



**PUC-SP**

**PONTÍFICA UNIVERSIDADE CATÓLICA DE SÃO PAULO**

**PROGRAMA DE ESTUDOS PÓS-GRADUADOS EM  
ADMINISTRAÇÃO/ PUC-SP**

**MÉTODOS QUANTITATIVOS**

**TRABALHO FINAL**

**1º SEMESTRE 2011**

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO**  
FEA - Faculdade de Economia e Administração

Programa de Estudos Pós-Graduados em Administração

## **MÉTODOS**

### **ANÁLISE EXPLORATÓRIA DE DADOS**

#### **TRABALHO E RENDA**

Disciplina: Métodos Quantitativos

Professor: Dr. Arnaldo Jose de Hoyos

**ROSIMAR PEREIRA BARBOSA**

## 1. INTRODUÇÃO

O presente trabalho tem por objetivo efetuar uma análise de tendência e projeções das variáveis temporais e quantitativas “População em Idade Ativa”, “População ocupada (mil)”, “População ocupada (%)”, “População não economicamente ativa (%)”, “Rendimento médio real efetivo (R\$)” e “Rendimento médio real habitual (R\$)”. Todos os dados utilizados são referentes ao Brasil nas regiões metropolitanas (Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre). Estas variáveis integram a pesquisa IPEA DATA. A análise de cada variável está dividida em três partes. A primeira refere-se a análise do comportamento histórico através de gráficos. A segunda trata da análise de tendências, buscando encontrar a função que melhor se adapte à cada uma das séries de variáveis analisadas. A terceira busca a elaboração e análise de projeções através de extrapolações estatísticas. O software estatístico utilizado é o **MINITAB 16** e o **XLSTAT**.

### 1.1 UMA BREVE DESCRIÇÃO DO IPEA DATA

Base de dados econômicos e financeiros mantida pelo IPEA incluindo séries estatísticas da economia brasileira e dos aspectos que lhe são mais pertinentes na economia internacional. Os dados são atualizados e documentados de forma sistemática e apresentados na mesma unidade monetária. Recursos disponíveis permitem a manipulação matemática e a extração dos resultados em planilhas ou gráficos.

O objetivo é demonstrar as direções e a intensidade das mudanças para os próximos anos, buscando identificar os fatores responsáveis por estas alterações. São utilizadas 6 variáveis, que foram definidas e classificadas previamente levando em consideração a relevância para as previsões e a disponibilidade de dados nos últimos 10 anos nos diversos países que a compõe. Trata-se de um trabalho contínuo que vem sendo atualizado ano a ano.

De 2003 a 2010, número de trabalhadores com carteira assinada aumentou 38,7%, enquanto população ocupada cresceu 18,9%.

Além de reduzir a taxa de desemprego a quase metade em oito anos, o Brasil conseguiu formalizar o mercado de trabalho num ritmo ainda maior que o da criação de vagas. De 2003 – o primeiro ano completo da série histórica do Instituto Brasileiro de Geografia (IBGE) – a 2010, o número de trabalhadores com carteira assinada aumentou 38,7%. No mesmo período, o população ocupada cresceu 18,9%.

“A grande característica do mercado de trabalho neste período de oito anos foi a formalização. A carteira assinada é quase que um passaporte para as classes menos favorecidas e cresceu mais entre os mais pobres e no Nordeste”, afirmou o coordenador da Pesquisa Mensal de Emprego do IBGE, Cimar Azeredo.

Das seis regiões metropolitanas investigadas pelo IBGE, o Recife foi a capital onde a formalização mais cresceu, num avanço de 57,3%. Belo Horizonte (53,21%), Salvador (48,5%), São Paulo (39,2%) Porto Alegre (33,6%) e Rio (26,4%) também mostram esta tendência.

Entre os segmentos pesquisados, a construção civil foi a que mais aumentou a formalização entre os trabalhadores. Em 2003, 25,5% dos empregados tinham carteira de trabalho assinada. O percentual saltou para 36,8% em 2010. No entanto, o setor continua sendo o mais informal entre os demais, com a maioria dos trabalhadores sem carteira.

A indústria, o setor da economia mais formalizado do mercado, a participação dos empregados com carteira passou de 60,7% para 66,7% em oito anos. Já no comércio, a participação dos trabalhadores formalizados cresceu de 39,7% para 49,2% – um grande salto segundo o IBGE. No segmento que abrange serviços prestados a empresas, aluguéis, atividades imobiliárias e intermediação financeira a formalização passou a atingir 67,8% dos trabalhadores – em 2003 a taxa era 60,3%.

Em oito anos, o IBGE contabilizou a criação de 3,5 milhões de postos de trabalho nas seis principais regiões metropolitanas do País. O aumento de quase 19% da população ocupada supera o aumento da População em Idade ativa (PIA), de 11% no mesmo período. O total de desempregados recuou de 2,6 milhões para 1,6 milhões de pessoas, enquanto a taxa de desemprego despencou de 12,5% para 6,7%.

## **1.2 AS VARIÁVEIS UTILIZADAS PELO IPEA DATA 2011**

Segue abaixo a relação de variáveis utilizadas pelo IPEA DATA 2011. As variáveis que são analisadas no presente trabalho são: População em Idade Ativa (mil); População Ocupada (mil); População Ocupada (%); População Não Economicamente Ativa (%); Rendimento Médio Real Efetivo (R\$) e Rendimento Médio Real Habitual (R\$)

Pode-se acessar o conteúdo completo desta pesquisa em:

<http://www.ipeadata.gov.br/Default.aspx>

## **2. ENTENDENDO OS DADOS**

### **2.1 OS INDIVÍDUOS**

Os dados são séries históricas referentes ao Brasil, portanto, trata-se de séries temporais. As séries vão de 2002 a 2011 para as variáveis

## 2.2 AS VARIÁVEIS

São 6 as variáveis desta pesquisa tendo a amostra de 109 indivíduos, explicados na Tabela 1. Ressaltamos que todos os dados desta pesquisa são referentes aos anos de 2002 à 2011.

<b>Variável</b>	<b>Significado</b>	<b>Tipo</b>	<b>Unidade de Medida</b>
Meses	É o mês a que se refere o dado de cada variável	Variável Categórica	N/A
População em idade ativa (mil)	Classificação etária que compreende o conjunto de todas as pessoas teoricamente aptas a exercer uma atividade econômica.	Variável Quantitativa	Mil
População ocupada (mil)	Compreende o potencial de mão de obra com que pode contar o setor produtivo.	Variável Quantitativa	Mil
População ocupada (%)	Compreende o potencial de mão de obra com que pode contar o setor produtivo.	Variável Quantitativa	%
População não economicamente ativa (%)	São as pessoas não classificadas como ocupadas ou desocupadas, ou seja, pessoas incapacitadas para o trabalho ou que desistiram de buscar trabalho ou não querem mesmo trabalhar.	Variável Quantitativa	%
Rendimento médio real efetivo (R\$)	Representa o impacto de uma variação do preço sobre a quantidade procurada de um determinado bem originada pela variação dos rendimentos reais dos consumidores	Variável Quantitativa	R\$
Rendimento médio real habitual (R\$)	Renda de um indivíduo ou grupo tendo em conta os efeitos da inflação sobre o poder de compra	Variável Quantitativa	R\$

## 2.3 A FONTE E O TAMANHO DA SÉRIE DE DADOS

### Fonte:

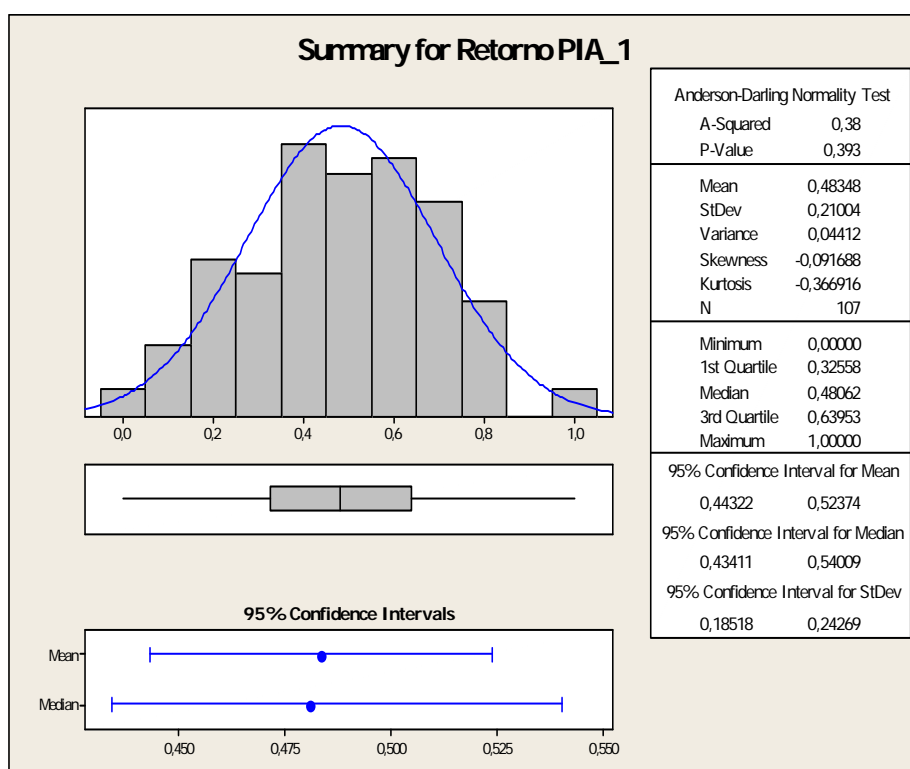
A maior parte dos dados desta pesquisa foram obtidos de base de dados dos estudos anteriores da própria IPEA DATA, sendo necessária apenas a atualização dos mesmos para alguns anos. Os dados para tal atualização foram obtidos do site do Banco Mundial, do Instituto Brasileiro de Geografia e Estatística (IBGE) e do Banco Central do Brasil (BCB).

### Tamanho da Série de Dados:

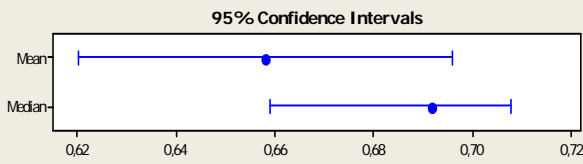
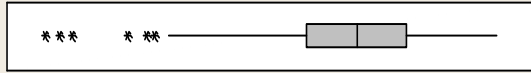
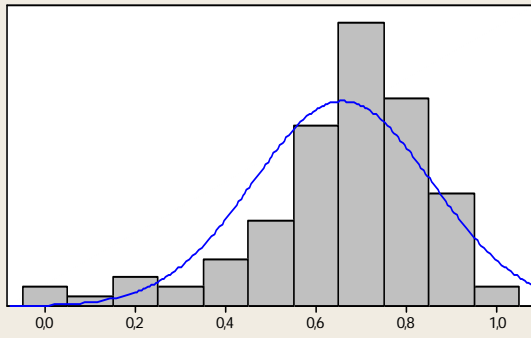
As três séries de dados podem ser consideradas satisfatórias para a realização desta pesquisa, uma vez que todas possuem dados de 10 anos ou mais, sem interrupções.

## 3. ANÁLISE DAS VARIÁVEIS

Segue abaixo quadro contendo Histograma, Curva de Densidade, Box-Plot, Intervalo de confiança da média e mediana, além das medidas numéricas como média, desvio-padrão, variância, quantidade de observações, valores mínimos, máximos, informações dos quartis e o teste de normalidade de Anderson-Darling (A-Squared e P-Value), para a variável RLV.



### Summary for Retorno PO\_1



#### Anderson-Darling Normality Test

A-Squared 1,98  
P-Value < 0,005

Mean 0,65812  
StDev 0,19706  
Variance 0,03883  
Skewness -1,14019  
Kurtosis 1,74443  
N 107

Minimum 0,00000  
1st Quartile 0,57832  
Median 0,69152  
3rd Quartile 0,79857  
Maximum 1,00000

#### 95% Confidence Interval for Mean

0,62035 0,69589

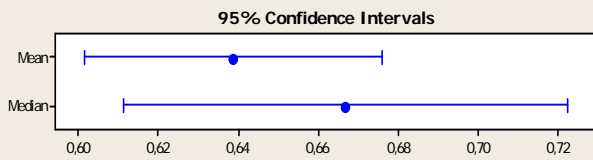
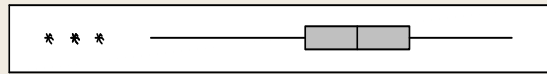
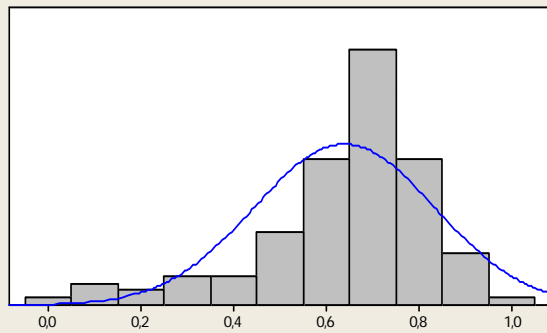
#### 95% Confidence Interval for Median

0,65899 0,70760

#### 95% Confidence Interval for StDev

0,17373 0,22769

### Summary for Retorno PO (%)\_1



#### Anderson-Darling Normality Test

A-Squared 2,63  
P-Value < 0,005

Mean 0,63863  
StDev 0,19401  
Variance 0,03764  
Skewness -1,11259  
Kurtosis 1,66107  
N 107

Minimum 0,00000  
1st Quartile 0,55556  
Median 0,66667  
3rd Quartile 0,77778  
Maximum 1,00000

#### 95% Confidence Interval for Mean

0,60144 0,67581

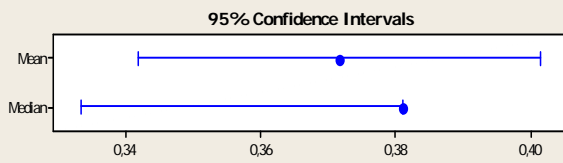
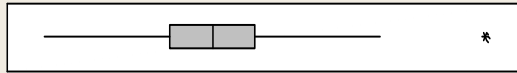
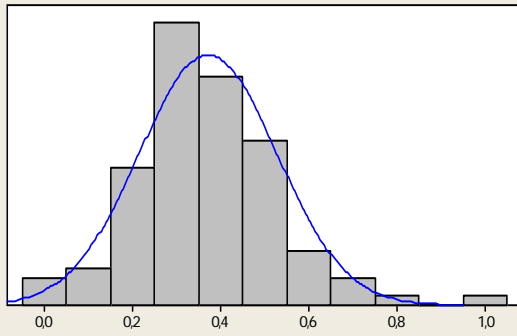
#### 95% Confidence Interval for Median

0,61111 0,72222

#### 95% Confidence Interval for StDev

0,17104 0,22417

### Summary for Retorno PNEA (%)\_1



#### Anderson-Darling Normality Test

A-Squared 1,05  
P-Value 0,009

Mean 0,37161  
StDev 0,15581  
Variance 0,02428  
Skewness 0,67179  
Kurtosis 1,96454  
N 107

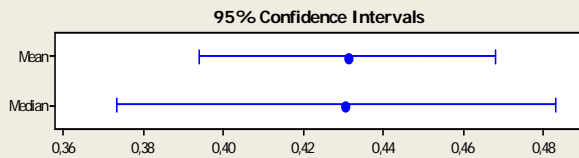
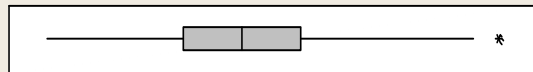
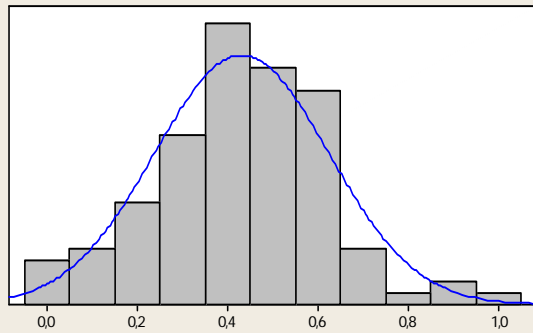
Minimum 0,00000  
1st Quartile 0,28571  
Median 0,38095  
3rd Quartile 0,47619  
Maximum 1,00000

95% Confidence Interval for Mean  
0,34174 0,40147

95% Confidence Interval for Median  
0,33333 0,38095

95% Confidence Interval for StDev  
0,13737 0,18003

### Summary for Retorno RMRH (R\$)\_1



#### Anderson-Darling Normality Test

A-Squared 0,40  
P-Value 0,353

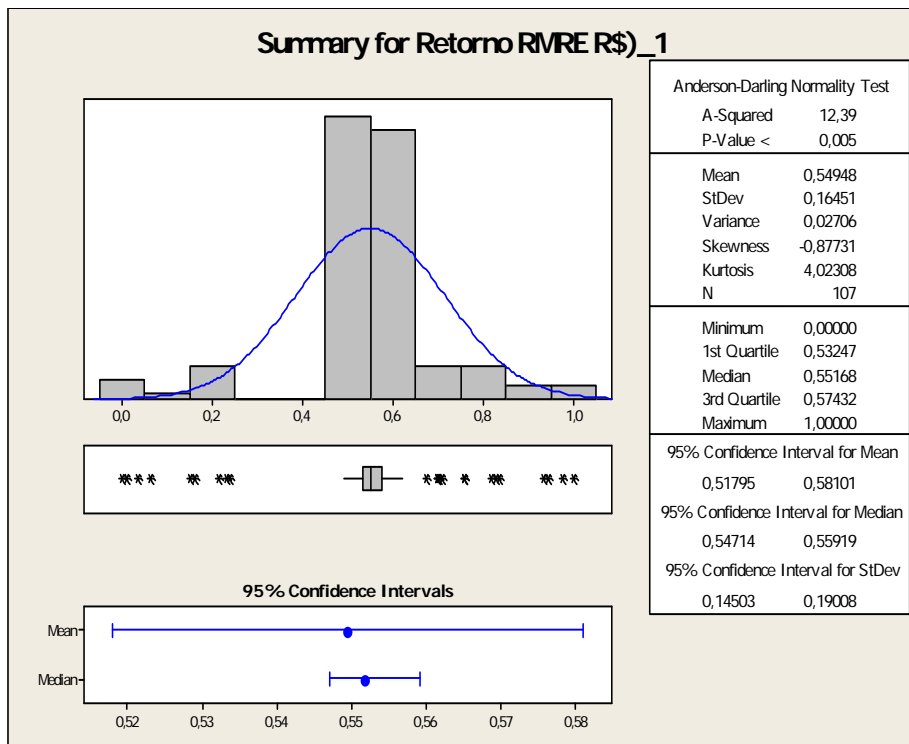
Mean 0,43102  
StDev 0,19281  
Variance 0,03718  
Skewness 0,095515  
Kurtosis 0,391270  
N 107

Minimum 0,00000  
1st Quartile 0,30125  
Median 0,43038  
3rd Quartile 0,56124  
Maximum 1,00000

95% Confidence Interval for Mean  
0,39406 0,46797

95% Confidence Interval for Median  
0,37317 0,48318

95% Confidence Interval for StDev  
0,16998 0,22278



Pelo teste de Anderson-Darling considerando 5% aproximadamente como parâmetro no teste de hipótese para as seis variáveis. Em quase todos os casos os valores de média e mediana estão relativamente próximos. Analisando a PIA, podemos verificar que a média da população está em 52,37% e a mediana em 54%; a População ocupada a média está em 69,58% e a mediana 70,76%; a População Ocupada a média 67,58% e a 72,22%; População Não Economicamente Ativa a média 40,14% e a mediana 38,09%; o Rendimento Médio Real Efetivo a média R\$ 58,10% e a mediana 55,9% e o Rendimento Médio Real Habitual a média 46,79% e a mediana 48,31%.

### 3.1 O COMPORTAMENTO DA VARIÁVEL

Conforme mencionado acima, cada variável será analisada utilizando gráficos para demonstrar o comportamento histórico da série, linhas de tendência, funções, erros das funções, além de extrapolações estatísticas.

Para análise do comportamento histórico da variável utilizaremos, inicialmente, gráficos que permitem uma verificação visual.

## 4. GRÁFICOS DE DISPERSÃO, DENDOGRAMAS, CORRELAÇÕES E REGRESSÕES.

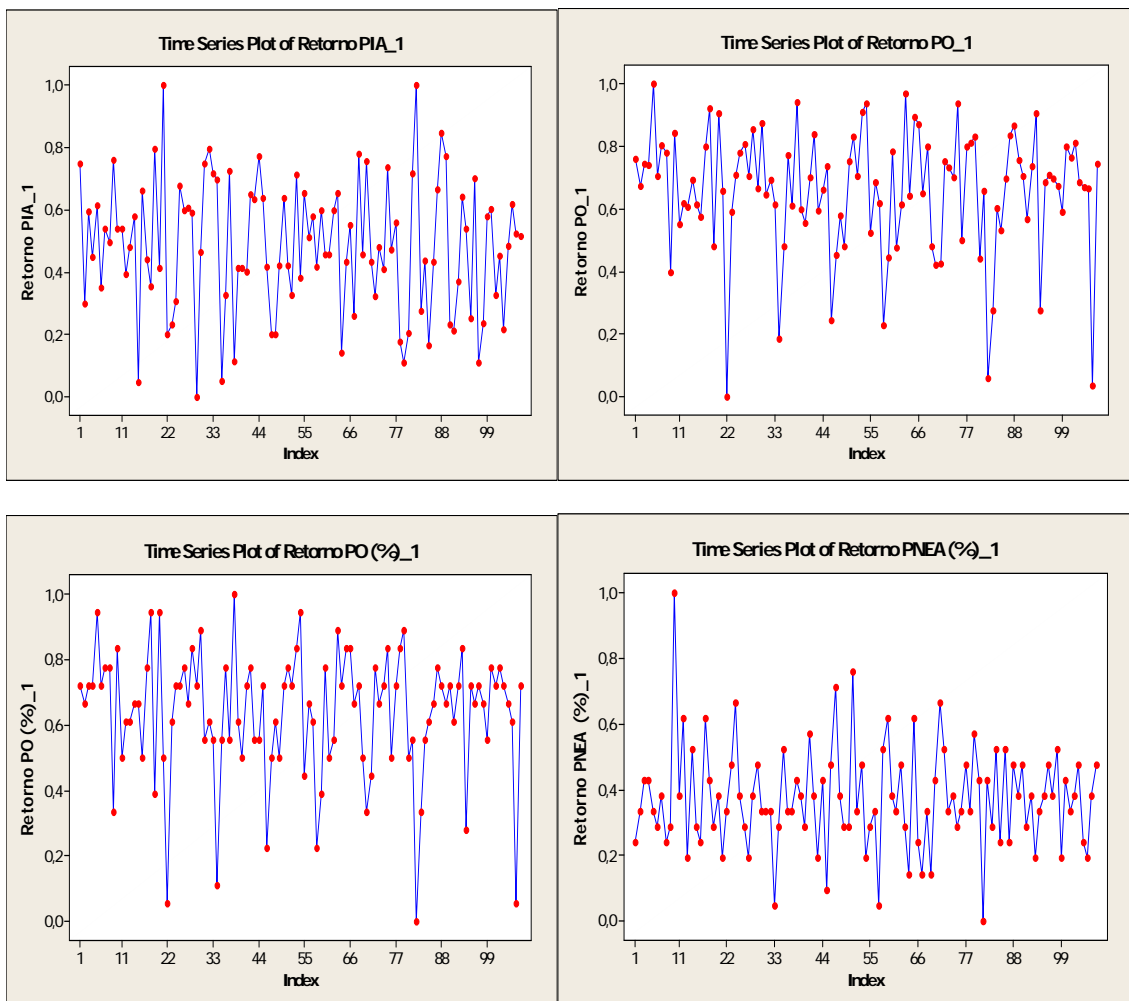
Através da correlação gerada através do software Minitab 16, é possível avaliar que existe uma forte correlação entre todas as variáveis e que todas apresentam alto grau de confiança.

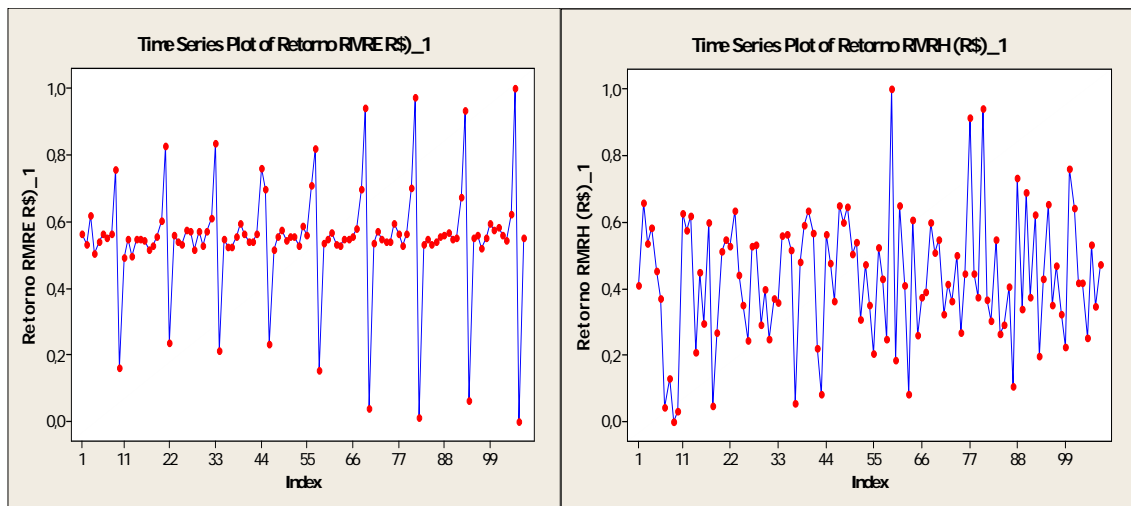
### 4.1 GRÁFICOS DE DISPERSÃO

Usado para explorar a relação entre um par de variáveis contínuas. Quando você cria um gráfico de dispersão, que geralmente mostra a variável resposta no eixo-y e a variável preditora no eixo-x para cada observação.

Os gráficos de dispersão que seguem demonstram essas correlações:

Analisaremos os gráficos de dispersão, sempre tendo no eixo Y a PIA.





**Segue abaixo uma breve análise dos gráficos acima:**

Direção: Grande parte das dispersões acima nos permitem verificar um padrão linear na maioria dos gráficos, demonstrando uma maior acentuação na parte superior. Mostrando-nos que há uma relação, ao menos de evidência visual. Em muitos casos o aumento de X é acompanhado por um forte aumento de Y, sendo que apenas nos casos de mortalidade materna e fertilidade na adolescência, onde não acompanha este tendência.

Intensidade: A intensidade é fortemente percebida nas comparações realizadas com a PIA, sendo que todos estão diretamente ligados.

Forma: De maneira visual apenas, os gráficos acima parecem denotar forma linear crescente e decrescente.

Valores Atípicos: Alguns gráficos indicam a existência de valores atípicos, ou seja, indivíduos ou estados, que estão localizados longe dos demais. Isto pode ser verificado também na análise exploratória que vimos no item anterior.

A matriz de correlação acima nos ajuda tirar conclusões mais precisas sobre a associação entre as variáveis.

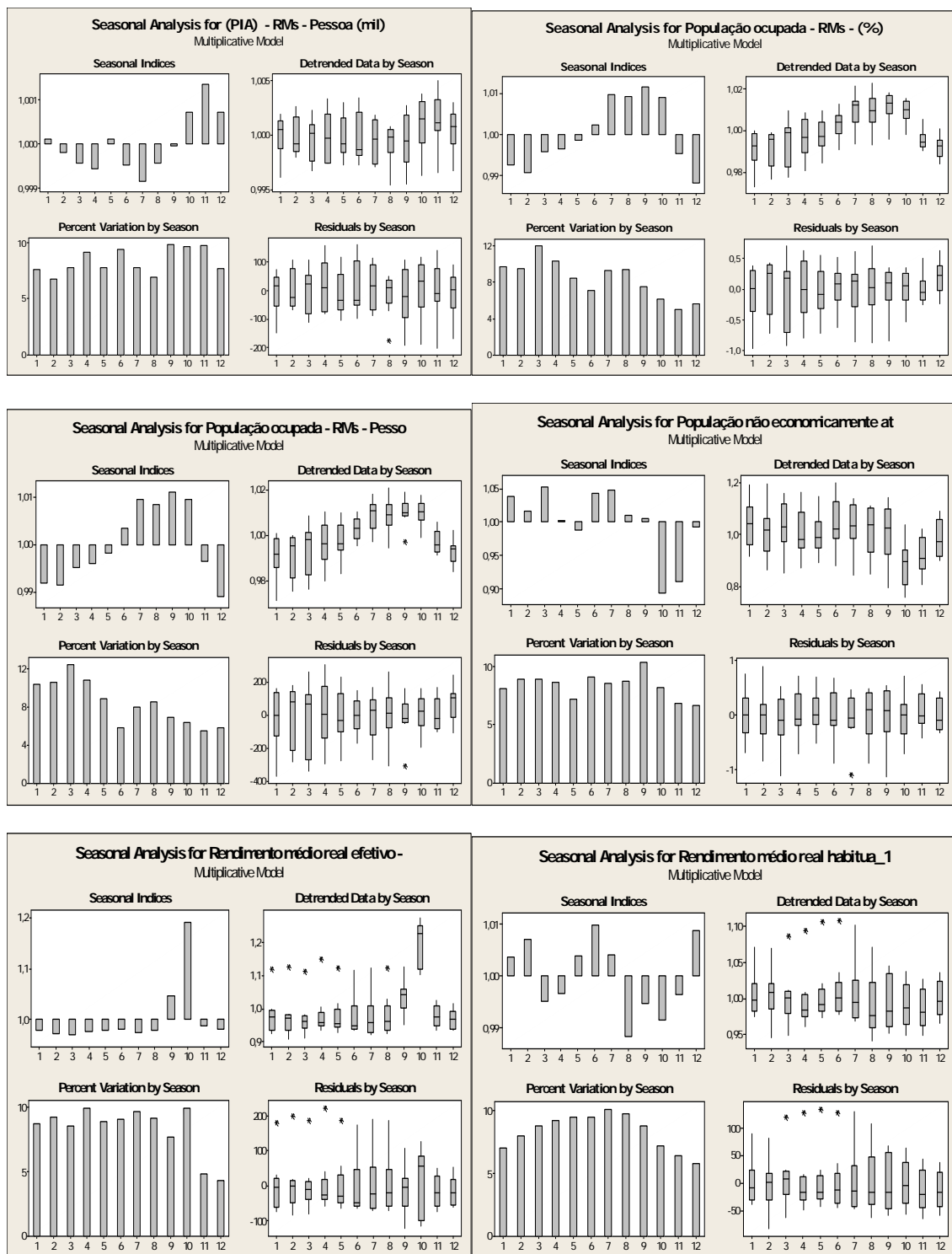
#### **4.2 DECOMPOSITION:**

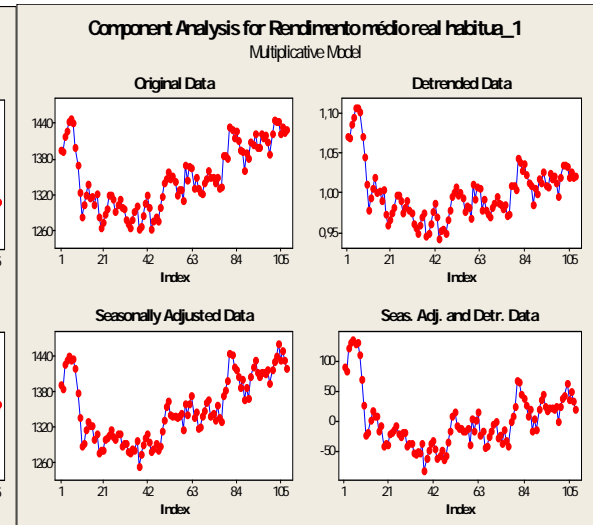
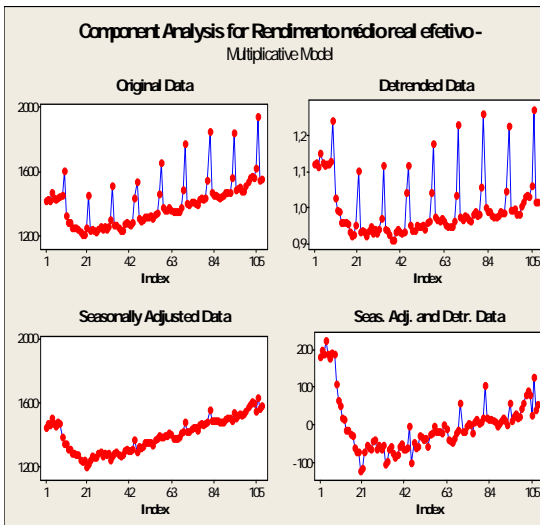
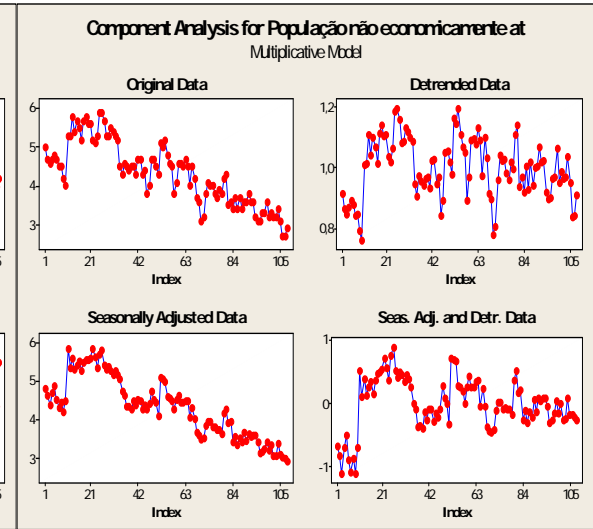
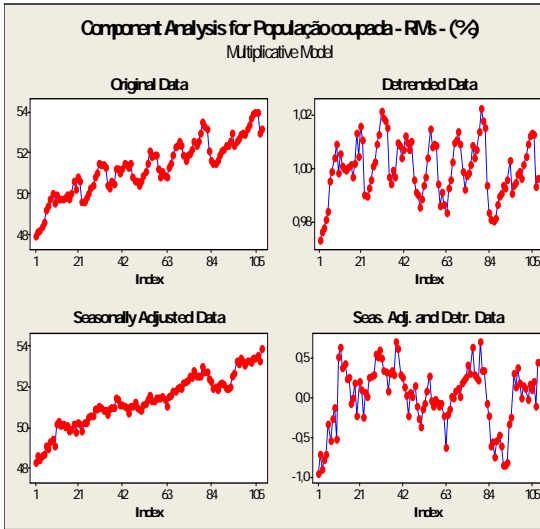
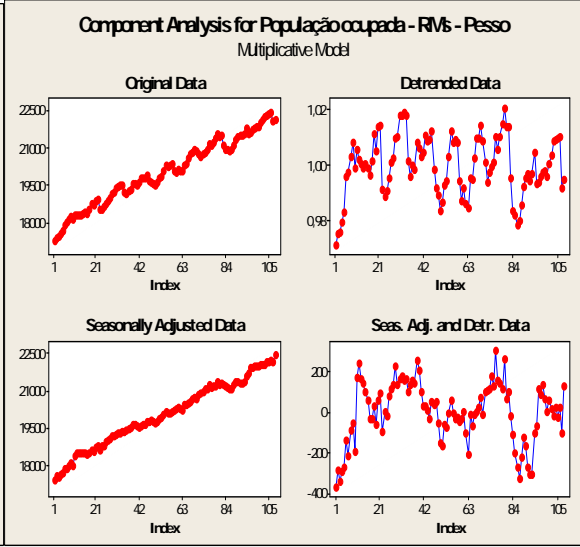
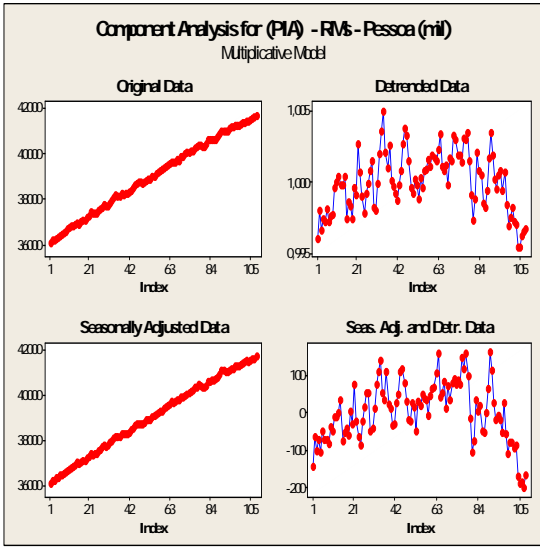
Para análise do comportamento histórico da variável utilizaremos, inicialmente, gráficos que permitem uma verificação visual.

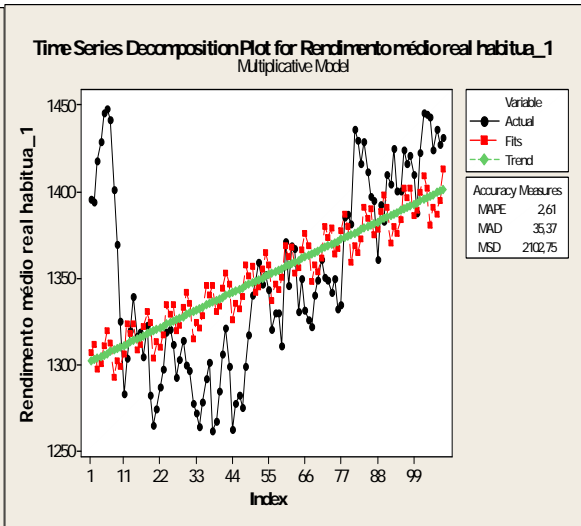
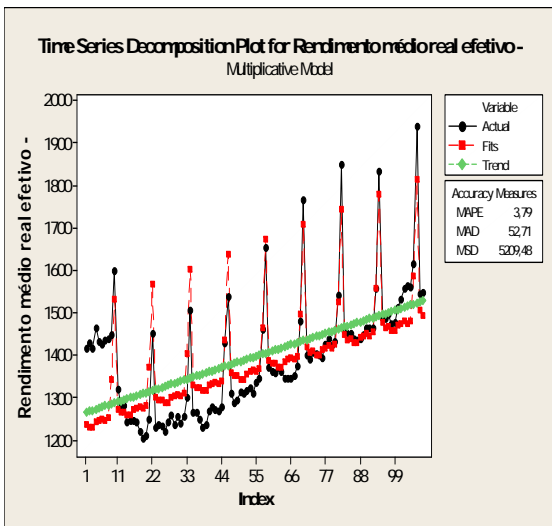
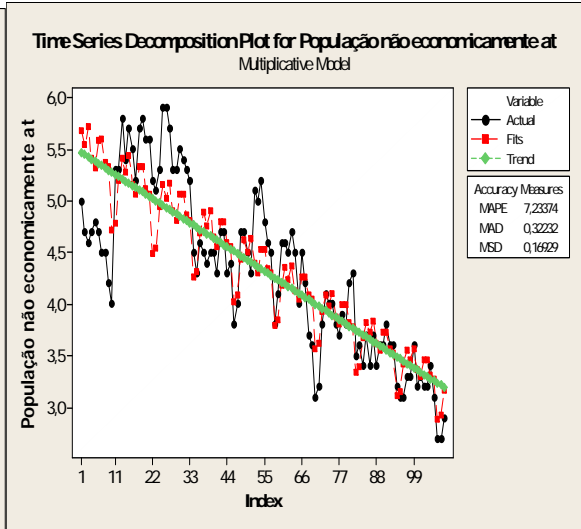
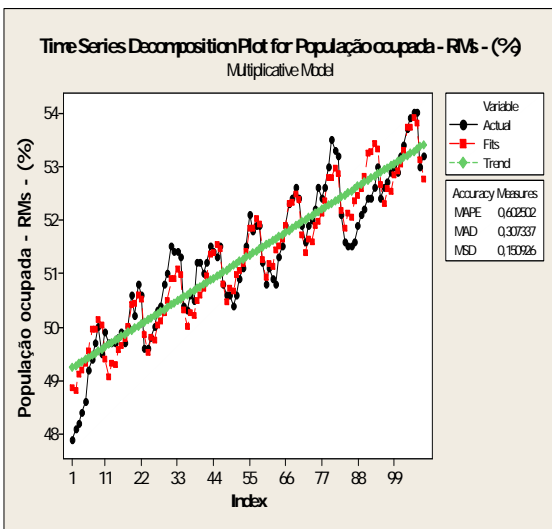
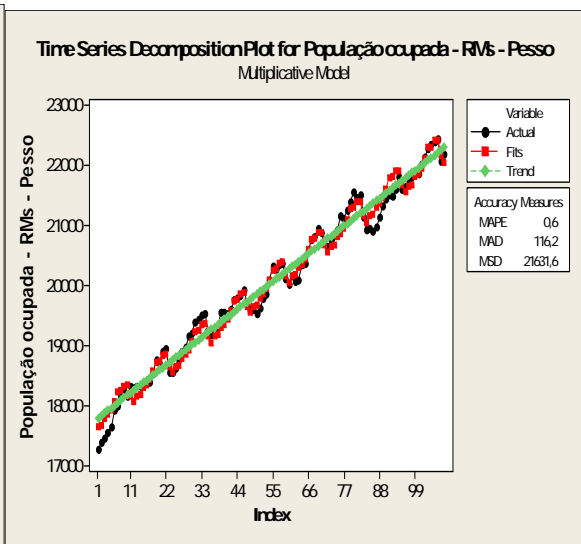
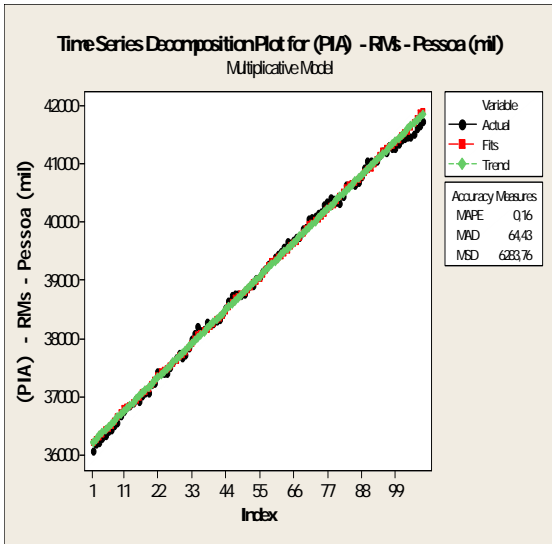
As medidas de precisão (análise de séries temporais) usamos as estatísticas para comparar ajustes de previsão e métodos de alisamento. O MINITAB calcula três medidas de precisão do modelo montado. MAPE, MAD e MSD. As três medidas não

são muito informativos por si só, mas você pode usá-las para comparar os ajustes obtidos por diferentes métodos, nesse caso os métodos utilizados são multiplicativo e linear. Para todas as três medidas, valores menores indicam um melhor modelo de ajuste.

#### 4.2.1 MULTIPLICATIVE







## Time Series Decomposition for (PIA) - RMs - Pessoa (mil)

Multiplicative Model

Data (PIA) - RMs - Pessoa (mil)  
Length 108

NMissing 0

Fitted Trend Equation

$$Y_t = 36161,0 + 52,7*t$$

Seasonal Indices

Period	Index
1	1,00011
2	0,99980
3	0,99956
4	0,99942
5	1,00011
6	0,99953
7	0,99915
8	0,99955
9	0,99997
10	1,00073
11	1,00136
12	1,00072

Accuracy Measures

MAPE	0,16
MAD	64,43
MSD	6283,76

### **Time Series Decomposition Plot for (PIA) - RMs - Pessoa (mil)**

### **Decomposition - Component Analysis for (PIA) - RMs - Pessoa (mil)**

### **Decomposition - Seasonal Analysis for (PIA) - RMs - Pessoa (mil)**

### **Time Series Decomposition for População ocupada - RMs - Pessoa**

Multiplicative Model

Data	População ocupada - RMs - Pessoa
Length	108
NMissing	0

Fitted Trend Equation

$$Y_t = 17746,9 + 42,2*t$$

Seasonal Indices

Period	Index
1	0,99189
2	0,99127
3	0,99511
4	0,99587
5	0,99811
6	1,00337
7	1,00965
8	1,00853
9	1,01112
10	1,00958

11 0,99652  
12 0,98897

#### Accuracy Measures

MAPE 0,6  
MAD 116,2  
MSD 21631,6

### **Time Series Decomposition Plot for População ocupada - RMs - Pessoa**

#### **Decomposition - Component Analysis for População ocupada - RMs - Pessoa**

#### **Decomposition - Seasonal Analysis for População ocupada - RMs - Pessoa**

### **Time Series Decomposition for População ocupada - RMs - (%)**

#### Multiplicative Model

Data População ocupada - RMs - (%)  
Length 108  
NMissing 0

#### Fitted Trend Equation

$$Y_t = 49,1960 + 0,0389*t$$

#### Seasonal Indices

Period	Index
1	0,99252
2	0,99079
3	0,99582
4	0,99662
5	0,99854
6	1,00235
7	1,00982
8	1,00930
9	1,01166
10	1,00905
11	0,99538
12	0,98813

#### Accuracy Measures

MAPE 0,602502  
MAD 0,307337  
MSD 0,150926

### **Time Series Decomposition Plot for População ocupada - RMs - (%)**

#### **Decomposition - Component Analysis for População ocupada - RMs - (%)**

#### **Decomposition - Seasonal Analysis for População ocupada - RMs - (%)**

### **Time Series Decomposition for População não economicamente at**

#### Multiplicative Model

Data População não economicamente at  
Length 108  
NMissing 0

Fitted Trend Equation

$$Y_t = 5,4870 - 0,021278*t$$

Seasonal Indices

Period	Index
1	1,03919
2	1,01703
3	1,05304
4	1,00160
5	0,98731
6	1,04310
7	1,04798
8	1,01039
9	1,00568
10	0,89295
11	0,90962
12	0,99211

Accuracy Measures

MAPE 7,23374  
MAD 0,32232  
MSD 0,16929

### **Time Series Decomposition Plot for População não economicamente at**

### **Decomposition - Component Analysis for População não economicamente at**

### **Decomposition - Seasonal Analysis for População não economicamente at**

### **Time Series Decomposition for Rendimento médio real efetivo -**

Multiplicative Model

Data Rendimento médio real efetivo -  
Length 108  
NMissing 0

Fitted Trend Equation

$$Y_t = 1262,2 + 2,46*t$$

Seasonal Indices

Period	Index
1	0,97778
2	0,97016
3	0,96841
4	0,97589
5	0,97679
6	0,97904
7	0,97348
8	0,97651

9	1,04480
10	1,19174
11	0,98693
12	0,97846

#### Accuracy Measures

MAPE	3,79
MAD	52,71
MSD	5209,48

### **Time Series Decomposition Plot for Rendimento médio real efetivo - Decomposition - Component Analysis for Rendimento médio real efetivo - Decomposition - Seasonal Analysis for Rendimento médio real efetivo –**

### **Time Series Decomposition for Rendimento médio real habitua\_1**

#### Multiplicative Model

Data	Rendimento médio real habitua_1
Length	108
NMissing	0

#### Fitted Trend Equation

$$Y_t = 1300,57 + 0,929*t$$

#### Seasonal Indices

Period	Index
1	1,00373
2	1,00709
3	0,99501
4	0,99663
5	1,00387
6	1,00996
7	1,00413
8	0,98821
9	0,99472
10	0,99133
11	0,99648
12	1,00885

#### Accuracy Measures

MAPE	2,61
MAD	35,37
MSD	2102,75

### **Time Series Decomposition Plot for Rendimento médio real habitua\_1**

### **Decomposition - Component Analysis for Rendimento médio real habitua\_1**

### **Decomposition - Seasonal Analysis for Rendimento médio real habitua\_1**

## 4.2.2 LINEAR:

### Trend Analysis for (PIA) - RMs - Pessoa (mil)

Data (PIA) - RMs - Pessoa (mil)  
 Length 108  
 NMissing 0

Fitted Trend Equation

$$Y_t = 36158,4 + 52,7 * t$$

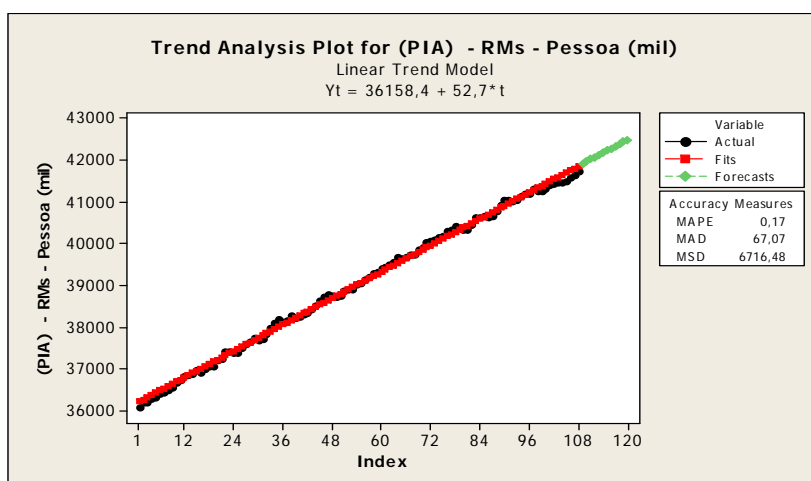
Accuracy Measures

MAPE 0,17  
 MAD 67,07  
 MSD 6716,48

Forecasts

Period	Forecast
109	41907,4
110	41960,1
111	42012,8
112	42065,6
113	42118,3
114	42171,1
115	42223,8
116	42276,6
117	42329,3
118	42382,1
119	42434,8
120	42487,5

### Trend Analysis Plot for (PIA) - RMs - Pessoa (mil)



	<b>MULTIPLICATIVE</b>	<b>LINEAR</b>
<b>MAPE</b>	0,16	0,17
<b>MAD</b>	64,43	67,07
<b>MSD</b>	6283,76	6716,48

## Trend Analysis for População ocupada - RMs - Pesso

Data População ocupada - RMs - Pesso  
 Length 108  
 NMissing 0

Fitted Trend Equation

$$Y_t = 17736,7 + 42,4 * t$$

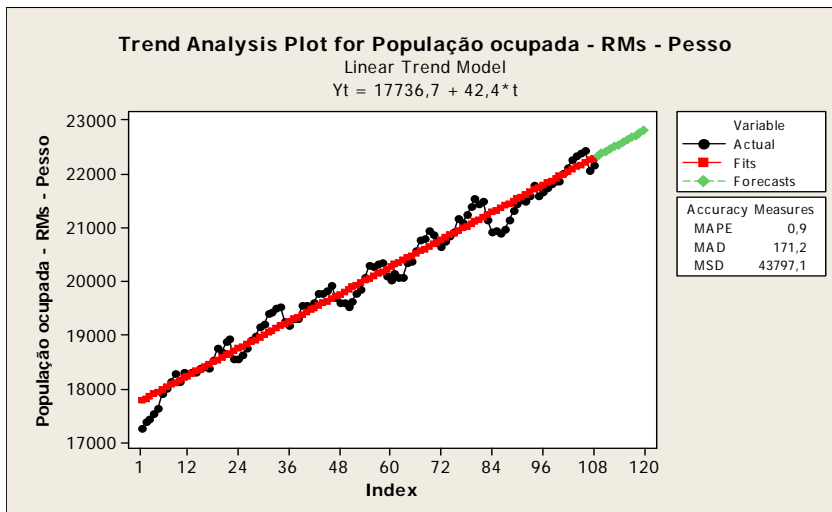
Accuracy Measures

MAPE 0,9  
 MAD 171,2  
 MSD 43797,1

Forecasts

Period	Forecast
109	22354,4
110	22396,7
111	22439,1
112	22481,4
113	22523,8
114	22566,2
115	22608,5
116	22650,9
117	22693,3
118	22735,6
119	22778,0
120	22820,4

## Trend Analysis Plot for População ocupada - RMs - Pesso



	<b>MULTIPLICATIVE</b>	<b>LINEAR</b>
<b>MAPE</b>	<b>0,6</b>	0,9
<b>MAD</b>	<b>116,2</b>	171,2
<b>MSD</b>	<b>21631,6</b>	43797,1

## Trend Analysis for População ocupada - RMs - (%)

Data População ocupada - RMs - (%)

Length 108  
 NMissing 0

Fitted Trend Equation

$$Y_t = 49,173 + 0,0393*t$$

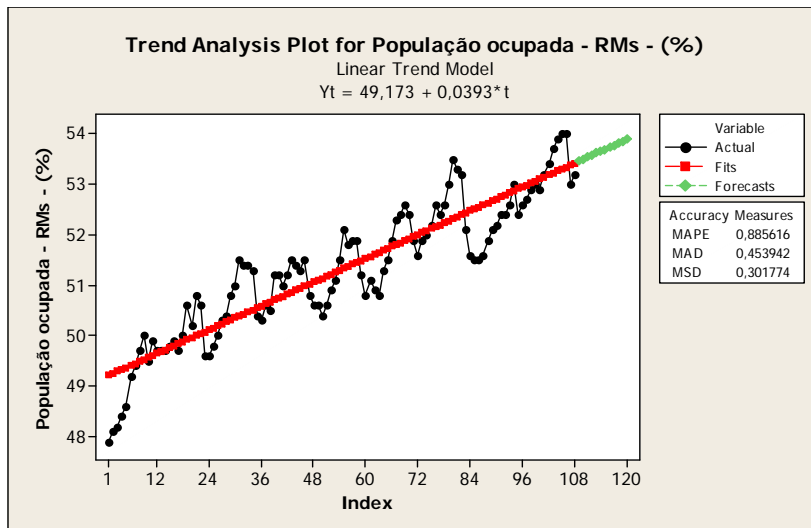
Accuracy Measures

MAPE 0,885616  
 MAD 0,453942  
 MSD 0,301774

Forecasts

Period	Forecast
109	53,4607
110	53,5001
111	53,5394
112	53,5788
113	53,6181
114	53,6575
115	53,6968
116	53,7361
117	53,7755
118	53,8148
119	53,8542
120	53,8935

### Trend Analysis Plot for População ocupada - RMs - (%)



	MULTIPLICATIVE	LINEAR
MAPE	0,602502	0,885616
MAD	0,307337	0,453942
MSD	0,150926	0,301774

### Trend Analysis for População não economicamente ativa

Data População não economicamente at  
 Length 108  
 NMissing 0

Fitted Trend Equation

$$Y_t = 5,5090 - 0,021657 * t$$

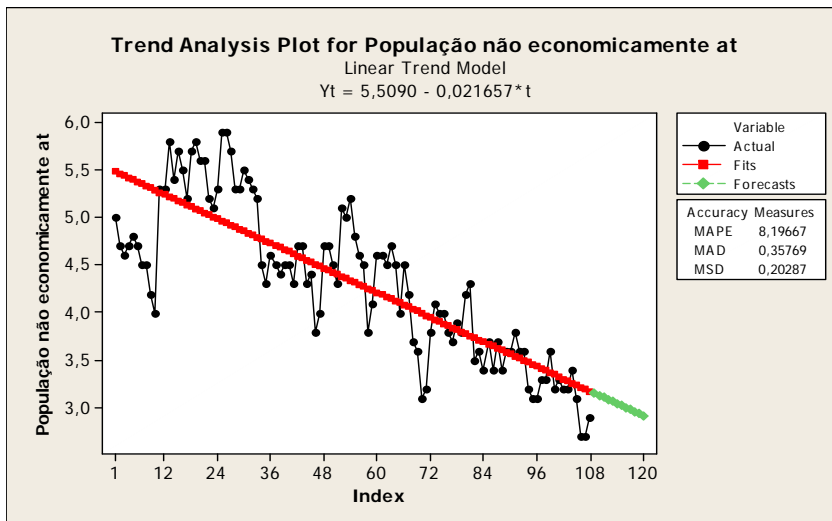
Accuracy Measures

MAPE 8,19667  
MAD 0,35769  
MSD 0,20287

Forecasts

Period	Forecast
109	3,14841
110	3,12675
111	3,10509
112	3,08344
113	3,06178
114	3,04012
115	3,01847
116	2,99681
117	2,97515
118	2,95350
119	2,93184
120	2,91018

### Trend Analysis Plot for População não economicamente ativa



	<b>MULTIPLICATIVE</b>	<b>LINEAR</b>
<b>MAPE</b>	7,23374	8,19667
<b>MAD</b>	0,32232	0,35769
<b>MSD</b>	0,16929	0,20287

### Trend Analysis for Rendimento médio real efetivo -

Data Rendimento médio real efetivo -  
Length 108  
NMissing 0

Fitted Trend Equation

$$Y_t = 1254,0 + 2,61 * t$$

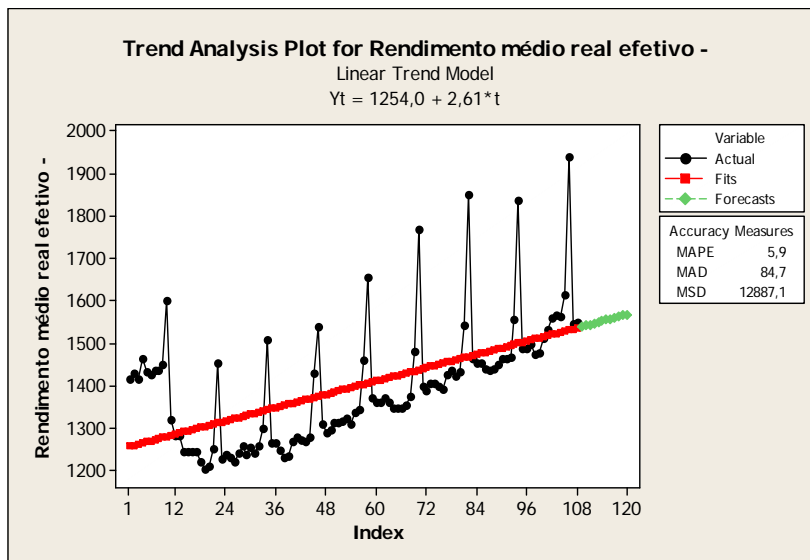
### Accuracy Measures

MAPE	5,9
MAD	84,7
MSD	12887,1

### Forecasts

Period	Forecast
109	1538,56
110	1541,17
111	1543,78
112	1546,39
113	1549,00
114	1551,61
115	1554,22
116	1556,84
117	1559,45
118	1562,06
119	1564,67
120	1567,28

### Trend Analysis Plot for Rendimento médio real efetivo –



	<b>MULTIPLICATIVE</b>	<b>LINEAR</b>
<b>MAPE</b>	3,79	5,9
<b>MAD</b>	52,71	84,7
<b>MSD</b>	5209,48	12887,1

### Trend Analysis for Rendimento médio real habitua\_1

Data Rendimento médio real habitua\_1  
 Length 108  
 NMissing 0

Fitted Trend Equation

$$Y_t = 1301,02 + 0,920 \cdot t$$

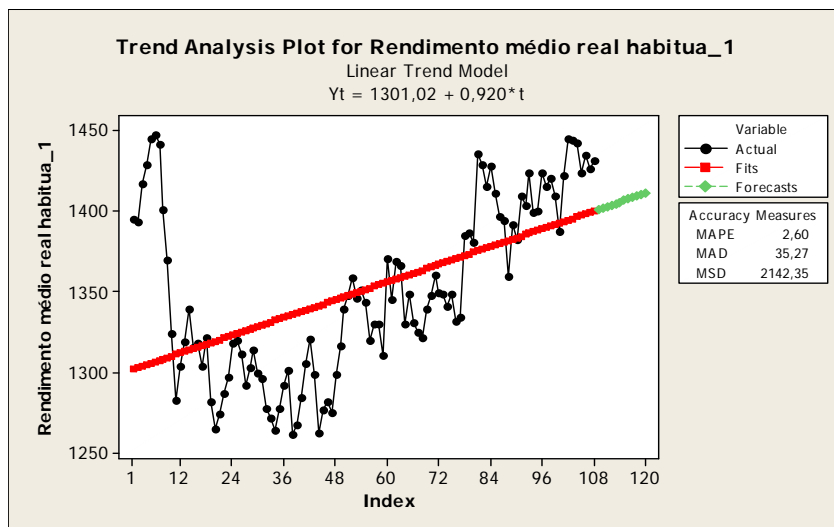
Accuracy Measures

MAPE	2,60
MAD	35,27
MSD	2142,35

Forecasts

Period	Forecast
109	1401,30
110	1402,22
111	1403,14
112	1404,06
113	1404,98
114	1405,90
115	1406,82
116	1407,74
117	1408,66
118	1409,58
119	1410,50
120	1411,42

### Trend Analysis Plot for Rendimento médio real habitua\_1



	MULTIPLICATIVE	LINEAR
MAPE	2,61	2,60
MAD	35,37	35,27
MSD	2102,75	2142,35

De acordo com o gráfico acima, percebemos tratar-se de uma série temporal com tendência ascendente, porém com muitas oscilações durante o período, não parecendo ser uma série linear. Apenas na variável Rendimento Médio Real Habitual podemos verificar uma série linear mas com pouquíssima diferença da multiplicativa.

Numa primeira análise, apenas visual, parece haver uma tendência de ascensão desta série, apesar de encontrarmos algumas irregularidades ou sazonalidades em alguns

períodos. Nessas considerações podemos observar que após os 48 meses apresenta ascensão.

### 4.3 DENDROGRAMA

Dendrogramas são estruturas gráficas em forma de árvore, utilizadas para representar as junções (métodos hierárquicos) ou divisões (métodos de partição) que ocorreram a partir de valores provenientes da matriz de distâncias (JOHNSON & WICHERN, 1988). De acordo com Bussab et al (1990), para construirmos um dendrograma utilizando os valores da matriz de distâncias com o objetivo de ilustrar as junções, devemos colocar no eixo horizontal os elementos, em uma ordem conveniente de acordo com os grupos formados, de onde partirá de cada um desses elementos uma linha vertical até a altura correspondente ao nível (o valor da distância) em que ocorreu a junção (a um outro elemento ou grupo). Essa altura é marcada no eixo vertical.

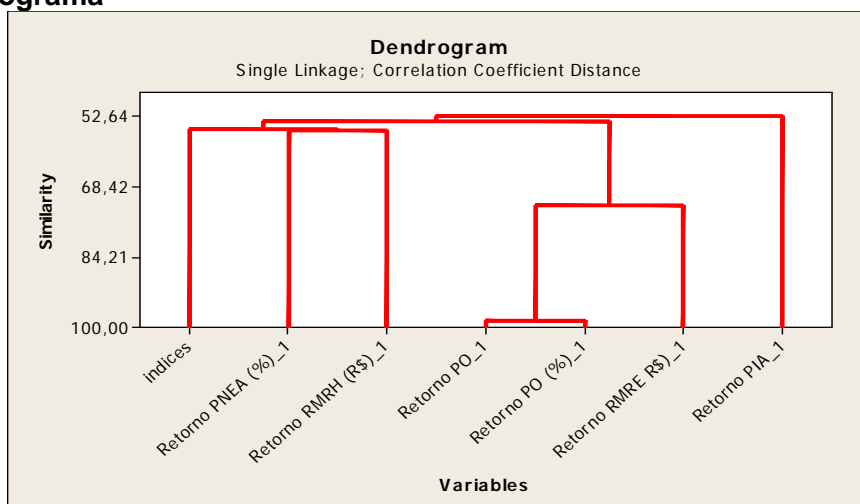
#### Cluster Analysis of Variables: indices; Retorno PIA\_; Retorno PO\_1; ...

Correlation Coefficient Distance, Single Linkage  
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	6	98,4094	0,031812	3	5	3
2	5	72,5788	0,548425	4	6	3
3	4	55,7828	0,884345	5	7	4
4	3	55,6350	0,887301	6	4	1
5	2	53,9103	0,921795	7	3	1
6	1	52,6357	0,947286	8	2	1

Vejamos abaixo o dendrograma que nada mais é do que as mesmas correlações acima, porém em forma gráfica, nos indicando mais claramente quais as variáveis que poderiam ser unidas.

#### Dendrograma



No gráfico acima podemos observar uma maior concentração e a similaridade entre as variáveis Retorno Médio Real Habitual e População Não Economicamente Ativa, aponta uma forte correlação entre estas variáveis do conjunto de dados estudado.

#### 4.4 CLUSTERS ANALYSIS OBSERVATIONS

##### Cluster Analysis of Observations: indices; Retorno PIA\_; Retorno PO\_1; ...

Euclidean Distance, Single Linkage  
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new cluster
1	106	97,7226	0,05978	49	61	49	2
2	105	97,1191	0,07562	40	49	40	3
3	104	96,8116	0,08369	6	52	6	2
4	103	96,5714	0,08999	2	95	2	2
5	102	96,0177	0,10453	2	90	2	3
6	101	95,6938	0,11303	85	91	85	2
7	100	95,6714	0,11362	32	89	32	2
8	99	95,6016	0,11545	92	101	92	2
9	98	95,5731	0,11620	4	74	4	2
10	97	95,5400	0,11707	25	32	25	3
11	96	95,3787	0,12130	98	103	98	2
12	95	95,3775	0,12133	26	42	26	2
13	94	95,3394	0,12233	2	92	2	5
14	93	95,1970	0,12607	16	55	16	2
15	92	95,1237	0,12800	3	107	3	2
16	91	95,0985	0,12866	25	96	25	4
17	90	95,0776	0,12921	27	50	27	2
18	89	94,9001	0,13387	4	36	4	3
19	88	94,7423	0,13801	3	86	3	3
20	87	94,5900	0,14201	11	40	11	4
21	86	94,5432	0,14323	30	102	30	2
22	85	94,4732	0,14507	30	66	30	3
23	84	94,4323	0,14614	3	4	3	6
24	83	94,3865	0,14734	12	41	12	2
25	82	94,3768	0,14760	72	98	72	3
26	81	94,2649	0,15054	46	58	46	2
27	80	94,2553	0,15079	2	3	2	11
28	79	94,2047	0,15212	1	27	1	3
29	78	94,1477	0,15361	33	81	33	2
30	77	94,1345	0,15396	6	73	6	3
31	76	94,1245	0,15422	25	31	25	5
32	75	94,1215	0,15430	43	99	43	2
33	74	94,0919	0,15508	15	97	15	2
34	73	94,0331	0,15662	15	72	15	5
35	72	93,9275	0,15939	8	60	8	2
36	71	93,8909	0,16035	2	12	2	13
37	70	93,7354	0,16444	17	51	17	2
38	69	93,6982	0,16541	15	24	15	6
39	68	93,6752	0,16602	26	87	26	3
40	67	93,6281	0,16725	20	30	20	4
41	66	93,6140	0,16762	33	57	33	3
42	65	93,6030	0,16791	16	19	16	3
43	64	93,5728	0,16870	2	23	2	14
44	63	93,5672	0,16885	20	26	20	7
45	62	93,5318	0,16978	1	2	1	17
46	61	93,4888	0,17091	1	15	1	23
47	60	93,4545	0,17181	1	25	1	28
48	59	93,4000	0,17324	1	6	1	31

49	58	93,3347	0,17495	1	56	1	32
50	57	93,3128	0,17553	43	104	43	3
51	56	93,2972	0,17594	1	43	1	35
52	55	93,2964	0,17596	77	100	77	2
53	54	93,2774	0,17646	1	37	1	36
54	53	93,2623	0,17686	20	75	20	8
55	52	93,2610	0,17689	28	53	28	2
56	51	93,0986	0,18115	1	45	1	37
57	50	93,0464	0,18252	1	39	1	38
58	49	92,9422	0,18526	20	63	20	9
59	48	92,8471	0,18775	1	67	1	39
60	47	92,7308	0,19080	1	62	1	40
61	46	92,5408	0,19579	20	28	20	11
62	45	92,4949	0,19700	1	14	1	41
63	44	92,3182	0,20164	20	54	20	12
64	43	92,2813	0,20260	1	84	1	42
65	42	92,2445	0,20357	7	20	7	13
66	41	92,1361	0,20642	11	76	11	5
67	40	92,1315	0,20654	1	13	1	43
68	39	92,0553	0,20854	5	7	5	14
69	38	92,0499	0,20868	1	8	1	45
70	37	91,5817	0,22097	1	33	1	48
71	36	90,9780	0,23682	38	79	38	2
72	35	90,9362	0,23791	1	29	1	49
73	34	90,9265	0,23816	1	68	1	50
74	33	90,8860	0,23923	35	48	35	2
75	32	90,8311	0,24067	1	64	1	51
76	31	90,6499	0,24543	38	78	38	3
77	30	90,5936	0,24690	1	44	1	52
78	29	90,5747	0,24740	35	85	35	4
79	28	90,4307	0,25118	5	77	5	16
80	27	90,0159	0,26207	1	105	1	53
81	26	90,0140	0,26212	11	35	11	9
82	25	89,7677	0,26858	5	88	5	17
83	24	89,7626	0,26872	11	47	11	10
84	23	89,7409	0,26929	46	94	46	3
85	22	89,4816	0,27609	11	16	11	13
86	21	89,2761	0,28149	5	18	5	18
87	20	89,0423	0,28762	5	38	5	21
88	19	88,6491	0,29794	5	17	5	23
89	18	88,3694	0,30529	5	65	5	24
90	17	88,2854	0,30749	46	70	46	4
91	16	87,7919	0,32044	34	46	34	5
92	15	87,7380	0,32186	11	83	11	14
93	14	86,9555	0,34240	1	21	1	54
94	13	86,7873	0,34681	9	11	9	15
95	12	86,7621	0,34747	34	106	34	6
96	11	85,7830	0,37318	9	71	9	16
97	10	85,0433	0,39259	5	93	5	25
98	9	84,6036	0,40413	9	69	9	17
99	8	84,1521	0,41598	9	80	9	18
100	7	83,7863	0,42558	22	34	22	7
101	6	83,2448	0,43980	9	22	9	25
102	5	81,9501	0,47378	9	59	9	26
103	4	81,5147	0,48521	9	82	9	27
104	3	77,1574	0,59958	5	10	5	26
105	2	61,6888	1,00561	1	9	1	81
106	1	61,5478	1,00931	1	5	1	107

Final Partition

Number of clusters: 3

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	54	5,95902	0,313906	0,623434
Cluster2	26	3,49990	0,336465	0,781177
Cluster3	27	6,18218	0,453689	0,879400

### Cluster Centroids

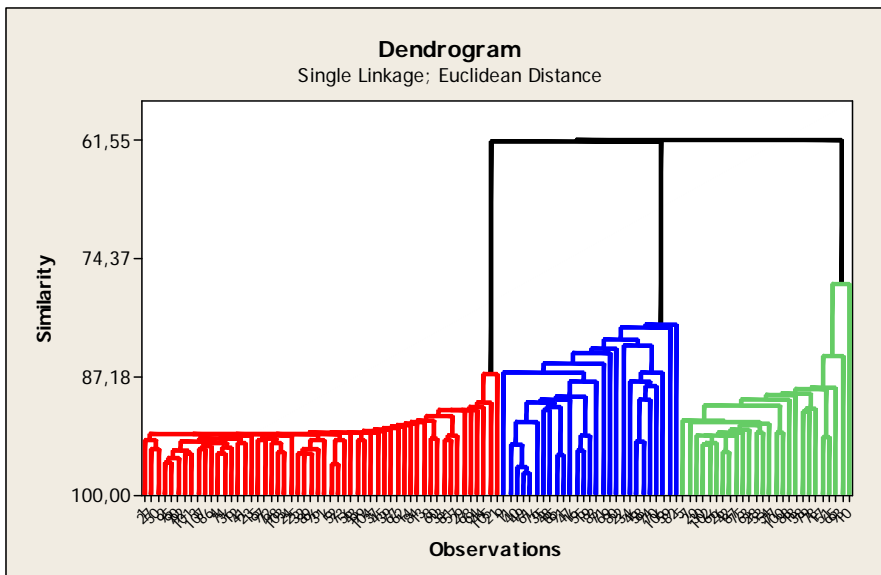
Variable	Cluster1	Cluster2	Cluster3	Grand centroid
indices	2,00000	3,00000	1,00000	1,99065
Retorno PIA_1	0,48019	0,51029	0,46425	0,48348
Retorno PO_1	0,68800	0,86995	0,39438	0,65812
Retorno PNEA (%)_1	0,34568	0,41941	0,37743	0,37161
Retorno PO (%)_1	0,66770	0,83761	0,38889	0,63863
Retorno RMRE R\$)_1	0,60033	0,54922	0,44804	0,54948
Retorno RMRH (R\$)_1	0,43308	0,37871	0,47726	0,43102

### Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3
Cluster1	0,00000	1,03629	1,09103
Cluster2	1,03629	0,00000	2,10982
Cluster3	1,09103	2,10982	0,00000

A tabela acima nos indica que, dada a grande similaridade de todas as observações, não faz sentido separarmos os dados por cluster. Caso contrário, teremos 1 cluster com muitas observações (quase 50% = 54 observações) e os demais com apenas 26 ou 27 observações.

### Dendrogram



Na análise Cluster x cluster observamos que a maior distancia entre os centróides se dá entre o cluster 2 e 3, e a menor, entre o cluster 1 e 2.

O grande centróide do indicador Retorno da PIA (mil) ficou em 48%; Retorno População Ocupada (mil) 68,8%; Retorno População Não Economicamente Ativa 34,56%; Retorno População Ocupada (%) 66,77%; Retorno Médio Real Efetivo 60% e Retorno Médio Relat Habitual 43,3%. No cluster um obtivemos 54 observações.

#### 4.5 REGRESSON:

##### Regression Analysis: Retorno PIA versus Retorno PO; Retorno PNEA; ...

The regression equation is

$$\begin{aligned} \text{Retorno PIA} = & 14,9 + 1,39 \text{ Retorno PO} - 21,6 \text{ Retorno PNEA (\%)} \\ & - 531 \text{ Retorno PO (\%)} - 0,0335 \text{ Retorno RMRE R\$)} \\ & - 0,200 \text{ Retorno RMRH (R\$)} \end{aligned}$$

Predictor	Coef	SE Coef	T	P
Constant	14,899	3,286	4,53	0,000
Retorno PO	1,39106	0,07664	18,15	0,000
Retorno PNEA (%)	-21,585	8,610	-2,51	0,014
Retorno PO (%)	-530,69	28,30	-18,75	0,000
Retorno RMRE R\$)	-0,03348	0,02663	-1,26	0,212
Retorno RMRH (R\$)	-0,1995	0,1258	-1,59	0,116

S = 25,0130    R-Sq = 79,7%    R-Sq(adj) = 78,7%

##### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	248100	49620	79,31	0,000
Residual Error	101	63191	626		
Total	106	311291			

Source	DF	Seq SS
Retorno PO	1	367
Retorno PNEA (%)	1	16418
Retorno PO (%)	1	227974
Retorno RMRE R\$)	1	1768
Retorno RMRH (R\$)	1	1573

##### Unusual Observations

Obs	Retorno PO	Retorno PIA	Fit	SE Fit	Residual	St Resid
10	169	67,00	26,64	12,11	40,36	1,84 X
15	18	-60,00	-9,51	4,61	-50,49	-2,05R
29	52	-72,00	-20,83	5,50	-51,17	-2,10R
40	-24	35,00	87,32	4,53	-52,32	-2,13R
82	-353	186,00	120,64	10,15	65,36	2,86R
97	72	-44,00	9,04	3,50	-53,04	-2,14R
101	115	12,00	66,08	4,36	-54,08	-2,20R

R denotes an observation with a large standardized residual.

X denotes an observation whose X value gives it large leverage.

## 4.6 STEPWISE

### Stepwise Regression: indices versus Retorno PIA\_1; Retorno PO\_1; ...

Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15

Response is indices on 6 predictors, with N = 107

Step	1	2
Constant	3,144	3,433
Retorno PO_1	-2,12	-1,76
T-Value	-5,91	-4,44
P-Value	0,000	0,000
Retorno RMRE R\$)_1		-0,96
T-Value		-2,02
P-Value		0,046
S	0,728	0,718
R-Sq	24,98	27,82
R-Sq(adj)	24,27	26,43
Mallows Cp	1,5	-0,4

O valor de R-Quadrado encontrado não é muito condizente com os tipos de dados dos indicadores, o índice mais alto com 79,7%.

As condições de inferência adicionada ao baixo valor de R-Quadrado nos denotam a impossibilidade de realizar inferências satisfatórias mediante a utilização da equação encontrada. O baixo valor do P-value nos mostra que as variáveis não são significativas para a regressão.

## 4.7 CORRELAÇÕES

### Correlations: indices; Retorno PIA\_1; Retorno PO\_1; Retorno PNEA; ...

	indices	Retorno PIA_1	Retorno PO_1
Retorno PIA_1	0,077 0,429		
Retorno PO_1	0,855 0,000	0,034 0,725	
Retorno PNEA (%)	0,093 0,341	-0,229 0,018	0,030 0,762
Retorno PO (%)_1	0,820 0,000	-0,187 0,054	0,968 0,000
Retorno RMRE R\$)	0,222 0,022	0,053 0,590	0,452 0,000
Retorno RMRH (R\$	-0,181 0,063	-0,217 0,025	-0,124 0,203

Retorno PNEA (%) Retorno PO (%)\_1 Retorno RMRE R\$)

Retorno PO (%)_1	0,064 0,513		
Retorno RMRE R\$)	-0,397 0,000	0,422 0,000	
Retorno RMRH (R\$	0,116 0,236	-0,087 0,373	0,078 0,423

Cell Contents: Pearson correlation  
P-Value

### Correlations: indices; Retorno PIA\_1

Pearson correlation of indices and Retorno PIA\_1 = 0,077  
P-Value = 0,429

### Correlations: indices; Retorno PO\_1

Pearson correlation of indices and Retorno PO\_1 = 0,855  
P-Value = 0,000

### Correlations: indices; Retorno PNEA (%)\_1

Pearson correlation of indices and Retorno PNEA (%)\_1 = 0,093  
P-Value = 0,341

### Correlations: indices; Retorno PO (%)\_1

Pearson correlation of indices and Retorno PO (%)\_1 = 0,820  
P-Value = 0,000

### Correlations: indices; Retorno RMRE R\$)\_1

Pearson correlation of indices and Retorno RMRE R\$)\_1 = 0,222  
P-Value = 0,022

### Correlations: indices; Retorno RMRH (R\$)\_1

Pearson correlation of indices and Retorno RMRH (R\$)\_1 = -0,181  
P-Value = 0,063

## 5. COMPARAÇÃO DE MÉDIAS, INTERVALOS DE CONFIANÇA, VARIÂNCIA E PAIREDT

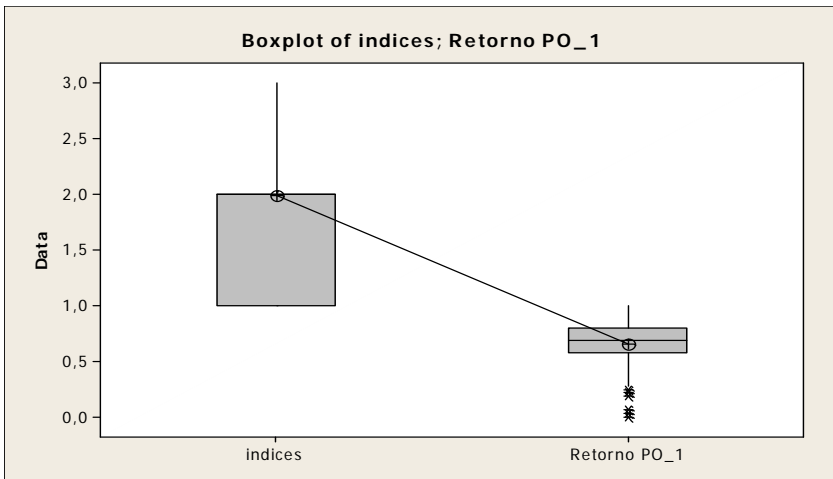
### One-way ANOVA: indices; Retorno PIA\_1

Source	DF	SS	MS	F	P
Factor	1	121,529	121,529	446,77	0,000
Error	212	57,667	0,272		
Total	213	179,196			

S = 0,5216    R-Sq = 67,82%    R-Sq(adj) = 67,67%

Individual 95% CIs For Mean Based on





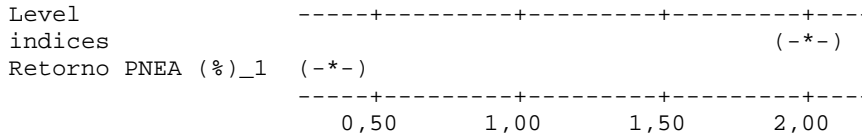
### One-way ANOVA: indices; Retorno PNEA (%)\_1

Source	DF	SS	MS	F	P
Factor	1	140,240	140,240	535,08	0,000
Error	212	55,564	0,262		
Total	213	195,804			

S = 0,5120 R-Sq = 71,62% R-Sq(adj) = 71,49%

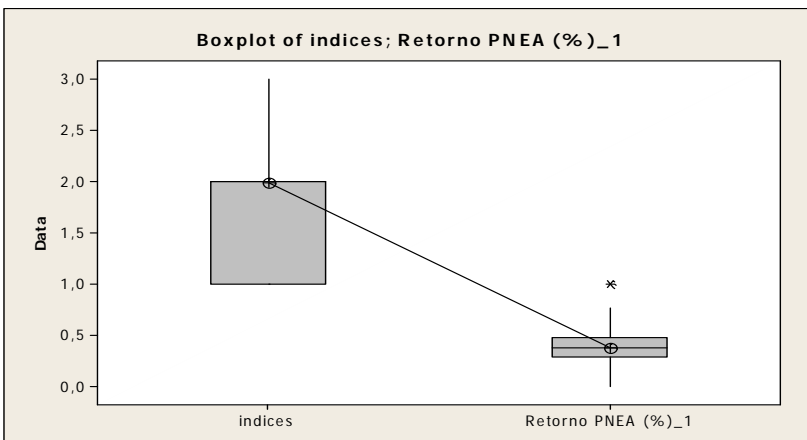
Level	N	Mean	StDev
indices	107	1,9907	0,7070
Retorno PNEA (%)_1	107	0,3716	0,1558

Individual 95% CIs For Mean Based on Pooled StDev



Pooled StDev = 0,5120

### Boxplot of indices; Retorno PNEA (%)\_1



### One-way ANOVA: indices; Retorno PO (%)\_1

Source	DF	SS	MS	F	P
Factor	1	97,796	97,796	363,86	0,000
Error	212	56,981	0,269		

Total 213 154,777

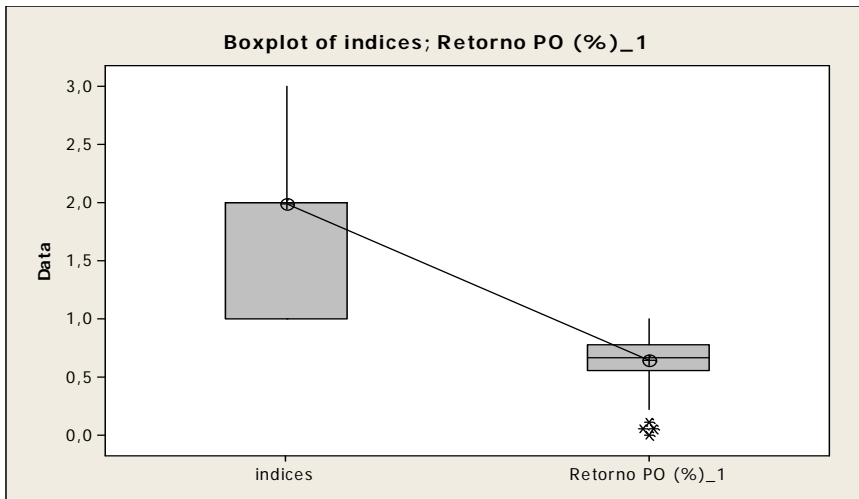
S = 0,5184 R-Sq = 63,19% R-Sq(adj) = 63,01%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
indices	107	1,9907	0,7070	(--*-)
Retorno PO (%)_1	107	0,6386	0,1940	(--*-)

0,80      1,20      1,60      2,00

Pooled StDev = 0,5184

### Boxplot of indices; Retorno PO (%)\_1



### One-way ANOVA: indices; Retorno RMRE R\$)\_1

Source	DF	SS	MS	F	P
Factor	1	111,118	111,118	421,72	0,000
Error	212	55,859	0,263		
Total	213	166,977			

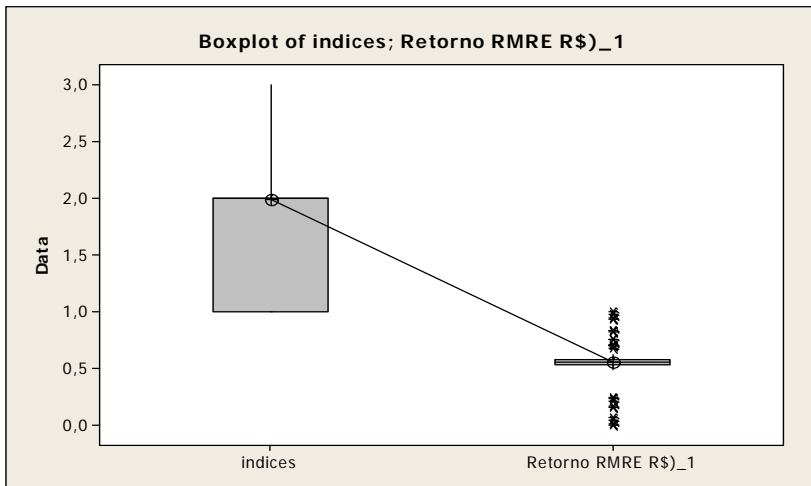
S = 0,5133 R-Sq = 66,55% R-Sq(adj) = 66,39%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
indices	107	1,9907	0,7070	(--*-)
Retorno RMRE R\$)_1	107	0,5495	0,1645	(--*-)

0,50      1,00      1,50      2,00

Pooled StDev = 0,5133

## Boxplot of indices; Retorno RMRE R\$)\_1



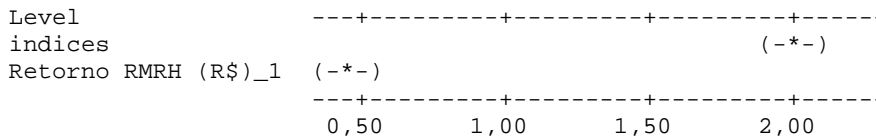
## One-way ANOVA: indices; Retorno RMRH (R\$)\_1

Source	DF	SS	MS	F	P
Factor	1	130,137	130,137	484,60	0,000
Error	212	56,931	0,269		
Total	213	187,068			

S = 0,5182    R-Sq = 69,57%    R-Sq(adj) = 69,42%

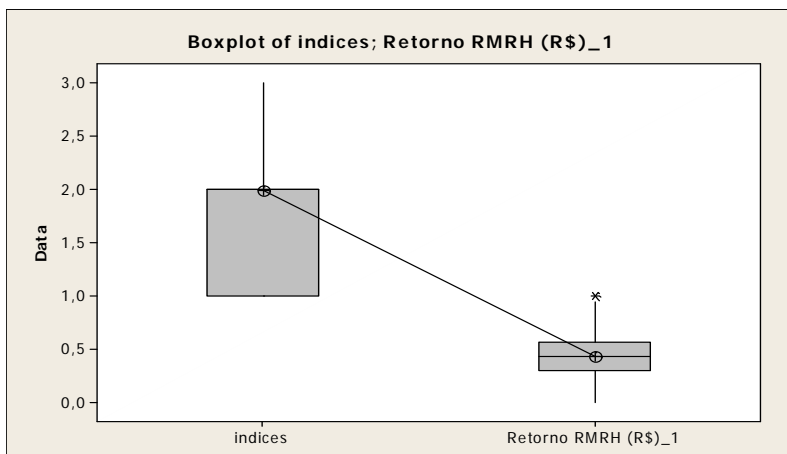
Level	N	Mean	StDev
indices	107	1,9907	0,7070
Retorno RMRH (R\$)_1	107	0,4310	0,1928

Individual 95% CIs For Mean Based on Pooled StDev



Pooled StDev = 0,5182

## Boxplot of indices; Retorno RMRH (R\$)\_1



Na tabela individual de intervalos de confiança de 95%, note que nenhum dos intervalos se sobrepõe, o que embasa a teoria de que as médias são estatisticamente diferentes. Entretanto, é necessário interpretar os resultados da comparação múltipla para verificar onde estão as diferenças nas médias dos centros de expedição.

## 6. ANÁLISE DOS COMPONENTES PRINCIPAIS

O objetivo deste tópico é, através da análise dos componentes principais, tentarmos reduzir o número de variáveis, ou seja, percebermos as relações entre as variáveis e a possibilidade de agruparmos as mesmas. Certamente a análise de correlações e dendogramas acima já nos dão uma idéia de que a possibilidade de agrupamento é grande pelos elevados índices de correlação entre todas as variáveis:

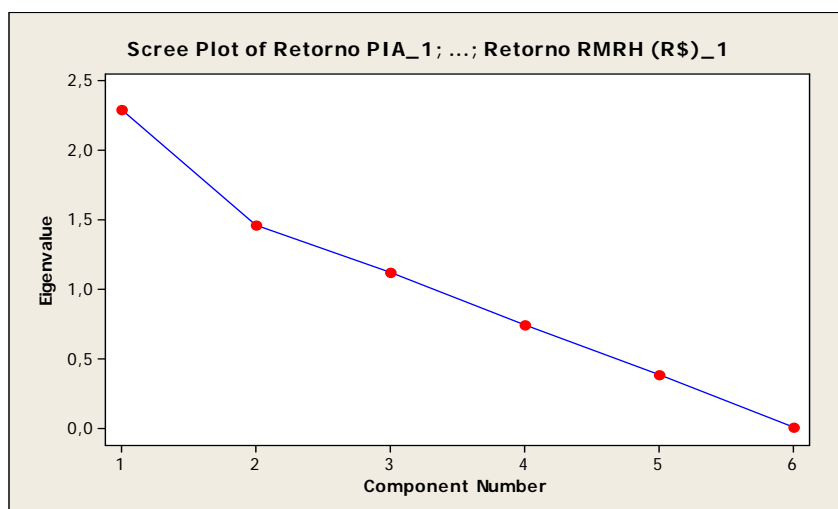
### Principal Component Analysis: Retorno PIA\_; Retorno PO\_1; Retorno PNEA; Retorno

Eigenanalysis of the Correlation Matrix

Eigenvalue	2,2898	1,4606	1,1211	0,7383	0,3837	0,0066
Proportion	0,382	0,243	0,187	0,123	0,064	0,001
Cumulative	0,382	0,625	0,812	0,935	0,999	1,000

Variable	PC1	PC2	PC3	PC4	PC5	PC6
Retorno PIA_1	-0,025	0,586	0,286	0,740	-0,020	-0,160
Retorno PO_1	0,629	-0,097	0,179	0,173	-0,222	0,695
Retorno PNEA (%)_1	-0,095	-0,647	0,365	0,379	0,544	-0,019
Retorno PO (%)_1	0,623	-0,217	0,120	-0,012	-0,246	-0,700
Retorno RMRE (R\$)_1	0,448	0,276	-0,455	-0,011	0,718	-0,017
Retorno RMRH (R\$)_1	-0,080	-0,325	-0,729	0,528	-0,280	-0,006

### Scree Plot of Retorno PIA\_1; ...; Retorno RMRH (R\$)\_1



Pela análise dos detalhes e gráfico acima percebemos que se juntarmos as 6 variáveis em apenas 1 (PC1) teremos um proporção de 38,2%, com 2 (PC1, PC2 e PC3)

chegamos a 81,2% e assim por diante. Isto é algo extremamente significativo, pois ao invés de trabalharmos com 6 variáveis poderíamos trabalhar com o índice PC1, PC2 e PC3, que já explica 81,2% das variáveis.

De acordo com todas as análises acima, percebemos claramente que o agrupamento de variáveis é bastante pertinente. Isto pôde ser observado inicialmente pelas matrizes de correlação e dendogramas e depois comprovados pela análise dos componentes principais.

## 7. ANÁLISE DE CORRESPONDÊNCIA

A análise de correspondência é um método de análise fatorial para variáveis categóricas. A AC, basicamente, 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 em dimensão reduzida, isto é, 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. A seguir, é apresentado o resultado da análise de correspondência para a tabela.

### Simple Correspondence Analysis: índices; Retorno PIA\_; Retorno PO\_1; Retorno PN

Analysis of Contingency Table

Axis	Inertia	Proportion	Cumulative	Histogram
1	0,0341	0,3852	0,3852	*****
2	0,0249	0,2812	0,6664	*****
3	0,0177	0,1993	0,8657	*****
4	0,0086	0,0974	0,9630	*****
5	0,0031	0,0353	0,9983	**
6	0,0001	0,0017	1,0000	
Total	0,0886			

Row Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	Row1	0,603	0,010	0,004	0,002	0,000	0,000	-0,145	0,603	0,008
2	Row2	0,739	0,009	0,004	0,037	0,036	0,000	0,162	0,703	0,010
3	Row3	0,705	0,010	0,001	0,064	0,699	0,001	-0,006	0,006	0,000
4	Row4	0,656	0,010	0,001	0,048	0,184	0,001	0,077	0,472	0,002
5	Row5	0,688	0,013	0,004	-0,139	0,656	0,007	-0,031	0,032	0,000
6	Row6	0,798	0,009	0,002	-0,098	0,600	0,003	0,056	0,198	0,001
7	Row7	0,897	0,011	0,013	-0,281	0,768	0,026	-0,116	0,130	0,006
8	Row8	0,751	0,009	0,007	-0,197	0,607	0,010	-0,096	0,145	0,003
9	Row9	0,550	0,006	0,028	0,110	0,031	0,002	-0,446	0,519	0,051
10	Row10	0,133	0,012	0,033	-0,176	0,123	0,011	-0,049	0,009	0,001
11	Row11	0,876	0,007	0,011	0,340	0,869	0,025	0,029	0,006	0,000
12	Row12	0,604	0,010	0,004	0,104	0,273	0,003	0,114	0,331	0,005
13	Row13	0,064	0,009	0,004	0,042	0,048	0,000	0,024	0,016	0,000
14	Row14	0,267	0,010	0,004	-0,045	0,054	0,001	-0,089	0,213	0,003
15	Row15	0,941	0,008	0,008	-0,134	0,203	0,004	0,256	0,738	0,022
16	Row16	0,567	0,007	0,010	0,162	0,208	0,005	-0,213	0,359	0,013
17	Row17	0,267	0,012	0,005	-0,043	0,055	0,001	0,083	0,212	0,003
18	Row18	0,888	0,011	0,016	-0,328	0,888	0,036	-0,006	0,000	0,000
19	Row19	0,827	0,007	0,016	0,237	0,276	0,011	-0,335	0,551	0,031

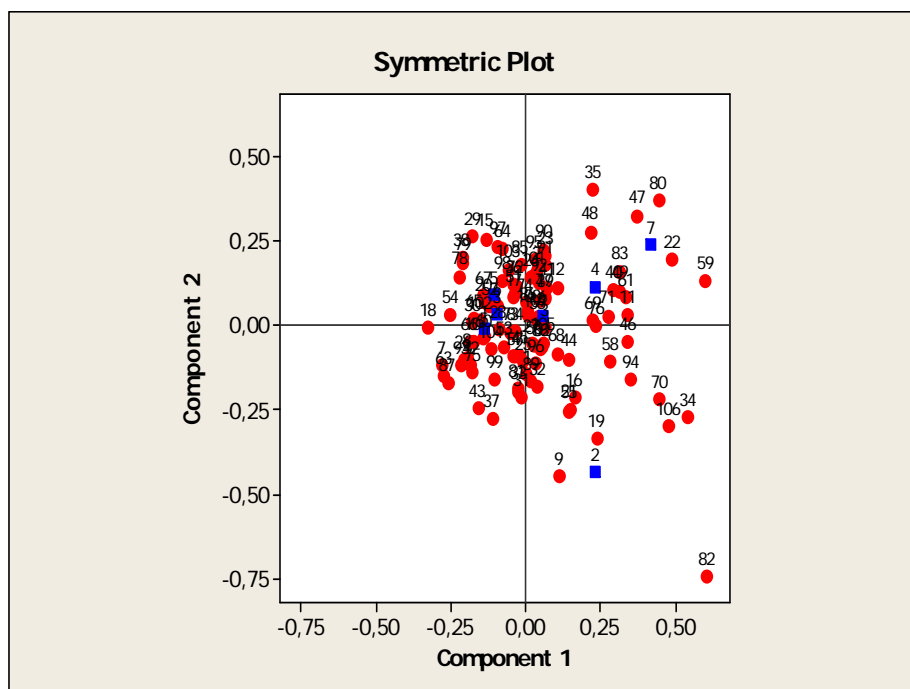
20	Row20	0,836	0,012	0,004	-0,139	0,670	0,007	0,069	0,167	0,002
21	Row21	0,681	0,010	0,014	0,147	0,178	0,007	-0,247	0,503	0,025
22	Row22	0,539	0,004	0,025	0,487	0,464	0,030	0,197	0,076	0,007
23	Row23	0,887	0,009	0,005	0,063	0,077	0,001	0,205	0,810	0,016
24	Row24	0,392	0,010	0,006	0,017	0,005	0,000	0,143	0,386	0,008
25	Row25	0,626	0,010	0,002	-0,010	0,005	0,000	-0,108	0,621	0,005
26	Row26	0,926	0,011	0,007	-0,207	0,752	0,014	-0,100	0,174	0,005
27	Row27	0,130	0,010	0,003	0,019	0,013	0,000	-0,055	0,117	0,001
28	Row28	0,394	0,012	0,003	-0,091	0,389	0,003	-0,011	0,006	0,000
29	Row29	0,825	0,009	0,012	-0,178	0,255	0,008	0,266	0,570	0,024
30	Row30	0,824	0,012	0,005	-0,180	0,821	0,011	0,011	0,003	0,000
31	Row31	0,961	0,009	0,005	-0,015	0,005	0,000	-0,212	0,956	0,017
32	Row32	0,951	0,010	0,004	0,039	0,043	0,000	-0,181	0,908	0,013
33	Row33	0,346	0,009	0,012	-0,027	0,006	0,000	-0,194	0,339	0,014
34	Row34	0,839	0,006	0,027	0,540	0,669	0,048	-0,272	0,170	0,017
35	Row35	0,877	0,007	0,019	0,224	0,206	0,010	0,404	0,672	0,044
36	Row36	0,720	0,010	0,002	-0,040	0,071	0,000	0,120	0,649	0,006
37	Row37	0,885	0,009	0,010	-0,113	0,127	0,003	-0,276	0,758	0,027
38	Row38	0,951	0,012	0,012	-0,214	0,502	0,016	0,203	0,449	0,020
39	Row39	0,716	0,009	0,002	0,065	0,284	0,001	0,080	0,432	0,002
40	Row40	0,737	0,007	0,011	0,291	0,650	0,018	0,106	0,086	0,003
41	Row41	0,689	0,010	0,003	0,068	0,185	0,001	0,113	0,505	0,005
42	Row42	0,897	0,012	0,007	-0,186	0,646	0,012	-0,116	0,251	0,006
43	Row43	0,929	0,008	0,009	-0,161	0,281	0,006	-0,244	0,648	0,020
44	Row44	0,837	0,010	0,004	0,146	0,574	0,007	-0,099	0,263	0,004
45	Row45	0,149	0,010	0,006	-0,020	0,007	0,000	-0,088	0,142	0,003
46	Row46	0,590	0,005	0,012	0,340	0,579	0,018	-0,048	0,011	0,000
47	Row47	0,818	0,007	0,025	0,373	0,464	0,030	0,325	0,354	0,031
48	Row48	0,817	0,007	0,012	0,216	0,307	0,010	0,278	0,510	0,022
49	Row49	0,778	0,007	0,011	0,314	0,703	0,021	0,103	0,076	0,003
50	Row50	0,281	0,010	0,002	0,027	0,048	0,000	-0,059	0,233	0,001
51	Row51	0,222	0,013	0,007	-0,039	0,034	0,001	0,093	0,188	0,004
52	Row52	0,868	0,009	0,002	-0,122	0,713	0,004	0,057	0,155	0,001
53	Row53	0,490	0,013	0,003	-0,072	0,283	0,002	-0,061	0,206	0,002
54	Row54	0,869	0,012	0,010	-0,252	0,858	0,022	0,030	0,012	0,000
55	Row55	0,604	0,007	0,011	0,144	0,148	0,004	-0,252	0,455	0,017
56	Row56	0,168	0,010	0,001	0,037	0,145	0,000	0,015	0,023	0,000
57	Row57	0,115	0,009	0,009	-0,037	0,016	0,000	-0,092	0,099	0,003
58	Row58	0,327	0,005	0,016	0,282	0,286	0,012	-0,107	0,042	0,002
59	Row59	0,993	0,008	0,036	0,597	0,946	0,087	0,134	0,047	0,006
60	Row60	0,563	0,009	0,004	-0,142	0,528	0,006	-0,037	0,035	0,001
61	Row61	0,856	0,007	0,012	0,336	0,803	0,024	0,086	0,053	0,002
62	Row62	0,525	0,009	0,001	0,055	0,223	0,001	-0,064	0,303	0,002
63	Row63	0,973	0,012	0,013	-0,275	0,754	0,026	-0,148	0,218	0,010
64	Row64	0,582	0,009	0,010	-0,077	0,060	0,002	0,228	0,522	0,018
65	Row65	0,536	0,012	0,008	-0,173	0,529	0,010	0,019	0,007	0,000
66	Row66	0,794	0,012	0,006	-0,187	0,748	0,012	-0,046	0,046	0,001
67	Row67	0,566	0,009	0,005	-0,143	0,410	0,005	0,088	0,156	0,003
68	Row68	0,596	0,011	0,004	0,104	0,366	0,003	-0,083	0,231	0,003
69	Row69	0,203	0,007	0,020	0,222	0,202	0,011	0,017	0,001	0,000
70	Row70	0,664	0,006	0,027	0,445	0,535	0,037	-0,219	0,129	0,012
71	Row71	0,377	0,007	0,016	0,274	0,373	0,015	0,029	0,004	0,000
72	Row72	0,559	0,010	0,003	-0,033	0,040	0,000	0,118	0,519	0,005
73	Row73	0,657	0,009	0,001	-0,053	0,584	0,001	-0,019	0,072	0,000
74	Row74	0,790	0,010	0,001	-0,001	0,000	0,000	0,069	0,790	0,002
75	Row75	0,922	0,012	0,008	-0,179	0,572	0,011	-0,140	0,350	0,010
76	Row76	0,627	0,007	0,007	0,233	0,627	0,011	0,001	0,000	0,000
77	Row77	0,246	0,013	0,006	0,062	0,087	0,001	0,083	0,159	0,004
78	Row78	0,850	0,011	0,010	-0,222	0,596	0,016	0,145	0,254	0,009
79	Row79	0,809	0,012	0,013	-0,211	0,453	0,015	0,188	0,356	0,016
80	Row80	0,925	0,008	0,031	0,443	0,543	0,044	0,372	0,382	0,043
81	Row81	0,231	0,010	0,017	-0,029	0,005	0,000	-0,186	0,226	0,013
82	Row82	0,801	0,005	0,066	0,603	0,319	0,055	-0,741	0,481	0,113
83	Row83	0,797	0,006	0,010	0,313	0,630	0,017	0,161	0,167	0,006
84	Row84	0,051	0,009	0,003	-0,035	0,042	0,000	-0,016	0,008	0,000
85	Row85	0,332	0,006	0,007	-0,017	0,003	0,000	0,181	0,329	0,008
86	Row86	0,147	0,010	0,001	0,005	0,002	0,000	0,043	0,145	0,001
87	Row87	0,983	0,011	0,013	-0,261	0,685	0,022	-0,172	0,298	0,013
88	Row88	0,267	0,013	0,004	0,049	0,093	0,001	-0,067	0,175	0,002
89	Row89	0,856	0,010	0,004	0,017	0,009	0,000	-0,163	0,847	0,011
90	Row90	0,970	0,010	0,006	0,059	0,063	0,001	0,225	0,907	0,020

91	Row91	0,398	0,007	0,007	0,065	0,046	0,001	0,180	0,352	0,009
92	Row92	0,798	0,010	0,003	0,043	0,084	0,001	0,126	0,714	0,006
93	Row93	0,737	0,012	0,011	-0,216	0,571	0,017	-0,117	0,166	0,007
94	Row94	0,603	0,005	0,015	0,352	0,500	0,019	-0,160	0,103	0,005
95	Row95	0,906	0,010	0,005	0,030	0,021	0,000	0,197	0,885	0,015
96	Row96	0,652	0,010	0,002	0,033	0,051	0,000	-0,113	0,601	0,005
97	Row97	0,998	0,009	0,006	-0,093	0,135	0,002	0,235	0,863	0,020
98	Row98	0,525	0,009	0,005	-0,079	0,141	0,002	0,130	0,384	0,006
99	Row99	0,795	0,009	0,004	-0,104	0,236	0,003	-0,160	0,558	0,009
100	Row100	0,053	0,013	0,003	0,007	0,002	0,000	0,034	0,051	0,001
101	Row101	0,724	0,010	0,003	0,024	0,019	0,000	0,149	0,705	0,009
102	Row102	0,712	0,012	0,004	-0,152	0,705	0,008	0,015	0,007	0,000
103	Row103	0,829	0,009	0,004	-0,060	0,097	0,001	0,166	0,732	0,010
104	Row104	0,695	0,009	0,003	-0,118	0,519	0,004	-0,069	0,176	0,002
105	Row105	0,075	0,010	0,009	0,056	0,041	0,001	-0,051	0,034	0,001
106	Row106	0,499	0,004	0,030	0,475	0,358	0,028	-0,299	0,142	0,015
107	Row107	0,236	0,010	0,001	0,030	0,162	0,000	0,020	0,074	0,000

#### Column Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	Column1	0,699	0,389	0,123	-0,139	0,693	0,222	-0,013	0,006	0,003
2	Column2	0,996	0,094	0,261	0,231	0,218	0,148	-0,437	0,778	0,722
3	Column3	0,450	0,128	0,033	-0,097	0,412	0,036	0,030	0,038	0,004
4	Column4	0,314	0,073	0,176	0,234	0,254	0,116	0,113	0,059	0,037
5	Column5	0,616	0,125	0,044	-0,106	0,356	0,041	0,090	0,261	0,041
6	Column6	0,043	0,107	0,114	0,057	0,034	0,010	0,029	0,009	0,004
7	Column7	0,878	0,084	0,248	0,416	0,664	0,428	0,237	0,214	0,189

#### Symmetric Plot



A análise da tabela de contingência mostra uma decomposição da inércia ( $\chi^2/n$ ). Do total da inércia da matriz de dados, 38,52% é contabilizada no primeiro componente, 28,12% é contabilizada no segundo componente e assim por diante.

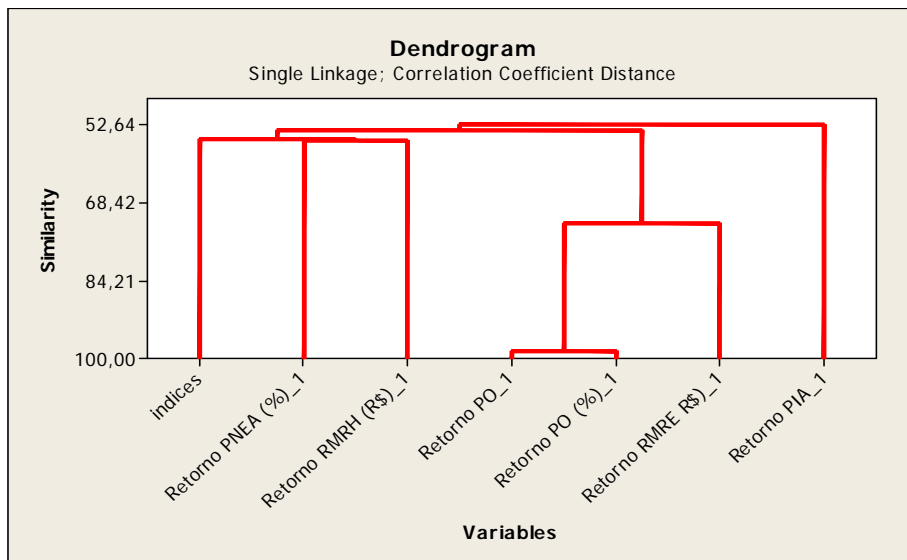
## 8. ANÁLISE DE CLUSTER E DENDOGRAMAS

### Cluster Analysis of Variables: indices; Retorno PIA\_; Retorno PO\_1; ...

Correlation Coefficient Distance, Single Linkage  
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	6	98,4094	0,031812	3	5	3
2	5	72,5788	0,548425	3	6	3
3	4	55,7828	0,884345	4	7	4
4	3	55,6350	0,887301	1	4	1
5	2	53,9103	0,921795	1	3	1
6	1	52,6357	0,947286	1	2	1

### Dendrogram



## 9. ANÁLISE DISCRIMINANTE

### 9.1 LINEAR:

#### Discriminant Analysis: indices versus Retorno PIA\_1; Retorno PO\_1; ...

Linear Method for Response: indices

Predictors: Retorno PIA\_1; Retorno PO\_1; Retorno PNEA (%)\_1; Retorno PO (%)\_1;  
Retorno RMRE R\$)\_1; Retorno RMRH (R\$)\_1

Group	1	2	3
Count	27	54	26

Summary of classification

	True Group		
Put into Group	1	2	3
1	22	0	0
2	5	50	0
3	0	4	26
Total N	27	54	26
N correct	22	50	26
Proportion	0,815	0,926	1,000

N = 107

N Correct = 98

Proportion Correct = 0,916

Squared Distance Between Groups

	1	2	3
1	0,0000	9,1758	26,9620
2	9,1758	0,0000	5,3925
3	26,9620	5,3925	0,0000

Linear Discriminant Function for Groups

	1	2	3
Constant	-38,52	-53,55	-69,09
Retorno PIA_1	70,57	71,42	73,33
Retorno PO_1	-213,82	-180,70	-155,81
Retorno PNEA (%)_1	35,66	34,42	35,26
Retorno PO (%)_1	251,70	251,45	252,60
Retorno RMRE R\$)_1	24,01	20,09	10,39
Retorno RMRH (R\$)_1	13,63	12,18	11,20

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
8**	2	3	1	21,943	0,000
			2	5,145	0,411
			3	4,424	0,589
11**	1	2	1	4,696	0,445
			2	4,260	0,553
			3	14,879	0,003
16**	1	2	1	6,788	0,173
			2	3,675	0,820
			3	13,277	0,007
25**	2	3	1	19,474	0,000
			2	3,129	0,304
			3	1,472	0,696
40**	1	2	1	8,886	0,487
			2	8,789	0,512
			3	21,263	0,001
48**	1	2	1	5,027	0,350
			2	3,793	0,648
			3	15,263	0,002
60**	2	3	1	20,227	0,000
			2	3,617	0,351
			3	2,391	0,649
89**	2	3	1	18,025	0,000
			2	3,331	0,496
			3	3,303	0,503
91**	1	2	1	5,576	0,336
			2	4,218	0,662
			3	15,852	0,002

Nesta análise comparativa discriminante, foi interessante observar que obtivemos um modelo quase perfeito considerando todas as variáveis. A proporção de correlação é de 91,6%

## 9.2 Quadrática:

### Discriminant Analysis: indices versus Retorno PIA\_1; Retorno PO\_1; ...

Quadratic Method for Response: indices

Predictors: Retorno PIA\_1; Retorno PO\_1; Retorno PNEA (%)\_1; Retorno PO (%)\_1;  
Retorno RMRE R\$)\_1; Retorno RMRH (R\$)\_1

Group	1	2	3
Count	27	54	26

Summary of classification

Put into Group	True Group		
	1	2	3
1	25	1	1
2	2	51	1
3	0	2	24
Total N	27	54	26
N correct	25	51	24
Proportion	0,926	0,944	0,923

N = 107

N Correct = 100

Proportion Correct = 0,935

From Generalized Squared Distance to Group

Group	1	2	3
1	-25,15	5,39	52,28
2	-20,85	-29,36	-16,20
3	-9,63	-18,81	-28,26

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
8**	2	3	1	-9,78	0,002
			2	-20,36	0,412
			3	-21,06	0,586
48**	1	2	1	-22,52	0,404
			2	-23,29	0,596
			3	3,86	0,000
60**	2	3	1	-11,70	0,001
			2	-22,70	0,329
			3	-24,12	0,669
78**	3	2	1	-12,10	0,010
			2	-20,03	0,553
			3	-19,57	0,437
91**	1	2	1	-21,21	0,319
			2	-22,73	0,681
			3	10,46	0,000
93**	3	1	1	-11,97	0,608
			2	8,79	0,000
			3	-11,10	0,392

105**	2	1	1	-17,15	0,872
			2	-13,31	0,128
			3	8,58	0,000

Nesta análise comparativa discriminante, foi interessante observar que obtivemos um modelo quase perfeito considerando todas as variáveis. A proporção de correlação é de 93,5%.

Entre os dois modelos (Linear e Quadrática), o melhor que representa a amostra é a Quadrática com 93,5%.

## 10. REGRESSÃO LOGÍSTICA:

### Ordinal Logistic Regression: indices versus Retorno PIA\_; Retorno PO\_1; ...

Link Function: Logit

Response Information

Variable	Value	Count
indices	1	54
	2	26
	3	27
Total		107

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower
Const(1)	-4,53105	1,99260	-2,27	0,023		
Const(2)	-3,17966	1,97256	-1,61	0,107		
Retorno PIA_1	-0,678646	2,10374	-0,32	0,747	0,51	0,01
Retorno PO_1	4,81109	8,59217	0,56	0,576	122,87	0,00
Retorno PNEA (%)_1	-0,417533	1,48150	-0,28	0,778	0,66	0,04
Retorno PO (%)_1	-0,782994	8,75586	-0,09	0,929	0,46	0,00
Retorno RMRE R\$)_1	4,11049	1,99455	2,06	0,039	60,98	1,22
Retorno RMRH (R\$)_1	-0,333514	1,05578	-0,32	0,752	0,72	0,09

Predictor	Upper
Const(1)	
Const(2)	
Retorno PIA_1	31,33
Retorno PO_1	2,52989E+09
Retorno PNEA (%)_1	12,02
Retorno PO (%)_1	12970406,16
Retorno RMRE R\$)_1	3040,38
Retorno RMRH (R\$)_1	5,67

Log-Likelihood = -96,316

Test that all slopes are zero: G = 29,147, DF = 6, P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	195,429	206	0,690
Deviance	192,632	206	0,739

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

Pairs	Number	Percent	Summary Measures
Concordant	2305	64,7	Somers' D 0,30
Discordant	1247	35,0	Goodman-Kruskal Gamma 0,30
Ties	12	0,3	Kendall's Tau-a 0,19
Total	3564	100,0	

### Ordinal Logistic Regression: indices versus Retorno PIA\_1; Retorno PO\_1

Link Function: Logit

Response Information

Variable	Value	Count
indices	1	54
	2	26
	3	27
Total		107

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Const(1)	-3,21082	0,889200	-3,61	0,000			
Const(2)	-1,92501	0,853369	-2,26	0,024			
Retorno PIA_1	-0,104840	0,910922	-0,12	0,908	0,90	0,15	5,37
Retorno PO_1	4,65925	1,14104	4,08	0,000	105,56	11,28	987,96

Log-Likelihood = -100,308

Test that all slopes are zero: G = 21,162, DF = 2, P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	195,465	210	0,756
Deviance	200,617	210	0,667

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

Pairs	Number	Percent	Summary Measures
Concordant	2160	60,6	Somers' D 0,21
Discordant	1404	39,4	Goodman-Kruskal Gamma 0,21
Ties	0	0,0	Kendall's Tau-a 0,13
Total	3564	100,0	

### Ordinal Logistic Regression: indices versus Retorno PNEA; Retorno PO (

Link Function: Logit

Response Information

Variable	Value	Count
indices	1	54
	2	26
	3	27
Total		107

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Const(1)	-0,327149	0,214511	-1,53	0,127			
Const(2)	0,975587	0,236913	4,12	0,000			
Retorno PNEA (%)	-0,989520	0,579996	-1,71	0,088	0,37	0,12	1,16
Retorno PO (%)	2,67502	0,652028	4,10	0,000	14,51	4,04	52,09

Log-Likelihood = -99,408

Test that all slopes are zero: G = 22,962, DF = 2, P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	177,847	152	0,074
Deviance	178,362	152	0,071

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	2313	64,9	Somers' D 0,30
Discordant	1236	34,7	Goodman-Kruskal Gamma 0,30
Ties	15	0,4	Kendall's Tau-a 0,19
Total	3564	100,0	

A representatividade do modelo Regressão Logística é de 64,9%.

De acordo com o Log-Likelihood, indicando um P-value de 0,000, não há evidência de que ao menos um dos coeficientes é diferente de zero, considerando o teste de hipótese nula de 100%.

O P-Value do teste de Pearson e Deviance estão com valor igual a 1, indicando que não há evidências suficientes para afirmar que o modelo não ajusta os dados adequadamente.

A tabela de frequências observadas e esperadas nos dá uma noção de qualidade do ajuste do modelo. Quanto mais próximos estão estes valores, melhor o ajuste e menores são os erros do modelo.

O Resumo das Medidas nos indicam uma habilidade de predição que varia de 0,19 e 0,30, o que não é satisfatório

## 11. ÁRVORES DE CLASSIFICAÇÃO

Além da fácil visualização e entendimento dos resultados, árvores de classificação possuem uma característica técnica que as diferenciam de outras técnicas: são ótimas para identificar padrões não lineares.

A maioria das técnicas multivariadas trabalha com pré-requisitos de linearidade dos dados. Em determinadas situações de negócios, esses pré-requisitos não são conseguidos. É nesse momento que uma árvore de classificação pode resolver o problema de negócio. Reproduzo abaixo uma situação-exemplo simplificada em duas dimensões de como NÃO É POSSÍVEL determinar uma reta para “separar” clientes que permanecem daqueles que abandonam a marca (*churn*). A solução nesse caso é possível com uma árvore de classificação e os perfis de classificação obtidos.

XLSTAT 2009.3.02 - Árvores de classificação e regressão - em 6/20/2011 às 12:05:03 AM

Y / Qualitativas: Documento = Árvore - trabalho final Rosimar.xls / Planilha = Plan1 / Intervalo = Plan1!\$K\$1:\$K\$108 / 107 linhas e 1 coluna

X / Quantitativas: Documento = Árvore - trabalho final Rosimar.xls / Planilha = Plan1 / Intervalo = Plan1!\$C\$1:\$H\$108 / 107 linhas e 6 colunas

Método: CHAID

Medida: Pearson

Profundidade máxima da árvore: 5

Nível de significância (%): 5

Umbra de separação (%): 5

Autorizar a redivisão: Correção de Bonferroni / Umbra de reagrupamento (%): 5

Tamanho mínimo de um pai: 2 / Tamanho mínimo de um filho: 1

Número de intervalos: 10

### Estatísticas descritivas:

Variável	Categorias	Freqüências	%
indices	1	54	50,467
	2	26	24,299
	3	27	25,234

Variável	Observações	Obs. com dados faltantes	Obs. sem dados faltantes	Mínimo	Máximo	Média	Desvio padrão
Retorno PIA	107	0	107	-72,000	186,000	52,738	54,191
Retorno PO	107	0	107	-392,981	273,849	45,873	131,406
Retorno PNEA (%)	107	0	107	-0,800	1,300	-0,020	0,327
Retorno PO (%)	107	0	107	-1,100	0,700	0,050	0,349
Retorno RMRE (R\$)	107	0	107	-394,728	325,890	1,240	118,547
Retorno RMRH (R\$)	107	0	107	-45,090	60,305	0,337	20,321

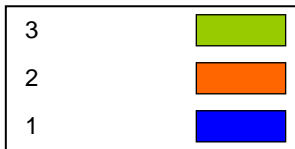
Matriz de correlação:

Variáveis	Retorno PIA	Retorno PO	Retorno PNEA (%)	Retorno PO (%)	Retorno RMRE R\$)	Retorno RMRH (R\$)
Retorno PIA	<b>1,000</b>	0,034	-0,229	-0,187	0,053	-0,217
Retorno PO	0,034	<b>1,000</b>	0,030	0,968	0,452	-0,124
Retorno PNEA (%)	-0,229	0,030	<b>1,000</b>	0,064	-0,397	0,116
Retorno PO (%)	-0,187	0,968	0,064	<b>1,000</b>	0,422	-0,087
Retorno RMRE R\$)	0,053	0,452	-0,397	0,422	<b>1,000</b>	0,078
Retorno RMRH (R\$)	-0,217	-0,124	0,116	-0,087	0,078	<b>1,000</b>

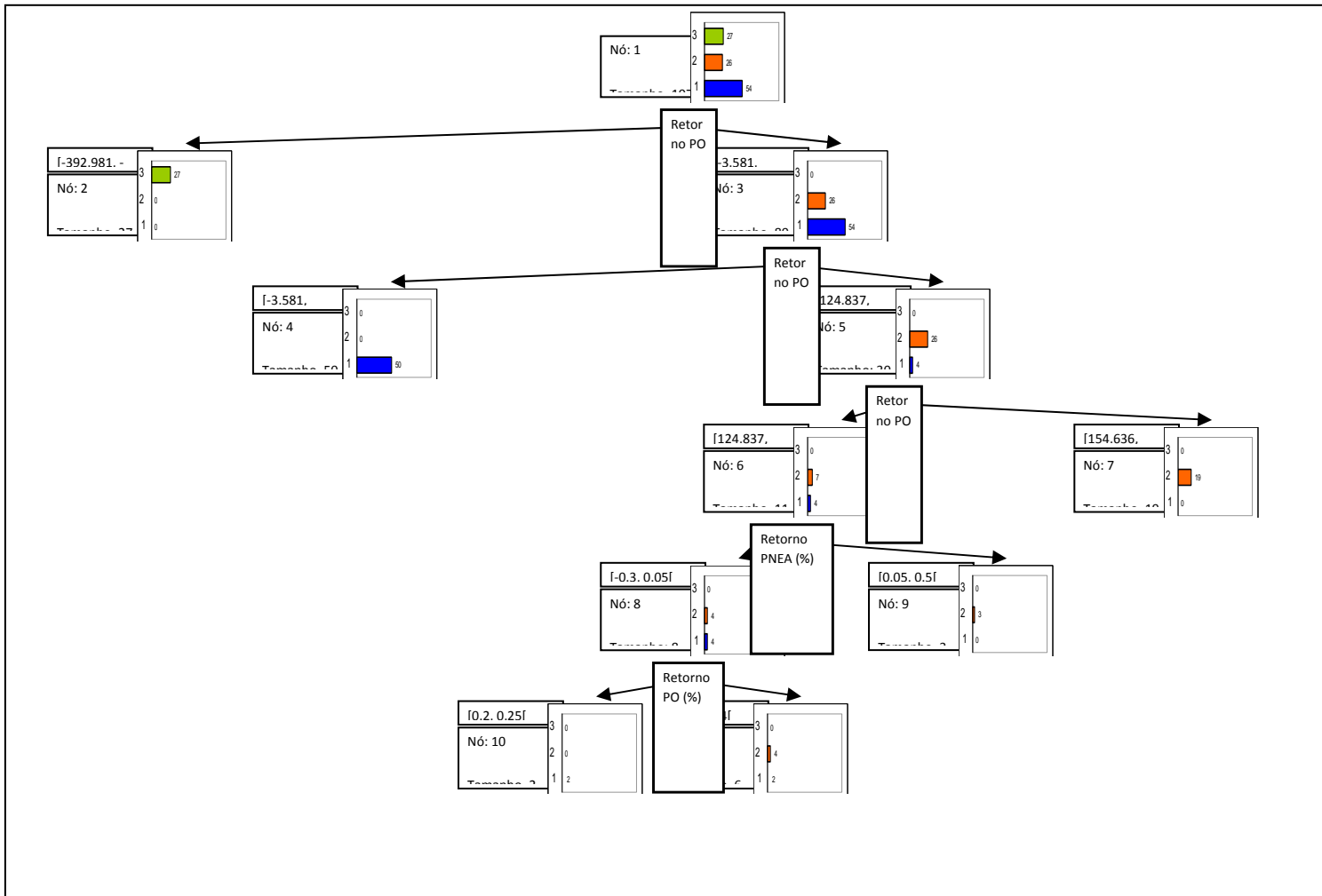
Estrutura da árvore:

Nó	p-valor	Objetos	%	Nó pai	Filhos	Variável de separação	Valores	Pureza
1	0,841	107	100,00%		2, 3			50,47%
2	0,000	27	25,23%	1		Retorno PO	[-392.981, -3.581[	100,00%
3	0,896	80	74,77%	1	4, 5	Retorno PO	[-3.581, 273.849[	67,50%
4	0,000	50	46,73%	3		Retorno PO	[-3.581, 124.837[	100,00%
5	0,515	30	28,04%	3	6, 7	Retorno PO	[124.837, 273.849[	86,67%
6	0,463	11	10,28%	5	8, 9	Retorno PO	[124.837, 154.636[	63,64%
7	0,000	19	17,76%	5		Retorno PO	[154.636, 273.849[	100,00%
8	0,577	8	7,48%	6	10, 11	Retorno PNEA (%)	[-0.3, 0.05[	50,00%
9	0,000	3	2,80%	6		Retorno PNEA (%)	[0.05, 0.5[	100,00%
10	0,000	2	1,87%	8		Retorno PO (%)	[0.2, 0.25[	100,00%
11	0,000	6	5,61%	8		Retorno PO (%)	[0.25, 0.4[	66,67%

## LEGENDA:



## ÁRVORE DE CLASSIFICAÇÃO



## 12 CONSIDERAÇÕES FINAIS

Foi possível observar que, tanto pelo aplicativo Minitab (pela Regressão Logística nominal) quanto pelo aplicativo XLSTAT (Árvore de classificação e Regressão), as variáveis que apresentam maior importância na separação dos grupos é a variável População Ocupada com 100% de pureza.

Podemos concluir e a ferramenta que melhor representa a análise é a Árvore de Classificação com 100% de acertos, em segundo lugar é a análise Discriminante representada no modelo Quadrática com 93,5%.