



ESCOLA SUPERIOR DE  
TECNOLOGIA DA SAÚDE  
DE LISBOA

Pós-Graduação em  
*Health Data Science*



Health  
Data Science

Atividade Prática

## **Explorar relações e padrões em dados cirúrgicos de ortopedia: Uma Abordagem Multivariada**

Análise Fatorial e Clusterização revelam fatores-chave para o  
sucesso e a satisfação na Artroplastia do Joelho

**Unidade Curricular: Estatística Multivariada**

**Docente:** Professora Ana Marreiros

**Realizado por:**

Ana Ripado, n.º 2023136

Raquel Quintão n.º 2023134

Rita Varandas n.º 2023112

Paulo Nunes n.º 2023129

*Lisboa, Maio de 2024*

## Respostas ao enunciado

- a) Considerando a sua base de dados, deverá apresentar uma descrição (explicar o que medem, representam, etc.) sucinta das suas variáveis de estudo mencionado a dimensão original da Base (1 val.).

### Variáveis de Identificação e Sociodemográficas:

- **Idade:** representa a idade dos utentes submetidos à cirurgia de prótese total do joelho, expressa em anos.
- **Sexo:** Indica o género, onde o valor 1 representa o género feminino e o valor 2 representa o masculino.
- **Peso:** refere-se ao peso dos utentes, medido em quilogramas (Kg).
- **Altura:** indica a altura, expressa em centímetros (cm).
- **IMC – Índice de Massa Corporal ( $\text{peso}/\text{altura}^2$ ):** evidencia a relação entre o peso e a altura dos utentes; quanto maior o IMC, maior a probabilidade de sobrepeso ou obesidade.

### Variáveis de Classificação do Alinhamento do Joelho:

- **Grupo pré e Grupos pós:** classificação do alinhamento do joelho antes e após a cirurgia, respetivamente, de I a IX com base no Coronal Plane Alignment of the Knee (CPAK). Por exemplo, um valor de 5 indica que o joelho foi classificado como CPAK Tipo V.

**Variáveis de Avaliação Pré-Cirúrgica:** parâmetros analíticos antes da intervenção cirúrgica.

- **Fle\_0:** ângulo máximo de flexão do joelho antes da cirurgia; o ângulo varia, podendo situar-se entre os 120 e os 140 graus.
- **EVA\_0:** representa a intensidade da dor percebida pelos utentes antes da cirurgia, medida através da Escala Visual Analógica (EVA).
  - A Escala Visual consiste numa linha de 10 centímetros onde "Sem Dor" está numa extremidade e "Dor Máxima/Pior Dor Possível" na outra. O utente marca a intensidade da dor e mede-se a distância entre essa marcação e o início da linha para obter a classificação numérica da dor.
- **PM6\_0:** distância percorrida durante a prova de marcha de 6 minutos antes da cirurgia, medida em metros (m).

- **WD\_0, WR\_0, WA\_0:** Representam a avaliação da dor, rigidez e atividades, respetivamente, antes da cirurgia, utilizando o questionário WOMAC.
- **WT\_0:** Pontuação total do questionário WOMAC antes da cirurgia. O questionário WOMAC é uma ferramenta utilizada na avaliação da osteoartrite do quadril e do joelho, composto por 24 itens divididos em 3 subescalas: dor, rigidez e função física. Pontuações mais elevadas indicam maior dor, rigidez e limitações funcionais.

#### **Avaliação após 90 dias: parâmetros analíticos pós-cirurgia**

- **Fle\_90:** indica flexão máxima do joelho após 90 dias da cirurgia.
- **EVA\_90:** refere-se à intensidade da dor percebida pelos pacientes após 90 dias da cirurgia, medida pela Escala Visual Analógica (EVA).
- **PM6\_90:** distância percorrida durante a prova de marcha de 6 minutos após 90 dias da cirurgia, em metros.
- **WD\_90, WR\_90, WA\_90:** Avaliação da dor, rigidez e atividades, respetivamente, após 90 dias da cirurgia, utilizando o questionário WOMAC.
- **WT\_90:** Pontuação total do questionário WOMAC após 90 dias da cirurgia.

#### **Variável Satisfação do utente após a intervenção cirúrgica:**

- **Satisfação:** escala de Likert permitiu avaliar o grau de satisfação dos utentes após a cirurgia da prótese total do joelho.
  - Os utentes indicam o seu nível de concordância com uma série de afirmações que podem variar de "discordo totalmente" a "concordo totalmente".
  - Foi importante notar que os valores numéricos já estavam atribuídos à escala ordinal na base de dados, permitindo assim o cálculo de médias e outras análises estatísticas para uma compreensão mais aprofundada dos dados da satisfação após a cirurgia.

#### **Dimensões da Matriz e Problema Estatístico**

- **Dimensão da matriz (álgebra):** A matriz original da base de dados, relativa à especialidade cirúrgica de ortopedia, é composta por 145 indivíduos e 22 variáveis, resultando numa dimensão de 145 x 22.
- **Dimensão computacional da matriz:** Ao multiplicar o número de indivíduos pelo número de variáveis, obtemos a dimensão computacional da matriz, que é de 3190 (145 indivíduos x 22 variáveis).

- **Dimensão geral do problema estatístico:** Antes da redução da dimensionalidade do problema, ele é representado por 22 variáveis. Cada indivíduo é representado por um vetor de dimensão 22, onde os valores das variáveis são apresentados no seguinte formato: ind1(67, 80, 158, 32.05, 5, 5, 86, 4, 324, 15, 6, 54, 75, 80, 1, 290, 5, 1, 9, 15, 8).
- **Para o processo de análise fatorial de componentes principais (AFCP):** A dimensão geral do problema estatístico é representada por X variáveis. Cada indivíduo será representado por um vetor de dimensão 16, onde são representadas as seguintes variáveis: ind1 (Idade, Sexo, Estatura, IMC, Grupo\_pré, Grupo\_pós, F\_0, PM\_0, EVA\_0, PM\_6, WOMAC\_0, F\_90, PM6\_90, WR\_90, WOMAC90 (Dor, Atividade e Total), Satisfação).

b) Pronuncie-se sobre a natureza (ou classificação) das variáveis que vai utilizar. Exemplo: género é uma variável qualitativa de escala nominal. (1 val.)

Classificação das Variáveis	
<b><i>Idade</i></b>	Quantitativa discreta na escala de razão
<b><i>Sexo</i></b>	Qualitativa nominal
<b><i>Peso</i></b>	Quantitativa contínua na escala de razão
<b><i>Altura</i></b>	Quantitativa contínua na escala de razão
<b><i>IMC</i></b>	Quantitativa contínua na escala de razão
<b><i>Grupo pré</i></b>	Qualitativa nominal
<b><i>Grupo pós</i></b>	Qualitativa nominal
<b><i>Fle_0 e Fle_90</i></b>	Quantitativa contínua na escala de intervalo
<b><i>EVA_0 e EVA_90</i></b>	Qualitativa ordinal
<b><i>PM6</i></b>	Quantitativa contínua na escala de razão
<b><i>WD_0 e WD_90</i></b>	Quantitativa discreta na escala de intervalo
<b><i>WR_0 e WR_90</i></b>	Quantitativa discreta na escala de intervalo
<b><i>WA_0 e WA_90</i></b>	Quantitativa discreta na escala de intervalo
<b><i>WT_0 e WT_90</i></b>	Quantitativa discreta na escala de intervalo
<b><i>Satisfação</i></b>	Qualitativa ordinal

c) Efetue uma análise descritiva simples, justificando as opções tomadas. Deverão constar todos os *outputs* em anexo e/ou ficheiro (1 val.)

### Variáveis quantitativas:

Nas variáveis quantitativas dividimos as variáveis demográficas das variáveis com valor clínico que se pretende comparar em dois momentos distintos, pré e pós cirúrgico, respetivamente momento **0** e **90**.

Para estas variáveis quantitativas é importante considerar como os dados se distribuem a partir das seguintes estatísticas descritivas para melhor compreendermos a distribuição da amostra de utentes submetidos a cirurgia do joelho.

Estatísticas descritivas									
Variaveis	Min	1º Q	Median	Mean	3º Q	Max	Sd	VC	Mode
Idade	48	65	70	70.48	76	86	7.6	10.79	70
Peso	49	67	77	77.66	85	135	13.94	17.95	80
Altura	140	155	162	162.6	170	198	9.76	6	155
IMC	19.14	26.12	28.73	29.35	32.39	42.32	4.45	15.15	23.44

Tabela1 - Estatísticas descritivas das variáveis demográficas

Estatísticas descritivas																		
Variaveis	Min		1º Q		Median		Mean		3º Q		Max		Sd		VC		Mode	
	0	90	0	90	0	90	0	90	0	90	0	90	0	90	0	90	0	90
F	46	70	90	95	105	100	103.60	104.5	120	115	140	153	19.80	13.50	19.11	12.92	100	90
PM6	0	121	188	267	256	320	259.10	317	328	370	522	543	103.06	82.40	39.78	25.99	240	360
WD	0	0	8	1	10	2	10.75	2.56	14	4	20	11	3.73	2.50	34.67	97.97	9	0
WR	0	0	2	0	3	1	2.90	1.27	4	2	8	4	2.09	1.20	71.96	94.83	2	0
WA	8	0	26	5	33	9	35.20	10.26	44	14	68	34	12.75	7.38	36.23	71.87	27	2
WT	8	1	37	7	47	12	48.86	14.08	59	19	96	46	16.78	9.66	34.34	68.59	47	12

Tabela 2 - Estatísticas descritivas de variáveis a comparar momento 0 (Pré-cirurgia) e momento 90 (Pós-cirurgia)

## Análise descritiva

### Variáveis quantitativas:

Na análise descritiva das variáveis quantitativas, optamos na representação gráfica por utilizar histogramas e boxplots devido à capacidade de fornecer uma representação visual clara da distribuição dos dados, incluindo tendência central, dispersão e presença de outliers. Os histogramas são eficazes para identificar a forma da distribuição e a densidade dos dados ao longo da escala de valores. Por sua vez, os boxplots destacam de forma sucinta as principais características da distribuição, como quartis, mediana e variação interquartil.

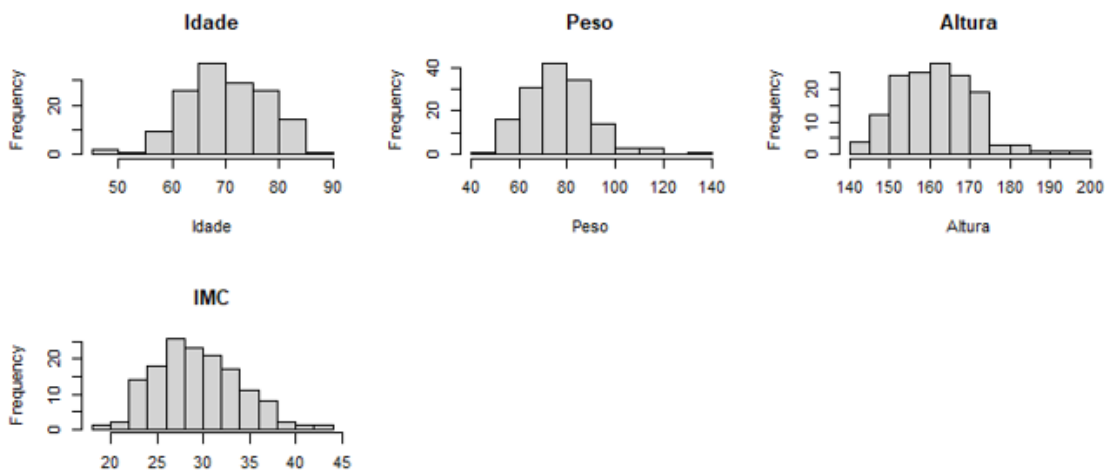


Gráfico 1 - Histograma variáveis demográficas

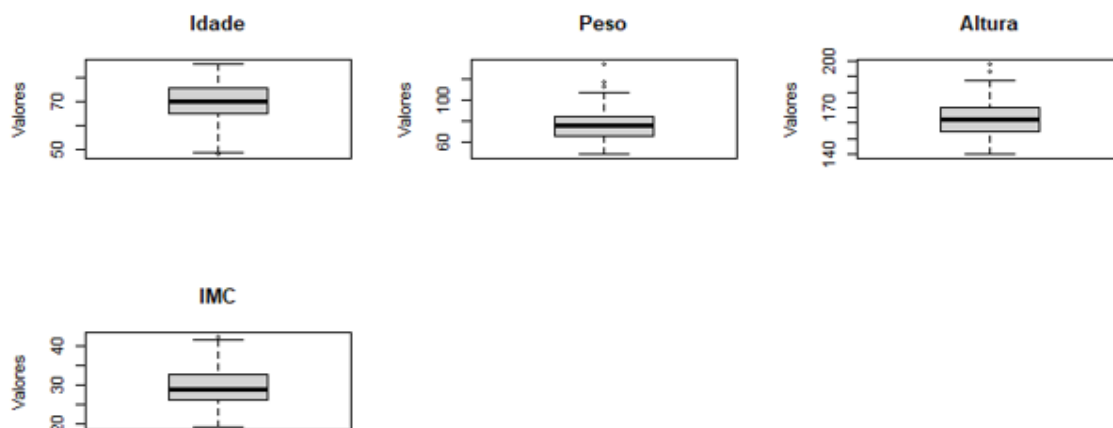


Gráfico 2 - Boxplot variáveis demográficas

Na análise das variáveis qualitativas, os gráficos de barras são mais adequados devido à sua capacidade de representar de forma simples e clara a distribuição de frequência das variáveis categorias. Desta forma permite comparar as proporções entre diferentes categorias, permitindo uma rápida visualização das diferenças na distribuição das variáveis qualitativas.

Analisando primeiramente as variáveis de dados demográficos que nos permitem conhecer as características dos utentes do estudo, observamos que:

Os utentes apresentam uma idade compreendida entre os 48 e os 86 anos, observa-se uma distribuição maioritariamente em idades dos 60 aos 80 anos.

O peso de 75% dos utentes submetidos, varia até 85kg. O utente com maior peso apresenta 135kg, mas representa-se como um dos outliers da distribuição dos valores desta variável. Existem 3 valores outliers, destacando a natureza incomum em relação à restante distribuição.

Observamos na altura dos utentes que a média dos mesmos é de 162 cm, com um desvio padrão que varia em torno de  $\pm 9.76$  cm deste valor. Além disso, o coeficiente de variação de 6% mostra a dispersão relativa das alturas em relação à média, o que sugere uma dispersão mais consistente das alturas em relação à média representando alguma homogeneidade nesta característica.

No IMC a mediana (28.73) e a média (29.35) estão próximas, sugerindo uma distribuição simétrica dos valores do IMC, que verificamos graficamente. Tem-se uma dispersão deste valor, relativa, de 15.15%.

#### Variáveis quantitativas que se pretende comparar:

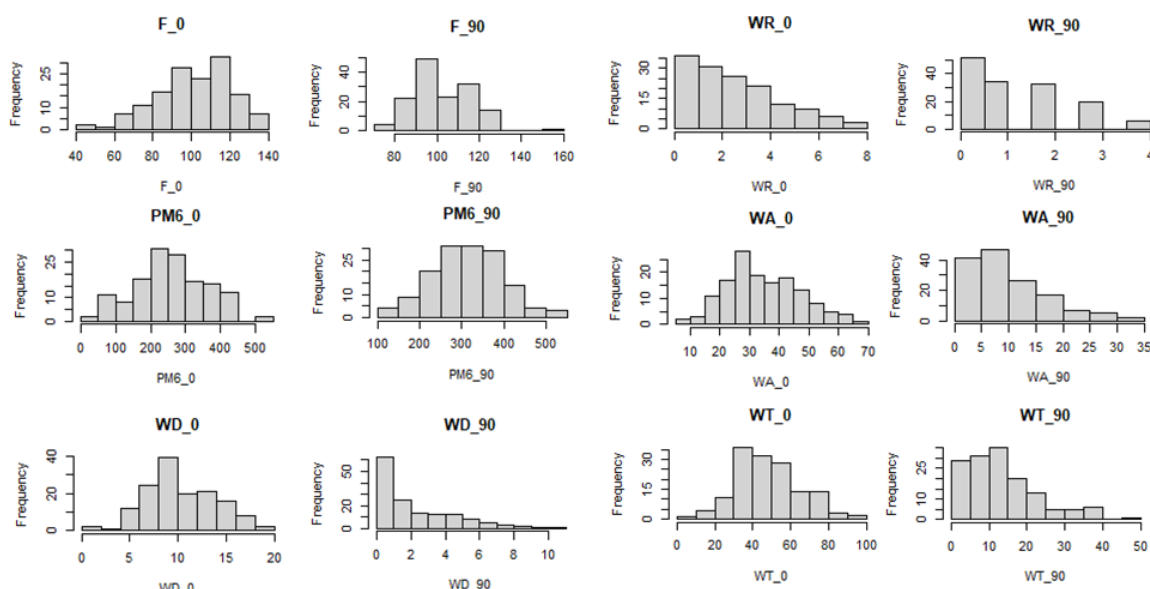


Gráfico 1 - Histogramas de variáveis a comparar momento 0 (Pré-cirurgia) e momento 90 (Pós-cirurgia)



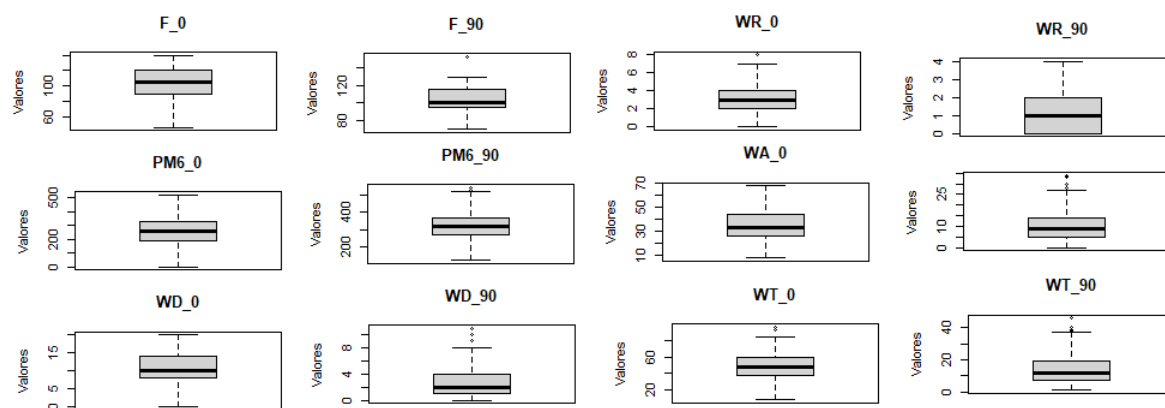


Gráfico 2 - Boxplot de variáveis a comparar momento 0 (Pré-cirurgia) e momento 90 (Pós-cirurgia)

### Variáveis qualitativas:

Nas variáveis qualitativas, como não há uma ordem específica entre as categorias, a medida de tendência central mais relevante é a moda, que representa a categoria mais frequente nos dados. As tabelas de frequências são apresentadas para cada variável categórica para fornecer uma visão clara da distribuição dos utentes nas diferentes categorias, permitindo a identificação das mais prevalentes na amostra.

### Variável qualitativa demográfica:

Moda – “sexo”: 1

Sexo	ni	fi (%)
Feminino (1)	94	64.83
Masculino (2)	51	35.17
Total	145	100

Tabela 3 - Tabela frequências relativa e absoluta - Sexo

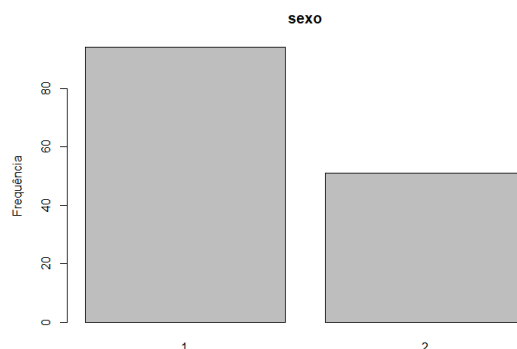


Gráfico 3 - Distribuição Sexo: Feminino - 1; Masculino - 2

A amostra é representada por uma maioria de utentes do sexo feminino, com 64.83%, para os restantes, do sexo masculino, 35.17%.

## Variáveis qualitativas a comparar:

### Grupo – Pré-cirúrgico e Pós-cirúrgico:

Moda – Grupo pré: 1

Moda – Grupo pós: 5

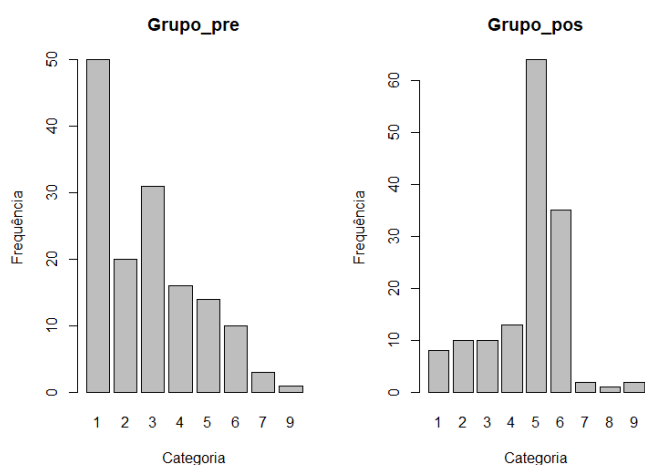


Tabela 4 - Tabela frequências Grupo-pré

Grupo-pós	ni	fi (%)
CPAK		
I	8	5.52
II	10	6.90
III	10	6.90
IV	13	8.97
V	64	44.14
VI	35	24.14
VII	2	1.38
VIII	1	0.69
IX	2	1.38
Total	145	100

Tabela 5 - Tabela frequências Grupo-pós

Grupo-pré	ni	fi (%)
CPAK		
I	50	34.48
II	20	13.79
III	31	21.38
IV	16	11.03
V	14	9.66
VI	10	6.90
VII	3	2.07
VIII	0	0.00
IX	1	0.69
Total	145	100

Gráfico 5 - Distribuição categoria CPAK no grupo, momento pré e pós.

Analisando estes dois momentos do grupo (pré e pós), é de destaque que no primeiro momento existe uma clara representação da categoria 1 (CPAK I), seguindo-se a categoria 3 (CPAK III). Verifica-se pouca representação dos grupos 5, 6, 7 e 9 (CPAK V, VI, VII e IX). Por fim, de notar que neste momento pré não se observou nenhum paciente que pertença à categoria 8 (CPAK VIII).

No segundo momento, destaca-se francamente uma maior frequência de utentes nas categorias 5 e 6 (CPAK V e VI), com uma representação de 44.14% e 24.14% respetivamente.

### Escala Visual Analógica:

Moda - EVA\_0: 8 e 5

Moda - EVA\_90: 0

EVA_0	ni	fi (%)
0	10	6.90
1	0	0.00
2	10	6.90
3	14	9.66
4	16	11.03
5	27	18.62
6	15	10.34
7	16	11.03
8	29	20.00
9	4	2.76
10	4	2.76
Total	145	100

EVA_90	ni	fi (%)
0	68	46.90
1	18	12.41
2	24	16.55
3	14	9.66
4	7	4.83
5	7	4.83
6	3	2.07
7	2	1.38
8	2	1.38
9	0	0
10	0	0
Total	145	100

Tabela 6 - Tabela frequências Dor momento pré e pós cirúrgico (0 e 90) - Escala Visual Analógica

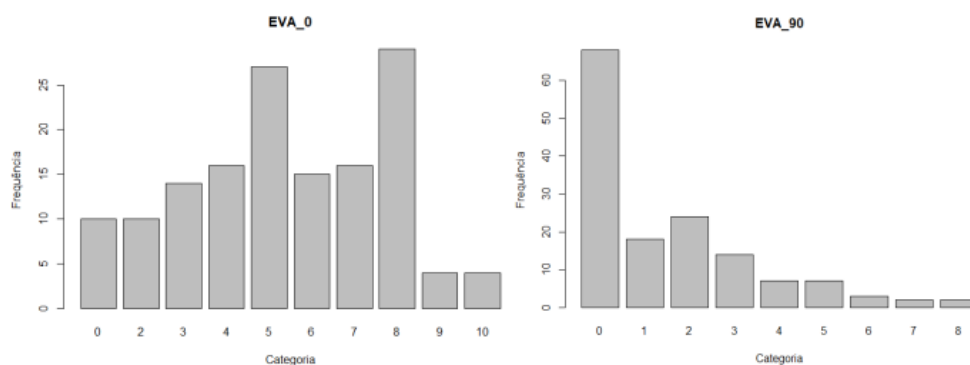


Tabela 7 - Distribuição EVA no grupo, momento pré e pós.

Através desta análise podemos destacar que no primeiro momento, a distribuição apresenta-se como bimodal assimétrica, com uma maior frequência da dor de categoria 8 e 5, representando 20% e 18.62% respetivamente.

Em contrapartida, no segundo momento verifica-se uma distribuição é assimétrica positiva com uma maior concentração dos dados à esquerda, representando níveis de dor mais baixos.

### Satisfação:

Moda - Satisfação: 10

Satisfação	ni	fi (%)
1	1	0.69
2	0	0
3	2	1.38
4	1	0.69
5	2	1.38
6	4	2.76
7	13	8.97
8	26	17.93
9	27	18.62
10	69	47.59
Total	145	100

Tabela 9 - Tabela frequências Satisfação

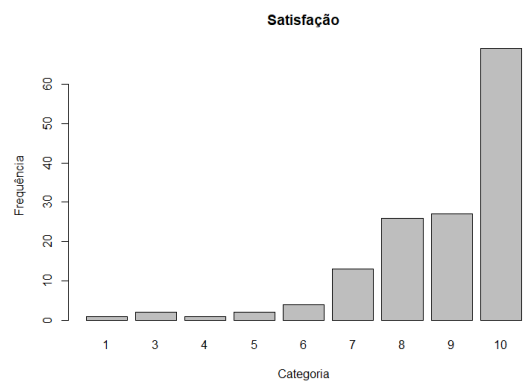


Tabela 10 – Gráfico de frequências absolutas - Satisfação

Através da variável satisfação, apenas presente no momento pós-cirúrgico, verifica-se uma distribuição assimétrica negativa, representando níveis de satisfação maioritariamente acima de 6. A moda apresenta-se distintamente no valor mais alto, 10.

**d) Execute computacionalmente em R uma AFCP + Análise de Clusters Hierárquicos com as coordenadas da factorial já no espaço de dimensão reduzida. Todos os objectos criados para assegurar os outputs deverão ser guardados e enviados em ficheiros R (10+2 val.).**

1º passo - Selecionar grupos de variáveis

2º passo - Normalizar os dados dos grupos de variáveis

3º Passo – Análise Fatorial dos Componentes Principais - Grupo variáveis pré-cirurgia

- PCA
- Analisar eigen value
- Percentagem de variância das dimensões

4º Passo - Validar a qualidade da AFCP - Grupo variáveis pré-cirurgia

- Determinante da matriz de correlações
- Teste KMO (Kaiser Meier Olkin)
- Bartlett Test

5º Passo – Análise dos Gráfico AFCP - Grupo variáveis pré-cirurgia

6º Passo – Análise Hierárquica de clusters para o espaço de dimensão reduzida – Grupo variáveis pré-cirurgia

- Dendrograma

7º Passo - Caracterização dos grupos de clusters - Grupo variáveis pré-cirurgia

#####REPLICAR PROCESSO PARA VARIÁVEIS POS CIRURGIA #####

8º Passo – Análise Fatorial dos Componentes Principais - Grupo variáveis pós-cirurgia

- PCA
- Analisar eigen value
- Percentagem de variância das dimensões

9º Passo - Validar a qualidade da AFCP - Grupo variáveis pós-cirurgia

- Determinante da matriz de correlações
- Teste KMO (Kaiser Meier Olkin)
- Bartlett Test

10º Passo – Análise dos Gráfico AFCP - Grupo variáveis pós-cirurgia

11º Passo – Análise Hierárquica de clusters para o espaço de dimensão reduzida – Grupo variáveis pós-cirurgia

- Dendrograma

12º Passo - Caracterização dos grupos de clusters - Grupo variáveis pós-cirurgia

**e) Fundamente a sua aplicação prática contextualizando os seus outputs (1 val.)**

**1º Passo – Selecionar as variáveis pré-cirurgia e pós-cirurgia**

De forma a garantir uma análise detalhada e comparativa entre as variáveis, decidimos realizar o trabalho desenvolvendo 2 AFCP entre as características pré-cirurgia e pós-cirurgia para compreender como estas variáveis se correlacionam entre si.

Consideramos útil agregar também os dados relativamente à idade, peso, altura e IMC em ambas as AFCP, uma vez que, poderíamos obter informações importante relativamente às correlações entre as variáveis.

Exclui-se as variáveis Sexo, Grupo-pré, Grupo-pós, EVA\_0, EVA\_90 e Satisfação por serem variáveis qualitativas.

**2º Passo – Normalizar as variáveis qualitativas**

Uma vez que as variáveis apresentam escalas de avaliação diferentes e desvios-padrões consideráveis entre si, considera-se pertinente realizar a normalização das variáveis analisadas.

**3ª Passo – Realizar e analisar a análise fatorial de componentes principais (AFCP) – Grupo Variáveis Pré-Cirurgia**

Com o objetivo de reduzir a dimensionalidade do problema estatístico, aplicou-se a AFCP e considerou-se a seguinte análise:

\$coord	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Idade	-0.009678863	-0.28572157	-0.5045974	0.627306947	0.365536327
Peso	0.008137307	0.94074782	-0.3161314	0.079856136	0.038722374
Altura	-0.237832891	0.63975849	0.3882410	0.564458088	-0.009362671
IMC	0.207248010	0.58498042	-0.6840924	-0.355576456	0.052264801
F_0	-0.456712558	0.05410079	0.2676670	-0.294512025	0.761075007
PM6_0	-0.529274869	0.39219630	0.5248478	-0.079611749	0.027749238
WD_0	0.868446959	0.04746456	0.1683409	-0.002132467	0.116465559
WR_0	0.547806729	0.17773226	0.3146994	-0.029169932	-0.211540283
WA_0	0.913380361	0.05061327	0.1464269	0.035664039	0.224229680
WT_0	0.955390508	0.07114847	0.1878858	0.022999771	0.169957607
\$cor	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Idade	-0.009678863	-0.28572157	-0.5045974	0.627306947	0.365536327
Peso	0.008137307	0.94074782	-0.3161314	0.079856136	0.038722374
Altura	-0.237832891	0.63975849	0.3882410	0.564458088	-0.009362671
IMC	0.207248010	0.58498042	-0.6840924	-0.355576456	0.052264801
F_0	-0.456712558	0.05410079	0.2676670	-0.294512025	0.761075007
PM6_0	-0.529274869	0.39219630	0.5248478	-0.079611749	0.027749238
WD_0	0.868446959	0.04746456	0.1683409	-0.002132467	0.116465559
WR_0	0.547806729	0.17773226	0.3146994	-0.029169932	-0.211540283
WA_0	0.913380361	0.05061327	0.1464269	0.035664039	0.224229680
WT_0	0.955390508	0.07114847	0.1878858	0.022999771	0.169957607

\$cos2					
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Idade	9.368038e-05	0.081636813	0.25461852	3.935140e-01	1.336168e-01
Peso	6.621576e-05	0.885006455	0.09993906	6.377002e-03	1.499422e-03
Altura	5.656448e-02	0.409290920	0.15073106	3.186129e-01	8.765961e-05
IMC	4.295174e-02	0.342202091	0.46798239	1.264346e-01	2.731609e-03
F_0	2.085864e-01	0.002926896	0.07164561	8.673733e-02	5.792352e-01
PM6_0	2.801319e-01	0.153817937	0.27546521	6.338031e-03	7.700202e-04
WD_0	7.542001e-01	0.002252884	0.02833866	4.547414e-06	1.356423e-02
WR_0	3.000922e-01	0.031588757	0.09903574	8.508849e-04	4.474929e-02
WA_0	8.342637e-01	0.002561703	0.02144085	1.271924e-03	5.027895e-02
WT_0	9.127710e-01	0.005062105	0.03530107	5.289895e-04	2.888559e-02

\$contrib					
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Idade	0.002763660	4.2600235	16.923817	4.183336e+01	15.62004667
Peso	0.001953428	46.1819628	6.642684	6.779211e-01	0.17528518
Altura	1.668705984	21.3578759	10.018693	3.387084e+01	0.01024757
IMC	1.267117041	17.8570044	31.105547	1.344091e+01	0.31933009
F_0	6.153495689	0.1527331	4.762093	9.220801e+00	67.71364009
PM6_0	8.264156647	8.0266242	18.309441	6.737781e-01	0.09001676
WD_0	22.249619678	0.1175614	1.883595	4.834228e-04	1.58568263
WR_0	8.853005193	1.6483844	6.582643	9.045517e-02	5.23127320
WA_0	24.611570840	0.1336764	1.425116	1.352146e-01	5.87770025
WT_0	26.927611839	0.2641539	2.346368	5.623538e-02	3.37677757

Tabela 11 – Output PCA Grupo\_Pre

Ao analisar os Eigenvalues é possível decidir quantas dimensões são possíveis reter, com base na quantidade de variância dos dados extraídos da AFCP.

Considerando a necessidade de eigen-value >1 , retêm-se no espaço de dimensão reduzida 3 dimensões com uma percentagem de variância de 68,1%.

#comp 1	3.389721	33.89721%
#comp 2	1.916347	53.06068%
#comp 3	1.504498	68.10566%
#com4	9.406703 <sup>-01</sup>	77.51236%
#com5	8.554187 <sup>-01</sup>	86.00655%

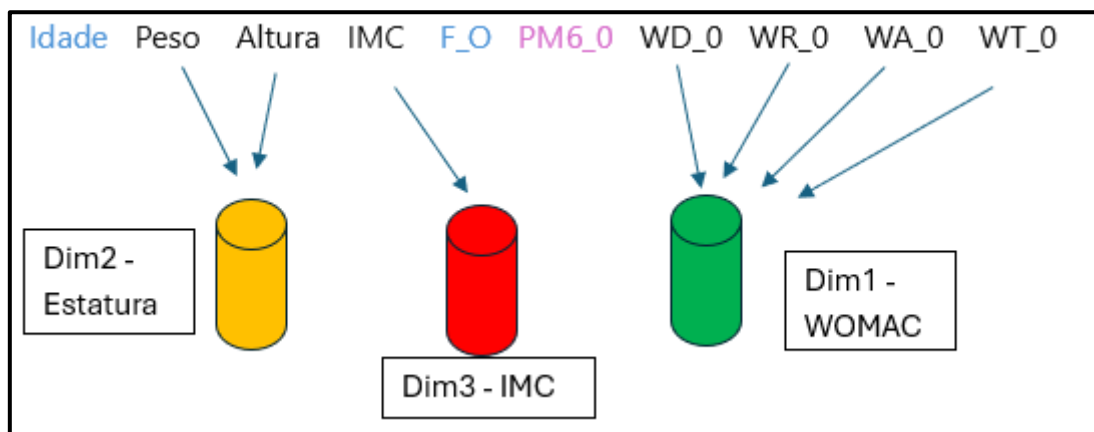


Figura 1 – Distribuição das variáveis pré-cirúrgicas pelas dimensões

Com base no output da realização da PCA é possível iniciar a interpretação dos componentes principais.

Considerando os resultados, é possível analisar que as variáveis WD\_0, WR\_0, WA\_0 e WT\_0 correlacionam-se e contribuem para a dimensão 1. Com isto, a dimensão 1 renomeia-se WOMAC.

Já as variáveis Peso e Altura agrupam-se na dimensão 2, renomeando esta dimensão para estatura do utente.

O IMC encontra-se correlacionado com a dimensão 3, sendo esta a única variável com alto valor de correlação com nesta dimensão.

A variável Idade apresenta um maior valor de correlação com a dimensão 4, que não é representativa da nossa AFCP e não se integra nas dimensões seleccionadas. O mesmo acontece com a variável F\_0 que possui o maior valor de correlação representado na dimensão 4.

A variável PM6\_0 demonstra valores de correlação bastante próximos entre a dimensão 1 e a dimensão 3 sendo considerada uma “variável camaleão”, ou seja, pode manifestar-se de formas diferentes consoante a análise. Com isto, também não relacionamos a variável PM6\_0 na análise AFCP.

#### 4ª Passo - Validar a qualidade da AFCP - Variáveis pré-cirurgia

##### 1. Avaliar o determinante da matriz de correlações das variáveis originais

A matriz de correlações entre as variáveis observadas fornece informações importantes sobre como as variáveis estão relacionadas umas com as outras. A análise deste output permite identificar padrões de associação entre as variáveis observadas, ou seja, alta correlação entre duas variáveis pode sugerir que elas estão medindo conceitos semelhantes e podem ser combinadas em um único fator.

	Idade	Peso	Altura	IMC	F_0
Idade	1.00000000	-0.065774469	-0.114233993	0.01700260	-0.05444092
Peso	-0.06577447	1.000000000	0.536698324	0.73168575	-0.02237473
Altura	-0.11423399	0.536698324	1.000000000	-0.17214492	0.09234052
IMC	0.01700260	0.731685754	-0.172144918	1.00000000	-0.10980527
F_0	-0.05444092	-0.022374734	0.092340524	-0.10980527	1.00000000
PM6_0	-0.26731936	0.157414681	0.414091700	-0.14988744	0.36449823
WD_0	-0.09938444	0.005156518	-0.092141250	0.09118575	-0.28178381
WR_0	-0.10808013	0.048918129	-0.005895889	0.05262153	-0.20695421
WA_0	-0.02449660	0.018126910	-0.105700843	0.11309336	-0.25104757
WT_0	-0.05416041	0.021015048	-0.101544357	0.11276972	-0.27918973
	PM6_0	WD_0	WR_0	WA_0	WT_0
Idade	-0.26731936	-0.099384438	-0.108080130	-0.02449660	-0.05416041
Peso	0.15741468	0.005156518	0.048918129	0.01812691	0.02101505
Altura	0.41409170	-0.092141250	-0.005895889	-0.10570084	-0.10154436
IMC	-0.14988744	0.091185751	0.052621533	0.11309336	0.11276972
F_0	0.36449823	-0.281783813	-0.206954209	-0.25104757	-0.27918973
PM6_0	1.00000000	-0.344869621	-0.013229709	-0.36055563	-0.35231207
WD_0	-0.34486962	1.000000000	0.375827283	0.76817096	0.85283671
WR_0	-0.01322971	0.375827283	1.000000000	0.40785523	0.51802860
WA_0	-0.36055563	0.768170960	0.407855231	1.00000000	0.98151624
WT_0	-0.35231207	0.852836712	0.518028600	0.98151624	1.00000000

Tabela 10 – Gráfico de frequências absolutas - Satisfação



```
#det matriz de correlações= -5.207493e-20
```

Sendo que determinante da matriz de correlações deve apresentar o valor mais próximo possível de 0, através deste output é possível garantir que as variáveis estão para o desenvolvido na AFCP.

## **2. Avaliar o valor de KMO (Kaiser Meier Olkin)**

```
#Overall MSA = 0.5
```

O teste KMO (Kaiser-Meyer-Olkin) é uma medida que avalia se os dados são adequados para serem submetidos à análise fatorial, ou seja, determina se os dados representam padrões ou estruturas subjacentes através da análise fatorial.

Para o teste de KMO ser considerado válido, o valor deve estar o mais próximo possível de 1 e superior a 0,5, para que os dados sejam considerados adequados para a análise fatorial.

Com este output, apesar do baixo valor de KMO, consideram-se os dados válidos para aplicação da AFCP.

## **3. Verificar os resultados do teste de Bartlett**

```
#RESULTADO: p-value < 2.2e-16 .
```

O intuito do teste de Bartlett é verificar se a matriz de correlações entre as variáveis originais é uma matriz de identidade, ou seja, se todas as variáveis são completamente não correlacionadas entre si pois, se a matriz de correlações for uma matriz de identidade, não há estrutura subjacente nos dados que justifique a redução dimensional e a AFCP não é uma análise apropriada.

O teste de Bartlett testa a hipótese nula de que a matriz de correlações entre as variáveis originais é uma matriz de identidade. Se o teste de Bartlett resultar em um valor-p significativamente baixo (geralmente abaixo de um nível de significância 0,05), isso indica que a matriz de correlações não é uma matriz de identidade e que há correlações significativas entre as variáveis originais.

Para a realização deste teste consideram-se os valores das variáveis originais.

O valor de p-value do Bartlett test considerando a Base de Dados Grupo\_pre apresenta valores >0.05, o que indica a matriz de correlação entre as variáveis originais não é uma matriz de identidade, e que os dados estão aptos para a realização da AFCP.

## 5º Passo – Gráficos Biplot

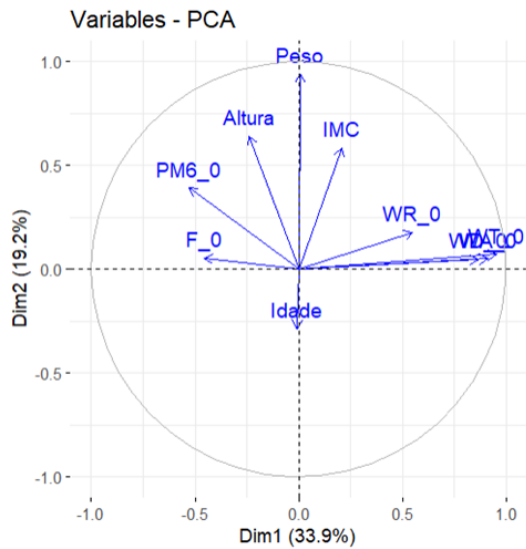


Gráfico 6 – Biplot (Dim2, Dim1)

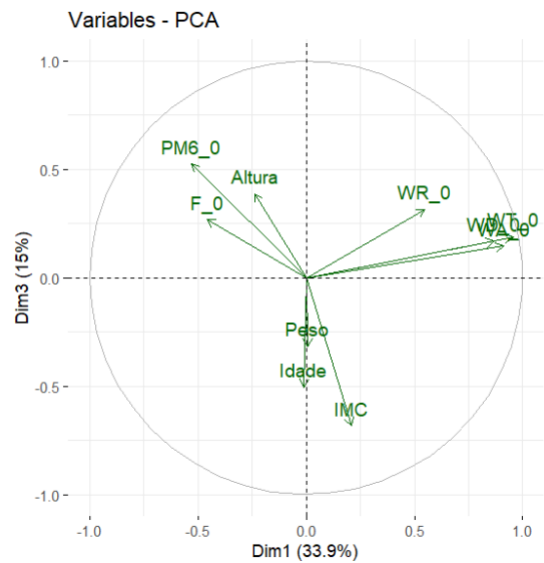


Gráfico 7 – Biplot (Dim3, Dim1)

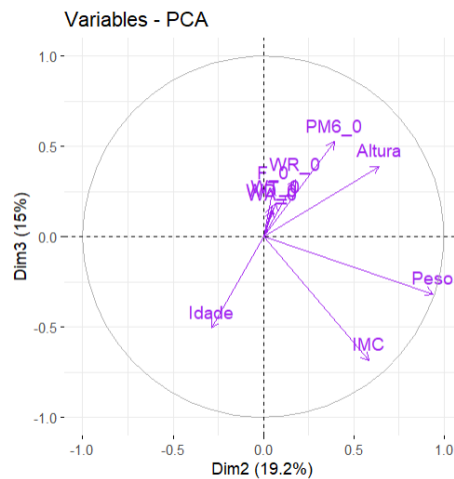


Gráfico 8 – Biplot (Dim3, Dim2)

A proximidade e os ângulos entre os vetores e os eixos e o comprimento dos vetores podem fornecer insights sobre como essas variáveis estão relacionadas entre si e com as dimensões.

Considerando a análise dos gráficos:

- WT\_0, WA\_0 e WD\_0 destacam-se com grande importância e grande correlação com a dimensão 1. O WR\_0 também se correlaciona com esta dimensão, mas representa menores valores de correlação e menor importância
- PM6\_0 apresenta ângulos próximos de 45° em todos os gráficos. Indicando fraca correlação com as dimensões nestas representações de gráficos biplot.

- Peso e Altura apresentam ângulos pequenos entre si, em todos os gráficos. Demonstram grande importância e correlação com a dimensão 2.
- O valor IMC, na análise da PCA, apresenta o maior valor absoluto na dimensão 3. Mas o gráfico 8 representa o IMC com um vetor de grande comprimento (significando grande importância) e com um ângulo próximo dos 45° entre a dimensão 2 e a dimensão 3.

Reanalizando a Análise PCA verifica-se o valor de IMC apresenta um valor significativo de correlação com o IMC, apesar de mais baixo. Confirmando o insight retirado o gráfico biplot.

## 6º Passo – Análises Hierárquicas Clusters – Pré-cirurgia

A Análise Hierárquica de Clusters no espaço de dimensão reduzida pode revelar padrões complexos nos dados que não seriam facilmente identificados no espaço original.

Os clusters identificados num espaço de dimensão reduzida podem ser mais facilmente interpretados e relacionados às variáveis originais. Esta metodologia facilita a interpretação dos resultados e a formulação de insights significativos.

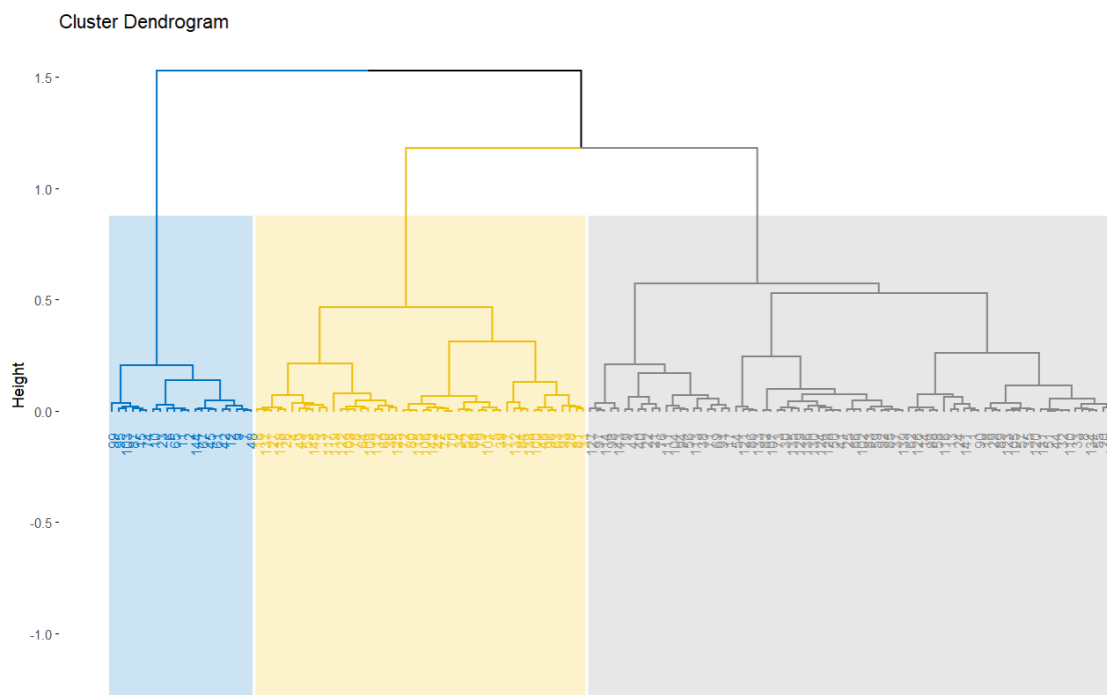


Gráfico 9 – Dendrograma de clusters pré-cirurgia

Para a análise do projeto e considerando a classificação do dendrograma e nível computacional definido, opta-se por selecionar 3 grupos de clusters diferentes.

## 7º Passo – Caracterização dos grupos de clusters

Considerando os grupos de clusters definidos no dendrograma, caracteriza-se os indivíduos e as suas variáveis originais por grupos de clusters.

O primeiro passo para realizar a caracterização dos grupos de clusters passou por gerar uma nova coluna na base de dados original que identificasse o cluster\_pre de cada indivíduo. Após esta análise considerou-se relevante quantificar quantos indivíduos se encontram representados em cada cluster\_pré-cirurgia e, para especificar a análise descritiva e estimativas amostrais de cada grupo de cluster pré, aplicou-se a função summary, agregada a cada grupo de cluster pré.

- Cluster 1 – 53 indivíduos
- Cluster 2 – 54 indivíduos
- Cluster 3 – 38 indivíduos

	Idade.Min.	Idade.1st Qu.	Idade.Median	Idade.Mean	Idade.3rd Qu.	Idade.Max.
1	48.00000	67.00000	74.00000	73.01887	79.00000	86.00000
2	49.00000	64.00000	70.00000	68.83333	73.00000	84.00000
3	52.00000	63.50000	70.00000	69.28947	76.00000	84.00000
	Peso.Min.	Peso.1st Qu.	Peso.Median	Peso.Mean	Peso.3rd Qu.	Peso.Max.
1	49.00000	62.00000	66.00000	67.71698	75.00000	88.00000
2	68.00000	81.25000	86.50000	89.35185	95.50000	135.00000
3	56.00000	70.00000	75.00000	74.89474	80.75000	97.00000
	IMC.Min.	IMC.1st Qu.	IMC.Median	IMC.Mean	IMC.3rd Qu.	IMC.Max.
1	19.14062	24.57787	26.72287	26.56976	28.04038	35.25076
2	22.20408	28.77829	31.44519	31.72585	34.03518	42.32428
3	23.12062	26.64738	30.07974	29.85180	31.92635	38.08864
	F_0.Min.	F_0.1st Qu.	F_0.Median	F_0.Mean	F_0.3rd Qu.	F_0.Max.
1	75.00000	100.00000	110.00000	110.27566	120.00000	140.00000
2	46.00000	98.50000	110.00000	108.30759	120.00000	140.00000
3	46.00000	75.75000	88.50000	87.63158	100.00000	120.00000
	EVA_0.Min.	EVA_0.1st Qu.	EVA_0.Median	EVA_0.Mean	EVA_0.3rd Qu.	EVA_0.Max.
1	0.000000	3.000000	5.000000	4.396226	6.000000	8.000000
2	0.000000	4.000000	5.000000	5.166667	7.000000	10.000000
3	0.000000	6.250000	8.000000	6.815789	8.000000	10.000000
	PM6_0.Min.	PM6_0.1st Qu.	PM6_0.Median	PM6_0.Mean	PM6_0.3rd Qu.	PM6_0.Max.
1	74.0000	220.0000	270.0000	264.1509	320.0000	450.0000
2	93.0000	250.5000	301.5000	311.2593	376.5000	522.0000
3	0.0000	104.7500	180.0000	177.7632	236.2500	402.0000
	WT_0.Min.	WT_0.1st Qu.	WT_0.Median	WT_0.Mean	WT_0.3rd Qu.	WT_0.Max.
1	8.00000	31.00000	39.00000	37.41509	46.00000	61.00000
2	29.00000	38.00000	45.50000	45.53704	51.75000	75.00000
3	52.00000	61.25000	70.00000	69.52632	75.75000	96.00000
	F_90.Min.	F_90.1st Qu.	F_90.Median	F_90.Mean	F_90.3rd Qu.	F_90.Max.
1	78.0000	95.0000	100.0000	105.0755	116.0000	130.0000
2	90.0000	95.0000	105.0000	106.1667	114.7500	153.0000
3	70.0000	90.0000	100.0000	101.4211	110.0000	130.0000
	EVA_90.Min.	EVA_90.1st Qu.	EVA_90.Median	EVA_90.Mean	EVA_90.3rd Qu.	EVA_90.Max.
1	0.000000	0.000000	0.000000	1.226415	2.000000	8.000000
2	0.000000	0.000000	1.000000	1.425926	2.750000	6.000000
3	0.000000	0.000000	1.500000	2.026316	3.000000	7.000000
	PM6_90.Min.	PM6_90.1st Qu.	PM6_90.Median	PM6_90.Mean	PM6_90.3rd Qu.	PM6_90.Max.
1	132.0000	267.0000	303.0000	303.1887	360.0000	450.0000
2	141.0000	294.0000	339.0000	339.1111	390.0000	528.0000
3	121.0000	240.0000	295.0000	304.8421	356.5000	543.0000
	WT_90.Min.	WT_90.1st Qu.	WT_90.Median	WT_90.Mean	WT_90.3rd Qu.	WT_90.Max.
1	1.00000	5.00000	11.00000	12.67925	18.00000	38.00000
2	1.00000	8.00000	12.00000	14.01852	18.00000	46.00000
3	2.00000	9.00000	14.00000	16.13158	22.25000	39.00000
	Satisfação.Min.	Satisfação.1st Qu.	Satisfação.Median	Satisfação.Mean	Satisfação.3rd Qu.	Satisfação.Max.
1	3.000000	9.000000	10.000000	9.113208	10.000000	10.000000
2	1.000000	8.000000	9.000000	8.444444	10.000000	10.000000
3	3.000000	8.000000	10.000000	8.894737	10.000000	10.000000

Tabela 11 – Função summary () agregada por cluster\_pre

Para visualização gráfica considerou-se pertinente realizar uma observação em boxplot para as variáveis quantitativas e em tabelas de frequências para as variáveis qualitativas. Os restantes gráficos encontram-se em anexo ao trabalho.

Opta-se por analisar todas as variáveis de forma a compreender como os indivíduos agregados através do cluster pré comportam após a cirurgia, considerando desta forma uma análise suplementar mas não representativa dos clusters.

Pode-se considerar que o grupo de indivíduos agregado no grupo pré caracteriza-se por:

### **Cluster 1:**

#### **Dados demográficos:**

- Idades médias mais avançadas
- Maioritariamente mulheres
- Menor estatura corporal com menor peso e menor valores de IMC médios

#### **Pré- Cirurgia:**

- Entre o Grupo 1 e 5 no Grupo Pré-Cirúrgico.
- A maioria dos indivíduos são capazes de garantir a amplitude de flexão do joelho pré-cirurgia
- Apresentam provas de marcha razoáveis
- Dor ligeira a moderada pré cirurgia.
- Menores valores na escala WOMAC pré-cirurgia comparativamente com os restantes clusters, significando menor dor, menos rigidez, mais capacidade de realização das atividades de vida (que resultam em menores valores na avaliação total WOMAC).
- Conclusão: Menor valores de WOMAC com marcha e flexão do joelho razoável e dor ligeira a moderada

#### **Pós-cirurgia:**

- Menor nível de flexão do joelho pós-cirurgia
- Melhores provas de marcha pós cirurgia
- Menor nível de dor pós-cirurgia
- Melhoria WT\_90
- 75% dos indivíduos avaliaram o nível de satisfação entre 9 e 10

### **Cluster 2:**

#### **Dados demográficos:**

- Idades média de 70 anos, com pequena amplitude interquartil
- Maior estatura corporal com maior peso e maiores valores de IMC médios.
- 50% dos indivíduos são do sexo feminino e 50% dos indivíduos do sexo masculino

#### **Pré- Cirurgia:**

- Maioritariamente entre o grupo 1 e 3 pré cirurgia
- A maioria dos indivíduos são capazes de garantir a amplitude de flexão do joelho pré-cirurgia
- Apresentam os melhores valores na prova de marcha
- Dor ligeira a moderada em 75% dos indivíduos e intensa em 25% dos indivíduos
- Escala WOMAC pré-cirurgia razoável mas aumentada comparativamente com o cluster 1.
- Conclusão: Indivíduos capazes de deambular mas com dor significativa. WOMAC razoável.

#### **Pós-cirurgia:**

- Menor nível de flexão do joelho pós-cirurgia
- Melhores provas de marcha pós cirurgia
- Menor nível de dor pós-cirurgia
- Melhoria WT\_90
- 75% dos indivíduos avaliaram o nível de satisfação entre 8 e 10

### **Cluster 3:**

#### **Dados demográficos:**

- Idades média de 70 anos, com grande amplitude interquartil
- Maioritariamente indivíduos do sexo feminino
- Menor estatura corporal com menor peso e menores valores de IMC médios.

#### **Pré- Cirurgia:**

- Maioritariamente entre o grupo 1 e 5 pré cirurgia
- Mais de 50% dos indivíduos na atingem um angulo de 90° do joelho.
- Apresentam os piores valores na prova de marcha
- Dor moderada a intensa em 75% dos indivíduos.
- Valores mais baixos na escala WOMAC Total que representam mais dor, mais rigidez, menor desempenho nas atividades de vida.
- Conclusão: Dor intensa, prova de marcha fraca, WOMAC aumentado

#### **Pós-cirurgia:**

- 75% dos indivíduos foram capazes de atingir flexão de 90° do joelho.
- Melhorias significativas na provas de marcha pós cirurgia, apesar que ainda apresentarem menores valores comparativamente com os restantes clusters.
- Menor nível de dor pós-cirurgia mas mais de 50% dos indivíduos mantem dor ligeira a moderada.
- Melhoria significativa da WT\_90
- Mais de 75% dos indivíduos avaliaram a cirurgia entre 8 e 10

### 8º Passo – Realizar e analisar a análise fatorial de componentes principais (AFCP) – Grupo Variáveis Pós-Cirurgia

Com o objetivo de reduzir a dimensionalidade do problema estatístico, aplicou-se a AFCP e considerou-se a seguinte análise:

\$coord	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Idade	0.24256675	-0.32927720	-0.2743829	0.54913047	0.59368142
Peso	0.04589172	0.82119947	-0.5537575	0.03397538	0.08025174
Altura	-0.30744504	0.69438600	0.1340101	0.50140006	0.11184376
IMC	0.28088531	0.39597758	-0.7643279	-0.36907268	0.01482116
F_90	-0.39352154	0.26037331	0.4096822	-0.41222113	0.55947724
PM6_90	-0.46292700	0.56086669	0.4260307	-0.05229804	0.01616066
WD_90	0.76777784	0.16004308	0.3313459	-0.24915719	0.16913567
WR_90	0.42308608	0.42644191	0.3159466	0.30138068	-0.35200798
WA_90	0.91329702	0.06933801	0.1218360	0.01586109	0.07897028
WT_90	0.94872825	0.14748604	0.2181374	-0.01482667	0.06021512
\$cor	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Idade	0.24256675	-0.32927720	-0.2743829	0.54913047	0.59368142
Peso	0.04589172	0.82119947	-0.5537575	0.03397538	0.08025174
Altura	-0.30744504	0.69438600	0.1340101	0.50140006	0.11184376
IMC	0.28088531	0.39597758	-0.7643279	-0.36907268	0.01482116
F_90	-0.39352154	0.26037331	0.4096822	-0.41222113	0.55947724
PM6_90	-0.46292700	0.56086669	0.4260307	-0.05229804	0.01616066
WD_90	0.76777784	0.16004308	0.3313459	-0.24915719	0.16913567
WR_90	0.42308608	0.42644191	0.3159466	0.30138068	-0.35200798
WA_90	0.91329702	0.06933801	0.1218360	0.01586109	0.07897028
WT_90	0.94872825	0.14748604	0.2181374	-0.01482667	0.06021512
\$cos2	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Idade	0.05883863	0.108423473	0.07528600	0.3015442741	0.3524576310
Peso	0.00210605	0.674368573	0.30664734	0.0011543265	0.0064403412
Altura	0.09452245	0.482171914	0.01795870	0.2514020194	0.0125090266
IMC	0.07889655	0.156798248	0.58419717	0.1362146450	0.0002196668
F_90	0.15485920	0.067794260	0.16783950	0.1699262627	0.3130147842
PM6_90	0.21430141	0.314571448	0.18150219	0.0027350854	0.0002611668
WD_90	0.58948282	0.025613788	0.10979009	0.0620793053	0.0286068761
WR_90	0.17900183	0.181852700	0.09982223	0.0908303172	0.1239096207
WA_90	0.83411145	0.004807759	0.01484401	0.0002515741	0.0062363055
WT_90	0.90008529	0.021752131	0.04758392	0.0002198303	0.0036258603
\$contrib	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Idade	1.89422831	5.3196891	4.6893398	29.66911078	41.59865674
Peso	0.06780138	33.0872189	19.1001464	0.11357483	0.76011843
Altura	3.04301981	23.6572823	1.1185935	24.73558613	1.47637236
IMC	2.53996557	7.6931491	36.3878957	13.40223556	0.02592607
F_90	4.98547802	3.3262575	10.4542211	16.71914058	36.94343211
PM6_90	6.89913773	15.4341332	11.3052291	0.26910659	0.03082410
WD_90	18.97758474	1.2567149	6.8384968	6.10801777	3.37631396
WR_90	5.76271663	8.9224207	6.2176284	8.93684601	14.62437844
WA_90	26.85306554	0.2358879	0.9245891	0.02475252	0.73603722
WT_90	28.97700228	1.0672465	2.9638600	0.02162922	0.42794056

Tabela 12 – Função PCA Grupo\_Pos\_Normalized



Ao analisar Eigen-Value e Aporte de informação retida (% de Variância) para cada dimensão, conclui-se:

#comp 1	3.106206e+01	31.06206
#comp 2	2.038154e+01	51.44360
#comp 3	1.605471e+01	67.49831
#comp 4	1.016358e+01	77.66189
#comp 5	8.472813e-01	86.13470

Considerando a necessidade de eigen-value  $>1$ , retêm-se no espaço de dimensão reduzida 4 dimensões com uma percentagem de variância de 77.66189%.

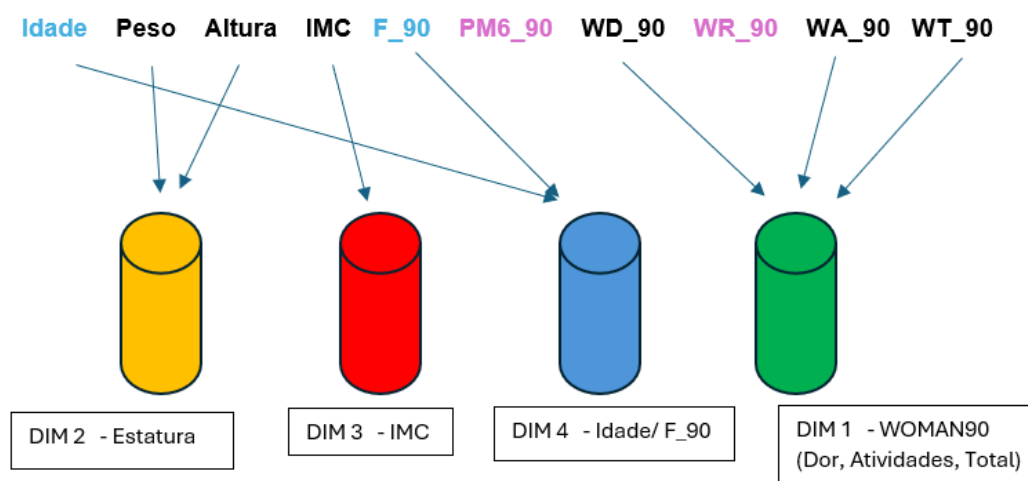


Figura 2 – Distribuição das variáveis pós- cirúrgicas pelas dimensões

Mantendo o raciocínio anterior e com base no output da realização da PCA, é possível analisar que as variáveis WD\_0, WA\_0 e WT\_0 correlacionam-se e contribuem para a dimensão 1. Com isto, a dimensão 1 renomeia-se WOMAC90 (Dor, Atividade e Total).

Neste grupo de variáveis, a WR\_90 destaca-se por se considerar “variável camaleão” que apresenta elevado nível de correlação com a dimensão 1 e 2, ou seja, pode manifestar-se com diferentes interpretações consoante a análise. Com isto, optamos por também não considerar a variável PM6\_0 na análise AFCP por apresentar correlação semelhante entre a dim2 e dim4.

Já as variáveis Peso e Altura agrupam-se na dimensão 2, mantendo a renomeação dimensão para estatura do utente.



O IMC encontra-se correlacionado com a dimensão 3, sendo esta a única variável com alto valor de correlação com nesta dimensão.

Relativamente à dim4, apesar de mencionada na PCA, é importante analisar tanto a variável Idade como F\_90, apresentarem valores de correlação elevados com a dim5 assim, poderão não ser representativas da variância deste componente.

## 9ª Passo - Validar a qualidade da AFCP - Variáveis pós-cirurgia

### 1.Avaliar o determinante da matriz de correlações das variáveis originais

```
#det matriz de correlações= -2.533659e-19
```

Sendo que determinante da matriz de correlações deve apresentar o valor mais próximo possível de 0, através deste output é possível garantir que as variáveis estão para o desenvolvido na AFCP.

	Idade	Peso	Altura	IMC	F_90
Idade	1.00000000	-0.065774469	-0.1142340	0.01700260	-0.191440774
Peso	-0.06577447	1.000000000	0.5366983	0.73168575	-0.001579049
Altura	-0.11423399	0.536698324	1.0000000	-0.17214492	0.192075645
IMC	0.01700260	0.731685754	-0.1721449	1.000000000	-0.144784815
F_90	-0.19144077	-0.001579049	0.1920756	-0.14478482	1.000000000
PM6_90	-0.32012710	0.171294837	0.4445031	-0.15595087	0.409748419
WD_90	0.03411860	-0.006459466	-0.1755504	0.11696032	-0.013772305
WR_90	-0.04313442	0.147152399	0.1595565	0.02392054	-0.107483058
WA_90	0.14609863	0.047426203	-0.1363091	0.14675747	-0.256944572
WT_90	0.11500510	0.052871069	-0.1296298	0.14530144	-0.213135915
	PM6_90	WD_90	WR_90	WA_90	WT_90
Idade	-0.32012710	0.034118604	-0.04313442	0.1460986	0.11500510
Peso	0.17129484	-0.006459466	0.14715240	0.0474262	0.05287107
Altura	0.44450310	-0.175550398	0.15955648	-0.1363091	-0.12962984
IMC	-0.15595087	0.116960319	0.02392054	0.1467575	0.14530144
F_90	0.40974842	-0.013772305	-0.10748306	-0.2569446	-0.21313592
PM6_90	1.00000000	-0.074909643	0.13418790	-0.3420995	-0.26386851
WD_90	-0.07490964	1.000000000	0.33355254	0.6448921	0.79274235
WR_90	0.13418790	0.333552538	1.00000000	0.3057949	0.44438288
WA_90	-0.34209950	0.644892087	0.30579492	1.0000000	0.96851223
WT_90	-0.26386851	0.792742353	0.44438288	0.9685122	1.00000000

Tabela 12 – Matriz de correlações\_Grupo\_Pós\_Normalized

### 2.Avaliar o valor de KMO (Kaiser Meier Olkin)

```
#Overall MSA = 0.5
```

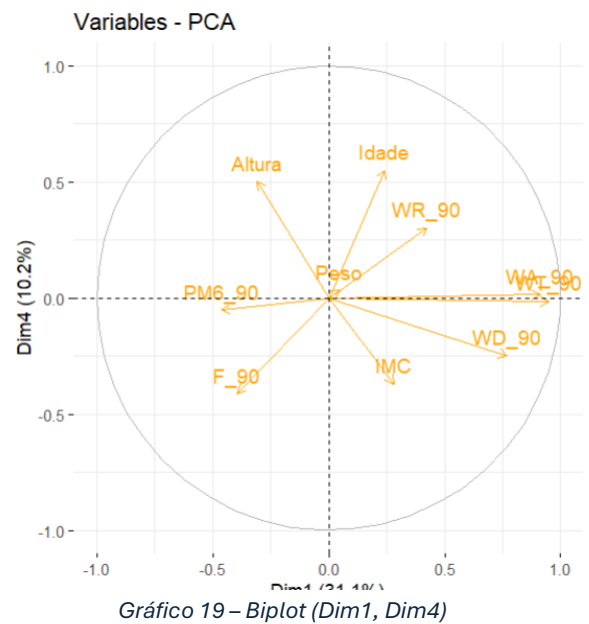
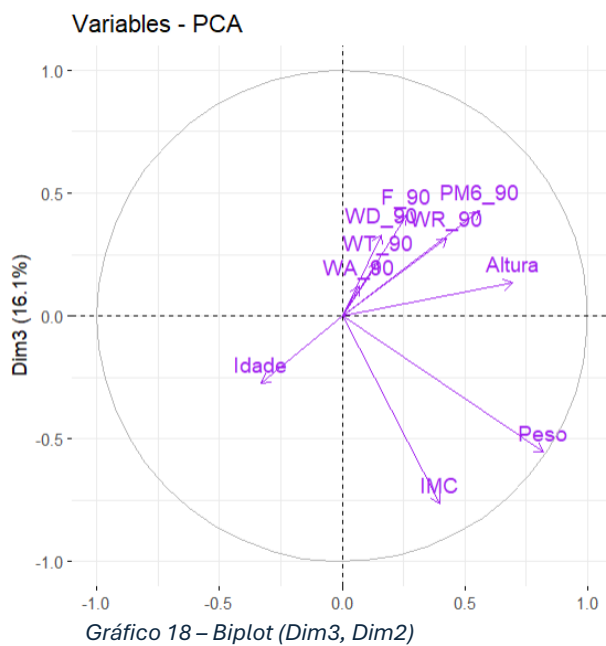
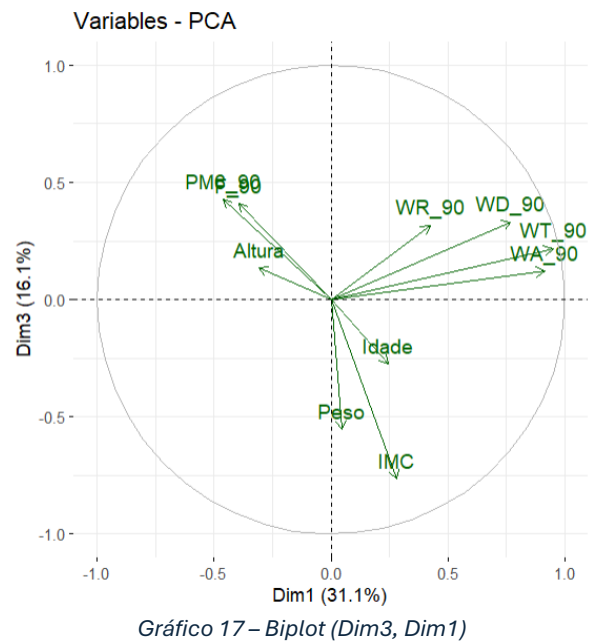
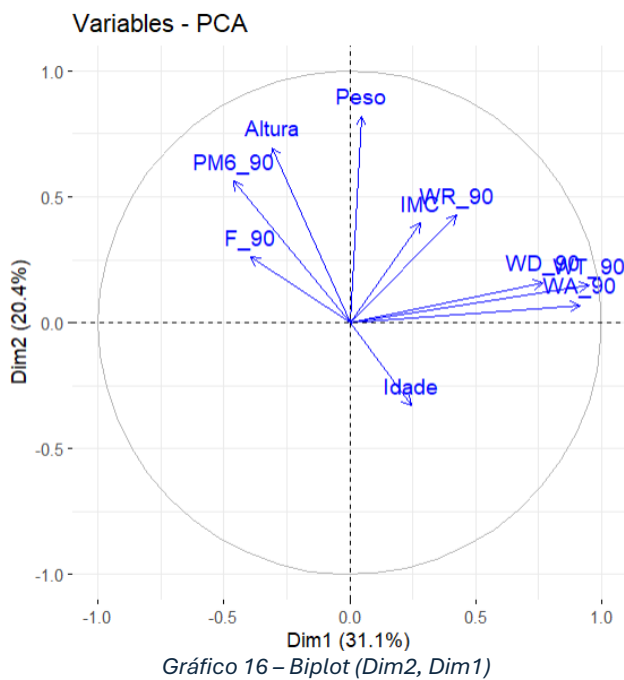
Com este output, apesar do baixo valor de KMO, consideram-se os dados válidos para aplicação da AFCP.

### 3. Verificar os resultados do teste de Bartlett

#RESULTADO:  $p\text{-value} < 2.2e-16$ .

O valor de  $p\text{-value}$  do Bartlett test considerando a Base de Dados Grupo\_pós apresenta valores  $>0.05$ , o que indica a matriz de correlação entre as variáveis originais não é uma matriz de identidade, e que os dados estão aptos para a realização da AFCP.

### 10º Passo – Gráficos Biplot -Variáveis Pós-Cirurgia



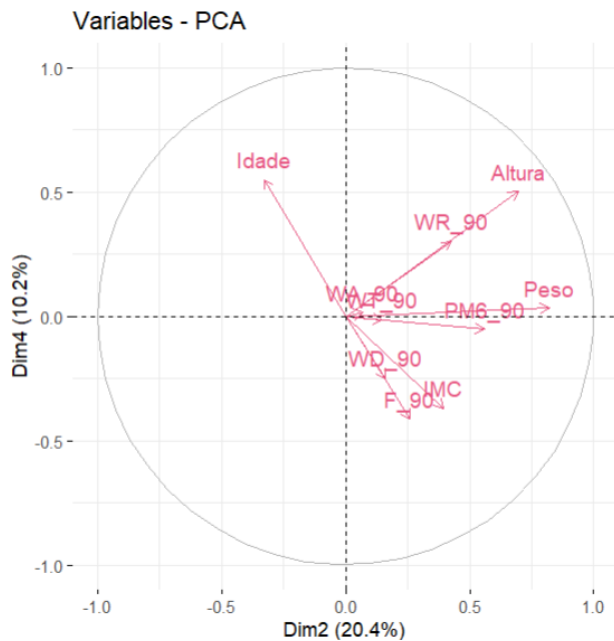


Gráfico 20 – Biplot (Dim2, Dim4)

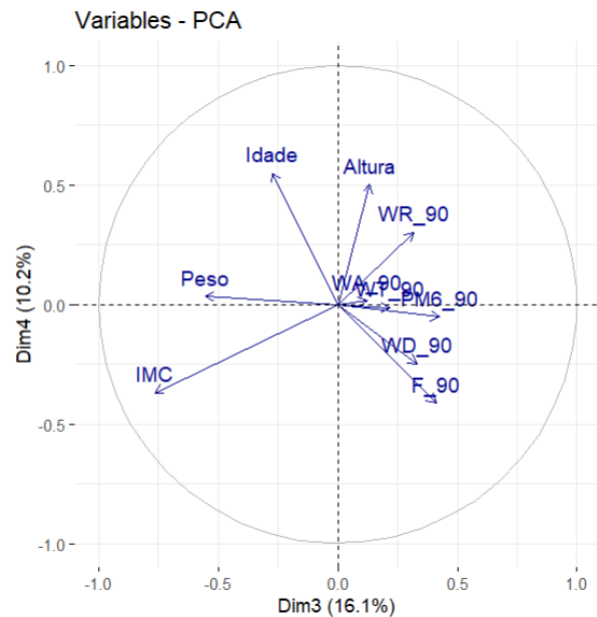


Gráfico 21 – Biplot (Dim3, Dim4)

Considerando a análise dos gráficos:

- WT\_0, WA\_0 e WD\_0 destacam-se com grande importância e grande correlação com a dimensão 1. O WR\_0 apresenta ângulos próximos de 45° em todos os gráficos.
- PM6\_0 apresenta também ângulos próximos de 45° em todos os gráficos. Indicando fraca correlação com as dimensões nestas representações de gráficos biplot.
- Peso e Altura destacam-se com grande importância e correlação com a dimensão 2.
- O valor IMC representa-se com um vetor de grande comprimento e pequeno ângulo com a dim3, demonstrando a sua correlação com este componente. Reanalizando a Análise PCA verifica-se o valor de IMC apresenta um valor significativo de correlação com o IMC, apesar de mais baixo. Confirmando o insight retirado o gráfico biplot.
- Analisando os gráficos que analisam a dim4 no eixo do x, é possível verificar que a variável Idade apresenta uma correlação positiva com a dim4 e a variável F\_90 uma correlação negativa com ligeira menor importância.

## 11º Passo – Análises Hierárquicas Clusters – Pré-cirurgia

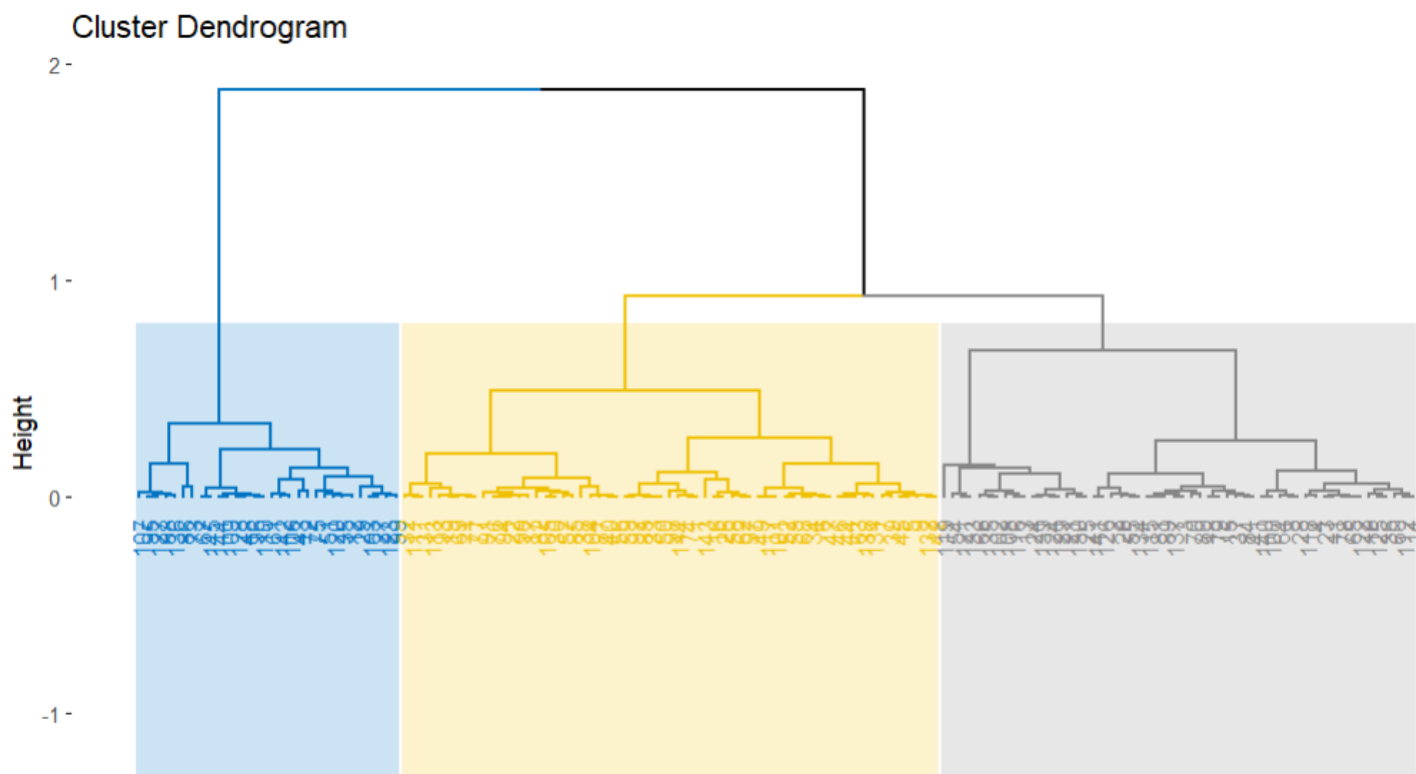


Gráfico 22 – Dendrograma de clusters pós-cirurgia

Para a análise do projeto e considerando a classificação do dendrograma e nível computacional definido, opta-se por selecionar 3 grupos de clusters diferentes.

## 12º Passo – Caracterização dos grupos de clusters

Considerando os grupos de clusters definidos no dendrograma, caracterizaram-se os indivíduos e as suas variáveis originais por grupos de clusters.

Mantendo o raciocínio como no grupo\_pre, o primeiro passo para realizar a caracterização dos grupos de clusters passou por gerar uma nova coluna na base de dados original que identificasse o cluster\_pós de cada indivíduo.

Após esta análise considerou-se relevante quantificar quantos indivíduos se encontram representados em cada cluster\_póscirurgia e, para especificar a análise descritiva e estimativas amostrais de cada grupo de cluster pré, aplicou-se a função summary, agregada a cada grupo de cluster pré.

- Cluster 1 – 47 indivíduos
- Cluster 2 – 65 indivíduos
- Cluster 3 – 33 indivíduos

	Idade.Min.	Idade.1st Qu.	Idade.Median	Idade.Mean	Idade.3rd Qu.	Idade.Max.
1	48.00000	63.00000	67.00000	67.06383	71.50000	83.00000
2	52.00000	69.00000	73.00000	72.50769	77.00000	84.00000
3	49.00000	67.00000	72.00000	71.36364	77.00000	86.00000
	Altura.Min.	Altura.1st Qu.	Altura.Median	Altura.Mean	Altura.3rd Qu.	Altura.Max.
1	155.0000	163.0000	170.0000	169.6170	174.0000	198.0000
2	140.0000	154.0000	159.0000	159.1846	164.0000	175.0000
3	144.0000	151.0000	158.0000	159.4242	167.0000	187.0000
	IMC.Min.	IMC.1st Qu.	IMC.Median	IMC.Mean	IMC.3rd Qu.	IMC.Max.
1	19.14062	24.28866	26.72287	27.49008	30.95365	37.42140
2	20.04745	27.47138	29.76190	30.50650	33.71488	41.32231
3	23.43750	26.92744	28.88889	29.72138	32.02037	42.32428
	F_0.Min.	F_0.1st Qu.	F_0.Median	F_0.Mean	F_0.3rd Qu.	F_0.Max.
1	70.0000	90.0000	110.0000	106.2340	120.0000	140.0000
2	46.0000	90.0000	103.6100	100.1726	112.0000	140.0000
3	65.0000	100.0000	110.0000	106.6364	120.0000	140.0000
	PM6_0.Min.	PM6_0.1st Qu.	PM6_0.Median	PM6_0.Mean	PM6_0.3rd Qu.	PM6_0.Max.
1	100.0000	256.0000	300.0000	312.9574	391.0000	522.0000
2	68.0000	180.0000	237.0000	228.2923	290.0000	420.0000
3	0.0000	162.0000	243.0000	242.8788	330.0000	450.0000
	EVA_0.Min.	EVA_0.1st Qu.	EVA_0.Median	EVA_0.Mean	EVA_0.3rd Qu.	EVA_0.Max.
1	0.000000	3.000000	5.000000	4.978723	7.000000	10.000000
2	0.000000	4.000000	6.000000	5.600000	7.000000	10.000000
3	0.000000	3.000000	5.000000	5.242424	8.000000	10.000000
	WT_0.Min.	WT_0.1st Qu.	WT_0.Median	WT_0.Mean	WT_0.3rd Qu.	WT_0.Max.
1	8.00000	35.50000	46.00000	47.08511	59.50000	75.00000
2	14.00000	38.00000	47.00000	47.44615	55.00000	84.00000
3	26.00000	39.00000	52.00000	54.15152	70.00000	96.00000
	EVA_90.Min.	EVA_90.1st Qu.	EVA_90.Median	EVA_90.Mean	EVA_90.3rd Qu.	EVA_90.Max.
1	0.0000000	0.0000000	0.0000000	0.8297872	1.0000000	7.0000000
2	0.0000000	0.0000000	0.0000000	1.0615385	2.0000000	8.0000000
3	0.0000000	2.0000000	3.0000000	3.3636364	5.0000000	8.0000000
	F_90.Min.	F_90.1st Qu.	F_90.Median	F_90.Mean	F_90.3rd Qu.	F_90.Max.
1	90.00000	110.00000	115.00000	114.17021	120.00000	153.00000
2	70.00000	91.00000	99.00000	100.63077	110.00000	128.00000
3	78.00000	90.00000	95.00000	98.45455	105.00000	116.00000
	PM6_90.Min.	PM6_90.1st Qu.	PM6_90.Median	PM6_90.Mean	PM6_90.3rd Qu.	PM6_90.Max.
1	284.0000	360.0000	400.0000	393.3191	420.0000	543.0000
2	141.0000	240.0000	283.0000	275.2615	312.0000	387.0000
3	121.0000	240.0000	300.0000	290.5152	346.0000	528.0000
	WT_90.Min.	WT_90.1st Qu.	WT_90.Median	WT_90.Mean	WT_90.3rd Qu.	WT_90.Max.
1	1.000000	3.000000	9.000000	8.574468	12.000000	20.000000
2	1.000000	8.000000	11.000000	10.876923	14.000000	23.000000
3	18.000000	23.000000	26.000000	28.242424	32.000000	46.000000
	Satisfação.Min.	Satisfação.1st Qu.	Satisfação.Median	Satisfação.Mean	Satisfação.3rd Qu.	Satisfação.Max.
1	5.000000	8.000000	9.000000	9.021277	10.000000	10.000000
2	6.000000	9.000000	10.000000	9.276923	10.000000	10.000000
3	1.000000	7.000000	8.000000	7.575758	10.000000	10.000000

Tabela 13 – Função summary () agregada por cluster\_pos

Para visualização gráfica considerou-se pertinente realizar uma observação em boxplot para as variáveis quantitativas e em tabelas de frequências para as variáveis qualitativas. Os restantes gráficos encontram-se em anexo ao trabalho.

Considera-se importante analisar a satisfação associada ao grupo de variáveis pós cirurgia pois pode estar influenciada com estas variáveis.

Pode-se considerar que o grupo de indivíduos agregado no grupo pós caracteriza-se por:

### **Cluster 1:**

#### **Dados demográficos:**

- Idades menores apresentando uma média e mediana mais baixas
- Maioritariamente do sexo masculino
- Destacam-se por serem indivíduos mais altos e mais pesados, mas associados a IMC mais baixos.

#### **Pré- Cirurgia:**

- Entre o Grupo 1 e 7 no Grupo Pré-Cirúrgico.
- A maioria dos indivíduos são capazes de garantir a amplitude de flexão do joelho pré-cirurgia
- Apresentam provas de marcha com grandes as maiores distâncias pré cirurgia.
- Dor ligeira a moderada pré cirurgia em 75% dos indivíduos
- Menores valores na escala WOMAC pré-cirurgia comparativamente com os restantes clusters, significando menor dor, menos rigidez, mais capacidade de realização das atividades de vida (que resultam em menores valores na avaliação total WOMAC).

#### **Pós-cirurgia:**

- Maioritariamente do grupo 5 pós cirurgia
- Melhoria significativa dos pós cirurgia, com 75% dos indivíduos mencionando um score de 0 e 1 no nível da dor.
- Mantem amplitude de flexão do joelho pós cirurgia.
- Melhorias nas provas de marcha pós cirurgia
- Melhoria significativa do WT\_90
- 75% dos indivíduos avaliaram o nível de satisfação entre 8 e 10
- Conclusão: Baixa dor, capazes de deambular e com melhoria significativa da escala WOMAC. Resulta em tendencialmente boa satisfação.

### **Cluster 2:**

#### **Dados demográficos:**

- Indivíduos mais velhos com media e mediana superior a 72 anos
- Indivíduos mais baixos, associados a IMC mais altos
- Maioritariamente do sexo feminino

#### **Pré- Cirurgia:**

- Maioritariamente do grupo 1 pré-cirurgia
- 75% dos indivíduos são capazes de garantir a amplitude de flexão do joelho pré-cirurgia



- Resultados na prova de marcha razoáveis
- Dor ligeira a moderada em 75% dos indivíduos e intensa em 25% dos indivíduos
- Escala WOMAC pré-cirurgia razoável, mas aumentada comparativamente com o cluster 1.

#### **Pós-cirurgia:**

- Menor nível de flexão do joelho pós-cirurgia
- Melhores provas de marcha pós cirurgia
- Menor nível de dor pós-cirurgia
- Melhoria no WT\_90
- Satisfação aumentada sendo que mais de 75% dos indivíduos apresentam scores superiores a 9
- Conclusão: Apesar da diminuição da flexão do joelho, os indivíduos apresentam melhorias nos resultados clínicos como a dor, WOMAC e flexão.

#### **Cluster 3:**

##### **Dados demográficos:**

- Idades média de 70 anos, com grande amplitude interquartil
- Maioritariamente indivíduos do sexo feminino
- Maior estatura corporal com maior peso e maiores valores de IMC médios.

##### **Pré- Cirurgia:**

- Maioritariamente entre o grupo 1 e 5 pré cirurgia
- Grande amplitude de flexão do joelho pré-cirurgia
- Resultados na prova de marcha de fraco a razoável
- Dor moderada a intensa em 75% dos indivíduos.
- Valores mais altos na escala WOMAC Total que representam mais dor, mais rigidez, menor desempenho nas atividades de vida.

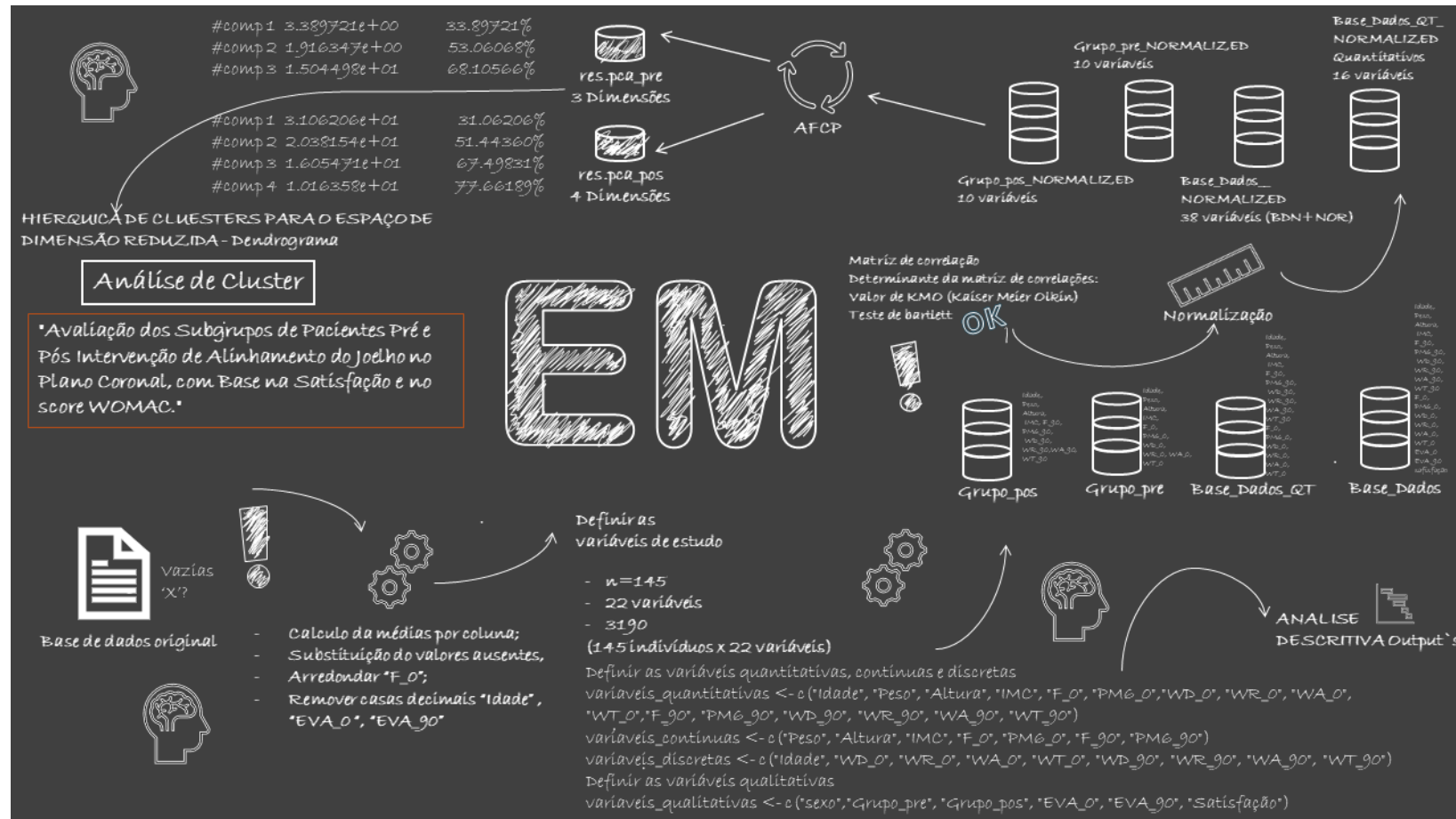
##### **Pós-cirurgia:**

- 75% dos indivíduos foram capazes de atingir flexão de 90° do joelho.
- Melhorias significativas nas provas de marcha pós cirurgia, apesar que ainda apresentarem menores valores comparativamente com os restantes clusters.
- Menor nível de dor pós-cirurgia, mas mais de 50% dos indivíduos mantem dor ligeira a moderada.
- Perda significativa da redução da amplitude interquartil
- Piores resultados na melhoria do WT no período pós cirurgia sendo que 50% dos indivíduos apresentam WT\_90 superior a 26.
- 75% dos indivíduos mantém níveis de dor após cirurgia desde ligeira a intensa.
- Conclusão: Diminuição da flexão do joelho associada, provas de marchas razoáveis. Mais dor e menor satisfação.

f) Descreva por infografia ponto de partida do seu estudo e os seus resultados (na verdade, é responder porque decidiu fazer este trabalho, com esta Base de Dados, e o que pretende responder?) (1+1 val.)

Problemas:

Como os dados demográficos e as variáveis clínicas pré e pós cirurgia influenciam a satisfação do utente ao realizar uma Prótese Total do Joelho?





## **Conclusão das relações entre as variáveis clínicas segundo caracterização dos clusters:**

### **Variáveis clínicas pré cirurgia:**

**Cluster 1** -. Alto WOMAC, com marcha e flexão do joelho razoável e dor ligeira a moderada.

**Cluster 2** - Indivíduos capazes de deambular, mas com dor significativa. WOMAC razoável.

**Cluster 3** - Dor intensa, prova de marcha fraca, flexão mantida, WOMAC aumentado

### **Variáveis clínicas pós cirurgia:**

**Cluster 1** - Baixa dor, capazes de deambular e com melhoria significativa da escala WOMAC. Resulta em tendencialmente boa satisfação.

**Cluster 2** - Apesar da diminuição da flexão do joelho, os indivíduos apresentam melhorias nos resultados clínicos como a dor, WOMAC e flexão.

**Cluster 3** - Diminuição da flexão do joelho associada, provas de marchas razoáveis. Mais dor e menor satisfação.

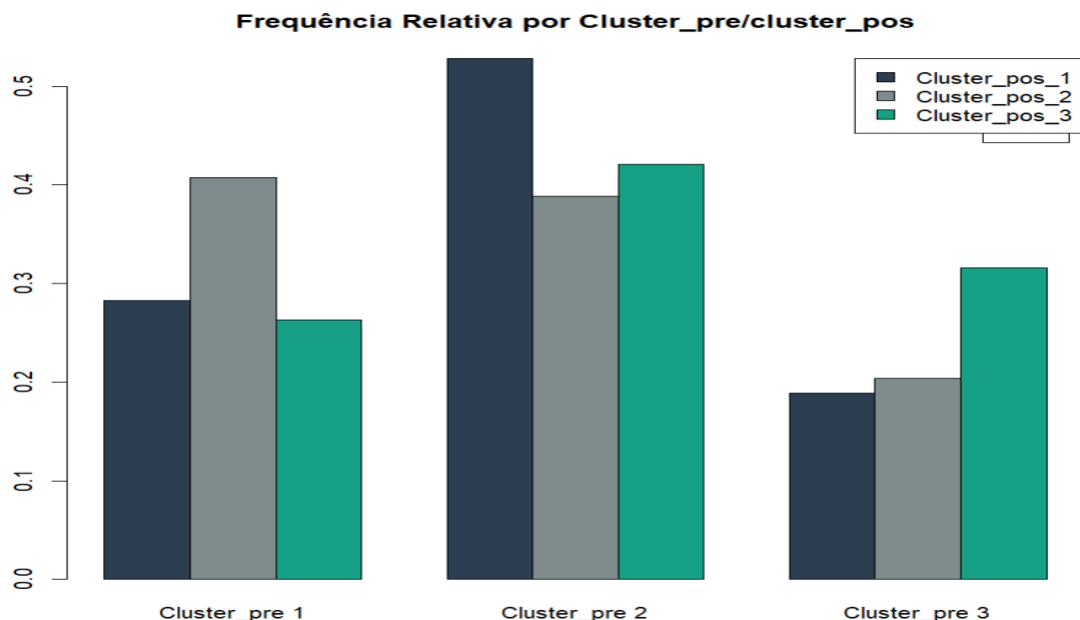
**g) Elabore um parágrafo conclusivo (máximo 5 a 7 linhas). (1 val.)**

Considerando a análise em estudo, decidimos analisar a forma como os indivíduos se relacionam entre as variáveis clínicas pré e pós cirurgia. Conclui-se que:

- Utentes que mantêm a dor após as cirurgias tendem a apresentar menores classificações de satisfação. Este grupo de indivíduos é identificado no cluster3. Grande parte destes indivíduos encontram-se previamente à cirurgia no cluster 2 ou 3, ou seja, caracterizam-se uma dor significativa e um WOMAC de razoável a aumentado.
- Sendo o cluster 1 pós cirurgia aquele que apresenta melhores resultados clínicos, menor dor e mais satisfação, este cluster é representado por 50% dos indivíduos do cluster 2. Ou seja, utentes com dor significativa e que melhorar a sua dor.

Clusters pré-cirurgia	Clusters pós-cirurgia		
	1	2	3
1	0.2830189	0.5283019	0.1886792
2	0.4074074	0.3888889	0.2037037
3	0.2631579	0.4210526	0.3157895

*Tabela 14 – Tabelas Frequências Relativas entre clusters*



*Gráfico 23 – Gráfico de Frequências Relativas entre clusters Pré e Pós Cirurgia*

**h) Adicionar, justificando, outros elementos para enriquecer o trabalho (1 val.)**

**Para utentes com uma dor pós-cirurgia superior ou igual a 3 segundo a escala EVA\_90, pode-se afirmar que o valor médio de satisfação é inferior a 7?**

1º Passo: Selecionado um subgrupo da amostra onde se incluem somente os indivíduos com Score de EVA90 superior a 3

- `Amostra_Dor3 <- Base_Dados[Base_Dados$EVA_90 >= 3, ]`

2º Passo: Avaliar a dimensão da amostra

- `nrow(Amostra_Dor3)` = 35 indivíduos ( $n > 30$ )

3º Passo: Formulação das hipóteses

- $H_0$ : valor médio de satisfação = 8
- $H_1$ : valor médio de satisfação  $< 8$

4º Passo: Condições de aplicabilidade

- X v.a.: nível de satisfação
- $N = 35$
- $X \cap$  Qualquer
- $\sigma$  - Desconhecido
- Nível de Significância – 0,05

➤ Utiliza-se o quantil normal padrão (Z) através da função `t.test()`

5º Passo: Função `t.test(Amostra_Dor3$Satisfação, mu = 7, alternative = "less")` para realizar o respetivo teste de hipóteses com um nível de significância de 0.05.

Com esta função, é possível verificar que:

- A média de idades deste subgrupo amostra é 7.91 (média estimada).
- Dado  $p\text{-value} = 0.986$ , ou seja,  $p\text{-value} > 0.05$  (valor do nível de significância), perante este teste de hipóteses não rejeitamos  $H_0$ .
- Desta forma, conclui-se que os dados da amostra não fornecem evidência significativa para afirmar que o valor médio da satisfação é significativamente inferior a 7, ou seja, o valor médio da satisfação dos indivíduos com `score EVA_90 => 3` não é significativamente inferior a 7.

# ANEXO 1

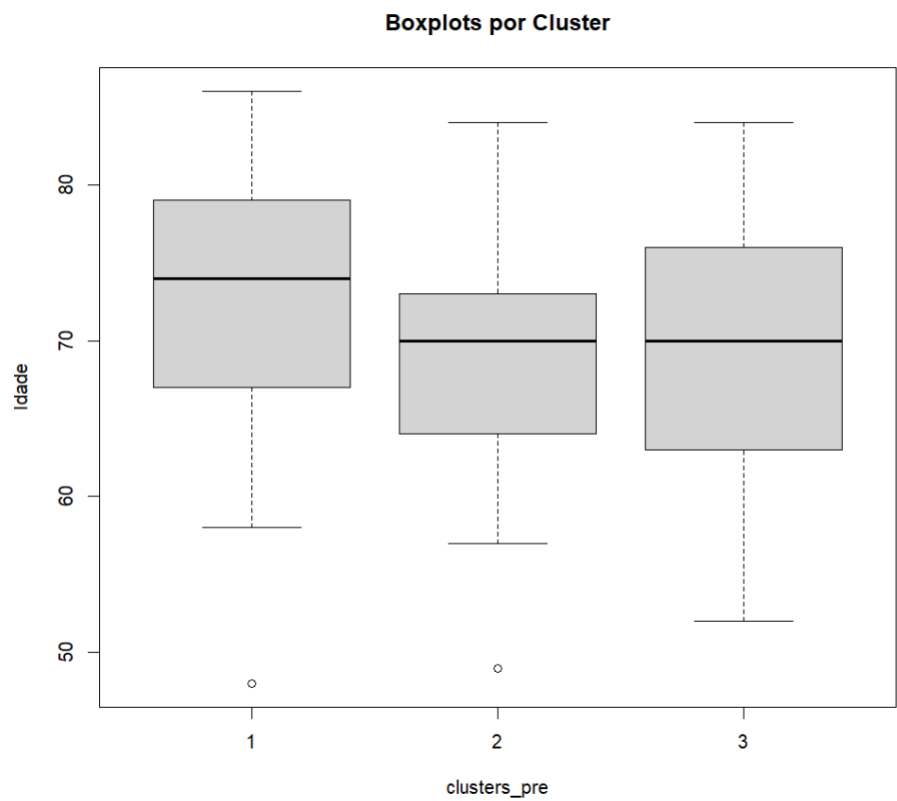


Gráfico 24 – Idades por Cluster Pré

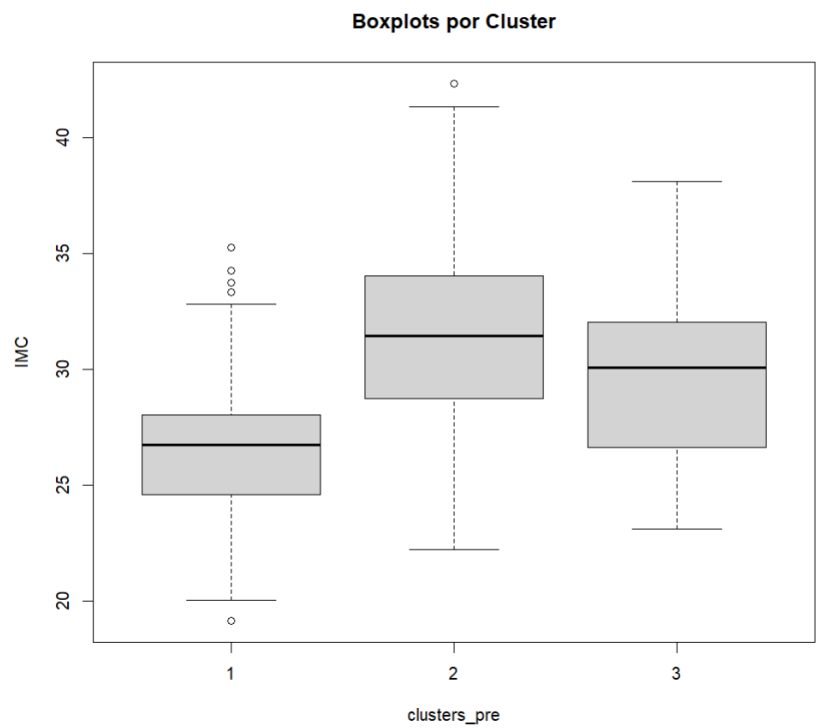


Gráfico 24 – IMC por Cluster Pré

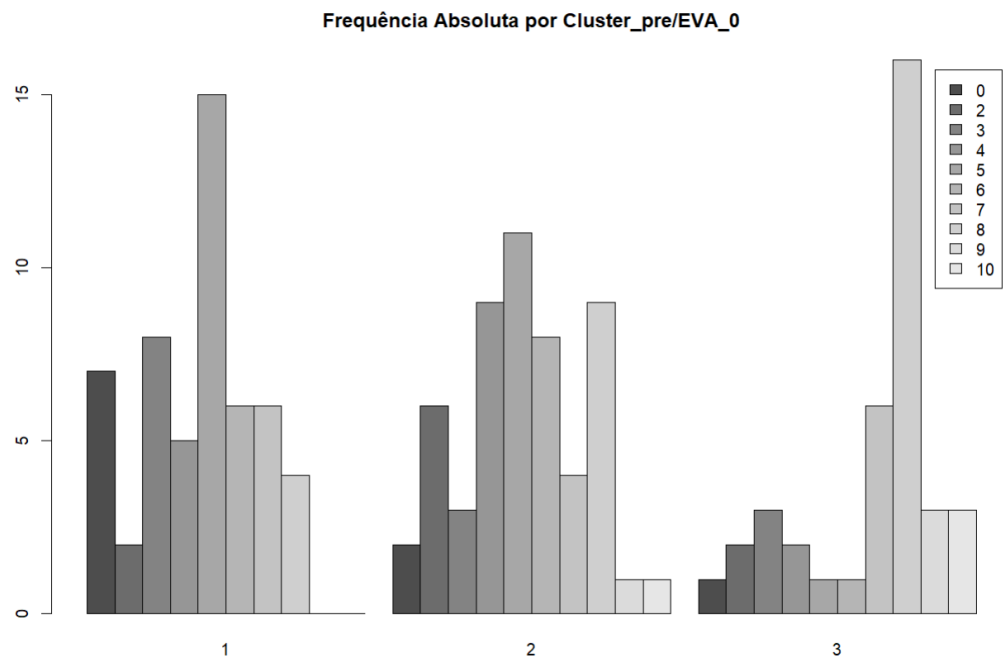


Gráfico 25 – EVA\_0 por Cluster Pré

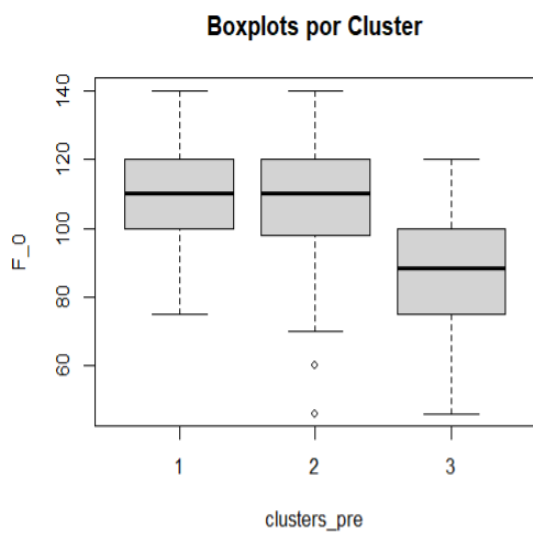


Gráfico 26 – BoxPlot F\_0 por Cluster Pré

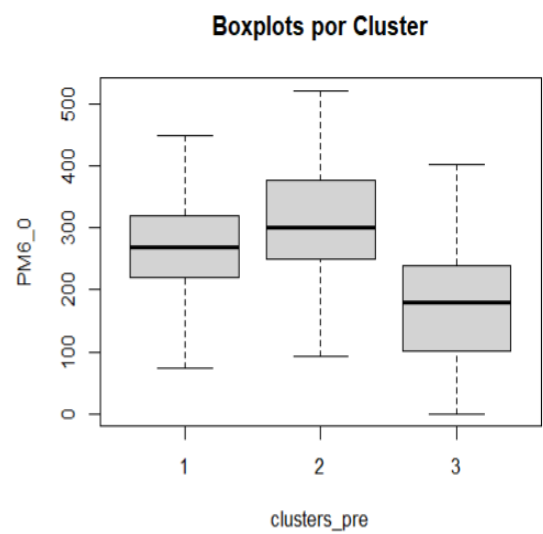


Gráfico 27 – BoxPlot F\_0 por Cluster Pré

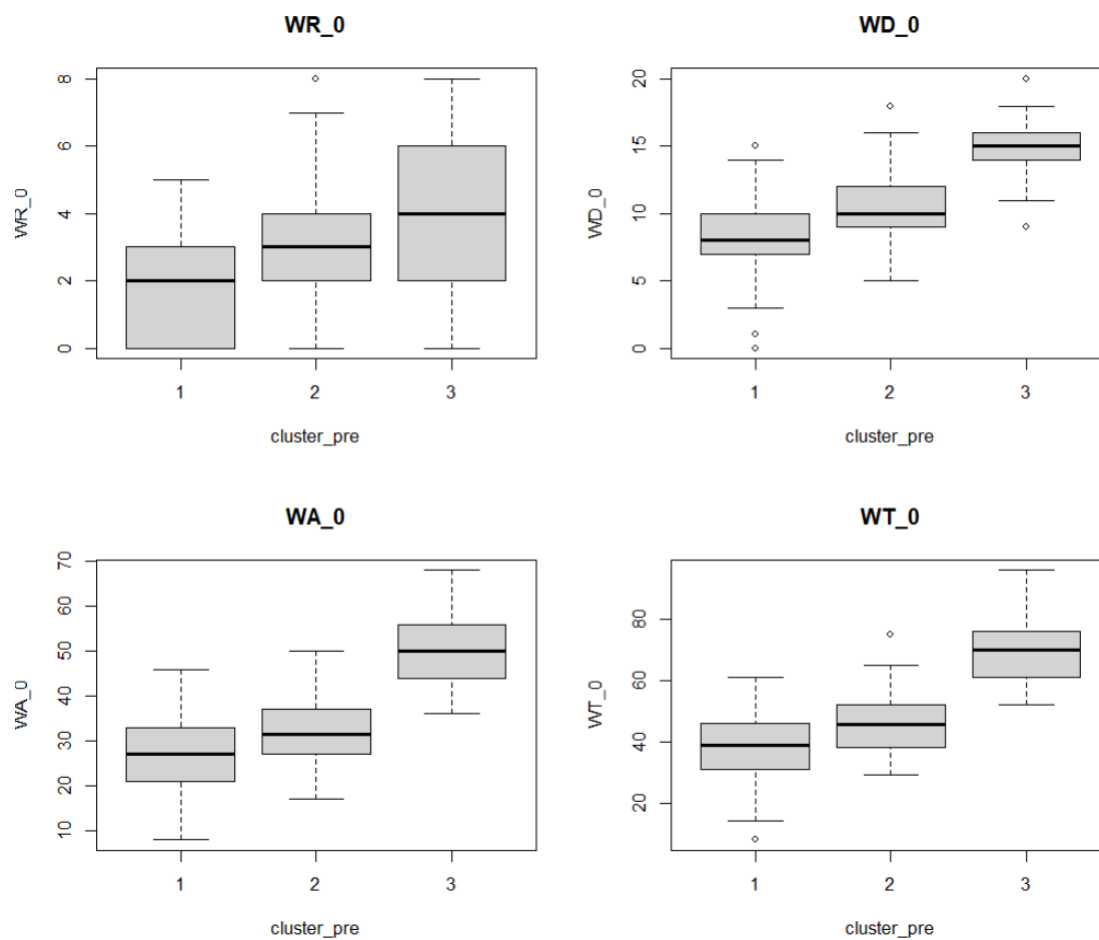


Gráfico 27 – BoxPlot WR\_0, WD\_0, WA\_0 e WT\_0 por Cluster Pré

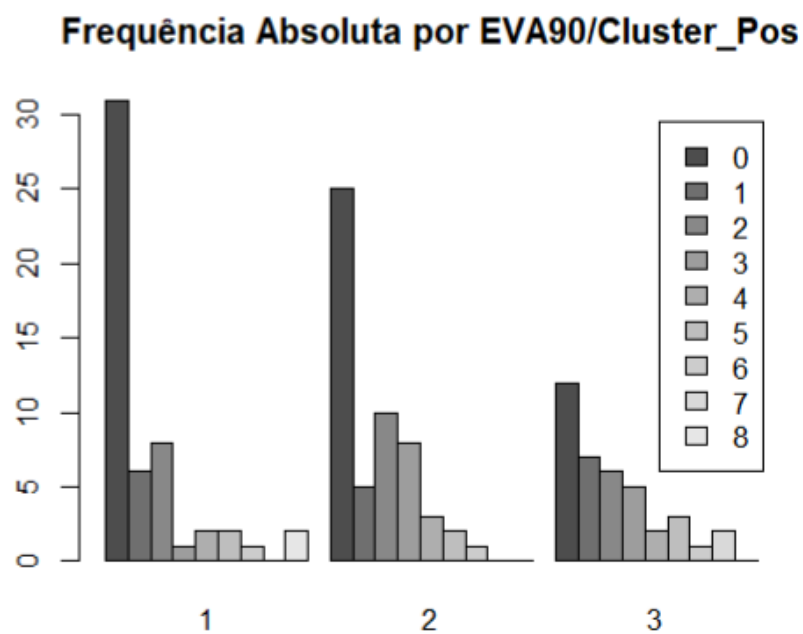


Gráfico 28 – Frequência Absoluta de EVA90/Cluster\_Pos

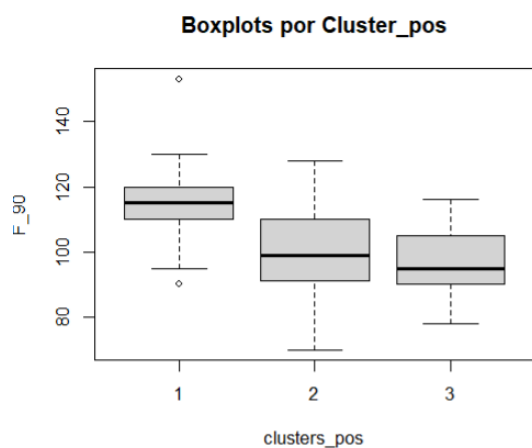


Gráfico 29 – BoxPlot F\_90 por Cluster Pós

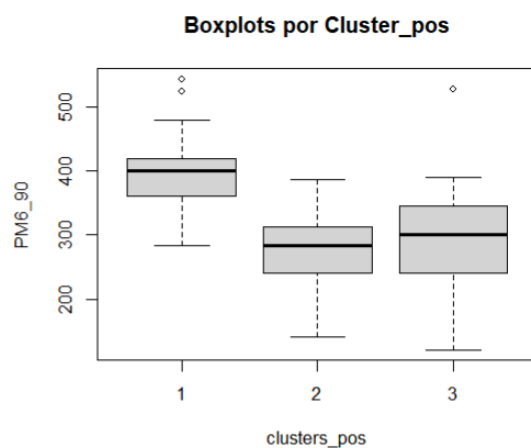


Gráfico 30 – BoxPlot PM\_90 por Cluster Pós

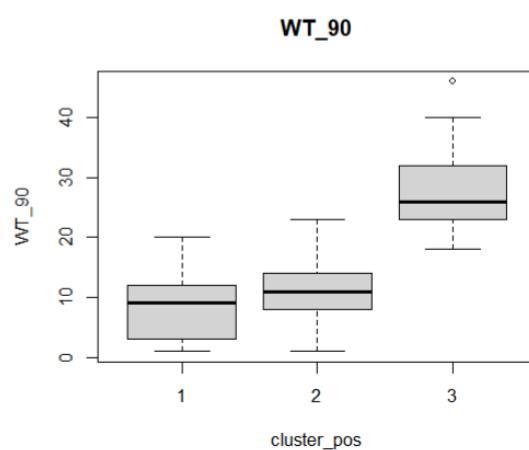
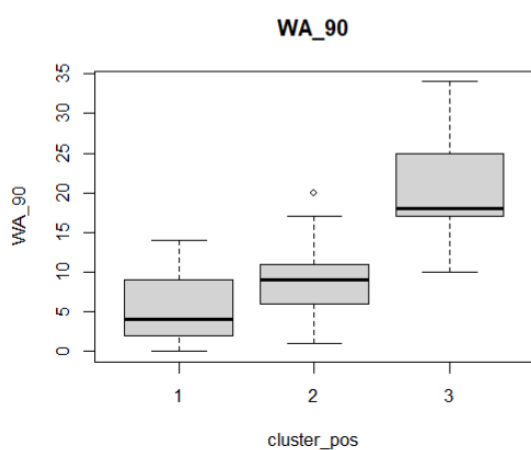
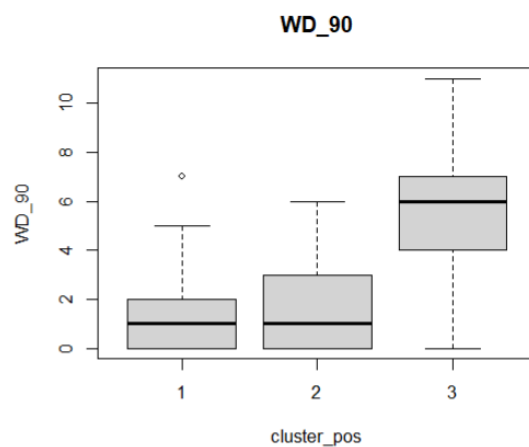
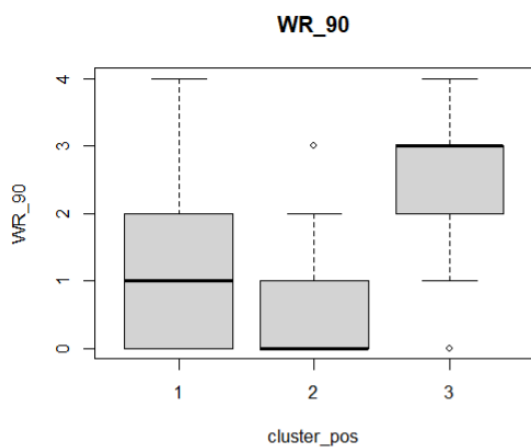


Gráfico 30 – BoxPlot WR\_90, WD\_90, WA\_90 e WT\_90\_90 por Cluster Pós

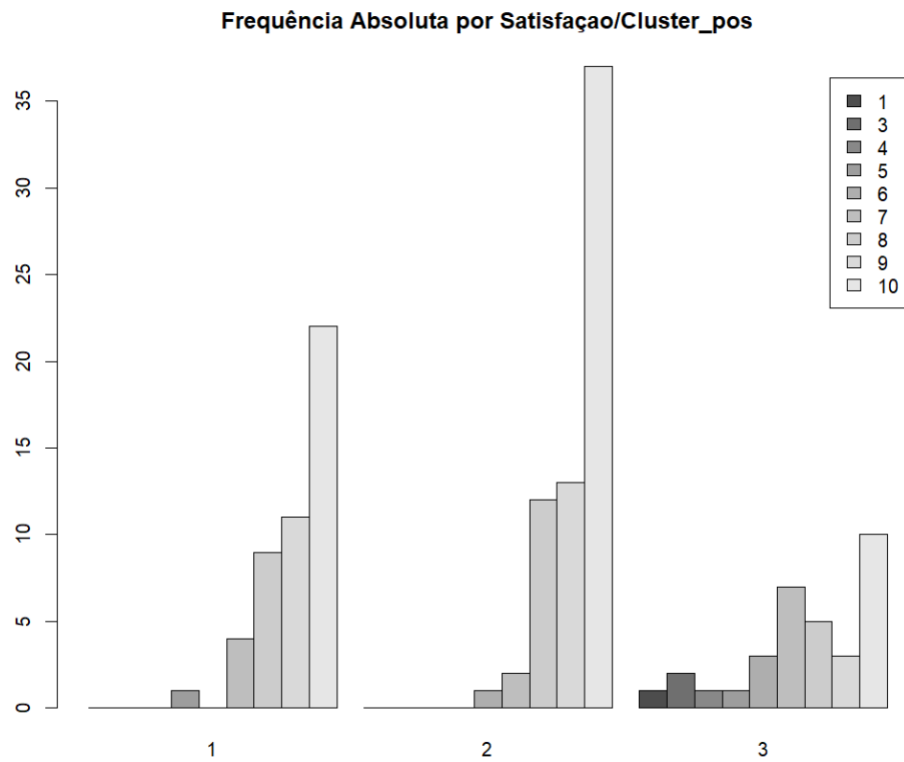


Gráfico 31 – Gráficos de Frequências Absolutas da Satisfação por Cluster\_Pós