



PUC - SP

**UM ESTUDO DA EVOLUÇÃO HISTÓRICA DOS
INDICADORES DAS 500 EMPRESAS MAIS VALIOSAS DO
MUNDO E DA POSIÇÃO DOS PAÍSES DO BRICS NO
RANKING**

**EDUARDO NOBORU KAWABUCHI
MESTRADO EM ADMINISTRAÇÃO DE EMPRESAS
MÉTODOS QUANTITATIVOS
PROF. DR. ARNOLDO HOYOS**

**SÃO PAULO
2011**

Conteúdo

1. INTRODUÇÃO	4
2. ENTENDENDO OS DADOS	4
2.1 Os Indivíduos	4
2.2 As Variáveis	4
2.3 A Tabela de Dados	5
2.4 Fonte de Dados	5
3. ANÁLISE DAS VARIÁVEIS	6
3.1 Variável Market Value \$m (Valor de Mercado)	6
3.2 Variável Turnover \$m	7
3.3 Variável Price \$ (Preço da Ação)	8
3.4 Variável PE ratio	10
3.5 Variável Dividend yield (%)	11
4. EVOLUÇÃO DOS PAÍSES DO BRICS NO RANKING DAS 500 EMPRESAS MAIS VALIOSAS DO MUNDO	12
5. ESTUDOS COMPLEMENTARES	15
5.1 Análise de Tendências	15
5.1.1 Países BRICS	16
5.1.2 Outros Países	18
5.1.3 Resultado	20
5.2 Estatística Descritiva	21
5.2.1 Variável Market Value \$m	21
5.2.2 Variável Turnover \$m	22
5.2.3 Variável Net Income \$m	22
5.2.4 Variável Total Assets \$m	23
5.2.5 Variável Employees	24
5.2.6 Variável Price \$	25
5.2.7 Variável PE ratio	26
5.2.8 Variável Dividend Yield (%)	27
5.3 Análise de Relação entre Variáveis	28
5.4 Regressões Múltiplas	30
5.5 Amostragem	32
5.5.1 Variável Market Value \$m	32
5.5.2 Variável Turnover Value \$m	36
5.5.3 Variável Turnover Value \$m	39
5.5.4 Variável Total Assets \$m	43
5.5.5 Variável Employees	46
5.5.6 Variável Price \$	49
5.5.7 Variável PE ratio	52
5.5.8 Variável Dividend Yield (%)	56
5.5.9 Análise de Correlação e Dendogramas	59
5.5.10 Considerações	61
5.6 Análise em Componentes Principais	62
5.7 Análise de Conglomerados	64
5.7.1 Estatística Descritiva / Pesquisa por Amostragem	64
5.7.2 Análise de Conglomerados	65
5.8 Análise Discriminante	69
5.8.1 Comparação de Média, Análise de Variância e Intervalo de Confiança	69
5.8.2 Análise Discriminante	72
5.9 Regressão Logística	75

5.10 Árvores de Classificação	78
5.11 Análise de Correspondências.....	80
6. CONSIDERAÇÕES FINAIS	81
6.1 Evolução Histórica dos Indicadores das 500 Empresas mais valiosas do mundo e da Posição dos Países do BRICS no Ranking	81
6.2 Estudos Complementares	82
6.2.1 Análise de Tendências	82
6.2.2 Relação entre Variáveis	83
6.2.3 Regressões Múltiplas	83
6.2.4 Amostragem	83
6.2.5 Análise em Componentes Principais.....	84
6.2.6 Análise de Conglomerados.....	84
6.2.7 Análise Discriminante	84
6.2.8 Regressão Logística	84
6.2.9 Árvores de Classificação	85
6.2.10 Análise de Correspondências.....	85

1. INTRODUÇÃO

Esse trabalho tem por objetivo efetuar uma análise comparativa das médias e intervalos de confiança, variância da evolução histórica de dados do Ranking das 500 Empresas mais valiosas do mundo publicada pela revista Financial Times e também analisar a evolução da posição dos países do BRICS nesse ranking. O software estatístico utilizado é o **MINITAB**.

2. ENTENDENDO OS DADOS

Os dados são referentes ao ano de 2005 e 2010. As variáveis são indicadores financeiros e relacionados de cada empresa.

Já o BRICS é um acrônimo que se refere aos países membros fundadores (Brasil, Rússia, Índia e China) e à África do Sul. A idéia dos BRICS foi formulada pelo economista-chefe da Goldman Sachs, Jim O'Neil, em estudo de 2001, intitulado "Building Better Global Economic BRICs". Fixou-se como categoria da análise nos meios econômico-financeiros, empresariais, acadêmicos e de comunicação. Em 2006, o conceito deu origem a um agrupamento, propriamente dito, incorporado à política externa de Brasil, Rússia, Índia e China. Em 2011, por ocasião da III Cúpula, a África do Sul passou a fazer parte do agrupamento, que adotou a sigla BRICS.

2.1 Os Indivíduos

Os indivíduos desta análise são as empresas reportadas no FT Global 500 2005 e 2010. São 500 empresas, classificadas em função do índice valor de mercado. Os dados analisados de cada empresa são as variáveis que descrevemos a seguir. Os dados foram coletados no banco de dados do Financial Times.

2.2 As Variáveis

São as variáveis desta pesquisa, incluindo o nome das empresas. As mesmas são melhor explicadas na Tabela 1.

Tabela 1. As Variáveis

Variável	Significado	Tipo	Unidade de Medida
Global Rank 2010	Ranking Global 2010	Variável Quantitativa	Número
Global Rank 2009	Ranking Global 2009	Variável Quantitativa	Número
Company	Nome da Empresa	Variável Categórica	N/A
Country	País de Origem	Variável Categórica	N/A
Continent	Continente	Variável Categórica	N/A
Sector	Setor de Atividade	Variável Categórica	N/A

Market value \$m	Valor de Mercado	Variável Quantitativa	\$m
Turnover \$m	Indicador fundamentalista que mede o grau de liquidez das ações de uma empresa no mercado. Expressa a relação entre a média de volume diário negociado no último mês e o free float da empresa, ambos medidos em números de ações. Free float é definido como a porcentagem do capital numa empresa que não se encontra em mãos de acionistas estratégicos (com participação superior a 5% do capital total da empresa). Assim a fórmula de cálculo do indicador é: Turnover = Volume Médio Diário (30 dias)/ Free float (em número de ações)	Variável Quantitativa	\$m
Net Income \$m	Lucro Líquido	Variável Quantitativa	\$m
Total Assets \$m	Ativo Total	Variável Quantitativa	\$m
Employees	Número de Empregados	Variável Quantitativa	Un
Price \$	Preço da Ação	Variável Quantitativa	\$
PE ratio	O Price - Earnings Ratio ou PER é um indicador de análise bolsista das empresas. É o coeficiente de capitalização do benefício. O Price - Earnings Ratio ilustra o relacionamento, por ação, entre o valor bolsista da empresa e os seus resultados líquidos. Cálculo: Price – Earnings Ratio = Cotação da ação / Resultado Líquido por Ação Na teoria, o PER representa o número de anos em que recuperaria o seu investimento sob a forma de dividendos.	Variável Quantitativa	Indicador
Dividend yield (%)	Rendimento dos Dividendos (%). É um índice criado para medir a rentabilidade dos dividendos de uma empresa em relação ao preço de suas ações. Este índice traz o benefício de poder comparar a rentabilidade dos dividendos entre empresas.	Variável Quantitativa	%
Year end	Fim do Ano Fiscal	Variável Quantitativa	Data

2.3 A Tabela de Dados

Vide em anexo arquivo em formato Excel.

2.4 Fonte de Dados

Todos os dados desta pesquisa foram obtidos em:

3. ANÁLISE DAS VARIÁVEIS

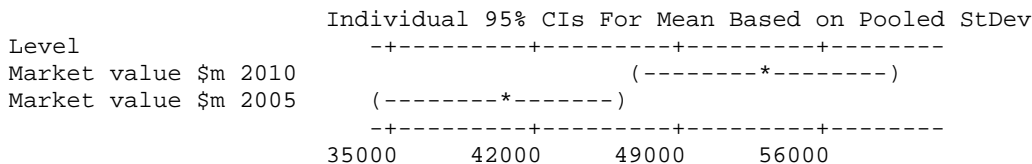
3.1 Variável Market Value \$m (Valor de Mercado)

One-way ANOVA: Market value \$m 2010; Market value \$m 2005

Source	DF	SS	MS	F	P
Factor	1	21412072746	21412072746	8.69	0.003
Error	510	1.25710E+12	2464910481		
Total	511	1.27852E+12			

S = 49648 R-Sq = 1.67% R-Sq(adj) = 1.48%

Level	N	Mean	StDev
Market value \$m 2010	256	53335	50816
Market value \$m 2005	256	40401	48452



Pooled StDev = 49648

O p-value obtido é de 0,003, abaixo de 0,05, indicando que, com intervalo de confiança de 95%, as médias populacionais dos anos 2010 e 2005 da variável são estatisticamente diferentes. Podemos observar que a média populacional de 2010 (\$m 40.401) é superior à 2005 (53.555) em 32,56%, demonstrando uma evolução.

Segue abaixo a análise do Two-Sample T-Test, que confirma o resultado obtido anteriormente:

Two-Sample T-Test and CI: Market value \$m 2010; Market value \$m 2005

Two-sample T for Market value \$m 2010 vs Market value \$m 2005

	N	Mean	StDev	SE Mean
Market value \$m 2010	256	53335	50816	3176
Market value \$m 2005	256	40401	48452	3028

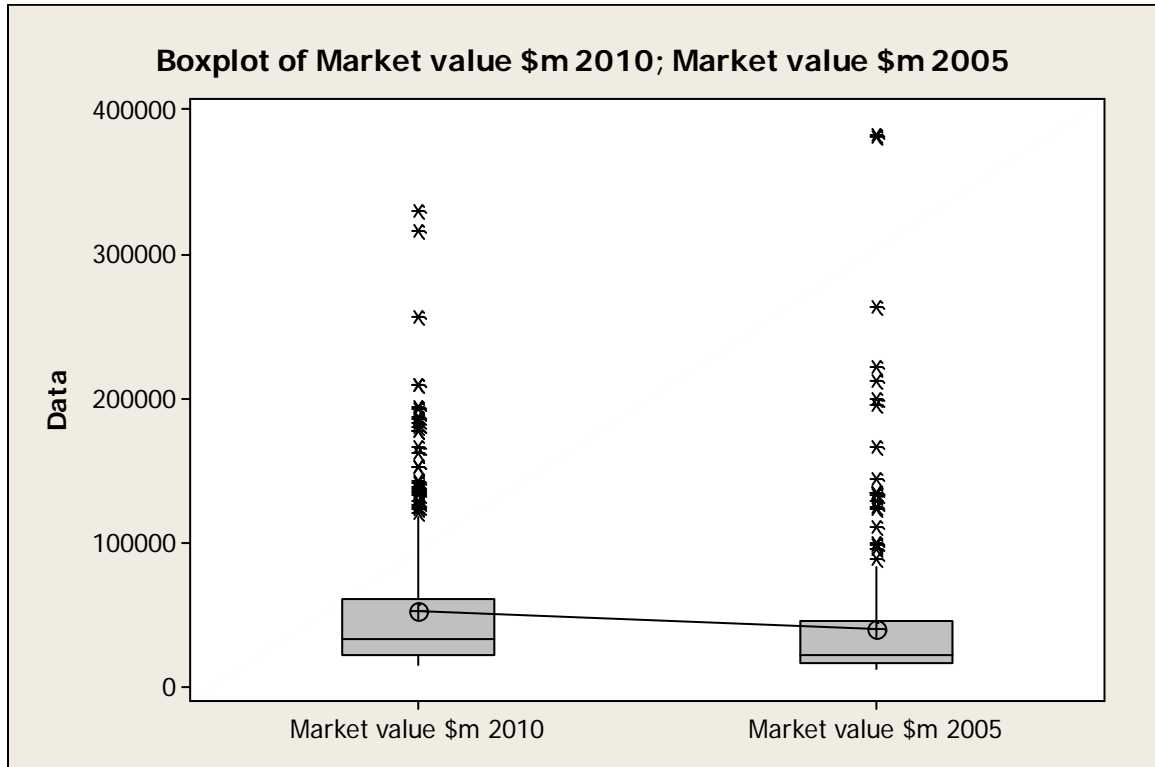
Difference = mu (Market value \$m 2010) - mu (Market value \$m 2005)

Estimate for difference: 12934

95% CI for difference: (4312; 21555)

T-Test of difference = 0 (vs not =): T-Value = 2.95 P-Value = 0.003 DF = 508

Abaixo o box-plot corroborando a as análises efetuadas acima.



3.2 Variável Turnover \$m

One-way ANOVA: Turnover \$m 2010; Turnover \$m 2005

Source	DF	SS	MS	F	P
Factor	1	12578452996	12578452996	5.90	0.015
Error	509	1.08487E+12	2131380423		
Total	510	1.09745E+12			

S = 46167 R-Sq = 1.15% R-Sq(adj) = 0.95%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
Turnover \$m 2010	255	40419	49562	(-----+-----+-----+-----+)
Turnover \$m 2005	256	30496	42517	(-----*-----)

30000 36000 42000 48000

Pooled StDev = 46167

O p-value obtido é de 0,015, abaixo de 0,05, indicando que, com intervalo de confiança de 95%, as médias populacionais dos anos 2010 e 2005 da variável são estatisticamente diferentes. Podemos observar que a média populacional de 2010 (\$m 30.496) é superior à 2005 (\$m 40.419) em 32.54%, demonstrando uma evolução.

Segue abaixo a análise do Two-Sample T-Test, que confirma o resultado obtido anteriormente:

Two-Sample T-Test and CI: Turnover \$m 2010; Turnover \$m 2005

Two-sample T for Turnover \$m 2010 vs Turnover \$m 2005

	N	Mean	StDev	SE Mean
Turnover \$m 2010	255	40419	49562	3104
Turnover \$m 2005	256	30496	42517	2657

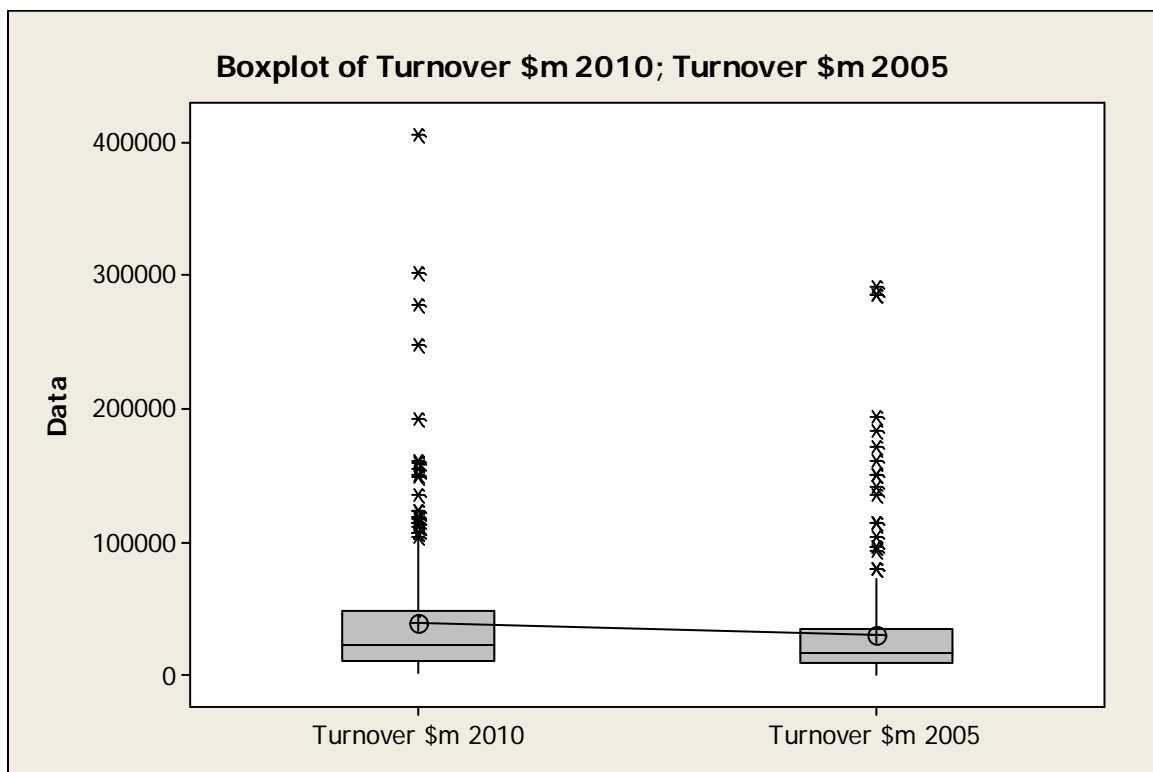
Difference = mu (Turnover \$m 2010) - mu (Turnover \$m 2005)

Estimate for difference: 9923

95% CI for difference: (1895; 17950)

T-Test of difference = 0 (vs not =): T-Value = 2.43 P-Value = 0.016 DF = 496

Abaixo o box-plot corroborando a as análises efetuadas acima.



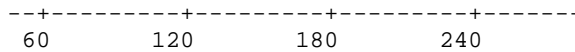
3.3 Variável Price \$ (Preço da Ação)

One-way ANOVA: Price \$ 2010; Price \$ 2005

Source	DF	SS	MS	F	P
Factor	1	80817	80817	0.11	0.742
Error	510	379150411	743432		
Total	511	379231228			

S = 862.2 R-Sq = 0.02% R-Sq(adj) = 0.00%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
Price \$ 2010	256	153.9	769.4	(-----*-----)
Price \$ 2005	256	179.0	946.0	(-----*-----)



Pooled StDev = 862.2

O p-value obtido é de 0,742, bastante superior à 0,05, indicando que, com intervalo de confiança de 95%, as médias populacionais dos anos 2010 e 2005 da variável não são estatisticamente diferentes.

Segue abaixo a análise do Two-Sample T-Test, que confirma o resultado obtido anteriormente:

Two-Sample T-Test and CI: Price \$ 2010; Price \$ 2005

Two-sample T for Price \$ 2010 vs Price \$ 2005

	N	Mean	StDev	SE Mean
Price \$ 2010	256	154	769	48
Price \$ 2005	256	179	946	59

Difference = mu (Price \$ 2010) - mu (Price \$ 2005)

Estimate for difference: -25.1

95% CI for difference: (-174.9; 124.6)

T-Test of difference = 0 (vs not =): T-Value = -0.33 **P-Value = 0.742** DF = 489

Abaixo o box-plot corroborando a as análises efetuadas acima.



3.4 Variável PE ratio

One-way ANOVA: PE ratio 2010; PE ratio 2005

Source	DF	SS	MS	F	P
Factor	1	1	1	0.01	0.926
Error	510	62170	122		
Total	511	62171			

S = 11.04 R-Sq = 0.00% R-Sq(adj) = 0.00%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
PE ratio 2010	256	21.15	11.73	(-----*-----)
PE ratio 2005	256	21.06	10.30	(-----*-----)

20.00 20.80 21.60 22.40

Pooled StDev = 11.04

O p-value obtido é de 0,926, bastante superior à 0,05, indicando que, com intervalo de confiança de 95%, as médias populacionais dos anos 2010 e 2005 da variável não são estatisticamente diferentes.

Segue abaixo a análise do Two-Sample T-Test, que confirma o resultado obtido anteriormente:

Two-Sample T-Test and CI: PE ratio 2010; PE ratio 2005

Two-sample T for PE ratio 2010 vs PE ratio 2005

	N	Mean	StDev	SE Mean
PE ratio 2010	256	21.2	11.7	0.73
PE ratio 2005	256	21.1	10.3	0.64

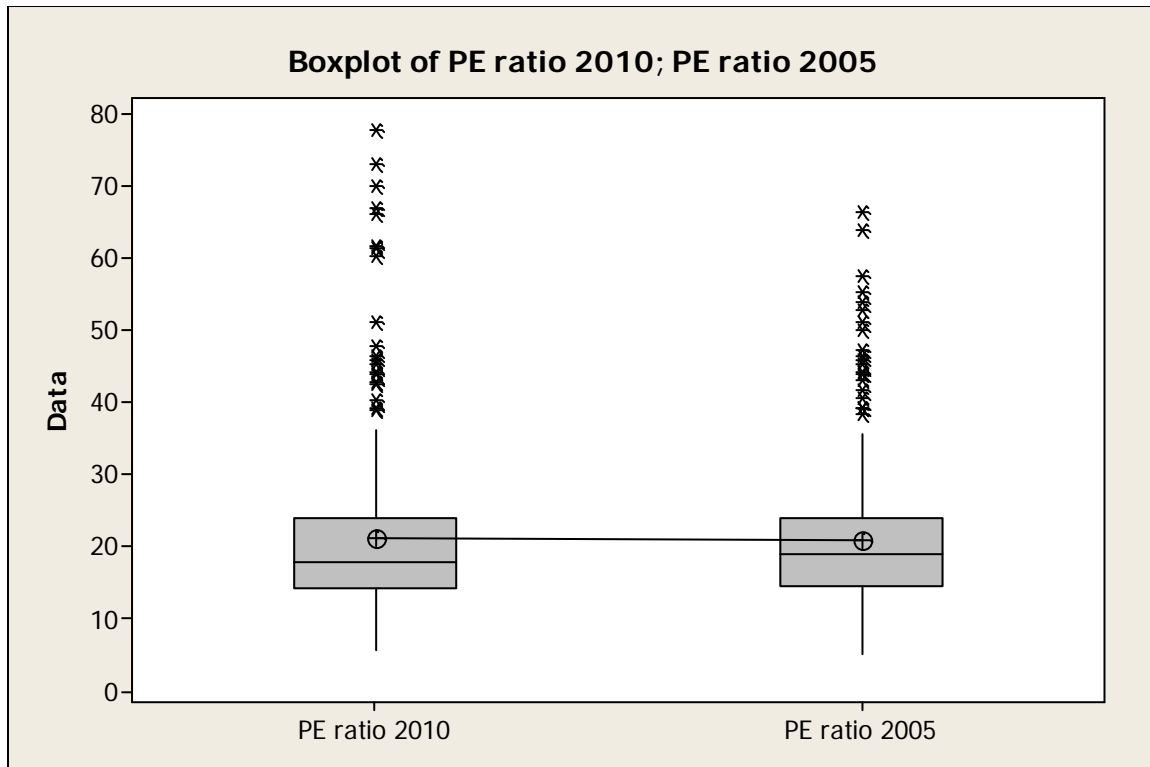
Difference = mu (PE ratio 2010) - mu (PE ratio 2005)

Estimate for difference: 0.090

95% CI for difference: (-1.827; 2.008)

T-Test of difference = 0 (vs not =): T-Value = 0.09 P-Value = 0.926 DF = 501

Abaixo o box-plot corroborando a as análises efetuadas acima.



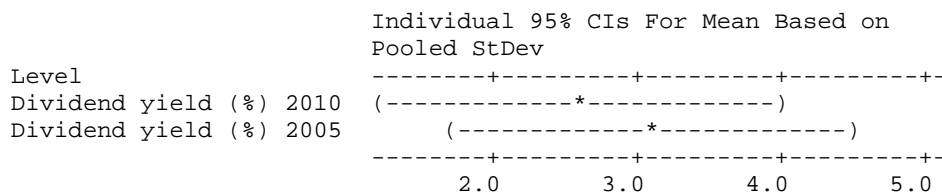
3.5 Variável Dividend yield (%)

One-way ANOVA: Dividend yield (%) 2010; Dividend yield (%) 2005

Source	DF	SS	MS	F	P
Factor	1	30	30	0.23	0.635
Error	510	67504	132		
Total	511	67534			

S = 11.50 R-Sq = 0.04% R-Sq(adj) = 0.00%

Level	N	Mean	StDev
Dividend yield (%) 2010	256	2.62	1.64
Dividend yield (%) 2005	256	3.11	16.19



Pooled StDev = 11.50

O p-value obtido é de 0,635, bastante superior à 0,05, indicando que, com intervalo de confiança de 95%, as médias populacionais dos anos 2010 e 2005 da variável não são estatisticamente diferentes.

Segue abaixo a análise do Two-Sample T-Test, que confirma o resultado obtido anteriormente:

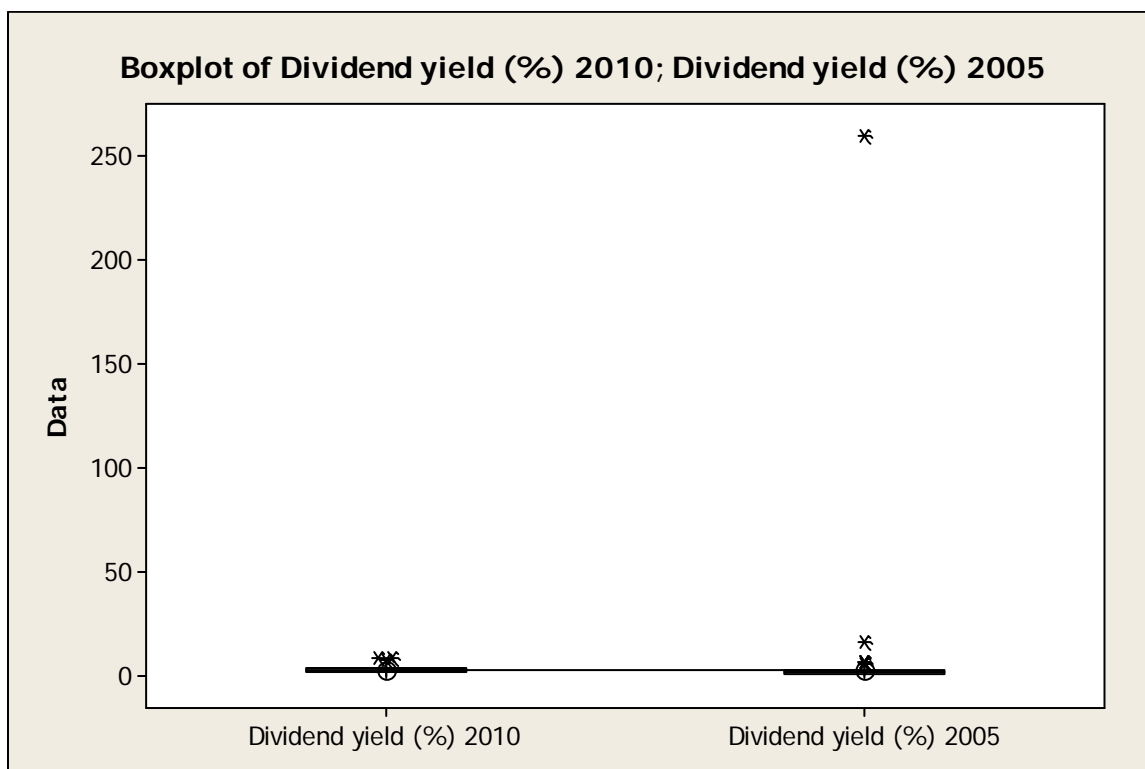
Two-Sample T-Test and CI: Dividend yield (%) 2010; Dividend yield (%) 2005

Two-sample T for Dividend yield (%) 2010 vs Dividend yield (%) 2005

	N	Mean	StDev	SE Mean
Dividend yield (%) 2010	256	2.62	1.64	0.10
Dividend yield (%) 2005	256	3.1	16.2	1.0

Difference = μ (Dividend yield (%) 2010) - μ (Dividend yield (%) 2005)
 Estimate for difference: -0.48
 95% CI for difference: (-2.49; 1.52)
 T-Test of difference = 0 (vs not =): T-Value = -0.48 **P-Value = 0.635** DF = 260

Abaixo o box-plot corroborando a as análises efetuadas acima.



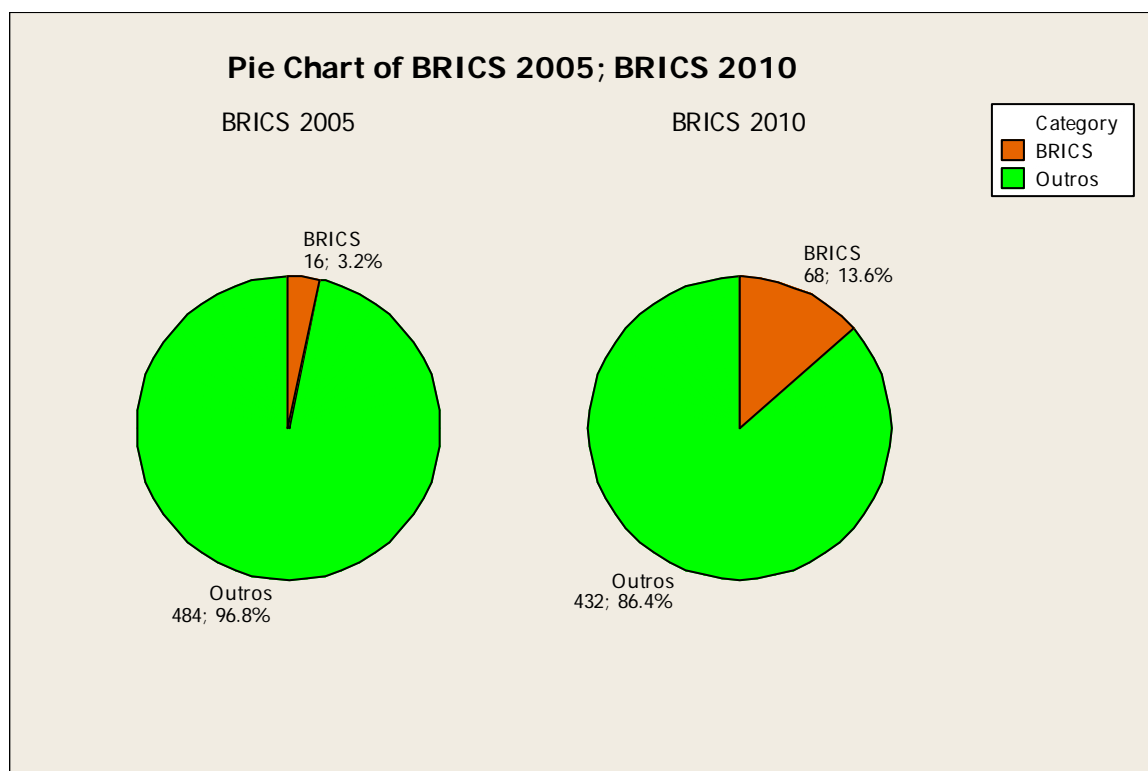
4. EVOLUÇÃO DOS PAÍSES DO BRICS NO RANKING DAS 500 EMPRESAS MAIS VALIOSAS DO MUNDO

Tabela 2. Evolução dos Países do BRICS no Ranking das 500 Empresas mais valiosas do mundo

País	2005			2010			Variação 2010/2005
	Qtde	% Total	% BRICS	Qtde	% Total	% BRICS	
Brasil	5	1.00%	31.25%	12	2.40%	17.65%	140.0%

Rússia	4	0.80%	25.00%	11	2.20%	16.18%	175.0%
Índia	5	1.00%	31.25%	16	3.20%	23.53%	220.0%
China	0	0.00%	0.00%	23	4.60%	33.82%	-
África do Sul	2	0.40%	12.50%	6	1.20%	8.82%	200.0%
Total BRICS	16	3.20%	-	68	13.60%	-	325.0%
Outros	484	96.80%	-	432	86.40%	-	-10.7%
Total	500	100.00%	-	500	100.00%	-	-

Podemos observar que houve uma evolução muito significativa dos países do BRICS no ranking das 500 empresas mais valiosas do mundo. Houve uma variação positiva de 325%, sendo que em 2005 haviam 16 empresas do BRICS no ranking, que representavam 3,2% e em 2010 há 68 empresas do BRICS entre as 500 empresas mais valiosas do mundo, que representam 13,6%.



Analisando isoladamente os países do BRICS, o país que teve maior evolução no ranking de 2005 para 2010 foi a China, que em 2005 não tinha nenhuma empresa no ranking e em 2010 tem 23 empresas. Sendo também a China o país do BRICS que possui mais empresas no ranking, representando 33,82%, seguido por Índia (16 empresas, 23,53%), Brasil (12 empresas, 17,65%), Rússia (11 empresas, 16,18%) e África do Sul (6 empresas, 8,82%).

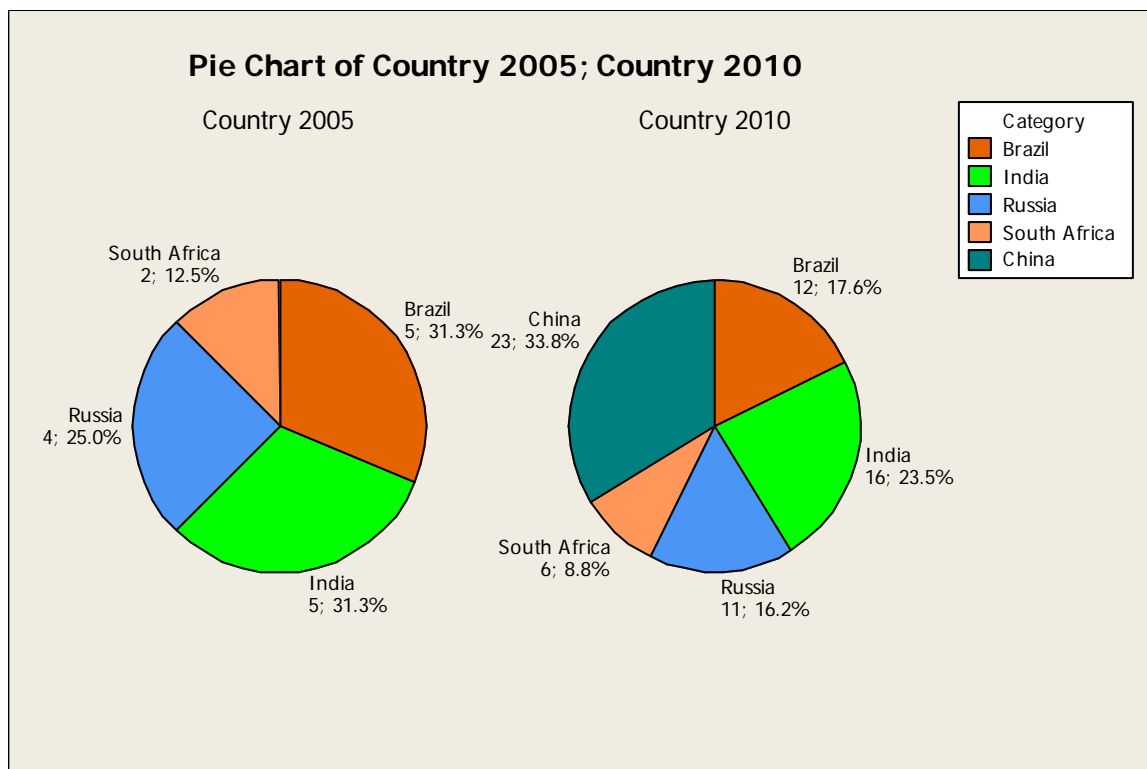


Tabela 3. Evolução do Market value \$m dos Países do BRICS no Ranking das 500 Empresas mais valiosas do mundo

País	2005			2010			Variação 2010/2005
	Market value \$m	% Total	% BRICS	Market value \$m	% Total	% BRICS	
Brasil	131,571.50	0.68%	33.61%	764,494.90	3.25%	20.93%	481.0%
Rússia	140,865.70	0.73%	35.98%	512,896.70	2.18%	14.04%	264.1%
Índia	89,731.30	0.46%	22.92%	485,434.70	2.07%	13.29%	441.0%
China	0.00	0.00%	0.00%	1,751,832.20	7.45%	47.95%	-
África do Sul	29,311.10	0.15%	7.49%	138,650.80	0.59%	3.80%	373.0%
Total BRICS	391,479.60	2.02%	-	3,653,309.30	15.54%	-	833.2%
Outros	18,948,604.10	97.98%	-	19,849,822.40	84.46%	-	4.8%
Total	19,340,083.70	100.00%	-	23,503,131.70	100.00%	-	-

Se analisarmos a evolução em relação ao Market value, podemos observar uma evolução mais significativa ainda dos países do BRICS no ranking das 500 empresas mais valiosas do mundo. Houve uma variação positiva de 833,20%, sendo que em 2005 representavam 2,02% e em 2010 representam 15,54%. A variação da média populacional foi 32,56%.

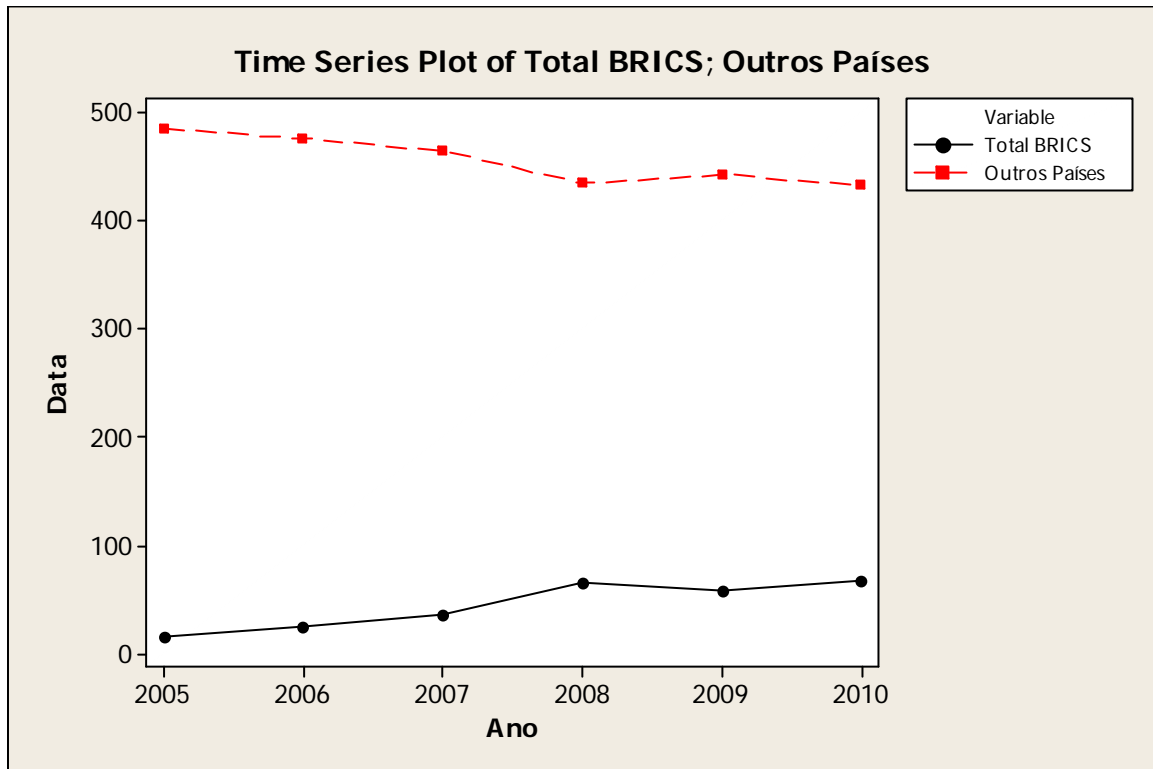
Outro dado muito interessante é que, em 2005 não havia nenhuma empresa do BRICS entre as 25 empresas mais valiosas do mundo e em 2010 existem 5 empresas entre as 25 mais valiosas do mundo, sendo que a empresa mais valiosa do mundo (PetroChina) é uma empresa da China, país pertencente ao BRICS. A China também possui a 4ª e a 11ª mais valiosa do mundo. O Brasil possui a 13ª (Petrobras) e a 22ª (Vale) empresa mais valiosa do mundo.

Global Rank 2005	Company	Country	Market value \$m	Global Rank 2010	Company	Country	Market value \$m
1	General Electric	US	382,233.10	1	PetroChina	China	329,259.70
2	ExxonMobil	US	380,567.20	2	Exxon Mobil	US	316,230.80
3	Microsoft	US	262,974.90	3	Microsoft	US	256,864.70
4	Citigroup	US	234,436.70	4	Industrial & Commercial Bank of China	China	246,419.80
5	BP	UK	221,365.30	5	Apple	US	213,096.70
6	Wal-Mart Stores	US	212,209.0	6	BHP Billiton	Australia/UK	209,935.10
7	Royal Dutch/Shell	Netherlands/UK	210,630.40	7	Wal-Mart Stores	US	209,000.70
8	Johnson & Johnson	US	199,711.40	8	Berkshire Hathaway	US	200,620.50
9	Pfizer	US	195,944.60	9	General Electric	US	194,246.20
10	Bank of America	US	178,765.40	10	China Mobile	Hong Kong	192,998.60
11	HSBC	UK	176,858.40	11	China Construction Bank	China	191,517.0
12	Vodafone	UK	171,416.90	12	Nestle	Switzerland	187,255.0
13	IBM	US	165,787.10	13	Petrobras	Brazil	186,239.50
14	Total	France	148,957.10	14	Procter & Gamble	US	183,773.40
15	Intel Corporation	US	144,695.40	15	Johnson & Johnson	US	179,708.10
16	American International Group	US	144,319.20	16	Bank of America	US	179,073.90
17	Altria	US	135,246.10	17	JP Morgan Chase	US	177,792.10
18	Toyota Motor	Japan	134,324.40	18	BP	UK	177,609.10
19	GlaxoSmithKline	UK	134,123.50	19	Royal Dutch Shell	UK	176,968.10
20	Berkshire Hathaway	US	134,067.20	20	HSBC	UK	176,573.50
21	Procter & Gamble	US	133,697.0	21	IBM	US	166,597.10
22	Saudi Basic Industries	Saudi Arabia	128,790.0	22	Vale	Brazil	162,829.90
23	Novartis	Switzerland	124,610.20	23	Wells Fargo & Co	US	161,454.50
24	ChevronTexaco	US	123,536.30	24	AT&T	US	152,509.60
25	JP Morgan Chase	US	123,261.10	25	Chevron	US	152,293.40

5. ESTUDOS COMPLEMENTARES

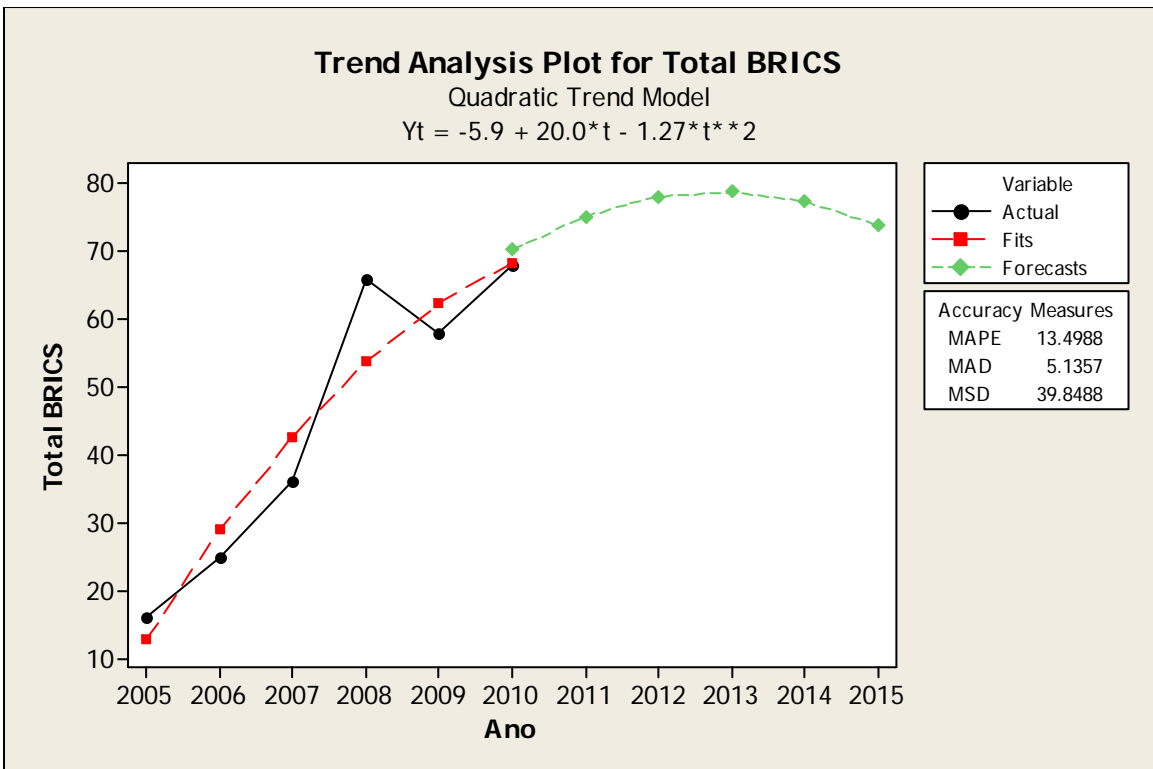
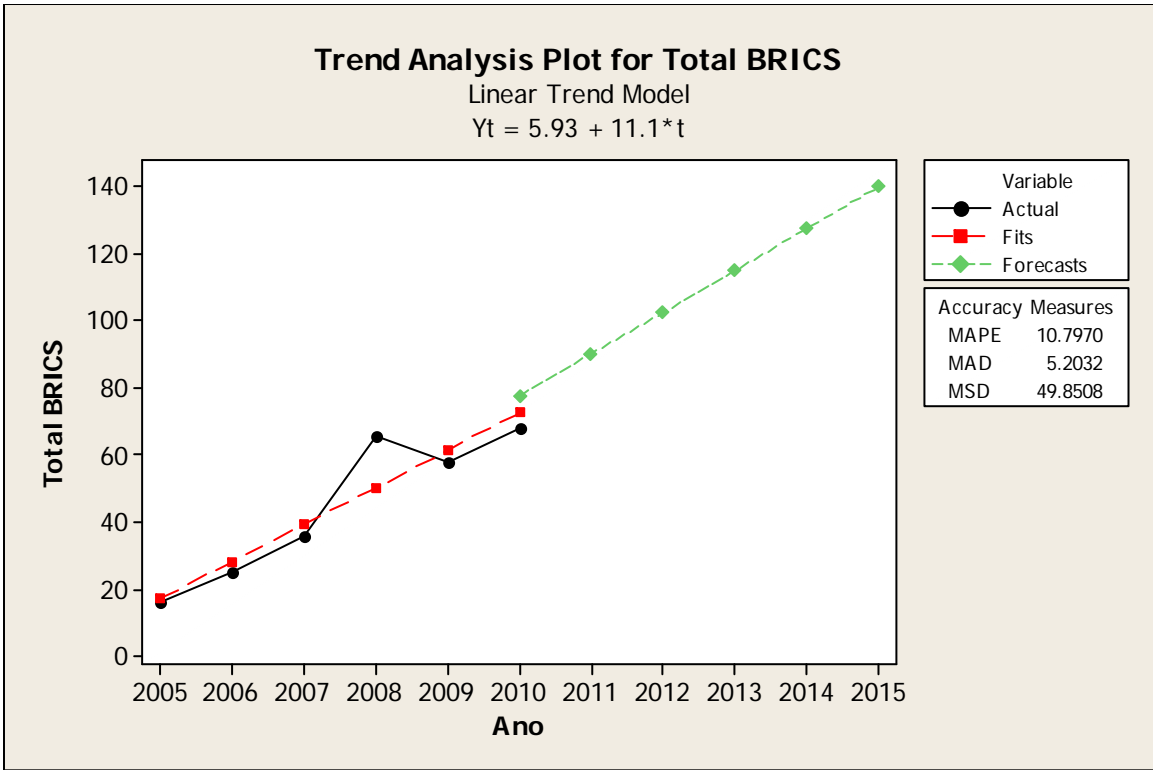
5.1 Análise de Tendências

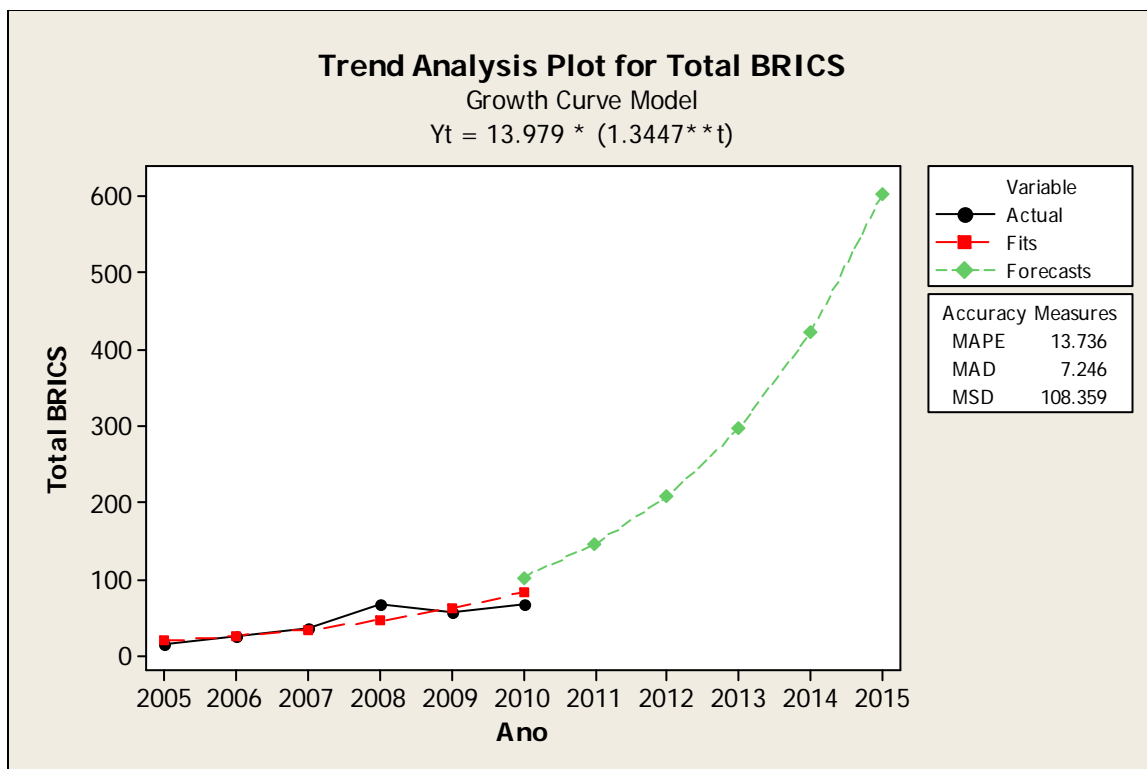
Efetuar uma análise de tendência e projeção da evolução das empresas dos países do BRICS no Ranking das 500 empresas mais valiosas do mundo.



Podemos observar pelo comportamento histórico das variáveis, que houve um crescimento muito grande na quantidade de empresas dos países do BRICS no ranking das 500 empresas mais valiosas do mundo de 2005 até 2008, em 2009 houve um leve queda e em 2010 o crescimento foi retomado. O mesmo comportamento, só que inverso, podemos observar na quantidade de empresas dos demais países do mundo. Vale observar que os dados divulgados pelo Financial Times tem como data base o ano anterior à divulgação, portanto essa variação no comportamento histórico ocorreu em função da crise financeira de 2008 e o comportamento de 2005 à 2008 deve ser mantido para os anos futuros.

5.1.1 Países BRICS



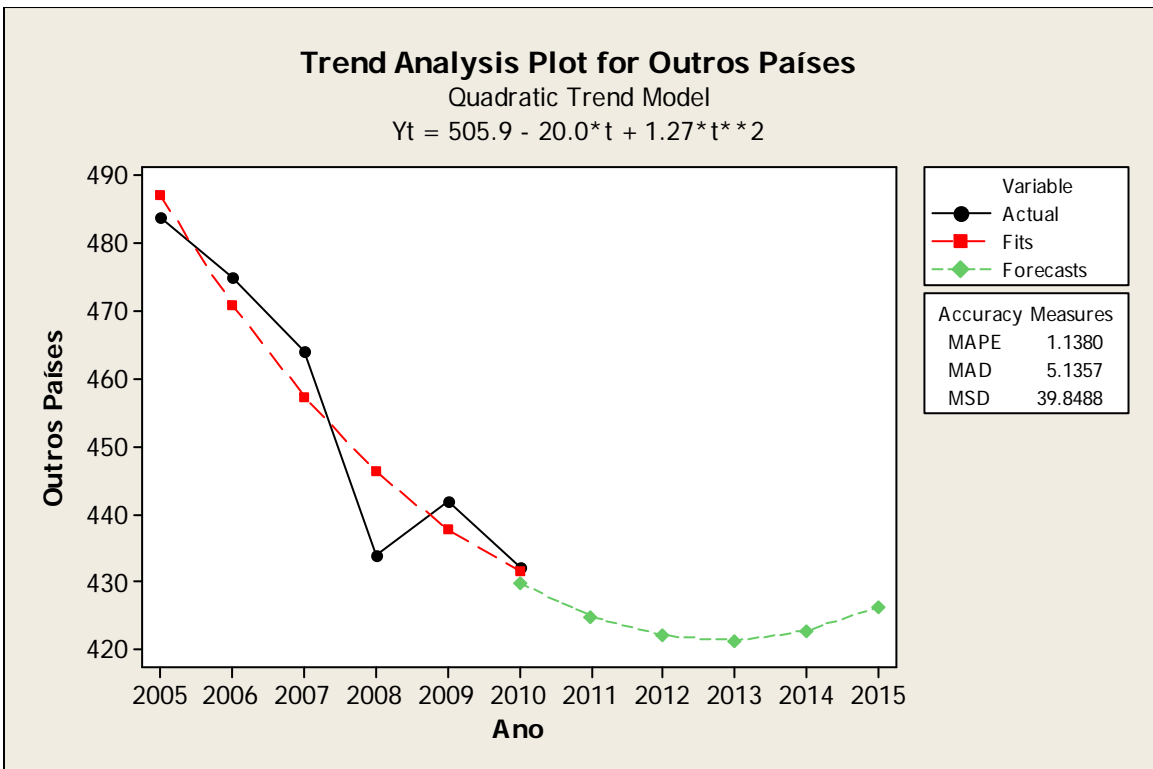
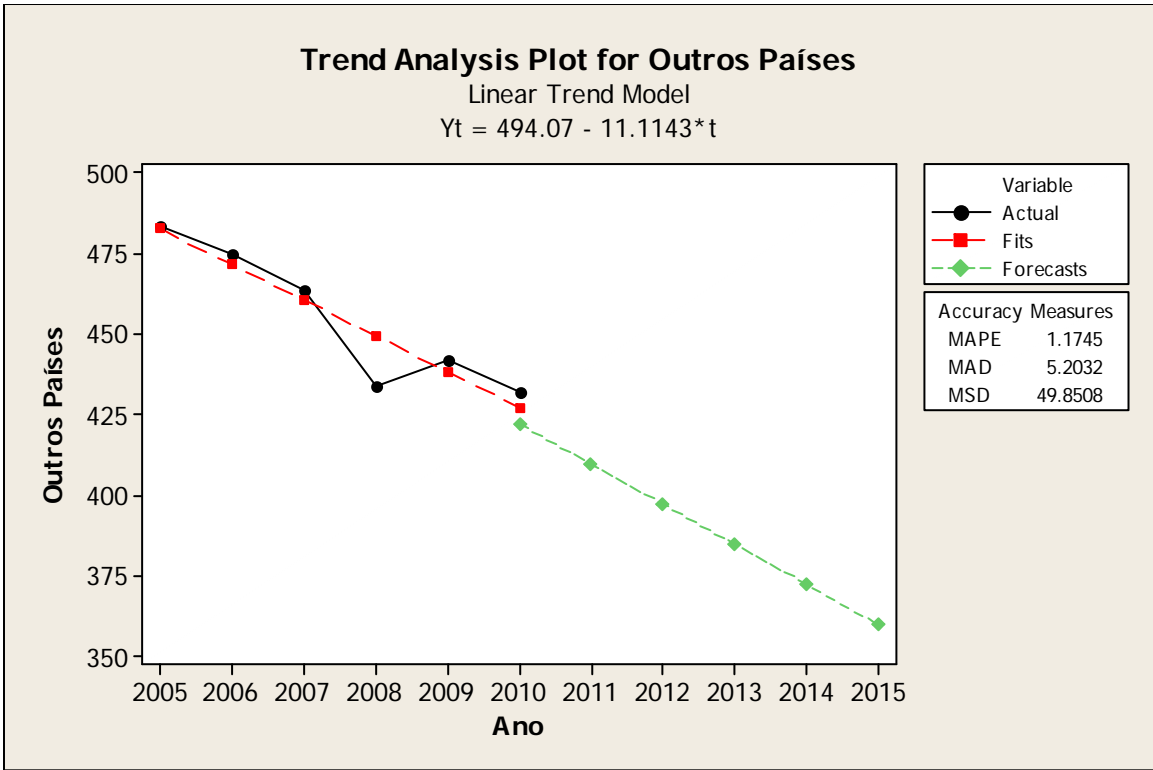


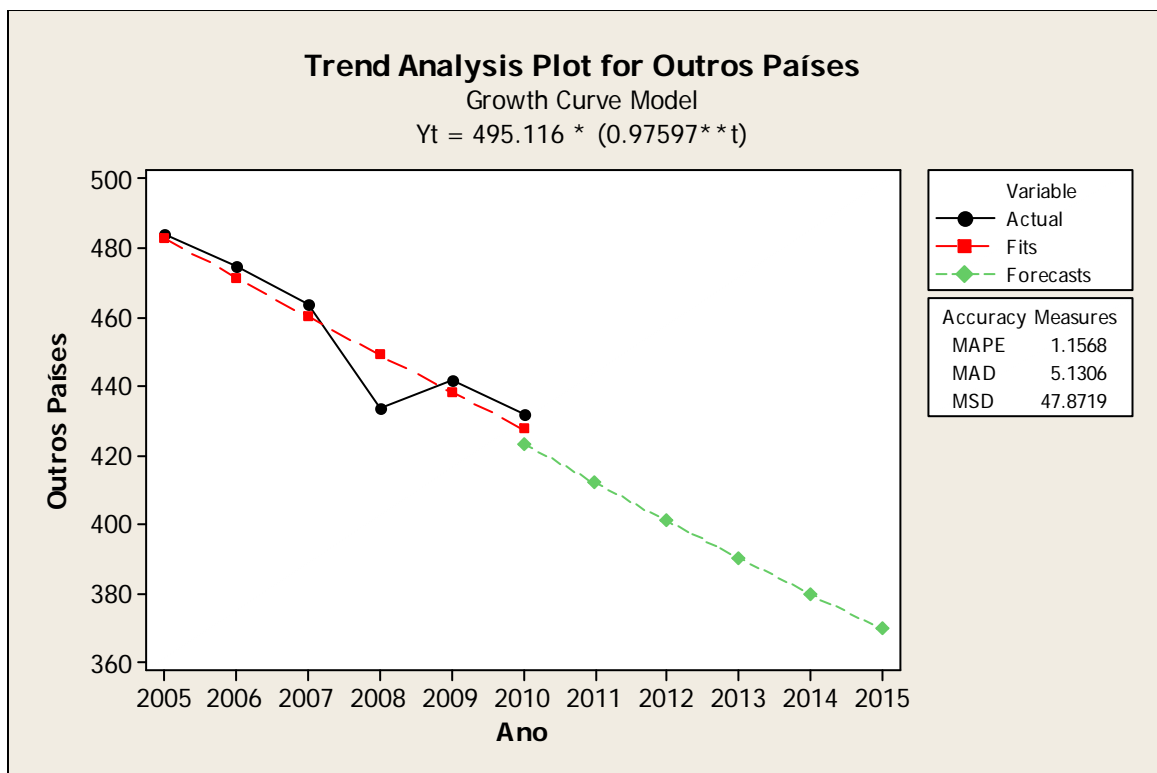
O cálculo da Função S-Curve não foi possível para esta série.

	Linear	Quadrática	Exponencial	S-Curve
MAPE	10,7970	13,4988	13,736	-
MAD	5,2032	5,1357	7,246	-
MSD	49,8508	39,8488	108,359	-

Com base nos resultados acima e considerando a variação do comportamento histórico em função da crise financeira de 2009, a função linear foi a que melhor se adaptou aos dados, pois a projeção da função quadrática sofre grande impacto em função da crise financeira.

5.1.2 Outros Países





O cálculo da Função S-Curve não foi possível para esta série.

	Linear	Quadrática	Exponencial	S-Curve
MAPE	1,1745	1,1380	1,1568	-
MAD	5,2032	5,1357	5,1306	-
MSD	49,8508	39,8488	47,8719	-

No caso da projeção dos demais países que não pertencem ao BRICS, também iremos considerar que a função linear foi a que melhor se adaptou aos dados, pois iremos utilizar o mesmo tipo de modelo para realizar as projeções. A projeção da função quadrática sofre grande impacto em função da crise financeira e a projeção da função exponencial não é aplicável para os países do BRICS.

5.1.3 Resultado

Analisando a planilha de resultados da análise de tendências abaixo, podemos observar que poderá haver um crescimento de 106% na quantidade de países do BRICS no ranking das 500 empresas mais valiosas do mundo, passando de 68 países em 2010 para 140 países em 2015. Já no caso dos outros países, poderá haver uma redução de 17%, passando de 432 países em 2010 para 360 países em 2015.

	Comportamento Histórico						Projeção				
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Quantidade											
Total BRICS	16	25	36	66	58	68	90	103	115	128	140
Outros	484	475	464	434	442	432	410	397	385	372	360

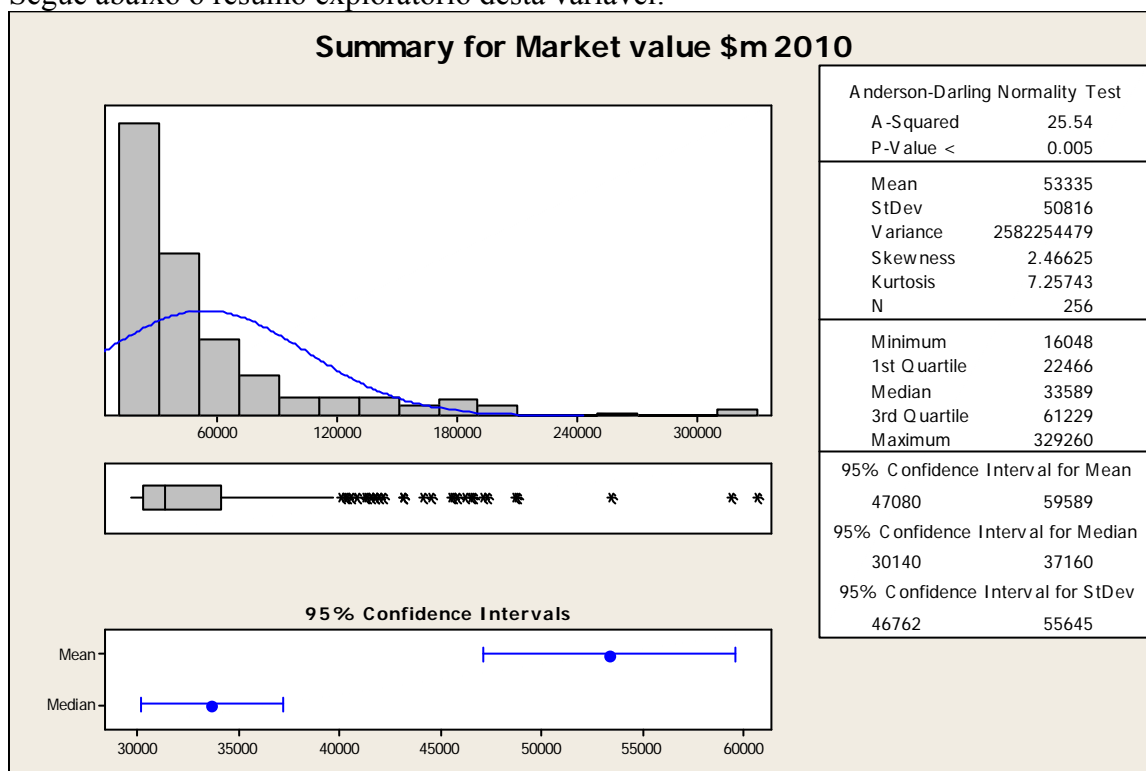
Total	500	500	500	500	500	500	500	500	500	500	500
Percentual											
Total BRICS	3%	5%	7%	13%	12%	14%	18%	21%	23%	26%	28%
Outros	97%	95%	93%	87%	88%	86%	82%	79%	77%	74%	72%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

5.2 Estatística Descritiva

Começamos com a análise das medidas e gráficos da estatística descritiva de cada uma das variáveis, a saber: Market Value \$m, Turnover \$m, Net Income \$m, Total Assets \$m, Employees, Price \$, PE ratio e Dividend Yield (%).

5.2.1 Variável Market Value \$m

Segue abaixo o resumo exploratório desta variável:



As principais observações que podemos fazer são:

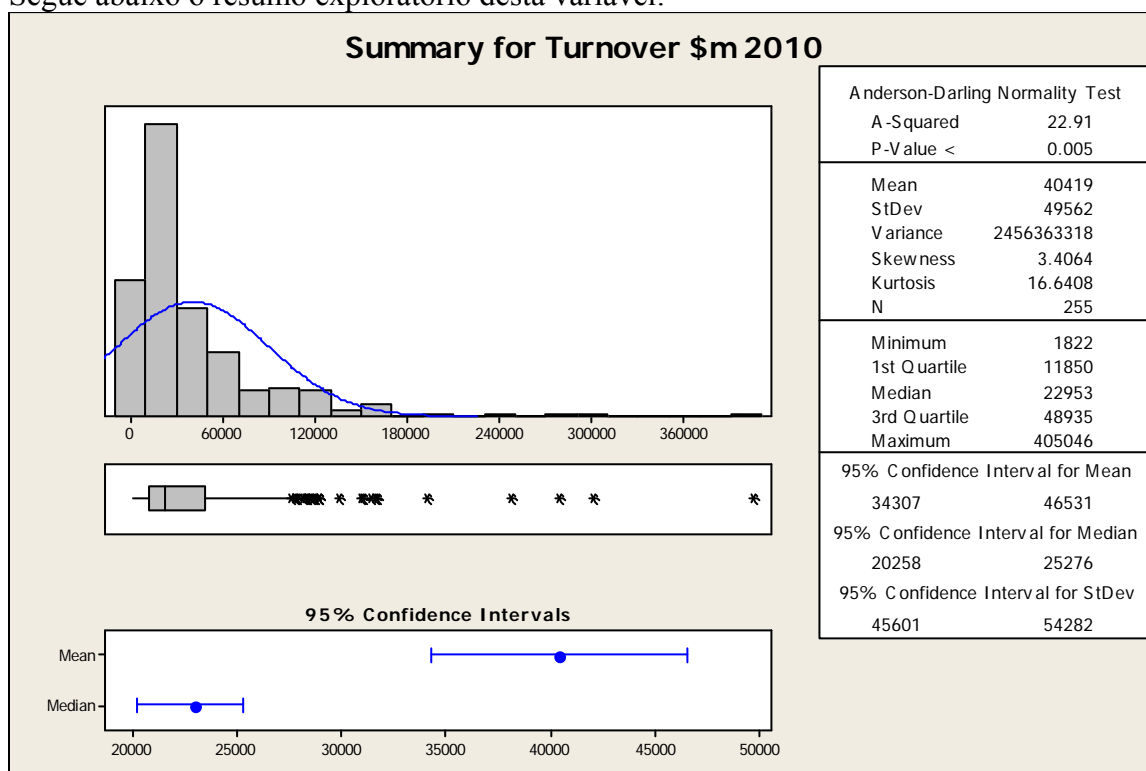
- Forma: o histograma permite verificar que trata-se de uma distribuição assimétrica para a direita, ou seja, a maior parte das empresas se concentra em uma faixa mais baixa que o ponto médio (172.654) entre o valor mínimo (16.048) e o valor máximo (329.260) registrado. O gráfico de caixa e a análise dos 5 números também nos deixa claro esta assimetria da distribuição. A posição da linha da mediana longe do centro do quadrado principal do gráfico de caixa nos confirma esta afirmativa.

- Valores Atípicos: Há valores atípicos no gráfico, com valores acima de 96.907,5 \$m. Os maiores valores são a PetroChina (1ª colocada) com 329.260 e a Exxon Mobil (2ª colocada) com 316.231.

- Centro e Dispersão: a mediana nos indica que aproximadamente metade das empresas possui o valor de mercado menor que 33.589 e metade maior que esse valor. A média do valor de mercado é de 53.335, e o desvio padrão é de 50.816 (alto). O valor mínimo foi 16.048 e o valor máximo foi 329.260. Com 95% de confiança, podemos afirmar que a média do valor de mercado das empresas encontra-se entre 47.080 e 59.589.

5.2.2 Variável Turnover \$m

Segue abaixo o resumo exploratório desta variável:



As principais observações que podemos fazer são:

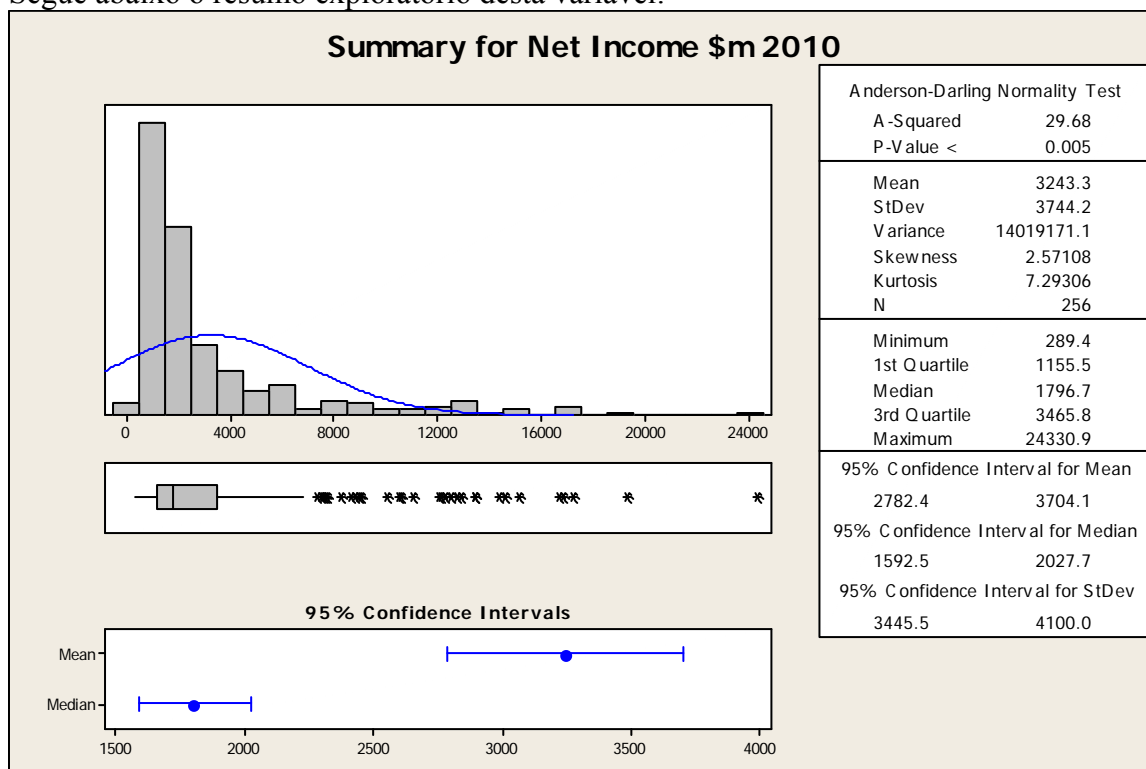
- Forma: o histograma permite verificar que trata-se de uma distribuição assimétrica para a direita, ou seja, a maior parte das empresas se concentra em uma faixa mais baixa que o ponto médio (203.434) entre o valor mínimo (1.822) e o valor máximo (405.046) registrado. O gráfico de caixa e a análise dos 5 números também nos deixa claro esta assimetria da distribuição. A posição da linha da mediana longe do centro do quadrado principal do gráfico de caixa nos confirma esta afirmativa.

- Valores Atípicos: Há valores atípicos no gráfico, com valores acima de 92.712,5 \$m. O maior valor é do Wal-Mart Stores (7ª colocada) com 405.046.

- Centro e Dispersão: a mediana nos indica que aproximadamente metade das empresas possui o turnover menor que 22.953 e metade maior que esse valor. A média do turnover é de 40.419, e o desvio padrão é de 49.562 (alto). O valor mínimo foi 1.822 e o valor máximo foi 405.046. Com 95% de confiança, podemos afirmar que a média do turnover das empresas encontra-se entre 34.307 e 46.531.

5.2.3 Variável Net Income \$m

Segue abaixo o resumo exploratório desta variável:



As principais observações que podemos fazer são:

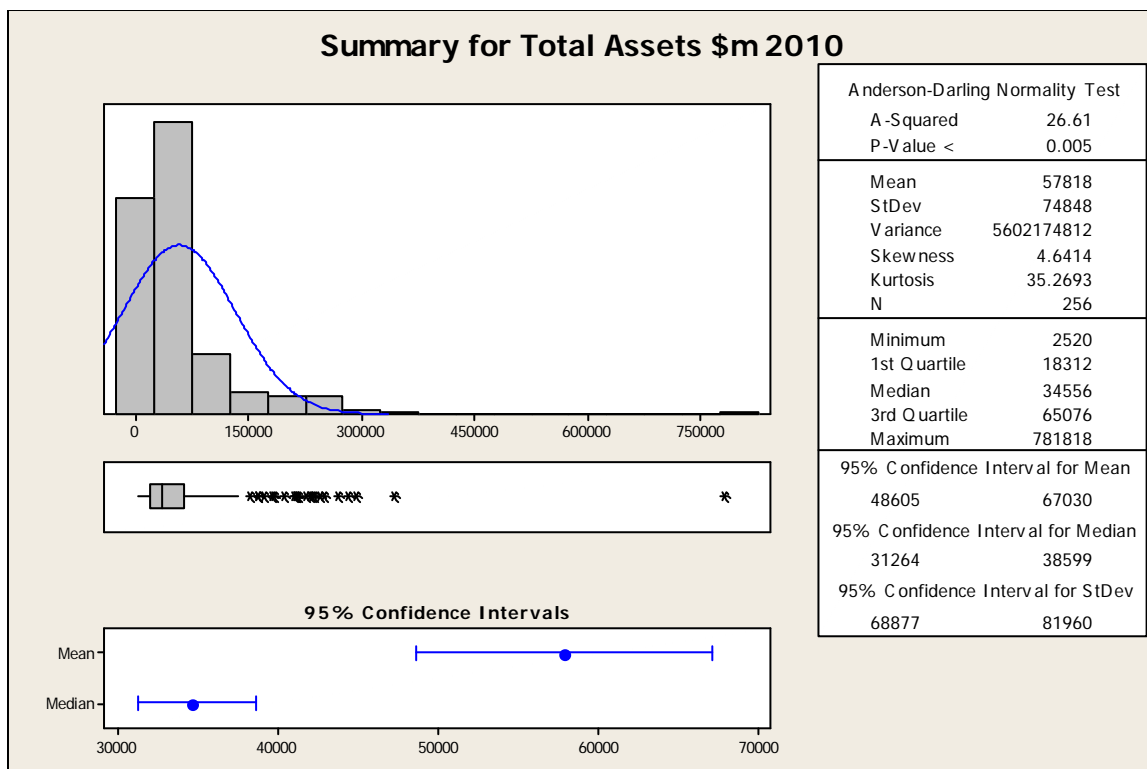
- Forma: o histograma permite verificar que trata-se de uma distribuição assimétrica para a direita, ou seja, a maior parte das empresas se concentra em uma faixa mais baixa que o ponto médio (12.310,15) entre o valor mínimo (289,4) e o valor máximo (24.330,9) registrado. O gráfico de caixa e a análise dos 5 números também nos deixa claro esta assimetria da distribuição. A posição da linha da mediana longe do centro do quadrado principal do gráfico de caixa nos confirma esta afirmativa.

- Valores Atípicos: Há valores atípicos no gráfico, com valores acima de 5.775,75 \$m. O maior valor é do Gazprom (33ª colocada) com 24.330,90.

- Centro e Dispersão: a mediana nos indica que aproximadamente metade das empresas possui o Net Income menor que 1.796,7 e metade maior que esse valor. A média do Net Income é de 3.243,3, e o desvio padrão é de 3.744,2 (alto). O valor mínimo foi 289,4 e o valor máximo foi 24.330,9. Com 95% de confiança, podemos afirmar que a média do Net Income das empresas encontra-se entre 2.782,4 e 3.704,1.

5.2.4 Variável Total Assets \$m

Segue abaixo o resumo exploratório desta variável:

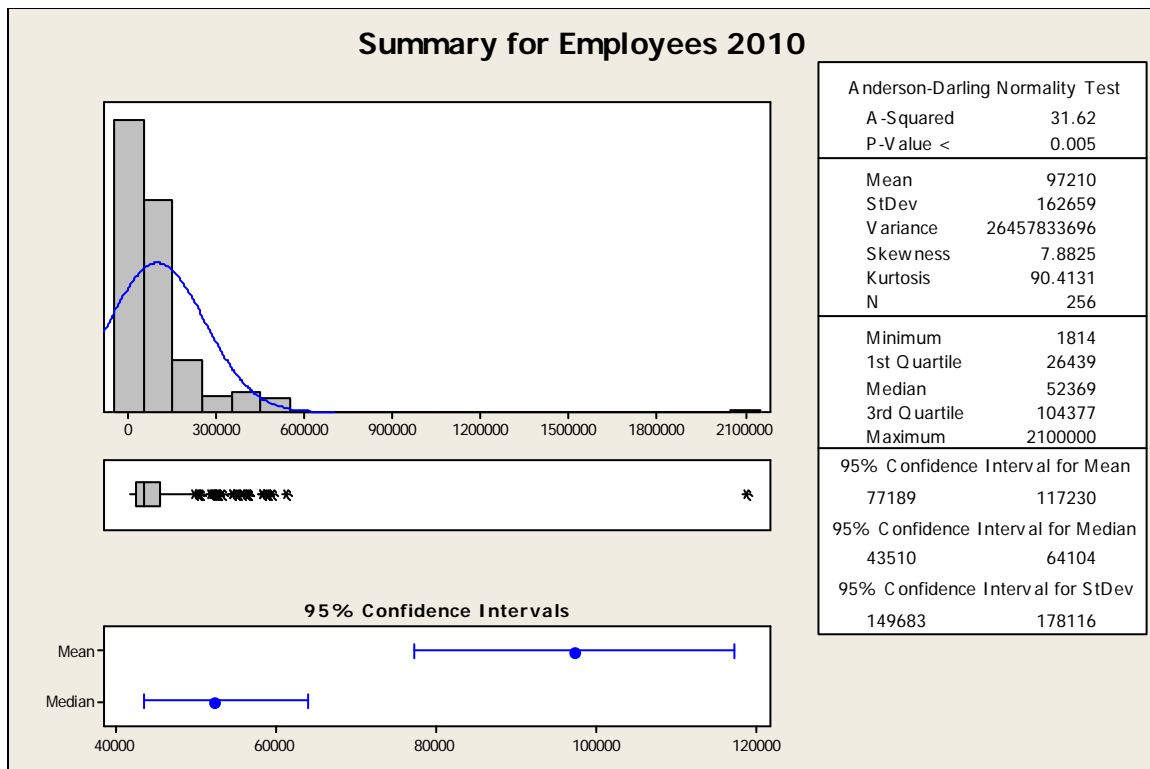


As principais observações que podemos fazer são:

- Forma: o histograma permite verificar que trata-se de uma distribuição assimétrica para a direita, ou seja, a maior parte das empresas se concentra em uma faixa mais baixa que o ponto médio (392.169) entre o valor mínimo (2.520) e o valor máximo (781.818) registrado. O gráfico de caixa e a análise dos 5 números também nos deixa claro esta assimetria da distribuição. A posição da linha da mediana longe do centro do quadrado principal do gráfico de caixa nos confirma esta afirmativa.
- Valores Atípicos: Há valores atípicos no gráfico, com valores acima de 116.910 \$m. O maior valor é da General Eletric (9ª colocada) com 781.818.
- Centro e Dispersão: a mediana nos indica que aproximadamente metade das empresas possui o Total Assets menor que 34.556 e metade maior que esse valor. A média do Total Assets é de 57.818, e o desvio padrão é de 74.848 (alto). O valor mínimo foi 2.520 e o valor máximo foi 781.818. Com 95% de confiança, podemos afirmar que a média do Total Assets das empresas encontra-se entre 48.605 e 67.030.

5.2.5 Variável Employees

Segue abaixo o resumo exploratório desta variável:

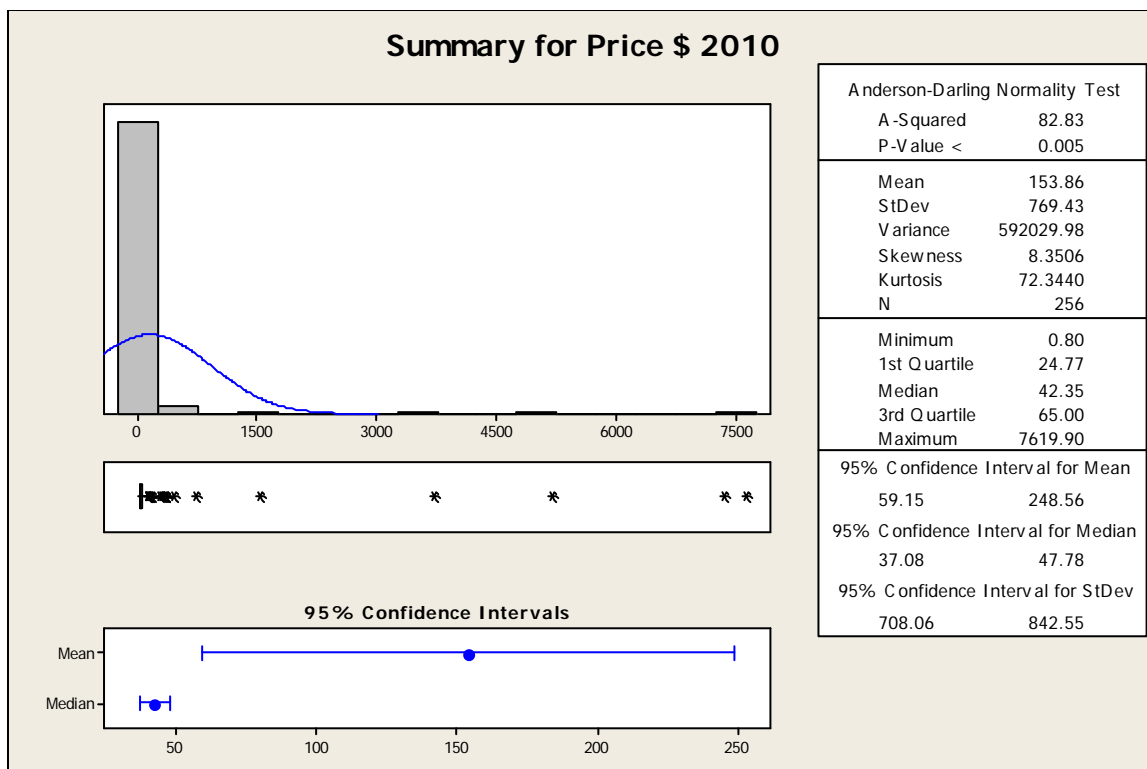


As principais observações que podemos fazer são:

- Forma: o histograma permite verificar que trata-se de uma distribuição assimétrica para a direita, ou seja, a maior parte das empresas se concentra em uma faixa mais baixa que o ponto médio (1.050.907) entre o valor mínimo (1.814) e o valor máximo (2.100.000) registrado. O gráfico de caixa e a análise dos 5 números também nos deixa claro esta assimetria da distribuição. A posição da linha da mediana longe do centro do quadrado principal do gráfico de caixa nos confirma esta afirmativa.
- Valores Atípicos: Há valores atípicos no gráfico, com valores acima de 197.345 \$m. O maior valor é do Wal-Mart Stores (7ª colocada) com 2.100.000 empregados.
- Centro e Dispersão: a mediana nos indica que aproximadamente metade das empresas possui quantidade de empregados menor que 52.369 e metade maior que esse valor. A média de empregados é de 97.210, e o desvio padrão é de 162.659 (alto). O valor mínimo foi 1.814 e o valor máximo foi 2.100.000. Com 95% de confiança, podemos afirmar que a média da quantidade de empregados das empresas encontra-se entre 77.189 e 117.230.

5.2.6 Variável Price \$

Segue abaixo o resumo exploratório desta variável:

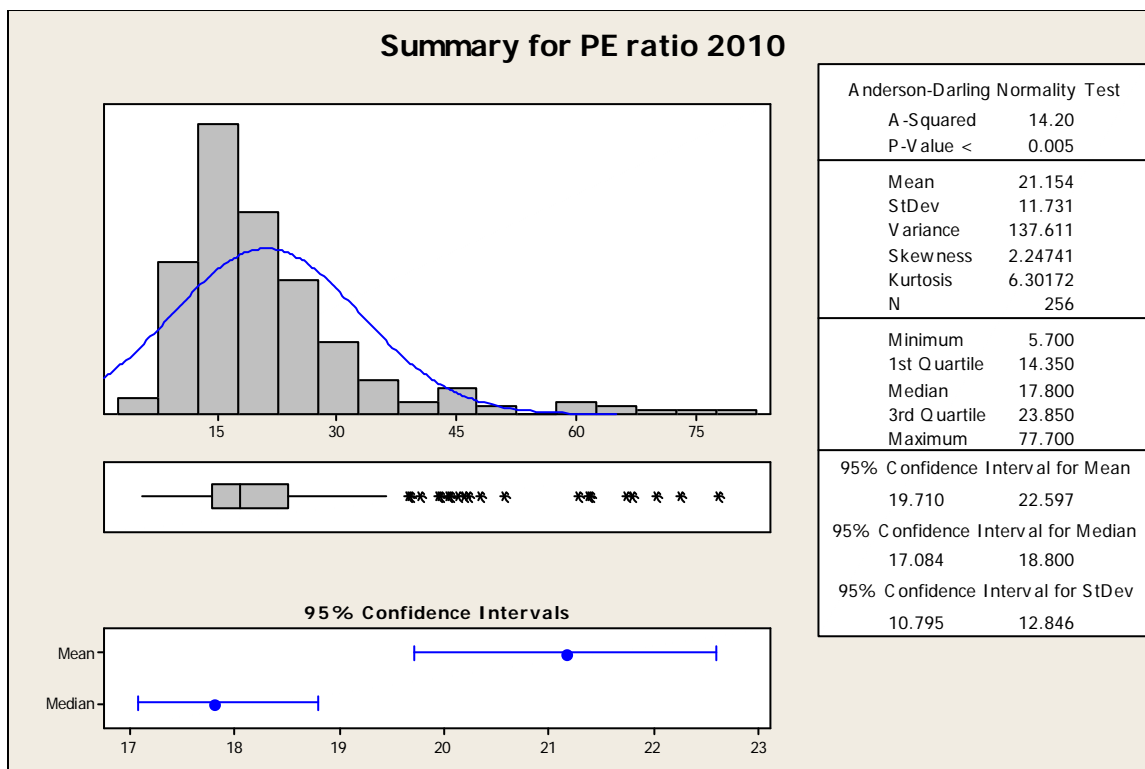


As principais observações que podemos fazer são:

- Forma: o histograma permite verificar que trata-se de uma distribuição assimétrica para a direita, ou seja, a maior parte das empresas se concentra em uma faixa mais baixa que o ponto médio (3.810,35) entre o valor mínimo (0,80) e o valor máximo (7.619,90) registrado. O gráfico de caixa e a análise dos 5 números também nos deixa claro esta assimetria da distribuição. A posição da linha da mediana longe do centro do quadrado principal do gráfico de caixa nos confirma esta afirmativa.
- Valores Atípicos: Há diversos valores atípicos no gráfico, com valores acima de 100,575 \$m. Os maiores valores são a Central Japan Railway (490ª colocada) com 7.619,90 e a Inpex (465ª colocada) com 7.341,60.
- Centro e Dispersão: a mediana nos indica que aproximadamente metade das empresas possui o preço da ação menor que 42.35 e metade maior que esse valor. A média do preço da ação é de 153,86, e o desvio padrão é de 769,43 (alto). O valor mínimo foi 0,80 e o valor máximo foi 7.619,90. Com 95% de confiança, podemos afirmar que a média do preço da ação das empresas encontra-se entre 59,15 e 248,56.

5.2.7 Variável PE ratio

Segue abaixo o resumo exploratório desta variável:

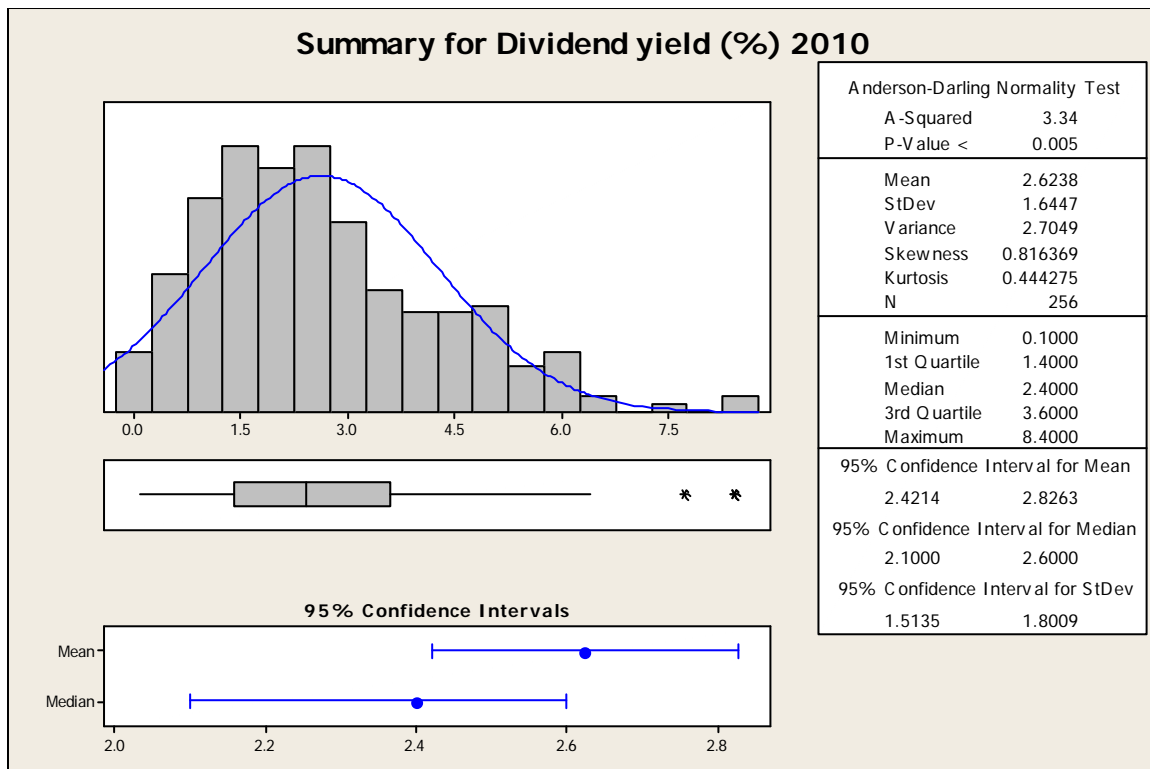


As principais observações que podemos fazer são:

- Forma: o histograma permite verificar que trata-se de uma distribuição assimétrica para a direita, ou seja, a maior parte das empresas se concentra em uma faixa mais baixa que o ponto médio (41,700) entre o valor mínimo (5,700) e o valor máximo (77,700) registrado. O gráfico de caixa e a análise dos 5 números também nos deixa claro esta assimetria da distribuição. A posição da linha da mediana longe do centro do quadrado principal do gráfico de caixa nos confirma esta afirmativa.
- Valores Atípicos: Há diversos valores atípicos no gráfico, com valores acima de 23,750 \$m.
- Centro e Dispersão: a mediana nos indica que aproximadamente metade das empresas possui o PE ratio menor que 17,800 e metade maior que esse valor. A média do PE ratio é de 21,154, e o desvio padrão é de 11,731 (alto). O valor mínimo foi 5,700 e o valor máximo foi 77,700. Com 95% de confiança, podemos afirmar que a média do PE ratio das empresas encontra-se entre 19,710 e 22,597.

5.2.8 Variável Dividend Yield (%)

Segue abaixo o resumo exploratório desta variável:



As principais observações que podemos fazer são:

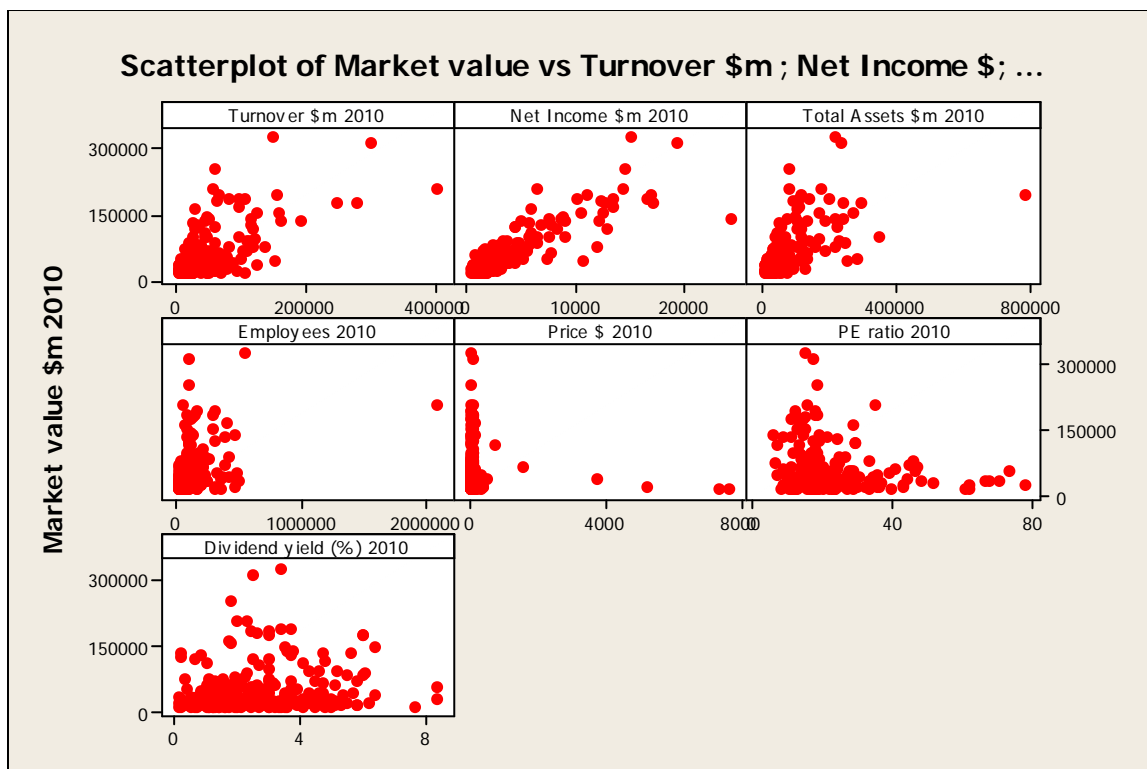
- Forma: o histograma permite verificar que trata-se de uma distribuição assimétrica para a direita, ou seja, a maior parte das empresas se concentra em uma faixa mais baixa que o ponto médio (4,25) entre o valor mínimo (0,10) e o valor máximo (8,40) registrado. O gráfico de caixa e a análise dos 5 números também nos deixa claro esta assimetria da distribuição. A posição da linha da mediana longe do centro do quadrado principal do gráfico de caixa nos confirma esta afirmativa.

- Valores Atípicos: Há 2 valores atípicos no gráfico, com valores acima de 5,50 \$m, sendo eles: a France Telecom (91ª colocada) com 8,40 e a Mobile Telesystems (470ª colocada) com 7,70.

- Centro e Dispersão: a mediana nos indica que aproximadamente metade das empresas possui o Dividend Yield (%) menor que 2,400 e metade maior que esse valor. A média do Dividend Yield (%) é de 2,6238, e o desvio padrão é de 1,6447 (alto). O valor mínimo foi 0,10 e o valor máximo foi 8,40. Com 95% de confiança, podemos afirmar que a média do Dividend Yield (%) das empresas encontra-se entre 2,4214 e 2,8263.

5.3 Análise de Relação entre Variáveis

Verificar a relação entre as variáveis de dados de indicadores relacionados ao Ranking das 500 empresas mais valiosas do mundo.



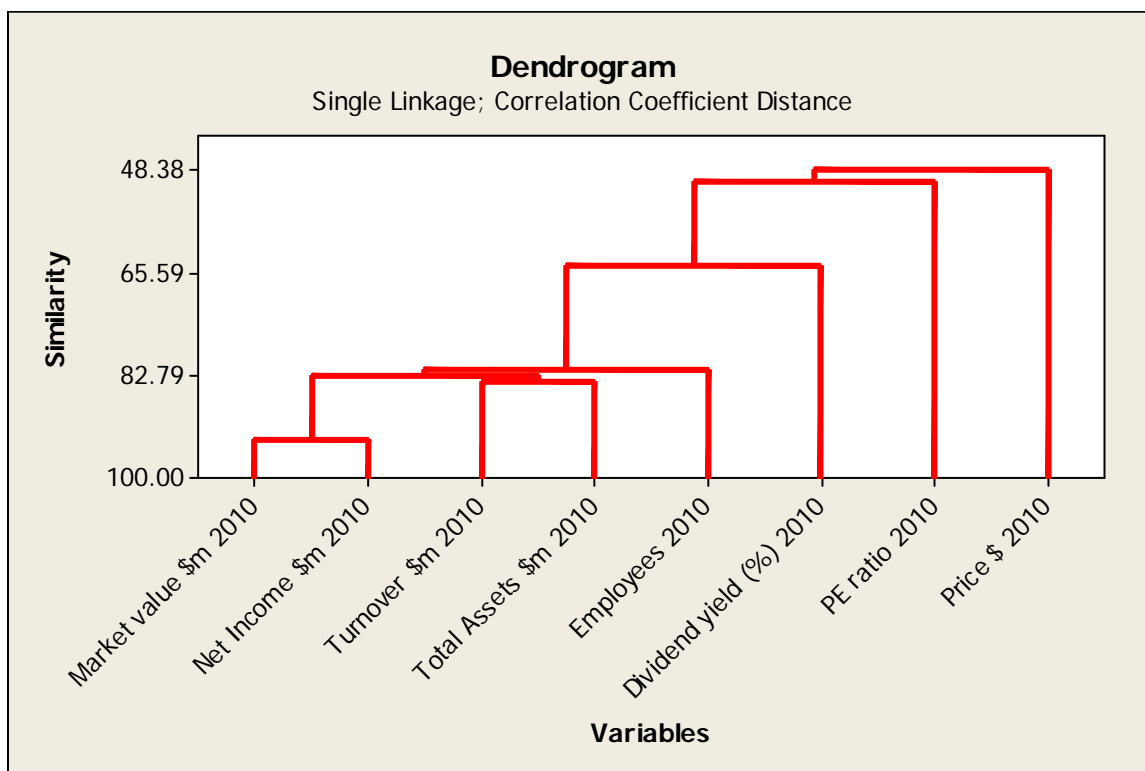
Podemos observar pela análise dos gráficos de dispersão que a variável Net Income \$m possui uma relação linear de associação positiva muito clara. Outras variáveis como o Turnover \$m e Total Assets \$m também possuem esta relação, porém com dados mais dispersos.

	Market Value \$m	Turnover \$m	Net Income \$m	Total Assets \$m	Employees	Price \$	PE ratio
Turnover \$m	0,654						
	0,000						
Net Income \$m	0,870	0,658					
	0,000	0,000					
Total Assets \$m	0,612	0,679	0,608				
	0,000	0,000	0,000				
Employees	0,366	0,639	0,347	0,306			
	0,000	0,000	0,000	0,000			
Price \$	-0,076	-0,032	-0,052	-0,043	-0,066		
	0,228	0,607	0,405	0,498	0,293		
PE ratio	-0,124	-0,111	-0,347	-0,134	0,005	-0,082	
	0,047	0,076	0,000	0,032	0,942	0,190	
Dividend yield	0,137	0,194	0,235	0,270	0,006	-0,098	-0,328
	0,028	0,002	0,000	0,000	0,920	0,116	0,000

Cell Contents: Pearson correlation
P-Value

Quase todas as correlações são significativas de acordo com o P-Value (para este trabalho consideraremos significativo P-Value menor ou igual a 0,10), exceto as correlações com a

variável Price \$ e PE ratio x Employees, Dividend yield x Employees. Assim como verificado no gráfico de dispersão, as variáveis Turnover \$m, NET Income \$m e Total Assets \$m possuem um coeficiente de correlação alto com a variável dependente Market Value \$m, sendo a variável Net Income \$m com maior correlação e a variável Price \$ com a menor correlação.



Através da observação e análise do Dendrograma, podemos observar um maior grau de similaridade da variável Market Value com o Net Income \$m. As variáveis Turnover \$m e Total Assets \$m também possuem um alto grau de similaridade. A variável Price \$ possui o menor grau de similaridade. Essas verificações estão de acordo com o observado previamente nos gráficos de dispersão e análise de correlação.

5.4 Regressões Múltiplas

Realizar uma análise de regressões múltiplas com dados de indicadores relacionados ao Ranking das 500 empresas mais valiosas do mundo, buscando encontrar as equações de regressão linear múltipla tendo como variável explicativa o Valor de Mercado.

Regression Analysis: Market value versus Turnover \$m ; Net Income \$; ...

The regression equation is
 Market value \$m 2010 = - 1851 + 0.0671 Turnover \$m 2010
 + 11.3 Net Income \$m 2010 + 0.0659 Total Assets \$m 2010
 + 0.0009 Employees 2010 - 1.06 Price \$ 2010
 + 739 PE ratio 2010 - 1389 Dividend yield (%) 2010

255 cases used, 1 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	-1851	4982	-0.37	0.711
Turnover \$m 2010	0.06706	0.05373	1.25	0.213
Net Income \$m 2010	11.3467	0.5743	19.76	0.000
Total Assets \$m 2010	0.06594	0.02887	2.28	0.023
Employees 2010	0.00086	0.01182	0.07	0.942
Price \$ 2010	-1.055	1.891	-0.56	0.577
PE ratio 2010	739.3	139.2	5.31	0.000
Dividend yield (%) 2010	-1389.1	967.7	-1.44	0.152

S = 22815.9 R-Sq = 80.5% R-Sq(adj) = 79.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	5.29880E+11	75697213299	145.41	0.000
Residual Error	247	1.28580E+11	520567290		
Total	254	6.58461E+11			

Source	DF	Seq SS
Turnover \$m 2010	1	2.82033E+11
Net Income \$m 2010	1	2.24561E+11
Total Assets \$m 2010	1	2595321650
Employees 2010	1	215084037
Price \$ 2010	1	524894264
PE ratio 2010	1	18878467084
Dividend yield (%) 2010	1	1072795435

Unusual Observations

Obs	Turnover \$m 2010	Market value \$m 2010	Fit	SE Fit	Residual	St Resid
1	149303	329260	200534	5645	128725	5.82R
2	301500	316231	261315	11153	54916	2.76RX
3	58437	256865	183425	6092	73439	3.34R
4	54661	209935	101872	3236	108063	4.78R
5	405046	209001	209253	18185	-252	-0.02 X
6	155777	194246	193853	17571	393	0.03 X
8	97084	187255	135263	3273	51992	2.30R
12	247788	177609	224581	8588	-46972	-2.22RX
13	278188	176968	182303	9331	-5334	-0.26 X
15	27823	162830	91823	2964	71007	3.14R
21	115246	137996	301755	10452	-163759	-8.07RX
23	41825	134040	76989	2730	57051	2.52R
25	23252	129041	83504	3244	45537	2.02R
28	35127	123130	71500	2078	51630	2.27R
33	95037	101072	99639	7005	1433	0.07 X
49	117216	74014	152892	4474	-78878	-3.53R
93	18808	45868	126610	5314	-80742	-3.64R
112	70168	37243	34947	7264	2296	0.11 X
127	123156	34038	65731	7352	-31694	-1.47 X
176	54135	24709	62453	7678	-37744	-1.76 X
187	35919	23230	27794	9477	-4564	-0.22 X
242	11052	17315	16425	13444	890	0.05 X
254	16127	16383	16373	13991	10	0.00 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

O R-Quadrado é de 80,5%, o que significa que 80,5% do comportamento do Market Value está sendo explicado pela equação apresentada. O P-value das variáveis Turnover \$m, Employees, Price e Dividend Yield (%) nos leva a descartá-las do modelo do modelo, o que é coerente com a análise do dendograma, exceto pela variável Turnover \$m.

Stepwise Regression: Market value versus Turnover \$m ; Net Income \$; ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is Market value \$m 2010 on 7 predictors, with N = 255
N(cases with missing observations) = 1 N(all cases) = 256

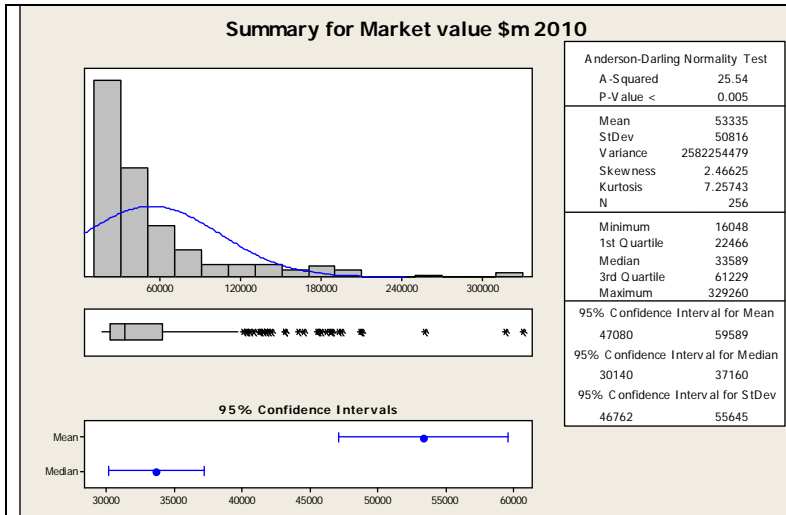
Step	1	2	3	4
Constant	14986	-6596	-6930	-6735
Net Income \$m 2010	11.81	12.76	11.78	11.39
T-Value	28.11	30.81	22.74	19.95
P-Value	0.000	0.000	0.000	0.000
PE ratio 2010		875	835	807
T-Value		6.62	6.38	6.13
P-Value		0.000	0.000	0.000
Total Assets \$m 2010			0.077	0.057
T-Value			3.09	2.06
P-Value			0.002	0.040
Turnover \$m 2010				0.069
T-Value				1.58
P-Value				0.116
S	25122	23234	22849	22782
R-Sq	75.75	79.34	80.10	80.29
R-Sq(adj)	75.66	79.18	79.86	79.98
Mallows Cp	55.7	12.3	4.7	4.3

Utilizando a função stepwise, as variáveis Employees, Price \$ e Dividend Yield (%) são excluídas, pelos motivos explicados na análise de regressão, com exceção da variável Turnover \$m, que foi mantida pela função stepwise. Os demais valores ficaram muito próximos da análise de regressão. O R-Quadrado é de 80,29%, o que significa que 80,29% do comportamento do Market Value está sendo explicado pela equação apresentada.

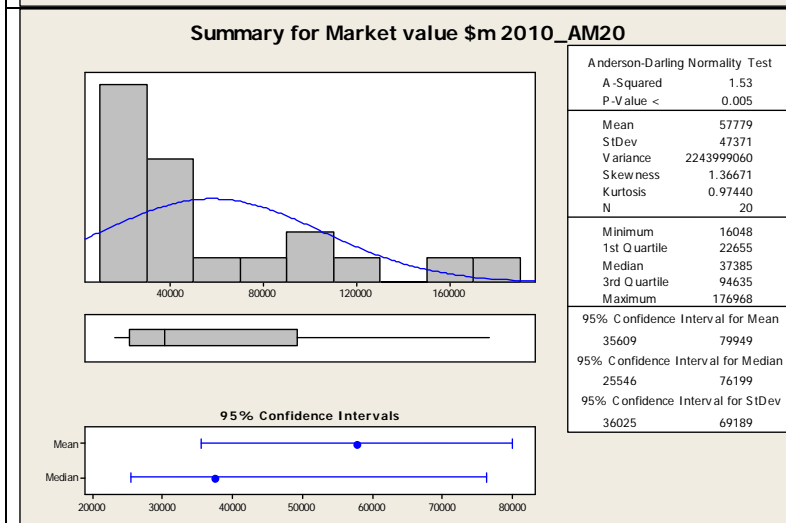
5.5 Amostragem

O presente tópico tem por objetivo efetuar uma pesquisa por amostragem e análise comparativa de médias e intervalos de confiança, variância dos dados de indicadores relacionados ao Ranking das 500 empresas mais valiosas do mundo.

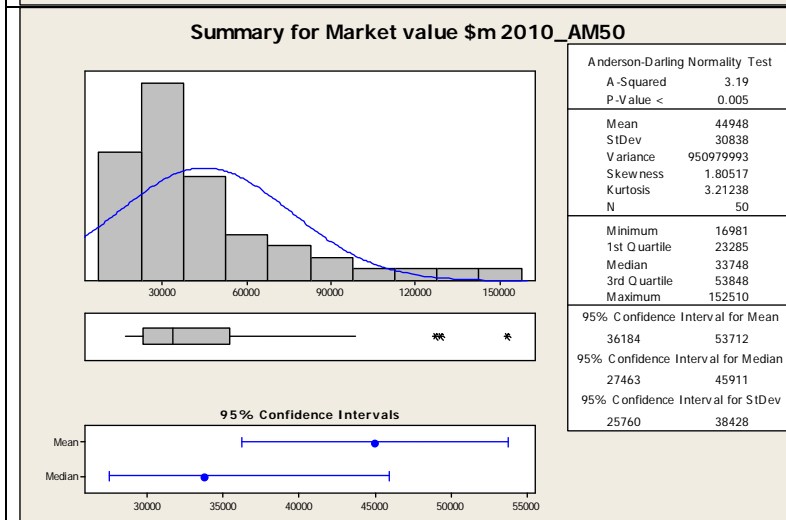
5.5.1 Variável Market Value \$m



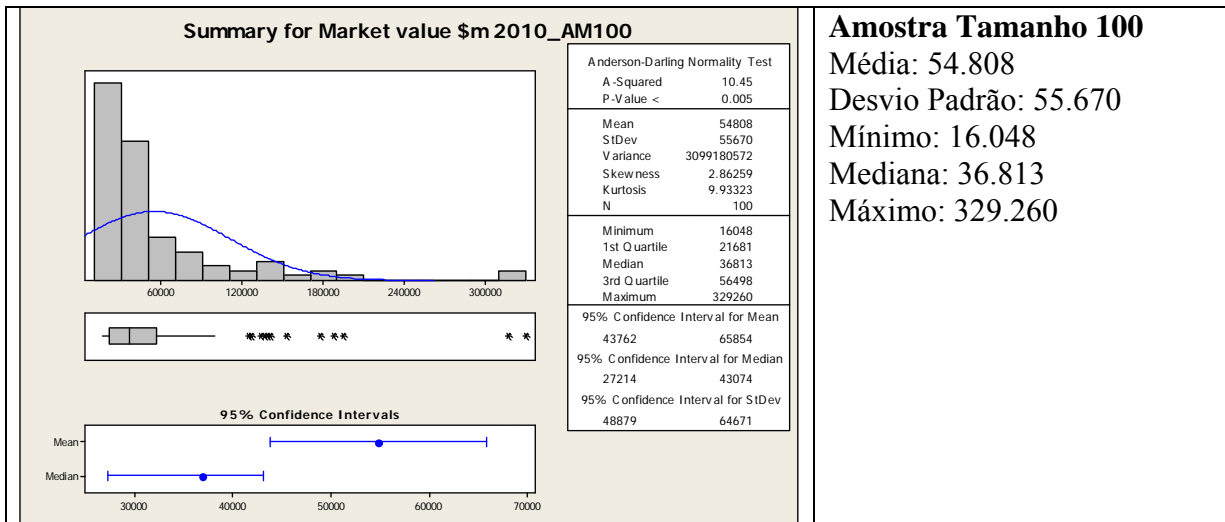
População Total
 Média: 53.335
 Desvio Padrão: 50.816
 Mínimo: 16.048
 Mediana: 33.589
 Máximo: 61.229



Amostra Tamanho 20
 Média: 57.779
 Desvio Padrão: 47.371
 Mínimo: 16.048
 Mediana: 37.385
 Máximo: 176.968



Amostra Tamanho 50
 Média: 44.958
 Desvio Padrão: 30.838
 Mínimo: 16.981
 Mediana: 33.748
 Máximo: 152.510



Amostra Tamanho 100

Média: 54.808

Desvio Padrão: 55.670

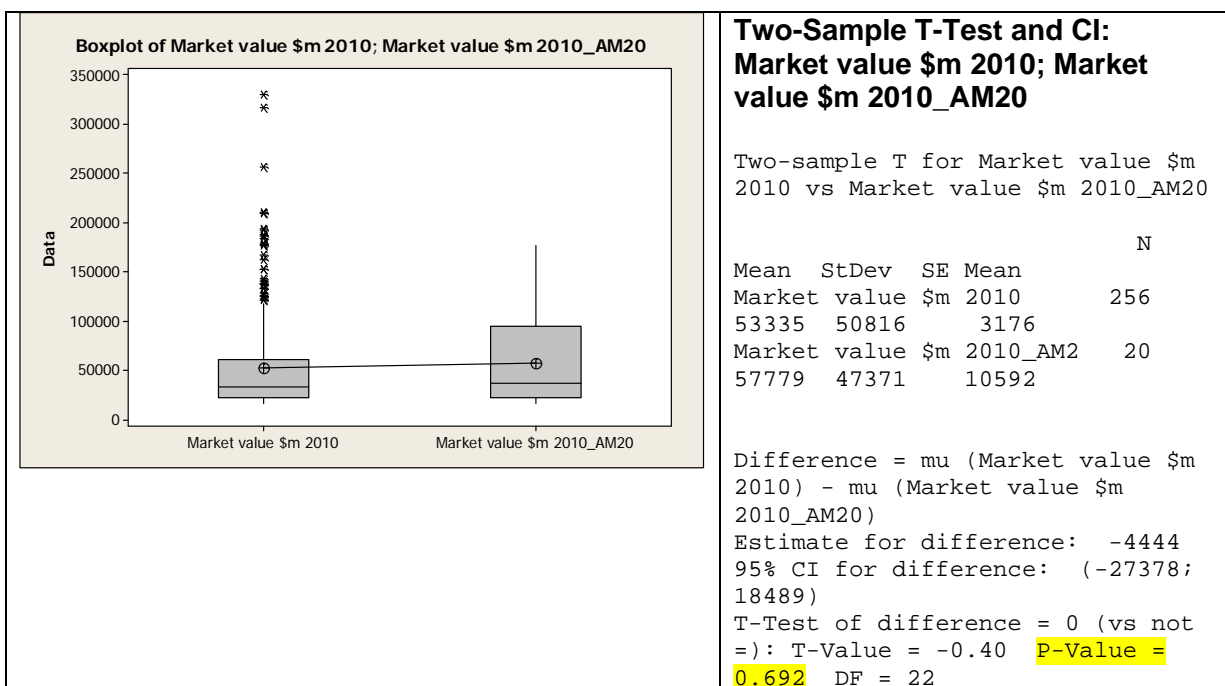
Mínimo: 16.048

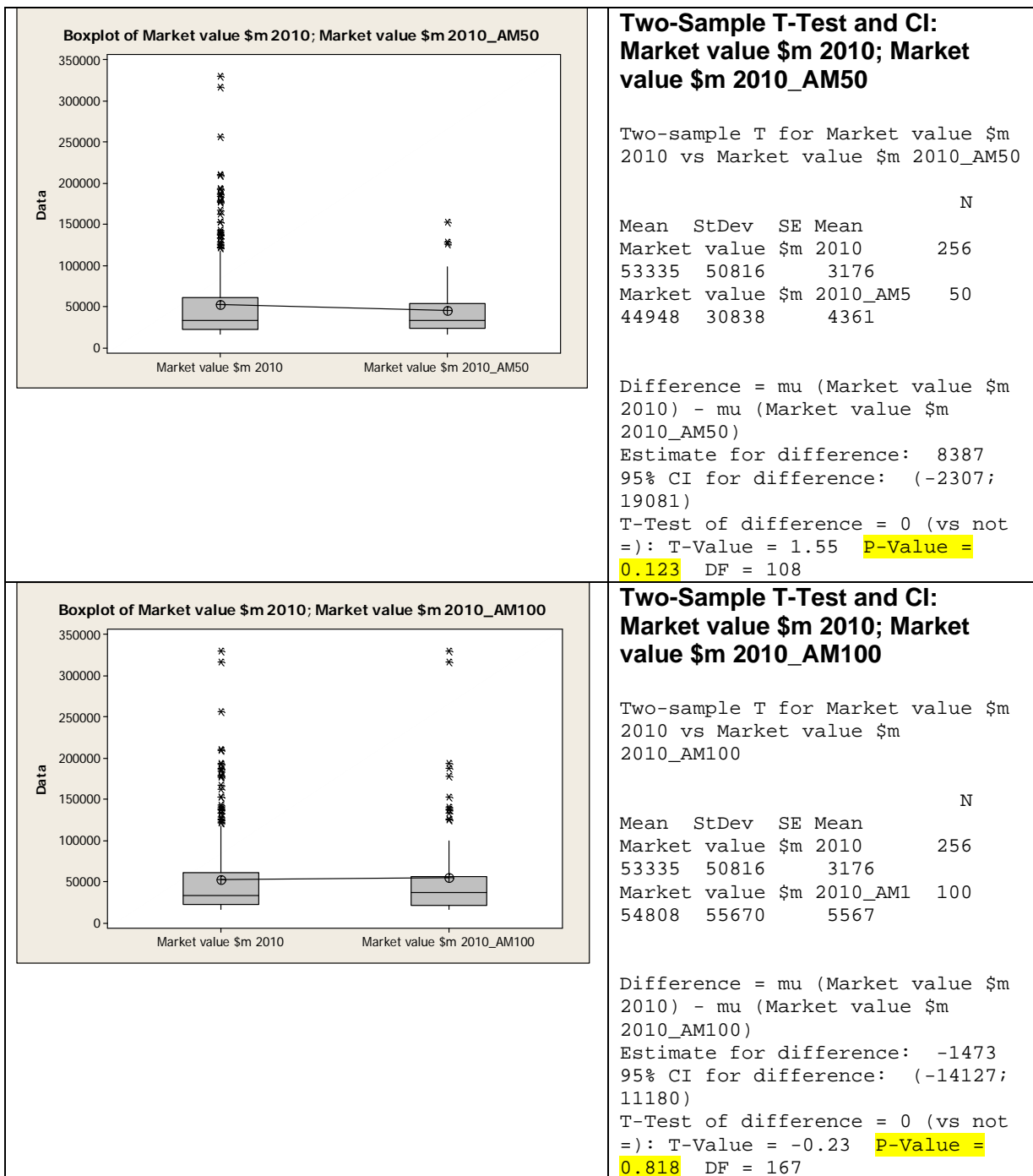
Mediana: 36.813

Máximo: 329.260

A amostra que mais representou a população total foi a de maior tamanho (100), cuja média é a mais próxima da população total, sendo, portanto a mais representativa.

O Box-Plot do Teste para as amostras confirma a hipótese nula de diferença de médias, para todas as amostras. As amostras 20, 50 e 100 são estatisticamente iguais, assumindo um intervalo de confiança igual a 95%, por que o valor de P é superior a 0,10.





Podemos também observar pelo teste One-way Anova que, conforme o tamanho das amostras aumenta, o desvio padrão diminui.

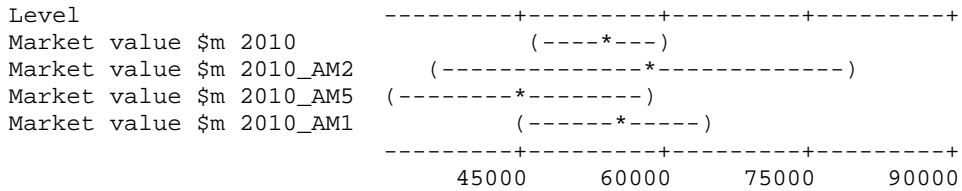
One-way ANOVA: Market value; Market value; Market value; Market value

Source	DF	SS	MS	F	P
Factor	3	4050595349	1350198450	0.54	0.655
Error	422	1.05453E+12	2498880973		
Total	425	1.05858E+12			

S = 49989 R-Sq = 0.38% R-Sq(adj) = 0.00%

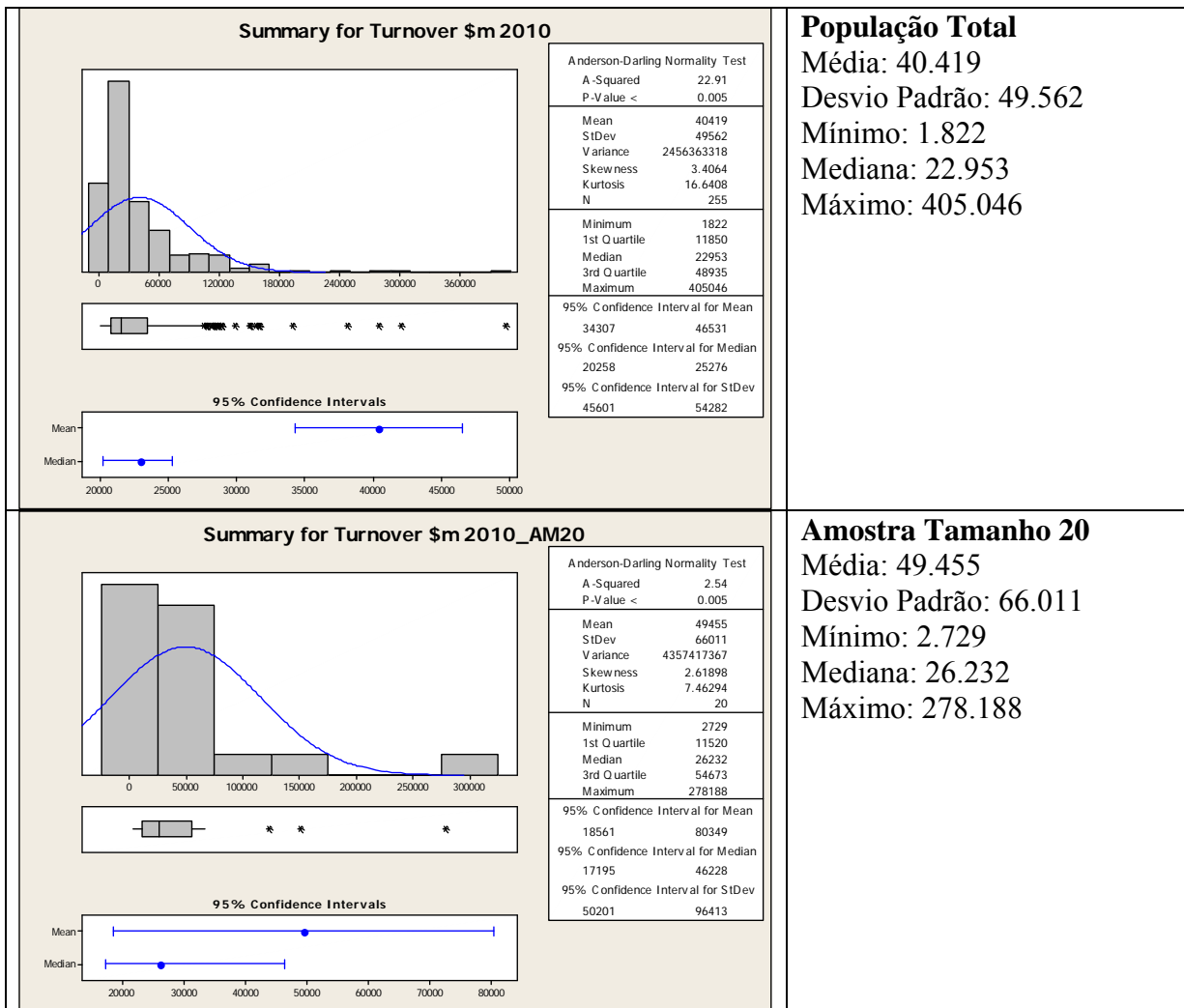
Level	N	Mean	StDev
Market value \$m 2010	256	53335	50816
Market value \$m 2010_AM2	20	57779	47371
Market value \$m 2010_AM5	50	44948	30838
Market value \$m 2010_AM1	100	54808	55670

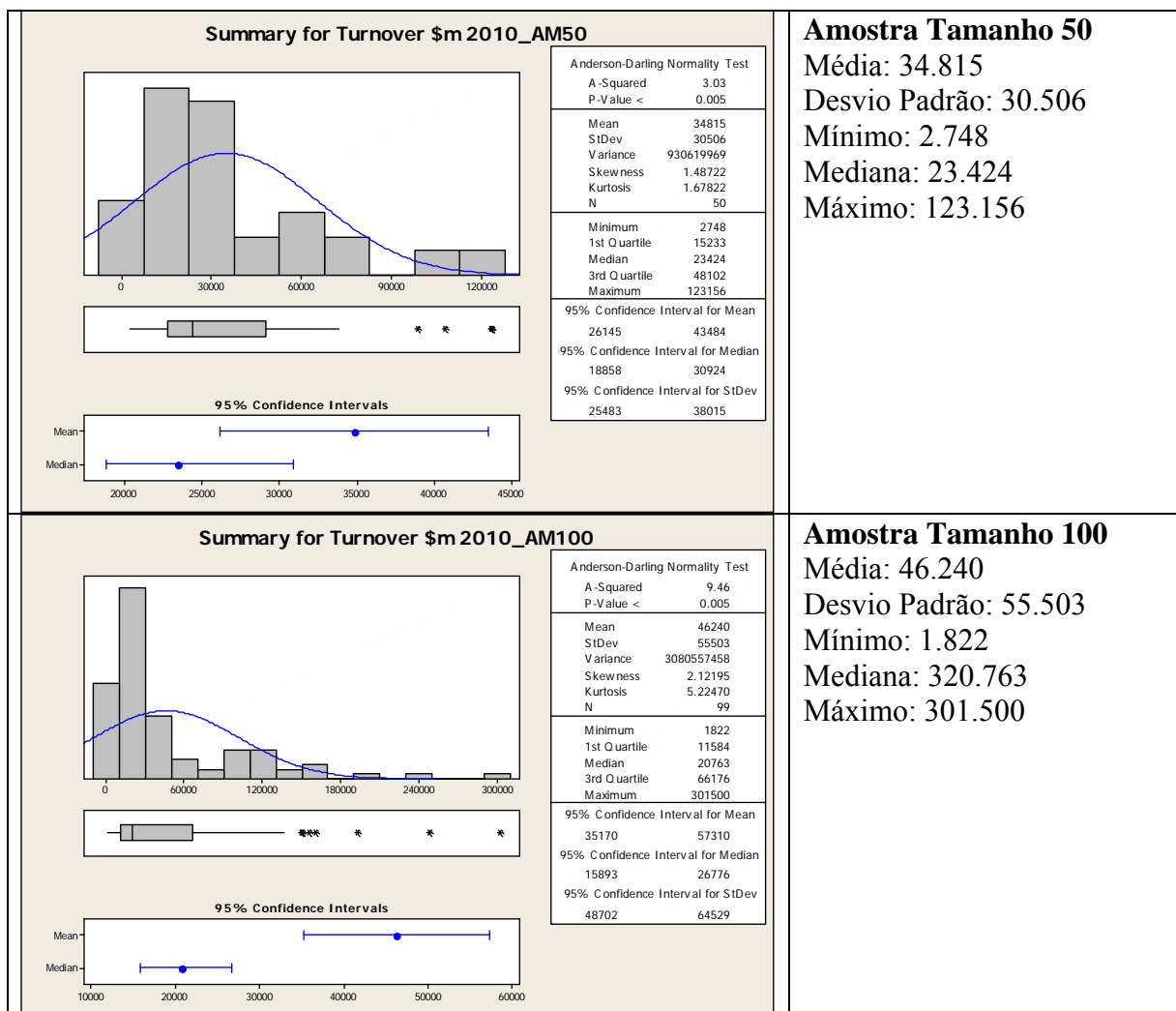
Individual 95% CIs For Mean Based on Pooled StDev



Pooled StDev = 49989

5.5.2 Variável Turnover Value \$m





Amostra Tamanho 50

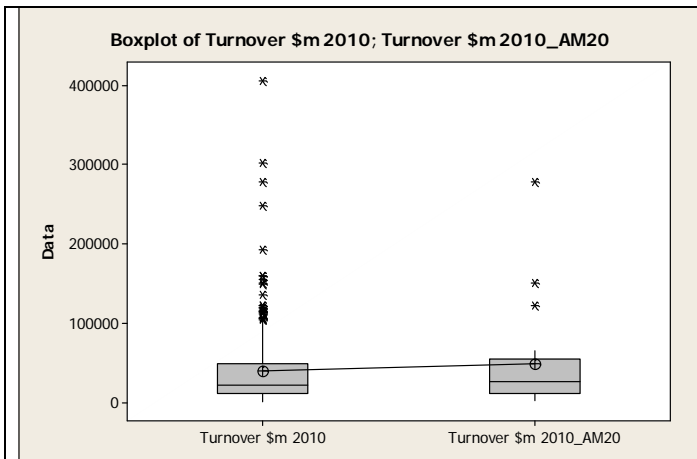
Média: 34.815
 Desvio Padrão: 30.506
 Mínimo: 2.748
 Mediana: 23.424
 Máximo: 123.156

Amostra Tamanho 100

Média: 46.240
 Desvio Padrão: 55.503
 Mínimo: 1.822
 Mediana: 320.763
 Máximo: 301.500

A amostra que mais representou a população total foi a de tamanho 50, cuja média é a mais próxima da população total, sendo, portanto a mais representativa. Talvez a amostra de tamanho 100 tenha contido valores aleatórios muito dispersos que tenham distorcido os seus resultados.

O Box-Plot do Teste para as amostras confirma a hipótese nula de diferença de médias, para todas as amostras. As amostras 20, 50 e 100 são estatisticamente iguais, assumindo um intervalo de confiança igual a 95%, por que o valor de P é superior a 0,10.



Two-Sample T-Test and CI: Turnover \$m 2010; Turnover \$m 2010_AM20

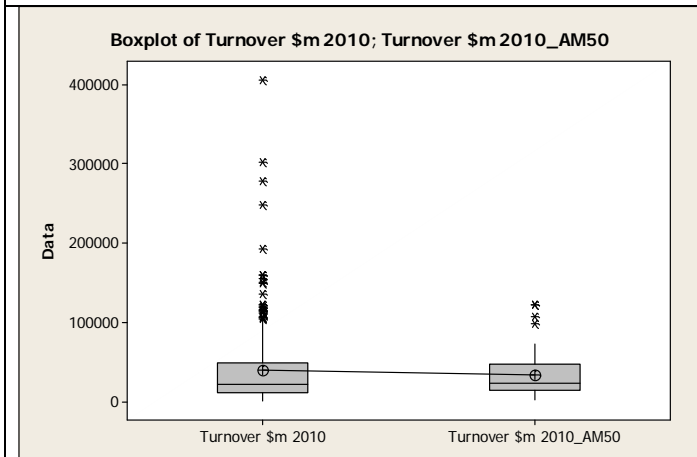
Two-sample T for Turnover \$m 2010
vs Turnover \$m 2010_AM20

	N	Mean	StDev	SE Mean
Turnover \$m 2010	255	40419	49562	3104
Turnover \$m 2010_AM20	20	49455	66011	14760

Difference = μ (Turnover \$m
2010) - μ (Turnover \$m
2010_AM20)

Estimate for difference: -9036
95% CI for difference: (-40499;
22427)

T-Test of difference = 0 (vs not
=): T-Value = -0.60 P-Value =
0.556 DF = 20



Two-Sample T-Test and CI: Turnover \$m 2010; Turnover \$m 2010_AM50

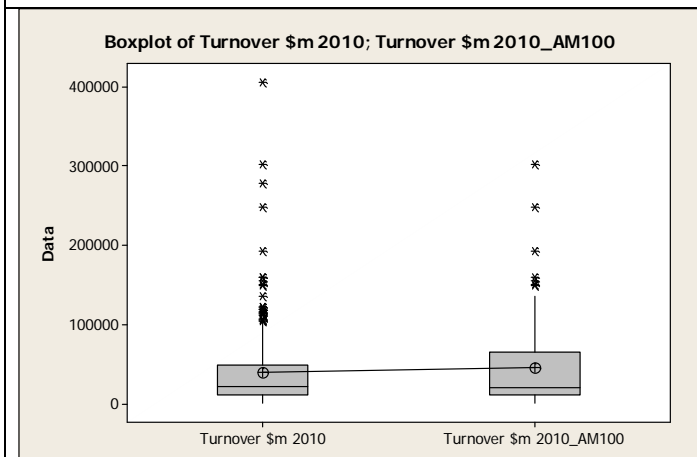
Two-sample T for Turnover \$m 2010
vs Turnover \$m 2010_AM50

	N	Mean	StDev	SE Mean
Turnover \$m 2010	255	40419	49562	3104
Turnover \$m 2010_AM50	50	34815	30506	4314

Difference = μ (Turnover \$m
2010) - μ (Turnover \$m
2010_AM50)

Estimate for difference: 5604
95% CI for difference: (-4931;
16140)

T-Test of difference = 0 (vs not
=): T-Value = 1.05 P-Value =
0.294 DF = 107



Two-Sample T-Test and CI: Turnover \$m 2010; Turnover \$m 2010_AM100

Two-sample T for Turnover \$m 2010
vs Turnover \$m 2010_AM100

	N	Mean	StDev	SE Mean
Turnover \$m 2010	255	40419	49562	3104
Turnover \$m 2010_AM100	99	46240	55503	5578

Difference = μ (Turnover \$m
2010) - μ (Turnover \$m
2010_AM100)

	Estimate for difference: -5821 95% CI for difference: (-18427; 6785) T-Test of difference = 0 (vs not =): T-Value = -0.91 P-Value = 0.363 DF = 162
--	--

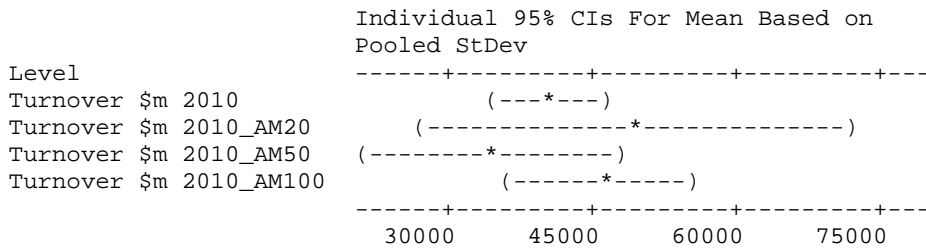
Podemos também observar pelo teste One-way Anova que, conforme o tamanho das amostras aumenta, o desvio padrão diminui, porém neste caso, a amostra maior não foi a que possui média mais próxima da população total.

One-way ANOVA: Turnover \$m ; Turnover \$m ; Turnover \$m ; Turnover \$m

Source	DF	SS	MS	F	P
Factor	3	6021976755	2007325585	0.80	0.495
Error	420	1.05420E+12	2510005291		
Total	423	1.06022E+12			

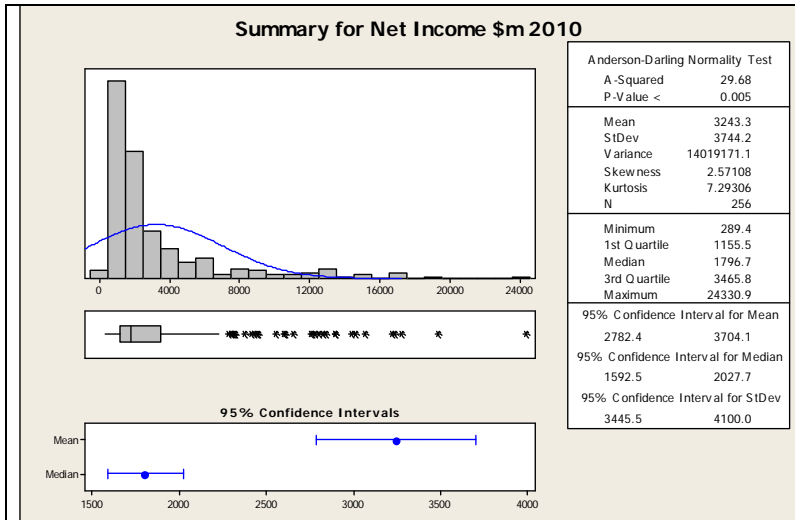
S = 50100 R-Sq = 0.57% R-Sq(adj) = 0.00%

Level	N	Mean	StDev
Turnover \$m 2010	255	40419	49562
Turnover \$m 2010_AM20	20	49455	66011
Turnover \$m 2010_AM50	50	34815	30506
Turnover \$m 2010_AM100	99	46240	55503

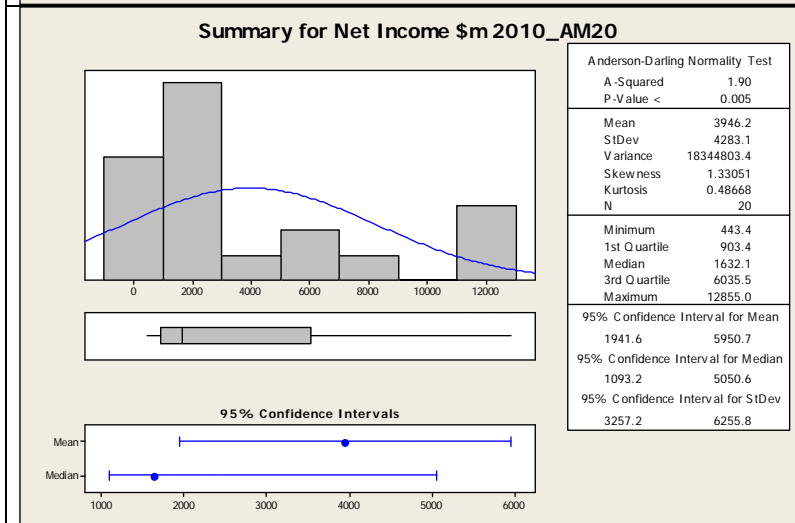


Pooled StDev = 50100

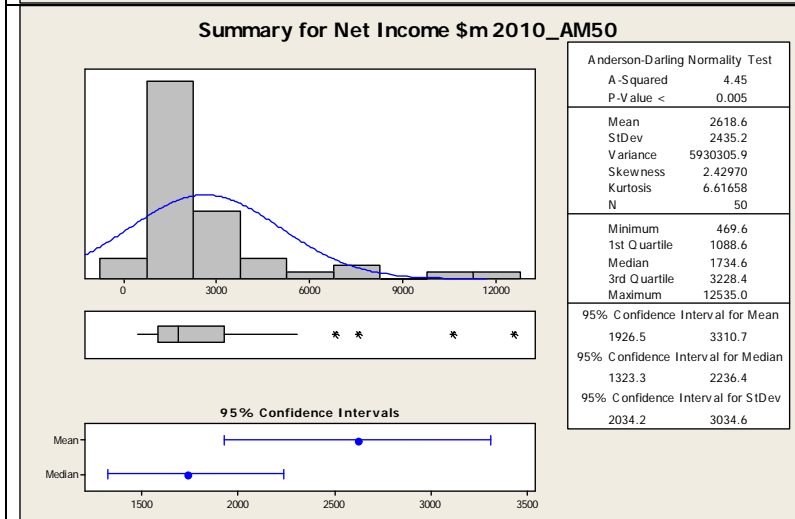
5.5.3 Variável Turnover Value \$m



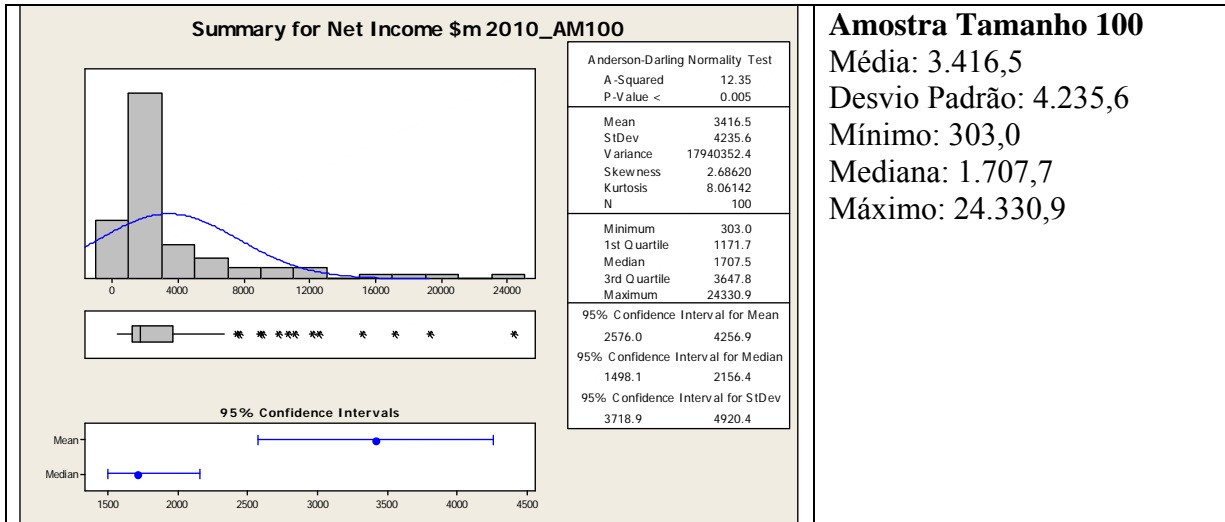
População Total
 Média: 3.243,3
 Desvio Padrão: 3.744,2
 Mínimo: 289,4
 Mediana: 1.796,7
 Máximo: 24.330,9



Amostra Tamanho 20
 Média: 3.946,2
 Desvio Padrão: 4.283,1
 Mínimo: 443,4
 Mediana: 1.632,1
 Máximo: 12.855,0



Amostra Tamanho 50
 Média: 2.618,6
 Desvio Padrão: 2.435,2
 Mínimo: 469,6
 Mediana: 1.734,6
 Máximo: 12.535,0

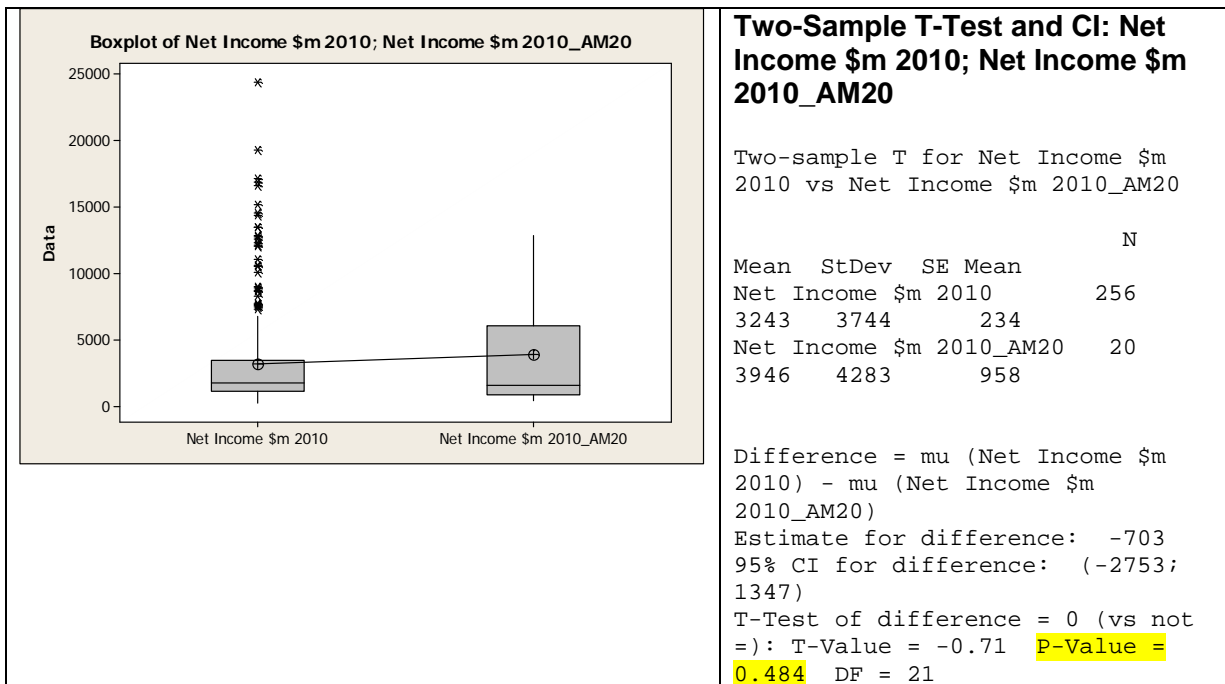


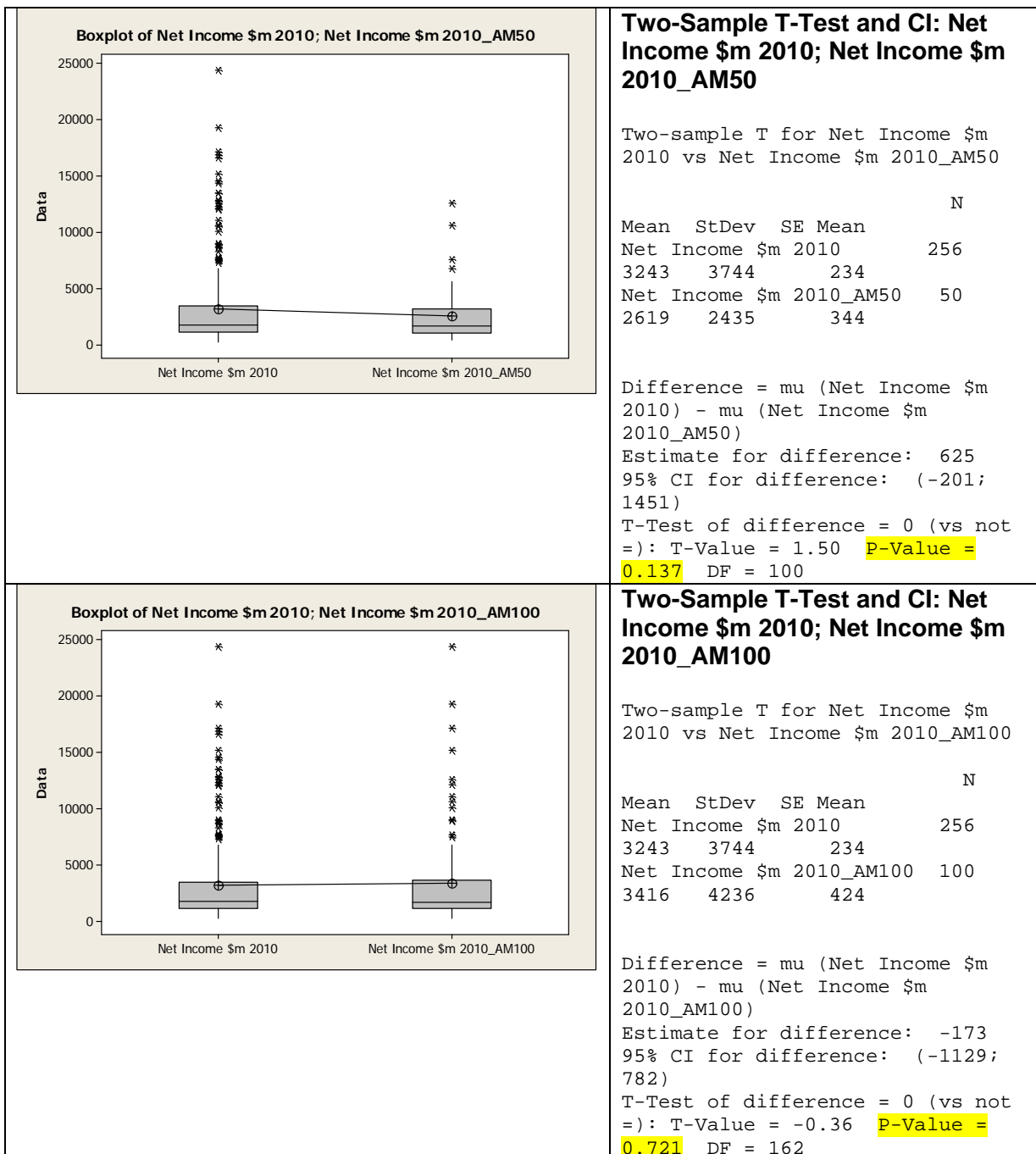
Amostra Tamanho 100

Média: 3.416,5
 Desvio Padrão: 4.235,6
 Mínimo: 303,0
 Mediana: 1.707,7
 Máximo: 24.330,9

A amostra que mais representou a população total foi a de maior tamanho (100), cuja média é a mais próxima da população total, sendo, portanto a mais representativa.

O Box-Plot do Teste para as amostras confirma a hipótese nula de diferença de médias, para todas as amostras. As amostras 20, 50 e 100 são estatisticamente iguais, assumindo um intervalo de confiança igual a 95%, por que o valor de P é superior a 0,10.





Podemos também observar pelo teste One-way Anova que, conforme o tamanho das amostras aumenta, o desvio padrão diminui, porém neste caso, a amostra maior não foi a que possui média mais próxima da população total.

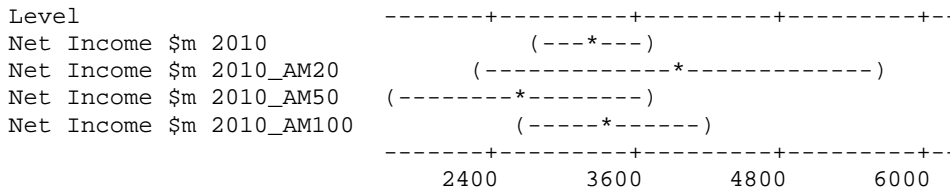
One-way ANOVA: Net Income \$; Net Income \$; Net Income \$; Net Income \$

Source	DF	SS	MS	F	P
Factor	3	32390802	10796934	0.76	0.517
Error	422	5990119766	14194597		
Total	425	6022510569			

S = 3768 R-Sq = 0.54% R-Sq(adj) = 0.00%

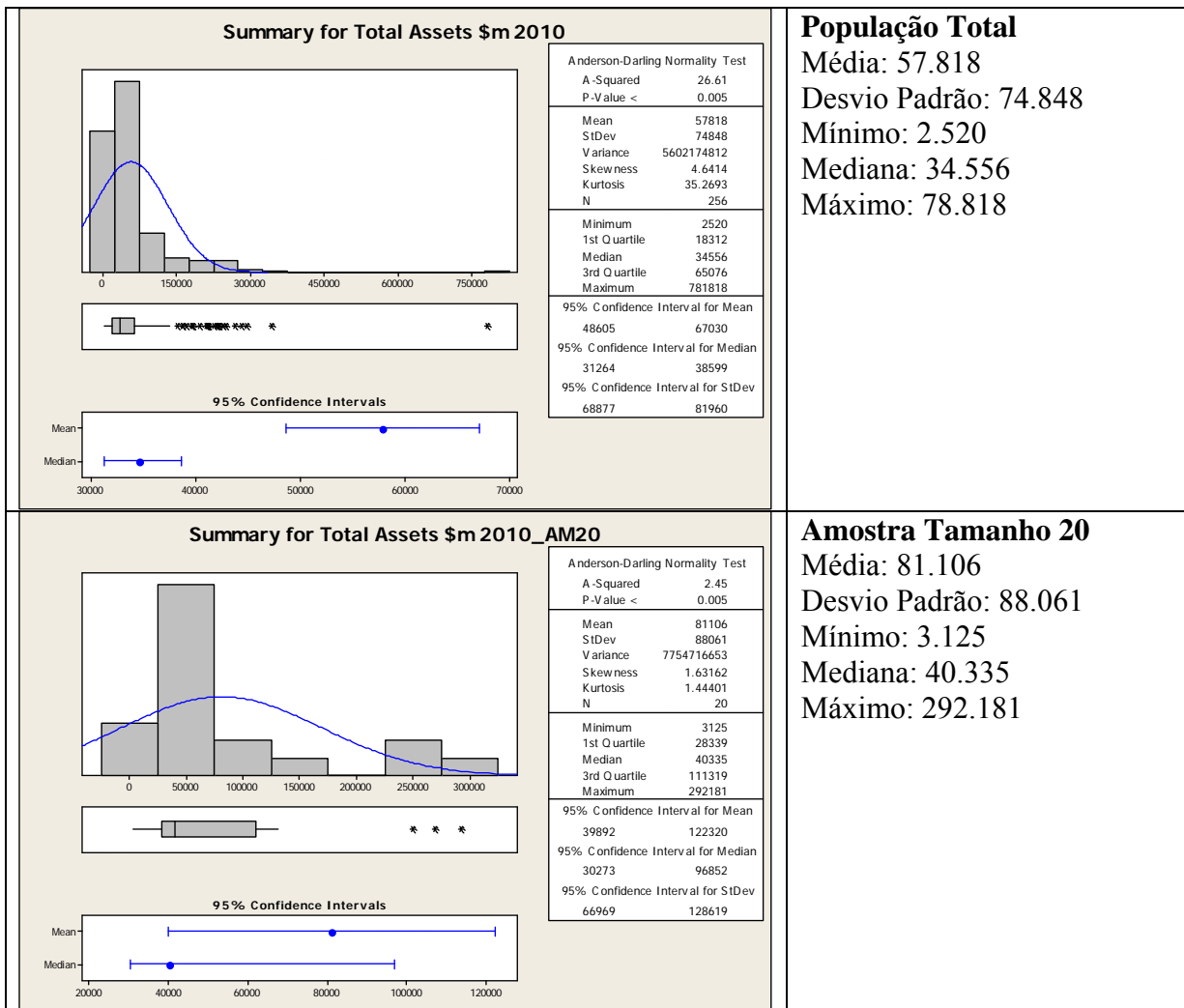
Level	N	Mean	StDev
Net Income \$m 2010	256	3243	3744
Net Income \$m 2010_AM20	20	3946	4283
Net Income \$m 2010_AM50	50	2619	2435
Net Income \$m 2010_AM100	100	3416	4236

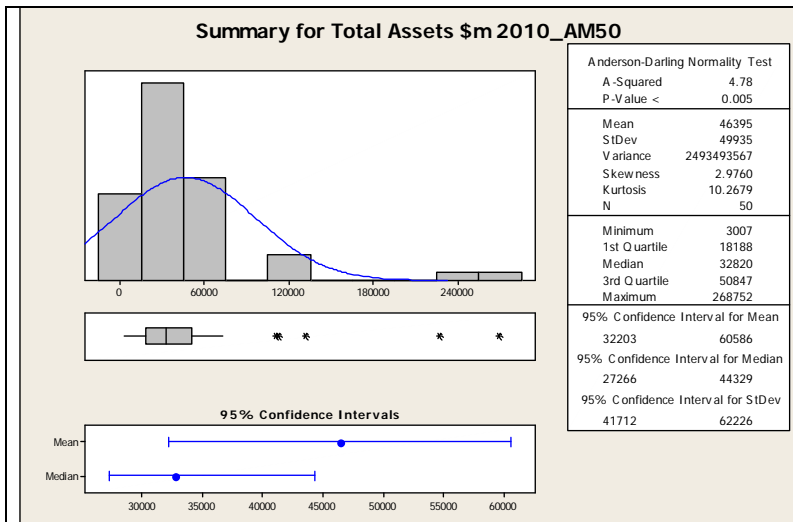
Individual 95% CIs For Mean Based on Pooled StDev



Pooled StDev = 3768

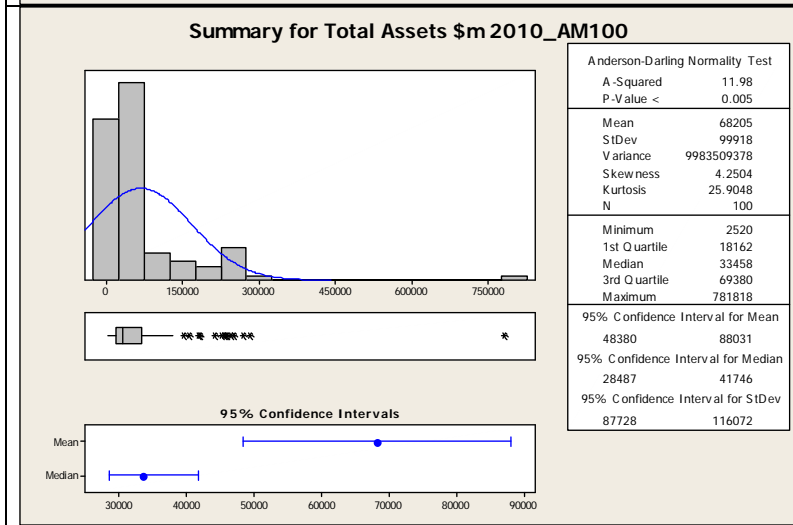
5.5.4 Variável Total Assets \$m





Amostra Tamanho 50

Média: 46.395
 Desvio Padrão: 49.935
 Mínimo: 3.007
 Mediana: 32.820
 Máximo: 268.752

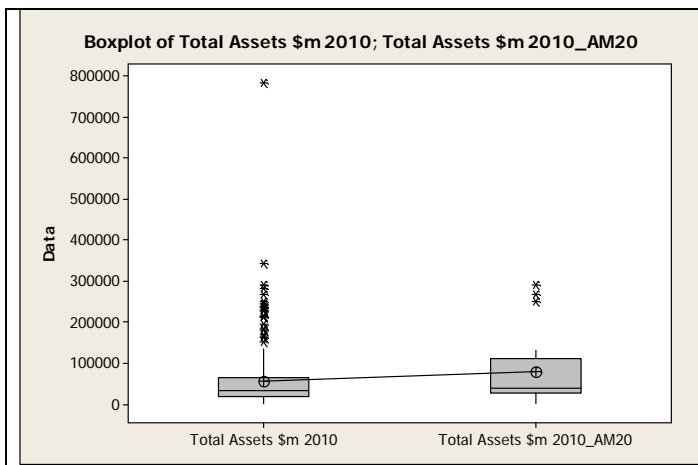


Amostra Tamanho 100

Média: 68.205
 Desvio Padrão: 99.918
 Mínimo: 2.520
 Mediana: 33.458
 Máximo: 781.818

A amostra que mais representou a população total foi a de maior tamanho (100), cuja média é a mais próxima da população total, sendo, portanto a mais representativa.

O Box-Plot do Teste para as amostras confirma a hipótese nula de diferença de médias, para todas as amostras. As amostras 20, 50 e 100 são estatisticamente iguais, assumindo um intervalo de confiança igual a 95%, por que o valor de P é superior a 0,10.

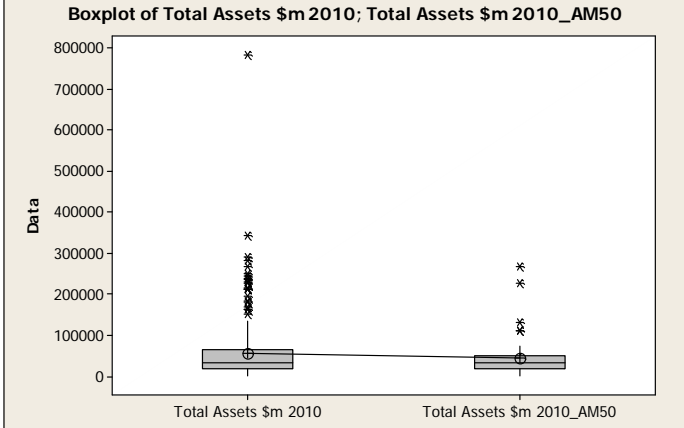
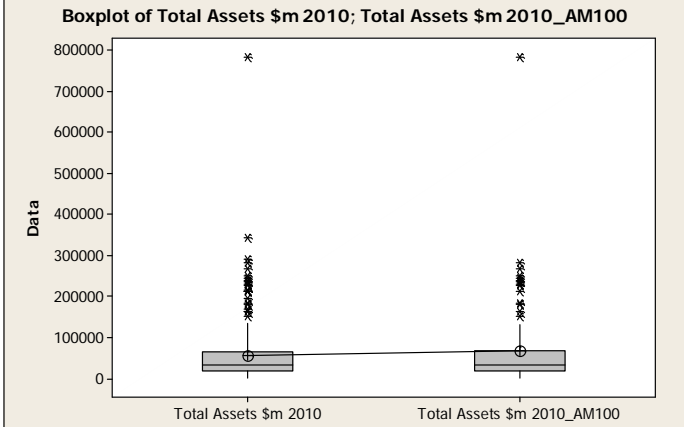


Two-Sample T-Test and CI: Total Assets \$m 2010; Total Assets \$m 2010_AM20

Two-sample T for Total Assets \$m 2010 vs Total Assets \$m 2010_AM20

	Mean	StDev	SE Mean	N
Total Assets \$m 2010	57818	74848	4678	256
Total Assets \$m 2010_AM20	81106	88061	19691	20

Difference = mu (Total Assets \$m

	<p>2010) - mu (Total Assets \$m 2010_AM20)</p> <p>Estimate for difference: -23288</p> <p>95% CI for difference: (-65378; 18801)</p> <p>T-Test of difference = 0 (vs not =): T-Value = -1.15 P-Value = 0.263 DF = 21</p>												
	<p>Two-Sample T-Test and CI: Total Assets \$m 2010; Total Assets \$m 2010_AM50</p> <p>Two-sample T for Total Assets \$m 2010 vs Total Assets \$m 2010_AM50</p> <table border="1" data-bbox="938 629 1423 786"> <thead> <tr> <th>Mean</th> <th>StDev</th> <th>SE Mean</th> <th>N</th> </tr> </thead> <tbody> <tr> <td>Total Assets \$m 2010</td> <td>57818</td> <td>74848</td> <td>256</td> </tr> <tr> <td>Total Assets \$m 2010_AM5</td> <td>46395</td> <td>49935</td> <td>50</td> </tr> </tbody> </table> <p>Difference = mu (Total Assets \$m 2010) - mu (Total Assets \$m 2010_AM50)</p> <p>Estimate for difference: 11423</p> <p>95% CI for difference: (-5389; 28235)</p> <p>T-Test of difference = 0 (vs not =): T-Value = 1.35 P-Value = 0.181 DF = 97</p>	Mean	StDev	SE Mean	N	Total Assets \$m 2010	57818	74848	256	Total Assets \$m 2010_AM5	46395	49935	50
Mean	StDev	SE Mean	N										
Total Assets \$m 2010	57818	74848	256										
Total Assets \$m 2010_AM5	46395	49935	50										
	<p>Two-Sample T-Test and CI: Total Assets \$m 2010; Total Assets \$m 2010_AM100</p> <p>Two-sample T for Total Assets \$m 2010 vs Total Assets \$m 2010_AM100</p> <table border="1" data-bbox="938 1328 1423 1485"> <thead> <tr> <th>Mean</th> <th>StDev</th> <th>SE Mean</th> <th>N</th> </tr> </thead> <tbody> <tr> <td>Total Assets \$m 2010</td> <td>57818</td> <td>74848</td> <td>256</td> </tr> <tr> <td>Total Assets \$m 2010_AM1</td> <td>68205</td> <td>99918</td> <td>100</td> </tr> </tbody> </table> <p>Difference = mu (Total Assets \$m 2010) - mu (Total Assets \$m 2010_AM100)</p> <p>Estimate for difference: -10388</p> <p>95% CI for difference: (-32194; 11419)</p> <p>T-Test of difference = 0 (vs not =): T-Value = -0.94 P-Value = 0.348 DF = 144</p>	Mean	StDev	SE Mean	N	Total Assets \$m 2010	57818	74848	256	Total Assets \$m 2010_AM1	68205	99918	100
Mean	StDev	SE Mean	N										
Total Assets \$m 2010	57818	74848	256										
Total Assets \$m 2010_AM1	68205	99918	100										

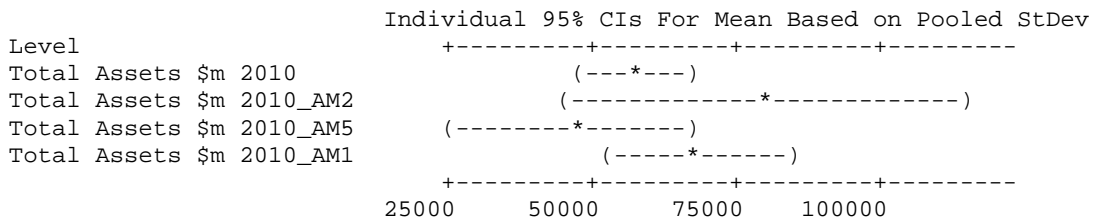
Podemos também observar pelo teste One-way Anova que, conforme o tamanho das amostras aumenta, o desvio padrão diminui.

One-way ANOVA: Total Assets; Total Assets; Total Assets; Total Assets

Source	DF	SS	MS	F	P
Factor	3	26116613421	8705537807	1.37	0.252
Error	422	2.68644E+12	6365978214		
Total	425	2.71256E+12			

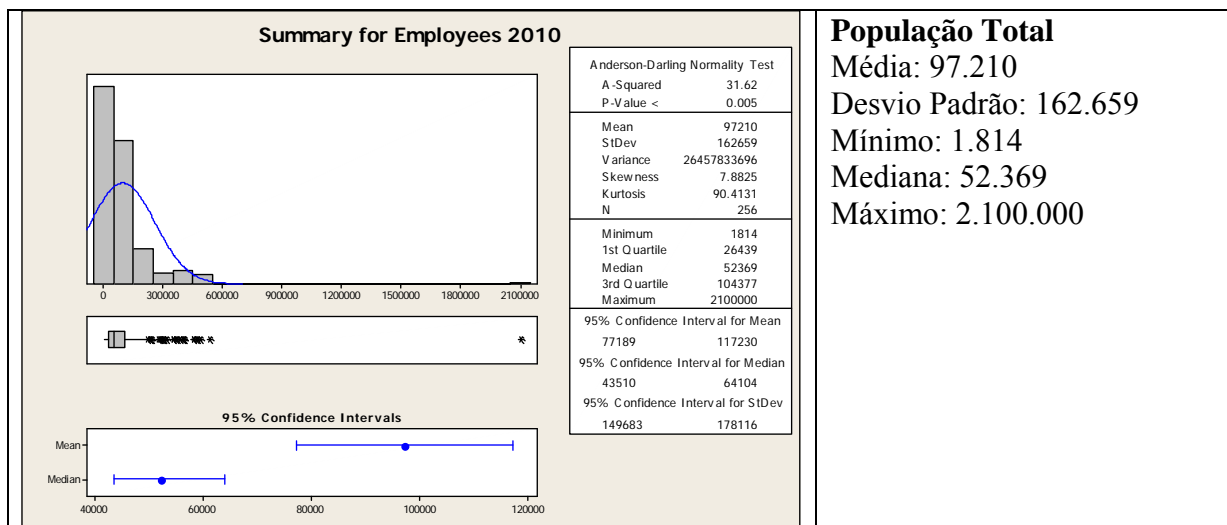
S = 79787 R-Sq = 0.96% R-Sq(adj) = 0.26%

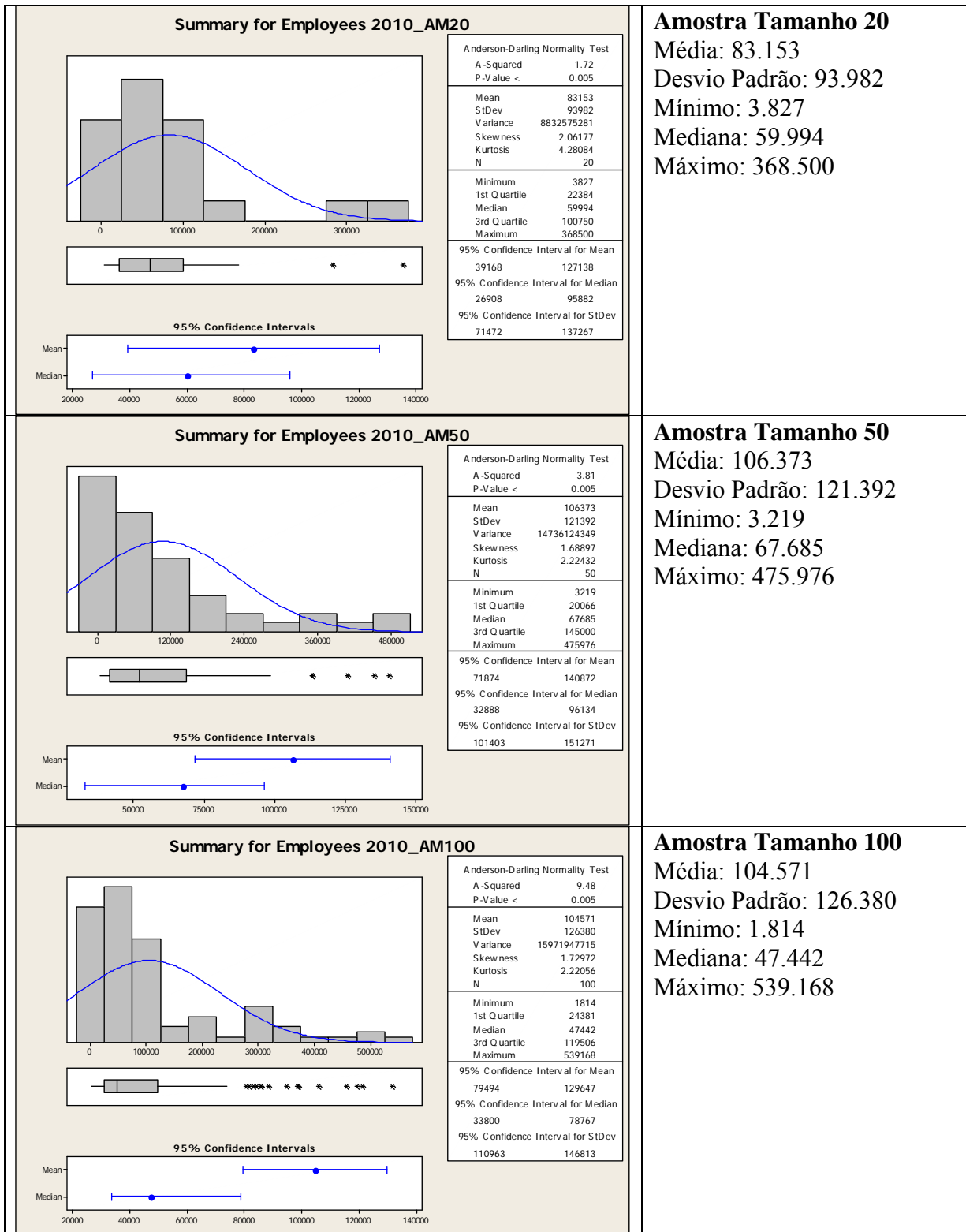
Level	N	Mean	StDev
Total Assets \$m 2010	256	57818	74848
Total Assets \$m 2010_AM2	20	81106	88061
Total Assets \$m 2010_AM5	50	46395	49935
Total Assets \$m 2010_AM1	100	68205	99918



Pooled StDev = 79787

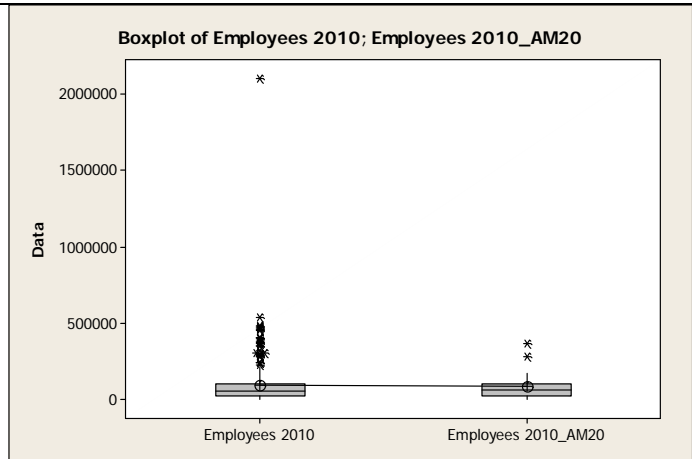
5.5.5 Variável Employees





A amostra que mais representou a população total foi a de maior tamanho (100), cuja média é a mais próxima da população total, sendo, portanto a mais representativa.

O Box-Plot do Teste para as amostras confirma a hipótese nula de diferença de médias, para todas as amostras. As amostras 20, 50 e 100 são estatisticamente iguais, assumindo um intervalo de confiança igual a 95%, por que o valor de P é superior a 0,10.

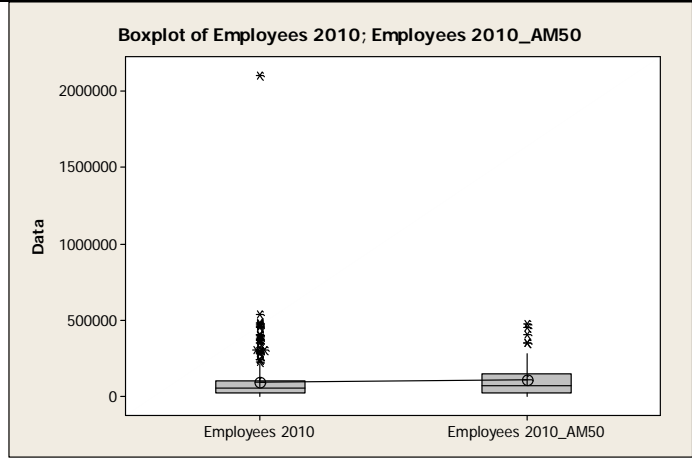


**Two-Sample T-Test and CI:
Employees 2010; Employees
2010_AM20**

Two-sample T for Employees 2010 vs Employees 2010_AM20

	StDev	SE Mean	N	Mean
Employees 2010	162659	10166	256	97210
Employees 2010_AM20	93982	21015	20	83153

Difference = mu (Employees 2010) - mu (Employees 2010_AM20)
 Estimate for difference: 14057
 95% CI for difference: (-33763; 61876)
 T-Test of difference = 0 (vs not =): T-Value = 0.60 P-Value = 0.552 DF = 28

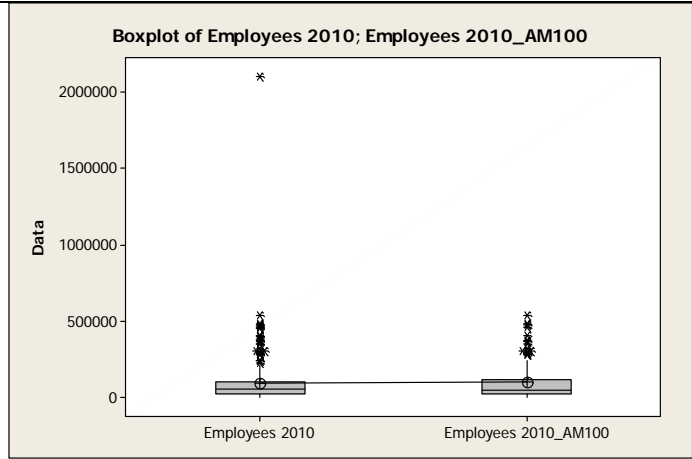


**Two-Sample T-Test and CI:
Employees 2010; Employees
2010_AM50**

Two-sample T for Employees 2010 vs Employees 2010_AM50

	StDev	SE Mean	N	Mean
Employees 2010	162659	10166	256	97210
Employees 2010_AM50	121392	17167	50	106373

Difference = mu (Employees 2010) - mu (Employees 2010_AM50)
 Estimate for difference: -9163
 95% CI for difference: (-48820; 30493)
 T-Test of difference = 0 (vs not =): T-Value = -0.46 P-Value = 0.647 DF = 87



**Two-Sample T-Test and CI:
Employees 2010; Employees
2010_AM100**

Two-sample T for Employees 2010 vs Employees 2010_AM100

	StDev	SE Mean	N	Mean
Employees 2010	162659	10166	256	97210
Employees 2010_AM100	126380	12638	100	104571

Difference = mu (Employees 2010) - mu (Employees 2010_AM100)

	Estimate for difference: -7361 95% CI for difference: (-39318; 24596) T-Test of difference = 0 (vs not =): T-Value = -0.45 P-Value = 0.650 DF = 231
--	--

Podemos também observar pelo teste One-way Anova que, conforme o tamanho das amostras aumenta, o desvio padrão diminui.

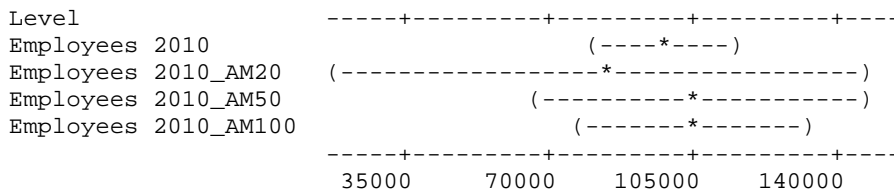
One-way ANOVA: Employees 20; Employees 20; Employees 20; Employees 20

Source	DF	SS	MS	F	P
Factor	3	11611466560	3870488853	0.18	0.912
Error	422	9.21786E+12	21843268814		
Total	425	9.22947E+12			

S = 147795 R-Sq = 0.13% R-Sq(adj) = 0.00%

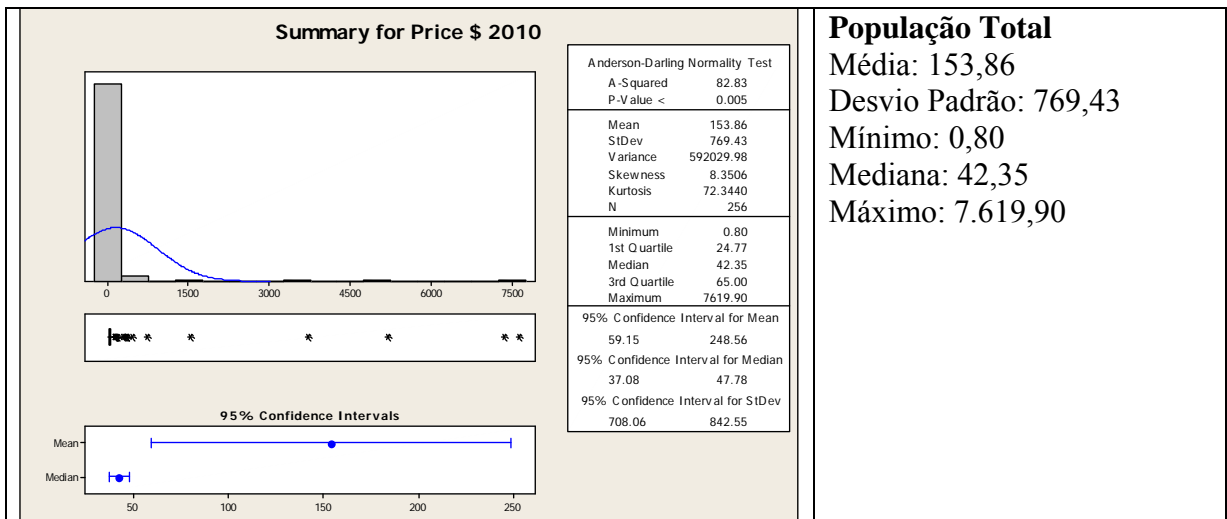
Level	N	Mean	StDev
Employees 2010	256	97210	162659
Employees 2010_AM20	20	83153	93982
Employees 2010_AM50	50	106373	121392
Employees 2010_AM100	100	104571	126380

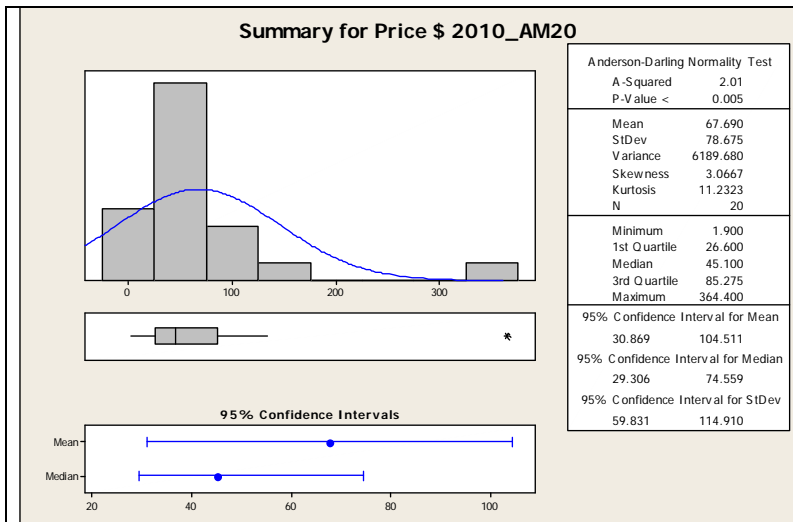
Individual 95% CIs For Mean Based on Pooled StDev



Pooled StDev = 147795

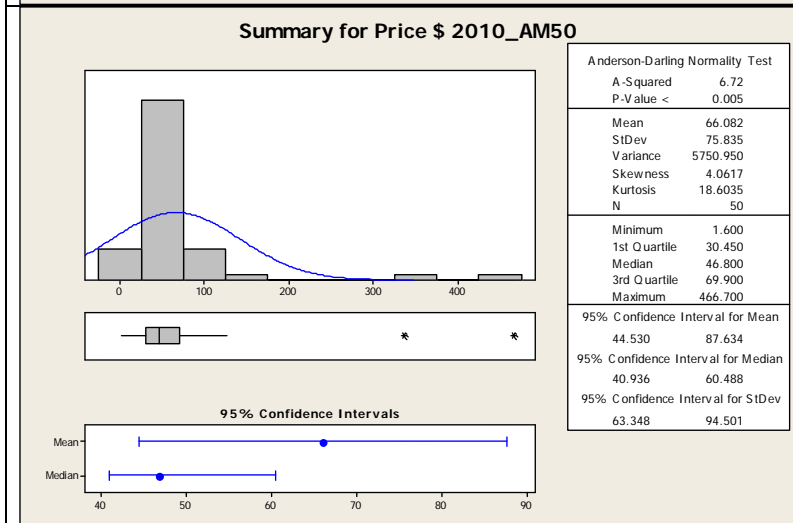
5.5.6 Variável Price \$





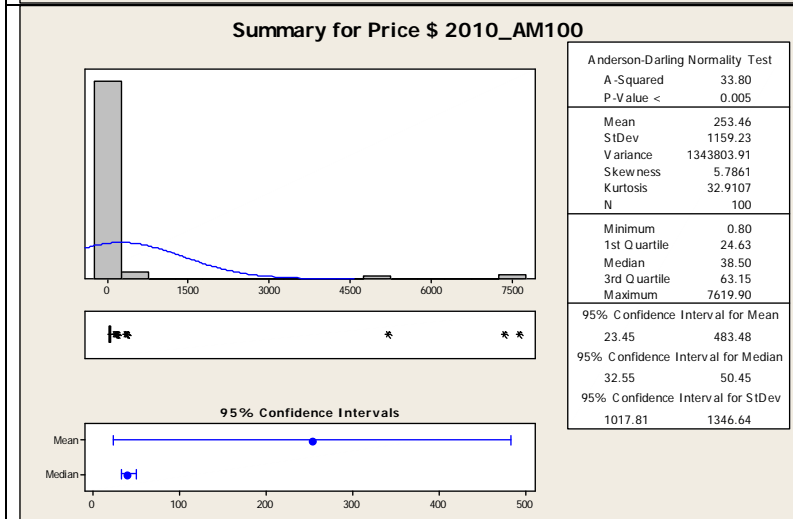
Amostra Tamanho 20

Média: 67,690
 Desvio Padrão: 78,675
 Mínimo: 1,90
 Mediana: 45,10
 Máximo: 364,40



Amostra Tamanho 50

Média: 66,082
 Desvio Padrão: 75,835
 Mínimo: 1,60
 Mediana: 46,80
 Máximo: 466,70

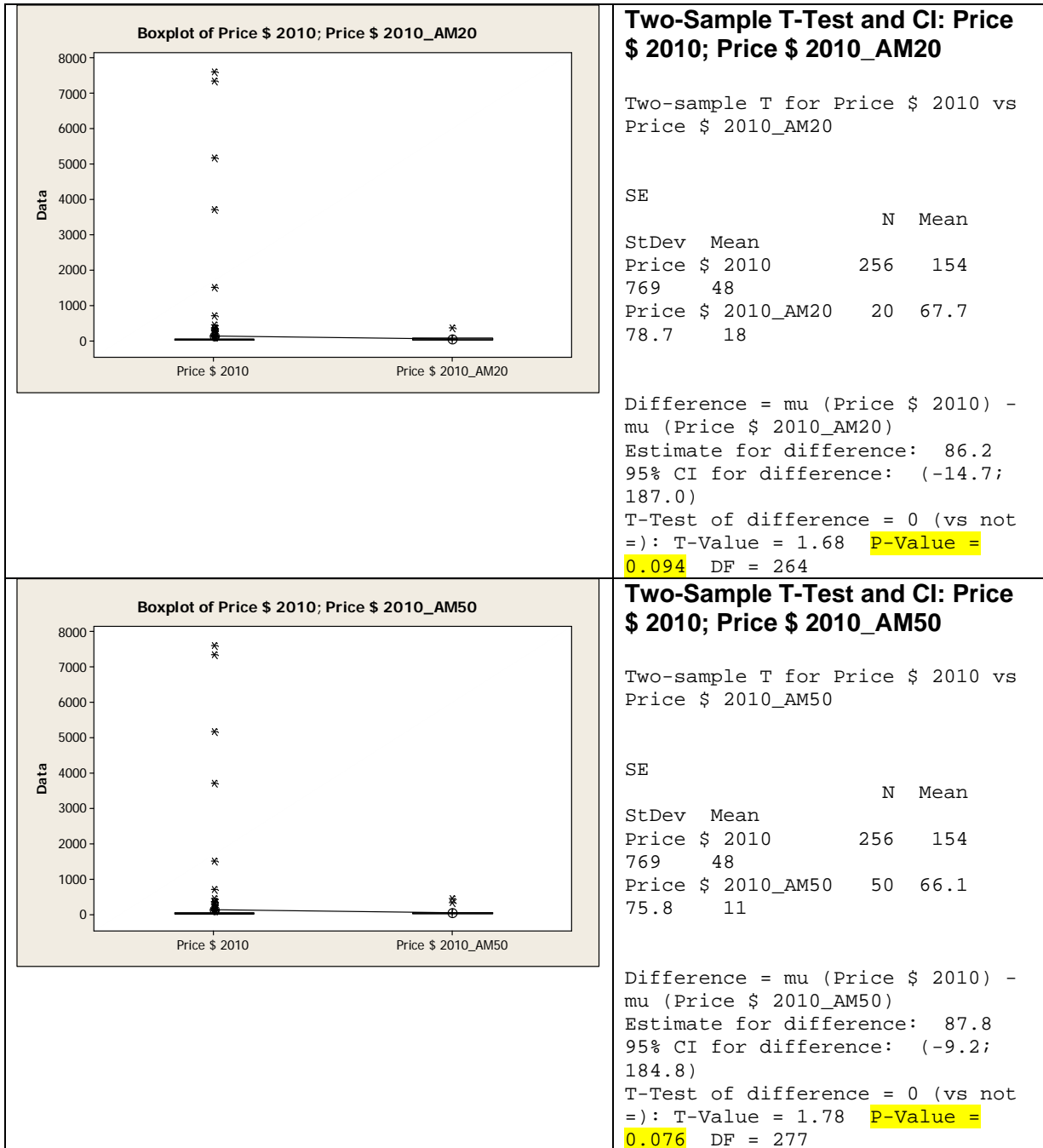


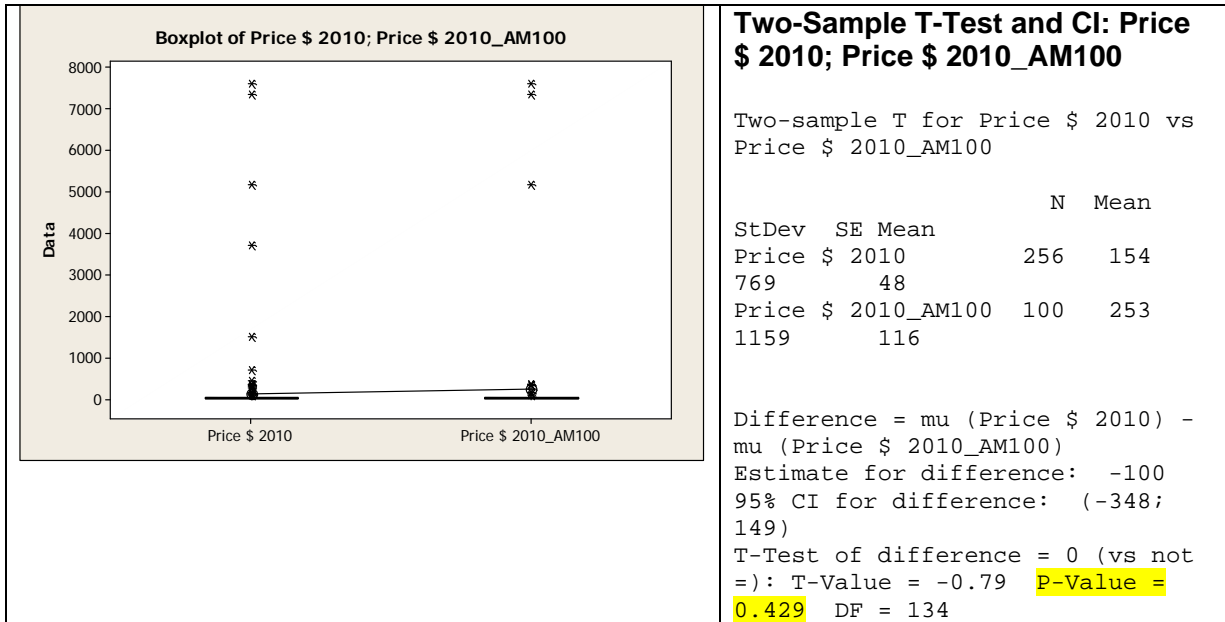
Amostra Tamanho 100

Média: 253,46
 Desvio Padrão: 1.159,23
 Mínimo: 0,80
 Mediana: 38,50
 Máximo: 7.619,90

A amostra que mais representou a população total foi a de tamanho 20, cuja média é a mais próxima da população total, sendo, portanto a mais representativa. Talvez as amostras de maior tamanho tenham contido valores aleatórios muito dispersos que tenham distorcido os seus resultados.

O Box-Plot do Teste para as amostras confirma a hipótese nula de diferença de médias somente para a amostra de tamanho 100, que é estatisticamente igual, assumindo um intervalo de confiança igual a 95%, por que o valor de P é superior a 0,10. Já as amostras 20 e 50 são estatisticamente diferente.





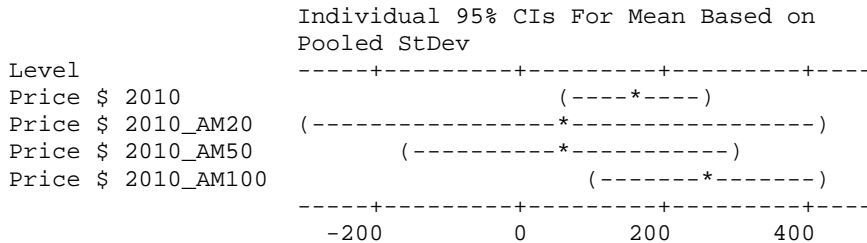
Podemos também observar pelo teste One-way Anova que, conforme o tamanho das amostras aumenta, o desvio padrão diminui, porém neste caso, a amostra maior não foi a que possui média mais próxima da população total.

One-way ANOVA: Price \$ 2010; Price \$ 2010; Price \$ 2010; Price \$ 2010

Source	DF	SS	MS	F	P
Factor	3	1491100	497033	0.74	0.530
Error	422	284403633	673942		
Total	425	285894733			

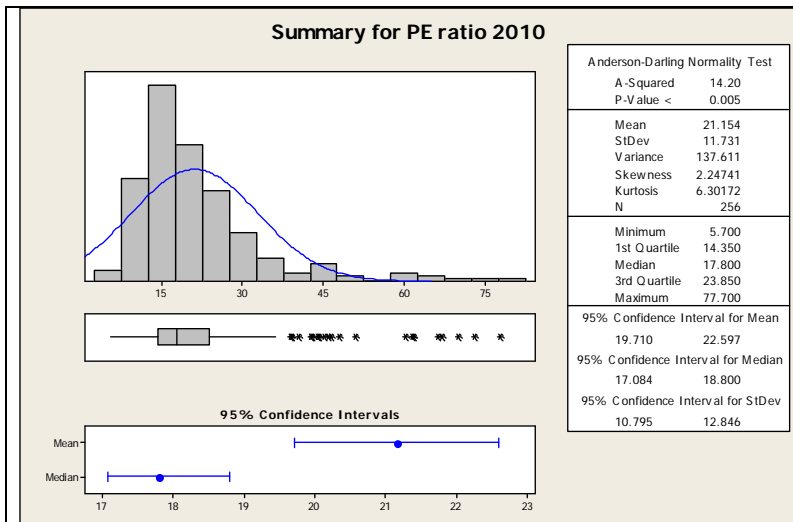
S = 820.9 R-Sq = 0.52% R-Sq(adj) = 0.00%

Level	N	Mean	StDev
Price \$ 2010	256	153.9	769.4
Price \$ 2010_AM20	20	67.7	78.7
Price \$ 2010_AM50	50	66.1	75.8
Price \$ 2010_AM100	100	253.5	1159.2

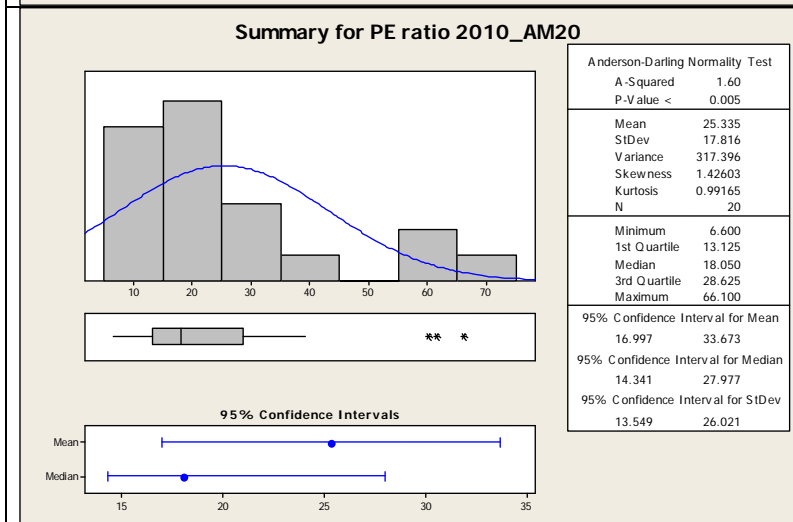


Pooled StDev = 820.9

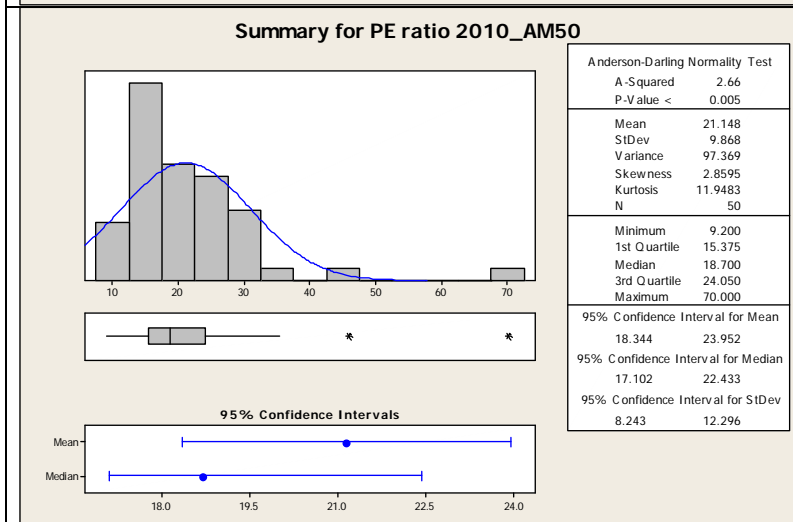
5.5.7 Variável PE ratio



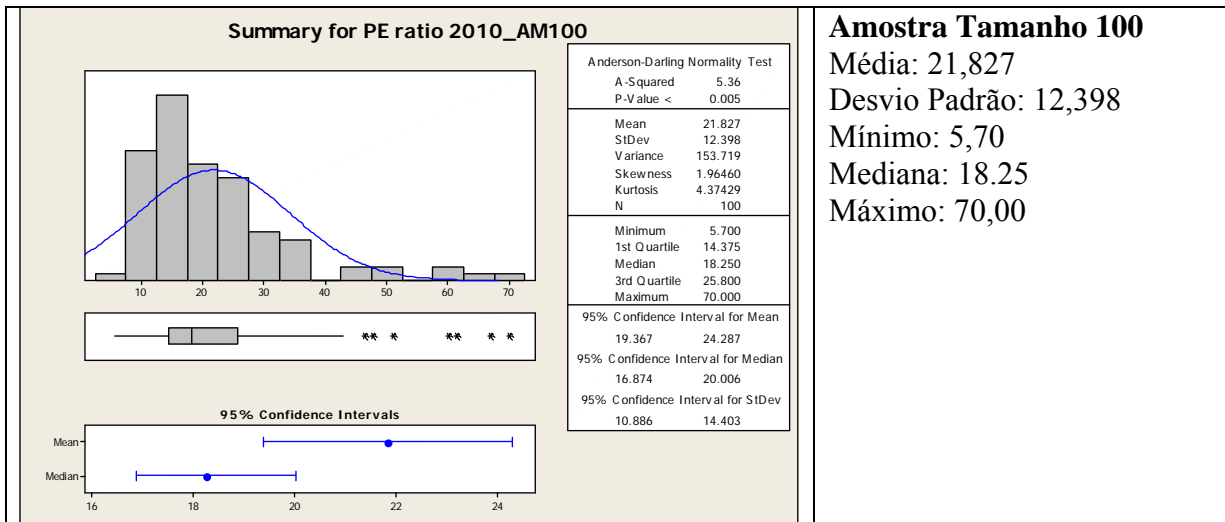
População Total
 Média: 21,154
 Desvio Padrão: 11,731
 Mínimo: 5,70
 Mediana: 17,80
 Máximo: 77,70



Amostra Tamanho 20
 Média: 25,335
 Desvio Padrão: 17,816
 Mínimo: 6,60
 Mediana: 18,05
 Máximo: 66,10



Amostra Tamanho 50
 Média: 21,148
 Desvio Padrão: 9,868
 Mínimo: 9,20
 Mediana: 18,70
 Máximo: 70,00



Amostra Tamanho 100

Média: 21,827

Desvio Padrão: 12,398

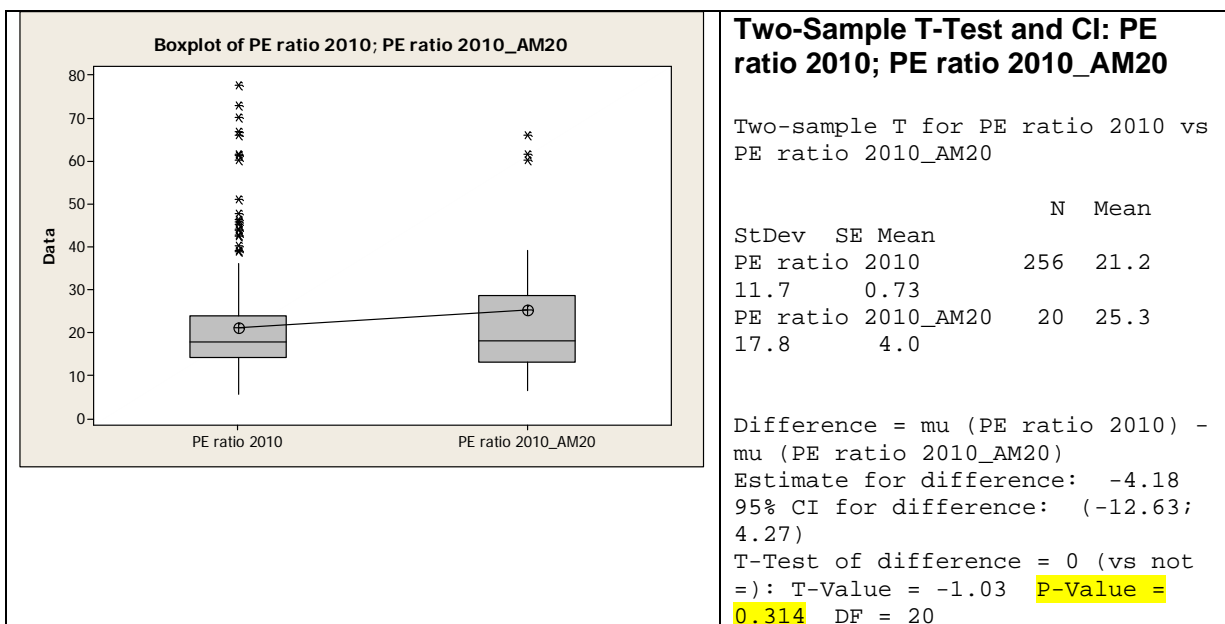
Mínimo: 5,70

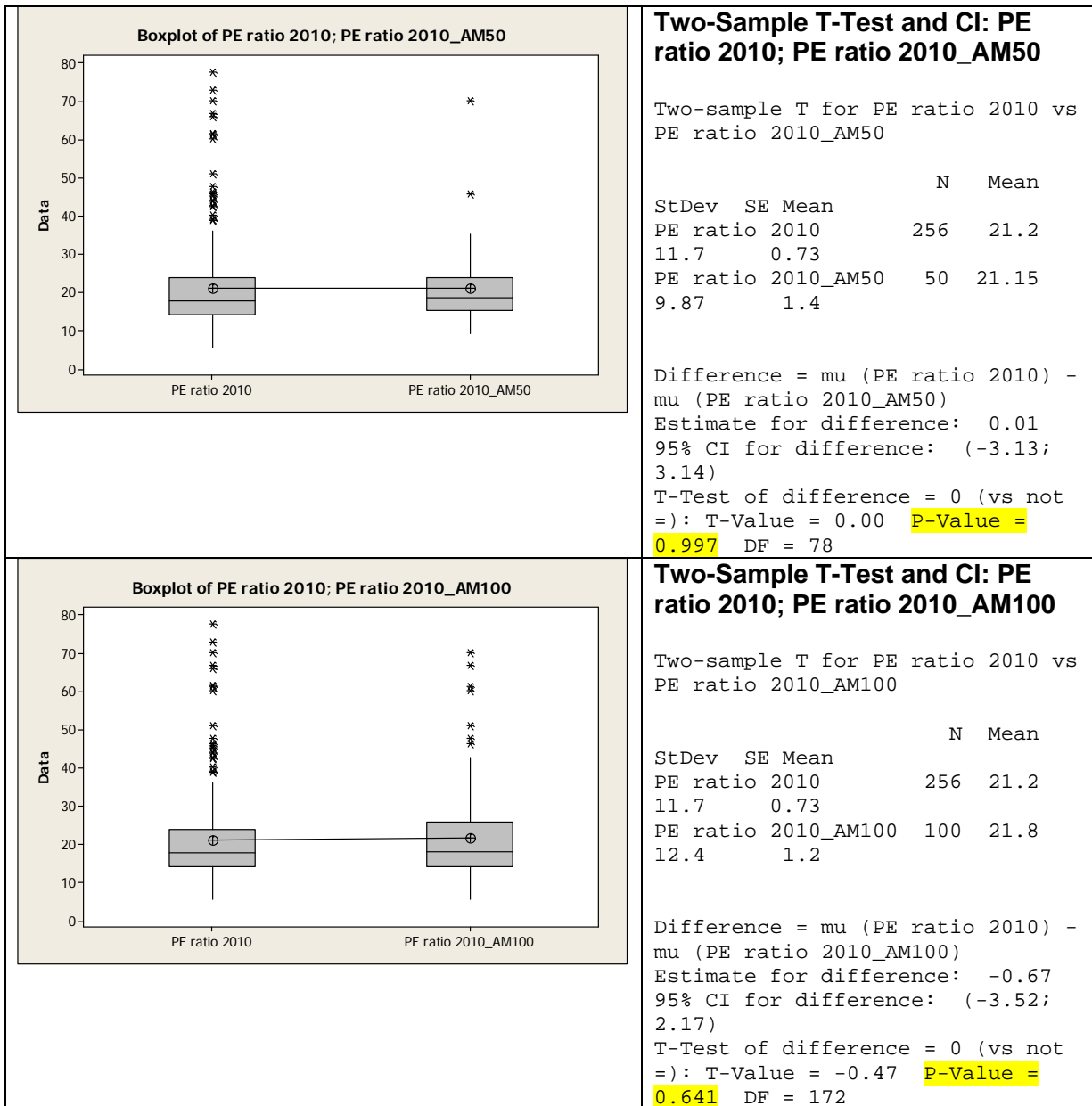
Mediana: 18.25

Máximo: 70,00

A amostra que mais representou a população total foi a de tamanho 50, cuja média é a mais próxima da população total, sendo, portanto a mais representativa. Talvez a amostra de maior tamanho tenha contido valores aleatórios muito dispersos que tenham distorcido os seus resultados.

O Box-Plot do Teste para as amostras confirma a hipótese nula de diferença de médias, para todas as amostras. As amostras 20, 50 e 100 são estatisticamente iguais, assumindo um intervalo de confiança igual a 95%, por que o valor de P é superior a 0,10.





Podemos também observar pelo teste One-way Anova que, conforme o tamanho das amostras aumenta, o desvio padrão diminui, porém neste caso, a amostra maior não foi a que possui média mais próxima da população total.

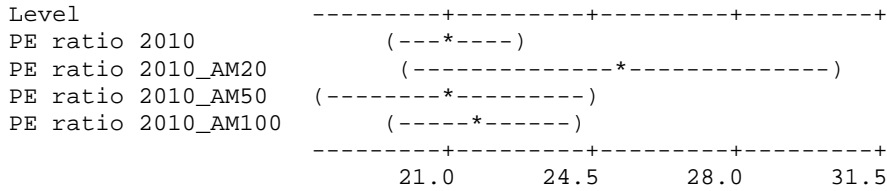
One-way ANOVA: PE ratio 201; PE ratio 201; PE ratio 201; PE ratio 201

Source	DF	SS	MS	F	P
Factor	3	342	114	0.79	0.502
Error	422	61111	145		
Total	425	61453			

S = 12.03 R-Sq = 0.56% R-Sq(adj) = 0.00%

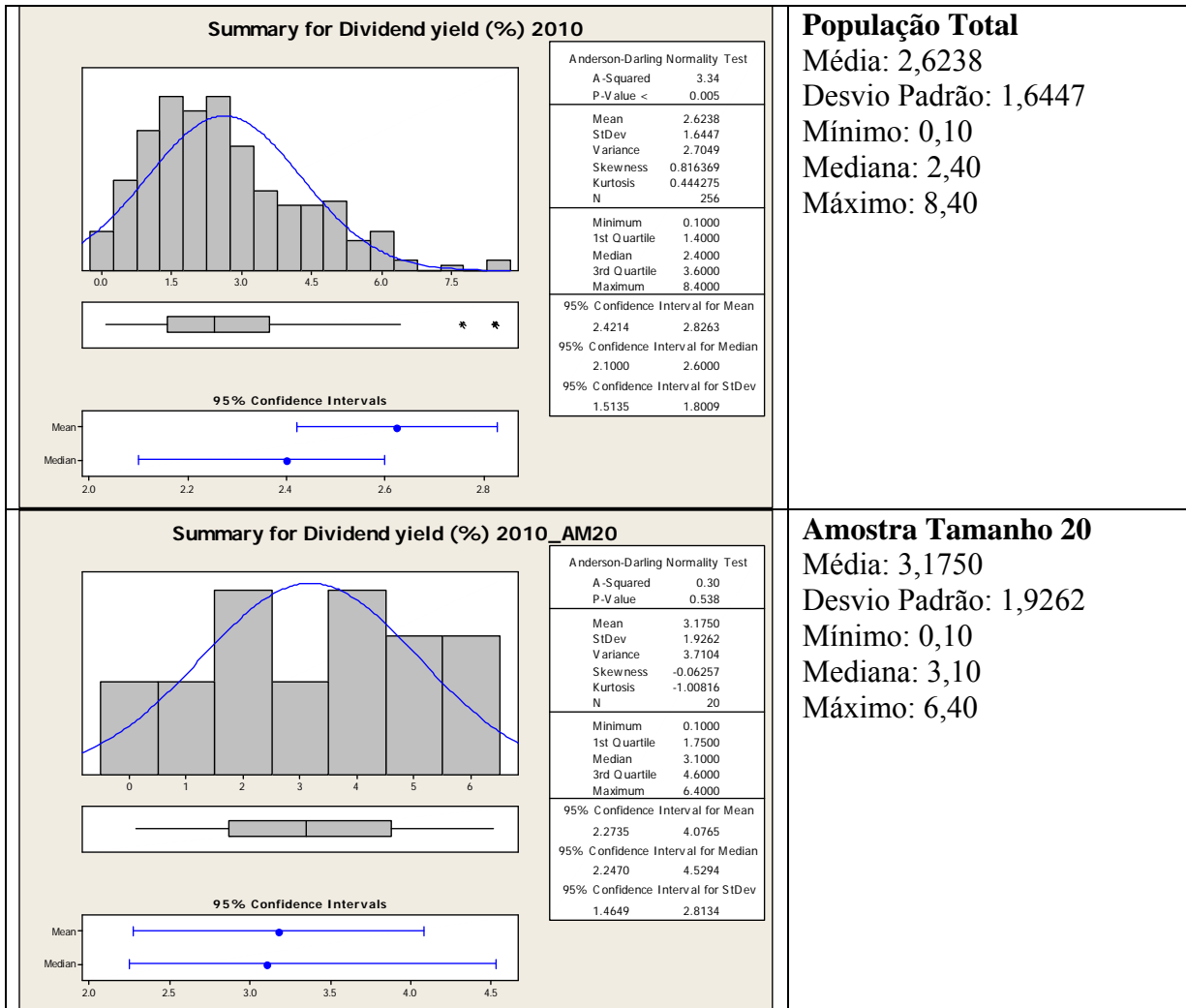
Level	N	Mean	StDev
PE ratio 2010	256	21.15	11.73
PE ratio 2010_AM20	20	25.34	17.82
PE ratio 2010_AM50	50	21.15	9.87
PE ratio 2010_AM100	100	21.83	12.40

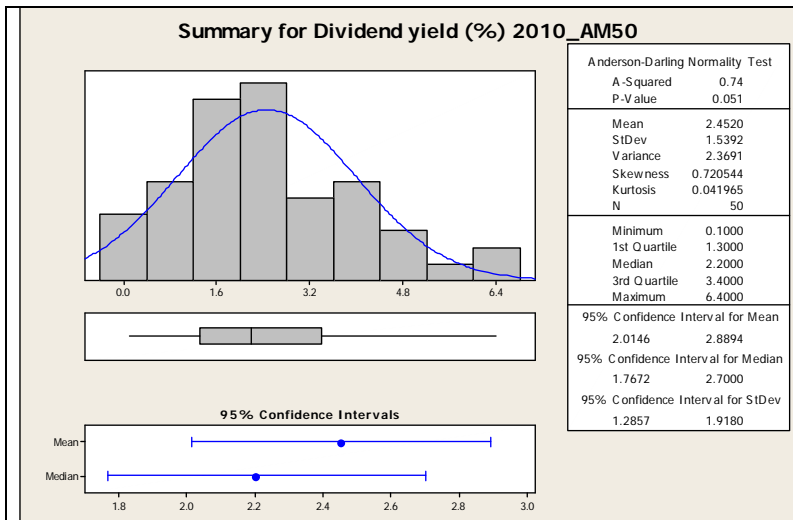
Individual 95% CIs For Mean Based on Pooled StDev



Pooled StDev = 12.03

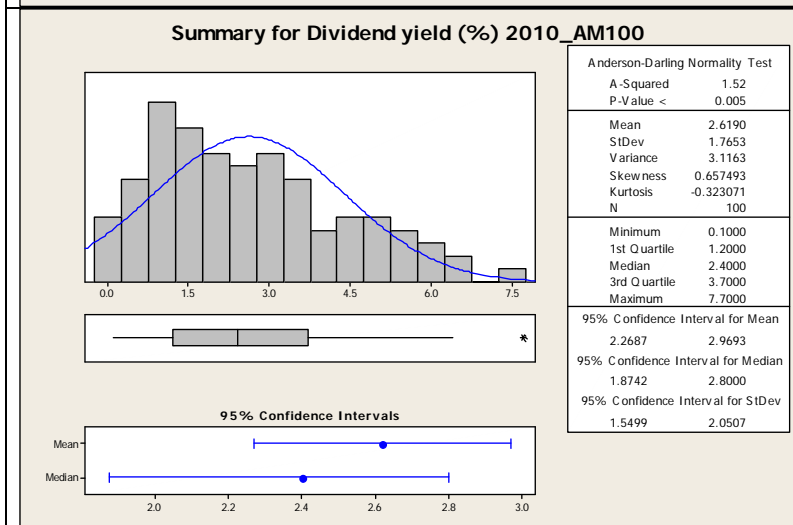
5.5.8 Variável Dividend Yield (%)





Amostra Tamanho 50

Média: 2,4520
 Desvio Padrão: 1,5392
 Mínimo: 0,10
 Mediana: 2,20
 Máximo: 6,40

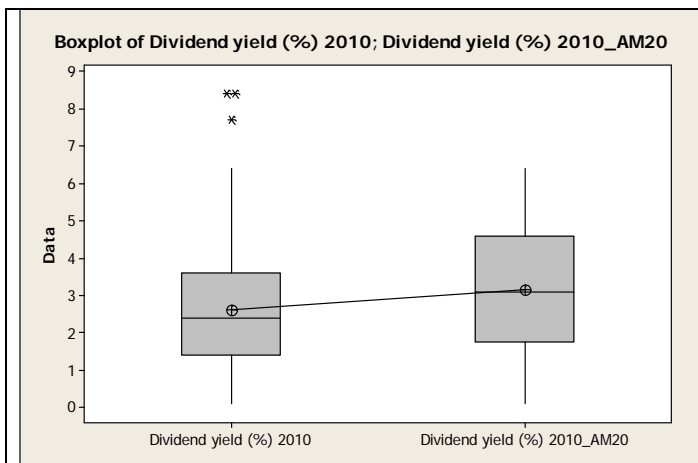


Amostra Tamanho 100

Média: 2,6190
 Desvio Padrão: 1,7653
 Mínimo: 0,10
 Mediana: 2,40
 Máximo: 7,70

A amostra que mais representou a população total foi a de maior tamanho (100), cuja média é a mais próxima da população total, sendo, portanto a mais representativa.

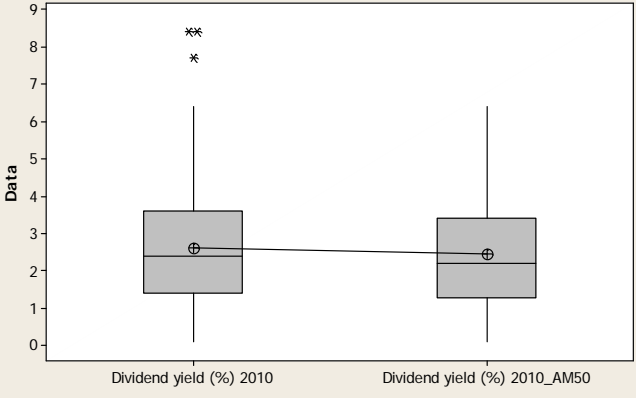
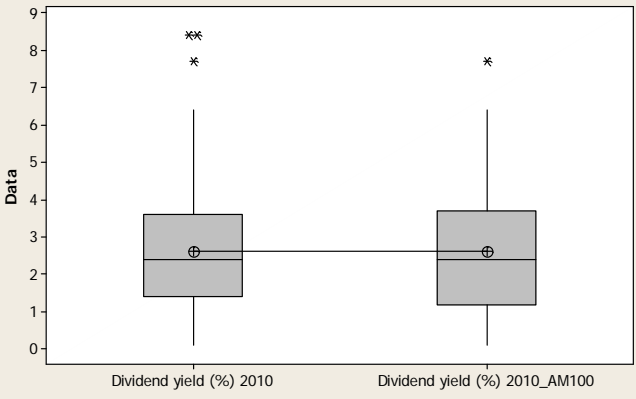
O Box-Plot do Teste para as amostras confirma a hipótese nula de diferença de médias, para todas as amostras. As amostras 20, 50 e 100 são estatisticamente iguais, assumindo um intervalo de confiança igual a 95%, por que o valor de P é superior a 0,10.



Two-Sample T-Test and CI: Dividend yield (%) 2010; Dividend yield (%) 2010_AM20

Two-sample T for Dividend yield (%) 2010 vs Dividend yield (%) 2010_AM20

	Mean	StDev	SE Mean	N
Dividend yield (%) 2010	2.62	1.64	0.10	256
Dividend yield (%) 2010_	3.18	1.93	0.43	20

	<p>Difference = mu (Dividend yield (%) 2010) - mu (Dividend yield (%) 2010_AM20) Estimate for difference: -0.551 95% CI for difference: (-1.472; 0.370) T-Test of difference = 0 (vs not =): T-Value = -1.24 P-Value = 0.227 DF = 21</p>															
<p>Boxplot of Dividend yield (%) 2010; Dividend yield (%) 2010_AM50</p> 	<p>Two-Sample T-Test and CI: Dividend yield (%) 2010; Dividend yield (%) 2010_AM50</p> <p>Two-sample T for Dividend yield (%) 2010 vs Dividend yield (%) 2010_AM50</p> <table border="1" data-bbox="928 683 1404 840"> <thead> <tr> <th></th> <th>Mean</th> <th>StDev</th> <th>SE Mean</th> <th>N</th> </tr> </thead> <tbody> <tr> <td>Dividend yield (%) 2010</td> <td>2.62</td> <td>1.64</td> <td>0.10</td> <td>256</td> </tr> <tr> <td>Dividend yield (%) 2010_</td> <td>2.45</td> <td>1.54</td> <td>0.22</td> <td>50</td> </tr> </tbody> </table> <p>Difference = mu (Dividend yield (%) 2010) - mu (Dividend yield (%) 2010_AM50) Estimate for difference: 0.172 95% CI for difference: (-0.308; 0.652) T-Test of difference = 0 (vs not =): T-Value = 0.71 P-Value = 0.478 DF = 72</p>		Mean	StDev	SE Mean	N	Dividend yield (%) 2010	2.62	1.64	0.10	256	Dividend yield (%) 2010_	2.45	1.54	0.22	50
	Mean	StDev	SE Mean	N												
Dividend yield (%) 2010	2.62	1.64	0.10	256												
Dividend yield (%) 2010_	2.45	1.54	0.22	50												
<p>Boxplot of Dividend yield (%) 2010; Dividend yield (%) 2010_AM100</p> 	<p>Two-Sample T-Test and CI: Dividend yield (%) 2010; Dividend yield (%) 2010_</p> <p>Two-sample T for Dividend yield (%) 2010 vs Dividend yield (%) 2010_AM100</p> <table border="1" data-bbox="928 1388 1404 1545"> <thead> <tr> <th></th> <th>Mean</th> <th>StDev</th> <th>SE Mean</th> <th>N</th> </tr> </thead> <tbody> <tr> <td>Dividend yield (%) 2010</td> <td>2.62</td> <td>1.64</td> <td>0.10</td> <td>256</td> </tr> <tr> <td>Dividend yield (%) 2010_</td> <td>2.62</td> <td>1.77</td> <td>0.18</td> <td>100</td> </tr> </tbody> </table> <p>Difference = mu (Dividend yield (%) 2010) - mu (Dividend yield (%) 2010_AM100) Estimate for difference: 0.005 95% CI for difference: (-0.398; 0.408) T-Test of difference = 0 (vs not =): T-Value = 0.02 P-Value = 0.981 DF = 169</p>		Mean	StDev	SE Mean	N	Dividend yield (%) 2010	2.62	1.64	0.10	256	Dividend yield (%) 2010_	2.62	1.77	0.18	100
	Mean	StDev	SE Mean	N												
Dividend yield (%) 2010	2.62	1.64	0.10	256												
Dividend yield (%) 2010_	2.62	1.77	0.18	100												

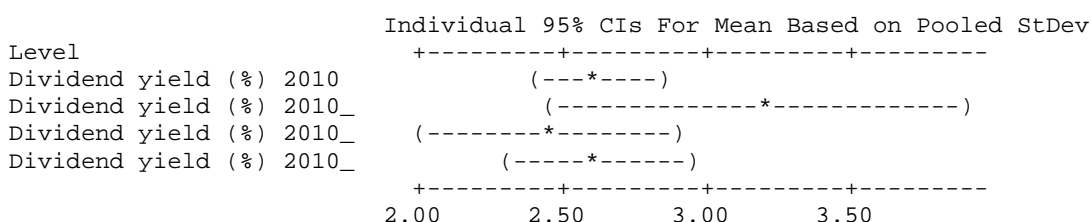
Podemos também observar pelo teste One-way Anova que, conforme o tamanho das amostras aumenta, o desvio padrão diminui.

One-way ANOVA: Dividend yie; Dividend yie; Dividend yie; Dividend yie

Source	DF	SS	MS	F	P
Factor	3	7.55	2.52	0.90	0.443
Error	422	1184.84	2.81		
Total	425	1192.39			

S = 1.676 R-Sq = 0.63% R-Sq(adj) = 0.00%

Level	N	Mean	StDev
Dividend yield (%) 2010	256	2.624	1.645
Dividend yield (%) 2010_	20	3.175	1.926
Dividend yield (%) 2010_	50	2.452	1.539
Dividend yield (%) 2010_	100	2.619	1.765



Pooled StDev = 1.676

5.5.9 Análise de Correlação e Dendogramas

População Total

	Market Value \$m	Turnover \$m	Net Income \$m	Total Assets \$m	Employees	Price \$	PE ratio
Turnover \$m	0,654						
	0,000						
Net Income \$m	0,870	0,658					
	0,000	0,000					
Total Assets \$m	0,612	0,679	0,608				
	0,000	0,000	0,000				
Employees	0,366	0,639	0,347	0,306			
	0,000	0,000	0,000	0,000			
Price \$	-0,076	-0,032	-0,052	-0,043	-0,066		
	0,228	0,607	0,405	0,498	0,293		
PE ratio	-0,124	-0,111	-0,347	-0,134	0,005	-0,082	
	0,047	0,076	0,000	0,032	0,942	0,190	
Dividend yield	0,137	0,194	0,235	0,270	0,006	-0,098	-0,328
	0,028	0,002	0,000	0,000	0,920	0,116	0,000

Amostra Tamanho 20

	Market Value \$m	Turnover \$m	Net Income \$m	Total Assets \$m	Employees	Price \$	PE ratio

Turnover \$m	0,691						
	0,001						
Net Income \$m	0,952	0,571					
	0,000	0,009					
Total Assets \$m	0,740	0,899	0,689				
	0,000	0,000	0,001				
Employees	0,460	0,580	0,389	0,752			
	0,041	0,007	0,090	0,000			
Price \$	-0,264	-0,176	-0,228	-0,173	-0,116		
	0,261	0,457	0,333	0,466	0,627		
PE ratio	-0,486	-0,160	-0,575	-0,318	-0,153	-0,062	
	0,030	0,500	0,008	0,172	0,520	0,796	
Dividend yield	0,651	0,433	0,709	0,565	0,293	-0,295	-0,641
	0,002	0,057	0,000	0,009	0,210	0,206	0,002

Amostra Tamanho 50

	Market Value \$m	Turnover \$m	Net Income \$m	Total Assets \$m	Employees	Price \$	PE ratio
Turnover \$m	0,387						
	0,006						
Net Income \$m	0,783	0,321					
	0,000	0,023					
Total Assets \$m	0,633	0,700	0,594				
	0,000	0,000	0,000				
Employees	0,123	0,650	0,057	0,295			
	0,396	0,000	0,694	0,038			
Price \$	-0,072	-0,123	-0,017	-0,119	-0,224		
	0,617	0,395	0,905	0,409	0,118		
PE ratio	-0,045	0,272	-0,344	-0,034	0,328	-0,196	
	0,756	0,056	0,014	0,815	0,020	0,173	
Dividend yield	0,338	0,388	0,470	0,587	0,206	-0,036	-0,065
	0,016	0,005	0,001	0,000	0,150	0,804	0,654

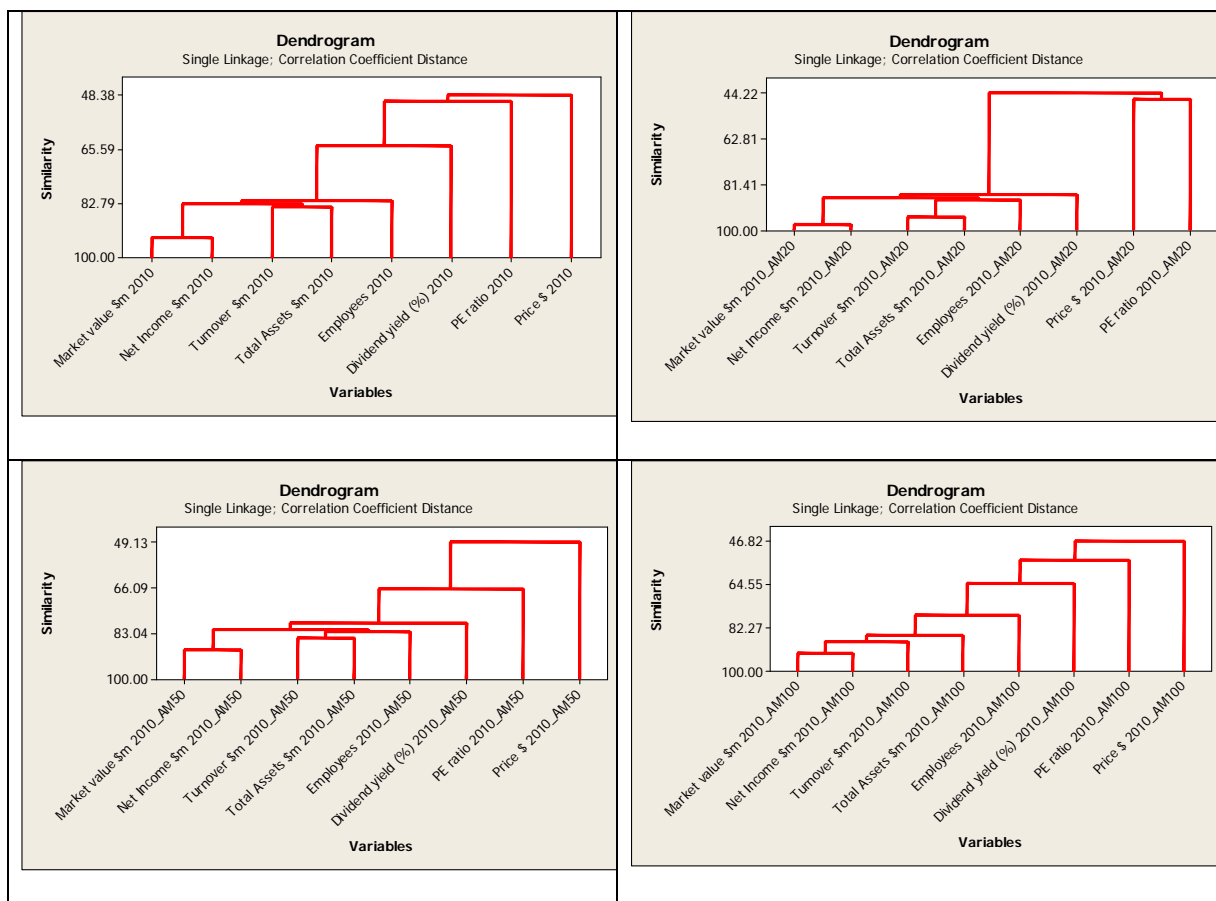
Amostra Tamanho 100

	Market Value \$m	Turnover \$m	Net Income \$m	Total Assets \$m	Employees	Price \$	PE ratio
Turnover \$m	0,757						
	0,000						
Net Income \$m	0,855	0,724					
	0,000	0,000					
Total Assets \$m	0,625	0,704	0,590				

	0,000	0,000	0,000				
Employees	0,439	0,539	0,390	0,425			
	0,000	0,000	0,000	0,000			
Price \$	-0,120	-0,093	-0,081	-0,066	-0,134		
	0,235	0,358	0,422	0,512	0,184		
PE ratio	-0,209	-0,138	-0,382	-0,160	0,094	-0,147	
	0,037	0,172	0,000	0,112	0,354	0,145	
Dividend yield	0,228	0,275	0,274	0,258	0,030	-0,121	-0,377
	0,023	0,006	0,006	0,010	0,766	0,232	0,000

Comentários:

- As variáveis nem sempre mantiveram o grau de correlação, na amostra aleatória, porém a correlação vai melhorando conforme o maior tamanho da amostra.



Comentários:

- O dendrograma das amostras aleatórias nem sempre manteve a mesma disposição das variáveis quando comparado ao dendrograma da população, porém conforme o tamanho da amostra aumenta, a disposição se torna cada vez mais parecida com a população.

5.5.10 Considerações

Era esperado que à medida que o tamanho da amostra aleatória aumenta, os valores da análise exploratória de dados também aumenta. Na maior parte dos casos foi possível verificar esta premissa, pela comparação dos valores de média, desvio padrão, mediana, mínimo e máximo entre os dados da população e da amostra, nas variáveis analisadas.

No caso de P-Value, podemos observar que a variável Price \$ apresentou valores abaixo de 0,10 nas amostras tamanho 20 e 50, portanto são estatisticamente diferentes da população. No caso das variáveis Market Value \$m, Net Income \$m, Total Assets \$m, Employees, Price \$ e Divident Yield (%), o P-Value é maior conforme o tamanho da amostra é maior. Nos demais casos, todas as amostras são estatisticamente iguais à população, porém nem sempre o P-Value é maior conforme o tamanho da amostra é maior.

P-Value do Teste – T para três Amostras

Variáveis	Amostra 20 Aleatória	Amostra 50 Estratificada	Amostra 100 Aleatória
Market Value \$m	0,692	0,123	0,818
Turnover \$m	0,556	0,294	0,363
Net Income \$m	0,484	0,137	0,721
Total Assets \$m	0,263	0,181	0,348
Employees	0,552	0,647	0,650
Price \$	0,094	0,076	0,429
PE ratio	0,314	0,997	0,641
Dividend Yield (%)	0,227	0,478	0,981

Na análise do teste One-way Anova, foi observado que o desvio padrão diminui conforme o tamanho da amostra aumenta e que a média das amostras vai se aproximando da média da população também conforme a amostra aumenta.

Na análise das correlações das variáveis, as amostras ficaram mais próximas das correlações da população total conforme o tamanho da amostra aumenta. Porém mesmo na amostra maior, de tamanho 100, algumas correlações não foram mantidas conforme a população.

No Dendograma as variáveis em geral mantiveram a disposição nas amostras. E foram ficando mais próximas da população conforme o tamanho da amostra aumenta.

Assim, podemos concluir que conforme o tamanho da amostra aumenta, melhor representa a população total. Porém no caso estudado, mesmo a maior amostra, de tamanho 100, não representa a população total, apresentando resultados diferentes em algumas análises.

5.6 Análise em 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. A análise de correlações e dendogramas realizada anteriormente já nos dão uma idéia de que a possibilidade de agrupamento existe pelos índices de correlação entre todas as variáveis:

Segue abaixo o resultado das análises dos componentes principais juntamente com o gráfico Scree Plot.

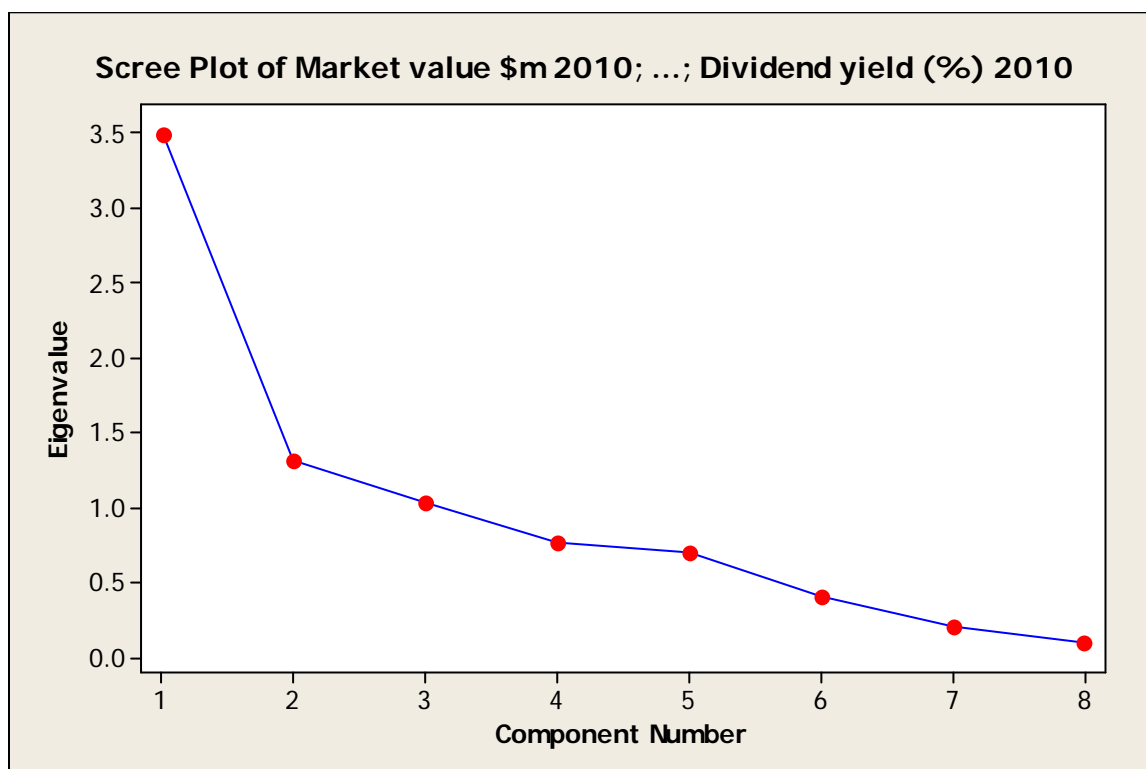
Principal Component Analysis: Market value; Turnover \$m ; Net Income \$; Total A

Eigenanalysis of the Correlation Matrix
255 cases used, 1 cases contain missing values

Eigenvalue	3.4884	1.3087	1.0378	0.7628	0.6946	0.4052	0.2034	0.0990
Proportion	0.436	0.164	0.130	0.095	0.087	0.051	0.025	0.012
Cumulative	0.436	0.600	0.729	0.825	0.912	0.962	0.988	1.000

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Market value \$m 2010	0.462	0.088	0.033	-0.422	-0.036	0.384	0.138
Turnover \$m 2010	0.467	0.191	0.053	0.237	0.071	-0.163	-0.807
Net Income \$m 2010	0.476	-0.092	0.064	-0.332	-0.174	0.298	0.026
Total Assets \$m 2010	0.432	-0.016	-0.009	-0.125	0.387	-0.707	0.383
Employees 2010	0.310	0.402	0.025	0.687	-0.260	0.147	0.424
Price \$ 2010	-0.048	-0.096	0.932	0.104	0.300	0.130	0.041
PE ratio 2010	-0.156	0.659	-0.164	-0.135	0.627	0.267	0.002
Dividend yield (%) 2010	0.180	-0.585	-0.308	0.372	0.513	0.357	0.029

Variable	PC8
Market value \$m 2010	-0.657
Turnover \$m 2010	-0.064
Net Income \$m 2010	0.728
Total Assets \$m 2010	0.028
Employees 2010	0.022
Price \$ 2010	0.002
PE ratio 2010	0.179
Dividend yield (%) 2010	-0.019



Pela análise dos detalhes e gráfico acima percebemos que se juntarmos as 6 variáveis em apenas 1 (PC1) teremos um proporção de 43,6%, com 2 (PC1 e PC2) chegamos a 60,0% com 3 (PC1, PC2 e PC3) chegamos a 72,9% e assim por diante. Isto é algo extremamente significativo, pois ao invés de trabalharmos com 8 variáveis poderíamos trabalhar com 3

(índice PC1, PC2 e PC3) que já explica 72,9% das variáveis. Se trabalharmos com 5 variáveis ao invés de 8, o nível de explicação chegará a 91,2%.

Pela análise dos valores / participação de cada variável no índice PC1, poderíamos denominá-lo índice complementado do valor de mercado. O índice PC2, poderíamos denominá-lo de contraste PE ratio versus rendimento de dividendos. Já o índice PC3, poderíamos denominá-lo de índice complementado do preço da ação.

5.7 Análise de Conglomerados

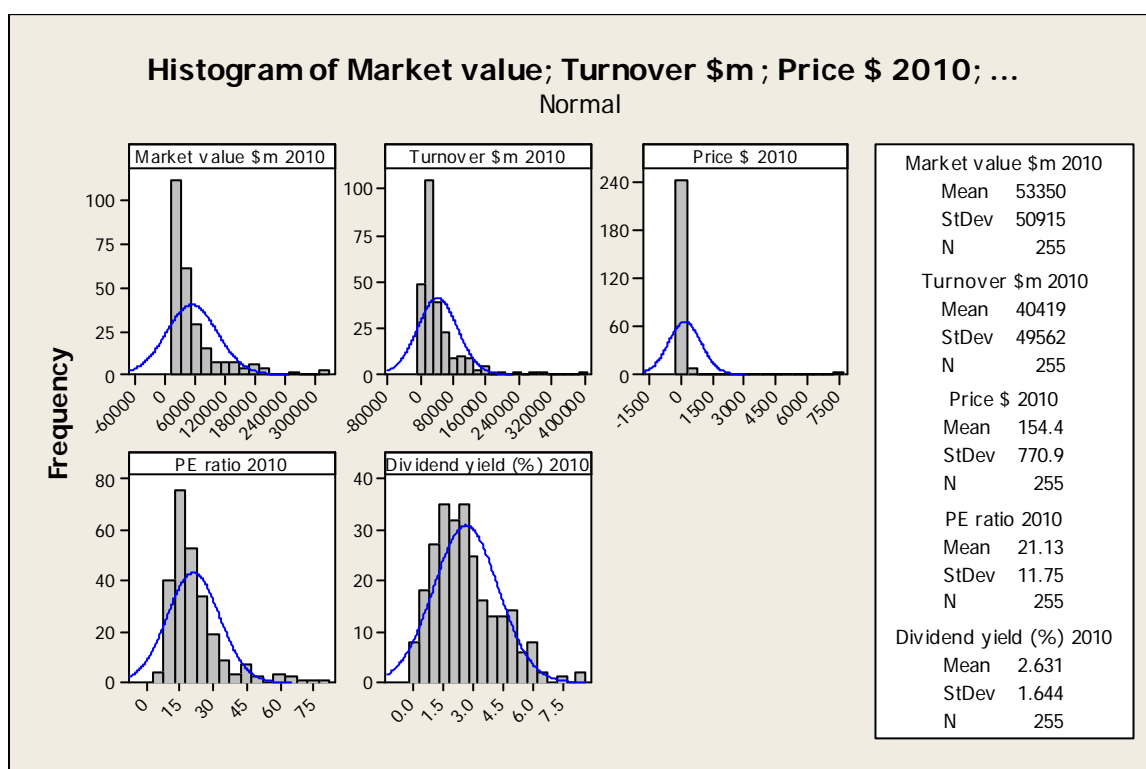
O objetivo deste tópico é efetuar uma análise de conglomerados através de uma pesquisa por amostragem de dados de indicadores relacionados ao Ranking das 500 Empresas mais valiosas do mundo. Para isso foi criada uma amostragem arbitrada de tamanho 60 e excluído posteriormente 3 outliers. Para viabilizar essa análise, a quantidade de variáveis foi reduzida para 5, sendo elas: Market Value \$m, Turnover \$m, Price \$, PE Ratio e Dividend Yield (%).

5.7.1 Estatística Descritiva / Pesquisa por Amostragem

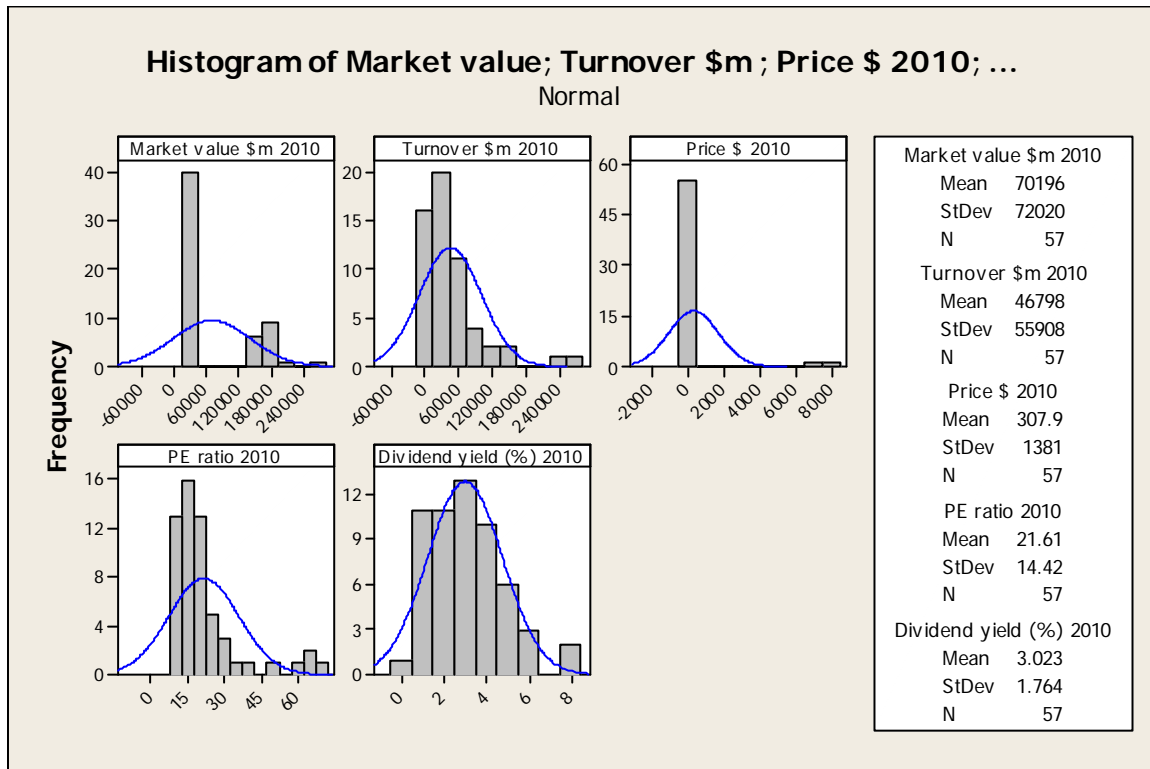
A pesquisa por amostragem foi feita em uma de 60 indivíduos inicialmente, e reduzida para 57, excluindo 3 outliers. Essa amostra foi arbitrada para viabilizar esta análise, sendo selecionados 20 empresas de maior valor, 20 empresas intermediárias e 20 empresas de menor valor.

Começamos com a análise das medidas e gráficos da estatística descritiva de cada variável da população total e por amostra.

Sumário População



No caso da população, nenhuma das variáveis apresenta curvas próximas à curva normal.



Para a amostra de 57 indivíduos observamos um resultado similar às curvas da população, isso demonstra que a amostra representa bem a população. Com exceção da variável Dividend Yield (%), que nessa amostra apresenta curva próxima à curva normal, ao contrário do observado na população.

5.7.2 Análise de Conglomerados

Cluster Analysis of Observations: Market value; Turnover \$m ; Price \$ 2010; ...

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	56	99.9221	248	53	54	2
2	55	99.8983	324	26	27	2
3	54	99.8791	385	40	46	2
4	53	99.7996	638	50	52	2
5	52	99.7691	735	41	50	3
6	51	99.7033	944	51	53	3
7	50	99.6801	1018	51	56	4
8	49	99.5813	1333	35	37	2
9	48	99.5757	1351	23	26	3
10	47	99.5742	1355	40	45	3
11	46	99.5616	1395	22	29	2
12	45	99.5198	1529	44	47	2
13	44	99.5025	1583	28	31	2
14	43	99.4498	1751	44	51	6

15	42	99.4392	1785	48	57	48	2
16	41	99.4246	1831	44	49	44	7
17	40	99.4169	1856	40	44	40	10
18	39	99.3950	1926	30	32	30	2
19	38	99.3650	2021	20	21	20	2
20	37	99.3346	2118	15	16	15	2
21	36	99.2943	2246	38	40	38	11
22	35	99.2529	2378	23	33	23	4
23	34	99.1128	2824	20	25	20	3
24	33	99.1033	2854	19	30	19	3
25	32	98.9832	3236	19	22	19	5
26	31	98.9809	3244	18	19	18	6
27	30	98.9232	3427	41	42	41	4
28	29	98.8495	3662	15	17	15	3
29	28	98.8403	3691	20	36	20	4
30	27	98.8134	3777	18	20	18	10
31	26	98.7049	4122	18	34	18	11
32	25	98.3856	5138	38	41	38	15
33	24	98.3767	5167	43	55	43	2
34	23	98.3607	5217	28	35	28	4
35	22	98.1455	5903	18	28	18	15
36	21	97.7384	7198	18	23	18	19
37	20	97.6952	7336	38	43	38	17
38	19	97.5471	7807	5	6	5	2
39	18	96.6146	10775	39	48	39	3
40	17	96.2411	11964	38	39	38	20
41	16	95.7148	13639	18	38	18	39
42	15	95.6184	13946	4	8	4	2
43	14	95.0270	15828	4	7	4	3
44	13	94.2228	18388	4	5	4	5
45	12	93.5668	20475	2	4	2	6
46	11	93.4960	20701	2	11	2	7
47	10	91.4186	27312	12	15	12	4
48	9	90.4463	30407	9	10	9	2
49	8	90.3590	30685	2	13	2	8
50	7	88.5729	36370	2	14	2	9
51	6	88.0396	38067	2	12	2	13
52	5	86.7700	42108	2	3	2	14
53	4	85.2074	47081	1	2	1	15
54	3	76.0844	76118	18	24	18	40
55	2	71.1086	91954	1	9	1	17
56	1	66.8502	105508	1	18	1	57

Final Partition

Number of clusters: 2

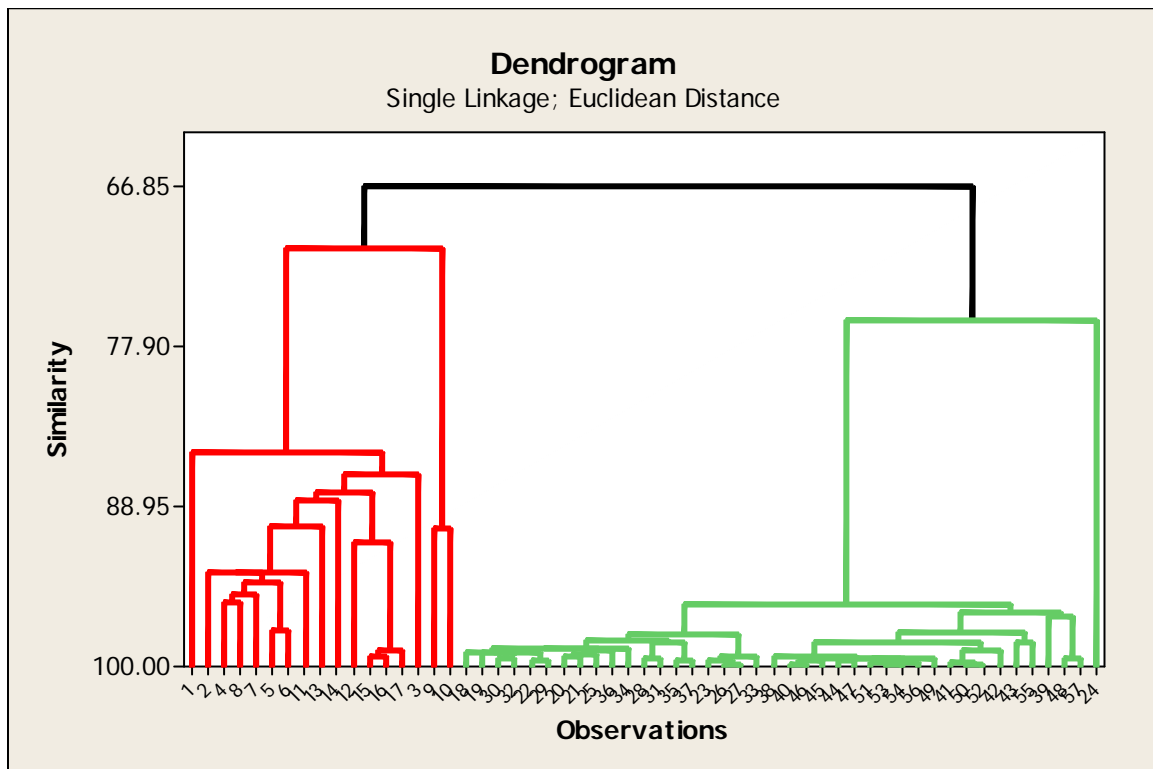
	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	17	9.45724E+10	60842.8	175039
Cluster2	40	1.99600E+10	17056.6	100716

Cluster Centroids

Variable	Cluster1	Cluster2	Grand centroid
Market value \$m 2010	176575	24984.3	70195.6
Turnover \$m 2010	103149	22848.3	46797.6
Price \$ 2010	49	418.0	307.9
PE ratio 2010	17	23.7	21.6
Dividend yield (%) 2010	4	2.8	3.0

Distances Between Cluster Centroids

	Cluster1	Cluster2
Cluster1	0	171546
Cluster2	171546	0



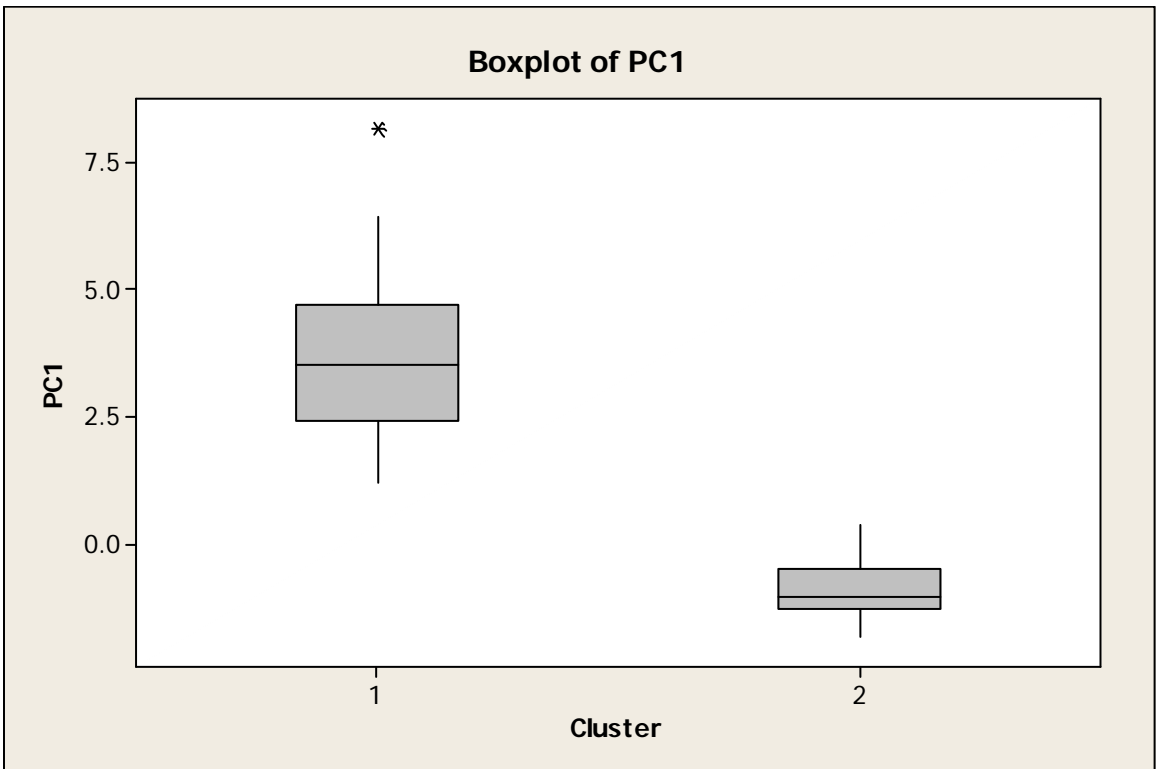
Podemos notar que para o número de conglomerados igual a 2 distribui de forma adequada a amostra de 57 indivíduos.

O conglomerado 1 possui 17 observações e o conglomerado 2 possui 40 observações.

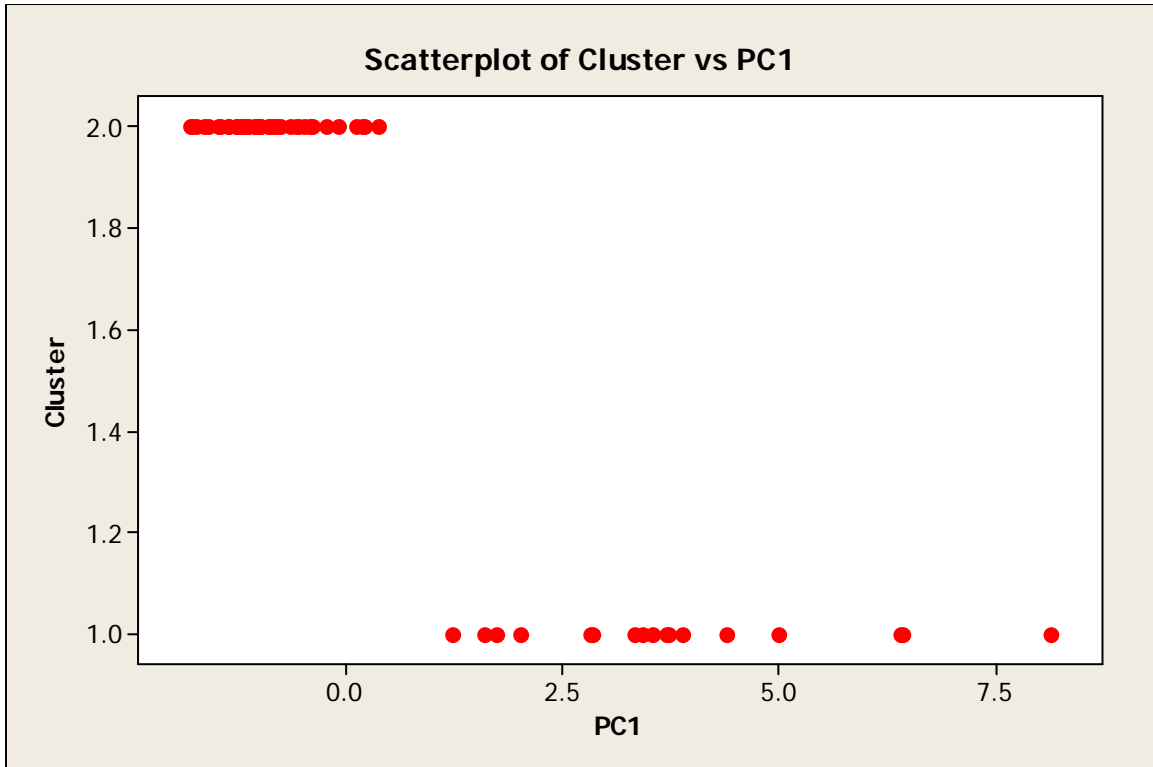
No conglomerado 1 se encontram as empresas mais valiosas e no conglomerado 2 se encontram as empresas intermediárias e menos valiosas. Foram realizados vários testes aumentando o número de conglomerados, porém as empresas intermediárias e menos valiosas não foram divididas.



O histograma mostra que o grupo mais importante é o 2.



O Box plot mostra que existe uma diferença visível entre os dois conglomerados.



Observamos nitidamente através do gráfico de dispersão a divisão entre os 2 conglomerados.

5.8 Análise Discriminante

O objetivo deste tópico é efetuar uma análise comparativa de médias, intervalos de confiança e regressões de dados de indicadores relacionados ao Ranking das 500 Empresas mais valiosas do mundo. Para isso foi criada uma amostragem arbitrada de tamanho 60 e excluído posteriormente 3 outliers. O principal propósito é comparar os conglomerados da amostra.

5.8.1 Comparação de Média, Análise de Variância e Intervalo de Confiança

One-way ANOVA: Market value \$m 2010 versus Cluster

Source	DF	SS	MS	F	P
Cluster	1	2.74145E+11	2.74145E+11	924.10	0.000
Error	55	16316476275	296663205		
Total	56	2.90462E+11			

S = 17224 R-Sq = 94.38% R-Sq(adj) = 94.28%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
1	17	176575	29315	(*-)
2	40	24984	8112	(*)

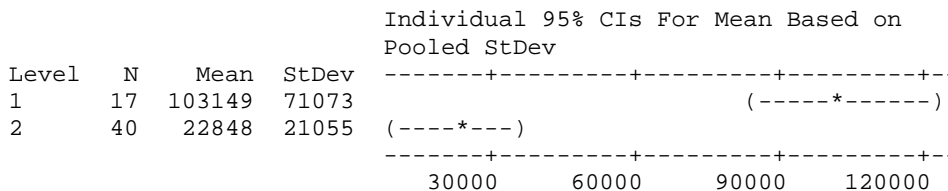
-----+-----+-----+-----+-----
50000 100000 150000 200000

Pooled StDev = 17224

One-way ANOVA: Turnover \$m 2010 versus Cluster

Source	DF	SS	MS	F	P
Cluster	1	76926169587	76926169587	43.12	0.000
Error	55	98110777945	1783832326		
Total	56	1.75037E+11			

S = 42235 R-Sq = 43.95% R-Sq(adj) = 42.93%

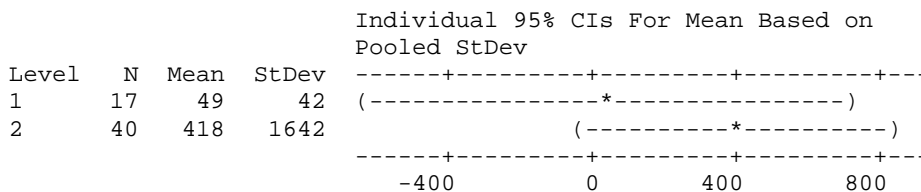


Pooled StDev = 42235

One-way ANOVA: Price \$ 2010 versus Cluster

Source	DF	SS	MS	F	P
Cluster	1	1624001	1624001	0.85	0.361
Error	55	105186768	1912487		
Total	56	106810770			

S = 1383 R-Sq = 1.52% R-Sq(adj) = 0.00%

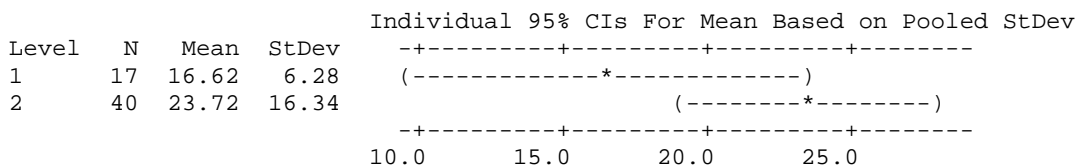


Pooled StDev = 1383

One-way ANOVA: PE ratio 2010 versus Cluster

Source	DF	SS	MS	F	P
Cluster	1	601	601	2.99	0.089
Error	55	11045	201		
Total	56	11646			

S = 14.17 R-Sq = 5.16% R-Sq(adj) = 3.44%



Pooled StDev = 14.17

One-way ANOVA: Dividend yield (%) 2010 versus Cluster

Source	DF	SS	MS	F	P
Cluster	1	5.79	5.79	1.89	0.175
Error	55	168.55	3.06		
Total	56	174.34			

S = 1.751 R-Sq = 3.32% R-Sq(adj) = 1.56%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
1	17	3.512	1.489	(-----*-----)
2	40	2.815	1.847	(-----*-----)

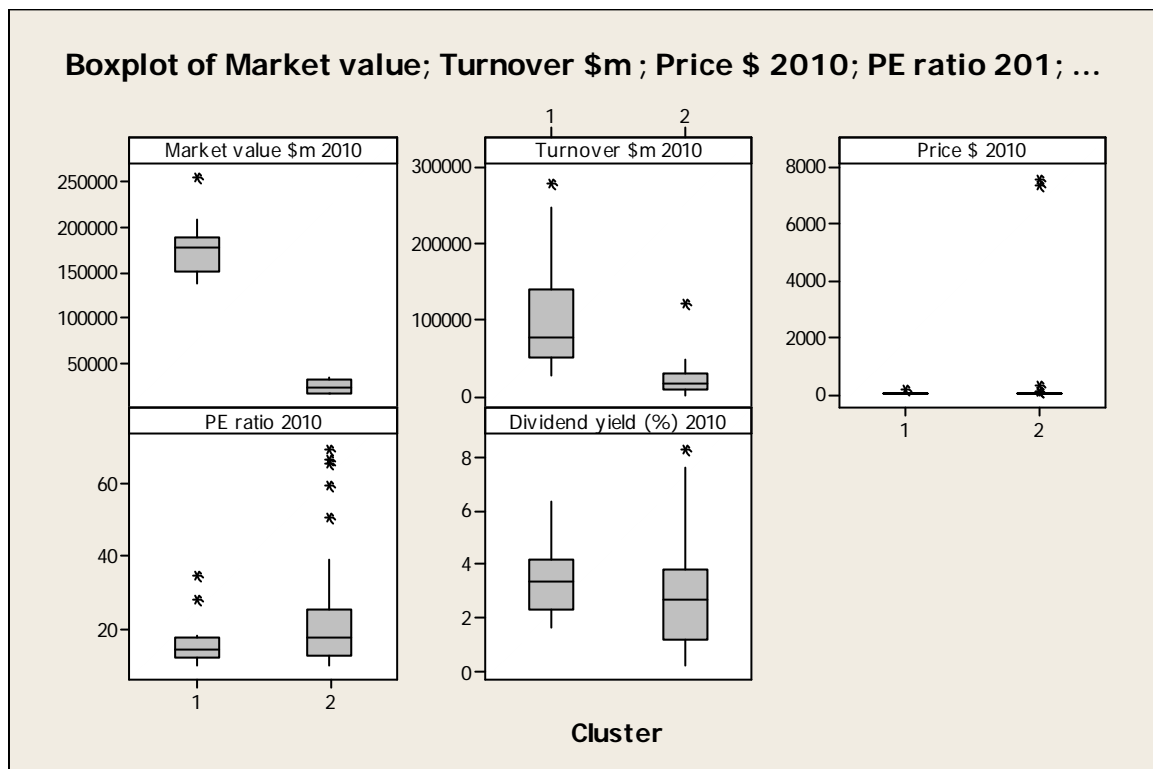
-----+-----+-----+-----+-----
2.40 3.00 3.60 4.20

Pooled StDev = 1.751

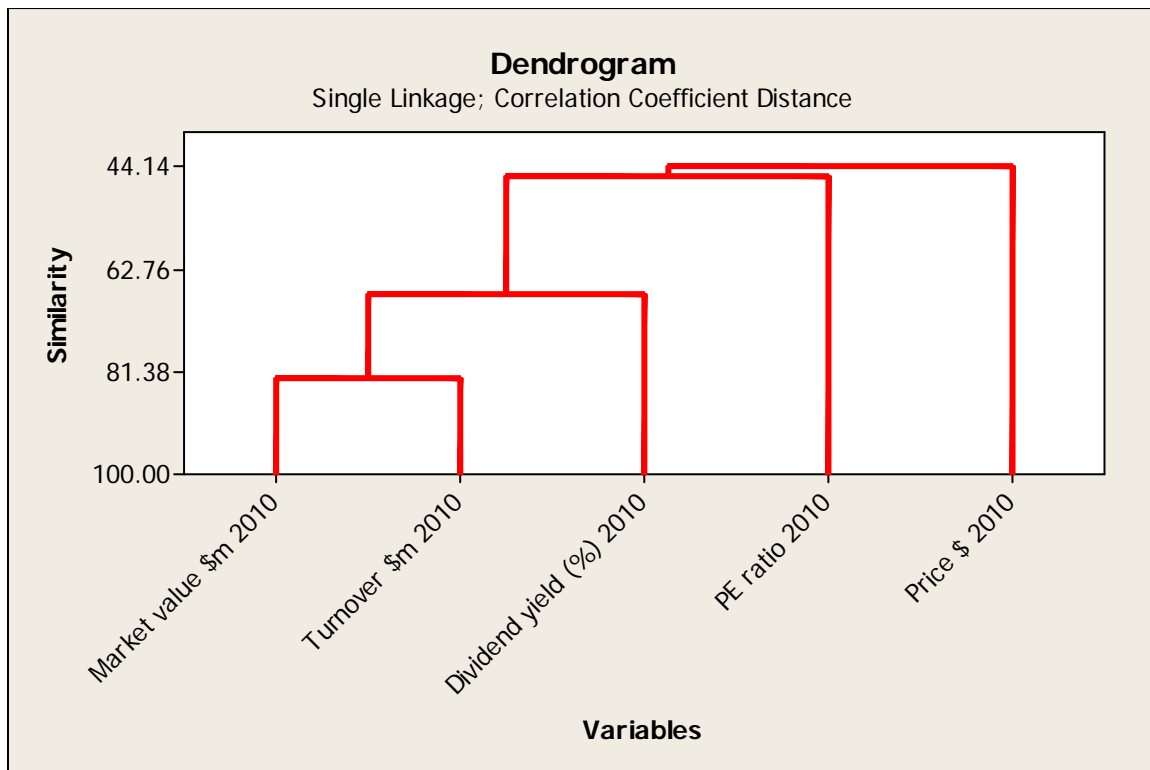
Os valores P-value das análises de variância acima nos confirmam que:

- Tendo como parâmetro o p-value menor ou igual a 5% para a rejeição da hipótese nula, podemos considerar que a média populacional das variáveis dos conglomerados são diferentes, com exceção das variáveis Price \$, Pe Ratio e Dividend Yield (%).
- Pela análise do valor F, percebemos que a maior diferença aparece na variável Market Value \$m.

Veremos abaixo o Box-Plot que nos permite visualizar mais claramente estas grandes diferenças entre os dois conglomerados.



Fica bastante notória as diferenças que separam os dois conglomerados nas 5 variáveis analisadas, com exceção das variáveis Price \$, PE ratio e Dividend Yield (%), conforme já explicado anteriormente.



Pelo dendrograma podemos observar que as variáveis Market Value \$m e Turnover \$m estão correlacionadas, e as variáveis Dividend Yield (%), PE ratio e Price \$ possui baixa correlação com as outras variáveis.

5.8.2 Análise Discriminante

A variável dependente de nossa análise será o conglomerado e para tentar explicar em qual conglomerado uma determinada empresa cai, utilizamos as 5 variáveis de indicadores relacionados ao ranking das 500 empresas mais valiosas do mundo.

Discriminant Analysis: Cluster versus Market value; Turnover \$m ; ...

Linear Method for Response: Cluster

Predictors: Market value \$m 2010; Turnover \$m 2010; Price \$ 2010;
PE ratio 2010; Dividend yield (%) 2010

Group	1	2
Count	17	40

Summary of classification

True Group

Put into Group	1	2
1	17	0
2	0	40
Total N	17	40
N correct	17	40
Proportion	1.000	1.000

N = 57 N Correct = 57

Proportion Correct = 1.000

Squared Distance Between Groups

	1	2
1	0.0000	84.2623
2	84.2623	0.0000

Linear Discriminant Function for Groups

	1	2
Constant	-60.382	-5.764
Market value \$m 2010	0.001	0.000
Turnover \$m 2010	0.000	-0.000
Price \$ 2010	0.001	0.001
PE ratio 2010	0.059	0.181
Dividend yield (%) 2010	2.113	1.707

Discriminant Analysis: Cluster versus Market value; Turnover \$m ; ...

Quadratic Method for Response: Cluster

Predictors: Market value \$m 2010; Turnover \$m 2010; Price \$ 2010;
PE ratio 2010; Dividend yield (%) 2010

Group	1	2
Count	17	40

Summary of classification

	True Group	
Put into Group	1	2
1	17	0
2	0	40
Total N	17	40
N correct	17	40
Proportion	1.000	1.000

N = 57 N Correct = 57

Proportion Correct = 1.000

From Generalized Squared Distance to Group

Group	1	2
1	53.19	462.98
2	142.99	58.91

A utilização de ambas as funções ajustou os dados a uma proporção de acerto de 100%, por isso vamos utilizar a função linear que é mais simples.

Veremos o que ocorre se excluirmos as variáveis com alto valor de p, utilizando a função quadrática:

Discriminant Analysis: Cluster versus Market value \$m ; Turnover \$m 2010

Linear Method for Response: Cluster

Predictors: Market value \$m 2010; Turnover \$m 2010

Group	1	2
Count	17	40

Summary of classification

Put into Group	True Group	
	1	2
1	17	0
2	0	40
Total N	17	40
N correct	17	40
Proportion	1.000	1.000

N = 57 N Correct = 57

Proportion Correct = 1.000

Squared Distance Between Groups

	1	2
1	0.0000	79.7746
2	79.7746	0.0000

Linear Discriminant Function for Groups

	1	2
Constant	-54.548	-1.167
Market value \$m 2010	0.001	0.000
Turnover \$m 2010	0.000	0.000

Não houve mudança no poder explicativo e o modelo fica mais simples, com duas variáveis ao invés de cinco. Veremos o que acontece se considerarmos apenas a variável Market Value \$m, variável com maior diferença entre os conglomerados. O objetivo é termos um modelo ainda mais simples e intuitivo:

Discriminant Analysis: Cluster versus Market value \$m 2010

Linear Method for Response: Cluster

Predictors: Market value \$m 2010

Group	1	2
Count	17	40

Summary of classification

Put into Group	True Group	
	1	2
1	17	0
2	0	40
Total N	17	40
N correct	17	40
Proportion	1.000	1.000

1	17	0
2	0	40
Total N	17	40
N correct	17	40
Proportion	1.000	1.000

N = 57 N Correct = 57

Proportion Correct = 1.000

Squared Distance Between Groups

	1	2
1	0.0000	77.4609
2	77.4609	0.0000

Linear Discriminant Function for Groups

	1	2
Constant	-52.549	-1.052
Market value \$m 2010	0.001	0.000

O poder explicativo se manteve em 100% e a análise fica ainda mais simples com apenas uma variável explicativa.

5.9 Regressão Logística

O objetivo deste tópico é efetuar uma análise de regressões múltiplas, logísticas binárias, logísticas ordinais de dados de indicadores relacionados ao Ranking das 500 Empresas mais valiosas do mundo. O principal propósito é comparar os conglomerados da amostra.

Binary Logistic Regression: Cluster versus Market value; Turnover \$m ; ...

```
* WARNING * Algorithm has not converged after 20 iterations.
* WARNING * Convergence has not been reached for the parameter estimates
              criterion.
* WARNING * The results may not be reliable.
* WARNING * Try increasing the maximum number of iterations.
```

Link Function: Logit

Response Information

Variable	Value	Count	
Cluster	2	40	(Event)
	1	17	
	Total	57	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95%
						CI
						Lower
Constant	32.4754	11872.5	0.00	0.998		
Market value \$m 2010	-0.0003777	0.0826762	-0.00	0.996	1.00	0.85
Turnover \$m 2010	0.0000313	0.135609	0.00	1.000	1.00	0.77
Price \$ 2010	-0.0006903	2.73929	-0.00	1.000	1.00	0.00
PE ratio 2010	0.0125430	281.823	0.00	1.000	1.01	0.00

Dividend yield (%) 2010 -0.111252 1783.43 -0.00 1.000 0.89 0.00

Predictor Upper
 Constant
 Market value \$m 2010 1.18
 Turnover \$m 2010 1.30
 Price \$ 2010 214.48
 PE ratio 2010 7.82247E+239
 Dividend yield (%) 2010 *

Log-Likelihood = -0.000
 Test that all slopes are zero: G = 69.468, DF = 5, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	0.0000000	51	1.000
Deviance	0.0000001	51	1.000
Hosmer-Lemeshow	0.0000000	8	1.000

Table of Observed and Expected Frequencies:
 (See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total	
	1	2	3	4	5	6	7	8	9	10		
2												
Obs	0	0	0	5	6	6	5	6	6	6	6	40
Exp	0.0	0.0	0.0	5.0	6.0	6.0	5.0	6.0	6.0	6.0	6.0	
1												
Obs	5	6	6	0	0	0	0	0	0	0	0	17
Exp	5.0	6.0	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Total	5	6	6	5	6	6	5	6	6	6	6	57

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

Pairs	Number	Percent	Summary Measures
Concordant	680	100.0	Somers' D 1.00
Discordant	0	0.0	Goodman-Kruskal Gamma 1.00
Ties	0	0.0	Kendall's Tau-a 0.43
Total	680	100.0	

A utilização da regressão logística binária chegou a um percentual de concordância de 100%, porém alguns erros ocorreram, o que torna o resultado não confiável.

Veremos o que ocorre se excluirmos as variáveis com alto valor de p:

Binary Logistic Regression: Cluster versus Market value; Turnover \$m

* WARNING * Algorithm has not converged after 20 iterations.
 * WARNING * Convergence has not been reached for the parameter estimates criterion.
 * WARNING * The results may not be reliable.
 * WARNING * Try increasing the maximum number of iterations.

Link Function: Logit

Response Information

Variable	Value	Count
Cluster	2	40 (Event)
	1	17
	Total	57

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds	95% CI	
					Ratio	Lower	Upper
Constant	32.4047	5092.39	0.01	0.995			
Market value \$m 2010	-0.0003788	0.0806930	-0.00	0.996	1.00	0.85	1.17
Turnover \$m 2010	0.0000299	0.138088	0.00	1.000	1.00	0.76	1.31

Log-Likelihood = -0.000

Test that all slopes are zero: G = 69.468, DF = 2, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	0.0000000	54	1.000
Deviance	0.0000001	54	1.000
Hosmer-Lemeshow	0.0000000	8	1.000

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total	
	1	2	3	4	5	6	7	8	9	10		
2												
Obs	0	0	0	5	6	6	5	6	6	6		40
Exp	0.0	0.0	0.0	5.0	6.0	6.0	5.0	6.0	6.0	6.0		
1												
Obs	5	6	6	0	0	0	0	0	0	0		17
Exp	5.0	6.0	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		
Total	5	6	6	5	6	6	5	6	6	6		57

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures		
Concordant	680	100.0	Somers' D		1.00
Discordant	0	0.0	Goodman-Kruskal	Gamma	1.00
Ties	0	0.0	Kendall's Tau-a		0.43
Total	680	100.0			

Não houve mudança no poder explicativo e o modelo fica mais simples, com duas variáveis ao invés de cinco, porém novamente alguns erros ocorreram, o que torna o resultado não confiável. Veremos o que acontece se considerarmos apenas a variável Market Value \$m, variável com maior diferença entre os conglomerados. O objetivo é termos um modelo válido e ainda mais simples e intuitivo:

Binary Logistic Regression: Cluster versus Market value \$m 2010

* WARNING * Algorithm has not converged after 20 iterations.
 * WARNING * Convergence has not been reached for the parameter estimates

criterion.
 * WARNING * The results may not be reliable.
 * WARNING * Try increasing the maximum number of iterations.

Link Function: Logit

Response Information

Variable	Value	Count
Cluster	2	40 (Event)
	1	17
	Total	57

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	32.7155	5082.36	0.01	0.995			
Market value \$m 2010	-0.0003695	0.0566962	-0.01	0.995	1.00	0.89	1.12

Log-Likelihood = -0.000

Test that all slopes are zero: G = 69.468, DF = 1, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	0.0000000	55	1.000
Deviance	0.0000001	55	1.000
Hosmer-Lemeshow	0.0000000	8	1.000

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total	
	1	2	3	4	5	6	7	8	9	10		
2												
Obs	0	0	0	5	6	6	5	6	6	6		40
Exp	0.0	0.0	0.0	5.0	6.0	6.0	5.0	6.0	6.0	6.0		
1												
Obs	5	6	6	0	0	0	0	0	0	0		17
Exp	5.0	6.0	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		
Total	5	6	6	5	6	6	5	6	6	6		57

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	680	100.0	Somers' D 1.00
Discordant	0	0.0	Goodman-Kruskal Gamma 1.00
Ties	0	0.0	Kendall's Tau-a 0.43
Total	680	100.0	

O poder explicativo se manteve em 100% e análise fica ainda mais simples com apenas uma variável explicativa. Porém novamente alguns erros ocorreram, o que torna o resultado não confiável, não sendo possível utilizar a regressão logística binária na amostra.

5.10 Árvores de Classificação

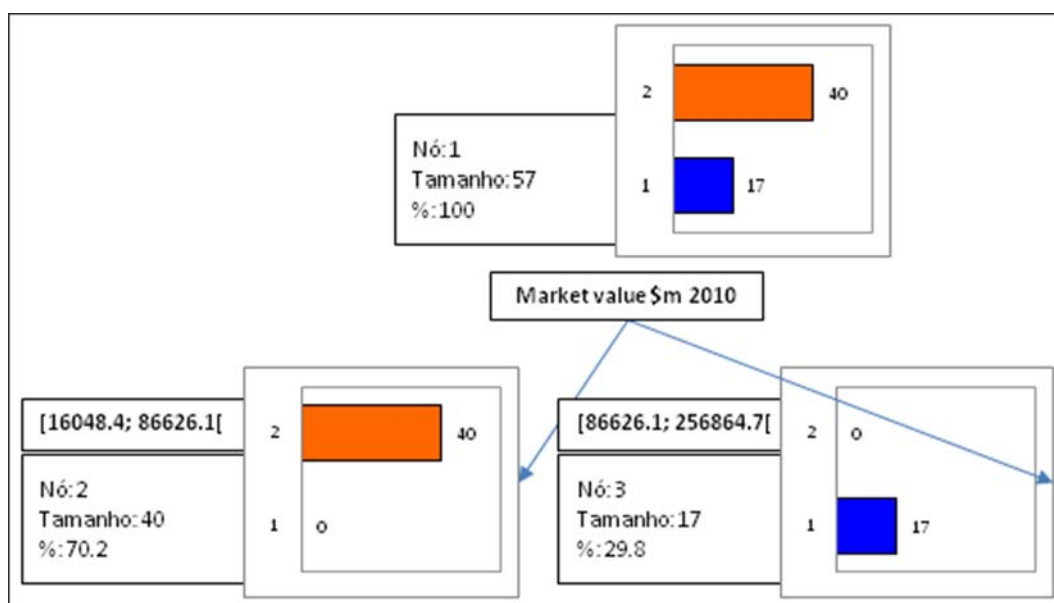
O objetivo deste tópico é efetuar uma análise de árvores de classificação de dados de indicadores relacionados ao Ranking das 500 Empresas mais valiosas do mundo. O principal propósito é comparar os conglomerados da amostra e verificar qual modelo oferece melhor análise de classificação.

Estatísticas descritivas:

Variável	Categorias	Freqüências	%
Cluster	1	17	29.825
	2	40	70.175

Estrutura da árvore:

Nó	p-valor	Objetos	%	Nó pai	Filhos	Variável de separação	Valores	Pureza
1	1.000	57	100.00%		2; 3			70.18%
2	0.000	40	70.18%	1		Market value \$m 2010	[16048.4; 86626.1[100.00%
3	0.000	17	29.82%	1		Market value \$m 2010	256864.7[100.00%



Foi possível observar pelo aplicativo XLSTAT (Árvore de classificação e regressão) que a única variável que apresenta importância na separação dos grupos é o Market Value \$m. A proporção de acerto foi de 100,0% (57 de 57 corretas).

Mesmo assim, a melhor opção ainda é a análise discriminante, pois também alcança 100% de proporção de acerto, utilizando somente uma variável (Market Value \$m) e o seu cálculo é mais simples que a árvore de classificação e regressão. A análise logística apresenta erros que tornam os resultados não confiáveis.

5.11 Análise de Correspondências

Realizar uma análise de correspondência (AC) – análise multivariada - com dados de indicadores relacionados ao Ranking das 500 empresas mais valiosas do mundo.

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 5 x 6.

Simple Correspondence Analysis: Carbon dioxi; Expected Yea; Expenditure ; GDP p

Analysis of Contingency Table

Axis	Inertia	Proportion	Cumulative	Histogram
1	0.0024	0.9531	0.9531	*****
2	0.0001	0.0397	0.9928	*
3	0.0000	0.0050	0.9978	
4	0.0000	0.0022	1.0000	
Total	0.0025			

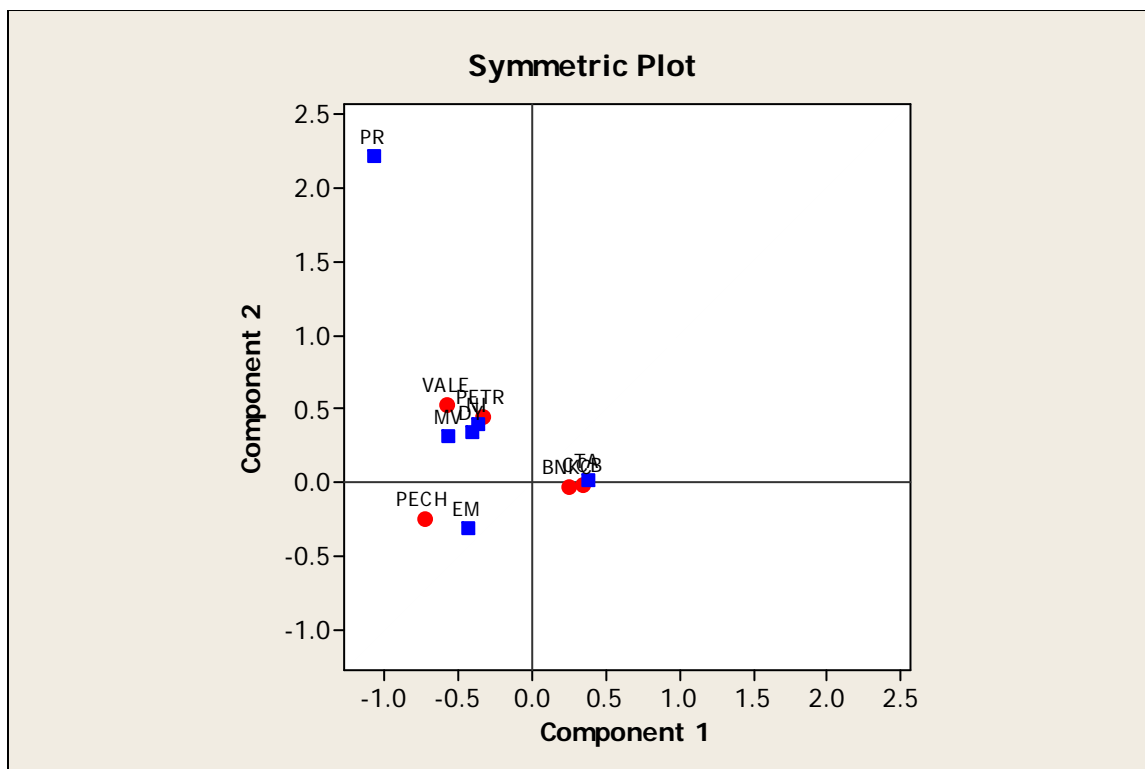
Row Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	BR	0.985	0.229	0.028	0.003	0.024	0.001	-0.017	0.961	0.667
2	CH	0.976	0.153	0.097	-0.039	0.958	0.097	0.005	0.019	0.046
3	IN	0.999	0.072	0.658	-0.152	0.999	0.690	0.002	0.000	0.004
4	RU	0.990	0.331	0.165	0.035	0.988	0.171	0.001	0.002	0.007
5	SA	0.957	0.214	0.052	0.021	0.747	0.041	0.011	0.210	0.276

Column Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	CDE	0.993	0.001	0.041	0.147	0.119	0.005	0.399	0.875	0.913
2	EYS	0.949	0.001	0.082	-0.384	0.941	0.081	0.037	0.009	0.018
3	EE	0.653	0.000	0.000	-0.437	0.645	0.000	0.048	0.008	0.000
4	GDP	1.000	0.990	0.008	0.004	0.994	0.008	-0.000	0.006	0.001
5	LEB	0.999	0.007	0.843	-0.554	0.999	0.884	-0.001	0.000	0.000
6	MYS	0.915	0.001	0.025	-0.257	0.806	0.021	0.095	0.109	0.068

Gráfico Symmetric Plot - Mostra a associação entre a categoria linha e coluna conforme a proximidade dos seus pontos no Biplot.



A análise da tabela de contingência mostra uma decomposição da inércia (χ^2/n). Do total da inércia da matriz de dados, 95,31% é contabilizada no primeiro componente, 3,97% é contabilizada no segundo componente e assim por diante.

No Symmetric Plot observa-se:

1. A empresa PetroChina, 1ª colocada no ranking está próxima ao Employees (Número de Empregados);
2. As empresas Industrial & Commercial Bank of China e China Construction Bank (4ª e 13ª colocadas no ranking) estão próximas ao indicador Total Assets \$m.
3. As empresas Petrobras e Vale (13ª e 22ª colocadas no ranking) estão entre os indicadores Market Value \$m, Net Income \$m e Dividend Yield (%).
4. Todas as empresas estão distantes do indicador Price \$.

6. CONSIDERAÇÕES FINAIS

6.1 Evolução Histórica dos Indicadores das 500 Empresas mais valiosas do mundo e da Posição dos Países do BRICS no Ranking

O presente trabalho efetuou uma análise das médias amostrais, intervalos de confiança para a média populacional e análise de variância referente a 5 variáveis do Ranking 2010 das 500 Empresas mais valiosas do mundo publicada pela revista Financial Times.

Constatamos que houve uma evolução em duas dessas variáveis, sendo elas: Market value \$m (Valor de Mercado) e Turnover \$m. No caso do Valor de Mercado, o valor médio passou de \$m 40.401 para \$m 53.555 (acréscimo de 32,56%) e no caso do Turnover, o valor médio passou de \$m 30.496 para \$m 40.419 (acréscimo de 32,54%). Isso demonstra que as

empresas se tornaram mais valiosas de 2005 para 2010. No caso das demais variáveis (Price \$ (Preço da Ação), PE ratio e Dividend yield %) foi constatado que as médias populacionais dos anos de 2010 e 2005 não são estatisticamente diferentes.

Já na análise da evolução dos países do BRICS no ranking das 500 empresas mais valiosas do mundo pudemos observar que houve uma evolução muito significativa. Houve uma variação positiva de 325% de 2005 para 2010, sendo que em 2005 haviam 16 empresas do BRICS no ranking, que representavam 3,2% e em 2010 há 68 empresas do BRICS entre as 500 empresas mais valiosas do mundo, que representam 13,6%.

Analisando isoladamente os países do BRICS, o país que teve maior evolução no ranking de 2005 para 2010 foi a China, que em 2005 não tinha nenhuma empresa no ranking e em 2010 tem 23 empresas. Sendo também a China o país do BRICS que possui mais empresas no ranking, representando 33,82%, seguido por Índia (16 empresas, 23,53%), Brasil (12 empresas, 17,65%), Rússia (11 empresas, 16,18%) e África do Sul (6 empresas, 8,82%).

Se analisarmos a evolução dos países do BRICS em relação ao Valor de Mercado, podemos observar uma evolução mais significativa ainda. Houve uma variação positiva de 833,20%, sendo que em 2005 representavam 2,02% e em 2010 representam 15,54%.

Outro dado muito interessante é que, em 2005 não havia nenhuma empresa do BRICS entre as 25 empresas mais valiosas do mundo e em 2010 existem 5 empresas entre as 25 mais valiosas do mundo, sendo que a empresa mais valiosa do mundo (PetroChina) é uma empresa da China, país pertencente ao BRICS. A China também possui a 4ª e a 11ª mais valiosa do mundo. O Brasil possui a 13ª (Petrobras) e a 22ª (Vale) empresa mais valiosa do mundo.

6.2 Estudos Complementares

6.2.1 Análise de Tendências

Analisando a planilha de resultados da análise de tendências abaixo, podemos observar que poderá haver um crescimento de 106% na quantidade de países do BRICS no ranking das 500 empresas mais valiosas do mundo, passando de 68 países em 2010 para 140 países em 2015. Já no caso do outros países, poderá haver uma redução de 17%, passando de 432 países em 2010 para 360 países em 2015.

	Comportamento Histórico						Projeção				
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Quantidade											
Total BRICS	16	25	36	66	58	68	90	103	115	128	140
Outros	484	475	464	434	442	432	410	397	385	372	360
Total	500	500	500	500	500	500	500	500	500	500	500
Percentual											
Total BRICS	3%	5%	7%	13%	12%	14%	18%	21%	23%	26%	28%
Outros	97%	95%	93%	87%	88%	86%	82%	79%	77%	74%	72%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

6.2.2 Relação entre Variáveis

A maior relação entre variáveis foi encontrada entre o Valor de Mercado e o Lucro Líquido. Isso é coerente, pois o Lucro Líquido é um indicador que tem um alto grau de influência no valor de mercado da empresa.

Já a menor relação entre variáveis contínuas foi encontrada entre o PE ratio e o Preço da Ação. Esse resultado sugere que o indicador PE ratio não tem muita influência sobre o preço da ação no mercado.

6.2.3 Regressões Múltiplas

Foi encontrado um valor de R-Quadrado por volta de 80%, considerando a variável dependente Valor de Mercado. Esse valor parece ser condizente com os tipos de dados, pois as demais variáveis estão diretamente relacionadas com o valor de mercado da empresa. Com isso é possível realizar inferências satisfatórias mediante a utilização da equação encontrada.

6.2.4 Amostragem

Era esperado que à medida que o tamanho da amostra aleatória aumenta, os valores da análise exploratória de dados também aumenta. Na maior parte dos casos foi possível verificar esta premissa, pela comparação dos valores de média, desvio padrão, mediana, mínimo e máximo entre os dados da população e da amostra, nas variáveis analisadas.

No caso de P-Value, podemos observar que a variável Price \$ apresentou valores abaixo de 0,10 nas amostras tamanho 20 e 50, portanto são estatisticamente diferentes da população. No caso das variáveis Market Value \$m, Net Income \$m, Total Assets \$m, Employees, Price \$ e Divident Yield (%), o P-Value é maior conforme o tamanho da amostra é maior. Nos demais casos, todas as amostras são estatisticamente iguais à população, porém nem sempre o P-Value é maior conforme o tamanho da amostra é maior.

P-Value do Teste – T para três Amostras

Variáveis	Amostra 20 Aleatória	Amostra 50 Estratificada	Amostra 100 Aleatória
Market Value \$m	0,692	0,123	0,818
Turnover \$m	0,556	0,294	0,363
Net Income \$m	0,484	0,137	0,721
Total Assets \$m	0,263	0,181	0,348
Employees	0,552	0,647	0,650
Price \$	0,094	0,076	0,429
PE ratio	0,314	0,997	0,641
Dividend Yield (%)	0,227	0,478	0,981

Na análise do teste One-way Anova, foi observado que o desvio padrão diminui conforme o tamanho da amostra aumenta e que a média das amostras vai se aproximando da média da população também conforme a amostra aumenta.

Na análise das correlações das variáveis, as amostras ficaram mais próximas das correlações da população total conforme o tamanho da amostra aumenta. Porém mesmo na amostra maior, de tamanho 100, algumas correlações não foram mantidas conforme a população.

No Dendograma as variáveis em geral mantiveram a disposição nas amostras. E foram ficando mais próximas da população conforme o tamanho da amostra aumenta.

Assim, podemos concluir que conforme o tamanho da amostra aumenta, melhor representa a população total. Porém no caso estudado, mesmo a maior amostra, de tamanho 100, não representa a população total, apresentando resultados diferentes em algumas análises.

6.2.5 Análise em Componentes Principais

De acordo com todas as análises efetuadas, percebemos que o agrupamento de variáveis é pertinente no caso das variáveis do Ranking das 500 empresas mais valiosas do mundo. Isto pôde ser observado inicialmente pelas matrizes de correlação e dendogramas e depois comprovados pela análise dos componentes principais.

Assim, ao invés de trabalharmos com um grupo grande de variáveis (8) poderíamos utilizar apenas três índices (PC1, PC2 e PC3) que as represente satisfatoriamente (72,9%).

6.2.6 Análise de Conglomerados

Foi possível agrupar as empresas da amostra em 2 conglomerados. Podemos observar que as empresas mais valiosas se encontram no conglomerado 1, enquanto as empresas intermediárias e menos valiosas não eram divididas e se encontram no conglomerado 2. As empresas intermediárias e menos valiosas não foram divididas em conglomerados, mesmo quando se aumentava o número de conglomerados. Isso pode acontecer em função das empresas intermediárias e menos valiosas não terem muita diferença de valor, ao contrário das empresas mais valiosas.

6.2.7 Análise Discriminante

De acordo com todas as análises realizadas, podemos constatar que a função linear se mostrou mais adequada para a amostra, pois apesar de ter obtido os mesmos resultados da função quadrática, ela é mais simples.

Também foi possível constatar que utilizando-se somente a variável Market Value \$m para divisão das empresas da amostra em conglomerados chegou-se a índices de proporção de acerto de 100%. Portanto podemos concluir que entre as variáveis de indicadores do ranking das 500 empresas mais valiosas do mundo, a variável Market Value \$m é a mais significativa e somente a sua utilização já é suficiente para dividir as empresas da amostra em conglomerados. As demais variáveis muito pouco ou nada acrescentam.

6.2.8 Regressão Logística

De acordo com todas as análises realizadas, podemos constatar que a análise discriminante é uma opção melhor que a análise de regressão logística binária, pois na última, ocorreram erros que tornaram os resultados não confiáveis, o que não ocorreu na análise discriminante.

Dentro da análise discriminante, a função linear se mostrou mais adequada.

Também foi possível constatar que utilizando-se somente a variável Market Value \$m para divisão das empresas da amostra em conglomerados chegou-se a índices de proporção de acerto de 100%. Portanto podemos concluir que entre as variáveis de indicadores das 500 empresas mais valiosas do mundo, a variável Valor de Mercado é a mais significativa e somente a sua utilização já é suficiente para dividir as empresas da amostra em conglomerados. As demais variáveis muito pouco ou nada acrescentam.

6.2.9 Árvores de Classificação

De acordo com todas as análises realizadas, podemos constatar que a análise discriminante e a árvore de classificação e regressão são uma opção melhor que a análise de regressão logística ordinal, pois na última, ocorreram erros que tornaram os resultados não confiáveis, o que não ocorreu nas demais análises.

A melhor opção fica empatada entre a análise logística (linear) e a árvore de classificação e regressão, pois ambas alcançam 100% de proporção de acerto e utiliza somente uma variável (Market Value \$m). Como o cálculo da análise logística linear é mais simples, esta se torna a melhor opção. A análise logística apresenta erros que tornam os resultados não confiáveis.

A análise pela árvore de classificação e regressão também chegou ao resultado de que a variável Market Value \$m é a que apresenta maior importância na separação de grupos, o que reforça a conclusão de que é a variável mais significativa e somente a sua utilização já é suficiente para dividir os países das amostras em conglomerados. As demais variáveis muito pouco ou nada acrescentam.

6.2.10 Análise de Correspondências

A análise da tabela de contingência mostra uma decomposição da inércia (χ^2/n). Do total da inércia da matriz de dados, 95,31% é contabilizada no primeiro componente, 3,97% é contabilizada no segundo componente e assim por diante.

De acordo com as análises efetuadas, observa-se:

1. A empresa PetroChina, 1ª colocada no ranking está próxima ao Employees (Número de Empregados);
2. As empresas Industrial & Commercial Bank of China e China Construction Bank (4ª e 13ª colocadas no ranking) estão próximas ao indicador Total Assets \$m.
3. As empresas Petrobras e Vale (13ª e 22ª colocadas no ranking) estão entre os indicadores Market Value \$m, Net Income \$m e Dividend Yield (%).
4. Todas as empresas estão distantes do indicador Price \$.