



**“PESQUISA NOSSA SÃO PAULO – QUALIDADE DE VIDA URBANA EM
SÃO PAULO” – Um estudo de correlações entre trabalho e renda e violência
urbana**

DANIEL RODRIGUES PIRES BEZERRA

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

ANÁLISE EXPLORATÓRIA DOS DADOS

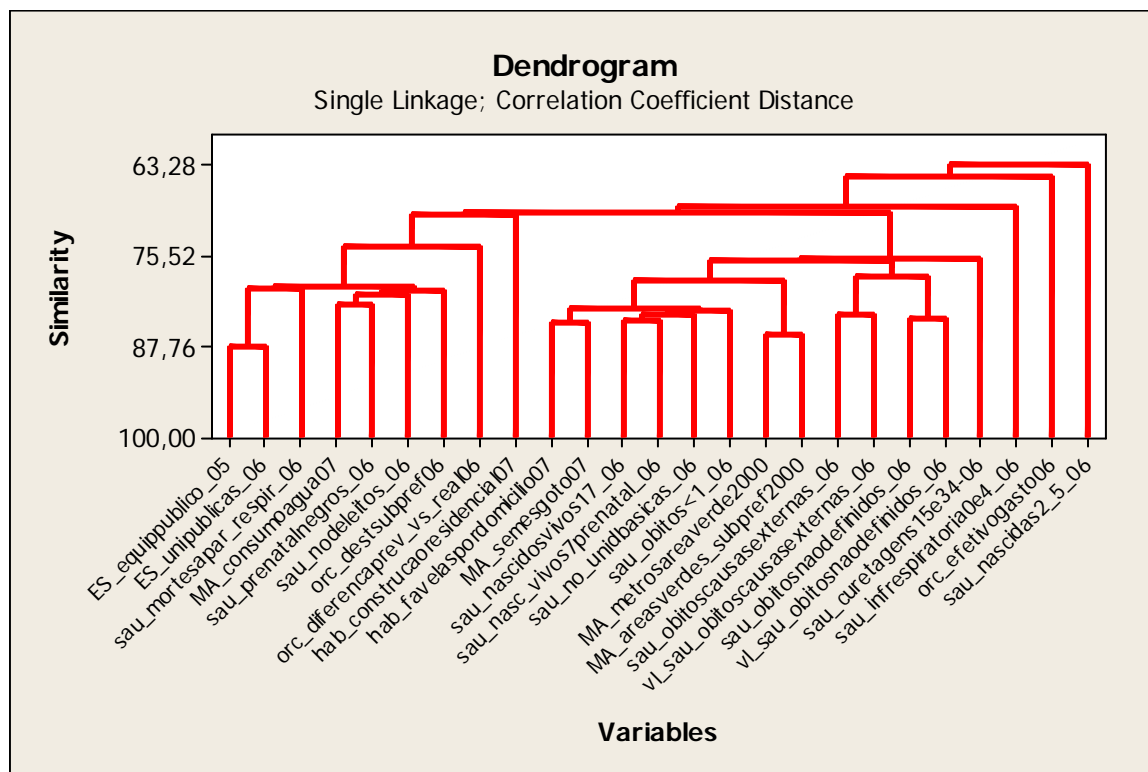
As análises presentes neste estudo foram baseadas na pesquisa efetuada pela organização Nossa São Paulo sobre qualidade de vida urbana na cidade de São Paulo. Foram feitas análises de diferentes tipos de gráfico e dados de correlação, entre outras.

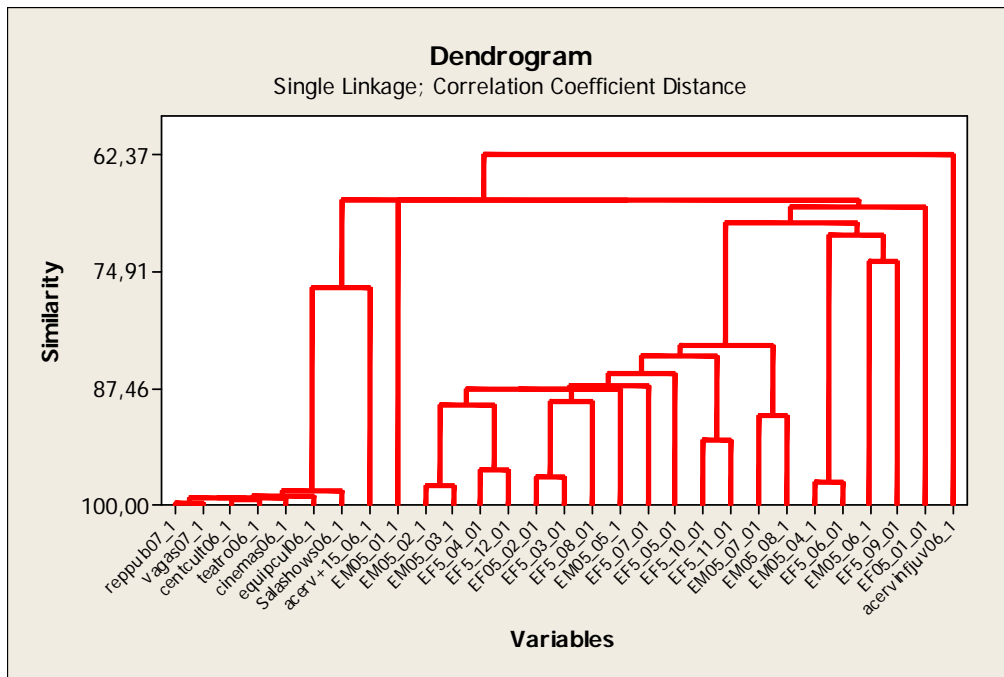
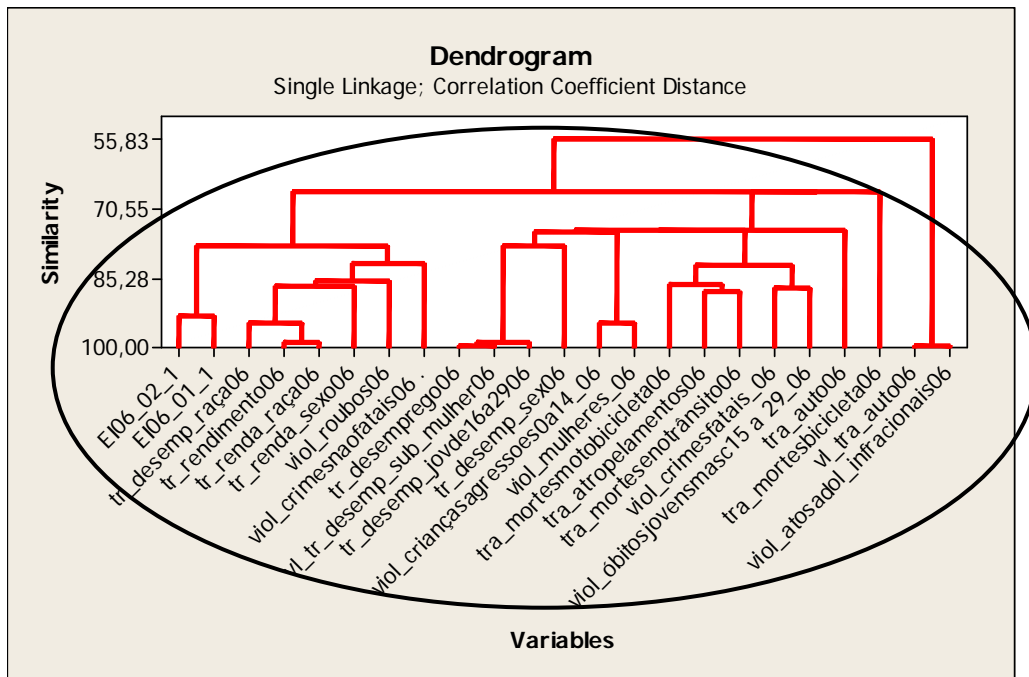
Os Indivíduos

Os indivíduos desta pesquisa são os distritos, que totalizam em 32.

As Variáveis

Dos indicadores originais da pesquisa, foram escolhidas 10, um categórico e nove quantitativas, e o critério de escolha dos indicadores dentro do total foi a semelhança e a área temática (foram buscadas correlações entre variáveis das áreas trabalho e renda e violência), que podemos verificar nos dendogramas abaixo. As variáveis são melhor explicadas na Tabela 1.





Foram descartadas as variáveis com semelhança muito elevada, de forma a possibilitar um estudo ótimo de componentes principais. Abaixo a análise de correlações e o dendrograma das variáveis escolhidas.

Correlations: tr_desempreg; tr_desemp_jo; tr_rendiment; tr_desemp_ra; ...

	tr_desemprego06	tr_desemp_jovdel	tr_rendimento06
tr_desemp_jovdel		0,976	0,000
tr_rendimento06		0,890	0,862

	0,000	0,000	
tr_desemp_raça06	-0,746 0,000	-0,717 0,000	-0,803 0,000
viol_atosadol_in	0,032 0,860	0,036 0,844	0,048 0,795
viol_crimesfatai	0,334 0,061	0,237 0,192	0,308 0,087
viol_roubos06	-0,707 0,000	-0,716 0,000	-0,649 0,000
viol_crimesnaofa	-0,583 0,000	-0,578 0,001	-0,419 0,017
viol_óbitosjuven	0,504 0,003	0,425 0,015	0,514 0,003

	tr_desemp_raça06	viol_atosadol_in	viol_crimesfatai
viol_atosadol_in	0,054 0,774		
viol_crimesfatai	-0,126 0,500	0,026 0,888	
viol_roubos06	0,465 0,008	-0,010 0,958	-0,146 0,424
viol_crimesnaofa	0,557 0,001	0,020 0,915	0,097 0,598
viol_óbitosjuven	-0,277 0,132	0,077 0,673	0,729 0,000

	viol_roubos06	viol_crimesnaofa
viol_crimesnaofa	0,775 0,000	
viol_óbitosjuven	-0,545 0,001	-0,272 0,131

Cell Contents: Pearson correlation
P-Value

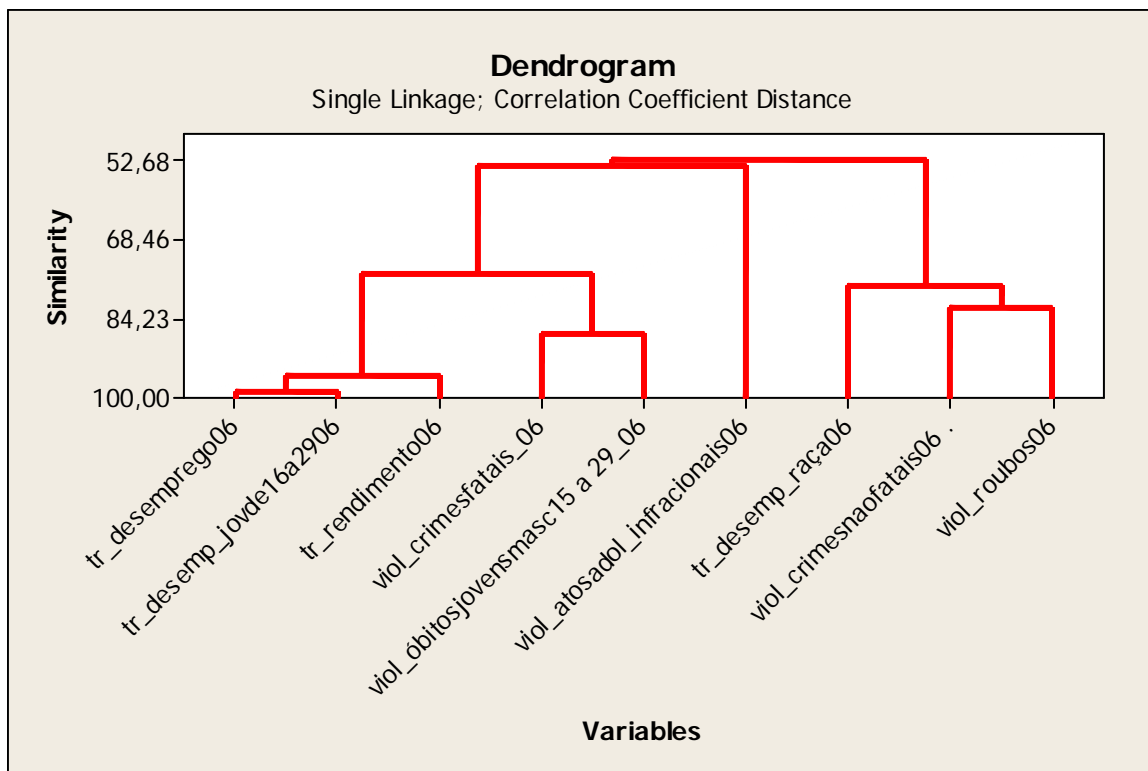


Tabela 1: Detalhamento dos indicadores

Áreas Temáticas (abreviação)	Indicadores	Descrição	Tipo de Variável	Período
Trabalho e Renda (tr)	tr_desemprego06	Desemprego	Variável Quantitativa	2006
	tr_desemprego06	Taxa média de desemprego de jovens de 16 a 29 anos.	Variável Quantitativa	2006
	tr_desemp_raça06	Proporção entre a taxa de desemprego de negros e não negros. Indisponível para a subprefeitura da Sé.	Variável Quantitativa	2006
	tr_rendimento06	Rendimento médio proveniente do trabalho, em reais.	Variável Quantitativa	2006

Violência (viol) histórico de 2003 a 2006	viol_atosadol_infracionais06	Porcentagem dos adolescentes envolvidos em ato infracional residentes em cada subprefeitura sobre o total de atos infracionais.	Variável Quantitativa	2006
	viol_crimesfatais_06	Número de crimes violentos fatais por cem mil habitantes, por local de ocorrência.	Variável Quantitativa	2006
	viol_crimesnaofatais06 .	Número de crimes violentos não fatais por cem mil habitantes, por local de ocorrência.	Variável Quantitativa	2006
	viol_óbitosjovensmasc15 a 29_06	Número de óbitos por homicídio de jovens do sexo masculino de 15 a 29 anos, por cem mil habitantes dessa faixa etária e sexo.	Variável Quantitativa	2006
	viol_roubos06	Número de roubos por cem mil habitantes, por local de ocorrência.	Variável Quantitativa	2006

Tabela de dados

Segue abaixo as tabelas de dados utilizadas nesta pesquisa:

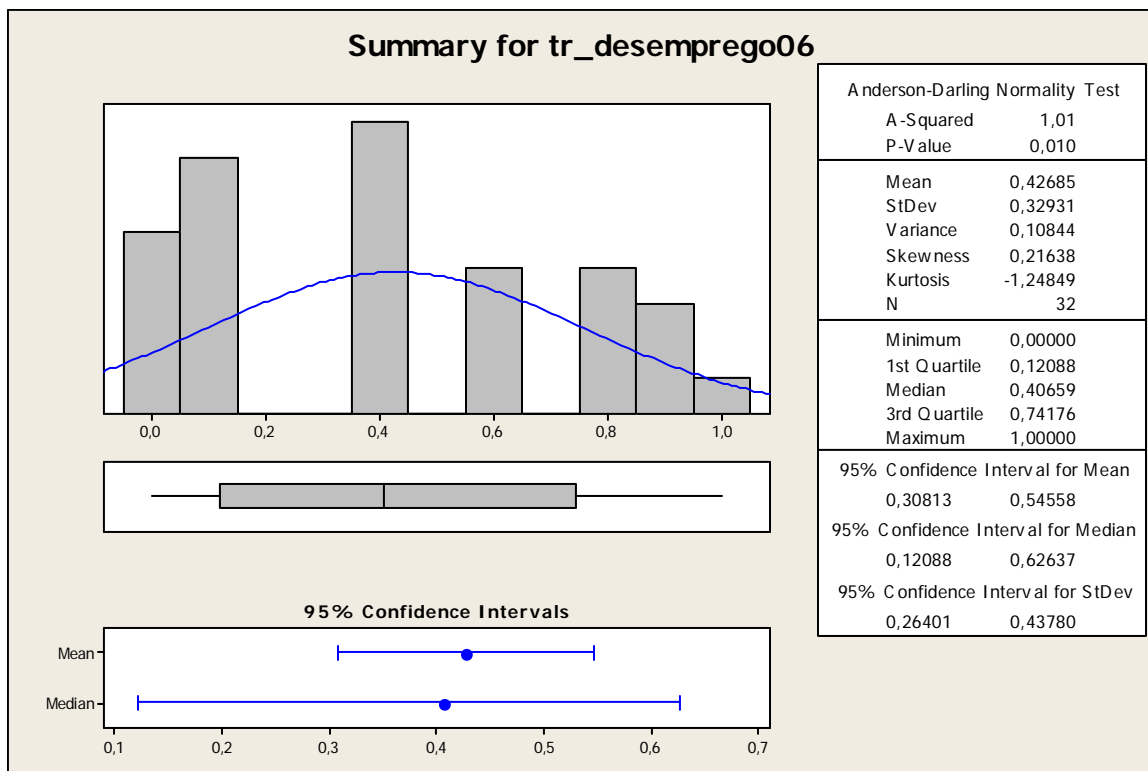
Região	tr_desemprego06	tr_desemp_jovde16a2906	tr_rendimento06	tr_desemp_raça06	viol_atosadol_infracionais06	viol_crimesfatais_06	viol_crimesnaofatais06 .	viol_roubos06	viol_óbitosjovensmasc15 a 29_06
São Paulo	0,41	0,36	0,26	0,50	0,00	0,65	0,81	0,76	0,58
Aricanduva	0,63	0,62	0,23	0,83	0,99	0,97	0,91	0,79	0,70
Butantã	0,95	0,88	1,00	0,00	0,98	0,84	0,83	0,73	0,81
Campo Limpo	0,00	0,10	0,00	0,92	0,94	0,55	0,92	0,83	0,47
Capela do Socorro	0,00	0,10	0,00	0,92	0,95	0,56	0,96	0,89	0,36
Casa Verde/Cachoeirinha	0,41	0,38	0,17	0,56	0,97	0,24	0,77	0,89	0,00

Cidade Ademar	0,00	0,10	0,00	0,92	0,96	0,55	0,92	0,83	0,55
Cidade Tiradentes	0,12	0,00	0,05	1,00	0,98	0,99	0,90	1,00	0,76
Ermelino Matarazzo	0,12	0,00	0,05	1,00	0,97	0,88	0,87	0,84	0,80
Freguesia/Brasilândia	0,41	0,38	0,17	0,56	0,96	0,45	0,80	0,94	0,19
Guaianases	0,12	0,00	0,05	1,00	0,97	0,75	0,94	0,93	0,71
Ipiranga	0,78	0,77	0,64	0,56	0,98	0,82	0,84	0,72	0,63
Itaim Paulista	0,12	0,00	0,05	1,00	0,97	0,80	0,96	0,88	0,68
Itaquera	0,12	0,00	0,05	1,00	0,95	0,74	0,88	0,84	0,68
Jabaquara	0,78	0,77	0,64	0,56	0,97	0,64	0,82	0,76	0,46
Jaçanã / Tremembé	0,44	0,36	0,25	0,25	0,98	0,51	0,78	0,91	0,44
Lapa	0,95	0,88	1,00	0,00	0,99	0,65	0,63	0,54	0,95
M'Boi Mirim	0,00	0,10	0,00	0,92	0,98	0,34	0,96	0,92	0,27
Moóca	0,63	0,62	0,23	0,83	0,96	0,57	0,43	0,52	0,83
Parelheiros	0,00	0,10	0,00	0,92	1,00	0,00	0,86	0,99	0,37
Penha	0,63	0,62	0,23	0,83	0,95	0,92	0,85	0,80	0,75
Perus	0,41	0,38	0,17	0,56	1,00	0,57	1,00	0,98	0,57
Pinheiros	0,95	0,88	1,00	0,00	1,00	0,93	0,54	0,25	0,98
Pirituba	0,41	0,38	0,17	0,56	0,98	0,76	0,91	0,93	0,72
Santana/Tucuruvi	0,44	0,36	0,25	0,25	0,99	0,94	0,79	0,79	0,83
Santo Amaro	0,78	0,77	0,64	0,56	1,00	0,61	0,76	0,11	0,93
São Mateus	0,12	0,00	0,05	1,00	0,97	0,42	0,93	0,91	0,46
São Miguel	0,12	0,00	0,05	1,00	0,97	0,73	0,88	0,87	0,65
Sé	1,00	1,00	0,46		0,94	0,51	0,00	0,00	0,75
Vila Maria/Vila Guilherme	0,44	0,36	0,25	0,25	0,98	0,62	0,70	0,77	0,74
Vila Mariana	0,78	0,77	0,64	0,56	1,00	1,00	0,85	0,48	1,00
Vila Prudente/Sapopemba	0,63	0,62	0,23	0,83	0,97	0,87	0,97	0,87	0,77

Análise individual das variáveis quantitativas

A análise deste tipo de variável permite a utilização de uma grande gama de ferramentas como histogramas, curvas de densidade, *box-plot*, além de medidas numéricas como média, desvio-padrão, variância, quantidade de observações, valor mínimos e máximos, informações dos quartis e teste de normalidade. Assim, segue abaixo a análise individual de cada variável.

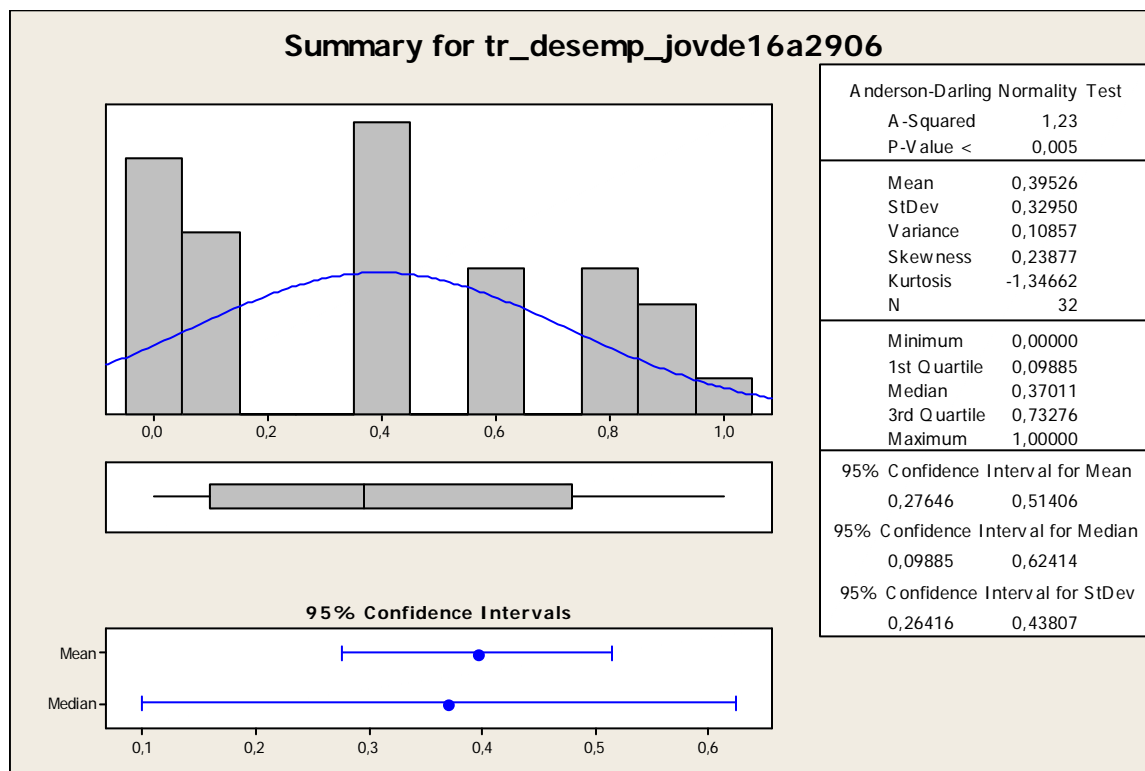
Desemprego



As principais observações que podemos fazer são:

- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição mais simétrica. Esta conclusão está comprovada pelo teste de normalidade de Anderson-Darling que indica que a distribuição pode ser considerada uma Normal. Isto pode ser explicado pela padronização dos dados feita com o intuito de homogeneizar as variáveis e possibilitar uma análise mais conclusiva. A distribuição tem um único pico, que representa distritos com valores próximos à 0,4.
- **Valores Atípicos:** Não há valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos distritos possui valores abaixo de 0,40659. O grau médio entre os distritos é de 0,42685, e o desvio-padrão (medida de dispersão) é relativamente elevado (0,32931). O distrito com o menor grau possui 0,0000, e o distrito com o maior grau possui 1,0000., demonstrando uma baixa amplitude. A mediana é 0,40659, encontrando considerável proximidade da média, que possui 95% de confiança de se encontrar entre 0,30813 e 0,53558.

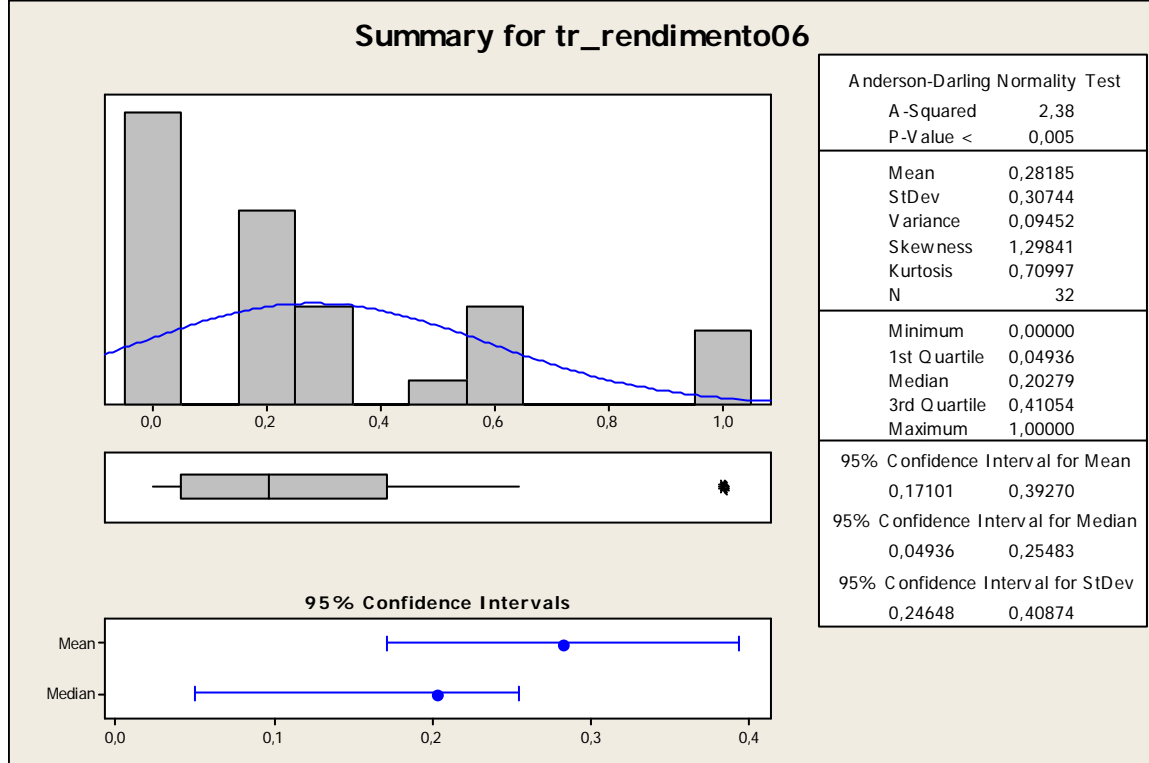
Desemprego – jovens entre 16 e 29 anos



As principais observações que podemos fazer são:

- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição levemente assimétrica para a direita. Esta conclusão está comprovada pelo teste de normalidade de Anderson-Darling que indica que a distribuição pode ser considerada uma Normal. Isto pode ser explicado pela padronização dos dados feita com o intuito de homogeneizar as variáveis e possibilitar uma análise mais conclusiva. A distribuição tem um único pico, que representa distritos com valores próximos a 0,4.
- **Valores Atípicos:** Não há valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos distritos possui valores abaixo de 0,37011. O grau médio entre os distritos é de 0,39526, e o desvio-padrão (medida de dispersão) é elevado (0,32950). O distrito com o menor grau possui 0,00000, e o distrito com o maior grau possui 1,00000, demonstrando uma baixa amplitude. A mediana é 0,37011, encontrando considerável proximidade da média, que possui 95% de confiança de se encontrar entre 0,27646 e 0,51406.

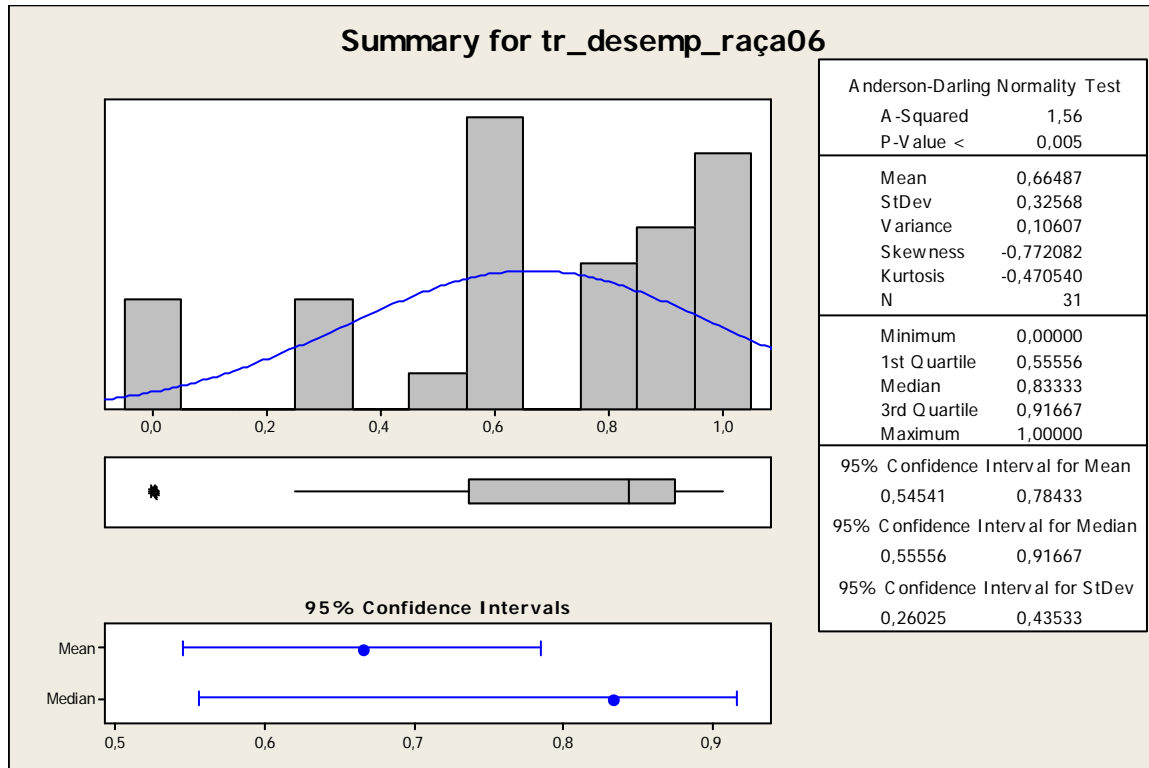
Rendimentos



As principais observações que podemos fazer são:

- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição assimétrica para a direita. Esta conclusão está comprovada pelo teste de normalidade de Anderson-Darling que indica que a distribuição não pode ser considerada uma Normal. Isto pode ser explicado pela padronização dos dados feita com o intuito de homogeneizar as variáveis e possibilitar uma análise mais conclusiva. A distribuição tem um único pico, que representa distritos com valores próximos à 0.
- **Valores Atípicos:** Há um único valor atípico.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos distritos possui valores abaixo de 0,20279. O grau médio entre os distritos é de 0,28185, e o desvio-padrão (medida de dispersão) é elevado (0,30744). O distrito com o menor grau possui 0,0000, e o distrito com o maior grau produz 1,0000, demonstrando uma pequena amplitude. A mediana é 0,20279, encontrando considerável proximidade da média, que possui 95% de confiança de se encontrar entre 0,17101 e 0,39270.

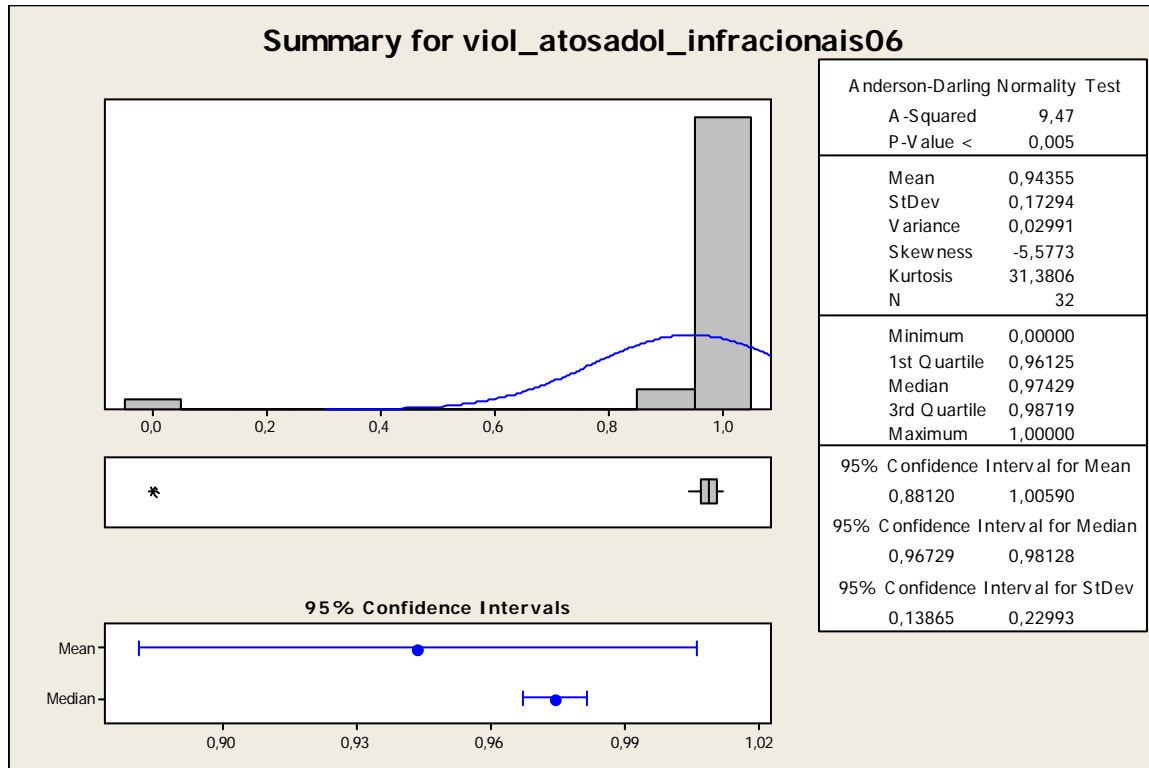
Relação de desemprego por raça



As principais observações que podemos fazer são:

- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição levemente assimétrica para a esquerda. Esta conclusão está comprovada pelo teste de normalidade de Anderson-Darling que indica que a distribuição pode ser considerada uma Normal. Isto pode ser explicado pela padronização dos dados feita com o intuito de homogeneizar as variáveis e possibilitar uma análise mais conclusiva. A distribuição tem um único pico, que representa distritos com valores próximos a 0,6.
- **Valores Atípicos:** Há um único valor atípico.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos distritos possui valores abaixo de 0,83333. O grau médio entre os distritos é de 0,66487, e o desvio-padrão (medida de dispersão) é elevado (0,32568). O distrito com o menor grau possui 0,0000, e o distrito com o maior grau possui 1,00000, demonstrando uma alta amplitude. A mediana é 0,83333, encontrando considerável proximidade da média, que possui 95% de confiança de se encontrar entre 0,54541 e 0,78433.

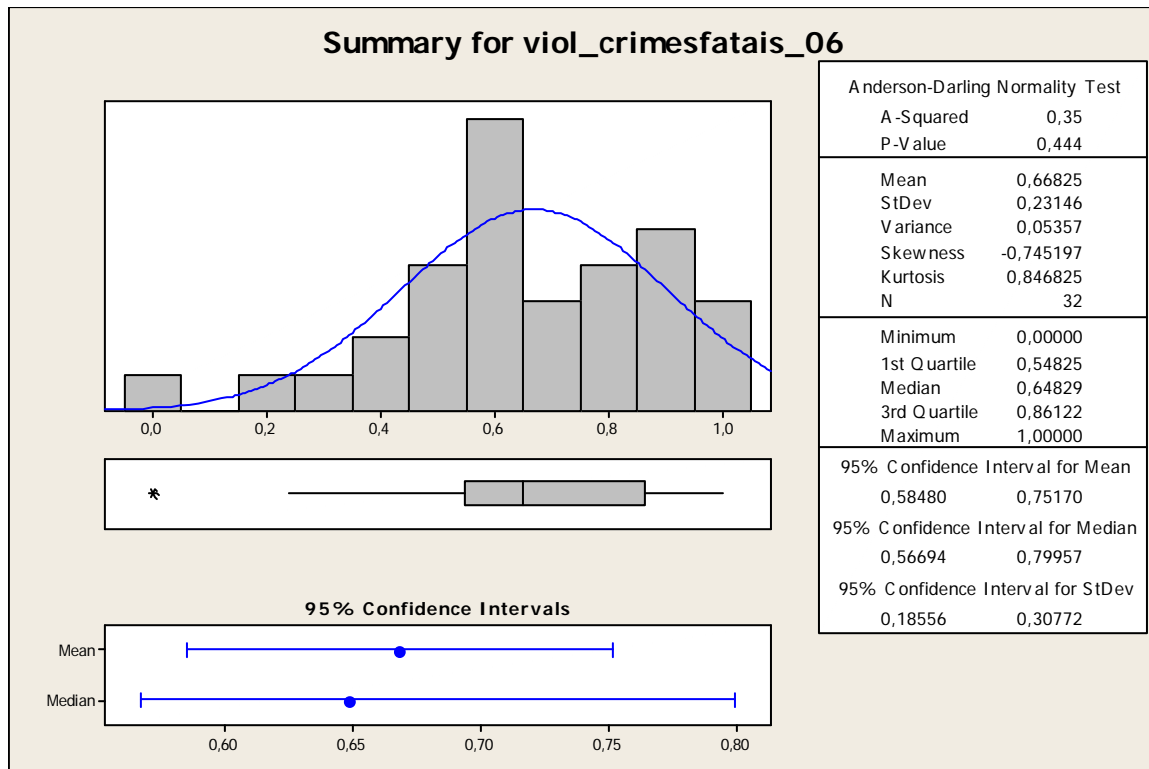
Atos infracionais



As principais observações que podemos fazer são:

- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição fortemente assimétrica para a esquerda. Esta conclusão está comprovada pelo teste de normalidade de Anderson-Darling que indica que a distribuição não pode ser considerada uma Normal, e pelo fato da observação atípica estar muito distantes da média. A distribuição tem um único pico, que representa distritos com valor próximo a 1.
- **Valores Atípicos:** Há um único valor atípico.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos distritos possui valores abaixo de 0,96125. O grau médio entre os distritos é de 0,94355, e o desvio-padrão (medida de dispersão) é baixo (0,17294). O distrito com o menor grau possui 0,6000, e o distrito com o maior grau possui 1,00000, demonstrando uma baixa amplitude. A mediana é 0,96125, encontrando considerável proximidade da média, que possui 95% de confiança de se encontrar entre 0,88120 e 1,00590.

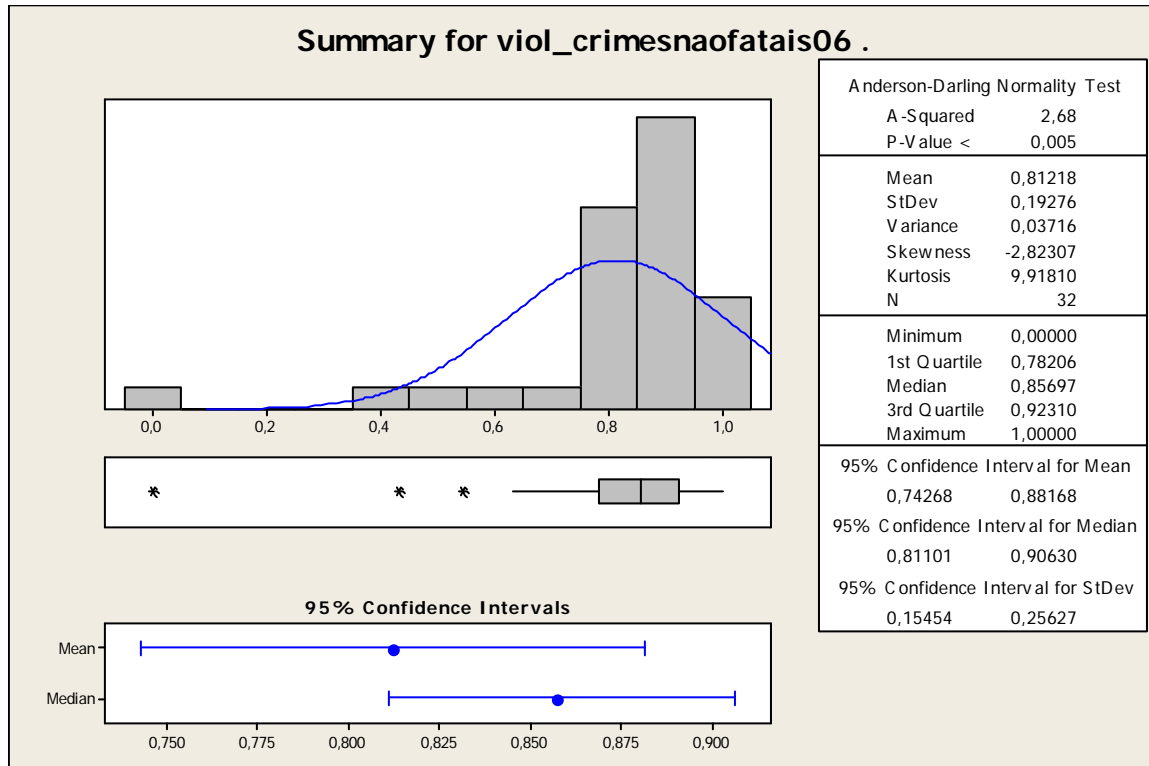
Crimes fatais



As principais observações que podemos fazer são:

- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição levemente assimétrica para esquerda. Esta conclusão está comprovada pelo teste de normalidade de Anderson-Darling que indica que a distribuição pode ser considerada uma Normal. Isto pode ser explicado pela padronização dos dados feita com o intuito de homogeneizar as variáveis e possibilitar uma análise mais conclusiva. A distribuição tem um único pico, que representa distritos com valores próximos a 0,6.
- **Valores Atípicos:** Há um único valor atípico.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos distritos possui valores abaixo de 0,64829. O grau médio entre os distritos é de 0,66825, e o desvio-padrão (medida de dispersão) é relativamente alto (0,23146). O distrito com o menor grau possui 0,00000, e o distrito com o maior grau possui 1,00000, demonstrando uma baixa amplitude. A mediana é 0,64829, encontrando considerável proximidade da média, que possui 95% de confiança de se encontrar entre 0,00000 e 1,00000.

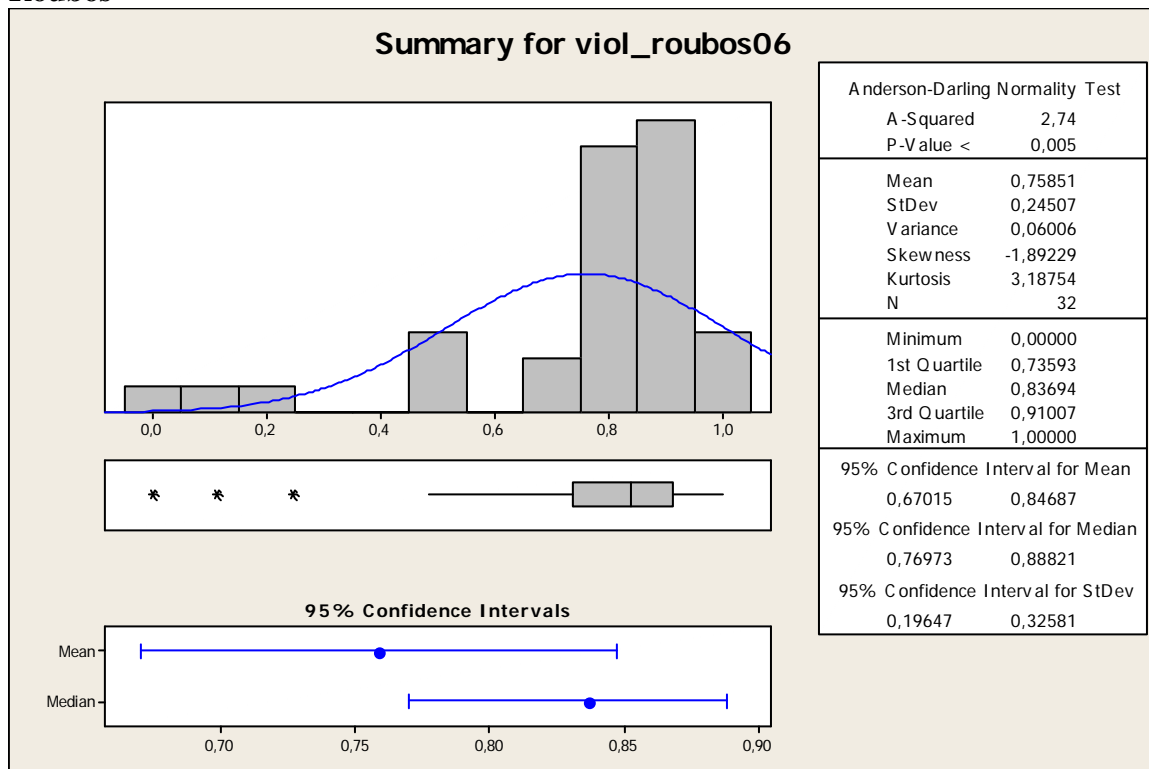
Crimes não-fatais



As principais observações que podemos fazer são:

- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição assimétrica para a esquerda. Esta conclusão está comprovada pelo teste de normalidade de Anderson-Darling que indica que a distribuição não pode ser considerada uma Normal. Isto pode ser explicado pela padronização dos dados feita com o intuito de homogeneizar as variáveis e possibilitar uma análise mais conclusiva. A distribuição tem um único pico, que representa distritos com valores próximos a 0,9.
- **Valores Atípicos:** Há três valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos distritos possui valores abaixo de 0,85697. O grau médio entre os distritos é de 0,81218, e o desvio-padrão (medida de dispersão) é alto (0,19276). O distrito com o menor grau possui 0,00000, e o distrito com o maior grau possui 1,00000, demonstrando uma baixa amplitude. A mediana é 0,85697, encontrando considerável proximidade da média, que possui 95% de confiança de se encontrar entre 0,74268 e 0,88168.

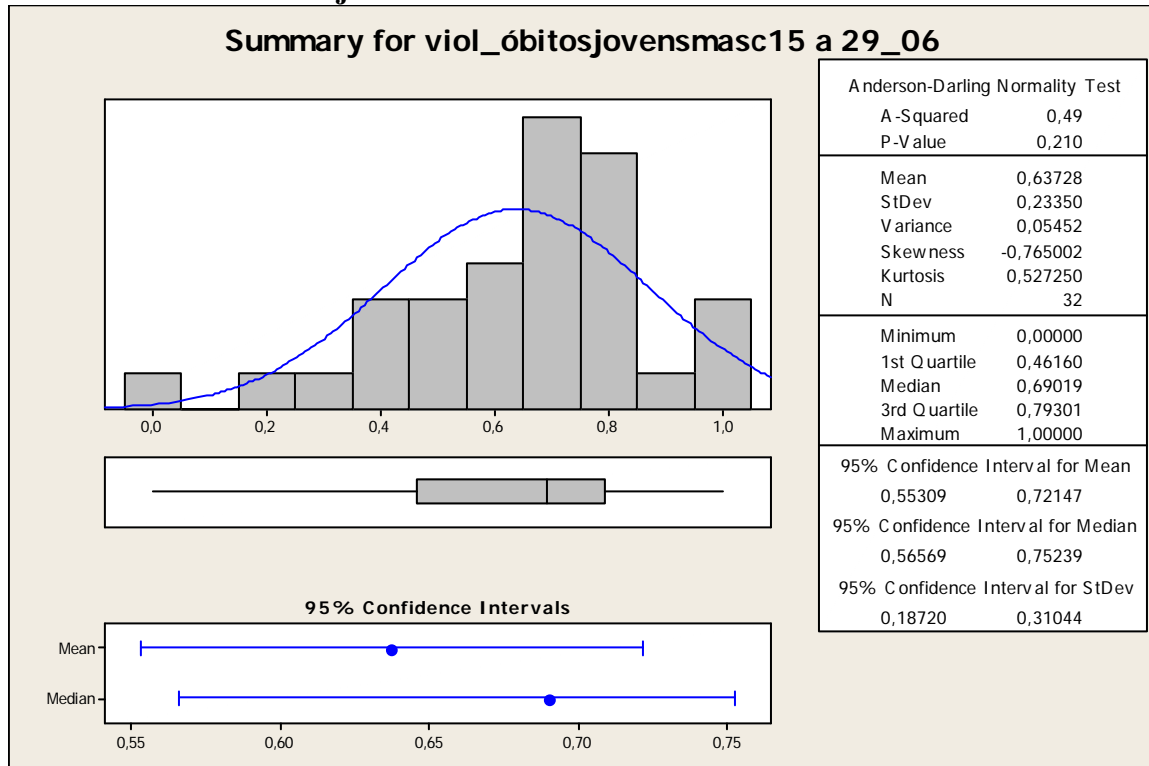
Roubos



As principais observações que podemos fazer são:

- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição assimétrica para a esquerda. Esta conclusão está comprovada pelo teste de normalidade de Anderson-Darling que indica que a distribuição pode ser considerada uma Normal. Isto pode ser explicado pela padronização dos dados feita com o intuito de homogeneizar as variáveis e possibilitar uma análise mais conclusiva. A distribuição tem um único pico, que representa distritos com valores próximos à 0,9.
- **Valores Atípicos:** Há três valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos distritos possui valores abaixo de 0,83694. O grau médio entre os distritos é de 0,75851, e o desvio-padrão (medida de dispersão) é alto (0,24507). O distrito com o menor grau possui 0,0000, e o distrito com o maior grau possui 1,00000, demonstrando uma baixa amplitude. A mediana é 0,83694, encontrando considerável proximidade da média, que possui 95% de confiança de se encontrar entre 0,67015 e 0,84687.

Número de óbitos entre jovens entre 15 e 29 anos



As principais observações que podemos fazer são:

- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição levemente assimétrica para a esquerda. Esta conclusão está comprovada pelo teste de normalidade de Anderson-Darling que indica que a distribuição pode ser considerada uma Normal. Isto pode ser explicado pela padronização dos dados feita com o intuito de homogeneizar as variáveis e possibilitar uma análise mais conclusiva. A distribuição um único pico, que representa distritos com valores próximos a 0,7.
- **Valores Atípicos:** Não há valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos distritos possui valores abaixo de 0,69019. O grau médio entre os distritos é de 0,63728, e o desvio-padrão (medida de dispersão) é alto (0,23350). O distrito com o menor grau possui 0,00000, e o distrito com o maior grau possui 1,00000, demonstrando uma baixa amplitude. A mediana é 0,69019, encontrando considerável proximidade da média, que possui 95% de confiança de se encontrar entre 0,55309 e 0,72147.

Análise das médias dos indicadores

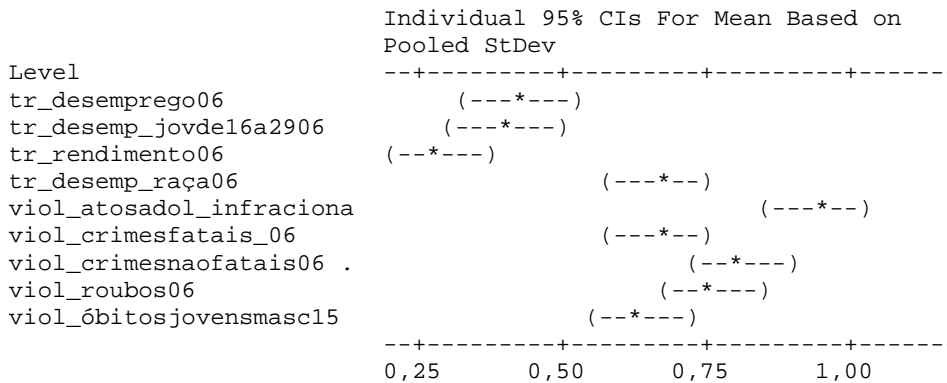
Para melhor compreensão das variáveis, foi feita a análise ANOVA para comparar as médias e os intervalos de confiança.

One-way ANOVA: tr_desempreg; tr_desemp_jo; tr_rendiment; tr_desemp_ra; ...

Source	DF	SS	MS	F	P
Factor	8	11,7610	1,4701	20,30	0,000
Error	278	20,1316	0,0724		
Total	286	31,8926			

S = 0,2691 R-Sq = 36,88% R-Sq(adj) = 35,06%

Level	N	Mean	StDev
tr_desemprego06	32	0,4269	0,3293
tr_desemp_jovdel16a2906	32	0,3953	0,3295
tr_rendimento06	32	0,2819	0,3074
tr_desemp_raça06	31	0,6649	0,3257
viol_atosadol_infraciona	32	0,9435	0,1729
viol_crimesfatais_06	32	0,6682	0,2315
viol_crimesnaofatais06 .	32	0,8122	0,1928
viol_roubos06	32	0,7585	0,2451
viol_óbitosjovensmasc15	32	0,6373	0,2335



Pooled StDev = 0,2691

Análise de componentes principais

Na busca da criação de um modelo a partir das variáveis, torna-se importante a busca da substituição de indicadores por componentes, para reduzir o número de variáveis através de índices substitutos.

Segue abaixo a análise de componentes principais:

Principal Component Analysis: tr_desempreg; tr_desemp_jo; tr_rendiment; tr_dese

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

Eigenvalue	4,9719	1,4269	0,9957	0,7430	0,3871	0,2600	0,1244	0,0764
Proportion	0,552	0,159	0,111	0,083	0,043	0,029	0,014	0,008
Cumulative	0,552	0,711	0,822	0,904	0,947	0,976	0,990	0,998

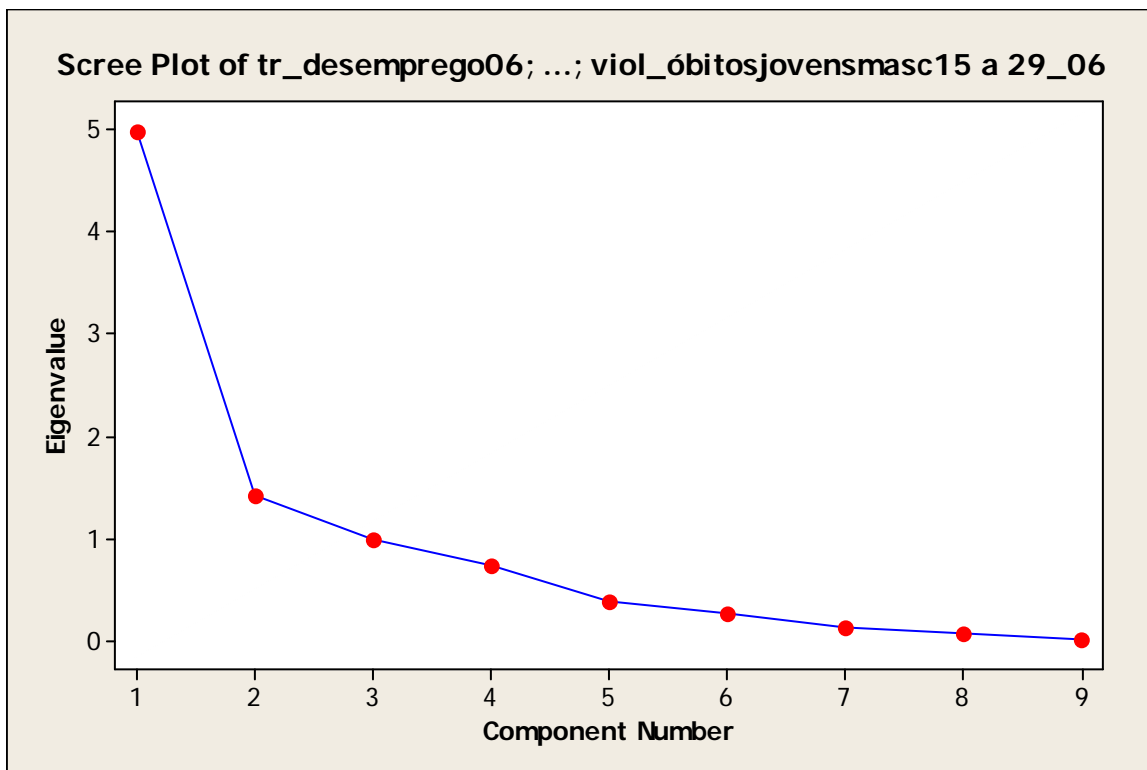
Eigenvalue	0,0146
Proportion	0,002
Cumulative	1,000

Variable	PC1	PC2	PC3	PC4	PC5	PC6
----------	-----	-----	-----	-----	-----	-----

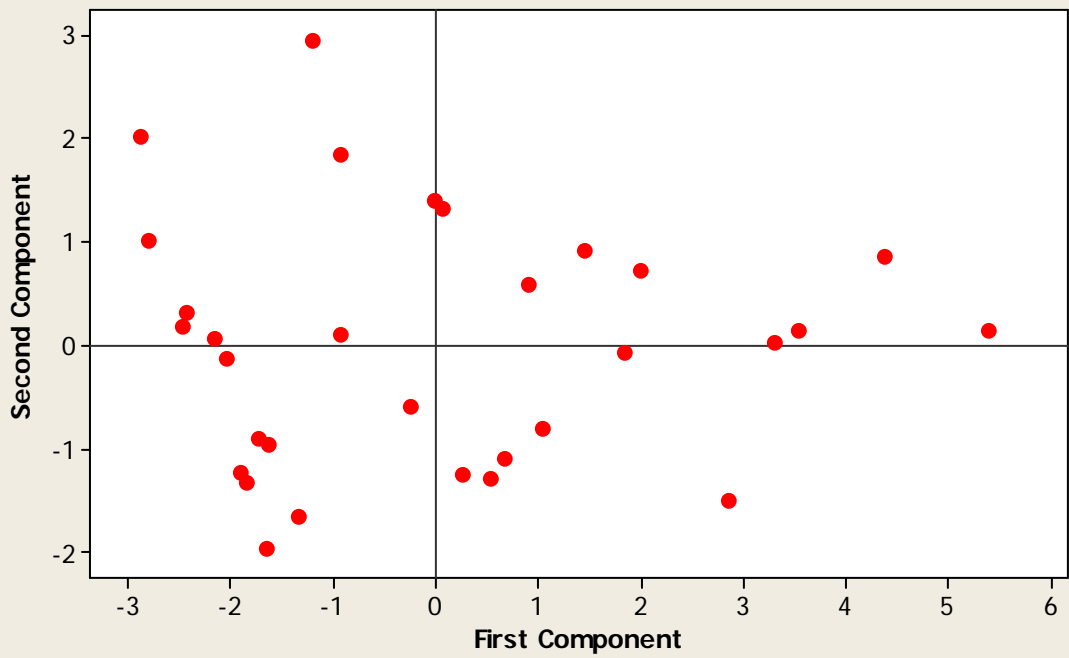
tr_desemprego06	0,425	0,053	-0,029	0,251	-0,151	0,343
tr_desemp_jovdel16a2906	0,410	0,120	-0,058	0,258	-0,322	0,353
tr_rendimento06	0,420	0,085	-0,049	0,212	-0,106	-0,291
tr_desemp_raça06	-0,351	-0,298	-0,024	-0,310	-0,567	0,441
viol_atosadol_infracionais06	0,015	-0,172	-0,978	-0,023	0,090	-0,005
viol_crimesfatais_06	0,198	-0,686	0,170	0,186	0,226	0,295
viol_crimesnaofatais06 .	-0,303	-0,298	-0,001	0,618	-0,450	-0,443
viol_roubos06	-0,366	0,010	-0,010	0,488	0,506	0,325
viol_óbitosjovensmasc15 a 29_06	0,294	-0,545	0,085	-0,273	0,151	-0,298

Variable	PC7	PC8	PC9
tr_desemprego06	0,173	0,055	0,761
tr_desemp_jovdel16a2906	0,214	0,302	-0,618
tr_rendimento06	-0,044	-0,808	-0,134
tr_desemp_raça06	0,161	-0,386	-0,002
viol_atosadol_infracionais06	-0,060	0,035	0,005
viol_crimesfatais_06	-0,529	-0,024	-0,084
viol_crimesnaofatais06 .	0,030	0,177	0,077
viol_roubos06	0,441	-0,254	-0,085
viol_óbitosjovensmasc15 a 29_06	0,646	0,074	-0,021

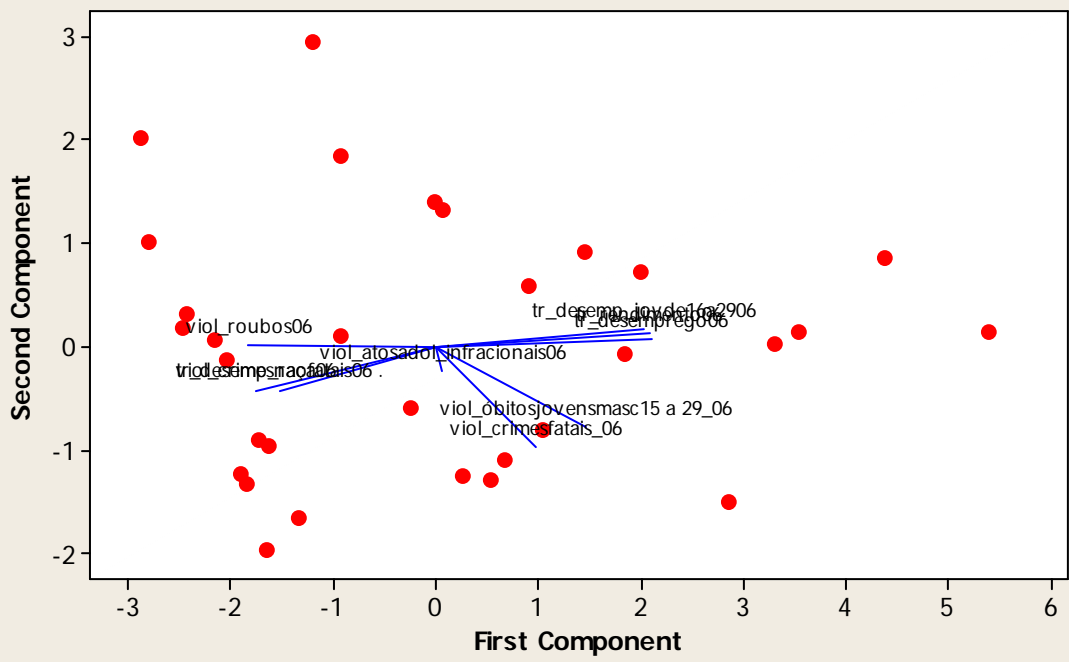
Considerando o Eigenvalue mínimo de 1,000 para a escolha das componentes, chegamos à duas componentes que explicam 82,2 % das variáveis. A primeira componente possui predominantemente cargas positivas, enquanto a segunda possui equilibrada distribuição de sinais. Enquanto a primeira componente possui pesos equilibrados entre todas as variáveis, a segunda possui pesos maiores em variáveis relacionadas com a violência urbana. Abaixo alguns gráficos que melhor ilustram estas conclusões:

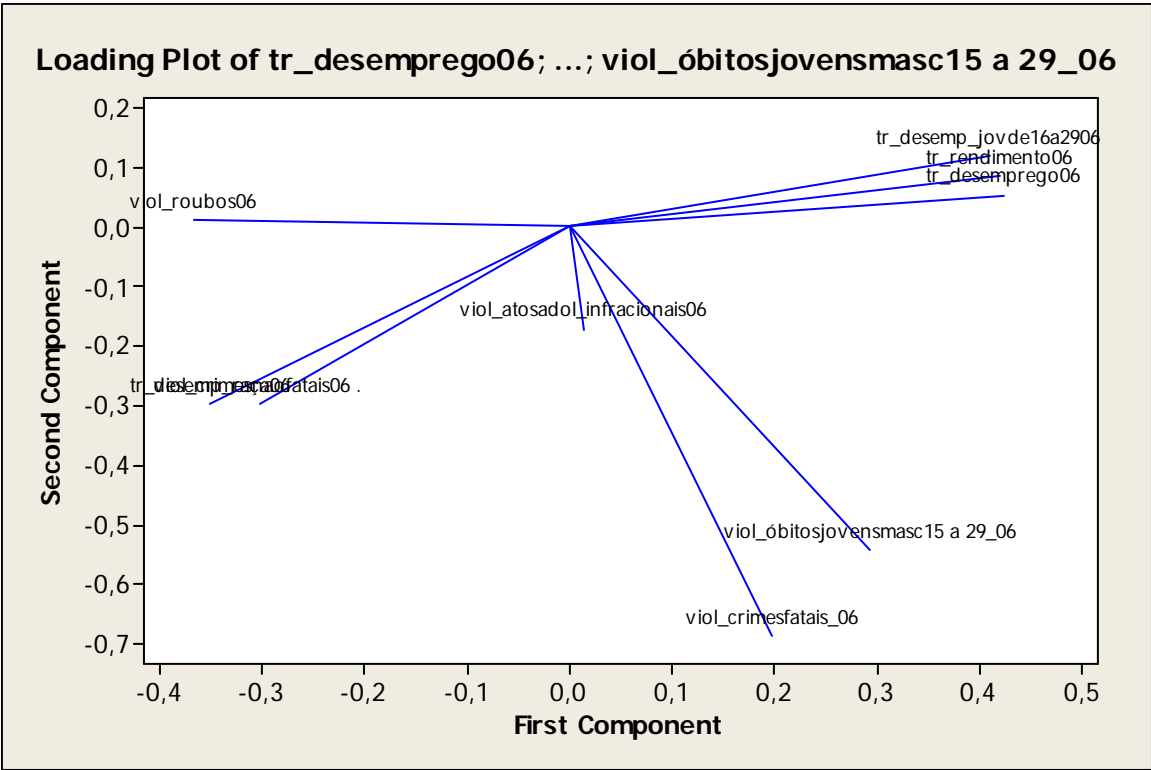


Score Plot of tr_desemprego06; ...; viol_óbitosjovensmasc15 a 29_06

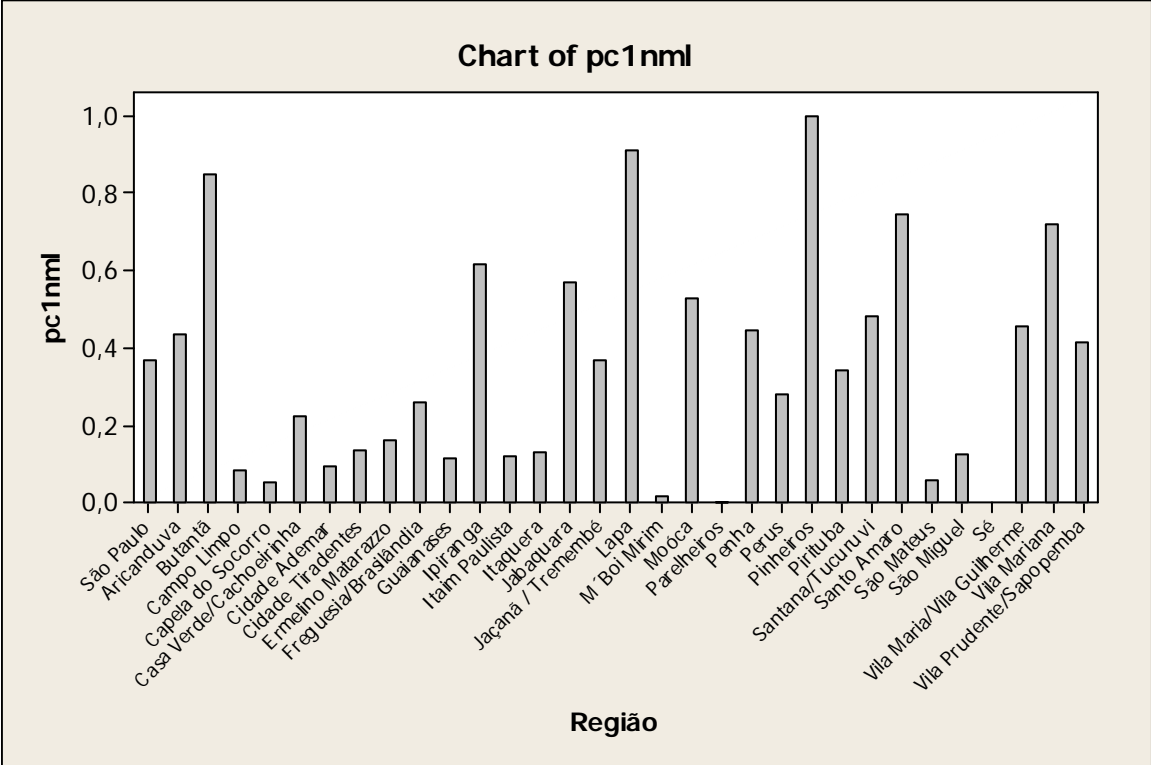


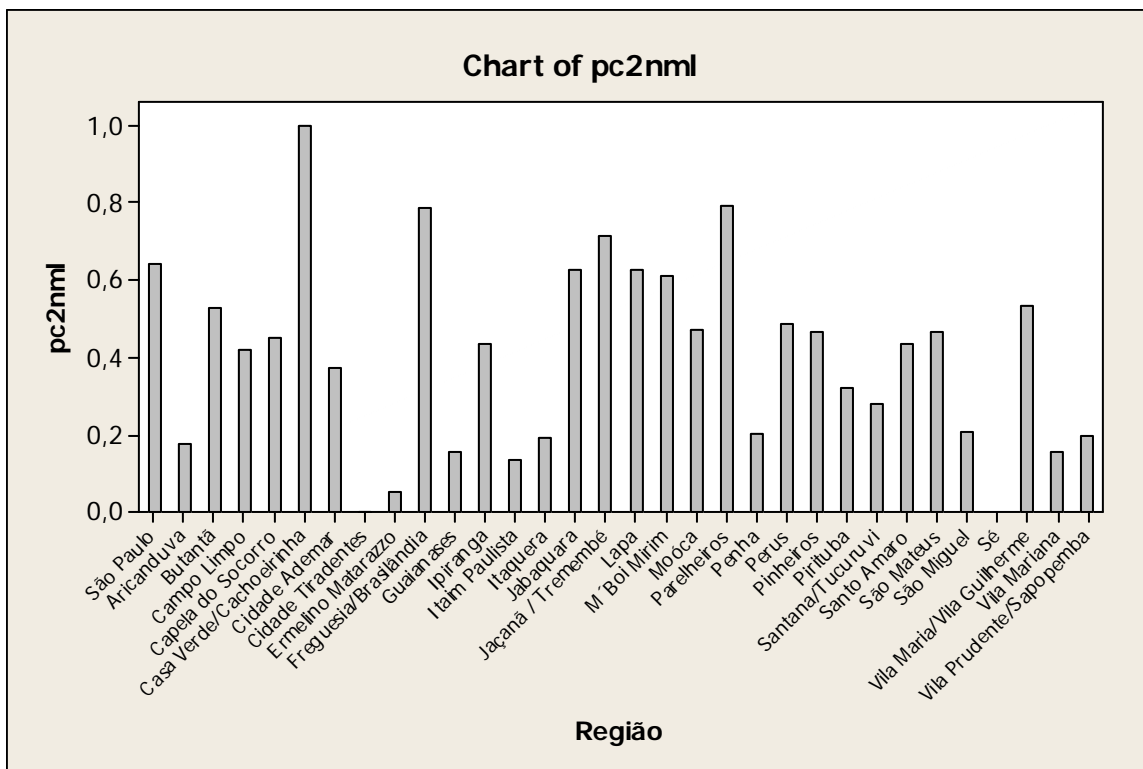
Biplot of tr_desemprego06; ...; viol_óbitosjovensmasc15 a 29_06





Abaixo temos o gráfico de barras com os índices pc1 e pc2 calculados e normalizados.





Podemos estabelecer PC1 como um índice que mede a relação entre assassinatos e desemprego, enquanto a PC2 ilustraria a violência urbana.

Conclusões finais

Há uma clara relação entre desemprego e violência, sendo que o cálculo do índice PC1 permite a explicitação desta relação, possibilitando a identificação das regiões que necessitam de uma melhoria mais urgente nas condições de empregabilidade. Podem até serem feitos estudos que analisem a relação destes índices com indicadores de educação, buscando uma relação causa-efeito entre eles. Já PC2 possibilita a quantificação da criminalidade nas regiões estudadas, possibilitando um melhor planejamento de ações de segurança.