



UNIVERSIDADE FEDERAL DO CEARÁ
CAEN - FEAAC

Marcelo Ponte Barbosa

**Uma análise exploratória dos diferenciais de salários entre
carreiras de nível superior**

Fortaleza - Ceará

2005



UNIVERSIDADE FEDERAL DO CEARÁ
CAEN - FEAAC

Marcelo Ponte Barbosa

**Uma análise exploratória dos diferenciais de salários entre
carreiras de nível superior**

Dissertação apresentada ao Curso de
Mestrado em Economia da
Universidade Federal do Ceará como
requisito para a obtenção do Título
de Mestre em Economia.

Orientador: Prof. Dr. José Raimundo Carvalho

Fortaleza – Ceará

2005

Marcelo Ponte Barbosa

**Uma análise exploratória dos diferenciais de salários entre
carreiras de nível superior**

Data de Aprovação: 13 de Julho de 2005

Banca Examinadora

Prof. Dr

José Raimundo Carvalho (orientador UFC)

Prof. Dr

Marcelo Lettieri Siqueira (membro EXTERNO)

Prof. Dr

Flávio Ataliba Flexa Daltro Barreto (membro UFC)

BARBOSA, Marcelo Ponte. Uma análise exploratória dos diferenciais de renda entre carreiras de nível superior. 2005. 123f. Dissertação (Mestrado). Universidade Federal do Ceará - UFC, CAEN. Fortaleza, 2005.

RESUMO

Este trabalho realiza uma análise exploratória dos diferenciais de risco e retorno entre vinte carreiras de nível superior. Como principal fonte de dados, utiliza-se a Amostra do Censo Demográfico do IBGE de 2000. Primeiramente, analisam-se as influências de características individuais sobre os diferenciais de salários para cada carreira. Explora-se o diferencial de salários entre diferentes grupos demográficos utilizando regressões de salários mincerianas. Em um segundo momento, critérios que consideram ambos o risco e o retorno são utilizados para comparar as distribuições de salários. Testes de dominância estocástica de primeira e segunda ordem são aplicados para estabelecer um *ranking* entre as profissões. Foi constatada a existência de uma grande heterogeneidade entre as diferentes carreiras no que concerne às características dos indivíduos em cada uma delas e a existência de um grande diferencial de salários por sexo, cor/raça e região. Constatou-se também que os salários de algumas carreiras sobem mais rapidamente com a experiência e chega ao equilíbrio mais rápido do que em outras. A carreira de Medicina domina estocasticamente todas as demais, enquanto Pedagogia é plenamente dominada. Os resultados encontrados sugerem que, para que políticas de educação superior direcionadas para a diminuição das desigualdades sejam efetivas, deve-se considerar a composição sócio-demográfica e os perfis de retornos das diferentes carreiras.

PALAVRAS CHAVES: Educação Superior, Diferenciais de Salários, Censo Demográfico, Regressão de Mincer, Dominância Estocástica.

Dedico este trabalho ao economista Raimundo Baima Barbosa

Agradecimentos

Aos meus pais, Raimundo e Rita, pela minha existência e por me ensinarem o caminho da perseverança, da honra e da verdade.

À Roberta, pela compreensão e apoio nos momentos mais difíceis e por todo amor que sempre me dedicou.

Aos meus irmãos, Andréa, Alexandre e Eduardo, meus grandes amigos, por contribuírem para o meu crescimento como pessoa e cidadão.

Ao meu orientador, Professor Doutor José Raimundo Carvalho, pela valiosa contribuição e dedicação na orientação desse trabalho, ao longo do qual recebi grande ensinamento, incentivo e valorização.

Aos Professores Doutores Flávio Ataliba e Marcelo Lettieri, pelas sugestões apresentadas quando da defesa desta dissertação.

Aos amigos e colegas de mestrado Jaqueline, Luiz Alberto, Daniel, Carlos Manso, David, Dílson, Débora e Nilo, pela amizade e companheirismo durante todo o período em que estivemos juntos.

Ao Kleber, pelas boas conversas nos momentos difíceis e aos funcionários do CAEN Carmem, Bibi, Constantino e Mônica, pela atenção que sempre dedicaram aos alunos da instituição.

À todos os que colaboraram para a realização desse trabalho.

ABSTRACT

This work carries through a exploratory analysis of the risk and return differentials between twenty higher education careers. As main source of data, it is used the Sample of the Demographic Census of the IBGE - 2000. At first, the influences of individual characteristics on wage differentials for each career are analyzed. Wage differentials between different demographic groups are explored using Mincer regressions. In a second moment, criterias that consider both risk and return are used to compare the distributions of wages. Tests of first and second orders stochastic dominance are applied to establish a ranking of the professions. The existence of a great heterogeneidade was evidenced between the careers in what concerns to the characteristics of the individuals. It was also evidenced the existence of a great differential of wages by gender, race/color and region. Wages for some careers go up more quickly with the experience and arrives at the equilibrium fastest than in others. Medicine dominates all other carriers while Pedagogy is fully dominated. The joined results suggest that, in order to be effective, higher education policies should consider the social-demographic composition and the profiles of returns of the different careers.

KEY-WORDS: Higher education, Wage differentials, Demografic census, Mincer regressions, Stochastic dominance.

SUMÁRIO

LISTA DE ILUSTRAÇÕES	x
INTRODUÇÃO	11
1. Aspectos Teóricos e Revisão da Literatura	15
1.1 O Investimento em Capital Humano	15
1.2 Regressão de Salários de Mincer	18
1.3 Diferenciais Salariais no Brasil	22
1.4 Escolha da Profissão como um Ativo de Risco	25
1.5 O Risco na Teoria Microeconômica	29
1.6 Teoria da Dominância Estocástica	32
2. Base de Dados e Descrição da Amostra	40
2.1 O Censo Demográfico do IBGE	40
2.1.1 O Censo de 2000	40
2.1.2 Aspectos da Amostragem do Censo	42
2.2 Descrição da Amostra	43
3. Ordenação de Distribuições utilizando Momentos Amostrais	52
4. Aplicação da Regressão Minceriana	56
5. Análise Não-Paramétrica dos Dados	64
5.1 Estimação Não-Paramétrica das Densidades	64
5.2 Aplicação do Teste de Normalidade	67
6. Teste Econométrico de Dominância Estocástica	71
6.1 O Teste de Davidson e Duclos	71
6.2 Resultados do Teste de Dominância Estocástica	76

CONCLUSÕES	82
REFERÊNCIAS BIBLIOGRÁFICAS	86
ANEXO 1 - REGRESSÃO DE SALÁRIOS (EXPERIÊNCIA = IDADE - IDADE MÉDIA DE FORMATURA)	91
ANEXO 2 - REGRESSÃO DE SALÁRIOS (EXPERIÊNCIA = IDADE - DURAÇÃO DA GRADUAÇÃO - 17 ANOS)	94
ANEXO 3 - DENSIDADES ESTIMADAS	97
ANEXO 4 - DISTRIBUIÇÕES AMOSTRAIS ACUMULADAS	101

LISTA DE ILUSTRAÇÕES

GRÁFICOS

Gráfico 1.1: Dominância Estocástica de Primeira Ordem (FSD)	38
Gráfico 1.2: Dominância Estocástica de Segunda Ordem (SSD)	39
Gráfico 3.1: Dispersão entre a média e o desvio-padrão dos salários/hora	53

FIGURAS

Figura 6.1: <i>Ranking</i> entre as dez carreiras segundo critérios de <i>FSD</i> e <i>SSD</i>	79
--	----

TABELAS

Tabela 2.1: Percentuais de indivíduos sem trabalho	44
Tabela 2.2: Distribuição dos indivíduos pela formação e região	45
Tabela 2.3: Distribuição dos indivíduos entre as diferentes formações por região	46
Tabela 2.4: Distribuição dos Indivíduos por Cor/Raça e Sexo	47
Tabela 2.5: Salário/Hora médio por Cor/Raça e Sexo	49
Tabela 2.6: Salário/hora médio por região	50
Tabela 2.7: Salário médio para diferentes grupos demográficos	51
Tabela 3.1: Ranking por Média, Desvio Padrão e Coeficiente de Variação	52
Tabela 3.2: Ranking sobre a utilidade esperada de Weiss	54
Tabela 4.1: Idade média dos participantes do Provão 1997-2001	58
Tabela 4.2: Duração média do curso superior	59
Tabela 5.1: Coeficientes de Assimetria e Curtose	67
Tabela 5.2: Teste de Normalidade de Jarque-Bera	69
Tabela 6.1: Relações de dominância estocástica entre as profissões	78
Tabela 6.2: <i>Ranking</i> das formações pelos momentos amostrais e utilidade esperada	80

INTRODUÇÃO

Ao escolher uma carreira, um indivíduo tenta quantificar o retorno e o risco embutido nesta escolha¹. Isto se torna um desafio, à medida em que o retorno financeiro do trabalho é uma realização com parte aleatória relevante, mesmo quando se comparam profissões similares ou indivíduos com características semelhantes. Neste contexto, a escolha de uma profissão pode ser comparada à escolha de um dentre vários ativos de risco, onde o retorno sobre o investimento educacional necessário para o desempenho profissional advém do rendimento futuro do trabalho².

Uma ordenação das carreiras de acordo com a atratividade dos retornos salariais, bem como a descrição das variáveis que afetam o retorno das diferentes carreiras são informações fundamentais na avaliação dos fatores que explicam o comportamento do mercado de trabalho. O alto nível de heterogeneidade encontrado entre as diversas profissões é confirmado pelas grandes diferenças entre perfis de salários e perfis demográficos dos indivíduos de acordo com a profissão³.

Baseado nessas constatações, pretende-se realizar uma análise exploratória das diferentes carreiras de nível superior quanto aos perfis de retornos financeiros advindos de cada uma delas. Assim, além de analisar a influência de diferentes variáveis individuais sobre o salário das diferentes profissões, pretende-se ordenar as distribuições de rendimentos segundo critérios que considerem o risco e o retorno salarial, estabelecendo-se relações de dominância entre os perfis de retornos. Considera-se, para tanto, as distribuições de salários de indivíduos de diferentes carreiras de nível superior, onde a carreira do indivíduo é definida segundo sua formação superior (graduação).

O objetivo geral deste trabalho é, portanto, explorar os diferenciais de risco e retorno entre carreiras de nível superior de duas formas:

¹ Levhari e Weiss (1974) analisa o efeito do risco sobre o investimento em capital humano. Weiss (1972) e King (1974) inserem a questão do risco no problema de escolha ocupacional.

² Ver Christiansen e Nielsen (2002) e Saks e Shore (2003).

³ Para efeito deste trabalho, os termos "profissão" e "carreira" são considerados sinônimos.

1. abordando os diferenciais de retornos médios entre as profissões considerando as diferenças nas características dos indivíduos, como sexo, cor e região, e
2. abordando os diferenciais de risco e retorno entre as profissões considerando todos os indivíduos (sem diferenciá-los segundo suas características individuais), chegando-se a um *ranking* das carreiras segundo a atratividade de suas distribuições de retornos.

Ao aplicar critérios de ordenação entre os perfis de salários das profissões, este trabalho procura auxiliar na mensuração do impacto do risco e retorno salarial sobre a escolha de um ativo educacional realizada pelo investidor em capital humano, ou seja, sobre a decisão sobre em qual curso de graduação ingressar. Assim, partindo-se de hipóteses mínimas sobre a utilidade dos indivíduos quanto aos retornos salariais, aplicam-se critérios de dominância estocástica e estabelece-se uma ordenação ou *ranking* entre os perfis de retorno das diferentes carreiras.

Algumas constatações motivaram este trabalho: i. o estudo dos diferenciais de salários entre carreiras de nível superior ainda é incipiente no Brasil, apesar da relevância do assunto para o planejamento de políticas de ensino superior; ii. as medidas usuais de desigualdade de rendimentos não levam em consideração toda a distribuição, mas somente medidas de estatísticas pontuais, tais como média e desvio-padrão⁴; iii. especificamente, há escassez de trabalhos aplicando testes de dominância estocástica em análise de desigualdades salariais; e iv. finalmente, a existência de uma rica base de microdados - Amostra do Censo Demográfico do IBGE -, que é subtilizada neste tipo de estudo econômico. Contudo, a relevância maior do trabalho revela-se nas possibilidades de aplicação dos resultados e das metodologias aqui expostos em estudos sobre o comportamento de escolha de carreiras e decisões de investimentos em educação e formação profissional.

⁴ Ver aplicações em Levhari e Weiss (1974), King (1974), Christiansen e Nielsen (2002), e comentário de Hause (1974).

Quanto a este último ponto, uma motivação de ordem de política pública deve ser ressaltada. As políticas educacionais brasileiras direcionadas para o ensino superior, como o sistema de cotas e financiamento educacional, não consideram as diferenças entre a atratividade dos perfis de retorno dos vários cursos. Espera-se, portanto, que os resultados deste trabalho proporcionem subsídios para o aperfeiçoamento de políticas educacionais.

O trabalho está dividido em seis seções. Na primeira seção, é realizada uma revisão da literatura relevante para o desenvolvimento da pesquisa. Primeiramente, descreve-se a relação entre investimento em educação e rendimento do trabalho segundo o enfoque da teoria do capital humano, para então, expor-se a abordagem de Mincer como uma extensão da teoria do capital humano. Em seguida, é feita uma breve revisão da literatura que trata dos diferenciais de salários no Brasil. Na subseção seguinte, discute-se a escolha da profissão como um ativo de risco, expondo-se os diferentes modelos propostos e suas limitações. Finalizando a seção, o risco é abordado segundo o enfoque da teoria microeconômica clássica, introduzindo-se, em seguida, a teoria da dominância estocástica.

Na segunda seção, descreve-se a base de dados utilizada e procede-se a uma análise exploratória dos dados da amostra, onde são apresentadas as composições de vinte carreiras (formação superior) e algumas estatísticas amostrais para as distribuições de rendimentos das mesmas. Na seção seguinte, exploram-se algumas medidas de risco e retorno tradicionalmente utilizadas para ordenar diferentes distribuições de retornos. Na quarta seção, explora-se o diferencial de salários entre diferentes grupos demográficos utilizando regressões de salários de Mincer, as quais controlam os resultados pela experiência de trabalho.

Na quinta seção, é realizada uma análise não paramétrica dos dados, estimando-se as densidades teóricas das distribuições salariais das profissões e aplicando-se, em seguida, um teste de normalidade sobre as distribuições. Na sexta e última seção, apresenta-se o teste de Davidson e Duclos (2000) para dominância estocástica, justificando-se sua utilização. Em seguida, são apresentados os resultados da aplicação dos testes de dominância estocástica de primeira e segunda ordens e estabelece-se o *ranking* entre as

profissões segundo os critérios de dominância estocástica. Por fim, compara-se este resultado com os provenientes dos outros critérios.

Finalmente, conclui-se o trabalho relacionando-se os resultados de todas as seções acima descritas. As principais conclusões da pesquisa são:

- A presença de indivíduos não-brancos exercendo carreiras de nível superior é bastante inferior a dos brancos. Já a presença das mulheres é ligeiramente superior a dos homens. Algumas carreiras são compostas principalmente por homens, enquanto outras são tipicamente femininas. A presença de indivíduos não-brancos e de mulheres é maior em carreiras com baixos retornos médios.
- O diferencial de salários por gênero privilegia os homens em todas as profissões. Este diferencial entre homens e mulheres é consideravelmente maior do que o diferencial de salários a favor dos brancos, salvo para poucas carreiras.
- Existe um grande diferencial de salários entre regiões para a maioria das carreiras, em especial entre as regiões Nordeste e Sudeste. A atratividade de uma carreira varia de acordo com a região considerada.
- A dominância estocástica é o critério mais robusto para a ordenação de distribuições de retornos das diferentes carreiras. A aplicação de outros critérios de ordenação de distribuições de retornos ou são inadequados ou podem apresentar resultados arbitrários. Assim, a aplicação de dominância estocástica apresenta *rankings* mais gerais, que espelham as preferências de qualquer indivíduo avesso ao risco, independente do grau de aversão e da forma da função de utilidade.
- A carreira de Medicina domina estocasticamente todas as demais, enquanto Pedagogia é dominada por todas as outras profissões. Assim, todo indivíduo maximizador de utilidade e avesso ao risco prefere Medicina à qualquer outra profissão, e prefere qualquer outra profissão à Pedagogia.

1. ASPECTOS TEÓRICOS E REVISÃO DA LITERATURA

1.1 O INVESTIMENTO EM CAPITAL HUMANO

A teoria do capital humano baseia-se na influência de atributos individuais sobre o lado da oferta de trabalho. Juntamente com as habilidades naturais do indivíduo, o nível educacional, o treinamento profissional e a experiência de trabalho são considerados os determinantes primários da renda salarial. Assim, as habilidades obtidas pelo investimento em capital humano, ou seja, em educação e treinamento, afetam positivamente os salários de um trabalhador. Esta afirmação baseia-se na idéia de que a oferta de trabalhadores com maior habilidade seria necessariamente menor do que a de trabalhadores com menor habilidade.

Muitas vezes a oferta de trabalho requer um alto investimento inicial por parte do trabalhador, que espera recuperá-lo dentro de determinado espaço de tempo. Assim, para modelar decisões sobre a oferta de trabalho deve-se incorporar, além do rendimento corrente e das condições do trabalho, o comportamento do investidor em capital humano. Segundo Becker (1964), um indivíduo escolhe uma profissão que maximize o valor presente dos benefícios financeiros e psicológicos a ser percebido durante toda sua vida. Ehrenberg e Smith (1993) argumentam que a hipótese de que o trabalhador procura maximizar sua utilidade implica que ele esteja interessado nos aspectos pecuniários e não pecuniários do trabalho, sendo as características específicas da ocupação e o modo como as preferências do trabalhador combinam tais especificidades, elementos críticos no processo de escolha da ocupação. Portanto, considerando a existência de diferentes tipos de ocupação e que os indivíduos apresentam diferentes preferências com respeito aos atributos de cada uma, espera-se que a escolha de uma profissão não dependa somente do nível de compensação financeira da ocupação.

Para Becker (1964) o indivíduo racional e bem informado investe na formação de capital humano somente se a taxa de retorno esperada for maior do que a soma da taxa de juros sobre os ativos sem risco e os prêmios de risco e de liquidez associados a esse

investimento. Como explicado pelo autor, o investimento em capital humano exige um prêmio de liquidez positivo, dada a impossibilidade de ser vendido e de servir de garantia para tomada de empréstimos. O verdadeiro retorno sobre este tipo de capital varia em torno de seu valor esperado, reflexo da existência de incerteza sobre vários fatores, como: a duração do tempo de vida ativa, variação nas habilidades entre indivíduos e o pouco conhecimento relativo a este fator, principalmente entre investidores mais jovens, além de inúmeros eventos não previsíveis. O longo período de tempo que é, em geral, exigido para a realização dos retornos sobre o investimento em capital humano limita a qualidade da informação disponível.

Ehrenberg e Smith (1993) explicam que a decisão de investir numa formação profissional, ou, mais especificamente, numa graduação, obriga o candidato a dispensar recursos em adiantado para, no futuro, receber os benefícios inerentes à carreira escolhida. Os benefícios esperados pelo investidor são o fator crítico para esta decisão, sendo que, por serem incertos, não podem ser perfeitamente previstos. Segundo os autores, como uma primeira aproximação, é razoável conjecturar que os retornos médios recebidos pelos graduados recentes exerçam uma importante influência sobre as decisões dos candidatos. No entanto, se por um lado, mudanças nos diferenciais dos ganho esperados entre profissões são um indicador útil das condições gerais do mercado de trabalho, por outro lado, os candidatos devem acessar suas próprias chances de sucesso em campos de atuação específicos. Fatores como relações de amizade, região, grupo étnico, sexo, dentre outros, são de grande importância nas decisões de capital humano, pois afetam o modo como os indivíduos vêem o risco, diante da escolha profissional. Dever-se á, portanto, lidar com a seguinte questão: que informações os candidatos a uma profissão consideram para a formação de suas expectativas quanto aos ganhos futuros?

Segundo os autores, o modo mais simples e ingênuo de prever os níveis futuros de salários é assumindo que o que é observado hoje é o que será observado no futuro. Esta hipótese é a base para o processo de ajustamento para o equilíbrio no mercado de trabalho

para indivíduos com educação superior, conhecido como *cobweb model*⁵. Outro modo de formar previsões sobre ganhos futuros é utilizando a abordagem das expectativas adaptativas, onde os salários esperados para o futuro seriam estabelecidos como uma ponderação entre salários presentes e passados. Já um modelo de expectativas racionais supõe que os agentes observam as mudanças no nível salarial para uma profissão específica como temporária, onde a oferta e a demanda naturalmente se reequilibram até o nível de equilíbrio compatível com as outras profissões. Neste caso, supõe-se que os trabalhadores se comportem como ótimos conhecedores da dinâmica do mercado de trabalho.

Investimentos em capital humano são estudados, tradicionalmente, dentro de uma perspectiva de ciclo de vida. No início do ciclo, os indivíduos alocam tempo para acumulação de capital humano, de modo que quanto maior for o tempo alocado para o investimento, maiores serão os ganhos futuros⁶. Adicionando algumas restrições, Mincer (1974) desenvolveu uma equação de salários visando estimar os retornos sobre o capital humano, onde o log do salário é regredido pelo número de anos de estudo, experiência e experiência ao quadrado, a taxa de retorno sobre a educação sendo dada pelo coeficiente da variável anos de estudo. Desde então, este modelo passou a servir de base para a maioria dos modelos de retornos sobre capital humano futuros. Teorias que se seguiram a de Mincer sugeriram fatores primários adicionais como determinantes dos ganhos salariais, como: sexo, raça e ocupação.

Teorias divergem umas das outras ao tentar explicar a relação entre salário e educação. Como foi visto, a teoria do capital humano afirma que mais educação leva a uma maior produtividade, que, por sua vez, resulta em maiores retornos para o trabalhador. Outra teoria, conhecida como *signalling theory*, parte da idéia de que pessoas com maiores habilidades teriam maior facilidade em obter educação, o que resultaria em maiores ganhos individuais. Tal abordagem trata o investimento em educação como uma ferramenta que

⁵ Segundo este modelo, o ajustamento da demanda por uma formação superior com relação a mudanças nos retornos sobre educação não é tão rápido, especialmente em campos de atuação altamente técnicos, que exigem um maior investimento de tempo e esforço. A hipótese básica para este modelo é que os indivíduos apresentam expectativas míopes sobre o comportamento futuro dos salários, onde se ignoram os efeitos das escolhas profissionais dos outros indivíduos. Ver Ehrenberg e Smith (1993).

⁶ Ver as abordagens de Becker (1964) e Mincer (1974).

indivíduos usam para sinalizar que possuem maiores habilidades. Isto é consistente com a teoria do capital humano, à medida em que parte da hipótese de que quanto menor for o custo psicológico ou intelectual de obter educação, maior será a demanda por esta⁷.

Uma abordagem alternativa a do capital humano, desenvolvida em Lester Thurow (1975), é conhecida como *job competition model*. Tal modelo argumenta que os fatores que explicam os ganhos salariais estão relacionados às características do trabalho e não ao nível de capital humano. Este modelo enfatiza a importância da experiência e do treinamento *on-the-job* para a formação do profissional, onde as firmas estariam dispostas a remunerar melhor os indivíduos mais treinados, de modo que a educação serviria apenas como uma sinalização da capacidade de aprendizado técnico. Esta visão de que os salários estariam relacionados ao trabalho em vez de ser fruto do investimento na formação de capital humano implica que um maior nível educacional muito pouco resultaria em maiores salários. Neste caso, o retorno social sobre a educação seria bem menor do que na abordagem do capital humano⁸.

1.2 REGRESSÃO DE SALÁRIOS DE MINCER

Investimentos em capital humano são realizados na expectativa de que os benefícios futuros superem o montante investido. Ao decidir sobre o investimento em uma unidade adicional de educação, o trabalhador (ou a firma) incorrerá em gastos diretos de educação (ou treinamento) e em custos de oportunidade sobre uma possível renda que deixará de ser obtida. Assim, para que um indivíduo tenha incentivo para investir em educação, ele deve ser compensado com uma renda mais alta o suficiente durante o período futuro de sua vida.

⁷ Segundo este modelo, o ajustamento da demanda por uma formação superior em resposta às mudanças nos retornos sobre educação não se dá tão rapidamente. A hipótese básica deste modelo é que os indivíduos apresentam expectativas míopes sobre o comportamento futuro dos salários, ignorando os efeitos das escolhas dos outros indivíduos sobre o mercado de trabalho.

⁸ Este modelo tem sofrido variações para considerar a teoria do capital humano, resultando no que é conhecido como "job matching theory", que considera que ambos o capital humano e as características do trabalho afetam os ganhos.

Por outro lado, para merecer um maior salário, trabalhadores com maior escolaridade devem ser mais produtivos do que aqueles com menor escolaridade. No equilíbrio competitivo de longo prazo, a relação entre educação e ganhos futuros deve ser tal que a oferta e a demanda por trabalhadores de cada nível educacional esteja em equilíbrio e nenhum trabalhador deseje alterar sua condição educacional.

Becker (1964) e Mincer (1958, 1974) foram pioneiros no estudo dos efeitos dos investimentos em educação e treinamento sobre os padrões de retornos dos trabalhadores durante suas vidas. Em especial, Mincer (1974) procura explicar tais retornos a partir do seguinte modelo empírico que relaciona os log dos retornos, um termo constante, um termo linear nos anos de estudo e termos linear e quadrático nos anos de experiência, ou seja,

$$\ln(w) = \beta_0 + rS + \beta_1 E + \beta_2 E^2 + \varepsilon, \quad (1.1)$$

onde w é o retorno do trabalho de um indivíduo com S anos de estudo e E anos de experiência no mercado de trabalho e ε é o erro.

Esta expressão é conhecida como *regressão de salários de Mincer*. Nesta equação, a variável “anos de estudo” é uma medida direta de capital humano advindo da educação, enquanto que “anos de experiência” é vista como uma *proxy* para o treinamento *on-the-job* (ou qualquer outro investimento após a escola, *post-school investment*). Como, em geral, dados de censos não incluem a experiência de trabalho dos indivíduos, Mincer sugere a seguinte aproximação para a variável, chamada de *experiência potencial*: $E = A - S - 6$, onde A é a idade do trabalhador, S é o número de anos de educação e 6 seria a idade média de ingresso na escola, capturando a interação entre escolaridade e experiência.

A especificação do modelo dado pela equação 1.1 é amplamente utilizada em pesquisas empíricas sobre o mercado de trabalho. Apesar de um conjunto de outros regressores serem usualmente adicionados à equação básica de Mincer, as três variáveis da equação 1.1 continuam aparecendo na maioria das estimativas empíricas de regressões de salários. Estudos sobre discriminação no mercado de trabalho e desigualdade de renda e

abordando capital humano como explicação dos diferenciais de salários e determinação dos salários, têm sido realizados dentro da estrutura da regressão de salários de Mincer. Particularmente, através desta especificação, busca-se uma maneira de obter estimativas da taxa de retorno da educação, informação fundamental para guiar decisões de políticas educacionais.

A equação de Mincer baseia-se em dois modelos formais de investimento em capital humano: *Compensating Differentials Model* (Mincer, 1958) e *Accounting-identity Model* (Mincer, 1974). O modelo *Compensating Differentials Model* assume que os indivíduos apresentam idênticos níveis de habilidade e de oportunidades, mas que as ocupações diferem na quantidade de treinamento requerido. O treinamento gera um custo de oportunidade que os indivíduos devem assumir por estarem distantes do mercado de trabalho durante o período de treinamento, não havendo custos diretos de educação. Desconsidera-se o investimentos *post-school* (incluindo treinamento *on-the-job*). Por serem *ex ante* idênticos, os indivíduos requerem uma compensação em diferencial de salários para trabalhar em ocupações que demandem um maior tempo investido. Este diferencial de salários é calculado pelo valor presente dos fluxos futuros de ganhos da ocupação líquido dos custos associados à mesma, estes não sofrendo variação durante o ciclo de vida do indivíduo.

Em particular, Mincer (1958) assinala que a parte referente ao retorno sobre educação na equação 1.1, $\ln(w) = K_0 + rS$, é uma condição de equilíbrio em um modelo onde agentes idênticos investiriam otimamente em capital humano para maximizar o valor presente de seus ganhos futuros. Neste modelo, o parâmetro r é a taxa de retorno de um ano adicional de educação e é igual à taxa de desconto, no equilíbrio.

O modelo *Accounting-identity Model* parte de hipóteses bastante diferentes, mas chega a uma especificação algebricamente similar a do primeiro modelo. Diferentemente do anterior, este modelo considera a dinâmica dos ganhos durante o ciclo de vida do

indivíduo⁹, relacionando ganhos observados e potenciais com o investimento em capital humano, este considerando tanto investimento em educação, como em investimento *post-school*¹⁰ (como treinamentos na local de trabalho). Mincer (1974) toma os ganhos observados como uma função dos ganhos potenciais líquidos dos custos com investimento em capital humano. Em qualquer período de tempo, os ganhos potenciais variam de acordo com os investimentos realizados em períodos anteriores¹¹. A principal contribuição deste modelo teórico para a formulação da equação de Mincer é a inclusão do investimento *post-school* (que inclui o treinamento *on-the-job*) como uma categoria de investimento em capital humano, aproximada pela variável *experiência potencial*.

Em seu primeiro trabalho, Mincer (1958) assinalou que o perfil idade-retorno era mais inclinado para trabalhadores mais educados, ou seja, log salários não seriam uma função estritamente separável para educação e idade. Assim, os resultados da estimação são diferentes taxas de retornos para cada grupo de idade. Em contraste, Mincer (1974) afirma que os perfis experiência-retorno eram relativamente paralelos para os diferentes grupos de educação. Assim, utilizando experiência potencial no lugar da idade na equação de salários é uma maneira parcimoniosa de capturar a forma e a inclinação do perfil de idade-retornos entre grupos de educação. Uma vantagem deste modelo é que, condicionado pelos anos de experiência potencial, existe uma simples taxa de retorno sobre educação no mercado de trabalho, dada por r .

Segundo Lemieux (2003), a especificação proposta por Mincer se encaixa muito bem aos dados na maior parte dos contextos. Utilizando dados em *cross-section* algumas importantes regularidades empíricas foram capturadas por Mincer (1974), como: salários crescentes com o nível educacional, concavidade do log dos salários com relação a experiência, o log dos salários são paralelos entre as diferentes faixas de educação e a variância interpessoal dos salários apresenta a forma de U.

⁹ Segundo Willis (1986), a equação de salários de Mincer representa um modo pragmático de incorporar algumas das maiores implicações dos modelos dinâmicos de investimento ótimo em capital humano em uma estrutura econométrica bastante simples, que pode ser aplicada a dados com informações limitadas, como em censos e outros do tipo *cross-section*.

¹⁰ Mincer afirma que investimentos em capital humano do tipo *full-time*, que é adquirido nas escolas, precede os investimentos do tipo *part-time*, que são conduzidos geralmente no trabalho.

¹¹ Para demonstração do modelo, ver Heckman (2003).

No entanto, para Willis (1986), a utilização de dados em painel seria ideal para estimar funções estatísticas de salários e taxa de retorno da educação, incluindo a idade de ingresso dos indivíduos no mercado de trabalho e aposentadoria, assim como, informações sobre os custos diretos da educação. Infelizmente, tais tipos de dados são dificilmente encontrados. Como opção, diversos estudos utilizam dados *cross-section* para estimar taxas de retornos. Tais dados contém informações sobre os indivíduos que fazem parte do mercado de trabalho naquele momento, como os salários correntes, idade e escolaridade, mas nenhuma informação sobre custos diretos com educação e idade de ingresso e saída do mercado de trabalho.

Por causa dessas limitações de dados, lança-se mão de um conjunto de hipóteses simplificadoras como forma de permitir estimativas da taxa de retorno da educação com os dados disponíveis. Segundo Lemieux (2003), evidências sugerem que o modelo básico de capital humano de Mincer permanece um modelo preciso e parcimonioso em ambientes estáveis, onde as realizações educacionais crescem suavemente entre os diferentes *cohorts*¹². Num ambiente menos estável, mudanças bruscas no estoque relativo dos diversos grupos com diferentes perfis de idade-educação podem induzir importantes mudanças na estrutura de salários considerada para a estimação da equação padrão de Mincer.

1.3 DIFERENCIAIS SALARIAIS NO BRASIL

Nesta subseção é realizada uma breve revisão da literatura relativa aos determinantes dos diferenciais salariais no mercado de trabalho brasileiro. Tradicionalmente, isto é feito a partir da estimação da equação de salários minceriana, a qual relaciona os salários a seus possíveis determinantes.

Um dos primeiros trabalhos a explorar os diferenciais de salários no Brasil deve-se a Langoni (1973). Este autor analisa o impacto do nível de escolaridade, setor de atividade,

¹² Um *cohort* define o grupo dos indivíduos nascidos em uma determinada época.

idade, sexo e região do trabalhador sobre os salários recebidos, sendo a primeira variável a de maior impacto. Assim, a relação entre a educação de um indivíduo e seu salário passou a ser um assunto amplamente estudado no Brasil.

Considerando apenas trabalhadores do setor formal da economia, de sexo masculino e de áreas urbanas, Senna (1976) estima uma taxa de retorno de aproximadamente 14% para cada ano adicional de estudo. Como variável de controle, o autor considera somente a experiência de trabalho dos indivíduos.

Tainnen (1991) estima uma taxa média de retorno na ordem de 13%, tomando como fonte de dados o Censo de 1980 relativos à indivíduos do sexo masculino. Quando os níveis de escolaridade são desagregados nos diferentes ciclos educacionais, as taxas de retorno para um ano adicional de estudo são de aproximadamente 23%, no ensino superior, e não mais de 16% nos outros ciclos.

Barros e Ramos (1994) avaliam os retornos da educação para os anos 1976 a 1989, considerando, além da educação e da experiência, a idade e a região geográfica como variáveis explicativas. Os autores mostram que enquanto os retornos da educação no primeiro grau apresentam uma queda durante o período estudado, nos níveis educacionais mais elevados (segundo e terceiro graus) o retorno sobre a educação são crescentes neste período.

Considerando o sexo dos trabalhadores, Strauss e Thomas (1996) encontram que os retornos sobre a educação é maior para os indivíduos do sexo masculino. Com relação à cor (ou raça) do trabalhador, Silva (1980) estima que a taxa de retorno sobre a educação para os brancos é significativamente maior do que para os não-brancos.

Quanto ao efeito da experiência sobre os salários, Senna (1976) encontra um padrão de retornos salariais decrescentes com relação à experiência. Este resultado é consistente com o encontrado por Mincer (1974), em seu estudo para os Estados Unidos.

Baseando-se nas PNADS dos anos 1981 a 1989, Barros, Ramos e Santos (1995) identificam a discriminação por gênero como o principal componente dos diferenciais de salários no Brasil. Kassouf (1998) aponta que esta discriminação é maior no setor informal da economia. Hoffmann (2001) encontra um menor diferencial de salários por sexo na agricultura, se comparada aos setores industrial e de serviços.

Utilizando dados para 1989, Cavalieri e Fernandes (1998) concluem que o diferencial salarial por sexo é maior entre os trabalhadores não-brancos e entre os mais educados, e são menores nas regiões mais desenvolvidas do país (Sudeste e Sul). Analisando o comportamento do diferencial salarial por cor/raça de acordo com outras características dos indivíduos (nível de educação, idade, sexo e região), os autores encontram que este diferencial pouco varia com a idade e com a região, mas a diferença a favor dos trabalhadores brancos cresce sensivelmente com o nível educacional e afeta mais as mulheres.

Considerando o período de 1976 a 1986, Savedoff (1991) identifica a existência de diferenciais de salários entre as regiões metropolitanas brasileiras. Segundo o autor, este diferencial varia segundo a ocupação dos trabalhadores e o ano considerado. Controlando pelos diferenciais de custo de vida, Azzoni e Servo (2001) confirmam este resultado para os anos de 1992, 1995 e 1997. Esses autores identificaram as regiões metropolitanas de Fortaleza e de Recife como apresentando os menores salários, enquanto as regiões metropolitanas de São Paulo e Brasília apresentaram os salários mais elevados.

Macedo (1985) e Foguel et al (2000) estudam os diferenciais de salários segundo a natureza jurídica do empregador. O primeiro autor, utilizando os dados da RAIS para o ano de 1981, encontra um grande diferencial favorecendo os trabalhadores de empresas estatais. Foguel et alii (2000), utilizando dados da PNAD para o ano de 1995, confirmam que o diferencial de salários entre trabalhadores do setor público e trabalhadores do setor privado, controlado pelo sexo, cor/raça, idade, educação e experiência, é significativamente alto em

favor dos primeiros. Contudo, este diferencial controlado varia bastante entre as diferentes regiões metropolitanas¹³.

Coelho e Corseuil (2002) realizam um levantamento da literatura acerca dos diferenciais de salários no Brasil. Alguns dos principais resultados encontrados nos diferentes trabalhos considerados foram:

- O efeito da educação sobre o salário é maior nos níveis mais altos de educação;
- A discriminação por cor e sexo aumentam com o nível educacional, sendo que esta última vem diminuindo;
- O diferencial salarial tende a favorecer trabalhadores atuando em setores de atividades que envolvam maior desenvolvimento tecnológico;
- As regiões mais desenvolvidas tendem a apresentar salários mais altos;
- Salvo algumas exceções, o setor público paga salários mais altos que o setor privado.

1.4 ESCOLHA DA PROFISSÃO COMO UM ATIVO DE RISCO

Desde o trabalho pioneiro de Mincer (1958), muitos estudos foram publicados abordando a relação entre educação, retornos econômicos e o processo de escolha educacional do indivíduo. Tal literatura trata desta escolha como uma decisão de investimento em capital humano.

Mincer (1974) aborda o problema da escolha profissional de um indivíduo supondo que este considera o valor presente esperado dos retornos de cada opção. Em seu trabalho, Mincer utiliza a variância dos resíduos de uma estimação de mínimos quadrados ordinários como medida de risco, controlando para as características dos indivíduos.

¹³ Enquanto a Região Metropolitana de São Paulo apresenta o maior diferencial a favor do setor privado, a região onde o setor público apresenta melhor resultado é a do Distrito Federal.

Weiss (1972) utiliza o coeficiente de variação dos salários para estimar o risco do investimento em educação sem, no entanto, controlar para características individuais. Para modelar o problema de escolha entre ocupações, o autor considera formas específicas de utilidade que são função do salário médio e desvio-padrão (e, obviamente, do coeficiente de variação), de modo que o impacto do risco sobre a decisão individual pode ser completamente explicado por tais medidas de retorno e risco.

No entanto, como observa Hause (1974), os resultados estimados podem variar consideravelmente dependendo da forma específica da função de utilidade aplicada. Além do mais, como o próprio Weiss (1972) explica, a utilização do coeficiente de variação como medida de risco só é válida se o grau de aversão ao risco for suposto moderado. Apesar das limitações expostas, trabalhos envolvendo investimento educacional passaram a considerar, em sua maioria, a média e o desvio-padrão dos retornos como fatores determinantes para tomada de decisão.

King (1974) estimou uma relação positiva entre o retorno médio e duas diferentes medidas de risco: o desvio-padrão e o coeficiente de assimetria da distribuição de retornos, o que indica a existência de aversão ao risco, ou seja, indivíduos necessitam de uma compensação em retornos médios em troca do risco assumido. Contudo, para Evans e Weinstein (1982), uma correlação positiva entre o retorno médio e o desvio padrão dos retornos entre ocupações não implica que a compensação pelo risco seja suficiente para igualar a atratividade entre todas as ocupações.

Christiansen e Nielsen (2002), investigam o investimento sobre o capital humano, relacionando-o com a abordagem de economia financeira. Partindo do *trade-off* entre média e variância em investimentos em capital humano, os autores calculam uma medida de performance que ordena tipos de educação, estabelecendo um guia para investimentos individuais. Para os autores, assim como os mercados para bens, trabalho e financeiros, o mercado educacional é caracterizado por uma entre muitas possibilidades de investimento, neste caso, em diferentes tipos e graus de educação. Cada indivíduo jovem escolhe a opção educacional que casa exatamente com sua combinação preferida de risco e retorno, em

termos de renda futura¹⁴. Neste caso, o tipo de educação seria uma escolha tão importante quanto o nível educacional.

Segundo Chen (2001), o risco envolvido numa decisão de investimento em uma graduação pode resultar da falta de conhecimento sobre aspectos como: (1) a habilidade individual, (2) a qualidade da instituição de ensino e (3) mudanças não-antecipadas nas condições do mercado de trabalho. As duas primeiras fontes de risco causam um "choque permanente" sobre os retornos futuros. A terceira fonte de risco causa um "choque transitório". O autor observa que o nível de desemprego é também uma importante fonte de risco para a decisão profissional, o que é deixado de lado pela a maior parte dos trabalhos.

Saks e Shore (2003), definem o risco da renda do trabalho como o componente involuntário da mudança no salário individual, de modo que as carreiras mais arriscadas apresentam uma maior variância entre os choques inesperados de renda. Segundo os autores, embora leve a uma dispersão entre os salários dos indivíduos numa carreira específica, é impossível afirmar se uma dispersão no *cross-section* dos salários reflete a presença de risco ou se é resultado da presença de diferenças de níveis de habilidade entre indivíduos. Assim, utilizando dados em painel, eles estimam a variância dos choques sobre a renda individual a partir das mudanças observadas no tempo.

Os trabalhos de Freeman (1976) e Siow (1984), utilizando séries temporais dos retornos agregados de uma determinada profissão, limitaram-se a estimar os retornos esperados, sem considerar a variabilidade desses retornos e, muito menos, os riscos inerentes às diferentes profissões. Tais modelos reconhecem a influência da oferta e demanda por trabalho na decisão de alocação profissional. Para Freeman (1976), as condições de demanda por trabalho afetam a oferta de novos candidatos a graduação, de modo que, ao decidir pelo ingresso num determinado campo profissional, os indivíduos comparam os salários presentes de cada profissão. Tendo em vista a possibilidade de mudança nas condições futuras de trabalho, os salários presentes podem ser fonte de erros

¹⁴ É importante citar outros fatores também relevantes para a escolha profissional, como: o *status* da profissão, a influência da família, religião, dentre outras variáveis sócio-demográficas.

de previsão, as quais levam a ciclos na oferta de novos estudantes de graduação. Como foi abordado anteriormente, modelos que consideram este comportamento cíclico da oferta são definidos *cobweb models*. Baseando-se nesta idéia, Freeman estimou modelos onde, no processo de escolha profissional, os indivíduos se baseavam nos salários iniciais para preverem as oportunidades de ganhos futuros das diferentes profissões.

Siow (1984) argumenta que, tendo em vista a incerteza dos salários futuros, o agente deve prevê-los de modo escolher uma profissão que proporcione o maior retorno esperado descontado. Segundo o autor, esta previsão seria racional, estando o indivíduo bem informado da estrutura geral dos mercados e dos efeitos da entrada de outros indivíduos nesses mercados. Assim, o autor vai além no modelo de Freeman pela imposição de que as expectativas futuras são previsões racionais.

Por outro lado, segundo Behrman et al (1998), os modelos agregados de oferta e demanda, em geral, não capturam as diferenças nas características dos participantes do mercado de trabalho e, portanto, não consideram as diferentes respostas às mudanças das condições externas. Os autores argumentam que modelos microeconômicos de escolha de carreiras devem considerar as diferenças entre habilidades, atitudes e circunstâncias, de modo testar hipóteses sobre a validade de tais diferenças para o processo de escolha.

Criticando a utilização de medidas pontuais de retorno e dispersão para a ordenação entre profissões quanto ao risco e como forma de evitar a especificação de uma mesma função de utilidade para todos os indivíduos, Evans e Weinstein (1982) propõem o uso da dominância estocástica. Segundo os autores, uma medida baseada na média e no desvio-padrão é muito simples para coincidir com a aversão ao risco da maior parte das pessoas. Como um exemplo, os autores afirmam que indivíduos bastante avessos ao risco preferem um índice que atribua maior peso ao risco, ao ordenar opções de investimento. Assim, uma ordenação das profissões deve reconhecer a diversidade dos níveis de aversão ao risco em vez de assumir que todos os trabalhadores apresentam funções de utilidade idênticas.

1.5 O RISCO NA TEORIA MICROECONÔMICA

O risco é um fenômeno inerente à vida econômica. A maior parte das decisões econômicas envolvem um elemento de risco, de modo que o assunto tem influenciado cada vez mais pesquisas em economia. A incorporação formal do risco e incerteza na teoria econômica deu-se a partir do trabalho de John von Neumann e Oscar Morgenstern, intitulado "Theory of Games and Economic Behavior", publicado em 1944. A grande contribuição desses autores foi a de estabelecer um fundamento racional para o processo de tomada de decisão sob condições de risco partindo de um tratamento axiomático das regras de utilidade esperada.

O conjunto de resultados advindos a partir dessa abordagem é denominado de Teoria da Utilidade Esperada, tendo como remotas origens o trabalho de Daniel Bernoulli (1738), que desenvolveu o conceito de utilidade esperada como uma solução para o conhecido *St. Petersburg Paradox*. Bernoulli sugeriu que um indivíduo, quando confrontado com uma aposta, baseava sua decisão na utilidade esperada de um jogo, e não no valor esperado deste¹⁵.

Dada sua adequação do contexto deste trabalho, utilizar-se-á a abordagem de von-Neumann and Morgenstern para o processo de escolha sob condições de risco. Esta abordagem considera um esquema onde as alternativas incertas de retorno são descritas por probabilidades observáveis, definidas sobre um conjunto de retornos possíveis. Tais representações de alternativas envolvendo risco são chamadas de 'loterias'. Em seu trabalho, von-Neumann and Morgenstern desenvolveram um conjunto de axiomas abordando a ordenabilidade, continuidade e independência das escolhas dos tomadores de decisão. Baseando-se nesses axiomas, os autores puderam derivar as propriedades da função de

¹⁵ Uma abordagem alternativa, cuja axiomatização foi proposta no trabalho de Savage, em 'The Foundations of Statistics' (1954), é conhecida como 'Teoria da Probabilidade Subjetiva'. Aqui, a hipótese da utilidade esperada é derivada sem a imposição de que as probabilidades das alternativas de risco sejam dadas objetivamente ao tomador de decisão. Em vez disso, permite-se que as probabilidades sejam determinadas de maneira subjetiva, as escolhas sendo baseadas nas crenças dos tomadores de decisão, as quais são reveladas pelo comportamento de escolha individual. Savage faz isso a partir de um conjunto de sete postulados necessários para que uma relação de preferências possa ser derivada das expectativas dos indivíduos.

utilidade esperada, para então, descrever as condições sob às quais a ordenação de preferências de um indivíduo confrontado com escolhas aleatórias corresponde à maximização de utilidade esperada. Assim, os autores provam que, sob algumas condições, pode-se representar as preferências individuais na conveniente forma de utilidade esperada.

Quando as alternativas de risco apresentam como possibilidades de retorno quantidades monetárias, a teoria da escolha sob incerteza pode ser estendida para o caso de um domínio infinito, que trata quantias monetárias como uma variável contínua. Tal abordagem deve vir acompanhada de uma particular estrutura de utilidade esperada. Pode-se descrever uma loteria monetária como uma função distribuição acumulada (cdf), $F: \mathbb{R} \rightarrow [0,1]$, significando que, para qualquer retorno monetário, x , a função $F(x)$ será a probabilidade de que o retorno realizado seja menor ou igual a x . Se a função distribuição de uma loteria possuir uma função densidade associada, $f(x)$, então pode-se relacionar as duas da seguinte maneira:

$$F(x) = \int_{-\infty}^x f(t)dt, \quad \forall x$$

A representação de uma loteria por uma função distribuição apresenta a vantagem de incluir também a possibilidade de análise de um conjunto discreto de possibilidades de retorno. Além disso, as funções distribuição preservam a estrutura linear das loterias, assim como o fazem as funções densidade. Dessa forma, pode-se considerar o espaço de loterias monetárias, L , como sendo o conjunto de todas as funções distribuição definidas sobre um intervalo $[a, \infty)$.

A aplicação do teorema da utilidade esperada neste contexto permite relacionar diferentes valores de utilidade, $u(x)$, às quantias não-negativas de moeda, com a importante propriedade de que qualquer função distribuição $F(\cdot)$ pode ser avaliada por uma função utilidade da forma $U(F) = \int u(x)dF(x)$, que corresponde a função utilidade esperada von Neumann-Morgenstern vista anteriormente. Note-se que esta função, é linear em $F(x)$ e é exatamente a esperança matemática dos valores, em utilidade $u(x)$, das diferentes realizações monetárias, x . Através dessa representação a utilidade de loterias monetárias se

torna sensível não somente à média, mas também, aos momentos de ordem superiores da distribuição de retornos monetários.

Para evitar confusão entre definições, utilizar-se a nomenclatura proposta por Mas-Colell et al (1995), onde a função de utilidade definida sobre loterias, $U(\cdot)$, é chamada de "função utilidade esperada von Neumann-Morgenstern", enquanto a função de utilidade definida sobre as quantias de moeda, $u(\cdot)$, é chamada de "função de utilidade de Bernoulli".

Note-se que os axiomas gerais que asseguram a representação das preferências na forma de utilidade esperada não restringem de nenhum modo a forma da função de utilidade de Bernoulli, $u(\cdot)$. Em grande parte, o poder analítico da formulação da utilidade esperada advém da possibilidade de se especificar esta função de utilidade, de modo a capturar importantes atributos econômicos relacionados ao comportamento de escolha. Assim, no contexto de loterias monetárias, faz sentido especificar a função de utilidade de Bernoulli, $u(\cdot)$, como sendo crescente e contínua, hipótese a ser mantida daqui em diante.

Após a axiomatização da hipótese da utilidade esperada por von Neumann e Morgenstern, os economistas passaram a procurar potenciais aplicações para a teoria da utilidade esperada, partindo-se de modelos simples com o retorno consistindo de somente um 'bem', renda ou riqueza. Como resultado, uma loteria passa a ser considerada como uma variável aleatória, z , tomando valores na reta real. Conseqüentemente, preferências sobre loterias passa a ser pensadas como preferências sobre distribuições de probabilidade alternativas, de modo que, denotando por F_z a distribuição de probabilidade acumulada associada à variável aleatória z , onde $F_z(x) = \Pr\{z \leq x\}$, então pode-se pensar em agentes fazendo escolhas sobre diferentes F_z .

Desse modo, as preferências sobre loterias são agora definidas sobre o espaço de funções de distribuição acumuladas. Fazendo a função de utilidade v.N-M, $U(\cdot)$, representar preferências sobre distribuições, tem-se que a loteria F_z é preferida à outra F_y ,

denotado por $F_z \succsim F_y$, se somente se $U(F_z) \geq U(F_y)$. Conseqüentemente, a decomposição da utilidade esperada de $U(F_z)$ passa a ser dada por

$$U(F_z) = \int_{\square} u(x) dF_z(x)$$

onde $u: \square \rightarrow \square$ é a função de utilidade de Bernoulli sobre os diferentes níveis de retorno, supondo que z segue uma distribuição contínua.

Deve-se a Arrow (1971) e Pratt (1964) a introdução de medidas de aversão ao risco no estudo do comportamento de escolha individual sob risco. Definindo U' e U'' como as derivadas primeira e segunda da função de utilidade *v.N-M*, respectivamente, e W , como a riqueza, os autores definiram as seguintes medidas de aversão ao risco:

- $R_A = -\frac{U''}{U'}$, é a medida de aversão absoluta ao risco;
- $R_R = -\frac{U''}{U'} W$, é a medida de aversão relativa ao risco.

Sob a hipótese de utilidade esperada, existe uma relação bijetora entre as preferências sobre retornos aleatórios e as medidas de aversão ao risco. À medida que a renda aumenta, um indivíduo passa a dar menor importância a uma unidade adicional de risco, então a medida de aversão absoluta ao risco é decrescente ($U''' \geq 0$).

1.6 TEORIA DA DOMINÂNCIA ESTOCÁSTICA

O passo seguinte para o desenvolvimento de uma teoria da decisão sob condições de risco foi o desenvolvimento de modelos para mensuração do risco. Os primeiros esforços nesta direção utilizaram medidas estatísticas como média e variância de retornos aleatórios para representar as preferências dos tomadores de decisão, de modo a se ter a utilidade como função dos dois primeiros momentos da distribuição de rendimentos.

Reconhecendo as limitações do uso da variância como medida de risco e buscando desenvolver modelos úteis para uma comparação mais geral entre alternativas de risco, foram publicados os trabalhos de Hadar e Russell (1969), Hanoch e Levy (1969) e Rothschild e Stiglitz (1970), dando origem à teoria da dominância estocástica e sua introdução na análise econômica. Esses autores apresentaram uma abordagem para o problema de mensuração do risco que independe da função de utilidade do tomador de decisão, fazendo uso de distribuições de probabilidade.

A Dominância Estocástica é uma área de estudo que vem se expandindo, com inúmeras aplicações em economia, finanças, estatística e pesquisa operacional. Levy (1992) cita as principais aplicações dos critérios de dominância estocástica, que foram primeiramente aplicados a finanças, como em diversificação de portfólios e políticas financeiras. Hoje, tais critérios são também utilizados em medição de desigualdade de renda, avaliação de opções de investimentos e em temas que envolvam escolha ótima de fatores de produção, como em vários estudos de economia agrícola.

Os critérios de Dominância Estocástica apresentam grandes vantagens sobre o popular critério de média-variância ou *M-V analysis*, à medida em que permite a avaliação de ordenações relativas de alternativas incertas para diferentes distribuições de retorno sem, no entanto, especificar uma particular função de utilidade do tomador de decisão. De acordo com o modelo de média-variância¹⁶ a decisão de investimento de um agente se baseia somente no retorno esperado e da variância do ativo em questão. Neste caso, o investimento preferido seria aquele com maior retorno esperado e menor variância, esta assumindo o papel de medida de risco¹⁷.

O perfil de preferências sobre risco e retorno só é compatível com funções de utilidade quadráticas, onde o valor médio e a variância dos salários apresentam efeitos positivo e negativo, respectivamente, sobre a utilidade do indivíduo. Segundo Christiansen e Nielsen (2002), o principal argumento contra a hipótese da função de utilidade quadrática

¹⁶ Proposto por Markowitz (1952).

¹⁷ Para críticas ao modelo, ver Rothschild e Stiglitz (1970).

é o fato dela apresentar aversão relativa ao risco crescente com relação a renda. Além do mais, sob esta hipótese, a função de utilidade esperada não captura momentos de ordem superior, como assimetria e curtose.

Segundo Hadar e Russell (1969), devido a utilidade esperada ser, em geral, uma função de todos os momentos da distribuição de probabilidade, regras de comparação que envolvam somente dois momentos são válidas para classes muito limitadas de funções de utilidade. Esta limitação impõe muitas restrições sobre a forma funcional da utilidade, o que gera severas perdas de generalidade. Assim, segundo o autor, a especificação de distribuições em termos de seus dois primeiros momentos provavelmente não asseguraria resultados fortes, pois informações sobre os momentos da distribuição não poderiam ser utilizadas eficientemente para a ordenação de alternativas incertas quando a função de utilidade não é conhecida.

Rothschild e Stiglitz (1970) afirmam que, apesar de proporcionar uma ordenação completa de distribuições, o critério de média-variância apresenta respostas espúrias, pela arbitrariedade das restrições impostas à forma da função de utilidade. Em seu trabalho, esses autores mostram que, caso a função de utilidade não assuma a forma quadrática, a ordenação pelo critério de média-variância pode apresentar um resultado diferente da ordenação por meio da utilidade esperada, levando a ordenações injustificadas. Diferentemente do critério de média-variância, a Dominância Estocástica utiliza-se de toda a distribuição de retornos de cada alternativa de risco, em vez de limitar-se à análise dos primeiros momentos da distribuição.

Como hipótese básica para a aplicação do critério de Dominância Estocástica, assume-se que o tomador de decisão seja maximizador de utilidade esperada e apresente função de utilidade da forma $v.N-M$. A partir disso, Hadar e Russell (1969) demonstram que a ordenação baseada na Dominância Estocástica é condição necessária e suficiente para a maximização da utilidade esperada. Assim, uma decisão baseada no critério de Dominância Estocástica assegura que a utilidade esperada do tomador de decisão seja maximizada. Se a distribuição de uma alternativa domina estocasticamente a de outra

alternativa então a primeira alternativa proporciona uma maior utilidade esperada para o indivíduo.

Entretanto, esta abordagem deve cercar-se de hipóteses complementares as mais fracas possíveis a respeito da forma da função de utilidade, garantindo uma maior adequação às diferentes formas de funções de utilidade individuais. Partindo das hipóteses mais fracas para hipóteses cada vez mais restritivas, Levy (1992) cita os três critérios mais utilizados de dominância estocástica: Dominância Estocástica de Primeira, Segunda e Terceira Ordens.

A Dominância Estocástica de Primeira Ordem (FSD) assume que a função de utilidade do tomador de decisão pertence a classe de funções U_1 , que inclui toda função u , com $u' \geq 0$, ou seja, u é uma função monotonicamente crescente, o que implica que o indivíduo prefere um maior retorno a um menor. O critério de Dominância Estocástica de Segunda Ordem (SSD) impõe uma restrição adicional, de modo que a função de utilidade individual passa a pertencer a classe de funções U_2 , com $u' \geq 0$ e $u'' \leq 0$, que implica na aversão individual ao risco. Finalmente, a Dominância Estocástica de Terceira Ordem (TSD), impõe uma classe de funções U_3 , com $u' \geq 0$, $u'' \leq 0$ e $u''' \geq 0$, o que implica em um indivíduo avesso ao risco, com aversão absoluta ao risco decrescente¹⁸.

O critério de FSD baseia-se no fato de que se uma distribuição acumulada F apresenta-se sempre à direita de uma outra G , para qualquer valor de x , então a probabilidade de se obter um valor maior ou igual a um dado x^* é maior sob a distribuição F do que sob a distribuição G . No entanto, uma limitação à aplicação do critério de FSD para comparação entre duas alternativas de risco se dá quando as distribuições acumuladas das duas alternativas se cruzam um número finito de vezes. Tal critério exige que uma distribuição supere sempre a outra, ou seja, para cada nível de probabilidade, obtenha-se sempre um melhor resultado esperado sob a distribuição dominante. Segundo Hadar e Russell (1969), esta é a condição mais fraca para garantir a preferência de uma alternativa

¹⁸ Este trabalho não fará uso do conceito de dominância de terceira ordem.

sobre outra, apesar de restringir a análise entre duas alternativas a indivíduos maximizadores de utilidade que preferem um maior retorno a um menor.

Entretanto, Hadar e Russell (1969) mostram que o conjunto de alternativas ordenáveis pode ser aumentado à medida que mais restrições sejam impostas sobre a função de utilidade individual. Neste caso, poder-se utilizar o critério de dominância estocástica de segunda ordem, SSD, para uma comparação entre duas distribuições que se cruzam, já que este critério envolve as áreas sob as distribuições acumuladas, em vez de confrontá-las ponto à ponto. O critério de SSD, ao admitir tomadores de decisão avessos ao risco, insere a questão da incerteza relacionada aos retornos de diferentes alternativas. O critério de SSD é baseado em duas hipóteses adicionais: a de que a derivada segunda da função utilidade é negativa (aversão ao risco) e a de que a integral da distribuição dominante, F , apresenta valores menores ou iguais à integral da distribuição dominada, G , para qualquer valor x , com desigualdade estrita para algum x . Combinando essas duas hipóteses com a hipótese inicial de utilidade marginal positiva, pode-se testar se a distribuição F domina estocasticamente em segunda ordem G , para todo x .

Os possíveis resultados dos testes de Dominância Estocástica serão: $F \succ_{sdi} G$, $G \succ_{sdi} F$ ou nenhum dos dois, onde o subscrito i refere-se a ordem da dominância. Neste último caso, as alternativas são classificadas como não-comparáveis, ou seja, nada pode ser dito sobre a ordenação entre as duas alternativas. A possibilidade de não-ordenação é devida a falta de restrições envolvendo as derivadas parciais de maiores ordens das funções de utilidade. Assim, o fato de duas alternativas não serem comparáveis no sentido da dominância estocástica revela que o tomador de decisão deve refinar mais ainda suas prioridades antes de tomar uma decisão racional, isto é, deve-se conhecer mais sobre sua função de utilidade. Conhecendo-se mais sobre o comportamento do tomador de decisão com relação ao risco, pode-se adicionar uma ou mais restrições sobre sua função de utilidade, limitando-se a classe de funções e permitindo-se a utilização de critérios de ordenação de ordens superiores. Por exemplo, ao usar o critério de SSD para comparar duas alternativas com distribuições acumuladas F e G , se o resultado do testes acusar a não-comparabilidade, pode-se impor uma restrição adicional sobre a derivada terceira da função

de utilidade do tomador de decisão, $u'' \geq 0$, de modo poder-se aplicar o critério de TSD, que não será abordado neste trabalho¹⁹.

Partindo das restrições necessárias relacionadas à forma funcional da utilidade do tomador de decisão, Levy (1992), resume os critérios de Dominância Estocástica no seguinte teorema:

Teorema 1.1 *Sejam F e G as distribuições acumuladas de duas alternativas incertas. F domina G estocasticamente em primeira ordem (FSD) e em segunda ordem (SSD), respectivamente, se e somente se,*

$$F(X) \leq G(X), \text{ para todo } X \text{ (FSD)}$$

$$\int_{-\infty}^x [G(t) - F(t)] dt \geq 0, \text{ para todo } X \text{ (SSD),}$$

onde ao menos uma desigualdade estrita deve ocorrer em todos os casos²⁰.

Levy (1992) prossegue relacionando as regras de Dominância Estocástica com as classes relevantes de preferências, representadas por U_i , da seguinte maneira:

$$\text{FSD: } F(X) \leq G(X) \forall X \Leftrightarrow E_F U(X) \geq E_G U(X) \forall u \in U_1$$

$$\text{SSD: } \int_{-\infty}^x F(t) dt \leq \int_{-\infty}^x G(t) dt \forall X \Leftrightarrow E_F U(X) \geq E_G U(X) \forall u \in U_2$$

onde U_1 e U_2 foram definidos anteriormente.

Graficamente, pode-se observar na Figura 1.1 que a dominância estocástica de primeira ordem requer que a distribuição de probabilidade acumulada $F(\cdot)$ esteja sempre à direita ou toque a distribuição de probabilidade acumulada $G(\cdot)$. Neste caso, a probabilidade de se alcançar uma renda até um particular valor, x^* , sob a distribuição $G(\cdot)$,

¹⁹ Ver Whitmore(1970).

²⁰ A demonstração da FSD e SSD pode ser vista em Hadar e Russell (1969).

é maior ou igual à probabilidade de alcançar até este mesmo valor sob a distribuição $F(\cdot)$. Raciocinando em direção oposta, a probabilidade de se obter uma renda que exceda x^* sob a distribuição $G(\cdot)$ é menor do que sob a distribuição $F(\cdot)$.

Quando as distribuições se cruzam não será possível encontrar dominância de primeira ordem, dado que há alguns níveis de renda onde a probabilidade de se obter valores maiores do que um nível de renda x^{**} será maior sob a distribuição $G(\cdot)$ do que sob a distribuição $F(\cdot)$.

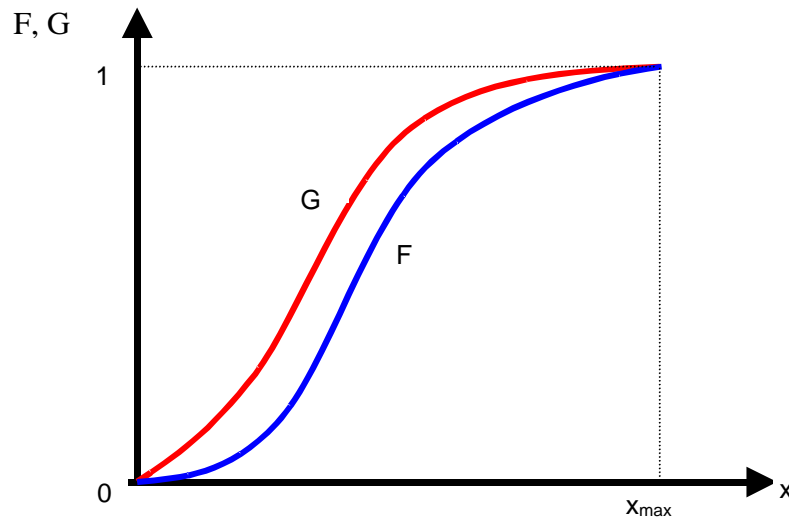


Gráfico 1.1: Dominância Estocástica de Primeira Ordem (FSD). **Fonte:** Elaborado pelo autor.

A dominância estocástica de segunda ordem possibilita que as distribuições acumuladas cruzem. O que este critério faz é assumir aversão ao risco e permitir que a utilidade marginal para níveis menores de renda sobreponha a utilidade dos incrementos de renda adicional para maiores níveis de renda. Neste caso, passa a ser importante observar se a área entre as distribuições acumuladas $F(\cdot)$ e $G(\cdot)$ se mantém positiva em todo o suporte, ou seja, se a distribuição $F(\cdot)$ mantém aquela vantagem advinda dos baixos valores de

renda x . Pode-se visualizar isto na Figura 1.2, onde F domina G somente se a área S_A for maior do que a área S_B .

Quando da aplicação de testes para dominância estocástica, mesmo partindo-se de tomadores de decisão avessos ao risco, não é necessário assumir, a priori, hipóteses adicionais como aversão absoluta ao risco decrescente ou qualquer outra classificação mais restritiva de aversão ao risco e, muito menos, uma forma particular de função de utilidade. Em vez disso, pode-se partir dos critérios mais gerais de FSD e SSD e, somente em casos de não-comparabilidade, empregar critérios que envolvam maiores restrições.

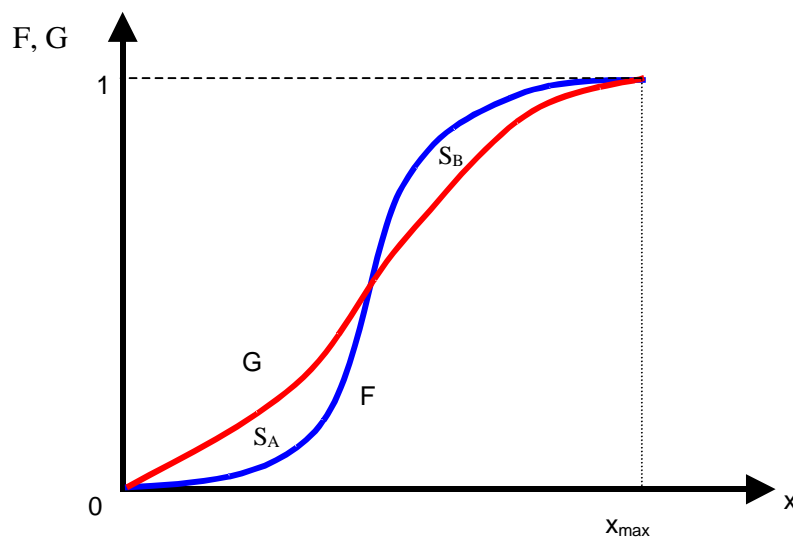


Gráfico 1.2: Dominância Estocástica de Segunda Ordem (SSD). **Fonte:** Elaborado pelo autor.

2. BASE DE DADOS E DESCRIÇÃO DA AMOSTRA

Para a realização deste trabalho foi necessário encontrar uma base de dados que incluísse informações sobre a formação superior dos indivíduos. Após analisar as principais bases de dados do Brasil, foi constatado que a base de dados que fornece a informação mais adequada é a Amostra do Censo Demográfico 2000, do IBGE. Dentre as bases de dados analisadas encontram-se a PNAD e a RAIS-CAGED, as quais apresentam somente um indicativo do tipo de graduação do indivíduo: a ocupação deste. Diferentemente dessas bases, a Amostra do Censo Demográfico do IBGE contém, dentre outras informações fundamentais para o estudo, o tipo de graduação do indivíduo, além de apresentar um grande número de observações, o que contribui para a robustez dos resultados estimados.

2.1 O CENSO DEMOGRÁFICO DO IBGE

A primeira contagem da população do Brasil foi realizada no ano de 1872, mas, a partir de 1890, o Censo Demográfico se tornou decenal. O Brasil foi o primeiro país a incluir questões sobre fecundidade no Censo Demográfico e um dos poucos da América Latina a pesquisar rendimento²¹.

2.1.1 O CENSO DE 2000

A coleta do Censo Demográfico 2000 foi realizada no período de 1º de agosto a 30 de novembro de 2000, abrangendo todo o Território Nacional, com exceção das representações do Brasil no exterior, totalizando 215.811 setores censitários, que constituíram as menores unidades territoriais da base operacional do censo²². A operação censitária mobilizou mais de 200 mil pessoas, em pesquisa a 54.265.618 domicílios nos 5.507 municípios existentes no ano 2000, das 27 Unidades da Federação. A investigação

²¹ Esta seção foi baseada na Documentação dos Microdados da Amostra, publicada em Novembro de 2000, pelo Instituto Brasileiro de Geografia e Estatística - IBGE. A documentação inclui as notas metodológicas e a descrição dos respectivos conteúdos.

²² Setor censitário é a unidade territorial criada para fins de controle cadastral da coleta. Para este censo, o Território Nacional foi dividido em 215.811 áreas contíguas, respeitando-se os limites da divisão político-administrativa, do quadro urbano e rural legal e de outras estruturas territoriais de interesse, além dos parâmetros de dimensão mais adequados à operação de coleta

dos domicílios e das pessoas neles residentes teve como data de referência o dia 1º de agosto de 2000.

Os arquivos de microdados são parte integrante do processo de disseminação dos resultados da Amostra do Censo Demográfico 2000. Os dados podem ser recuperados até o nível de distrito, subdistrito ou área de ponderação, desde que satisfeita a condição de possuírem na amostra pelo menos 400 domicílios particulares ocupados.

As variáveis da Amostra do Censo Demográfico 2000 foram classificadas e descritas nos seguintes grupos:

- Variáveis Geográficas: incluem a Unidade da Federação, mesorregião, microrregião, município, Região Metropolitana e região geográfica;
- Variáveis de Domicílios: incluem as características apresentadas pelos domicílios na data de referência do Censo, visando conhecer as condições de moradia, os níveis de qualidade de vida da população;
- Variáveis de Pessoas: investigam as características das pessoas residentes no Território Nacional na data de referência do Censo, como: sexo, idade, condição na família e no domicílio, deficiência física ou mental, capacidade de caminhar, de enxergar e de ouvir, raça ou cor, e religião ou culto professado;
- Migração: investigam a mobilidade espacial da população, identificando sua nacionalidade, naturalidade, migrações intramunicipais, intermunicipais e internacionais, local de trabalho ou estudo;
- Nupcialidade: investigam a condição dos indivíduos em relação ao fato de viverem em companhia de cônjuge, em decorrência de casamento civil, religioso, civil e religioso ou de união consensual estável;
- Fecundidade: investiga a fecundidade das mulheres de 10 anos ou mais, através de quesitos referentes ao número de filhos(as) tidos(as) até 31 de julho de 2000, dos filhos (as) que teve, quantos estavam vivos até essa mesma data, qual a data de nascimento do(a) último(a) filho(a) nascido(a) vivo(a) até 31 de julho de 2000, se

este(a) filho(a) estava vivo(a) na mesma data e quantos filhos(as) nascidos(as) mortos(as) teve até 31 de julho de 2000;

- Instrução: investiga as características de instrução, abrangendo os seguintes aspectos: alfabetização, anos de estudo, frequência à escola, curso e série concluídos, incluindo cursos de nível superior²³;
- Trabalho e rendimento: investiga a composição da força de trabalho do País, distinguindo as pessoas que procuram trabalho e as que têm trabalho, identifica as principais características do trabalho, tais como ocupação e horas trabalhadas, e retrata o nível de rendimento da população, pesquisando a existência e o valor dos rendimentos de trabalho e de outras fontes das pessoas de 10 anos ou mais de idade. Para a finalidade do Censo Demográfico de 2000, trabalho em atividade econômica é o exercício de trabalho remunerado, trabalho sem remuneração e trabalho na produção para o próprio consumo, na semana de referência - 23 a 29 de julho de 2000.

2.1.2 ASPECTOS DA AMOSTRAGEM DO CENSO

Desde 1960 vem sendo utilizada a técnica de amostragem aleatória na coleta do Censo Demográfico do Brasil. O desenho amostral adotado compreende a seleção sistemática e com equiprobabilidade, dentro de cada setor censitário, de uma amostra dos domicílios particulares e das famílias ou componentes de grupos conviventes recenseados em domicílios coletivos, com fração amostral constante para setores de um mesmo município. Para a realização do Censo Demográfico de 2000, da mesma forma que no Censo de 1991, foram definidas duas frações amostrais distintas: 10% para os municípios com população estimada superior a 15.000 habitantes e 20% para os demais municípios.

Na coleta das informações do Censo 2000, foram usados dois modelos de questionário:

²³ Algumas variáveis, que ainda se encontram em tratamento de crítica e codificação, não estão disponíveis e serão incorporadas em uma segunda versão do arquivo de microdados, dentre elas, a espécie do curso superior ou de mestrado ou doutorado codificada a três dígitos, segundo nova classificação por área de afinidade.

- Um questionário básico aplicado nas unidades não selecionadas para a amostra e contendo perguntas referentes às características que foram investigadas para 100% da população;
- Um segundo questionário aplicado somente nos domicílios selecionados para a amostra contendo, além das perguntas que também constam do questionário básico, outras perguntas mais detalhadas sobre características do domicílio e de seus moradores, referentes aos temas religião, cor ou raça, deficiência, migração, escolaridade, fecundidade, nupcialidade, trabalho e rendimento.

Em todo o território nacional foram selecionados aleatoriamente 5.304.711 domicílios para responder ao questionário da amostra, o que significou uma fração amostral da ordem de 11,7%. Nesses domicílios foram levantadas as informações para todos os seus moradores, totalizando 20.274.412 pessoas.

2.2 DESCRIÇÃO DA AMOSTRA

Utiliza-se os Dados da Amostra do Censo Demográfico 2000 - IBGE, de onde foram consideradas as seguintes variáveis: Estado, Sexo, Idade, Cor ou Raça, Estado Civil, Curso Mais Elevado Concluído, "Exerceu trabalho remunerado?", "Era aposentado de instituto de previdência oficial?", Rendimento do Trabalho Principal, Rendimento dos Demais Trabalhos, Horas Trabalhadas por Semana no Trabalho Principal e Horas Trabalhadas por Semana nos Demais Trabalhos.

Inicialmente, foram selecionados os indivíduos com formação superior, com idade entre 18 e 60 anos e que não fossem aposentados por instituto de previdência, sendo consideradas vinte formações²⁴: Agronomia, Veterinária, Biologia, Farmácia, Medicina, Odontologia, Eng. Civil, Eng. Elétrica, Eng. Mecânica, Eng. Química, Física, Matemática, Química, Administração, Economia, Comunicação, Direito, Pedagogia, Psicologia e Letras,

²⁴ As vinte formações escolhidas neste trabalho são as cobertas no Exame do Provão de 2000, coordenado pelo MEC.

totalizando 339.004 observações. Deste total, foi calculado o número de indivíduos sem trabalho, para cada formação, o que é mostrado na Tabela 2.1.

<i>Formação</i>	<i>Todos</i>	<i>Trabalham</i>	<i>Não Trabalham</i>	<i>Não Trabalham (%)</i> ²⁵
Agronomia	7.284	6.686	598	8%
Veterinária	3.455	3.144	311	9%
Biologia	9.936	8.522	1.414	14%
Farmácia	5.168	4.789	379	7%
Medicina	19.700	18.734	966	5%
Odontologia	12.787	11.993	794	6%
Eng. Civil	13.539	12.513	1.026	8%
Eng. Elétrica	7.472	7.027	445	6%
Eng. Mecânica	7.753	7.266	487	6%
Eng. Química	2.788	2.493	295	11%
Física	1.057	969	88	8%
Matemática	12.169	11.222	947	8%
Química	2.988	2.680	308	10%
Administração	57.847	50.655	7.192	12%
Economia	15.490	13.672	1.818	12%
Comunicação	15.201	12.485	2.716	18%
Direito	51.522	45.025	6.497	13%
Pedagogia	47.618	40.434	7.184	15%
Psicologia	12.793	10.167	2.626	21%
Letras	32.437	27.830	4.607	14%
Total	339.004	298.306	40.698	12%

Tabela 2.1: Percentuais de indivíduos sem trabalho. **Fonte:** Elaborado pelo autor.

Pode-se observar na Tabela 2.1 que a formação com menor número de pessoas sem trabalho é Medicina, com 5% dos médicos não trabalhando. Outras formações apresentam resultados semelhantes a este são: Odontologia (6%), Engenharias Elétrica e Mecânica (6%) e Farmácia (7%). Um segundo grupo de formações vem logo em seguida: Agronomia (8%), Engenharia Civil, Física e Matemática (8%), Veterinária (9%) e Química (10%). Em torno da taxa média de 12% de indivíduos sem trabalho, encontram-se: Engenharia Química (11%), Administração e Economia (12%) e Direito (13%). Por outro lado, Psicologia apresenta o maior número de indivíduos sem trabalho, com 21% dos psicólogos

²⁵ Inclui-se na classificação "não trabalham" aqueles indivíduos que declararam não estar trabalhando nem estar temporariamente afastado do trabalho no período de referência da pesquisa.

estando sem trabalho. Comunicação também apresenta um número bastante alto, 18% dos formados sem trabalho. Outras formações com resultados relativamente altos são: Biologia e Letras (14%) e Pedagogia (15%).

Foi calculada a relação salário/hora trabalhada, considerando a proporção de horas trabalhadas em cada trabalho. Foi realizado um corte inferior e outro superior no valor do salário/hora, excluindo da amostra indivíduos com salário/hora menor do que R\$ 0,50 e acima de R\$300,00 - o que representou 0,24% do total da amostra e não mais de 0,36% em cada uma das formações. Após este corte e a retirada de observações com *missing data*, a amostra passou a apresentar 296.474 observações, divididas entre as 20 formações segundo a Tabela 2.2.

<i>Formação</i>	<i>Sul</i>	<i>Sudeste</i>	<i>C.oeste</i>	<i>Nordeste</i>	<i>Norte</i>	<i>Total</i>
Agronomia	1.691	2.480	864	1.244	363	6.642
Veterinária	771	1.387	373	474	115	3.120
Biologia	1.337	4.844	618	1.403	263	8.465
Farmácia	1.117	2.303	419	728	192	4.759
Medicina	2.967	10.510	1.326	3.071	727	18.601
Odontologia	1.843	7.341	891	1.488	373	11.936
Eng. Civil	1.952	6.823	950	2.211	495	12.431
Eng. Elétrica	895	4.748	436	689	231	6.999
Eng. Mecânica	1.067	5.187	168	639	155	7.216
Eng. Química	471	1.541	54	364	58	2.488
Física	135	617	67	117	30	966
Matemática	1.884	6.888	709	1.313	372	11.166
Química	390	1.858	112	268	45	2.673
Administração	8.008	32.587	3.170	5.289	1.266	50.320
Economia	2.268	7.473	1.210	2.083	545	13.579
Comunicação	1.549	8.562	681	1.385	219	12.396
Direito	6.896	27.302	3.817	5.441	1.294	44.750
Pedagogia	7.287	21.018	3.942	6.489	1.449	40.185
Psicologia	1.313	6.866	594	1.154	180	10.107
Letras	4.948	14.900	2.034	4.930	863	27.675
Total	48.789	175.235	22.435	40.780	9.235	296.474

Tabela 2.2: Distribuição dos indivíduos pela formação e região. **Fonte:** Elaborado pelo autor.

A região Sudeste abriga mais da metade do total de indivíduos com formação superior no país, aproximadamente 59%. As regiões Sul, Centro-Oeste, Nordeste e Norte representam aproximadamente 16%, 8%, 14% e 3%, respectivamente. As formações que apresentam maiores números de observações são Administração, Direito e Pedagogia, que juntas respondem por aproximadamente 46% da amostra. Por outro lado, as formações que apresentam menor número de observações são Física, Engenharia Química e Química, que juntas correspondem a aproximadamente 2% da amostra.

<i>Formação</i>	<i>Sul</i>	<i>Sudeste</i>	<i>Coeste</i>	<i>Nordeste</i>	<i>Norte</i>	<i>Brasil</i>
Agronomia	3%	1%	4%	3%	4%	2%
Veterinária	2%	1%	2%	1%	1%	1%
Biologia	3%	3%	3%	3%	3%	3%
Farmácia	2%	1%	2%	2%	2%	2%
Medicina	6%	6%	6%	8%	8%	6%
Odontologia	4%	4%	4%	4%	4%	4%
Eng. Civil	4%	4%	4%	5%	5%	4%
Eng. Elétrica	2%	3%	2%	2%	3%	2%
Eng. Mecânica	2%	3%	1%	2%	2%	2%
Eng. Química	1%	1%	*	1%	1%	1%
Física	*	*	*	*	*	*
Matemática	4%	4%	3%	3%	4%	4%
Química	1%	1%	*	1%	*%	1%
Administração	16%	19%	14%	13%	14%	17%
Economia	5%	4%	5%	5%	6%	5%
Comunicação	3%	5%	3%	3%	2%	4%
Direito	14%	16%	17%	13%	14%	15%
Pedagogia	15%	12%	18%	16%	16%	14%
Psicologia	3%	4%	3%	3%	2%	3%
Letras	10%	9%	9%	12%	9%	9%

Tabela 2.3: Distribuição dos indivíduos entre as diferentes formações por região.

Fonte: Elaborado pelo autor. *Participação menor do que 1% da amostra

A Tabela 2.3 apresenta a participação de cada formação no total de indivíduos com formação superior de cada região. Tomando Direito como exemplo, vê-se que esta

formação representa 17% dos indivíduos com formação superior na região Centro-Oeste, 13% na região Nordeste e 15% no Brasil. Já para a formação Administração, nota-se uma grande diferença entre as regiões Sudeste, com 19% de participação, e Nordeste, com 13%, bastante abaixo da média nacional de 17%. Os percentuais de indivíduos formados em Pedagogia é bem maior nas regiões Nordeste, Centro-Oeste e Norte, do que nas regiões Sul e Sudeste.

Formação	Total	Cor/Raça		Sexo	
		Branco	Não-branco	Homem	Mulher
Agronomia	6642	83%	17%	89%	11%
Veterinária	3120	87%	13%	71%	29%
Biologia	8465	81%	19%	26%	74%
Farmácia	4759	84%	16%	33%	67%
Medicina	18601	85%	15%	62%	38%
Odontologia	11936	87%	13%	45%	55%
Eng. Civil	12431	85%	15%	84%	16%
Eng. Elétrica	6999	83%	17%	92%	8%
Eng. Mecânica	7216	86%	14%	97%	3%
Eng. Química	2488	86%	14%	71%	29%
Física	966	78%	22%	72%	28%
Matemática	11166	79%	21%	42%	58%
Química	2673	82%	18%	56%	44%
Administração	50320	85%	15%	62%	38%
Economia	13579	83%	17%	68%	32%
Comunicação	12396	86%	14%	38%	62%
Direito	44750	85%	15%	60%	40%
Pedagogia	40185	78%	22%	8%	92%
Psicologia	10107	88%	12%	12%	88%
Letras	27675	78%	22%	14%	86%
Total	296474	83%	17%	48%	52%

Tabela 2.4: Distribuição dos Indivíduos por Cor/Raça e Sexo. **Fonte:** Elaborado pelo autor.

Os indivíduos da amostra foram classificados segundo critérios de Cor/Raça (Branco e Não-branco) e Sexo (Homem e Mulher). Pode-se observar na Tabela 2.4 que os indivíduos brancos são a grande maioria da amostra, representando aproximadamente 83% da mesma. As profissões com maior participação de indivíduos não-brancos são

Matemática (21%) e Física, Pedagogia e Letras (22%). Quanto ao sexo dos indivíduos da amostra, as mulheres estão ligeiramente à frente, com 52% do total de indivíduos formados. Deve-se notar que algumas formações são tipicamente femininas, como Biologia, Psicologia, Pedagogia e Letras e outras tipicamente masculinas, como Agronomia, Veterinária, Física e todas as Engenharias.

Considerando os indivíduos de todas as carreiras, nota-se que enquanto a parcela de mulheres graduadas é ligeiramente maior do que de homens, os não-brancos apresentam uma participação bastante inferior a dos brancos (somente 17% da amostra é composta por não-brancos).

Observando-se as diferenças nas composições das formações segundo região, sexo e cor/raça, pode-se constatar uma grande heterogeneidade entre as diferentes carreiras. Algumas carreiras são tipicamente femininas e indivíduos não-brancos têm participação mais acentuada em algumas carreiras específicas. Observou-se também que o percentual de indivíduos sem trabalho varia significativamente entre profissões, o que é um importante fator de diferencial de risco entre profissões.

Para explorar as diferenças entre os salários médios das várias formações segundo o sexo e cor/raça, toma-se a Tabela 2.5. Inicialmente, observa-se, uma grande variação do salário/hora médio dos indivíduos segundo a formação. É importante notar as formações com maior percentual de mulheres, Biologia, Pedagogia e Letras, são também as que apresentam os menores salários médios. Além disso, das quatro profissões com maior percentual de indivíduos não-brancos, três apresentam salários médios mais baixos: Pedagogia, Letras e Matemática (quarto menor salário médio).

Nas colunas *Diferenças* da Tabela 2.5, pode-se observar quais formações apresentam maiores diferenças entre salários por cor e sexo. Na última linha da tabela, vê-se que, em média, os indivíduo não-brancos ganha um salário 19% menor do que os brancos e que as mulheres recebem, em média 41% do salário dos homens. As formações que apresentam maiores diferenças salariais médias entre brancos e não-brancos são:

Engenharias Civil e Química (21% e 25% de diferença, respectivamente), Economia e Psicologia (ambas com uma diferença de 22%) e Administração (23%). Já as formações com maiores diferenças entre homens e mulheres são: Engenharias Civil, Mecânica e Química (38%, 39% e 36%, respectivamente), Comunicação (32%) e Administração e Economia (ambas com 40% de diferença).

Formação	Todos	Cor/Raça			Sexo		
		Branco	Não-Branco	Diferença	Homem	Mulher	Diferença
Agronomia	14,83	15,28	12,74	17%	15,25	11,35	26%
Veterinária	14,05	14,20	13,03	8%	15,17	11,24	26%
Biologia	9,45	9,75	8,15	16%	12,09	8,51	30%
Farmácia	14,12	14,20	13,70	4%	17,01	12,70	25%
Medicina	30,21	30,80	26,80	13%	33,53	24,75	26%
Odontologia	19,13	19,35	17,68	9%	21,28	17,33	19%
Eng. Civil	19,28	19,92	15,72	21%	20,50	12,64	38%
Eng. Elétrica	17,82	18,26	15,70	14%	18,19	13,38	26%
Eng. Mecânica	18,07	18,60	14,92	20%	18,32	11,09	39%
Eng. Química	16,86	17,49	13,04	25%	18,84	12,14	36%
Física	14,75	15,10	13,55	10%	15,62	12,54	20%
Matemática	10,15	10,35	9,37	9%	12,08	8,77	27%
Química	12,34	12,77	10,30	19%	14,15	10,04	29%
Administração	13,12	13,58	10,50	23%	15,45	9,23	40%
Economia	15,79	16,41	12,75	22%	18,11	10,79	40%
Comunicação	12,82	13,12	10,94	17%	15,91	10,90	32%
Direito	17,54	17,74	16,45	7%	19,76	14,27	28%
Pedagogia	7,74	7,91	7,12	10%	10,03	7,55	25%
Psicologia	13,60	13,98	10,90	22%	15,91	13,28	16%
Letras	8,85	9,13	7,87	14%	10,97	8,51	22%
Total	14,47	14,95	12,07	19%	18,42	10,84	41%

Tabela 2.5: Salário/hora médio por Cor/Raça e Sexo. **Fonte:** Elaborado pelo autor.

A Tabela 2.6 mostra os salários médios de cada profissão de acordo com a região geográfica. Vê-se que a região Nordeste apresenta o menor salário médio, enquanto a região Centro-oeste apresenta o maior resultado.

Para simplificar a análise, considera-se neste trabalho somente os diferenciais de salários entre as regiões Nordeste e Sudeste. A última coluna da Tabela 2.6 apresenta as

diferenças entre os salários médios das duas regiões, tomando como base o salário médio da região Sudeste. Note-se que valores negativos aparecem quando o salário/hora médio para o Nordeste é superior ao do Sudeste, o que somente ocorre em três formações: Farmácia, Odontologia e Direito. Para todas as outras formações, o Nordeste apresenta salários médios inferiores, principalmente para as formações em Biologia e Física (31% a menos), Matemática e Economia (27% a menos), Pedagogia e Letras, com 24% e 29% a menos, respectivamente. Considerando todas as formações, indivíduos atuando na região Nordeste recebem um salário/hora aproximadamente 10% menor do que os aqueles que atuam no Sudeste.

Formação	Região					Diferença Sudeste x Nordeste
	Sul	Sudeste	Coeste	Nordeste	Norte	
Agronomia	14,05	15,53	16,68	13,18	14,97	15%
Veterinária	14,43	14,08	14,76	12,42	15,46	12%
Biologia	7,88	10,49	9,07	7,24	11,08	31%
Farmácia	13,55	13,75	15,55	15,10	15,13	-10%
Medicina	30,02	30,33	32,34	29,69	27,72	2%
Odontologia	20,40	18,40	20,93	19,76	20,23	-7%
Eng. Civil	18,48	20,01	20,35	17,78	17,00	11%
Eng. Elétrica	17,48	17,85	19,61	17,78	15,30	*
Eng. Mecânica	17,65	18,40	20,11	15,54	18,24	16%
Eng. Química	15,72	17,40	18,82	16,60	11,78	5%
Física	12,03	15,93	17,97	11,05	9,92	31%
Matemática	7,80	11,12	11,20	8,16	9,02	27%
Química	10,37	12,95	12,91	11,04	10,31	15%
Administração	12,13	13,45	13,56	12,35	12,83	8%
Economia	13,06	17,24	17,73	12,66	14,91	27%
Comunicação	11,70	13,11	17,16	10,24	12,08	22%
Direito	17,93	16,46	20,97	19,39	20,48	-18%
Pedagogia	6,86	8,40	8,03	6,35	8,07	24%
Psicologia	13,84	13,84	14,52	11,00	15,99	21%
Letras	8,00	9,76	8,75	6,91	9,31	29%
Todas	13,56	14,80	15,63	13,39	14,77	10%

Tabela 2.6: Salário/hora médio por Região. **Fonte:** Elaborado pelo autor.

* diferença próxima a zero

A Tabela 2.7 resume os diferenciais de salários/hora médios entre indivíduos de diferentes raças, sexo e regiões, para qualquer formação superior. Independente da região a categoria com salário/hora mais alto é formada por homens brancos, seguida dos homens não-brancos, das mulheres brancas e, por último, das mulheres não-brancas, que recebem menos da metade do salário médio do homem branco. Observa-se também que a maior diferença entre os salários médios das regiões encontra-se no grupo das mulheres não-brancas (14,26%). Um interessante resultado é que o salário médio dos homens não-brancos do Nordeste é 5,89% maior do que os do Sudeste.

<i>Região</i>	<i>Homem</i>		<i>Mulher</i>		<i>Qualquer Cor e Sexo</i>
	<i>Branco</i>	<i>Não-branco</i>	<i>Branco</i>	<i>Não-branco</i>	
Nordeste	19,07	15,45	10,53	8,30	13,39
Sudeste	19,15	14,59	11,52	9,68	14,80
Diferença	0,42%	-5,89%	8,59%	14,26%	9,53%

Tabela 2.7: Salário médio para diferentes grupos demográficos. **Fonte:** Elaborado pelo autor.

Nas duas seções que seguem são utilizados métodos paramétricos para explorar os diferenciais de salários entre as carreiras. Na Seção 3, as diferentes carreiras são ordenadas a partir dos momentos amostrais das distribuições de retorno (tomando-se o desvio-padrão como medida de risco). Na Seção 4, a partir de uma regressão de salários, estima-se os efeitos líquidos das variáveis demográficas consideradas nesta seção sobre o diferencial de salários entre os indivíduos, controlando pela experiência de trabalho e pela influência do fato de um indivíduo estar ou não no serviço público.

3. ORDENAÇÃO DE DISTRIBUIÇÕES UTILIZANDO MOMENTOS AMOSTRAIS

O método mais comum de ordenação de distribuições de retornos consiste na simples comparação de medidas estatísticas pontuais, como média, desvio-padrão e coeficiente de variação. Para uma comparação inicial, a Tabela 3.1 apresenta a ordenação das profissões segundo tais medidas.

<i>Formação</i>	<i>Min.</i>	<i>Max.</i>	<i>Média</i>	<i>Rank</i>	<i>D.Padrão</i>	<i>Rank</i>	<i>Coef. Var.</i>	<i>Rank</i>
Agronomia	0,50	287,67	14,83	9°	20,71	4°	1,40	20°
Veterinária	0,55	268,49	14,05	12°	16,46	14°	1,17	11°
Biologia	0,50	268,49	9,45	18°	10,95	17°	1,16	8°
Farmácia	0,58	255,71	14,12	11°	16,53	13°	1,17	10°
Medicina	0,51	299,18	30,21	1°	28,51	1°	0,94	1°
Odontologia	0,69	294,58	19,13	3°	19,12	6°	1,00	2°
Eng. Civil	0,50	299,18	19,28	2°	23,08	2°	1,20	15°
Eng. Elétrica	0,51	255,71	17,82	5°	17,89	9°	1,00	3°
Eng. Mecânica	0,50	287,67	18,07	4°	18,23	7°	1,01	4°
Eng. Química	0,50	287,67	16,86	7°	17,97	8°	1,07	6°
Física	0,64	287,67	14,75	10°	17,45	10°	1,18	12°
Matemática	0,58	287,67	10,15	17°	10,68	18°	1,05	5°
Química	0,52	293,34	12,34	16°	15,10	16°	1,22	16°
Administração	0,50	295,89	13,12	14°	17,27	11°	1,32	19°
Economia	0,50	287,67	15,79	8°	19,63	5°	1,24	17°
Comunicação	0,50	276,16	12,82	15°	16,81	12°	1,31	18°
Direito	0,50	287,67	17,54	6°	20,91	3°	1,19	14°
Pedagogia	0,50	268,49	7,74	20°	8,48	20°	1,09	7°
Psicologia	0,52	287,67	13,60	13°	16,12	15°	1,19	13°
Letras	0,50	276,16	8,85	19°	10,34	19°	1,17	9°

Tabela 3.1: Ranking por Média, Desvio Padrão e Coeficiente de Variação²⁶. **Fonte:** Elaborado pelo autor.

Pode-se observar que as ordenações apresentam uma forte correspondência entre média e desvio-padrão, o que é um indício de que o risco inerente a uma formação (neste caso, medido pelo desvio-padrão) é compensado por um salário médio mais alto, ou seja, existe uma relação risco-retorno positiva²⁷. Este fato pode ser visualizado na Figura 3.1 e

²⁶ Quanto menor o valor do coeficiente de variação, melhor a classificação da formação no *ranking*.

²⁷ Este resultado foi também encontrado por King (1974) ao comparar um conjunto de ocupações.

confirmado pelo coeficiente de correlação positivo de 0,917 entre os *rankings* baseados nas duas medidas, significativa ao nível de 0,01.

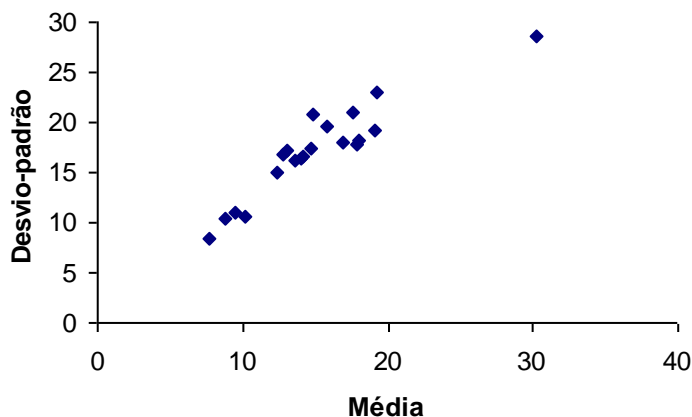


Gráfico 3.1: Dispersão entre a média e o desvio-padrão dos salários/hora. **Fonte:** Elaborado pelo autor.

Utilizando o coeficiente de variação como critério, vê-se que Medicina apresenta o melhor resultado, apesar de apresentar o maior desvio-padrão, e Agronomia aparece na última colocação, apesar de apresentar o 9º maior salário médio e o 4º maior desvio-padrão. Como se vê, o coeficiente de variação consiste de uma medida arbitrária de risco/retorno, podendo não representar o comportamento do indivíduo frente ao risco (aversão ao risco).

Weiss (1972) sugere uma análise de escolha de capital humano baseada na hipótese de que as distribuições dos retornos são lognormais ao longo do tempo²⁸ e que a função de utilidade dos indivíduos assume a seguinte forma²⁹

$$u(y) = \frac{1}{1-\alpha} y^{1-\alpha}, \quad (3.1)$$

onde α é a medida de aversão relativa ao risco de Arrow e y é o retorno do indivíduo. Weiss utiliza o seguinte índice de utilidade esperada como medida, onde c é o coeficiente de variação e μ é a média dos retornos³⁰

²⁸ Na análise de distribuições de retornos de um período (com dados *cross-section*), supõe-se que distribuições dos retornos são lognormais ao longo das observações.

²⁹ Trata-se de uma função de utilidade de Arrow, a qual apresenta aversão relativa ao risco constante.

$$v(\mu, c) = \ln \mu - \frac{1}{2} \alpha \ln(1 + c^2). \quad (3.2)$$

Como explica Evans e Weinstein (1982), o grau de aversão relativa ao risco pode ter uma influência significativa sobre a ordenação das ocupações. Assim, a limitação desta medida para mensuração do risco percebido pelos indivíduos se deve ao fato de que, mesmo supondo uma mesma forma funcional para a utilidade dos diferentes indivíduos, dependendo o grau de aversão ao risco suposto, ter-se-á ordenações diferentes. A Tabela 3.2 mostra os valores do índice de Weiss para $\alpha = 0.1$ e $\alpha = 0.9$:

Formação	$\alpha = 0.1$		$\alpha = 0.9$	
	Valor	Ranking	Valor	Ranking
Agronomia	2,64	10°	2,21	13°
Veterinária	2,60	12°	2,25	11°
Biologia	2,20	18°	1,86	18°
Farmácia	2,60	11°	2,26	10°
Medicina	3,38	1°	3,12	1°
Odontologia	2,92	2°	2,64	2°
Eng. Civil	2,91	3°	2,56	5°
Eng. Elétrica	2,85	5°	2,57	4°
Eng. Mecânica	2,86	4°	2,58	3°
Eng. Química	2,79	7°	2,48	6°
Física	2,65	9°	2,30	9°
Matemática	2,28	17°	1,98	17°
Química	2,47	16°	2,10	16°
Administração	2,52	14°	2,12	14°
Economia	2,71	8°	2,34	8°
Comunicação	2,50	15°	2,10	15°
Direito	2,82	6°	2,47	7°
Pedagogia	2,01	20°	1,69	20°
Psicologia	2,57	13°	2,21	12°
Letras	2,14	19°	1,79	19°

Tabela 3.2: Ranking sobre a utilidade esperada de Weiss. **Fonte:** Elaborado pelo autor.

³⁰ Evans e Weinstein (1982) aplicam este critério sobre as distribuições de retornos entre diferentes ocupações nos Estados Unidos, utilizando o Censo de 1970.

Embora os *rankings* para $\alpha = 0,1$ e $\alpha = 0,9$ apresentem uma correlação alta (0,985, com nível de significância de 0,01), pode-se notar que algumas formações podem apresentar um resultado melhor ou pior, dependendo da aversão ao risco do agente. Tomando a distribuição dos retornos para Agronomia como exemplo, vê-se com indivíduos avessos ao risco classificam esta formação em 10º lugar no *ranking* de utilidade esperada quando a aversão relativa ao risco é baixa ($\alpha = 0,1$). Contudo, quando a função de utilidade é modificada para englobar indivíduos altamente avessos ao risco ($\alpha = 0,9$), esta formação cai três posições, ficando na 13º colocação.

Apesar de uma alta correlação entre *rankings* ser um fator positivo para um critério, esta correlação não garante que todos os indivíduos tenham suas preferências exatamente representadas por qualquer um dos critérios abordados nesta seção. Segundo Hause (1974), deve-se reconhecer a diversidade de níveis de aversão ao risco entre os indivíduos, em vez de assumir que todos os membros envolvidos têm funções de utilidade idênticas.

4. APLICAÇÃO DA REGRESSÃO MINCERIANA

Uma diferença básica deste estudo com relação a maior parte dos estudos sobre investimento em capital humano é que se considera aqui indivíduos com mesmo nível de escolaridade (todos com formação superior completa). Assim, como o número de anos de estudo é considerado constante entre todos os indivíduos com determinada formação superior, não há necessidade de se incluir a variável *anos de estudo* na regressão de salários de cada formação. Contudo, na regressão que considera indivíduos de todas as carreiras, esta variável deve ser incluída, pois diferentes formações exigem diferentes quantidades de anos de estudo. Considerando esta particularidade (S sendo constante entre os diferentes indivíduos dentro ou não de uma mesma profissão), propõe-se o seguinte modelo econométrico:

$$\ln(w) = \beta_0 + \beta_1 \text{anos estudo} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + \beta_4 \text{sexo} + \beta_5 \text{cor} + \beta_6 \text{setor pub} + \beta_7 \text{região} + \varepsilon \quad (4.1)$$

onde a variável *exper* é a experiência de trabalho do indivíduo, *cor* é uma variável *dummy* assumindo valores 1 (branco) ou 0 (outros), *setor pub* é uma variável *dummy* para a natureza jurídica do empregador, assumindo valor 1 (serviço público) e 0 caso contrário e *região* define um conjunto de quatro variáveis *dummy*, uma para cada região geográfica do Brasil (Região Sudeste é omitida).

Para a regressão que considera todas as profissões, a variável *anos estudo* foi calculada do seguinte modo:

$$\text{anos estudo} = \text{duração do curso} + 11, \quad (4.2)$$

ou seja, o total de anos de estudo do indivíduo é igual ao número de anos de estudo do ensino fundamental, médio e superior³¹. Como foi explicado anteriormente, a variável

³¹ A Lei de Diretrizes e Bases nº 9.394, de 20 de Dezembro de 1996, estabelece que o ensino fundamental deve durar oito anos e o ensino médio, três anos.

relativa aos anos de estudo dos indivíduos (*anos estudo*) deve ser desconsiderada na regressão de salários para cada formação particular.

O objetivo desta subseção é estimar os retornos das vinte diferentes profissões, comparando os diferenciais de renda segundo o sexo, cor ou raça, região geográfica e natureza jurídica do empregador dos indivíduos, controlando pela experiência de trabalho. Entretanto, a base de dados utilizada (Censo de 2000) infelizmente não contém informação sobre a idade de ingresso no mercado de trabalho dos indivíduos, nem a idade com a qual este se graduou, sendo impossível obter diretamente valores para a variável *exper*.

Como foi visto anteriormente, uma opção para aproximar a experiência do indivíduo é utilizar a experiência potencial do mesmo, calculada como a diferença entre sua idade e a o número de anos de estudo adicionado de seis ($exper = idade - anos\ de\ estudo - 6$). No entanto, se diferentes profissões apresentarem diferentes perfis de mercado de trabalho, os quais apresentam diferentes exigências quanto ao tempo de formatura ou de treinamento *post school*, então a aplicação generalizada dessa fórmula pode não capturar peculiaridades importantes de cada profissão. A seguir expõe-se duas opções para aproximar a variável *exper*.

A primeira opção para aproximar a experiência de trabalho dos indivíduos de diferentes formações superiores seria considerar a idade média de formatura nos diferentes cursos superiores, o que seria aplicado no cálculo da experiência da seguinte forma:

$$Experiência = Idade - Idade\ de\ Formatura . \quad (4.3)$$

Esta última informação pode obtida a partir da base de dados do Relatório Final da Análise dos dados do PROVÃO (CESOP-UNICAMP), que fornece a idade média do formando daquele ano de vinte diferentes cursos, segundo a Tabela 4.1.

<i>Curso</i>	<i>Idade média do formando</i>	<i>Número de observações</i>	<i>Desvio-padrão</i>
Administração	26,54	180620	5,67
Direito	27,86	192483	7,49
Eng Civil	25,62	27474	4,53
Eng Química	24,72	6730	3,20
Veterinária	24,52	12015	3,09
Odontologia	23,45	36943	2,81
Matemática	28,75	38325	7,23
Comunicação	24,91	18495	5,07
Letras	28,92	76689	7,41
Eng Elétrica	25,60	16390	3,98
Eng Mecânica	25,67	9427	3,80
Medicina	24,83	22500	2,69
Economia	27,34	24014	5,63
Física	27,05	2995	6,10
Química	27,21	6201	5,90
Biologia	26,86	21019	6,72
Agronomia	24,96	6234	3,70
Psicologia	27,29	17933	7,21
Farmácia	23,79	8066	3,81
Pedagogia	31,59	43752	8,22
Total	27,17	768305	6,60

Tabela 4.1: Idade média dos participantes do Provão 1997-2001.

Fonte: Análise dos dados do ENC (PROVÃO), Relatório Final, CESOP - UNICAMP

Inserindo as idades médias de formatura para os 20 cursos analisados e calculando a experiência se acordo com a equação 4.3, pode-se constatar que um grande número de observações apresentaram um valor negativo para experiência, pois muitos indivíduos apresentam idade inferior à idade média de formatura. Para contornar este problema, considerou-se duas possibilidades: excluir todas as observações apresentando experiência negativa ou retirar da amostra todas as observações que apresentam idade inferior a 32 anos, deste modo excluindo qualquer possibilidade de ocorrência de experiência negativa, já que idade média de formatura máxima é de 31,59 anos (referente a formação Pedagogia (ver Tabela 4.1).

Utilizando-se a primeira possibilidade, que é a que elimina menos observações, obteve-se um corte de 34.208 observações e a amostra passou a apresentar um total de 259.955 observações. Estimou-se o modelo de regressão dado pela equação 4.1 por mínimos quadrados ordinários para cada uma das profissões, com os resultados podendo ser visualizados no Anexo 1.

A segunda opção para aproximar a experiência de trabalho dos indivíduos para as diferentes formações superiores seria considerar o tempo médio de duração dos diferentes cursos superiores, o qual seria aplicado no cálculo da experiência da seguinte forma:

$$\text{Experiência} = \text{Idade} - \text{Duração do Curso} - 17, \quad (4.3)$$

onde supõe-se que os indivíduos terminem o ensino médio aos dezessete anos de idade. A Tabela 4.2 mostra a duração média dos cursos superiores abordados neste trabalho.

<i>Curso Superior</i>	<i>Duração</i>	<i>Curso Superior</i>	<i>Duração</i>
Agronomia	5	Física	4
Veterinária	5	Matemática	4
Biologia	4	Química	4
Farmácia	4	Administração	4
Medicina	6	Economia	4
Odontologia	5	Comunicação	4
Eng. Civil	5	Direito	5
Eng. Elétrica	5	Pedagogia	4
Eng. Mecânica	5	Psicologia	4
Eng. Química	5	Letras	4

Tabela 4.2: Duração média do curso superior.

Fonte: Guia Abril do Estudante, 2000.

Assim como a primeira opção para aproximação da experiência, encontra-se também alguns valores negativos para a variável *exper*, só que em um número irrisório de observações (somente 381 observações), a amostra passando a apresentar um total de 296.093 observações após o corte. Estimou-se o modelo de regressão dado na equação 4.1

por mínimos quadrados ordinários para cada uma das profissões, com os resultados podendo ser visualizados no Anexo 2.

É importante notar que nas regressões deste trabalho não são considerados alguns dos possíveis problemas econométricos pertinentes à regressão de salários, como o viés de seleção da amostra³² e o viés de habilidade. Isto é devido à limitação dos dados em *cross-section* utilizados, além de que o objetivo desta seção é somente obter uma idéia geral das diferenças salariais dentro de cada profissão, deixando-se o tratamento desses possíveis problemas para outro trabalho.

Por conta do menor descarte de observações, escolheu-se analisar somente os resultados desta última regressão. Assim, daqui em diante no trabalho, deve-se considerar os resultados contidos no Anexo 2. Como a variável dependente (salário/hora) está em log, os coeficientes das regressões podem ser interpretados como uma variação percentual no log do salário/hora devido à variação em uma unidade na variável independente.

Primeiramente, pode-se observar baixos coeficientes de significância global, R^2 , para todas as regressões, em especial aquelas para formações específicas. Isto é uma característica típica de estudos utilizando *cross-section* de indivíduos, os quais apresentam um nível de heterogeneidade alto o bastante para tornar impossível aproximar-se de todas as possíveis variáveis que explicam as diferenças individuais. Segundo Wooldridge (2000), o fato das variáveis incluídas na regressão explicarem somente um pequeno percentual da variação nos salários não necessariamente significa que a equação estimada não seja útil. Segundo o autor, "mesmo tais variáveis não explicando coletivamente a maior parte da variação nos salários, ainda assim é possível que as estimativas de mínimos quadrados ordinários sejam estimativas confiáveis dos efeitos *ceteris paribus* da cada variáveis independente sobre o salário" (Wooldridge, 2000, pg. 83). É importante notar que, enquanto na regressão que considera todas as profissões as variáveis independentes explicam os salários em aproximadamente 20%, em todas as regressões (que consideram

³² Problema devido às diferenças de estratégias de *job-search* entre os indivíduos dos diferentes tipos de carreiras. Heckman (1979) aborda esta questão.

cada formação) o R^2 é bastante baixo, chegando a 0,071 na regressão para a formação em Psicologia e 0,089 para Agronomia. Assim, ao restringir as observações a uma formação específica, as variáveis consideradas passaram a explicar menos os retornos salariais.

Analisando as estimativas da regressão para o total das formações, vê-se que a única variável não-significativa é a *dummy* para a região norte, todas as outras apresentando resultados significativos a 1%. Observando a variável *anos estudo*, a qual assume valores entre 15 e 17 anos, vê-se que cada ano de estudo a mais implica num aumento significativo do salário/hora (42,24% do valor do salário/hora). Assim, formações que demandam um maior tempo para conclusão estão relacionadas a um salário médio mais elevado. Todos os outros resultados são os esperados: o fato do indivíduo graduado ser do sexo masculino ou ser branco afeta positivamente o salário; o indivíduo graduado estando na região nordeste afeta negativamente o salário, se comparado à região sudeste.

Observando-se os resultados para todas as regressões, pode-se constatar que os coeficientes para a experiência (*exper* e *exper2*) apresentam resultados compatíveis com Mincer (1974), ou seja, a relação entre o log do salário/hora e os anos de experiência consiste de uma parábola. Percebe-se que a contribuição de um ano a mais de experiência causa um impacto maior sobre os salários em algumas profissões do que em outras. Por exemplo, um ano a mais de experiência tem um impacto inicial bem maior sobre os salários de profissões como Medicina (9%), Odontologia (7,9%) e Engenharia Civil (6,9%) do que em profissões como Letras (4%), Química (4,5%) e Pedagogia (4,8%). No entanto, à medida que o indivíduo vai ganhando mais experiência o termo quadrático, *exper2*, vai reduzindo este impacto, já que o coeficiente de *exper2* é negativamente maior para Medicina (-0,0017), Odontologia (-0,0017) e Engenharia Civil (-0,0010) do que para Letras (-0,0005), Química (-0,0006) e Pedagogia (-0,0008). Esta constatação nos permite conjecturar a existência de outro aspecto que diferencia as carreiras umas das outras: há carreiras (como Medicina e Odontologia) onde o salário sobe mais rapidamente com a experiência e se equilibra (chega ao máximo da parábola) do que outras profissões.

Outras variáveis também apresentam alguns resultados esperados, independente da profissão:

- o fato de o indivíduo ser do sexo masculino afeta positivamente os salários, principalmente para Engenharia Mecânica, Administração e Economia, onde o efeito desta variável é responsável por um salário/hora 46,4%, 34,4% e 37,9% mais altos, respectivamente. Este efeito é bem menor para outras profissões, como Odontologia (12,4%), Química (14,1%), Psicologia (9,8%) e Letras (14,7%);
- o fato de o indivíduo ser branco tem efeito positivo sobre os salários de todas as profissões, com exceção de Física, que apresenta coeficiente não significativo. Este impacto também varia de acordo com a profissão, sendo maior para profissões como Administração, Economia, Comunicação e Engenharias (todas apresentando um impacto acima de 20% do salário/hora) e menor para profissões como Veterinária, Matemática, Pedagogia e Letras (todas apresentando um impacto inferior a 10% do salário/hora).

Analisando a variável *setorpub*, observa-se que o fato de um indivíduo encontrar-se no setor público pode apresentar efeitos positivos ou negativos sobre os salários recebidos, dependendo da profissão. Dentre todas as profissões, sete apresentaram coeficientes não significativos. Já para as formações em Medicina, Engenharia Química, Física, Matemática, Química e Psicologia este efeito é negativo, ou seja, o fato de um indivíduo encontrar-se no setor público está relacionado a um salário mais baixo. Por outro lado, para as formações em Veterinária, Farmácia, Odontologia, Administração, Economia, Direito e Pedagogia o fato de encontrar-se no setor público afeta positivamente o salário recebido. Em especial, este fator é bastante forte para a formação em Direito (salário 32,8% mais alto), provavelmente devido ao fato de o Judiciário pagar salários bem mais altos que o Executivo.

Os coeficientes da variável *dummy* para a região nordeste são significativos para a maioria das formações, com exceção de Farmácia, Engenharia Civil e Engenharia Elétrica. Odontologia e Direito são as únicas profissões onde o fato de se estar no nordeste afeta positivamente os salários, com salários 11,3% maiores para Direito e 4,52% maiores para

Odontologia. Para o restante das profissões, os coeficientes para a região nordeste são negativos, causando um impacto negativo sobre os salários, principalmente nas formações de Economia (-25,7%), Veterinária (-27,6%), Pedagogia (-31,9%), Biologia (-39,8), Letras (-40,7%) e Matemática (-41,4%).

Constata-se, portanto, que as variáveis consideradas nesta seção podem exercer uma maior ou menor influência sobre o salário do indivíduo, dependendo da carreira considerada. Assim, a ordenação das profissões por atratividade dos salários pode variar bastante de acordo com o sexo, cor/raça e região do indivíduo e indivíduos de diferentes grupos sócio-demográficos podem apresentar diferentes perfis de preferências sobre o retorno e o risco de cada profissão.

Nas seções seguintes utilizam-se métodos não paramétricos para a análise dos retornos das profissões, o que inclui a estimação das densidades das distribuições de salários e a aplicação de um teste de dominância estocástica para ordenar as distribuições de salários das diferentes carreiras.

5. ANÁLISE NÃO-PARAMÉTRICA DOS DADOS

Como foi explicado na seção anterior, uma potencial limitação dos métodos paramétricos utilizados é devida às hipóteses de log dos salários e erro normalmente distribuídos. Enquanto os critérios de ordenação das preferências sobre distribuições de retornos baseados na média e desvio-padrão dependem substancialmente da hipótese de normalidade do log dos salários, o critério de dominância estocástica apresenta a grande vantagem de não impor restrições sobre as possíveis formas das distribuições de renda. Para reforçar a utilização do último critério neste trabalho, convém analisar de maneira menos restritiva possível a distribuição do log dos salários para cada profissão.

Segundo Neal e Rosen (2000), é bastante comum para grandes populações, encontrar-se um padrão regular entre distribuições observadas de salários: tais distribuições, em geral, apresentam uma longa cauda direita e uma medida positiva de assimetria. Apesar de comumente serem aproximadas pela densidade lognormal tais distribuições observadas tendem a apresentar caudas mais longas e espessas do que a lognormal.

Há diferentes modos de checar a normalidade de distribuições. Métodos gráficos, como histogramas e outros estimadores de densidade, permitem visualizar diferenças entre a distribuição empírica e a teórica. Apesar de visualmente atrativos, tais métodos não nos provém de um critério objetivo para determinar a normalidade da distribuição. Por outro lado, um método numérico consiste de um critério objetivo para testar uma hipótese, neste caso, a de normalidade. Nas subseções seguintes, os dois métodos serão aplicados para explorar as distribuições observadas de salários para as diferentes profissões.

5.1 ESTIMAÇÃO NÃO-PARAMÉTRICA DAS DENSIDADES

Os estimadores de núcleo aproximam a função densidade de probabilidade, $f(x)$, a partir dos valores observados de x . O procedimento básico se dá pela divisão do suporte da distribuição em intervalos não-sobrepostos e contagem do número de observações

encontradas nos diferentes intervalos³³. O histograma é um tipo de estimador de núcleo, onde a estimativa é apresentada como um conjunto de barras, centradas no ponto médio de cada intervalo, onde a altura de cada uma delas reflete a frequência com que as observações aparecem em cada intervalo.

Algumas características básicas separam estimadores gerais de núcleo de um simples histograma. No caso geral, permite-se que os intervalos sejam definidos de modo a poderem se sobrepor. Além disso, na contagem do número de observações são estabelecidos pesos, entre 0 e 1, baseados na distância entre cada observação e o centro de seu respectivo intervalo, para somente então, somar-se os valores ponderados. A função que atribui tais pesos às observações é chamada de núcleo ou *kernel*.

Com o intuito de apresentar o estimador geral de núcleo, considera-se a seguinte estrutura. Seja uma amostra com n observações retiradas independentemente, X_1, \dots, X_n , identicamente distribuídas com função densidade de probabilidade, f , a ser estimada. Então, o estimador geral de núcleo é dado por

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K \left\{ \frac{x - X_i}{h} \right\}, \quad (5.1)$$

onde K e h são o *kernel* e o *bandwidth* (largura da janela) do estimador, respectivamente.. Para obter uma estimativa suavizada de f , deve-se utilizar uma densidade K que torne o estimador \hat{f} contínuo, ou seja, $K: \mathbb{R} \rightarrow \mathbb{R}$ deve ser uma *fdp* contínua e diferenciável. Obtém-se, assim, uma classe geral de estimadores de densidade, conhecidos como Estimadores de Densidade de Núcleo ou *Kernel Density Estimators*.

Diferentes tipos de núcleo, usualmente simétricos, aparecem na literatura. No entanto, segundo Silverman (1986), há muito pouca diferença entre os vários tipos de *kernel*, quanto a critérios de eficiência assintótica. Para o autor, é legítimo considerar-se

³³ Esta seção baseou-se em Silverman (1986).

outros critérios para a escolha da função a ser utilizada, como seu grau de diferenciabilidade e o esforço computacional envolvido. Um *kernel* bastante utilizado e de fácil computação é o núcleo gaussiano, dado por:

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right],$$

o qual é utilizado neste trabalho.

A escolha do valor do *bandwidth*, h , afeta o número de observações incluídas na estimação de f em torno de cada ponto x do suporte da distribuição. Um pequeno valor de h implica que somente observações muito próximas a x são utilizadas para a estimação de $f(x)$. Como observações mais próximas a x apresentam maior chance de trazer informação a respeito do comportamento da densidade naquele ponto, a precisão do estimador de densidade deve aumentar com a diminuição da largura da janela, diminuindo o viés na estimação. Por outro lado, menos observações serão utilizadas para a estimação de $f(x)$, o que aumenta a variância do estimador.

Utilizando o software *Stata8*®, estimou-se as densidades das distribuições dos logaritmos dos salários-hora para as vinte profissões listadas. Utilizou-se o núcleo gaussiano e a largura ótima da janela calculada pelo programa (aproximadamente 0,12)³⁴. No Anexo 3 são apresentadas as densidades estimadas e as distribuições normais baseadas nos momentos amostrais para as vinte profissões.

Pode-se constatar que a maioria das formações apresentam distribuições do log dos salários aproximadamente normal, com exceção de Engenharia Elétrica, Engenharia Mecânica e Medicina, que são visualmente assimétricas à esquerda, especialmente a última.

³⁴ Para maiores detalhes, ver Silverman (1986)

5.2 APLICAÇÃO DO TESTE DE NORMALIDADE

Como foi argumentado anteriormente, as densidades estimadas não compreendem a mensuração e teste de hipóteses distribucionais, sendo mais utilizadas como uma ferramenta de análise exploratória dos dados inicial, não proporcionando maiores recursos à análise.

Os coeficientes de assimetria e curtose são medidas que descrevem o quão simétrica é uma distribuição e espessas são as caudas desta. Se uma variável for normalmente distribuída, os coeficientes de assimetria e curtose apresentam valores zero e três, respectivamente. Apesar de proporcionar alguma idéia da proximidade de distribuições amostrais com relação à normal, tais medidas, por si só, não compreende um método conclusivo para interpretação dos dados.

A Tabela 5.1 mostra os coeficientes de assimetria e curtose para as distribuições do log do salário/hora para as vinte profissões. Note-se que a distribuição para os graduados em Direito apresenta medidas aparentemente compatíveis com uma distribuição normal.

<i>Formação</i>	<i>Assimetria</i>	<i>Curtose</i>	<i>Formação</i>	<i>Assimetria</i>	<i>Curtose</i>
Agronomia	0,1653	3,5719	Física	-0,0333	3,4697
Veterinária	0,1928	3,5239	Matemática	0,0413	3,5791
Biologia	0,1430	3,5735	Química	0,1003	3,6183
Farmácia	0,2477	3,5009	Administração	0,1091	3,2608
Medicina	-0,4095	3,1624	Economia	0,0002	3,1557
Odontologia	-0,1085	3,3488	Comunicação	0,1804	3,3323
Eng. Civil	0,0101	3,6270	Direito	0,0379	3,0989
Eng.Elétrica	-0,3114	3,8338	Pedagogia	0,0972	3,7283
Eng.Mecânica	-0,2798	3,7180	Psicologia	0,0673	3,3023
Eng.Química	-0,1888	3,5729	Letras	0,1233	3,6662

Tabela 5.1: Coeficientes de Assimetria e Curtose. **Fonte:** Elaborado pelo autor.

Um dos teste de normalidade mais conhecidos é o teste de Jarque-Bera ou *JB test*³⁵. A estatística do teste é função dos coeficientes de assimetria, S , e curtose, K , computados da distribuição observada de retornos. Para o caso de uma distribuição normal, os valores de S e K são, respectivamente, iguais a 0 e 3. A estatística do teste é dada por

$$JB = n \left[\frac{S^2}{6} + \frac{(K-3)^2}{24} \right],$$

com coeficientes de assimetria e curtose sendo calculados, respectivamente, a partir de

$$S = \frac{\hat{\mu}_3}{(\hat{\mu}_2)^{3/2}} \text{ e } K = \frac{\hat{\mu}_4}{(\hat{\mu}_2)^2},$$

onde μ_j , $j = 2, 3, 4$ é o j -ésimo momento de medida central da distribuição teórica, que são estimados por

$$\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^j.$$

Por resultar da soma dos quadrados de duas distribuições assintoticamente *normal-padrão* (independentes), a distribuição assintótica da estatística do teste *JB* é uma chi-quadrado com dois graus de liberdade. Assim, se a estatística do teste for tal que $JB > \chi^2_{1-\alpha}(2)$, então a hipótese nula de normalidade não é aceita, para um nível de significância α .

A principal vantagem do uso do teste *JB* neste trabalho é que ele permite a utilização de uma quantidade grande de observações³⁶. Além do mais, o teste *JB* é mais confiável que um outro teste bastante utilizado, o *teste Kolmogorov-Smirnov*, o qual tende a rejeitar a hipótese nula à medida que cresce o número de observações.

³⁵ Ver Jarque e Bera (1987). Thadewald e Büning (2004) abordam testes alternativos, comparando o poder desses com o do *JB test*.

³⁶ O teste de Shapiro-Wilk e de Shapiro-Francia para normalidade somente podem ser usados para um número restrito de observações: até 2000 observações para o primeiro e até 5000 observações para o último.

Foi utilizado o software *Stata8*® para a realização do teste de normalidade sobre as distribuições dos log dos salários/hora para todas as profissões³⁷. Os resultados encontram-se na Tabela 5.2. Se o valor de *Prob>chi2* for suficientemente baixo, pode-se rejeitar a hipótese de normalidade da distribuição de retornos. Se, ao contrário, *Prob>chi2* for relativamente alto, a hipótese de normalidade não pode ser rejeitada.

Pode-se observar, pela Tabela 5.2, que a hipótese de normalidade do log do salário/hora é rejeitada para todas as distribuições, com exceção de Física, que é significativa ao nível de 0,0327.

Formação	<i>Pr(Skewness)</i>	<i>Pr(Kurtosis)</i>	<i>joint</i>	
			<i>chi2(2)</i>	<i>Prob>chi2</i>
Agronomia	0,000	0,000	86,55	0,0000
Veterinária	0,000	0,000	42,68	0,0000
Biologia	0,000	0,000	100,89	0,0000
Farmácia	0,000	0,000	80,48	0,0000
Medicina	0,000	0,000	499,61	0,0000
Odontologia	0,000	0,000	68,05	0,0000
Eng. Civil	0,645	0,000	121,71	0,0000
Eng.Elétrica	0,000	0,000	213,94	0,0000
Eng.Mecânica	0,000	0,000	178,19	0,0000
Eng.Química	0,000	0,000	36,49	0,0000
Física	0,671	0,010	6,84	0,0327
Matemática	0,075	0,000	99,66	0,0000
Química	0,034	0,000	30,85	0,0000
Administração	0,000	0,000	211,78	0,0000
Economia	0,994	0,001	12,07	0,0024
Comunicação	0,000	0,000	108,99	0,0000
Direito	0,001	0,000	27,44	0,0000
Pedagogia	0,000	0,000	553,98	0,0000
Psicologia	0,006	0,000	37,31	0,0000
Letras	0,000	0,000	365,30	0,0000

Tabela 5.2: Teste de Normalidade de Jarque-Bera. **Fonte:** Elaborado pelo autor.

³⁷ No *Stata*® utiliza-se o teste *Skewness/Kurtosis test* (dado pela função *sktest* do programa), que é equivalente ao teste de Jarque-Bera.

Constata-se, portanto, que a suposição de normalidade das distribuições dos log dos salários, a qual é base para a aplicação de critérios baseados nos dois primeiros momentos amostrais, não é sustentada pelos dados utilizados neste trabalho, o que reforça a utilização de critérios de dominância estocástica para a ordenação de distribuições de salários para indivíduos com educação superior.

6. TESTE ECONOMÉTRICO DE DOMINÂNCIA ESTOCÁSTICA

Como em qualquer outro estudo empírico, a aplicação dos critérios de dominância estocástica está também sujeita a erros amostrais. Mais especificamente, podem ocorrer dois problemas: 1. haver dominância de primeira ordem na população, mas observar-se cruzamentos entre as duas distribuições amostrais; 2. haver cruzamentos na população que não sejam observados nas distribuições amostrais. Portanto, como em qualquer análise estatística, as regras de dominância estocástica estão sujeitas a erros do tipo I ou do tipo II.

Vários métodos têm sido propostos para testar dominância estocástica. Segundo Tse e Zhang (2003), pode-se dividir esses testes em dois grupos. No primeiro grupo, os testes baseiam-se na comparação entre distribuições para um número pré-estabelecido e finito de pontos, *grid points*, do suporte das distribuições. Os trabalhos de Anderson (1996) e de Davidson e Duclos (2000) estão inseridos neste grupo. No segundo grupo, os testes partem de estatísticas de *inf* e *sup* para todo o suporte das distribuições. Os testes de McFadden (1989) e de Kaur et al (1994) seguem esta linha.

Utiliza-se, neste trabalho, o teste descrito em Davidson e Duclos (2000), a partir daqui, teste *DD*. De acordo com Tse e Zhang (2003), que utilizaram experimentos de Monte Carlo para comparar a performance de diferentes testes de dominância estocástica, o teste *DD* é o que apresenta melhor desempenho no que concerne ao *poder* de teste.

6.1 O TESTE DE DAVIDSON E DUCLOS

Seguindo Davidson e Duclos (2000), consideram-se duas distribuições de renda, caracterizadas pelas funções de distribuição acumuladas F_A e F_B , com suporte não-negativo. Seja $D_A^1(x) = F_A(x)$, com $x \geq 0$, e define-se

$$D_A^s(x) = \int_0^x D_A^{(s-1)}(y) dy. \quad (6.1)$$

Define-se, analogamente, $D_B^s(x)$. Pode-se mostrar, por indução, que

$$D_i^s(x) = \frac{1}{(s-1)!} \int_0^x (x-y)^{s-1} dF_i(y), \quad i = A, B. \quad (6.2)$$

Diz-se que a distribuição B domina estocasticamente a distribuição A em ordem s , se $D_A^s(x) \geq D_B^s(x)$, para todo $x \in \square_+$. Para dominância estrita, a desigualdade deve ser estrita em algum intervalo. Desse modo, dominância estocástica de primeira ordem da distribuição B sobre a distribuição A implica em $F_A(x) \geq F_B(x)$, com $x \leq z$, onde z é o limite superior das distribuições³⁸. Isto equivale a afirmar que a proporção de indivíduos com renda abaixo da linha de pobreza, z , é sempre maior sob a distribuição A do que sob a B , sendo válido também para qualquer linha de pobreza z' inferior a z .

A dominância estocástica de segunda ordem da distribuição B sobre a distribuição A , para um limite superior z , implica em $D_A^2(x) \geq D_B^2(x)$. Pela equação 6.2, isto equivale dizer que, para todo $x \leq z$,

$$\int_0^x (x-y) dF_A(y) \geq \int_0^x (x-y) dF_B(y). \quad (6.3)$$

Supõe-se uma amostra com N observações independentes y_i , com $i = 1, 2, \dots, N$, retiradas de uma dada população. Partindo da equação 6.2 e suprimindo-se o índice i , vê-se que um estimador natural para $D^s(x)$ pode ser obtido por³⁹

$$\begin{aligned} \hat{D}^s(x) &= \frac{1}{(s-1)!} \int_0^x (x-y)^{s-1} d\hat{F}(y) \\ &= \frac{1}{N(s-1)!} \sum_{i=1}^N (x-y_i)^{s-1} I(y_i \leq x) \\ &= \frac{1}{N(s-1)!} \sum_{i=1}^N (x-y_i)_+^{s-1}, \end{aligned} \quad (6.4)$$

³⁸ Davidson e Duclos (2000) chamam este limite superior de *linha de pobreza*, nomenclatura comum em estudos de desigualdade.

³⁹ $(x-y)_+$ é utilizado para denotar $\max(x-y, 0)$.

onde \hat{F} denota a função de distribuição acumulada empírica da amostra e $I(\cdot)$ é uma função indicadora igual a 1 quando seu argumento é verdadeiro e 0 caso contrário. Note-se que, para $s=1$, a equação 6.4 simplesmente estima a função distribuição acumulada populacional a partir da distribuição empírica. Para um valor arbitrário para s , este estimador apresenta a conveniente propriedade de ser a soma de variáveis aleatórias identicamente e independentemente distribuídas.

Para testar a igualdade de $D_A^s(x)$ e $D_B^s(x)$, Davidson e Duclos (2000) propuseram a seguinte estatística normalizada

$$T^s(x) = \frac{\hat{D}_A^s(x) - \hat{D}_B^s(x)}{\sqrt{\text{var}(\hat{D}_A^s(x) - \hat{D}_B^s(x))}} \quad (6.5)$$

onde

$$\text{var}(\hat{D}_A^s(x) - \hat{D}_B^s(x)) = \text{var}(\hat{D}_A^s(x)) + \text{var}(\hat{D}_B^s(x)) - 2\text{cov}(\hat{D}_A^s(x), \hat{D}_B^s(x)). \quad (6.6)$$

O seguinte teorema torna possível realizar inferência estatística a partir de $T^s(x)$ ⁴⁰:

Teorema 6.1 *Considere como finitos os momentos de ordem $2s-s$ da distribuição conjunta das populações y^A e y^B . Então, $N^{1/2}(\hat{D}_K^s(x) - D_K^s(x))$ é assintoticamente normal com média zero, para $K = A, B$, e matriz de covariância assintótica dada por $(K, L = A, B)$*

$$\lim_{N \rightarrow \infty} N \text{cov}(\hat{D}_A^s(x), \hat{D}_B^s(x)) = \frac{1}{((s-1)!)^2} E\left((x - y^K)_+^{s-1} (x - y^L)_+^{s-1}\right) - D_K^s(x) D_L^s(x). \quad (6.7)$$

A covariância assintótica pode ser estimada consistentemente de uma maneira *distribution-free*, usando os equivalentes amostrais. Portanto, $D^s(x)$ é estimado por $\hat{D}^s(x)$, e a esperança na equação 6.7 é estimada por

⁴⁰ Teorema e prova estão disponíveis em Davidson e Duclos (2000).

$$\frac{1}{N} \sum_{i=1}^N (x - y_i^K)_+^{s-1} (x - y_i^L)_+^{s-1}. \quad (6.8)$$

Desse modo, substituindo a equação 6.8 e os estimadores $\hat{D}_K^s(x)$ e $\hat{D}_L^s(x)$ na equação 6.7, o estimador da matriz de covariância assintótica, denotado por $\hat{V}_{K,L}^s(x)$, será

$$\hat{V}_{K,L}^s(x) = \frac{1}{N} \left[\frac{1}{((s-1)!)^2} \frac{1}{N} \sum_{i=1}^N (x - y_i^K)_+^{s-1} (x - y_i^L)_+^{s-1} - \hat{D}_K^s(x) \hat{D}_L^s(x) \right]. \quad (6.9)$$

Fazendo $K = L = A$ na equação 6.9 para obter o estimador da variância de $\hat{D}_A^s(x)$, $K = L = B$ para obter o estimador da variância de $\hat{D}_B^s(x)$, e $K = A$ e $L = B$ para obter a covariância entre $\hat{D}_A^s(x)$ e $\hat{D}_B^s(x)$, tem-se que

$$\text{var}(\hat{D}_A^s(x)) = \frac{1}{N} \left[\frac{1}{((s-1)!)^2} \frac{1}{N} \sum_{i=1}^N (x - y_i^A)_+^{2(s-1)} - (\hat{D}_A^s(x))^2 \right], \quad (6.10)$$

$$\text{var}(\hat{D}_B^s(x)) = \frac{1}{N} \left[\frac{1}{((s-1)!)^2} \frac{1}{N} \sum_{i=1}^N (x - y_i^B)_+^{2(s-1)} - (\hat{D}_B^s(x))^2 \right], \quad (6.11)$$

e

$$\text{cov}(\hat{D}_A^s(x), \hat{D}_B^s(x)) = \frac{1}{N} \left[\frac{1}{((s-1)!)^2} \frac{1}{N} \sum_{i=1}^N (x - y_i^A)_+^{(s-1)} (x - y_i^B)_+^{(s-1)} - \hat{D}_A^s(x) \hat{D}_B^s(x) \right]. \quad (6.12)$$

Neste trabalho, fazem-se comparações entre diferentes populações A e B , com amostras independentemente retiradas N_A e N_B . Neste caso, permite-se que $N_A \neq N_B$ e substitui-se N por N_A , na equação 6.10, e por N_B , na equação 6.11. A covariância entre as

duas populações deve ser zero, devendo-se, assim, desconsiderar o último termo na equação 6.6, permanecendo válido o resultado de normalidade assintótica do teorema 6.1.⁴¹

No teste para dominância estocástica, H_0 deveria considerar todo o suporte, o que é empiricamente impossível. O que é possível fazer é testar H_0 para um número finito de valores pré-estabelecidos de x , o que envolve testar hipóteses múltiplas. Utilizando um procedimento de comparações múltiplas⁴², considera-se os valores fixos x_1, x_2, \dots, x_K e suas respectivas estatísticas $T^s(x_i)$, para $i = 1, 2, \dots, K$, pode-se definir as seguintes hipóteses:

1. $H_0 : D_A^s(x_i) = D_B^s(x_i)$, para todo x_i ,
2. $H_A : D_A^s(x_i) \neq D_B^s(x_i)$, para algum x_i ,
3. $H_{A1} : A \succ_s B$, ou seja, A domina B em ordem s ,
4. $H_{A1} : B \succ_s A$, ou seja, B domina A em ordem s .

A hipótese nula global, H_0 , é a interseção lógica de várias hipóteses, uma para cada x_i , a mesma idéia devendo ser estendida para a hipótese alternativa global, H_A . Para controlar pela probabilidade de rejeição da hipótese nula global, Bishop, Formby e Thistle (1992) sugerem a utilização da estatística *studentized maximum modulus*, M_{∞}^K , com graus de liberdade K e ∞ . Denota-se o percentil $1-\alpha$ de M_{∞}^K por $M_{\infty, \alpha}^K$. A partir desta estatística, pode-se definir as seguintes regras de decisão⁴³:

1. Se $|T^s(x_i)| < M_{\infty, \alpha}^K$ para $i = 1, \dots, K$, então aceita-se H_0 .
2. Se $-T^s(x_i) > M_{\infty, \alpha}^K$ para algum i e $T^s(x_i) < M_{\infty, \alpha}^K$ para todo i , então aceita-se H_{A1} .
3. Se $T^s(x_i) > M_{\infty, \alpha}^K$ para algum i e $-T^s(x_i) < M_{\infty, \alpha}^K$ para todo i , então aceita-se H_{A2} .
4. Se $T^s(x_i) > M_{\infty, \alpha}^K$ para algum i e $-T^s(x_i) > M_{\infty, \alpha}^K$ para algum i , então aceita-se H_{A2} .

⁴¹ Tse e Zhang (2003) partem desta mesma situação ao comparar a performance de alguns testes disponíveis na literatura.

⁴² Proposto por Bishop, Formby e Thistle (1992), que utilizaram um teste do tipo *union-intersection*.

⁴³ Este teste, baseado no resultado da distribuição assintótica de $T^s(x_i)$ derivada por Davidson e Duclos (2000) e no procedimento comparações múltiplas proposto por Bishop, Formby e Thistle (1992), é exposto no trabalho de Tse e Zhang (2003).

6.2 RESULTADOS DO TESTE DE DOMINÂNCIA ESTOCÁSTICA

O objetivo desta seção é apresentar um *ranking* das profissões utilizando um dos critério mais gerais para ordenação e comparar este ranking com os obtidos na Seção 3. Para isto, testou-se dominância estocástica de primeira e segunda ordens entre dez profissões escolhidas: Administração, Agronomia, Biologia, Direito, Economia, Eng. Civil, Farmácia, Matemática, Medicina e Pedagogia.

Vale lembrar que dominância estocástica de primeira ordem (*FSD*) implica em dominância de ordens superiores. Porém, como foi explicado na Subseção 1.6, ao utilizar o critério de dominância estocástica de segunda ordem (*SSD*) deve-se supor que os indivíduos são aversos ao risco, o que é uma hipótese bastante razoável. Para estabelecer uma relação de preferências entre as distribuições de salários de duas carreiras A e B, combina-se os dois critérios (*FSD* e *SSD*), seguindo-se os seguintes passos :

1. Primeiramente, testa-se *FSD*. Os possíveis resultados são:

- A domina B, ou seja, todo indivíduo maximizador de utilidade esperada preferirá a carreira A à carreira B; ou
- B domina A, ou seja, todo indivíduo maximizador de utilidade esperada preferirá a carreira B à carreira A; ou ainda
- A e B se cruzam, de modo que nenhuma distribuição mantém um melhor resultado esperado por todo o suporte. Assim, alguns indivíduos maximizadores de utilidade esperada preferirão a carreira A, enquanto outros escolherão a B, e alguns serão indiferentes. Neste caso, supõe-se que os indivíduos são aversos ao risco e segue-se para o passo 2 .

2. Testa-se *SSD* . Os possíveis resultados são:

- A domina B, ou seja, todo indivíduo avesso ao risco preferirá a carreira A à carreira B; ou
- B domina A, ou seja, todo indivíduo avesso ao risco preferirá a carreira B à carreira A; ou ainda

- A e B são não comparáveis, ou seja, alguns indivíduos avessos ao risco preferirão a carreira A, enquanto outros escolherão a B e alguns serão indiferentes. Neste caso, nenhuma conclusão pode ser tomada.

Apesar de todas as vantagens dos critérios de dominância estocástica já abordados anteriormente, tais critérios raramente geram uma ordenação completa das distribuições estudadas. Isto ocorre neste trabalho, onde a ordenação entre as carreiras não é completa, diferentemente do que ocorre ao utilizar-se outros critérios, como os mostrados na Seção 3. Entretanto, apesar de gerar somente uma ordenação parcial, a análise aqui realizada é bastante útil para testar a validade das outras técnicas de ordenação.

Iniciando a análise a partir da comparação visual entre as distribuições, o Anexo 4 apresenta as distribuições amostrais acumuladas dos salários/hora para as dez carreiras⁴⁴. Nota-se que em todos os gráficos é difícil analisar as distâncias entre as distribuições para valores de salários/hora ou muito baixos ou muito altos. Entretanto, a partir dos gráficos pode-se ter uma boa idéia do que será a ordenação. Por exemplo, enquanto Medicina parece dominar todas as outras carreiras, aparentemente Pedagogia é dominada por todas as demais. Alguns pares são de difícil comparação visual, pois suas distribuições amostrais apresentam vários cruzamentos, em especial: *Administração x Farmácia*, *Agronomia x Farmácia*, *Direito x Economia*, *Economia x Agronomia*, *Economia x Farmácia* e *Matemática x Biologia*.

Apesar de intuitivo, a mera comparação visual entre distribuições acumuladas amostrais pode levar o pesquisador a tirar conclusões bastante inadequadas. Como foi comentado na Subseção 6.1, a aplicação dos critérios de dominância estocástica está também sujeita a erros amostrais. Muitas vezes um par de distribuições amostrais apresenta cruzamentos não existentes nas distribuições populacionais, outras vezes ocorre cruzamentos na população que não são observados nas distribuições amostrais. Para evitar

⁴⁴ Como o teste de dominância estocástica demanda uma variável com suporte positivo, utilizou-se aqui as distribuições de frequências acumulada dos salários/hora, e não log dos salários/hora.

tais erros, utiliza-se o teste *DD*, o qual estima os possíveis cruzamentos entre os pares de distribuições (*crossing points*).

Utilizou-se o software *DAD 4.3*.⁴⁵ para testar *FSD* e *SSD* entre as distribuições de salários das dez carreiras⁴⁶. Os resultados podem ser visualizados na Tabela 6.1.

Nota-se que, do total de quarenta e cinco pares possíveis de distribuições, nove pares apresentaram não comparabilidade. Esta *não-comparabilidade* ocorre quando a distribuição que começa dominando nos menores valores do salário/hora não mantém a vantagem advinda dos baixos valores dos salários⁴⁷. Para que se possa tirar alguma conclusão neste caso, deve-se impor restrições adicionais quanto ao modo como os indivíduos enxergam o risco, o que pode causar uma perda de generalidade na ordenação.

Carreira	Adm.	Agron.	Biol.	Dir.	Econ.	E.Civil	Farm.	Mat.	Med.	Ped.
Administ.	✘	< _{SSD}	<i>nc</i>	< _{SSD}	< _{SSD}	< _{FSD}	< _{SSD}	<i>nc</i>	< _{FSD}	> _{FSD}
Agronomia	> _{SSD} ¹	✘	<i>nc</i>	< _{SSD}	< _{SSD}	< _{FSD}	<i>nc</i>	<i>nc</i>	< _{SSD}	> _{FSD}
Biologia	<i>nc</i> ³	<i>nc</i>	✘	< _{FSD}	< _{FSD}	< _{FSD}	< _{SSD}	<i>nc</i>	< _{FSD}	> _{FSD}
Direito	> _{SSD}	> _{SSD}	> _{FSD}	✘	> _{SSD}	< _{FSD}	<i>nc</i>	> _{FSD}	< _{FSD}	> _{FSD}
Economia	> _{SSD}	> _{SSD}	> _{FSD}	< _{SSD}	✘	< _{FSD}	<i>nc</i>	> _{SSD}	< _{FSD}	> _{FSD}
Eng. Civil	> _{FSD}	> _{FSD}	> _{FSD}	> _{FSD}	> _{FSD}	✘	<i>nc</i>	> _{FSD}	< _{SSD}	> _{FSD}
Farmacia	> _{SSD}	<i>nc</i>	> _{SSD}	<i>nc</i>	<i>nc</i>	<i>nc</i>	✘	> _{SSD}	< _{FSD}	> _{SSD}
Matemática	<i>nc</i>	<i>nc</i>	<i>nc</i>	< _{FSD}	< _{SSD}	< _{FSD}	< _{SSD}	✘	< _{FSD}	> _{FSD}
Medicina	> _{FSD}	> _{SSD}	> _{FSD}	> _{FSD}	> _{FSD}	> _{SSD}	> _{FSD}	> _{FSD}	✘	> _{FSD}
Pedagogia	< _{FSD} ²	< _{FSD}	< _{FSD}	< _{FSD}	< _{FSD}	< _{FSD}	< _{SSD}	< _{FSD}	< _{FSD}	✘

Tabela 6.1: Relações de dominância estocástica entre as profissões. **Fonte:** Elaborado pelo autor.

¹ Agronomia domina estocasticamente em segunda ordem Administração.

² Pedagogia é dominada estocasticamente em primeira ordem por Administração.

³ As distribuições não são comparáveis.

Tomando-se a Tabela 6.1, percebe-se que a carreira Medicina domina todas as outras, enquanto Pedagogia é plenamente dominada, o que está de acordo com os gráficos contendo as duas profissões (ver Anexo 4). Nota-se também que, enquanto Direito e

⁴⁵ Programa desenvolvido por Duclos, Araar e Fortin (2004).

⁴⁶ Como foi comentado na Subseção 6.1, o teste utilizado neste trabalho compara as distribuições em um número pré-estabelecido e finito de pontos no suporte das distribuições, chamados de *grid points*. Para este estudo, foi estabelecido um intervalo de 0,25 entre os *grid points*.

⁴⁷ Isto pode ser visualizado na Figura 1.2 da Seção 1.4.

Engenharia Civil são carreiras predominantemente dominantes, Administração, Biologia e Matemática aparecem como carreiras predominantemente dominadas.

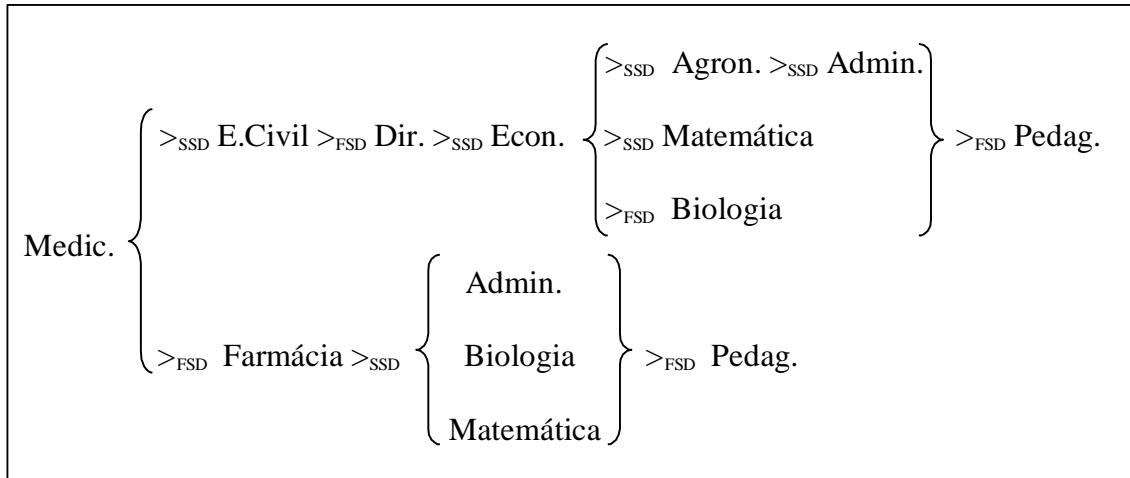


Figura 6.1: Ranking entre as dez carreiras segundo os critérios de *FSD* e *SSD*. **Fonte:** Elaborado pelo autor.

A partir das comparações entre os diferentes pares de distribuições, procurou-se estabelecer um ranking entre os perfis de retorno das diferentes profissões. A Figura 6.1 apresenta as diferentes possibilidades de ordenação das carreiras, onde a presença de parênteses é devido à *não-comparabilidade* em alguns pares de distribuições.

Para testar a adequação de outras técnicas de ordenação, compara-se o *ranking* entre as dez carreiras segundo dado na Figura 6.1 com as ordenações baseadas nos momentos amostrais das distribuições. Adaptando-se as Tabelas 3.1 e 3.2 para considerar somente as dez profissões, constrói-se a Tabela 6.2.

De acordo com a Tabela 6.2, o *ranking* baseado na média amostral é suportado pelo *ranking* baseado na utilidade esperada de Weiss com $\alpha = 0,1$ (medida de aversão relativa ao risco de Arrow). O *ranking* baseado na utilidade esperada com $\alpha = 0,9$ apresenta uma pequena inversão de posição entre Agronomia e Farmácia. Por outro lado, ao utilizar somente o critério de risco (desvio-padrão) a ordenação passa a apresentar uma grande mudança, o mesmo ocorrendo para a utilização do coeficiente de variação como critério.

Formação	Momentos Amostrais			Utilidade Esperada	
	Média	D. Padrão ⁴⁸	Coef. Var.	$\alpha = 0,1$	$\alpha = 0,9$
Medicina	1°	1°	1°	1°	1°
Eng. Civil	2°	2°	7°	2°	2°
Direito	3°	3°	6°	3°	3°
Economia	4°	5°	8°	4°	4°
Agronomia	5°	4°	10°	5°	6°
Farmácia	6°	7°	5°	6°	5°
Administração	7°	6°	9°	7°	7°
Matemática	8°	9°	2°	8°	8°
Biologia	9°	8°	4°	9°	9°
Pedagogia	10°	10°	3°	10°	10°

Tabela 6.2: *Ranking* das formações pelos momentos amostrais e utilidade esperada. **Fonte:** Elaborado pelo autor.

Tomando-se como exemplo a ordenação entre as carreiras de Agronomia e Farmácia, vê-se que a primeira é preferível à segunda pelo critério de salário médio, mas que segundo o critério que considera somente a medida de risco (desvio-padrão), Farmácia passa a ser preferível a Agronomia. Este resultado se mantém ao se utilizar o coeficiente de variação como critério. Considerando o critério de utilidade esperada, vê-se que Agronomia é preferida à Farmácia para $\alpha = 0,1$, mas que ocorre o inverso quando $\alpha = 0,9$. Esta inversão no resultado é devida aos diferentes pesos atribuídos à medida de risco (maior aversão ao risco) na utilidade esperada. Por outro lado, de acordo com a Tabela 6.1, as duas distribuições não são comparáveis segundo o critério de *SSD*, havendo portanto, a possibilidade de se especificar uma função de utilidade factível que leve um indivíduo avesso ao risco a rejeitar ambos os *rankings* baseados na utilidade esperada.

De acordo com o *ranking* sob os critérios de dominância estocástica, a distribuição de salários para Pedagogia é dominada por todas as outras distribuições, o que significa que todo indivíduo avesso ao risco prefere qualquer outra carreira a Pedagogia. Isto mostra a

⁴⁸ *Ranking* pelo valor do desvio-padrão. Portanto, a primeira colocação equivale à profissão com maior desvio-padrão (medido de risco), sendo, portanto, o pior resultado.

inadequação do uso dos critérios de desvio-padrão e coeficiente de variação, segundo os quais Pedagogia domina a maioria das carreiras. Neste caso, a baixa medida de dispersão nos salários desta profissão pode não representar um menor risco (maior segurança), mas uma certeza de salários baixos.

Outros exemplos podem ser tomados, todos convergindo para a utilização do critérios de dominância estocástica de primeiro e segunda ordens. Ao considerar a atratividade das distribuições dos retornos segundo qualquer indivíduo maximizador de utilidade e avesso ao risco, a *FSD* e a *SSD* servem como referência para a aplicação de outros critérios para ordenação de distribuições de salários. Assim, mesmo os critérios de dominância estocástica podendo apresentar *rankings* incompletos, tais critérios apresentam a vantagem de serem gerais o suficiente para não permitir ordenações arbitrárias.

CONCLUSÕES

O objetivo geral deste trabalho foi o de explorar os diferenciais de risco e retorno entre carreiras de nível superior. Isto foi feito de duas formas:

- abordando os diferenciais de retornos médios entre as profissões considerando as diferenças nas características dos indivíduos, como sexo, cor e região, e
- abordando os diferenciais de risco e retorno entre carreiras, considerando todos os indivíduos (sem diferenciá-los segundo suas características individuais).

Assim, primeiramente, constatou-se a existência de uma grande heterogeneidade entre as diferentes carreiras no que concerne as características dos indivíduos em cada uma delas. Viu-se que algumas carreiras, como Biologia, Psicologia, Pedagogia e Letras, apresentam maior participação feminina, enquanto outras são predominantemente masculinas (Agronomia, Veterinária, Física e Engenharias). Além disso, os indivíduos não-brancos escolhem algumas profissões com maior frequência do que outras, como Matemática, Física, Pedagogia e Letras.

Constatou-se também que, com exceção de Odontologia, as carreiras com retornos médios mais altos têm predominância masculina (Medicina, Engenharias e Direito), enquanto que carreiras onde as mulheres são maioria, como Pedagogia, Letras e Biologia, apresentam um retorno bastante inferior. Além disso, as profissões mais escolhidas pelos indivíduos não-brancos são aquelas que apresentam menor retorno médio (Pedagogia, Letras).

A existência de grandes diferenciais de salários por sexo e cor/raça e região reflete-se na ordenação das carreiras pelo retorno médio, que varia de acordo com as características dos indivíduos. Tais diferenças são confirmadas na regressão minceriana da Seção 4, a qual captura os efeitos líquidos dessas variáveis sobre o log do salário/hora. Foi constatado também que há carreiras, como Medicina e Odontologia, onde o salário sobe mais rapidamente com a experiência e chega ao equilíbrio mais rápido do que em outras

carreiras. O diferencial de salários por gênero privilegia os homens em todas as profissões. Este diferencial entre homens e mulheres é consideravelmente maior do que o diferencial de salários a favor dos brancos, salvo para poucas carreiras. Tais resultados reforçam aqueles encontrados nos trabalhos de Barros, Ramos e Santos (1995), onde foi encontrado que a discriminação por gênero como o principal componente dos diferenciais de salários no Brasil, de Strauss e Thomas (1996), onde os retornos sobre a educação é maior para os indivíduos do sexo masculino, e de Silva (1980), onde a taxa de retorno sobre a educação para os brancos é significativamente maior do que para os não-brancos.

Um dos resultados mais claros da estudo foi a existência de um grande diferencial de salários entre regiões para a maioria das carreiras, em especial entre as regiões Nordeste e Sudeste. Para a maior parte das profissões, o fato de se estar na Região Sudeste afeta positivamente os salários. Verificou-se também que a atratividade de uma carreira varia de acordo com a região considerada, por exemplo: os salários médios das carreiras de Direito e Farmácia são maiores no Nordeste do que no Sudeste, o que não ocorre para outras profissões.

Para abordar os diferenciais de risco e retorno entre as profissões considerando todos os indivíduos, compara-se as diferentes carreiras segundo critérios diversos de ordenação das distribuições de retornos. Mostrou-se que é possível ordenar parcialmente as distribuições de retornos de um conjunto de carreiras utilizando os critérios de dominância estocástica de primeira e segunda ordens (*FSD* e *SSD*). Comparou-se os *rankings* obtidos da aplicação de critérios baseados nos momentos amostrais e na utilidade esperada (especificada por Weiss) com o da Figura 6.1, baseado na *FSD* e *SSD*.

Constatou-se que o desvio-padrão e o coeficiente de variação são medidas inadequadas para ordenação de distribuições de retornos, pois contradizem as relações de dominância estabelecidas pelos critérios de *FSD* e *SSD*. Além do mais, ordenações das carreiras a partir do retorno médio e da utilidade esperada baseada em funções de utilidade específicas podem apresentar resultados arbitrários, ou seja, pode estabelecer ordenações que, apesar de completas, diferem entre si. Assim, por ser aceito por todos os indivíduos

avessos ao risco, o *ranking* da Figura 6.1 é o mais adequado ao se considerar indivíduos com diferentes preferências com relação ao risco.

Apesar deste trabalho ser eminentemente exploratório e empírico, algumas conjecturas podem ser feitas no sentido de justificar a posição de algumas carreiras no *ranking* da figura. 6.1. Por exemplo, tomando como parâmetro o grau de substituíbilidade de cada profissão no mercado de trabalho, chega-se a uma justificativa das ótimas colocações das carreiras de Medicina, Eng. Civil e Direito no *ranking*, as quais apresentam pouca ou nenhuma substituíbilidade. O mesmo pode estar ocorrendo com a profissão de Farmácia, onde boa parte do mercado de trabalho é de exercício exclusivo dos indivíduos com esta formação. Quanto à superioridade da carreira de Engenharia Civil sobre a de Direito, apesar desta última contar com o imenso campo de trabalho representado pelo setor público (mais especificamente, pelo Judiciário), é de conhecimento geral que as muitas demandas tecnológicas da economia torna o mercado de trabalho muito fértil para as profissões de engenharia.

Análises mais aprofundadas a respeito da dinâmica dos mercados de trabalho para as diferentes profissões foge do escopo deste trabalho, apesar de se poder utilizar as técnicas e informações aqui contidas para tal objetivo.

Algumas considerações de ordem de política pública podem ser apontadas a partir dos resultados aqui contidos. Baseados nos perfis demográficos das diferentes profissões e no *ranking* das distribuições de retornos, pode-se constatar que as profissões com maior participação relativa de indivíduos não-brancos são as que apresentam piores posições no *ranking*. Assim, uma política efetiva de diminuição das desigualdades entre brancos e não-brancos deveria atribuir uma maior importância às profissões que apresentem distribuições de retornos melhores com melhor colocação no *ranking*, seja por meio de incentivos ou por meio de cotas.

Por fim, propõe-se que esta análise seja aplicada a um número maior de profissões e que se considere amostras de outros anos, como os censos dos anos 1980 e 1991, de modo

observar como as mudanças de cunho econômico e tecnológico afetaram os diferenciais salariais entre os diferentes grupos demográficos e a ordenação entre as diferentes carreiras. Sugere-se também um trabalho futuro que venha a considerar os efeitos dos possíveis problemas econométricos advindos das regressões mincerianas aqui estimadas.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERSON, G. "Nonparametric Tests of Stochastic Dominance in Income Distributions". *Econometrica*, Vol. 64, nº 5, 1183-1193. September, 1996.
- AZZONI, C., SERVO, L. Education, cost of living and regional wage inequality in Brazil. *Papers in Regional Science*, 2001.
- BARROS, R. P. de, RAMOS, L. A note on the temporal evolution of the relationship between wages and education among Brazilian prime-age males —1976/89. In: MENDONÇA, R., URANI, A. (org.). *Estudos sociais e do trabalho*. Rio de Janeiro: IPEA, 1994.
- BARROS, R. P. de, RAMOS, L., SANTOS, E. Gender differences in Brazilian labor market. In: SCHULTZ, P. *Investments in women human capital*. Chicago University Press, 1995.
- BECKER, G. "Human Capital". Columbia University Press, New York, 1964 (edição de 1993).
- BEHRMAN, J. R.; KLETZER, L. G.; MCPHERSON, M. S.; SCHAPIRO, M. O. "The Microeconomics of College Choices, Careers, and Wages: Measuring the Impact of Higher Education". *The Annals of the American Academy of Political and Social Science: The Changing Educational Quality of the Workforce*, 12-23. September 1998.
- BERGER, M. C. "Predicted Future Earnings and Choice of College Major". *Industrial and Labor Relations Review*, Vol. 41, No 3, 418-429. April 1988.
- BISHOP, J.A., J.P.FORMBY, AND P.D. THISLE "Convergence of the South and non-South income distributions, 1969-1979". *American Economic Review* 82, 262-272. 1992
- CAVALIERI, C., FERNANDES, R. Diferenciais de salários por gênero e cor: uma comparação entre as regiões metropolitanas brasileiras. *Revista de Economia Política*, v. 18, n. 1, 1998.
- CESOP - UNICAMP "Análise dos Dados do Exame Nacional de Cursos - Provão". *Projeto Estudos Socioeconômicos em Educação - Relatório Final*. 2002.
- CHEN, S. H. "Is Investing in College Education Risky ?". *Economics Working Paper Archive at WUSTL*, #0202001, 2001.
- CHRISTIANSEN, C. e NIELSEN, H. S. "The Educational Asset Market: A Finance Perspective on Human Capital Investment". The Danish Social Research Foundation, 2002.

- COELHO, A. M., CORSEUIL, C. H. Diferenciais salariais no Brasil: um breve panorama. *Texto para Discussão*. IPEA, n. 898, 2002.
- DAVIDSON, R.; DUCLOS, J. Y. "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality". *Econometrica*, Vol. 68, No 6, 1435-1464, Nov 2000.
- DUCLOS, J. Y. , ABDELKRIM, A. e FORTIN, C. "DAD 4.3: Distributive Analysis user's manual". Université Laval, January 2004.
- EHRENBERG, R. G., SMITH, R. S. "Modern labor economics: theory and public policy". 5th.ed. New York: HarperCollins, 1994.
- EVANS, R.; WEINSTEIN, R. "Ranking Occupations as Risky Income Prospects". *Industrial and Labor Relations Review*, Vol. 35, No 2, January 1982.
- FOGUEL, M., GILL, I., MENDONÇA, R., BARROS, R. P. de The public-private wage gap in Brazil. *Texto para Discussão*. IPEA, n. 754, 2000.
- FREEMAN, R. B. "A Cobweb Model of the Supply and Starting Salary of New Engineers". *Industrial and Labour Relations Review*. Vol.30, No.2, pp.236-48, 1976.
- GUIA ABRIL DO ESTUDANTE, 2000. Disponível no site www.vestibulandoweb.com.br.
- HADAR, J. & RUSSELL, W. R. "Rules for Ordering Uncertain Prospects". *American Economic Review*, Vol. 59, 25-34, 1969.
- HANOCH, G. e LEVY, H. "The Efficiency Analysis of Choices Involving Risk". *Review of Economic Studies*, nº36, 335-346.
- HAUSE, J. C. "The Risk Element in Occupational and Educational Choices: Comment". *Journal of Political Economy*, Vol. 82, No 4, 803-807, 1974.
- HECKMAN, J. "Sample Selection Bias as a Specification Error". *Econometrica*, Vol. 47(1), p.153-161, 1979.
- HECKMAN, J. , LOCHNER, L. e TODD, P. "Fifty Years of Mincer Earnings Regressions", National Bureau of Economic Research, Working Paper nº 9732, May 2003.
- HOFFMANN, R. Income distribution in Brazil and the regional and sectorial contrasts. In: GUILHOTO, J. J. M., HEWINGS, G. J. D. (Org.). *Structure and structural change in the Brazilian Economy*. Ashgate, 2001.
- IBGE - Instituto Brasileiro de Geografia e Estatística. "Documentação dos Microdados da Amostra - Censo Demográfico 2000". Novembro, 2002.

- INEP - Instituto Nacional de Estudos e Pesquisas Educacionais. "Exame Nacional de Cursos: relatório-síntese 2000". Brasília, 2000.
- JARQUE, C. M. and BERA, A.K. "A test for normality of observations and regression residuals", *International Statistical Review*, 55, pp. 163 –172, 1987.
- KASSOUF, A. L. Wage gender discrimination and segmentation in the Brazilian labor market. *Economia Aplicada*, v. 2, 1998.
- KAUR, A., RAO, P. AND SINGH, H. "Testing for second-order stochastic dominance of two distributions". *Econometric Theory* 10, 849-866. 1994
- KING, A. "Occupational Choice, Risk Aversion, and Wealth", *Industrial and Labour Relations Review*. July, 586-596, 1974.
- LANGONI, C. G. *Distribuição da renda e desenvolvimento econômico do Brasil*. Rio de Janeiro: Expressão e Cultura, 1973.
- LEMIEUX, T. "The Mincer Equation Thirty Years after *Schooling, Experience, and Earnings*". Center for Labor Economics/University of California, Working Paper nº62.
- LEVHARI, D. & WEISS, Y. "The Effect of Risk on the Investment in Human Capital". *American Economic Review*, Vol. 64, No. 6, 950-63. December 1974.
- LEVY, H. "Stochastic Dominance and Expected Utility: Survey and Analysis". *Management Science*, Vol. 38, No 4, April 1992.
- LEVY, H. "The efficient analysis of choice involving risk". *Review of Economic Studies*, Vol.26, No.3, 25-34, 1969.
- MACEDO, R. Diferenciais de salários entre empresas privadas e estatais no Brasil. *Revista Brasileira de Economia*, v. 39, n. 4, p. 437-448, 1985.
- MAS-COLELL, A.; WHINSTON, M. D.; GREEN, J. R. "Microeconomic Theory". Oxford University Press, 1995.
- MCCARL, B. A. "Stochastic Dominance Notes". Department of Agricultural Economics, Texas A&M University. July 1996
- MCFADDEN, D. "Testing for Stochastic Dominance". *Studies in the Economics of Uncertainty*, New York: Springer-Verlag, 1989.
- MEC - Ministério da Educação e Cultura, " Lei de Diretrizes e Bases nº 9.394" de 20 de Dezembro de 1996.
- NEAL, D. & ROSEN, S. "Theories of the Distribution of Earnings". *Handbook of Income Distribution*, Vol. 1, 379-427. Edited by A. B. Atkinson and F. Bourguignon, 2000.

- RIPPLE, R. D. "Stochastic Dominance, Major Project Investment Analysis and the Range of Dominance". School of Finance and Business Economics Working Paper Series, Working Paper 98.4, April 1998.
- ROTHSCHILD, M. & STIGLITZ, J. "Increasing Risk I: A Definition". *Journal of Economic Theory*, Vol. 2, No 3, September 1970.
- SAKS, R. & SHORE, S. H. "Risk and Career Choice". January 2003. website: kuznets.fas.harvard.edu/~rsaks/papers/career503.pdf
- SAVEDOFF, W. Wage dynamics in urban Brazil: evidence of regional segmentation or national markets. *Revista de Econometria*, v. 11, n. 2, 1991.
- SENNA, J. Escolaridade, experiência no trabalho e salários no Brasil. *Revista Brasileira de Economia*, v. 30, n. 2, 1976.
- SHORROCKS, A. F. "Ranking Income Distributions". *Economica*, Vol. 50, 3-17, 1983.
- SILVA, N. V. O preço da cor: diferenciais raciais na distribuição da renda no Brasil. *Pesquisa e Planejamento Econômico*, Rio de Janeiro, v. 10, n. 2, 1980.
- SILVERMAN, B. W. "Density Estimation for Statistics and Data Analysis". New York: Chapman and Hall. 1986.
- SLOW, A. "Occupational Choice Under Uncertainty". *Econometrica*, Vol. 52, No.3, May 1984.
- STRAUSS, J., THOMAS, D. Wages, schooling and background: investments in men and women in urban Brazil. *Opportunity foregone: education in Brazil*. Washington: BID, p. 193-214, 1996.
- TANNEN, M. B. New estimates of the returns to schooling in Brazil. *Economics of Education Review*, v. 10, n. 2, p. 123-135, 1991.
- THADEWALD, T. , BÜNING, H. "Jarque-Bera Test and its Competitors for Testing Normality - A Power Comparison".
- THURLOW, L. "Generating Inequality: Mechanisms of Distribution in the U.S. Economy. New York: Basic Books, 1975.
- TSE, Y.K.; ZHANG, X. "A Monte Carlo Investigation of Some Tests for Stochastic Dominance". Monash Econometrics and Business Statistics Working Papers, Working Paper 7/2003, March 2003.
- WEISS, Y. "The Risk Element in Occupational and Educational Choices". *Journal of Political Economy*, Vol. 80, No 6, 1203-13, 1972.

- WHITMORE, G. A. "Third-degree stochastic dominance. *American Economic Review* 60(3), pp. 457-59, 1970.
- WILLIS, R. J. "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions", in Orley Ashenfelter and Richard Layard (eds), *Handbook of Labour Economics*, North Holland: Amsterdam and New York, 1986.
- WOOLDRIDGE, J. M. "Introductory Econometrics: A Modern Approach", South-Western College Pub; 2nd edition, 2002.

ANEXO 1 - Regressão de Salários (Experiência = Idade - Idade Média de Formatura)

ln sh	Total	Agron.	Veter.	Biol.	Farm.	Medic.	Odont.
anosest.	0.3663** (0.0027)	dropped	dropped	dropped	dropped	dropped	dropped
exper	0.0392** (0.0007)	0.0506** (0.0048)	0.0455** (0.0061)	0.0351** (0.0039)	0.0430** (0.0045)	0.0792** (0.0027)	0.0712** (0.0029)
exper2	-0.0006** (0.0000)	-0.0008** (0.0001)	-0.0008** (0.0002)	-0.0006** (0.0001)	-0.0008** (0.0001)	-0.0016** (0.0001)	-0.0016** (0.0001)
sexo	0.2627** (0.0034)	0.2119** (0.0360)	0.2185** (0.0334)	0.2525** (0.0190)	0.1955** (0.0249)	0.1924** (0.0134)	0.1239** (0.0146)
branco	0.1844** (0.0045)	0.1598** (0.0299)	0.0933* (0.0445)	0.1239** (0.0226)	0.1199** (0.0327)	0.1850** (0.0185)	0.1217** (0.0215)
setorpub	0.0281** (0.0041)	-0.0241 (0.0353)	0.0836 (0.0446)	-0.0239 (0.0178)	0.2085** (0.0366)	-0.0690** (0.0175)	0.1015** (0.0249)
norte	-0.0074 (0.0095)	0.0727 (0.0508)	-0.0487 (0.0780)	0.0681 (0.0505)	0.0352 (0.0608)	-0.0999** (0.0338)	0.0803 (0.0417)
nordeste	-0.1653** (0.0049)	-0.1416** (0.0313)	-0.2798** (0.0441)	-0.4000** (0.0240)	-0.0070 (0.0346)	-0.0494** (0.0182)	0.0405 (0.0225)
coeste	0.0555** (0.0063)	0.1100** (0.0354)	0.0130 (0.0477)	-0.0832* (0.0338)	0.1265** (0.0431)	0.0388 (0.0254)	0.1572** (0.0279)
sul	-0.1213** (0.0045)	-0.0500 (0.0280)	-0.0494 (0.0365)	-0.2379** (0.0238)	-0.0474 (0.0295)	-0.0300 (0.0180)	0.1125** (0.0205)
Constant	-3.9649** (0.0417)	1.4529** (0.0545)	1.7009** (0.0611)	1.6202** (0.0326)	1.7585** (0.0425)	2.0813** (0.0268)	1.9089** (0.0285)
Amostra	261618	6469	2998	7573	4538	18282	11492
Adj. R ²	0.1763	0.0733	0.0839	0.1109	0.0854	0.1233	0.0945

Standard errors in parentheses

* significant at 5%; ** significant at 1%

ANEXO 1 (continuação)

ln sh	E. Civil	E. Elét.	E. Mecan.	E. Quím.	Física	Matem.	Química
anosest.	dropped	dropped	dropped	dropped	dropped	dropped	dropped
exper	0.0587** (0.0034)	0.0398** (0.0044)	0.0476** (0.0042)	0.0511** (0.0068)	0.0534** (0.0135)	0.0345** (0.0036)	0.0350** (0.0076)
exper2	-0.0010** (0.0001)	-0.0006** (0.0001)	-0.0009** (0.0001)	-0.0009** (0.0002)	-0.0012** (0.0004)	-0.0007** (0.0001)	-0.0005 (0.0003)
sexo	0.2704** (0.0213)	0.2014** (0.0374)	0.5019** (0.0529)	0.2760** (0.0364)	0.2383** (0.0616)	0.2374** (0.0153)	0.1394** (0.0342)
branco	0.2459** (0.0213)	0.2150** (0.0259)	0.2015** (0.0274)	0.2748** (0.0479)	0.1219 (0.0702)	0.0614** (0.0191)	0.1451** (0.0447)
setorpub	-0.0089 (0.0253)	-0.0571 (0.0409)	-0.0338 (0.0437)	-0.1423* (0.0620)	-0.1968** (0.0629)	-0.0793** (0.0154)	-0.1147** (0.0408)
norte	-0.0238 (0.0390)	-0.0740 (0.0548)	-0.0935 (0.0658)	-0.3949** (0.1090)	-0.2678 (0.1548)	-0.2187** (0.0434)	0.1582 (0.1327)
nordeste	-0.0276 (0.0206)	-0.0368 (0.0331)	-0.1549** (0.0337)	-0.1286** (0.0475)	-0.2052* (0.0881)	-0.4155** (0.0241)	-0.1467** (0.0546)
coeste	0.0950** (0.0288)	0.0567 (0.0403)	0.0490 (0.0628)	0.0170 (0.1108)	0.2746* (0.1162)	-0.0038 (0.0314)	0.0768 (0.0820)
sul	-0.0453* (0.0212)	-0.0090 (0.0293)	-0.0837** (0.0271)	-0.1932** (0.0425)	-0.1599 (0.0832)	-0.2880** (0.0204)	-0.1651** (0.0469)
Constant	1.5552** (0.0349)	1.8236** (0.0498)	1.5018** (0.0628)	1.6700** (0.0671)	1.7477** (0.1241)	1.8093** (0.0282)	1.7359** (0.0656)
Amostra	11877	6644	6899	2391	866	9445	2362
Adj. R ²	0.1173	0.0687	0.0776	0.1218	0.0870	0.1005	0.0650

Standard errors in parentheses

* significant at 5%; ** significant at 1%

ANEXO 1 (continuação)

ln sh	Admin.	Econ.	Comunic.	Direito	Pedag.	Psicol.	Letras
anosest.	dropped	dropped	dropped	dropped	dropped	dropped	dropped
exper	0.0454** (0.0018)	0.0413** (0.0037)	0.0493** (0.0033)	0.0275** (0.0019)	0.0228** (0.0020)	0.0265** (0.0042)	0.0211** (0.0022)
exper2	-0.0008** (0.0001)	-0.0007** (0.0001)	-0.0011** (0.0001)	-0.0004** (0.0001)	-0.0004** (0.0001)	-0.0003 (0.0001)	-0.0002* (0.0001)
sexo	0.3683** (0.0086)	0.3939** (0.0173)	0.2691** (0.0164)	0.2313** (0.0093)	0.1971** (0.0148)	0.0912** (0.0263)	0.1559** (0.0139)
branco	0.2399** (0.0115)	0.2071** (0.0213)	0.2138** (0.0238)	0.2053** (0.0127)	0.0916** (0.0103)	0.2035** (0.0270)	0.0881** (0.0124)
setorpub	0.0525** (0.0138)	0.1152** (0.0241)	-0.0047 (0.0294)	0.3090** (0.0103)	0.0567** (0.0082)	-0.0649** (0.0241)	-0.0098 (0.0100)
norte	-0.0191 (0.0256)	-0.0646 (0.0405)	-0.1170 (0.0602)	0.2508** (0.0268)	-0.0324 (0.0226)	-0.0123 (0.0651)	-0.0856** (0.0292)
nordeste	-0.0946** (0.0134)	-0.2447** (0.0226)	-0.2431** (0.0261)	0.1147** (0.0141)	-0.3102** (0.0118)	-0.2375** (0.0280)	-0.4129** (0.0136)
coeste	0.0795** (0.0169)	0.0830** (0.0282)	0.2702** (0.0355)	0.2222** (0.0164)	-0.0682** (0.0141)	0.1403** (0.0379)	-0.1067** (0.0198)
sul	-0.1532** (0.0112)	-0.2910** (0.0217)	-0.1353** (0.0245)	0.0704** (0.0128)	-0.1985** (0.0110)	0.0343 (0.0267)	-0.2161** (0.0131)
Constant	1.4075** (0.0158)	1.6071** (0.0333)	1.5744** (0.0302)	1.8122** (0.0179)	1.6602** (0.0145)	1.8450** (0.0368)	1.7436** (0.0174)
Amostra	44573	12456	11160	38242	30745	9059	23547
Adj. R ²	0.1099	0.1151	0.0867	0.0757	0.0546	0.0448	0.0757

Standard errors in parentheses

* significant at 5%; ** significant at 1%

ANEXO 2 - Regressão de Salários (Experiência = Idade - Duração da Graduação - 17 anos)

	Total	Agron.	Veter.	Biol.	Farm.	Medic.	Odont.
anosest.	0.4224** (0.0026)	dropped	dropped	dropped	dropped	dropped	dropped
exper	0.0573** (0.0007)	0.0621** (0.0050)	0.0544** (0.0061)	0.0515** (0.0037)	0.0532** (0.0047)	0.0901** (0.0028)	0.0791** (0.0029)
exper2	-0.0010** (0.0000)	-0.0010** (0.0001)	-0.0009** (0.0002)	-0.0008** (0.0001)	-0.0009** (0.0001)	-0.0017** (0.0001)	-0.0017** (0.0001)
sexo	0.2817** (0.0032)	0.2050** (0.0353)	0.2157** (0.0322)	0.2513** (0.0181)	0.1954** (0.0243)	0.1905** (0.0133)	0.1241** (0.0144)
setorpub	0.0200** (0.0038)	-0.0159 (0.0350)	0.0856* (0.0435)	-0.0109 (0.0169)	0.2088** (0.0361)	-0.0619** (0.0174)	0.1159** (0.0244)
branco	0.1849** (0.0042)	0.1615** (0.0296)	0.0919* (0.0434)	0.1141** (0.0214)	0.1160** (0.0320)	0.1824** (0.0184)	0.1216** (0.0212)
norte	-0.0126 (0.0089)	0.0702 (0.0505)	-0.0428 (0.0768)	0.0503 (0.0466)	0.0350 (0.0600)	-0.0965** (0.0335)	0.0779 (0.0413)
nordeste	-0.1759** (0.0046)	-0.1469** (0.0311)	-0.2762** (0.0433)	-0.3977** (0.0228)	-0.0110 (0.0338)	-0.0474** (0.0180)	0.0452* (0.0222)
coeste	0.0412** (0.0058)	0.1023** (0.0346)	0.0161 (0.0458)	-0.0979** (0.0312)	0.1184** (0.0415)	0.0364 (0.0251)	0.1621** (0.0273)
sul	-0.1299** (0.0042)	-0.0579* (0.0276)	-0.0480 (0.0355)	-0.2198** (0.0225)	-0.0573* (0.0284)	-0.0332 (0.0179)	0.1072** (0.0201)
Constant	-5.1641** (0.0401)	1.2417** (0.0585)	1.5408** (0.0635)	1.3124** (0.0362)	1.5886** (0.0463)	1.8885** (0.0287)	1.7755** (0.0288)
Amostra	296093	6632	3117	8461	4755	18585	11901
Adj. R ²	0.2055	0.0874	0.1001	0.1359	0.1038	0.1394	0.1143

Standard errors in parentheses

* significant at 5%; ** significant at 1%

ANEXO 2 (continuação)

	E. Civil	E. Elét.	E. Mecan.	E. Quím.	Física	Matem.	Química
anosest.	dropped	dropped	dropped	dropped	dropped	dropped	dropped
exper	0.0694** (0.0034)	0.0508** (0.0043)	0.0572** (0.0042)	0.0607** (0.0068)	0.0614** (0.0120)	0.0525** (0.0031)	0.0447** (0.0069)
exper2	-0.0010** (0.0001)	-0.0007** (0.0001)	-0.0009** (0.0001)	-0.0010** (0.0002)	-0.0010** (0.0003)	-0.0009** (0.0001)	-0.0006** (0.0002)
sexo	0.2593** (0.0203)	0.1882** (0.0355)	0.4690** (0.0505)	0.2660** (0.0352)	0.2209** (0.0578)	0.2350** (0.0141)	0.1406** (0.0317)
setorpub	-0.0115 (0.0248)	-0.0523 (0.0403)	-0.0341 (0.0430)	-0.1398* (0.0604)	-0.1505* (0.0588)	-0.0475** (0.0141)	-0.1004** (0.0379)
branco	0.2496** (0.0208)	0.2169** (0.0253)	0.2063** (0.0268)	0.2721** (0.0468)	0.0855 (0.0664)	0.0567** (0.0177)	0.1242** (0.0414)
norte	-0.0234 (0.0379)	-0.0992 (0.0532)	-0.0970 (0.0643)	-0.4021** (0.1054)	-0.2990* (0.1513)	-0.2195** (0.0387)	0.0579 (0.1191)
nordeste	-0.0276 (0.0200)	-0.0475 (0.0322)	-0.1548** (0.0330)	-0.1327** (0.0467)	-0.2368** (0.0842)	-0.4135** (0.0223)	-0.1578** (0.0526)
coeste	0.0973** (0.0278)	0.0534 (0.0390)	0.0674 (0.0613)	0.0156 (0.1078)	0.1638 (0.1039)	-0.0189 (0.0283)	0.0806 (0.0757)
sul	-0.0399 (0.0205)	-0.0153 (0.0284)	-0.0807** (0.0263)	-0.1973** (0.0412)	-0.1540* (0.0765)	-0.2733** (0.0186)	-0.1715** (0.0438)
Constant	1.2993** (0.0368)	1.6226** (0.0515)	1.3107** (0.0634)	1.4935** (0.0710)	1.4798** (0.1299)	1.4213** (0.0313)	1.4845** (0.0687)
Amostra	12417	6985	7202	2486	966	11164	2671
Adj. R ²	0.1473	0.0898	0.0964	0.1414	0.1114	0.1353	0.0912

Standard errors in parentheses

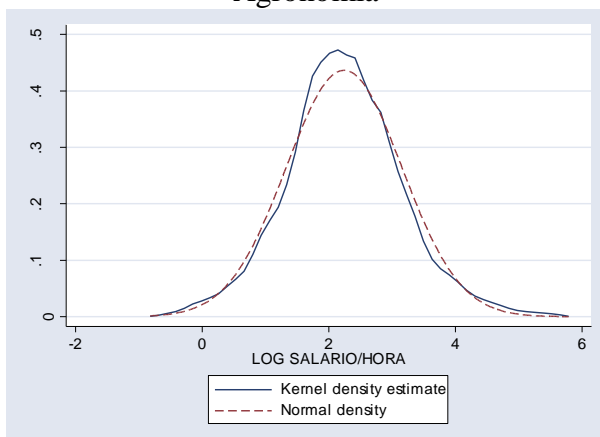
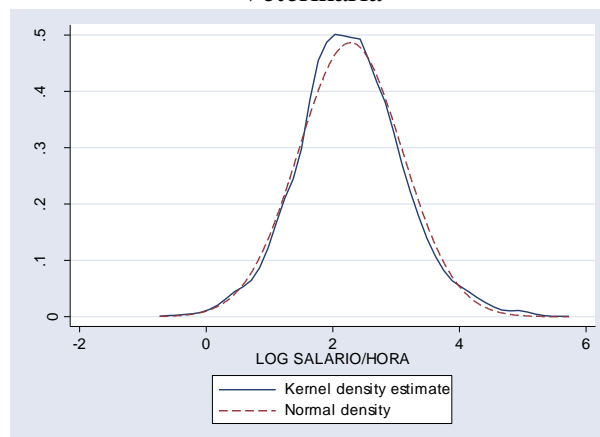
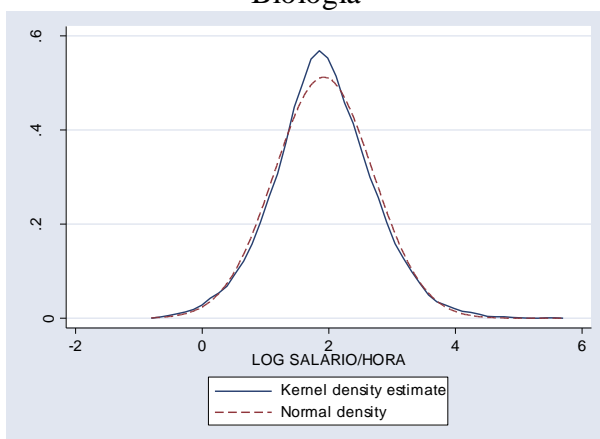
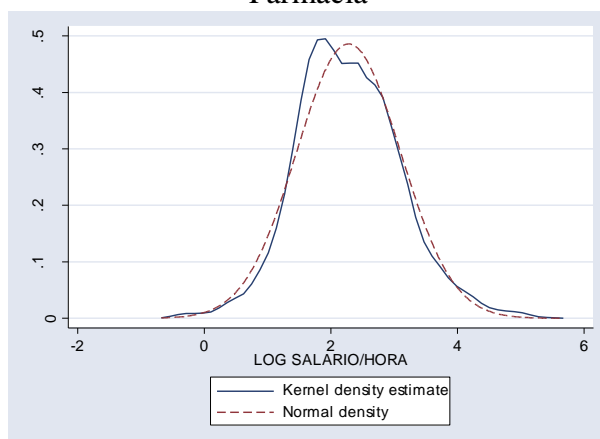
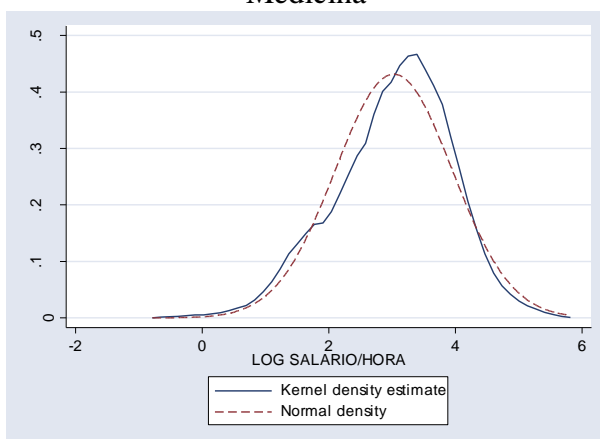
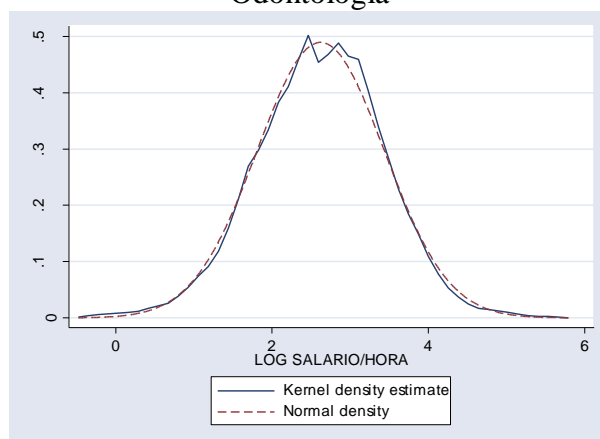
* significant at 5%; ** significant at 1%

ANEXO 2 (continuação)

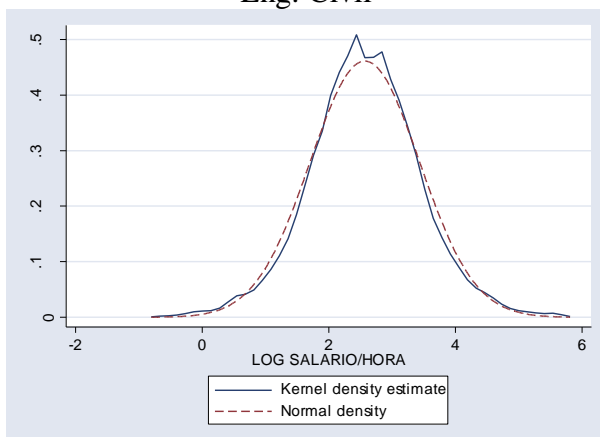
	Admin.	Econ.	Comunic.	Direito	Pedag.	Psicol.	Letras
anosest.	dropped	dropped	dropped	dropped	dropped	dropped	dropped
exper	0.0600** (0.0017)	0.0579** (0.0035)	0.0686** (0.0031)	0.0580** (0.0016)	0.0481** (0.0016)	0.0490** (0.0041)	0.0396** (0.0020)
exper2	-0.0009** (0.0000)	-0.0009** (0.0001)	-0.0014** (0.0001)	-0.0010** (0.0000)	-0.0008** (0.0000)	-0.0007** (0.0001)	-0.0005** (0.0001)
sexo	0.3438** (0.0079)	0.3786** (0.0162)	0.2562** (0.0155)	0.2261** (0.0085)	0.1798** (0.0132)	0.0983** (0.0255)	0.1469** (0.0130)
setorpub	0.0483** (0.0132)	0.1190** (0.0232)	-0.0038 (0.0285)	0.3276** (0.0096)	0.0720** (0.0072)	-0.0567* (0.0233)	0.0060 (0.0093)
branco	0.2376** (0.0107)	0.2064** (0.0203)	0.2083** (0.0225)	0.1963** (0.0118)	0.0922** (0.0090)	0.1961** (0.0256)	0.0880** (0.0114)
norte	-0.0157 (0.0242)	-0.0636 (0.0386)	-0.1096 (0.0574)	0.2310** (0.0247)	-0.0340 (0.0193)	-0.0171 (0.0625)	-0.0779** (0.0264)
nordeste	-0.1053** (0.0126)	-0.2571** (0.0216)	-0.2493** (0.0246)	0.1130** (0.0130)	-0.3185** (0.0102)	-0.2274** (0.0266)	-0.4066** (0.0126)
coeste	0.0514** (0.0157)	0.0665* (0.0267)	0.2616** (0.0333)	0.2043** (0.0150)	-0.0553** (0.0123)	0.1312** (0.0354)	-0.0957** (0.0176)
sul	-0.1628** (0.0105)	-0.3078** (0.0204)	-0.1387** (0.0229)	0.0500** (0.0117)	-0.1805** (0.0095)	0.0294 (0.0249)	-0.2113** (0.0122)
Constant	1.0989** (0.0167)	1.2505** (0.0377)	1.2813** (0.0311)	1.3924** (0.0177)	1.1891** (0.0159)	1.4801** (0.0417)	1.4008** (0.0201)
Amostra	50251	13575	12377	44698	40118	10105	27627
Adj. R ²	0.1434	0.1460	0.1203	0.1299	0.1031	0.0702	0.0997

Standard errors in parentheses

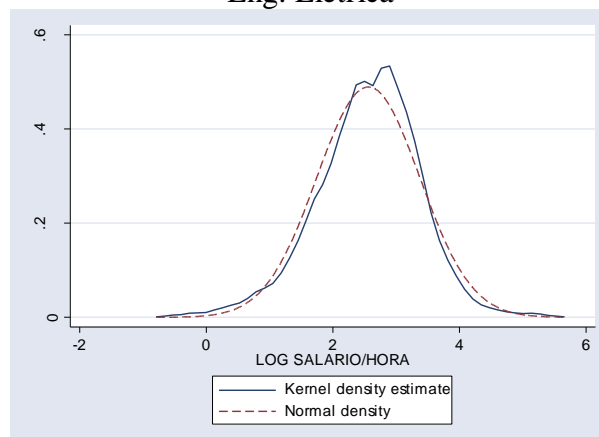
* significant at 5%; ** significant at 1%

ANEXO 3: Densidades Estimadas**Agronomia****Veterinária****Biologia****Farmácia****Medicina****Odontologia**

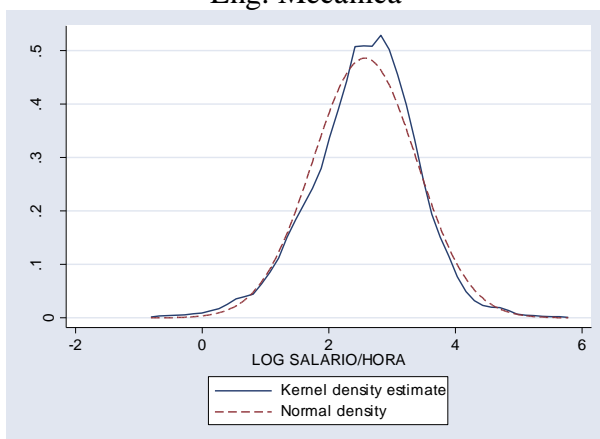
Eng. Civil



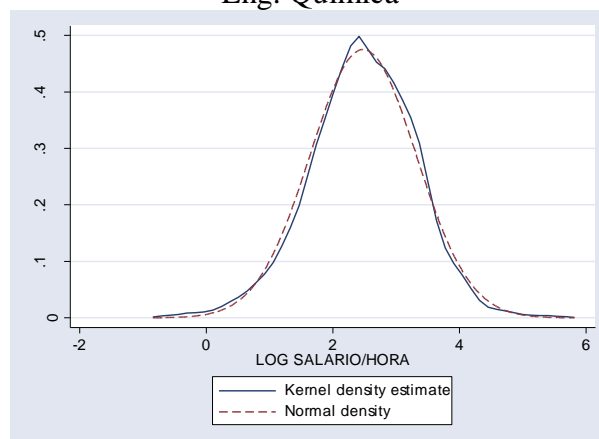
Eng. Elétrica



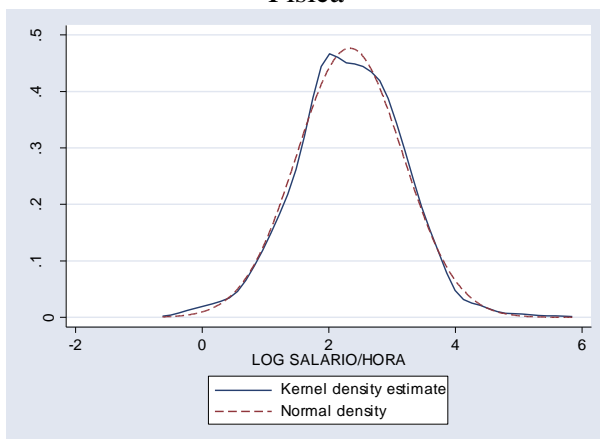
Eng. Mecânica



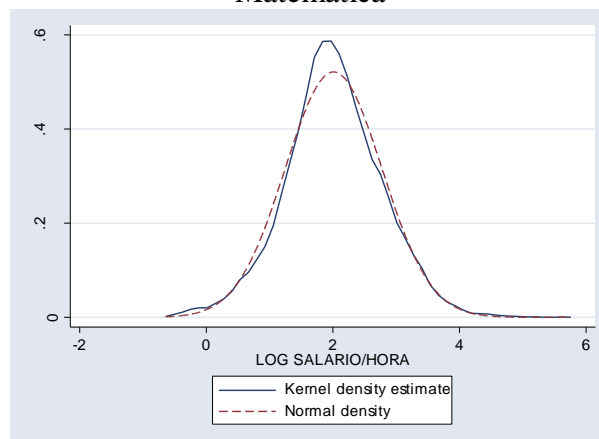
Eng. Química



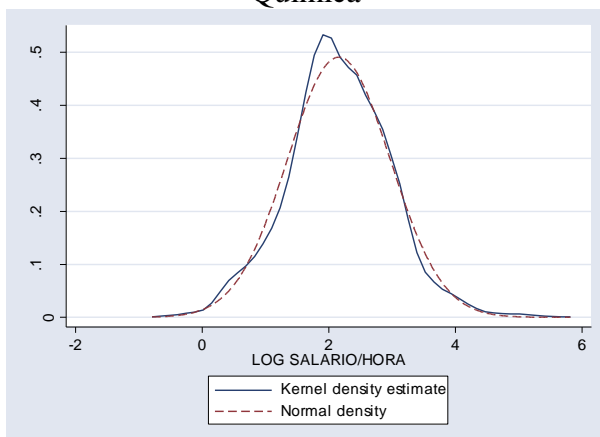
Física



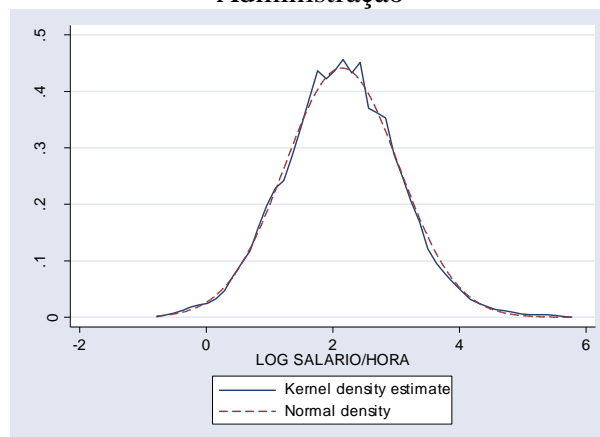
Matemática



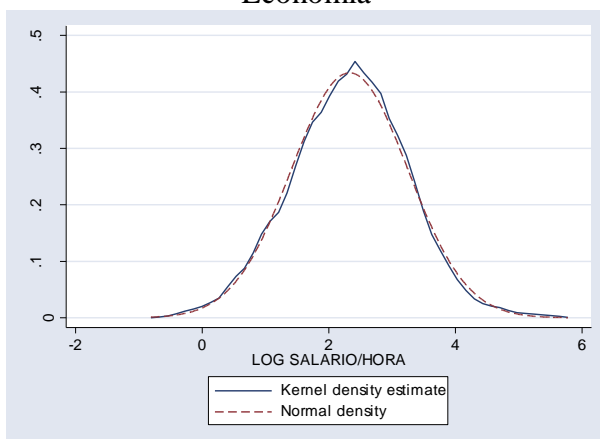
Química



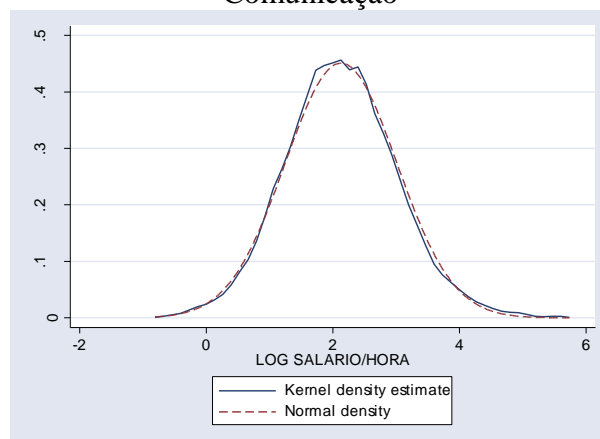
Administração



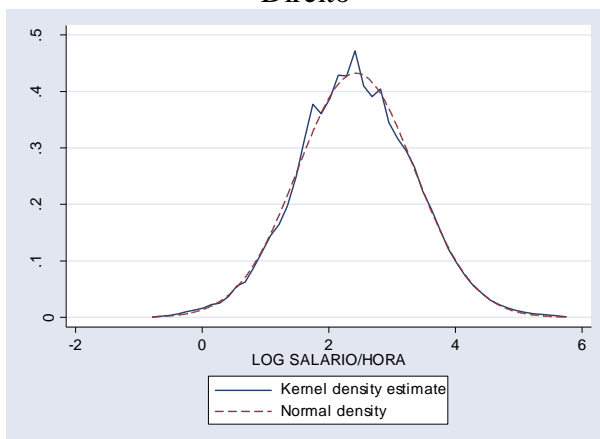
Economia



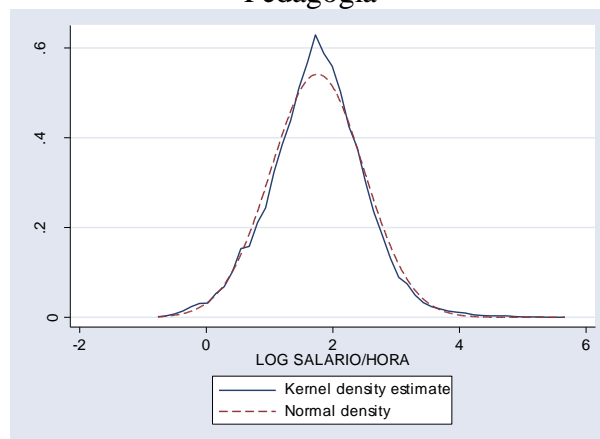
Comunicação



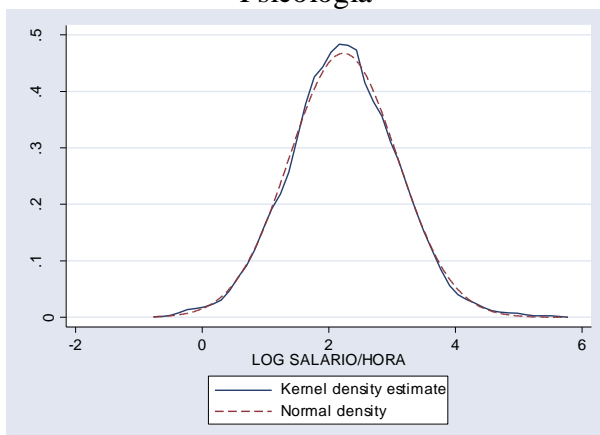
Direito



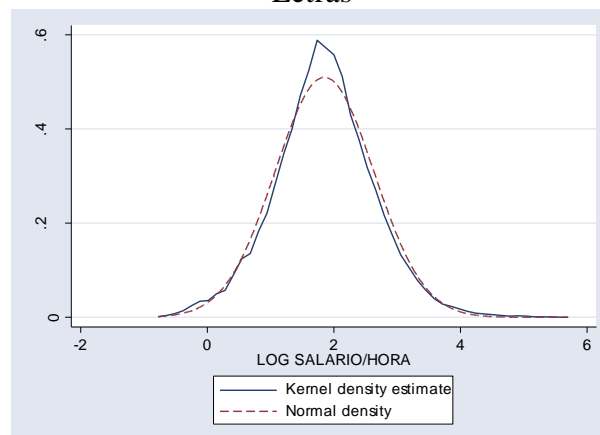
Pedagogia

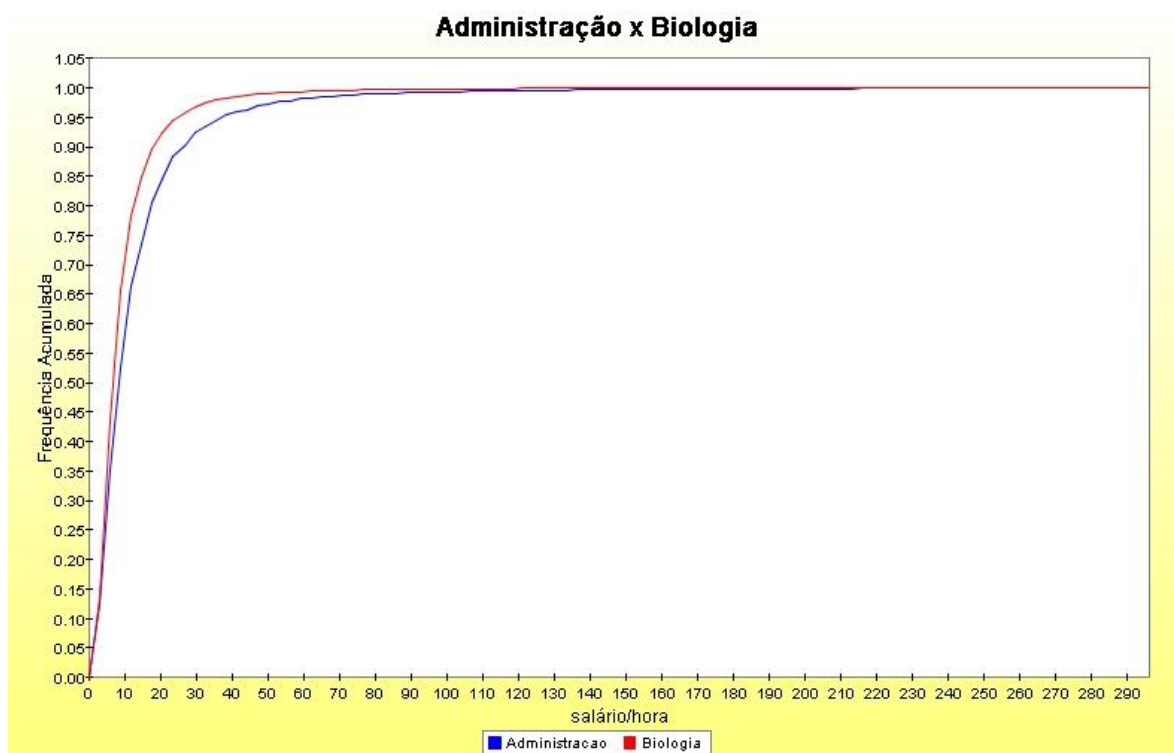
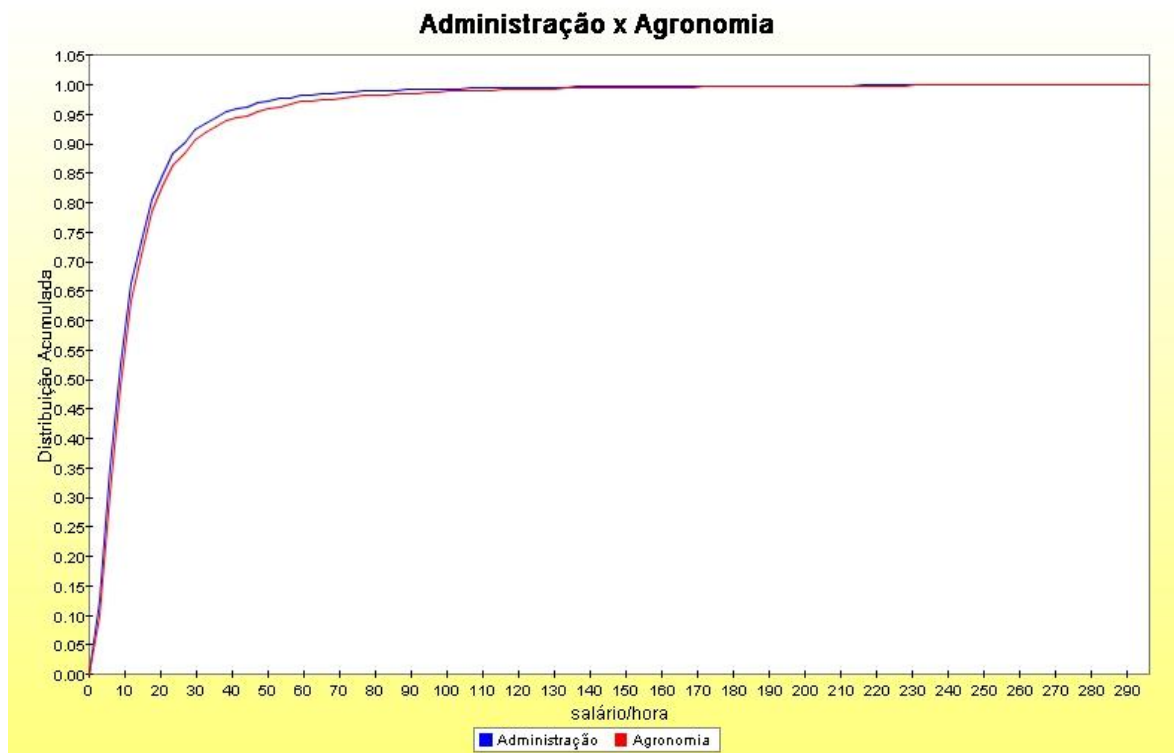


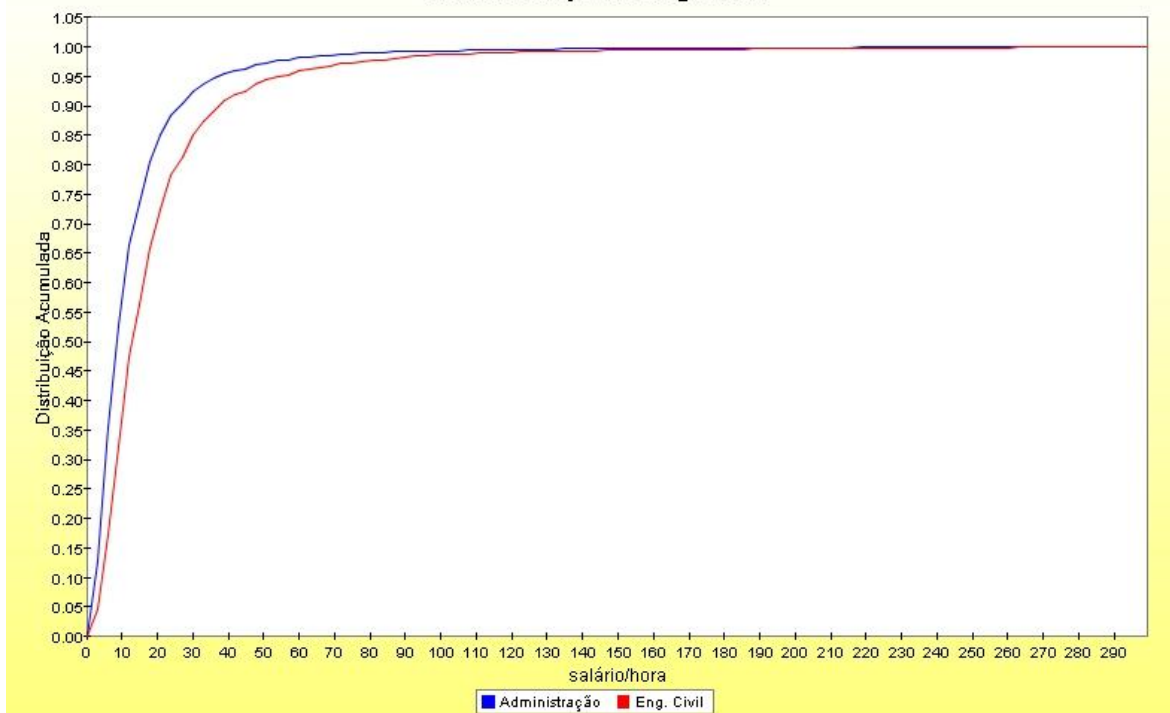
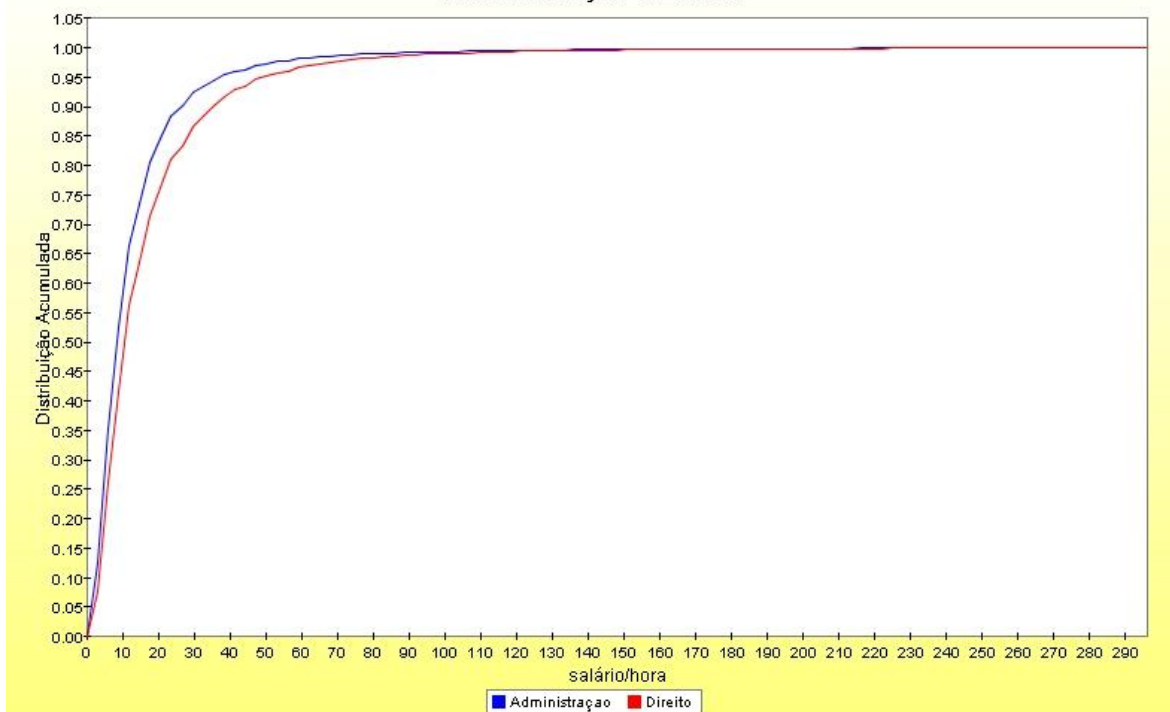
Psicologia

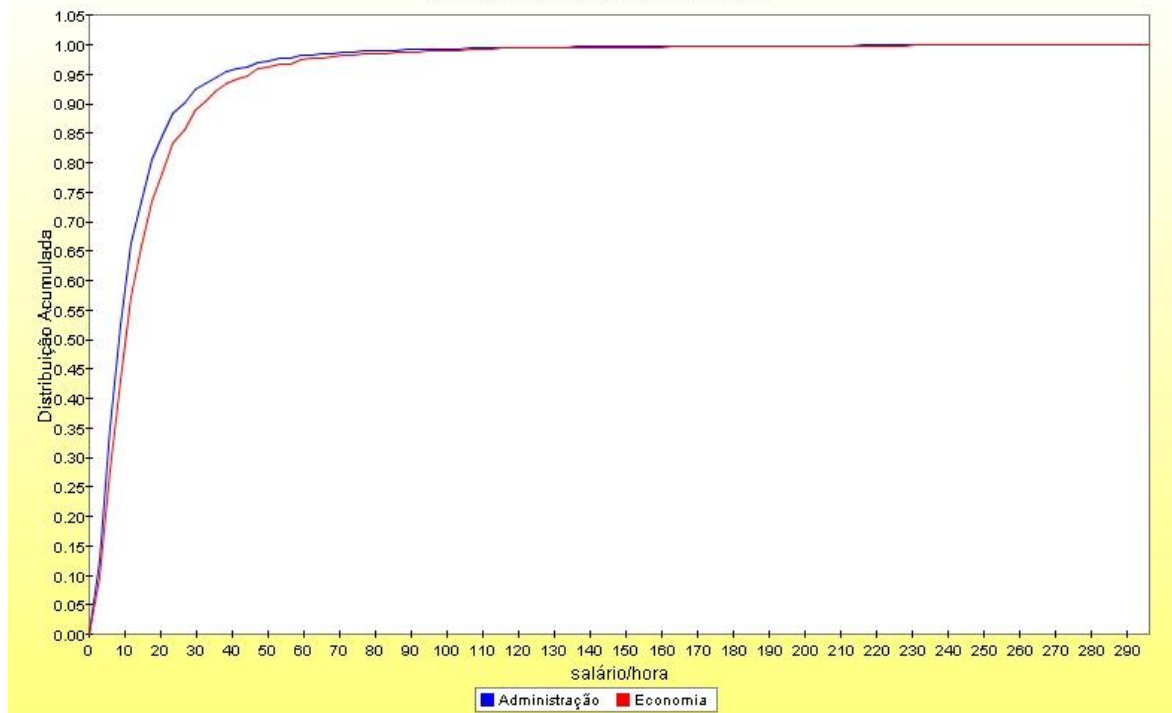
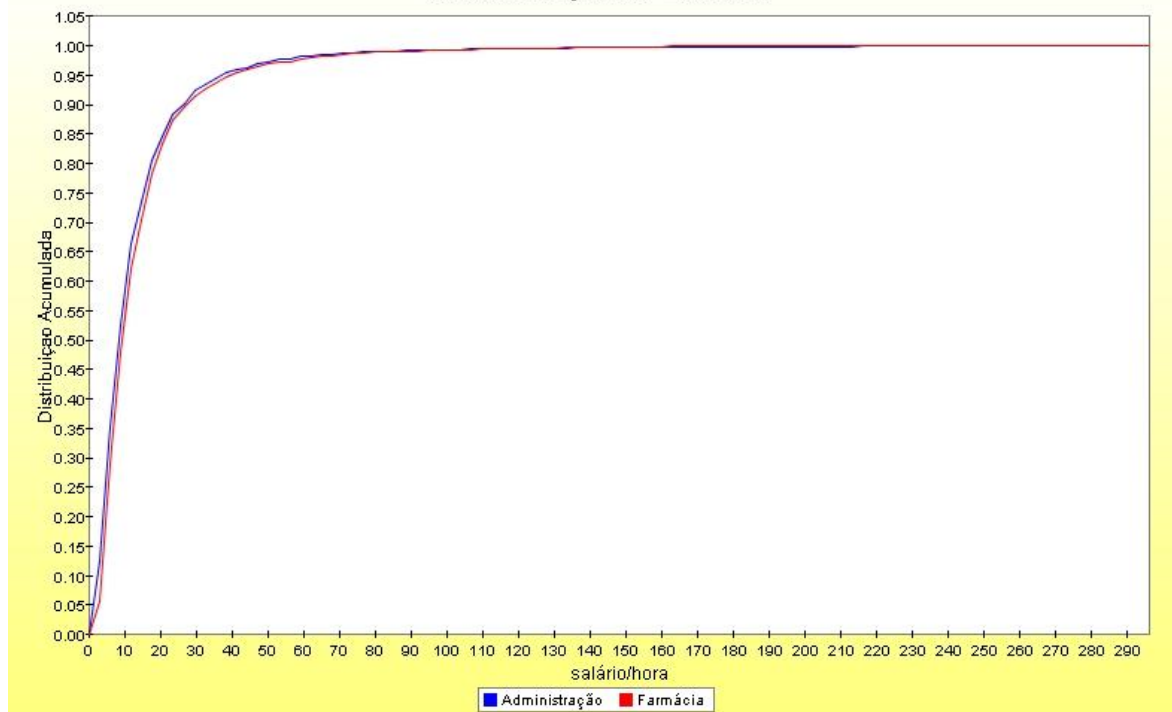


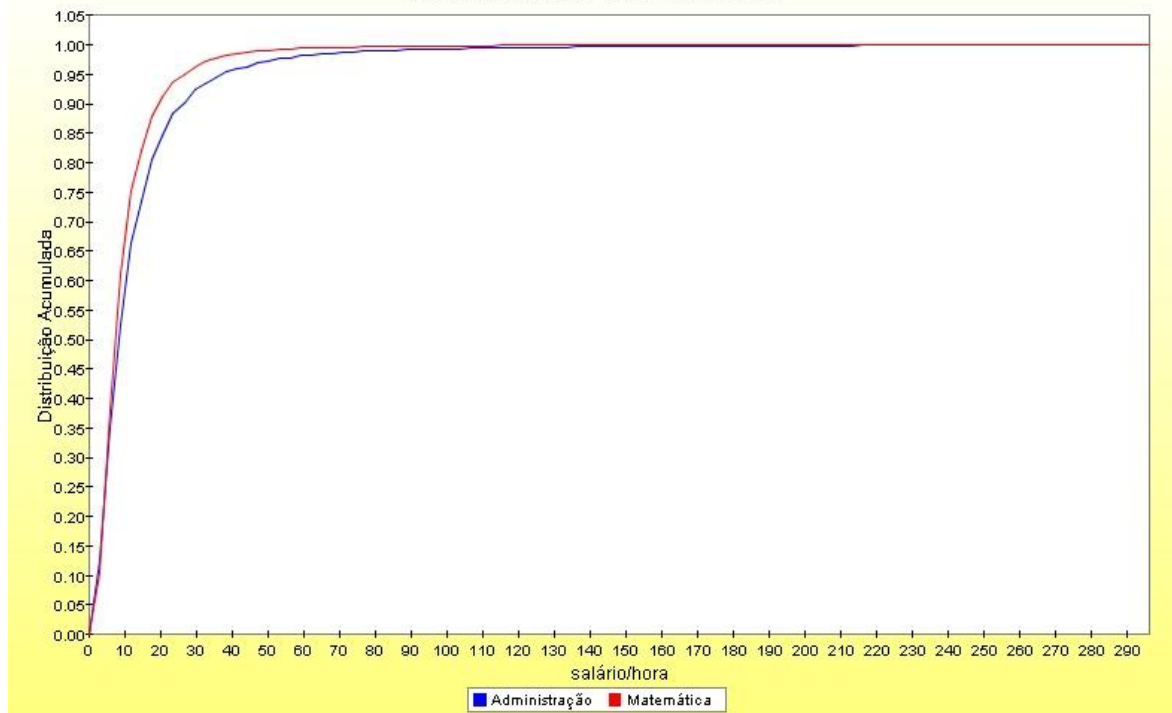
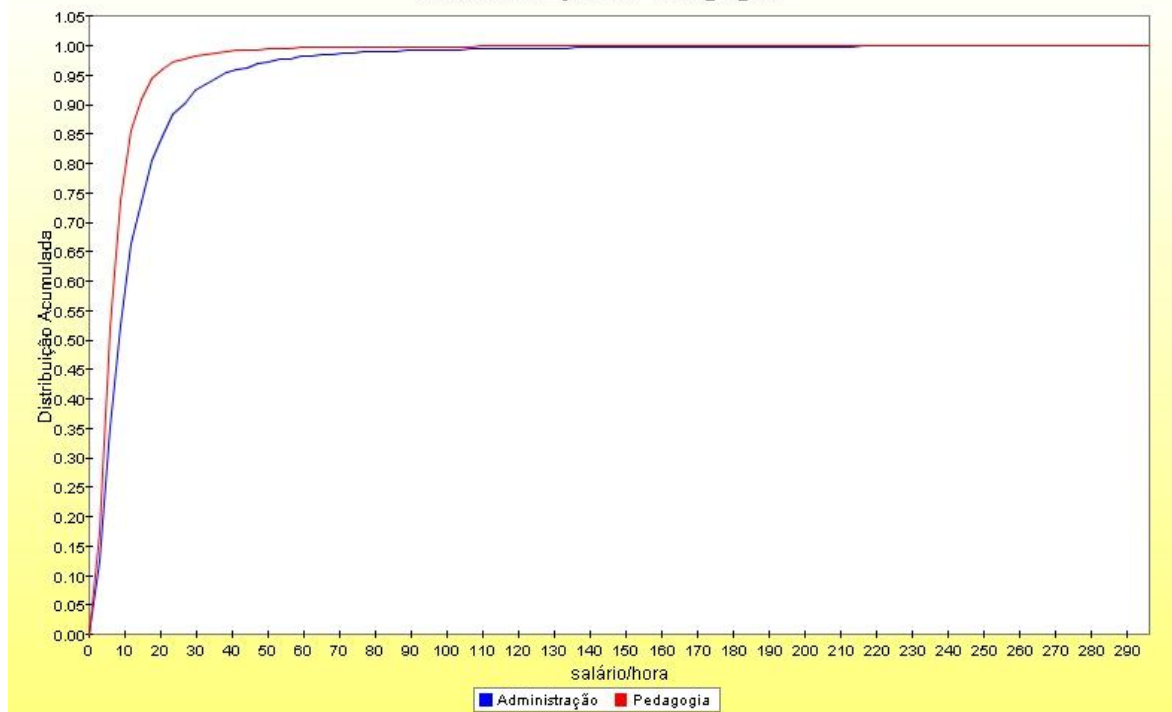
Letras

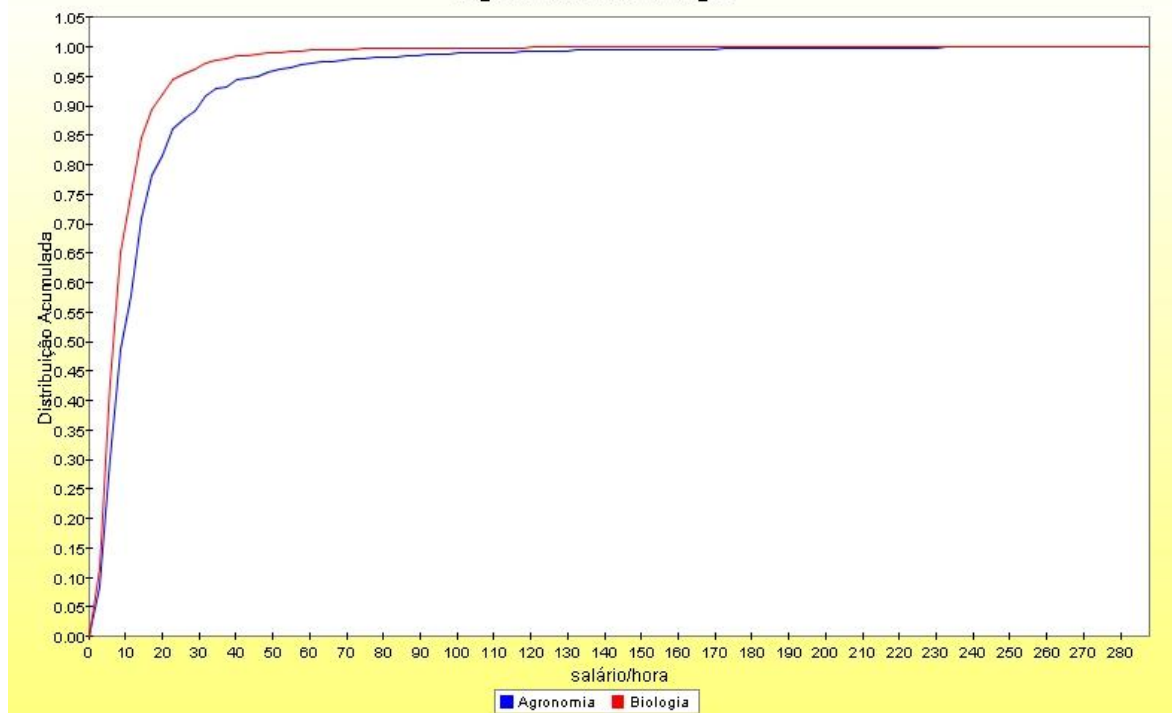
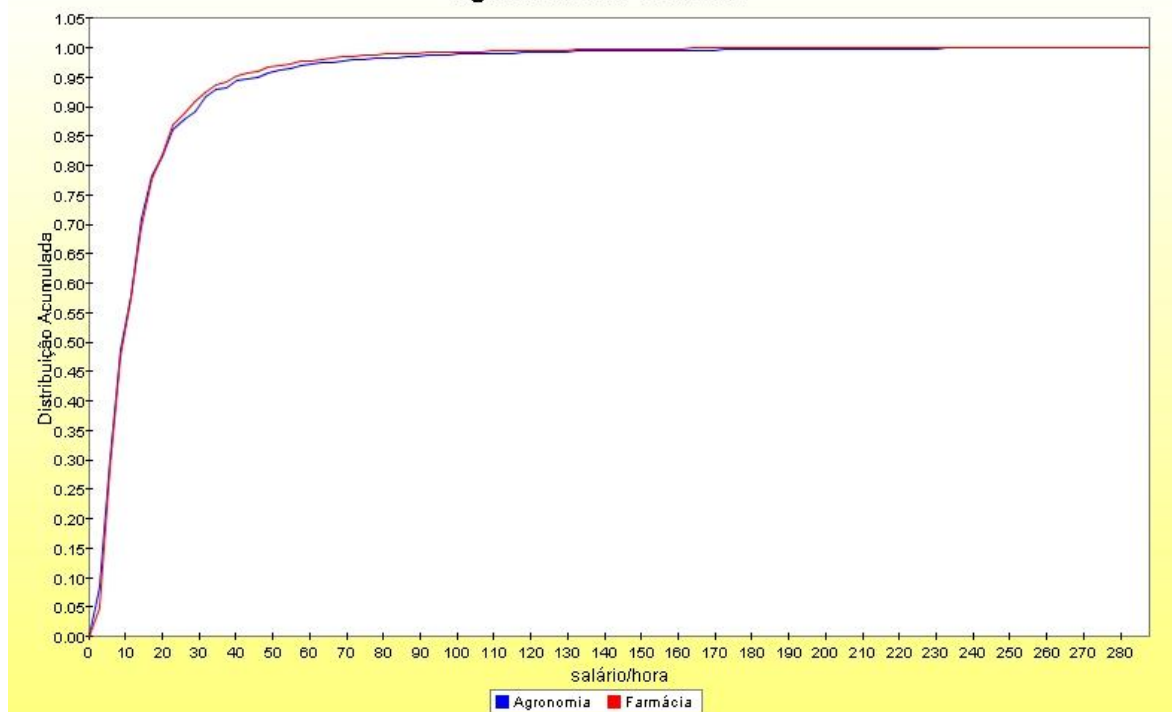


ANEXO 4: Distribuições Amostrais Acumuladas

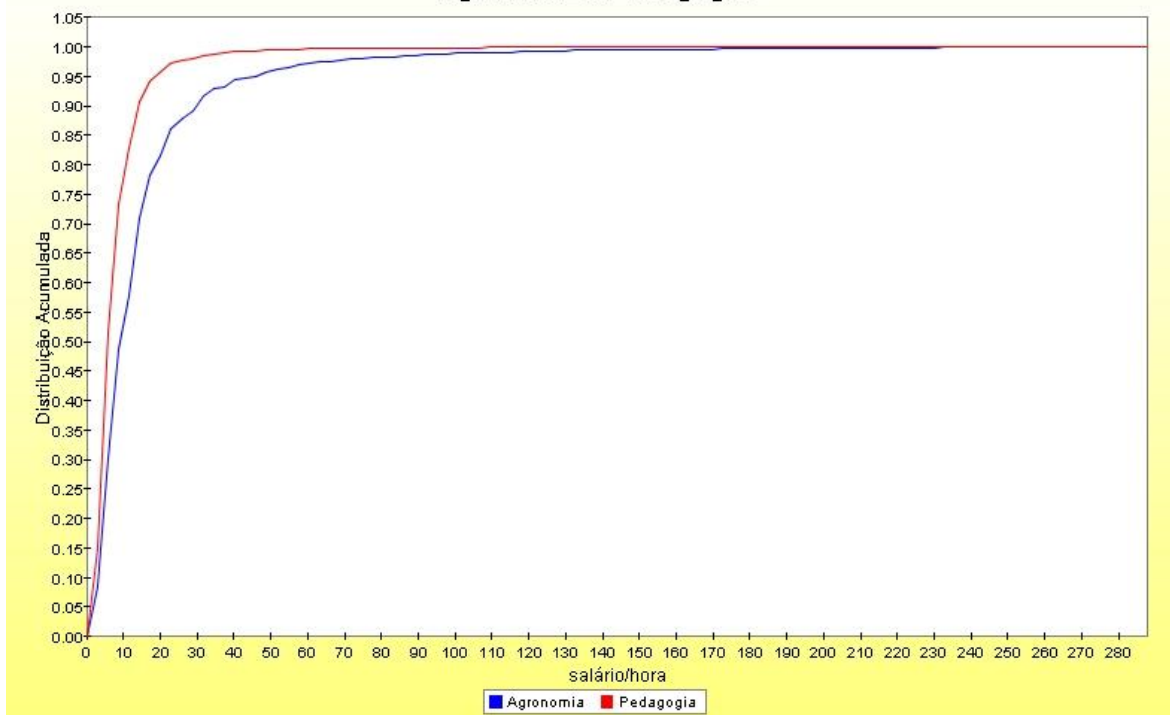
Administração x Eng. Civil**Administração x Direito**

Administração x Economia**Administração x Farmácia**

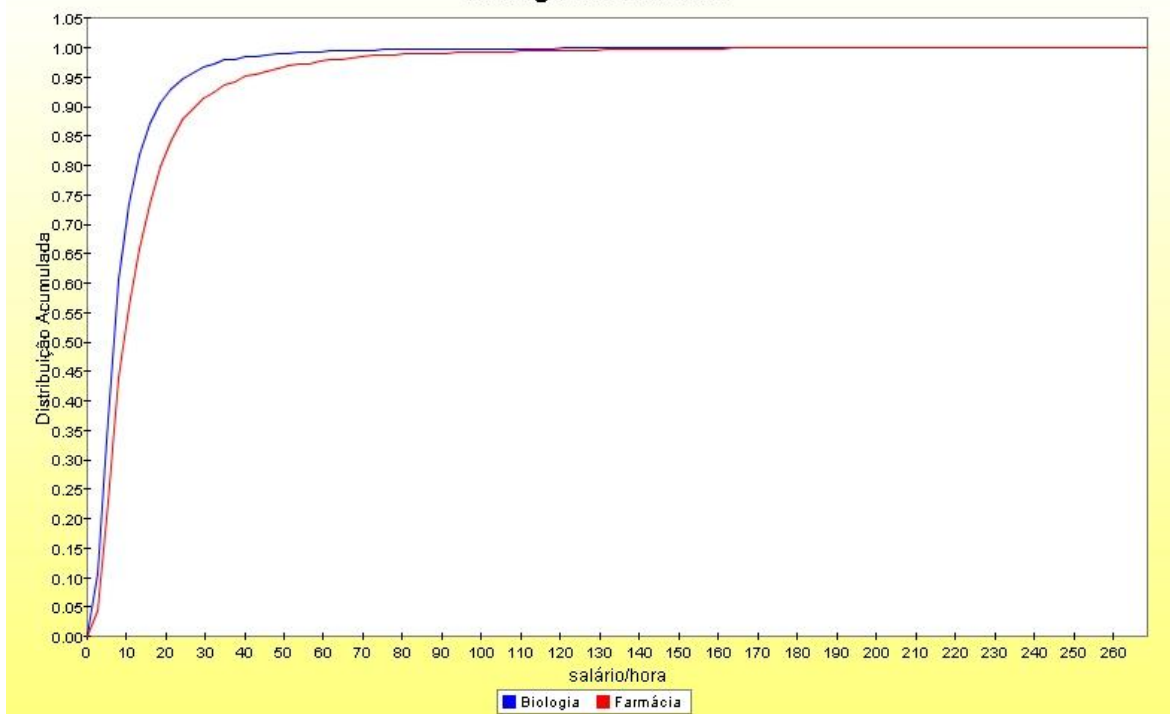
Administração x Matemática**Administração x Pedagogia**

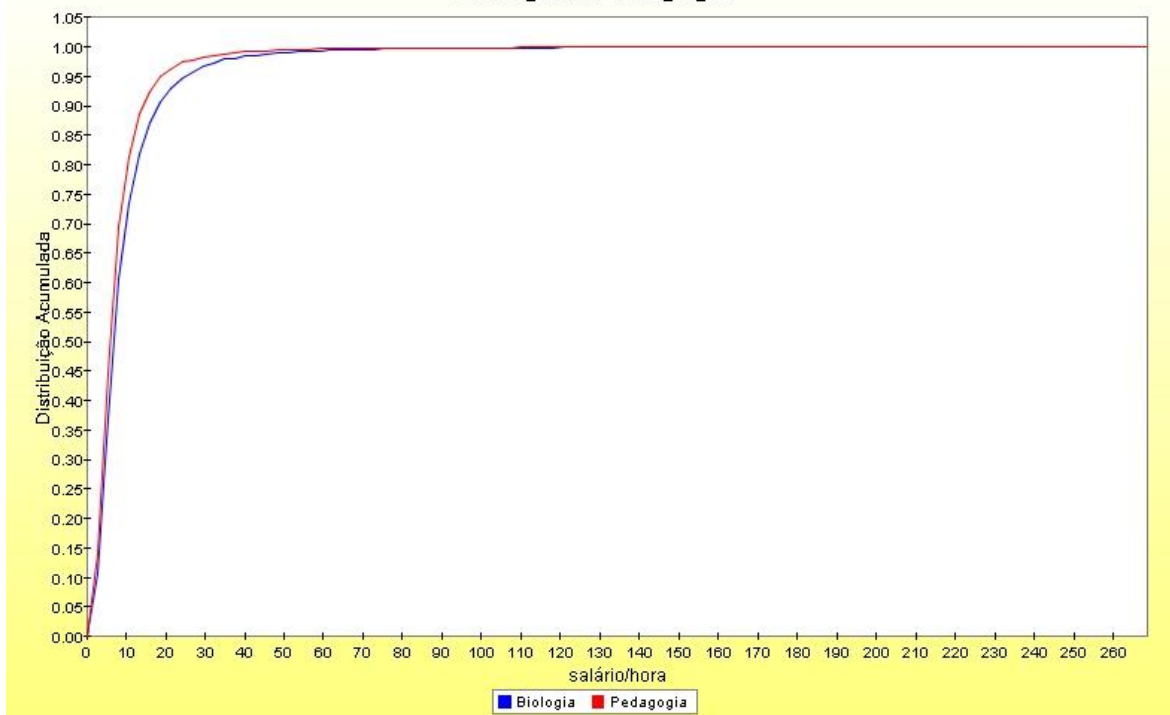
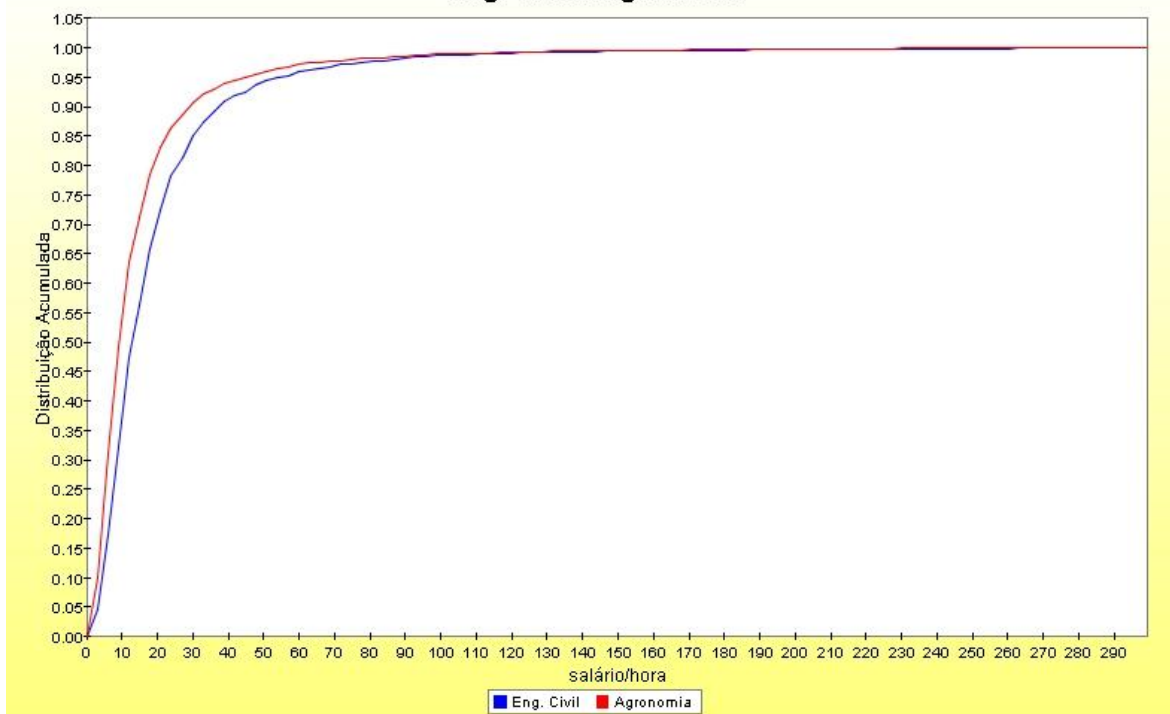
Agronomia x Biologia**Agronomia x Farmácia**

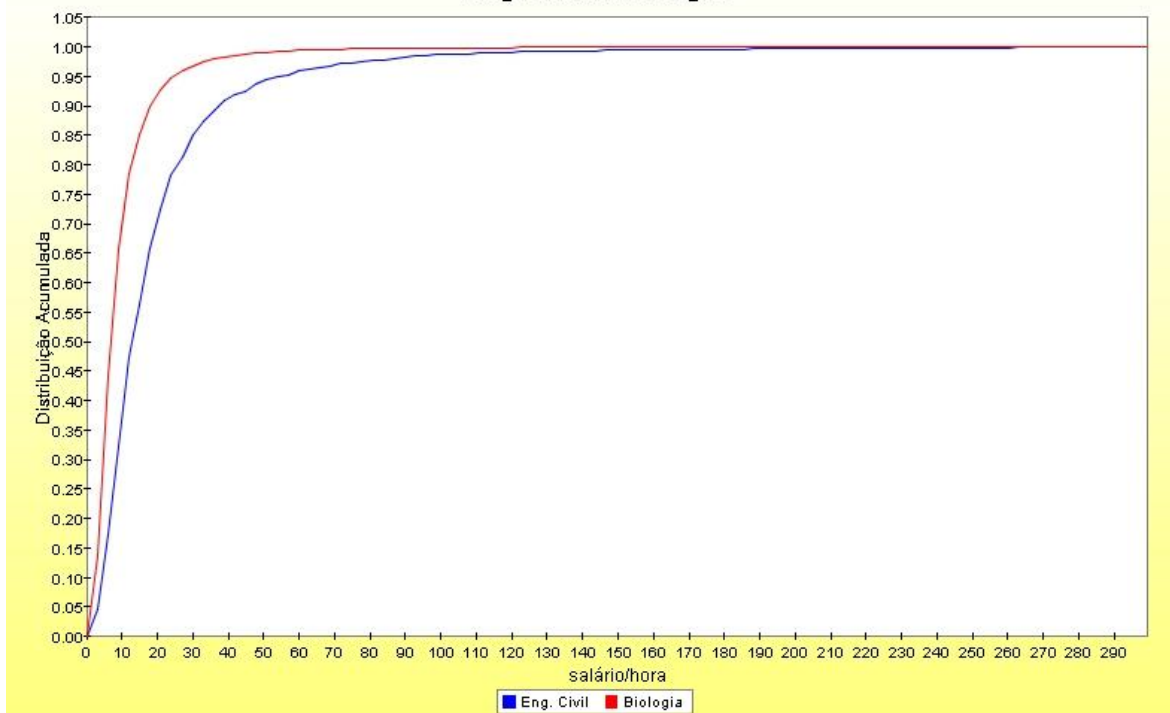
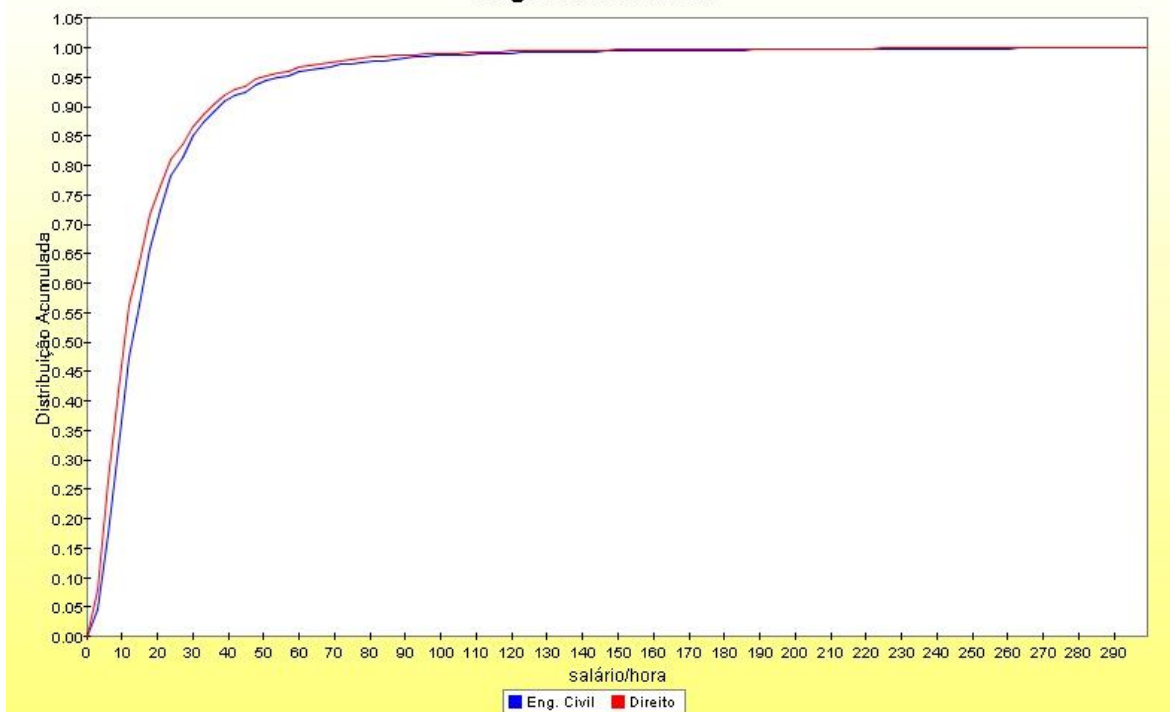
Agronomia x Pedagogia

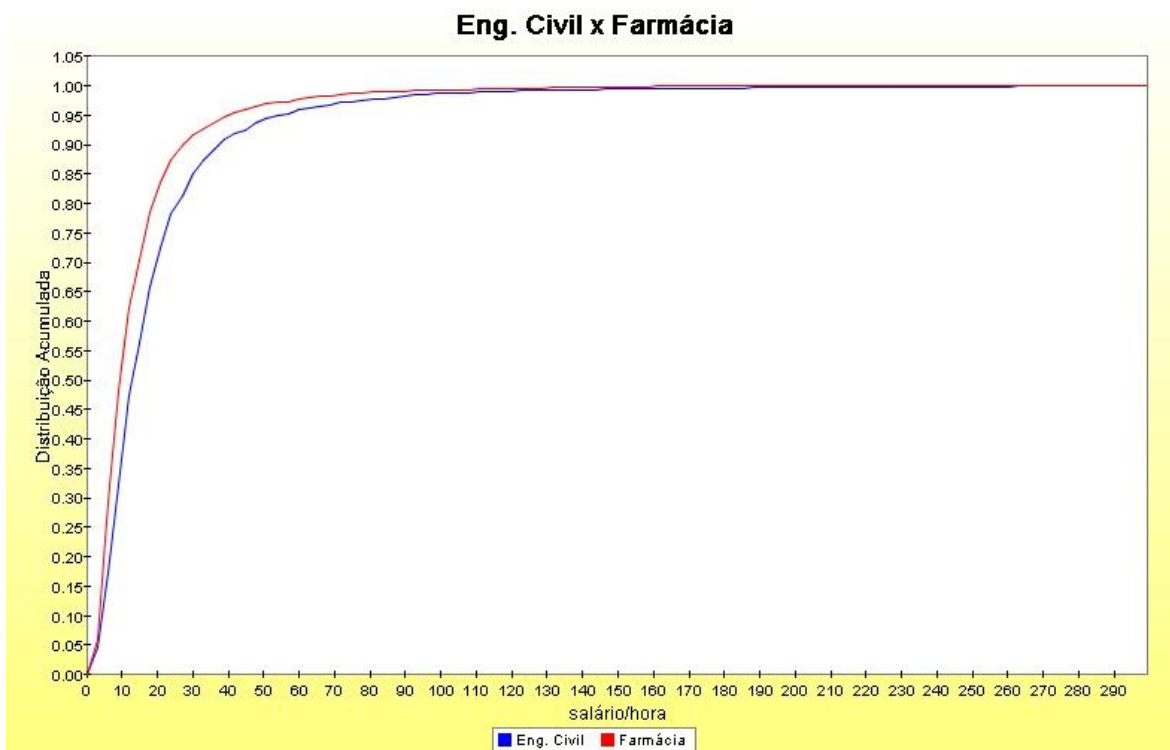
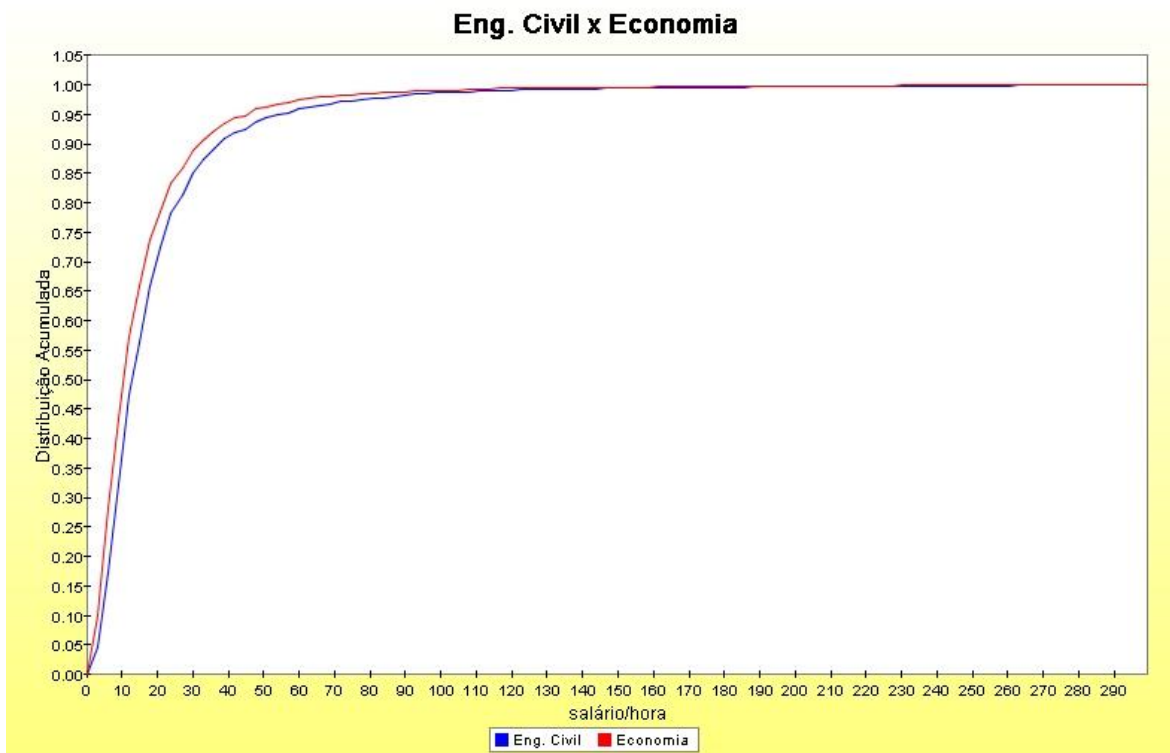


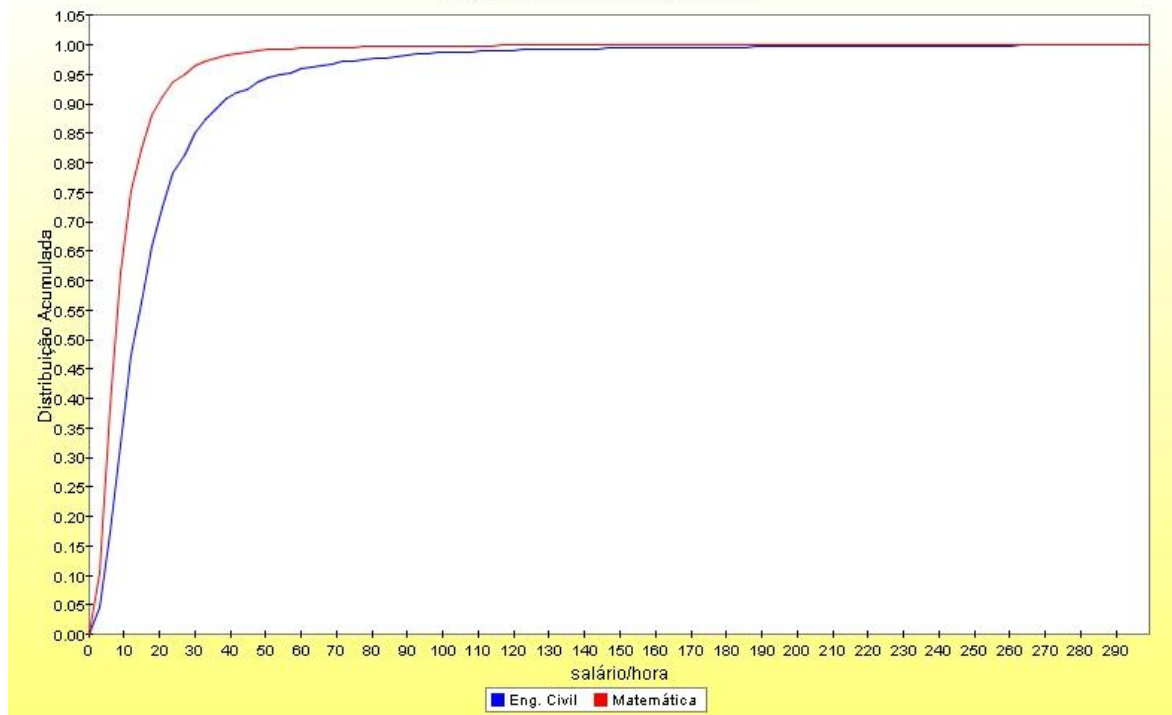
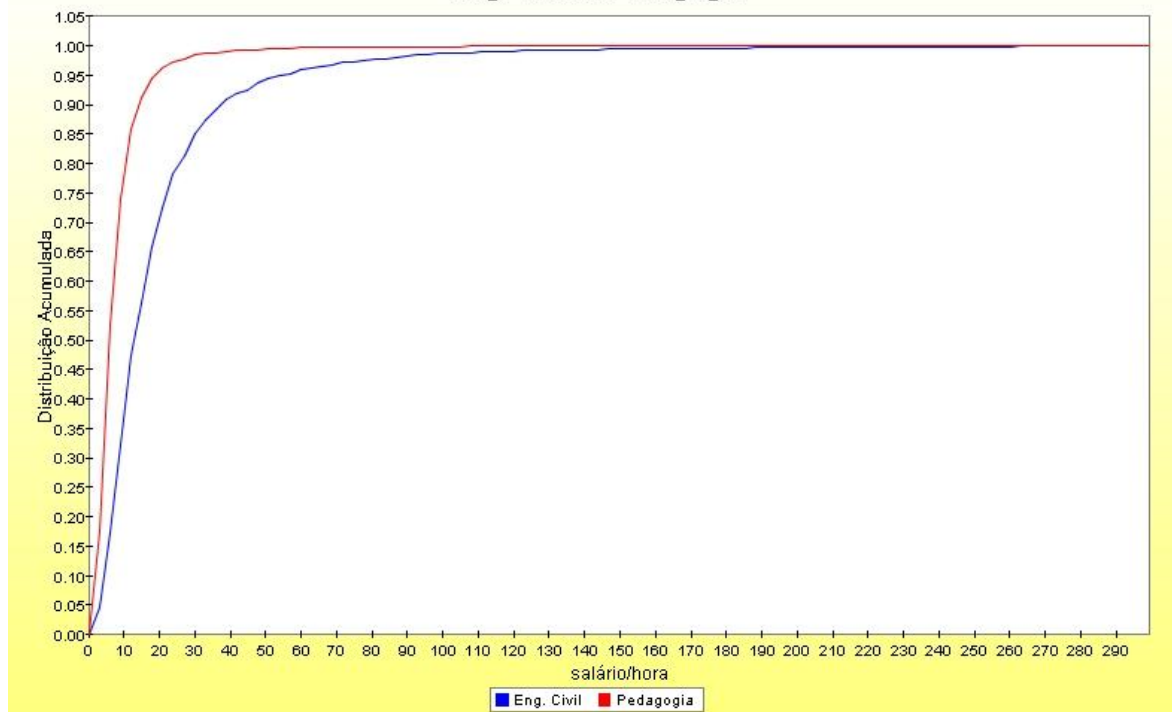
Biologia x Farmácia

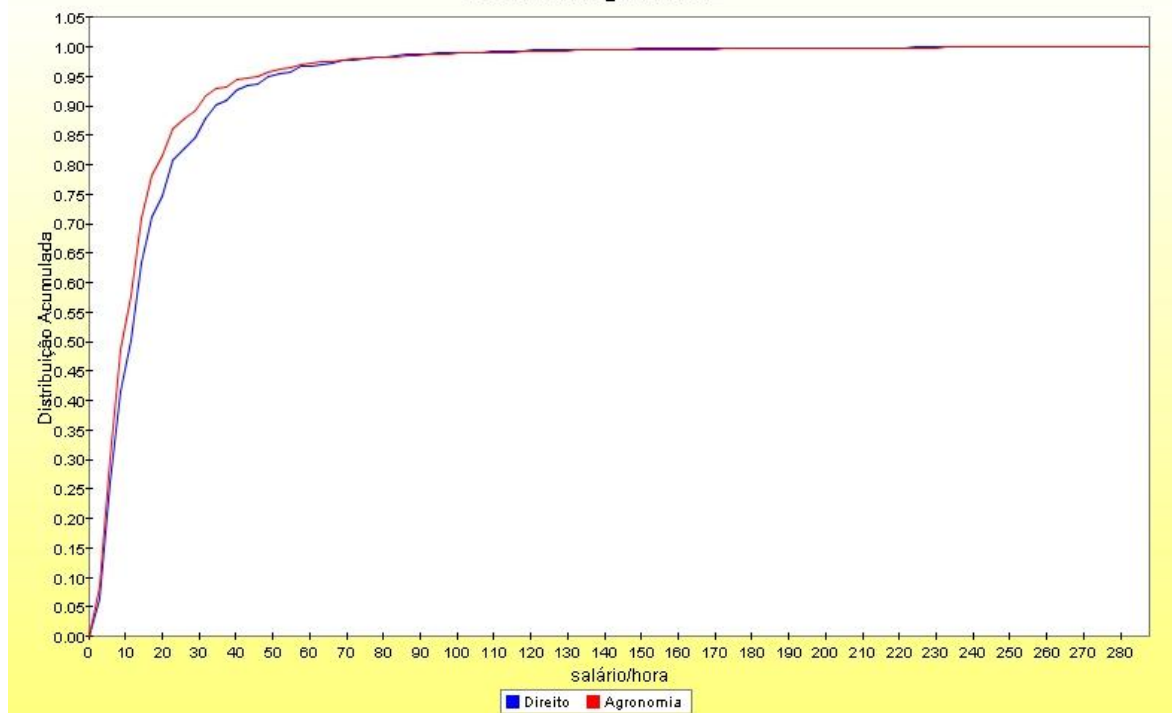
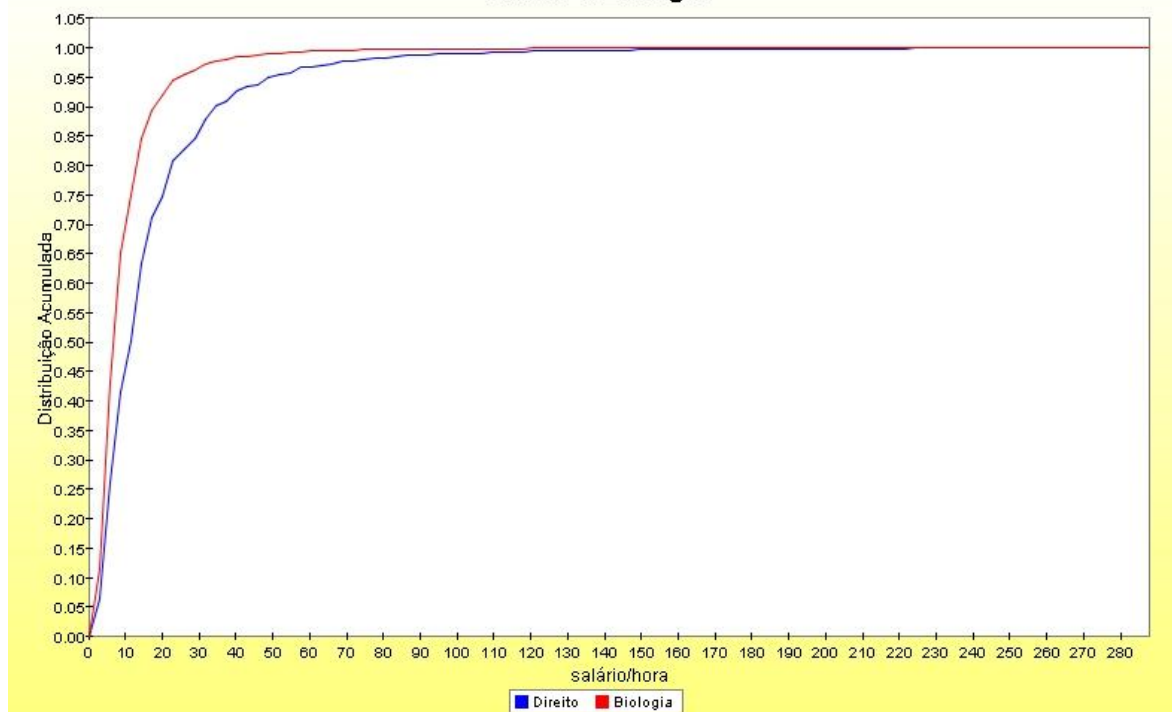


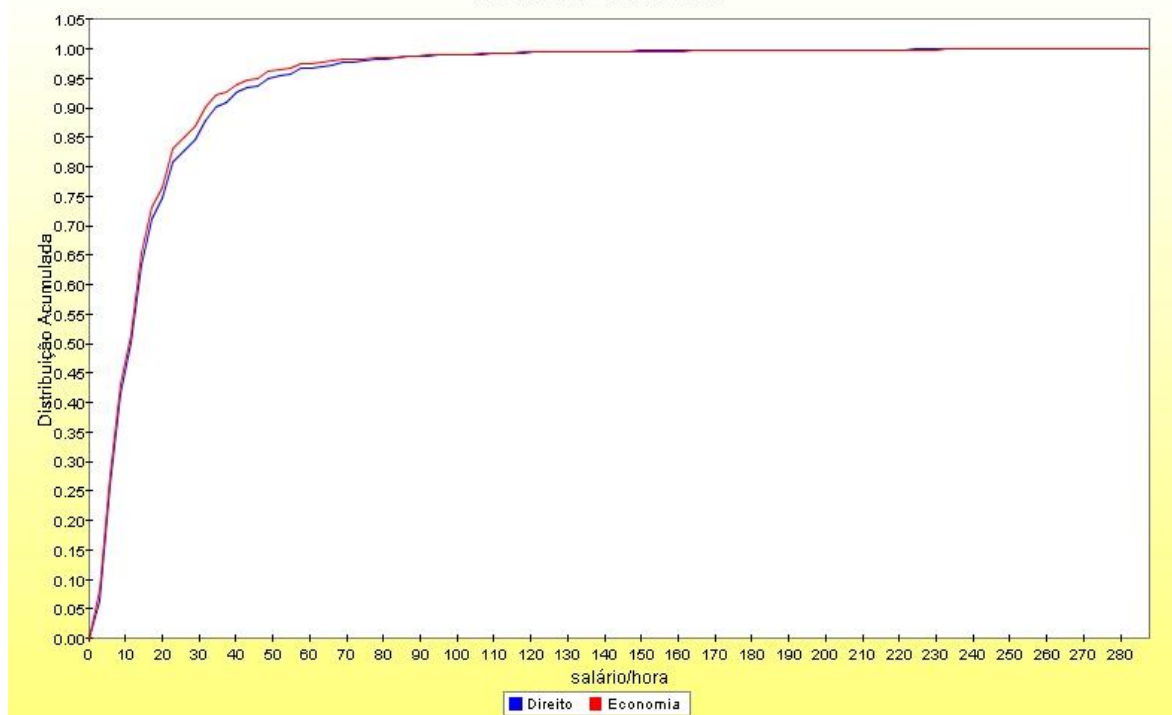
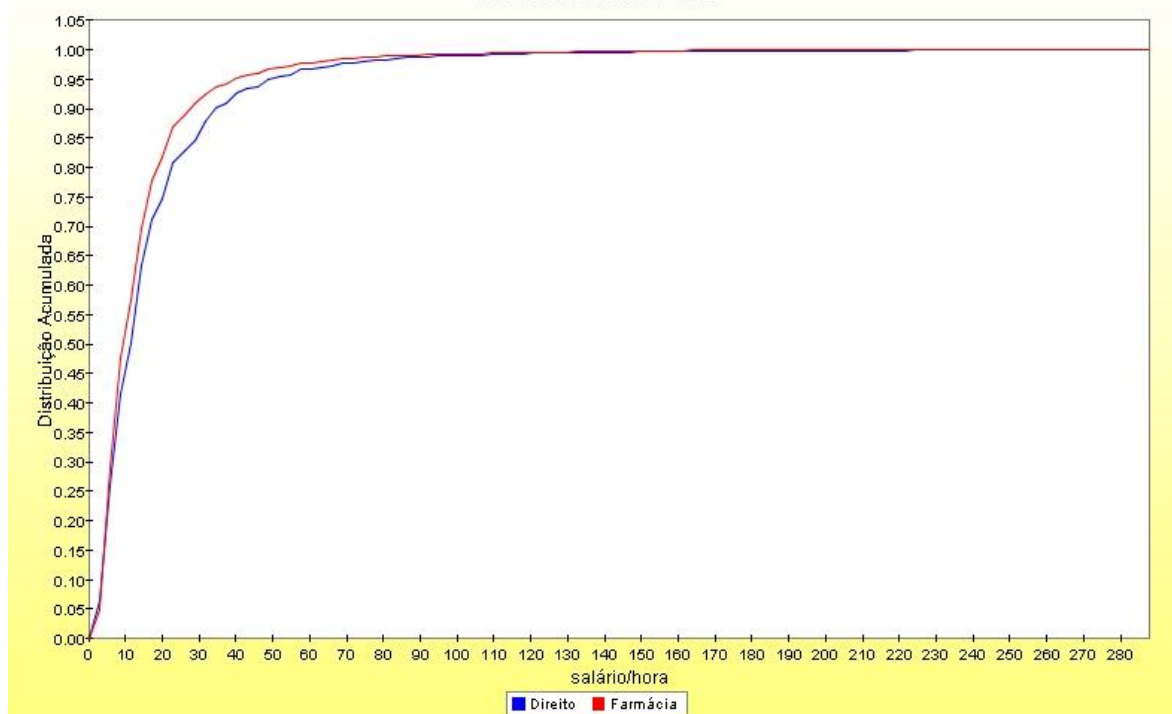
Biologia x Pedagogia**Eng. Civil x Agronomia**

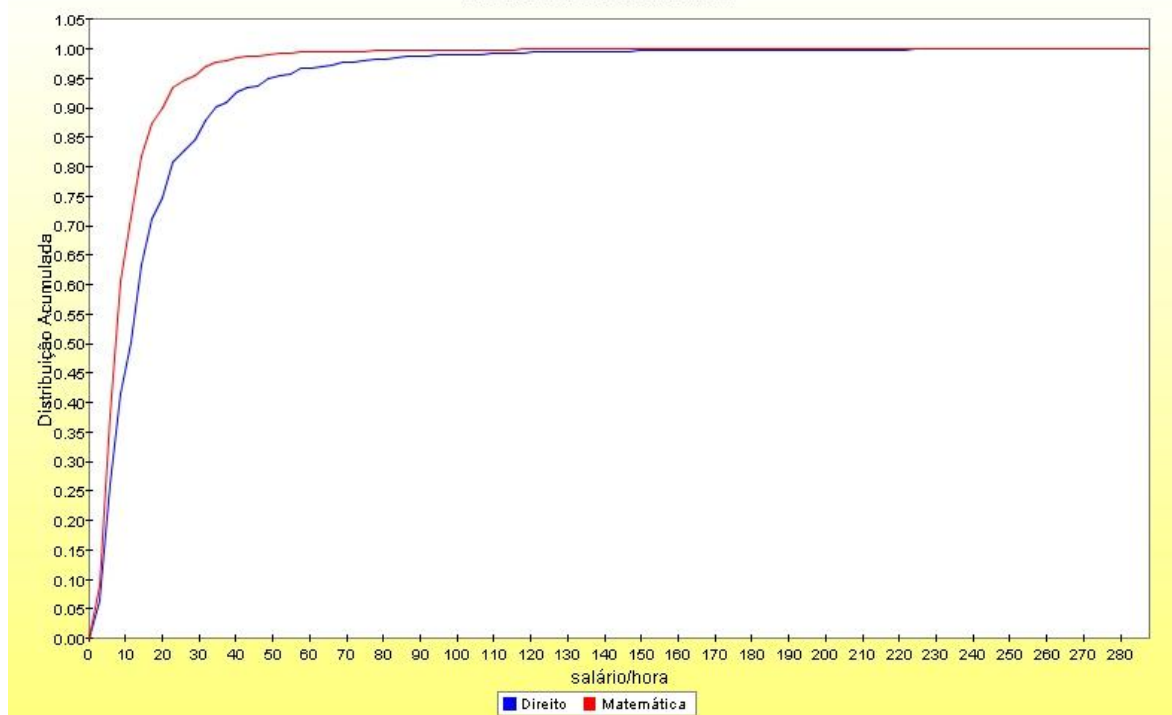
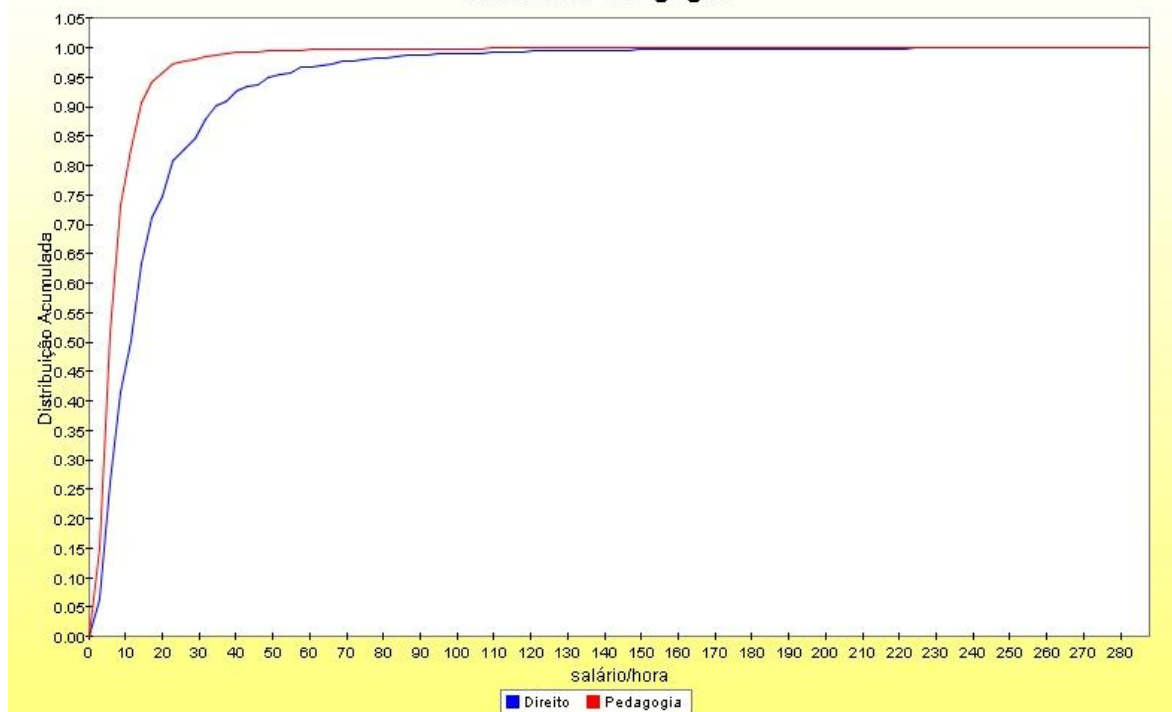
Eng. Civil x Biologia**Eng. Civil x Direito**

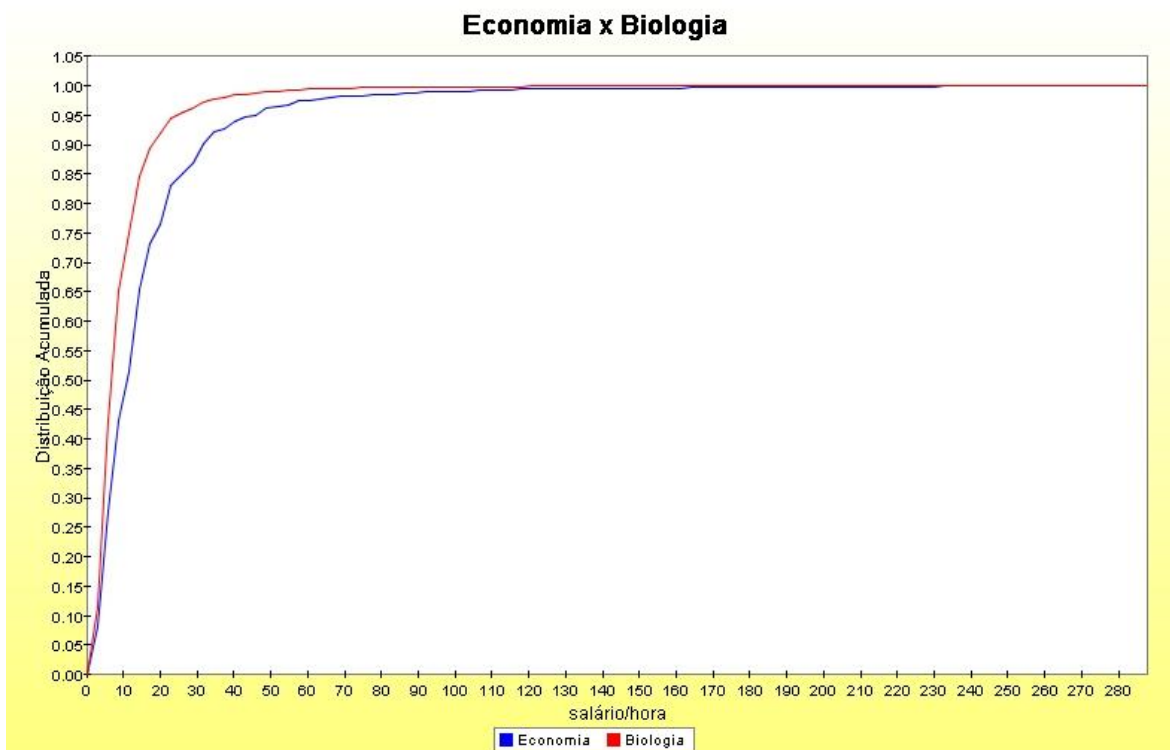
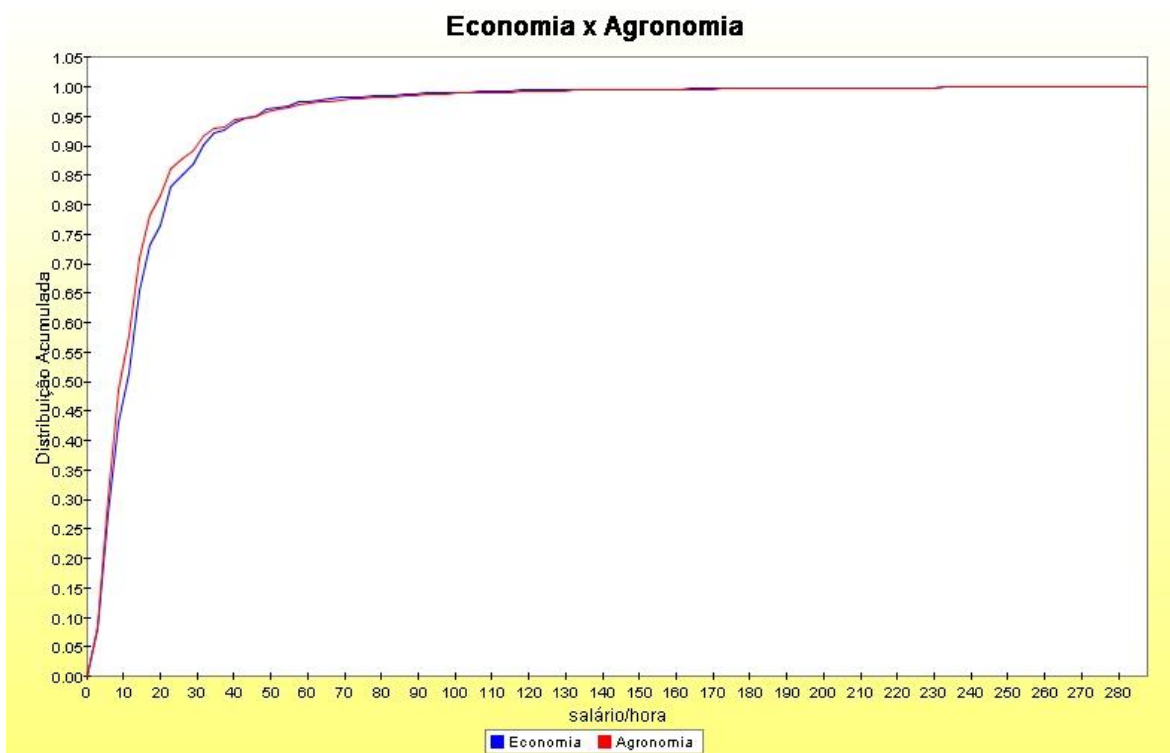


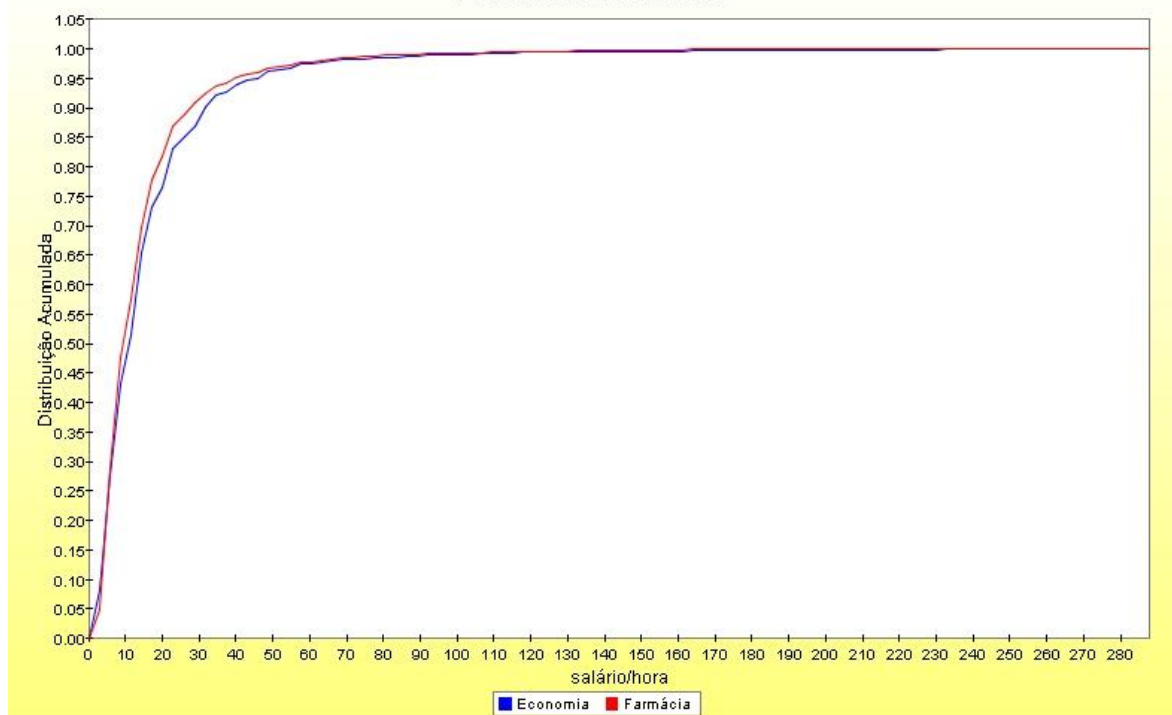
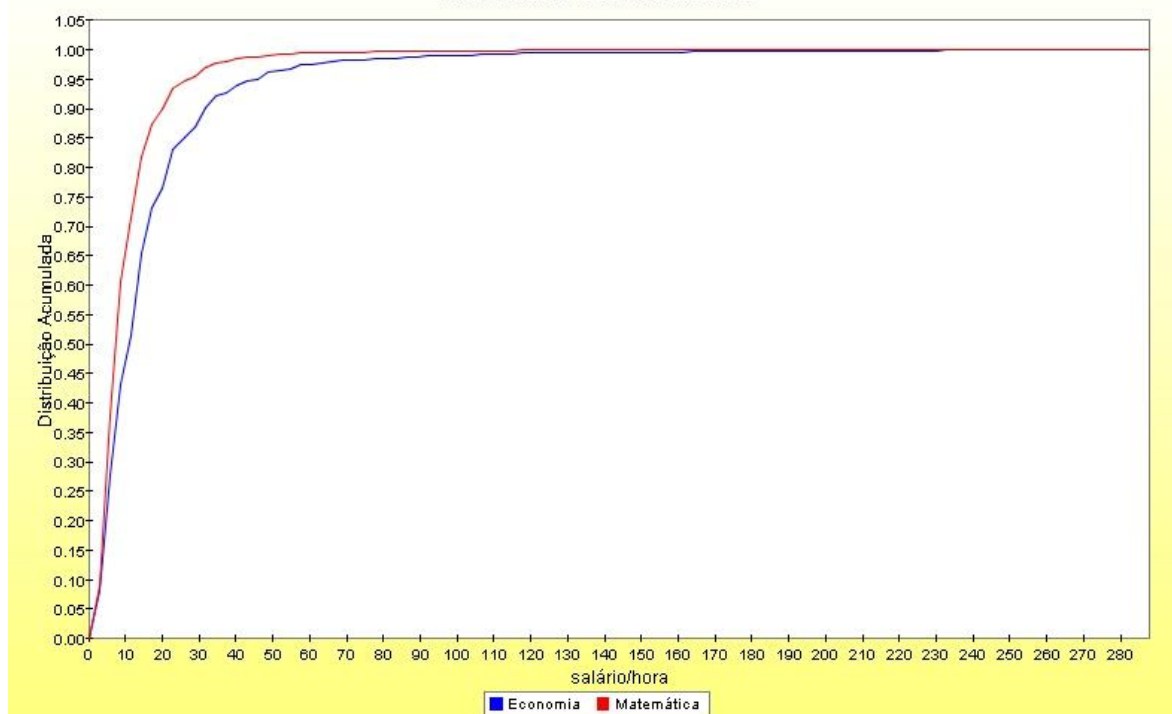
Eng. Civil x Matemática**Eng. Civil x Pedagogia**

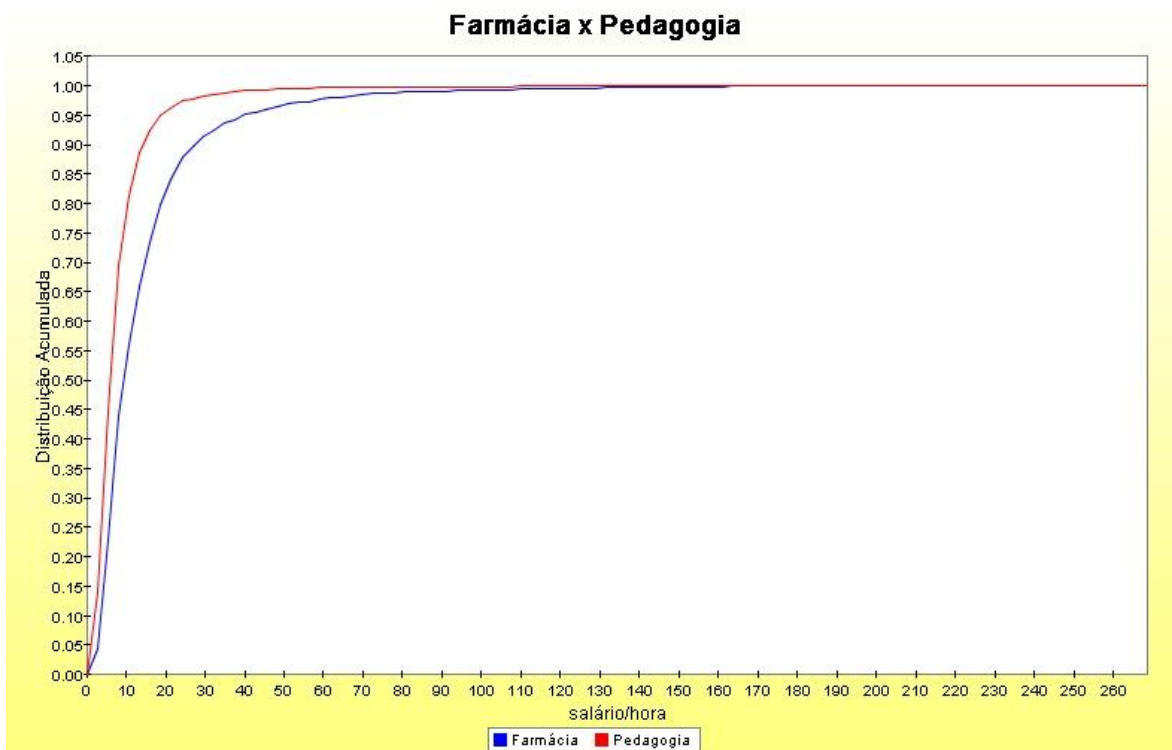
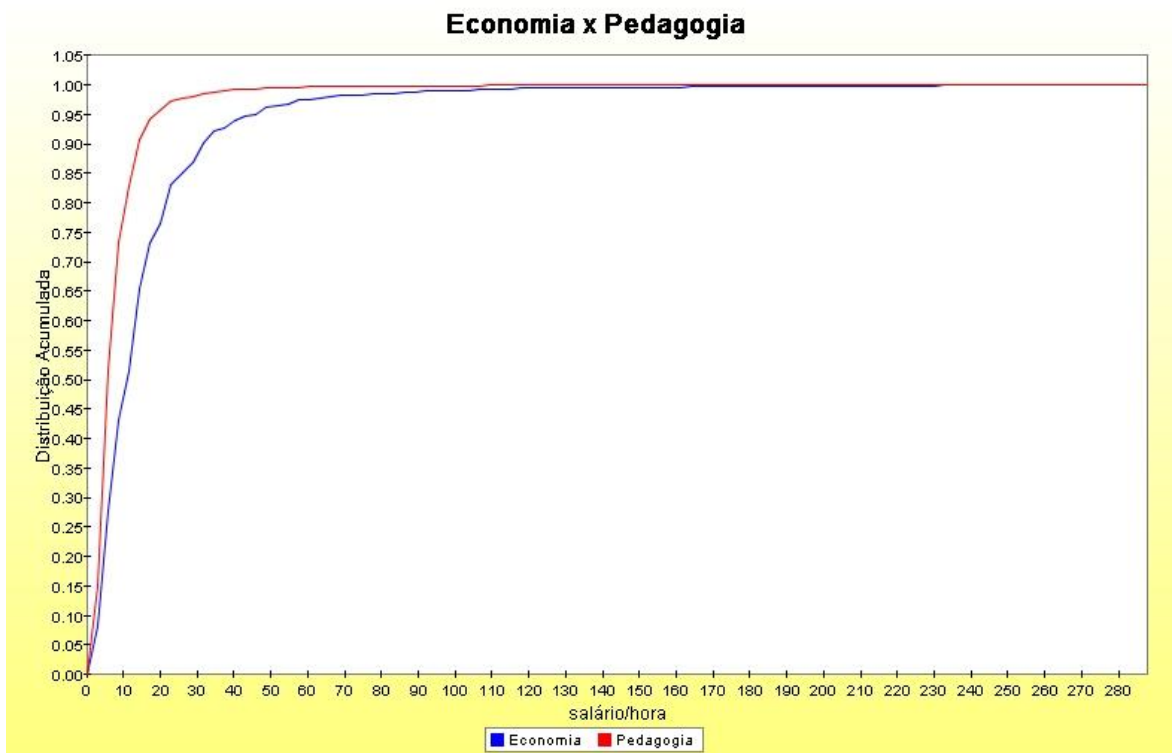
Direito x Agronomia**Direito x Biologia**

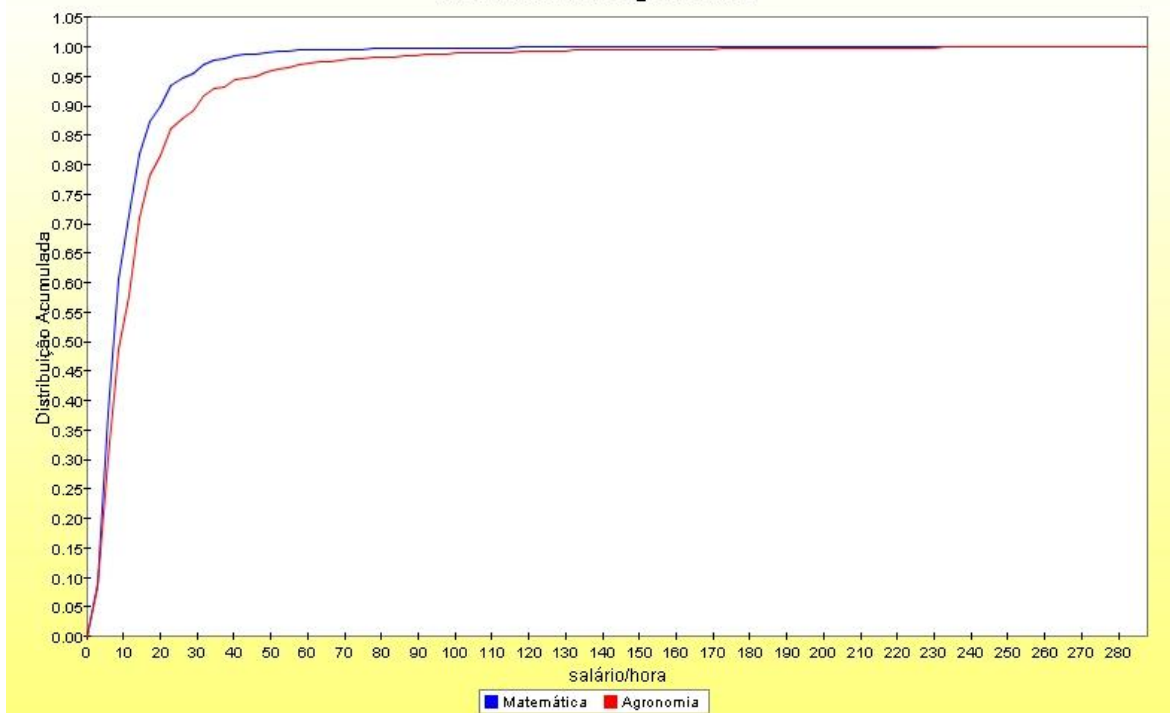
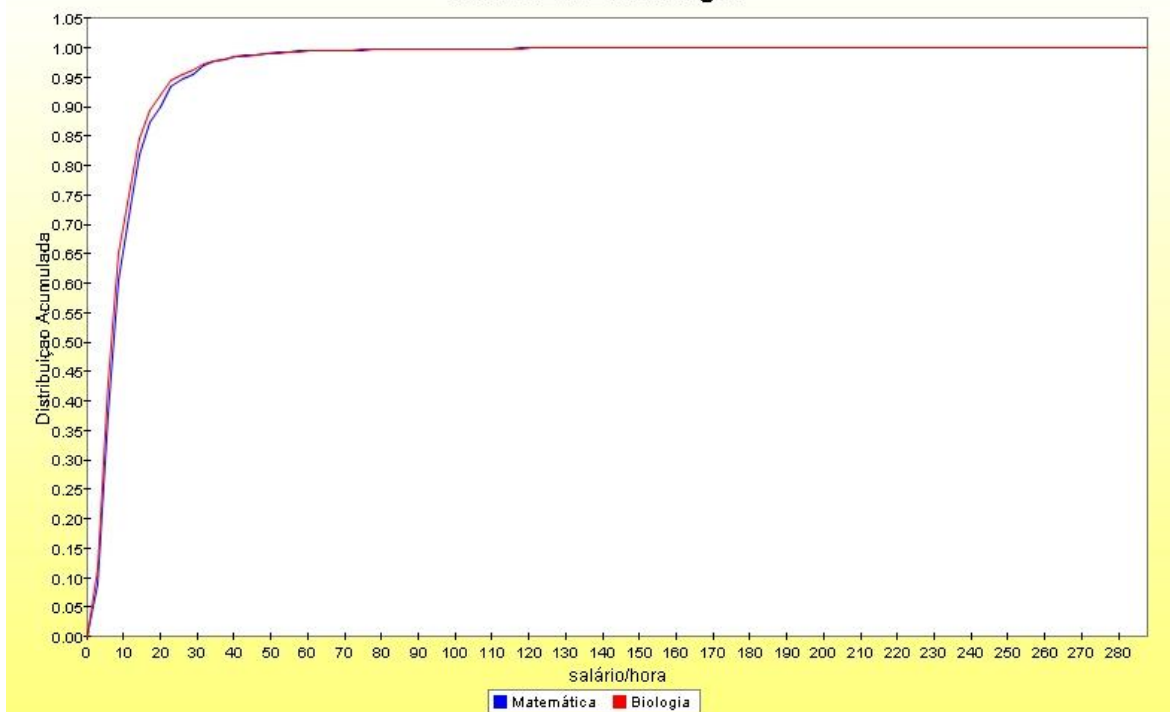
Direito x Economia**Direito x Farmácia**

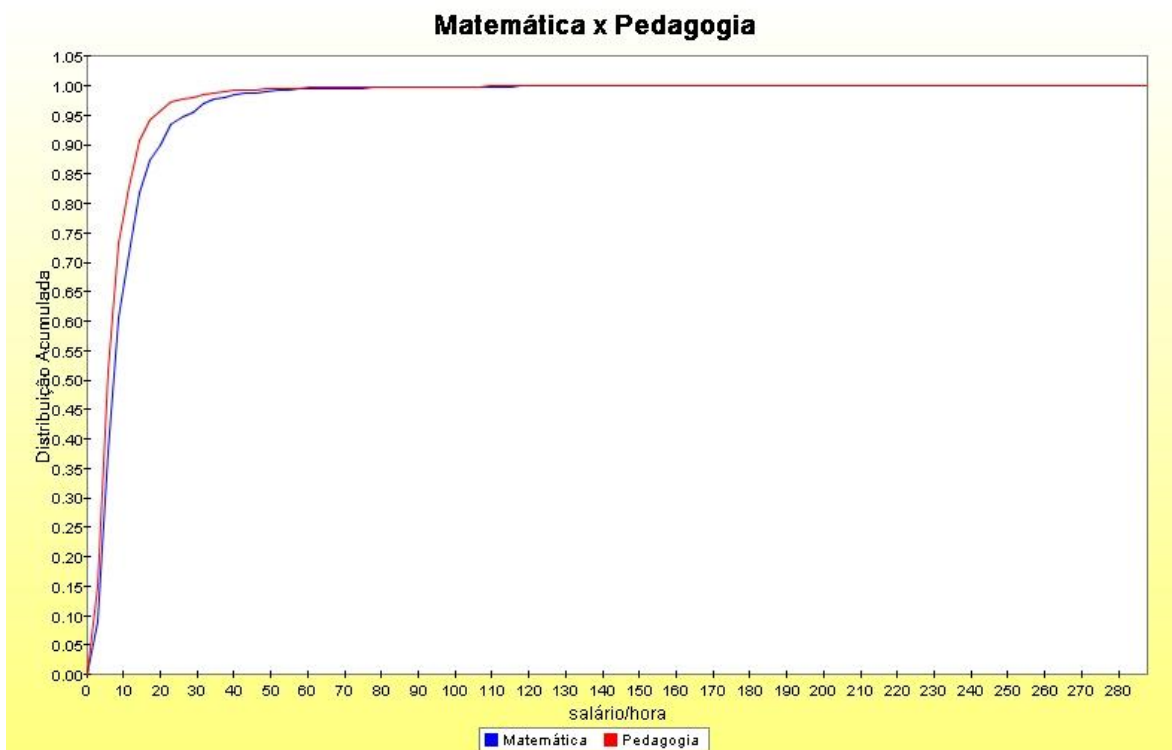
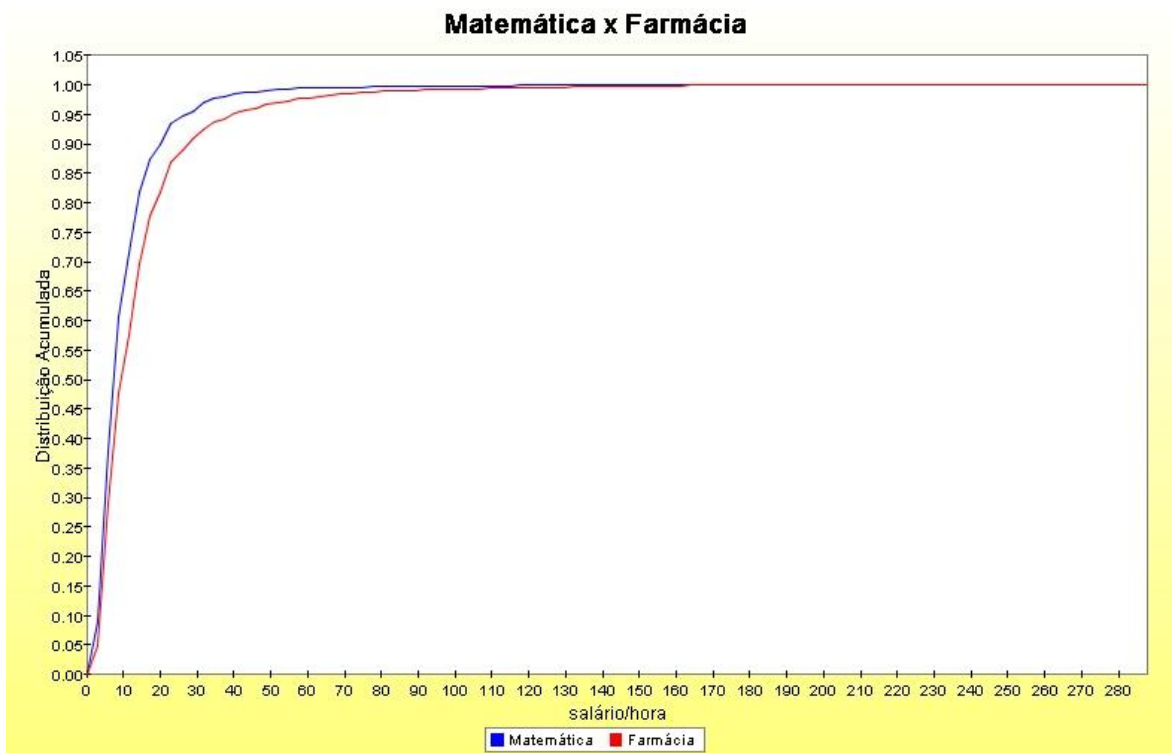
Direito x Matemática**Direito x Pedagogia**

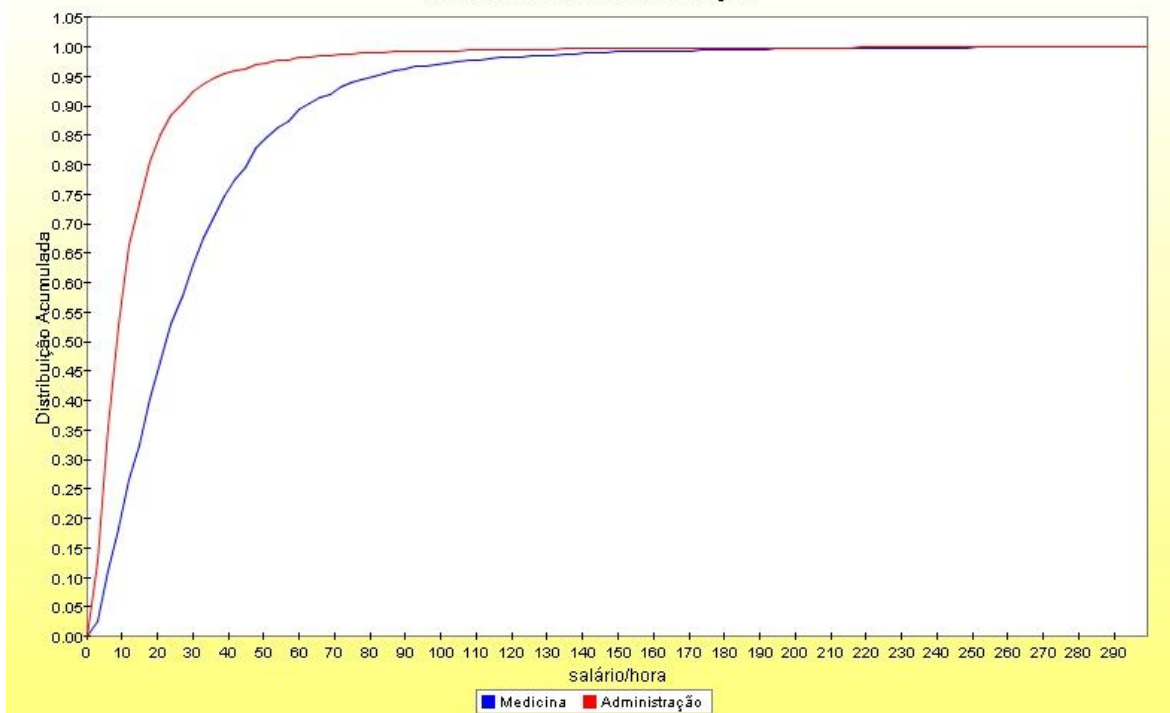
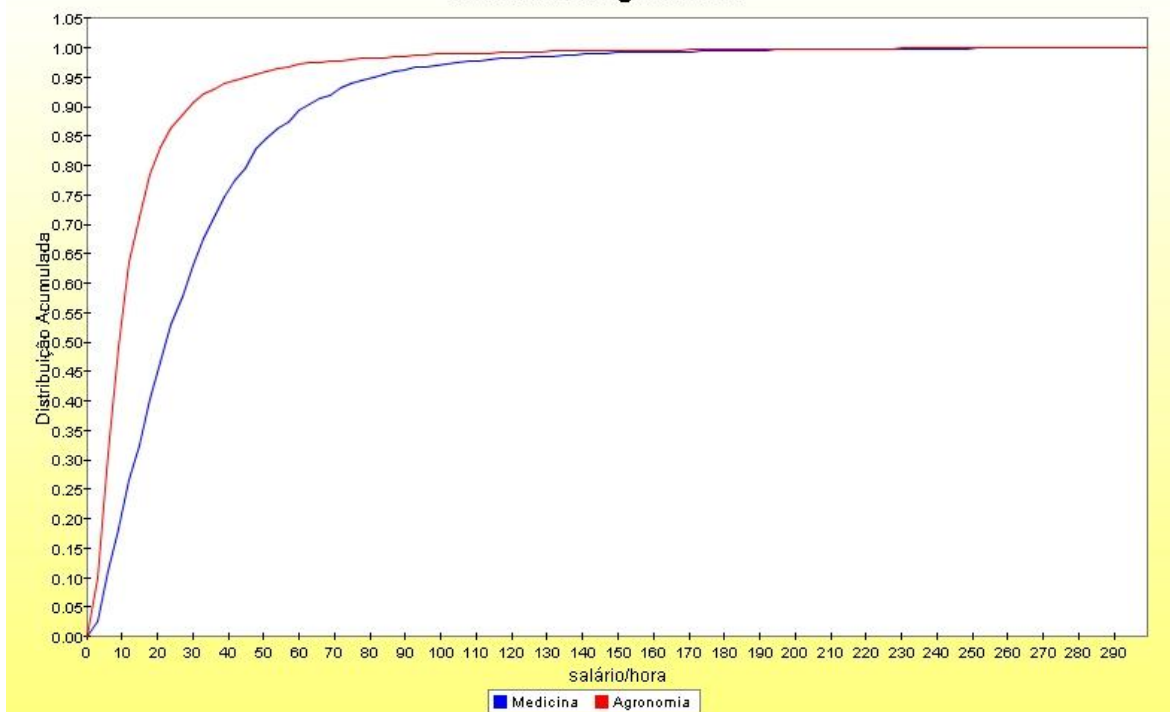


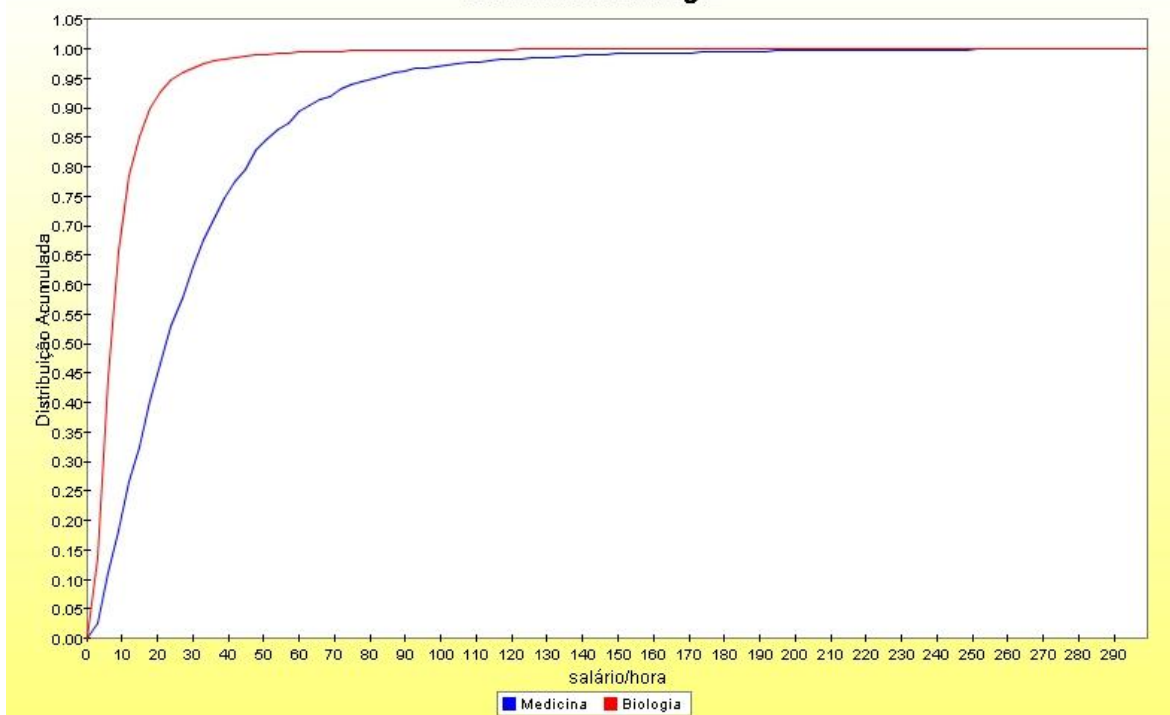
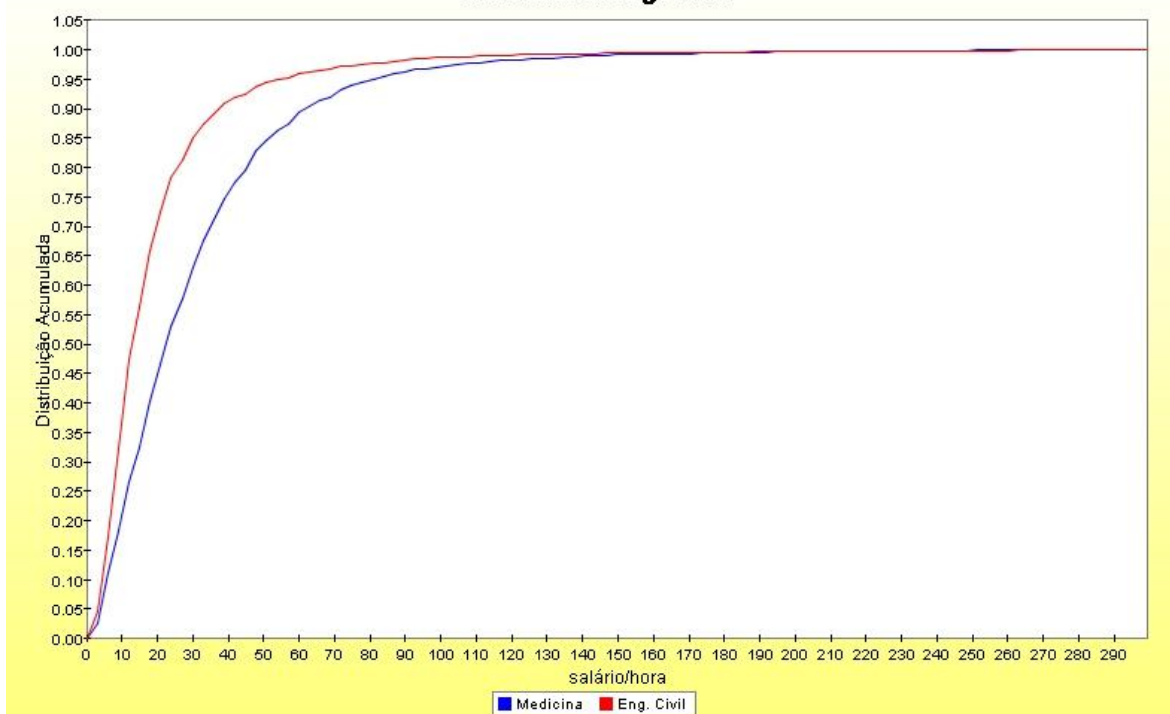
Economia x Farmácia**Economia x Matemática**

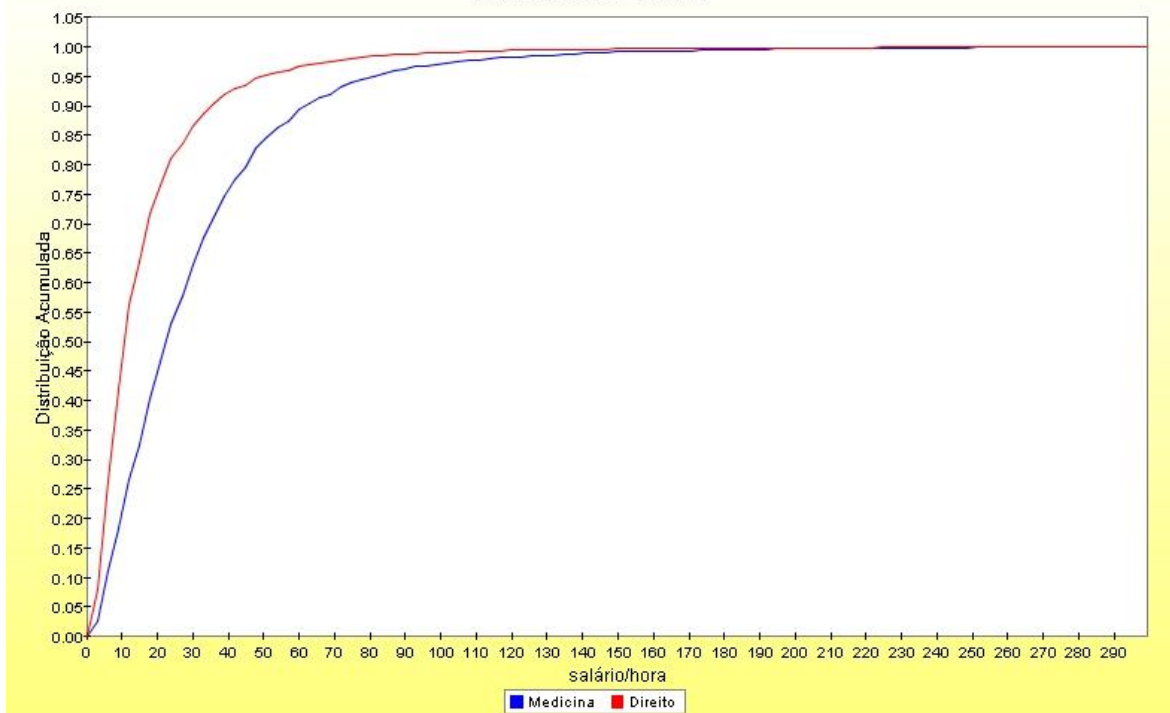
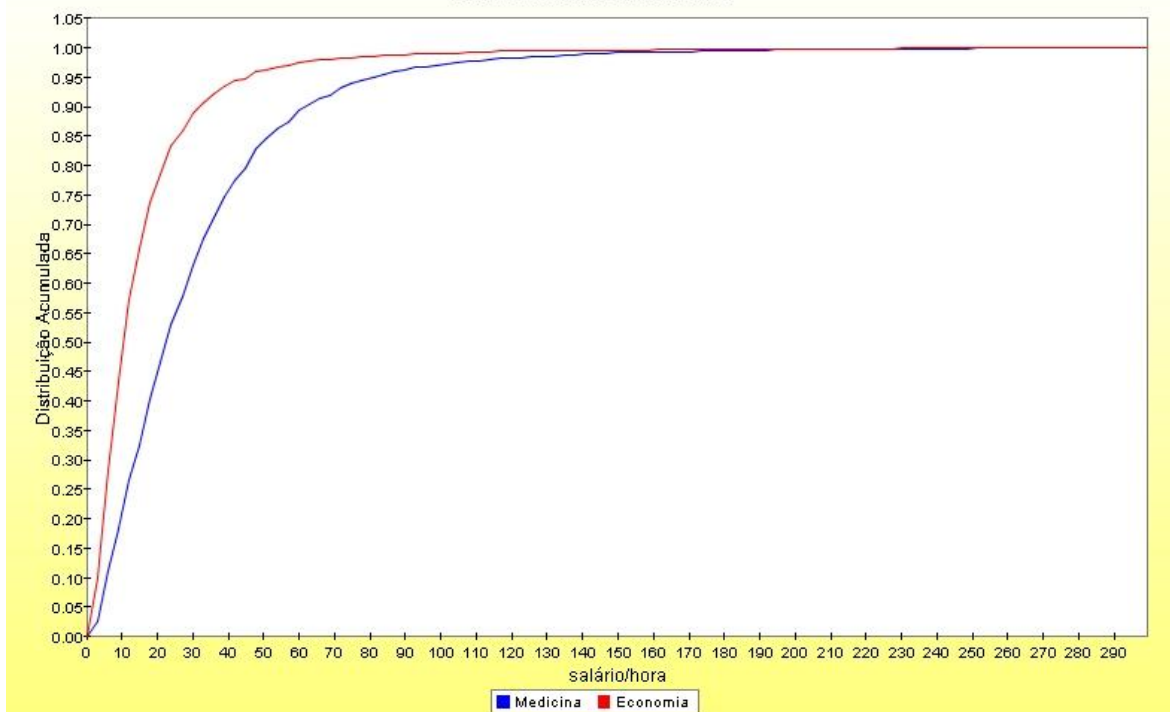


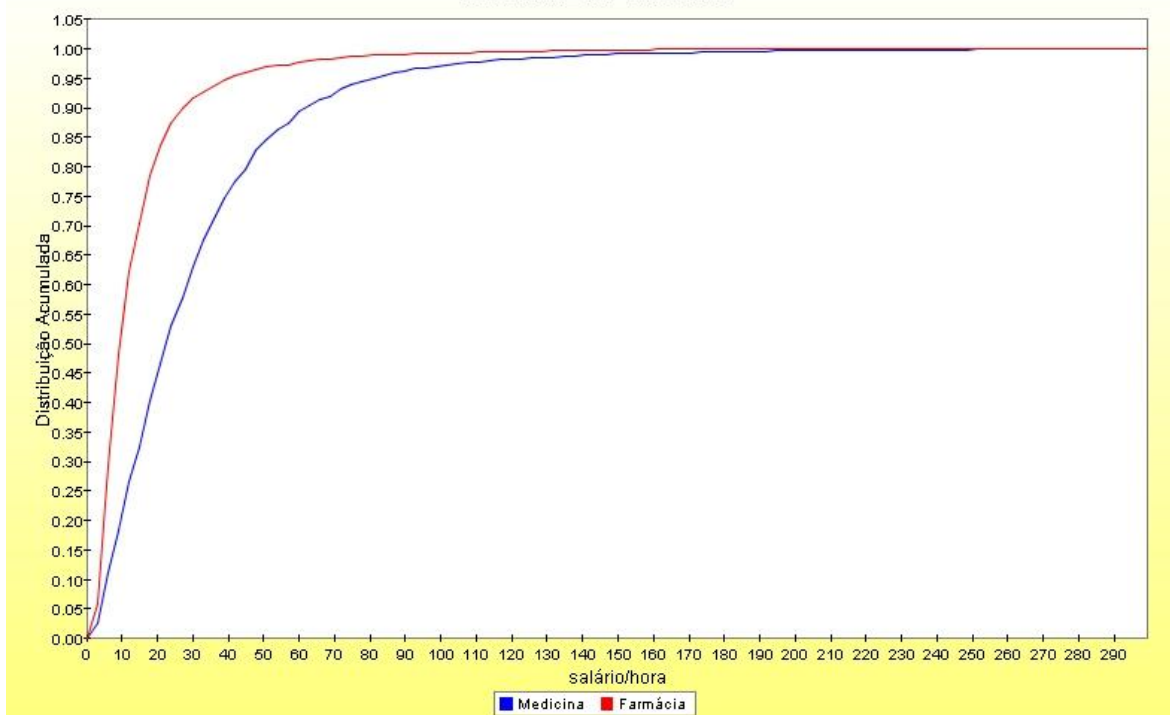
Matemática x Agronomia**Matemática x Biologia**



Medicina x Administração**Medicina x Agronomia**

Medicina x Biologia**Medicina x Eng. Civil**

Medicina x Direito**Medicina x Economia**

Medicina x Farmácia**Medicina x Matemática**