



CONSUMO DE ÁLCOOL POR ESTUDANTES

GRUPO 4

Augusto Xavier
Gustavo Rogério
Henrique Tavares
Juan Kineipe
Wallace Ramon

Prof. Dra. Ana Amélia Benedito Silva

Quem?

- Estudantes de Português no Ensino Médio

Onde?

- Portugal

Quem fez o levantamento dos dados?

- University of California, Irvine

Quando o levantamento foi feito?

- Outubro de 2016

Quantas linhas e quantas colunas tem?

- 649 linhas e 33 colunas

Variáveis

Qualitativa Nominal	Qualitativa Ordinal	Quantitativa Discreta	Quantitativa Contínua
Tipo de trabalho da mãe	Grau de educação da mãe	Tempo de viagem (da casa até a escola)	Idade
Tipo de trabalho do pai	Grau de educação do pai	Tempo de estudo semanal	
Razão (pela qual escolheu a escola)	Relação familiar	Reprovações	
Guardião (quem possui a guarda do estudante)	Tempo livre (depois da escola)	Faltas	
Suporte extra educacional da escola	Passeio (se sai com os amigos)	G1 (nota do primeiro período)	
Suporte extra educacional da família	Consumo de álcool em dia de semana	G2 (nota do segundo período)	
Aulas extras pagas de portugueses	Consumo de álcool em fim de semana	G3 (nota do terceiro período)	
Atividades extra curriculares	Saúde	Tamanho da família	
Enfermaria (se já foi atendido)			
Superior (se deseja cursar)			
Internet (se tem acesso em casa)			
Relacionamento			
Escola			
Gênero			
Tipo de endereço			
Situação dos pais			

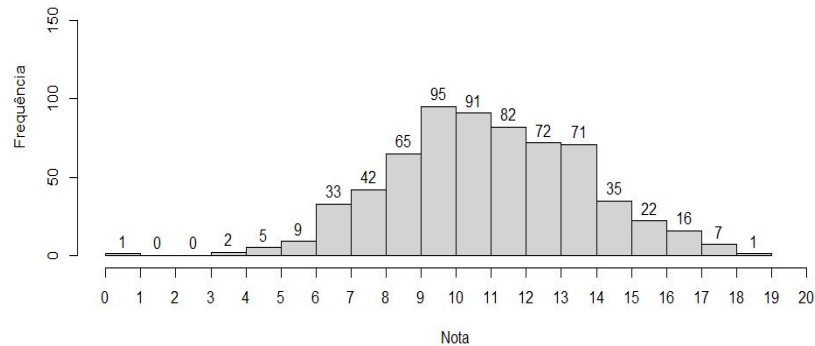
Variáveis utilizadas

- Notas do primeiro período (G1);
- Grau de educação da mãe;
- Grau de educação do pai;
- Relação familiar;
- Tempo livre;
- Passeio;
- Consumo de álcool em dia de semana;
- Consumo de álcool em fim de semana;
- Saúde;
- Tempo de viagem;
- Tempo de estudo semanal;
- Reprovações;
- Faltas;
- Idade.

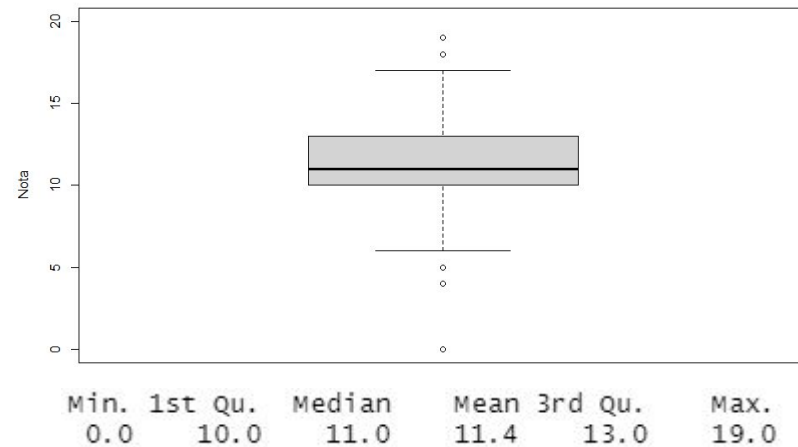
Pergunta a ser respondida

- *Dentre os alunos com notas baixas menores que 10, na primeira prova (G1), quais são os fatores relevantes que explicam tal nota?*

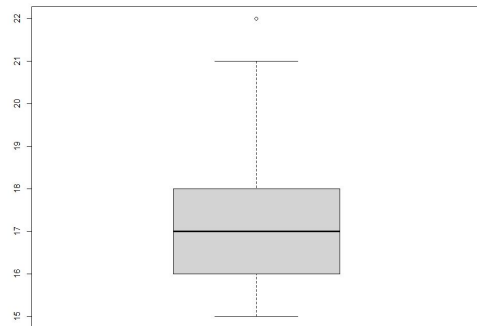
Notas G1



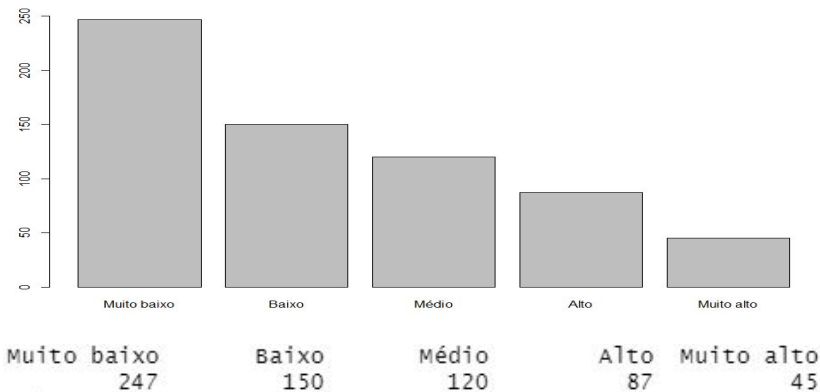
Notas G1



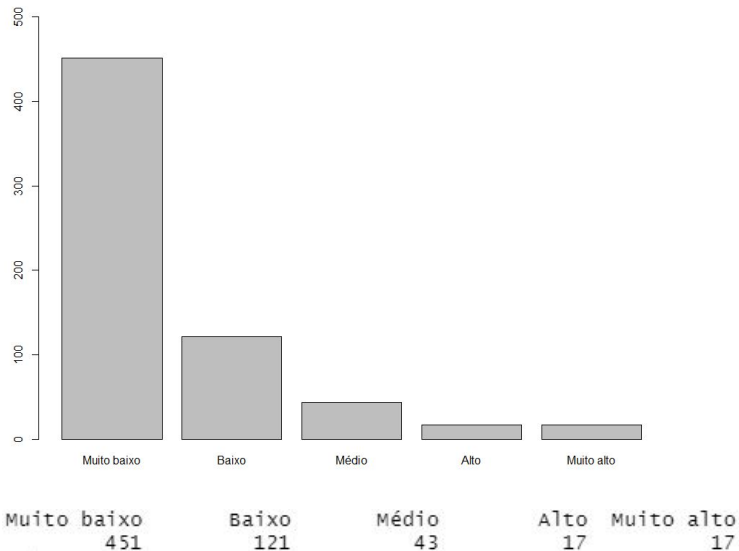
Idade



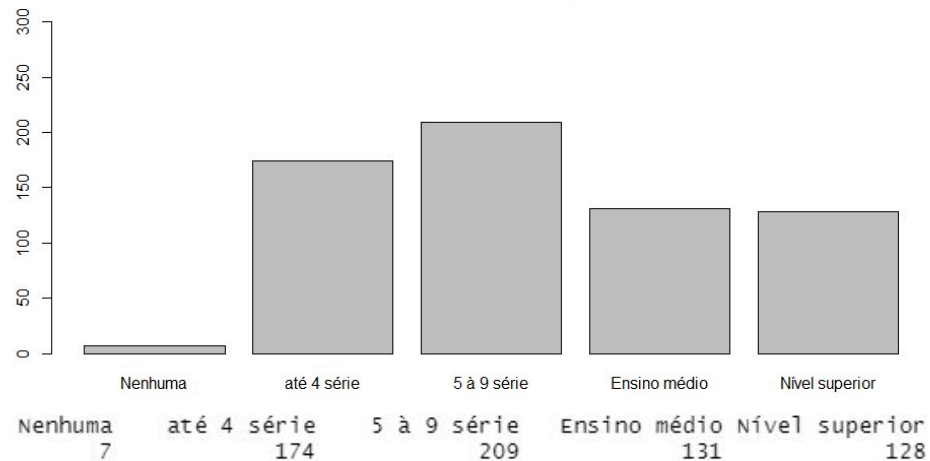
Álcool durante fim de semana



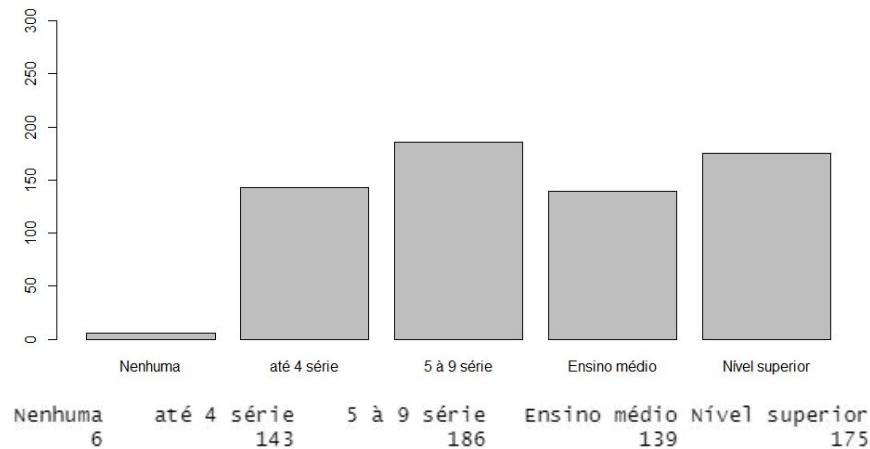
Álcool durante dias úteis

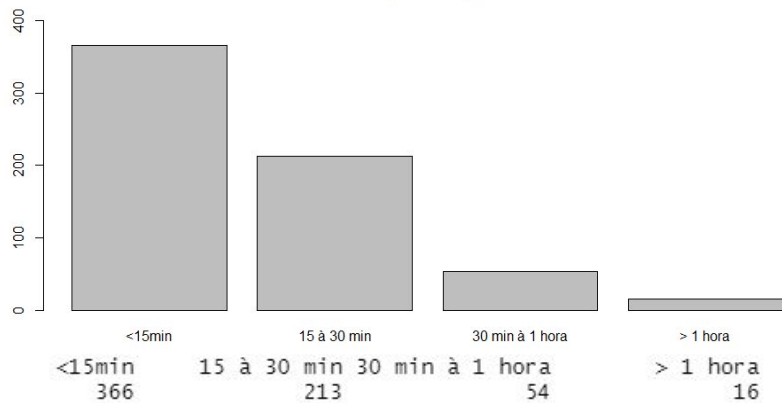
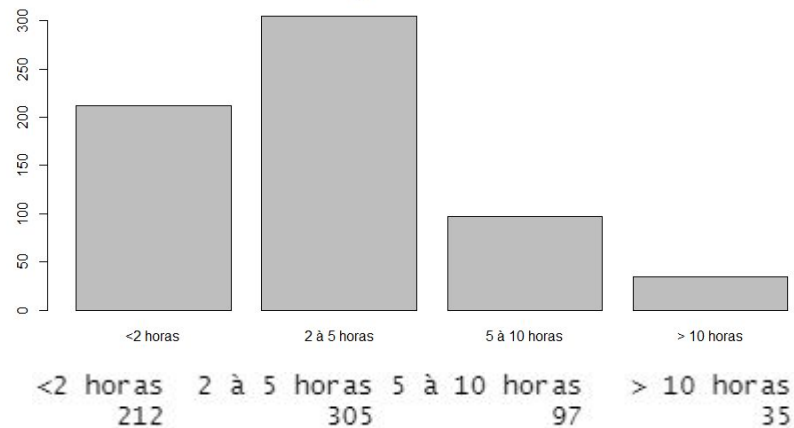
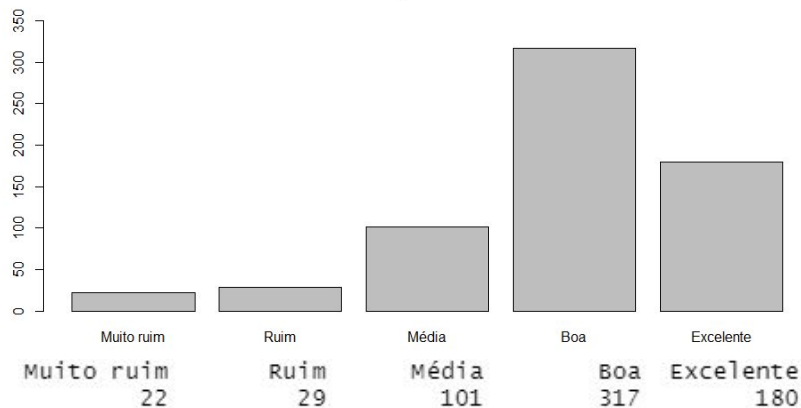
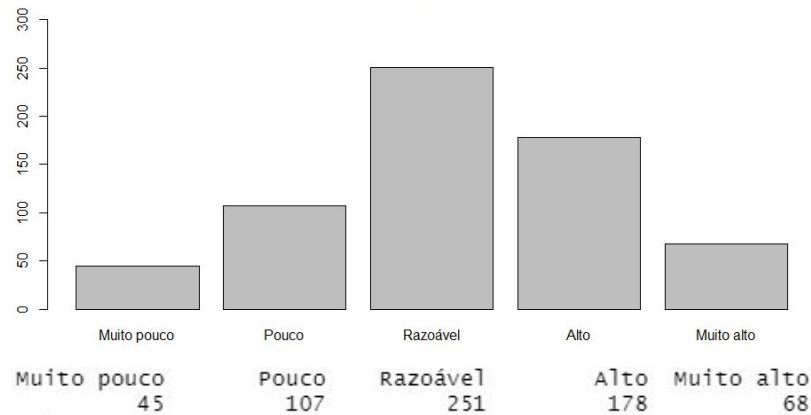


Nível escolar do pai

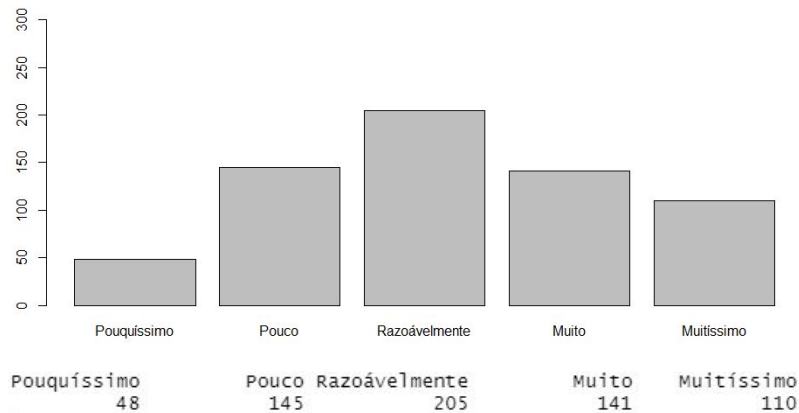


Nível escolar da mãe

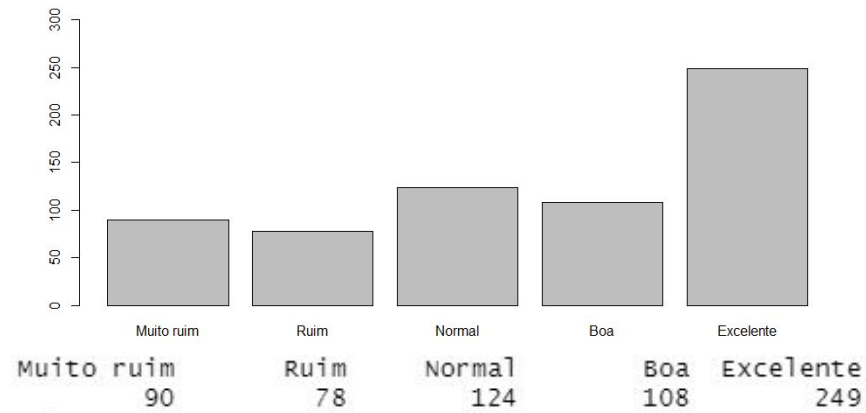


Tempo de viagem**Tempo de estudo semanal****Relação familiar****Tempo livre**

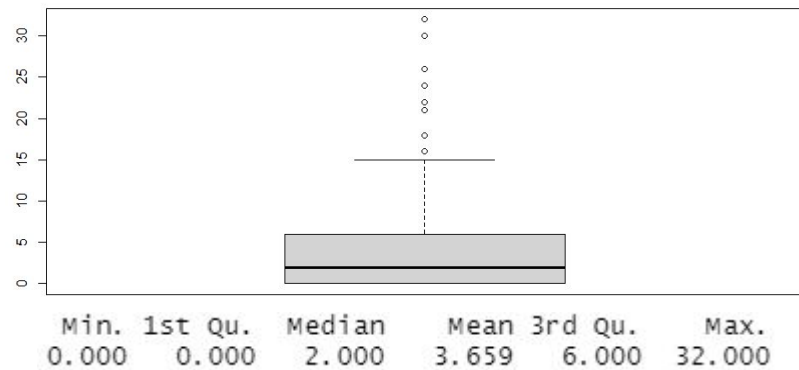
Sai com os amigos



Saúde



Faltas



Análise Fatorial

- A análise fatorial é uma técnica que é usada para reduzir um grande número de variáveis em um número menor de fatores a fim de facilitar o estudo com os dados. Sendo assim, essa técnica extrai a variância máxima comum de todas as variáveis, formando grupos com as variáveis que são fortemente correlacionadas entre si e consequentemente separando variáveis que não possuem correlação.

Análise Fatorial - Exemplo

Imagine o contexto de uma pesquisa em uma empresa sobre barreiras de compra para um cliente em potencial. A seguir, estão as possíveis barreiras à compra de um determinado produto:

- O preço é proibitivo;
- Custos gerais de implementação;
- O produto não é consistente com nossa estratégia de negócios;
- Estamos presos a um contrato com outro produto;
- Os benefícios do produto não superam o custo;
- Nosso departamento de TI não pode dar suporte ao seu produto;
- Não temos recursos técnicos suficientes;
- Seu produto não tem um recurso que exigimos.

A análise fatorial pode revelar as tendências de como essas questões se moverão juntas. A seguir estão as definições de 3 fatores englobando essas variáveis e os nomes dos fatores:

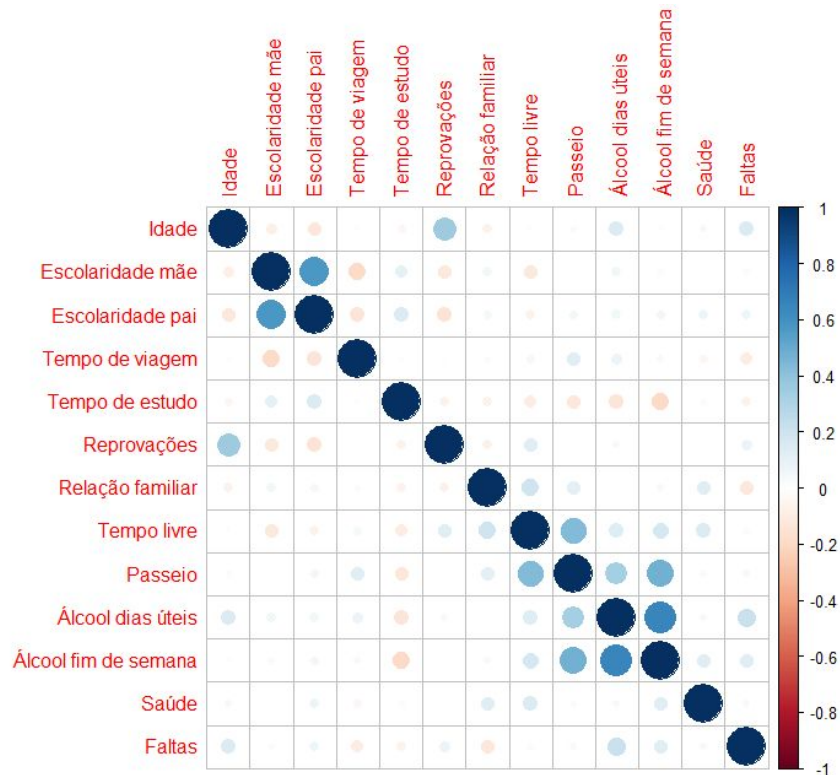
- Variáveis relacionadas ao custo - CUSTO
- Variáveis relacionadas à TI - TI
- Variáveis relacionadas aos fatores organizacionais - ORG.

Simplificar os dados usando a análise fatorial auxilia no foco e no esclarecimento dos resultados, além de reduzir o número de dimensões em que as variáveis estão agrupadas.

Análise Fatorial - Etapas

1. Matriz de correlação entre as variáveis;
2. Cálculo e interpretação do coeficiente *KMO (Kaiser-Meyer-Olkin)*;
3. Teste de Bartlett;
4. Cálculo e interpretação de *MSA (Measure of Sampling Adequacy)* para cada variável;
5. Determinação do número de fatores através do <Critério escolhido, justificando decisão>;
6. Rotação dos fatores mostrando as cargas antes e depois da rotação;
7. Interpretação dos fatores obtidos;

Análise Fatorial - Matriz de Correlação



Sobre a matriz de correlações ao lado, as correlações em azul são positivas e as vermelhas, negativas. Além disso, o tom mais forte dessas cores indicam uma proximidade maior do valor do valor 1 ou -1.

Análise Fatorial - KMO das variáveis envolvidas

Essa técnica basicamente avalia a adequacidade da análise fatorial através da seguinte fórmula matemática:

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2}$$

Abaixo, os resultado obtido com o R, em que é possível ver que, individualmente, a maioria dos valores estão no espectro 0.5 - 0.6, em que são considerados **maus**; destaque para *health* que obteve um valor menor que 0.5, sendo considerado **inaceitável** e *studytime*, 0.75, que indica uma adequação **média**. Com isso, o valor do *KMO* geral deu 0.58.

```
Kaiser-Meyer-Olkin factor adequacy
```

```
call: KMO(r = dados_uteis)
```

```
overall MSA = 0.58
```

```
MSA for each item =
```

Idade	Escolaridade mãe	Escolaridade pai	Tempo de viagem	Tempo de estudo
0.53	0.53	0.56	0.56	0.75
Reprovações	Relação familiar	Tempo livre	Passeio	Alcool dias úteis
0.56	0.62	0.56	0.63	0.60
Alcool fim de semana	Saúde	Faltas		
0.58	0.43	0.64		

Análise Fatorial - KMO das variáveis envolvidas

- Removendo algumas variáveis cujo *KMO* individual obtido foi baixo, tem-se, então, um novo conjunto com 7 variáveis, em que o *KMO* geral obtido foi de 0.64, conforme a saída gerado pelo R.

```
Kaiser-Meyer-Olkin factor adequacy
```

```
Call: KMO(r = dados_uteis)
```

```
Overall MSA = 0.64
```

```
MSA for each item =
```

Tempo de estudo	Relação familiar	Tempo livre	Passeio	Alcool dias úteis
0.80	0.62	0.61	0.67	0.63
Alcool fim de semana	Faltas			
0.61	0.64			

- Com isso, a adequação amostral se torna mais aceitável, passando de um valor para *KMO* geral de 0.58 (considerado mal) para 0.64 (considerado razoável).

Análise Fatorial - Teste de Bartlett

Também conhecido como Teste de Esfericidade de Bartlett, é um teste que verifica a hipótese de que as variáveis não são correlacionadas na população. Portanto:

H_0 : Matriz das correlações é a matriz identidade com determinante igual a 1;

H_1 : Matriz das correlações não é a matriz identidade com determinante igual a 1;

Assim, para prosseguir com a Análise Fatorial, espera-se que a hipótese H_0 seja rejeitada, ou melhor, o nível de significância p seja menor que 0,05:

Abaixo, a saída gerada pelo *R* indica que os dados se mostram adequados para uma análise fatorial:

```
$chi sq
[1] 291.6485

$ p.value
[1] 1.58641e-49

$df
[1] 21
```


Análise Fatorial - Cálculo e Interpretação de MSA

Variável	MSA
<i>studytime</i> (tempo de estudo)	0.80
<i>famrel</i> (qualidade das relações familiares)	0.62
<i>freetime</i> (tempo livre depois da escola)	0.61
<i>goout</i> (sair com os amigos)	0.67
<i>dalc</i> (consumo de álcool nos dias de trabalho)	0.63
<i>walc</i> (consumo de álcool nos fins de semana)	0.61
<i>absences</i> (número de faltas na escola)	0.64

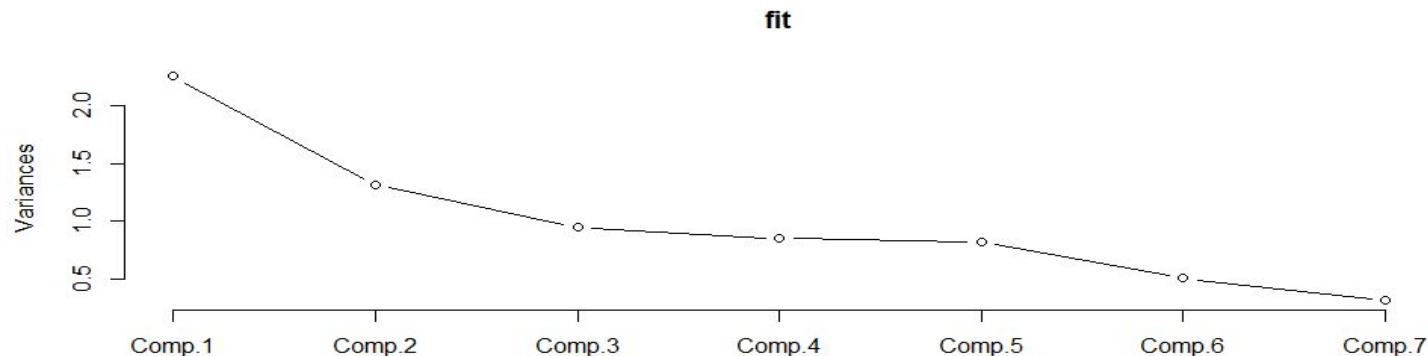
Todas as variáveis tiveram **MSA** > 0.5, portanto, todas serão consideradas no modelo e é possível também dizer que a Análise Fatorial é uma técnica adequada para o conjunto de dados em análise.

* Dentro os valores de *MSA* obtidos, a variável *studytime*, de longe, obteve o maior o valor, sendo 0.8 (considerado bom). Ao passo que as demais entraram na faixa 0.6 - 0.7 (razoável).

Análise Fatorial - Determinação do número de fatores

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Standard deviation	1.5011349	1.1464735	0.9726051	0.9221826	0.9035706	0.71171240	0.55931803
Proportion of Variance	0.3219151	0.1877716	0.1351372	0.1214887	0.1166343	0.07236208	0.04469095
Cumulative Proportion	0.3219151	0.5096868	0.6448240	0.7663127	0.8829470	0.95530905	1.00000000



Para a determinação do número de fatores foi utilizado o **critério do gráfico de scree**, na qual mostra que a variância única começa a dominar a estrutura de variância comum a partir do componente 2, ou melhor, a reta começa a ficar horizontal a partir do componente 2, portanto, será utilizado 2 fatores para a análise fatorial. É válido pontuar também que a escolha desse critério para definir o número de fatores se deu pela facilidade de identificar o início da horizontalidade do gráfico junto ao fato de que a amostra possui mais de 200 registros, o que torna o gráfico de scree um critério confiável de acordo com a literatura.

Análise Fatorial - Sem rotação

Principal Components Analysis

Call: principal(r = dados_uteis, nfactors = 2, rotate = "none",
scores = TRUE)

standardized loadings (pattern matrix) based upon correlation matrix

	PC1	PC2	h2	u2	com
Tempo de estudo	-0.34	-0.01	0.12	0.88	1.0
Relação familiar	0.16	0.66	0.46	0.54	1.1
Tempo livre	0.50	0.55	0.55	0.45	2.0
Passeio	0.74	0.25	0.61	0.39	1.2
Alcool dias úteis	0.76	-0.34	0.69	0.31	1.4
Alcool fim de semana	0.83	-0.21	0.73	0.27	1.1
Faltas	0.22	-0.60	0.41	0.59	1.3

	PC1	PC2
SS loadings	2.25	1.31
Proportion Var	0.32	0.19
Cumulative Var	0.32	0.51
Proportion Explained	0.63	0.37
Cumulative Proportion	0.63	1.00

Mean item complexity = 1.3

Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is 0.13
with the empirical chi square 176 with prob < 7.1e-34

Fit based upon off diagonal values = 0.71

Análise Fatorial - Com rotação (Varimax)

Principal Components Analysis

Call: principal(r = dados_uteis, nfactors = 2, rotate = "varimax",
scores = TRUE)

Standardized loadings (pattern matrix) based upon correlation matrix

	RC1	RC2	h2	u2	com
Tempo de estudo	-0.34	-0.07	0.12	0.88	1.1
Relação familiar	0.05	0.67	0.46	0.54	1.0
Tempo livre	0.40	0.62	0.55	0.45	1.7
Passeio	0.70	0.36	0.61	0.39	1.5
Alcool dias úteis	0.80	-0.22	0.69	0.31	1.1
Alcool fim de semana	0.85	-0.07	0.73	0.27	1.0
Faltas	0.32	-0.56	0.41	0.59	1.6

	RC1	RC2
SS loadings	2.23	1.34
Proportion Var	0.32	0.19
Cumulative Var	0.32	0.51
Proportion Explained	0.62	0.38
Cumulative Proportion	0.62	1.00

Mean item complexity = 1.3

Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is 0.13
with the empirical chi square 176 with prob < 7.1e-34

Fit based upon off diagonal values = 0.71

Interpretação e Conclusão

Interpretação

Através da análise fatorial, podemos dizer que há 2 fatores relevantes que explicam uma nota menor que 10 na prova G1:

- Fator 1: Consumo de álcool;
- Fator 2: Saúde mental.

Dentro do fator 1, existem as variáveis:

- Consumo de álcool durante o fim de semana e durante os dias úteis;
- Passeio (Se sai com os amigos);

E dentro do fator 2:

- Relação familiar;
- Tempo livre;

Conclusão

A análise fatorial foi excelente para responder a pergunta inicial, pois agrupa as variáveis determinantes em fatores relevantes.

Para o nosso caso, em que a pergunta inicial foi:

Dentre os alunos com notas baixas menores que 10, na primeira prova (G1), quais são os fatores relevantes que explicam tal nota?

Obtivemos que o consumo de álcool e a saúde mental do aluno são fatores importantes que explicam, em parte, o motivo de sua baixa performance na prova G1.



OBRIGADO!

