

# Efeitos do álcool nos estudos

Grupo 04:  
Maria Eduarda Garcia  
Mirela Mei

Prof.<sup>a</sup> Ana Amélia

# Dataset

## DESCRIÇÃO

Acompanha o desempenho estudantil na educação secundária em duas escolas portuguesas. Consideram-se matérias de matemática e português. Na presente análise, estudaremos apenas o primeiro caso.

## VARIÁVEIS

Os atributos incluem as notas dos alunos, dispositivos demográficos, sociais e relacionados à escola e foram coletados com o uso de relatórios e questionários escolares.

## DADOS COLETADOS

- UCI Machine Learning Repository
- 395 linhas e 33 colunas
- Sem missing data

# Dataset

## VARIÁVEIS

- Escola em que estuda
- Gênero
- Idade
- Área em que reside
- Tamanho da família
- Estado de coabitação dos pais
- Nível de escolaridade da mãe
- Nível de escolaridade do pai
- Trabalho do pai
- Trabalho da mãe
- Razão pela qual escola foi escolhida
- Quem possui a guarda
- Tempo da escola para casa
- Tempo de estudo semanal
- Repetências
- Suporte escolar extra
- Suporte familiar
- Classes extra pagas
- Atividades extracurriculares
- Frequentou escola de enfermagem
- Pretende cursar ensino superior
- Acesso à internet
- Participa de relação romântica
- Qualidade da relação familiar
- Tempo livre após a escola
- Frequência com que sai com amigos
- Consumo de álcool diário
- Consumo de álcool aos finais de semana
- Estado de saúde
- Número de faltas
- Nota do primeiro período
- Nota do segundo período
- Nota final

# Análise Fatorial

## DESCRIÇÃO

A análise fatorial é utilizada para tornar mais simples um estudo complexo, reduzindo um grande número de variáveis correlacionadas em fatores com baixa correlação entre si. Ela estabelece a correlação das variáveis observáveis e as organiza em fatores, que por si só são variáveis não observáveis.

Portanto, os dois principais usos da análise fatorial são resumo e redução dos dados, que podem ser muito úteis à medida que o número de variáveis utilizadas em técnicas multivariadas aumenta.

# Análise Fatorial

## EXPLORATÓRIA (EFA) X CONFIRMATÓRIA (CFA)

Na AFE, deixamos os dados observados determinarem o modelo fatorial subjacente (raciocínio indutivo para inferir um modelo a partir dos dados observados). Já na AFC, derivamos um modelo fatorial a priori (raciocínio dedutivo para fazer hipóteses de uma estrutura antecipadamente). Nesse sentido, uma técnica exploratória “deixa os dados falarem por eles mesmos”, não existe uma intervenção do pesquisador predeterminando uma estrutura.

# Análise Fatorial

## EXEMPLO

O dono do estabelecimento gostaria de entender o que faz com que os clientes visitem o local uma vez e não voltem depois, preferindo comprar em um concorrente.

Para isso, ele realiza uma pesquisa qualitativa e constata que os clientes reclamam de vários fatores, tais como atendentes de mau humor, o aspecto das verduras, produtos que estão quase chegando na data de validade, formas de pagamento disponíveis dentre outras reclamações.

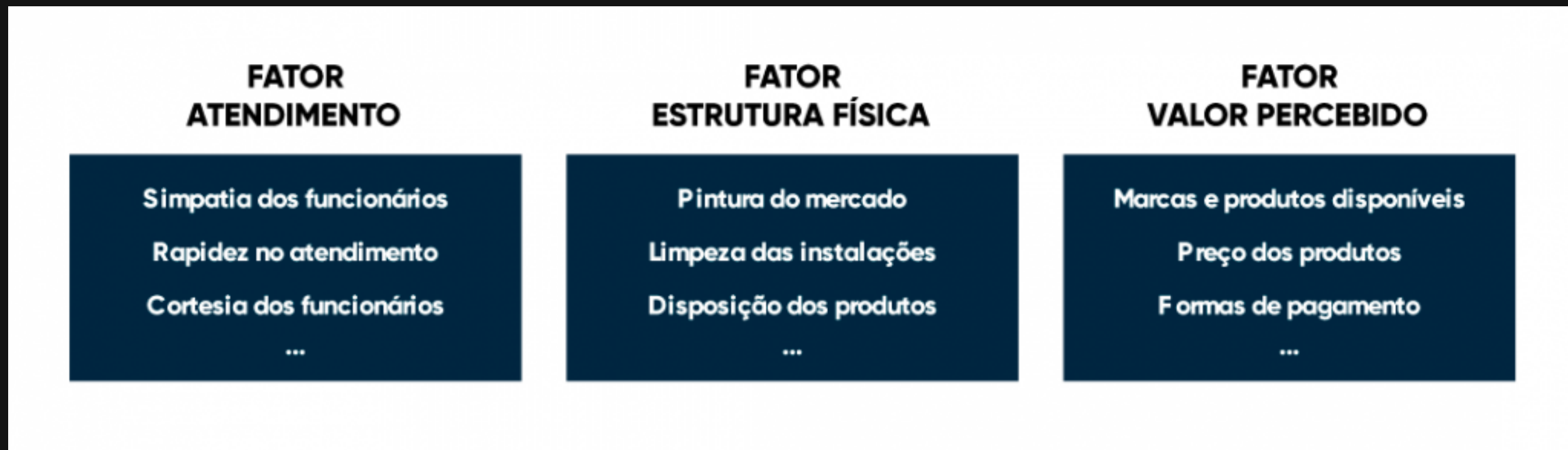
06 Ao final da pesquisa, ele percebe que possui mais de cinquenta tipos de reclamações diferentes, e simplesmente não tem tempo para analisar como cada uma delas influencia na sua perda de vendas. Diante disso, decide fazer uma pesquisa quantitativa e utilizar a análise fatorial exploratória.



# Análise Fatorial

## EXEMPLO

Após aplicar um questionário medindo a satisfação dos clientes em relação aos mais de cinquenta tipos de reclamações, utilizou a análise fatorial e verificou que havia correlação entre algumas reclamações. Dessa forma, ele pôde nomear cada grupo de variáveis e assim compreender quais são os principais fatores que influenciam nas vendas do seu mercado, entender quais são os seus pontos fortes e onde precisa melhorar.



# Análise Fatorial

As variáveis escolhidas para a análise foram:

- idade;
- grau de educação da mãe;
- grau de educação do pai;
- tempo da escola para casa;
- tempo estudando;
- reprovações;
- tempo livre;
- frequência que sai de casa;
- consumo de álcool diário;
- consumo de álcool aos fins de semana;
- saúde;
- nota do primeiro período (G1);
- nota do segundo período (G2);



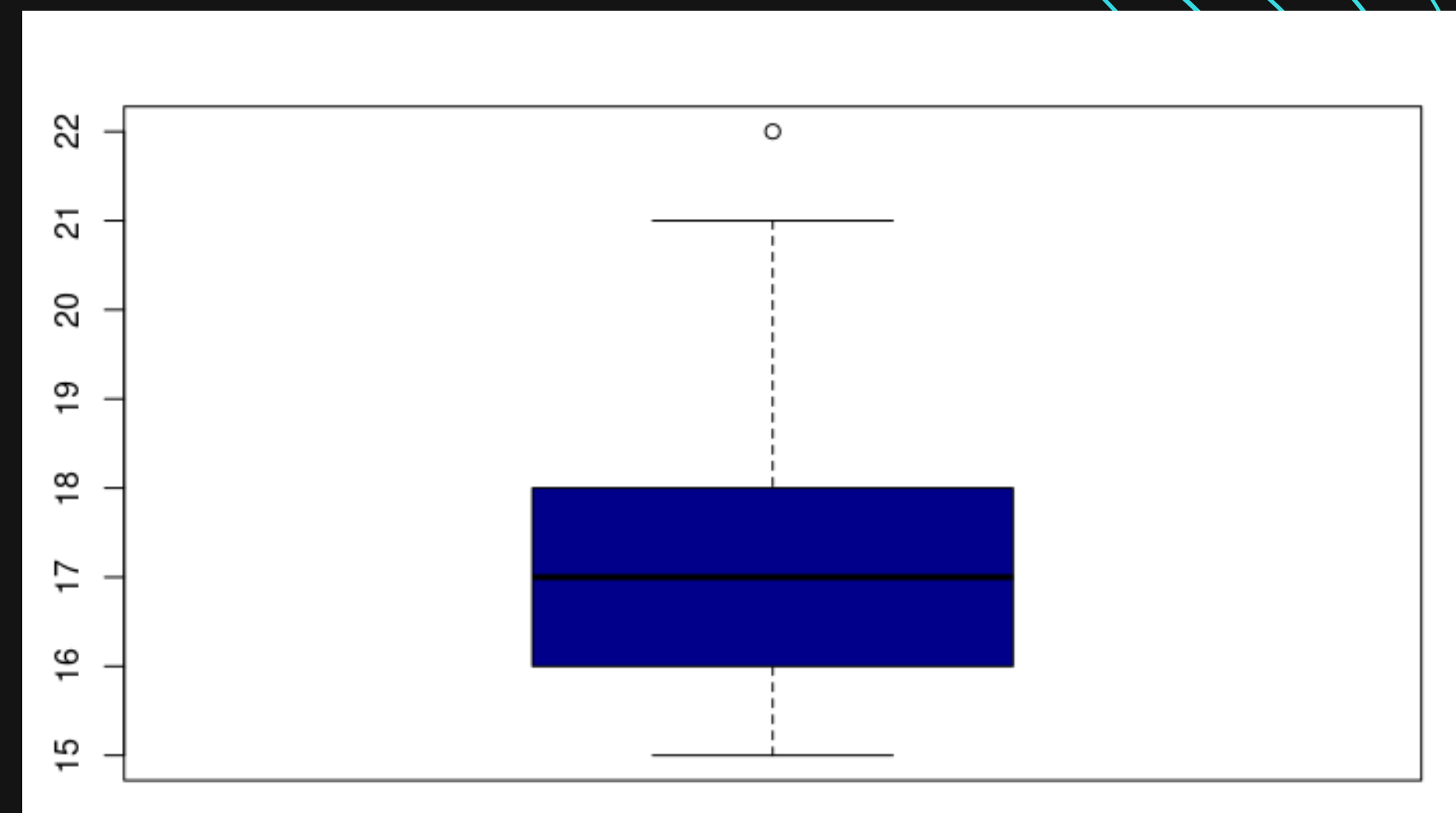
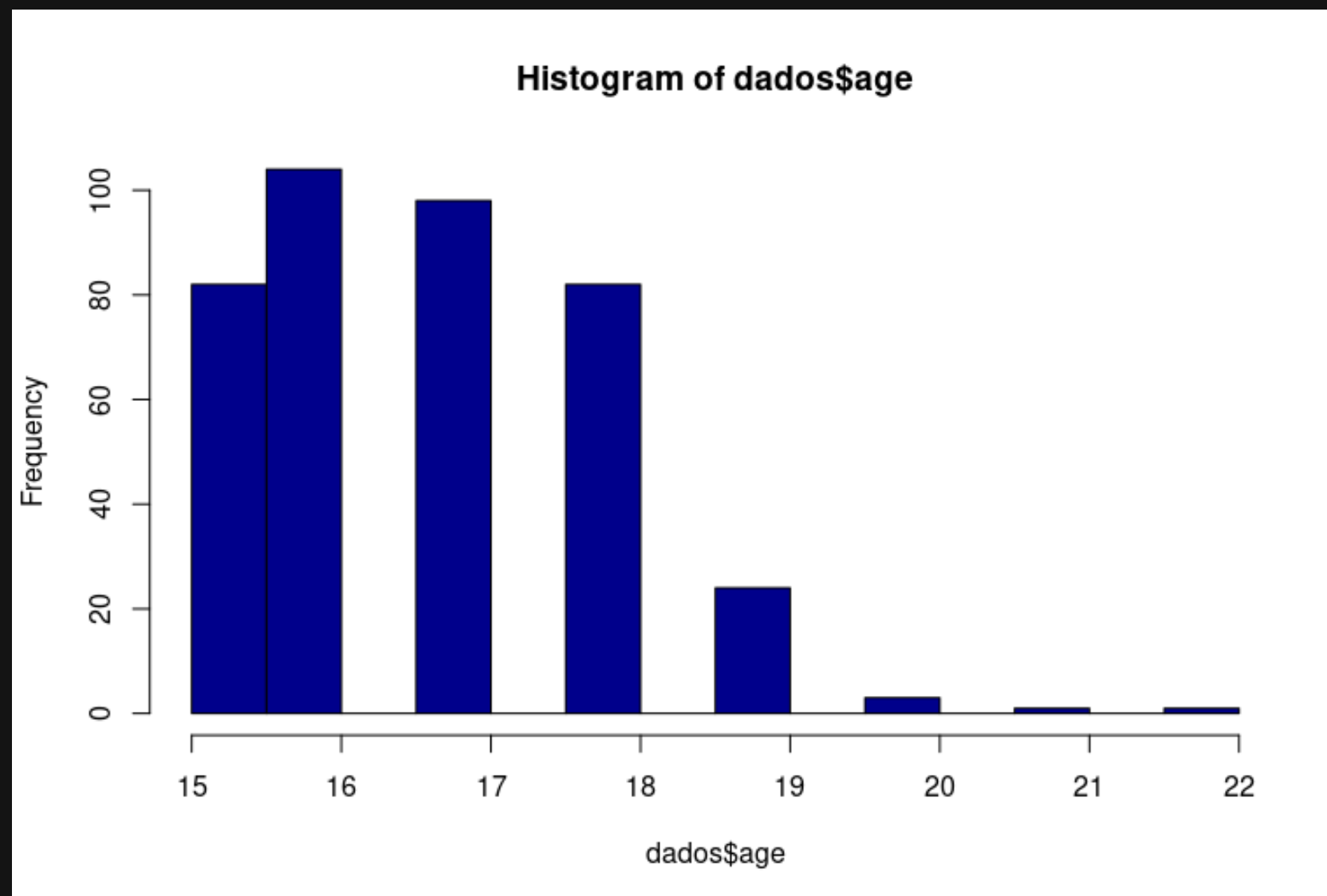
# Análise Fatorial

## PERGUNTA DA ANÁLISE

Quais fatores podem explicar as médias finais (G3) dos alunos da turma de matemática?

# Estatística Descritiva

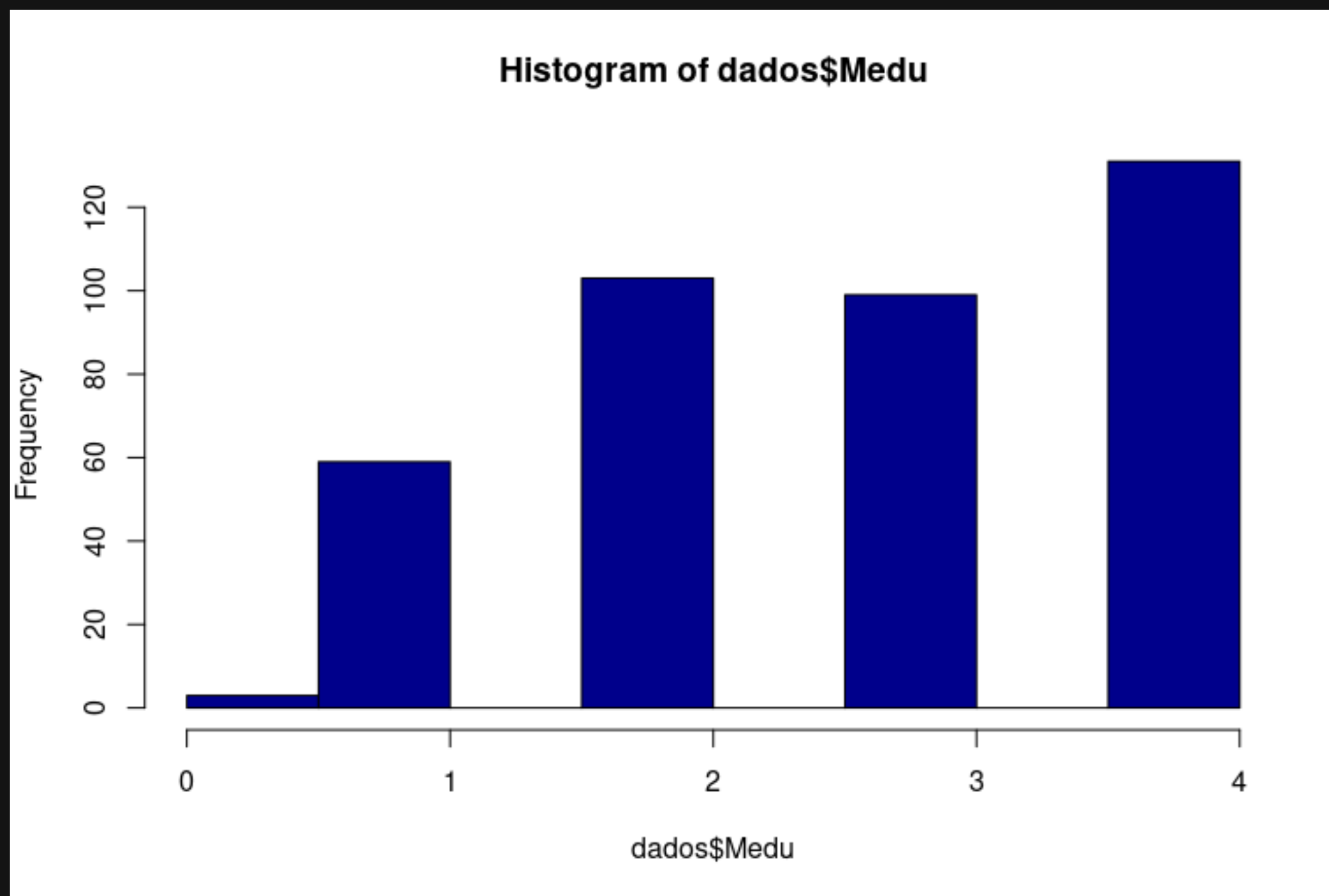
## IDADE



# Estatística Descritiva

## GRAU DE EDUCAÇÃO DA MÃE

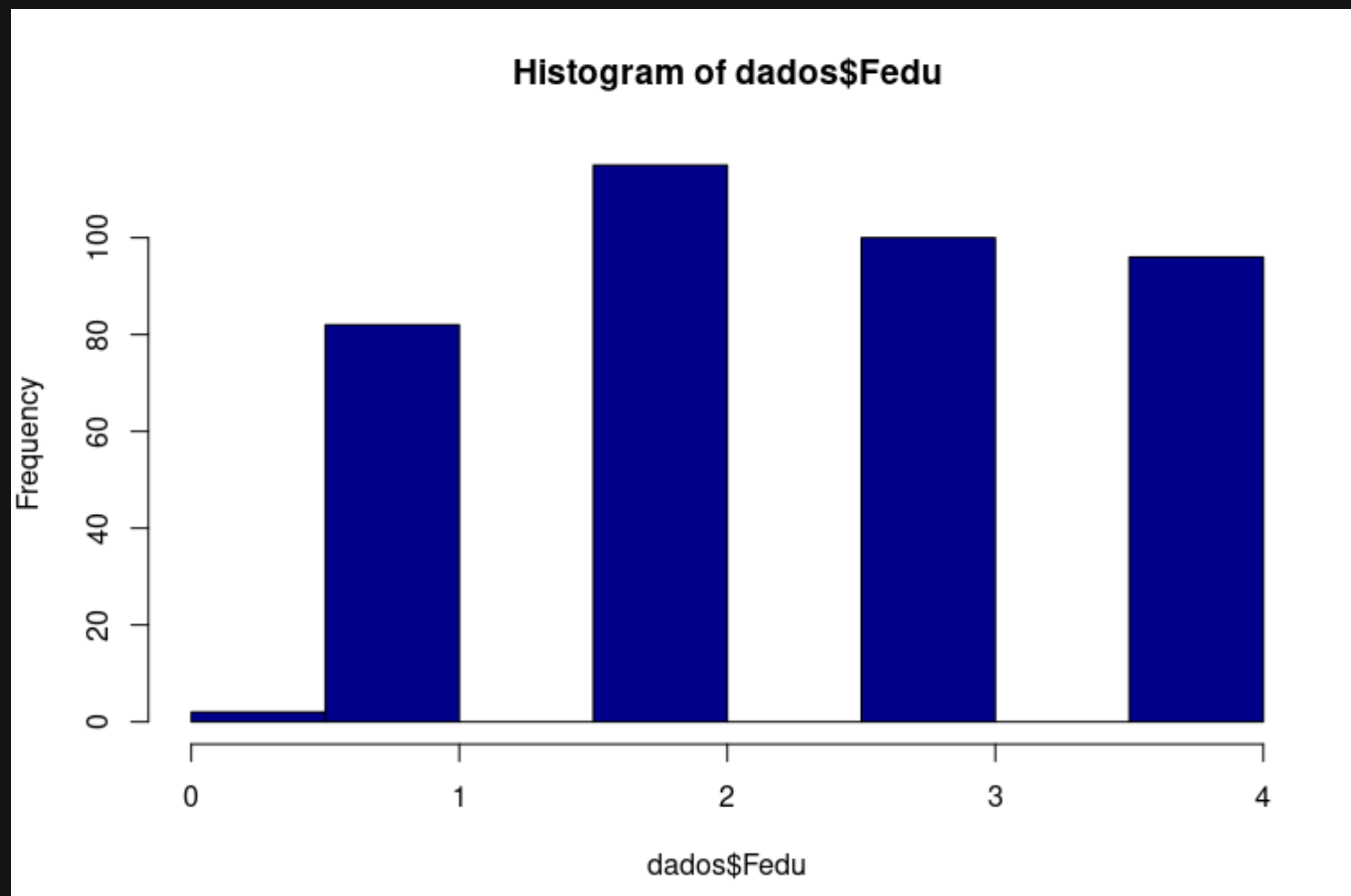
0 - nenhuma  
1 - até 4ª série  
2 - 5ª á 9ª série  
3 - ensino médio  
4 - nível superior



# Estatística Descritiva

## GRAU DE EDUCAÇÃO DO PAI

0 - nenhuma  
1 - até 4ª série  
2 - 5ª á 9ª série  
3 - ensino médio  
4 - nível superior



# Estatística Descritiva

## TEMPO DA ESCOLA PARA CASA

```
> table(dados$traveltime)
```

1	2	3	4
257	107	23	8

- 1 - menos de 15 min
- 2 - 15 a 30 min
- 3 - 30 min a 1 hora
- 4 - mais de 1 hora

## TEMPO ESTUDANDO

```
> table(dados$studytime)
```

1	2	3	4
105	198	65	27

- 1 - menos de 2 horas
- 2 - 2 a 5 horas
- 3 - 5 a 10 horas
- 4 - mais de 10 horas

# Estatística Descritiva

## REPROVAÇÕES

```
> table(dados$failures)
```

0	1	2	3
312	50	17	16

## TEMPO LIVRE

```
> table(dados$freetime)
```

1	2	3	4	5
19	64	157	115	40

1 - baixíssimo  
2 - baixo  
3 - médio  
4 - alto  
5 - altíssimo



# Estatística Descritiva

## FREQUÊNCIA QUE SAI DE CASA

```
> table(dados$goout)
```

1	2	3	4	5
23	103	130	86	53

- 1 - baixíssimo
- 2 - baixo
- 3 - médio
- 4 - alto
- 5 - altíssimo

## CONSUMO DE ÁLCOOL DIÁRIO

```
> table(dados$Dalc)
```

1	2	3	4	5
276	75	26	9	9

- 1 - baixíssimo
- 2 - baixo
- 3 - médio
- 4 - alto
- 5 - altíssimo

# Estatística Descritiva

## CONSUMO DE ÁLCOOL FINS DE SEMANA

```
> table(dados$Walc)
```

1	2	3	4	5
151	85	80	51	28

- 1 - baixíssimo
- 2 - baixo
- 3 - médio
- 4 - alto
- 5 - altíssimo

## ESTADO DE SAÚDE

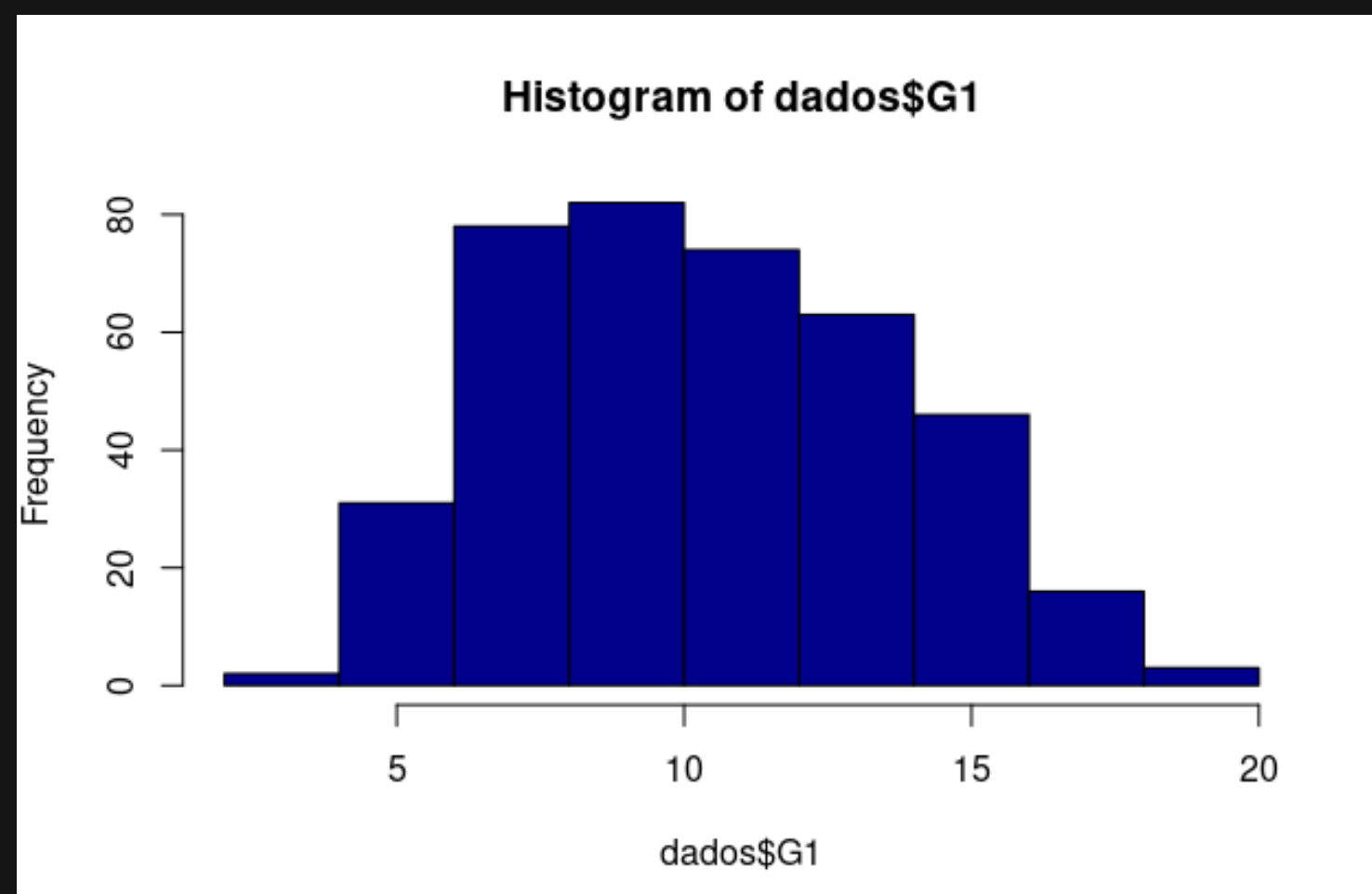
```
> table(dados$health)
```

1	2	3	4	5
47	45	91	66	146

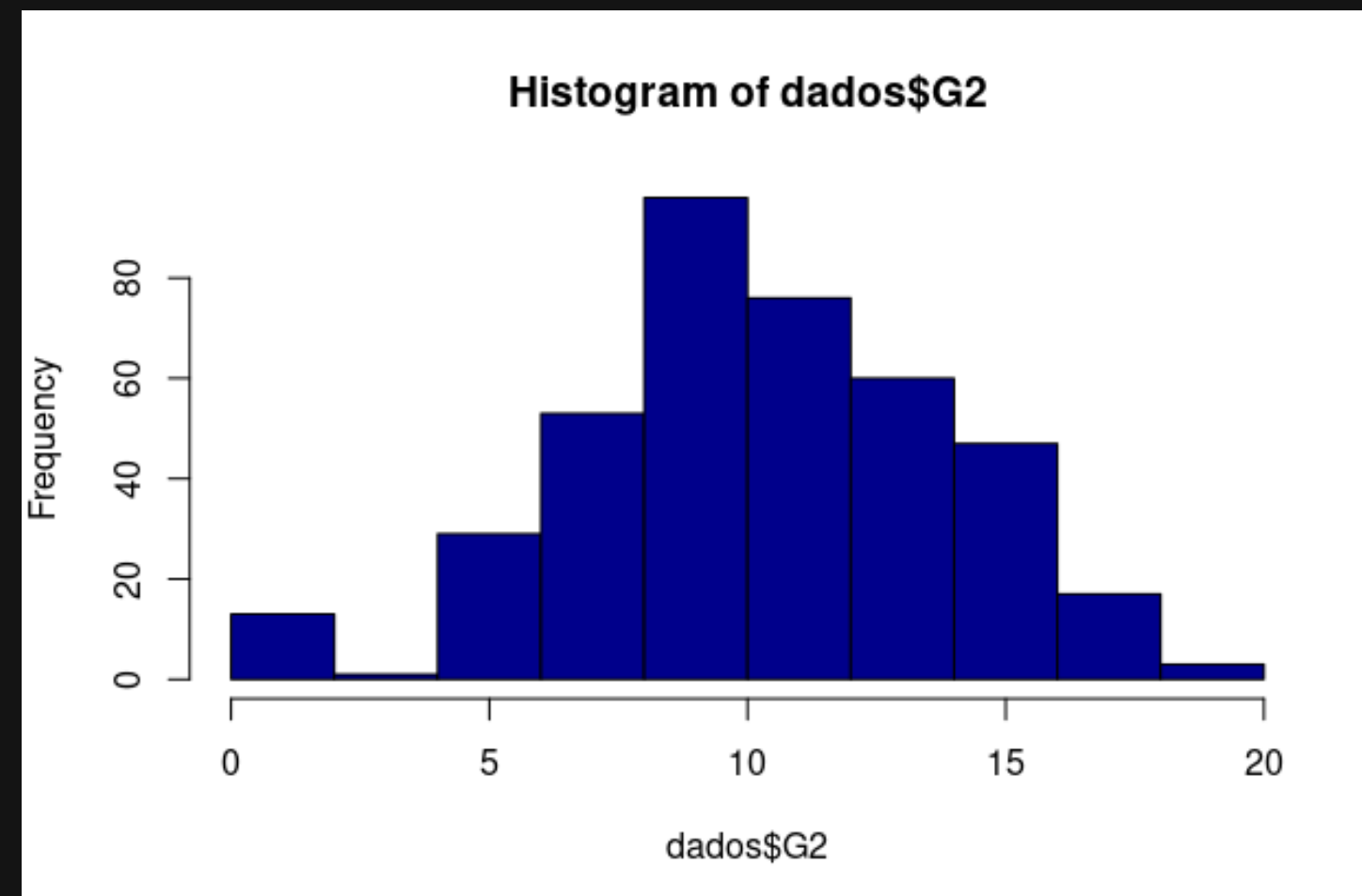
- 1 - baixíssimo
- 2 - baixo
- 3 - médio
- 4 - alto
- 5 - altíssimo

# Estatística Descritiva

## NOTAS PRIMEIRO PERÍODO

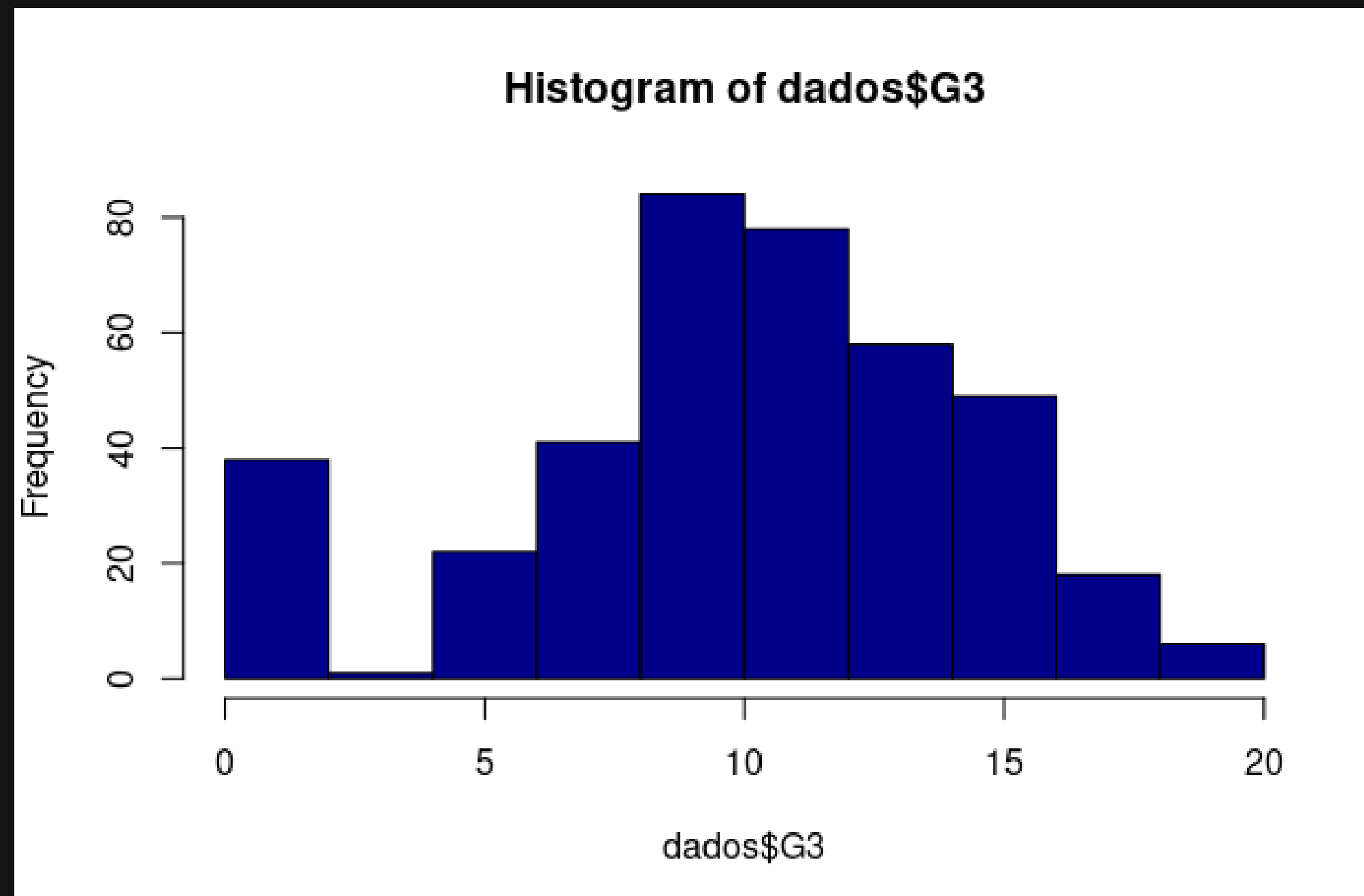


## NOTAS SEGUNDO PERÍODO



# Estatística Descritiva

## NOTAS FINAIS



# 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;
6. Rotação dos fatores mostrando as cargas antes e depois da rotação;
7. Interpretação dos resultados;

# MATRIZ DE CORRELAÇÃO

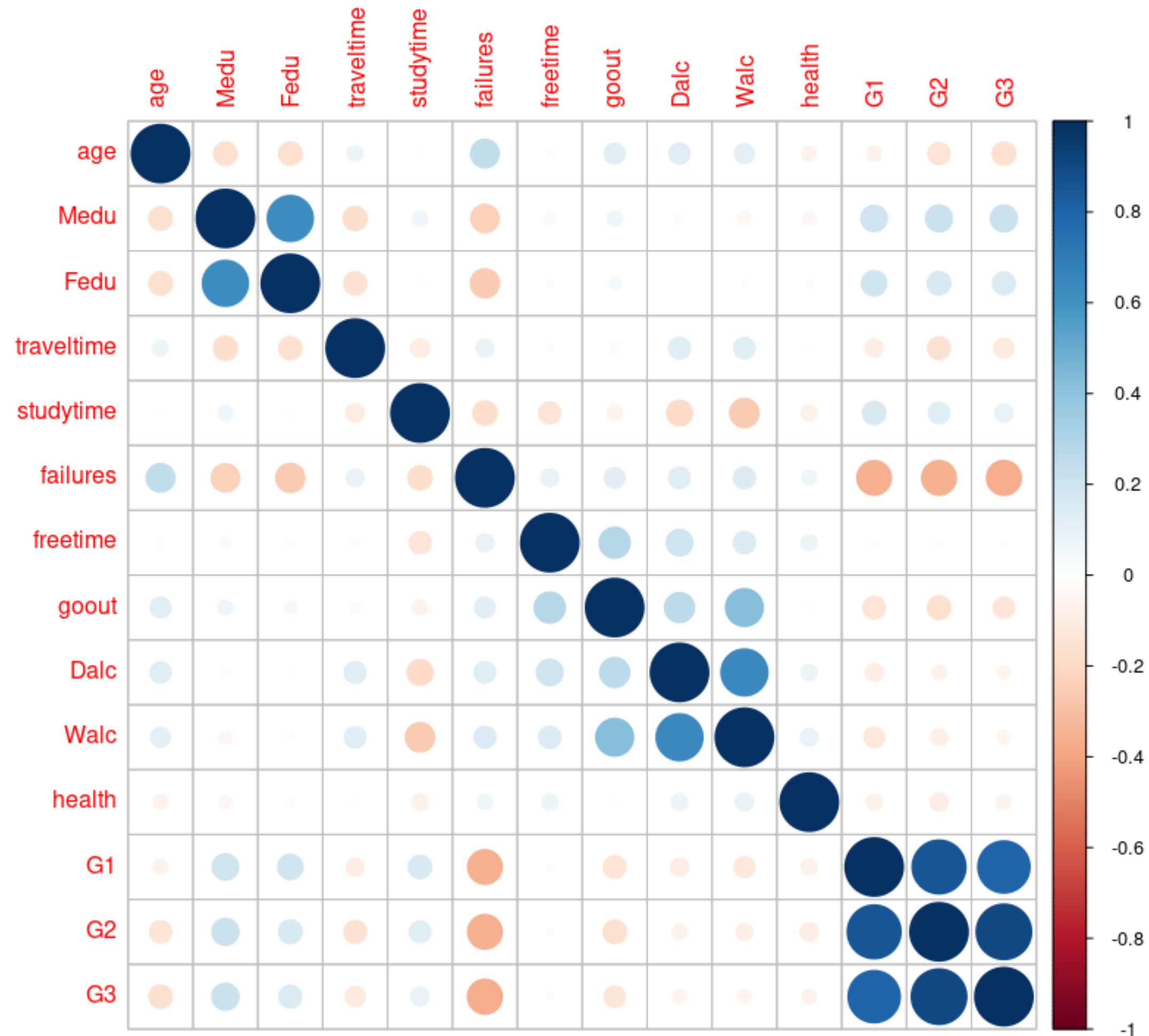
Para que a análise fatorial seja adequada, as variáveis devem ser correlacionadas. A Matriz de Correlação indica o nível de correlação entre as variáveis.

O triângulo inferior da matriz exibe as correlações simples,  $r$ , entre todos os pares possíveis de variáveis incluídas na análise, enquanto os elementos da diagonal, que são todos iguais a 1, em geral são omitidos.



```
matcor <- cor(dados2)
print(matcor, digits = 2)
corrplot(matcor, method="circle")
```





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.

# KMO (KAISER-MEYER-OLKIN)

A medida de adequacidade da amostra de KMO compara as magnitudes dos coeficientes de correlação observados com as magnitudes dos coeficientes de correlação parcial. Pequenos valores de KMO indicam que as correlações entre os pares de variáveis não podem ser explicadas por outras variáveis, indicando que a análise fatorial não é adequada.



```
p = 14
idiag <- seq(1, by = p + 1, length = p)
somar2 <- sum((as.numeric(matcor)[-idiag])^2)
cat("\n KMO = ",somar2 / (somar2 + sum((as.numeric(matcorp)[-idiag])^2)))
```

# TESTE DE BARTLETT

O teste de esfericidade de Bartlett mede se a análise fatorial é adequada ao problema. Em outras palavras, deve-se verificar se existe correlação suficientemente forte para que a análise fatorial possa ser aplicada

Temos aqui um teste de hipóteses:

- H0: a matriz de correlação é uma matriz identidade, não há correlação suficiente entre as variáveis. Análise não é adequada;
- H1: a análise é adequada, existe correlação.



```
bartlett.test(list(dados2$age, dados2$Medu, dados2$Fedu, dados2$traveltime,  
                  dados2$studytime, dados2$failures, dados2$freetime,  
                  dados2$goout, dados2$Dalc, dados2$Walc,  
                  dados2$health, dados2$G1, dados2$G2, dados2$G3))
```

# TESTE DE BARTLETT

Bartlett's K-squared = 4571.8, df = 13, p-value < 2.2e-16

Com base nos valores acima, é possível observar que o p-value é menor do que 5%, portanto rejeitamos  $H_0$  e a análise é considerada adequada.

# MSA - MEASURE OF SAMPLING ADEQUACY

MAA 1 = 0.6866743  
MAA 2 = 0.6267143  
MAA 3 = 0.5965955  
MAA 4 = 0.7389038  
MAA 5 = 0.7121953  
MAA 6 = 0.887226  
MAA 7 = 0.556829  
MAA 8 = 0.6515071  
MAA 9 = 0.6172841  
MAA 10 = 0.5914579  
MAA 11 = 0.55028  
MAA 12 = 0.829523  
MAA 13 = 0.7078214  
MAA 14 = 0.763046

```
for (j in 1:p) {  
  somar2j <- sum(matcor[j, -j]^2)  
  cat("\n MAA", j, "=", somar2j / (somar2j + sum(matcorp[j, -j]^2)))  
}
```

MSA indica o grau de explicação dos dados a partir dos fatores encontrados na Análise Fatorial. Caso o MSA indique um grau de explicação menor do que 0.50 significa que os fatores encontrados na análise não conseguem descrever, satisfatoriamente, as variações dos dados originais. É possível notar que a maior parte dos valores ficou entre médio e ruim, com exceção das variáveis 6 (reprovações) e 12 (G1) que obtiveram valores considerados bons.

# DETERMINAÇÃO DO NÚMERO DE FATORES

O objetivo da extração de fatores é encontrar um conjunto de fatores que formem uma combinação linear das variáveis originais ou da matriz de correlações.



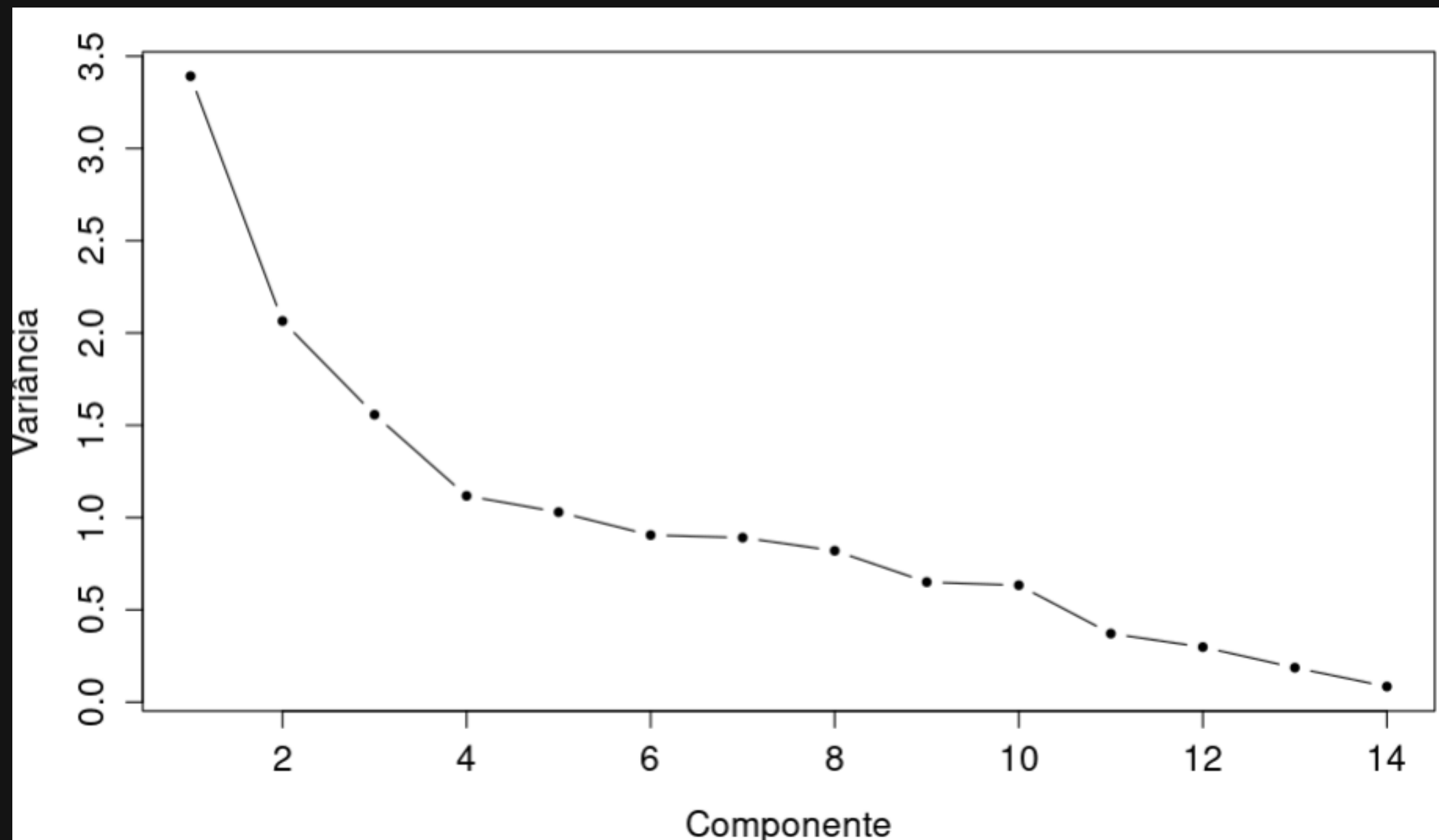
```
acpcor <- prcomp(dados2, scale = TRUE)  
summary(acpcor)
```



# DETERMINAÇÃO DO NÚMERO DE FATORES

## GRÁFICO DE DECLIVE (SCREE PLOT)

Utilizamos o método de representação gráfica (scree plot) para determinar o número de fatores. A escolha deve-se ao fato de ser um método de fácil observação, pois o ponto em que a inclinação suaviza indica o número de fatores a ser usados, que em geral é superior ao revelado pelos autovalores.



De acordo com a plotagem dos componentes calculados pela variância, pode-se observar uma estabilidade ao redor do fator 4, que foi o fator escolhido para realização de nossa análise.

# CÁLCULO DAS CARGAS FATORIAIS SEM ROTAÇÃO

De acordo com a escolha de 4 fatores, foi calculada a correlação de cada variável com cada um dos fatores, a carga fatorial. Se essa carga assume um valor positivo, significa que a variável está positivamente correlacionada com o fator, e, se assume valor negativo, essa correlação é negativa. As cargas fatoriais medem o grau de adaptação das variáveis aos fatores.



```
carfat<-principal(dados2, nfactors=4, rotate="none", score=TRUE)
```

# CÁLCULO DAS CARGAS FATORIAIS SEM ROTAÇÃO

Principal Components Analysis  
Call: principal(r = dados2, nfactors = 4, rotate = "none", scores = TRUE)  
Standardized loadings (pattern matrix) based upon correlation matrix

	PC1	PC2	PC3	PC4	h2	u2	com
age	-0.30	0.07	0.32	0.54	0.49	0.506	2.3
Medu	0.43	0.29	-0.67	0.13	0.73	0.267	2.2
Fedu	0.39	0.28	-0.71	0.02	0.72	0.277	1.9
traveltime	-0.27	0.06	0.30	-0.20	0.21	0.794	2.8
studytime	0.28	-0.32	-0.01	0.44	0.38	0.621	2.6
failures	-0.58	0.00	0.16	0.07	0.37	0.627	1.2
freetime	-0.12	0.45	0.01	0.00	0.22	0.780	1.1
goout	-0.30	0.56	-0.13	0.36	0.55	0.453	2.5
Dalc	-0.30	0.73	0.11	-0.04	0.63	0.368	1.4
Walc	-0.35	0.75	0.12	-0.03	0.70	0.302	1.5
health	-0.13	0.10	-0.08	-0.65	0.46	0.542	1.2
G1	0.83	0.21	0.33	0.03	0.85	0.149	1.5
G2	0.86	0.24	0.34	-0.01	0.92	0.077	1.5

A matriz não rotada apresenta, muitas vezes, dificuldades para ser interpretada pelo fato de que, os fatores, em geral, são correlacionados com muitas variáveis.

# CÁLCULO DA ROTAÇÃO VARIMAX

Uma rotação fatorial é o processo de manipulação ou de ajuste dos eixos fatoriais para conseguir uma solução fatorial mais simples e pragmaticamente mais significativa, cujos fatores sejam mais facilmente interpretáveis. O método Varimax é um método de rotação ortogonal, sendo o mais comumente utilizado dentre os métodos ortogonais, que procura minimizar o número de variáveis que apresentam altas cargas em cada fator.



```
carfatr<-principal(dados2, nfactors=4, rotate="varimax", score=TRUE)
```

# CÁLCULO DA ROTAÇÃO VARIMAX

Principal Components Analysis  
Call: principal(r = dados2, nfactors = 4, rotate = "varimax", scores = TRUE)  
Standardized loadings (pattern matrix) based upon correlation matrix

	RC1	RC2	RC3	RC4	h2	u2	com
age	-0.13	0.24	-0.30	0.57	0.49	0.506	2.1
Medu	0.13	0.10	0.84	0.00	0.73	0.267	1.1
Fedu	0.08	0.09	0.83	-0.11	0.72	0.277	1.1
traveltime	-0.06	0.15	-0.39	-0.16	0.21	0.794	1.7
studytime	0.12	-0.35	0.13	0.47	0.38	0.621	2.2
failures	-0.43	0.21	-0.37	0.08	0.37	0.627	2.5
freetime	0.02	0.47	0.03	-0.05	0.22	0.780	1.0
goout	-0.19	0.65	0.15	0.27	0.55	0.453	1.7
Dalc	-0.01	0.78	-0.09	-0.10	0.63	0.368	1.1
Walc	-0.04	0.82	-0.11	-0.10	0.70	0.302	1.1
health	-0.07	0.07	-0.07	-0.67	0.46	0.542	1.1
G1	0.91	-0.04	0.10	0.09	0.85	0.149	1.0
G2	0.95	-0.03	0.10	0.05	0.92	0.077	1.0

Com o processo da rotação, a matriz de fatores resulta numa matriz mais simples, sendo que a rotação não afeta as comunalidades e a porcentagem da variância explicada.

# INTERPRETAÇÃO DOS FATORES

Na matriz rotada, tem-se:

- O fator 1 apresenta altos coeficientes para as variáveis G1 e G2;
- O fator 2 apresenta altos coeficientes para as variáveis traveltime, failures, freetime, goout, Dalc (consumo de álcool diário), Walc (consumo de álcool em finais de semana), health;
- O fator 3 apresenta altos coeficientes para as variáveis Medu (grau de escolaridade da mãe), Fedu (grau de escolaridade do pai);
- O fator 4 apresenta altos coeficientes para as variáveis age, studytime;

Pode-se dizer, portanto, que a descrição de cada um dos fatores é:

- Fator 1: Notas finais dos dois anos anteriores
- Fator 2: Características pessoais, como o quanto viaja, sai, quanto tempo tem livre, saúde e consumo de álcool
- Fator 3: Estrutura familiar e educacional da família
- Fator 4: Características pessoais diretamente relacionadas ao rendimento escolar, como idade e tempo de estudo



# RESPOSTA À PERGUNTA INICIAL

A pergunta inicial a ser respondida foi: Quais fatores podem explicar as médias finais (G3) dos alunos da turma de matemática?

Entende-se que, pela análise fatorial, esses fatores são:

- As notas finais dos anos anteriores, ou seja, o desempenho do aluno até então e o quanto absorveu os anos anteriores ao final
- Características pessoais
- Características familiares
- Características relacionadas ao rendimento

Esses fatores parecem apresentar verossimilhança com a realidade e aparentemente explicar o fenômeno com medidores relevantes.

Observação: Algumas variáveis poderiam estar melhor distribuídas entre os fatores, como reprovações (failures) pertencendo ao fator 4 e idade (age) ao fator 2.

# CONCLUSÃO

Assim, a análise fatorial foi uma boa escolha para responder à pergunta inicial, uma vez que apresentou resultados consistentes.

# Bibliotecas utilizadas



```
library(corrplot)  
library(dplyr)  
library(psych)  
library(GPArotation)  
library(Rcmdr)
```

# Referências

OLIVEIRA, B. **Análise Fatorial: Uma importante técnica multivariada**. Disponível em: <<https://statplace.com.br/blog/analise-fatorial/>>.

ABUD, D. et al. **Análise fatorial Metodologias COLEÇÃO**. [s.l: s.n.]. Disponível em: <<https://repositorio.enap.gov.br/bitstream/1/4790/1/Livro%20An%C3%A1lise%20Fatorial.pdf>>.

BATTISTI, I. D. E.; SMOLSKI, F. M. DA S. **Capítulo 2 Análise Fatorial | Software R: curso avançado**. [s.l: s.n.].

**Análise fatorial em R**. [s.l: s.n.]. Disponível em: <<http://wiki.icmc.usp.br/images/b/b5/Factanalysis.pdf>>. Acesso em: 12 dez. 2022.

MATOS, Daniel; RODRIGUES, Erica. **Análise fatorial**. Disponível em <<https://repositorio.enap.gov.br/bitstream/1/4790/1/Livro%20Análise%20Fatorial.pdf>>.

AMARAL, Ernesto. **Aula 28 - Correlação e análise fatorial**. Disponível em <<https://www.ernestoamaral.com/docs/dcp046-111/Aula28.pdf>>

SILVA, Naje; FERREIRA, Wederson; CIRILLO, Marcelo; SCALON, João. **O uso da análise fatorial na descrição e identificação dos perfis característicos de municípios de Minas Gerais**. Disponível em <[http://jaguar.fcav.unesp.br/RME/fasciculos/v32/v32\\_n2/A3\\_Naje\\_Wederson.pdf](http://jaguar.fcav.unesp.br/RME/fasciculos/v32/v32_n2/A3_Naje_Wederson.pdf)>

PEREIRA, André; PALUDO, Berenice; VIEIRA, Manoel; CERBARO, Rodolfo. **Apostila Análise Fatorial**. Disponível em <[https://www.upf.br/\\_uploads/Conteudo/cepeac/textos-discussao/texto-02-2019.pdf](https://www.upf.br/_uploads/Conteudo/cepeac/textos-discussao/texto-02-2019.pdf)>.

REIS, Denise. **Capítulo 2: Análise Fatorial**. Dispoível em <<https://smolski.github.io/livroavancado/analysf.html>>



Obrigada!