

**UNIVERSIDADE ESTADUAL PAULISTA  
INSTITUTO DE GEOCIÊNCIAS E CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA  
CAMPUS DE RIO CLARO**

**DISTRIBUIÇÃO ESTATÍSTICA DO  
FATURAMENTO DAS EMPRESAS BRASILEIRAS  
TÉCNICA DE ZIPF**

Edna Sakon Banin  
Orientador: Hari Moham Gupta

Dissertação apresentada ao  
Programa de Pós Graduação em  
Física como requisito para a  
obtenção do título de Mestre na  
Área de Concentração em Física  
Aplicada

Rio Claro  
S.P. – Brasil.  
2003

**UNIVERSIDADE ESTADUAL PAULISTA  
INSTITUTO DE GEOCIÊNCIAS E CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA  
CAMPUS DE RIO CLARO**

**DISTRIBUIÇÃO ESTATÍSTICA DO  
FATURAMENTO DAS EMPRESAS BRASILEIRAS  
TÉCNICA DE ZIPF**

Edna Sakon Banin  
Orientador: Hari Moham Gupta

---

Prof. Dr. Hari Moham Gupta - Orientador  
IGCE/RC

---

Prof. Dr. José Roberto Campanha  
IGCE/RC

---

Prof. Dr. Elbert E.N. Macau  
INPE-Instituto Nacional de  
Pesquisas Espaciais/SJC

Rio Claro  
S.P. – Brasil  
2003

*“A ausência é para o amor o que o vento é para o fogo:  
apaga o pequeno e aviva o forte” (Johan)*

*Dedico este trabalho a Deus,  
minha mãe,  
meu esposo e minhas filhas,  
que avivaram o nosso amor,  
nos momentos em que estive  
ausente para realização  
deste trabalho.*

## Agradecimentos

Ao Prof. Dr. Hari Moham Gupta, mestre e orientador, pela oportunidade de ter sido sua aluna, pela direção e atenção com que sempre dispensou a minha pessoa.

Ao Prof. Dr. José Roberto Campanha, pela sensatez, paciência e longanimidade com que sempre me atendeu.

Ao Prof. Dr. Elbert E. N. Macau, do Instituto Nacional de Pesquisas Espaciais-INPE-S.J.Campos, pelas sugestões para o aprimoramento deste trabalho.

Ao Instituto de Geociências e à Coordenação do Departamento de Pós-Graduação em Física da UNESP de Rio Claro, pelo apoio e suporte.

A CAPES pela assistência e suporte financeiro.

A FIPECAFI, ao prof. Ariovaldo dos Santos e Nivaldo Gomes Lamac pela forma gentil com que nos atendeu fornecendo os dados para a realização deste estudo.

Aos meus amigos Gislaine, Danilo, Érika e Fúlvia que me incentivaram e auxiliaram neste trabalho, e especialmente ao Fábio Chavarette pelo seu apoio e atenção.

A todo o pessoal do Departamento de Física, pela gentileza com que sempre me trataram, em especial à Rosana e Maristela.

A direção e coordenação da Escola Municipal Agrícola de Ensino Fundamental e Escola Nazarena de Ensino Fundamental, pelo apoio e compreensão a mim dispensados.

Agradeço pelas orações, pelas palavras de incentivo e pelo suporte emocional às minhas amigas Edna, Da.Geni e Izabel.

Ao carinho e apoio da minha irmã Tânia, Isaque e minhas queridas sobrinhas Sarah e Gabriela.

A Ivana por inspirar-me na busca por este título.

As queridas amigas Flávia, Gisele, Larissa e em especial Luciana pelas palavras de estímulo nos momentos em que mais precisei.

Obrigada a todos.

*“Eu não me envergonho de corrigir e mudar as minhas opiniões”,  
Porque não me envergonho de raciocinar e aprender”.*

*(Alexandre Herculano)*

# S U M Á R I O

Sumário	vi
Índice	vii
Resumo	ix
Abstract	x
Lista de Figuras	xi
Capítulo 1 - Introdução	12
Capítulo 2 – Sistemas Complexos	14
Capítulo 3 - Distribuições Estatísticas	23
Capítulo 4 - Sistema Econômico como um Sistema Complexo	34
Capítulo 5 – Distribuição do Faturamento de Empresas Brasileiras	39
Capítulo 6 – Conclusões Gerais	53
Referências Bibliográficas	55
Anexo 1	61

# ÍNDICE

<b>Capítulo 1</b>	12
Introdução	12
<b>Capítulo 2</b>	14
Sistemas Complexos	14
2.1-Introdução	14
2.2-Características de um Sistema Complexo	19
2.2.1- Geometria Fractal	19
2.2.2- Distribuição de Lei de Potência	20
2.2.3- Comportamento 1/f	21
2.2.4- Lei de Zipf	22
<b>Capítulo 3</b>	23
Distribuições Estatísticas	23
3.1 – Distribuição Normal	23
3.2 – Distribuição com Caudas Longas: Pareto e Distribuição Log-normal	26
3.2.1- Exemplos de Distribuição Log-normal	29
3.3 – Distribuição de Lei de Potência Gradualmente Truncada	29
3.4 – Distribuição Log-normal Gradualmente Truncada	31
3.5 – Técnica de Zipf	32
<b>Capítulo 4</b>	34
4.1 – Sistema Econômico como um Sistema Complexo	34
4.2 – Distribuição do Tamanho de Firms	35
4.3 – Distribuição do Tamanho de Firms Manufatureiras dos EUA	36
<b>Capítulo 5</b>	39
Distribuição do Faturamento de Empresas Brasileiras	39
5.1 – Análise Empírica	39

Tabela 1 – Tabela de Parâmetros	40
Classificação das Empresas Brasileiras	41
5.2 – Análise e Discussão	51
<b>Capítulo 6</b>	<b>53</b>
6.1 - Conclusões Gerais	53
Referências Bibliográficas	55
Anexo	61



## RESUMO

No presente trabalho estudamos a distribuição do tamanho de empresas brasileiras nos anos de 1996, 1997, 1998, 1999 e 2000, utilizando a técnica de Zipf. Consideramos como tamanho de empresas o faturamento das mesmas. Gibrat (1931) averiguou que esta distribuição é do tipo Log-normal, mas recentes trabalhos demonstram que o faturamento nas empresas de grande porte é menor do que prevê este tipo de distribuição. Num trabalho recente Gupta e Campanha mostraram que distribuições como **Log-normal e lei de potência** devem ser **gradualmente truncadas** após um certo valor crítico, que depende diretamente da limitação física do sistema analisado. Diante desta concepção utilizamos a distribuição **Log-normal gradualmente truncada** para estudar a distribuição por faturamento de empresas brasileiras nos anos acima mencionados. Observamos, no entanto, que a distribuição Log-normal define suficientemente esta distribuição exceto para as duas ou três empresas de maior faturamento.

Finalmente discutiremos a possível razão da distribuição obtida e justificação dos parâmetros.

**Palavras-chave:** Distribuição Log-normal, gradualmente truncada, faturamento, ranking.

**Abstract:**

In the present work, we studied size distribution of Brazilian firms for years 1996, 1997, 1998, 1999 and 2000, using technique of Zipf plot. We measure size of firm through its sell. Gibrat in 1931 speculated that distribution must be Log-normal. However the sell of very big firms is less than what is predicted through Log-normal distribution. Recently Gupta and Campanha show that distributions of type Log-normal or power law must be truncated after certain critical value, depending on physical limitations of the systems. We observe that size distribution of Brazilians firms is well given through Log-normal distribution except first two or three firms. Finally we discuss possible reason of this distribution and justified values of parameters

**Key Words:** gradually truncated Log-normal distribution, size of firms, ranking.

## Lista de Figuras:

<b>Figura 1:</b>	Classificação das Firms Manufatureiras dos EUA	38
<b>Figura 2:</b>	Gráfico da Curva Empírica da Classificação de Empresas Brasileiras Ano de 1996	41
<b>Figura 3:</b>	Gráfico da Curva Empírica da Classificação de Empresas Brasileiras Ano de 1997	42
<b>Figura 4:</b>	Gráfico da Curva Empírica da Classificação de Empresas Brasileiras Ano de 1998	43
<b>Figura 5:</b>	Gráfico da Curva Empírica da Classificação de Empresas Brasileiras Ano de 1999	44
<b>Figura 6:</b>	Gráfico da Curva Empírica da Classificação de Empresas Brasileiras Ano de 2000	45
<b>Figura 7:</b>	Gráfico da Curva Empírica e Teórica da Classificação de Empresas Brasileiras. Ano de 1996	46
<b>Figura 8:</b>	Gráfico da Curva Empírica e Teórica da Classificação de Empresas Brasileiras. Ano de 1997	47
<b>Figura 9:</b>	Gráfico da Curva Empírica e Teórica da Classificação de Empresas Brasileiras. Ano de 1998	48
<b>Figura 10:</b>	Gráfico da Curva Empírica e Teórica da Classificação de Empresas Brasileiras. Ano de 1999	49
<b>Figura 11:</b>	Gráfico da Curva Empírica e Teórica da Classificação de Empresas Brasileiras. Ano de 2000	50

# CAPÍTULO 1

## Introdução

Propomos tratar neste trabalho o comportamento da distribuição das Empresas Brasileiras usando a técnica de Zipf (Zipf, 1949) motivados por diversos estudos nesta área (Stanley et al, 1999; Axtell, 2001; Arthur, 1999). Utilizamos dados sobre o faturamento de Empresas Brasileiras dos anos de 1996, 1997, 1998, 1999 e 2000 fornecidos pela Fipecafi (Fundação Instituto de Pesquisas Contábeis, Atuariais e Financeiras da Universidade de São Paulo) entidade responsável pela pesquisa da Revista Exame. A partir destas informações estudamos as distribuições estatísticas obtidas, seus parâmetros, e discutimos os resultados.

Diante de relevantes trabalhos sobre o faturamento de empresas produzidos por vários estudiosos nesta área como Gibrat (Gibrat, 1931), Stanley e seus colaboradores (Stanley, 1995), Axtell (Axtell, 2001) e posteriormente Gupta e Campanha (Gupta, Campanha, 2001) utilizando os mesmos dados de Stanley, ficamos animados em trabalhar com dados das empresas brasileiras. Curiosos por saber qual seria o comportamento desta distribuição realizamos este estudo comparando os resultados obtidos com a distribuição de empresas americanas.

Para realizar tal intento, no próximo capítulo, faremos algumas explicações introdutórias sobre os principais conceitos e características dos sistemas complexos. O capítulo três apresentará informações básicas sobre as técnicas estatísticas mencionadas ao longo deste trabalho. No capítulo quatro, com base nos trabalhos de Gupta e Campanha descreveremos o modelo que utilizamos assim como as hipóteses e comparações entre os dois estudos. A análise da distribuição e seus resultados foram desenvolvidos no capítulo cinco. No capítulo seis finalizamos com as conclusões gerais.

# CAPÍTULO 2

## SISTEMAS COMPLEXOS

### 2.1 Introdução

*“Desde há algum tempo é crescente o interesse científico no estudo dos chamados Sistemas Complexos. Estes sistemas, a princípio caracterizados por mostrarem um comportamento de difícil compreensão, se apresentam como derradeiros bastiões que desafiam o esforço da ciência em apresentar uma visão unificada e homogênea do universo que nos rodeia”* (Macau, 2002).

A ciência tem promovido grandes avanços ao longo do tempo por meio do conhecimento científico nos permitindo uma compreensão cada vez mais profunda da Natureza. Assim o desafio do estudo de sistemas complexos vem sendo aos poucos vencido. Para que saibamos como essa evolução vem ocorrendo, faremos algumas explicações gerais para sua compreensão. Inicialmente, é fundamental sabermos o conceito de sistema: trata-se de uma parte do universo que se pretende estudar, descrever ou prognosticar, distinguindo-se do restante deste universo, não por fronteiras visíveis, mas pela organização de relação entre seus componentes. Como exemplos de sistemas podemos citar o sistema Solar, a Terra, o sistema de um relógio e outros. Um sistema sendo

identificado pode ser descrito por suas propriedades, podendo também fazer parte de outro sistema, apenas como um de seus componentes internos.

Para compreensão e elaboração de previsões de comportamento de um sistema aplica-se o método científico, em geral, de acordo com critérios de execução, segundo quatro itens de ação:

1. Observação, descrição e classificação de um fenômeno natural ou grupo de fenômenos;
2. Dedução das regularidades e leis que permitam a explicação do fenômeno por meio de hipóteses;
3. Uso dessas hipóteses para fazer previsões sobre a ocorrência do fenômeno e de fenômenos correlacionados;
4. Comparação das previsões com os resultados dos experimentos e observações subseqüentes, empreendidos por experimentadores e observadores independentes. (Einstein, 1976).

Caso a aplicação dos itens acima sobre um determinado fenômeno, experimento ou observações se ajustem ao resultado das hipóteses apresentadas, tornam-se teoria ou leis da Natureza, caso contrário, devem ser rejeitadas ou reavaliadas.

Utilizando o método científico encontramos o comportamento de um sistema, preferencialmente, por meio da comparação entre modelos, ou seja, uma confrontação para verificar se um modelo é *mais* ou *menos* adequado que outro em relação aos objetivos que se desejam atingir ou provar com o estudo do sistema. Esta concepção está estreitamente ligada a filosofia do determinismo laplaciano, segundo o conceito de que “um matemático infinitamente inteligente seria capaz de prever com a certeza que desejasse o estado futuro de um sistema a partir de observações do estado atual e do conhecimento das leis da Natureza” introduzido por Laplace no século XIX.

Em sistemas físicos as *equações matemáticas* são utilizadas para representar os modelos. A matemática é fundamental na aplicação do método científico, pois freqüentemente as leis obtidas por meio de observações da Natureza são expressas através de equações matemáticas, permitindo a realização de previsões minuciosas, que podem ser comparadas com resultados de experimentos subseqüentes.

O século XIX e XX presenciou o apogeu do método científico, associado ao enfoque *reducionista*, em que os sistemas devem ser observados sob um nível crescente de resolução na busca por seus constituintes elementares. Em decorrência disto a matéria foi considerada como uma formação sucessiva de moléculas, átomos, núcleos e quarks, constituindo a base de tudo o que existe na Natureza.

Mas nosso mundo não é composto apenas por sistemas que podem ser definidos sob o enfoque reducionista para previsão de comportamentos futuros. Os terremotos, por exemplo, são um caso em que teoricamente poderíamos entender seu comportamento, desde que atribuíssemos condições iniciais a cada partícula elementar, medindo posições e velocidades associadas a trilhões e trilhões dessas partículas e depois computar individualmente a trajetória e estado de cada uma delas inserindo-as num sistema de equações diferenciais de ordem extremamente elevada, num esforço impraticável na busca por soluções numéricas.

A termodinâmica por sua vez, introduziu novos conceitos de modelos e previsões utilizando apenas variáveis que se mostrem relevantes para uma descrição satisfatória do comportamento de um sistema, ou seja, a utilização de variáveis macroscópicas para descrever satisfatoriamente o comportamento do sistema. Como no caso dos gases, através da utilização das variáveis macroscópicas de estado, pressão, volume e temperatura é possível descrever ou predizer seu estado de equilíbrio, por meio da “lei dos gases perfeitos”. Esta descrição permite a utilização de um modelo simples e tratável, no nível microscópico, possibilitando também fazer previsões do comportamento futuro, associadas ao comportamento do sistema como um todo, no nível macroscópico.

Somente em meados do século XIX com a introdução da mecânica estatística por Maxwell, Boltzman and Gibbs houve alteração sobre o conceito laplaciano de predição, que se mostrava ineficaz em vários casos, assim como no exemplo da previsão do comportamento dos gases ou terremoto citado anteriormente. Neste caso aplica-se o conceito de *predição probabilística*, calculando-se apenas a distribuição de probabilidade das variáveis e não a solução exata associada ao conjunto de condições iniciais.

Para descrever um sistema através da mecânica estatística se faz necessário aceitar que todas as possíveis combinações associadas aos processos rápidos, ou seja, processos



que são muito mais velozes do que a escala de tempo de nossas observações, ocorrem de acordo com uma distribuição fixa de probabilidade. Assim, por meio de uma equação de movimento, calcula-se a distribuição de probabilidade associada a esta equação de movimento e suas propriedades. Conseguimos descrever na mecânica estatística, o comportamento de cristais que possuem regularidade periódica ou os gases, onde todas as partículas são estatisticamente iguais. Recentemente Tsallis (Tsallis, 1999) introduziu uma extensão da estatística de Boltzman conseguindo explicar o comportamento de alguns sistemas complexos sob o ponto de vista microscópico, considerando interações de longa distância entre seus elementos e memória microscópica de longa duração.

Porém estes conceitos de predição têm aplicação limitada para determinados sistemas. Consideremos alguns fenômenos que nos rodeiam: O furacão que pode durar poucas horas até três semanas e provocar muitos danos, tem como fato comum apresentar uma trajetória fora dos padrões de previsão, permitindo apenas estimativas probabilísticas de sua trajetória, típico exemplo de um sistema complexo em que a previsão de comportamento desafia os modelos existentes. Os tornados também são outro exemplo de sistema complexo, pois há grande dificuldade em se prever quando ocorrerá e a sua trajetória, trata-se de um caso ainda mais crítico que o dos furacões, pois se desenvolvem com muita rapidez, produzindo muita destruição especialmente em áreas habitadas. É extremamente pequena a capacidade de prevê-los por meio de modelos sendo, portanto, um fenômeno complexo.

Outro caso de um fenômeno altamente complexo são os relâmpagos, pois também não existem modelos que possibilitem predizer o local de ocorrência deles, caso contrário, seria possível armazenar a energia gerada pelos relâmpagos.

A vida biológica evoluiu na Terra, com muitas espécies diferentes, muitas com bilhões de indivíduos, competindo e interagindo uns com os outros num mesmo ambiente. No fim de um pequenino ramo da biologia nós nos encontramos e podemos reconhecer outros seres humanos porque todos nós somos diferentes. O corpo humano é formado por um emaranhado arranjo de células que interagem entre si. O conjunto dos neurônios no cérebro é um exemplo pleno de sistema complexo que se deve à particular e intrincada forma de interligação entre eles, e da evolução dinâmica do sistema, onde um neurônio A influencia outro B que por sua vez influencia outro C, e assim por diante. Entretanto apesar

da simplicidade de sua unidade formadora fundamental o cérebro humano é sem dúvida o sistema complexo mais conhecido e fascinante, pois ainda é capaz de formar representações do complexo mundo exterior. Nossa história, por meio de registros de guerras, revoluções, religiões e sistemas políticos, constitui outro nível de complexidade; economias compostas por consumidores, produtores, governos, políticos e especuladores também caracterizam sistemas complexos, criados pelo homem.

Todos estes sistemas apesar de manifestarem comportamento complexo, apresentam *leis universais* de comportamento. A ciência tem canalizado seus esforços na convicção da existência destas leis, que seriam aplicáveis a todas as classes de sistemas, (Badii, 1997). O grande propósito seria o de identificá-las.

Entender o comportamento de sistemas complexos significa compreender como suas diversas partes agem em conjunto de forma a produzirem o comportamento do todo, em decorrência surgem os fenômenos coletivos e as propriedades que não estão presentes nas partes quando analisadas separadamente.

O estudo do comportamento dos sistemas complexos provocou o surgimento de novas linhas de pesquisas em áreas da física, geografia, geologia, economia, biologia, etc. Cada uma trabalhando seus próprios objetos e conceitos com a finalidade de compreender melhor o comportamento complexo. Não há evidentemente, especialistas que tenham uma visão geral e profunda do comportamento de todos os fenômenos naturais que apresentam sistemas complexos, sendo um estudo fundamentalmente interdisciplinar, porém não podem ser definidas dentro de um domínio científico específico, se faz necessário então, uma teoria geral ou alguns mecanismos que possibilitem explicar estas observações em todas as áreas. Técnicas matemáticas e estatísticas vêm sendo desenvolvidas na física com objetivo de utilizá-las em todos os sistemas complexos.

Comparar aspectos estatísticos com padrões gerais é uma confrontação essencial para qualquer esforço científico; não seria diferente para sistemas complexos onde a conjugação entre a abstração e a natureza estatística e probabilística poderia parecer estranha para alguns cientistas que esperam obter uma caracterização exata dos fenômenos reais. No entanto, a teoria da complexidade é de natureza estatística, pois estuda os fenômenos reais, por meio de teorias probabilísticas, através da confrontação entre teoria e

experimentos ou observações, comparando as características estatísticas com padrões gerais.

## **2.2 Características dos Sistemas Complexos**

Enumeramos algumas características que nos permitem identificar o comportamento de um sistema complexo (Poon, 1995, Macau, 2002):

1. Um comportamento ao longo do tempo que não se mostra nem completamente ordenado e previsível, nem completamente aleatório e imprevisível;
2. A evolução espaço-temporal do sistema revela padrões e estruturas coerentes, que se desenvolvem em várias escalas, sem que apresentem um relacionamento elementar entre si;
3. Estas estruturas podem apresentar um relacionamento hierárquico, o que significa que o aparecimento de estruturas não triviais que se estendem por uma ampla gama de escalas de espaço de tempo.

Através de observações empíricas estabelecidas em diversas áreas, algumas características são onipresentes em Sistemas Complexos podendo ser destacadas algumas delas, como:

### **2.2.1 Geometria Fractal**

“Fractal” o termo significa fragmentado, quebrado, irregular (latim: fractus). Utilizado pela primeira vez por Mandelbrot para definir estruturas que possuem irregularidade geométrica fundamental e invariante em diferentes escalas. A geometria fractal descreve em geral objetos de dimensões fracionárias (Mandelbrot, 1983). As partes constituintes dos fractais relacionam-se entre si e apresentam características como auto-similaridade, e lacunaridade. Um estudo sobre o comprimento da costa da Noruega, demonstra um conjunto hierárquico de fiordes, numa sucessão em todas as escalas. Assim, uma medida do comprimento dessa costa depende da escala usada para medi-la.

Muito trabalho foi feito na caracterização geométrica dos fractais, mas o problema da dinâmica da origem de fractais ainda persiste.

### 2.2.2 Distribuições do tipo Lei de Potência

Alguns sistemas complexos podem apresentar comportamentos catastróficos, devido a natureza interativa de seus elementos, onde um certo movimento afeta os componentes do sistema provocando um efeito dominó. Este comportamento é definido como realimentação positiva, em que um evento produz outro, aumentando sua magnitude, como por exemplo, a energia liberada por terremotos em uma determinada região. (Mandelbrot, 1983; Levy, 1937 apud Tsallis, 2000)

Tal comportamento poderia ser explicado pela **distribuição do tipo Lei de Potência**, dada por:

$$P(s) = As^{-a} \quad (2.1)$$

onde  $P(s)$  é a probabilidade de ocorrência de  $s$ ;  $a$  é o expoente de distribuição da Lei de Potência e  $A$  é constante.

Gráficos logarítmicos (log-log) que apresentam linhas retas descrevem a distribuição de “Lei de Potência”, uma certa quantidade  $N$  pode ser expressa como proporcional a potência de uma outra quantidade  $s$ :

$$\log N(s) = -a \log(s) + A \quad (2.2)$$

onde  $s$  poderia ser a potência de energia liberada por um terremoto,  $N(s)$  o número de terremotos com esta energia e  $A$  é uma constante. O gráfico  $\log N(s)$  versus  $\log(s)$  é uma linha reta e o expoente  $\alpha$  define a inclinação da reta.

### 2.2.3 Ruído do tipo $1/f$

Ruído do tipo  $1/f$  é uma das manifestações da Lei de Potência e tem sido observado em diversos sistemas, tais como o nível de água do Rio Nilo (Hurst, 1951), luzes de quasares e tráfego em rodovias (Mandelbrot, 1999). Estes fenômenos apresentam detalhes em todas as escalas de tempo. J.Hurst passou parte de sua vida estudando o nível de água do rio Nilo, em várias escalas de tempo, (de minutos até anos) observou que esta série temporal poderia ser vista graficamente como uma superposição de sinais de todas as variações do nível de água do rio, ou seja, como uma superposição de sinais periódicos de todas as frequências. O espectro da frequência por meio da transformada de Fourier é equivalente a uma superposição de sinais periódicos em todas as frequências. Os sinais  $1/f$  possuem também picos de todas as durações. Podemos dizer que são fractais no tempo.

A potência da componente de frequência é maior para as frequências menores, sendo inversamente proporcional à frequência, é por isso que o chamamos comportamento  $1/f$ , embora o correto seria tratá-lo como sinal e não como comportamento.

Sinais  $1/f$  são diferentes de ruídos brancos aleatórios, onde não existe correlação entre os valores que descrevem o sinal e, geralmente, incluem todas as frequências em iguais quantidades.

Existem casos em que o espectro de frequência não pode ser distribuído como sinal  $1/f$ , mas como  $1/f^\alpha$ , onde  $0 \leq \alpha \leq 2$ , embora continue com a denominação  $1/f$ .

### 2.2.4 Lei de Zipf

Em seu livro: “O comportamento humano e a lei do menor esforço”, o professor George Kingsley Zipf (Zipf, 1949) da Universidade Harvard fez várias observações para

algumas regularidades simples de sistemas na área de ciências humanas. Utilizando dados como o número de habitantes de cidades em que traçou um gráfico logarítmico *da classificação* (em ordem decrescente) pelo *número de habitantes*, obteve uma reta com inclinação muito próxima a 1. Observou similarmente outras cidades encontrando o mesmo comportamento gráfico.

Zipf observou também, com que frequência determinada palavra era usada na obra “Ulisses” de James Joyce, e numa coleção de jornais americanos, utilizando o mesmo critério logarítmico entre a *colocação* (ranking) pelo *número de frequência das palavras* e obteve o mesmo gráfico de regularidade: uma linha reta, com queda próxima da unidade 1, tal tipo de comportamento é hoje conhecido como a Lei de Zipf (Zipf, 1949).

As leis mencionadas são expressas como funções de distribuição para quantidades mensuráveis, portanto explicar as características físicas observadas em sistemas complexos seria explicar matematicamente as leis de potências implícitas, mais especificamente os valores dos expoentes. Em seu livro: “How nature works”, Bak (Bak,1997) utiliza o conceito de “*criticalidade auto-organizada*” para explicar as propriedades de sistemas complexos.

## CAPÍTULO 3

### DISTRIBUIÇÕES ESTATÍSTICAS

As distribuições estatísticas são de grande interesse para a Física e outras disciplinas, pois têm servido como base para o estudo de sistemas compostos por grande número de unidades.

Procuraremos demonstrar, neste capítulo técnicas e algumas distribuições estatísticas, comumente usadas para definir o estudo dos sistemas complexos.

#### 3.1 Distribuição Normal

No século XVIII a distribuição Normal foi considerada como uma aproximação da distribuição Binomial (Moivre,1733). Somente reconhecida no século XIX, devido aos trabalhos de Laplace e Gauss (Laplace, 1781; Gauss, 1816) e utilizada desde então como uma distribuição normal para qualquer distribuição estatística. Sendo hoje, base para a maioria dos trabalhos estatísticos.

Laplace estabeleceu pela primeira vez o teorema do Limite Central e fez grandes aplicações da Distribuição Normal em várias questões na teoria das probabilidades.

No Teorema do Limite Central (CLT) (Mantegna, 1994; Stanley, 1999) se  $x_1, x_2, x_3, \dots, x_n$  são valores independentes de uma variável  $x$ , com valor médio e desvio padrão finitos então a distribuição de  $S$ , onde

$$S = \sum_{i=1}^n x_i \quad (3.1)$$

será uma Distribuição Normal para  $n \gg \mathcal{G}$ .

Durante muito tempo, por influência dos trabalhos de Gauss e Laplace considerava-se como axioma que todas as distribuições estatísticas se aproximavam da distribuição normal e o desvio de qualquer variável aleatória de sua média era visto como um “erro” sujeito à Lei dos Erros expressa pela distribuição Normal, dada por:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\langle x \rangle)^2}{2\sigma^2}} \quad (3.2)$$

onde  $P(x)$  representa a densidade de probabilidade da variável  $x$ .  $\langle x \rangle$  representa a média aritmética da distribuição; e  $\sigma$  representa o desvio padrão da distribuição.

As estimativas do valor médio e o desvio padrão são dados por:

$$\langle x \rangle = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.3)$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \langle x \rangle)^2 \quad (3.4)$$



onde  $n$  representa o número de observações de  $x$ .

Considerando a variável:

$$U = \frac{x - \langle x \rangle}{\sigma} \quad (3.5)$$

$P(U)$  é dado por:

$$P(U) = \frac{1}{\sqrt{2\pi}} e^{-\frac{U^2}{2}} \quad (3.6)$$

para calcular a probabilidade de  $U$  menor que um valor  $\theta$ , definimos:

$$f(q) = P[0 \leq U \leq q] = \frac{1}{\sqrt{2\pi}} \int_0^q e^{-\left(\frac{1}{2}U^2\right)} dU \quad (3.7)$$

$\Phi(q)$  resulta na probabilidade da variável  $U$  ser menor do que  $q$ .

Geralmente, as bibliografias estatísticas (Magalhães, 2001; Vieira, 1980) consideram apenas valores positivos de  $q$ , pois  $P(q) = P(-q)$ .

### 3.2 Distribuições com Caudas Longas: Pareto e Distribuição Log-Normal

Se investigarmos uma distribuição de rendimentos dos indivíduos em um grande país, encontraremos um significativo número de pessoas com rendimentos 10 vezes maior que a média dessa distribuição e ainda algumas poucas pessoas com rendimentos 100 vezes maior que a média. Tais distribuições possuem caudas longas quando representadas graficamente e foram primeiramente estudadas por um italiano economista social Pareto que coletou dados estatísticos de rendimentos e riqueza individual em muitos países e em diversas épocas da história.

A distribuição de Pareto tem a forma

$$1 - p(x) = ax^{-\nu} \quad (3.8)$$

de modo que

$$\log [ 1 - p(x) ] = \log a - \nu \log x \quad (3.9)$$

e se o número de pessoas com rendimento maior ou igual a  $x$  for traçado no gráfico como uma função de  $x$  em log-log, o resultado da curva será uma linha reta com inclinação  $-\nu$ .

Em todos lugares e épocas, com economias estáveis, quando a origem das medições era de alta renda, a distribuição de rendimentos foi dada pela fórmula empírica  $y = ax^{-\nu}$  onde  $y$  representa o número de pessoas com rendimento maior ou igual a  $x$  e  $\nu$  aproximadamente igual a 1,5.

H.T. Davis (Davis, 1941) coletou e analisou dados de: ( i ) número de pessoas que ganham dinheiro com diferentes habilidades; (i.i) número de pessoas que escreveram

vários números de trabalhos matemáticos e (i.i.i.) o número de pessoas que atingiram vários escores no bilhar em 50 turnos. Notaram que em todos os casos, os dados foram bem definidos pela distribuição de Pareto.

B. Mandelbrot (Mandelbrot, 1960, 1963) investigou vários mecanismos que estão por trás da Distribuição de Pareto.

Outra distribuição que é frequentemente associada a distribuições com caudas longas é a distribuição Log-normal.

Que é definida por

$$p(x) = \frac{1}{x\mathbf{s}\sqrt{2\mathbf{p}}} e^{-\frac{(\log x - \langle \log x \rangle)^2}{2\mathbf{s}^2}} \quad (3.10)$$

Onde  $\langle \log x \rangle$  é valor médio de  $\log x$ .

Note que

$$\log p(x) = -\frac{1}{2} \log(2\mathbf{p}\mathbf{s}^2) - \frac{(\log x - \langle \log x \rangle)^2}{2\mathbf{s}^2} - \log x \quad (3.11)$$

Shockley (Shockley, 1957) propôs o seguinte mecanismo para explicar porque caudas longas existem em uma distribuição de progressão que exija um bem sucedido resultado para diversos empreendimentos, onde o fracasso de um provocaria o fracasso do

projeto. Ele usou a publicação de documentos técnicos como exemplo. Considerou como importantes algumas habilidades abaixo mencionadas.

1. Habilidade de levantar um bom tema;
2. Habilidade para trabalhar nele;
3. Habilidade para reconhecer soluções que valham a pena;
4. Habilidade de tomar decisões de quando parar ou levantar resultados;
5. Habilidade de redigir adequadamente;
6. Habilidade de aproveitar-se construtivamente das críticas;
7. Determinação para apresentar o trabalho em jornais;
8. Disposição para agüentar julgamentos de oposição.

Então, ele relata que a probabilidade de um pesquisador produzir um trabalho de sucesso em tempo determinado seria o produto de um conjunto de probabilidades que provocaria o sucesso do empreendimento, ou seja, possuindo cada uma das habilidades dos itens da relação anterior o sucesso seria evidente.

$$p = p_1 p_2 \dots p_8 \quad (3.12)$$

O aspecto log normal torna-se aparente ao considerarmos logaritmos em (3.12).

Logo:

$$\log p = \log p_1 + \log p_2 + \dots + \log p_8 \quad (3.13)$$

Desde que  $\log P$  seja a soma de um conjunto de variáveis, cada qual com sua própria distribuição de função, o teorema do limite central é aplicável de modo que a distribuição da função  $\log P$  poderia ser Gaussiana.

### 3.2.1 Exemplos de distribuição Log-Normal

**Geologia e mineração:** Na crosta terrestre, a concentração de elementos e sua radioatividade geralmente seguem a distribuição Log-normal (Razumovsky, 1940).

**Medicina:** Uma variedade de exemplos da medicina ajusta-se a distribuição Log-normal. Períodos latentes de doenças infecciosas têm sido freqüentemente demonstrados como Log-normal (Sartwell, 1950; Sartwell, 1952; Sartwell, 1996; Kondo, 1977): aproximadamente 70% de 86 exemplos publicados por Kondo (Kondo,1977) aparentam ser Log-normais. Sartwell (Sartwell, 1950; Sartwell, 1952; Sartwell, 1966) documenta 37 casos que se ajustaram à distribuição Log-normal.

**Meio Ambiente:** A distribuição de partículas químicas e orgânicas no meio ambiente são freqüentemente Log-normal (Biondini, 1976).

**Ecologia:** Na espécie animal (pássaros, peixes, insetos) e vegetal, a maioria de suas comunidades seguem a distribuição Log-normal (truncada) (Magurran, 1988).

**Ciências Sociais e Econômicas:** Exemplos de distribuição Log-normal em Ciências Sociais incluem: idade de casamento, tamanho de fazendas e rendimentos, em Economia, tamanho de firmas (Aitchison, Brown, 1957; Gibrat, 1931).

### 3.3 Distribuição De Lei de Potência Gradualmente Truncada

A distribuição de Lei de Potência possui desvio padrão infinito, embora os sistemas físicos reais apresentem desvio padrão finito. Ao aplicarmos então, a distribuição de Lévy, Pareto ou ainda em alguns casos de Log normal, precisamos truncar distribuições após um determinado valor, a fim de evitar um desvio padrão infinito.

Pareto explicou a distribuição de Lei de Potência baseando-se na realimentação positiva. De uma forma simples, um exemplo da economia seria de que o “dinheiro faz dinheiro”, ou seja, pessoas com grande rendimento pessoal conseguem em suas aplicações

financeiras melhores taxas de retorno, conseguindo dessa forma multiplicarem os seus rendimentos. Gupta e Campanha (Stanley et al., 1995; Gupta, Campanha, 2002) consideram que a validade da Lei de Potência tem um limite devido à capacidade física do sistema, portanto a realimentação positiva também deveria cessar após algum valor da variável. Como em sistemas complexos temos várias interações e grande número de componentes interagindo de maneiras diferentes, esperamos que o truncamento desta probabilidade seja gradual após certo ponto crítico. A distribuição proposta por eles é denominada ‘Distribuição de Lei de Potência Gradualmente Truncada’ e é dada por:

$$P_{GT}(x) = P(x)f(x) \quad (3.14)$$

onde  $P(x)$  é a distribuição da Lei de Potência e,  $f(x)$  é dada por

$$f(x) = \begin{cases} 1 & \text{se } x \leq x_c \\ \exp\left(-\left(\frac{x-x_c}{k}\right)^{2-a}\right) & \text{se } x \geq x_c \end{cases} \quad (3.15)$$

$x_c$  é o ponto crítico onde a suposta distribuição, i.é, a distribuição log-normal ou a distribuição de lei de potência começa a descolar da curva dos dados empíricos.  $\alpha$  é índice da lei de potência e  $k$  uma constante de truncamento gradual. Para valores de  $k$  menores, o truncamento será mais rápido. Este parâmetro foi ajustado para melhor adequação à curva empírica. Consideramos para a distribuição Log-normal  $\alpha = 0$ .

Esta distribuição tem desvio padrão finito e também variância finita, no limite obtém-se uma Distribuição Normal como exigido pelo Teorema do limite Central (Gupta, Campanha, 2000). Esta distribuição também obedece a Lei de Potência em sua parte central e decaimento exponencial nos valores extremos de  $x$ . Podemos considerá-la

equivalente à distribuição de Tsallis (Tsallis, 2000), obtidas pela termodinâmica para sistemas complexos. A Distribuição Gradualmente Truncada torna mais simples a extração de informações úteis que descrevem o sistema real (Gupta, Campanha, 2002).

A mesma teoria também foi aplicada à distribuição da bolsa de valores de NovaYork (S&P 500) (Gupta, Campanha, 2000), à distribuição das notas para o Vestibular da Unesp (Chavarette, 2002) e à distribuição do faturamento de firmas dos E.U.A. (log-normal gradualmente truncado). Em todos os casos houve concordância entre o modelo teórico e os dados empíricos.

### 3.4 Distribuição Log-normal gradualmente truncada

Na hipótese de truncar gradualmente a distribuição por causa da limitação física do sistema a distribuição log-normal discutida anteriormente, será dada por:

$$P(\ln x) = P_{\log\text{-normal}}(\ln x) f(x) \quad (3.16)$$

onde  $f(x)$  é dado por (3.16) considerando  $\alpha = 0$

Então,

$$P(\log x) = \frac{C}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log x - \langle \log x \rangle)^2}{2\sigma^2}\right\} f(x) \quad (3.17)$$

onde  $C$  é uma constante de normalização dada por

$$\int_{-\infty}^{\infty} P(\log x) d(\log x) = 1 \quad (3.18)$$

Assim como truncamos somente valores de  $\log x$ , onde  $P(x)$  é muito pequeno, o valor de  $C$  é muito próximo de 1.  $\sigma$  é o desvio padrão de  $\log x$  e  $f(x)$  é dado por

$$f(x) = \begin{cases} 1 & \text{se } x \leq x_c \\ \exp\left(-\left(\frac{x-x_c}{k}\right)^2\right) & \text{se } x > x_c \end{cases} \quad (3.19)$$

### 3.5 Técnica de Zipf (gráfico do tipo Zipf)

Infelizmente, temos poucos dados para valores extremos em distribuições estatísticas. Isso dificulta saber que tipo de distribuição ajustam-se estes valores, pois qualquer distribuição parece cabível dentro do limite de erro. Além disso, no caso de distribuição do faturamento de empresas, geralmente temos somente informações sobre as vendas das maiores empresas. Neste caso empregamos a técnica de Zipf, uma técnica estatística muito usada na Física e bastante útil (Zipf, 1949).

Supomos que  $x_1, x_2 \dots x_i$  seja um conjunto de  $n$  observações de uma variável  $x$ , cuja distribuição de probabilidade é dada por  $P(x)$ , e que estas observações sejam alinhadas em ordem decrescente de valores, então o índice  $i$  de  $x$  nos proporciona uma classificação das observações, que pode ser calculada por:

$$i = N \int_{x_i}^{\infty} P(x) dx \quad (3.20)$$



onde  $\int_{x_i}^{\infty} P(x)$  nos dá a probabilidade de observações que têm valores acima de  $x_i$ .

Multiplicando por  $N$  obtemos o número de observações acima de  $x_i$ , i.é., a classificação das observações de  $x_i$ .

Num gráfico do tipo Zipf usamos como abscissa  $\log i$  e por ordenada  $\log x_i$ , em seguida calculamos a curva teórica por meio da Equação (3.17) supondo uma distribuição para estas observações verificando assim o valor do parâmetro e o tipo de distribuição das mesmas, somente para alguns valores extremos.

Em casos da distribuição log-normal gradualmente truncada, esta relação será:

$$i = N \int_{x_i}^{\infty} \frac{C}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log x - \langle \log x \rangle)^2}{2\sigma^2}\right\} f(x) d(\log x) \quad (3.21)$$

onde:

$$f(x) = \begin{cases} 1 & \text{se } x \leq x_c \\ \exp\left(-\left(\frac{x-x_c}{k}\right)^2\right) & \text{se } x > x_c \end{cases} \quad (3.22)$$

Conhecendo  $\langle \log x \rangle$ ,  $\sigma$ ,  $x_c$  e  $k$  podemos ajustar a curva teórica entre  $\log i$  e  $\log x_i$ , podendo então confirmar a validade da teoria comparando-a com a curva empírica.

## CAPÍTULO 4

### 4.1 Sistema Econômico como um Sistema Complexo

A economia também não pode ser dissociada e estudada em partes. Neste sistema não existe a possibilidade de estudar o efeito de um único parâmetro em laboratório, como fazemos na física, pois temos fatores como política, psicologia, o desenvolvimento tecnológico e outros que sempre estão interagindo sem a possibilidade de separação.

Portanto a economia possui um sistema muito complicado sendo geralmente estudado de forma qualitativa. Somente após o desenvolvimento da Física de Sistemas Complexos, passou a ser estudada de forma semelhante a usada em sistemas físicos.

Há alguns anos, estudiosos da Física Estatística, iniciaram pesquisas na área econômica e financeira (Stanley, 1999; Axtell, 2001; Artur, 1999; Stanley et. al, 1995) sendo que dois tópicos causaram grande interesse entre eles: *o comportamento estatístico sobre a evolução de preços de mercado e a distribuição do crescimento e dinâmica de desenvolvimento de empresas.*

Mandelbrot (Mandelbrot, 1963) demonstrou que logaritmos de preços de mercado poderiam ser dados pela distribuição de Lei de Potência. Recentemente Gupta e Campanha (Gupta, Campanha, 2002) mostraram que os preços de mercado possuem propriedades fractais e que sua distribuição é dada pela Lei de Potência Gradualmente Truncada tanto

para curtos períodos de tempo em observações sucessivas, i.é, de um minuto como para longo período de tempo, ou seja, uma semana.

## 4.2 Distribuição do Tamanho de Firms

Estudos sobre a distribuição do tamanho de firmas foram iniciados em 1931, quando o economista francês Gibrat (Gibrat, 1931) sugeriu um modelo simples para explicar a distribuição do tamanho de empresas. As premissas básicas de seu modelo foram:

1. A taxa de crescimento de uma empresa ( $R$ ) é independente do tamanho desta. Na economia esta suposição é conhecida como “Lei de Efeito Proporcional”;
2. A taxa de desenvolvimento sucessivo de uma empresa independe do tempo;
3. Empresas não interagem entre si:

Este modelo pode ser escrito em linguagem matemática pelo processo estocástico:

$$S_{t+\Delta t} = S_t (1 + \epsilon) \quad (4.1)$$

onde  $S_t$  e  $S_{t+\Delta t}$  são respectivamente tamanho de empresas em relação ao tempo  $t$  e  $t + \Delta t$ .  $\epsilon$  é um número aleatório independente do tempo.

Este modelo proporciona uma distribuição Log-normal que também foi observado em muitos outros casos.

Stanley e seus colaboradores (Stanley et al, 1995) determinaram a distribuição do tamanho de empresas manufatureiras de capital aberto para o ano de 1993 nos EUA e

observaram um bom ajuste com a distribuição Log-normal exceto para as 100 maiores firmas.

Hart e Prais (Hart, Prais, 1956) estudaram 80.000 empresas na Inglaterra, e determinaram a mesma distribuição Log-normal. Axtell (Axtell, 2001) estudou todas as empresas do EUA e observou que neste caso a distribuição era dada pela Lei de Potência.

### **4.3 Distribuição do Tamanho de Firmas Manufatureiras dos EUA**

Relatamos agora, o trabalho de Gupta e Campanha (Gupta, Campanha, 2001), que utilizando os dados de Stanley e colaboradores (Stanley et al, 1995) observaram as vendas do ano de 1993 dos EUA conforme o anuário fornecido pela Compustat firms in SIC codes 2000-3999. Da distribuição Log-normal das vendas destas firmas, consideraram a média  $\mu = 7,71$  e o desvio padrão  $\sigma = 1,18$ , como havia sido considerado por eles também. Tanto na distribuição Log-normal como na distribuição de Lei potência, a probabilidade de ocorrer um valor muito alto é significativa, por isso é necessário truncar gradualmente esses valores, bem como na distribuição Log-normal Gradualmente Truncada como foi apresentado na seção 3.4 do capítulo anterior.

Gupta e Campanha consideraram  $x_c$  igual a  $8 \times 10^9$  como desvio inicial destes pontos e  $k$  igual a  $1,5 \times 10^{11}$ . Representaram graficamente o log da classificação pelo log das vendas. O resultado da análise se apresentou de acordo com a teoria, para efeito de comparação traçaram também a distribuição Log-normal (Figura 1). Concluíram que o fator limitante era a capacidade física do mercado e as condições favoráveis que contam os dirigentes das grandes empresas. Ao considerarmos a distribuição Log-normal, a projeção das vendas destas firmas situa-se na ordem de  $30 \times 10^{12}$  dólares, cerca de três vezes a produção nacional (valor bruto) dos EUA por ano, o que seria impossível. Considerando que

- (i) A produção Nacional inclui todos os serviços e manufaturados;
  
- (ii) A maior parte das indústrias não manufatura o produto final (Muitas vezes o produto final de algumas empresas é o inicial de outras). No modelo de distribuição 'Log-normal gradualmente truncada' as vendas situam-se na ordem de  $10 \times 10^{12}$  dólares, o que parece razoável, pois consta entre a ordem de produção nacional americana.

Concluimos então que as vendas de muitas empresas não dependem somente do seu próprio potencial de desenvolvimento, mas também da disposição do mercado. Julgamos então que como na distribuição de Lei de Potência a distribuição Log-normal deveria ser também gradualmente truncada devido aos limites do sistema físico.

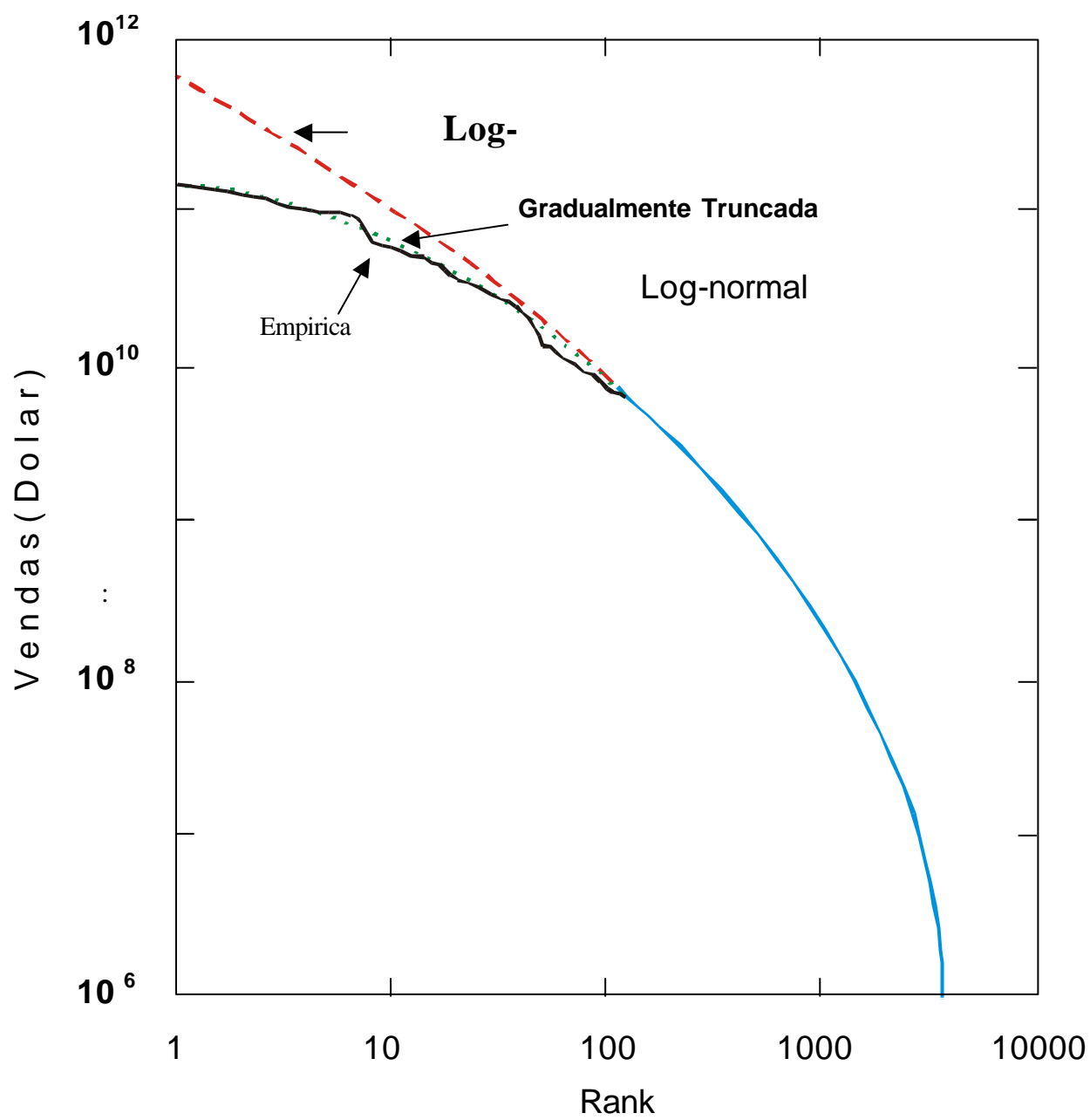


Figura 1: Classificação de Firmas Manufatureiras dos EUA.  
Ano de 1993.

## Capítulo 5

### Distribuição do Faturamento de Empresas Brasileiras

Neste capítulo, estudaremos a Distribuição Estatística do tamanho de empresas brasileiras, considerando como tamanho destas o faturamento em dólares. Esses dados foram fornecidos pela FIPECAFI (Fundação Instituto de Pesquisas Contábeis, Atuariais e Financeiras) da Universidade de São Paulo, responsável pelas análises financeiras da Revista Exame, Edição Especial: “Melhores e Maiores – As 500 maiores empresas do Brasil”. Somente os dados de 500 empresas, eram insuficientes para este trabalho, portanto solicitamos a FIPECAFI o número total de empresas pesquisadas, no que fomos gentilmente atendidos.

No Brasil não existe um banco de dados com o faturamento de todas as empresas, pois se trata de informação sigilosa. Consideramos então apenas empresas de Capital Aberto, subtraindo as empresas estatais, devido à reserva de mercado que possuem, pois não entram em competição real com as demais. Obtivemos então, os dados de aproximadamente 2000 empresas para os anos de 1996, 1997, 1998, 1999 e 2000.

#### 5.1 Análise empírica

Como não tínhamos informações sobre o faturamento de todas as empresas brasileiras, fomos levados a utilizar a Técnica de Zipf (como discutimos anteriormente

usamos esta técnica quando temos informações somente dos maiores elementos da distribuição).

Para realizarmos as análises, utilizamos o software *Mathematica*® 4.1 (Anexo 1) para descrever os parâmetros e as propriedades das distribuições, tais como localização, dispersão e forma.

Como procedimento da Técnica de Zipf classificamos as firmas em ordem decrescente de faturamento para cada ano e representamos graficamente o *log* do faturamento pelo *log* da classificação, demonstrados nas Figuras de 2 a 6.

Uma observação simples destas figuras mostra que as distribuições são definidas pela Log-normal. Calculamos a média e o desvio padrão de *log x* por meio das observações empíricas. Utilizando estes valores desenhamos então, a curva teórica, onde  $x_c$  é o ponto crítico, ou seja, o início do descolamento entre a curva com os dados empíricos e a curva teórica da Distribuição Log-normal Gradualmente Truncada. Ajustamos o valor de *k* para uma melhor adaptação da curva teórica com a empírica.

Para firmas dentro da classificação até 2000 empresas, ajustamos o parâmetro de distribuição gradualmente truncada, ou seja,  $x_c$  e *k*, para definir melhor o ajustamento. Na figura 7 a 11 comparamos as curvas teóricas com as empíricas.

Utilizamos os seguintes parâmetros:

Ano	$m = \langle \log x \rangle$	$\sigma$	$x_c$	<i>k</i>
1996	7,72	0,70	$7,94 \times 10^9$	$5 \times 10^9$
1997	7,74	0,66	$7,94 \times 10^9$	$5 \times 10^9$
1998	7,80	0,66	$7,94 \times 10^9$	$5 \times 10^9$
1999	7,74	0,68	$3,98 \times 10^9$	$5 \times 10^9$
2000	7,75	0,70	$5,01 \times 10^9$	$5 \times 10^9$

Tabela 1: Parâmetros utilizados nos gráficos de Classificação de Empresas Brasileiras para os anos de 1996, 1997, 1998, 1999 e 2000.



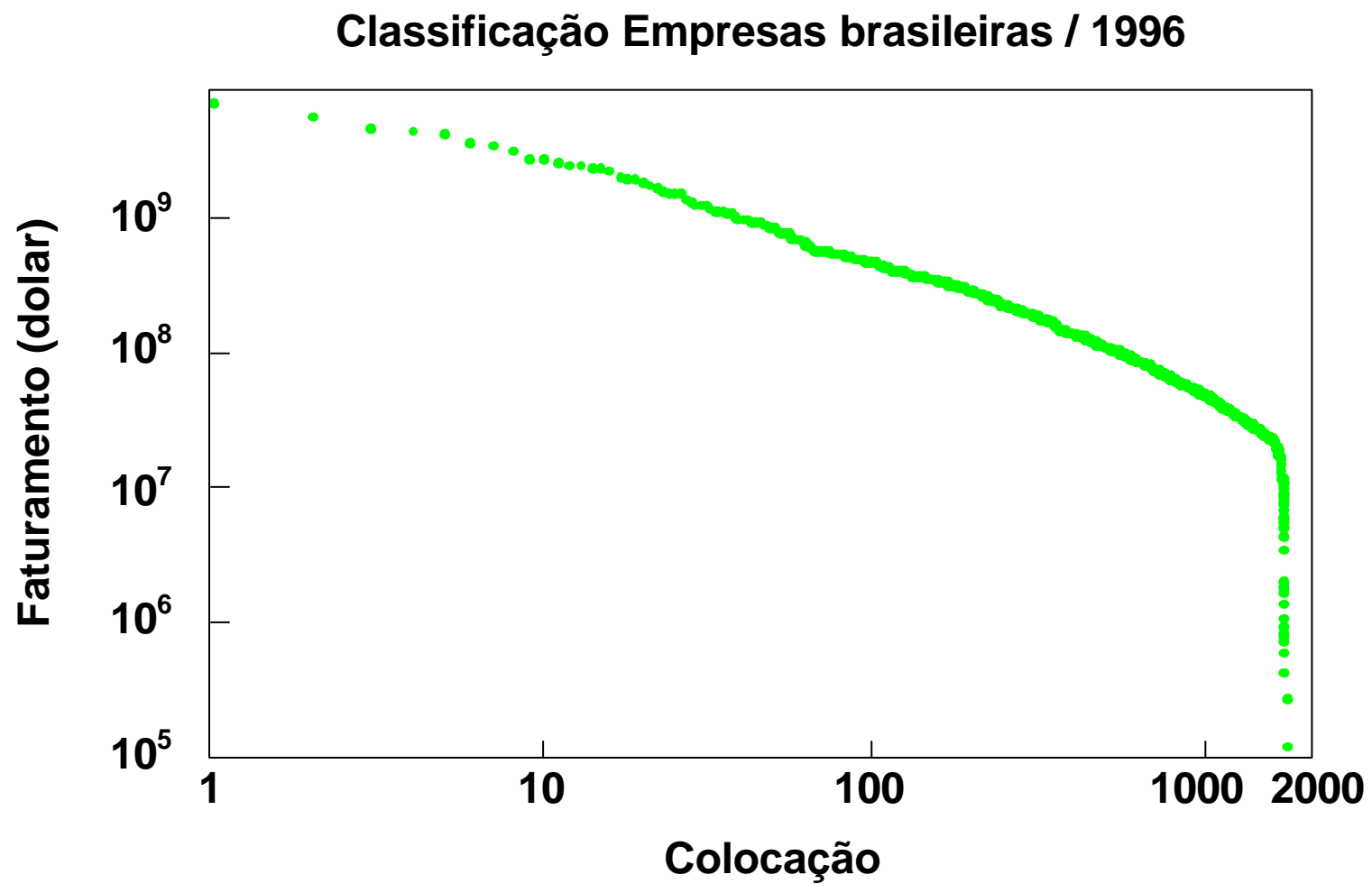


Figura 2: Curva Empírica da Classificação de Empresas Brasileiras no ano de 1996.  
Faturamento das Empresas versus a colocação das mesmas

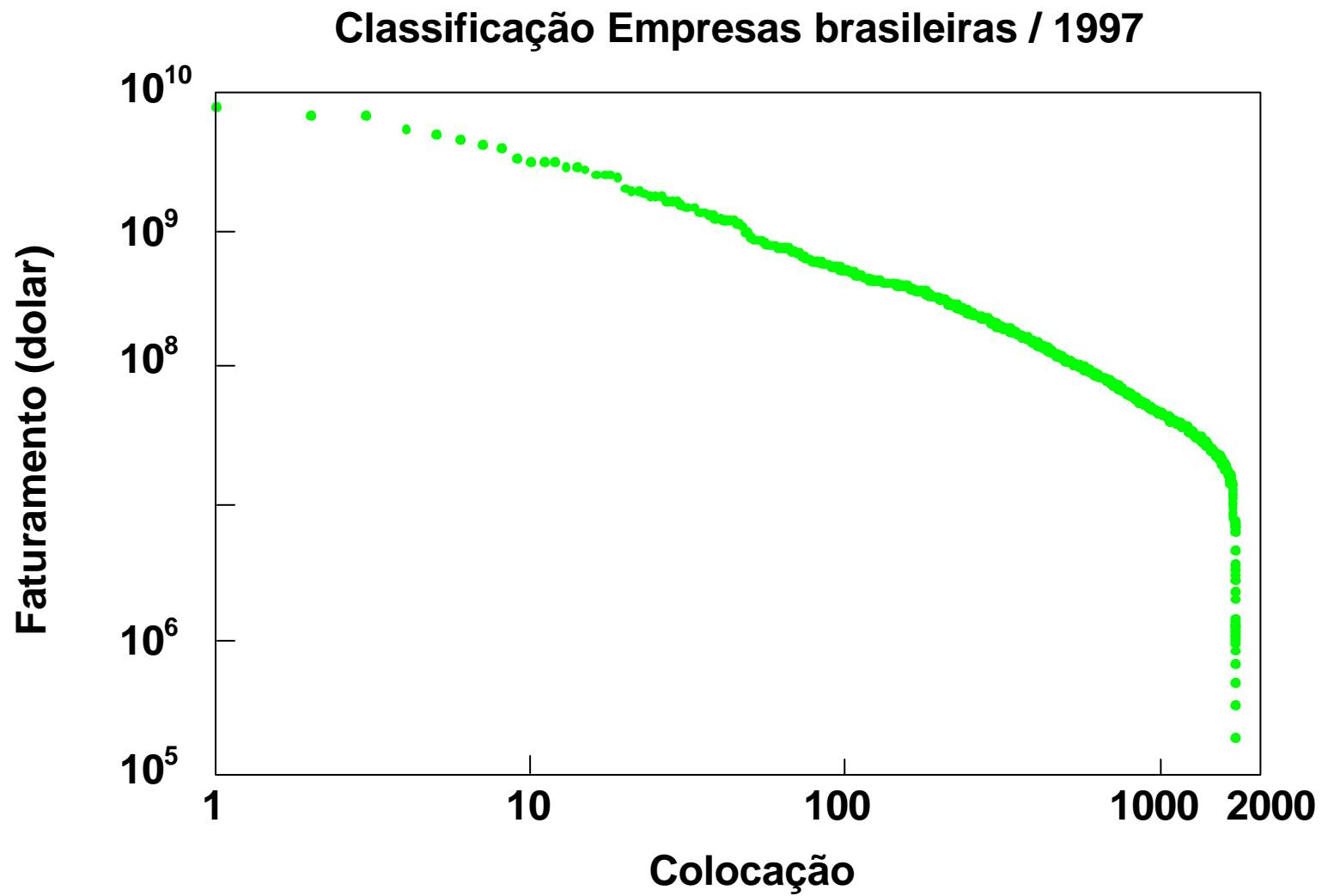


Figura 3: Curva Empírica da Classificação de Empresas Brasileiras no ano de 1997. Faturamento das Empresas versus a colocação das mesmas.

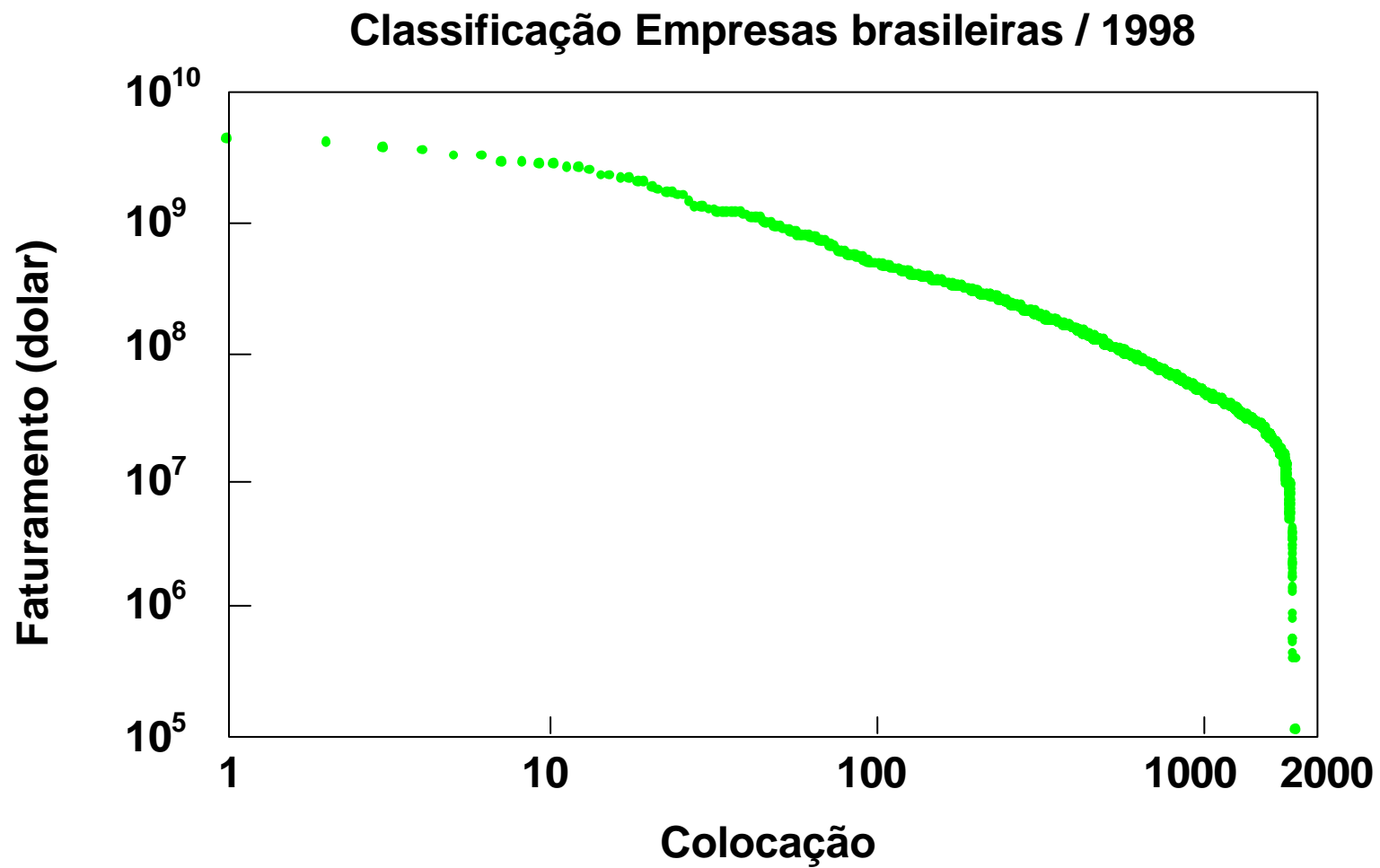


Figura 4: Curva Empírica da Classificação de Empresas Brasileiras no ano de 1998.  
Faturamento das Empresas versus a colocação das mesmas.

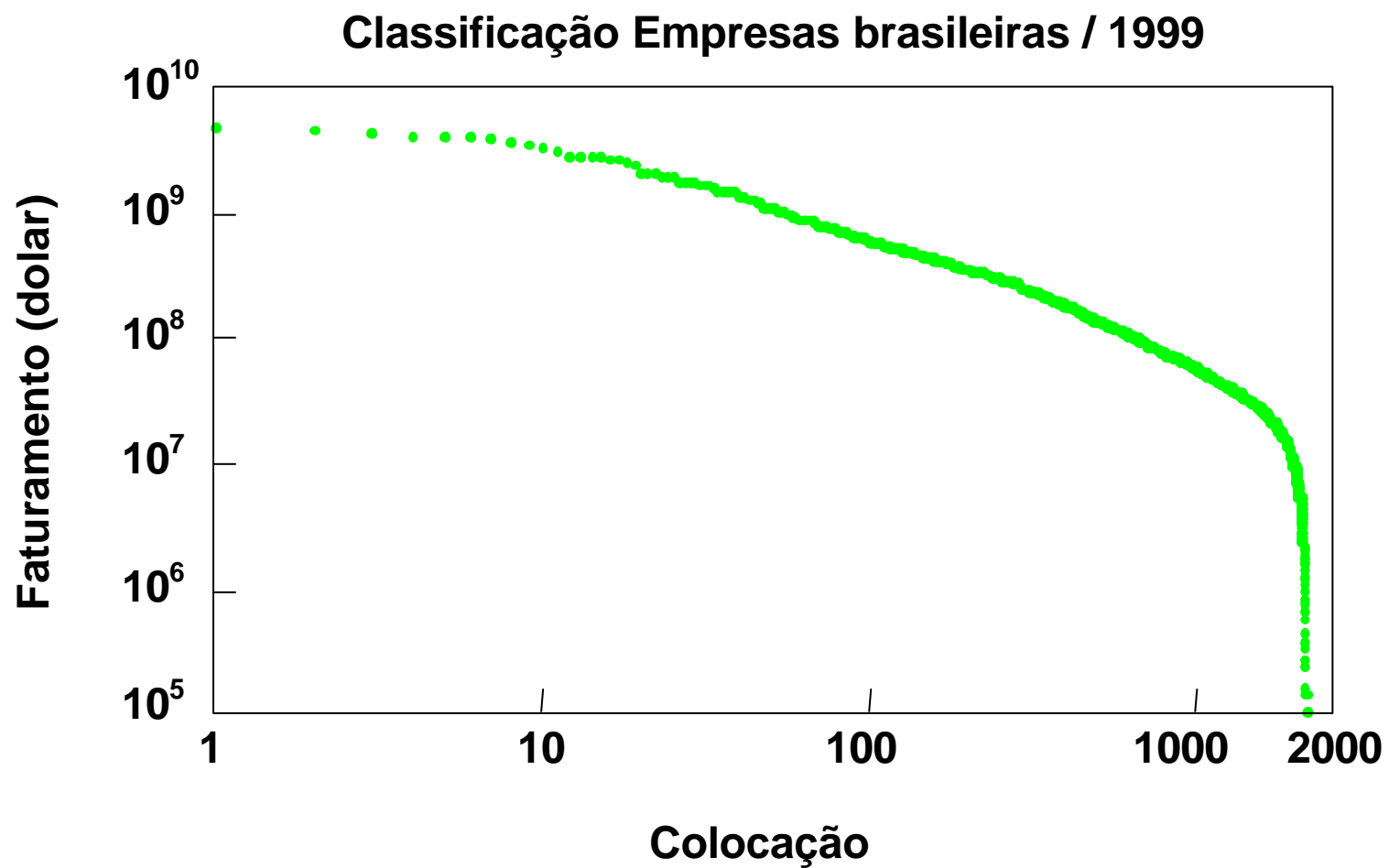


Figura 5: Curva Empírica da Classificação de Empresas Brasileiras no ano de 1999. Faturamento das Empresas versus a colocação das mesmas.

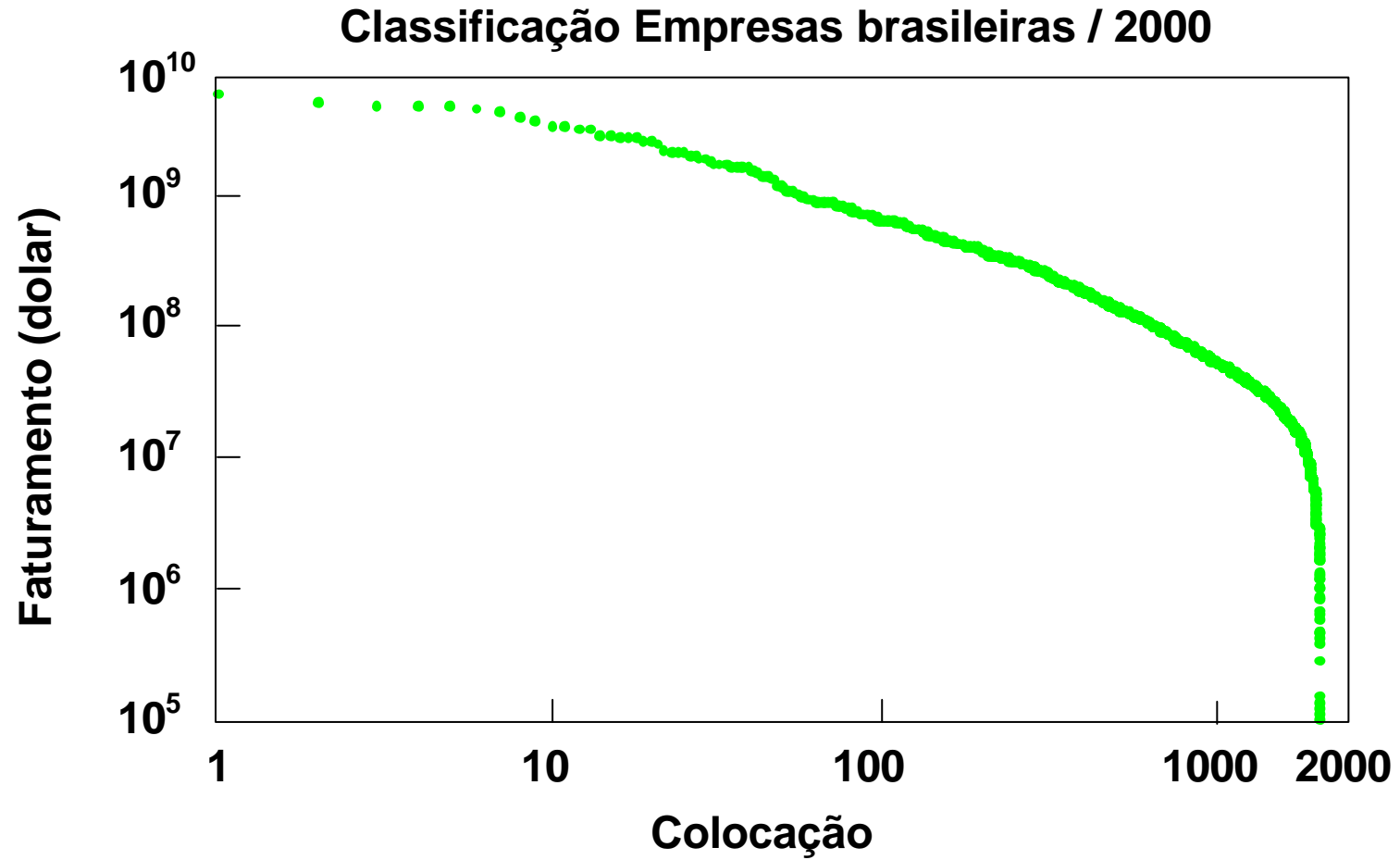


Figura 6: Curva Empírica da Classificação de Empresas Brasileiras no ano de 2000. Faturamento das Empresas versus a colocação das mesmas.

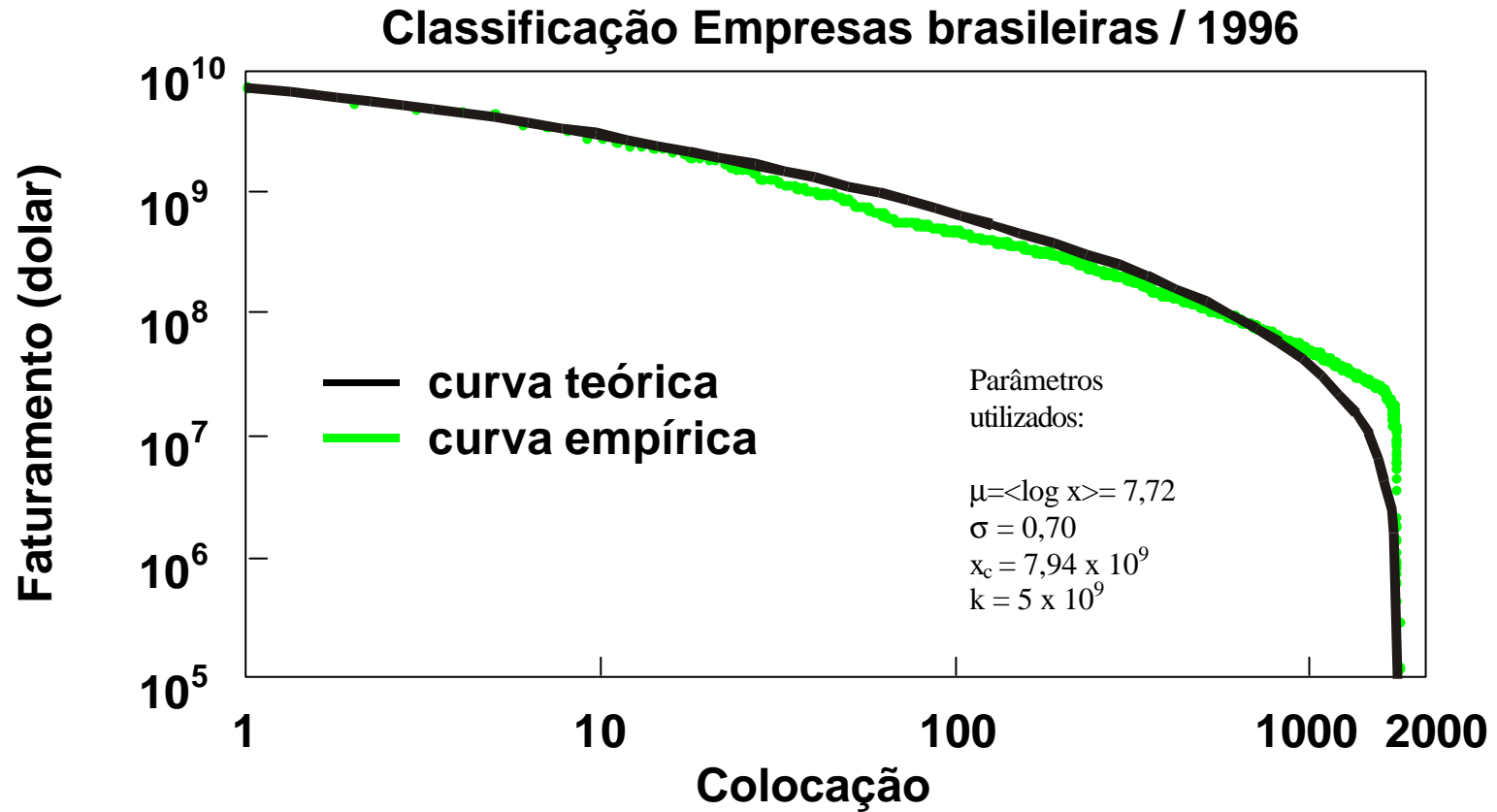


Figura 7: Curva Empírica e Teórica da Classificação de Empresas Brasileiras do ano de 1996. Faturamento das Empresas versus a colocação das mesmas.

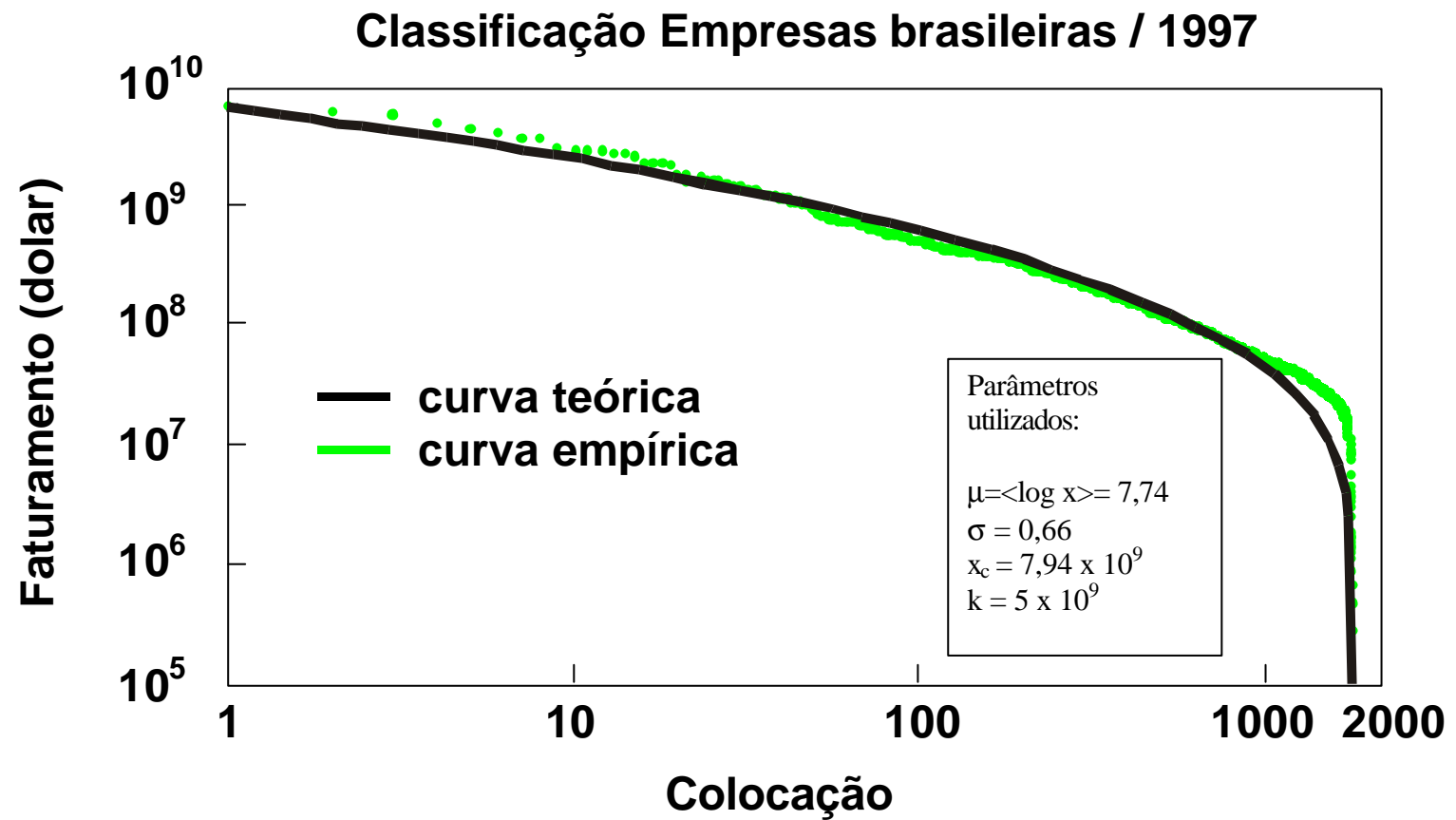


Figura 8: Curva Empírica e Teórica da Classificação de Empresas Brasileiras do ano de 1997. Faturamento das Empresas versus a colocação das mesmas.

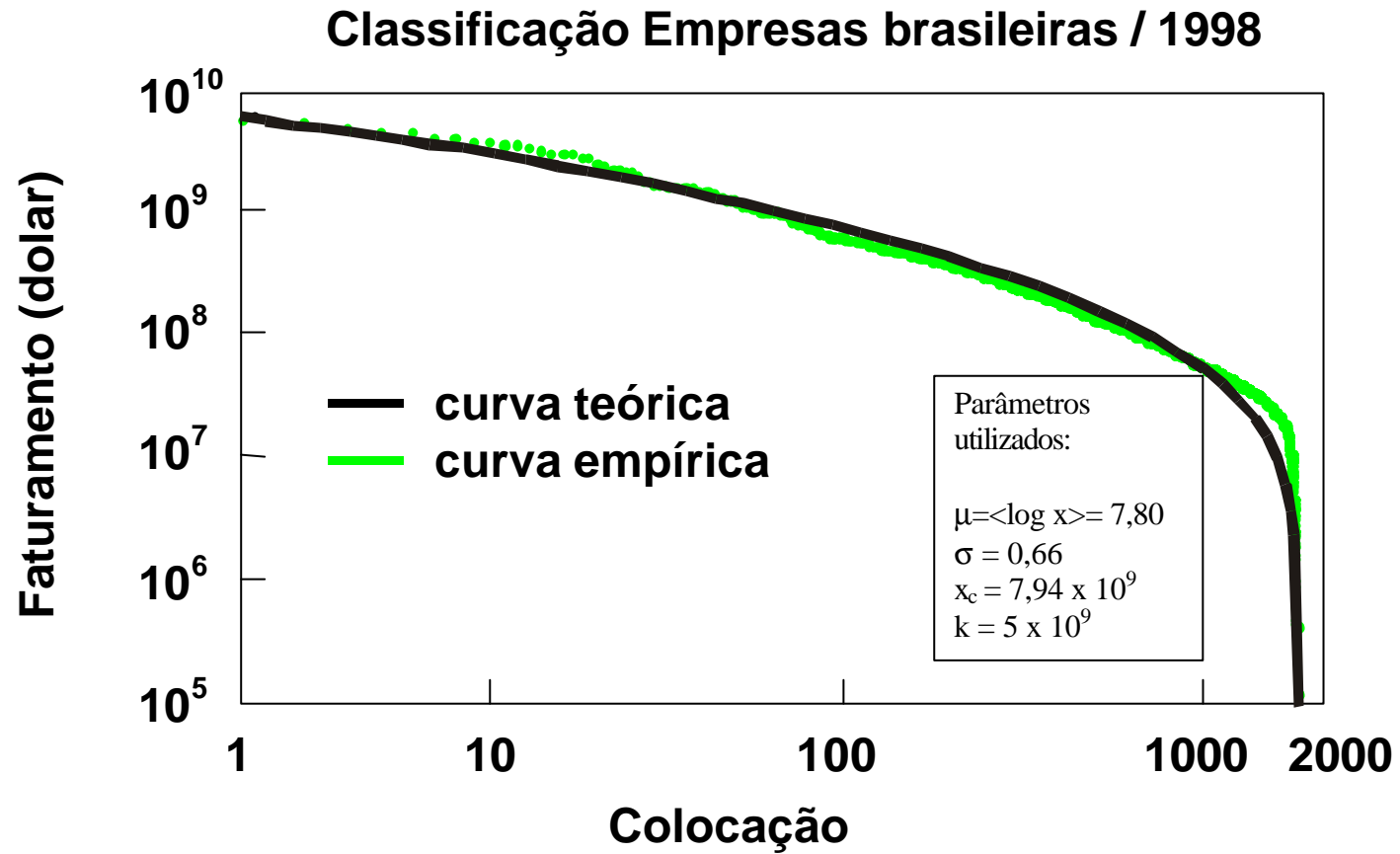


Figura 9: Curva Empírica e Teórica da Classificação de Empresas Brasileiras do ano de 1998. Faturamento das Empresas versus a colocação das mesmas.



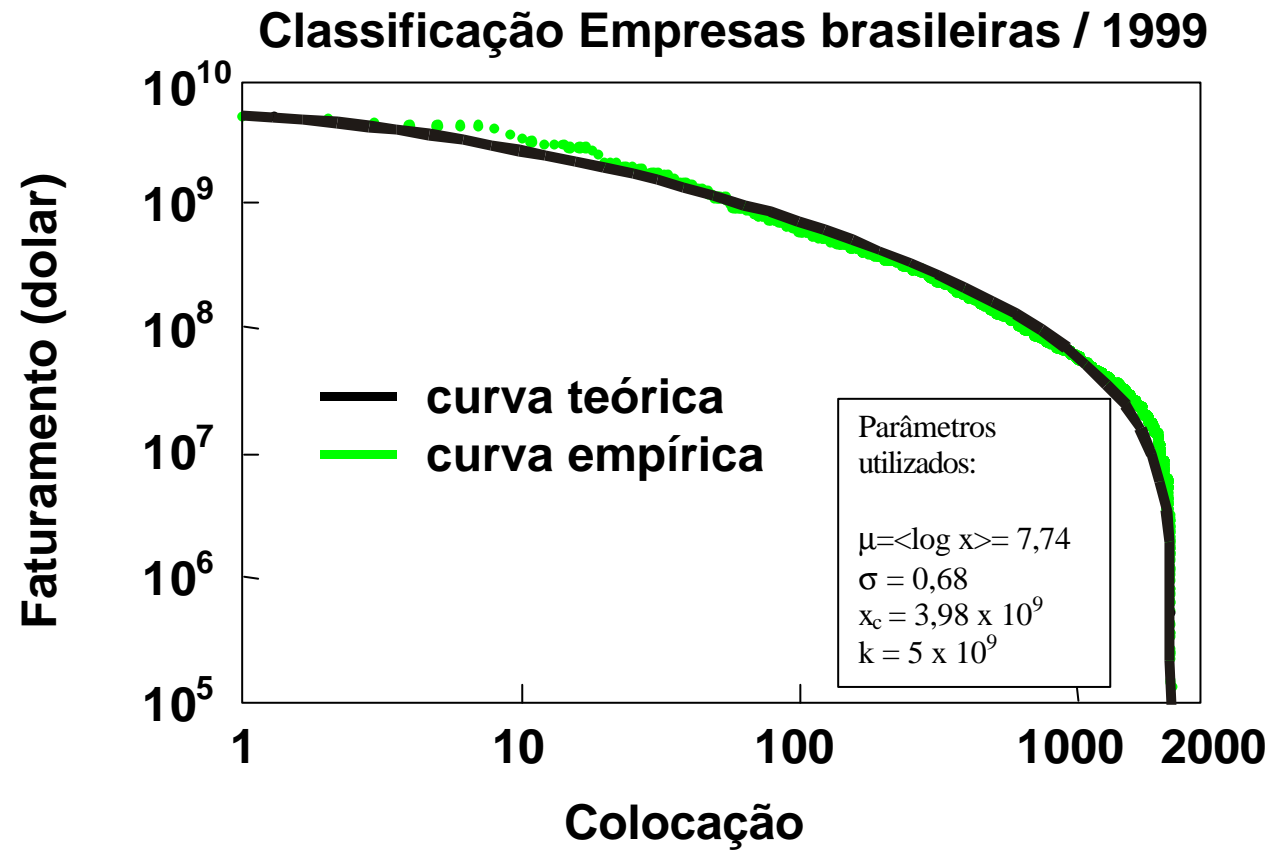


Figura 10: Curva Empírica e Teórica da Classificação de Empresas Brasileiras do ano de 1999. Faturamento das Empresas versus a colocação das mesmas.

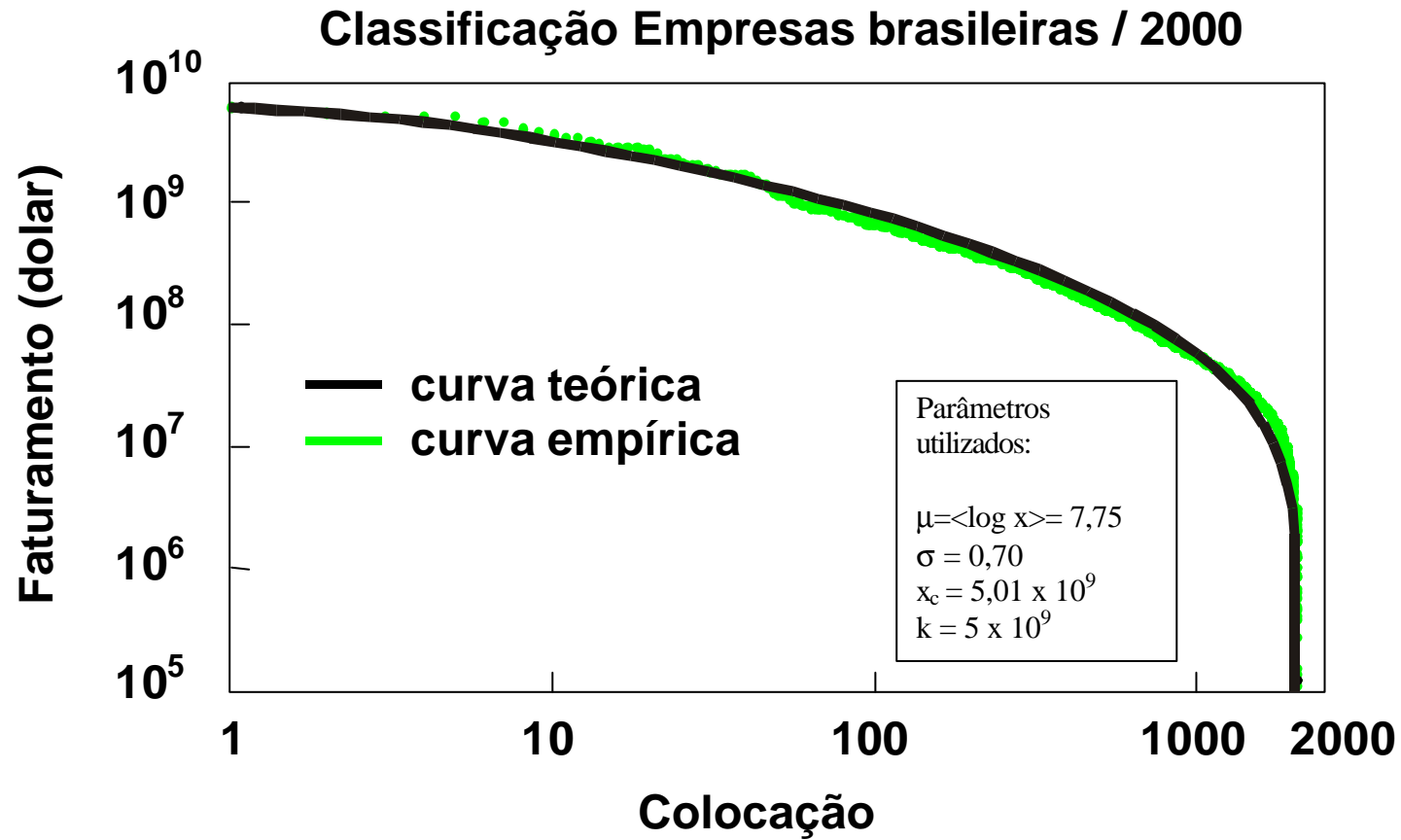


Figura 11: Curva Empírica e Teórica da Classificação de Empresas Brasileiras do ano de 2000. Faturamento das Empresas versus a colocação das mesmas.

## 5.2 Análise e discussão

Observamos que para os 5 anos estudados a distribuição é bem definida pela distribuição Log-normal e não pela distribuição de Lei de Potência como foi proposto por Axtell. Se a distribuição fosse pela lei de Potência o gráfico do log do faturamento e o log da colocação seria uma linha reta e não uma curva o que não é o caso. Estes dados estão de acordo com a teoria de Gibrat em que o crescimento proporcional de firmas é aleatório e dado pela distribuição Log-normal.

Notamos que o valor médio de  $\log x$  e o desvio padrão são praticamente os mesmos em todos os anos analisados, i.é, o valor médio de  $\log x$  fica em torno de 7,75 que corresponde ao faturamento de  $5,62 \times 10^7$  dólares e desvio padrão 0,68. Os mesmos valores para distribuição de empresas americanas em 1993 foram discutidos no capítulo anterior com 7,71 e 1,18 para média e desvio padrão respectivamente.

Concluimos que o desvio padrão da distribuição das empresas brasileiras é bem menor que o das empresas americanas. Fato que talvez se deva à política econômica mais agressiva e competitiva dos EUA. Podemos ilustrar como exemplo, a variedade de rendimentos dos comerciantes, de zero à milhões de dólares, provocando desvio padrão elevado devido a acirrada competição comercial. Por outro lado temos o salário dos professores, que em geral não apresenta grande variação, uma vez que não existe competição tão acentuada nesta profissão, como acontece entre comerciantes, jogadores de futebol, músicos etc.

O valor de corte  $x_c$  da Classificação das Empresas Brasileiras é da ordem de  $7 \times 10^9$  aproximando-se do valor de corte da distribuição das empresas americanas que foi de  $8 \times 10^9$ . Como o valor de corte da classificação das empresas brasileiras também é muito próximo do faturamento das empresas colocadas em 1º e 2º lugar, não observamos claramente o truncamento gradual da distribuição em questão, portanto, não estamos confiantes sobre o valor dos parâmetros  $x_c$  e  $k$ , unicamente podemos afirmar que  $x_c$  deveria ser acima de  $5 \times 10^9$ .

Na distribuição de classificação de empresas americanas, aproximadamente 100 delas possuíam faturamento superior ao valor de corte sendo visível o efeito de truncamento gradual. A economia brasileira é bem menor que a economia americana (corresponde a aproximadamente a 10%). As sedes das maiores empresas multinacionais estão nos EUA ou na Europa, sendo assim, empresas com faturamento na ordem de  $10^{11}$  dólares/ano que figuram somente no mercado econômico dos EUA e Europa ao contrário do Brasil e outros países menos desenvolvidos. Na distribuição de empresas brasileiras o faturamento da 1ª colocada está na ordem de  $10^{10}$  dólares/ano o que justifica não observarmos o efeito de truncamento gradual nitidamente nesta distribuição, pois os dados são insuficientes para definir graficamente a curva.

## CAPÍTULO 6

### **Conclusões Gerais**

Normalmente as grandes indústrias são multinacionais, por exemplo, aqui no Brasil os primeiros colocados são Volkswagen, Carrefour, Telefônica, Shell, General Motors, etc. Essas indústrias possuem excelentes administradores de empresas, assim têm condições favoráveis para conquistar o mercado de outros países. Na teoria da distribuição gradualmente truncada proposta anteriormente observamos que o truncamento gradual é devido a limitação física do mercado e ou das condições oferecidas aos dirigentes por parte das empresas. Como tais fatores são os mesmos tanto para empresas americanas como para brasileiras já que trabalham no mesmo ambiente mundial, prevíamos o mesmo valor de corte observado em torno de  $7 \times 10^9$  dólares. Neste sentido concluímos que a distribuição por faturamento de empresas privadas seguiria o comportamento da distribuição Log-normal Gradualmente Truncada. Presumimos que a representação gráfica da distribuição Log-normal Gradualmente Truncada deve ser mais evidente para a distribuição de empresas do Japão e da Europa, que possuem suas economias bem desenvolvidas e muitas empresas com faturamento anual acima de 10 bilhões de dólares. Por enquanto, para a distribuição de empresas brasileiras a distribuição Log-normal, dada por Gibrat é suficiente para descrevê-la.

Sob o ponto de vista da Física, consideramos distribuição de empresas como um sistema complexo, discutimos as técnicas da Física estatística, usamos a técnica de Zipf para analisar a distribuição por faturamento de empresas brasileiras e comparamos com o estudo da distribuição das empresas americanas.

## Referências:

AITCHISON, J.; BROWN, J.A.C. **The Log-normal Distribution**. Cambridge: Cambridge United Press, 1957.

ARTHUR, W.B. Complexity and the Economy. **Science**, Washington, v.284, n.5411, p.107-109, 1999.

AXTELL, R.L. Zipf Distribution of U.S. Firm Size. **Science**. Washington, DC, v.293, n.5536, p.1818-1820, 2001.

AYRES, M.; AYRES JR, M.; AYRES, D. L.; SANTOS, A.S. **Bioestat 2.0**: aplicações estatísticas nas áreas das ciências biológicas e médicas. Belém: Sociedade Civil Mamirauá; MCT- CNPq, 2000.

BADII, R.; POLITI, A. **Complexity–Hierarchical structures and scaling in physics**. Cambridge, UK, 1997

BAK, P. **How nature works**: The science of self-organized criticality. Oxford: University Press, 1997.

BIONDINI, R. Cloud motion and rainfall statistics. **Journal of Applied Meteorology**, Charlottesville, v.15, p. 205-224, 1976.

CHAVARETTE, F. R. et al. Um breve comentário sobre o vestibular da UNESP: Um exemplo de modelo dinâmico aleatório de sistema complexo. IN: CONGRESSO TEMÁTICO DE DINÂMICA, CONTROLE E APLICAÇÕES DA SOCIEDADE BRASILEIRA DE MATEMÁTICA APLICADA E COMPUTACIONAL, 1., 2002, São José do Rio Preto. **Anais...** São José do Rio Preto: SBMAC, 2002. p.879-886.

CHAVARETTE, F. R. **A Distribuição de Lei de Potência Gradualmente Truncada na Educação: Exame de Vestibular da UNESP.** 2002. 60 f. dissertação (mestrado em Física Aplicada) - Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Rio Claro, 2002.

DAVIS, H. T. **The Theory of Econometrics.** USA: Bloomington Indiana, 1941.

EINSTEIN, A., INFELD, L., **The Evolution of Physics: From Early Concepts to Relativity and Quanta.** USA: Simin & Schuster, 1976.

FECHNER, G.T. **Elemente der Psychophysik.** Leipzig: Breitkopf und Härtel, 1860.

FECHNER, G.T. **Kollektivmasslehre.** Leipzig: Engelmann, 1897.

FELLER, W. **An introduction to probability theory and its applications.** New York: Wiley, 1971.

GADDUM, J. H. **Reports on biological standards III:** methods of biological assay depending on a quantal response. London: Medical Research Council, 1945.

GAUSS, C. F. Bestimmung der genauigkeit der beobachtungen. **Zeitschrift für Astronomie**, Berlin, v.1, p.185-197, 1816.



GIBRAT, R. **Les inégalités économiques**. Paris: Librairie du Recueil Sirey, 1931.

GUPTA, H. M.; CAMPANHA, J. R. The gradually truncated Lévy flight: stochastic process for complex systems. **Physica A**, New York, v. 268, p.231-239, 1999

GUPTA, H. M.; CAMPANHA, J. R. The gradually truncated Lévy flight: stochastic process for complex systems. **Physica A**, New York, v.275, p.531-543, 2000.

GUPTA, H.M.; CAMPANHA, J. R. Tsallis statistics and gradually truncated Lévy flight – distribution of an economical index. **Physica A**, New York, v. 309, p. 381-387, 2002.

GUPTA, H. M.; CAMPANHA, J. R. **Gradually truncated log-normal distribution – size distribution of firms**. arXiv.org e-Print archive, v.1, 30 nov. 2001. Disponível em: <http://arxiv.org/pdf/cond-mat/0111579>.

HART, P.E.; PRAIS, S. J. 1956, The analysis of business concentration: A statistical approach. **Journal of the Royal Statistical Society**, Series A, v.119, 150-181.

HURST, H. E. Long-term storage of reservoirs. **Transactions of the American Society of Civil Engineers**, Reston, v.116, p.770-779, 1951.

KAPTEYN, J. C.; VAN UEN, M. J. **Skew frequency curves in biology and statistics**. Groningen: Astronomical Laboratory at Groningen, 1916.

KONDO, K. The log-normal distribution of the incubation time of exogenous diseases. **Japanese Journal of Human Genetic**, Nigata, v.21, n. 4, p.217-237, 1977.

LAPLACE, P. S. Mémoire sur les probabilités. **Histoire de l'Académie Royale de Sciences**, Paris, v.9, p.227-332, 1781.

LÉVY, P. **Théorie de l'addition des Variables Aléatoires**. Paris: Gauthier-Villars, 1937.

MACAU, E.E.N. Sistemas Complexos. **Anais do I Congresso de Dinâmica e Aplicações**. Rio Claro, v.1, p.29-49, 2002.

MAGALHÃES, M. V.; LIMA, A.C.P. **Noções de probabilidade e estatística**. 3.ed. São Paulo: IME/USP, 2001.

MAGURRAN, A.E. **Ecological Diversity and its Measurement**. London: Croom Helm, 1988.

MANDELBROT, B. The variation of certain speculative prices. **Journal of Business of the University of Chicago**, Chicago, v.36, p.394, 1963.

MANDELBROT, B. **The fractal geometry of nature**. New York: Freeman, 1983.

MANDELBROT, B. **Multifractals and 1/f Noise: Wild Self-Affinity in Physics**. Springer-verlag, 1999.

MANTEGNA, R.; STANLEY, H. Stochastic Process with Ultraslow Convergence to a Gaussian: The truncated Lévy Flight. **Physical Review Letters**. N.Y., v.73, n.22, p. 2946-2949, 1994.

Mc-ALLISTER, D. The law of the geometrical mean. **Royal Society of London**, London, v.29, p.369-375, 1879.

MOIVRE, A. de. **Approximatio ad summam terminorum binomii(a+b)<sup>n</sup> in seriem expansi**. London: Supplement to Miscellanea Analytica, 1733.

PARETO, V. **Cours d'Economic Politique**. Reprint as a volume. Ouvres Completes. Droz Geneur, 1896-1965.

POON, L.; GREBOGI, C. Controlling Complexity. **Physical Review Letters**. N.Y., v.75, p. 4023-4026, 1995.

RAZUMOVSKY, N. K. Distribution of metal values in ore deposits. **Doklady Akademii**, Moscou, v.28, n.9, p.814-816, 1940.

SARTWELL, P.E. The distribution of incubation periods of infectious disease. **American Journal of Hygiene**, Baltimore, v.51, n. 3, p.310-318, 1950.

SARTWELL, P.E. The incubation period of poliomyelitis. **American Journal of Public Health and the National Health**, Washington, v.42, n. 11, p.1403-1408, 1952.

SARTWELL, P.E. The incubation period and the dynamics of infectious disease. **American Journal of Epidemiology**, Baltimore, v. 83, n. 2, p. 204-216, 1966.

SHOCKLEY W. **Log-Normal Distribution Citation index** Proceeding of. IRE, USA, v **45**, p. **279**, 1957

SINGH, A; WHITTINGTON, G. The size and Growth of firms. **Review of Economic Studies**, Bristol, v.42, n. 1, p.15-26, 1975.

STANLEY, H.E. Scaling, universality, and renormalization: Three pillars of modern critical phenomena. **Reviews of Modern Physics**, v.71, n. 2, p. 5358-5366, 1999.

STANLEY, M.H.R.; BULDYREV, S.V.; HAVLIN, S.; MANTEGNA, R.N.; SALINGER, M.A.; STANLEY, H.E. Zipf plots and the size distribution of firms. **Economics Letters**, Boston, Boston University, v.49, p.453-457, 1995.

TSALLIS, C. Nonextensive Statistics: Theoretical, Experimental and Computational Evidences and Connections. **Brazilian Journal of Physics**, SP, v.29. p.1-45, 1999.

TSALLIS, C. As distribuições de Lévy. **Revista Brasileira de Ensino de Física**, São Paulo, v.22, p.156-162, 2000.

VIEIRA, S. **Introdução à bioestatística**. 4. ed. São Paulo: Campus, 2000.

WEBER, H. **De pulsa resorptione auditu et tactu. Annotationes anatomicae et physiologicae**. Leipzig: Koehler, 1834.

ZIPF, G. K. **Human behavior and the principle of least effort**. Cambridge: Addison-Wesley, 1949.

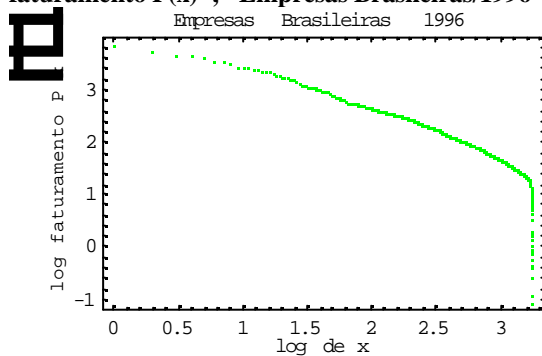
## Anexo 1

Software Matemática® 4.1 utilizado para desenvolver a análise dos dados da Distribuição do Faturamento de Empresas Brasileiras.

```
<< Statistics`DescriptiveStatistics`  
<< Statistics`DataManipulation`  
<< Graphics`Graphics`  
SetDirectory "C:\vedn" |  
priv=ReadList["96priv.dat"]*1000;
```

```
empilog = Reverse List 10, priv |  
eixox=Range[1776];  
X=Log[10,eixox]/N;
```

```
transpriv=Transpose[{X,empilog}];  
graph0=ListPlot[transpriv,PlotRange->All,PlotStyle->RGBColor[0,1,0]];  
Show[graph0, Frame->True, AxesOrigin->{-2,0}, Axes->False, FrameLabel->{"log de x", "log  
faturamento P(x)", "Empresas Brasileiras/1996", ""}]
```



...Graphics ...

```
LocationReport[empilog]
```

```
Mean@ 1.83992, HarmonicMean@ 1.84524, Median@ 1.76117
```

```
DispersionReport[empilog]
```

```
Variance@ 0.247751, StandardDeviation@ 0.497746, SampleRange@ 4.91086,  
MeanDeviation@ 0.374824, MedianDeviation@ 0.288555, QuartileDeviation@ 0.30365
```

```
A= 1776
```

```
B= 2.507
```

```
Fun1 x_?media_>, desvio_?desvio_> := desvio_?desvio_> - media_?media_> * desvio_?desvio_>^2
```

```
Fun2 x_?media_>, desvio_?desvio_> := Integrate Fun1 x_?media_>, desvio_?desvio_>, {x, 10
```

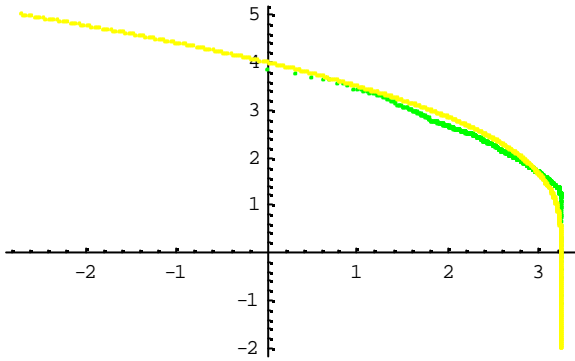
```
listax=Table Fun2 x_?media_>, {1.78, 0.68, 2, 5, 0.01}
```

```
listalog=Log 10 listax
```

```
listax=Table {1.78, 0.68, 2, 5, 0.01}
```

```
trans1=Transpose listalog, listax
```

```
Show graph0 graph1
```



...Graphics...

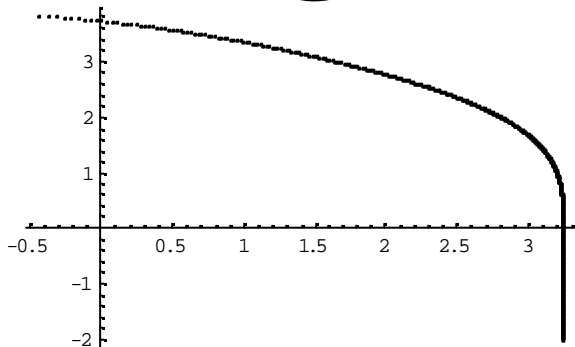
A= 1776

B= 2.507

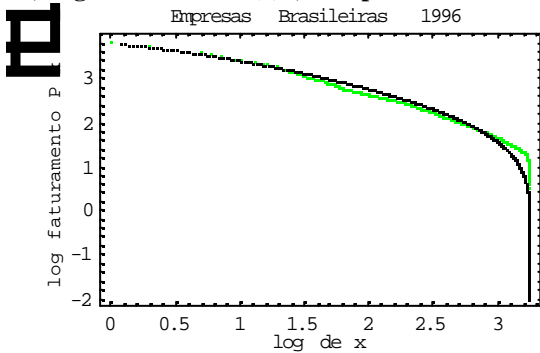
```

Fun1 := (m, a, desvio, xc) := desvio
Fun2 x := xc, k := which x < xc, x > xc, Exp - 10^xc
Fun50 x := media, desvio, xc, k := Fun1 x, media, desvio
Fun3 x := media, desvio, xc, k := Integrate Fun50 x, media, desvio, xc, k, 1, 5
lista1 = Table Fun3 x, {1.78, 0.62, 3.8, 1600}, {-2, 3.8, 0.01}
listalog = Log 10 lista1
listax = Table i, {3.8, 0.01}
trans1 = Transpose listalog, listax
Short[trans1]
graph1 = ListPlot trans1

```



Show[graph0, graph13, Frame->True, AxesOrigin->{-2,0}, Axes->False, FrameLabel->{"log de x", "log faturamento P(x)", "Empresas Brasileiras/1996", ""}]



...Graphics...