



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA

ANÁLISE DA DISTRIBUIÇÃO DE RENDA NO SETOR TURISMO DO NORDESTE

BÁRBARA LOPES FRANCO

Brasília

2013

BÁRBARA LOPES FRANCO

ANÁLISE DA DISTRIBUIÇÃO DE RENDA NO SETOR TURISMO DO NORDESTE

Relatório apresentado à disciplina Estágio Supervisionado II do curso de graduação em Estatística, Instituto de Ciências Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Orientador: Prof. Dr. Raul Y. Matsushita

Brasília

2013

Dedico este trabalho aos meus pais, pelo apoio e dedicação por toda a vida, aos meus irmãos e todos os familiares.

Bárbara Lopes Franco

AGRADECIMENTOS

Agradeço primeiramente a Deus, por me dar força e iluminar meu caminho ao longo dessa longa caminhada.

A toda a minha família pelo amor e apoio.

Ao professor Raul Yukihiro Matsushita por sua orientação.

Ao Instituto de Pesquisa Econômica Aplicada, por ter cedido o banco de dados para realização das análises.

À minha equipe de trabalho, em especial à Margaria Hatem pelo apoio e por todo conhecimento a mim passado.

Ao SAS *Institute* Brasil por possibilitar a utilização desse software por meio de parceria acadêmica com o Departamento de Estatística da UnB.

LISTA DE FIGURAS

Figura 1: estimativas kernel das densidades para o logaritmo da renda para cada categoria de gênero nos anos de 2007 a 2010, nordeste, em que GENEROCAT = 1 se refere aos homens e GENEROCAT = 2 se refere às mulheres.	16
Figura 2: estimativas kernel das densidades para o logaritmo da renda para cada categoria de grau de instrução nos anos de 2007 a 2010, nordeste, em que GRINSTRUCAOCAT é a variável referente a grau de instrução, de modo que ‘1’ se refere aos trabalhadores com nível de escolaridade até 4ª série, ‘2’ aos trabalhadores com nível de escolaridade de 5ª a 8ª série e 2º grau incompleto, ‘3’ aos trabalhadores com 2º grau completo e superior incompleto e ‘4’ aos trabalhadores ensino superior completo.	17
Figura 3: estimativas kernel das densidades para o logaritmo da renda para cada categoria de porte do estabelecimento nos anos de 2007 a 2010, nordeste, em que TAMESTABCAT é a variável que se refere ao tamanho do estabelecimento, de modo que ‘1’ se refere aos estabelecimentos com até 9 empregados, ‘2’ aos estabelecimentos de 10 a 99 empregados, ‘3’ estabelecimento com 100 a 499 empregados e ‘4’ aos estabelecimentos com 500 ou mais empregados.	18
Figura 4: estimativas kernel das densidades para o logaritmo da renda para cada categoria de horas contratuais nos anos de 2007 a 2010, nordeste, em que HORASCONTRCAT é a variável que se refere a horas contratuais, de modo que ‘1’ denota trabalhadores que cumprem até 20 horas, ‘2’ representa os trabalhadores que cumprem de 21 a 40 horas e ‘3’ diz respeito aos trabalhadores que cumprem 41 horas ou mais.	19
Figura 5: estimativas kernel das densidades para o logaritmo da renda para cada categoria de idade nos anos de 2007 a 2010, nordeste, em que IDADECAT é a variável referente a idade, sendo que ‘1’ se refere aos trabalhadores com até 24 anos, ‘2’ representa os trabalhadores com idade de 25 a 49 anos e ‘3’ denota os trabalhadores com 50 anos ou mais.	20
Figura 6: curvas de Lorenz relativos à distribuição de renda do setor de turismo do nordeste em 2007 (superior à esquerda), 2008 (superior à direita), 2009 (inferior à esquerda) e 2010 (inferior à direita).	25
Figura 10: Relação do índice de Theil com o coeficiente de assimetria, Nordeste.	32
Figura 11: Relação do índice de Theil com o coeficiente de variação, Nordeste.	33
Figura 12: Relação do índice de Gini com o índice de Atkinson, Nordeste.	33
Figura 13: Relação do índice de Gini com a curtose, Nordeste.	34

Figura 14: Relação do índice de Gini com o coeficiente de assimetria, Nordeste	34
Figura 15: Relação do índice de Gini com o coeficiente de variação, Nordeste.....	35
Figura 16: Relação do índice de Atkinson com a curtose, Nordeste.....	35
Figura 17: Relação do índice de Atkinson com o coeficiente de assimetria, Nordeste.....	36
Figura 18: Relação do índice de Atkinson com o coeficiente de variação, Nordeste	36
Figura 19: Relação da curtose com o coeficiente de assimetria, Nordeste.....	37
Figura 21: Relação do coeficiente de assimetria com o coeficiente de variação, Nordeste	38

LISTA DE TABELAS

Tabela 1: Distribuição de ocupados por grupo, Nordeste	13
Tabela 2: Índice de Gini, Nordeste.....	21
Tabela 3: Índice de Theil, Nordeste.....	26
Tabela 4: Coeficiente de variação, Nordeste.....	28
Tabela 5: Correlações de Pearson entre os índices, Nordeste	29

SUMÁRIO

1	INTRODUÇÃO	1
1.1	Motivação	1
1.2	A Relação Anual de Informações Sociais (RAIS)	3
1.3	Turismo	4
2	METODOLOGIA	6
2.1	Estimação da densidade pelo método do kernel	7
2.2	Índice de Gini	8
2.3	Índice de Theil	9
2.4	Índice de Atkinson	10
2.5	Coefficiente de Assimetria	11
2.6	Curtose	11
2.7	O Coeficiente de Variação	12
3	RESULTADOS E DISCUSSÕES	13
3.1	Perfil dos Ocupados	13
3.2	Distribuição Empírica da Renda	14
3.3	Análise dos indicadores de desigualdade	21
3.3	Comparações entre os índices	29
4	CONCLUSÃO	38
5	REFERÊNCIAS BIBLIOGRÁFICAS	41
	APÊNDICE A – Lista de CNAES referentes ao Turismo	43
	APÊNDICE B	48
	APÊNDICE C	53
	APÊNDICE D	57
	APÊNDICE E	60
	APÊNDICE F	63
	APÊNDICE G	64

1 INTRODUÇÃO

1.1 Motivação

A inovação tecnológica – um dos braços da globalização – proporcionou significativa redução de custos de transporte e de comunicação, e tem impulsionado o rápido crescimento do turismo de massa em todo o mundo (Swarbrooke e Horner, 2002). Isso permite a utilização do turismo como uma estratégia de desenvolvimento econômico, o que pode ser importante não apenas para países menos favorecidos economicamente (Ajala, 2008), mas também para o Brasil¹.

Como o turismo é o maior empregador de mão de obra no setor terciário na economia mundial, estrategicamente, tanto as organizações públicas como as privadas devem avaliar os impactos econômicos do turismo em âmbito nacional, estadual e local, como a necessidade de investimento em infraestruturas, superestruturas e a geração de emprego e renda (Stynes, 1997). Independentemente de a atração turística ser natural ou artificial, o destino turístico deve oferecer uma infraestrutura adequada em transportes (estradas, ferrovias, portos, aeroportos, estacionamentos), serviços de utilidade pública (eletricidade, água e comunicações) e outros serviços (saúde e segurança). Já a superestrutura é constituída por atividades do setor privado como a hospedagem, o comércio varejista e todos os demais serviços. Em cada uma dessas atividades associadas diretamente ou indiretamente com o turismo, espera-se gerar emprego e renda à medida que esse setor cresce (Cooper, 2001).

Hoje, o mundo está mais atento ao Brasil, por diversos motivos. Mais do que ser um país do futebol e do carnaval, ele se destaca no cenário mundial como potência emergente, ao lado de Rússia, Índia e China. Aos poucos, o país tornou-se mais confiável para receber investimentos externos. Recentemente, o Brasil sediou a Conferência das Nações Unidas sobre Desenvolvimento Sustentável (CNUDS) — conhecida também como Rio+20, em 2012 na cidade do Rio de Janeiro — e a Copa das Confederações da FIFA em junho de 2013, e

¹ veja, por exemplo, o Plano Nacional de Turismo 2011-2015 do Ministério do Turismo em http://www.turismo.gov.br/export/sites/default/turismo/o_ministerio/publicacoes/downloads_publicacoes/Documento_Referencial_Turismo_no_Brasil_2011-2014.pdf

receberá a Jornada Mundial da Juventude e será o anfitrião da Copa do Mundo de FIFA em 2014 e dos Jogos Olímpicos e Paraolímpicos em 2016. Localmente e momentaneamente, esses eventos atraem turistas, mas espera-se que eles ajudem a despertar a curiosidade do mundo para que se aumente o fluxo perene de turistas para o Brasil.

Mas os recursos econômicos necessários para investimento em turismo são finitos, de modo que pode haver transferência de recursos de outros setores já existentes para os setores associados ao turismo. Se essa transferência representa uma distribuição de riquezas, será que houve redução da desigualdade na distribuição da renda nesse setor à medida que o país cresceu economicamente nesses últimos anos?

Pant (2011) observou que no setor de turismo a desigualdade de renda sofreu reduções em diversos países, incluindo o Brasil. Ele também concluiu que, em comparação com o turismo internacional, o doméstico contribui mais para a redução da desigualdade na distribuição de renda.

Neste trabalho investigaremos esse assunto com base nos dados da Relação Anual de Informações Sociais (RAIS), considerando o perfil de trabalhadores empregados em atividades ligadas diretamente ou indiretamente com o setor do turismo na região nordeste do Brasil. Estudaremos o padrão de desigualdade da distribuição da renda com a utilização de indicadores clássicos como o índice de Gini, de Theil, de Atkinson, e os coeficientes de assimetria, de curtose e o de variação.

O restante do trabalho se organiza da seguinte forma. A Seção 1.2 apresenta brevemente a RAIS e a Seção 1.3 descreve o setor do turismo para se delimitar as atividades características desse setor para o nosso estudo. Na Seção 2 trata da metodologia, contemplando a estimação de densidades por kernel e os índices de Gini de Theil, de Atkinson, e os coeficientes de assimetria, de curtose e de variação. Na Seção 3 são apresentados os resultados e a análise, e a Seção 4 conclui o trabalho. O Apêndice A mostra a relação das atividades econômicas relacionadas ao turismo que foram considerados nesse estudo, com base na classificação nacional de atividades econômicas (CNAE), e os Apêndices

B, C, D, E, F e G apresentam parte do programa SAS utilizado para a operacionalização das estatísticas.

1.2 A Relação Anual de Informações Sociais (RAIS)

A Relação Anual de Informações Sociais (RAIS) é uma base de dados a partir da qual é possível descrever, analisar e avaliar o mercado formal do trabalho no Brasil. Ela representa um censo do mercado formal de trabalho, uma vez que todos os estabelecimentos legalmente constituídos devem, anualmente, fornecer ao Ministério do Trabalho e Emprego (MTE), por meio da RAIS, as informações referentes a cada um de seus empregados.

Assim, são obrigados a declarar os estabelecimentos inscritos no cadastro nacional de pessoa jurídica (CNPJ) — caso o estabelecimento não tenha empregados ou mantenha suas atividades paralisadas durante o ano-base, ele deve entregar a RAIS negativa —; os empregadores (conforme a CLT); as pessoas jurídicas de direito privado (inclusive as empresas públicas domiciliadas no país, com registro, ou não, nas juntas comerciais, no Ministério da Fazenda, nas secretarias de finanças ou da fazenda dos governos estaduais e nos cartórios de registro de pessoa jurídica); as empresas individuais (inclusive aquelas que não possuem empregados); os cartórios extrajudiciais e consórcios de empresas; os empregadores urbanos pessoas físicas (autônomos e profissionais liberais que mantiveram empregados no ano-base); os órgãos da administração direta e indireta dos governos federal, estadual ou municipal (inclusive as fundações supervisionadas e as entidades criadas por lei, com atribuições de fiscalização do exercício das profissões liberais); os condomínios e as sociedades civis; os empregadores rurais pessoas físicas que mantiveram empregados no ano-base; as filiais, agências, sucursais, representações ou quaisquer outras formas de entidades vinculadas à pessoa jurídica domiciliada no exterior.

Cada empresa deve relacionar, no que lhe for pertinente: *a)* os empregados contratados por empregadores, pessoa física ou jurídica, sob o regime da CLT, por prazo indeterminado ou determinado, inclusive a título de experiência; *b)* os servidores da administração pública direta ou indireta, federal, estadual ou municipal, bem como das fundações supervisionadas; *c)* os trabalhadores avulsos (aqueles que prestam serviços de natureza urbana ou rural a

diversas empresas, sem vínculo empregatício, com a intermediação obrigatória do órgão gestor de mão-de-obra, nos termos da lei ou do sindicato da categoria); *d*) os empregados de cartórios extrajudiciais; trabalhadores temporários, conforme a lei; *e*) trabalhadores com contrato de trabalho por prazo determinado, segundo a legislação; *f*) os diretores sem vínculo empregatício, para os quais o estabelecimento tenha optado pelo recolhimento do FGTS; *g*) os servidores públicos não-efetivos (demissíveis *ad nutum* ou admitidos por meio de legislação especial, não regidos pela CLT); *h*) os trabalhadores regidos pelo Estatuto do Trabalhador Rural (Lei nº 5.889/1973); *h*) o aprendiz (maior de 14 anos e menor de 24 anos), contratado nos termos da legislação pertinente; *i*) os servidores e trabalhadores licenciados; servidores públicos cedidos e requisitados; *j*) dirigentes sindicais.

Por outro lado, não devem ser relacionados os diretores sem vínculo empregatício para os quais não se recolhe o FGTS, os autônomos os trabalhadores eventuais, os ocupantes de cargos eletivos (como os governadores, deputados, prefeitos, vereadores) — desde que não tenham feito opção pelos vencimentos do órgão de origem a partir da data da posse —, os estagiários, os empregados domésticos, os cooperados e os cooperativados.

A seção seguinte descreve o setor do turismo para que possamos delimitar as atividades características desse setor para o nosso estudo.

1.3 Turismo

De acordo com as recomendações internacionais para estatísticas do turismo — IRTS 2008 (UNWTO, 2010) — existem dois métodos para mensurar o emprego relacionado ao turismo. O primeiro considera todas as ocupações nas atividades características do turismo (ACTs), independentemente de essas ocupações estarem relacionadas ao turista. O segundo consiste em contabilizar apenas o emprego estritamente relacionado aos bens e serviços adquiridos por visitantes de acordo com a nomenclatura da Organização Mundial do Turismo (OMT), o que representa o “emprego nas indústrias do turismo ou emprego nas ACTs”, mas que não se restringe apenas às atividades características do turismo.

De acordo com o primeiro método, para dimensionar o mercado de trabalho no turismo, consideram-se, por exemplo, todos os empregos na ACT *Alimentação*. Pelo segundo método, são contabilizados apenas aqueles empregos relacionados aos bens e serviços consumidos pelos turistas, ou seja, apenas uma parcela dos empregos na ACT *Alimentação* (Sakowski, 2013).

Portanto, é importante definir claramente quais são os critérios para se delimitar as atividades que compõem o setor. A inclusão ou exclusão de uma única atividade pode gerar estimativas bastante diferentes, causando divergências nos números do mercado de trabalho do turismo (Sakowski, 2013).

Aa primeira pesquisa de campo² do IPEA – Instituto de Pesquisa Econômica Aplicada - em 2004/2005 acerca do emprego no turismo seguiu as recomendações da OMT dessa época, e contemplou sete atividades características (ACTs): *alojamento, transporte, alimentação, agências de viagem, auxiliar de transporte, aluguel de transporte e cultura e lazer* (Árias e Zamboni, 2009).

A partir de 2006, o IPEA considerou as mudanças de CNAE — quando houve mudança da CNAE 1.0 para a CNAE 2.0. Essa mudança permitiu delimitar com maior precisão as ACTs do turismo no Brasil. A atividade *Transporte*, foi subdividida em três ACTs: *transporte aéreo; transporte terrestre e transporte aquaviário* (Sakowski, 2013).

Houve um esforço para adequar as CNAES às novas recomendações internacionais para estatísticas do turismo (IRTS 2008) de modo a assegurar a acurácia na delimitação das atividades características do turismo, contando com o apoio da OMT, do Ministério do Turismo e do IBGE.

Entre as mudanças ocorridas no recorte de CNAEs para adequação às recomendações da OMT, vale ressaltar a exclusão da ACT *auxiliar de transporte*, de CNAEs referentes a *catering* e de CNAES associadas ao *transporte local*. Essa última alteração ensejou a junção

² http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=2971&Itemid=33

das ACTs *transporte terrestre internacional, interestadual e intermunicipal* e *transporte local* na ACT *transporte terrestre*.

Assim, atualmente, o recorte final do setor turismo delimitado pelo Ipea contempla oito ACTs: *alojamento, transporte aéreo, transporte terrestre, transporte aquaviário, alimentação, agências de viagem, aluguel de transporte e cultura e lazer* (Sakowski, 2013).

Em nosso estudo, utilizaremos o método em que consideram todas as ocupações nas atividades características do turismo (ACTs), independentemente de essas ocupações estarem relacionadas ao consumo de bens e serviços dos turistas.

Com base nos dados da RAIS, foram selecionados apenas os trabalhadores empregados em dezembro dos anos de referência 2007 a 2010 que apresentaram rendas não nulas. Além disso, concentraremos nosso estudo na região nordeste, por ser, tradicionalmente, um pólo de atração do turismo nacional e internacional.

O Apêndice A mostra a relação das atividades econômicas relacionadas ao turismo que foram considerados nesse estudo, com base na classificação nacional de atividades econômicas (CNAE).

2 METODOLOGIA

Neste trabalho estudaremos o padrão de desigualdade da distribuição da renda do setor do turismo no nordeste mediante uso de indicadores clássicos como o índice de Gini (Seção 2.2), de Theil (Seção 2.3), de Atkinson (Seção 2.4), e os coeficientes de assimetria (Seção 2.5), de curtose (Seção 2.6) e o de variação (Seção 2.7).

Para o estudo descritivo da distribuição do logaritmo da renda utilizaremos o método de suavização por kernel (Seção 2.1) para a estimação da sua função de densidade.

Os cálculos computacionais e a produção dos gráficos serão feitos no software SAS³ 9.3, e parte da programação nesse ambiente se encontra nos Apêndices B, C, D, E, F e G.

2.1 Estimação da densidade pelo método do kernel

Seja X_1, \dots, X_n uma amostra aleatória simples extraída de uma população com função de densidade $f(x)$. Define o estimador de $f(x)$ pelo método do kernel como (Silverman, 1986):

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (1.1)$$

em que, $h > 0$ é a constante de suavização, K é uma função real positiva que se denomina kernel, e que satisfaz a condição $\int_{-\infty}^{\infty} K(x) dx = 1$.

A escolha mais comum da função K é o kernel gaussiana dado por (Sheater, 2004)

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \quad (1.2)$$

Note que esse kernel satisfaz à condição

$$\int_{-\infty}^{\infty} K(x) dx = 1, \quad (1.3)$$

de modo que $K(x)$ é uma função de densidade.

Se a constante de suavização h for muito pequena, a curva será mais rugosa, tendendo a interpolar perfeitamente os dados. Por outro lado, um valor de h muito grande proporciona maior suavidade, porém com maior possibilidade de viés. Portanto, a determinação da constante h deve considerar um equilíbrio entre a rugosidade e o viés. Por exemplo, Scott (1992) recomenda minimizar o erro quadrático médio integrado assintótico (AMISE, *asymptotic mean integrated square error*) em relação a h , em que

³ www.sas.com

$$AMISE(\hat{f}) = \frac{1}{4}h^4 \left(\int_t t^2 K(t) dt \right)^2 \int_x (f''(x))^2 dx + \frac{1}{nh} \int_t K(t)^2 dt, \quad (1.4)$$

Com base nesse critério, estimaremos as funções de densidade do logaritmo das rendas dos trabalhadores do setor de turismo do nordeste.

2.2 Índice de Gini

O índice de Gini foi proposto em 1914 por Corrado Gini. O intervalo de resultados possíveis para esse índice varia de zero a um.

O valor nulo representa o grau máximo de igualdade, e só ocorrerá se todas as unidades apresentarem a mesma renda (M/n), em que M representa o montante total da renda e n representa o tamanho da população.

Por outro lado, quando o índice for igual a um, ele remete ao grau máximo de desigualdade, o que ocorre se apenas um indivíduo detiver toda a renda (M), tendo as demais unidades renda zero (Jenkins e Van Kerm, 2008).

O índice de Gini é um indicador clássico que se associa diretamente com a curva de Lorenz. Essa curva é desenhada pelo gráfico de dispersão entre a distribuição cumulada do rendimento e a distribuição acumulada da população. Nesse gráfico, se a distribuição for equânime, a curva de Lorenz será uma reta de 45° a partir da origem. Em contraste, se um único indivíduo concentrar toda a renda (total desigualdade), a curva ficará ao longo do eixo horizontal (Jenkins e Van Kerm, 2008).

Para se definir o índice de Gini, primeiramente considere uma variável aleatória discreta X_i ($i = 1, \dots, n$) cujos valores se encontrem em ordem crescente. Ao admitir que os n valores sejam igualmente prováveis, observa-se que a proporção acumulada do número de elementos até o i -ésimo elemento é dada por (Hoffman, 1980)

$$P = \frac{i}{n}, \quad (2.1)$$

em que $i = 1, \dots, n$. Assim, a correspondente proporção cumulada da renda até o i -ésimo elemento é:

$$\phi_i = \frac{\sum_{j=1}^i X_j}{\sum_{j=1}^n X_j}. \quad (2.2)$$

Desse modo, se $X_i < X_{i+1}$, a razão ϕ_i representa a fração da renda total apropriada pelos indivíduos com renda igual ou inferior a X_i . O índice de Gini é dado por (Hoffman, 1980):

$$G = 1 - \frac{1}{n} \sum_{i=1}^n (\phi_{i-1} + \phi_i), \quad (2.3)$$

em que ϕ_i é a proporção acumulada da renda dentro de cada estrato, $\phi_0 = 0$, e n é a quantidade de indivíduos da população.

Aqui, cabe uma observação importante. O índice de Gini não é invariante a transposições. Por exemplo, considere que o conjunto de valores $\{20, 23, 18, 25, 28, 18\}$ represente uma distribuição de renda. O índice de Gini dessa distribuição é aproximadamente igual a 0,112. Se adicionarmos uma constante, por exemplo, 10, teremos o conjunto de valores $\{30, 33, 28, 35, 38, 28\}$. Note que, embora a variância e o aspecto da distribuição permaneçam os mesmos, o coeficiente de Gini diminui para 0,077.

2.3 Índice de Theil

Segundo Hoffman (1980), o índice T de Theil também é conhecido como redundância de distribuição de renda, pois ele remete à medida de entropia da teoria da informação. O índice de Theil é medida em nits, podendo variar de 0 a $\ln(n)$. O valor nulo indica que a distribuição seja perfeitamente equânime, enquanto o extremo $\ln(n)$ representa completa desigualdade.

Considerando uma população com n indivíduos em que cada um recebe uma fração não negativa ($y_i \geq 0$, com $i = 1, \dots, n$) da renda total, de modo que

$$\sum_{i=1}^n y_i = 1, \quad (3.1)$$

define-se o índice de Theil como (Theil, 1967); Hoffman, 1980)

$$T = \sum_{i=1}^n y_i \log (ny_i). \quad (3.2)$$

Do ponto de vista da teoria da informação, a redundância corresponde à esperança do valor informativo de uma mensagem incerta, em que as probabilidades *a posteriori* são as frações da renda total apropriadas pelos indivíduos e as probabilidades *a priori* são iguais a $\frac{1}{n}$, ou seja, iguais à fração de renda total que caberia a cada indivíduo no caso de haver perfeita igualdade da distribuição (Hoffman, 1980).

2.4 Índice de Atkinson

O índice de Atkinson permite variar a sensibilidade para as desigualdades em diferentes partes da distribuição de renda. No texto *The Economics of Inequality*, Atkinson (1970) observou que a desigualdade "não pode, em geral, ser medida sem a introdução de julgamentos sociais. Medidas como o índice de Gini não são puramente estatísticas, e elas incorporam implicitamente julgamentos sobre o peso a ser anexado à desigualdade em diferentes pontos da escala de renda".

Assim, Atkinson (1970) propôs que um índice na forma:

$$A = 1 - \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i}{\bar{x}} \right)^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}}, \quad (4.1)$$

em que x_i é a renda da i -ésima pessoa, \bar{x} é a renda média, e ϵ representa o parâmetro de aversão à desigualdade, tal que $\epsilon > 0$ e $\epsilon \neq 1$.

O índice de Atkinson varia de 0 (zero) e 1 (um). Quanto mais próximo de zero estiver o índice, mais equânime será a distribuição das rendas.

Para um valor ϵ próximo de zero, o observador é indiferente à natureza da distribuição de renda, enquanto para ϵ tendendo a infinito, o observador se preocupa apenas com a posição da renda do grupo com renda mais baixa. Na prática, os valores de ϵ mais usados são 1/2, 1, 3/2 ou 2 (Jenkins e Van Kerm, 2008). Em nosso trabalho, utilizaremos $\epsilon = 2$.

2.5 Coeficiente de Assimetria

Define-se o coeficiente de assimetria como o terceiro momento central (Cramér, 1951):

$$CA = \frac{n \sum_{i=1}^n \{(x_i - \bar{x})/s\}^3}{(n-1)(n-2)}, \quad (5.1)$$

em que s representa o desvio padrão amostral e x_i é a renda do i -ésimo indivíduo.

Um número negativo implica assimetria à esquerda, o que indica maior cauda à esquerda. Se o coeficiente for zero, a distribuição é perfeitamente simétrica em torno da média (ou seja, a média, a mediana e a moda coincidem). Um valor positivo indica que a distribuição é assimétrica à direita, ou seja, que sua cauda à direita é mais pesada do que a cauda à esquerda.

O coeficiente de assimetria é uma medida útil quando, por exemplo, a distribuição de renda segue uma distribuição lognormal com três parâmetros (Bendel et al., 1989).

2.6 Curtose

A curtose descreve o achatamento da curva, sendo definida pelo quarto momento de ordem central (Cramér, 1951):

$$Ks = \frac{1}{n} \sum_{i=1}^n \{(x_i - \bar{x})/s\}^4 - 3, \quad (6.1)$$

em que s representa o desvio padrão amostral e x_i é a renda do indivíduo i .

Uma distribuição com coeficiente de curtose nula é considerada mesocúrtica, nem alongada nem achatada. Se o coeficiente de curtose for negativo, será considerada platicúrtica (achatada); e se for maior que zero, será considerada leptocúrtica (alongada).

2.7 O Coeficiente de Variação

Embora a variância represente a dispersão dos dados, seu uso para medir desigualdade de renda esbarra em um problema. Novamente, por exemplo, considere que o conjunto de valores {20, 23, 18, 25, 28, 18} represente uma distribuição de renda. Da Seção 2.2, o índice de Gini dessa distribuição é aproximadamente igual a 0,112. Se adicionarmos uma constante igual a 10, o conjunto de valores {30, 33, 28, 35, 38, 28} apresentará o coeficiente de Gini igual a 0,077. A variância de ambos os conjuntos de dados são iguais.

O problema de se usar a variância para medir a desigualdade de renda é que, se todas as rendas crescem proporcionalmente, a desigualdade aumenta. Para contornar esse problema, utiliza-se o coeficiente de variação.

O coeficiente de variação é uma medida de desigualdade calculada pela razão entre o desvio padrão da distribuição de renda pela sua média. Em sociedades com rendas igualitárias, o desvio padrão será pequeno e, assim conclui-se que quanto menor o coeficiente de variação, menor será a desigualdade.

$$CV = \frac{100 \times \sigma}{\mu}, \quad (7.1)$$

Tal medida possui algumas limitações: ela não possui um limite superior, ao contrário do índice de Gini, tornando a interpretação difícil. Os seus componentes, desvio padrão e média, podem ser muito influenciados por rendimentos muito baixos ou muito altos (De Maio, 2007). Mas a sensibilidade maior do CV aos valores da renda na cauda superior, em comparação com o índice de Gini, por exemplo, pode ser interessante (Bendel et al., 1989) quando se deseja uma medida invariante a mudança de escala. Ou seja, o CV de uma distribuição X é igual ao CV da distribuição ρX , em que $\rho > 0$. Mas o CV não é invariante à transposição. Considerando novamente o exemplo dado no início desta seção, o CV da distribuição inicial é aproximadamente igual a 0,18, e o da distribuição transposta com 10 unidades adicionais é aproximadamente 0,13.

A seguir, apresentaremos os resultados da aplicação da metodologia que aqui foi descrita nos dados de renda fornecidos pela RAIS para o setor do turismo no nordeste.

3 RESULTADOS E DISCUSSÕES

3.1 Perfil dos Ocupados

Tabela 1: Distribuição de ocupados por grupo, Nordeste

Grupo	Categoria	2007	2008	2009	2010
Gênero	Homem	60,1	59,4	58,9	57,7
	Mulher	39,9	40,6	41,1	42,3
Grau de Instrução	Até 4ª série	7,9	6,7	6,1	5,6
	De 5ª a 8ª série e segundo grau incompleto	42,2	40,1	37,7	34,9
	2º grau completo e superior incompleto	47,2	49,7	52,8	55,8
	Superior completo	2,7	3,5	3,5	3,7
Tamanho do estabelecimento	Até 9 empregados	27,2	26,9	26,9	26,4
	De 10 a 99 empregados	55,5	55,6	56,1	55,4
	De 100 a 499 empregados	14,5	13,9	13,1	13,3
	500 empregados ou mais	2,9	3,6	3,9	4,8
Horas contratuais	Até 20 horas	0,6	0,6	0,9	0,8
	De 21 a 40 horas	4,5	4,4	4,4	4,4
	41 horas ou mais	94,9	95,0	94,7	94,8
Idade	Ate 24 anos	18,3	18,4	17,7	17,6
	De 25 a 49 anos	75,0	74,7	75,1	74,8
	50 anos ou mais	6,7	6,9	7,2	7,6

Em dezembro do ano de 2007 haviam 218.986 empregados formais no turismo; em 2008, 234.422 empregados; em 2009, 252.647 empregados e em 2010, 281.241 empregados.

A Tabela 1 mostra que, de todos os empregados no turismo no nordeste, em todos os anos, a maioria era do sexo masculino. Observa-se, porém, um sensível aumento na participação das mulheres como mão-de-obra nesse setor. Pinheiro e Galiza (2011) pontuam que na década de 90 a participação feminina no mercado de trabalho elevou-se por alguns fatores como transformações culturais, redução da fecundidade e aumento no nível de escolaridade.

A Tabela 1 também mostra que predominam pessoas com segundo grau completo e superior incompleto; trabalhadores em empresas de pequeno porte (de 10 a 99 empregados); com carga de trabalho superior a 41 horas ou mais, e força de trabalho com idades entre 25 a 49 anos.

3.2 Distribuição Empírica da Renda

O perfil da distribuição do logaritmo da renda, por categoria e ano, foi obtido com base na estimação da densidade por kernel, comparando-o com a distribuição normal (Figuras 1 a 5).

Nota-se pela Figura 1 que a curva para os homens, em todos os anos, se posiciona mais à direita que a curva para as mulheres, o que indica que a renda dos homens, em geral, é maior que a das mulheres. Em todos os anos a curva para mulheres é mais estreita e com pico mais acentuado (mais leptocúrtica) que a curva para homens.

Como se espera, os trabalhadores com ensino superior completo recebem melhor que os outros trabalhadores, uma vez que a curva está mais deslocada para a direita para todos os anos, como mostra a Figura 2. O impacto da educação na desigualdade de renda depende de dois fatores. Primeiramente, a escolaridade não está distribuída igualmente entre os indivíduos de uma sociedade. Há desigualdade na distribuição de educação. Em segundo lugar, em função das ofertas e demandas por trabalho qualificado, a remuneração de mercado aos

atributos que são adquiridos por meio da escolaridade formal pode ser maior ou menor. O balanço dos dois efeitos estabelecerá o impacto da educação sobre a desigualdade de renda. Quanto pior for a distribuição de escolaridade e quanto mais o mercado de trabalho remunerar a educação, maior será o impacto desta na desigualdade (Barbosa Filho e Pessoa, 2009).

A Figura 3 mostra que as curvas que se deslocam para a direita à medida que o porte aumenta, sendo bem parecidas para estabelecimentos com 100 a 499 empregados e 500 ou mais empregados. Como que mais da metade dos trabalhadores se concentram em empresas de pequeno porte (de 10 a 99 empregados), percebe-se que a distribuição da renda pode ser impactada pela assimetria da distribuição dos trabalhos segundo o porte das empresas.

Na Figura 4, observa-se que a renda dos trabalhadores que cumprem até 20 horas é menor do que a dos demais, cuja apresentando distribuições mais platicúrticas. Por outro lado, as distribuições para os que trabalham 41 horas ou mais são mais leptocúrticas.

A Figura 5 mostra que os trabalhadores com renda menor, em geral, são aqueles com até 24 anos. De um modo geral, as figuras mostram que as distribuições apresentam um aspecto unimodal para as categorias das variáveis gênero, grau de instrução, tamanho do estabelecimento e idade.

Mas em alguns casos, como as curvas para horas contratuais da Figura 4, mostram um comportamento bimodal para trabalhadores que cumprem até 20 horas nos anos de 2007 e 2008 e para trabalhadores que cumprem de 21 a 40 horas e unimodal para os outros trabalhadores.

Notamos que, de um ano para outro, as mudanças ocorridas no padrão distributivo foram muito pequenas (ou, pelo menos, imperceptíveis no ponto de vista descritivo). No entanto, mudanças sensíveis podem ser detectadas nos indicadores, como veremos nas próximas seções.

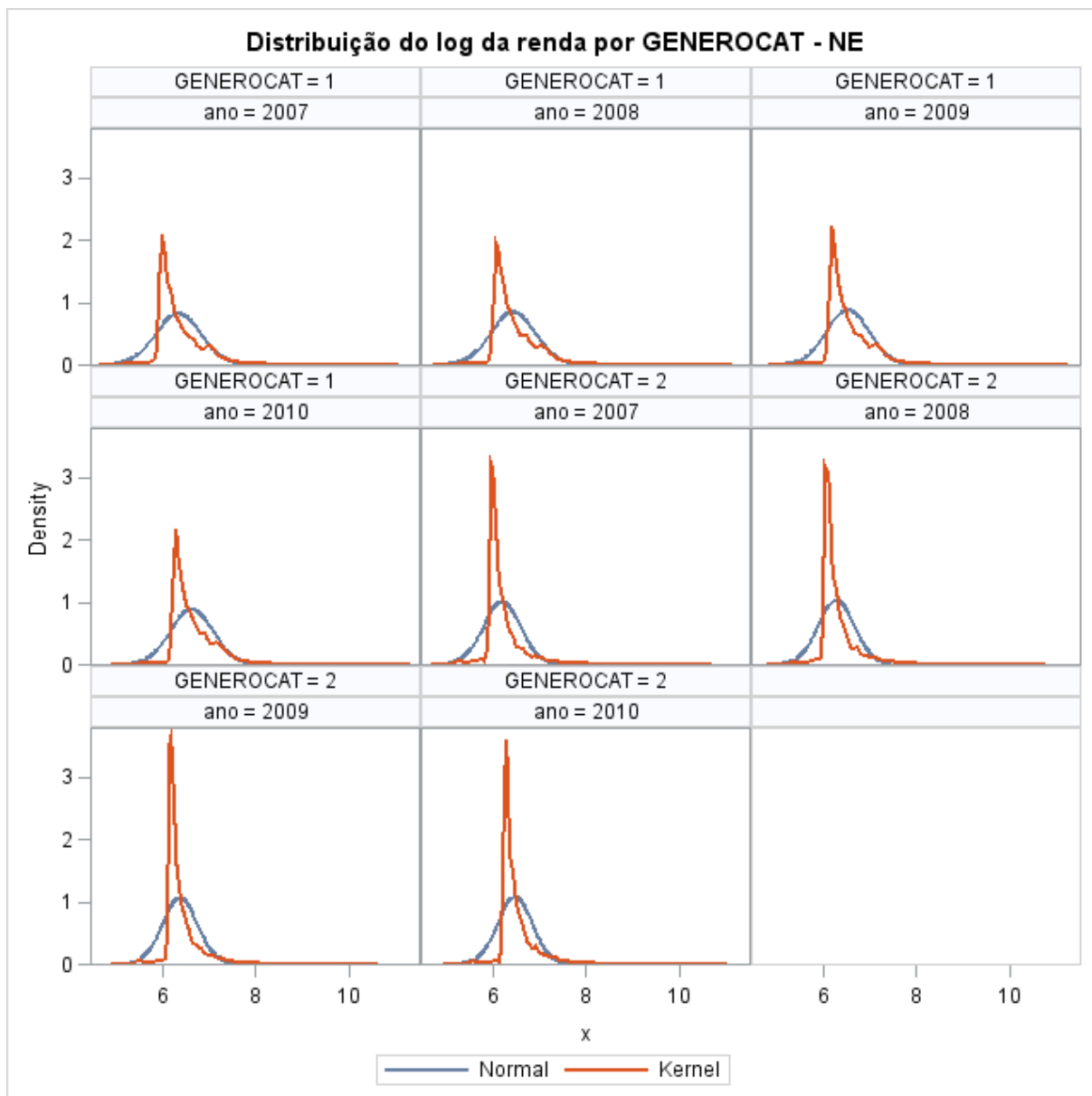


Figura 1: estimativas kernel das densidades para o logaritmo da renda para cada categoria de gênero nos anos de 2007 a 2010, nordeste, em que GENEROCAT = 1 se refere aos homens e GENEROCAT = 2 se refere às mulheres.

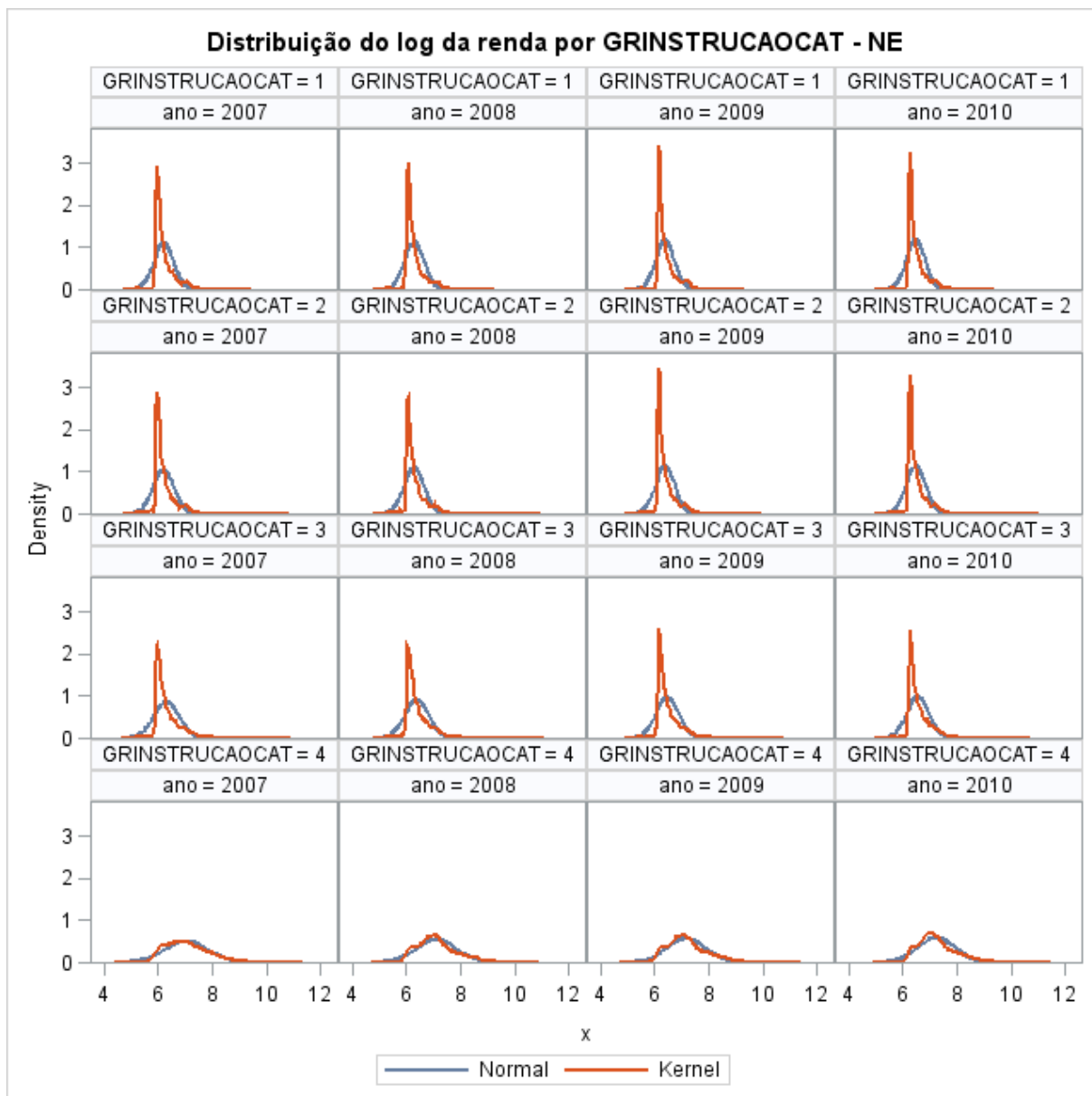


Figura 2: estimativas kernel das densidades para o logaritmo da renda para cada categoria de grau de instrução nos anos de 2007 a 2010, nordeste, em que GRINSTRUCAOCAT é a variável referente a grau de instrução, de modo que ‘1’ se refere aos trabalhadores com nível de escolaridade até 4ª série, ‘2’ aos trabalhadores com nível de escolaridade de 5ª a 8ª série e 2º grau incompleto, ‘3’ aos trabalhadores com 2º grau completo e superior incompleto e ‘4’ aos trabalhadores ensino superior completo.

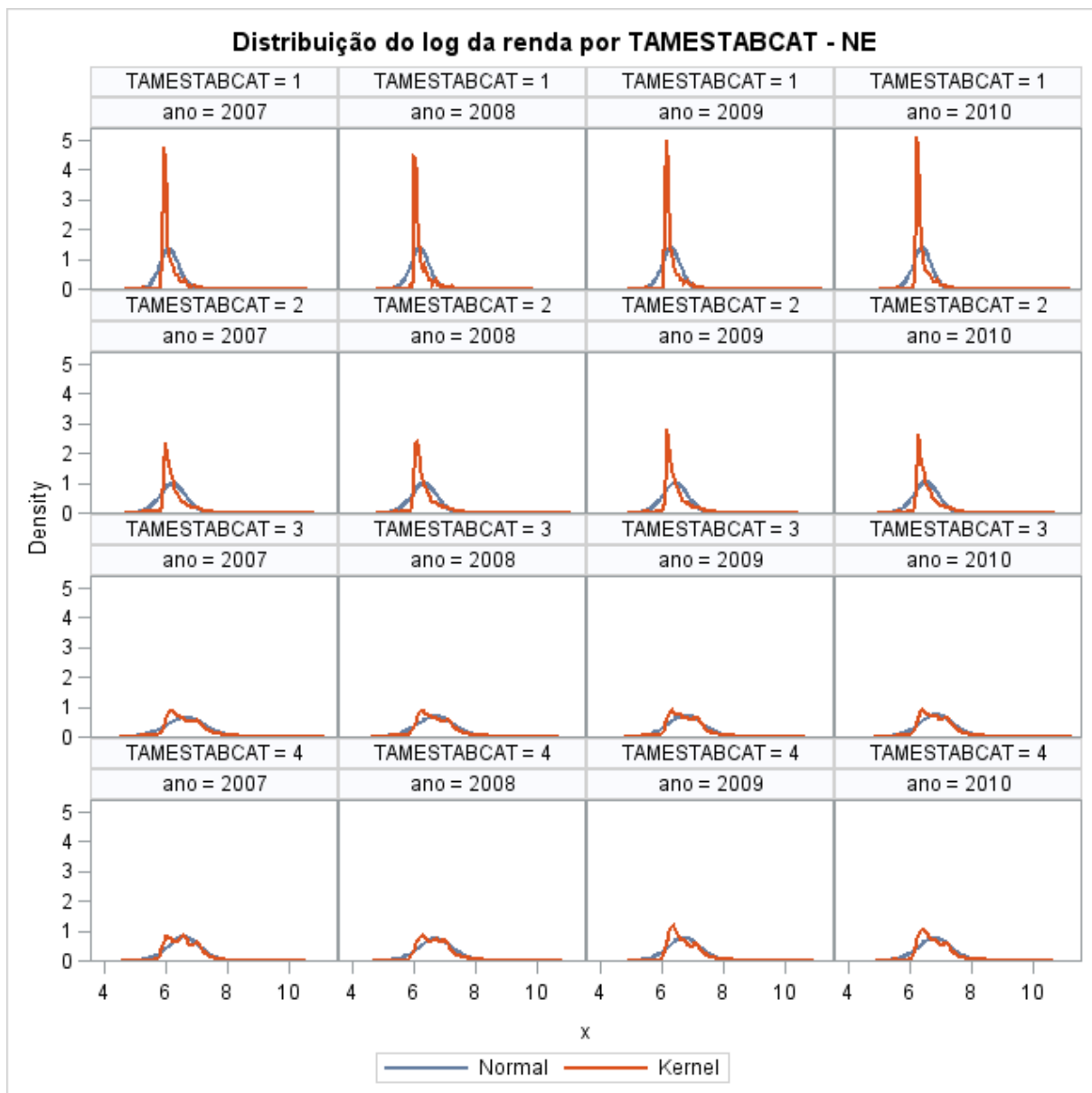


Figura 3: estimativas kernel das densidades para o logaritmo da renda para cada categoria de porte do estabelecimento nos anos de 2007 a 2010, nordeste, em que TAMESTABCAT é a variável que se refere ao tamanho do estabelecimento, de modo que '1' se refere aos estabelecimentos com até 9 empregados, '2' aos estabelecimentos de 10 a 99 empregados, '3' estabelecimento com 100 a 499 empregados e '4' aos estabelecimentos com 500 ou mais empregados.

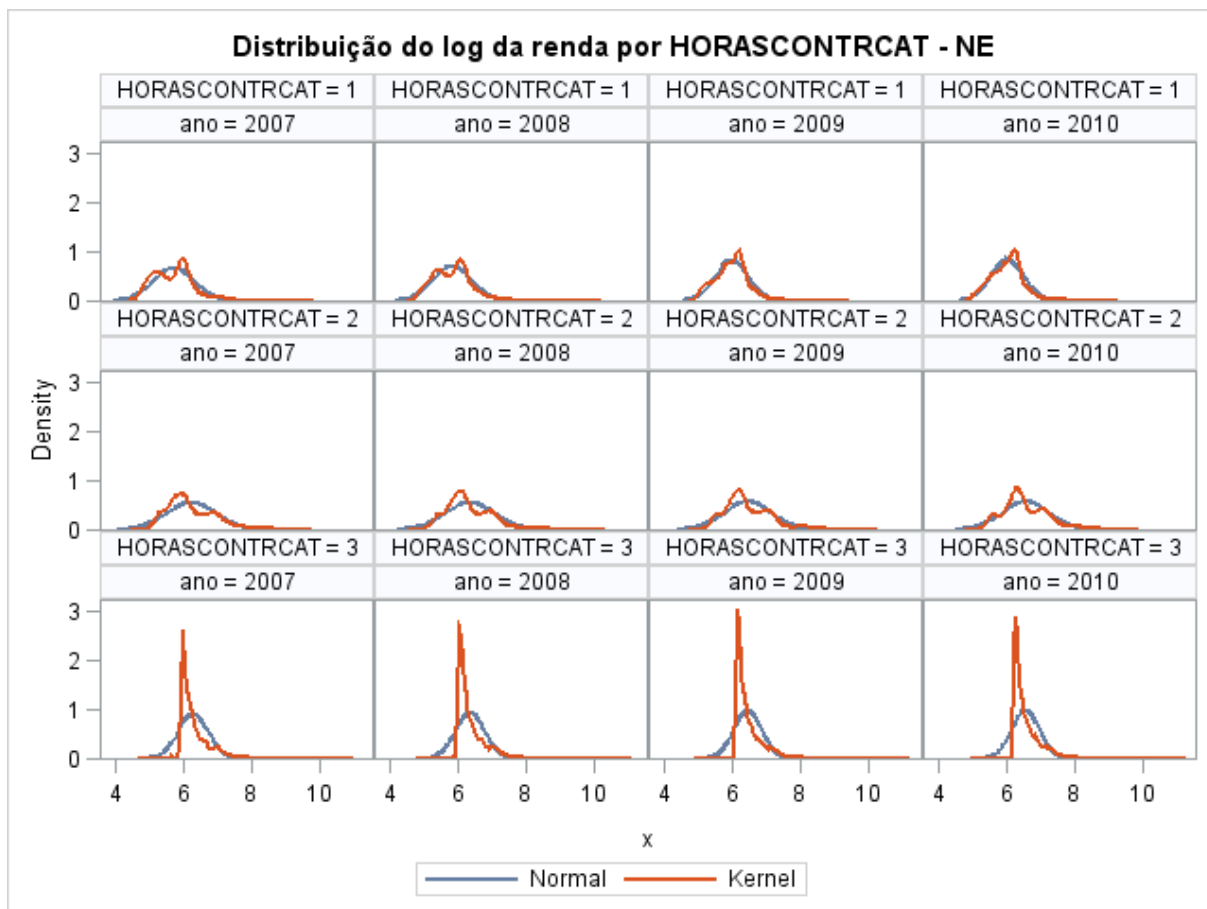


Figura 4: estimativas kernel das densidades para o logaritmo da renda para cada categoria de horas contratuais nos anos de 2007 a 2010, nordeste, em que HORASCONTRCAT é a variável que se refere a horas contratuais, de modo que ‘1’ denota trabalhadores que cumprem até 20 horas, ‘2’ representa os trabalhadores que cumprem de 21 a 40 horas e ‘3’ diz respeito aos trabalhadores que cumprem 41 horas ou mais.

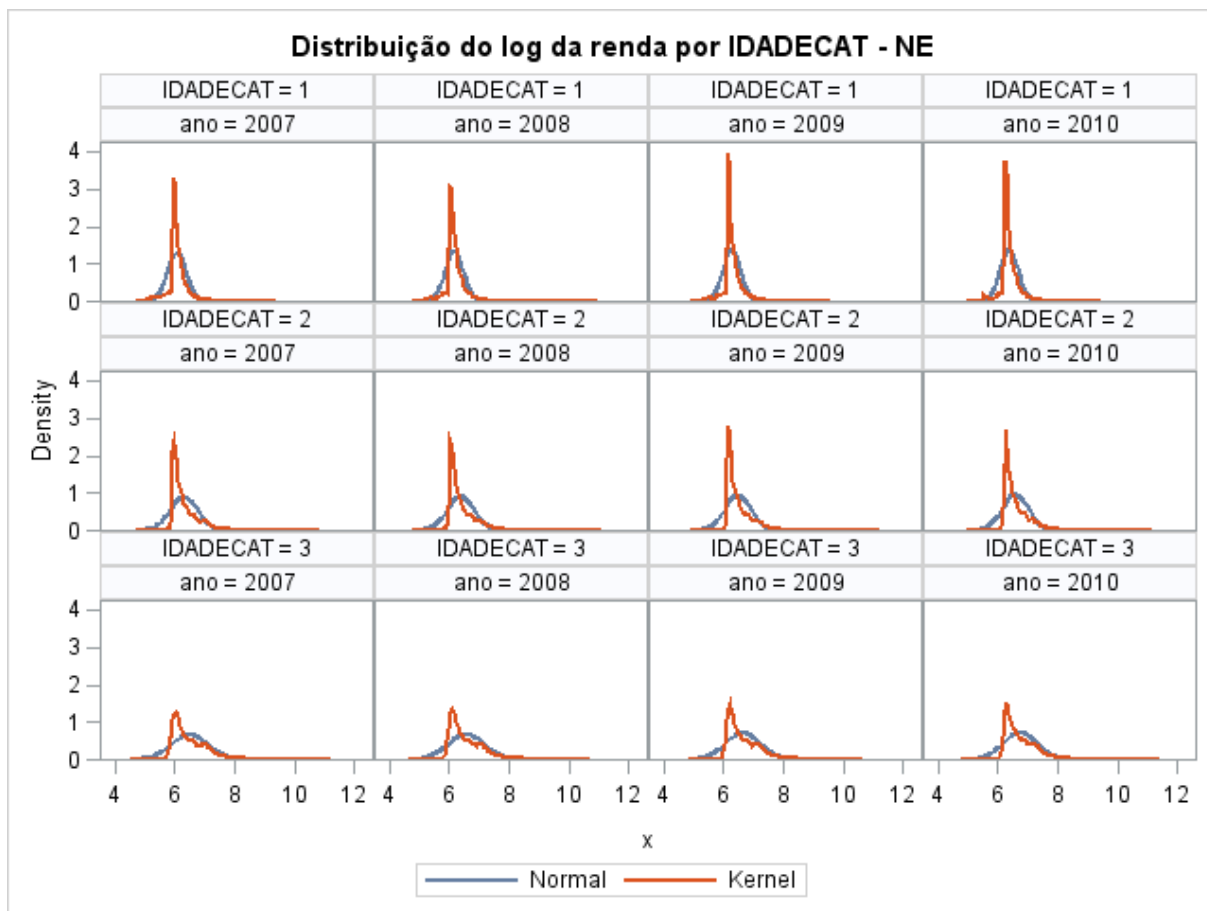


Figura 5: estimativas kernel das densidades para o logaritmo da renda para cada categoria de idade nos anos de 2007 a 2010, nordeste, em que IDADECAT é a variável referente a idade, sendo que ‘1’ se refere aos trabalhadores com até 24 anos, ‘2’ representa os trabalhadores com idade de 25 a 49 anos e ‘3’ denota os trabalhadores com 50 anos ou mais.

3.3 Análise dos indicadores de desigualdade

Tabela 2: Índice de Gini, Nordeste

		2007	2008	2009	2010
Gênero	Homem	0,31	0,30	0,28	0,28
	Mulher	0,24	0,23	0,22	0,22
Grau de Instrução	Até 4ª série	0,21	0,21	0,20	0,19
	De 5ª a 8ª série e segundo grau incompleto	0,23	0,21	0,20	0,20
	2º grau completo e superior incompleto	0,29	0,27	0,25	0,25
	Superior completo	0,48	0,44	0,43	0,42
Tamanho do estabelecimento	Até 9 empregados	0,17	0,16	0,16	0,16
	De 10 a 99 empregados	0,25	0,24	0,24	0,23
	De 100 a 499 empregados	0,39	0,36	0,34	0,33
	500 empregados ou mais	0,31	0,34	0,33	0,32
Horas contratuais	Até 20 horas	0,37	0,36	0,27	0,28
	De 21 a 40 horas	0,45	0,45	0,43	0,42
	41 horas ou mais	0,28	0,27	0,25	0,25
Idade	Ate 24 anos	0,16	0,16	0,15	0,16
	De 25 a 49 anos	0,29	0,28	0,26	0,26
	50 anos ou mais	0,40	0,39	0,37	0,36
Geral		0,29	0,28	0,26	0,26

Da Tabela 2, nota-se que o índice de Gini apresenta tendência de queda para a maioria das categorias. Na variável gênero, o índice de Gini para homens é maior que para mulheres, indicando que a desigualdade de renda é maior entre homens, em todos os anos. De qualquer forma, os índices são menores que o índice geral médio do país, que foi de 0,53 em 2010. Como comparação, nesse ano, o índice de Gini geral dos Estados Unidos da América foi de 0,42.

A desigualdade de renda entre homens e mulheres, através das diferenças de rendimentos por gênero, apresenta-se como uma característica injusta, ou seja, por mais que as mulheres tenham iguais atributos produtivos elas recebem uniformemente menos do que os homens (Gomes e Wajnman, 2005). Talvez, para o empregador, a mão-de-obra feminina custe um pouco mais por causa da licença maternidade, o que, porém, não desobriga o estado a buscar políticas para buscar uma distribuição mais igualitária entre homens e mulheres.

Para o grau de instrução a desigualdade é maior no de trabalhadores com ensino superior completo, em todos os anos. A menor desigualdade se encontra entre os trabalhadores com até 4ª série e de 5ª série a 8ª série e segundo grau completo. De fato, os indicadores de escolaridade são amplamente usados para explicar diferenças salariais, de modo que não parece restar dúvidas acerca da associação entre escolaridade e a distribuição de renda. Pela Figura 2 podemos admitir que o rendimento e a escolaridade se correlacionem positivamente, uma vez que o aumento da escolaridade implique, em geral, rendas maiores. Porém, a dispersão também parece maior, o que sugere que quanto maior for a dispersão dos níveis de escolaridade (digamos, heterogeneidade educacional) na população, maior deve ser a desigualdade na distribuição de renda. A Tabela 2 mostra que as categorias abaixo do nível superior apresentam, uniformemente, rendas menores (com índices de Gini entre 0,20 e 0,30), enquanto no nível superior a desigualdade é maior.

Por que será? esperávamos o resultado contrário. Como há a possibilidade de se substituir a mão-de-obra não qualificada por qualificada no processo de produção, a força de trabalho mais qualificada produz um deslocamento de ocupações de baixos salários para as de salários maiores. O problema é que esse deslocamento não parece suficiente para reduzir os diferenciais salariais; ao contrário, há aumento substancial.

Por alguma razão, não há uniformidade nesse deslocamento, o que nos permite concluir que a escolaridade, ao mesmo tempo que aumenta a renda esperada do trabalhador, também expõe o trabalhador a uma desigualdade maior na distribuição das rendas.

Nossa observação contrasta com a visão de que “a redução das desigualdades ocorreria pela elevação dos salários mais baixos à medida que o nível educacional (da camada com menor escolarização) se eleva” (Da Mata, 1979). Isso poderia ser verdade desde que a elevação dos salários ocorresse uniformemente para todos os trabalhadores. Aqui cabem algumas especulações. Por exemplo, a falta de postos de trabalhos para algumas ocupações de nível superior poderia fazer com que um profissional com curso superior ocupe um posto de nível inferior, com salário também inferior. Talvez, o problema não seja a escolaridade, mas sim o nível de especialização. No Brasil, há falta de profissionais como médicos, engenheiros civis, estatísticos, mas há excesso de profissionais de algumas outras áreas. Em nossa opinião,

é preciso combinar as políticas educacionais de acesso às faculdades com políticas econômicas e sociais que permitam impulsionar o setor produtivo.

Quanto ao porte do estabelecimento, as maiores desigualdades se encontram na distribuição de renda dos trabalhadores de empresas com mais de 100 empregados. Por outro lado, pela Tabela 1, mais da metade das pessoas trabalham em pequenas empresas (de 10 a 99 empregados). Assim, as pequenas empresas são as que mais contratam, apresentando regime de remuneração mais uniforme do que as empresas maiores, o que é natural, uma vez que o perfil de trabalhadores tende a variar mais nas empresas de maior porte. Aparentemente há maior concentração de renda em níveis mais baixos nas empresas com menos de 100 empregados, como ilustra a Figura 3, mas, ao mesmo tempo, a Tabela 2 mostra que os índices de Gini para essas categorias são menores. Por isso, a estrutura do emprego é um fator de influência sobre a concentração de renda. Nesse caso, não parece difícil identificar a influência equalizadora. Em geral, as pequenas empresas se compõem de atividades mais homogêneas (em termos de produtividade da mão-de-obra e de tecnologia), o que provavelmente explica menores diferenciais de renda. Considerando que essas empresas contratam mais da metade do contingente de trabalhadores, a maior homogeneidade do setor de pequenas empresas talvez explique a razão de o índice de Gini da distribuição de renda do nordeste ser inferior ao seu correspondente nacional (0,56). Esse resultado contrasta com a ideia — ou talvez seja apenas um contraexemplo — de que a desigualdade de renda seja promovida pelo setor terciário (Da Mata, 1979).

Com respeito ao número de horas de trabalho, a maior desigualdade de renda foi encontrada na categoria dos que trabalham de 21 a 40 horas. Nessa categoria, os índices foram maiores do que o índice geral do setor de turismo no nordeste (última linha da Tabela 2). Esse resultado, talvez, se explique pela bimodalidade na distribuição que se observa na Figura 4. Essa bimodalidade pode ser consequência de uma mistura de perfis, o que gera heterogeneidade da distribuição segundo o número de horas de trabalho. Naturalmente, o número de horas de trabalho produz uma grande influência na renda, e talvez seu efeito se potencialize (como um efeito multiplicativo) dependendo das outras características do trabalhador (gênero, escolaridade, porte e idade).

Finalmente, quanto à idade, a categoria de empregados com idade de 50 anos ou mais apresentou maior desigualdade, e a que teve menor índice foi a categoria de empregados com até 24 anos. Esse efeito idade reflete as implicações tradicionais da teoria de capital humano que prevê uma crescente dispersão de rendimentos à medida que os indivíduos envelhecem. Isto se relaciona a diferenciados retornos à experiência entre os indivíduos, principalmente entre aqueles com níveis de escolaridade diferentes (Menezes-Filho, Fernandes e Pichetti, 2000).

Contudo, segundo a Tabela 1, a categoria predominante é aquela formada por trabalhadores com 25 a 49 anos, com 75% de participação. Observe que os índices de Gini para essa faixa etária coincidem com os índices gerais (última linha da Tabela 2), o que indica que, para o nosso estudo particular, o fato de a força de trabalho ser predominantemente jovem no setor de turismo, essa variável demográfica não é relevante.

A Figura 6 mostra as curvas de Lorenz relativas à distribuição de renda do setor de turismo do nordeste em 2007 (superior à esquerda), 2008 (superior à direita), 2009 (inferior à esquerda) e 2010 (inferior à direita), correspondendo aos índices de Gini 0,29; 0,28; 0,26 e 0,26, respectivamente. O fato de a curva de Lorenz se aproximar sensivelmente da reta de 45° ao longo do tempo, indica, graficamente, que a desigualdade de renda diminuiu de 2007 a 2010.

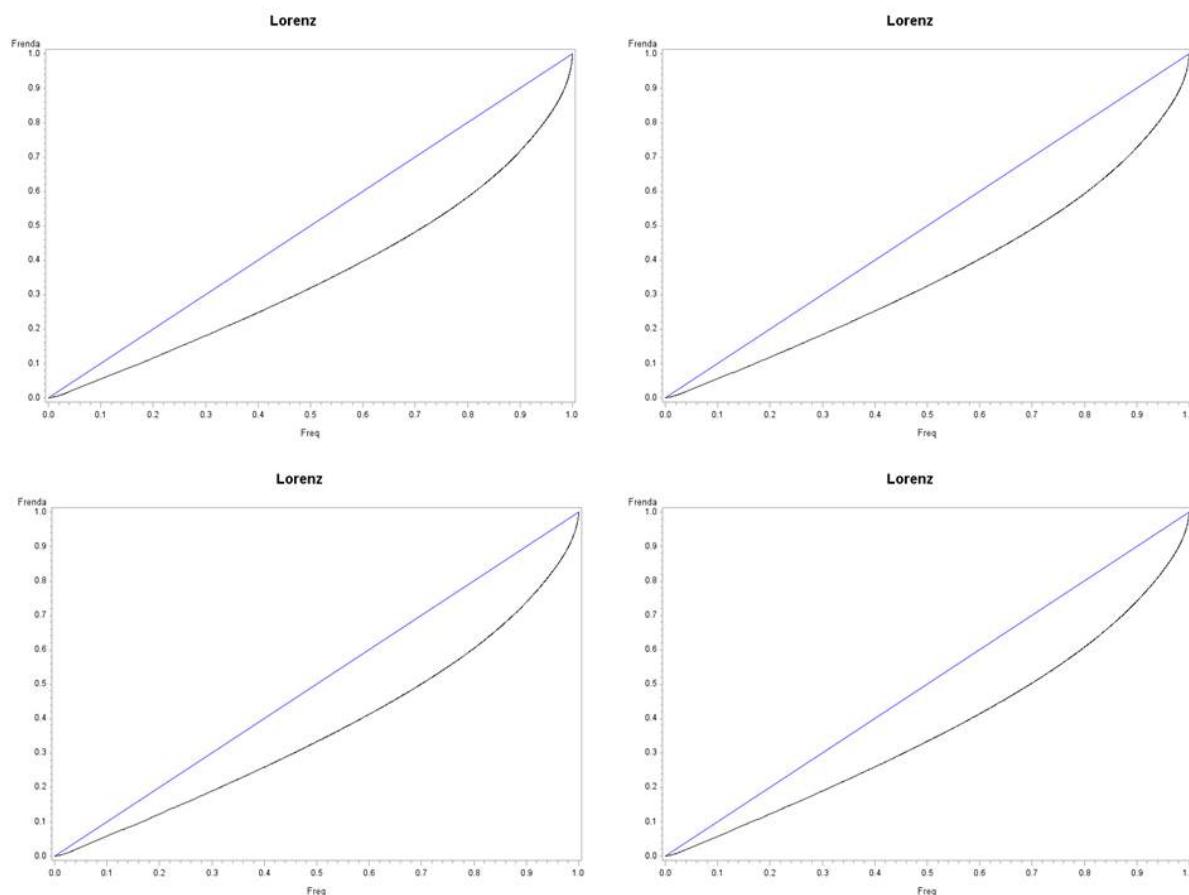


Figura 6: curvas de Lorenz relativas à distribuição de renda do setor de turismo do nordeste em 2007 (superior à esquerda), 2008 (superior à direita), 2009 (inferior à esquerda) e 2010 (inferior à direita).

As Tabelas 3 e 4 mostram os índices de Theil e de Atkinson com $\epsilon = 2$. Os resultados fornecidos por essas tabelas são análogos com os que foram apresentados pelo índice de Gini na Tabela 1. Na Seção 3.4 apresentaremos uma comparação entre os indicadores de desigualdade de renda, e mostremos que, de fato, os índices de Gini, de Theil e de Atkinson se correlacionam.

Quanto ao índice de Atkinson, em nosso estudo a escolha da constante ϵ não é questão relevante, uma vez que, independentemente da sua escolha, os resultados se relacionam com o índice de Gini, fornecendo interpretações análogas.

Tabela 3: Índice de Theil, Nordeste

		2007	2008	2009	2010
Gênero	Homem	0,23	0,21	0,19	0,18
	Mulher	0,17	0,15	0,14	0,13
Grau de Instrução	Até 4ª série	0,10	0,09	0,08	0,08
	De 5ª a 8ª série e segundo grau incompleto	0,12	0,10	0,09	0,09
	2º grau completo e superior incompleto	0,20	0,17	0,15	0,14
	Superior completo	0,45	0,38	0,36	0,35
Tamanho do estabelecimento	Até 9 empregados	0,09	0,08	0,08	0,09
	De 10 a 99 empregados	0,16	0,16	0,14	0,14
	De 100 a 499 empregados	0,32	0,25	0,23	0,23
	500 empregados ou mais	0,21	0,27	0,25	0,22
Horas contratuais	Até 20 horas	0,32	0,32	0,15	0,15
	De 21 a 40 horas	0,41	0,41	0,36	0,35
	41 horas ou mais	0,20	0,18	0,16	0,16
Idade	Ate 24 anos	0,07	0,06	0,05	0,05
	De 25 a 49 anos	0,21	0,18	0,17	0,16
	50 anos ou mais	0,35	0,34	0,31	0,31
Geral		0,22	0,19	0,17	0,17

Tabela 4: Índice de Atkinson, Nordeste

		2007	2008	2009	2010
Gênero	Homem	0,23	0,22	0,20	0,20
	Mulher	0,16	0,15	0,14	0,14
Grau de Instrução	Até 4ª série	0,13	0,12	0,11	0,11
	De 5ª a 8ª série e segundo grau incompleto	0,15	0,13	0,12	0,12
	2º grau completo e superior incompleto	0,21	0,19	0,17	0,17
	Superior completo	0,48	0,43	0,41	0,39
Tamanho do estabelecimento	Até 9 empregados	0,10	0,09	0,09	0,09
	De 10 a 99 empregados	0,17	0,16	0,16	0,15
	De 100 a 499 empregados	0,34	0,30	0,28	0,27
	500 empregados ou mais	0,24	0,28	0,25	0,25
Horas contratuais	Até 20 horas	0,32	0,30	0,21	0,21
	De 21 a 40 horas	0,42	0,42	0,39	0,39
	41 horas ou mais	0,19	0,18	0,17	0,17
Idade	Ate 24 anos	0,10	0,09	0,08	0,09
	De 25 a 49 anos	0,21	0,19	0,18	0,18
	50 anos ou mais	0,33	0,32	0,30	0,29
Geral		0,21	0,20	0,18	0,18

Tabela 5: Coeficiente de assimetria, Nordeste

		2007	2008	2009	2010
Gênero	Homem	1,59	1,51	1,50	1,45
	Mulher	2,09	2,07	2,08	2,01
Grau de Instrução	Até 4ª série	1,48	1,44	1,35	1,36
	De 5ª a 8ª série e segundo grau incompleto	1,31	1,23	1,18	1,15
	2º grau completo e superior incompleto	1,67	1,56	1,57	1,52
	Superior completo	0,62	0,66	0,69	0,81
Tamanho do estabelecimento	Até 9 empregados	2,36	2,22	2,30	2,30
	De 10 a 99 empregados	1,63	1,69	1,68	1,60
	De 100 a 499 empregados	0,98	0,81	0,82	0,89
	500 empregados ou mais	0,96	1,10	1,33	1,05
Horas contratuais	Até 20 horas	0,92	0,95	0,35	0,50
	De 21 a 40 horas	0,96	0,99	0,91	0,85
	41 horas ou mais	1,96	1,86	1,89	1,84
Idade	Ate 24 anos	0,74	0,59	0,53	0,54
	De 25 a 49 anos	1,76	1,67	1,67	1,63
	50 anos ou mais	1,38	1,42	1,44	1,47
Geral		1,76	1,70	1,70	1,65

Tabela 6: Curtose, Nordeste

		2007	2008	2009	2010
Gênero	Homem	4,61	4,33	4,31	4,19
	Mulher	8,24	7,81	8,04	7,97
Grau de Instrução	Até 4ª série	4,07	4,07	3,33	3,86
	De 5ª a 8ª série e segundo grau incompleto	4,28	3,94	3,28	3,68
	2º grau completo e superior incompleto	4,65	4,20	4,49	4,35
	Superior completo	0,54	0,57	0,61	0,85
Tamanho do estabelecimento	Até 9 empregados	12,22	10,45	11,20	12,12
	De 10 a 99 empregados	6,01	6,47	6,08	6,01
	De 100 a 499 empregados	2,02	1,60	1,79	1,93
	500 empregados ou mais	2,27	2,85	3,52	2,39
Horas contratuais	Até 20 horas	2,41	3,11	1,10	1,32
	De 21 a 40 horas	0,89	1,19	1,02	0,96
	41 horas ou mais	6,24	5,65	5,80	5,70
Idade	Ate 24 anos	6,14	6,18	5,06	5,11
	De 25 a 49 anos	5,23	4,66	4,75	4,63
	50 anos ou mais	2,66	2,79	3,04	3,41
Geral		5,58	5,26	5,31	5,22

Tabela 4: Coeficiente de variação, Nordeste

		2007	2008	2009	2010
Gênero	Homem	7,70	7,37	6,98	6,84
	Mulher	6,47	6,19	5,89	5,75
Grau de Instrução	Até 4ª série	5,86	5,65	5,38	5,23
	De 5ª a 8ª série e segundo grau incompleto	6,22	5,87	5,56	5,50
	2º grau completo e superior incompleto	7,38	6,86	6,46	6,27
	Superior completo	11,33	10,42	9,97	9,49
Tamanho do estabelecimento	Até 9 empregados	4,96	4,77	4,66	4,59
	De 10 a 99 empregados	6,61	6,37	6,18	5,98
	De 100 a 499 empregados	9,30	8,73	8,19	7,95
	500 empregados ou mais	7,71	8,13	7,66	7,74
Horas contratuais	Até 20 horas	10,43	9,75	8,05	7,96
	De 21 a 40 horas	11,69	11,29	10,71	10,53
	41 horas ou mais	7,02	6,72	6,35	6,21
Idade	Até 24 anos	5,10	4,85	4,63	4,62
	De 25 a 49 anos	7,28	6,97	6,60	6,46
	50 anos ou mais	9,39	9,14	8,61	8,28
Geral		7,34	7,03	6,66	6,51

Sobre a distribuição do logaritmo da renda, foram calculadas as estatísticas acerca da assimetria (Tabela 5), da curtose (Tabela 6) e o coeficiente de variação (Tabela 7)

A Tabela 4 indica que todas as categorias apresentam assimetria à direita (ou positiva), o que sugere que as medianas das rendas sejam inferiores à média das rendas nas suas distribuições correspondentes.

A assimetria, para cada variável, foi maior entre as mulheres, empregados com 2º grau completo e superior incompleto, empregados de estabelecimentos com até 9 empregados, que trabalham 41 horas ou mais e que possuem de 25 a 49 anos. A assimetria diminui sensivelmente ao longo dos anos, com exceção das categorias de empregados com ensino superior completo, de estabelecimentos de grande porte e para os trabalhadores com 50 anos ou mais.

Quanto à curtose, ela se mostrou positiva para todas as categorias, o que indicaria que as distribuições são mais achatadas do que a distribuição gaussiana.

Mas é necessário fazer o seguinte alerta. A curtose só se relaciona com o grau de achatamento em distribuições simétricas. Em nosso caso, as distribuições são assimétricas, e a Figura 1 mostra que algumas distribuições apresentam picos, o que indicaria leptocurtose em vez de platicurtose. Como não há uma compreensão clara do significado da medida de curtose em distribuições assimétrica, nos limitaremos a comparar essa medida com as demais na Seção 3.4.

Quanto ao coeficiente de variação, a Tabela 7 mostra um coeficiente de variação cujos valores são todos menores que 100, o que indica que o desvio padrão de cada distribuição é menor que a sua média. Mostraremos na seção seguinte como o coeficiente de variação se relaciona com os índices de Gini, de Theil e de Atkinson.

3.3 Comparações entre os índices

Tabela 5: Correlações de Pearson entre os índices, Nordeste

	Índice de Theil	Índice de Gini	Índice de Atkinson	Curtose	Coeficiente de Assimetria	Coeficiente de Variação
Índice de Theil	1,00	0,93	0,93	-0,13	0,10	0,87
Índice de Gini	0,93	1,00	0,99	-0,35	-0,10	0,97
Índice de Atkinson	0,93	0,99	1,00	-0,34	-0,12	0,98
Curtose	-0,13	-0,35	-0,34	1,00	0,73	-0,39
Coeficiente de Assimetria	0,10	-0,10	-0,12	0,73	1,00	-0,18
Coeficiente de Variação	0,87	0,97	0,98	-0,39	-0,18	1,00

A Tabela 8 mostra uma comparação mútua entre os índices com base no coeficiente de Correlação de Pearson. Essa tabela mostra que os índices de Theil, Gini, Atkinson e o coeficiente de variação podem ser considerados, de certa forma, equivalentes.

Já o coeficiente de assimetria se associa com a curtose, o que sugere que esses índices descrevem características de desigualdade não perceptíveis pelos outros índices.

As Figuras 7 a 21 mostram a dispersão entre os índices. O índice de Theil possui forte associação linear positiva com os índices de Gini (Figura 7) e de Atkinson (Figura 8) e com o coeficiente de variação (Figura 11). Os valores do coeficiente de correlação foram de 0,93, 0,93 e 0,87, respectivamente. Esses valores, analisados conjuntamente com as figuras indicam que o índice de Theil cresce quase na mesma proporção que o índice de Gini, de Atkinson e Coeficiente de variação. Também foram altos os valores do coeficiente de correlação do índice de Gini com o índice de Atkinson (0,99) e com o coeficiente de variação (0,97), revelando forte associação linear positiva como mostram as Figuras 12 e 15. O índice de Atkinson é fortemente associado com o coeficiente de variação (Figura 18), com coeficiente de correlação igual a 0,98.

A curtose só é fortemente associada com o coeficiente de assimetria, e vice-versa, (Figura 9), com coeficiente de correlação igual a 0,73, mas eles possuem baixa correlação com todos demais índices. Isso mostra que esses indicadores podem contribuir com informações diferentes daquelas fornecidas pelos outros índices.

O coeficiente de assimetria e a curtose são invariantes a transformações de escala, ou seja, $g(kY) = kg(Y)$, e de locação, i.e., $g(k + Y) = k + g(Y)$, em que y representa a distribuição da renda, $k > 0$, e $g(.)$ é um indicador. Já o coeficiente de variação e o índice de Gini são apenas invariantes a transformações de escala. O coeficiente de variação é mais sensível à cauda direita da distribuição, enquanto o índice de Gini é menos sensível (sendo uma medida mais robusta, conforme Bendel et al. (1989)).

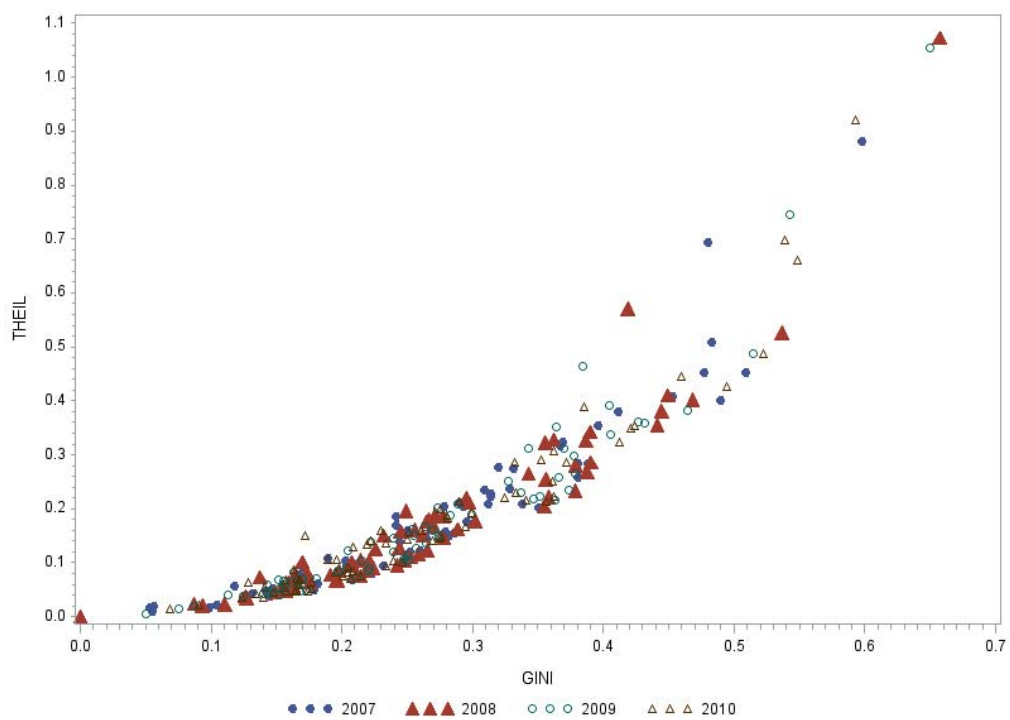


Figura 7: Relação do índice de Theil com o índice de Gini, Nordeste

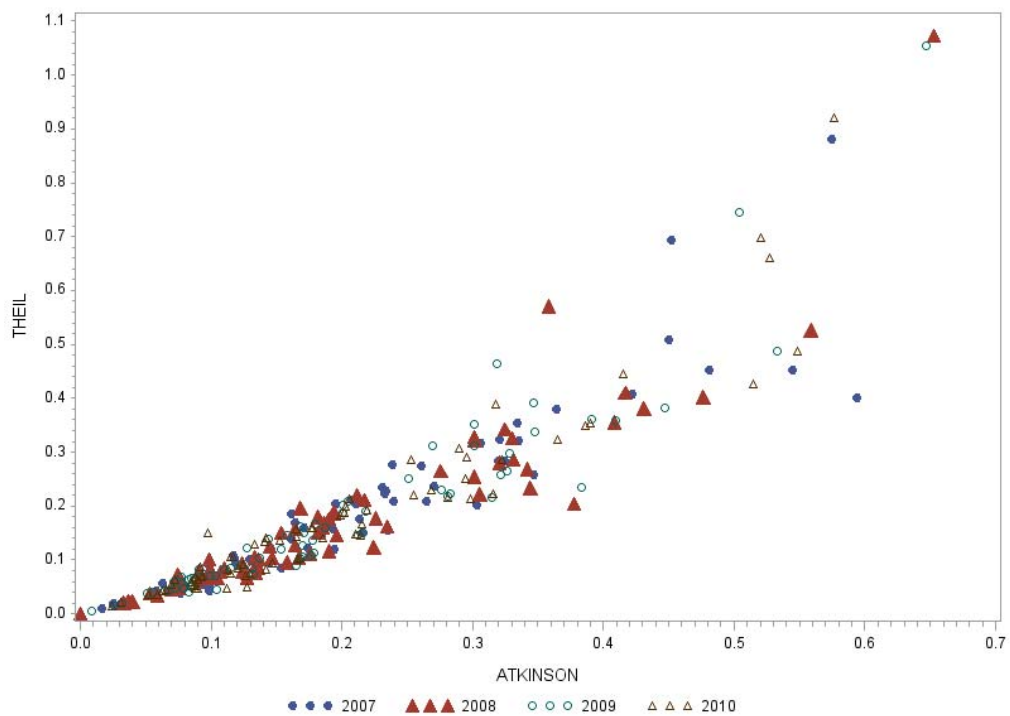


Figura 8: Relação do índice de Theil com o índice de Atkinson, Nordeste

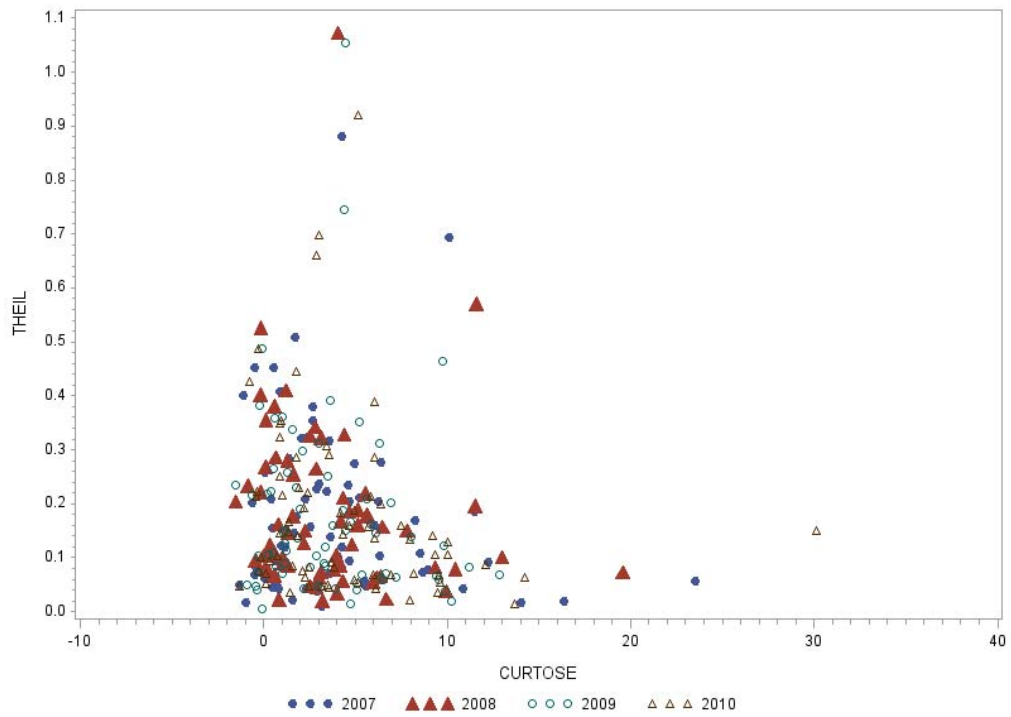


Figura 9: Relação do índice de Theil com a curtose, Nordeste

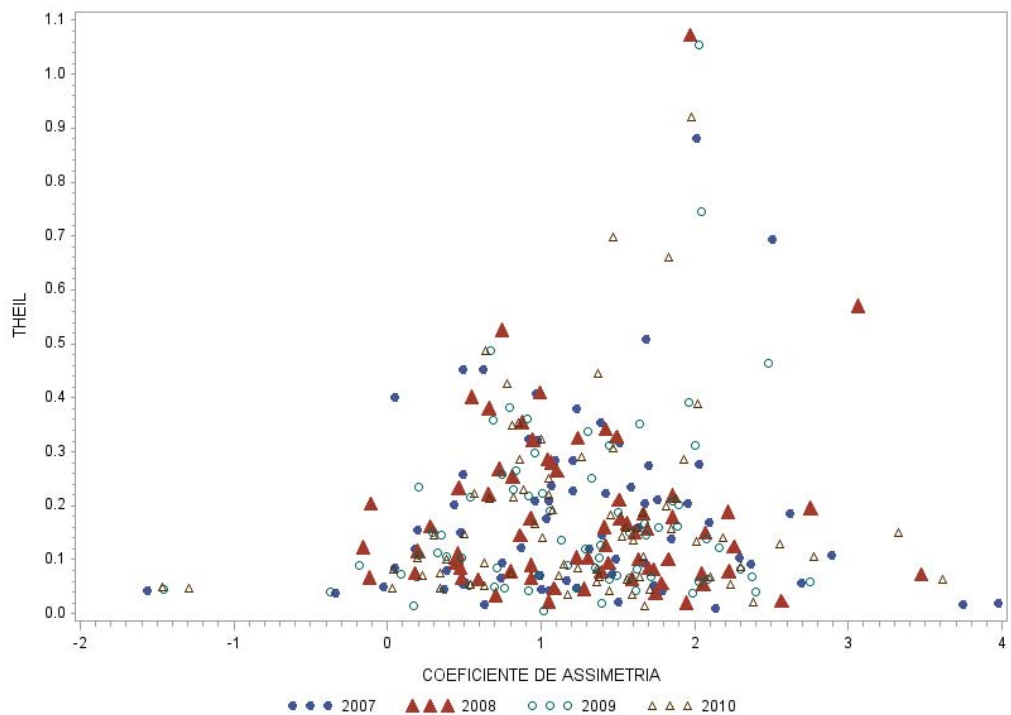


Figura 7: Relação do índice de Theil com o coeficiente de assimetria, Nordeste

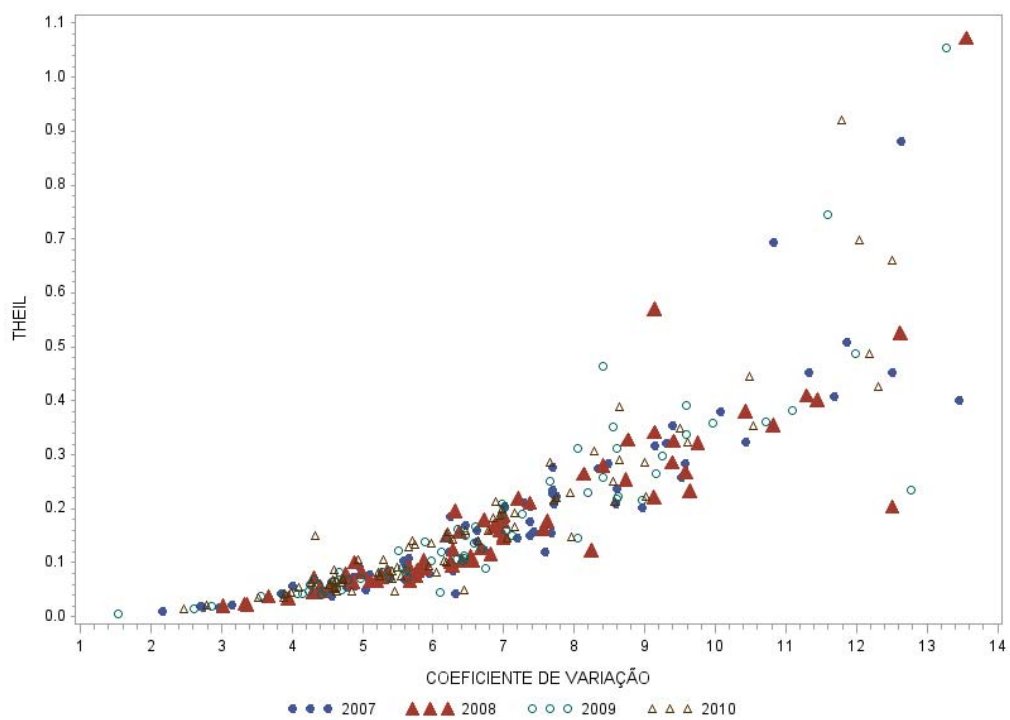


Figura 8: Relação do índice de Theil com o coeficiente de variação, Nordeste

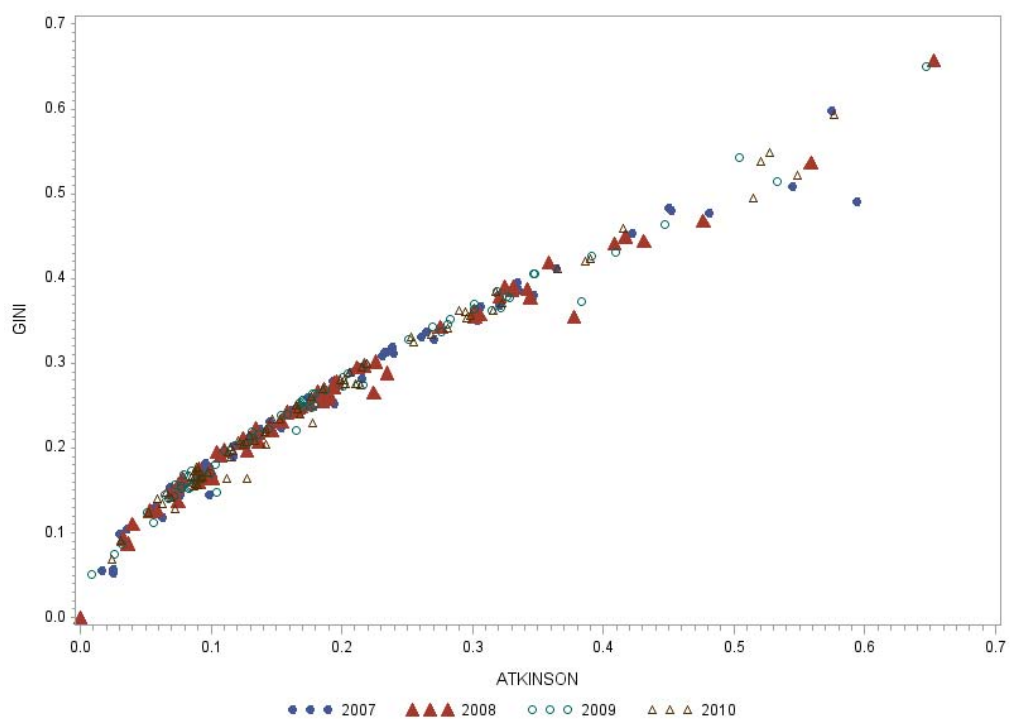


Figura 9: Relação do índice de Gini com o índice de Atkinson, Nordeste

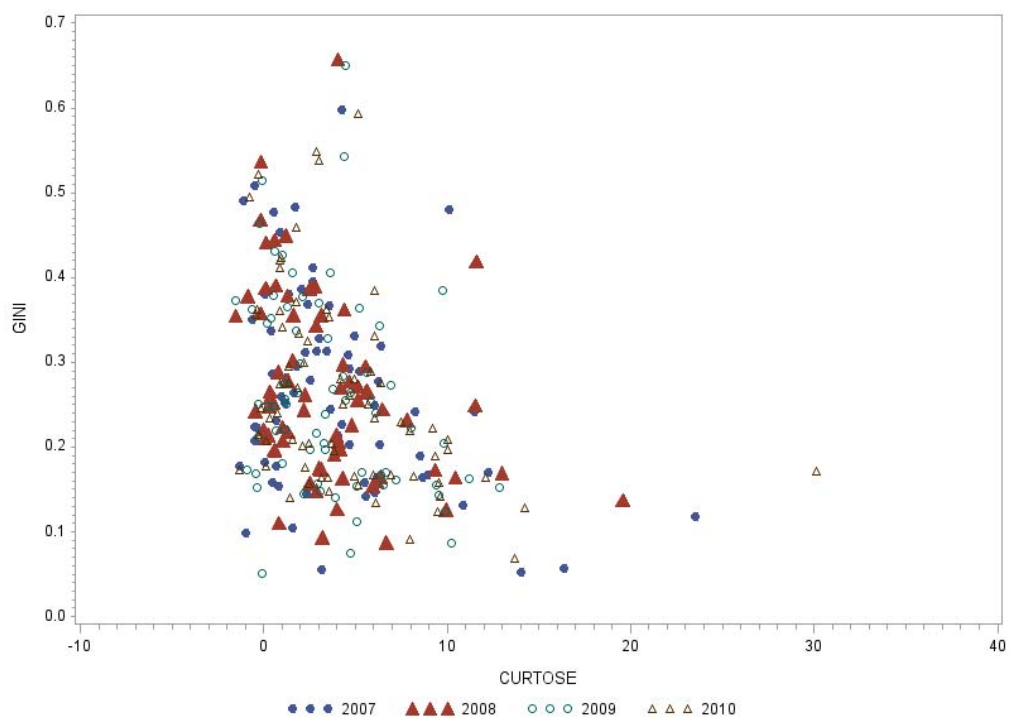


Figura 10: Relação do índice de Gini com a curtose, Nordeste

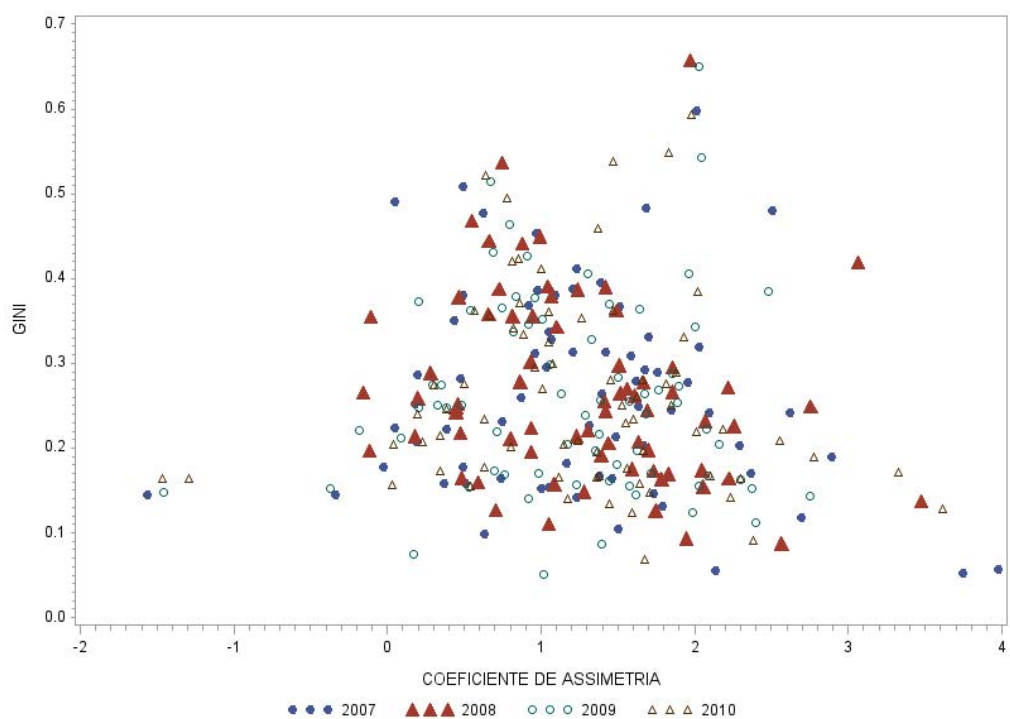


Figura 11: Relação do índice de Gini com o coeficiente de assimetria, Nordeste

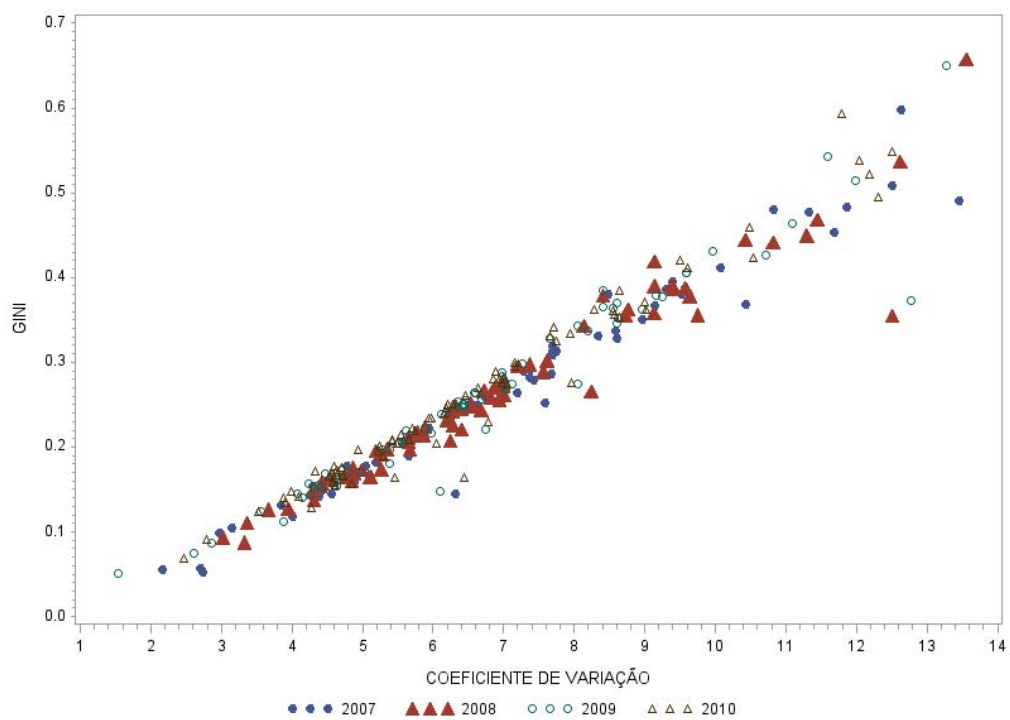


Figura 12: Relação do índice de Gini com o coeficiente de variação, Nordeste

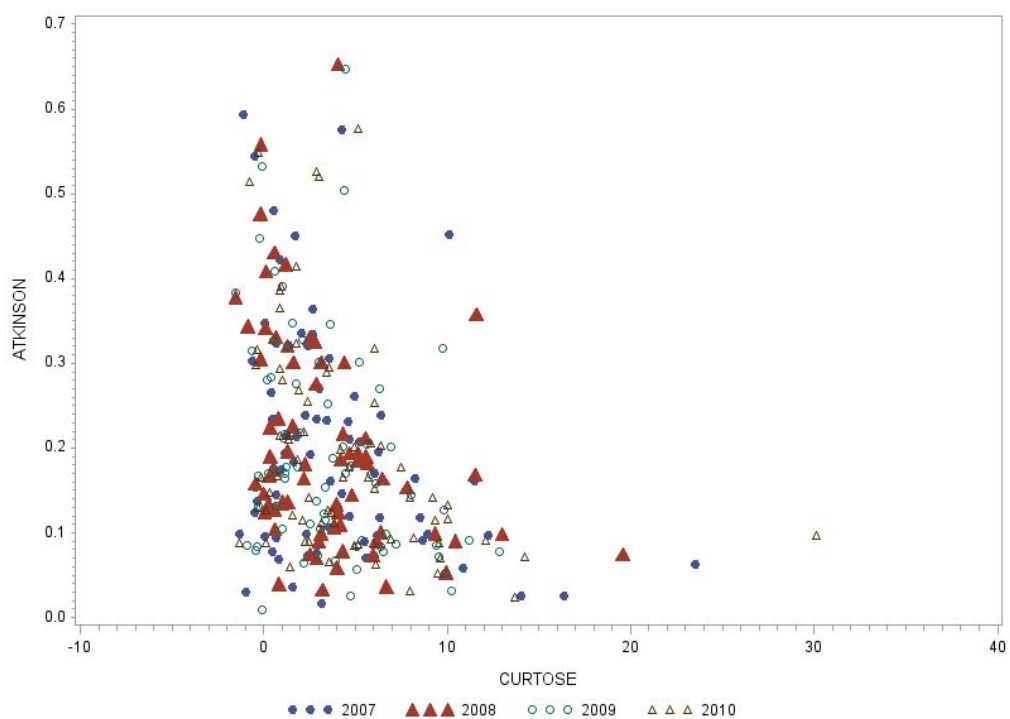


Figura 13: Relação do índice de Atkinson com a curtose, Nordeste

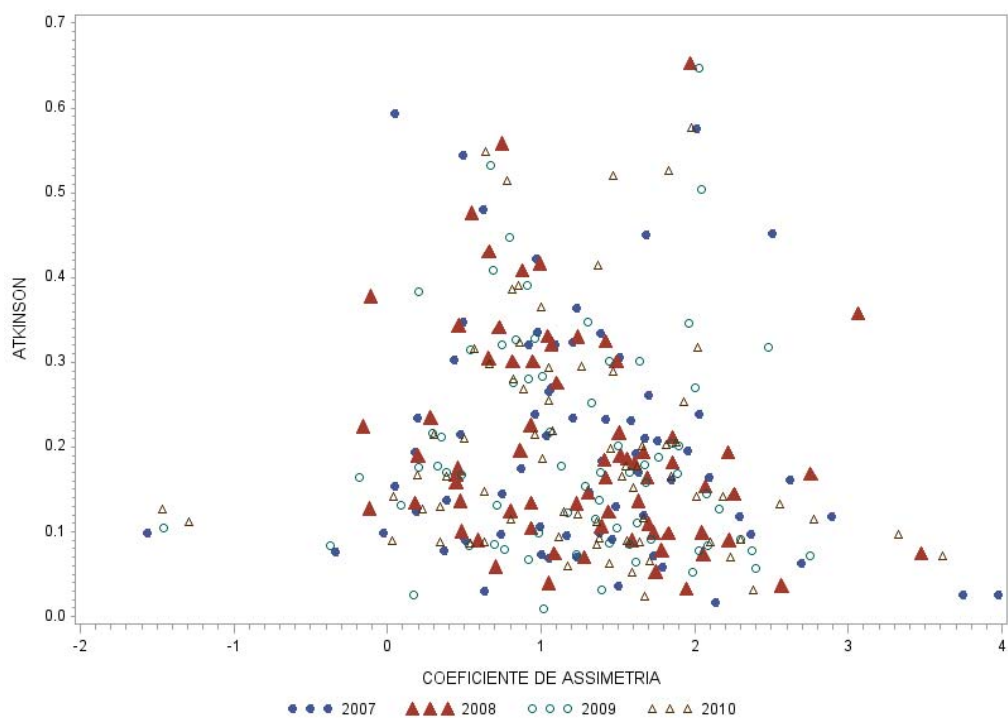


Figura 14: Relação do índice de Atkinson com o coeficiente de assimetria, Nordeste

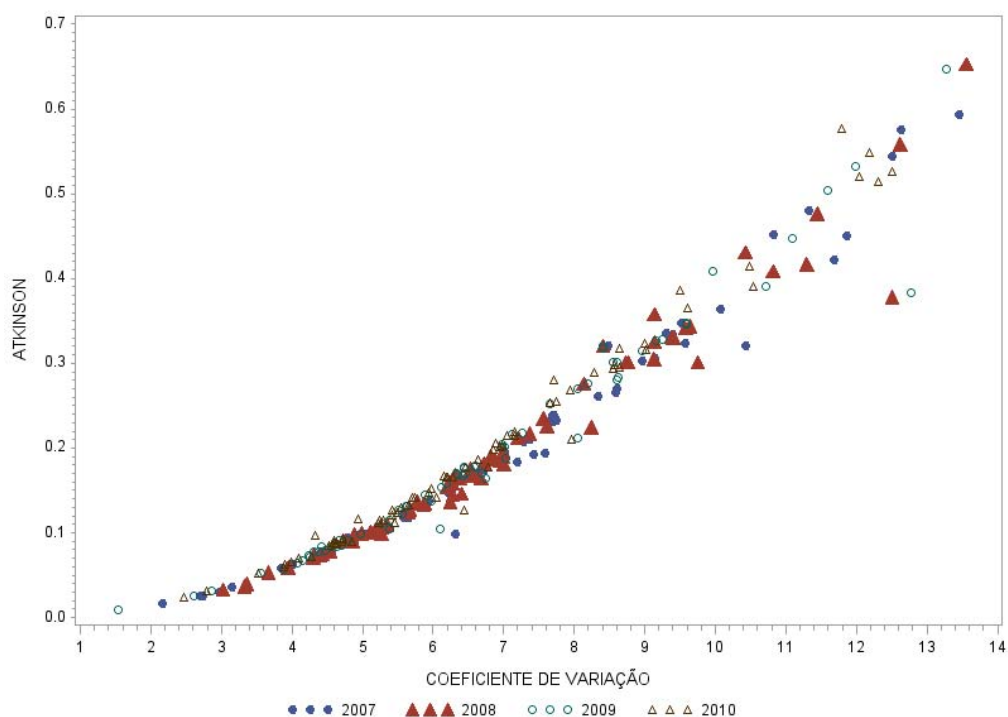


Figura 15: Relação do índice de Atkinson com o coeficiente de variação, Nordeste

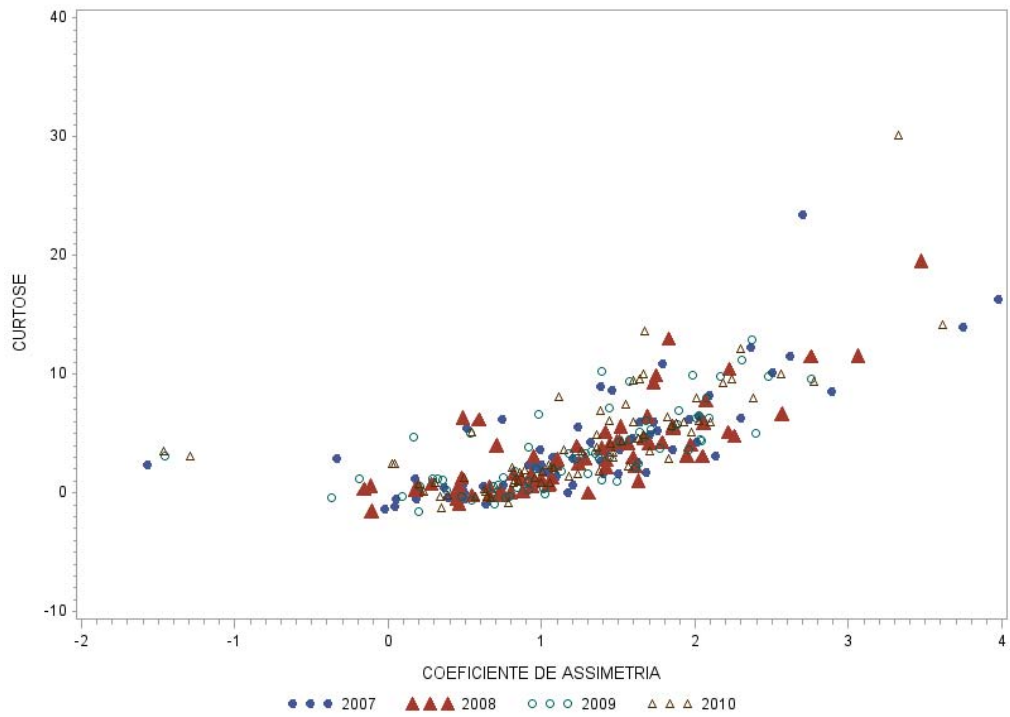


Figura 16: Relação da curtose com o coeficiente de assimetria, Nordeste

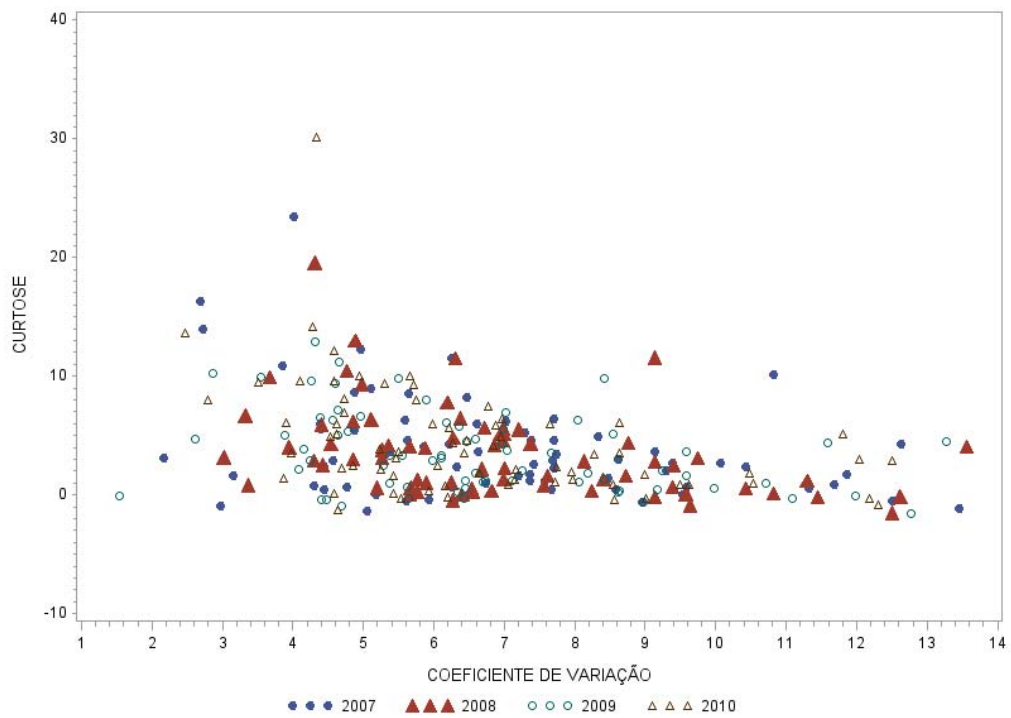


Figura 20: Relação da curtose com o coeficiente de variação, Nordeste

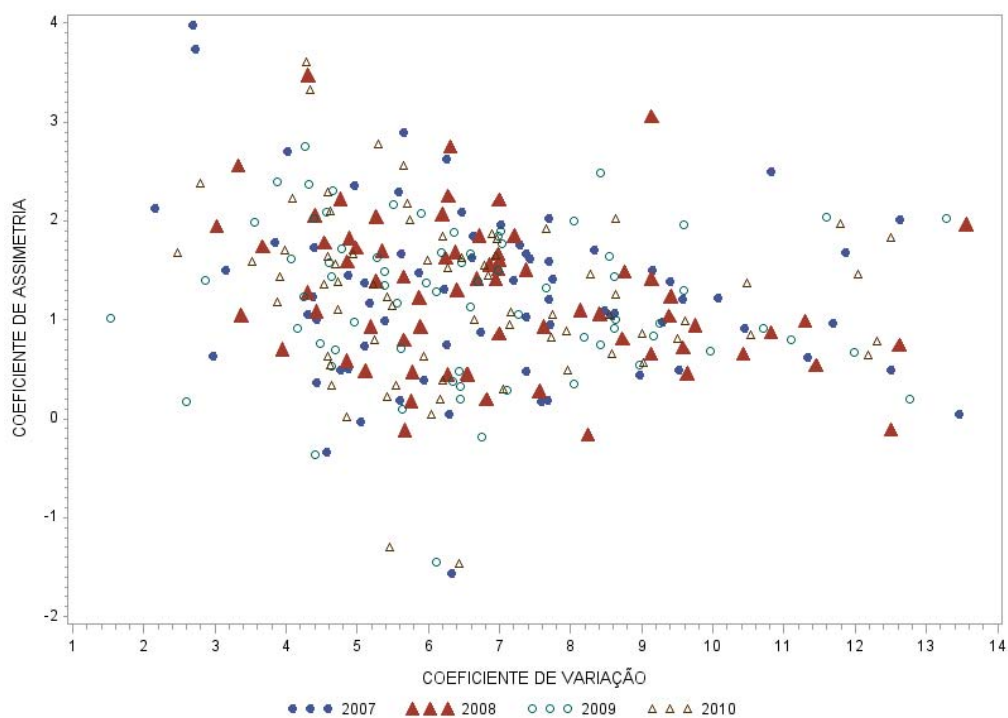


Figura 17: Relação do coeficiente de assimetria com o coeficiente de variação, Nordeste

4 CONCLUSÃO

Neste trabalho, estudamos brevemente alguns aspectos referentes à desigualdade de renda no setor de turismo na região do nordeste brasileiro de 2007 a 2010. Utilizamos a base nos dados da Relação Anual de Informações Sociais (RAIS), considerando o perfil de trabalhadores empregados em atividades ligadas diretamente ou indiretamente com o setor do turismo na região nordeste do Brasil. O banco de dados da RAIS foi gentilmente cedido pelo IPEA para uso restrito nesse trabalho.

Grande parte do tempo investido neste estudo concentrou-se na organização de dados censitários, o que exigiu manipulação cuidadosa e demorada de um grande banco de dados com tamanho aproximado de 18,4 GB de memória. Para lidar com a imensa quantidade de dados, a análise foi efetuada no SAS 9.3, cuja programação se encontra nos Apêndices B, C, D, E, F e G.

Estudamos o padrão de desigualdade da distribuição da renda com a utilização de indicadores clássicos como o índice de Gini, de Theil, de Atkinson, e os coeficientes de assimetria, de curtose e o de variação.

Entre os resultados interessantes, destacamos o fato de a desigualdade de renda no setor do turismo dessa região se encontra em níveis compatíveis aos que são encontrados em países desenvolvidos como os EUA (0,42), e muito menor do que o índice geral médio do país, que foi de 0,53 em 2010.

Encontramos resultados esperados, como uma tendência sensível de queda nos indicadores de desigualdade, e diferenças nesses índices por conta do gênero do trabalhador. Mas nos surpreendemos com uma desigualdade de renda maior no grupo de trabalhadores com ensino superior completo, em todos os anos. Por alguma razão, não deve haver uniformidade no deslocamento de ocupações de baixos salários para as de salários maiores, o que nos leva à conclusão de que a escolaridade, ao mesmo tempo que aumenta a renda esperada do trabalhador, também expõe o trabalhador a uma desigualdade maior na distribuição das rendas. Esse resultado contrasta com a visão de que a redução das desigualdades ocorre pela elevação dos salários mais baixos à medida que o nível educacional aumenta (Da Mata, 1979).

Com respeito à estrutura do emprego sobre a concentração de renda, identificamos uma influência equalizadora gerada por pequenas empresas. Estas se compõem de atividades mais homogêneas (em termos de produtividade da mão-de-obra e de tecnologia), o que provavelmente explica menores diferenciais de renda. Nosso resultado contrasta com a hipótese de que a desigualdade de renda seja promovida pelo setor terciário (Da Mata, 1979).

Com respeito ao número de horas de trabalho, a maior desigualdade de renda foi encontrada na categoria dos que trabalham de 21 a 40 horas. Nessa categoria, os índices foram maiores do que o índice geral do setor de turismo no nordeste (última linha da Tabela 2). Esse resultado, talvez, se explique pela bimodalidade na distribuição que se observa na Figura 4. Essa bimodalidade pode ser consequência de uma mistura de perfis, o que gera

heterogeneidade da distribuição segundo o número de horas de trabalho. Naturalmente, o número de horas de trabalho produz uma grande influência na renda, e talvez seu efeito se potencialize (como um efeito multiplicativo) dependendo das outras características do trabalhador (gênero, escolaridade, porte e idade).

E no que se refere à idade, embora ela se relacione naturalmente à diferenciação salarial (Menezes-Filho, Fernandes e Pichetti, 2000) nosso estudo conclui que esse efeito é pouco relevante no setor de turismo no nordeste. Isso ocorre porque, segundo a Tabela 1, a categoria predominante é aquela formada por trabalhadores com 25 a 49 anos, com 75% de participação. Nesse caso, os índices de Gini para essa faixa etária coincidiram com os índices gerais (última linha da Tabela 2).

Por outro lado, para estudos posteriores, é recomendável que se divida a faixa etária dessa classe para se ter uma ideia mais clara acerca do efeito da idade na distribuição de renda.

Nosso estudo também comparou mutuamente os índices de desigualdade de renda. Os resultados mostram que os índices de Theil, Gini, Atkinson e o coeficiente de variação são aproximadamente proporcionais. Já o coeficiente de assimetria se associa com a curtose, o que sugere que esses índices descrevem características de desigualdade não perceptíveis a outros índices.

5 REFERÊNCIAS BIBLIOGRÁFICAS

- International Recommendations for Tourism Statistics (IRTS 2008)*. (2008). Nova York.
- Ajala, O. A. (2008). Employment and Income Potentiality of Tourism Development in Amhara Region Ethiopia. *Ethiopian Journal of Environmental Studies and Management*, p. 74.
- Árias, A. R., Zamboni, R. A. (2009). *Sistema integrado de informações sobre o mercado de trabalho no setor turismo no Brasil*. Brasília: IPEA.
- Atkinson, A. B. (1983). *The economics of inequality*. Londres: Clarendon Press.
- Barbosa Filho, F. H., Pessôa, S. (2009). *Educação, Crescimento e Distribuição de Renda: A experiência Brasileira em perspectiva histórica*. http://www.alfaebeto.com.br/documentos/artigo_20091805.pdf.
- Bendel, R. B. (1989). Comparison of skewness coefficient, coefficient of variation, an Gini coefficient as inequality measures within populations. *Oecologia*, 394-400.
- Coelho, A., & Corseuil, C. (2002). *Diferenciais salariais no Brasil: um breve panorama*. Rio de Janeiro: IPEA.
- Coelho, A., Corseuil, C., Santos, D., Filho, N., & Fernandes, R. (2002). *Estrutura Salarial: aspectos conceituais e novos resultados para o Brasil*. Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada.
- Cooper, C. (2001). *Turismo, princípios e prática*. Porto Alegre: Bookman.
- Cramér, H. (1951). *Mathematical methods of statistics*. Princeton: Princeton University Press.
- Da Mata, M. (1979). *Concentração de Renda, Desemprego e Pobreza no Brasil*. Rio de Janeiro: IPEA.
- De Maio, F. G. (2007). Income Inequality Measures. *Journal of Epidemiol Community Health*, pp. 849-852.
- Gomes, A., & Wajnman, S. (2005). Diferenciais salariais segundo posição de homens e mulheres nas famílias. Seminário as Famílias e as Políticas Públicas no Brasil. In: *Seminário as Famílias e as Políticas Públicas no Brasil*. Belo Horizonte: <http://www.abep.nepo.unicamp.br/docs/anais/outros/FamPolPublicas/GomesWajnman.pdf>.
- Hoffman, R. (1980). *Estatística para economistas*. São Paulo: Livraria Pioneira .
- Jenkins, S. P., & Van Kerm, P. (2008). The Measurement of Economic Inequality. *Oxford Handbook on Economic Inequality*.
- Menezes-Filho, N., Fernandes, R., & Pichetti, P. (2000). A evolução da distribuição de salários no Brasil: fatos estilizados para as décadas de 80 e 90. In: H. R. (ed.), *Desigualdade e Pobreza no Brasil*. (pp. 231-250). Rio de Janeiro: IPEA.
- Organization, U. -U. (2001). *Cuenta satélite de turismo: recomendaciones sobre el marco conceptual*. Luxemburgo, Madrid, Nova York, Paris.
- Pant, S. (2011). The Impact of Tourism on Income Inequality: An Econometric Assessment. *The UCLA ungergraduate*, 48-99.
- Pinheiro, L., Galiza, M., & Fontoura, N. (2011). Novos arranjos familiares, velhas convenções sociais de gênero: a licença parental como política pública para lidar com estas tensões. In: A. L. BONETTI, & M. A. ABREU, *Faces da Desigualdade de Gênero e Raça no Brasil* (pp. 45-60). Brasília: IPEA.
- Sakowski, P. A. (2013). *Aspectos metodológicos do sistema integrado de informações sobre o mercado de trabalho no setor turismo*. Rio de Janeiro: IPEA.

- Scott, D. W. (1992). *Multivariate density estimation*. Nova York: Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics.
- Sheater, S. J. (2004). Density Estimation. *Statistical Science*, pp. 588-597.
- Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. *Monographs on Statistics and Applied probability*.
- Stynes, J. D. (1997). ECONOMIC IMPACTS OF TOURISM: A Handbook for Tourism Professionals. *Tourism Research Laboratory at the University of Illinois at Urbana-Champaign*, p. 3.
- Swarbrooke, J., & Horner, S. (2002). *O comportamento do consumidor no turismo*. São Paulo: Aleph.
- Theil, H. (1967). *Economics and information theory*. Amsterdã: North-Holland.

APÊNDICE A – Lista de CNAES referentes ao Turismo

1. Alojamento

55.10-8 - Hotéis e similares

5510-8/01 Hotéis

5510-8/02 Apart-hotéis

5510-8/03 Motéis

55.90-6 - Outros tipos de alojamento não especificados anteriormente

5590-6/01 Albergues, exceto assistenciais

5590-6/02 Campings

590-6/03 Pensões (alojamento)

5590-6/99 Outros alojamentos não especificados anteriormente

2. Alimentação

56.11-2 - Restaurantes e outros estabelecimentos de serviços de alimentação e bebidas

5611-2/01 Restaurantes e similares

5611-2/02 Bares e outros estabelecimentos especializados em servir bebidas

5611-2/03 Lanchonetes, casas de chá, de sucos e similares

56.12-1 - Serviços ambulantes de alimentação

5612-1/00 Serviços ambulantes de alimentação

3. Transporte terrestre

49.23-0 - Transporte rodoviário de táxi

4923-0/01 Serviço de táxi

4923-0/02 Serviço de transporte de passageiros- locação de automóveis com motorista

49.29-9 - Transporte rodoviário coletivo de passageiros, sob regime de fretamento, e outros transportes rodoviários não especificados anteriormente

4929-9/01 Transporte rodoviário coletivo de passageiros, sob regime de fretamento, municipal

4929-9/03 Organização de excursões em veículos rodoviários próprios, municipal

4929-9/99 Outros transportes rodoviários de passageiros não especificados anteriormente

49.50-7 - Trens turísticos, teleféricos e similares

4950-7/00 Trens turísticos, teleféricos e similares

49.22-1 - Transporte rodoviário coletivo de passageiros, com itinerário fixo, intermunicipal, interestadual e internacional

4922-1/01 Transporte rodoviário coletivo de passageiros, com itinerário fixo, intermunicipal, exceto em região metropolitana

4922-1/02 Transporte rodoviário coletivo de passageiros, com itinerário fixo, interestadual

4922-1/03 Transporte rodoviário coletivo de passageiros, com itinerário fixo, internacional

49.29-9 - Transporte rodoviário coletivo de passageiros, sob regime de fretamento, e outros transportes rodoviários não especificados anteriormente

4929-9/02 Transporte rodoviário coletivo de passageiros, sob regime de fretamento intermunicipal, interestadual e internacional

4929-9/04 Organização de excursões em veículos rodoviários próprios, intermunicipal, interestadual e internacional

4. Transporte aquaviário

50.11-4 - Transporte marítimo de cabotagem

5011-4/02 Transporte marítimo de cabotagem - passageiros

50.12-2 - Transporte marítimo de longo curso

5012-2/02 Transporte marítimo de longo curso- passageiro

50.22-0 - Transporte por navegação interior de passageiros em linhas regulares

5022-0/01 Transporte por navegação interior de passageiros em linhas regulares, municipal, exceto travessia

5022-0/02 Transporte por navegação interior de passageiros em linhas regulares, intermunicipal, interestadual e internacional, exceto travessia

50.91-2 - Transporte por navegação de travessia

5091-2/01 Transporte por navegação de travessia, municipal

5091-2/02 Transporte por navegação de travessia, intermunicipal

50.99-8 - Transporte aquaviários não especificados anteriormente

5099-8/01 Transporte aquaviário para passeios turísticos

5099-8/99 Outros transportes aquaviários não especificados anteriormente

5. Transporte aéreo

51.11-1 - Transporte aéreo de passageiros regular

5111-1/00 Transporte aéreo de passageiros regular

51.12-9 - Transporte aéreo de passageiros não regular

5112-9/01 Serviço de táxi aéreo e locação de aeronaves com tripulação

5112-9/99 Outros serviços de transporte de passageiros não regular

6. Aluguel de Transporte

77.11-0 - Locação de automóveis sem condutor

7711-0/00 Locação de automóveis sem condutor

7. Agências de Viagem

79.11-2 - Agências de viagens

7911-2/00 Agências de viagens

79.12-1 - Operadores turísticos

7912-1/00 Operadores turísticos

79.90-2 - Serviços de reserva e outros serviços de turismo não especificados anteriormente

7990-2/00 Serviços de reserva e outros serviços de turismo não especificados anteriormente

8. Cultura e Lazer

90.01-9 - Artes cênicas, espetáculos e atividades complementares

9001-9/01 Produção teatral

9001-9/02 Produção musical

9001-9/03 Produção de espetáculos de dança

9001-9/04 Produção de espetáculos circenses, de marionetes e similares

9001-9/05 Produção de espetáculos de rodeios, vaquejadas e similares

9001-9/99 Artes cênicas, espetáculos e atividades complementares não especificadas anteriormente

91.02-3 - Atividades de museus e de exploração, restauração artística e conservação de lugares e prédios históricos e atrações similares

9102-3/01 Atividades de museus e de exploração de lugares e prédios históricos e atrações similares

91.03-1 - Atividades de jardins botânicos, zoológicos, parques nacionais, reservas ecológicas e áreas de proteção ambiental

9103-1/00 Atividades de jardins botânicos, zoológicos, parques nacionais, reservas ecológicas e áreas de proteção ambiental

92.00-3 - Atividades de exploração de jogos de azar e apostas

9200-3/01 Casas de bingo

9200-3/02 Exploração de apostas em corridas de cavalos

9200-3/99 Exploração de jogos de azar e apostas não especificados anteriormente

93.19-1 - Atividades esportivas não especificadas anteriormente

9319-1/01 Produção e promoção de eventos esportivos

9319-1/99 Outras atividades esportivas não especificadas anteriormente

93.21-2 - Parques de diversão e parques temáticos

9321-2/00 Parques de diversão e parques temáticos

93.29-8 - Atividades de recreação e lazer não especificadas anteriormente

9329-8/01 Discotecas, danceterias, salões de dança e similares

9329-8/02 Exploração de boliches

9329-8/03 Exploração de jogos de sinuca, bilhar e similares

9329-8/04 Exploração de jogos eletrônicos recreativos

9329-8/99 Outras atividades de recreação e lazer não especificadas anteriormente

APÊNDICE B

Programas em SAS 9.3 da preparação de dados e análises

```
libname R 'C:\Users\Barbara\Desktop\todos';
libname S 'C:\Users\Barbara\Desktop\Tur';
libname T 'C:\Users\Barbara\Desktop\Monog';
libname G 'C:\Users\Barbara\Desktop\Gini';
libname A 'C:\Users\Barbara\Desktop\ANALISE';
libname F 'C:\Users\Barbara\Desktop\FREQ';

/*****1ª PARTE : SELEÇÃO DE VARIÁVEIS DE INTERESSE*****/

%macro TURISMO(UF= , ANO= );
data S. &uf&ano;
set R. &uf&ano;
    if SBCLAS20='5510801' or SBCLAS20='5510802' or SBCLAS20='5510803' or
SBCLAS20='5590601' or SBCLAS20='5590602' or SBCLAS20='5590603' or
SBCLAS20='5590699' or SBCLAS20='5611201' or
    SBCLAS20='5611202' or SBCLAS20='5611203' or SBCLAS20='5612100' or
SBCLAS20='4923001' or SBCLAS20='4923002' or SBCLAS20='4929901' or
SBCLAS20='4929903' or SBCLAS20='4929999' or
    SBCLAS20='4950700' or SBCLAS20='4922101' or SBCLAS20='4922102' or
SBCLAS20='4922103' or SBCLAS20='4929902' or SBCLAS20='4929904' or
SBCLAS20='5011402' or SBCLAS20='5012202' or SBCLAS20='5022001' or
    SBCLAS20='5022002' or SBCLAS20='5091201' or SBCLAS20='5091202' or
SBCLAS20='5099801' or SBCLAS20='5099899' or SBCLAS20='5111100' or
SBCLAS20='5112901' or SBCLAS20='5112999' or
    SBCLAS20='7711000' or SBCLAS20='7911200' or SBCLAS20='7912100' or
SBCLAS20='7990200' or SBCLAS20='9001901' or SBCLAS20='9001902' or
SBCLAS20='9001903' or SBCLAS20='9001904' or
    SBCLAS20='9001905' or SBCLAS20='9001999' or SBCLAS20='9102301' or
SBCLAS20='9103100' or SBCLAS20='9200301' or SBCLAS20='9200302' or
SBCLAS20='9200399' or SBCLAS20='9319101' or
    SBCLAS20='9319199' or SBCLAS20='9321200' or SBCLAS20='9329801' or
SBCLAS20='9329802' or SBCLAS20='9329803' or SBCLAS20='9329804' or
SBCLAS20='9329899' then output; /*cnaes do turismo */
run;
%mend TURISMO;

%TURISMO(UF= AL , ano= 2007); %TURISMO(UF= BA, ano = 2007); %TURISMO(UF=
CE, ano= 2007); %TURISMO(UF= MA, ano= 2007);
%TURISMO(UF= PB, ano= 2007); %TURISMO(UF= PE, ano= 2007); %TURISMO(UF= PI,
ano = 2007); %TURISMO(UF= RN, ano= 2007);
%TURISMO(UF= SE, ano= 2007); %TURISMO(UF= AL , ano= 2008); %TURISMO(UF= BA,
ano = 2008); %TURISMO(UF= CE, ano= 2008);
%TURISMO(UF= MA, ano= 2008); %TURISMO(UF= PB, ano= 2008); %TURISMO(UF= PE,
ano= 2008); %TURISMO(UF= PI, ano = 2008);
%TURISMO(UF= RN, ano= 2008); %TURISMO(UF= SE, ano= 2008); %TURISMO(UF= AL ,
ano= 2009); %TURISMO(UF= BA, ano = 2009);
%TURISMO(UF= CE, ano= 2009); %TURISMO(UF= MA, ano= 2009); %TURISMO(UF= PB,
ano= 2009); %TURISMO(UF= PE, ano= 2009);
%TURISMO(UF= PI, ano = 2009); %TURISMO(UF= RN, ano= 2009); %TURISMO(UF= SE,
ano= 2009); %TURISMO(UF= AL , ano= 2010);
%TURISMO(UF= BA, ano = 2010); %TURISMO(UF= CE, ano= 2010); %TURISMO(UF= MA,
ano= 2010); %TURISMO(UF= PB, ano= 2010);
```



```

%TURISMO(UF= PE, ano= 2010); %TURISMO(UF= PI, ano = 2010); %TURISMO(UF= RN,
ano= 2010); %TURISMO(UF= SE, ano= 2010);

%macro AL(ANO);
data S. AL&ano; set S. AL&ano; if REMDEZR_ >= 0 then UF = 'AL'; run;
%mend AL;
%AL(2007); %AL(2009);
%macro BA(ANO);
data S. BA&ano; set S. BA&ano; if REMDEZR_ >= 0 then UF = 'BA'; run;
%mend BA;
%BA(2007); %BA(2009);
%macro CE(ANO);
data S. CE&ano; set S. CE&ano; if REMDEZR_ >= 0 then UF = 'CE'; run;
%mend CE;
%CE(2007); %CE(2009);
%macro MA(ANO);
data S. MA&ano; set S. MA&ano; if REMDEZR_ >= 0 then UF = 'MA'; run;
%mend MA;
%MA(2007); %MA(2009);
%macro PB(ANO);
data S. PB&ano; set S. PB&ano; if REMDEZR_ >= 0 then UF = 'PB'; run;
%mend PB;
%PB(2007); %PB(2009);
%macro PE(ANO);
data S. PE&ano; set S. PE&ano; if REMDEZR_ >= 0 then UF = 'PE'; run;
%mend PE;
%PE(2007); %PE(2009);
%macro PI(ANO);
data S. PI&ano; set S. PI&ano; if REMDEZR_ >= 0 then UF = 'PI'; run;
%mend PI;
%PI(2007); %PI(2009);
%macro RN(ANO);
data S. RN&ano; set S. RN&ano; if REMDEZR_ >= 0 then UF = 'RN'; run;
%mend RN;
%RN(2007); %RN(2009);
%macro SE(ANO);
data S. SE&ano; set S. SE&ano; if REMDEZR_ >= 0 then UF = 'SE'; run;
%mend SE;
%SE(2007); %SE(2009);

%macro ARQUIVO(UF= , ano= );

data T. &uf&ano; set S. &uf&ano;
      x = log(REMDEZR_);
      if EMPEM3112=1 or EMPPEM3112='SIM' then output; /* apenas empregados
em 31/12 */
run;
data T. &uf&ano; set T. &uf&ano;
      if REMDEZR_ > 0 then output;
run;
data T. &uf&ano; set T. &uf&ano;
      if IDADE < 24 then IDADECAT = 1;
      else if IDADE >= 24 and IDADE < 50 then IDADECAT = 2;
      else if IDADE >= 50 then IDADECAT = 3;
      if HORASCONTR < 20 then HORASCONTRCAT = 1;
      else if HORASCONTR >= 20 and HORASCONTR < 40 then HORASCONTRCAT = 2;
      else if HORASCONTR >= 40 then HORASCONTRCAT = 3;
      if TAMESTAB = '1' or TAMESTAB = '2' then TAMESTABCAT = 1;

```

```

else if TAMESTAB = '3' or TAMESTAB = '4' or TAMESTAB = '5' then
TAMESTABCAT = 2;
else if TAMESTAB = '6' or TAMESTAB = '7' then TAMESTABCAT = 3;
else if TAMESTAB = '8' or TAMESTAB = '9' then TAMESTABCAT = 4;
if GRINSTRUCAO = '2' or GRINSTRUCAO = '3' then GRINSTRUCAOCAT = 1;
else if GRINSTRUCAO = '4' or GRINSTRUCAO = '5' or GRINSTRUCAO = '6'
then GRINSTRUCAOCAT = 2;
else if GRINSTRUCAO = '7' or GRINSTRUCAO = '8' then GRINSTRUCAOCAT =
3;
else if GRINSTRUCAO = '9' then GRINSTRUCAOCAT = 4;
if GENERO = 'MASCULINO' then GENEROCAT = 1;
else if GENERO = 'FEMININO' then GENEROCAT = 2;

if SBCLAS20='5510801' then CNAE = 1;
else if SBCLAS20='5510802' then CNAE = 2;
else if SBCLAS20='5510803' then CNAE = 3;
else if SBCLAS20='5590601' then CNAE = 4;
else if SBCLAS20='5590602' then CNAE = 5;
else if SBCLAS20='5590603' then CNAE = 6;
else if SBCLAS20='5590699' then CNAE = 7;
else if SBCLAS20='5611201' then CNAE = 8;
else if SBCLAS20='5611202' then CNAE = 9;
else if SBCLAS20='5611203' then CNAE = 10;
else if SBCLAS20='5612100' then CNAE = 11;
else if SBCLAS20='4923001' then CNAE = 12;
else if SBCLAS20='4923002' then CNAE = 13;
else if SBCLAS20='4929901' then CNAE = 14;
else if SBCLAS20='4929903' then CNAE = 15;
else if SBCLAS20='4929999' then CNAE = 16;
else if SBCLAS20='4950700' then CNAE = 17;
else if SBCLAS20='4922101' then CNAE = 18;
else if SBCLAS20='4922102' then CNAE = 19;
else if SBCLAS20='4922103' then CNAE = 20;
else if SBCLAS20='4929902' then CNAE = 21;
else if SBCLAS20='4929904' then CNAE = 22;
else if SBCLAS20='5011402' then CNAE = 23;
else if SBCLAS20='5012202' then CNAE = 24;
else if SBCLAS20='5022001' then CNAE = 25;
else if SBCLAS20='5022002' then CNAE = 26;
else if SBCLAS20='5091201' then CNAE = 27;
else if SBCLAS20='5091202' then CNAE = 28;
else if SBCLAS20='5099801' then CNAE = 29;
else if SBCLAS20='5099899' then CNAE = 30;
else if SBCLAS20='5111100' then CNAE = 31;
else if SBCLAS20='5112901' then CNAE = 32;
else if SBCLAS20='5112999' then CNAE = 33;
else if SBCLAS20='7711000' then CNAE = 34;
else if SBCLAS20='7911200' then CNAE = 35;
else if SBCLAS20='7912100' then CNAE = 36;
else if SBCLAS20='7990200' then CNAE = 37;
else if SBCLAS20='9001901' then CNAE = 38;
else if SBCLAS20='9001902' then CNAE = 39;
else if SBCLAS20='9001903' then CNAE = 40;
else if SBCLAS20='9001904' then CNAE = 41;
else if SBCLAS20='9001905' then CNAE = 42;
else if SBCLAS20='9001999' then CNAE = 43;
else if SBCLAS20='9102301' then CNAE = 44;
else if SBCLAS20='9103100' then CNAE = 45;
else if SBCLAS20='9200301' then CNAE = 46;

```

```

else if SBCLAS20='9200302' then CNAE = 47;
else if SBCLAS20='9200399' then CNAE = 48;
else if SBCLAS20='9319101' then CNAE = 49;
else if SBCLAS20='9319199' then CNAE = 50;
else if SBCLAS20='9321200' then CNAE = 51;
else if SBCLAS20='9329801' then CNAE = 52;
else if SBCLAS20='9329802' then CNAE = 53;
else if SBCLAS20='9329803' then CNAE = 54;
else if SBCLAS20='9329804' then CNAE = 55;
else if SBCLAS20='9329899' then CNAE = 56;

run;
data T. &uf&ano; set T. &uf&ano;
      IF GRINSTRUCAOCAT = '1' OR GRINSTRUCAOCAT = '2' OR GRINSTRUCAOCAT =
'3' OR GRINSTRUCAOCAT = '4' THEN OUTPUT;
run;
data T. &uf&ano; set T. &uf&ano;
      if generocat = '1' or generocat = '2' then ano = &ano;
run;
data T. &uf&ano; set T. &uf&ano;
      keep UF GENEROCAT GRINSTRUCAOCAT TAMESTABCAT HORASCONTRCAT IDADECAT
CNAE REMDEZR_ x ano;
run;
%mend ARQUIVO;
%ARQUIVO(UF= AL , ano= 2007); %ARQUIVO(UF= BA, ano= 2007); %ARQUIVO(UF= CE,
ano= 2007); %ARQUIVO(UF= MA, ano= 2007);
%ARQUIVO(UF= PB, ano= 2007); %ARQUIVO(UF= PE, ano= 2007); %ARQUIVO(UF= PI,
ano= 2007); %ARQUIVO(UF= RN, ano= 2007);
%ARQUIVO(UF= SE, ano= 2007); %ARQUIVO(UF= AL , ano= 2008); %ARQUIVO(UF= BA,
ano= 2008); %ARQUIVO(UF= CE, ano= 2008);
%ARQUIVO(UF= MA, ano= 2008); %ARQUIVO(UF= PB, ano= 2008); %ARQUIVO(UF= PE,
ano= 2008); %ARQUIVO(UF= PI, ano= 2008);
%ARQUIVO(UF= RN, ano= 2008); %ARQUIVO(UF= SE, ano= 2008); %ARQUIVO(UF= AL ,
ano= 2009); %ARQUIVO(UF= BA, ano= 2009);
%ARQUIVO(UF= CE, ano= 2009); %ARQUIVO(UF= MA, ano= 2009); %ARQUIVO(UF= PB,
ano= 2009); %ARQUIVO(UF= PE, ano= 2009);
%ARQUIVO(UF= PI, ano= 2009); %ARQUIVO(UF= RN, ano= 2009); %ARQUIVO(UF= SE,
ano= 2009); %ARQUIVO(UF= AL , ano= 2010);
%ARQUIVO(UF= BA, ano= 2010); %ARQUIVO(UF= CE, ano= 2010); %ARQUIVO(UF= MA,
ano= 2010); %ARQUIVO(UF= PB, ano= 2010);
%ARQUIVO(UF= PE, ano= 2010); %ARQUIVO(UF= PI, ano= 2010); %ARQUIVO(UF= RN,
ano= 2010); %ARQUIVO(UF= SE, ano= 2010);

%macro MERGE(ES = , UF1 = , UF2 = , UF3 = , UF4 = , UF5 = , UF6 = , UF7 = ,
UF8 = , UF9 = , ano = );
data T. &ES&ano;
      merge T. &UF1&ano T. &UF2&ano T. &UF3&ano T. &UF4&ano T. &UF5&ano T.
&UF6&ano T. &UF7&ano T. &UF8&ano T. &UF9&ano;
      by UF;
run;
%mend MERGE;
%MERGE(ES = NE, UF1 = AL, UF2 = BA, UF3 = CE, UF4 = MA, UF5 = PB, UF6 = PE,
UF7 = PI, UF8 = RN, UF9 = SE, ano = 2007);
%MERGE(ES = NE, UF1 = AL, UF2 = BA, UF3 = CE, UF4 = MA, UF5 = PB, UF6 = PE,
UF7 = PI, UF8 = RN, UF9 = SE, ano = 2008);
%MERGE(ES = NE, UF1 = AL, UF2 = BA, UF3 = CE, UF4 = MA, UF5 = PB, UF6 = PE,
UF7 = PI, UF8 = RN, UF9 = SE, ano = 2009);
%MERGE(ES = NE, UF1 = AL, UF2 = BA, UF3 = CE, UF4 = MA, UF5 = PB, UF6 = PE,
UF7 = PI, UF8 = RN, UF9 = SE, ano = 2010);

```

```

%macro MERGE(UF = , ano1 = , ano2 = , ano3 = , ano4 = );
data A. &uf.ano;
    merge T. &uf&ano1 T. &uf&ano2 T. &uf&ano3 T. &uf&ano4;
    by ano;
run;
%mend MERGE;
%MERGE(UF = NE, ano1 = 2007, ano2 = 2008, ano3 = 2009, ano4 = 2010);

/*****2ª PARTE : ANÁLISE EXPLORATÓRIA DE DADOS*****/

%macro ANALISE(UF = , VARCAT = );
ods html gpath = 'C:\Users\Barbara\Desktop\ANALISE';
proc sgpanel data = A. &uf.ano; title "Distribuição do log da renda por
&varcat - &uf";
panelby &varcat ano / onepanel;
    density x;
    density x / type=kernel;
run;
ods html close;
%mend ANALISE;
%ANALISE(UF= NE, VARCAT= GENEROCAT); %ANALISE(UF= NE, VARCAT=
GRINSTRUCAOCAT); %ANALISE(UF= NE, VARCAT= TAMESTABCAT);
%ANALISE(UF= NE, VARCAT= HORASCONTRCAT); %ANALISE(UF= NE, VARCAT=
IDADECAT);

%macro FREQ(UF = , ANO = ,VARCAT = );
proc freq data = T. &uf&ano; tables &varcat / out = &uf&ano&varcat; run;
proc print &uf&ano&varcat; run;
data F. &uf&ano&varcat; set &uf&ano&varcat; run;
%mend FREQ;
%macro FREQF(varcat);
%FREQ(UF = NE, ANO = 2007,VARCAT = &varcat); %FREQ(UF = NE, ANO =
2008,VARCAT = &varcat);
%FREQ(UF = NE, ANO = 2009,VARCAT = &varcat); %FREQ(UF = NE, ANO =
2010,VARCAT = &varcat);
%mend FREQF;
%FREQF( GENEROCAT );%FREQF( GRINSTRUCAOCAT );%FREQF( TAMESTABCAT );%FREQF( HORASCO
NTRCAT );%FREQF( IDADECAT );
%FREQF( CNAE );

```

APÊNDICE C

Programação em SAS 9.3 do cálculo do Índice de Gini

```
%macro GINI(UF = , ANO = , VAR = );
/*****GINI*****/
data teste; set T. &UF&ANO (KEEP = &VAR REMDEZR_);
    if REMDEZR_ > 0;
run;
proc sort data=teste; by REMDEZR_; run;
data testel teste2 teste3 teste4 teste5 teste6 teste7 teste8 teste9 teste10
testel1 testel2 testel3 testel4 testel5
    testel6 testel7 testel8 testel9 teste20 teste21 teste22 teste23
teste24 teste25 teste26 teste27 teste28 teste29
    teste30 teste31 teste32 teste33 teste34 teste35 teste36 teste37
teste38 teste39 teste40 teste41 teste42 teste43
    teste44 teste45 teste46 teste47 teste48 teste49 teste50 teste51
teste52 teste53 teste54 teste55 teste56;
set teste;
    if &VAR = '1' then output testel;
    else if &VAR = '2' then output teste2;    else if &VAR = '3' then
output teste3;
    else if &VAR = '4' then output teste4;    else if &VAR = '5' then
output teste5;
    else if &VAR = '6' then output teste6;    else if &VAR = '7' then
output teste7;
    else if &VAR = '8' then output teste8;    else if &VAR = '9' then
output teste9;
    else if &VAR = '10' then output testel0; else if &VAR = '11' then
output testel1;
    else if &VAR = '12' then output testel2; else if &VAR = '13' then
output testel3;
    else if &VAR = '14' then output testel4; else if &VAR = '15' then
output testel5;
    else if &VAR = '16' then output testel6; else if &VAR = '17' then
output testel7;
    else if &VAR = '18' then output testel8; else if &VAR = '19' then
output testel9;
    else if &VAR = '20' then output teste20; else if &VAR = '21' then
output teste21;
    else if &VAR = '22' then output teste22; else if &VAR = '23' then
output teste23;
    else if &VAR = '24' then output teste24; else if &VAR = '25' then
output teste25;
    else if &VAR = '26' then output teste26; else if &VAR = '27' then
output teste27;
    else if &VAR = '28' then output teste28; else if &VAR = '29' then
output teste29;
    else if &VAR = '30' then output teste30; else if &VAR = '31' then
output teste31;
    else if &VAR = '32' then output teste32; else if &VAR = '33' then
output teste33;
    else if &VAR = '34' then output teste34; else if &VAR = '35' then
output teste35;
    else if &VAR = '36' then output teste36; else if &VAR = '37' then
output teste37;
```

```

        else if &VAR = '38' then output teste38; else if &VAR = '39' then
output teste39;
        else if &VAR = '40' then output teste40; else if &VAR = '41' then
output teste41;
        else if &VAR = '42' then output teste42; else if &VAR = '43' then
output teste43;
        else if &VAR = '44' then output teste44; else if &VAR = '45' then
output teste45;
        else if &VAR = '46' then output teste46; else if &VAR = '47' then
output teste47;
        else if &VAR = '48' then output teste48; else if &VAR = '49' then
output teste49;
        else if &VAR = '50' then output teste50; else if &VAR = '51' then
output teste51;
        else if &VAR = '52' then output teste52; else if &VAR = '53' then
output teste53;
        else if &VAR = '54' then output teste54; else if &VAR = '55' then
output teste55;
        else output teste56;
run;
%let i = i;
%do i = 1 %to 56;
data TESTE&i; set TESTE&i;
        RENDACUM+REMDEZR_;
run;
%END;
data teste; set testel1 teste2 teste3 teste4 teste5 teste6 teste7 teste8
teste9 testel10 testel11 testel12 testel13 testel14
        testel15        testel16 testel17 testel18 testel19 teste20 teste21 teste22
teste23 teste24 teste25 teste26 teste27 teste28
        teste29        teste30 teste31 teste32 teste33 teste34 teste35 teste36
teste37 teste38 teste39 teste40 teste41 teste42
        teste43        teste44 teste45 teste46 teste47 teste48 teste49 teste50
teste51 teste52 teste53 teste54 teste55 teste56;
run;
proc sql;
        create table testefi as select &VAR,
        REMDEZR_,
        RENDACUM/SUM(REMDEZR_) as FI,
        from teste
        group by &VAR;
quit;
proc sort data=testefi; by FI; run;;
data testefi1 testefi2 testefi3 testefi4 testefi5 testefi6 testefi7
testefi8 testefi9 testefi10 testefi11 testefi12
        testefi13 testefi14 testefi15 testefi16 testefi17 testefi18 testefi19
testefi20 testefi21 testefi22 testefi23
        testefi24 testefi25 testefi26 testefi27 testefi28 testefi29 testefi30
testefi31 testefi32 testefi33 testefi34
        testefi35 testefi36 testefi37 testefi38 testefi39 testefi40 testefi41
testefi42 testefi43 testefi44 testefi45
        testefi46 testefi47 testefi48 testefi49 testefi50 testefi51 testefi52
testefi53 testefi54 testefi55 testefi56;
set testefi;
        if &VAR = '1' then output testefi1;
        else if &VAR = '2' then output testefi2; else if &VAR = '3' then
output testefi3;
        else if &VAR = '4' then output testefi4; else if &VAR = '5' then
output testefi5;

```

```

        else if &VAR = '6' then output testefi6; else if &VAR = '7' then
output testefi7;
        else if &VAR = '8' then output testefi8; else if &VAR = '9' then
output testefi9;
        else if &VAR = '10' then output testefi10;        else if &VAR = '11'
then output testefi11;
        else if &VAR = '12' then output testefi12;        else if &VAR = '13'
then output testefi13;
        else if &VAR = '14' then output testefi14;        else if &VAR = '15'
then output testefi15;
        else if &VAR = '16' then output testefi16;        else if &VAR = '17'
then output testefi17;
        else if &VAR = '18' then output testefi18;        else if &VAR = '19'
then output testefi19;
        else if &VAR = '20' then output testefi20;        else if &VAR = '21'
then output testefi21;
        else if &VAR = '22' then output testefi22;        else if &VAR = '23'
then output testefi23;
        else if &VAR = '24' then output testefi24;        else if &VAR = '25'
then output testefi25;
        else if &VAR = '26' then output testefi26;        else if &VAR = '27'
then output testefi27;
        else if &VAR = '28' then output testefi28;        else if &VAR = '29'
then output testefi29;
        else if &VAR = '30' then output testefi30;        else if &VAR = '31'
then output testefi31;
        else if &VAR = '32' then output testefi32;        else if &VAR = '33'
then output testefi33;
        else if &VAR = '34' then output testefi34;        else if &VAR = '35'
then output testefi35;
        else if &VAR = '36' then output testefi36;        else if &VAR = '37'
then output testefi37;
        else if &VAR = '38' then output testefi38;        else if &VAR = '39'
then output testefi39;
        else if &VAR = '40' then output testefi40;        else if &VAR = '41'
then output testefi41;
        else if &VAR = '42' then output testefi42;        else if &VAR = '43'
then output testefi43;
        else if &VAR = '44' then output testefi44;        else if &VAR = '45'
then output testefi45;
        else if &VAR = '46' then output testefi46;        else if &VAR = '47'
then output testefi47;
        else if &VAR = '48' then output testefi48;        else if &VAR = '49'
then output testefi49;
        else if &VAR = '50' then output testefi50;        else if &VAR = '51'
then output testefi51;
        else if &VAR = '52' then output testefi52;        else if &VAR = '53'
then output testefi53;
        else if &VAR = '54' then output testefi54;        else if &VAR = '55'
then output testefi55;
        else output testefi56;
run;
%do i = 1 %to 56;
data testefi&i; set testefi&i;
    F11 = LAG(FI); IF F11 = '.' THEN F11 = 0;
    SUMFI = FI+F11;
run;
%end;

```

```

data testesumfi; set testefi1 testefi2 testefi3 testefi4 testefi5 testefi6
testefi7 testefi8 testefi9 testefi10 testefi11 testefi12
    testefi13 testefi14 testefi15 testefi16 testefi17 testefi18 testefi19
testefi20 testefi21 testefi22 testefi23
    testefi24 testefi25 testefi26 testefi27 testefi28 testefi29 testefi30
testefi31 testefi32 testefi33 testefi34
    testefi35 testefi36 testefi37 testefi38 testefi39 testefi40 testefi41
testefi42 testefi43 testefi44 testefi45
    testefi46 testefi47 testefi48 testefi49 testefi50 testefi51 testefi52
testefi53 testefi54 testefi55 testefi56;
run;
proc sql;
    create table testegh as select distinct &VAR,
        1 - (1/COUNT(&VAR))*SUM(SUMFI) as Gini
    from testesumfi
    group by &VAR;
quit;
data G. G&UF&ANO&VAR; SET testegh; run;
%mend GINI;

%macro GINIFINAL(varcat);
%GINI(UF = NE, ANO = 2007, VAR = &varcat); %GINI(UF = NE, ANO = 2008, VAR =
&varcat);
%GINI(UF = NE, ANO = 2009, VAR = &varcat); %GINI(UF = NE, ANO = 2010, VAR =
&varcat);
%mend GINIFINAL;

%GINIFINAL(GENEROCAT); %GINIFINAL(GRINSTRUCAOCAT); %GINIFINAL(TAMESTABCAT);
%GINIFINAL(HORASCONTRCAT);
%GINIFINAL(IDADECAT); %GINIFINAL(CNAE);

```


APÊNDICE D

Programação em SAS 9.3 do cálculo do Índice de Theil

```
%macro THEIL(UF = , ANO = , VAR = );
/*****THEIL*****/
data teste; set T. &UF&ANO (KEEP = &VAR REMDEZR_);
    if REMDEZR_ > 0;
run;
proc sort data=teste; by REMDEZR_; run;
data testel teste2 teste3 teste4 teste5 teste6 teste7 teste8 teste9 testel0
testel1 testel2 testel3 testel4 testel5
    testel6 testel7 testel8 testel9 teste20 teste21 teste22 teste23
teste24 teste25 teste26 teste27 teste28 teste29
    teste30 teste31 teste32 teste33 teste34 teste35 teste36 teste37
teste38 teste39 teste40 teste41 teste42 teste43
    teste44 teste45 teste46 teste47 teste48 teste49 teste50 teste51
teste52 teste53 teste54 teste55 teste56;
set teste;
    if &VAR = '1' then output testel;
    else if &VAR = '2' then output teste2;    else if &VAR = '3' then
output teste3;
    else if &VAR = '4' then output teste4;    else if &VAR = '5' then
output teste5;
    else if &VAR = '6' then output teste6;    else if &VAR = '7' then
output teste7;
    else if &VAR = '8' then output teste8;    else if &VAR = '9' then
output teste9;
    else if &VAR = '10' then output testel0; else if &VAR = '11' then
output testel1;
    else if &VAR = '12' then output testel2; else if &VAR = '13' then
output testel3;
    else if &VAR = '14' then output testel4; else if &VAR = '15' then
output testel5;
    else if &VAR = '16' then output testel6; else if &VAR = '17' then
output testel7;
    else if &VAR = '18' then output testel8; else if &VAR = '19' then
output testel9;
    else if &VAR = '20' then output teste20; else if &VAR = '21' then
output teste21;
    else if &VAR = '22' then output teste22; else if &VAR = '23' then
output teste23;
    else if &VAR = '24' then output teste24; else if &VAR = '25' then
output teste25;
    else if &VAR = '26' then output teste26; else if &VAR = '27' then
output teste27;
    else if &VAR = '28' then output teste28; else if &VAR = '29' then
output teste29;
    else if &VAR = '30' then output teste30; else if &VAR = '31' then
output teste31;
    else if &VAR = '32' then output teste32; else if &VAR = '33' then
output teste33;
    else if &VAR = '34' then output teste34; else if &VAR = '35' then
output teste35;
    else if &VAR = '36' then output teste36; else if &VAR = '37' then
output teste37;
```

```

        else if &VAR = '38' then output teste38; else if &VAR = '39' then
output teste39;
        else if &VAR = '40' then output teste40; else if &VAR = '41' then
output teste41;
        else if &VAR = '42' then output teste42; else if &VAR = '43' then
output teste43;
        else if &VAR = '44' then output teste44; else if &VAR = '45' then
output teste45;
        else if &VAR = '46' then output teste46; else if &VAR = '47' then
output teste47;
        else if &VAR = '48' then output teste48; else if &VAR = '49' then
output teste49;
        else if &VAR = '50' then output teste50; else if &VAR = '51' then
output teste51;
        else if &VAR = '52' then output teste52; else if &VAR = '53' then
output teste53;
        else if &VAR = '54' then output teste54; else if &VAR = '55' then
output teste55;
        else output teste56;
run;
%let i = i;
%do i = 1 %to 56;
data TESTE&i; set TESTE&i;
        RENDACUM+REMDEZR_;
run;
%END;
data teste; set testel1 teste2 teste3 teste4 teste5 teste6 teste7 teste8
teste9 testel10 testel11 testel12 testel13 testel14
        testel15        testel16 testel17 testel18 testel19 teste20 teste21 teste22
teste23 teste24 teste25 teste26 teste27 teste28
        teste29        teste30 teste31 teste32 teste33 teste34 teste35 teste36
teste37 teste38 teste39 teste40 teste41 teste42
        teste43        teste44 teste45 teste46 teste47 teste48 teste49 teste50
teste51 teste52 teste53 teste54 teste55 teste56;
run;
proc sort data=teste; by &VAR; run;
proc sql;
        create table TESTESOM as select
        &VAR,
        REMDEZR_,
        sum(REMDEZR_) as TOTRENDA
        from TESTE
        group by &VAR;
quit;
proc sql;
        create table TESTETERMOSTH as select
        &VAR,
        REMDEZR_/TOTRENDA as Yi,
        count(&VAR) as N
        from TESTESOM
        group by &VAR;
quit;
proc sql;
        create table TESTETH as select distinct
        &VAR,
        sum((Yi)*(log(N*(Yi)))) as THEIL
        from TESTETERMOSTH
        group by &VAR;
quit;

```

```

data G. T&uf&ano&VAR; set TESTETH; run;
%mend THEIL;

%macro THEILFINAL(varcat);
%THEIL(UF = NE, ANO = 2007, VAR = &varcat); %THEIL(UF = NE, ANO = 2008, VAR
= &varcat);
%THEIL(UF = NE, ANO = 2009, VAR = &varcat); %THEIL(UF = NE, ANO = 2010, VAR
= &varcat);
%mend THEILFINAL;

%THEILFINAL(GENEROCAT); %THEILFINAL(GRINSTRUCAOCAT);
%THEILFINAL(TAMESTABCAT);
%THEILFINAL(HORASCONTRCAT); %THEILFINAL(IDADECAT); %THEILFINAL(CNAE);

```

APÊNDICE E

Programação em SAS 9.3 do cálculo do Índice de Atkinson

```
%macro ATKINSON(UF = , ANO = , VAR = );
/*****ATKINSON*****/
*****/
data teste; set T. &UF&ANO (KEEP = &VAR REMDEZR_);
    if REMDEZR_ > 0;
run;
proc sort data=teste; by REMDEZR_; run;
data testel teste2 teste3 teste4 teste5 teste6 teste7 teste8 teste9 testel0
testel1 testel2 testel3 testel4 testel5
    teste16 teste17 teste18 teste19 teste20 teste21 teste22 teste23
teste24 teste25 teste26 teste27 teste28 teste29
    teste30 teste31 teste32 teste33 teste34 teste35 teste36 teste37
teste38 teste39 teste40 teste41 teste42 teste43
    teste44 teste45 teste46 teste47 teste48 teste49 teste50 teste51
teste52 teste53 teste54 teste55 teste56;
set teste;
    if &VAR = '1' then output testel;
    else if &VAR = '2' then output teste2;    else if &VAR = '3' then
output teste3;
    else if &VAR = '4' then output teste4;    else if &VAR = '5' then
output teste5;
    else if &VAR = '6' then output teste6;    else if &VAR = '7' then
output teste7;
    else if &VAR = '8' then output teste8;    else if &VAR = '9' then
output teste9;
    else if &VAR = '10' then output testel0; else if &VAR = '11' then
output testel1;
    else if &VAR = '12' then output testel2; else if &VAR = '13' then
output testel3;
    else if &VAR = '14' then output testel4; else if &VAR = '15' then
output testel5;
    else if &VAR = '16' then output teste16; else if &VAR = '17' then
output teste17;
    else if &VAR = '18' then output teste18; else if &VAR = '19' then
output teste19;
    else if &VAR = '20' then output teste20; else if &VAR = '21' then
output teste21;
    else if &VAR = '22' then output teste22; else if &VAR = '23' then
output teste23;
    else if &VAR = '24' then output teste24; else if &VAR = '25' then
output teste25;
    else if &VAR = '26' then output teste26; else if &VAR = '27' then
output teste27;
    else if &VAR = '28' then output teste28; else if &VAR = '29' then
output teste29;
    else if &VAR = '30' then output teste30; else if &VAR = '31' then
output teste31;
    else if &VAR = '32' then output teste32; else if &VAR = '33' then
output teste33;
    else if &VAR = '34' then output teste34; else if &VAR = '35' then
output teste35;
```

```

        else if &VAR = '36' then output teste36; else if &VAR = '37' then
output teste37;
        else if &VAR = '38' then output teste38; else if &VAR = '39' then
output teste39;
        else if &VAR = '40' then output teste40; else if &VAR = '41' then
output teste41;
        else if &VAR = '42' then output teste42; else if &VAR = '43' then
output teste43;
        else if &VAR = '44' then output teste44; else if &VAR = '45' then
output teste45;
        else if &VAR = '46' then output teste46; else if &VAR = '47' then
output teste47;
        else if &VAR = '48' then output teste48; else if &VAR = '49' then
output teste49;
        else if &VAR = '50' then output teste50; else if &VAR = '51' then
output teste51;
        else if &VAR = '52' then output teste52; else if &VAR = '53' then
output teste53;
        else if &VAR = '54' then output teste54; else if &VAR = '55' then
output teste55;
        else output teste56;
run;
proc sql;
    create table TESTESOM as select
    &VAR,
    REMDEZR_,
    sum(REMDEZR_) as TOTRENDA,
    count(&VAR) as N
    from TESTE
    group by &VAR;
quit;
proc sql;
    create table ATKINSON as select
    &VAR,
    REMDEZR_,
    N,
    TOTRENDA/N as MU
    from TESTESOM
    group by &VAR;
quit;
proc sql;
    create table ATKINSONF as select distinct
    &VAR,
    1-(N/sum(MU/REMDEZR_)) as AT /*CONSIDERANDO E=2*/
    from ATKINSON

    group by &VAR;
quit;
data G. A&f&ano&VAR; set ATKINSONF; run;
%mend ATKINSON;

%macro ATKINSONFINAL(varcat);
%ATKINSON(UF = NE, ANO = 2007, VAR = &varcat); %ATKINSON(UF = NE, ANO =
2008, VAR = &varcat);
%ATKINSON(UF = NE, ANO = 2009, VAR = &varcat); %ATKINSON(UF = NE, ANO =
2010, VAR = &varcat);
%mend ATKINSONFINAL;

```

%ATKINSONFINAL(GENEROCAT) ; **%ATKINSONFINAL**(GRINSTRUCAOCAT) ;
%ATKINSONFINAL(TAMESTABCAT) ;
%ATKINSONFINAL(HORASCONTRCAT) ; **%ATKINSONFINAL**(IDADECAT) ;
%ATKINSONFINAL(CNAE) ;

APÊNDICE F

Programação em SAS 9.3 do cálculo da curtose, coeficiente de assimetria e coeficiente de variação

```
%macro CURTASS(UF = , ANO = , VAR = );  
/*****CURTOSE E  
ASSIMETRIA*****/  
proc summary data=t. &uf&ano; class &var / GROUPINTERNAL; var x;  
    output out = G. CA&uf&ano&var kurt=curtose skew=assimetria cv=cv;  
run;  
data G. CA&uf&ano&var; set G. CA&uf&ano&var; run;  
%mend CURTASS;  
  
%macro CURTASSFINAL(varcat);  
%CURTASS(UF = NE, ANO = 2007, VAR = &varcat); %CURTASS(UF = NE, ANO = 2008,  
VAR = &varcat);  
%CURTASS(UF = NE, ANO = 2009, VAR = &varcat); %CURTASS(UF = NE, ANO = 2010,  
VAR = &varcat);  
%mend CURTASSFINAL;  
  
%CURTASSFINAL(GENEROCAT); %CURTASSFINAL(GRINSTRUCAOCAT);  
%CURTASSFINAL(TAMESTABCAT);  
%CURTASSFINAL(HORASCONTRCAT); %CURTASSFINAL(IDADECAT); %CURTASSFINAL(CNAE);
```

APÊNDICE G

Programação em SAS 9.3 plotagem dos gráficos de dispersão e coeficiente de correlação de Pearson

```
ods html gpath = 'C:\Users\Barbara\Desktop\GF';

symbol1 value=dot
           color=_style_;
symbol2 value=C
           font=marker
           color=_style_ ;
symbol3 value=circle
           color=_style_;
symbol4 value=triangle
           color=_style_;

axis1 label = (a=90 'COEFICIENTE DE ASSIMETRIA'); axis2 label =
('COEFICIENTE DE ASSIMETRIA');
axis3 label = (a=90 'COEFICIENTE DE VARIAÇÃO'); axis4 label = ('COEFICIENTE
DE VARIAÇÃO');
axis5 label = (a=90 'ATKINSON'); axis6 label = ('ATKINSON');
axis7 label = (a=90 'CURTOSE'); axis8 label = ('CURTOSE');
axis9 label = (a=90 'THEIL'); axis10 label = ('THEIL');
axis11 label = (a=90 'GINI'); axis12 label = ('GINI');

legend1 label=none value=('2007' '2008' '2009' '2010');

proc gplot data = GF. T_G;
  plot T2007*G2007 T2008*G2008 T2009*G2009 T2010*G2010 / overlay
vaxis=axis9 haxis=axis12 legend=legend1;
run;

proc gplot data = GF. T_A;
  plot T2007*A2007 T2008*A2008 T2009*A2009 T2010*A2010 / overlay
vaxis=axis9 haxis=axis6 legend=legend1;
run;

proc gplot data = GF. T_CURT;
  plot T2007*C2007 T2008*C2008 T2009*C2009 T2010*C2010 / overlay
vaxis=axis9 haxis=axis8 legend=legend1;
run;

proc gplot data = GF. T_ASS;
  plot T2007*AS2007 T2008*AS2008 T2009*AS2009 T2010*AS2010 / overlay
vaxis=axis9 haxis=axis2 legend=legend1;
run;

proc gplot data = GF. T_CV;
  plot T2007*CV2007 T2008*CV2008 T2009*CV2009 T2010*CV2010 / overlay
vaxis=axis9 haxis=axis4 legend=legend1;
run;
```



```

proc gplot data = GF. G_A;
    plot G2007*A2007 G2008*A2008 G2009*A2009 G2010*A2010 / overlay
vaxis=axis11 haxis=axis6 legend=legend1;
run;

proc gplot data = GF. G_CURT;
    plot G2007*C2007 G2008*C2008 G2009*C2009 G2010*C2010 / overlay
vaxis=axis11 haxis=axis8 legend=legend1;
run;

proc gplot data = GF. G_ASS;
    plot G2007*AS2007 G2008*AS2008 G2009*AS2009 G2010*AS2010 / overlay
vaxis=axis11 haxis=axis2 legend=legend1;
run;

proc gplot data = GF. G_CV;
    plot G2007*CV2007 G2008*CV2008 G2009*CV2009 G2010*CV2010 / overlay
vaxis=axis11 haxis=axis4 legend=legend1;
run;

proc gplot data = GF. A_CURT;
    plot A2007*C2007 A2008*C2008 A2009*C2009 A2010*C2010 / overlay
vaxis=axis5 haxis=axis8 legend=legend1;
run;

proc gplot data = GF. A_ASS;
    plot A2007*AS2007 A2008*AS2008 A2009*AS2009 A2010*AS2010 / overlay
vaxis=axis5 haxis=axis2 legend=legend1;
run;

proc gplot data = GF. A_CV;
    plot A2007*CV2007 A2008*CV2008 A2009*CV2009 A2010*CV2010 / overlay
vaxis=axis5 haxis=axis4 legend=legend1;
run;

proc gplot data = GF. CURT_ASS;
    plot C2007*AS2007 C2008*AS2008 C2009*AS2009 C2010*AS2010 / overlay
vaxis=axis7 haxis=axis2 legend=legend1;
run;

proc gplot data = GF. CURT_CV;
    plot C2007*CV2007 C2008*CV2008 C2009*CV2009 C2010*CV2010 / overlay
vaxis=axis7 haxis=axis4 legend=legend1;
run;

proc gplot data = GF. ASS_CV;
    plot AS2007*CV2007 AS2008*CV2008 AS2009*CV2009 AS2010*CV2010 /
overlay vaxis=axis1 haxis=axis4 legend=legend1;
run;

ods html close;

proc corr data = GF. TODOS outp = GF. CORR;
    var THEIL GINI ATKINSON CURTOSE ASSOCIAC COEF_VAR;
run;

```