



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO  
FEA - Faculdade de Economia e Administração  
Programa de Estudos Pós-Graduados em Administração

## **TRABALHO FINAL**

**“Análise Estatística da Pesquisa de Clima aplicada em operadores  
de uma empresa do segmento alimentício”**

**Juliana De Moura Jorge Graminho  
Mestrado em Administração de Empresas  
Disciplina de Métodos Quantitativos  
Prof. Dr. Arnaldo Hoyos**

**SÃO PAULO  
2011**

## 1. INTRODUÇÃO

O presente trabalho tem por objetivo analisar, com o apoio de ferramentas e modelos estatísticos, a pesquisa de clima organizacional aplicada em abril de 2011 em população de operadores de uma indústria do segmento alimentício situada na cidade de São Paulo. Para tal, iniciamos com o entendimento dos dados, incluindo a definição dos indivíduos e das variáveis, suas classificações em categóricas, discretas ou contínuas e os significados, além da apresentação da tabela de dados. Na seqüência, analisamos cada uma das variáveis separadamente quanto a sua forma de distribuição, os valores atípicos, medidas de centro e dispersão, a partir de gráficos (*pie chart*, barras, histogramas, gráficos de ramos, box-plot e gráficos resumo da estatística básica) e de medidas numéricas (média, mediana, quartis, desvio-padrão, variância, intervalo de confiança e teste de normalidade de Anderson-Darling). Também buscamos comparar as análises efetuadas para cada variável por subcategoria. Não será possível realizar a análise de séries temporais, já que se trata aplicação única da pesquisa, tal análise será possível a partir de futuras aplicações do mesmo instrumento nesta mesma população. Efetuou-se também uma **análise de regressões múltiplas** com dados extraídos, incluindo matriz de correlação. Utilizou-se também **análise dos componentes principais**, **análise dos clusters** e análise gráfica (Dendogramas). Em seguida passamos para a análise da comparação de médias e variância através da One-Way Analysis of Variance. Na seqüência, buscamos 2 tipos de regressão, utilizando a **regressão logística** e a **análise discriminante**. Além disso, a **análise de correspondências** e **árvore de classificação**. Por fim, buscamos avaliar qual regressão foi capaz de produzir um modelo mais simples (parcimonioso) que explique o comportamento da população em relação às variáveis pesquisadas. Ao final apresentam-se as considerações finais. Os softwares estatísticos utilizados são o **MINITAB** e **XLSTAT**.

## 2. ENTENDENDO OS DADOS

Para obtenção de dados foi aplicado o “Formulário Completo de Coleta de Dados” (anexo 1). Para este estudo foram selecionadas 9 perguntas agrupadas por similaridade em 6 temas conforme abaixo, para as quais trabalhou-se com os valores de mediana.

Pergunta	TEMA \ VARIÁVEL
Sou tratado pelo meu chefe com respeito e educação	Relação com chefia
Aqui todos os funcionários são ouvidos pela chefia	Relação com chefia
Não temos uma carga estressante de trabalho	Sem stress
Gosto da refeição que é servida na empresa	Refeição
Nesta empresa o bom profissional pode crescer e progredir	Chance de Progredir
Sou solicitado a ajudar a resolver problemas e propor soluções	Faço parte
Sinto que sou parte importante da empresa	Faço parte
Recebo orientação suficiente para executar minha função	Treinamento
O treinamento que recebo me prepara para o trabalho que faço	Treinamento

### 2.1. Os Indivíduos

Os indivíduos desta análise foram 104 indivíduos (operadores) de uma população total de 153 indivíduos, ou seja, obteve-se 70% de indivíduos respondentes, dos quais 90 foram pesquisas consideradas válidas (58,8% do total). As pesquisas consideradas não válidas foram aquelas para as quais algumas das variáveis categóricas (Escolaridade, Sexo, Faixa Etária, Tempo de

Empresa) foram deixadas em branco. A partir deste percentual de participação pode-se considerar a amostra estatisticamente significativa. Todos os indivíduos participantes da pesquisa possuem regime de contrato CLT no formato de “horistas”, com atividades essencialmente ligadas à produção. Além das respostas fornecidas na pesquisa não é possível identifica os indivíduos já que se trata de pesquisa anônima.

## 2.2. As Variáveis

As variáveis desta pesquisa estão descritas e detalhadas na Tabela abaixo.

Variável	Detalhe	Tipo	Unidade de Medida
Idade	Faixa etária	Variável Discreta	1 - menos de 25 anos 2 - 25 a 34 anos 3 - 35 a 44 anos 4 - 45 a 54 anos 5 - 55 ou mais
Sexo		Variável Categórica	f - Feminino m - Masculino
Turno	Corresponde ao turno no qual cada indivíduo pesquisado trabalha	Variável Categórica	a - Turno A b - Turno B c - Turno C d - Turno D adm - Turno Administrativo
Grau de escolaridade	Corresponde ao grau de escolaridade até onde cada indivíduo pesquisado cursou	Variável Categórica	1 - Fundamental 2 - Médio Incompleto 3 - Médio Completo 4 - Superior Incompleto 5 - Superior Completo 6 - Pós
Tempo de empresa	Corresponde à faixa de tempo trabalhado na empresa de cada indivíduo pesquisado	Variável Discreta	1. menos de 1 ano 2. 1 a 2 anos 3. 3 a 5 anos 4. 6 a 10 anos 5. 11 a 20 anos 6. mais de 20 anos
“Relação com chefia”	Corresponde à mediana dos resultados obtidos para as perguntas da pesquisa relacionadas com o tema “Relação com chefia”	Variável Quantitativa	1. Discordo totalmente 2. Discordo 3. nem concordo em discordo 4. Concordo 5. Concordo totalmente
“Sem stress”	Corresponde à mediana dos resultados obtidos para as perguntas da pesquisa relacionadas com o tema “Sem stress”	Variável Quantitativa	1. Discordo totalmente 2. Discordo 3. nem concordo em discordo 4. Concordo 5. Concordo totalmente
“Refeição”	Corresponde à mediana dos resultados obtidos para as perguntas da pesquisa relacionadas com o tema “Refeição”	Variável Quantitativa	1. Discordo totalmente 2. Discordo 3. nem concordo em discordo 4. Concordo 5. Concordo totalmente
“Chance de Progredir”	Corresponde à mediana dos resultados obtidos para as perguntas da pesquisa relacionadas com o tema “Chance de progredir”	Variável Quantitativa	1. Discordo totalmente 2. Discordo 3. nem concordo em discordo 4. Concordo 5. Concordo totalmente

“Faço parte”	Corresponde à mediana dos resultados obtidos para as perguntas da pesquisa relacionadas com o tema “Faço parte”	Variável Quantitativa	1. Discordo totalmente 2. Discordo 3. nem concordo em discordo 4. Concordo 5. Concordo totalmente
“Treinamento”	Corresponde à mediana dos resultados obtidos para as perguntas da pesquisa relacionadas com o tema “Treinamento”	Variável Quantitativa	1. Discordo totalmente 2. Discordo 3. nem concordo em discordo 4. Concordo 5. Concordo totalmente
“Mediana Total”	Corresponde à mediana de todos resultados entre as 6 variáveis estudadas “Relação com Líder”, “Sem stress”, “Refeição”, “Chance de Progredir”, “Faço parte” e “Treinamento”	Variável Quantitativa	1. Discordo totalmente 2. Discordo 3. nem concordo em discordo 4. Concordo 5. Concordo totalmente

### 2.3. A Tabela de dados

no pesq	Relac Líder	Sem stress	Refeição	Chance progredir	Faço parte	Treina-mento	Mediana	Tempo Empresa	Sexo	Turno	Idade	Escolar
1	4.5	3	1	1	2.5	4	2.75	4	m	b	2	3
2	5	5	2	5	3.5	1	4.25	1	m	b	1	5
3	5	5	4	1	2	4.5	4.25	4	m	b	2	3
4	5	4	3	4	4	4	4	3	m	b	1	3
5	3.5	2	2	2	3	2.5	2.25	1	m	b	2	3
6	4.5	4	4	5	5	4	4.25	2	m	b	4	5
7	3	4	1	2	3	5	3	3	m	b	2	3
8	3.5	2	1	1	2	2	2	2	m	b	2	3
10	3.5	4	2	4	4	4	4	5	m	b	3	3
11	3.5	2	2	3	3.5	3.5	3.25	2	m	b	3	3
13	3.5	1	1	1	1	1	1	5	m	b	3	3
14	5	2	2	2	3	4	2.5	2	m	b	1	4
15	4	5	3	5	4	4.5	4.25	4	m	b	3	3
17	4.5	1	3	4	4	3.5	3.75	4	m	b	3	3
18	4	4	3	2	3.5	4	3.75	2	m	b	1	3
19	4	5	1	4	5	5	4.5	3	m	b	2	3
20	3.5	4	4	4	3.5	4	4	5	m	b	4	5
21	3.5	4	3	2	3	3.5	3.25	3	m	b	4	3
23	4	3	5	3	4.5	4	4	2	m	b	2	3
24	3.5	5	1	1	1	3	2	4	m	b	3	3
25	3	5	1	1	5	1	2	5	m	c	3	3
26	3	1	2	1	3	2.5	2.25	1	m	c	2	3
27	4.5	4	1	4	3	5	4	3	m	c	2	3
28	4.5	3	4	3	4.5	3.5	3.75	5	m	c	4	3
29	1	1	1	1	3	4	1	4	m	c	3	3
30	5	4	4	4	4.5	5	4.25	3	m	c	2	3
31	3.5	3	4	2	3	2.5	3	2	m	c	1	3
32	1	2	1	3	2.5	4	2.25	2	m	adm	3	3
33	4	3	3	4	4	3.5	3.75	1	m	c	1	3
34	4.5	3	1	3	4.5	3.5	3.25	5	m	c	3	3
35	2	3	3	1	2	4	2.5	4	m	adm	3	3

36	3	1	1	3	2.5	3	2.75	2	f	c	2	3
37	1.5	1	2	1	1.5	4	1.5	3	m	adm	3	3
38	3.5	4	2	5	5	3	3.75	2	m	c	2	3
39	4.5	1	2	2	4	4	3	4	m	c	3	3
40	2.5	2	4	1	2.5	2	2.25	2	m	c	1	3
41	4	2	1	3	3.5	2.5	2.75	4	m	c	3	4
42	4	4	2	2	4	3.5	3.75	3	m	c	2	1
43	4.5	5	3	5	4	4.5	4.5	1	m	c	1	3
44	4	3	1	4	3.5	3.5	3.5	4	m	c	4	3
48	4	3	3	3	4	3.5	3.25	2	m	c	3	3
49	4.5	4	4	4	4.5	4	4	3	m	c	1	3
52	4	3	5	5	4	4.5	4.25	4	m	c	4	3
53	4.5	3	5	4	3.5	4.5	4.25	2	m	c	1	3
54	2.5	1	1	1	4	4	1.75	4	m	d	2	2
55	3.5	2	2	4	2	2	2	1	m	d	2	3
56	1	2	1	1	1	2	1	4	m	d	2	3
58	3.5	2	3	2	4	4	3.25	3	m	d	3	4
59	3.5	3	4	2	3	3.5	3.25	2	m	d	3	3
60	3	5	4	1	2.5	4.5	3.5	4	m	d	2	3
61	4	2	4	3	3.5	3.5	3.5	2	m	d	2	3
62	1	1	1	1	1	1	1	2	m	adm	1	3
64	3.5	2	2	4	5	4.5	3.75	5	m	d	3	3
65	4.5	3	2	4	5	5	4.25	5	m	d	3	2
67	3.5	3	4	3	4	4	3.75	4	m	d	3	3
68	1	3	3	1	3	3.5	3	1	f	d	3	6
69	2	1	4	1	2.5	1.5	1.75	2	m	d	1	3
70	4	4	4	5	5	4.5	4.25	5	m	d	4	3
72	1	1	1	1	1	3	1	5	m	b	3	2
74	4.5	2	2	4	4	4	4	2	m	d	3	3
75	4	3	2	4	4	4	4	2	m	adm	1	2
76	3	2	2	3	3	4	3	2	f	a	1	3
77	5	2	4	4	1	2	3	4	m	a	2	3
78	4	3	3	3	4	4	3.5	3	m	a	2	3
79	3	1	4	3	3.5	3	3	4	m	a	3	4
80	2	5	2	2	4.5	2.5	2.25	4	m	a	2	3
81	4.5	5	2	2	2.5	4	3.25	5	f	a	3	2
82	4.5	5	2	2	2.5	4	3.25	5	f	a	5	1
83	3	2	1	2	2	2.5	2	3	m	a	3	3
84	3	2	3	3	2	4	3	2	m	a	2	3
85	3	1	1	2	5	4.5	2.5	5	m	a	4	2
86	4.5	4	4	5	4.5	2.5	4.25	1	m	a	3	3
87	3	2	2	3	4	3.5	3	2	m	a	1	2
88	4.5	4	3	3	4	4	4	4	m	b	3	5
89	4.5	5	3	4	4	5	4.25	2	m	a	1	3
90	5	5	2	4	4	5	4.5	2	m	a	1	3
91	4.5	4	1	1	2	3	2.5	5	m	a	3	3
92	4	2	1	4	4.5	3.5	3.75	4	m	a	3	3
93	4.5	2	4	4	4	4	4	5	m	a	3	3
94	4.5	2	2	3	4	1.5	2.5	1	m	a	2	3
95	3	2	3	1	3	1.5	2.5	1	m	a	1	3
96	3.5	4	4	4	3	3.5	3.75	5	m	a	4	1
97	4.5	5	4	5	4.5	3	4.5	1	m	a	1	3
98	4.5	4	2	4	4.5	3	4	6	m	a	4	3

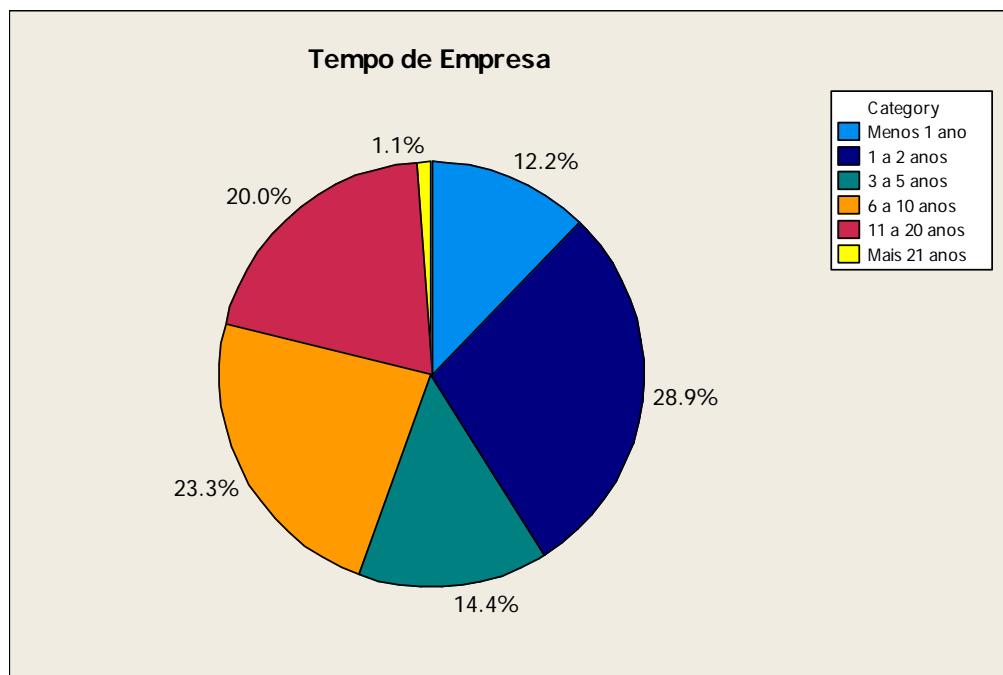
99	3.5	3	4	3	4	3	3.25	5	m	c	3	5
100	4.5	4	2	2	3.5	4	3.75	2	m	d	3	5
101	5	1	5	5	5	5	5	4	m	adm	3	3
102	5	5	1	5	5	5	5	5	m	adm	3	3
103	5	2	3	3	4	4	3.5	3	f	a	2	5
104	5	4	4	3	4	4.5	4	2	m	adm	1	4

Para os dados não informados das variáveis quantitativas adotamos o valor central “3” (não concordo, nem discordo) – que representa neutralidade quanto à concordância.

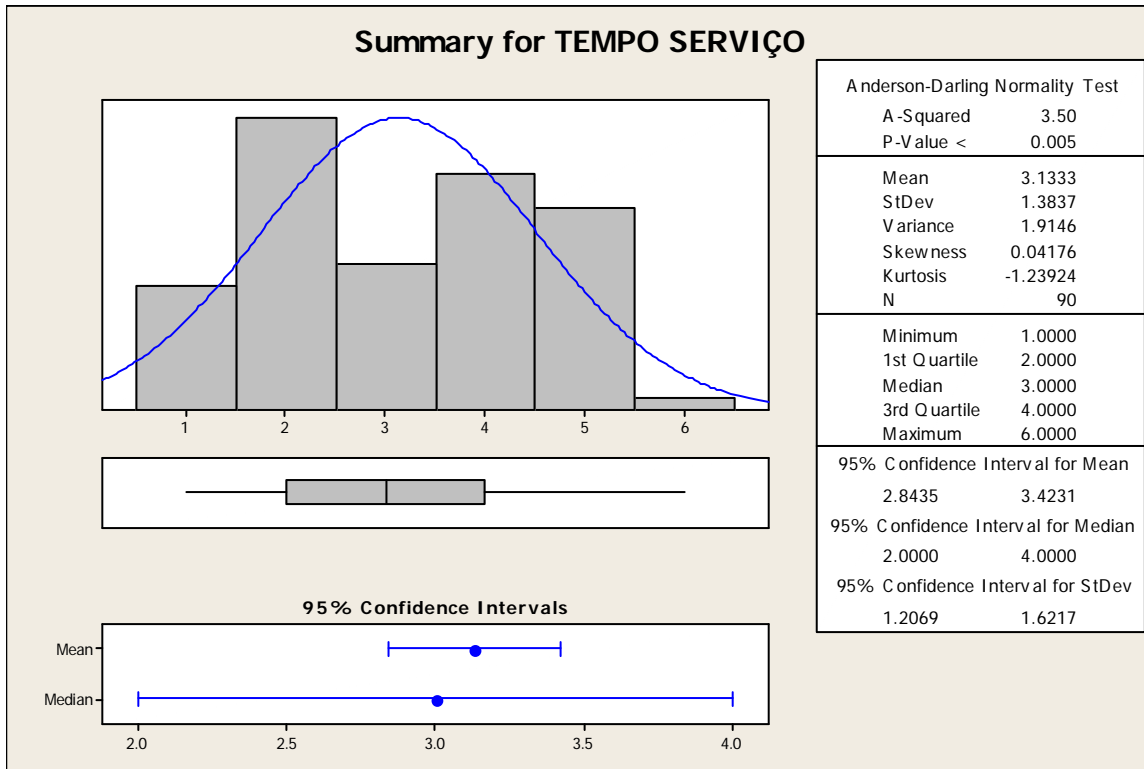
### 3. ANÁLISE EXPLORATÓRIA

#### 3.1. ANÁLISE DAS VARIÁVEIS CATEGÓRICAS E DISCRETAS

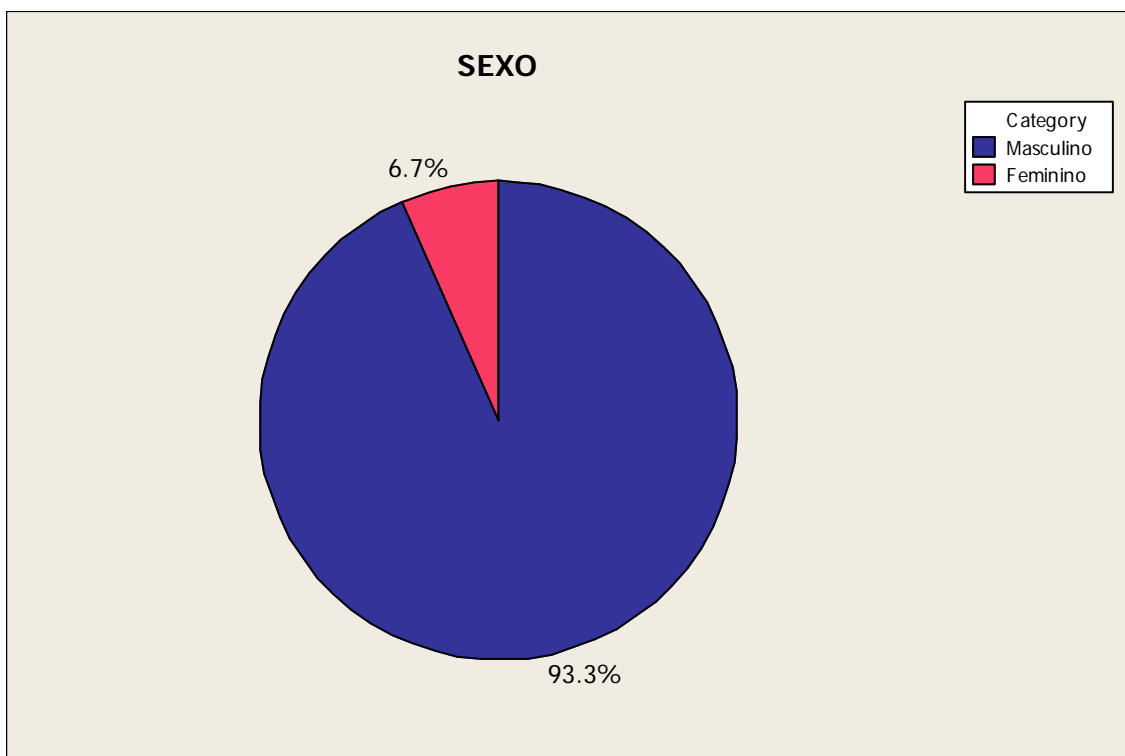
##### 3.1.1. Tempo de Empresa



Daqueles que informaram o tempo de empresa, observamos boa parte da população (41,1%) com menos de 2 anos de empresa, na contrapartida temos 44,4% da população com mais de 6 anos de empresa. O que denota dois grandes grupos existentes nesta população.

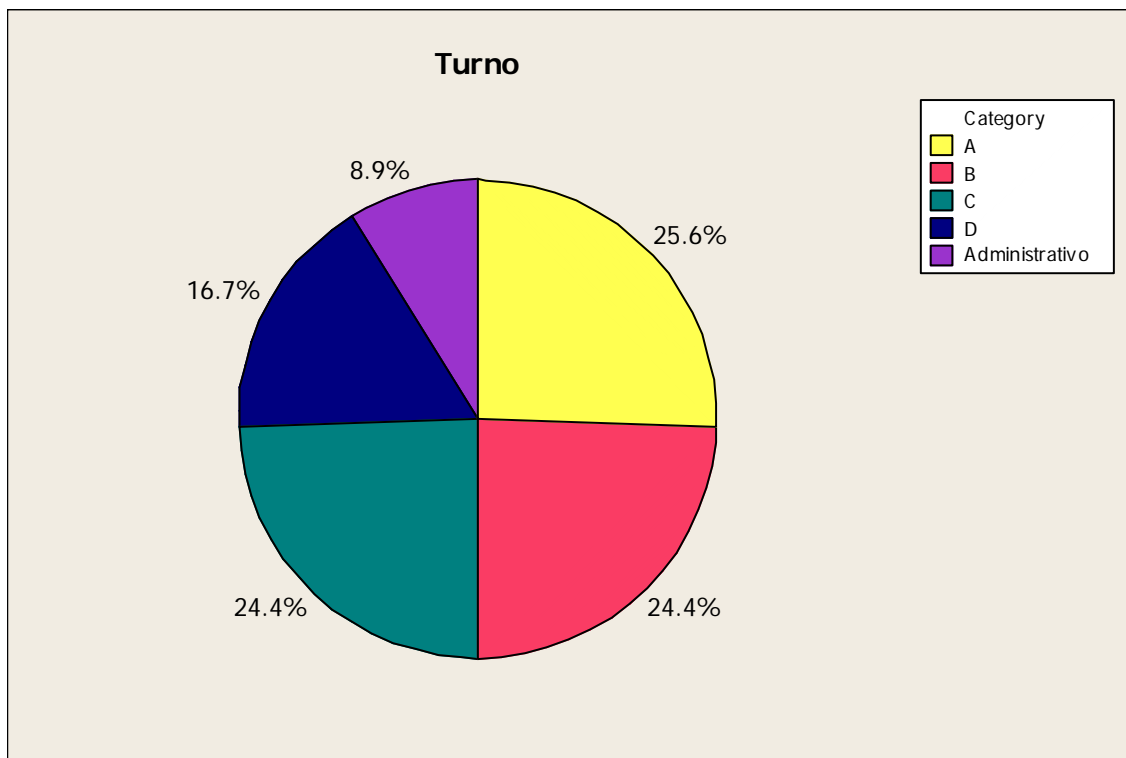


### 3.1.2. Sexo



Temos uma população com altíssima predominância de homens (93,3%).

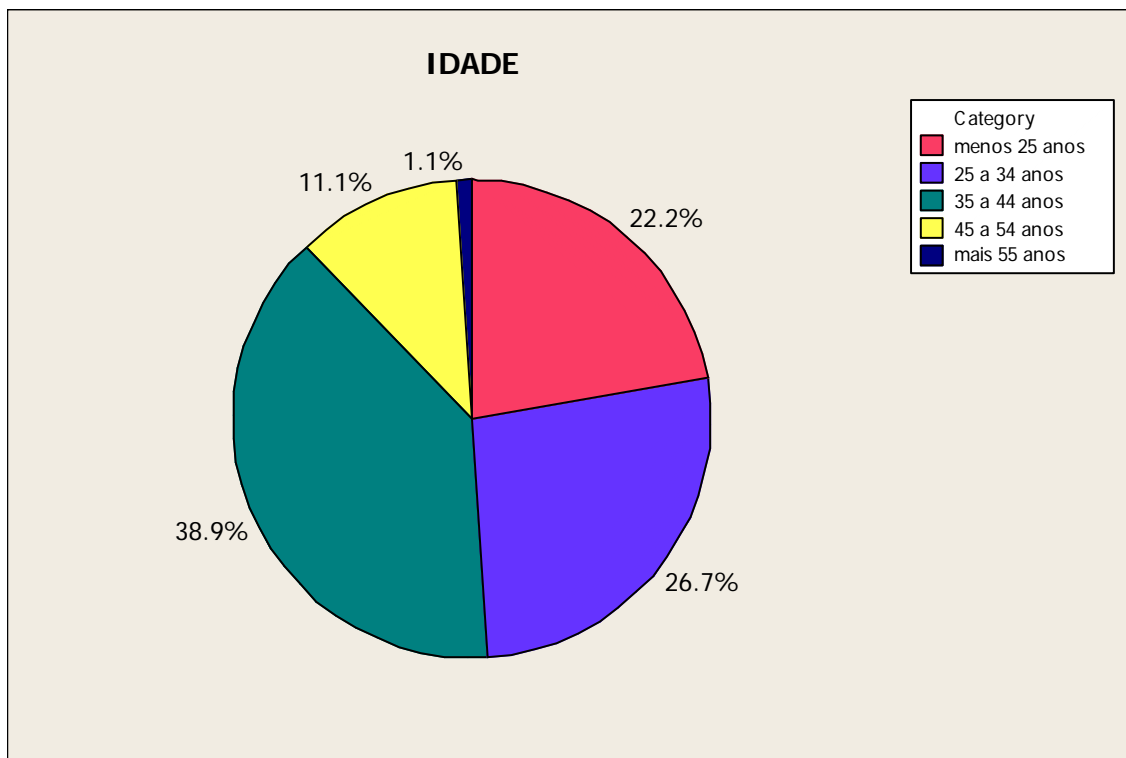
## 3.1.3. Turno de Trabalho



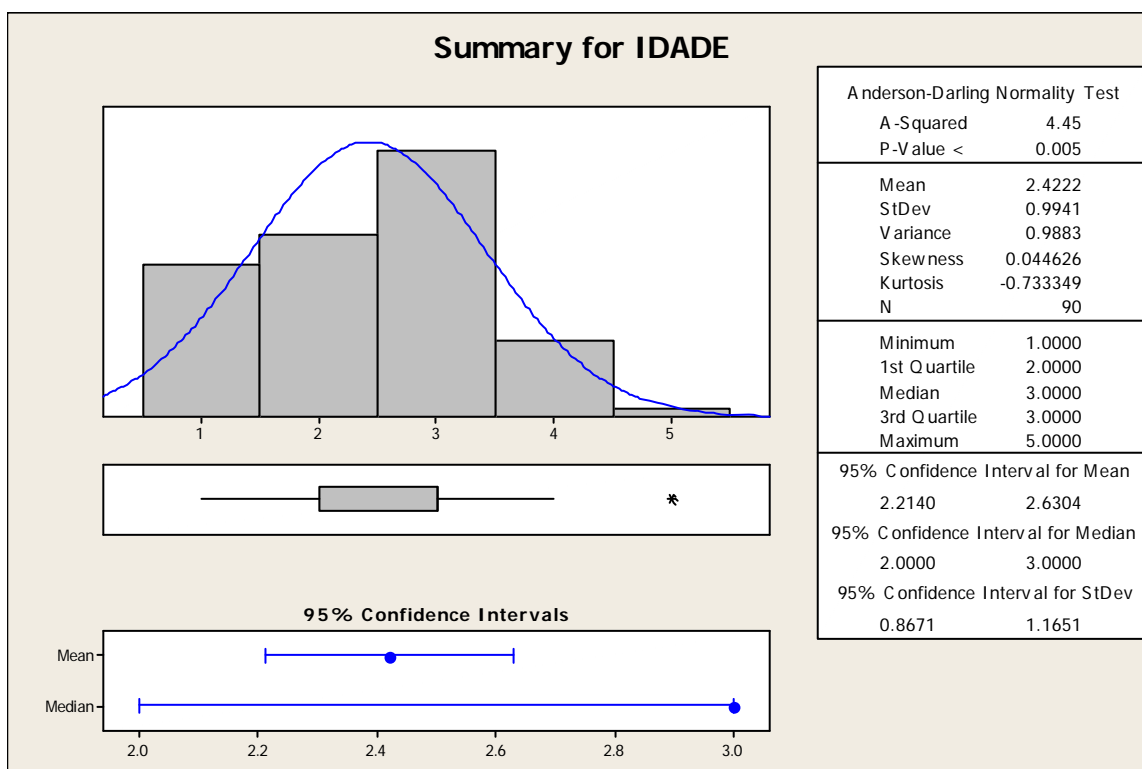
Temos uma participação equivalente nos turnos A, B e C. Chama atenção a menor participação do turno D, pois a população total é numericamente equivalente aos turnos A, B e C. A menor participação denota um menor engajamento daquela equipe e pode significar um pior clima no turno B em relação aos demais. O turno Administrativo possui realmente menor quantidade de indivíduos.



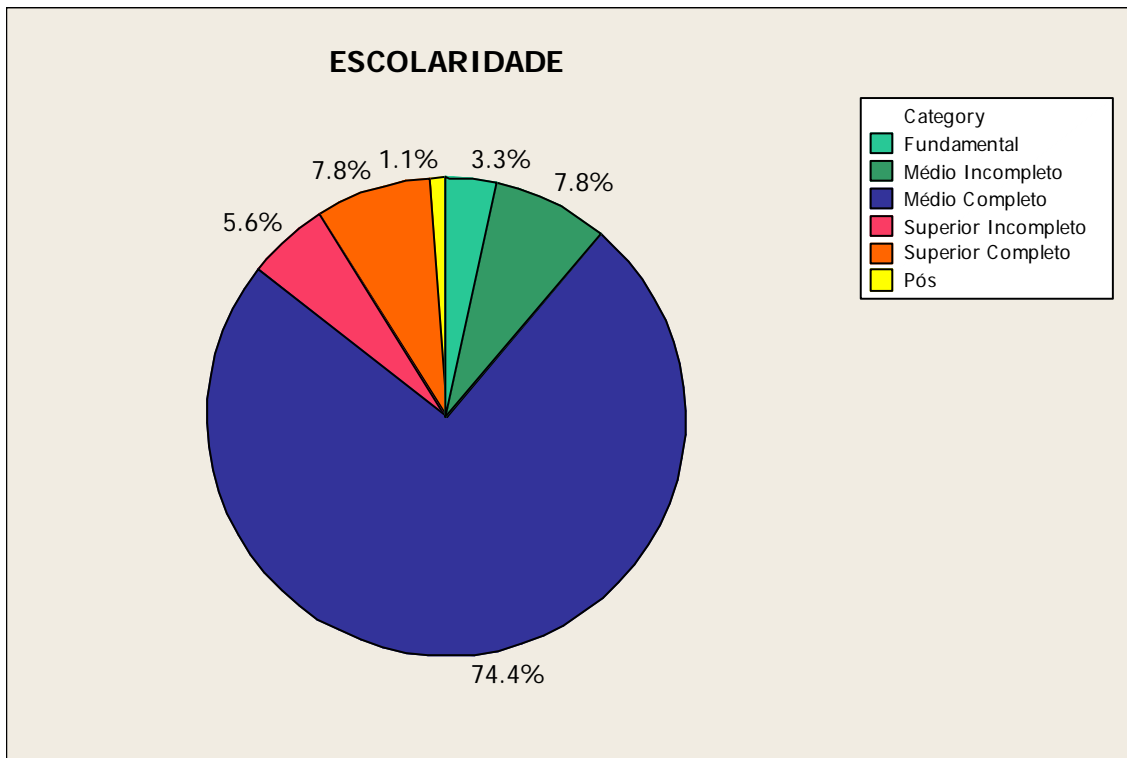
## 3.1.4. Faixa Etária



A grande maioria (65,6%) dos indivíduos pesquisados que informaram sua faixa etária possui idade entre 25 e 44 anos.



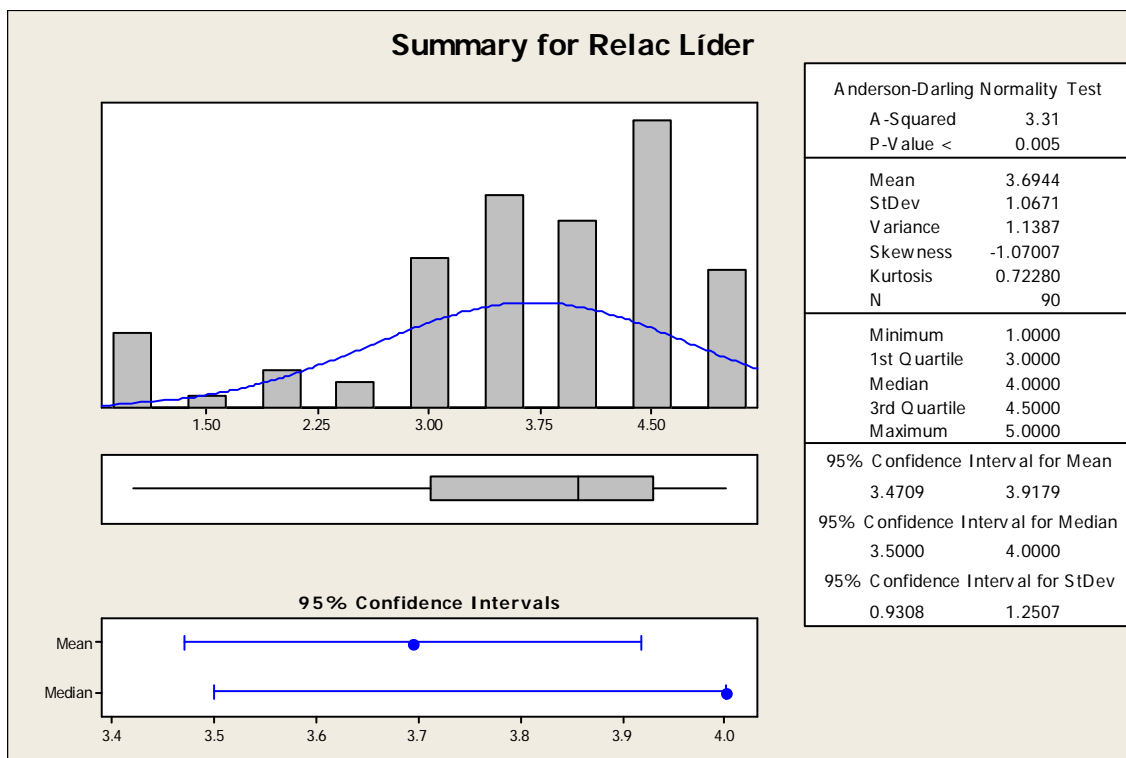
## 3.1.5. Escolaridade



A grande maioria (74,4%) dos indivíduos pesquisados possui ensino médio completo.

## 3.2. ANÁLISE DAS VARIÁVEIS QUANTITATIVAS

### 3.2.1. Variável “Relação com Líder”



Os dados da variável “Relação com Líder” referem-se ao ano de 2011.

Quanto à forma: A distribuição encontra-se assimétrica positiva levemente à esquerda, com coeficiente de assimetria igual à  $-1,07007$ .

Quanto ao centro e dispersão: Pelo histograma, observa-se que na metade dos valores encontra-se acima de 4,00 (mediana) e a outra metade abaixo desse valor. Verifica-se uma certa distância entre a média (3,6944) e a mediana (4,0), o que significa que existem algumas distorções da média.

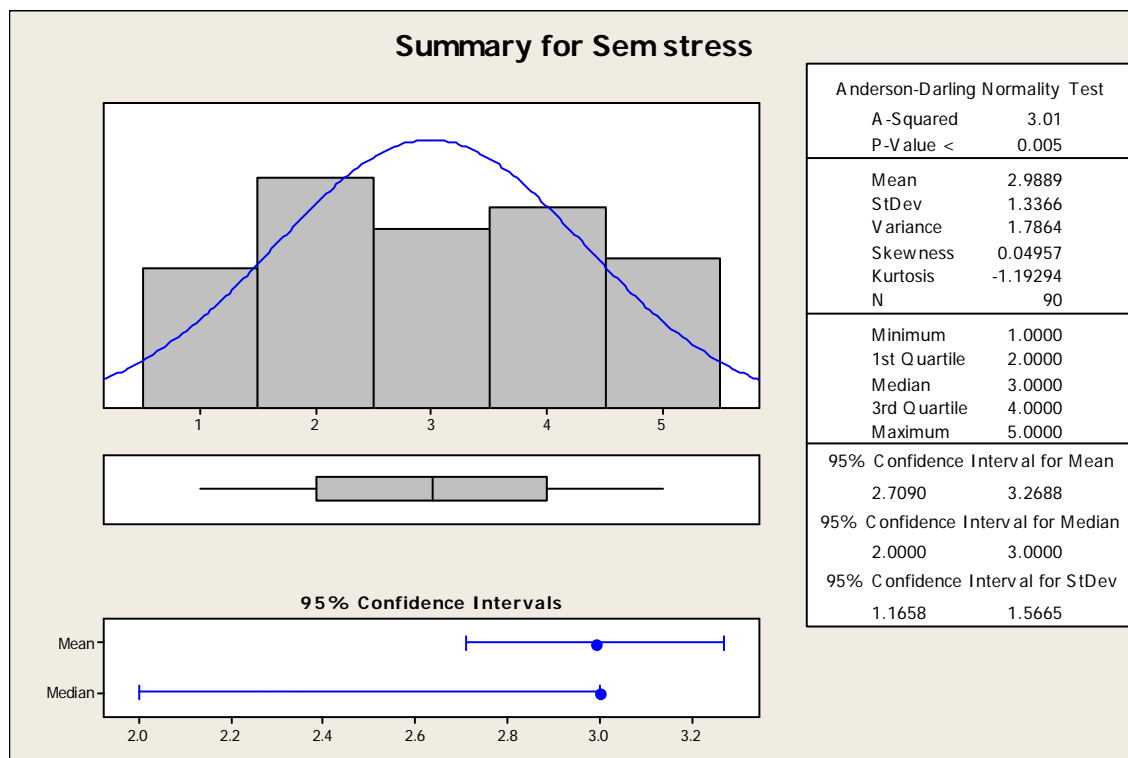
Quanto à Dispersão: esta variável apresenta amplitude máxima dos valores respondidos, já que o menor valor identificado é 1,0 e o maior valor é 5,0.

Considerando a probabilidade de ocorrência, pode-se dizer que com 95% de confiança existe a probabilidade de uma observação assumir um valor médio entre 3,4709 e 3,9179.

Quanto ao desvio padrão, que mede o grau de dispersão em torno do valor médio (3,6944), podemos indicar um grau de dispersão forte, pois o desvio padrão (1,0671) é maior que 10%.

Pelos quartis identificamos que 25 % dos indivíduos estudados responderam num valor abaixo de 3,0 e outros 25% acima 4,5, sendo que 50% das respostas dadas encontram-se dentro dessa faixa.

### 3.2.2. Variável “Sem stress”



Os dados da variável “Sem stress” referem-se ao ano de 2011.

Quanto à forma: A distribuição encontra-se assimétrica positiva levemente à direita, com coeficiente de assimetria igual à 0,04957.

Quanto ao centro e dispersão: Pelo histograma, observa-se que na metade dos valores encontra-se acima de 3,00 (mediana) e a outra metade abaixo desse valor. Verifica-se bastante proximidade entre a média (2,9889) e a mediana (3,0), o que significa que existem pouquíssimas distorções da média.

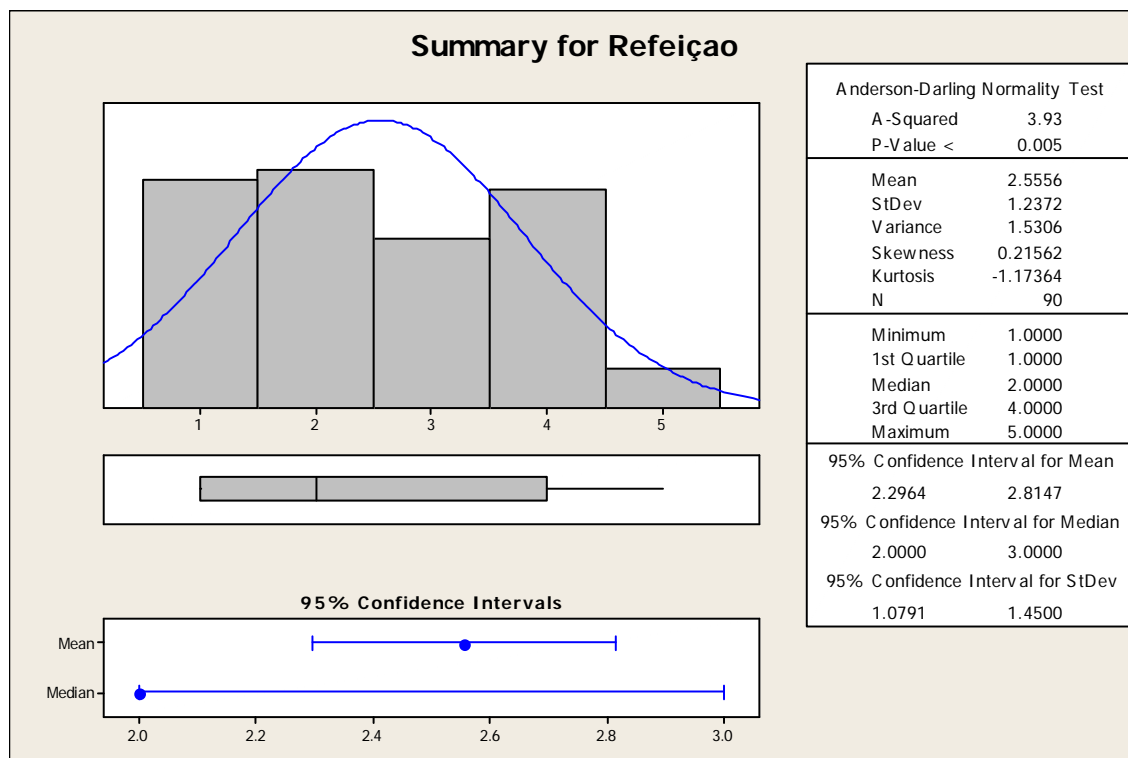
Quanto à Dispersão: esta variável apresenta amplitude máxima dos valores respondidos, já que o menor valor identificado é 1,0 e o maior valor é 5,0.

Considerando a probabilidade de ocorrência, pode-se dizer que com 95% de confiança existe a probabilidade de uma observação assumir um valor médio entre 2,7090 e 3,2688.

Quanto ao desvio padrão, que mede o grau de dispersão em torno do valor médio (2,9889), podemos indicar um grau de dispersão forte, pois o desvio padrão (1,3140) é maior que 10%.

Pelos quartis identificamos que 25 % dos indivíduos estudados responderam num valor abaixo de 2,0 e outros 25% acima de 4,0, sendo que 50% das respostas dadas encontra-se dentro dessa faixa.

### 3.2.3. Variável “Refeição”



Os dados da variável “Refeição” referem-se ao ano de 2011.

Quanto à forma: A distribuição encontra-se assimétrica positiva levemente à direita, com coeficiente de assimetria igual à 0,21562.

Quanto ao centro e dispersão: Pelo histograma, observa-se que na metade dos valores encontra-se acima de 2,00 (mediana) e a outra metade abaixo desse valor. Verifica-se pouca proximidade entre a média (2,5556) e a mediana (2,0), o que significa que existem distorções da média.

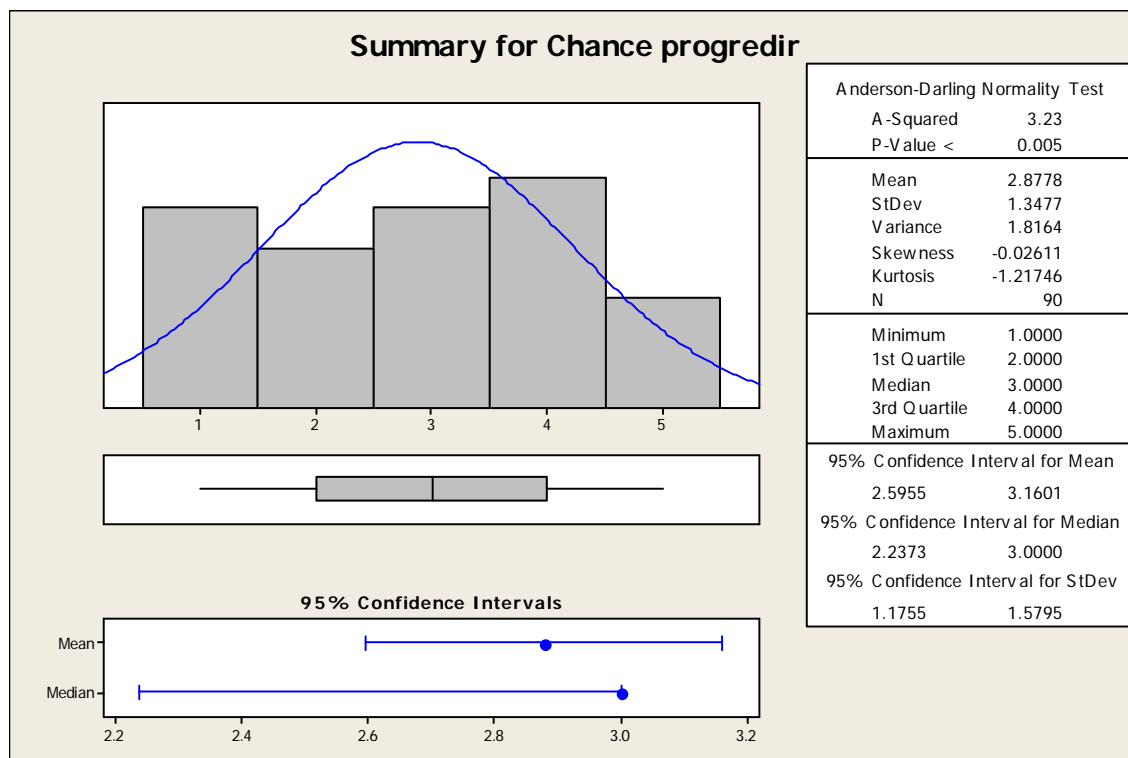
Quanto à Dispersão: esta variável apresenta amplitude máxima dos valores respondidos, já que o menor valor identificado é 1,0 e o maior valor é 5,0.

Considerando a probabilidade de ocorrência, pode-se dizer que com 95% de confiança existe a probabilidade de uma observação assumir um valor médio entre 2,2964 e 2,8147.

Quanto ao desvio padrão, que mede o grau de dispersão em torno do valor médio (2,5556), podemos indicar um grau de dispersão forte, pois o desvio padrão (1,2372) é bem maior que 10%.

Pelos quartis identificamos que 25 % dos indivíduos estudados responderam num valor abaixo de 1,0 e outros 25% acima 4,0 sendo que 50% das respostas dadas encontram-se dentro dessa faixa.

### 3.2.4. Variável “Chance de progredir”



Os dados da variável “Chance de Progredir” referem-se ao ano de 2011.

Quanto à forma: A distribuição encontra-se assimétrica positiva levemente à esquerda, com coeficiente de assimetria igual à -0,002611.

Quanto ao centro e dispersão: Pelo histograma, observa-se que na metade dos valores encontra-se acima de 3,00 (mediana) e a outra metade abaixo desse valor. Verifica-se alguma proximidade entre a média (2,8778) e a mediana (3,0), o que significa que existem poucas distorções da média.

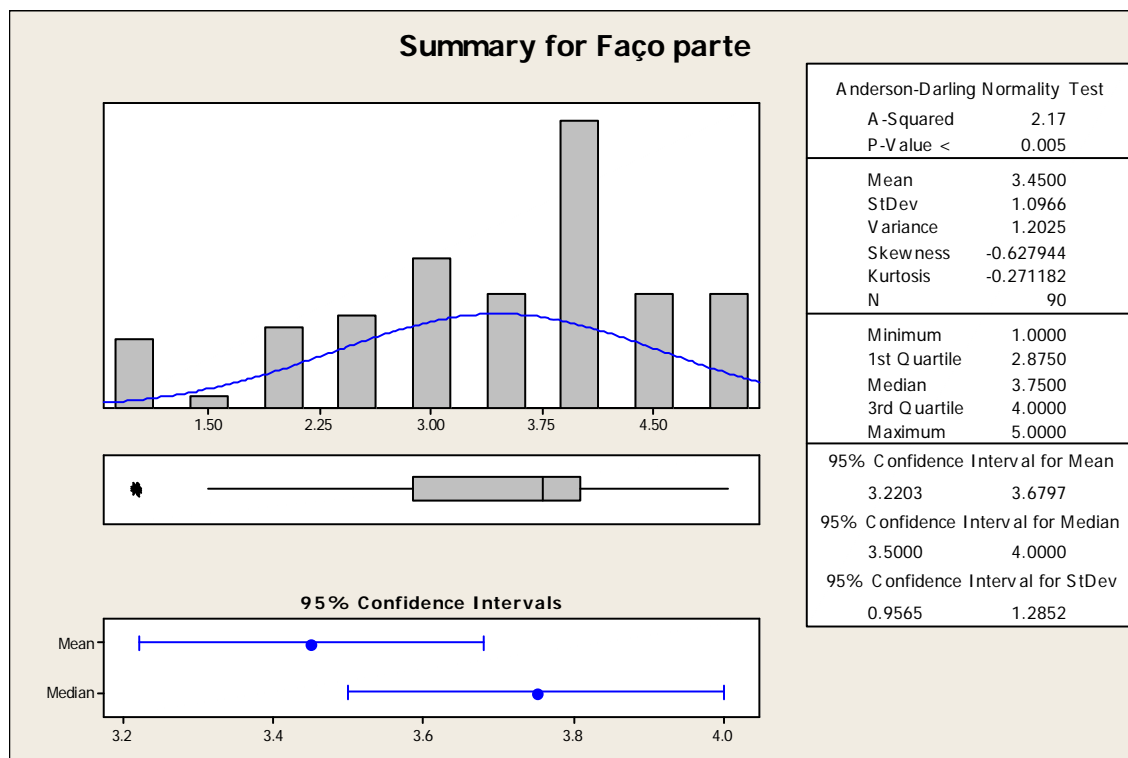
Quanto à Dispersão: esta variável apresenta amplitude máxima dos valores respondidos, já que o menor valor identificado é 1,0 e o maior valor é 5,0.

Considerando a probabilidade de ocorrência, pode-se dizer que com 95% de confiança existe a probabilidade de uma observação assumir um valor médio entre 2,5955 e 3,1601.

Quanto ao desvio padrão, que mede o grau de dispersão em torno do valor médio (2,8778), podemos indicar um grau de dispersão forte, pois o desvio padrão (1,3477) é maior que 10%.

Pelos quartis identificamos que 25 % dos indivíduos estudados responderam num valor abaixo de 2,0 e outros 25% acima 4,0 sendo que 50% das respostas dadas encontram-se dentro dessa faixa.

## 3.2.5. Variável “Faço parte”



Os dados da variável “Faço parte” referem-se ao ano de 2011.

Quanto à forma: A distribuição encontra-se assimétrica positiva à esquerda, com coeficiente de assimetria igual à  $-0,627944$ .

Quanto ao centro e dispersão: Pelo histograma, observa-se que na metade dos valores encontra-se acima de 3,7500 (mediana) e a outra metade abaixo desse valor. Verifica-se alguma proximidade entre a média (3,4500) e a mediana (3,7500), o que significa que existem poucas distorções da média.

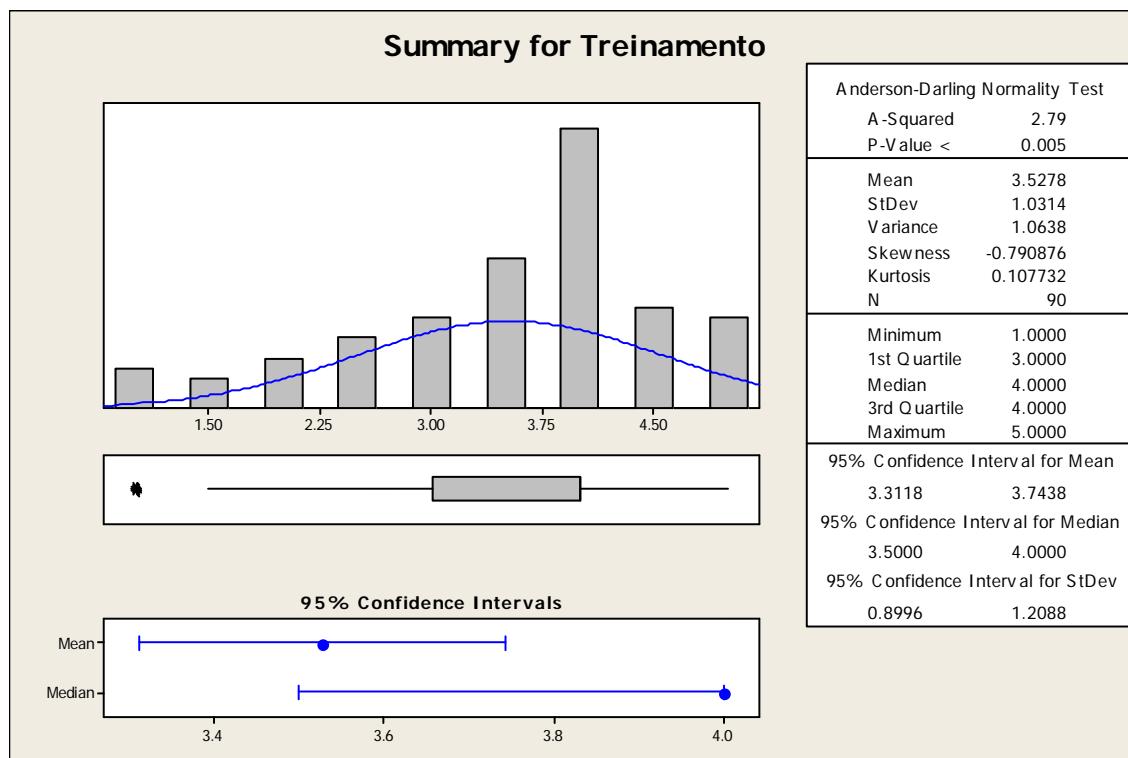
Quanto à Dispersão: esta variável apresenta amplitude máxima dos valores respondidos, já que o menor valor identificado é 1,0 e o maior valor é 5,0.

Considerando a probabilidade de ocorrência, pode-se dizer que com 95% de confiança existe a probabilidade de uma observação assumir um valor médio entre 3,2203 e 3,6797

Quanto ao desvio padrão, que mede o grau de dispersão em torno do valor médio (3,4500), podemos indicar um grau de dispersão forte, pois o desvio padrão (1,0966) é maior que 10%.

Pelos quartis identificamos que 25 % dos indivíduos estudados responderam num valor abaixo de 2,8750 e outros 25% acima 4,0 sendo que 50% das respostas dadas encontram-se dentro dessa faixa.

### 3.2.6. Variável “Treinamento”



Os dados da variável “Treinamento” referem-se ao ano de 2011.

Quanto à forma: A distribuição encontra-se assimétrica positiva à esquerda, com coeficiente de assimetria igual à  $-0,790876$ .

Quanto ao centro e dispersão: Pelo histograma, observa-se que na metade dos valores encontra-se acima de 4,0 (mediana) e a outra metade abaixo desse valor. Verifica-se alguma proximidade entre a média (3,5278) e a mediana (4,0), o que significa que existem poucas distorções da média.

Quanto à Dispersão: esta variável apresenta amplitude máxima dos valores respondidos, já que o menor valor identificado é 1,0 e o maior valor é 5,0.

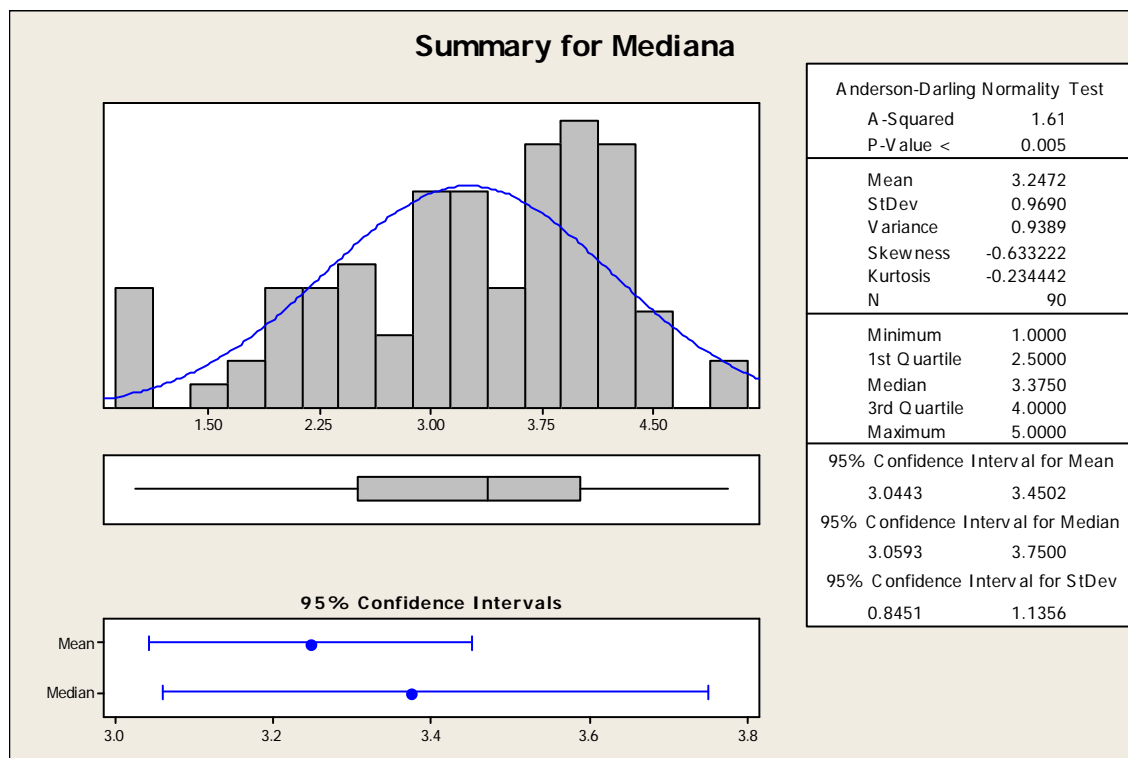
Considerando a probabilidade de ocorrência, pode-se dizer que com 95% de confiança existe a probabilidade de uma observação assumir um valor médio entre 3,3118 e 3,7438.

Quanto ao desvio padrão, que mede o grau de dispersão em torno do valor médio (3,5278), podemos indicar um grau de dispersão forte, pois o desvio padrão (1,0314) é maior que 10%.

Pelos quartis identificamos que 25 % dos indivíduos estudados responderam num valor abaixo de 3,0 e outros 25% acima 4,0 sendo que 50% das respostas dadas encontram-se dentro dessa faixa.



## 3.2.7. Variável “Mediana Total”



Os dados da variável “Mediana Total” referem-se ao ano de 2011.

Quanto à forma: A distribuição encontra-se assimétrica positiva à esquerda, com coeficiente de assimetria igual à -0,633222.

Quanto ao centro e dispersão: Pelo histograma, observa-se que na metade dos valores encontra-se acima de 3,3750 (mediana) e a outra metade abaixo desse valor. Verifica-se alguma proximidade entre a média (3,2472) e a mediana (3,3750), o que significa que existem poucas distorções da média.

Quanto à Dispersão: esta variável apresenta amplitude máxima dos valores respondidos, já que o menor valor identificado é 1,0 e o maior valor é 5,0.

Considerando a probabilidade de ocorrência, pode-se dizer que com 95% de confiança existe a probabilidade de uma observação assumir um valor médio entre 3,0443 e 3,4502.

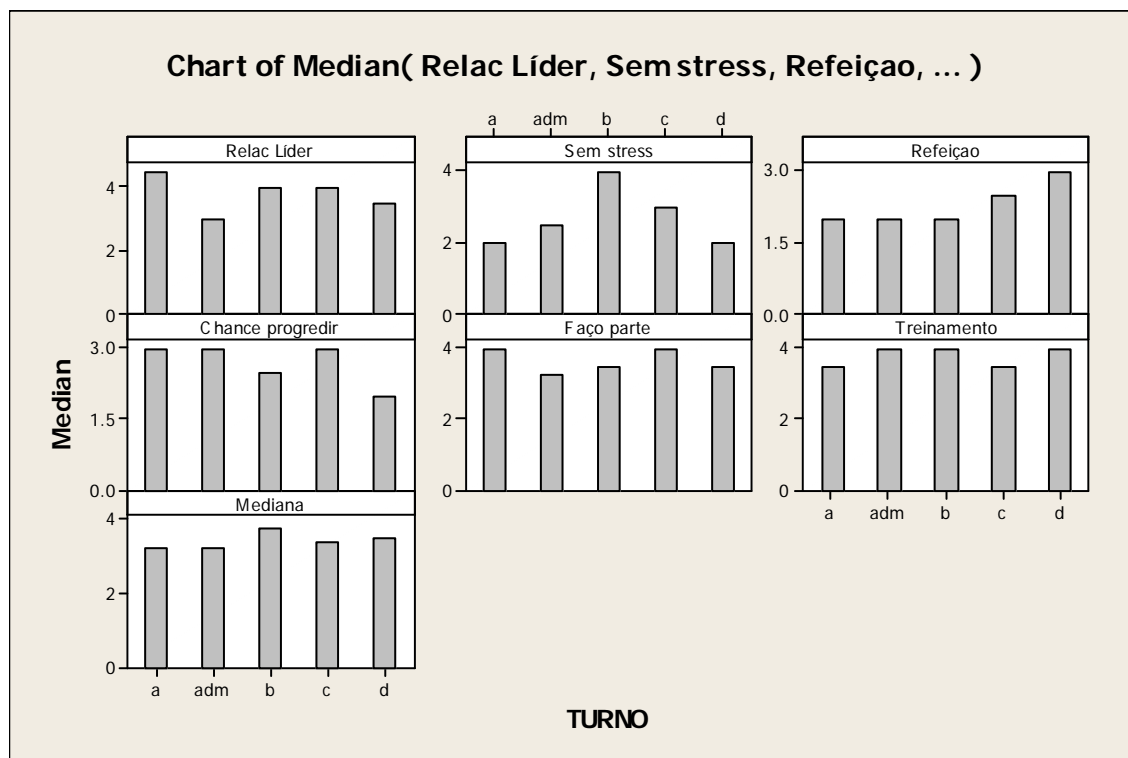
Quanto ao desvio padrão, que mede o grau de dispersão em torno do valor médio (3,2472), podemos indicar um grau de dispersão forte, pois o desvio padrão (0,9690) é maior que 10%.

Pelos quartis identificamos que 25 % dos indivíduos estudados responderam num valor abaixo de 2,5 e outros 25% acima 4,0 sendo que 50% das respostas dadas encontram-se dentro dessa faixa.

### 3.3. Análise das Variáveis em Subcategorias

#### 3.3.1. Mediana por turno

A análise percentual entre medianas por turnos permite comparar o resultado de cada variável estudada, conforme abaixo. Chamam atenção o baixo resultado de “Relação com Líder” para o turno administrativo e os baixos resultados de “Sem stress” e para o turno A e D. Além disso, o baixo valor de “Chance de Progredir” para o turno D. A melhor mediana de “Relação com Líder” aparece no turno A, para “Sem stress” no turno B, para “Refeição” no turno D. A variável “Mediana Total” que consolida todas as variáveis tem a melhor mediana no turno B.



Legenda eixo X

a - Turno A  
 b - Turno B  
 d - Turno C  
 d - Turno D  
 adm - Turno Administrativo

#### 3.3.2. Mediana por Tempo de Empresa

A análise percentual e comparativa entre medianas por tempo de empresa demonstra os valores melhores para “Relação com Líder” em funcionários com mais de 20 anos de empresa.

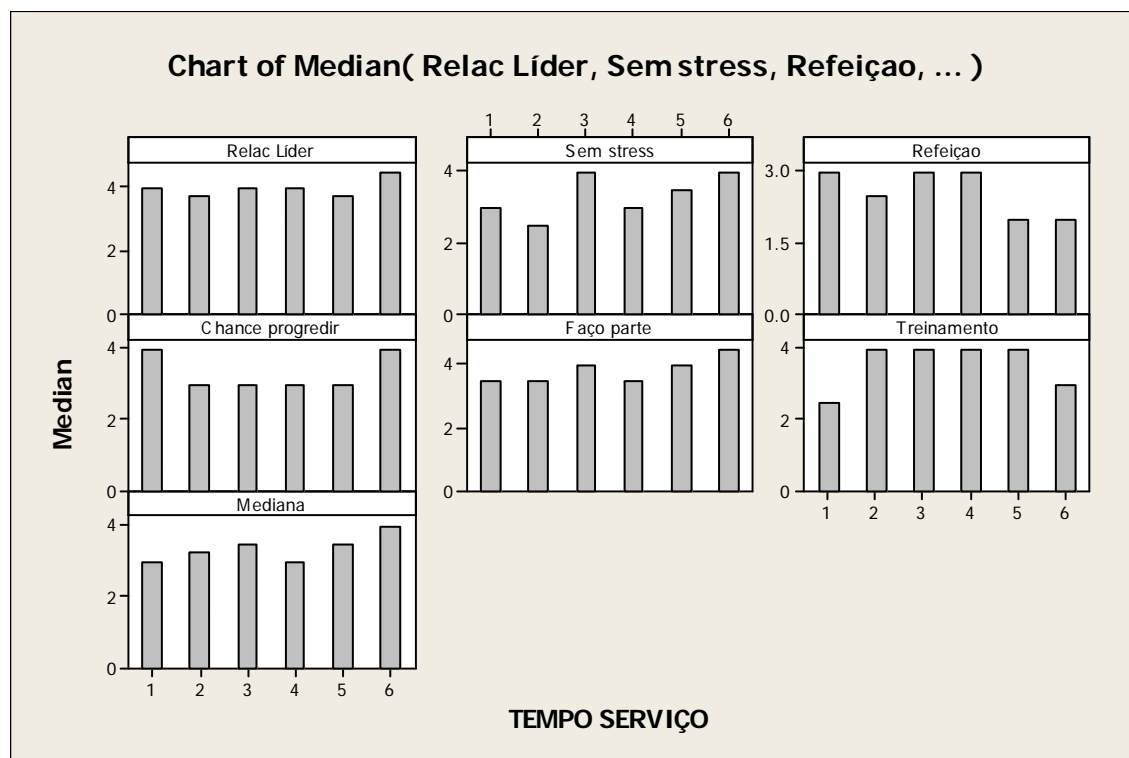
A mediana da variável “Sem stress” apresenta seus maiores valores para funcionários entre 3 e 5 anos de empresa e também para aqueles que possuem mais de 20 anos. Os piores valores para esta variável encontra-se nos funcionários que possuem 1 a 2 anos de empresa.

A mediana da variável “Refeição” é pior para funcionários mais antigos (11 a 20 anos e mais que 20 anos). Os demais, com exceção de funcionários entre 1 e 2 anos, possuem valores altos e equivalentes.

Os valores de mediana da variável “Chance de Progredir” aparece destacadamente melhores para funcionários com menos de 1 ano e para aqueles com mais de 20 anos.

A mediana da variável “Faço Parte” possui seu melhor valor para funcionários com mais de 20 anos de empresa.

Por fim, os valores de mediana da variável “Treinamento” apresenta os piores valores para funcionários com menos de 1 ano de empresa, seguidos por aqueles que estão com mais de 20 anos de empresa. A variável “Mediana Total” que consolida todas as variáveis tem os melhores valores para funcionários com mais de 20 anos de empresa.



Legenda eixo X

1. menos de 1 ano
2. 1 a 2 anos
3. 3 a 5 anos
4. 6 a 10 anos
5. 11 a 20 anos
6. mais de 20 anos

### 3.3.3. Mediana por Escolaridade

A análise percentual e comparativa entre medianas por escolaridade demonstra que para a variável “Relação com Líder” os valores crescem conforme aumenta a escolaridade, exceto aqueles em funcionários pós graduação que possuem os piores valores para esta variável e para aqueles com apenas Fundamental que já possuem valores mais altos.

A mediana da variável “Sem stress” apresenta seus maiores valores para “Fundamental” e para “Superior Completo”. Os piores valores para esta variável encontram-se nos funcionários superior incompleto e médio incompleto.

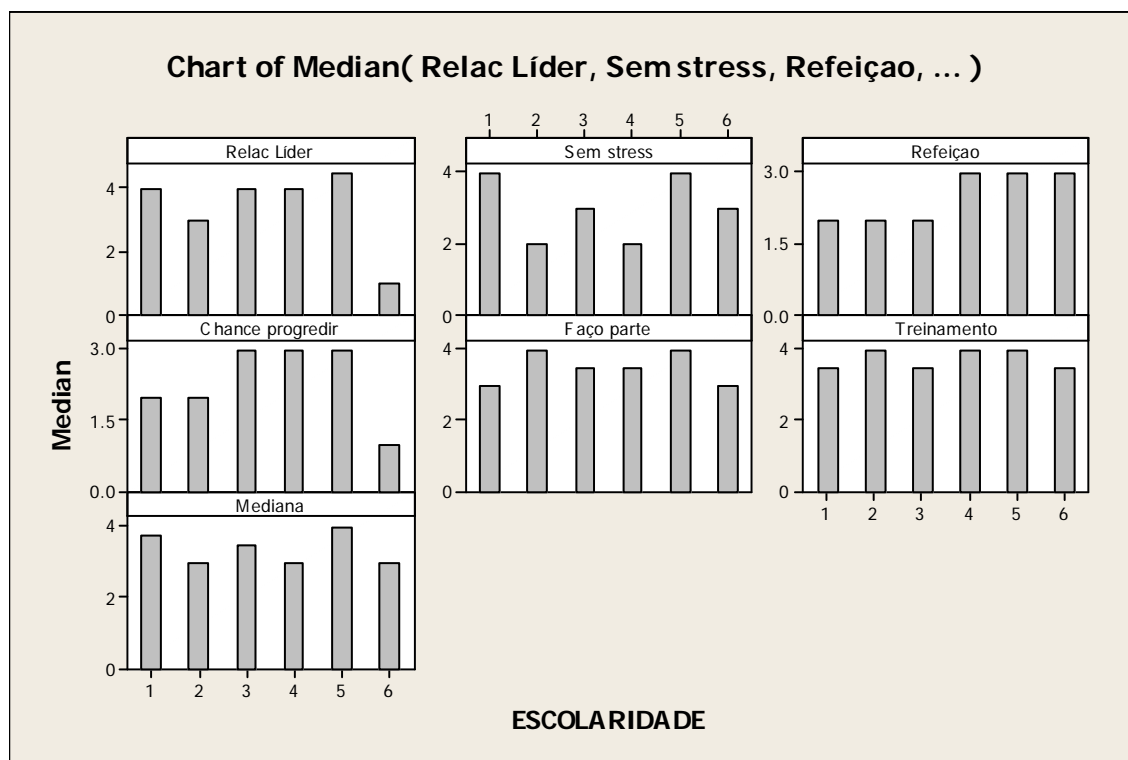
A mediana da variável “Refeição” é pior para menor escolaridade (Fundamental, Médio Incompleto e Médio Completo) e melhor para Superior Incompleto, Superior Completo e Pós. Os valores de mediana da variável “Chance de Progredir” aparece destacadamente melhores para os funcionários com maior escolaridade e piores valores para Fundamental e Médio

Incompleto, com exceção do grau de escolaridade Pós graduação que apresentam a pior mediana.

A mediana da variável “Faço Parte” mostra seus piores valores para Fundamental e Pós, na contrapartida possui melhores resultados para Superior Completo e Médio Incompleto.

Os valores de mediana da variável “Treinamento” apresenta os piores valores para funcionários com Fundamental, Médio Completo e Pós; e melhores valores para Médio Incompleto, Superior Incompleto e Superior Completo.

A variável “Mediana Total” que consolida todas as variáveis tem os melhores valores para funcionários com Superior Completo.



Legenda eixo X

- 1 - Fundamental
- 2 - Médio Incompleto
- 3 - Médio Completo
- 4 - Superior Incompleto
- 5 - Superior Completo
- 6 - Pós

### 3.3.4. Mediana por por faixa etária

A análise percentual e comparativa entre medianas por faixa etária demonstra valores melhores para “Relação com Líder” para indivíduos com menos de 25 anos para aqueles com mais de 55 anos. Os piores resultados dessa variável aparecem em indivíduos entre 35 a 44 anos.

A mediana da variável “Sem stress” melhora os resultados conforme aumenta a faixa etária, com exceção de indivíduos que possuem menos de 25 anos.

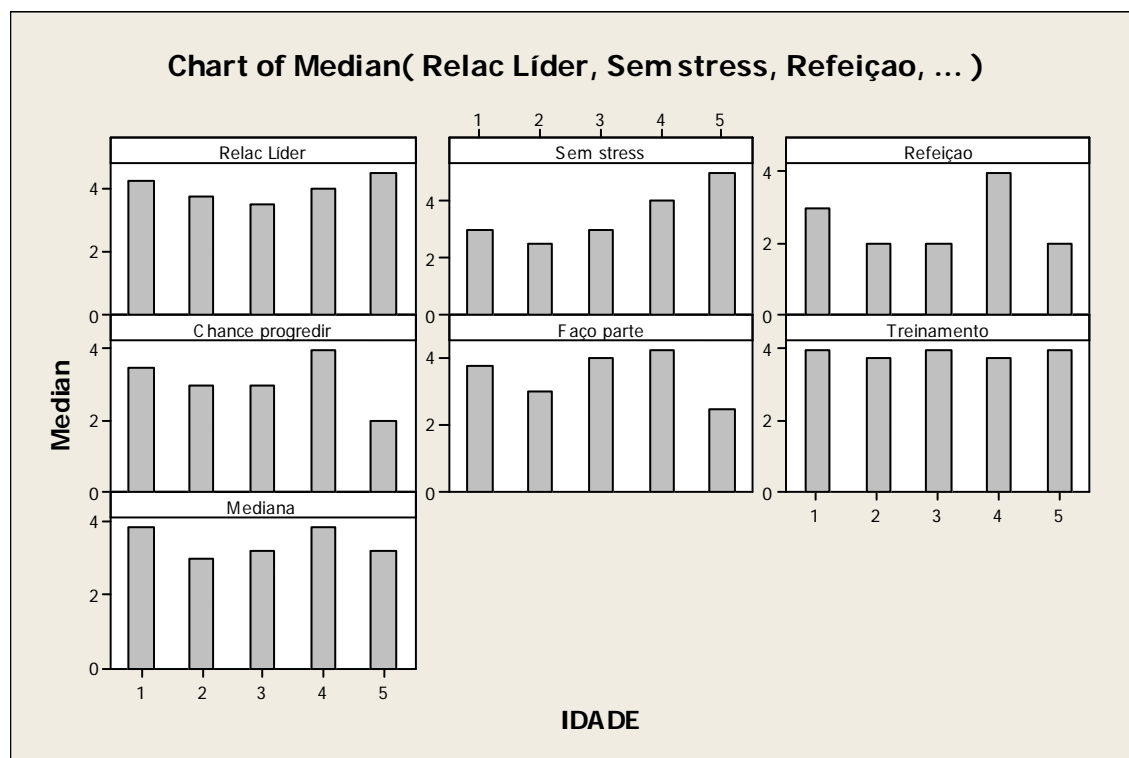
A mediana da variável “Refeição” apresenta valores significativamente melhores para indivíduos com 45 a 54 anos. Os piores valores apresentam-se naqueles com 25 a 34 anos, 35 a 44 anos e mais de 55 anos.

Os valores de mediana da variável “Chance de Progredir” aparecem piores para indivíduos com mais de 55 anos; e os melhores resultados para aqueles entre 44 e 55 anos.

Os piores valores para “Faço Parte” são de indivíduos com mais de 55 anos, na contrapartida melhores valores para aqueles com 45 a 54 anos.

Por fim, os valores de mediana da variável “Treinamento” apresenta valores similares para todas as faixas etárias.

A variável “Mediana Total” que consolida todas as variáveis tem os melhores valores para funcionários com menos de 25 anos e para aqueles entre 45 e 54 anos; os piores estão entre 25 e 34 anos.



Legenda eixo X

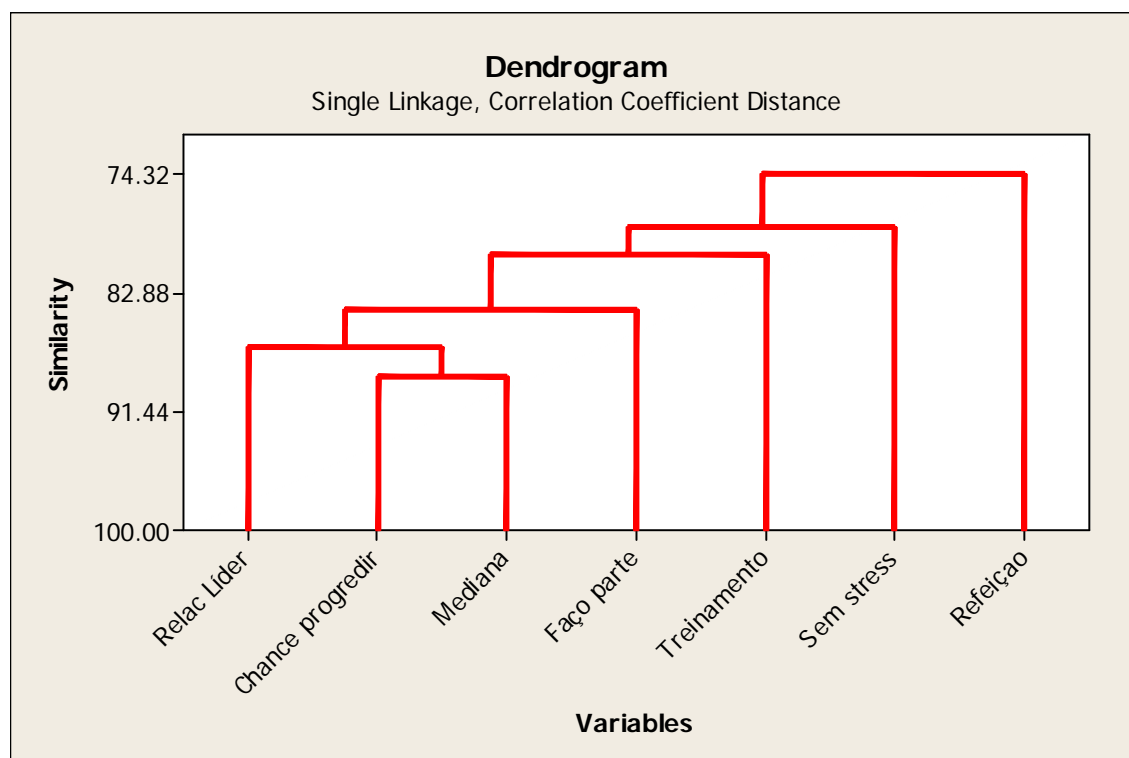
- 1 - menos de 25 anos
- 2 - 25 a 34 anos
- 3 - 35 a 44 anos
- 4 - 45 a 54 anos
- 5 - 55 ou mais

#### 4. MATRIZ DE CORRELAÇÃO

	Relac Líder	Sem stress	Refeição
Sem stress	0.419 0.000		
Refeição	0.279 0.008	0.133 0.212	
Chance progredir	0.564 0.000	0.305 0.003	0.317 0.002
Faço parte	0.457 0.000	0.276 0.009	0.236 0.025
Treinamento	0.322 0.002	0.290 0.006	0.181 0.087
Mediana	0.737 0.000	0.562 0.000	0.486 0.000
	Chance progredir	Faço parte	Treinamento
Faço parte	0.615 0.000		
Treinamento	0.350 0.001	0.409 0.000	
Mediana	0.778 0.000	0.681 0.000	0.603 0.000
Cell Contents: Pearson correlation P-Value			

As maiores correlações apresentam-se entre a “Mediana Total” e, respectivamente, as variáveis “Chance de Progredir” (0,778), “Relação com Líder” (0,737), “Faço Parte” (0,681) e “Treinamento” (0,603). Além disso, há correlação importante entre as variáveis “Chance de Progredir” e “Relação com o Líder” (0,564); também entre “Chance de Progredir” e “Faço Parte” (0.615). Além disso, é possível observar que todas os valores de correlação mencionados apresentam intervalo de confiança bastante altos, pois o P-Value apresenta valores inferiores a 5%, neste caso todos iguais a 0,00.

## 5. DENDOGRAMA



O dendrograma comprova a maior similaridade entre as variáveis “Mediana Total” e “Chance de Progredir”, e ainda próxima aparece a variável “Relação com Líder”. Ao passo que, as variáveis “Sem stress” e “Refeição” apresenta a menor similaridade dentre todas as variáveis.

## 6. SEGMENTAÇÃO DA POPULAÇÃO

### 6.1. Segmentação em Grupos a partir dos resultados para uma variável

Primeiramente para segmentar a amostra, adotamos a variável “Sem stress” como referência, pois parece ser aquela que melhor representa o clima organizacional. Assim, obteve-se 3 grupos: 1) **Grupo 1 (Discordância)**: resultados para a variável “Sem stress” em 1 (“Discordo Totalmente”) ou 2 (“Discordo”); 2) **Grupo 2 (Neutralidade)**: resultados para a variável “Sem stress” em 3 (“Nem concordo, nem discordo”); e 3) **Grupo 3 (Concordância)**: resultados para a variável “Sem stress” em 4 (“Concordo”) ou em 5 (“Concordo Totalmente”), conforme detalhamento abaixo. Não foi necessário definir intervalos para a criação dos grupos, visto que todas as respostas para a variável “Sem Stress” são numeros inteiros.

no pesq	Relac Líder	Sem stress	Refeição	Chanc e prog	Faço parte	Treina mento	Medi ana	Tempo Empr	Sexo	Turn o	Idade	Escol	Grupo de “Sem Stress”
5	3.5	2	2	2	3	2.5	2.25	1	m	b	2	3	1
8	3.5	2	1	1	2	2	2	2	m	b	2	3	1
11	3.5	2	2	3	3.5	3.5	3.25	2	m	b	3	3	1
13	3.5	1	1	1	1	1	1	5	m	b	3	3	1
14	5	2	2	2	3	4	2.5	2	m	b	1	4	1
17	4.5	1	3	4	4	3.5	3.75	4	m	b	3	3	1

26	3	1	2	1	3	2.5	2.25	1	m	c	2	3	1
29	1	1	1	1	3	4	1	4	m	c	3	3	1
32	1	2	1	3	2.5	4	2.25	2	m	adm	3	3	1
36	3	1	1	3	2.5	3	2.75	2	f	c	2	3	1
37	1.5	1	2	1	1.5	4	1.5	3	m	adm	3	3	1
39	4.5	1	2	2	4	4	3	4	m	c	3	3	1
40	2.5	2	4	1	2.5	2	2.25	2	m	c	1	3	1
41	4	2	1	3	3.5	2.5	2.75	4	m	c	3	4	1
54	2.5	1	1	1	4	4	1.75	4	m	d	2	2	1
55	3.5	2	2	4	2	2	2	1	m	d	2	3	1
56	1	2	1	1	1	2	1	4	m	d	2	3	1
58	3.5	2	3	2	4	4	3.25	3	m	d	3	4	1
61	4	2	4	3	3.5	3.5	3.5	2	m	d	2	3	1
62	1	1	1	1	1	1	1	2	m	adm	1	3	1
64	3.5	2	2	4	5	4.5	3.75	5	m	d	3	3	1
69	2	1	4	1	2.5	1.5	1.75	2	m	d	1	3	1
72	1	1	1	1	1	3	1	5	m	b	3	2	1
74	4.5	2	2	4	4	4	4	2	m	d	3	3	1
76	3	2	2	3	3	4	3	2	f	a	1	3	1
77	5	2	4	4	1	2	3	4	m	a	2	3	1
79	3	1	4	3	3.5	3	3	4	m	a	3	4	1
83	3	2	1	2	2	2.5	2	3	m	a	3	3	1
84	3	2	3	3	2	4	3	2	m	a	2	3	1
85	3	1	1	2	5	4.5	2.5	5	m	a	4	2	1
87	3	2	2	3	4	3.5	3	2	m	a	1	2	1
92	4	2	1	4	4.5	3.5	3.75	4	m	a	3	3	1
93	4.5	2	4	4	4	4	4	5	m	a	3	3	1
94	4.5	2	2	3	4	1.5	2.5	1	m	a	2	3	1
95	3	2	3	1	3	1.5	2.5	1	m	a	1	3	1
101	5	1	5	5	5	5	5	4	m	adm	3	3	1
103	5	2	3	3	4	4	3.5	3	f	a	2	5	1
1	4.5	3	1	1	2.5	4	2.75	4	m	b	2	3	2
23	4	3	5	3	4.5	4	4	2	m	b	2	3	2
28	4.5	3	4	3	4.5	3.5	3.75	5	m	c	4	3	2
31	3.5	3	4	2	3	2.5	3	2	m	c	1	3	2
33	4	3	3	4	4	3.5	3.75	1	m	c	1	3	2
34	4.5	3	1	3	4.5	3.5	3.25	5	m	c	3	3	2
35	2	3	3	1	2	4	2.5	4	m	adm	3	3	2
44	4	3	1	4	3.5	3.5	3.5	4	m	c	4	3	2
48	4	3	3	3	4	3.5	3.25	2	m	c	3	3	2
52	4	3	5	5	4	4.5	4.25	4	m	c	4	3	2
53	4.5	3	5	4	3.5	4.5	4.25	2	m	c	1	3	2
59	3.5	3	4	2	3	3.5	3.25	2	m	d	3	3	2
65	4.5	3	2	4	5	5	4.25	5	m	d	3	2	2
67	3.5	3	4	3	4	4	3.75	4	m	d	3	3	2
68	1	3	3	1	3	3.5	3	1	f	d	3	6	2
75	4	3	2	4	4	4	4	2	m	adm	1	2	2
78	4	3	3	3	4	4	3.5	3	m	a	2	3	2
99	3.5	3	4	3	4	3	3.25	5	m	c	3	5	2
2	5	5	2	5	3.5	1	4.25	1	m	b	1	5	3
3	5	5	4	1	2	4.5	4.25	4	m	b	2	3	3
4	5	4	3	4	4	4	4	3	m	b	1	3	3
6	4.5	4	4	5	5	4	4.25	2	m	b	4	5	3



7	3	4	1	2	3	5	3	3	m	b	2	3	3
10	3.5	4	2	4	4	4	4	5	m	b	3	3	3
15	4	5	3	5	4	4.5	4.25	4	m	b	3	3	3
18	4	4	3	2	3.5	4	3.75	2	m	b	1	3	3
19	4	5	1	4	5	5	4.5	3	m	b	2	3	3
20	3.5	4	4	4	3.5	4	4	5	m	b	4	5	3
21	3.5	4	3	2	3	3.5	3.25	3	m	b	4	3	3
24	3.5	5	1	1	1	3	2	4	m	b	3	3	3
25	3	5	1	1	5	1	2	5	m	c	3	3	3
27	4.5	4	1	4	3	5	4	3	m	c	2	3	3
30	5	4	4	4	4.5	5	4.25	3	m	c	2	3	3
38	3.5	4	2	5	5	3	3.75	2	m	c	2	3	3
42	4	4	2	2	4	3.5	3.75	3	m	c	2	1	3
43	4.5	5	3	5	4	4.5	4.5	1	m	c	1	3	3
49	4.5	4	4	4	4.5	4	4	3	m	c	1	3	3
60	3	5	4	1	2.5	4.5	3.5	4	m	d	2	3	3
70	4	4	4	5	5	4.5	4.25	5	m	d	4	3	3
80	2	5	2	2	4.5	2.5	2.25	4	m	a	2	3	3
81	4.5	5	2	2	2.5	4	3.25	5	f	a	3	2	3
82	4.5	5	2	2	2.5	4	3.25	5	f	a	5	1	3
86	4.5	4	4	5	4.5	2.5	4.25	1	m	a	3	3	3
88	4.5	4	3	3	4	4	4	4	m	b	3	5	3
89	4.5	5	3	4	4	5	4.25	2	m	a	1	3	3
90	5	5	2	4	4	5	4.5	2	m	a	1	3	3
91	4.5	4	1	1	2	3	2.5	5	m	a	3	3	3
96	3.5	4	4	4	3	3.5	3.75	5	m	a	4	1	3
97	4.5	5	4	5	4.5	3	4.5	1	m	a	1	3	3
98	4.5	4	2	4	4.5	3	4	6	m	a	4	3	3
100	4.5	4	2	2	3.5	4	3.75	2	m	d	3	5	3
102	5	5	1	5	5	5	5	5	m	adm	3	3	3
104	5	4	4	3	4	4.5	4	2	m	adm	1	4	3

### 6.1.1. Mediana por turno por Grupo de “Sem Stress”

A análise percentual e comparativa entre medianas por grupo de “Sem Stress” demonstra valores melhores para “Relação com Líder” para indivíduos do Grupo 3, valores intermediários para o grupo 2; e menores valores para o grupo 1.

A mediana da variável “Sem stress” obviamente apresenta correspondência de valores.

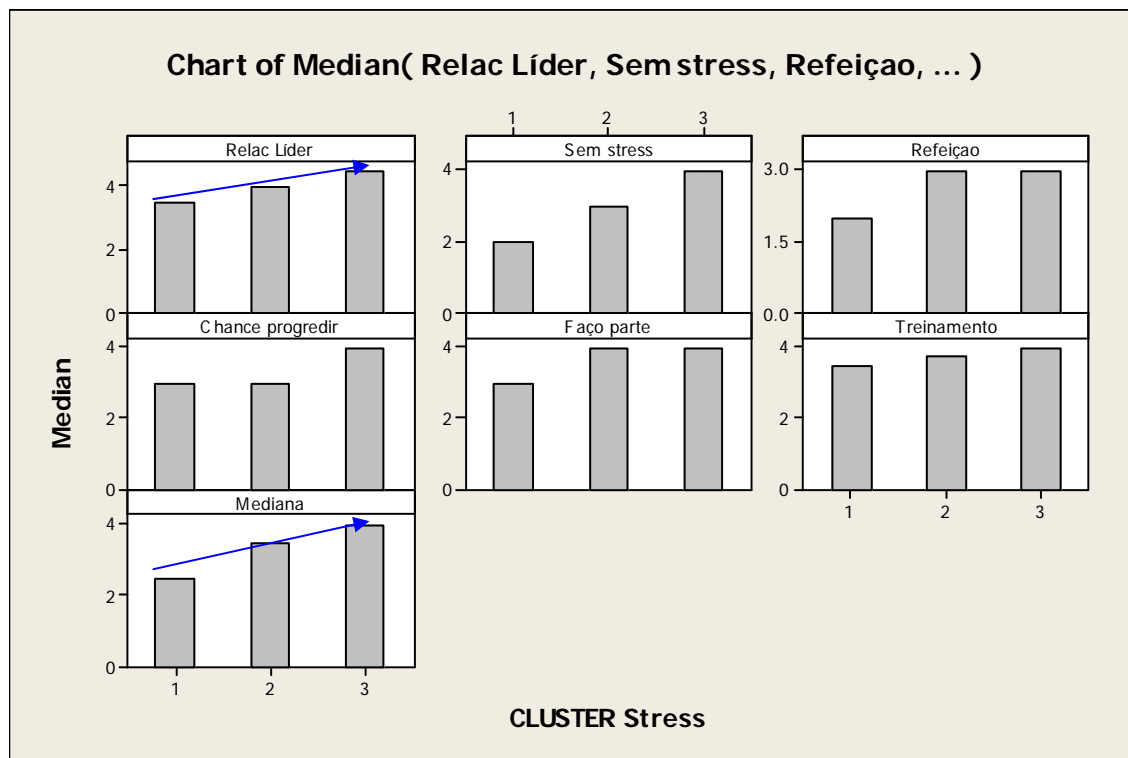
A mediana da variável “Refeição” apresenta valores menores para o grupo 1 (“Discordância”); e valores altos para os grupos 2 (“Neutralidade”) e 3 (“Concordância”).

Os valores de mediana da variável “Chance de Progredir” apresenta valores intermediários para os grupos 1; e valores altos para os grupos 2 e 3.

Os piores valores para “Faço Parte” apresenta valores intermediários para os grupos 1 e 2; e valores altos para o grupo 3.

Os valores de mediana da variável “Treinamento” apresenta valores similares para todas as faixas etárias.

A variável “Mediana Total” que consolida todas as variáveis ” apresenta valores crescentes do grupo 1 ao 3.



Legenda eixo X

1 – “Grupo Discordância” à “Sem Stress”: respostas 1 (“Discordo Totalmente”) ou 2 (“Discordo”) para a variável “Sem stress”  
 2 – “Grupo Neutralidade” à “Sem Stress”: respostas 3 (“Nem concordo, nem discordo”)  
 3 – “Grupo Concordância” à “Sem Stress”: respostas 4 (“Concordo”) ou 5 (“Concordo Totalmente”) para a variável “Sem stress”

Através da análise visual dos “Charts” das diferentes variáveis por grupo de “Sem Stress” concluímos que as variáveis “Relação com Líder” e “Mediana Total” são aquelas que mais acompanham os grupos de “Sem stress”, fora é lógico a própria variável “Sem stress”.

## 7. COMPARAÇÃO DE MÉDIA, ANÁLISE DE VARIÂNCIA E INTERVALO DE CONFIANÇA.

### 7.1. One-Way A-Nova (Unstacked) entre todas as variáveis

#### One-way ANOVA: Relac Líder, Sem stress, Refeição, Chance progr, Faço parte, ...

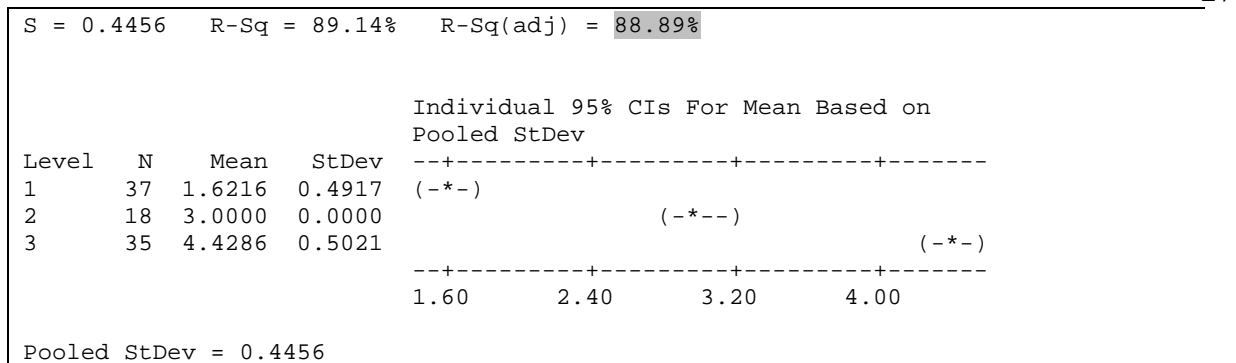
Source	DF	SS	MS	F	P
Factor	6	88.19	14.70	10.86	0.000
Error	623	843.48	1.35		
Total	629	931.67			

S = 1.164    R-Sq = 9.47%    R-Sq(adj) = 8.59%

Individual 95% CIs For Mean Based on Pooled StDev

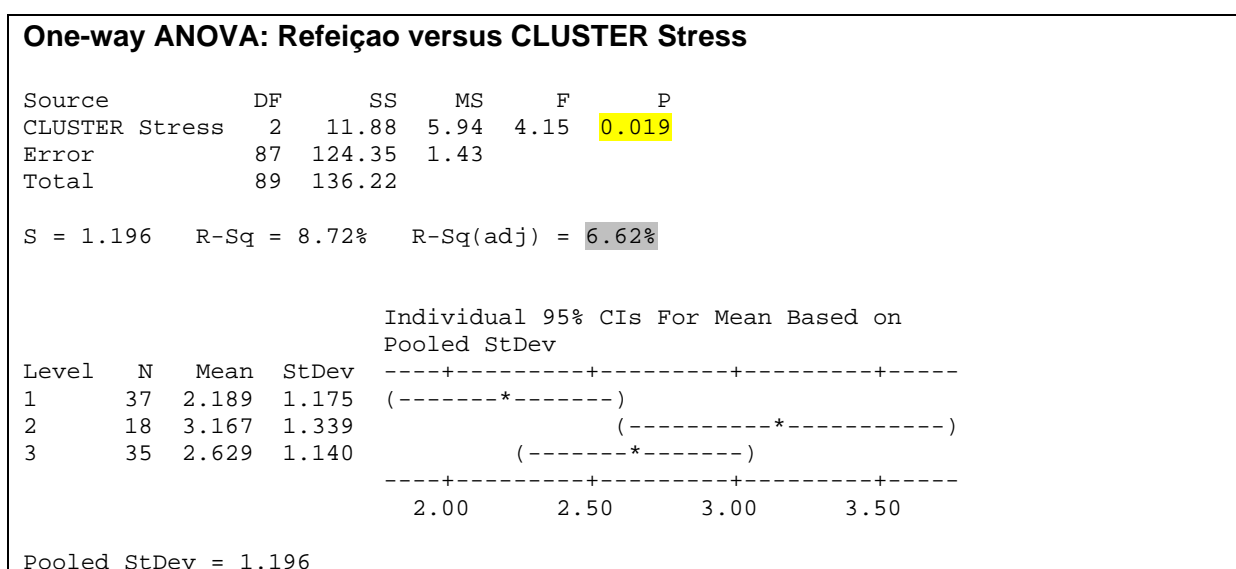
Level	N	Mean	StDev	CI
Relac Líder	90	3.694	1.067	(----*----)





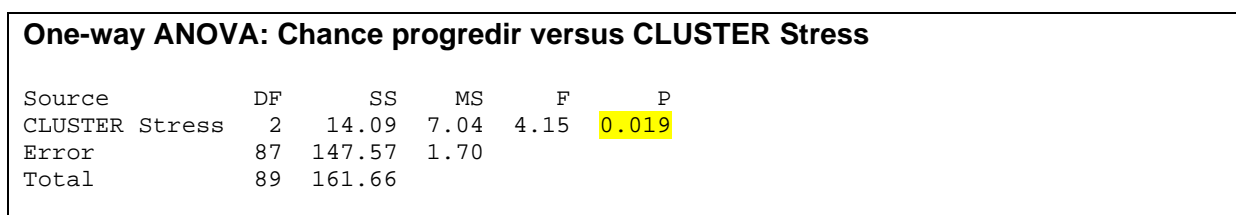
Se calcularmos a média de “Sem stress” por grupo de “Sem stress” obteremos maiores valores próximos e crescentes para os grupos 1 (1,6216), 2 (3,0000) e 3 (4,4286). Observa-se alto grau de confiabilidade das médias já que o valor de P-Value é inferior a 5% (0,0%). O alto valor de R-Sq maior que 50% (88,89%) demonstra obviamente que há alta correlação entre a variável “Sem Stress” e os resultados dos grupos de “Sem Stress”.

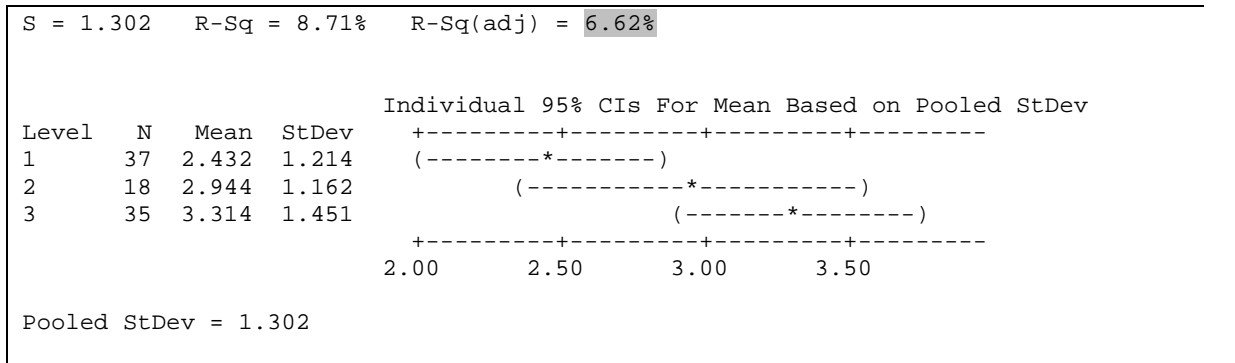
#### 7.4. One-Way A-Nova para “Refeição” por grupo de “Sem Stress”



Se calcularmos a média de “Refeição” por grupo de “Sem stress” obteremos valores desordenados para os grupos 1 (2,189) 2 (3,167) e 3 (2,629). Observa-se alto grau de confiabilidade das médias já que o valor de P-Value é inferior a 5% (1,9%). O baixíssimo valor de R-Sq, bastante inferior a 50% (6,62%) demonstra que não há correlação entre a variável “Refeição” e os resultados dos grupos de “Sem Stress”.

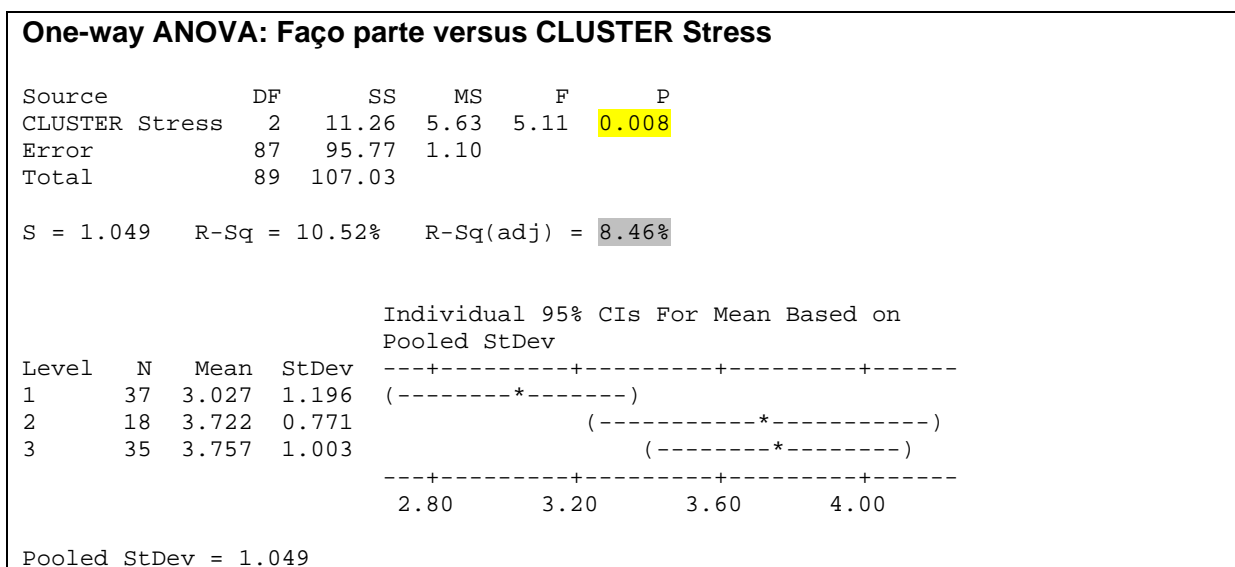
#### 7.5. One-Way A-Nova para “Chance de Progredir” por grupo de “Sem Stress”





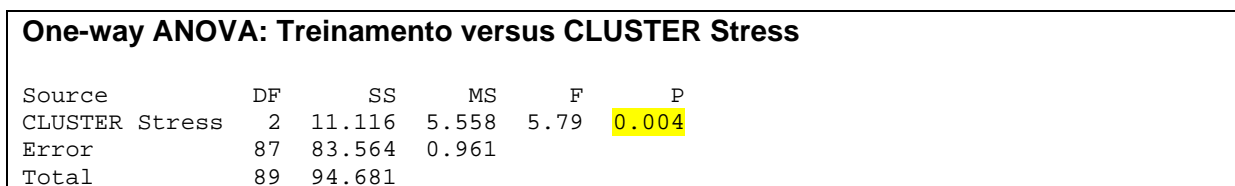
Se calcularmos a média de “Chance de Progredir” por grupo de “Sem stress” obteremos valores crescentes para os grupos 1 (2,432), 2 (2,944) e 3 (3,314). Observa-se alto grau de confiabilidade das médias já que o valor de P-Value é inferior a 5% (1,9%). O baixíssimo valor de R-Sq, bastante inferior a 50% (6,62%) demonstra que não há correlação entre a variável “Chance de Progredir” e os resultados dos grupos de “Sem Stress”.

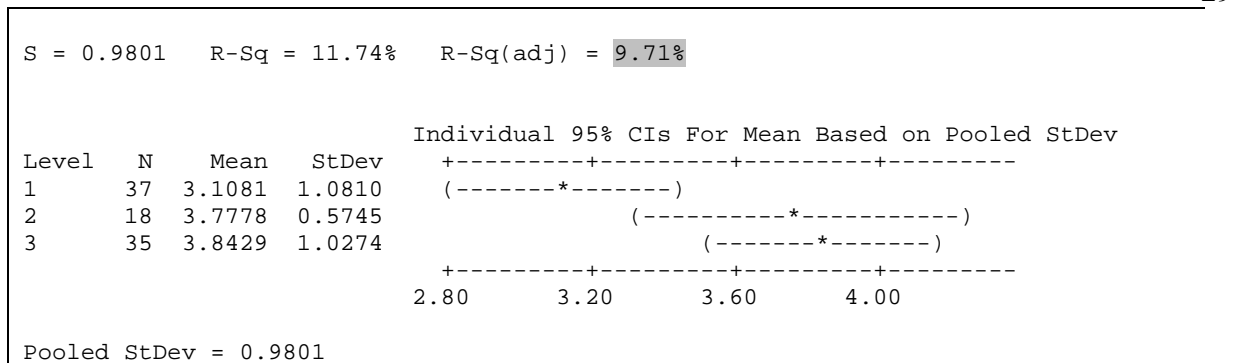
### 7.6. One-Way A-Nova para “Faço Parte” por grupo de “Sem Stress”



Se calcularmos a média de “Faço Parte” por grupo de “Sem stress” obteremos valores levemente crescentes para os grupos 1 (3,027), 2 (3,722) e 3 (3,757). Observa-se alto grau de confiabilidade das médias já que o valor de P-Value é inferior a 5% (0,8%). O baixíssimo valor de R-Sq, bastante inferior a 50% (8,46%) demonstra que não há correlação entre a variável “Faço Parte” e os resultados dos grupos de “Sem Stress”.

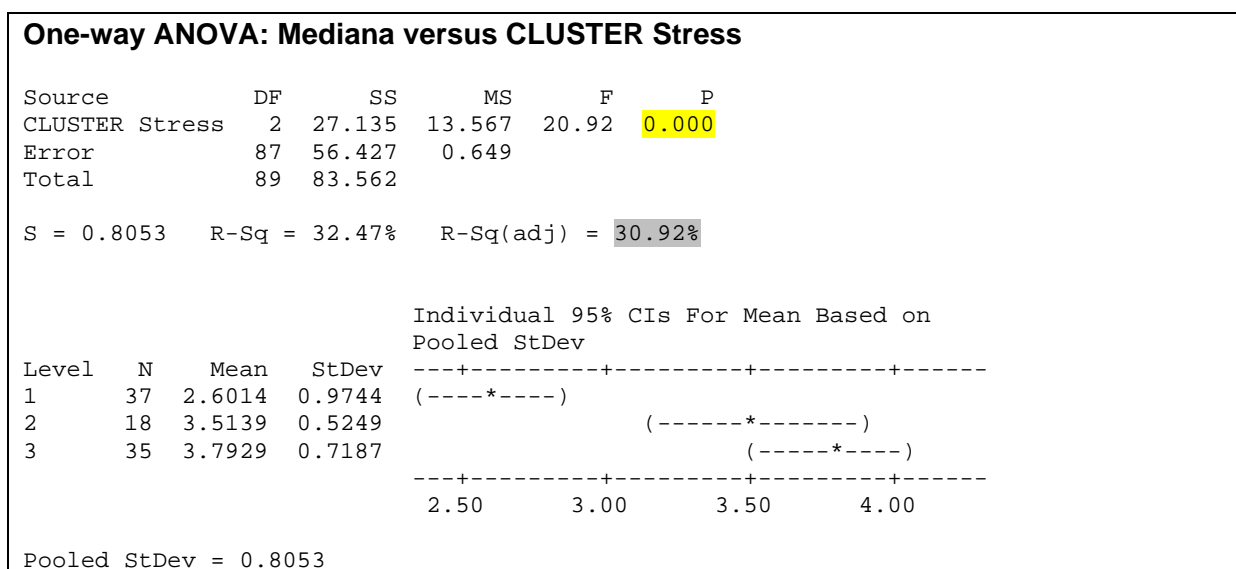
### 7.7. One-Way A-Nova para “Treinamento” por grupo de “Sem Stress”





Se calcularmos a média de “Treinamento” por grupo de “Sem stress” obteremos valores levemente crescentes para os grupos 1 (3,1081), 2 (3,7778) e 3 (3,8429). Observa-se alto grau de confiabilidade das médias já que o valor de P-Value é inferior a 5% (0,4%). O baixíssimo valor de R-Sq, bastante inferior a 50% (9,71%) demonstra que não há correlação entre a variável “Treinamento” e os resultados dos grupos de “Sem Stress”.

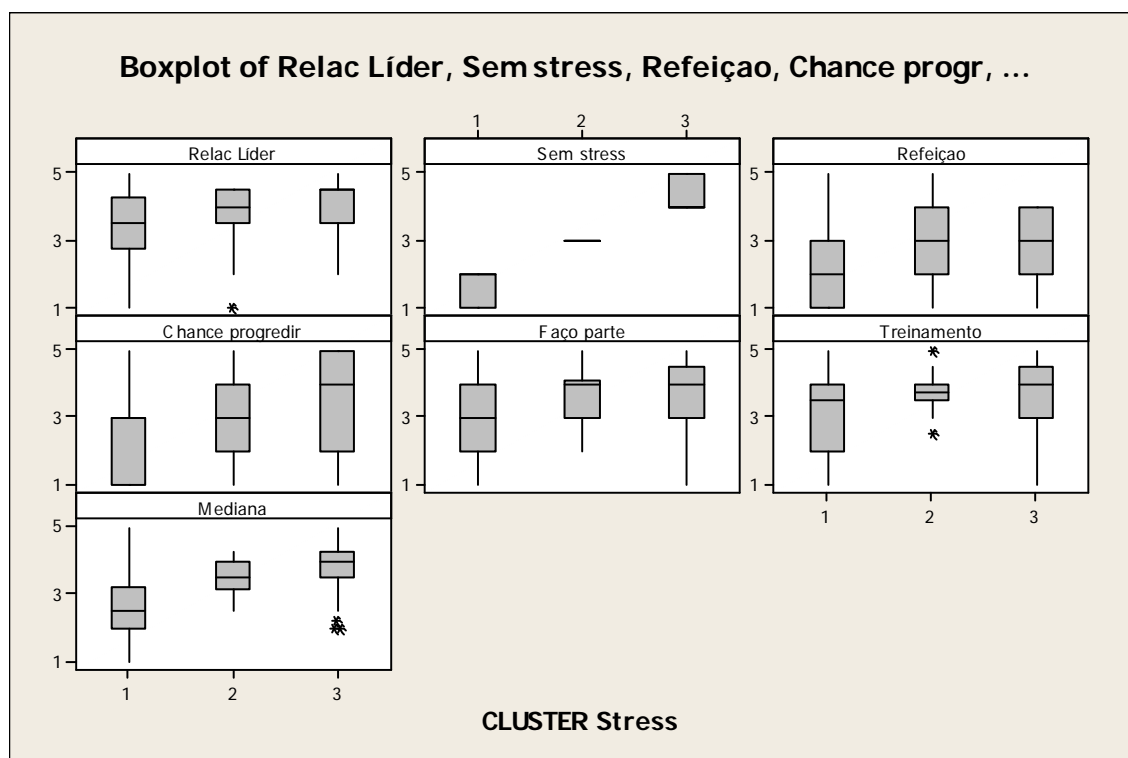
### 7.8. One-Way A-Nova para “Mediana Total” por grupo de “Sem Stress”



Se calcularmos a média da variável “Mediana Total” por grupo de “Sem stress” obteremos valores crescentes para os grupos 1 (2,6014), 2 (3,5139) e 3 (3,7929). Observa-se alto grau de confiabilidade das médias já que o valor de P-Value é inferior a 5% (0,0%). O baixo valor de R-Sq, inferior a 50% (30,92%) demonstra que há pouca correlação entre a variável “Mediana Total” e os resultados dos grupos de “Sem Stress”.

- Pela análise do valor F de cada variável, percebemos que os clusters são bastante próximos, com exceção de “Sem stress” que é própria variável na qual baseou-se a divisão dos grupos de “Sem Stress” (1, 2 e 3).

### 7.9. “Box Plot” de todas as variáveis segmentando grupos de “Sem Stress”



Através da análise visual do Box Plot acima, percebemos que as variáveis que melhor distinguem os grupos são obviamente a própria “Sem stress” e “Mediana Total”, além disso, apesar do alto desvio padrão estaria “Chance de Progredir”.

As variáveis “Refeição”, “Relação com Líder”, “Faço Parte” e “Treinamento” não possibilitam boa distinção dos grupos (1, 2 e 3) pois, apresentam valores muitíssimo próximos entre os grupos.

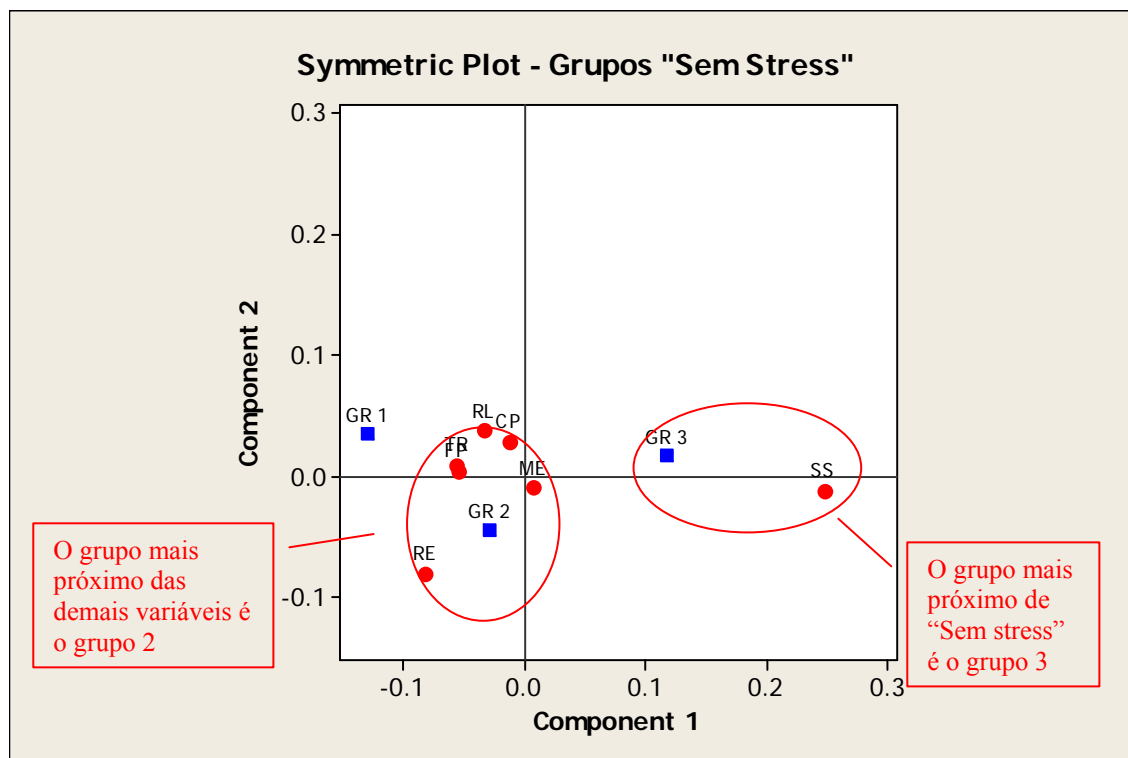
## 8. ANÁLISE DE CORRESPONDENCIA

A partir das médias obtidas para cada variável por grupo (dispostas na tabela abaixo) é possível avaliar a “Análise de Correspondência” entre as mesmas. A análise de correspondência é um método de análise fatorial para variáveis categóricas, que converte uma tabela de dados não negativos de duas ou múltiplas entradas em um tipo de representação gráfica em que as linhas e as colunas são simultaneamente representadas por pontos no gráfico. Este método permite mostrar como as variáveis dispostas em linhas e colunas estão relacionadas e não somente se a relação existe.

Variável \ Grupos "Sem stress"	Média Grupo 1 "Discordância"	Média Grupo 2 "neutralidade"	Média Grupo 3 "Concordância"
"Relação com Líder"	3.2297	3.7500	4.1571
"Sem stress"	1.6216	3.0000	4.4286
"Refeição"	2.1890	3.1670	2.6290
"Chance de Progredir"	2.4320	2.9440	3.3140
"Faço Parte"	3.0270	3.7220	3.7570
"Treinamento"	3.1081	3.7778	3.8429
"Mediana"	2.6014	3.5139	3.7929

Abaixo é apresentado o resultado da análise de correspondência para os 3 Grupos de “Sem Stress” x 7 variáveis estudadas.

### 8.1. Symetric Plot (Variáveis x grupos de “Sem Stress”)



No gráfico acima percebemos que dentre as variáveis estudadas as mais próximas ao Grupo 2 são “Relação com Líder”, “Chance de Progredir”, “Treinamento”, “Faço Parte” e “Refeição”. A variável “Sem stress” obviamente aparece mais próxima do Grupo 3, que é aquele que foi formado pelos indivíduos que responderam os maiores valores para esta mesma variável.



## 8.2. Tabela de Análise de Contigência e Contribuições

Simple Correspondence Analysis: GR 1, GR 2, GR 3										
Analysis of Contingency Table										
Axis	Inertia	Proportion	Cumulative	Histogram						
1	0.0101	0.8982	0.8982	*****						
2	0.0011	0.1018	1.0000	***						
Total	0.0112									
Row Contributions										
ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	RL	1.000	0.164	0.038	-0.034	0.441	0.019	0.038	0.559	0.208
2	SS	1.000	0.133	0.725	0.247	0.998	0.805	-0.011	0.002	0.015
3	RE	1.000	0.117	0.138	-0.082	0.515	0.079	-0.080	0.485	0.655
4	CP	1.000	0.128	0.011	-0.012	0.144	0.002	0.029	0.856	0.094
5	FP	1.000	0.154	0.042	-0.055	0.992	0.046	0.005	0.008	0.003
6	TR	1.000	0.158	0.045	-0.056	0.970	0.048	0.010	0.030	0.013
7	ME	1.000	0.146	0.002	0.008	0.401	0.001	-0.009	0.599	0.011
Column Contributions										
ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	GR 1	1.000	0.268	0.430	-0.130	0.933	0.447	0.035	0.067	0.285
2	GR 2	1.000	0.351	0.089	-0.029	0.293	0.029	-0.045	0.707	0.620
3	GR 3	1.000	0.381	0.480	0.118	0.980	0.524	0.017	0.020	0.095

Através da tabela da “Análise de Contingência”, que mostra decomposição da inércia ( $\chi^2/n$ ), identificamos que, do total da inércia da matriz de dados, 89,82% é contabilizada no primeiro componente; e 10,18 % é contabilizada no segundo componente. Assim os dois primeiros já contabilizam 100%. Detalhando temos que:

- A coluna “Coord” traz a coordenada (distância gráfica) dos componentes em relação ao zero.
- A coluna “Corr” representa a contribuição do componente para a inércia da linha (dos 3 grupos de “Sem Stress” no nosso caso) ou da coluna (das 7 variáveis).
- A coluna “Contr”, significa a contribuição de cada linha (do grupo de “Sem stress” em nosso caso) para o eixo da inércia ou de cada coluna (das 7 variáveis).

Dessa forma, analisando as variáveis (linhas), o Componente 1 contribui mais para a inércia de “Sem Stress”, “Faço Parte” e “Treinamento” (Corr = 0.998, 0.992 e 0.970, respectivamente). A variável “Sem Stress” é a única com altos valores de “Corr” e “Contr”; assim as variáveis “Faço Parte” e “Treinamento” apesar desses valores dos altos valores de “Corr” não apresentam contribuição significativa ao eixo da inércia (“Contr”), sendo assim desconsideradas para efeitos de distinção entre componentes. Já o Componente 2 contribui mais para a inércia de “Chance de Progredir” e “Relação com Líder” (Corr = 0.856, 0.559, respectivamente), mas explica pouco da inércia das demais variáveis. Vale ressaltar que “Refeição” é a única que contribui de forma destacada para o eixo de inércia do Componente 2, pois as variáveis “Relação com Líder” e “Chance de Progredir” apesar do alto valor de

“Corr” não apresentam contribuição significativa ao eixo da inércia, sendo assim desconsideradas para efeitos de distinção entre componentes.

Seguimos agora analisando os grupos (colunas), o Componente 1 contribui mais para a inércia dos grupos 3 e 1 (Corr = 0.980 e 0.933, respectivamente), mas comparativamente aos anteriores explica pouco da inércia do grupo 2 (Corr = 0.293). Na contrapartida, o Componente 2 contribui mais para a inércia de do grupo 2 (Corr = 0.707), mas explica pouco da inércia do grupo 1 e 3 (Corr = 0.067 e 0.020, respectivamente).

Resumindo, temos que a variável “Sem stress” destacadamente mais contribui para o **componente 1**. Os grupos 1 e 3, que encontram-se em oposição entre si, são os que mais contribuem para o mesmo componente (1). Na contrapartida, “Refeição” é a variável que mais interfere no **componente 2**. O grupo 2 é o que mais contribui para o componente 2. Assim poderíamos denominar o componente 1 como “Estudo do stress nos grupos 1 e 3” e o componente 2 como “Estudo da refeição no grupo 2”.

A alta contribuição da variável “Sem Stress” ocorre justamente porque os grupos foram formados por distintas repostas desta variável; onde valores 4 (“Concordo”) e 5 (“Concordo Totalmente”) formam o grupo 3; e valores 2 (“Discordo”) e 1 (“Discordo Totalmente”) formam o grupo 1. Assim sendo, a análise de correspondência, com exceção da **conexão encontrada entre o grupo 2 e “Refeição”**, não se mostra muito elucidativa do comportamento da amostra em relação as variáveis estudadas.

## 9. ANÁLISE DISCRIMINANTE

A Análise Discriminante é usada para classificar observações em 2 ou mais grupos se há uma amostra com grupos conhecidos. Também pode ser usada para investigar como as variáveis contribuem para a separação dos grupos.

Através da “Análise Discriminante”, analisaremos qual modelo (linear ou quadrático) e quais agrupamento de variáveis podem explicar melhor e de maneira mais simples a população estudada.

Para esta análise especificamente, necessitou-se excluir a variável “Sem stress”, pois dela derivam os Grupos selecionados (1, 2 e 3) para este estudo, a aderência observada entre a referida variável e os grupos não significa um modelo preditivo, e sim uma obviedade. Da mesma forma, foi excluída a variável “Mediana Total”, já que é calculada através da mediana entre todas as variáveis, incluindo a variável “Sem stress”, o que acaba resultando no mesmo problema mencionado, não formando parte portanto de um modelo preditivo.

**Todas as inúmeras tentativa de combinar variáveis (tanto no modelo linear quanto quadrático) não obtiveram muito êxito, resultando em uma aderência entre 38,9% e 61%, conforme exposto nos itens 9.1 e 9.2 abaixo.**

## 9.1. “Análise Discriminante – Modelo Linear”

O melhor resultado para o modelo linear resultou em aderência modesta de 61%, através da utilização de todas as variáveis, exceto “Faço Parte”, conforme abaixo demonstrado.

Discriminant Analysis: CLUSTER Stress versus Relac Líder, Refeição, ...					
Linear Method for Response: CLUSTER Stress					
Predictors: Relac Líder, Refeição, Chance progredir, Treinamento					
Group	1	2	3		
Count	37	18	35		
Summary of classification					
	True Group				
Put into Group	1	2	3		
1	23	1	5		
2	6	11	9		
3	8	6	21		
Total N	37	18	35		
N correct	23	11	21		
Proportion	0.622	0.611	0.600		
N = 90	N Correct = 55		Proportion Correct = 0.611		
Squared Distance Between Groups					
	1	2	3		
1	0.00000	1.08280	1.20417		
2	1.08280	0.00000	0.52953		
3	1.20417	0.52953	0.00000		
Linear Discriminant Function for Groups					
	1	2	3		
Constant	-8.968	-13.349	-14.058		
Relac Líder	2.714	3.001	3.443		
Refeição	0.882	1.485	0.963		
Chance progredir	-0.418	-0.512	-0.320		
Treinamento	2.657	3.242	3.209		
Summary of Misclassified Observations					
Observation	True Group	Pred Group	Group	Squared Distance	Probability
1**	2	3	1	7.870	0.280
			2	8.342	0.221
			3	6.710	0.500
2**	3	1	1	13.89	0.504
			2	17.13	0.099
			3	14.36	0.397
3**	3	2	1	13.540	0.078
			2	9.521	0.583
			3	10.610	0.338
12**	1	3	1	6.057	0.179
			2	5.224	0.271
			3	3.809	0.550
14**	1	3	1	2.1153	0.216
			2	1.3260	0.321
			3	0.5947	0.463
15**	3	2	1	2.657	0.213
			2	1.193	0.443
			3	1.705	0.343
17**	3	2	1	3.590	0.192
			2	1.585	0.524
			3	2.815	0.283
18**	3	2	1	1.1485	0.313
			2	0.5576	0.421
			3	1.4785	0.266
20**	3	1	1	2.834	0.548

			2	5.052	0.181
			3	4.237	0.272
21**	3	1	1	5.723	0.831
			2	10.570	0.074
			3	10.063	0.095
29**	2	3	1	1.5917	0.265
			2	1.0902	0.341
			3	0.8008	0.394
30**	2	3	1	3.469	0.319
			2	4.904	0.156
			3	2.469	0.526
34**	3	1	1	5.421	0.453
			2	7.188	0.187
			3	5.879	0.360
35**	1	3	1	3.804	0.225
			2	3.258	0.295
			3	2.285	0.480
40**	2	3	1	3.544	0.378
			2	5.455	0.145
			3	3.078	0.477
42**	3	2	1	3.593	0.130
			2	1.013	0.472
			3	1.358	0.397
48**	1	2	1	1.9150	0.255
			2	0.7388	0.460
			3	1.6930	0.285
50**	3	2	1	7.910	0.161
			2	5.042	0.677
			3	7.899	0.162
51**	1	2	1	2.5685	0.198
			2	0.6723	0.512
			3	1.8106	0.290
53**	1	3	1	3.242	0.267
			2	3.064	0.292
			3	2.239	0.441
54**	2	3	1	5.007	0.134
			2	3.668	0.262
			3	1.993	0.604
56**	2	1	1	7.469	0.543
			2	8.129	0.390
			3	11.643	0.067
58**	3	2	1	5.829	0.126
			2	3.139	0.484
			3	3.576	0.389
60**	1	3	1	2.7182	0.213
			2	2.5497	0.232
			3	0.8083	0.554
61**	2	3	1	2.1413	0.267
			2	2.2607	0.251
			3	0.9612	0.482
63**	1	3	1	7.324	0.266
			2	6.797	0.346
			3	6.567	0.388
65**	1	2	1	2.825	0.336
			2	2.089	0.485
			3	4.078	0.179
66**	3	1	1	1.846	0.706
			2	4.495	0.188
			3	5.632	0.106
70**	1	2	1	1.6191	0.305
			2	0.9195	0.433
			3	1.9314	0.261
78**	1	3	1	3.544	0.378
			2	5.455	0.145
			3	3.078	0.477
79**	1	2	1	3.593	0.130
			2	1.013	0.472
			3	1.358	0.397
82**	3	2	1	3.225	0.241
			2	1.805	0.490
			3	3.001	0.269
87**	1	2	1	9.563	0.044
			2	4.506	0.553
			3	5.136	0.403
89**	1	3	1	4.030	0.139
			2	2.179	0.350
			3	1.424	0.511
90**	3	2	1	6.391	0.080
			2	2.749	0.497
			3	3.070	0.423

## 9.2. “Análise Discriminante – Modelo Quadrático”

Para o modelo quadrático obtivemos o melhor resultado através da utilização de todas as variáveis, o que resultou em 60% de aderência.

### Discriminant Analysis: CLUSTER Stress versus Relac Líder, Refeição, ...

Quadratic Method for Response: CLUSTER Stress

Predictors: Relac Líder, Refeição, Chance progredir, Faça parte, Treinamento

Group	1	2	3
Count	37	18	35

#### Summary of classification

Put into Group	True Group		
	1	2	3
1	21	2	2
2	11	13	13
3	5	3	20
Total N	37	18	35
N correct	21	13	20
Proportion	0.568	0.722	0.571

N = 90

N Correct = 54

Proportion Correct = 0.600

#### From Generalized Squared Distance to Group

Group	1	2	3
1	-0.109	0.424	1.917
2	0.889	-2.371	0.181
3	0.776	-1.900	-0.521

#### Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
1**	2	3	1	14.479	0.006
			2	8.529	0.120
			3	4.560	0.874
4**	3	2	1	2.47220	0.150
			2	-0.01885	0.522
			3	0.90854	0.328
6**	3	2	1	6.693	0.041
			2	1.441	0.563
			3	2.147	0.396
9**	3	2	1	2.592	0.224
			2	1.279	0.432
			3	1.740	0.343
10**	1	2	1	0.2884	0.287
			2	-0.7204	0.476
			3	0.6721	0.237
12**	1	3	1	8.986	0.038
			2	5.347	0.233
			3	3.060	0.730
14**	1	2	1	1.7747	0.164
			2	-0.2766	0.458
			3	0.1038	0.378
15**	3	2	1	3.2532	0.103
			2	-0.1883	0.577
			3	0.9881	0.320
17**	3	2	1	4.2349	0.123
			2	0.7314	0.708
			3	3.6004	0.169
18**	3	2	1	1.538	0.164
			2	-1.228	0.652
			3	1.299	0.184
27**	2	1	1	3.486	0.424
			2	3.741	0.374
			3	4.970	0.202
35**	1	2	1	4.420	0.176
			2	2.441	0.473
			3	3.033	0.352
37**	1	3	1	3.520	0.407
			2	6.996	0.072
			3	3.022	0.522
38**	3	2	1	2.1182	0.240
			2	0.7599	0.473
			3	1.7525	0.288

40**	2	3	1	4.253	0.301
			2	4.827	0.226
			3	3.348	0.473
42**	3	2	1	3.2819	0.087
			2	-0.8750	0.699
			3	1.4962	0.214
48**	1	2	1	2.4713	0.219
			2	0.6373	0.547
			3	2.3372	0.234
50**	3	2	1	12.670	0.026
			2	6.120	0.675
			3	7.740	0.300
51**	1	2	1	2.356	0.121
			2	-1.143	0.698
			3	1.553	0.181
53**	1	3	1	4.504	0.299
			2	5.448	0.187
			3	3.424	0.514
54**	2	3	1	4.216	0.343
			2	5.183	0.211
			3	3.686	0.446
56**	2	1	1	7.273	0.606
			2	8.135	0.394
			3	20.703	0.001
60**	1	2	1	2.3660	0.135
			2	-0.2493	0.499
			3	0.3757	0.365
65**	1	2	1	4.0803	0.158
			2	0.9884	0.742
			3	4.9960	0.100
66**	3	1	1	9.622	0.488
			2	14.727	0.038
			3	9.682	0.474
73**	1	2	1	1.578	0.337
			2	1.078	0.432
			3	2.332	0.231
74**	3	2	1	1.764	0.137
			2	-1.187	0.601
			3	0.474	0.262
78**	1	3	1	5.520	0.128
			2	3.621	0.331
			3	2.644	0.540
79**	1	2	1	3.0083	0.096
			2	-0.8107	0.645
			3	1.0077	0.260
80**	1	3	1	7.855	0.318
			2	18.168	0.002
			3	6.336	0.680
82**	3	1	1	4.406	0.400
			2	4.501	0.381
			3	5.613	0.219
86**	3	2	1	4.850	0.089
			2	1.525	0.469
			3	1.643	0.442
87**	1	2	1	8.503	0.106
			2	5.362	0.512
			3	5.950	0.382
88**	3	2	1	9.360	0.133
			2	6.988	0.434
			3	6.989	0.434
89**	1	2	1	3.1460	0.163
			2	0.6500	0.568
			3	2.1435	0.269
90**	3	2	1	5.928	0.130
			2	3.371	0.468
			3	3.676	0.402

## 10. REGRESSAO LOGÍSTICA

O modelo de regressão logística ordinal apresenta concordância de 71,6%, mostra-se pouco útil porque há algumas variáveis (“Refeição”, “Chance de Progredir” e “Faço Parte”) com valores de P superiores a 0,05 (0,661, 0,986 e 0,513, respectivamente), significando que os valores apresentados possuem pouca previsibilidade devido à enorme variância.

**Ordinal Logistic Regression: CLUSTER Stre versus Relac Líder, Refeição, ...**

Link Function: Logit								
Response Information								
Variable	Value	Count						
CLUSTER Stress	1	37						
	2	18						
	3	35						
	Total	90						
Logistic Regression Table								
Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI		
Const(1)	4.37521	1.21703	3.60	0.000				
Const(2)	5.36540	1.26117	4.25	0.000				
Relac Líder	-0.621929	0.271909	-2.29	0.022	0.54	0.32	0.91	
Refeição	-0.0804203	0.183460	-0.44	0.661	0.92	0.64	1.32	
Chance progredir	0.0038239	0.221662	0.02	0.986	1.00	0.65	1.55	
Faço parte	-0.173360	0.264817	-0.65	0.513	0.84	0.50	1.41	
Treinamento	-0.470803	0.239538	-1.97	0.049	0.62	0.39	1.00	
Log-Likelihood = -84.164								
Test that all slopes are zero: G = 21.501, DF = 5, P-Value = 0.001								
Goodness-of-Fit Tests								
Method	Chi-Square	DF	P					
Pearson	179.657	171	0.310					
Deviance	168.329	171	0.543					
Measures of Association:								
(Between the Response Variable and Predicted Probabilities)								
Pairs	Number	Percent	Summary Measures					
Concordant	1856	71.6	Somers' D	0.44				
Discordant	715	27.6	Goodman-Kruskal Gamma	0.44				
Ties	20	0.8	Kendall's Tau-a	0.28				
Total	2591	100.0						

Sendo assim, buscaremos novo modelo de regressão logística ordinal excluindo as estas variáveis com valores altos para P-value. Dessa forma, num segunda tentativa, foi possível obter boa concordância de 70.9%, com P-Value da “Tabela de Regressão Logística” com resultados menores de 5% (0,003 para “Relação com Líder” e 0,020 para “Treinamento”), significando que os valores apresentados possuem boa previsibilidade devido à pequena variância. Além disso, o P-Value do Teste “Goodness-of-Fit”, que analisa a diferença entre os valores observados e os valores esperados pelo modelo, obteve 0,222 para Pearson e 0,361 para “Deviance”, indicando que **não há evidências para afirmar que o modelo não se ajusta adequadamente aos dados**. Se o P-Value fosse menor que 5%, o teste rejeitaria a hipótese nula que o modelo devesse explicar os dados adequadamente.

### Ordinal Logistic Regression: CLUSTER Stre versus Relac Líder, Treinamento

Link Function: Logit

Response Information								
Variable	Value	Count						
CLUSTER Stress	1	37						
	2	18						
	3	35						
	Total	90						
Logistic Regression Table								
Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI		
Const(1)	4.07839	1.11998	3.64	0.000				
Const(2)	5.06394	1.16445	4.35	0.000				
Relac Líder	-0.697801	0.238212	-2.93	0.003	0.50	0.31	0.79	
Treinamento	-0.534611	0.229878	-2.33	0.020	0.59	0.37	0.92	
Log-Likelihood = -84.550								
Test that all slopes are zero: G = 20.729, DF = 2, P-Value = 0.000								
Goodness-of-Fit Tests								
Method	Chi-Square	DF	P					
Pearson	91.4794	82	0.222					
Deviance	85.9589	82	0.361					
Measures of Association: (Between the Response Variable and Predicted Probabilities)								
Pairs	Number	Percent	Summary Measures					
Concordant	1836	70.9	Somers' D		0.44			
Discordant	688	26.6	Goodman-Kruskal Gamma		0.45			
Ties	67	2.6	Kendall's Tau-a		0.29			
Total	2591	100.0						

## 11. ANÁLISE DOS COMPONENTES PRINCIPAIS

O objetivo deste tópico é criar indicadores (componentes principais) que expliquem as relações entre as variáveis e tornem possível o agrupamento e das mesmas.

Agrupando-se as 7 variáveis em apenas 1 indicador (PC1) teremos uma aderência de 53,1% à amostra, somando-se os 2 indicadores (PC1 + PC2) chegamos a um valor acumulado de 65,8%. Agregando ainda PC3 obteremos um percentual de 77,0%, o que já é bastante significativo, pois ao invés de trabalharmos com 7 variáveis podemos adotar apenas com os índices PC1, PC2 e PC3 para explicar 77,0% da amostra conforme abaixo.

### Principal Component Analysis: Relac Líder, Sem stress, Refeição, Chance progr,

Eigenanalysis of the Correlation Matrix



Eigenvalue	3.7151	0.8939	0.7787	0.7156	0.4795	0.3533	0.0638
Proportion	0.531	0.128	0.111	0.102	0.069	0.050	0.009
Cumulative	0.531	0.658	0.770	0.872	0.940	0.991	1.000
Variable	PC1	PC2	PC3				
Relac Líder	0.400	-0.066	-0.202				
Sem stress	0.300	-0.529	-0.652				
Refeição	0.251	0.801	-0.369				
Chance progredir	0.421	0.122	0.212				
Faço parte	0.390	0.008	0.457				
Treinamento	0.321	-0.242	0.377				
Mediana	0.505	0.014	-0.039				

Pela análise da tabela de variáveis (“Relação com Líder”, “Sem Stress”, etc.) versus indicadores (PC1, PC2 e PC3) acima, percebemos que aquelas que mais contribuem para PC1 são “Relação com Líder”, “Chance de Progredir” e “Mediana Total”; já para PC2 aquelas que mais influenciam são “Refeição” (positiva) e “Sem stress” (negativa). Finalmente para PC3, há forte influencia positiva de “Faço Parte” e forte influencia negativa de “Faço Parte”.

Identificamos aí, a oposição de “Refeição” e “Sem stress” (PC2). Também, a oposição de “Faço Parte” e “Sem stress” (PC3).

**Visto que os 3 primeiros indicadores “Análise de Componentes Principais” explicam 77% da amostra; visto que a “Análise de Correspondência” trouxe poucos dados elucidativos do comportamento da amostra em relação as variáveis estudadas; visto que o melhor resultado da “Análise Discriminante” obtido pelo método linear resultou em apenas 61% de aderência; e visto também que a “Regressão Logística” trouxe uma concordância de razoável 70.9%, buscaremos outra maneira de segmentar a amostra, daquela utilizada até o momento neste trabalho (a saber, o agrupamento dos indivíduos através dos distintos resultados para a variável “Sem Stress”).**

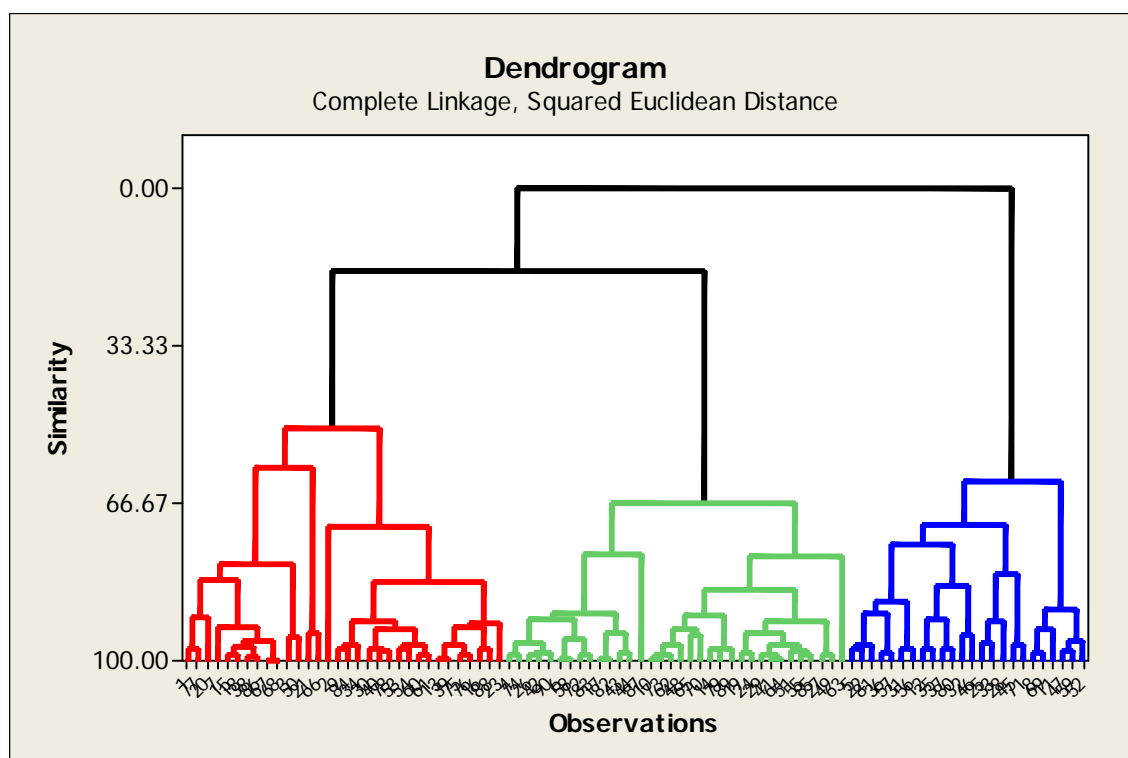
## 12. ANÁLISE DE CONGLOMERADOS – BUSCANDO NOVOS CLUSTERS

### 12.1. Escolhendo Método de Ligação (Linkage)

A divisão em clusters através dos métodos “Single Linkage”, “Mediana”, “Média” e “Centróide” pela mensuração de distância Euclidiana **não se mostraram possíveis**, pois sempre produziram um único grande cluster com diversos outliers; e mesmo quando estes foram sendo excluídos, sempre resultava num único grande cluster com novos outliers. Este método de “Single linkage” seria uma boa escolha se os clusters fossem claramente separados, quando as variáveis aparentam estar próximas, este método tende a identificar grandes clusters – como pudémos observar.

Outra opção seria o método de ligação “Ward's”, no qual a distancia entre 2 clusters é a soma dos quadrados dos desvios entre os pontos e os centróides. Este método tende a produzir clusters com número similar de variáveis, com a vantagem de ser sensível a outliers. Para a população e variáveis deste estudo, este método, através da mensuração de distância Euclidiana, produziu clusters numericamente equilibrados, mas com % de similaridade pequeno (cluster 1 = 12,76%; cluster 2 = - 52,20%; e cluster 3 = -53,78%).

O método “Complete Linkage” pela mensuração de distância Euclidiana resultou em um número adequado de indivíduos por cluster, mas também trouxe % de similaridade pequeno por cluster (cluster 1 = 29,84%; cluster 2 = 42,27%; e cluster 3 = 38,56%). Finalmente, através do método “Complete Linkage” pela mensuração de distância “Squared Euclidean Distance”, a distribuição de indivíduos foi numericamente coerente e percentuais de similaridade dos clusters aceitáveis (cluster 1 = 50,79%; cluster 2 = 66,67%; e cluster 3 = 62,24%), conforme é possível constatar abaixo.



## 12.2. Análise de Clusters

**Cluster Analysis of Observations: Relac Líder, Sem stress, Refeição, ...**

Squared Euclidean Distance, Complete Linkage  
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	89	100.000	0.0000	67 68	67	2
2	88	99.674	0.3125	41 64	41	2
3	87	99.674	0.3125	13 39	13	2
4	86	99.479	0.5000	6 58	6	2
5	85	99.414	0.5625	17 82	17	2
6	84	99.414	0.5625	10 73	10	2
7	83	99.219	0.7500	38 86	38	2
8	82	98.958	1.0000	15 18	15	2
9	81	98.893	1.0625	27 49	27	2
10	80	98.698	1.2500	55 85	55	2
11	79	98.698	1.2500	4 74	4	2
12	78	98.698	1.2500	10 62	10	3
13	77	98.698	1.2500	60 61	60	2
14	76	98.633	1.3125	72 83	72	2
15	75	98.633	1.3125	75 76	75	2
16	74	98.633	1.3125	26 42	26	2
17	73	98.633	1.3125	29 41	29	3
18	72	98.438	1.5000	8 69	8	2
19	71	98.438	1.5000	43 44	43	2
20	70	98.372	1.5625	19 24	19	2
21	69	98.177	1.7500	26 90	26	3
22	68	98.177	1.7500	36 57	36	2
23	67	98.112	1.8125	51 55	51	3
24	66	97.917	2.0000	47 59	47	2
25	65	97.852	2.0625	40 78	40	2
26	64	97.656	2.2500	16 88	16	2
27	63	97.656	2.2500	9 84	9	2
28	62	97.656	2.2500	5 22	5	2
29	61	97.591	2.3125	14 79	14	2
30	60	97.591	2.3125	1 77	1	2
31	59	97.396	2.5000	31 56	31	2
32	58	97.396	2.5000	53 54	53	2
33	57	97.396	2.5000	30 40	30	3
34	56	97.396	2.5000	12 35	12	2
35	55	97.331	2.5625	14 89	14	3
36	54	97.331	2.5625	37 80	37	2
37	53	97.135	2.7500	15 38	15	4
38	52	96.810	3.0625	45 71	45	2
39	51	96.810	3.0625	53 60	53	4
40	50	96.810	3.0625	9 34	9	3
41	49	96.549	3.3125	5 81	5	3
42	48	96.549	3.3125	13 75	13	4
43	47	96.549	3.3125	29 51	29	6
44	46	96.549	3.3125	10 48	10	4
45	45	96.289	3.5625	4 26	4	5
46	44	96.094	3.7500	25 33	25	2
47	43	95.833	4.0000	15 67	15	6
48	42	95.833	4.0000	47 52	47	3
49	41	95.313	4.5000	6 72	6	4
50	40	94.987	4.8125	3 50	3	2
51	39	94.727	5.0625	32 46	32	2
52	38	94.531	5.2500	65 70	65	2
53	37	94.271	5.5000	17 43	17	4
54	36	94.206	5.5625	21 66	21	2
55	35	94.206	5.5625	19 29	19	8
56	34	93.490	6.2500	10 65	10	6
57	33	93.424	6.3125	30 53	30	7
58	32	93.229	6.5000	8 11	8	3
59	31	92.904	6.8125	13 16	13	6
60	30	92.904	6.8125	7 15	7	7
61	29	92.122	7.5625	13 23	13	7
62	28	91.862	7.8125	25 28	25	3
63	27	91.862	7.8125	19 27	19	10
64	26	91.602	8.0625	9 30	9	10

65	25	91.146	8.5000	12	37	12	4
66	24	91.081	8.5625	4	6	4	9
67	23	90.820	8.8125	1	20	1	3
68	22	90.365	9.2500	10	14	10	9
69	21	90.039	9.5625	4	17	4	13
70	20	89.844	9.7500	5	36	5	5
71	19	89.323	10.2500	8	47	8	6
72	18	87.435	12.0625	5	31	5	7
73	17	85.156	14.2500	10	19	10	19
74	16	84.375	15.0000	12	32	12	6
75	15	83.268	16.0625	9	13	9	17
76	14	82.747	16.5625	1	7	1	10
77	13	81.510	17.7500	25	45	25	5
78	12	79.622	19.5625	1	3	1	12
79	11	77.865	21.2500	10	63	10	20
80	10	77.344	21.7500	4	87	4	14
81	9	75.521	23.5000	5	12	5	13
82	8	71.875	27.0000	2	9	2	18
83	7	71.094	27.7500	5	25	5	18
84	6	66.667	32.0000	4	10	4	34
85	5	62.240	36.2500	5	8	5	24
86	4	59.049	39.3125	1	21	1	14
87	3	50.781	47.2500	1	2	1	32
88	2	17.448	79.2500	1	4	1	66
89	1	0.000	96.0000	1	5	1	90

Final Partition

Number of clusters: 3

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	32	217.623	2.46566	4.44213
Cluster2	34	141.956	1.92277	3.69138
Cluster3	24	151.438	2.43548	3.30128

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Grand centroid
Relac Líder	4.07813	4.07353	2.64583	3.69444
Sem stress	4.09375	2.91176	1.62500	2.98889
Refeição	1.93750	3.67647	1.79167	2.55556
Chance progredir	3.06250	3.55882	1.66667	2.87778
Faço parte	3.68750	3.83824	2.58333	3.45000
Treinamento	3.85938	3.75000	2.77083	3.52778
Mediana	3.64844	3.72059	2.04167	3.24722

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3
Cluster1	0.00000	2.16964	3.88608
Cluster2	2.16964	0.00000	4.02233
Cluster3	3.88608	4.02233	0.00000

### 13. COMPARAÇÃO DE MÉDIAS, ANÁLISE DE VARIÂNCIA E INTERVALO DE CONFIANÇA (CLUSTERS “COMPLETE LINKAGE”)

#### 13.1. One-Way A-Nova “Relação com Líder” por Clusters “Complete Linkage”

**One-way ANOVA: Relac Líder versus CLUSTER Complete**

Source	DF	SS	MS	F	P
CLUSTER Complete	2	35.987	17.993	23.95	0.000
Error	87	65.360	0.751		
Total	89	101.347			

S = 0.8668    R-Sq = 35.51%    R-Sq(adj) = 34.03%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
1	32	4.0781	0.6969	3.3812	4.7750
2	34	4.0735	0.6529	3.4196	4.7274
3	24	2.6458	1.2552	1.3896	3.9020

Pooled StDev = 0.8668

#### 13.2. One-Way A-Nova “Sem Stress” por Clusters “Complete Linkage”

**One-way ANOVA: Sem stress versus CLUSTER Complete**

Source	DF	SS	MS	F	P
CLUSTER Complete	2	83.910	41.955	48.62	0.000
Error	87	75.079	0.863		
Total	89	158.989			

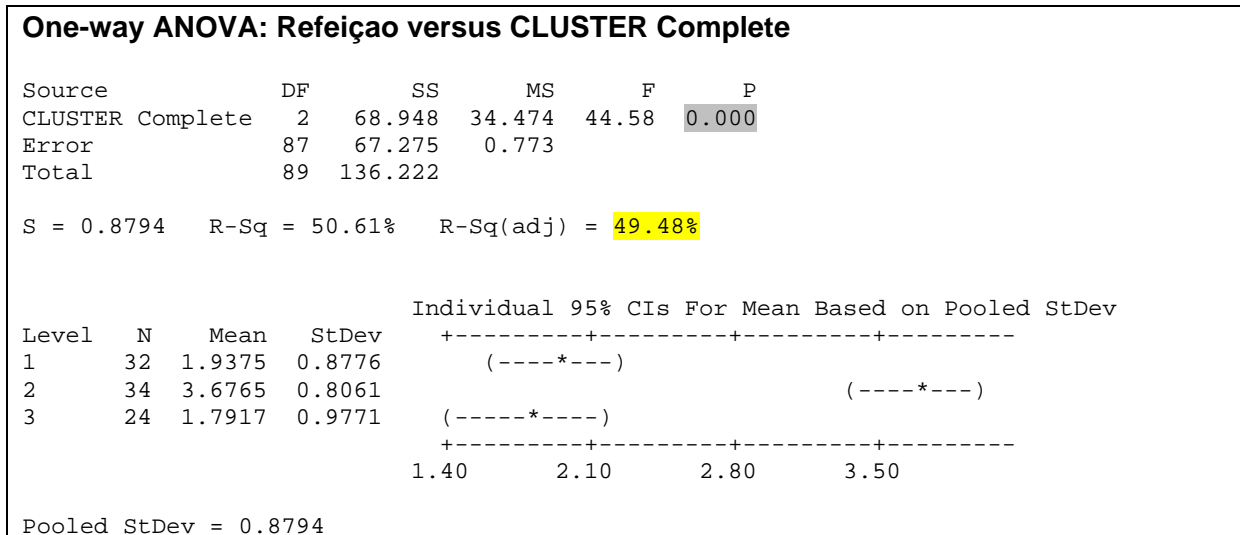
S = 0.9290    R-Sq = 52.78%    R-Sq(adj) = 51.69%

Individual 95% CIs For Mean Based on Pooled StDev

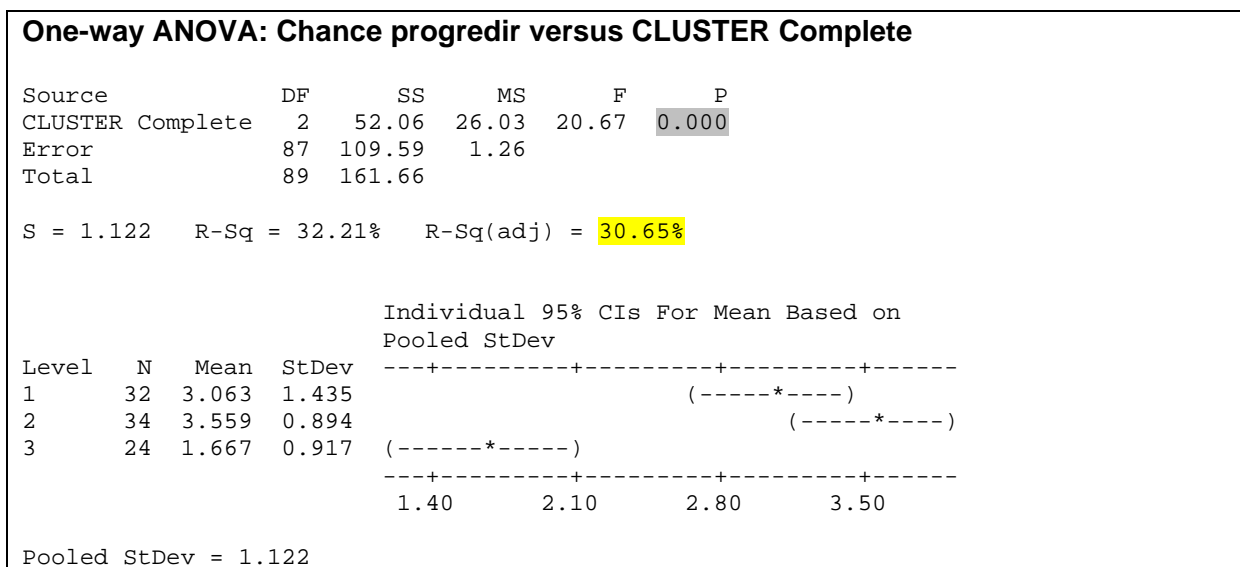
Level	N	Mean	StDev	CI Lower	CI Upper
1	32	4.0938	0.9955	3.0983	5.0893
2	34	2.9118	1.0260	1.8858	3.9378
3	24	1.6250	0.6469	1.0781	2.1719

Pooled StDev = 0.9290

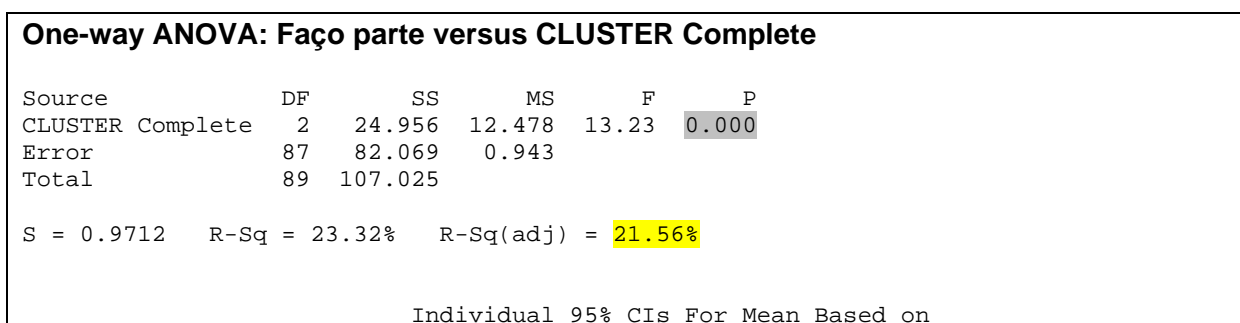
### 13.3. One-Way A-Nova “Refeição” por Clusters “Complete Linkage”



### 13.4. One-Way A-Nova “Chance de Progredir” por Clusters “Complete Linkage”



### 13.5. One-Way A-Nova “Faço Parte” por Clusters “Complete Linkage”



Level	N	Mean	StDev	Pooled StDev
1	32	3.6875	1.0453	(-----*-----)
2	34	3.8382	0.8139	(-----*-----)
3	24	2.5833	1.0700	(-----*-----)

Pooled StDev = 0.9712

### 13.6. One-Way A-Nova “Treinamento” por Clusters “Complete Linkage”

One-way ANOVA: Treinamento versus CLUSTER Complete					
Source	DF	SS	MS	F	P
CLUSTER Complete	2	18.949	9.474	10.88	0.000
Error	87	75.732	0.870		
Total	89	94.681			

S = 0.9330    R-Sq = 20.01%    R-Sq(adj) = 18.17%

Individual 95% CIs For Mean Based on Pooled StDev					
Level	N	Mean	StDev		
1	32	3.8594	1.0336	(-----*-----)	
2	34	3.7500	0.6657	(-----*-----)	
3	24	2.7708	1.1031	(-----*-----)	

Pooled StDev = 0.9330

### 13.7. One-Way A-Nova “Mediana” por Clusters “Complete Linkage”

One-way ANOVA: Mediana versus CLUSTER Complete					
Source	DF	SS	MS	F	P
CLUSTER Complete	2	47.650	23.825	57.72	0.000
Error	87	35.911	0.413		
Total	89	83.562			

S = 0.6425    R-Sq = 57.02%    R-Sq(adj) = 56.04%

Individual 95% CIs For Mean Based on Pooled StDev					
Level	N	Mean	StDev		
1	32	3.6484	0.7429	(---*---)	
2	34	3.7206	0.5177	(---*---)	
3	24	2.0417	0.6580	(---*---)	

Pooled StDev = 0.6425

- Observa-se em todas as variáveis alto grau de confiabilidade da médias já que para todas elas o valor de P-Value é inferior a 5% (0,0%).

- As variáveis “Relação com Líder”, “Chance de Progredir”, “Faço Parte” e “Treinamento” apresentam valores de R-Sq (Adj) inferiores a 50% (34,03%, 30,65%, 21,56% e 18,17%, respectivamente), e portanto, demonstram menor correlação com os resultados dos “Clusters Complete Linkage”.
- As variáveis “Sem stress”, “Refeição”, “Mediana Total” apresentam valores de R-Sq (Adj) próximos ou superiores a 50% (51,69%, 49,48%, 56,04%, respectivamente), e portanto, demonstram boa correlação com os resultados dos “Clusters Complete Linkage”.
- Comparando-se visualmente as médias de cada variável, percebemos a “Sem stress” é a que melhor divide os clusters, pois nela os resultados aparecem bem demarcados, com pequeno desvio padrão.
- As demais variáveis não são capazes de dividir bem os clusters, pois ou as médias estão muito próximas; ou embora as médias sejam distantes, há grande desvio padrão. Todas as variáveis acabam sobrepondo as médias nos clusters 1 e 2, com exceção de “Refeição” que sobrepõe as médias nos clusters 1 e 3.
- Pela análise do valor F de cada variável, percebemos que os clusters são bastante próximos em todas as variáveis, com exceção de “Sem stress” e “Refeição”, embora nesta última ocorra o problema de sobreposição de médias que detalhamos no tópico anterior.

## 14. ANÁLISE DISCRIMINANTE

Através da “Análise Discriminante”, analisaremos qual modelo (linear ou quadrático) e quais variáveis representam a melhor a população.

### 14.1. Análise Discriminante – Modelo Linear

A “Análise Discriminante” pelo método linear com todas as variáveis, demonstra uma aderência de 94,4%. Nesta análise apenas 5 indivíduos ficaram “mal classificados”.

#### Discriminant Analysis: CLUSTER Comp versus Relac Líder, Sem stress, ...

Linear Method for Response: CLUSTER Complete

Predictors: Relac Líder, Sem stress, Refeição, Chance progredir, Faço parte, Treinamento, Mediana

Group	1	2	3
Count	32	34	24

Summary of classification



Put into Group	True Group		
	1	2	3
1	31	1	0
2	1	30	0
3	0	3	24
Total N	32	34	24
N correct	31	30	24
Proportion	0.969	0.882	1.000

N = 90                      N Correct = 85                      Proportion Correct = 0.944

#### Squared Distance Between Groups

	1	2	3
1	0.0000	9.4385	13.2665
2	9.4385	0.0000	9.1670
3	13.2665	9.1670	0.0000

#### Linear Discriminant Function for Groups

	1	2	3
Constant	-28.449	-28.593	-12.251
Relac Líder	4.073	4.738	3.846
Sem stress	4.246	2.476	1.690
Refeição	0.403	4.333	2.489
Chance progredir	-0.470	1.215	0.496
Faço parte	2.213	2.744	2.442
Treinamento	3.462	4.045	3.532
Mediana	0.560	-4.107	-4.800

#### Summary of Misclassified Observations

Observation	True Group	Pred Group	Squared Distance	Probability	
10**	2	3	1	6.145	0.199
			2	5.561	0.267
			3	4.171	0.534
53**	1	2	1	8.936	0.222
			2	6.889	0.617
			3	9.562	0.162
62**	2	3	1	8.215	0.068
			2	5.910	0.214
			3	3.493	0.718
73**	2	3	1	8.221	0.056
			2	5.437	0.225
			3	3.110	0.719
74**	2	1	1	1.977	0.830
			2	5.167	0.168
			3	14.249	0.002

Após inúmeras tentativas de combinar variáveis, o melhor resultado de 94,4% obtido com todas as variáveis foi mantido utilizando-se todas elas, exceto “Faço Parte”, o que torna o modelo um pouco mais parcimonioso. Segue abaixo o detalhe desta análise.

## Discriminant Analysis: CLUSTER Comp versus Relac Líder, Sem stress, ...

Linear Method for Response: CLUSTER Complete

Predictors: Relac Líder, Sem stress, Refeição, Chance progredir, Treinamento, Mediana

Group	1	2	3
Count	32	34	24

### Summary of classification

	True Group		
Put into Group	1	2	3
1	31	1	0
2	1	30	0
3	0	3	24
Total N	32	34	24
N correct	31	30	24
Proportion	0.969	0.882	1.000

N = 90

N Correct = 85

Proportion Correct = 0.944

### Squared Distance Between Groups

	1	2	3
1	0.0000	9.2713	13.2353
2	9.2713	0.0000	9.1131
3	13.2353	9.1131	0.0000

### Linear Discriminant Function for Groups

	1	2	3
Constant	-27.000	-26.366	-10.487
Relac Líder	3.732	4.315	3.469
Sem stress	4.009	2.182	1.428
Refeição	-0.113	3.693	1.920
Chance progredir	-0.345	1.369	0.633
Treinamento	3.315	3.863	3.370
Mediana	2.974	-1.113	-2.135

### Summary of Misclassified Observations

Observation	True Group	Pred Group	Squared Distance	Probability	
10**	2	3	1	6.136	0.189
			2	5.308	0.286
			3	4.098	0.524
53**	1	2	1	7.591	0.287
			2	6.325	0.540
			3	8.596	0.173
62**	2	3	1	8.015	0.060
			2	5.179	0.246
			3	3.104	0.694
73**	2	3	1	7.564	0.065
			2	5.276	0.203
			3	2.709	0.732
74**	2	1	1	1.876	0.836
			2	5.158	0.162
			3	14.230	0.002

## 14.2. Análise Discriminante – Modelo Quadrático

A “Análise Discriminante” pelo método quadrático com todas as variáveis, demonstra uma aderência de 98,9%. Nesta análise apenas 1 indivíduo fica “mal classificado”.

Discriminant Analysis: CLUSTER Comp versus Relac Líder, Sem stress, ...				
Quadratic Method for Response: CLUSTER Complete				
Predictors: Relac Líder, Sem stress, Refeição, Chance progredir, Faça parte, Treinamento, Mediana				
Group	1	2	3	
Count	32	34	24	
Summary of classification				
	True Group			
Put into Group	1	2	3	
1	32	1	0	
2	0	33	0	
3	0	0	24	
Total N	32	34	24	
N correct	32	33	24	
Proportion	1.000	0.971	1.000	
N = 90	N Correct = 89		Proportion Correct = 0.989	
From Generalized Squared Distance to Group				
Group	1	2	3	
1	-4.194	3.244	14.332	
2	7.372	-7.568	8.423	
3	12.104	4.624	-3.883	
Summary of Misclassified Observations				
Observation	True Group	Pred Group	Squared Distance	Probability
74**	2	1	1 -0.8277	0.676
			2 0.6442	0.324
			3 13.1070	0.001

Após inúmeras tentativas de combinar variáveis, o melhor resultado de 98,9% obtido com todas as variáveis foi mantido utilizando-se todas elas, exceto “Faço Parte”, o que torna o modelo um pouco mais parcimonioso. Segue abaixo o detalhe desta análise.

## Discriminant Analysis: CLUSTER Comp versus Relac Líder, Sem stress, ...

Quadratic Method for Response: CLUSTER Complete

Predictors: Relac Líder, Sem stress, Refeição, Chance progredir, Treinamento, Mediana

Group	1	2	3
Count	32	34	24

Summary of classification

Put into Group	True Group		
	1	2	3
1	32	1	0
2	0	33	0
3	0	0	24
Total N	32	34	24
N correct	32	33	24
Proportion	1.000	0.971	1.000

N = 90

N Correct = 89

Proportion Correct = 0.989

From Generalized Squared Distance to Group

Group	1	2	3
1	-3.302	2.637	15.046
2	7.844	-6.288	9.115
3	10.695	5.044	-3.044

Summary of Misclassified Observations

Observation	True Group	Pred Group	Squared Distance	Probability
74**	2	1	-1.045	0.651
			2	0.208
			3	13.946

## 15. REGRESSAO LOGÍSTICA

### 15.1. Ordenando os clusters

Retomando a “Análise de Clusters” (item 12.2, parcialmente ilustrada abaixo) pode-se afirmar, através da comparação de valores dos centróides, que o pior cluster é o 3. Embora os outros 2 clusters sejam bastante equivalentes, o cluster 2 é ainda melhor já que possui valores levemente maiores.

Cluster Centroids

Variable	Cluster1	Melhor	Pior	Grand centroid
		Cluster2	Cluster3	
Relac Líder	4.07813	4.07353	2.64583	3.69444
Sem stress	4.09375	2.91176	1.62500	2.98889
Refeição	1.93750	3.67647	1.79167	2.55556
Chance progredir	3.06250	3.55882	1.66667	2.87778
Faço parte	3.68750	3.83824	2.58333	3.45000
Treinamento	3.85938	3.75000	2.77083	3.52778
Mediana	3.64844	3.72059	2.04167	3.24722

## 15.2. Regressão Logística Ordinal

**Ordinal Logistic Regression: CLUSTER Comp versus Relac Líder, Sem stress, ...**

Link Function: Logit

Response Information

Variable	Value	Count
CLUSTER Complete	1	32
	2	34
	3	24
Total		90

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds	95% CI	
					Ratio	Lower	Upper
Const(1)	-11.7074	2.32066	-5.04	0.000			
Const(2)	-7.26475	1.83937	-3.95	0.000			
Relac Líder	0.0980251	0.384623	0.25	0.799	1.10	0.52	2.34
Sem stress	1.36616	0.332110	4.11	0.000	3.92	2.04	7.52
Refeição	-1.56117	0.365340	-4.27	0.000	0.21	0.10	0.43
Chance progredir	-0.672055	0.400820	-1.68	0.094	0.51	0.23	1.12
Faço parte	-0.264648	0.356285	-0.74	0.458	0.77	0.38	1.54
Treinamento	-0.200945	0.425953	-0.47	0.637	0.82	0.35	1.88
Mediana	4.02410	1.18527	3.40	0.001	55.93	5.48	570.89

Log-Likelihood = -46.237  
 Test that all slopes are zero: G = 103.346, DF = 7, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	115.442	169	0.999
Deviance	92.473	169	1.000

Measures of Association:  
 (Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	2515	94.1	Somers' D 0.88
Discordant	155	5.8	Goodman-Kruskal Gamma 0.88
Ties	2	0.1	Kendall's Tau-a 0.59
Total	2672	100.0	

Apesar do modelo de regressão logística ordinal apresentar concordância alta de 94,1 %, mostra-se inválido, pois grande parte dos valores de P são todos superiores a 0,05, significando que os valores apresentados possuem pouca previsibilidade devido à enorme variância. Assim sendo, buscaremos eliminar do modelo as variáveis que apresentaram valores de P-Value maiores que 0,05.

### Ordinal Logistic Regression: CLUSTER Comp versus Sem stress, Refeição, ...

Link Function: Logit

#### Response Information

Variable	Value	Count
CLUSTER Complete	1	32
	2	34
	3	24
Total		90

#### Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds	95% CI	
					Ratio	Lower	Upper
Const(1)	-11.3147	2.00835	-5.63	0.000			
Const(2)	-6.92252	1.46351	-4.73	0.000			
Sem stress	1.51088	0.320495	4.71	0.000	4.53	2.42	8.49
Refeição	-1.40264	0.332999	-4.21	0.000	0.25	0.13	0.47
Mediana	2.66927	0.555736	4.80	0.000	14.43	4.86	42.88

Log-Likelihood = -48.347

Test that all slopes are zero: G = 99.125, DF = 3, P-Value = 0.000

#### Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	108.542	133	0.941
Deviance	96.694	133	0.992

#### Measures of Association:

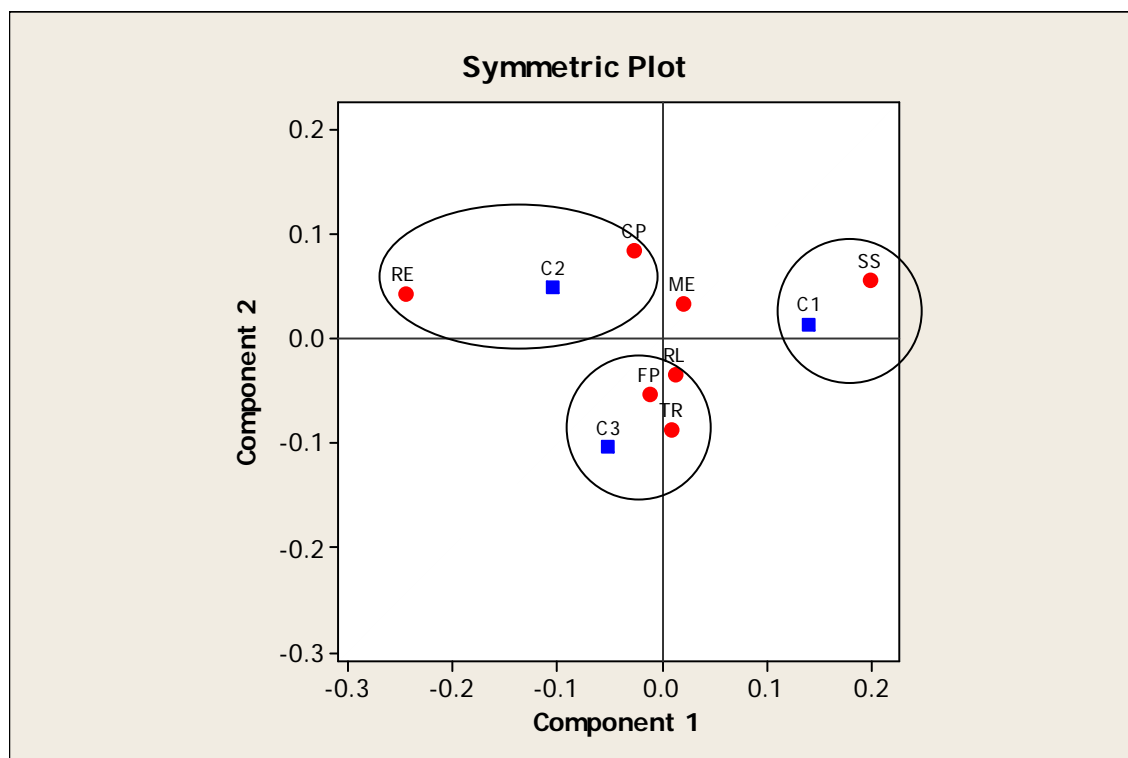
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	2476	92.7	Somers' D 0.85
Discordant	196	7.3	Goodman-Kruskal Gamma 0.85
Ties	0	0.0	Kendall's Tau-a 0.57
Total	2672	100.0	

Nessa nova tentativa de utilizar o modelo de regressão logística ordinal excluindo as variáveis com valores altos para P-value, obtivemos alta concordância de 92,7%, com P-Value da “Tabela de Regressão Logística” com resultados menores de 5% (todos 0.00), significando que os valores apresentados possuem boa previsibilidade devido à pequena variância. Além disso, o P-Value do Teste “Goodness-of-Fit”, que analisa a diferença entre os valores observados e os valores esperados pelo modelo, obteve 0,941 para Pearson e 0,992 para “Deviance”, indicando que **não há evidências para afirmar que o modelo não se ajusta adequadamente aos dados**. Sendo assim, o modelo é bastante válido.

### 15.3. Symmetric Plot (Variáveis x Clusters “Complete Linkage”)

Abaixo é apresentado o resultado da análise de correspondência para os 3 novos clusters criados através de “Complete Linkage” x 7 variáveis estudadas.



No gráfico acima percebemos que dentre as variáveis estudadas as mais próximas ao Cluster 1 é “Sem stress”; do Cluster 2 são “Refeição” e “Chance de Progredir”; e do Cluster 3, “Relação com Líder”, “Treinamento” e “Faço Parte”. Não avaliaremos a variável “Mediana Total” que parece estar bastante equidistante dos 3 clusters.

### 15.4. Tabela de Análise de Contigência e Contribuições

#### Simple Correspondence Analysis: Média C1, Média C2, Média C3

Analysis of Contingency Table

Axis	Inertia	Proportion	Cumulative	Histogram
1	0.0123	0.7788	0.7788	*****
2	0.0035	0.2212	1.0000	*****
Total	0.0158			

Row Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	RL	1.000	0.166	0.014	0.012	0.118	0.002	-0.034	0.882	0.054
2	SS	1.000	0.133	0.358	<b>0.199</b>	<b>0.927</b>	<b>0.426</b>	0.056	0.073	0.118
3	RE	1.000	0.114	0.446	<b>-0.245</b>	<b>0.971</b>	<b>0.556</b>	0.042	0.029	0.058
4	CP	1.000	0.127	0.063	-0.026	0.090	0.007	<b>0.084</b>	<b>0.910</b>	<b>0.258</b>
5	FP	1.000	0.155	0.029	-0.011	0.043	0.002	-0.053	0.957	0.126
6	TR	1.000	0.160	0.077	0.010	0.012	0.001	<b>-0.087</b>	<b>0.988</b>	<b>0.342</b>
7	ME	1.000	0.145	0.014	0.021	0.295	0.005	0.033	0.705	0.044

Column Contributions										
ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	C1	1.000	0.375	0.476	<b>0.141</b>	<b>0.991</b>	<b>0.605</b>	0.014	0.009	0.020
2	C2	1.000	0.393	0.326	-0.104	0.822	0.345	0.048	0.178	0.263
3	C3	1.000	0.233	0.198	-0.052	0.199	0.050	<b>-0.104</b>	<b>0.801</b>	<b>0.717</b>

Através da tabela da “Análise de Contingência”, que mostra decomposição da inércia ( $\chi^2/n$ ), identificamos que, do total da inércia da matriz de dados, 77,88% é contabilizada no primeiro componente; e 22,12 % é contabilizada no segundo componente. Assim os dois primeiros já contabilizam 100%. Detalhando temos que:

- Analisando as variáveis (linhas), o Componente 1 contribui mais para a inércia de “Sem Stress” e “Refeição” (Corr = 0.927 e 0.971, respectivamente) em direções opostas (valores de “Coord” positivo e negativo, respectivamente). Já o Componente 2 contribui mais para a inércia de “Chance de Progredir” e “Treinamento” (Corr = 0.910, 0.988, respectivamente) também em direções opostas. As demais variáveis não possuem valores de “Corr” e “Contr” relevantes.
- Analisando os clusters (colunas), o Componente 1 contribui mais para a inércia dos cluster 1 (Corr = 0.991). O Componente 2 contribui mais para a inércia do cluster 3 (Corr = 0.801).

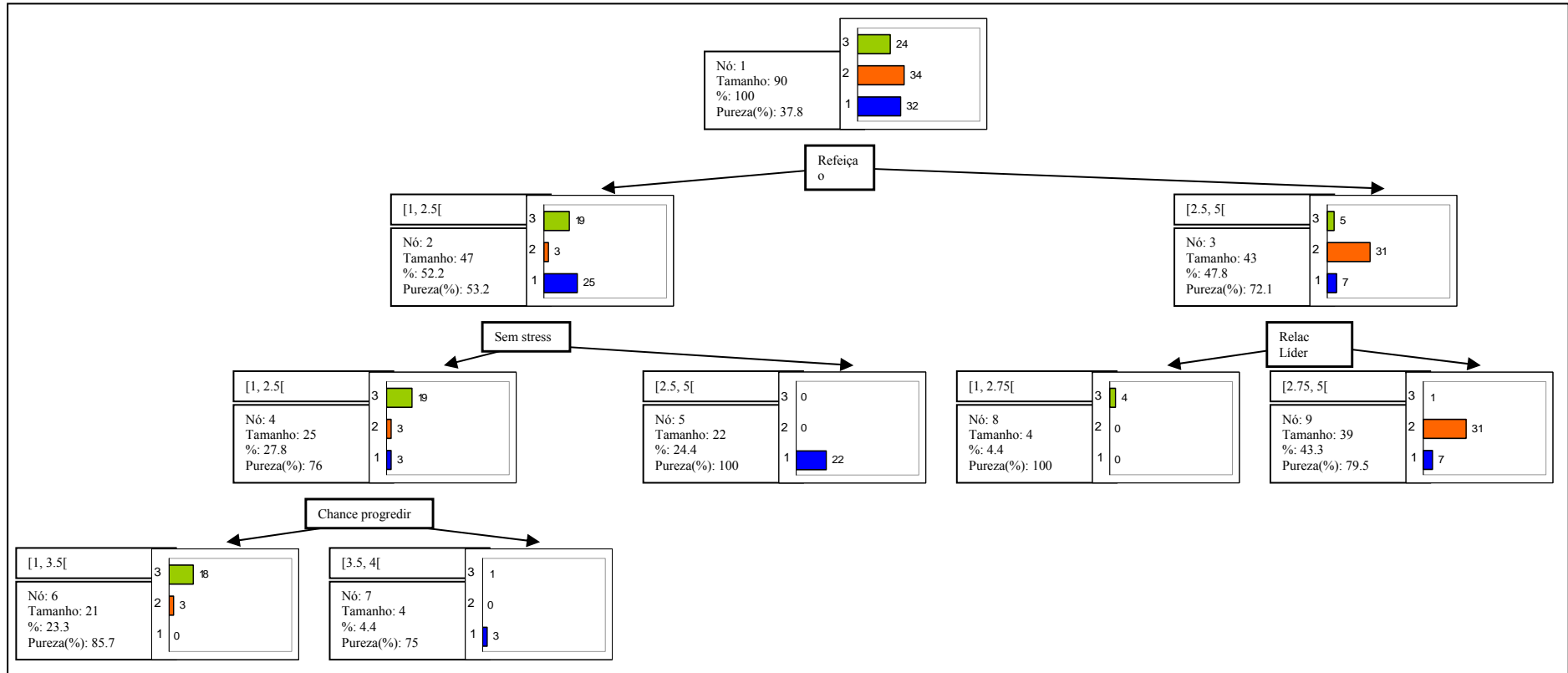
Resumindo, temos que as variáveis “Sem stress” (positiva) e “Refeição” (negativa) e o cluster 1 mais contribuem para o **componente 1**. As variáveis “Chance de Progredir” (positiva) e “Treinamento” (negativa) e o cluster 3 mais contribuem para o **componente 2**. O cluster 2 e as demais variáveis mostram interferir pouco nestes componentes.

## 16. ÁRVORE DE CLASSIFICAÇÃO

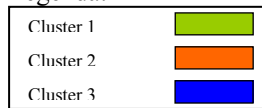
Para obter a “Árvore de Classificação” utilizaremos o aplicativo XLSTAT. Esse aplicativo indica qual a variável melhor separa os grupos e classifica as variáveis por ordem de importância na separação dos grupos. A seguir é demonstrado o teste desse modelo.



### 14.1 Árvore de Classificação – Clusters “Complete Linkage”



Legenda:



## 14.2 Estatísticas Descritivas

Variável	Observações	Obs. com	Obs. sem	Mínimo	Máximo	Média	Desvio padrão
		dados faltantes	dados faltantes				
Relac Líder	90	0	90	1.000	5.000	3.694	1.067
Sem stress	90	0	90	1.000	5.000	2.989	1.337
Refeição	90	0	90	1.000	5.000	2.556	1.237
Chance progredir	90	0	90	1.000	5.000	2.878	1.348
Faço parte	90	0	90	1.000	5.000	3.450	1.097
Treinamento	90	0	90	1.000	5.000	3.528	1.031

Variável	Categorias	Frequências	%
CLUSTER Complete	1	32	35.556
	2	34	37.778
	3	24	26.667

Variáveis	Relac	Sem stress	Refeição	Chance	Faço parte	Treinamento
	Líder			progredir		
Relac Líder	<b>1.000</b>	0.419	0.279	0.564	0.457	0.322
Sem stress	0.419	<b>1.000</b>	0.133	0.305	0.276	0.290
Refeição	0.279	0.133	<b>1.000</b>	0.317	0.236	0.181
Chance progredir	0.564	0.305	0.317	<b>1.000</b>	0.615	0.350
Faço parte	0.457	0.276	0.236	0.615	<b>1.000</b>	0.409
Treinamento	0.322	0.290	0.181	0.350	0.409	<b>1.000</b>

## 14.3 Estrutura da Árvore

Nó	p-valor	Objetos	%	Nó pai	Filhos	Variável de separação	Valores	Pureza
1	0.569	90	100.00%		2, 3			37.78%
2	0.740	47	52.22%	1	4, 5	Refeição	[1, 2.5[	53.19%
3	0.742	43	47.78%	1	8, 9	Refeição	[2.5, 5[	72.09%
4	0.713	25	27.78%	2	6, 7	Sem stress	[1, 2.5[	76.00%
5	<b>0.000</b>	22	24.44%	2		Sem stress	[2.5, 5[	<b>100.00%</b>
6	<b>0.000</b>	21	23.33%	4		Chance progredir	[1, 3.5[	<b>85.71%</b>
7	<b>0.000</b>	4	4.44%	4		Chance progredir	[3.5, 4[	<b>75.00%</b>
8	<b>0.000</b>	4	4.44%	3		Relac Líder	[1, 2.75[	<b>100.00%</b>
9	0.841	39	43.33%	3		Relac Líder	[2.75, 5[	79.49%

Analisando-se os nós obtidos pela árvore de classificação, podemos considerar válidos apenas os nós 5, 6, 7 e 8 – que apresentam P-Value menores que 5%, conforme do Teste de normalidade de Anderson-Darling como parâmetro no teste de hipótese. Assim sendo, os demais nós (1, 2, 3, 4 e 9) não seriam valores válidos para diferenciar os clusters apresentados. Nenhum dos nós tidos como válidos pelo valor de P-Value é descartado devido ao baixo grau de pureza (todos acima de 75%).

### 14.4 Matriz de Confusão

A “Matriz de confusão” abaixo é capaz de estimar 86,67% da divisão desta população em clusters:

de \ a	1	2	3	Total	% correto
1	25	7	0	32	78.13%
2	0	31	3	34	91.18%
3	1	1	22	24	91.67%
Total	26	39	25	90	86.67%

### 14.5 Análise das Réguas

Através da análise das réguas abaixo, confirma-se que os nós válidos são 5 e 6 devido á alta frequência e alta pureza. Os nós 7 e 8 apesar de boa pureza possuem baixa frequência.

Nó	Pred(CLUSTER Complete)	Frequência	Pureza	Réguas
Nó1	2.000	34	37.78%	
Nó2	1.000	25	53.19%	Se Refeição em [1, 2.5[ então CLUSTER Complete = 1 em 53.2% dos casos
Nó3	2.000	31	72.09%	Se Refeição em [2.5, 5[ então CLUSTER Complete = 2 em 72.1% dos casos
Nó4	3.000	19	76.00%	Se Sem stress em [1, 2.5[ e Refeição em [1, 2.5[ então CLUSTER Complete = 3 em 76% dos casos
Nó5	1.000	22	100.00%	Se Sem stress em [2.5, 5[ e Refeição em [1, 2.5[ então CLUSTER Complete = 1 em 100% dos casos
Nó6	3.000	18	85.71%	Se Chance progredir em [1, 3.5[ e Sem stress em [1, 2.5[ e Refeição em [1, 2.5[ então CLUSTER Complete = 3 em 85.7% dos casos
Nó7	1.000	3	75.00%	Se Chance progredir em [3.5, 4[ e Sem stress em [1, 2.5[ e Refeição em [1, 2.5[ então CLUSTER Complete = 1 em 75% dos casos
Nó8	3.000	4	100.00%	Se Relac Líder em [1, 2.75[ e Refeição em [2.5, 5[ então CLUSTER Complete = 3 em 100% dos casos
Nó9	2.000	31	79.49%	Se Relac Líder em [2.75, 5[ e Refeição em [2.5, 5[ então CLUSTER Complete = 2 em 79.5% dos casos

Da árvore analisada conclui-se que as variáveis “Refeição”, “Chance de Progredir” e “Sem stress” são as que melhor podem prever a divisão de clusters, conforme abaixo.

- 100% de chance de ser cluster 1 quando “Refeição” tiver valores entre 2.5 e 5.
- 85.71% de chance de ser cluster 3 quando “Chance de Progredir” tiver valores entre 1 e 3.5 e quando “Sem stress” oscilar entre 1 e 2.5 e “Refeição” estiver entre 1 e 2.5.

## CONCLUSÕES

Curiosamente a mediana da variável “Sem stress” melhora os resultados conforme aumenta a faixa etária, com exceção de indivíduos que possuem menos de 25 anos.

Observamos significativa correlação entre as variáveis “Chance de Progredir” e “Relação com o Líder” (0,564); também entre “Chance de Progredir” e “Faço Parte” (0.615), denotando que a qualidade da relação com o líder possa ter influencia na percepção de chance para progredir e sentimento de pertencer à empresa.

A análise das medianas das variáveis em subcategorias demonstra:

- Os valores melhores para “Relação com Líder” e “Faço Parte” em funcionários com mais de 20 anos de empresa; e em “Chance de Progredir” os melhores valores para funcionários com menos de 1 ano e também para aqueles com mais de 20 anos;
- Que em “Relação com Líder” os valores crescem conforme aumenta a escolaridade, exceto para pós graduação e Fundamental; e que em “Chance de Progredir” os resultados são destacadamente melhores para os funcionários com maior escolaridade, exceto também Pós graduação;
- Os valores melhores para “Relação com Líder” em indivíduos com menos de 25 anos e para aqueles com mais de 55 anos; e em “Chance de Progredir” e “Faço Parte” os piores valores para indivíduos com mais de 55 anos, e os melhores resultados para aqueles entre 44 e 55 anos.

A fim de buscar um modelo preditivo para as variáveis utilizou-se diferentes modelos para 2 opções de segmentação da amostra:

- 1) Agrupamento dos indivíduos através dos distintos resultados para a variável “Sem Stress”: grupo 1 ‘Discordancia’ com respostas 1 (discordo totalmente) e 2 (discordo); grupo 2 ‘neutralidade’ com respostas 3 (nem concordo, nem discordo); e grupo 3 ‘concordancia’ com respostas 4 (concordo) e 5 (concordo totalmente);
- 2) Formação de Clusters através do método “Complete Linkage” pela mensuração de distância Euclidiana.

A primeira opção de segmentação (Grupos de “Sem Stress”) mostrou-se pouco elucidativa, conforme os resultados para cada método preditivo utilizado abaixo:

- Em “*Análise de Componentes Principais*” os 3 primeiros indicadores explicam 77% da amostra;
- A “*Análise de Correspondência*” trouxe poucos dados relevantes em relação as variáveis estudadas;
- O melhor resultado da “*Análise Discriminante*” pelo método linear resultou em apenas 61% de aderência;
- O melhor resultado da “*Análise Discriminante*” pelo método quadrático resultou 60% de aderência;

- A “*Regressão Logística Ordinal*”, excluindo as variáveis com valores altos para P-value, trouxe uma concordância de razoável 70.9%.

A segunda opção de segmentação (Clusters por “Complete Linkage”) mostrou-se muito mais elucidativa, conforme os resultados para cada método preditivo utilizado abaixo:






- O melhor resultado da “*Análise Discriminante*” pelo método linear resultou 94,4% de aderência, utilizando-se todas as variáveis, exceto “Faço Parte”;
- O melhor resultado da “*Análise Discriminante*” pelo método quadrático resultou 98,9% de aderência, obtido com todas as variáveis, exceto “Faço Parte”;
- A “*Regressão Logística Ordinal*”, excluindo as variáveis com valores altos para P-value, resultou alta concordância de 92,7%;
- A “*Análise de Contingência*” mostra decomposição da inércia de 77,88% no primeiro componente, o qual mais influencia “Sem Stress” (“Corr” = 0,927 e “Contr” = 0,426), “Refeição” (“Corr” = 0,971 e “Contr” = 0,556) e o Cluster 1 (“Corr” = 0,991 e “Contr” = 0,605); e inércia de 22,12 % contabilizada no segundo componente que mais influencia “Chance de Progredir” (“Corr” = 0,910 e “Contr” = 0,258), “Treinamento” (“Corr” = 0,988 e “Contr” = 0,342) e Cluster 3 (“Corr” = 0,801 e “Contr” = 0,717);
- Na “*Análise de Componentes Principais*” os 3 primeiros indicadores explicam 77% da amostra;
- Na “*Árvore de Classificação*” a “Matriz de Confusão” é capaz de estimar 86,67% da divisão dos clusters;

Assim sendo, **o modelo que melhor explica esta população é a “*Análise Discriminante*” pelo método quadrático** que resultou 98,9% de aderência.

## Anexo 1 – Formulário Completo de Coleta de Dados

**PESQUISA CLIMA INTERNO**

Assinale com X o quanto você concorda ou discorda:	Discordo totalmente	Discordo	Não concordo nem discordo	Concordo	Concordo Totalmente
--	---------------------	----------	---------------------------	----------	---------------------

					
Sou tratado pelo meu chefe com respeito e educação					
Eu não faço muitas horas extras					
Na minha área as pessoas costumam assumir erros					
Nesta empresa o bom profissional pode crescer e progredir					
Sou solicitado a ajudar a resolver problemas e propor soluções					
Estou satisfeito com o turno 12x36					
Gosto da refeição que é servida na empresa					
Sinto que sou parte importante da empresa					
Na minha área as pessoas cooperam umas com as outras					
Na minha área há poucos desperdícios					
Aqui as pessoas cuidam bem das máquinas e equipamentos					
A empresa sempre busca alta produtividade e qualidade					
Aqui todos os funcionários são ouvidos pela chefia					
Recebo orientação suficiente para executar minha função					
Eu compreendo o sistema de consequência da empresa					
A consequência ajuda as pessoas serem mais comprometidas					
Não temos uma carga estressante de trabalho					
O treinamento que recebo me prepara para o trabalho que faço					
Estou satisfeito em poder participar de uma pesquisa como esta					

Quantos anos você tem de serviço?

- ( ) menos de 1 ano  
 ( ) 1 a 2 anos  
 ( ) 3 a 5 anos  
 ( ) 6 a 10 anos  
 ( ) 11 a 20 anos  
 ( ) mais de 20 anos

Qual seu sexo?

- ( ) Feminino  
 ( ) Masculino

Qual seu turno?

- ( ) A  
 ( ) B  
 ( ) C  
 ( ) D  
 ( ) Administrativo

Qual sua escolaridade?

- Fundamental
- Ensino Médio Incompleto
- Ensino Médio Completo
- Superior Incompleto
- Superior Completo
- Pós

Sua idade?

- menos de 25 anos
- entre 25 e 34 anos
- entre 35 e 44 anos
- entre 45 e 54 anos
- 55 ou mais