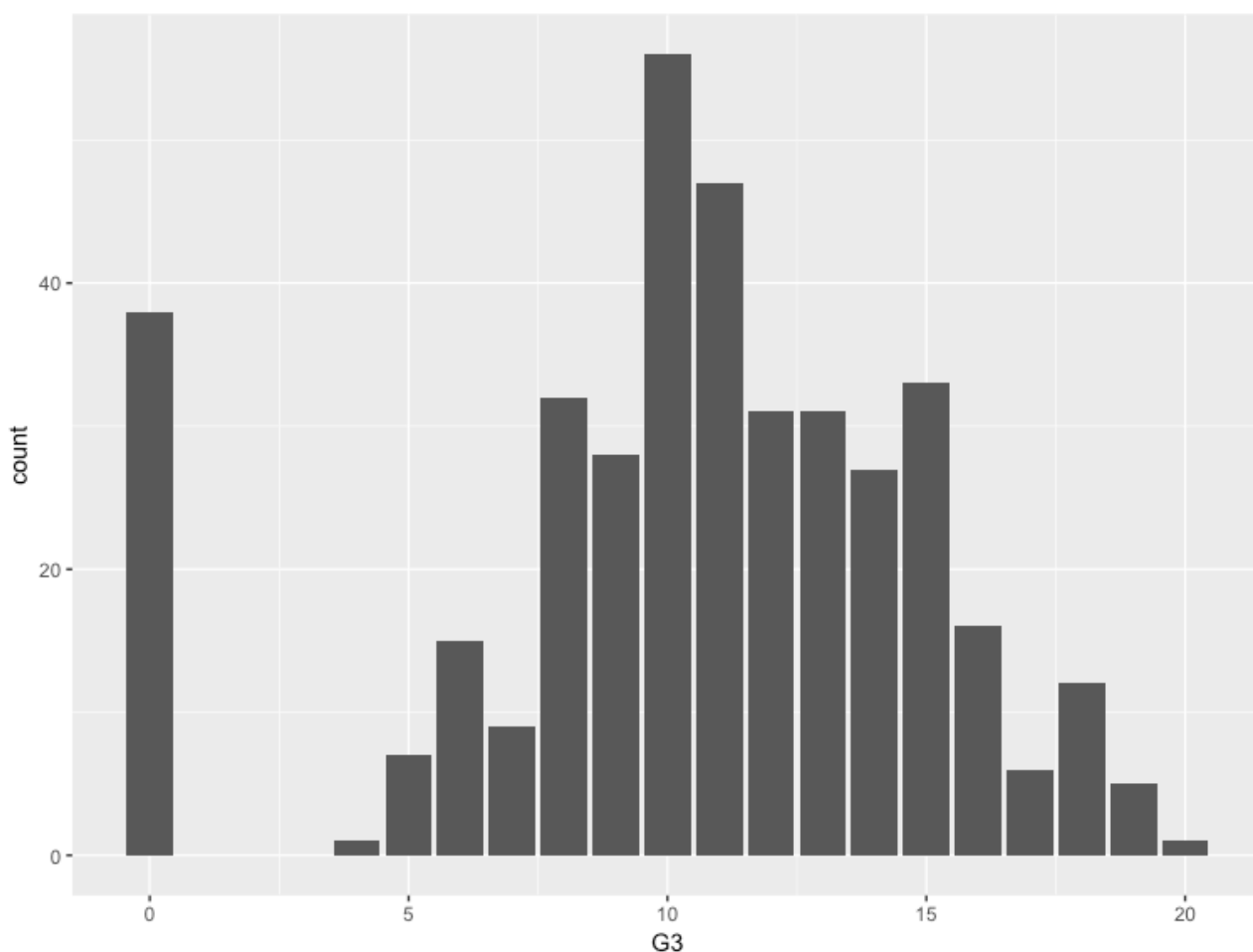
- Qualidade das relações familiares
- Quanto tempo livre possui fora da escola
- Frequência com que sai com os amigos
- Frequência com que bebe em dias de semana
- Frequências com que bebe em fins de semana
- Saúde
- Número de faltas
- Nota no primeiro período
- Nota no segundo período
- Nota final

Existem dois arquivos no .zip fornecido pelo Kaggle. O primeiro possui a nota dos alunos somente de matemática e possui 395 alunos. O segundo possui a nota somente de português, com 649 alunos. Dentre os alunos da base, 382 deles foram entrevistados para ambos.

Amostra base de dados 1:

school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltme	studytme	failures	schoolsup	famsup	paid	activities	nursery	higher	internet	romantic	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
GP	F	18	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	yes	no	no	no	yes	yes	no	no	4	3	4	1	1	3	4	0	11	11
GP	F	17	U	GT3	T	1	1	at_home	other	course	father	1	2	0	no	yes	no	no	no	yes	yes	no	5	3	3	1	1	3	2	9	11	11
GP	F	15	U	LE3	T	1	1	at_home	other	other	mother	1	2	0	yes	no	no	no	yes	yes	yes	no	4	3	2	2	3	3	6	12	13	12

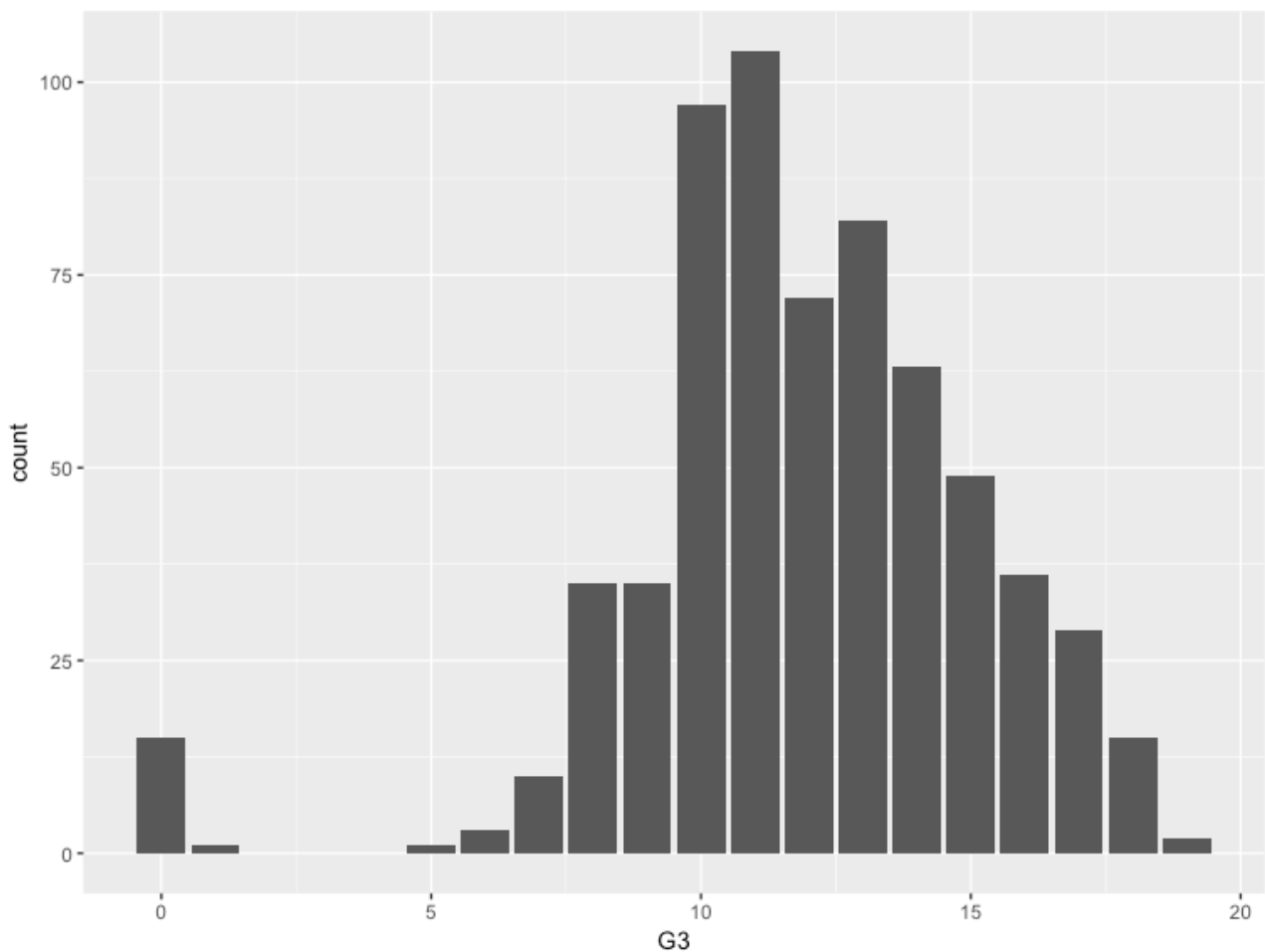
Distribuição das notas finais da base 1:



Amostra da base de dados 2:

school	sex	age	address	famsize	Pstatus	Medu	Fedu	MJob	Fjob	reason	guardian	traveltme	studytme	failures	schoolsup	famsup	paid	activities	nursery	higher	internet	romantic	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
GP	F	18	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	yes	no	no	no	yes	yes	no	no	4	3	4	1	1	3	6	5	6	6
GP	F	17	U	GT3	T	1	1	at_home	other	course	father	1	2	0	no	yes	no	no	no	yes	yes	no	5	3	3	1	1	3	4	5	5	6
GP	F	15	U	LE3	T	1	1	at_home	other	other	mother	1	2	3	yes	no	yes	no	yes	yes	yes	no	4	3	2	2	3	3	10	7	8	10

Distribuição das notas finais da base 2:



3. Planejamento:

Para o desenvolvimento do trabalho serão necessárias 3 principais etapas: agregação das bases em uma só, montagem do classificador e teste dele.

Com relação à agregação, ela será simples pois o Kaggle disponibiliza quais são as colunas que, juntas, indentificam um mesmo estudante na base 1 e na base 2.

Com relação ao classificador, será usado o algoritmo Random Forest uma vez que ele é um dos mais precisos na tarefa.

