



PONTÍFICIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
Faculdade de Economia, Administração, Contabilidade e Atuariais.

ATLAS BRASIL 2013

DIMENSÃO GESTÃO FISCAL DOS MUNICÍPIOS

Disciplina: Métodos Quantitativos

Professor: Dr. Arnaldo Jose de Hoyos

Silmara C. Gomes

1º Semestre 2014

1. INTRODUÇÃO

O presente trabalho tem por objetivo efetuar uma análise Multivariada, uma análise de conglomerados, análise discriminante, regressão lógica, análise de correspondências e arvores de classificação dos dados da Pesquisa Firjan/FGV sobre o Desenvolvimento dos Municípios nos períodos de 2000 e 2010. Inicia-se com o entendimento dos dados, incluindo a definição dos indivíduos e das variáveis, suas classificações em variáveis categóricas ou quantitativas, os significados e unidades de medida, além da apresentação da tabela de dados.

O software estatístico utilizado é o **MINITAB16**.

2. ENTENDENDO OS DADOS

2.1 Os Indivíduos

Os indivíduos desta análise são os municípios brasileiros ano 2012. Trata-se de um total de 5565 municípios distribuídos em 27 unidades federativas, sendo 26 estados e um distrito federal. Os dados analisados de cada município são as variáveis que descrevemos a seguir.

2.2 As Variáveis

São 7 as variáveis desta pesquisa, incluindo o nome dos municípios. As mesmas são melhor explicadas na Tabela 1. Ressaltamos que todos os dados desta pesquisa são referentes ao ano de 2012.

Tabela 1. As Variáveis

Variável	Significado	Tipo	Unidade de Medida
UF	Abreviação de Unidade Federativa (ou Unidade da Federação) do Brasil. As UF do Brasil são entidades autônomas, com governo e constituição próprias, que em seu conjunto constituem a República Federativa do Brasil. (IBGE, 2013)	Variável Categórica	N/A
Município	O município é a divisão administrativa autônoma da UF. São as unidades de menor hierarquia dentro da organização político administrativa do Brasil, criadas através de leis ordinárias das Assembléias Legislativas de cada Unidade da Federação e sancionadas pelo Governador. (IBGE, 2013)	Variável Categórica	N/A

UF2	Apresenta a sigla que representa as Unidades Federativas (ou Unidades da Federação) do Brasil.	Variável Categórica	N/A
IFGF	Índice Firjan de Gestão Fiscal. O IFGF é composto por cinco indicadores: Receita Própria, Gastos com Pessoal, Investimentos, Liquidez e Custo da Dívida. Os quatro primeiros possuem peso 22,5% e o último 10,0% no resultado final do Índice. A seguir, a descrição de cada um deles. (FIRJAN, 2013)	Variável Quantitativa	0 - 1 (qto + próx. de 1 melhor)
Receita Própria	É, de acordo com a FIRJAN (2013), “a dependência de transferências intergovernamentais engessa o orçamento dos municípios, uma vez que é pautada por incertezas e está sujeita á programação das administrações superiores. Assim, um município deve ser capaz de gerar um alto nível de receita própria para ter condições de colocar em prática projetos de longo prazo. Este indicador tem o objetivo de verificar o grau de autonomia das receitas do município. A receita própria do município é formada por IPTU, ISS, IRFF e Outras Receitas Próprias”.	Variável Quantitativa	0 - 1 (qto + próx. de 1 melhor)
Pessoal	Refere-se aos Gastos com Pessoal, ou seja, é "a despesa com pessoal é o principal item da despesa do setor público. Este indicador busca avaliar o comprometimento das receitas com as despesas de pessoal, a formula de cálculo leva em consideração, bases da Lei de Responsabilidade Fiscal – LRF". (FIRJAN, 2013)	Variável Quantitativa	0 - 1 (qto + próx. de 1 melhor)
Investimentos	“o objetivo deste indicador é medir a parcela dos investimentos nos orçamentos municipais. Estabeleceu-se que a nota de corte seria investir 20% da RCL – Receita Corrente Líquida”. (FIRJAN, 2013)	Variável Quantitativa	0 - 1 (qto + próx. de 1 melhor)
Liquidez	"o indicador procura verificar se o município possui recursos financeiros suficientes para fazer frente ao montante de restos a pagar. Se o município apresentar mais restos a pagar do que ativos financeiros disponíveis a pontuação será zero". (FIRJAN, 2013)	Variável Quantitativa	0 - 1 (qto + próx. de 1 melhor)
Custo da Dívida	"este indicador avalia o peso dos encargos da dívida em relação às receitas líquidas reais, destacando-se que o limite é de 13%". (FIRJAN, 2013)	Variável Quantitativa	0 - 1 (qto + próx. de 1 melhor)

2.3 A Tabela de Dados

Tabela 2. Tabela de Dados

UF	Município	UF2	IFGF	Receita Própria	Pessoal	Investimentos	Liquidez	Custo da Dívida
Goiás	Abadia de Goiás	GO	0,68	0,36	0,45	0,84	0,93	1,00
Minas Gerais	Abadia dos Dourados	MG	0,35	0,22	0,43	0,55	0,00	0,82
Goiás	Abadiânia	GO	0,54	0,48	0,61	0,16	0,75	0,86
Minas Gerais	Abaeté	MG	0,35	0,28	0,45	0,41	0,07	0,83
Pará	Abaetetuba	PA	0,16	0,13	0,00	0,20	0,00	0,81
Ceará	Abaiara	CE	0,46	0,06	0,85	0,77	0,00	0,80
Bahia	Abaíra	BA	0,53	0,13	0,55	0,29	0,99	0,93
Bahia	Abaré	BA	0,21	0,11	0,00	0,42	0,00	0,88
Paraná	Abatiá	PR	0,46	0,27	0,52	0,19	0,80	0,61
Santa Catarina	Abdon Batista	SC	0,69	0,09	0,86	0,90	0,80	0,93
Pará	Abel Figueiredo	PA	0,51	0,22	0,60	0,88	0,22	0,80
Santa Catarina	Abelardo Luz	SC	0,53	0,29	0,61	1,00	0,13	0,75
Minas Gerais	Abre Campo	MG	0,41	0,35	0,67	0,48	0,00	0,69
Pernambuco	Abreu e Lima	PE	0,36	0,22	0,71	0,32	0,00	0,83
Tocantins	Abreulândia	TO	0,40	0,15	0,82	0,43	0,00	0,81
Minas Gerais	Acaiaca	MG	0,41	0,10	0,46	0,06	0,80	0,88
Maranhão	Açailândia	MA	0,45	0,26	0,73	0,61	0,09	0,67
Bahia	Acajutiba							
Pará	Acará	PA	0,19	0,10	0,00	0,39	0,00	0,74
Ceará	Acarape	CE	0,27	0,09	0,41	0,36	0,00	0,76
Ceará	Acaraú	CE	0,39	0,18	0,58	0,52	0,00	1,00
Rio Grande do Norte	Acari	RN	0,54	0,12	0,46	0,44	0,96	0,89
Piauí	Acauã	PI	0,38	0,03	0,49	0,72	0,00	1,00
Rio Grande do Sul	Aceguá	RS	0,69	0,17	0,54	1,00	0,93	0,97
Ceará	Acopiara	CE	0,38	0,17	0,58	0,57	0,10	0,62

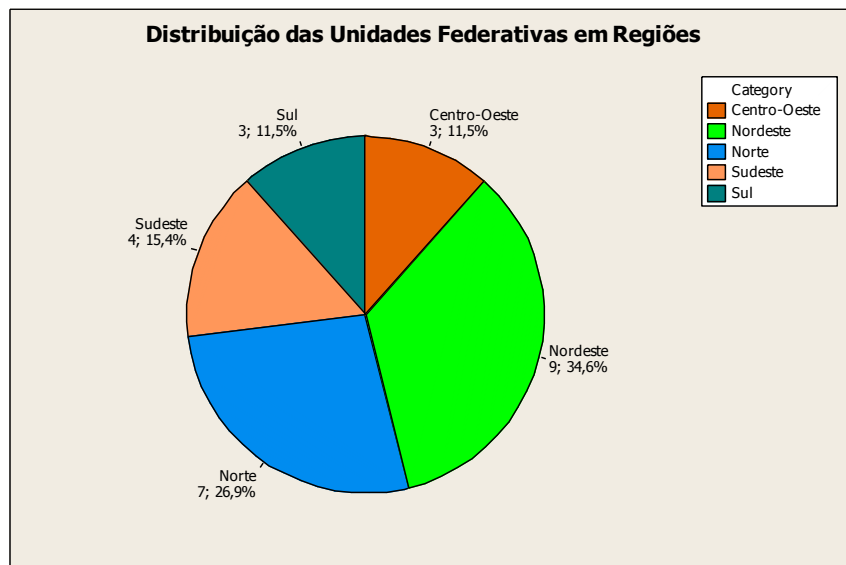
3. ANÁLISE DAS VARIÁVEIS

3.1 Variáveis Categóricas

Este tipo de variável indica que o foco de concentração deve ser a análise de gráficos do tipo pie chart e barras.

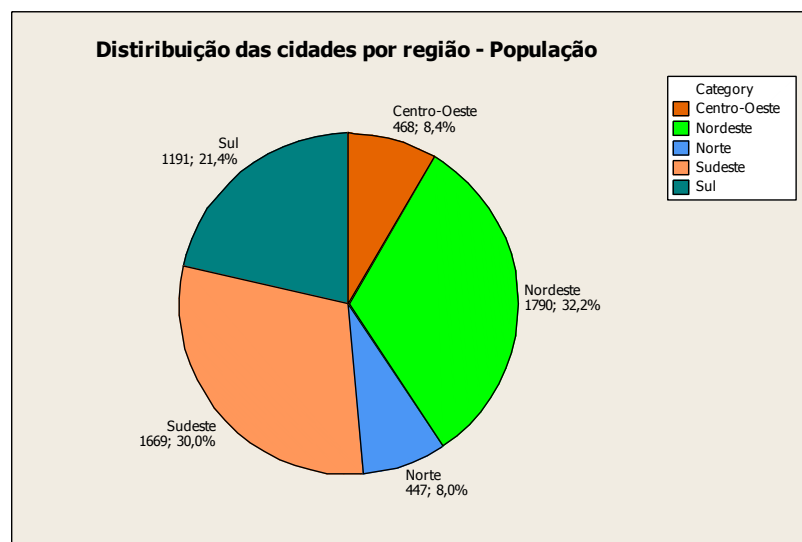
3.1.1 Variável: “UF” e “UF2”

Nossa amostra totaliza 26 unidades federativas e 1 distrito federal. As unidades federativas estão distribuídas em 5 regiões.



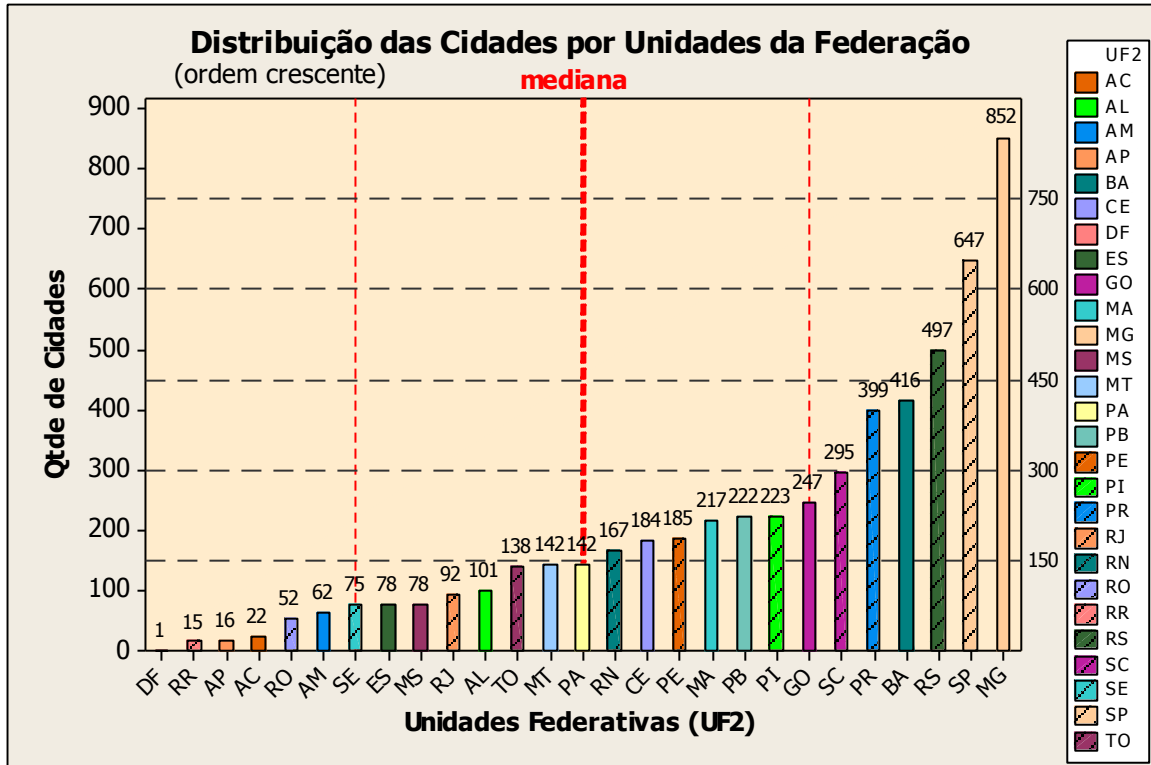
3.1.2 Variável: “Municípios”

Os gráficos abaixo nos ajudam a entender melhor o comportamento desta variável



No que diz respeito a relação regiões e cidades pode-se observar no gráfico acima que as regiões Nordeste (32,2%), Sudeste (30,0%) e Sul (21,4%) concentram 83, 6% dos municípios do território nacional, enquanto as demais regiões, Norte (8,0%) e Centro-Oeste (8,4%) somam apenas 16, 4% dos municípios. Além da concentração dos municípios brasileiros, as três regiões tem em comum o fato de serem as três regiões banhadas

significativamente pelo oceano Atlântico. Fato este, que ajuda a entender a concentração nestas regiões.



3.2 VARIÁVEIS QUANTITATIVAS

A análise deste tipo de variável permite a utilização de uma maior gama de ferramentas de análise como histogramas, curvas de densidade, gráfico de ramos, box-plot e dot-plot, além de informações numéricas como média, desvio-padrão, mediana, quartis, 5 números, intervalo de confiança e teste de normalidade de Anderson-Darling.

3.2.1. CORRELAÇÃO LINEAR

CORRELAÇÃO DA VARIÁVEL IFGF E VARIÁVEIS RP, PES, INVEST, LIQ e CD.

A matriz de correlação, apresentada abaixo, inclui o teste de significância P-Value. Nesta matriz foi relacionado, o índice de Pearson. Este índice de correlação não implica em haver uma relação de causa e efeito entre as variáveis. Para melhor apresentação dos dados na matriz, adotou-se o uso de abreviações para o desenvolvimento da matriz, a saber: IFGF (IFGF), Receita Própria (RP), Despesa com Pessoal (PES), Investimentos (INVEST), Custo da Dívida (CD) e Liquidez (LIQ).

Correlations: IFGF; RP; PES; INVEST; LIQ; CD

	IFGF	RP	PES	INVEST	LIQ
RP	0,463 0,000				
PES	0,557 0,000	0,123 0,000			
INVEST	0,602 0,000	0,074 0,000	0,201 0,000		
LIQ	0,766 0,000	0,229 0,000	0,216 0,000	0,173 0,000	
CD	0,138 0,000	-0,146 0,000	0,080 0,000	0,038 0,004	0,043 0,001

Cell Contents: Pearson correlation
P-Value

As correlações marcadas em amarelo são significativas de acordo com o P-Value (para este trabalho consideraremos significativo P-Value $\geq 0,050$). Apenas as variáveis INVEST e LIQ possuem um coeficiente de correlação satisfatoriamente forte com a variável dependente IFGF.

A correlação é sempre um número entre zero e um, e mede a intensidade de relações lineares. A correlação entre as variáveis analisadas é positiva praticamente em todos os casos. No que tange as correlações entre as variáveis da Gestão Fiscal dos municípios, as que apresentam índices de correlação mais fortes e positivos, são entre as variáveis o IFGF e as demais variáveis, exceção feita ao Custo da Dívida (CD), sendo as mais expressivas LIQ (Liquidez) e INVEST (Investimentos). Portanto, podemos afirmar que IFGF e as variáveis LIQ e INVEST são relações lineares.

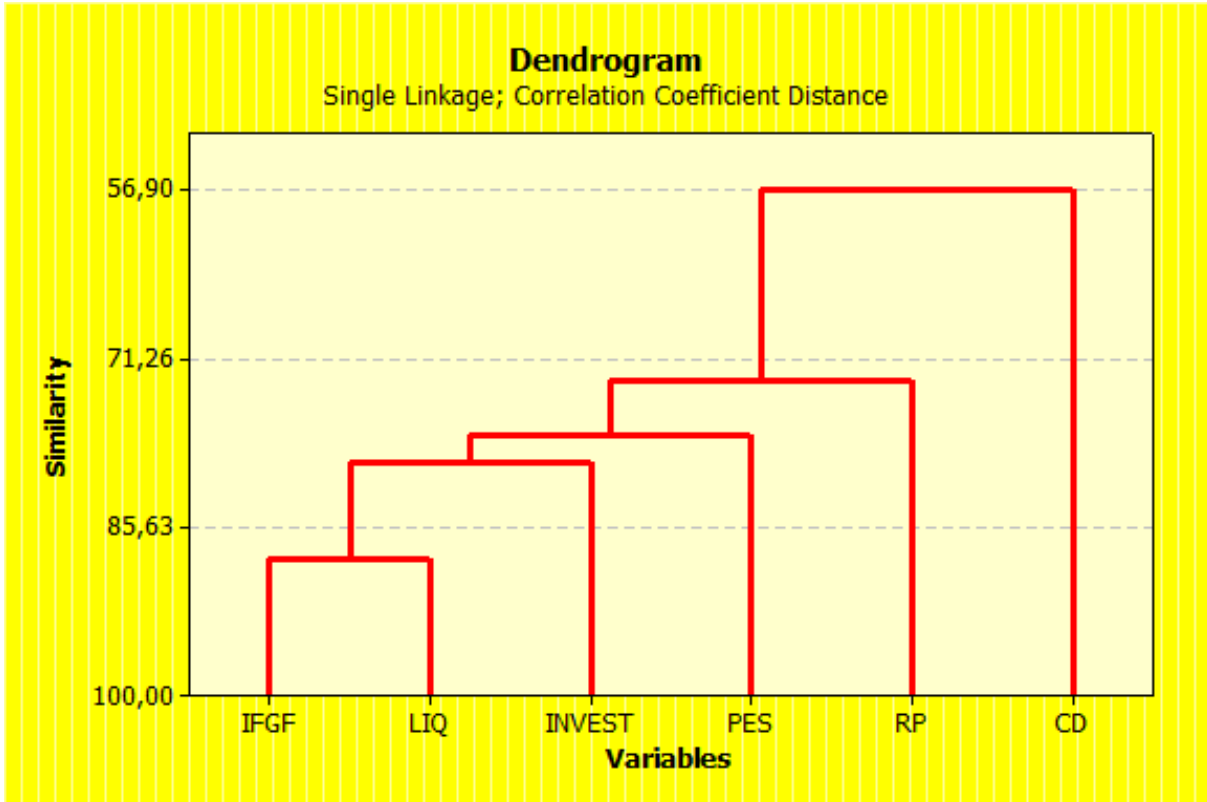
3.2.2. DENDOGRAMA

Um Dendrograma (dendr(o) = árvore) é um tipo específico de diagrama ou representação icônica que organiza determinados fatores e variáveis. É um diagrama de similaridade.

A interpretação de um dendrograma de similaridade entre amostras fundamenta-se na intuição: duas amostras próximas devem ter também valores semelhantes para as variáveis medidas. Ou seja, elas devem ser próximas matematicamente no espaço multidimensional. Portanto, quanto maior a proximidade entre as medidas relativas às amostras, maior a similaridade entre elas o dendrograma hierarquiza esta similaridade de modo que podemos ter uma visão bidimensional da similaridade ou dissimilaridade de todo o conjunto de amostras utilizado no estudo.

Segue abaixo o Dendrograma das variáveis analisadas:

Figura 2 – Dendrograma das variáveis da Gestão Fiscal

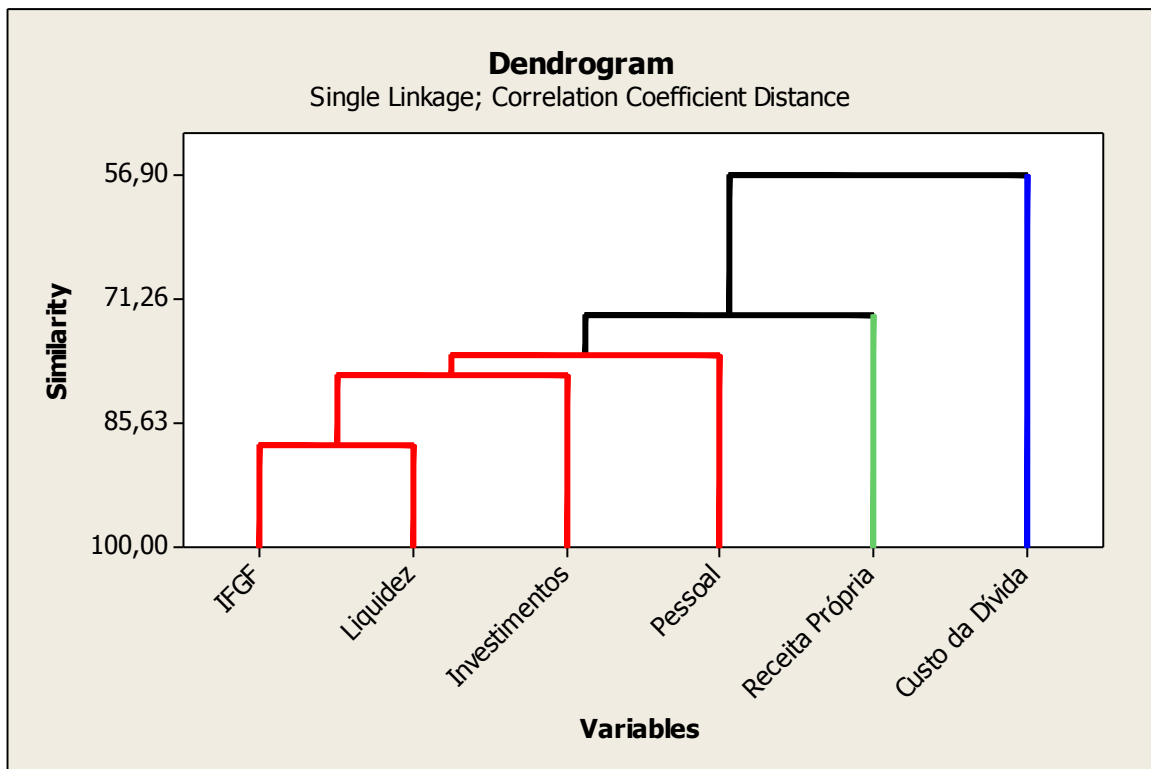


Pode-se concluir pelo Dendrograma que toda a variável, exceto CD (Custo da Dívida), tem um grau significativo de similaridade entre elas, sendo em ordem decrescente: IFGF (IFGF), Liquidez (LIQ), Investimentos (INVEST). Despesa com Pessoal (PES) e Receita Própria (RP).

Podemos concluir então segundo esta visão do Dendrograma, que a variável do primeiro grupo tem uma relação mais próxima da variável: IFGF (IFGF), Liquidez (LIQ), Investimentos (INVEST). Despesa com Pessoal (PES) e Receita Própria (RP).

Correlation Coefficient Distance, Single Linkage
Amalgamation Steps

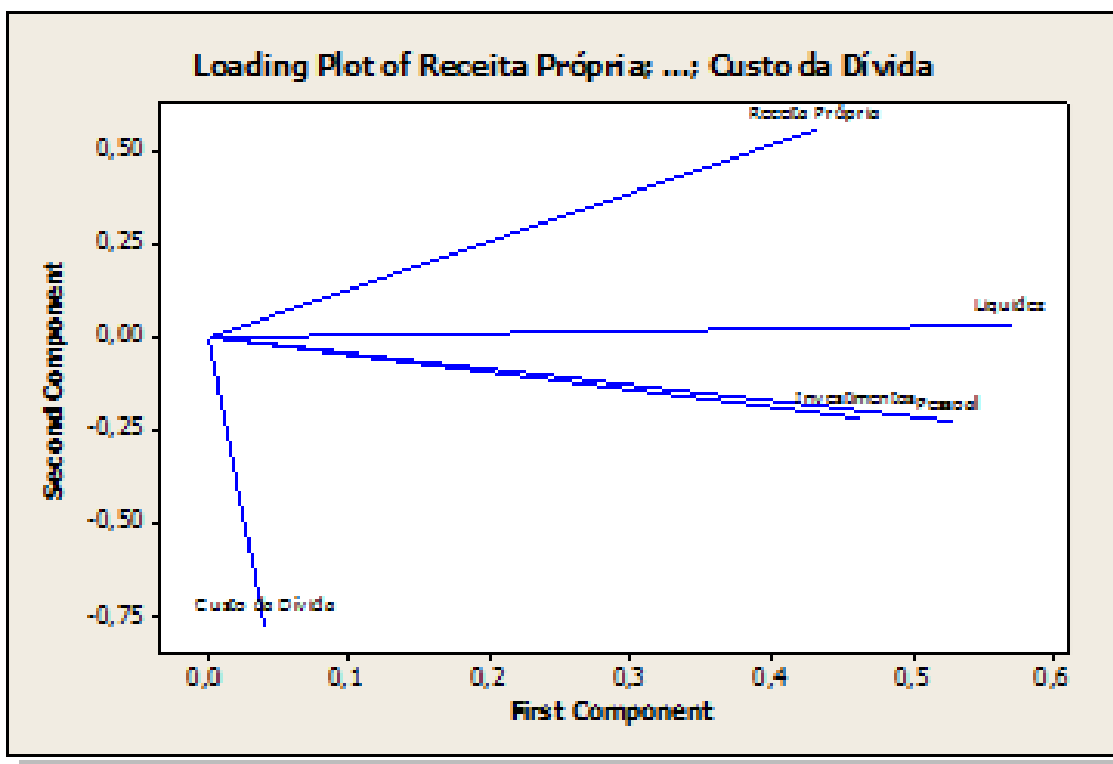
Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	5	88,3036	0,233929	5	1	2
2	4	80,0923	0,398153	4	1	3
3	3	77,8700	0,442601	3	1	4
4	2	73,1337	0,537326	2	1	5
5	1	56,8961	0,862079	1	1	6



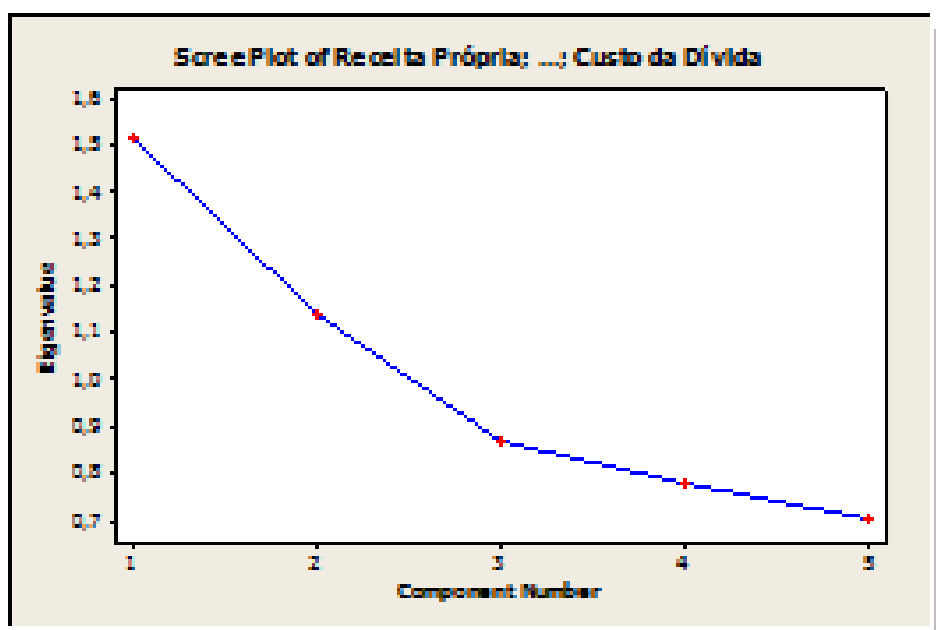
Dendrograma dos agrupamentos das variáveis por similaridade

Neste gráfico pode-se visualizar que as variáveis que possuem um percentual de similaridade forte, na faixa de 80% são IFGF, LIQUIDEZ, INVESTIMENTOS E PESSOAL, que se encontram em um nível de similaridade pouco abaixo de 80%.

3.2.3. PRINCIPAIS COMPONENTES



Podemos observar 3 grupos de dados sendo o primeiro composto pelas variáveis: Liquidez, Investimentos e Despesa com Pessoal, o segundo pela variável: Receita própria, e o terceiro pela variável: Custo da Dívida.



Existe um peso muito grande da primeira variável e as demais estão distantes. A variável 2 possui peso maior que 1, e as demais não dá para aproveitar pois estão abaixo de 1.

Principal Component Analysis: Receita Próp; Pessoal; Investimento; Liquidez; Custo de dívida

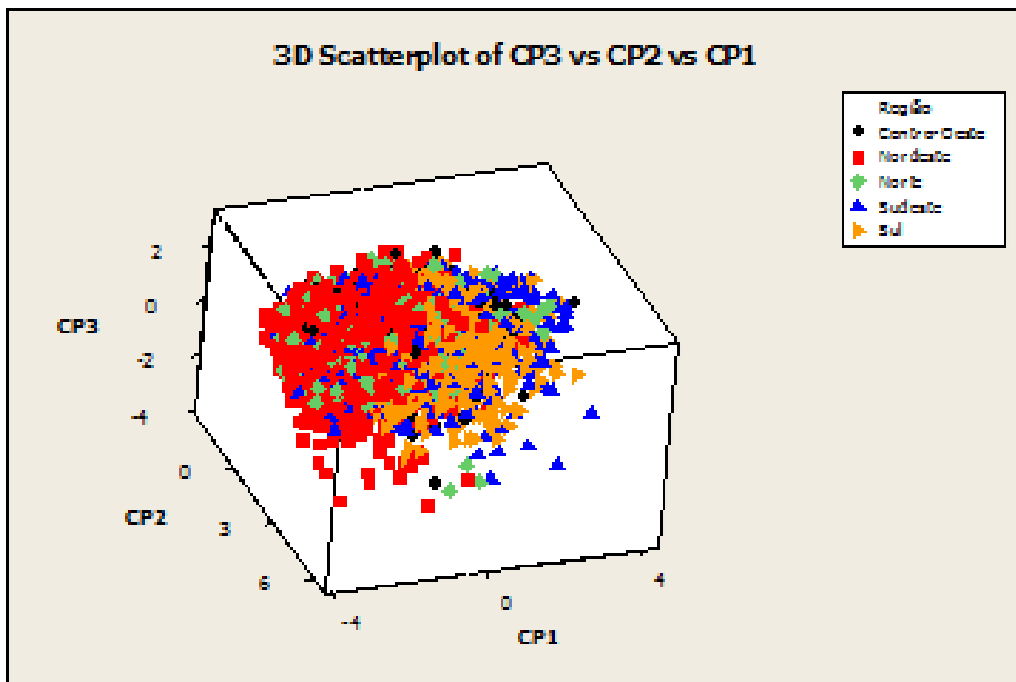
Eigenanalysis of the Correlation Matrix

Eigenvalue	1,5162	1,1371	0,8694	0,7758	0,7016
Proportion	0,303	0,227	0,174	0,155	0,140
Cumulative	0,303	0,531	0,705	0,860	1,000

(*) Eigenvalue maior que 1 (um)

Variable	PC1	PC2	PC3	PC4
Receita Própria	0,431	0,549	0,299	-0,081
Pessoal	0,526	-0,232	-0,037	0,817
Investimentos	0,460	-0,219	-0,749	-0,385
Liquidez	0,569	0,032	0,380	-0,367
Custo da Dívida	0,042	-0,772	0,452	-0,209

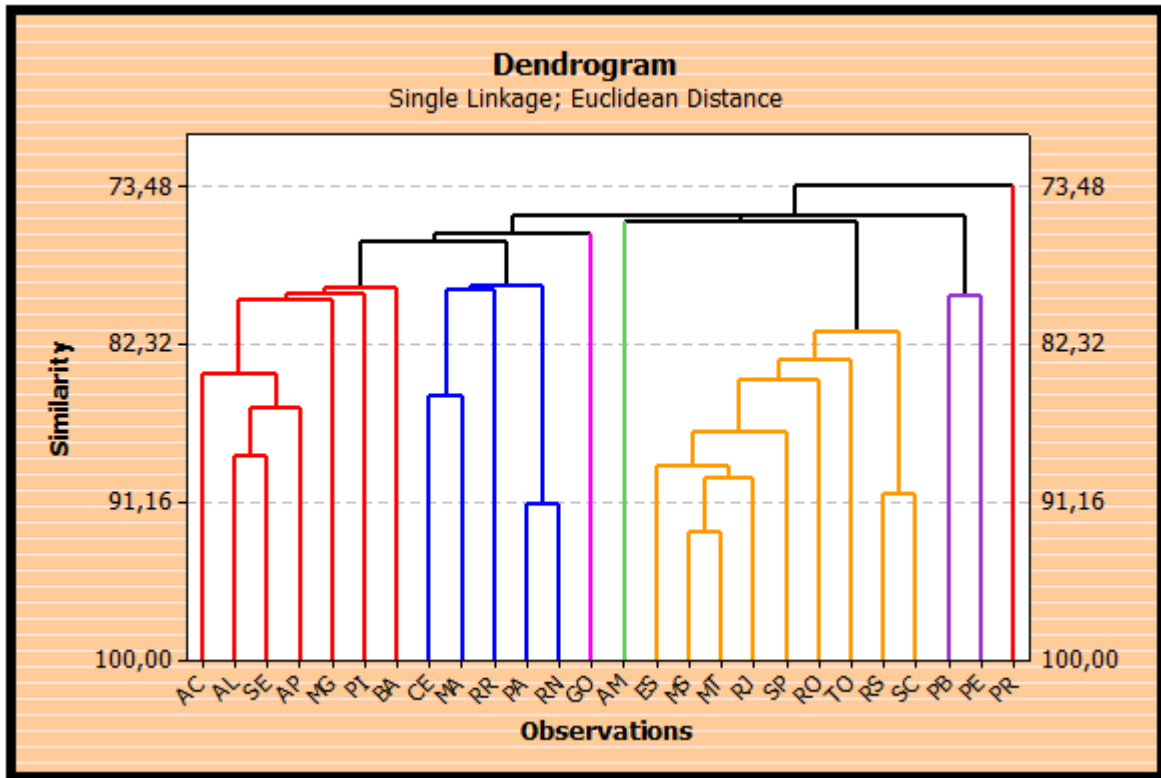
Os gráficos abaixo apresentam uma visão tridimensional dos agrupamentos das colunas em 3 variáveis CP1, CP2 e CP3, e está agrupado por região. Dentro do minitab é possível movimentar o gráfico ou girá-lo em várias rotações, para encontrar a melhor visualização.



Como os dados dos municípios são muito grandes (5565), dificulta um pouco a visualização.

4. DENDOGRAMA DE IFGF POR ESTADO (-DF)

O Dendograma permite uma análise do grau de similaridade dos dados para geramos o Dendrograma de IFGF por Estado



Dendrograma da variável IFGF por estados do Brasil (classificação não supervisionada)

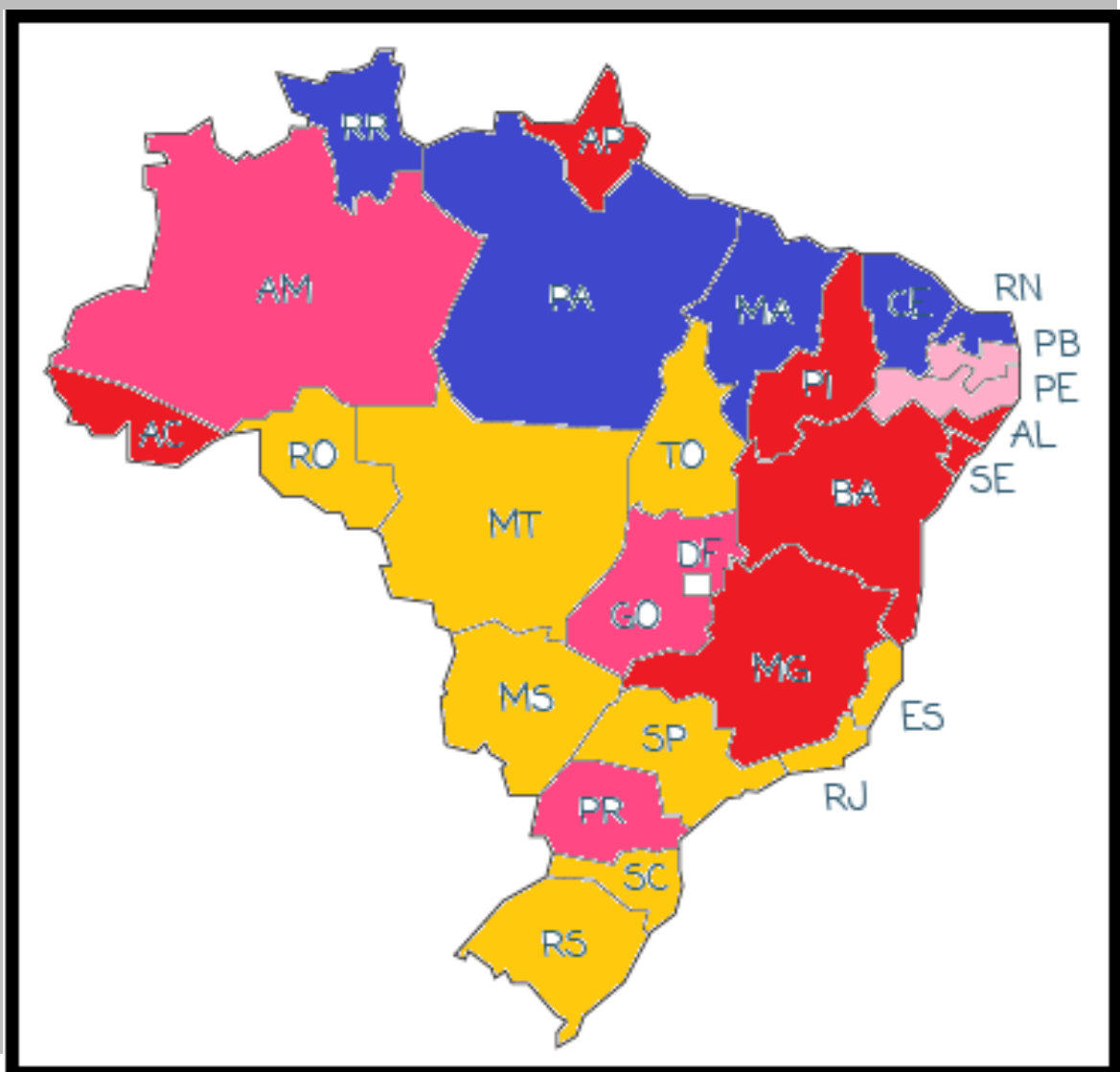
Na figura acima podemos verificar três grandes grupos de variáveis, agrupadas pela similaridade dos dados. O nível de similaridade dos dados destes estados está entre 73,48 a 81,68%, conforme indicado na escala apresentada no eixo Y do gráfico.

Os 3 grandes agrupamentos de dados, compostos pelos grupos de 5 a 9 estados do Brasil, além de cinco estados que ficaram isolados por não terem seus dados em similaridade com os outros estados. Estes estados isolados são: Goiás, Amazonas, Paraná, além da Paraíba e Pernambuco que apresentaram certa similaridade.

Na classificação não supervisionada não se tem informações prévias sobre estes grupos. Não se tem informações sobre os por quês ou os critérios de agrupamento utilizados neste agrupamento.

Podemos observar estados com alto nível de similaridade o que significa que a desigualdade é baixa. O menor nível de desigualdade se encontra nos estados mais próximos

do eixo X, por exemplo, Mato Grosso e Mato Grosso do Sul, que tem um nível de similaridade próximo de 92,88%.



Mapa do Brasil representando os três grupos de estados por similaridades de IFGF (classificação não supervisionada)

Quando o nível de desigualdade é baixo poderíamos erroneamente dizer que a situação é boa. Isso não é verdade. Baixa desigualdade não significa que as coisas vão bem, e sim que existe um padrão nos municípios do estado em termos de IFGF, uma maior similaridade entre estes municípios, e não é possível responder se esta similaridade é boa ou não.

Na figura acima (figura 3) podemos observar o agrupamento por similaridade de IFGF do dendrograma no mapa político do Brasil.

4.1 ANÁLISES DAS VARIÂNCIAS DE LIQUIDEZ E IFGF POR UF (– DF)

A análise das variâncias permite a verificação e visualização das médias e desvios padrões da variável a ser analisada. O gráfico BOXPLOT ilustra os agrupamentos, o seu tamanho (largura) varia de acordo com a quantidade de dados de cada grupo e amplitude dos dados (comprimento), e também é possível visualizar as ocorrências de outliers (marcas fora das caixas) dentro de um grupo de dados.

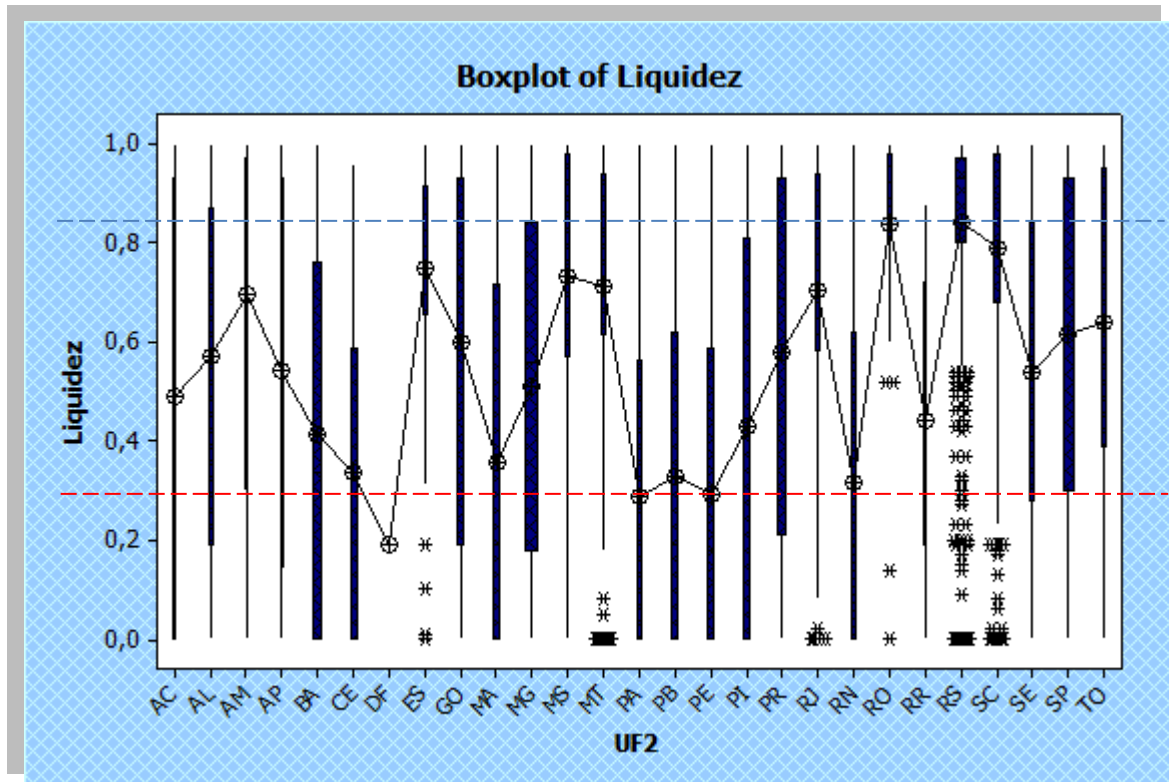


Gráfico BOXPLOT de LIQUIDEZ por Unidade Federativa

Pode-se visualizar no gráfico da figura 4, uma grande variabilidade sobre as médias de liquidez por unidades federativas. A UF que apresenta maior variabilidade dos dados é Acre. E o Rio Grande do Sul apresenta uma baixa variabilidade dos dados de liquidez, embora tenha muitos outliers (o maior de todos) que são os dados muito distantes das médias.

O resultado deste comando não fica armazenado na base de dados, é necessário copiar da área session para a área worksheet, para cada variável gerada. Com isso tem-se os dados dos 5565 municípios do Brasil, resumidos pela média e pelo desvio padrão. A partir destes dados resumidos, fica mais fácil trabalhar os dados, uma vez que estando resumido se torna mais simples a sua manipulação e análise.

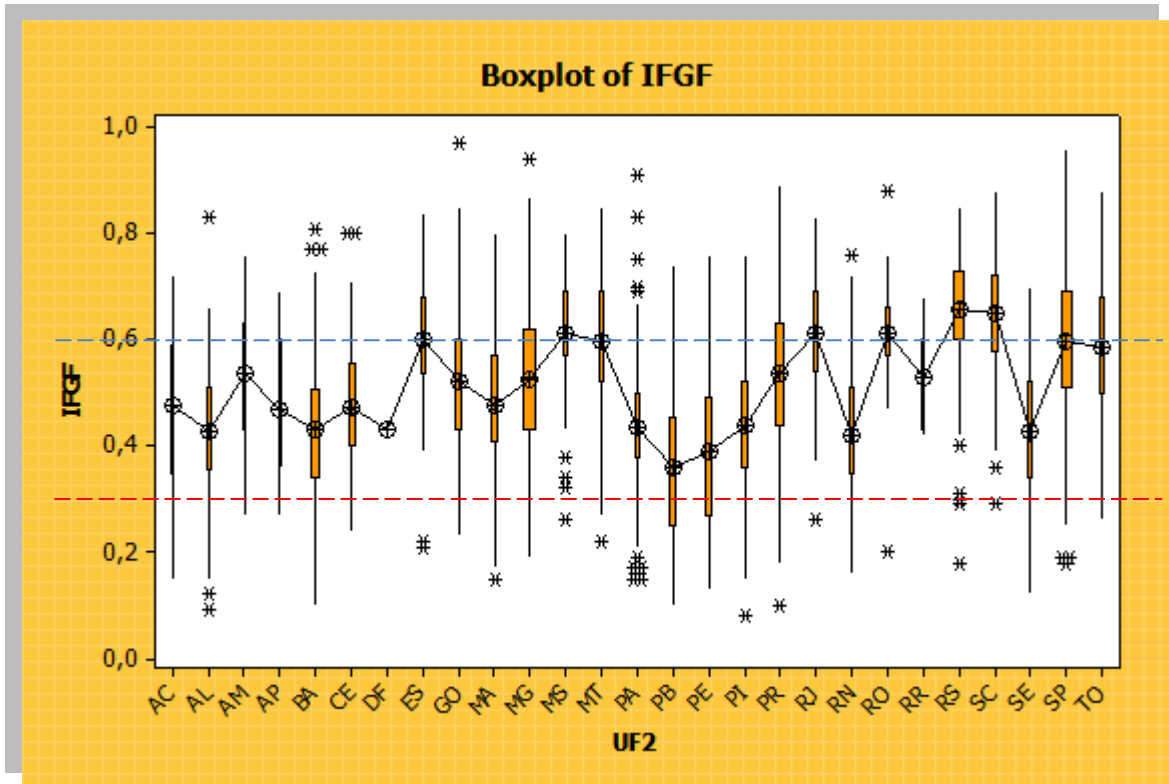


Gráfico BOXPLOT de IFGF por Unidade Federativa

Podemos verificar na figura 5 que existe uma variação grande entre as médias das UFs do Brasil, no que diz respeito à IFGF. O tamanho das caixas de cada estado representa a variância dos dados de IFGF de cada UF, e os sinais * representam o outliers ou pontos fora da curva, que são dados ou muito acima ou abaixo da média dos dados do estado. O estado que apresenta a maior média de IFGF é também o Rio Grande do Sul (acima de 0,6579), e o estado que apresenta a menor média é Pará, pouco acima de 0,3604.

Abaixo podemos visualizar os dados descritivos gerados pelo comando, para a variável IFGF e, na sequência, pela liquidez.

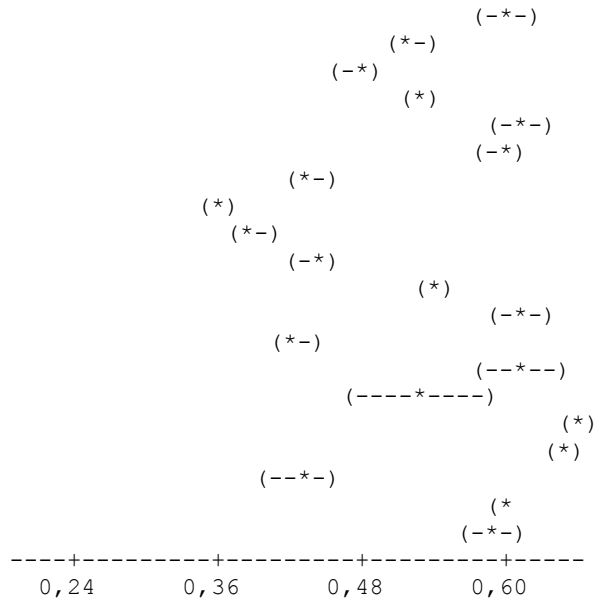
One-way ANOVA: IFGF versus UF2

Source	DF	SS	MS	F	P
UF2	26	40,5810	1,5608	106,53	0,000
Error	5538	81,1359	0,0147		
Total	5564	121,7169			

S = 0,1210 R-Sq = 33,34% R-Sq(adj) = 33,03%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
AC	22	0,4782	0,1469	(---*---)
AL	101	0,4261	0,1260	(-*)
AM	62	0,5373	0,1164	(---*)
AP	16	0,4687	0,1373	(----*----)
BA	416	0,4302	0,1245	(*)
CE	184	0,4723	0,1106	(*--)

ES	78	0,6003	0,1236
GO	247	0,5211	0,1259
MA	217	0,4778	0,1177
MG	852	0,5269	0,1219
MS	78	0,6129	0,1069
MT	142	0,5961	0,1232
PA	142	0,4340	0,1254
PB	222	0,3605	0,1403
PE	185	0,3889	0,1436
PI	223	0,4387	0,1200
PR	399	0,5381	0,1260
RJ	92	0,6103	0,1189
RN	167	0,4215	0,1220
RO	52	0,6110	0,0969
RR	15	0,5300	0,0951
RS	497	0,6579	0,0966
SC	295	0,6484	0,1013
SE	75	0,4292	0,1261
SP	647	0,5965	0,1271
TO	138	0,5872	0,1235

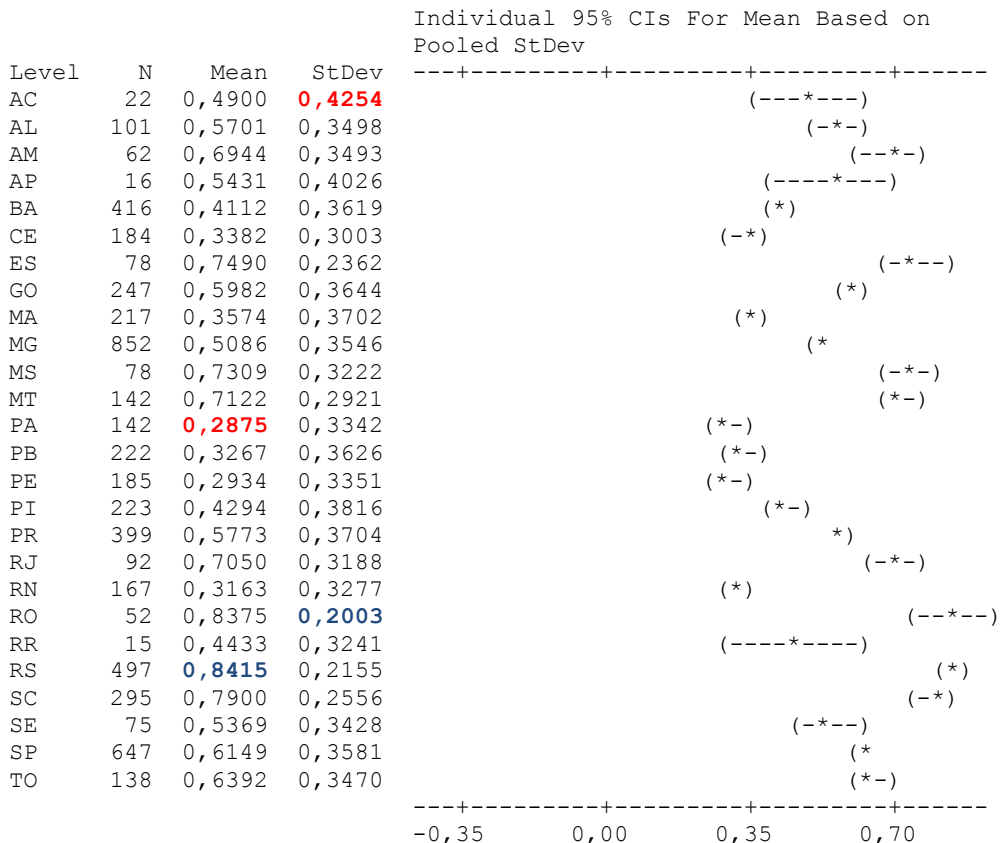


Pooled StDev = 0,1210

One-way ANOVA: Liquidez versus UF2

Source	DF	SS	MS	F	P
UF2	26	152,708	5,873	52,16	0,000
Error	5538	623,566	0,113		
Total	5564	776,273			

S = 0,3356 R-Sq = 19,67% R-Sq(adj) = 19,29%



Pooled StDev = 0,3356

Assim como podemos observar nos gráficos de Boxplot, nas tabelas de dados, as unidades da federação que apresentam maior amplitude dos dados, ou seja, possuem alta variabilidade dos dados em relação à média, e os que se destacam nestes casos são: AC, PB e PE, enquanto as que apresentaram menor variabilidade do IFGF são: RS, SC e RO. No que tange a liquidez, os maiores desvios padrões ficaram com os estados do AP e AC, enquanto os menores foram dos estados de RO e RS..

Observando ainda os dados descritivos, notamos que alguns estados possuem valores médios maiores do IFGF, como RS, e outros com menor média PB. O mesmo pode ser notado com os dados da liquidez, onde a maior média é do RS e a menor média é do AC.

As análises comparativas dos dados permitem um resumo dos dados através de cálculos específicos como médias e desvios padrões, tornando a análise dos dados mais fácil e simples. Os gráficos de Boxplot e Dendograma são excelentes figuras visuais para que se possa analisar e interpretar os diferentes comportamentos dos dados. No dendograma pode-se analisar as similaridades dos dados e no Boxplot pode-se ver as relações entre as médias e as variâncias dos agrupamentos analisados. Trata-se de ferramentas úteis para análise de grandes volumes de dados.

5. ANÁLISE DISCRIMINANTE LINEAR POR REGIÃO

A análise discriminante é uma técnica da estatística multivariada utilizada para discriminar e classificar objetos, e estuda a separação de objetos de uma população em duas ou mais classes. Neste caso queremos discriminar os valores de IFGF dos municípios do Brasil, e utilizaremos inicialmente a variável categórica Região. Para geração de análise discriminante utilizaremos o comando do Minitab:

```
STAT >> MULTIVARIATE >> DISCRIMINANT ANALYSIS
```

Obtendo como resultado os seguintes dados:

Discriminant Analysis: Região versus IFGF; Receita Própria; ...

```
After subtracting group means,
IFGF is highly correlated with other predictors.
```

```
After subtracting group means,
Liquidez is highly correlated with other predictors.
```

```
Linear Method for Response: Região
```

Predictors: IFGF; Receita Própria; Pessoal; Investimentos; Liquidez; Custo da Dívida

Group	Centro-Oeste	Nordeste	Norte	Sudeste	Sul
Count	468	1790	447	1669	1191

Summary of classification

Put into Group	True Group				
	Centro-Oeste	Nordeste	Norte	Sudeste	Sul
Centro-Oeste	157	168	61	348	186
Nordeste	69	971	110	301	102
Norte	72	381	134	271	104
Sudeste	74	78	51	358	124
Sul	96	192	91	391	675
Total N	468	1790	447	1669	1191
N correct	157	971	134	358	675
Proportion	0,335	0,542	0,300	0,214	0,567

N = 5565 N Correct = 2295 **Proportion Correct = 0,412**

A região que errou mais é Sudeste (0,214 ou 21,4%) e a que acertou mais é o Sul (0,675 ou 67,5%). O gráfico exibe o cruzamento de dados entre as regiões. Por exemplo, a região Sudeste possui 1669 municípios e apenas 352 correspondem a região, sendo que 391 são semelhantes aos dados da região Sul. O nome desta matriz é confusion matrix ou matriz de confusão. Pode-se concluir que o agrupamento por região não é uma boa escolha segundo esta avaliação, pois o percentual de acerto é muito baixo.

3.2.2. ANÁLISE DISCRIMINANTE LINEAR POR “4 BRASIS”

Esta segunda análise está interessada em verificar os possíveis agrupamento de dados utilizando a variável 4 Brasis, calculada no exercício anterior, e demonstra os agrupamentos do Brasil segundo sua proximidade de dados de Gestão Fiscal. Em decorrência da análise do dendrograma do exercício anterior, chegou-se a quatro grupos por níveis de similaridade, os quais agrupam as unidades federativas como descritas abaixo:

Grupo	Unidades Federativas	Obs.
1	AC, AL, SE, AP, MG, PI, BA	Nível de similaridade é de 79,17%
2	CE, MA, RR, PA, RN	Nível de similaridade é de 79,06%
3	ES, MS, MT, RJ, SP, RO, TO, RS, SC	Nível de similaridade é de 81,68%
4	GO, AM, PB, PE, PR	São as UF que não formaram clusters, similaridade um pouco maior que 73%.

Seguindo o procedimento descrito no item anterior no MiniTab, encontrou-se o seguinte resultado:

Discriminant Analysis: 4Brasis versus IFGF; Receita Própria; ...

After subtracting group means,
IFGF is highly correlated with other predictors.

After subtracting group means,
Liquidez is highly correlated with other predictors.

Linear Method for Response: 4Brasis

Predictors: IFGF; Receita Própria; Pessoal; Investimentos; Liquidez;
Custo da Dívida

Group	1	2	3	4
Count	1705	725	2019	1115

5564 cases used, 1 cases contain missing values

Summary of classification

		True Group			
Put into Group	1	2	3	4	
1	444	119	198	213	
2	535	396	247	298	
3	390	78	1372	271	
4	336	132	202	333	
Total N	1705	725	2019	1115	
N correct	444	396	1372	333	
Proportion	0,260	0,546	0,680	0,299	

N = 5564 N Correct = 2545 Proportion Correct = 0,457

Apesar do maior nível de similaridade geral, 45,7%, dos “4Brasis” quando comparado com a das “Regiões”, o nível de similaridade contínua muito baixo, principalmente, nos grupos 1 e 4, respectivamente, 26% e 29,9%.

3.2.3. ANÁLISE DISCRIMINANTE LINEAR POR “3 BRASIS”

Esta terceira análise está interessada em verificar os possíveis agrupamento de dados utilizando a variável 4 Brasis, reduzidas em 3Brasis, calculada com base nos dados do item anterior, ela demonstra os agrupamentos do Brasil segundo sua proximidade de dados de Gestão Fiscal. Em decorrência da análise dos dados do item anterior, chegou-se a 3 grupos por níveis de similaridade, agrupando o grupo 1 e 4, cujos o nível de acerto foram os mais baixos. As unidades federativas ficaram distribuídas em 3 grupos como descritas abaixo:

Grupo	Unidades Federativas	Obs.
1	AC, AL, SE, AP, MG, PI, BA - GO, AM, PB, PE, PR	Nível de similaridade entre 73% e 79,17%
2	CE, MA, RR, PA, RN	Nível de similaridade é de 79,06%
3	ES, MS, MT, RJ, SP, RO, TO, RS, SC	Nível de similaridade é de 81,68%

Seguindo o procedimento descrito nos itens anteriores no MiniTab, encontramos o seguinte resultado:

Discriminant Analysis: test versus IFGF; Receita Própria; ...

After subtracting group means,
IFGF is highly correlated with other predictors.

After subtracting group means,
Liquidez is highly correlated with other predictors.

Linear Method for Response: test

Predictors: IFGF; Receita Própria; Pessoal; Investimentos; Liquidez;
Custo da Dívida

Group	1	2	3
Count	2820	725	2019

5564 cases used, 1 cases contain missing values

Summary of classification

Put into Group	True Group		
	1	2	3
1	1132	200	294
2	958	437	269
3	730	88	1456
Total N	2820	725	2019
N correct	1132	437	1456
Proportion	0,401	0,603	0,721

N = 5564 N Correct = 3025 Proportion Correct = 0,544

A tarefa da análise discriminante é encontrar a melhor função discriminante linear de um conjunto de variáveis que reproduza, tanto quanto possível, um agrupamento a priori de casos considerados.

Um procedimento em passos é utilizado nesse programa, e em cada passo a variável mais poderosa é introduzida na função discriminante. A função critério para selecionar a próxima variável depende do número de grupos especificados (o número de grupos varia de 2 a 20).

Quando o número de variáveis é maior do que dois, então o critério de seleção de variáveis é o traço do produto da matriz de covariância para as variáveis envolvidas e a matriz de covariância interclasse em um passo particular.

Os cálculos podem ser realizados em toda a população ou em amostra de dados ou mesmo em dados previamente agrupados.

Nos nossos exemplos com as variáveis de gestão fiscal, utiliza-se a análise discriminante linear e conseguimos um resultado de 54,4% de proporção correta.

6. REGRESSÃO LOGÍSTICA

Stat >> Regression >> Nominal Logistical Regression

Nominal Logistic Regression: Região versus IFGF; Receita Própria; ...

Response Information

Variable	Value	Count	
Região	Sul	1191	(Reference Event)
	Sudeste	1669	
	Norte	447	
	Nordeste	1790	
	Centro-Oeste	468	
	Total	5565	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio
Logit 1: (Sudeste/Sul)					
Constant	0,704277	0,290471	2,42	0,015	
IFGF	-6,33147	7,05955	-0,90	0,370	0,00
Receita Própria	2,61715	1,58175	1,65	0,098	13,70
Pessoal	-2,25628	1,63428	-1,38	0,167	0,10
Investimentos	1,14480	1,57164	0,73	0,466	3,14
Liquidez	-0,200741	1,56644	-0,13	0,898	0,82
Custo da Dívida	4,29920	0,788078	5,46	0,000	73,64
Logit 2: (Norte/Sul)					
Constant	0,126758	0,453969	0,28	0,780	
IFGF	21,0740	8,03236	2,62	0,009	1,42009E+09
Receita Própria	-7,86312	1,81497	-4,33	0,000	0,00
Pessoal	-9,48776	1,90045	-4,99	0,000	0,00
Investimentos	-5,06017	1,76797	-2,86	0,004	0,01
Liquidez	-6,04933	1,76679	-3,42	0,001	0,00
Custo da Dívida	2,32168	0,975614	2,38	0,017	10,19
Logit 3: (Nordeste/Sul)					
Constant	5,84708	0,310244	18,85	0,000	
IFGF	-10,8124	6,81598	-1,59	0,113	0,00
Receita Própria	-4,43755	1,54540	-2,87	0,004	0,01
Pessoal	-3,17906	1,58954	-2,00	0,046	0,04
Investimentos	0,583120	1,51336	0,39	0,700	1,79
Liquidez	0,122850	1,50748	0,08	0,935	1,13
Custo da Dívida	3,31384	0,782636	4,23	0,000	27,49
Logit 4: (Centro-Oeste/Sul)					
Constant	-1,44389	0,448093	-3,22	0,001	
IFGF	2,05099	8,82476	0,23	0,816	7,78
Receita Própria	0,309447	1,97061	0,16	0,875	1,36
Pessoal	-4,46234	2,06304	-2,16	0,031	0,01
Investimentos	-2,01398	1,96640	-1,02	0,306	0,13
Liquidez	-1,24474	1,94541	-0,64	0,522	0,29
Custo da Dívida	5,16057	1,03021	5,01	0,000	174,26

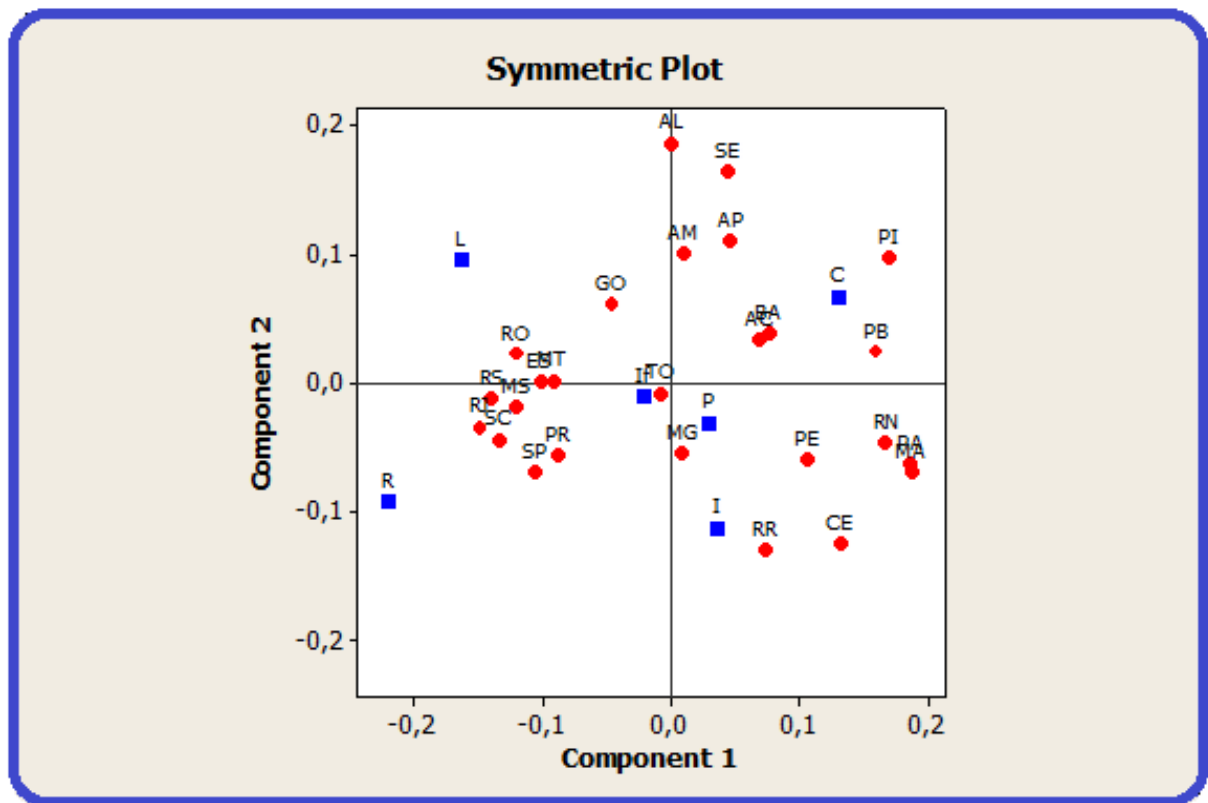
95% CI		
Predictor	Lower	Upper
Logit 1: (Sudeste/Sul)		
Constant		
IFGF	0,00	1817,08
Receita Própria	0,62	304,09
Pessoal	0,00	2,58
Investimentos	0,14	68,39
Liquidez	0,04	17,63
Custo da Dívida	15,71	345,09
Logit 2: (Norte/Sul)		
Constant		
IFGF	206,61	9,76062E+15
Receita Própria	0,00	0,01
Pessoal	0,00	0,00
Investimentos	0,00	0,20
Liquidez	0,00	0,08
Custo da Dívida	1,51	68,98
Logit 3: (Nordeste/Sul)		
Constant		
IFGF	0,00	12,76
Receita Própria	0,00	0,24
Pessoal	0,00	0,94
Investimentos	0,09	34,79
Liquidez	0,06	21,70
Custo da Dívida	5,93	127,46
Logit 4: (Centro-Oeste/Sul)		
Constant		
IFGF	0,00	2,52569E+08
Receita Própria	0,03	64,83
Pessoal	0,00	0,66
Investimentos	0,00	6,30
Liquidez	0,01	13,04
Custo da Dívida	23,14	1312,57
Log-Likelihood = -6845,829		
Test that all slopes are zero: G = 2632,996, DF = 24, P-Value = 0,000		

Enquanto método de predição para variáveis categóricas, a regressão logística é comparável às técnicas supervisionadas propostas em aprendizagem automática (árvores de decisão, redes neurais, etc.), ou ainda a análise discriminante preditiva em estatística exploratória. É possível de colocá-la em concorrência para escolha do modelo mais adaptado para certo problema preditivo a resolver.

7. ANÁLISE DE CORRESPONDÊNCIA DAS VARIÁVEIS DE GESTÃO FISCAL

Nesta análise serão trabalhados os estados e as médias dos indicadores de gestão fiscal por unidades federativas (-DF). Na análise de correspondência será gerado um mapa contendo quais estados estão mais próximos e quais variáveis tem a ver entre si. O comando para gerar o gráfico é:

STAT >> MULTIVARIATE >> SIMPLE CORRESPONDENCE ANALISYS



Nesta análise trabalhamos com as variáveis IFGF (If), Receita Própria (R), Despesas com Pessoal (P), Investimentos (I), Liquidez (L), e Custo da Dívida (C), representadas pelos quadrados azuis. Os círculos vermelhos são as médias das variáveis de gestão fiscal por estado, e podem ser identificados pela sigla. Observa-se Tocantins, e Minas Gerais levemente afastada, ocupando uma posição central no gráfico, e a esquerda, em torno do eixo "X", um agrupamento de estados (RO, MT, MS, GO, RS, SC, PR, ES, SP, RJ), predominantemente das regiões Centro-oeste, Sul e Sudeste, exceção feita a Rondônia, os quais ficam equidistantes de três indicadores (If, L e R), e no lado direito do gráfico, muito mais disperso, os demais estados, sendo estes da região Norte e Nordeste, próximos dos outros três indicadores (I, P e C).

Simple Correspondence Analysis: IFGF_M1; Rec-M1; Pes-M1; Invest-M1; Liq -M1; Cu

Contingency Table

	If	R	P	I	L	C	Total
AC	48	17	52	55	49	88	309
AL	43	15	45	35	57	83	278
AM	54	12	58	58	69	88	339
AP	47	18	45	48	54	93	305
BA	43	15	59	41	41	76	275
CE	47	16	57	68	34	78	300
ES	60	27	58	70	75	83	373
GO	52	26	57	48	60	87	330
MA	48	11	56	69	36	88	308
MG	53	24	58	65	51	83	334
MS	61	33	61	68	73	84	380
MT	60	29	62	65	71	85	372
PA	43	15	52	57	29	85	281
PB	36	10	37	47	33	74	237
PE	39	22	34	52	29	80	256
PI	44	8	56	46	43	90	287
PR	54	26	67	58	58	68	331
RJ	61	37	63	64	71	80	376
RN	42	13	52	55	32	80	274
RO	61	22	57	74	84	79	377
RR	53	16	62	76	44	75	326
RS	66	30	69	74	84	81	404
SC	65	31	66	77	79	79	397
SE	43	15	43	40	54	89	284
SP	60	38	63	65	61	83	370
TO	59	22	68	68	64	89	370
Total	1342	548	1457	1543	1435	2148	8473

Analysis of Contingency Table

Axis	Inertia	Proportion	Cumulative	Histogram
1	0,0124	0,5325	0,5325	

2	0,0057	0,2456	0,7780	*****
3	0,0034	0,1448	0,9228	*****
4	0,0018	0,0770	0,9998	****
5	0,0000	0,0002	1,0000	
Total	0,0234			

Simple Correspondence Analysis: IFGF_M1; Rec-M1; Pes-M1; Invest-M1; Liq -M1; Cu

Row Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	AC	0,965	0,036	0,010	0,070	0,780	0,014	0,034	0,186	0,007
2	AL	0,983	0,033	0,050	0,001	0,000	0,000	0,186	0,983	0,198
3	AM	0,490	0,040	0,036	0,010	0,005	0,000	0,101	0,485	0,071
4	AP	0,775	0,036	0,028	0,046	0,116	0,006	0,110	0,659	0,075
5	BA	0,333	0,032	0,031	0,077	0,265	0,016	0,039	0,068	0,009
6	CE	0,983	0,035	0,051	0,132	0,521	0,050	-0,124	0,462	0,096
7	ES	0,822	0,044	0,023	-0,101	0,822	0,036	0,001	0,000	0,000
8	GO	0,503	0,039	0,020	-0,046	0,179	0,007	0,062	0,324	0,026
9	MA	0,924	0,036	0,068	0,189	0,817	0,104	-0,068	0,107	0,030
10	MG	0,865	0,039	0,006	0,009	0,024	0,000	-0,054	0,841	0,020
11	MS	0,933	0,045	0,030	-0,119	0,912	0,051	-0,018	0,021	0,003
12	MT	0,989	0,044	0,016	-0,091	0,989	0,029	0,001	0,000	0,000
13	PA	0,960	0,033	0,058	0,188	0,867	0,094	-0,062	0,093	0,022
14	PB	0,869	0,028	0,036	0,159	0,848	0,057	0,025	0,020	0,003
15	PE	0,294	0,030	0,066	0,107	0,226	0,028	-0,059	0,068	0,018
16	PI	0,896	0,034	0,063	0,171	0,677	0,080	0,097	0,219	0,056
17	PR	0,585	0,039	0,030	-0,087	0,418	0,024	-0,055	0,167	0,021
18	RJ	0,846	0,044	0,052	-0,148	0,804	0,078	-0,034	0,042	0,009
19	RN	0,988	0,032	0,042	0,168	0,921	0,073	-0,045	0,067	0,012
20	RO	0,511	0,044	0,055	-0,119	0,493	0,051	0,023	0,018	0,004
21	RR	0,760	0,038	0,048	0,075	0,191	0,017	-0,129	0,569	0,112
22	RS	0,899	0,048	0,044	-0,138	0,893	0,073	-0,012	0,007	0,001
23	SC	0,931	0,047	0,042	-0,132	0,840	0,066	-0,043	0,090	0,015
24	SE	0,939	0,034	0,044	0,045	0,065	0,005	0,165	0,874	0,158
25	SP	0,603	0,044	0,049	-0,105	0,425	0,039	-0,068	0,178	0,035
26	TO	0,055	0,044	0,004	-0,007	0,022	0,000	-0,008	0,032	0,000

Column Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	If	0,841	0,158	0,004	-0,020	0,673	0,005	-0,010	0,168	0,003
2	R	0,653	0,065	0,239	-0,219	0,554	0,249	-0,092	0,099	0,096
3	P	0,185	0,172	0,074	0,030	0,092	0,013	-0,031	0,093	0,028
4	I	0,765	0,182	0,143	0,037	0,074	0,020	-0,113	0,690	0,402
5	L	0,930	0,169	0,281	-0,164	0,693	0,365	0,096	0,237	0,271
6	C	0,905	0,254	0,259	0,131	0,715	0,348	0,067	0,190	0,201

8. ÁRVORES DE CLASSIFICAÇÃO DAS VARIÁVEIS DE GESTÃO FISCAL

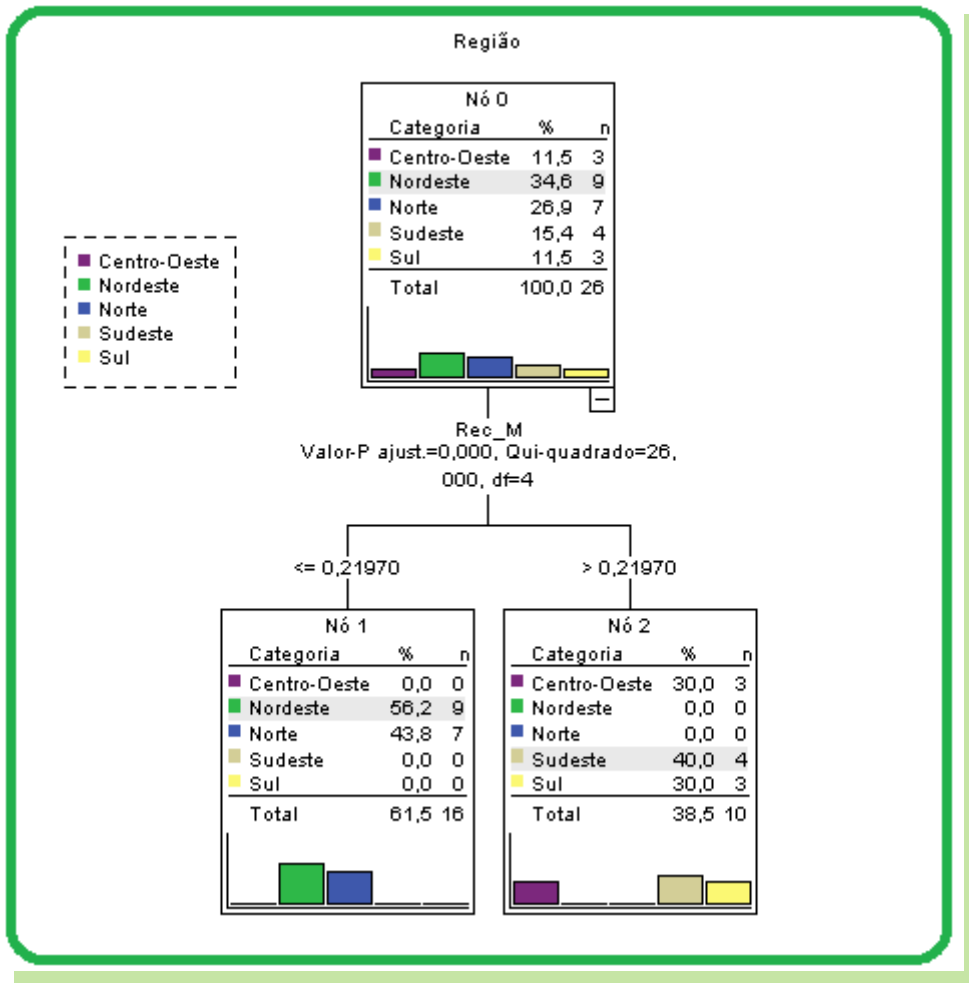
COMANDO SPSS:

ANALISAR >> CLASSIFICAR >> ARVORE

Este resultado se refere à variável dependente região e as variáveis independente: IFGF, Receita Própria, Despesas com Pessoal, Investimentos, Liquidez e Custo da Dívida.

Resumo do modelo

	Método de crescimento	CHAID	
	Variável dependente	Região	
	Variáveis independentes	IFGF_M, Rec_M, Pes_M, Invest_M, Liq_M, Custo_M	
Especificações	Validação	Nenhum	
	Profundidade de árvore máxima		3
	Casos mínimos em nó pai		2
	Casos mínimos em nó filho		1
	Variáveis independentes incluídas	Rec_M	
Resultados	Número de nós		3
	Número de nós de terminal		2
	Profundidade		1



Risco

Estimativas	Modelo padrão
,500	,098

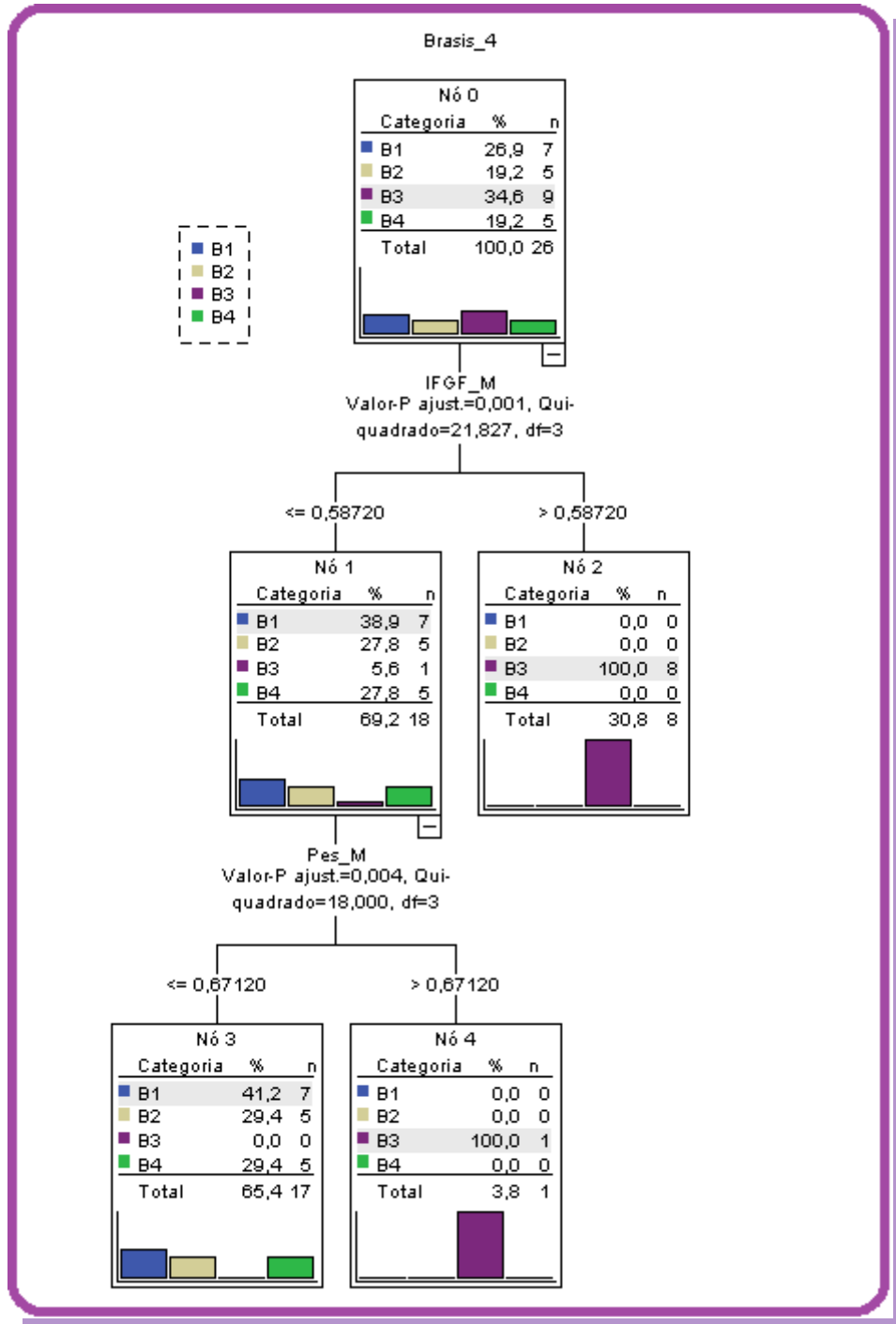
Método de crescimento: CHAID

Variável dependente: Região

Posto

Observado	Previsto					Porcentagem Correta
	Centro-Oeste	Nordeste	Norte	Sudeste	Sul	
Centro-Oeste	0	0	0	3	0	0,0%
Nordeste	0	9	0	0	0	100,0%
Norte	0	7	0	0	0	0,0%
Sudeste	0	0	0	4	0	100,0%
Sul	0	0	0	3	0	0,0%
Porcentagem global	0,0%	61,5%	0,0%	38,5%	0,0%	50,0%

Este resultado se refere à variável dependente Brasis_4 num e as variáveis independente: IFGF, Receita Própria, Despesas com Pessoal, Investimentos, Liquidez e Custo da Dívida.



Resumo do modelo

Especificações	Método de crescimento	CHAID		
	Variável dependente	Brasis_4		
	Variáveis independentes	IFGF_M, Rec_M, Pes_M, Invest_M, Liq_M, Custo_M		
	Validação	Nenhum		
	Profundidade de árvore máxima		3	
	Casos mínimos em nó pai		2	
	Casos mínimos em nó filho		1	
	Variáveis independentes incluídas	IFGF_M, Pes_M		
	Resultados	Número de nós		5
		Número de nós de terminal		3
Profundidade			2	

Risco

Estimativas	Modelo padrão
,385	,095

Método de crescimento: CHAID

Variável dependente: Brasis_4

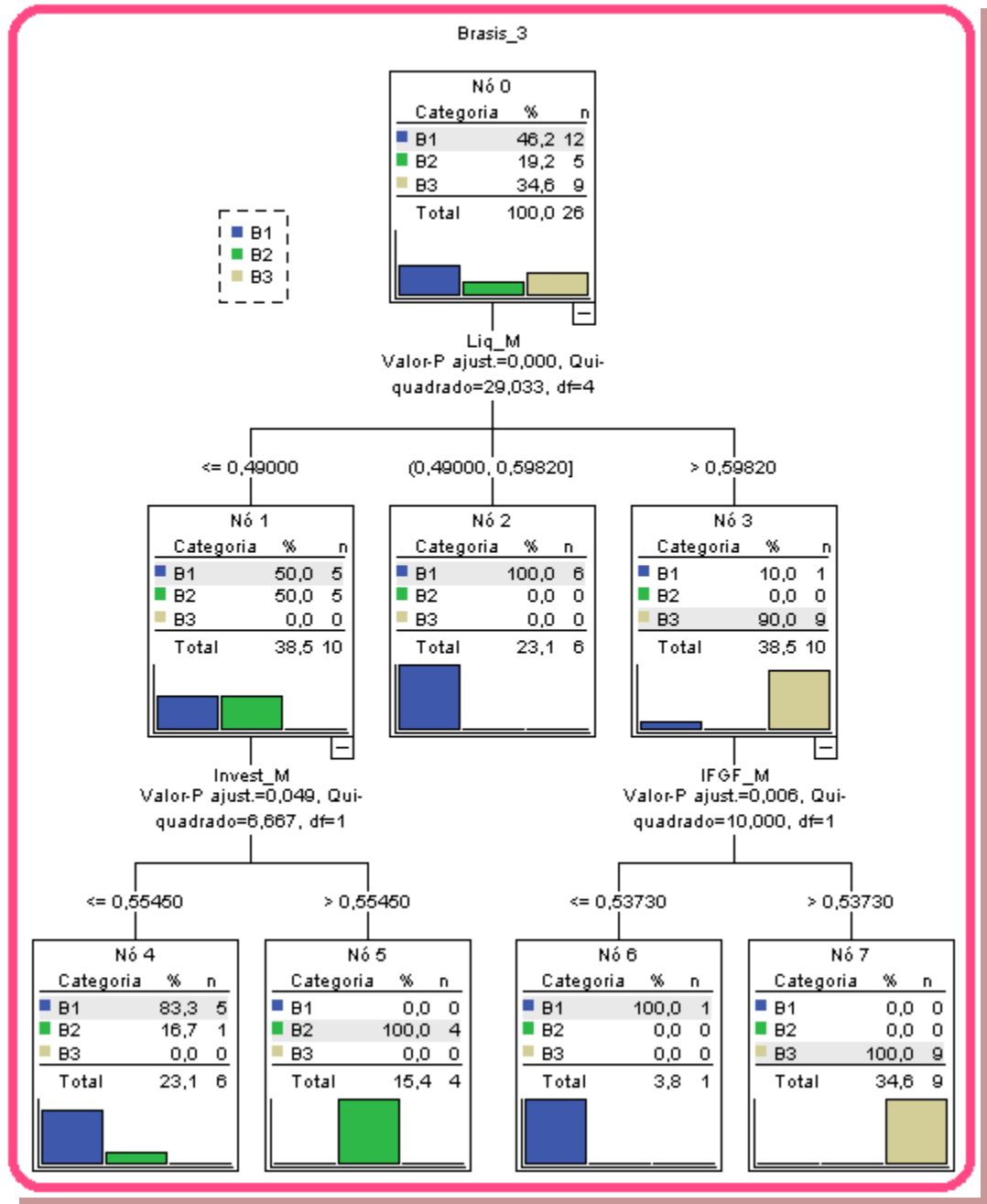
Posto

Observado	Previsto				Porcentagem Correta
	B1	B2	B3	B4	
B1	7	0	0	0	100,0%
B2	5	0	0	0	0,0%
B3	0	0	9	0	100,0%
B4	5	0	0	0	0,0%
Porcentagem global	65,4%	0,0%	34,6%	0,0%	61,5%

Método de crescimento: CHAID

Variável dependente: Brasis_4

Este resultado se refere à variável dependente Brasis_3 num e as variáveis independente: IFGF, Receita Própria, Despesas com Pessoal, Investimentos, Liquidez e Custo da Dívida.



Resumo do modelo

Especificações	Método de crescimento	CHAID		
	Variável dependente	Brasis_3		
	Variáveis independentes	IFGF_M, Rec_M, Pes_M, Invest_M, Liq_M, Custo_M		
	Validação	Nenhum		
	Profundidade de árvore máxima		3	
	Casos mínimos em nó pai		2	
	Casos mínimos em nó filho		1	
	Variáveis independentes incluídas	Liq_M, Invest_M, IFGF_M		
	Resultados	Número de nós		8
		Número de nós de terminal		5
Profundidade			2	

Risco

Estimativas	Modelo padrão
,038	,038

Método de crescimento: CHAID

Variável dependente: Brasis_3

Posto

Observado	Previsto			Porcentagem Correta
	B1	B2	B3	
B1	12	0	0	100,0%
B2	1	4	0	80,0%
B3	0	0	9	100,0%
Porcentagem global	50,0%	15,4%	34,6%	96,2%

Método de crescimento: CHAID

Variável dependente: Brasis_3

Entre as três variáveis categóricas apresentadas, observou-se um índice de previsibilidade na dos 3 Brasis (Brasis_3), que alcançou 96,2% de acerto, contra 61,5% do 4Brasis (Brasis_4) e 50% das Regiões (Região).

Referência Bibliográfica

INFONET. Disponível em: <http://www.infonet.com.br/economia/ler.asp?id=125792> . Acesso em: 05/03/13.

FIRJAN. Disponível em:
http://www.firjan.org.br/IFGF/download/anexo_metodologico_IFGF.pdf . Acesso em:
05/03/13

IBGE. Disponível em: <http://www.ngb.ibge.gov.br/Default.aspx?pagina=divisao> . Acesso em
14/04/13