



Instituto Politécnico de Setúbal

Escola Superior de Tecnologia do Barreiro

**Projeto de “Análise e Tratamento de Dados  
Multivariados”**

Licenciatura em Bioinformática

**Análise ao Burnout**

janeiro de 2022

Grupo 5

João Yanga (202000167)

Pedro Neves (201400949)

Samuel Correia (202000094)

# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Metodologia</b>	<b>1</b>
2.1	Tratamento de dados . . . . .	1
2.2	Testes e métodos . . . . .	1
<b>3</b>	<b>Desenvolvimentos</b>	<b>3</b>
3.1	Análise descritiva univariada . . . . .	3
3.2	Análise descritiva bivariada . . . . .	16
3.3	Análise Inferencial . . . . .	19
3.4	Regressão Linear Múltipla . . . . .	22
3.5	Análise fatorial . . . . .	25
<b>4</b>	<b>Conclusões</b>	<b>30</b>


# 1 Introdução

Este relatório tem o objetivo de avaliar o nível de Burnout dos estudantes da ESTBarreiro e perceber os efeitos que este pode ter na vida escolar. Atualmente a população estudantil enfrenta diferentes tipos de pressões: a nível académico, na relação com os colegas, na capacidade de concretização dos seus objetivos e no aproveitamento escolar, bem como num futuro profissional. Todos estes fatores apresentam-se numa população onde a ocorrência de Burnout poderá limitar fortemente o bem-estar psicossocial, a eficácia profissional e o rendimento escolar dos alunos.

Em outubro de 2021 recolheram-se questionários de uma amostra representativa de 135 estudantes da ESTBarreiro. Os inqueritos foram feitos em sala de aula onde se escolheram respostas de diferentes cursos e anos curriculares para que de facto a amostra fosse o mais representativa possível. As respostas foram completamente anónimas. O questionário estava dividido em 2 partes: Informação geral/habitos do estudante e 15 afirmações classificadas na escala de burnout de Maslach. Estas afirmações têm a finalidade de classificar emocionalmente o percurso académico do estudante. (As afirmações eram classificadas de 0 a 6 em que 0 corresponde a "nunca" e 6 corresponde a "sempre").

## 2 Metodologia

### 2.1 Tratamento de dados

Todas as análises estatísticas foram executadas no programa RStudio  *Version 1.4.1717*. Foi necessário aplicar uma limpeza e tratamentos dos dados recolhidos, pois existiam valores inválidos e variáveis que não estavam representadas da melhor forma.

#### **Variável: Sexo**

Foram modificados os valores "1" e "2" para "Feminino" e "Masculino".

#### **Variável: Escolhi, opcao1 e PMentoria**

Foram modificados os valores "1" e "2" para "Sim" e "Não".

#### **Variável: HorasRedes, HorasTV**

Considerou-se impossível assistir televisão ou estar nas redes sociais durante 24 horas diárias e um dia não tem mais de 24 horas, logo foram removidos todos valores iguais ou superiores a 24.

### 2.2 Testes e métodos

Foram usados vários testes e métodos de diferentes áreas de estatística para responder as nossas problemáticas:

### **Análise descritiva univariada**

Fez-se uma análise descritiva de todas as variáveis individualmente:

- Média, moda, mediana, quartis, desvio-padrão, variância e desvio-padrão;
- Diagrama de extremos e quartis, tabelas de frequências, gráfico circular, histograma e gráfico de barras;
- Coeficiente de assimetria de pearson;
- Coeficiente de dispersão;

### **Análise descritiva bivariada**

Na análise bivariada procurou-se saber se existia alguma correlação entre duas variáveis:

- Coeficiente de correlação de spearman;
- Coeficiente de correlação de pearson;
- Coeficiente de associação de cramer;

### **Análise inferencial**

Na análise inferencial procurou-se responder às seguintes questões de investigação:

- Será que os estudantes do sexo feminino se sentem emocionalmente mais exaustos que os estudantes do sexo masculino depois de um dia de aulas?
- Será que os estudantes se sentem melhores alunos com o progredir do curso?
- Será que os alunos do curso de Bioinformática tem tido mais interesse comparativamente com os alunos dos outros cursos?

Teste usados:

- Teste de Mann-Whitney;
- Teste de Kruskal-Wallis;

### **Regressão Linear Múltipla**

Pretendeu-se criar um modelo que prevê as horas de estudo de um estudante através das horas que este passa nas redes sociais e na televisão.

Teste usados:

- Teste de Shapiro-Wilk;
- Teste de Breusch-Pagan;
- Teste de One-Way Anova;

### **Análise fatorial**

Pretendeu-se criar um modelo que resumia as variáveis burnout.

Teste usados:

- Medida de adequabilidade de Kaiser-Mayer-Olkin;
- Teste de esfericidade de Bartlett;

### 3 Desenvolvimento

#### 3.1 Análise descritiva univariada

Variável: Idade

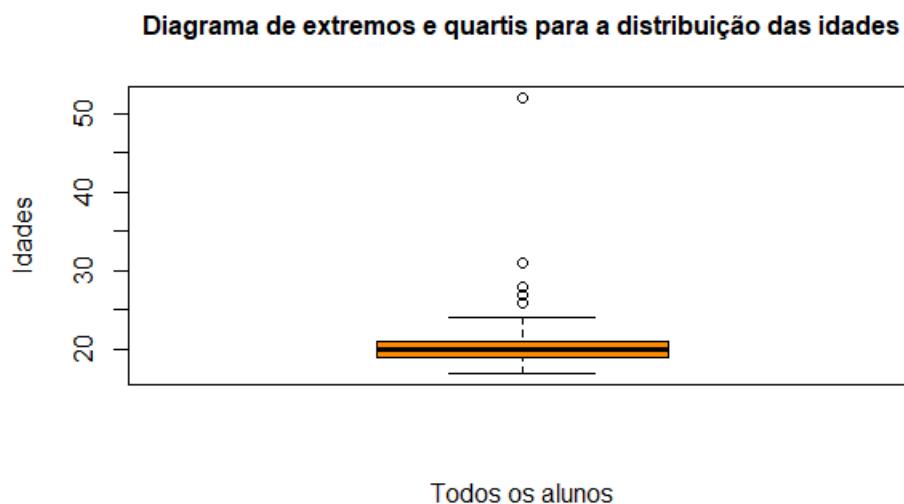


Figura 1: Diagrama de extremos e quartis para a distribuição das idades.

O diagrama de extremos e quartis para a distribuição das idades é apresentada na **Figura 1**. Dos 133 estudantes que responderam ao parâmetro idade, verifica-se que as idades variam entre 17 anos a 52 anos, sendo que existem 5 estudantes com idades muito superiores à média de 20.55 anos com um desvio padrão de 3.498. O coeficiente de assimetria de Pearson é de 0.157 e na **Figura 1** demonstram que existe uma distribuição simétrica, sendo que os estudantes estão relativamente concentrados na zona da idade média do diagrama de extremos e quartis de **Figura 1**.

#### SEM OUTLIERS

Existem 5 idades que fogem muito da normalidade por serem muito superiores comparado com as idades dos restantes 128 estudantes. Para evitar anormalidades na nossa análise e ter uma melhor visibilidade foi criado o diagrama de extremos e quartis para a distribuição das idades sem outliers da **Figura 2**. Neste diagrama as idades variam entre 17 anos a 24 anos, pois foram removidos 5 estudantes com idades superiores a 24 anos, o maior valor retirado corresponde a um estudante de 52 anos. A remoção destas idades fez com que a idade

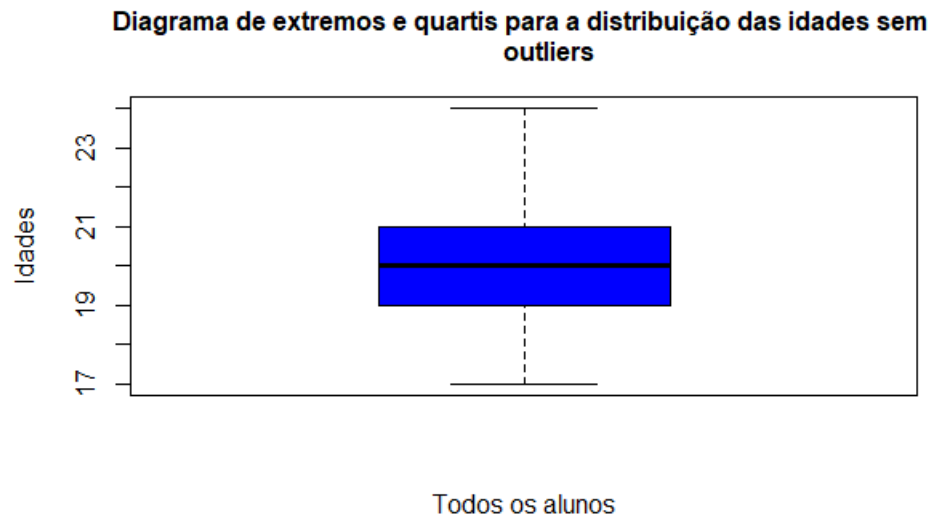


Figura 2: Diagrama de extremos e quartis para a distribuição das idades sem outliers.

média e o desvio padrão dos estudantes diminuíssse para 19.97 e 1.486, existe uma maior aproximação e uma distribuição das idades perto da média de uma forma simétrica suportado pelo o coeficiente de assimetria de Pearson de -0.02 e a **Figura 2**.

**Variável: Sexo**

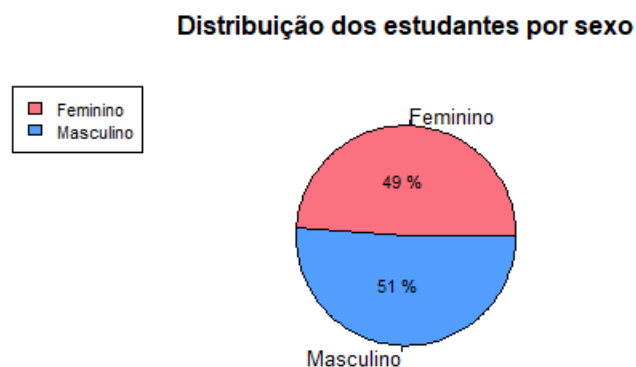


Figura 3: Distribuição dos estudantes por sexo.

A distribuições dos estudantes por sexo está apresentado na **Figura 3**. Dos 133 estudantes envolvidos no estudo que responderam ao parâmetro sexo, verifica-se que as percentagem de estudantes de ambos os sexos é bastante próxima tendo-se, no entanto, registado uma maior predominância de estudantes do sexo masculino (51%).

**Variável: Curso**

	Freq.Absoluta	Freq.Relativa
CTeSP TLQB	17	0.13
L. Bioinformática	31	0.23
L. Biotecnologia	85	0.63

Figura 4: Tabela de frequências dos cursos.

A tabela de frequências dos cursos está apresentado na **Figura 4**. Dos 133 estudantes que responderam ao parametro curso, verifica-se que existe uma grande variação entre as quantidades de estudantes em cada curso, sendo o curso mais

frequentado o de biotecnologia com 85 estudantes ( 63%), e o menos frequentado é o de CTeSP TLQB com apenas 17 estudantes (13%).

**Variável: Ano Curricular**

	Freq.Absoluta	Freq.Relativa
1º ano	63	0.47
2º ano	37	0.28
3ºano	32	0.24

Figura 5: Tabela de frequências dos anos.

A tabela de frequências dos anos apresentado na **Figura 5**. Dos 133 estudantes que responderam ao parametro ano curricular, verifica-se que existe alguma variação entre o 1º ano até ao 3º ano, sendo o ano mais frequentado o 1º ano com 63 estudantes (47%), e o menos frequentado é o 3º ano com 32 estudantes (24%).

**Variável: Opção 1**

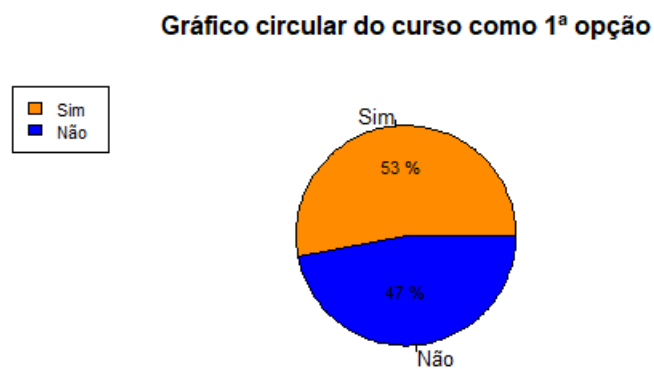


Figura 6: Gráfico circular do curso como 1ª opção.



O gráfico circular do curso como 1ª opção está apresentado na **Figura 6**. Dos 134 estudantes envolvidos no estudo que responderam se o curso que estão foi a primeira escolha, verifica-se que as percentagens de estudantes que responderam sim comparados ao que responderam não é bastante próxima tendo-se, no entanto, registado uma maior predominância de estudantes que responderam sim (53%).

**Variável: Escolhi**



Figura 7: Gráfico circular da escolha pessoal do curso.

O gráfico circular da escolha pessoal do curso está apresentado na **Figura 7**. Dos 134 estudantes envolvidos no estudo que responderam se o curso foi uma escolha própria, verifica-se que as percentagens de estudantes que responderam sim comparados ao que responderam não é muito desequilibrado tendo-se, registado uma predominância de estudantes que responderam sim (99%).

Variável: Tempo Deslocação

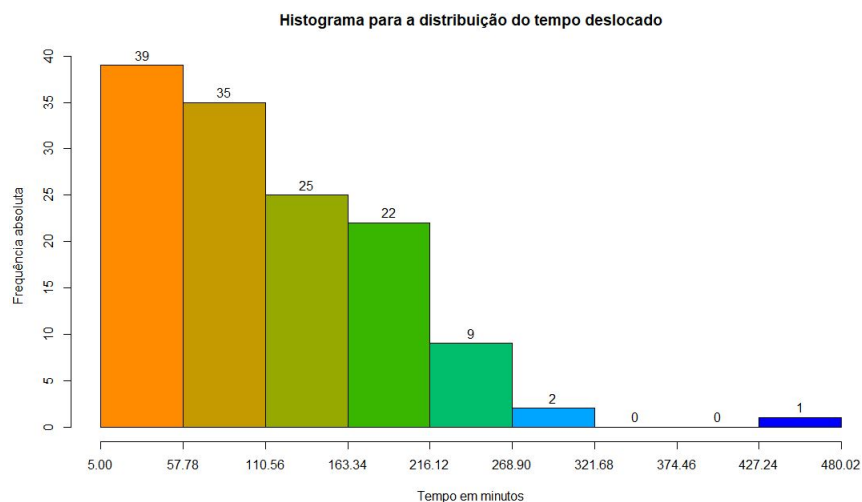


Figura 8: Histograma para a distribuição do tempo deslocado.

O histograma para a distribuição do tempo deslocado está apresentado na **Figura 8**. Dos 133 estudantes que responderam ao tempo total de deslocação entre casa e escola em minutos (ida e volta), verifica-se que o tempo varia entre 5 minutos a 427.24 minutos, existe alguma distanciação dos tempos demorados pelos estudantes, tendo-se no entanto, registado uma predominância do tempo deslocado demorar entre 5 minutos a 57.68 minutos (39 estudantes). A média do tempo de deslocação é 109.54 minutos e o desvio padrão é muito grande tendo um valor de 75.629, isto significa que os estudantes têm tempos de deslocação muito diferentes da média e comparados uns com outros, podendo estar em lugares geográficos diferentes ou usarem transportes diferentes. Conforme o coeficiente de assimetria de Pearson de 0.583 e a **Figura 8** existe uma distribuição assimétrica positiva já que estudantes estão mais concentrados na zona esquerda do histograma de **Figura 8**.

Variável: Horas Estudo

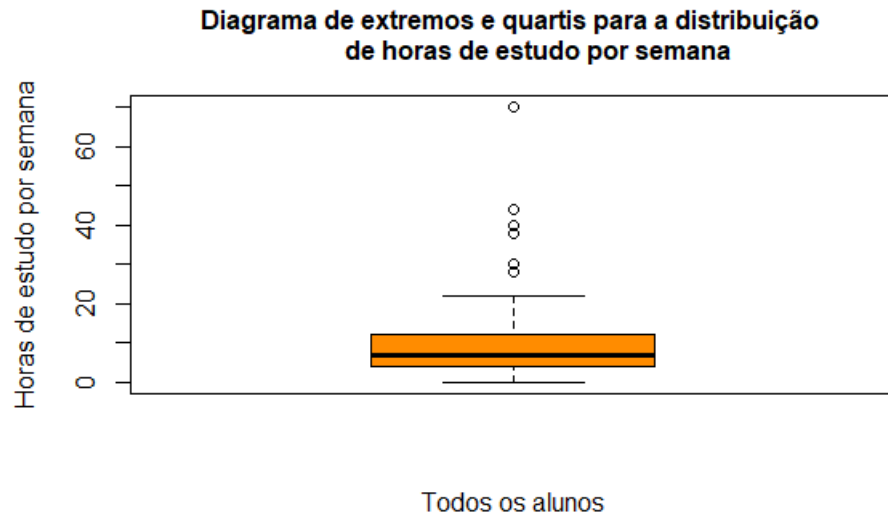


Figura 9: Diagrama de extremos e quartis para a distribuição de horas de estudo por semana.

O diagrama de extremos e quartis para a distribuição de horas de estudo por semana é apresentada na **Figura 9**. Dos 130 estudantes que responderam ao número de horas que estudam por semana, verifica-se que as horas de estudos variam entre 0 horas a 70 horas por semana, sendo que existem 6 estudantes com horas de estudo muito superiores á média de 9.37 horas com um desvio padrão de 9.215. O coeficiente de assimetria de Pearson de 0.583 e a **Figura 9** as horas de estudo dos alunos desviam-se um pouco do valor da média, criando uma distribuição assimétrica positiva, sendo que os estudantes estão mais concentrados na zona inferior do diagrama de extremos e quartis de **Figura 9**.

## SEM OUTLIERS

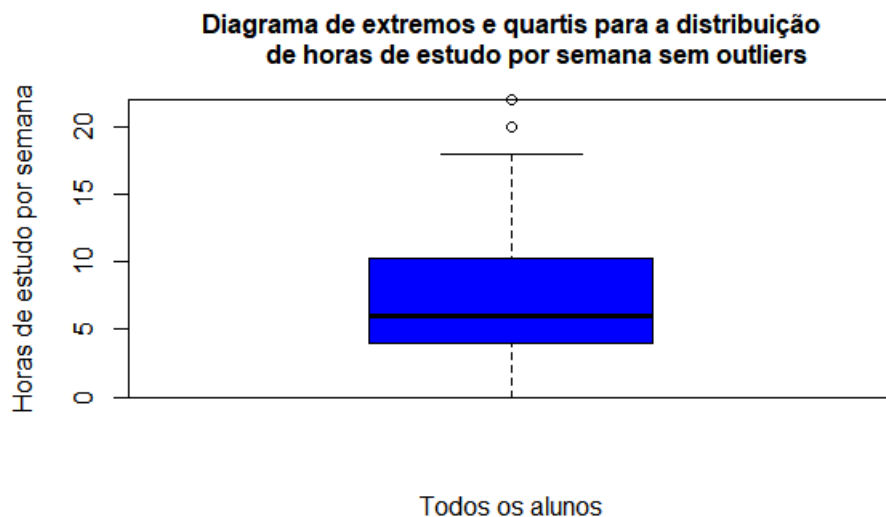


Figura 10: Diagrama de extremos e quartis para a distribuição de horas de estudo por semana sem outliers.

Existem 6 horas de estudo que fogem muito da normalidade por serem muito superiores comparado com as horas de estudo dos restantes 124 estudantes. Para evitar anormalidades na nossa análise e ter uma melhor visibilidade foi criado o diagrama de extremos e quartis para a distribuição de horas de estudo por semana sem outliers da **Figura 10**. Neste diagrama as horas de estudo variam de 0 horas a 22 horas, pois foram removidos os 6 estudantes com horas de estudo superiores a 22h, o maior valor retirado corresponde a um estudante que estuda 70 horas por semana. A remoção destas horas fez com que a hora de estudo média e o desvio padrão dos estudantes diminuísse para 7.24 e 5.136. O diagrama de extremos e quartis de **Figura 10** suportado pelo coeficiente de assimetria de Pearson de 0.742 demonstra uma distribuição assimétrica positiva maior, sendo que os estudantes estão mais concentrados na parte inferior do diagrama de extremos e quartis.

Variável: Horas Redes

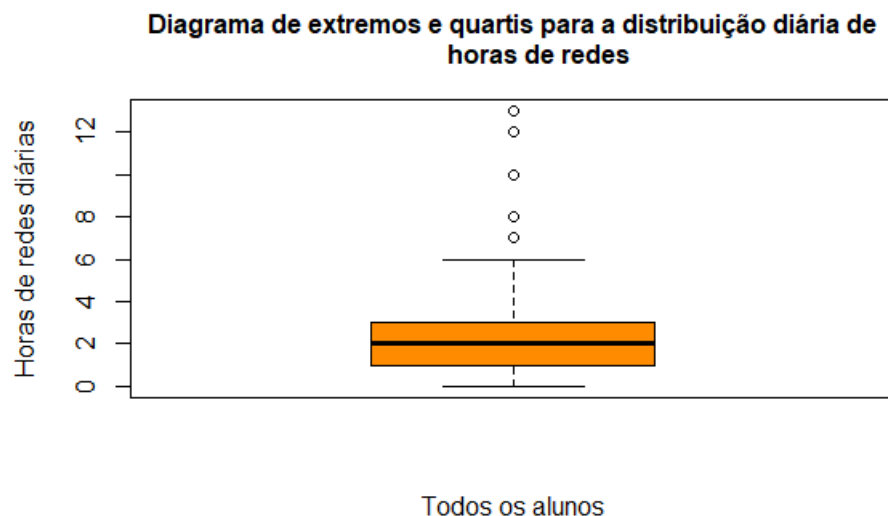


Figura 11: Diagrama de extremos e quartis para a distribuição diária de horas de redes.

O diagrama de extremos e quartis para a distribuição diária de horas de redes **Figura 11**. Dos 130 estudantes que responderam ao número de horas diárias que se dedica às redes sociais, verifica-se que as horas de redes variam entre 0 horas a 13 horas diárias, sendo que existem 7 estudantes com horas de redes muito superiores à média de 2.7 horas com um desvio padrão. O coeficiente de assimetria de Pearson de 0.310 e a **Figura 11** demonstram que existe uma distribuição simétrica, sendo que os estudantes estão relativamente concentrados na zona das horas de redes média do diagrama de extremos e quartis de **Figura 11**.

## SEM OUTLIERS

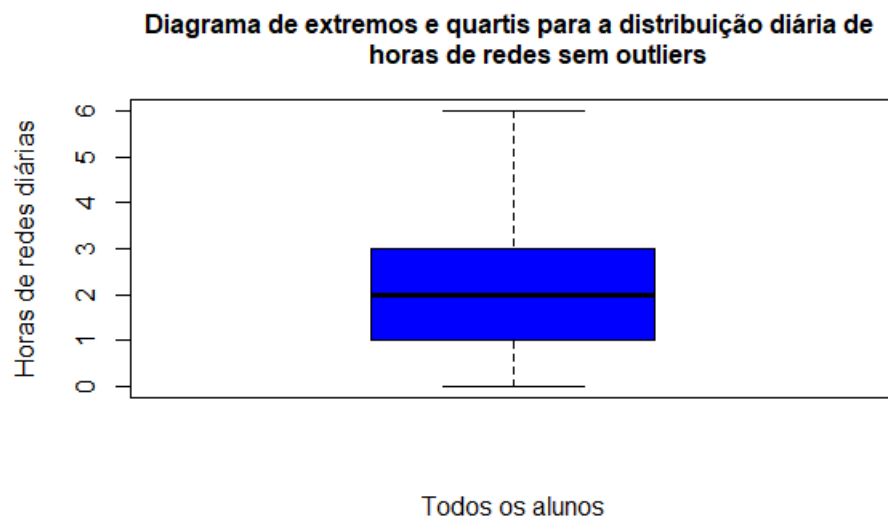


Figura 12: Diagrama de extremos e quartis para a distribuição diária de horas de redes sem outliers.

Existem 7 horas de redes que fogem muito da normalidade por serem muito superiores comparado com as horas de estudo dos restantes 123 estudantes. Para evitar anormalidades na nossa análise e ter uma melhor visibilidade foi criado o diagrama de extremos e quartis para a distribuição de horas diárias de horas de redes sem outliers da **Figura 12**. Neste diagrama as horas de redes variam de 0 horas a 6 horas, pois foram removidos os 7 estudantes com horas de estudo superiores a 6, o maior valor retirado corresponde a um estudante que está nas redes 13 horas diariamente. A remoção destas horas fez com que a hora de estudo média dos estudantes diminuísse ligeiramente para 2.3 com um desvio padrão de 1.362. O diagrama de extremos e quartis de **Figura 12** suportado pelo coeficiente de assimetria de Pearson de 0.22 demonstra uma distribuição simétrica, sendo que os estudantes estão concentrados perto da média de uma forma simétrica no do diagrama de extremos e quartis .

Variável: Horas TV

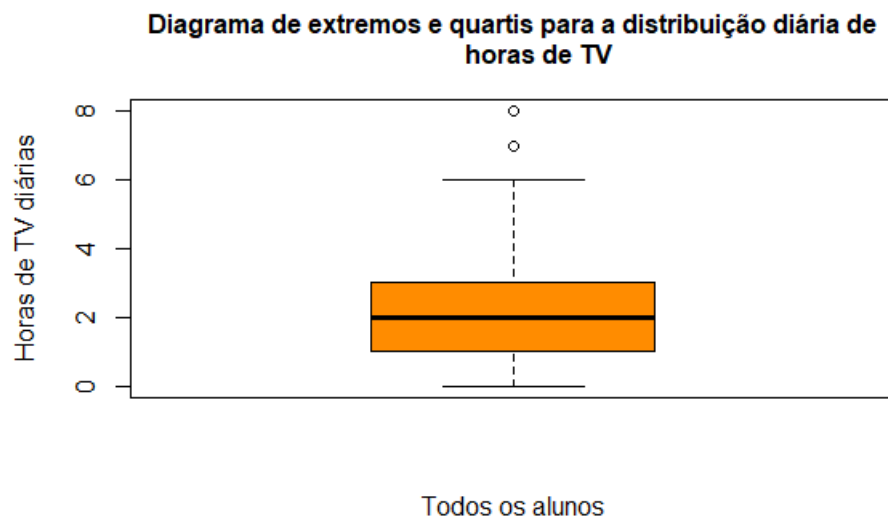


Figura 13: Diagrama de extremos e quartis para a distribuição diária de horas de TV.

O diagrama de extremos e quartis para a distribuição diária de horas de TV **Figura 13**. Dos 128 estudantes que responderam ao número de horas diárias que se dedica á tv, verifica-se que a quantidade de horas de TV variam entre 0 horas a 8 horas diárias, sendo que existem 2 estudantes com horas de TV muito superiores á média de 1.91 horas com um desvio padrão de 1.438. O Coeficiente de assimetria de Pearson de 0.633 e a **Figura 13** demonstram que existe uma distribuição ligeiramente simétrica, sendo que os estudantes estão concentrados na zona das horas de TV média do diagrama de extremos e quartis de **Figura 13**.

## SEM OUTLIERS

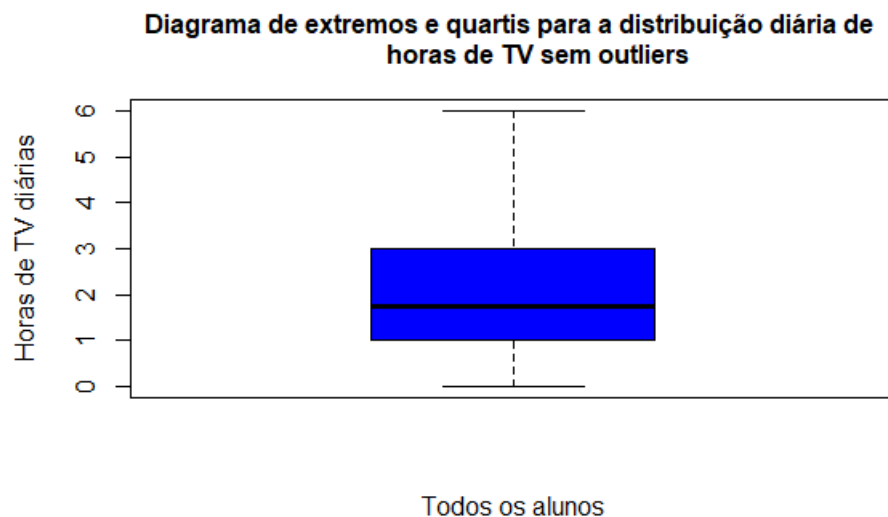


Figura 14: Diagrama de extremos e quartis para a distribuição diária de horas de TV sem outliers.

Existem 2 horas de TV que fogem muito da normalidade por serem muito superiores comparado com as horas de TV dos restantes 126 estudantes. Para evitar anormalidades na nossa análise e ter uma melhor visibilidade foi criado o diagrama de extremos e quartis para a distribuição de horas diárias de horas de TV sem outliers da **Figura 14**. Neste diagrama as horas de TV variam de 0 horas a 6 horas, pois foram removidos os 2 estudantes com horas de TV superiores a 6, o maior valor retirado corresponde a um estudante que vê TV 8 horas diariamente. A remoção destas horas fez com que a hora de TV média dos estudantes diminuísse ligeiramente para 1,81, existe um pequeno aumento na aproximação das horas de estudo com a média de horas de estudo com um desvio de padrão de 1.26. O diagrama de extremos e quartis de **Figura 14** suportado pelo coeficiente de assimetria de Pearson de 0.651 demonstra uma distribuição ligeiramente assimétrica, sendo que os estudantes estão concentrados na zona inferior do diagrama de extremos e quartis.



Variável: Horas Sono

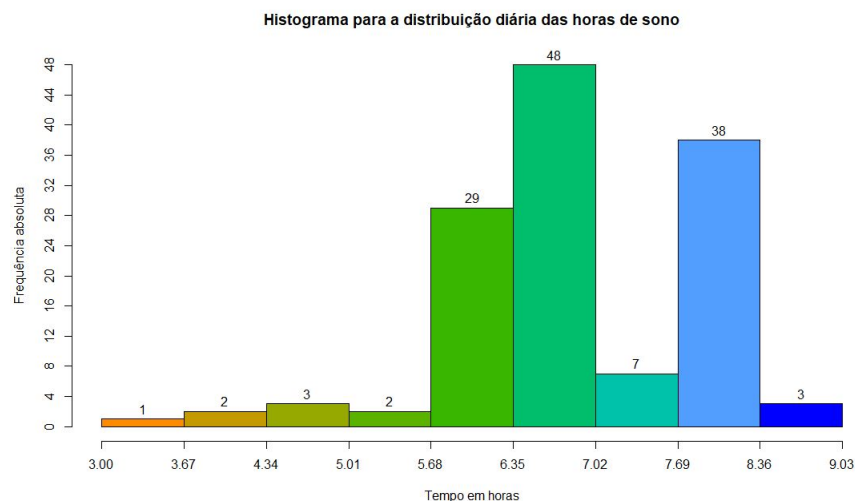


Figura 15: Histograma para a distribuição diária das horas de sono.

O histograma para a distribuição diária das horas de sono está apresentado na **Figura 15**. Dos 133 estudantes que responderam ao tempo que dormem diariamente em horas, verifica-se que o tempo varia entre 3 horas a 9.03 horas, existe alguma variação das horas de sono, tendo-se no entanto, registado uma predominância de os estudantes dormirem 6.35 horas a 7.02 horas (48 estudantes). A média das horas de sono é 6.96 horas e o desvio padrão é muito pequeno tendo um valor de 0.988, isto significa que os estudantes tendem a dormir horas com valor muito parecido a da média de 6.96 horas. Conforme o coeficiente de assimetria de Pearson de -0.04 e a **Figura 15** existe uma distribuição simétrica já que estudantes estão concentrados na zona da média do histograma da **Figura 8** simetricamente.

Variável: PMentoria

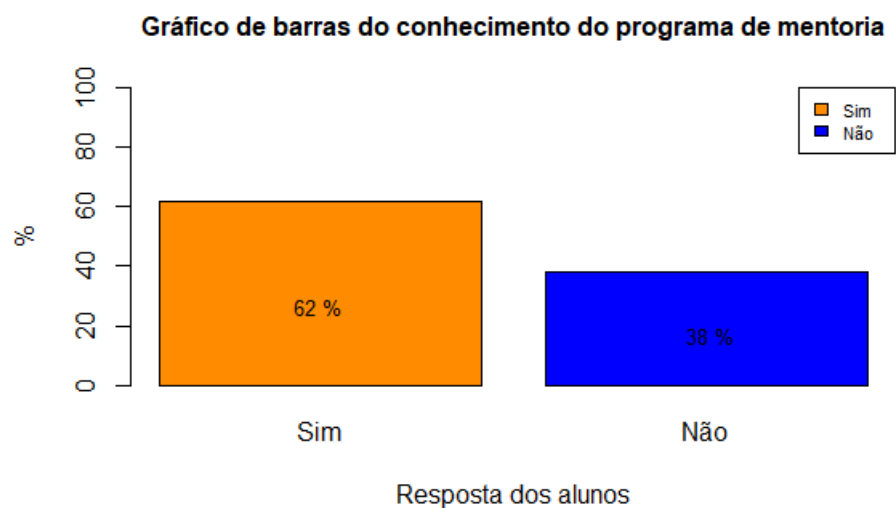


Figura 16: Gráfico de barras do conhecimento do programa de mentoria.

O gráfico de barras do conhecimento do programa de mentoria está apresentado na **Figura 16**. Dos 124 estudantes envolvidos no estudo que responderam se tinham conhecimento sobre a existência do programa de mentoria, verifica-se que as percentagens de estudantes que responderam sim comparados ao que responderam não é ligeiramente desequilibrada, registado uma predominância de estudantes que responderam sim (62%).

### 3.2 Análise descritiva bivariada

Para análise descritiva bivariada foi feito analisado as seguintes correlações:

- A exaustão sentida pelos estudantes de sexo diferente;
- O entusiasmo sentido pelo estudantes sobre os estudos com o ano curricular;
- As horas de estudo semanais com as horas de sono diárias;

### Correlação da exaustão sentida pelos estudantes de sexo diferente

	Burnout1	Sexo	Freq
1	Nunca	Feminino	0
2	Quase Nunca	Feminino	6
3	Algumas vezes	Feminino	22
4	Regularmente	Feminino	3
5	Muitas vezes	Feminino	11
6	Quase sempre	Feminino	14
7	Sempre	Feminino	9
8	Nunca	Masculino	1
9	Quase Nunca	Masculino	17
10	Algumas vezes	Masculino	18
11	Regularmente	Masculino	10
12	Muitas vezes	Masculino	11
13	Quase sempre	Masculino	6
14	Sempre	Masculino	4

Figura 17: Tabela de contingência da questão 1 do burnout pelo sexo dos estudantes.

A tabela de contingência do burnout 1 com o sexo está apresentado na **Figura 17**. Foi feito o coeficiente de associação de cramer com um valor de 0.343, sendo que existe baixa associação entre a questão 1 do burnout e o sexo dos estudantes.

### Correlação do entusiasmo sentido pelos estudantes sobre os estudos com o ano curricular

	Burnout7	Ano Curricular	Freq
1	Nunca	1	9
2	Quase Nunca	1	14
3	Algumas vezes	1	22
4	Regularmente	1	8
5	Muitas vezes	1	7
6	Quase sempre	1	0
7	Sempre	1	2
8	Nunca	2	3
9	Quase Nunca	2	7
10	Algumas vezes	2	11
11	Regularmente	2	8
12	Muitas vezes	2	5
13	Quase sempre	2	2
14	Sempre	2	1
15	Nunca	3	5
16	Quase Nunca	3	5
17	Algumas vezes	3	8
18	Regularmente	3	6
19	Muitas vezes	3	6
20	Quase sempre	3	1
21	Sempre	3	1

Figura 18: Tabela de contigência da questão 7 do burnout pelo ano curricular dos estudantes.

A tabela de contigência da questão 7 do burnout pelo ano curricular dos estudantes está apresentado na **Figura 18**. Para análise da figura foi usado o **coeficiente de correlação de spearman** e foi obtido um valor de 0.125, conforme o valor obtido existe uma correlação muito fraca, sendo que o nível de entusiasmo sentido pelos estudantes sobre os estudos é muito pouco afetado pelo ano curricular que estão.

## Correlação das horas de estudo semanais com as horas de sono diárias

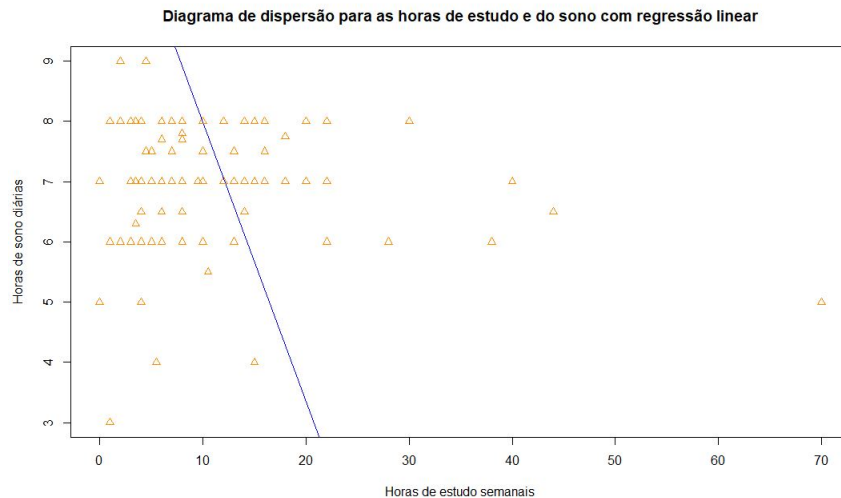


Figura 19: Regressão linear das horas de estudo com as de sono.

O gráfico da regressão linear das horas de estudo com as de sono está apresentado na **Figura 19**. Para melhor análise da correlação foi usado o **coeficiente de correlação de pearson** e foi obtido um valor de -0.0492, conforme o valor e o gráfico obtido pode-se dizer que não existe correlação, sendo que as horas de estudo semanais não vão afetar as horas de sono dos estudantes.

### 3.3 Análise Inferencial

Escolheram-se três questões de investigação com a finalidade de as responder usando a inferência estatística.

As questões escolhidas foram:

- Será que os estudantes do sexo feminino se sentem emocionalmente mais exaustos que os estudantes do sexo masculino depois de um dia de aulas?(Variáveis: **sexo**, **burnout 1**)
- Será que os estudantes se sentem melhores alunos com o progredir do curso?(Variáveis: **ano curricular**, **burnout 12**)
- Será que os alunos do curso de Bioinformática tem tido mais interesse comparativamente com os alunos dos outros cursos?(Variáveis: **curso**, **burnout 14**) (Variáveis: **curso**, **burnout 14**)

**1ª questão:** Será que os estudantes do sexo feminino se sentem emocionalmente mais exaustos que os estudantes do sexo masculino depois de um dia de aulas?

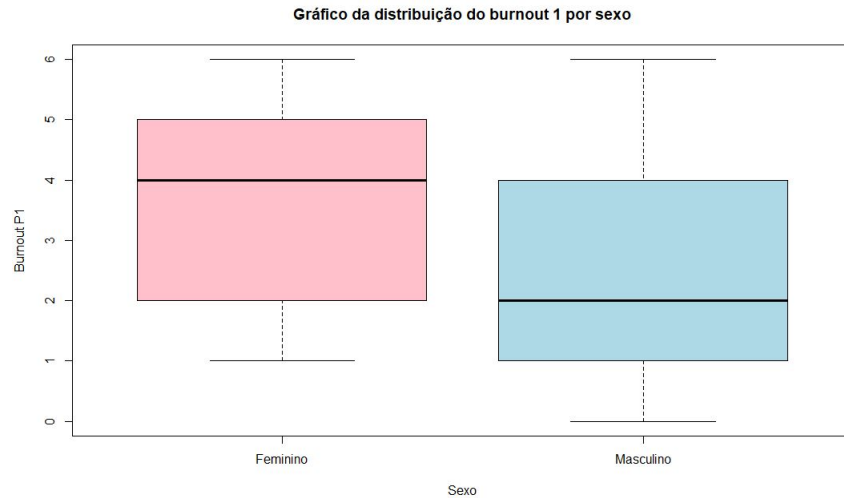


Figura 20: Gráfico da distribuição do burnout 1 por sexo.

Para verificar a existência de diferenças no nível de exaustão emocional depois de estudar entre os sexos foi feito o teste **Mann-Whitney** e o gráfico da distribuição do burnout 1 por sexo representado na **Figura 20**. O gráfico apresenta uma variação da exaustão entre o sexo feminino e masculino, tendo que o sexo feminino fica mais exausto que o sexo masculino ( $W = 38004.5$ ;  $p = 2.2e-16$ ).

**2ª questão:** Será que os estudantes se sentem melhores alunos com o progredir do curso?

Para avaliar se existiam diferenças no nível de auto-estima entre os alunos que frequentam os 3 anos do curso de licenciatura foi usado o teste de **Kruskal-Wallis**, tendo em conta que a variável ano curricular contempla o três anos curriculares assumiu-se que cada ano era uma população.

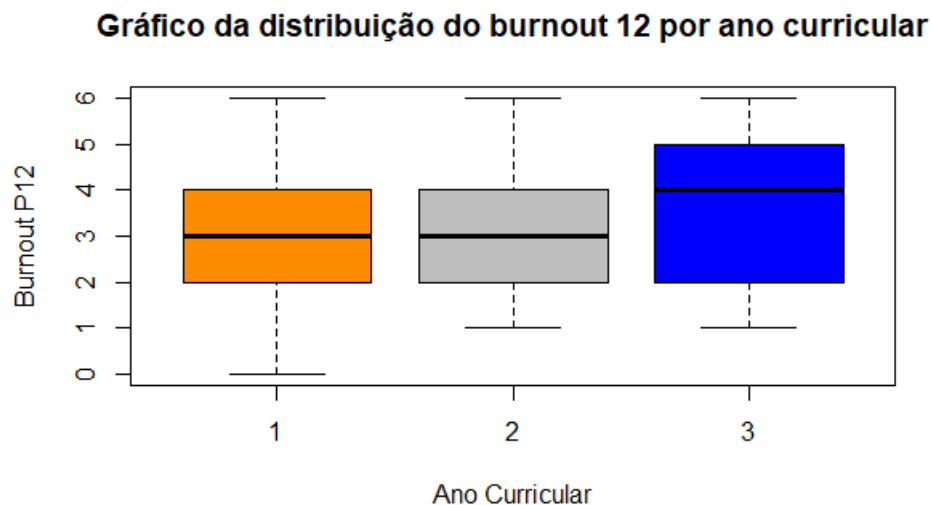


Figura 21: Gráfico da distribuição do burnout 12 por ano curricular.

Conforme o teste de **Kruskal-Wallis** ( $DF = 2$ ;  $p = 0.112$ ) não existe mudanças no sentimento de sucesso dos estudantes, enquanto avançam nos cursos. Isto é apoiado pelo gráfico da distribuição do burnout 12 por ano curricular representado na **Figura 21**

**3ª questão:** Será que os alunos do curso de Bioinformática tem tido mais interesse comparativamente com os alunos dos outros cursos?

Usou-se novamente o teste de **Kruskal-Wallis** para avaliar se existe variância no nível de interesse nos alunos de bioinformática relativamente aos outros. Considerou os três diferentes cursos (Licenciatura em Bioinformática, Licenciatura em Biotecnologia e o CTeSP Tecnologias de Laboratório Químico e Biológico) como diferentes populações.

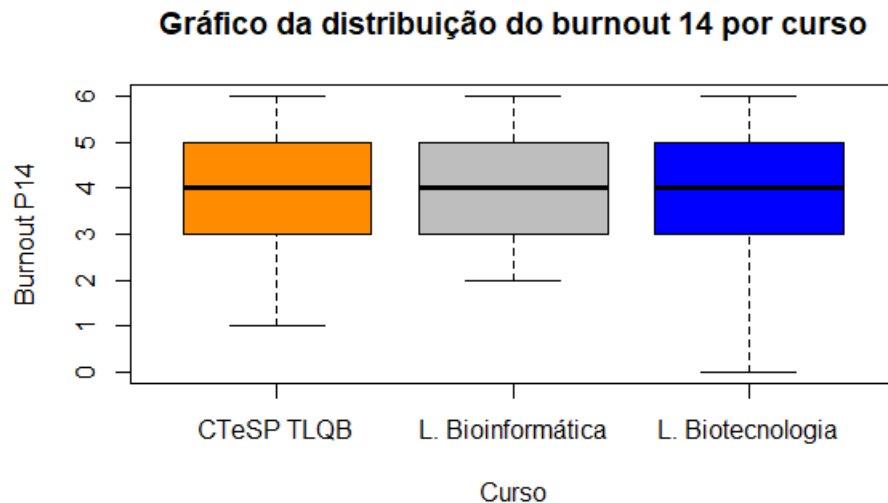


Figura 22: Gráfico da distribuição do burnout 14 por curso.

Conforme o teste de **Kruskal-Wallis** ( $DF = 2$ ,  $p = 0.9645$ ) não existe variação no nível de interesse entre os estudantes de cursos diferentes. Para complementar o teste, foi feito o gráfico da distribuição do burnout 14 por curso representado na **Figura 22** que comprova a falta de variação.

### 3.4 Regressão Linear Múltipla

Pretende-se construir um modelo de regressão múltipla que permita prever o número de horas de estudo a partir do número de horas dedicadas às redes sociais e a assistir a programas de TV. Para este modelo ser válido terá que ter estes pressupostos:

- Relação linear entre variáveis;
- Normalidade dos resíduos;
- Homogeneidade;
- Ausência de Outliers;
- Ausência de Multicolinearidade

Será também feito o teste da **One-Way Anova** para verificar a significância estatística do modelo de regressão construído.



## Relação linear entre variáveis

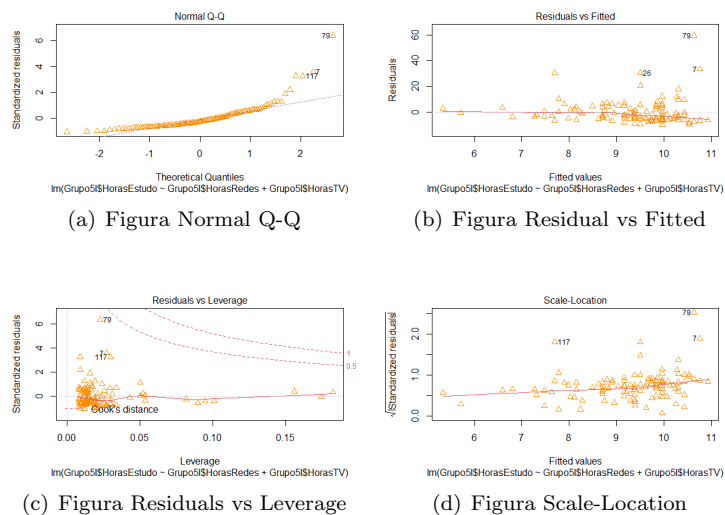


Figura 23: Figuras a, b, c e d

Os gráficos da regressão linear múltipla estão representados na **Figura 23**. Nesta figura existem valores 3 outliers presentes em todas as figuras (a, b, c e d) correspondentes aos estudantes 7, 79 e 117.

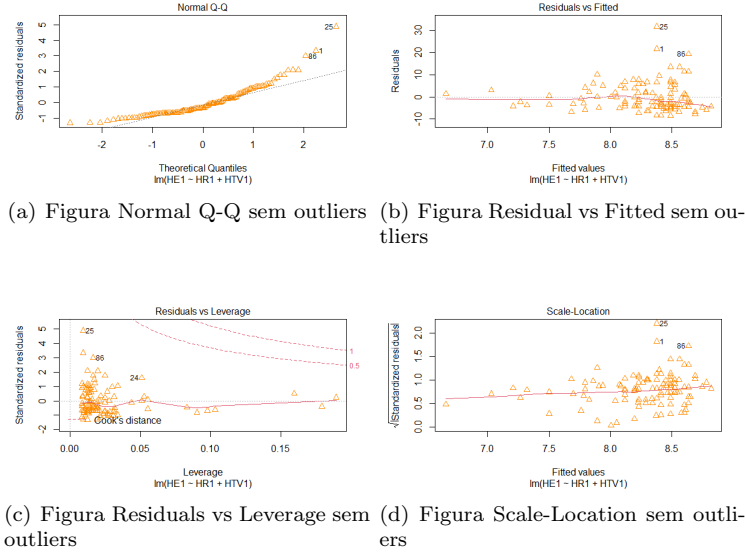


Figura 24: Figuras a, b, c e d

Para obter um modelo de regressão linear múltipla melhor e linear foram removido os 3 estudantes considerados como outliers e criado os gráficos da regressão linear múltipla sem outliers **Figura 24**. Apesar de se remover os outliers anteriores ainda existem outliers no nosso modelo, existem também problemas com a linha do modelo na **Figura 24 a**, sendo que existe a possibilidade que este modelo seja de má qualidade.

#### Normalidade dos resíduos

Para verificar a normalidade dos resíduos foi feito o teste de **Shapiro** ( $W = 0.85445$ ;  $p = 1.675e-09$ ), que demonstra que o modelo não têm uma distribuição normal, logo existe a possibilidade que este modelo não seja bom.

#### Homogeneidade

Para verificar a homogeneidade dos resíduos foi o teste de **Breusch-Pagan** ( $BP = 1.5554$ ;  $DF = 2$ ;  $p = 0.4595$ ), que demonstra que o modelo não é homogêneo, o que significa que o modelo é não homogêneo.

#### Ausência de Outliers

Como mostrado anteriormente na **Relação linear entre variáveis**, mesmo ao remover 3 estudantes ainda existia outliers, comprovado pela distribuição dos valores serem de  $[-1.309347, 4.869080]$  e para não existir outliers os valores devem estar entre  $[-3, 3]$ , é previsto que este modelo não seja usável.

## Ausência de Multicolinearidade

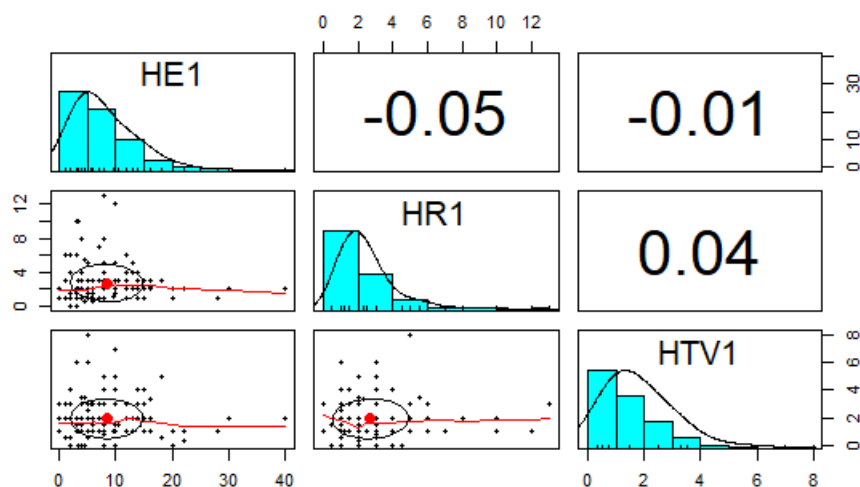


Figura 25: Ausência de multicolinearidade.

O gráfico de ausência de multicolinearidade está representada na **Figura 25**. Para existir ausência de multicolinearidade o  $r$  não poderá ter valores inferiores a 0.9, o que demonstrado pelo gráfico acontece, logo o nosso modelo apresenta multicolinearidade, é previsto que este modelo não seja usável.

### Análise do teste One-Way Anova

Foi feito o modelo  $\text{HorasEstudo} = 8.8935 - 0.1459 \cdot \text{HorasRede} - 0.1122 \cdot \text{HorasTV}$  que foi avaliado pelo teste **One-Way Anova** ( $DF = 117$ ;  $p = 0.8285$ ). Como previsto nos parâmetros anteriores este modelo não possibilita a previsão das horas de sono através das horas de TV e horas de rede, pois não é estatisticamente significante. Este modelo será descartado, pois é impossível de usar.

## 3.5 Análise fatorial

Será feito uma análise fatorial com o objetivo de resumir e demonstrar correlações entre as questões de burnout usando um modelo fatorial. Para este modelo ser válido terá que ser feito cumprir os seguintes passos:

- Verificação de correlação entre as variáveis;
- Verificação de homogeneidade;

- Verificação da existência de um nível de correlação substancial;
- Seleção do número de fatores;
- Criação do modelo fatorial;
- Verificação da sua consistência;
- Criação de índices;

#### Verificação de correlação entre as variáveis

	Br_P1	Br_P2	Br_P3	Br_P4	Br_P5	Br_P6	Br_P7	Br_P8	Br_P9	B_P10	B_P11	B_P12
Burnout_P1	1.00											
Burnout_P2	0.65	1.00										
Burnout_P3	0.45	0.69	1.00									
Burnout_P4	0.54	0.61	0.65	1.00								
Burnout_P5	0.66	0.69	0.61	0.70	1.00							
Burnout_P6	0.44	0.46	0.59	0.61	0.57	1.00						
Burnout_P7	0.35	0.39	0.48	0.47	0.42	0.78	1.00					
Burnout_P8	0.31	0.41	0.51	0.57	0.49	0.62	0.61	1.00				
Burnout_P9	0.29	0.25	0.46	0.57	0.44	0.56	0.57	0.74	1.00			
Burnout_P10	-0.05	-0.03	0.05	0.00	-0.03	0.00	0.04	0.00	-0.02	1.00		
Burnout_P11	0.19	0.09	0.08	0.15	0.24	0.11	0.04	0.14	0.12	0.30	1.00	
Burnout_P12	0.12	0.11	0.04	-0.03	0.13	0.05	-0.09	-0.06	-0.12	0.24	0.29	1.00
Burnout_P13	0.01	0.11	0.06	-0.07	0.07	-0.15	-0.29	-0.03	-0.14	0.12	0.27	0.40
Burnout_P14	-0.02	0.02	-0.10	-0.12	0.03	-0.33	-0.37	-0.33	-0.33	0.12	0.27	0.25
Burnout_P15	-0.29	-0.19	-0.29	-0.37	-0.39	-0.32	-0.18	-0.19	-0.22	0.18	0.07	0.34
	B_P13	B_P14	B_P15									
Burnout_P13	1.00											
Burnout_P14	0.41	1.00										
Burnout_P15	0.27	0.32	1.00									

Figura 26: Correlação das variáveis burnout.

Os valores das correlações das variáveis do burnout estão apresentadas na **Figura 26**. É concluído que existe correlação entre as variáveis, mas terá de ser avaliado a sua relevância.

#### Verificação de homogeneidade

Para verificar a homogeneidade das variáveis foi feita a medida de adequabilidade de Kaiser-Mayer-Olkin, onde foi verificado que todos os valores são superiores a 0.50 e muitos perto de 1, sendo que existe homogeneidade.

#### Verificação da existência de um nível de correlação substancial

Para verificara se o nível da correlação é substancial foi feita o teste de esfericidade de Bartlett. Como o valor obtido do teste foi inferior a 0.5 existe um nível de correlação substancial ( $DF = 105$ ;  $p = 6.278e-159$ ).

#### Seleção do número de fatores

Para escolher o número de fatores foi usado os critérios de Kaiser e um gráfico de linhas apresentado na **Figura 27**.

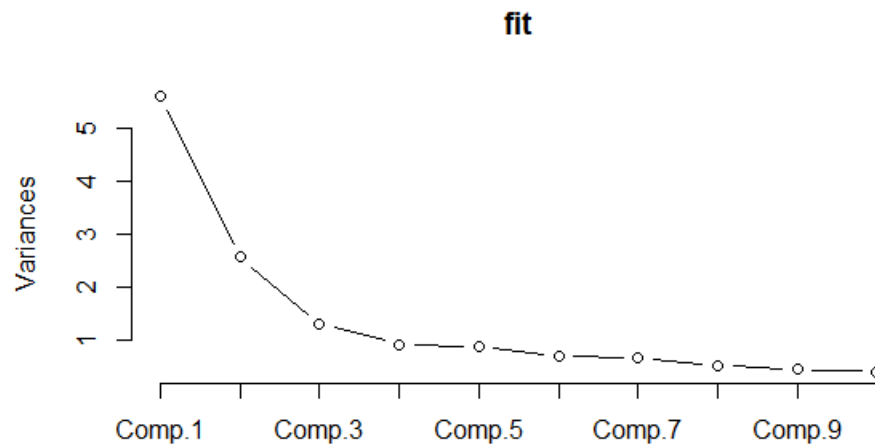


Figura 27: Gráfico de linhas do Fit.

O gráfico recomenda 4 fatores, contudo os critérios de Kaiser recomendam 3 fatores. Por isso foi feita uma comparação entre os valores de comunalidade, apresentada na **Figura 28**.

```
Principal Components Analysis
Call: principal(r = burnout, nfactors = 3, rotate = "none", n.obs = nrow(Grupo5),
  scores = TRUE)
Standardized loadings (pattern matrix) based upon correlation matrix
      PC1  PC2  PC3  h2  u2  com
Burnout_P1  0.66  0.26 -0.35 0.63 0.37 1.9
Burnout_P2  0.72  0.32 -0.35 0.74 0.26 1.9
Burnout_P3  0.78  0.16 -0.10 0.64 0.36 1.1
Burnout_P4  0.83  0.08 -0.12 0.71 0.29 1.1
Burnout_P5  0.80  0.29 -0.27 0.80 0.20 1.5
Burnout_P6  0.83 -0.09  0.21 0.74 0.26 1.2
Burnout_P7  0.74 -0.23  0.36 0.73 0.27 1.7
Burnout_P8  0.76 -0.08  0.34 0.71 0.29 1.4
Burnout_P9  0.71 -0.19  0.36 0.67 0.33 1.6
Burnout_P10 -0.02  0.38  0.54 0.43 0.57 1.8
Burnout_P11 0.16  0.58  0.29 0.45 0.55 1.6
Burnout_P12 -0.03  0.69  0.21 0.52 0.48 1.2
Burnout_P13 -0.12  0.71 -0.03 0.52 0.48 1.1
Burnout_P14 -0.29  0.68 -0.24 0.60 0.40 1.6
Burnout_P15 -0.44  0.38  0.42 0.52 0.48 3.0

      PC1  PC2  PC3
SS loadings      5.54 2.44 1.41
Proportion var    0.37 0.16 0.09
Cumulative var    0.37 0.53 0.63
Proportion Explained 0.59 0.26 0.15
Cumulative Proportion 0.59 0.85 1.00

Mean item complexity = 1.6
Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.07
with the empirical chi square 151.35 with prob < 3.1e-09

Fit based upon off diagonal values = 0.96
```

(a) Gráfico com os valores da comunalidade de 3 fatores

```
Principal Components Analysis
Call: principal(r = burnout, nfactors = 4, rotate = "none", n.obs = nrow(Grupo5),
  scores = TRUE)
Standardized loadings (pattern matrix) based upon correlation matrix
      PC1  PC2  PC3  PC4  h2  u2  com
Burnout_P1  0.66  0.26 -0.35 -0.09 0.64 0.36 1.9
Burnout_P2  0.72  0.32 -0.35  0.17 0.76 0.24 2.0
Burnout_P3  0.78  0.16 -0.10  0.12 0.66 0.34 1.2
Burnout_P4  0.83  0.08 -0.12 -0.09 0.72 0.28 1.1
Burnout_P5  0.80  0.29 -0.27 -0.10 0.81 0.19 1.5
Burnout_P6  0.83 -0.09  0.21  0.08 0.74 0.26 1.2
Burnout_P7  0.74 -0.23  0.36  0.11 0.74 0.26 1.7
Burnout_P8  0.76 -0.08  0.34  0.15 0.73 0.27 1.5
Burnout_P9  0.71 -0.19  0.36  0.04 0.67 0.33 1.6
Burnout_P10 -0.02  0.38  0.54 -0.45 0.64 0.36 2.8
Burnout_P11 0.16  0.58  0.29 -0.50 0.70 0.30 2.6
Burnout_P12 -0.03  0.69  0.21  0.26 0.58 0.42 1.5
Burnout_P13 -0.12  0.71 -0.03  0.27 0.59 0.41 1.3
Burnout_P14 -0.29  0.68 -0.24 -0.07 0.61 0.39 1.7
Burnout_P15 -0.44  0.38  0.42  0.48 0.74 0.26 3.9

      PC1  PC2  PC3  PC4
SS loadings      5.54 2.44 1.41 0.93
Proportion var    0.37 0.16 0.09 0.06
Cumulative var    0.37 0.53 0.63 0.69
Proportion Explained 0.54 0.24 0.14 0.09
Cumulative Proportion 0.54 0.77 0.91 1.00

Mean item complexity = 1.8
Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is 0.07
with the empirical chi square 136.37 with prob < 1e-09

Fit based upon off diagonal values = 0.96
```

(b) Gráfico com os valores da comunalidade de 4 fatores

Figura 28: Figuras a, b

Como confirmado pela figura é escolhido 4 fatores, pois existem valores inferior a 0.60 com 3 fatores na coluna "h2".

### Criação do modelo fatorial

Para fazer o modelo fatorial foi feita a representação gráfica dos 4 fatores em varimax que esta apresentado na **Figura 29**.

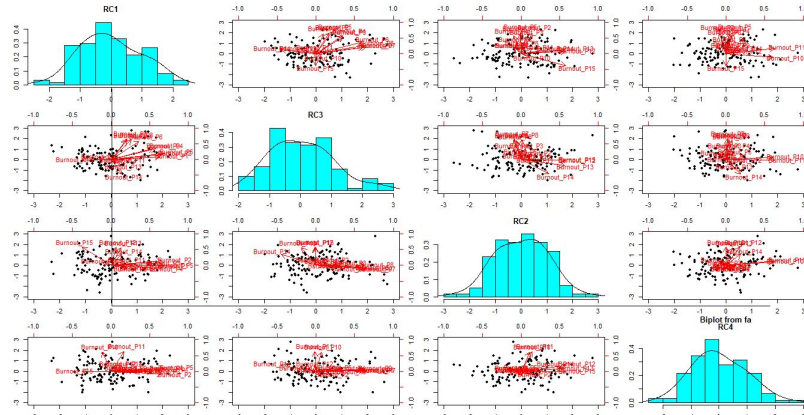


Figura 29: Representação gráfica de 4 fatores varimax.

Como existe uma certa organização foi criado o modelo fatorial que está apresentado na **Figura 30**.

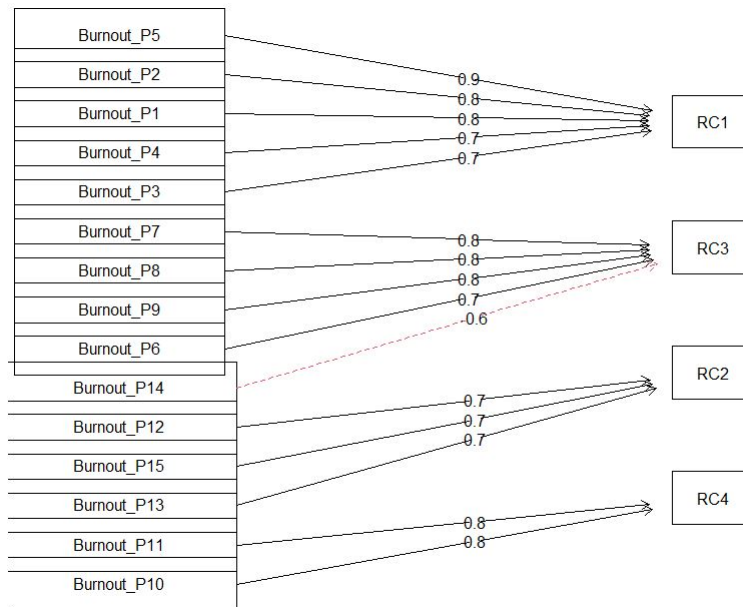


Figura 30: Modelo fatorial.

#### Verificação da sua consistência

Foi verificado o modelo fatorial, sendo que geralmente existe uma boa consistência estando por volta dos valores 0.7 e 0.9, com o resto estando na zona razoável.

#### Criação de índices

O modelo fatorial foi bem sucedido permitindo resumir informação nos seus fatores:

- O fator 1 representa a exaustão dos estudantes relativos aos seus estudos e aulas;
- O fator 2 representa a auto estimas dos estudantes relativos ao seu sucesso académico;
- O fator 3 representa a falta de entusiasmos e motivação dos estudantes perante o seu curso;
- O fator 4 representa a autonomia e participação do estudantes perante as aulas;

## 4 Conclusões

Resumindo, nas principais características da análise univariada da nossa amostra vemos que:



- A média das idades dos estudantes é de aproximadamente de 21 anos.
- O curso mais frequentado é a Licenciatura em Biotecnologia.
- O 1º ano é ano curricular com mais alunos.
- 99% dos estudantes escolheram o curso por vontade própria.
- A média das horas de estudo (semanais) e horas de sono dos estudantes são de 9 e 7 horas respetivamente. (Valores aproximados)

Em relação à análise bivariada concluímos que:

- A associação entre a exaustão emocional e o sexo dos estudantes é baixa.
- Existe uma correlação muito fraca entre o entusiasmo dos estudantes em relação às matérias e ano curricular que estes frequentam.
- A correlação entre as horas de estudo e as horas de sono, tem pouca significância.

Na inferência estatística em relação à primeira questão de investigação chega-se à conclusão que os estudantes do sexo feminino não se sentem mais exaustos (emocionalmente) que os estudantes do sexo masculino.

Na segunda questão viu-se que os alunos não aumentam necessariamente a sua auto-estima em relação ao sucesso escolar ao longo do seu percurso académico. Por fim na terceira questão observou-se que não existe propriamente um maior interesse do curso pelos alunos de bioinformática em relação aos alunos dos outros cursos.

A regressão linear obtida para prever as horas de um estudante através das horas que este passa nas redes sociais e na televisão não serve de modelo, não tem significância estatística.

Foi possível criar um modelo fatorial que resume as variáveis do burnout em 4 fatores diferentes. O primeiro representa a verificação de correlação entre as variáveis, o segundo a verificação de homogeneidade, o terceiro a verificação da existência de um nível de correlação substancial e o quarto a eleição do número de fatores.

Pode-se então concluir que em termos estatísticos o projeto não leva a verificar algum padrão ou tendência relativamente ao nível emocional dos alunos nos seus estudos, no entanto o projeto foi extremamente importante para a consolidar de uma forma integral os conhecimentos necessários para esta disciplina. Além disso foi possível verificar de uma forma prática a utilidade da análise e tratamento de dados multivariados.