

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS

CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS

CURSO DE INFORMÁTICA

**AvaDisPro: Projeto e Implementação de uma Ferramenta de  
Apoio à Escolha de Distribuições de Probabilidade**

**Claudio Kinzel**

Prof. MS Ernesto Lindstaedt

Orientador

*Monografia submetida como requisito  
parcial para a obtenção do título de  
Bacharel em Informática*

São Leopoldo, novembro de 2003

*O milagre não é dar vida ao corpo extinto,  
Ou luz ao cego, ou eloquência ao mundo...  
Nem mudar água pura em vinho tinto...  
Milagre é acreditar em nisso tudo!*  
**Mário Quintana**

## **Dedicatória**

*À minha prima, amiga, namorada e mulher, Adriana.*

## Agradecimentos

*Agradeço a todos que me ajudaram na realização deste trabalho.  
A todos que, se não ajudaram, ao menos não atrapalharam, também.  
Não cito nomes aqui (é uma forma elegante de não esquecer ninguém),  
e também não farei desta seção um compêndio de lamúrias  
(o meu leitor merece mais que isso...).*

*Não agradeço a Deus, pois sou ateu;  
Não agradeço também ao meu 'anjo da guarda',  
muito menos a Alá, Maomé, Sidarta Gautama  
e a todos aqueles adorados pela choldra...*

*Também não agradeço ao 'absurdo',  
nem a R. Descartes, I. Kant, F. W. Nietzsche  
e todos os demais que julgaram entender a vida ou,  
ao menos, 'tudo que há'.*

## Resumo

O presente trabalho descreve a ferramenta AvaDisPro (Avaliador de Distribuições de Probabilidade), software que aplica os testes Qui-Quadrado ( $\chi^2$ ) e Kolmogorov-Smirnov (K-S) para determinar se uma variável segue uma ou mais entre as seguintes distribuições de probabilidade: Exponencial, Log-Logistic, Gama, Normal, Triangular, Uniforme e Weibull, entre as contínuas; e Binomial, Binomial Negativa, Geométrica, Poisson e Uniforme Discreta, entre as discretas.

O AvaDisPro foi projetado para suprir uma lacuna em relação a softwares estatísticos da área, que impõe restrições com relação à licença de uso, custo e utilização, não sendo – dessa forma – apropriados ao uso acadêmico, em atividades como ensino ou pesquisa.

Para ser uma ferramenta extensível, o projeto de software foi elaborado de modo a proporcionar uma maior facilidade para um desenvolvimento futuro. A inclusão de novas distribuições de probabilidade, bem como a leitura de dados em arquivos de diversos formatos são exemplos de extensões possíveis.

Neste trabalho é apresentada, além do projeto da ferramenta, uma análise de alguns softwares estatísticos, bem como as distribuições de probabilidade implementadas pelo AvaDisPro, os testes de adequação  $\chi^2$  e K-S, uma demonstração em forma de tutorial do AvaDisPro e o resultado de uma pesquisa de avaliação da ferramenta com alunos de 3º grau.

**Palavras-chave:** Estatística, Distribuições de Probabilidade, Testes de Adequação, Simulação de Sistemas.

# Abstract

The present work describes the tool AvaDisPro (Assessor of Probability Distributions), software that applies the fit tests Chi-Square ( $\chi^2$ ) and Kolmogorov-Smirnov (K-S) to determine if a variable follows one or more between the following probability distributions: Exponential, Log-Logistic, Gamma, Normal, Triangular, Uniform and Weibull, among the continuous ones; and Binomial, Negative Binomial, Geometric, Poisson and Discrete Uniform, among the discrete ones.

AvaDisPro was projected to supply a gap in relation to statistical softwares of the area, that imposes restrictions with relationship of use license, cost and use, not being – in that way – adapted to the academic use, for activities as teaching or research.

To be an extensible tool, the software project was elaborated in way to provide a larger easiness for a future development. The inclusion of new probability distributions, as well as the reading of data in files of several formats are examples of possible extensions.

In this work is described, besides the tool project, an analysis of some statistical softwares, the probability distributions implemented by AvaDisPro, the  $\chi^2$  and K-S goodness-of-fit tests, an AvaDisPro demonstration in tutorial form and the result of a tool evaluation research with college students.

**Keywords:** Statistics, Probability Distributions, Goodness-of-Fit Tests, Systems Simulation.

# Sumário

<b>1</b>	<b>Introdução.....</b>	<b>13</b>
1.1	Motivação.....	13
1.2	Objetivos .....	14
1.3	Estrutura.....	15
1.4	Sítio do AvaDisPro .....	16
<b>2</b>	<b>Análise de Ferramentas .....</b>	<b>17</b>
2.1	Métricas.....	17
2.2	Critérios .....	18
2.3	Softwares Avaliados .....	18
2.3.1	Arena – Módulo Input Analyzer.....	18
2.3.2	DataPlot .....	19
2.3.3	Excel .....	20
2.3.4	OpenStat .....	20
2.3.5	SPSS .....	21
2.3.6	StatCalc.....	21
2.3.7	Statistica .....	21
2.3.8	ViSta.....	22
2.4	Resultados .....	22
<b>3</b>	<b>Distribuições de Probabilidade.....</b>	<b>25</b>
3.1	Conceito de Distribuição Discreta .....	25
3.2	Conceito de Distribuição Contínua.....	26
3.3	Função Densidade (ou Massa).....	26
3.4	Função Cumulativa.....	26
3.5	Parametrização.....	27
3.6	Distribuições Contínuas.....	28
3.6.1	Exponencial .....	28
3.6.2	Gama.....	29
3.6.3	Log-Logistic .....	30
3.6.4	Normal.....	30
3.6.5	Triangular .....	31

3.6.6	Uniforme.....	31
3.6.7	Weibull .....	32
<b>3.7</b>	<b>Distribuições Discretas.....</b>	<b>33</b>
3.7.1	Binomial .....	33
3.7.2	Binomial Negativa.....	33
3.7.3	Geométrica .....	34
3.7.4	Poisson.....	35
3.7.5	Uniforme Discreta .....	35
<b>4</b>	<b>Testes de Adequação.....</b>	<b>37</b>
4.1	Qui-Quadrado ( $\chi^2$ ).....	37
4.1.1	Região Crítica.....	37
4.2	Kolmogorov-Smirnov (K-S).....	38
<b>5</b>	<b>Projeto do AvaDisPro .....</b>	<b>40</b>
5.1	Arquitetura .....	40
5.2	Ferramentas Utilizadas.....	41
5.3	Diagrama de Caso de Uso.....	41
5.3.1	Caso de Uso: Escolha de Tipo de Dados.....	41
5.3.2	Caso de Uso: Leitura de Dados Brutos.....	42
5.3.3	Caso de Uso: Leitura de Dados Agrupados.....	42
5.3.4	Caso de Uso: Definição de Agrupamento .....	43
5.3.5	Caso de Uso: Definição de Parâmetros .....	43
5.3.6	Caso de Uso: Teste de Distribuição.....	43
5.3.7	Caso de Uso: Visualização de Gráfico .....	43
5.3.8	Caso de Uso: Execução de Teste de Adequação .....	44
5.4	Diagramas de Seqüência.....	44
5.4.1	Seqüência: Escolha de Tipo de Dado .....	44
5.4.2	Seqüência: Leitura de Dados Brutos .....	45
5.4.3	Seqüência: Leitura de Dados Agrupados .....	45
5.4.4	Seqüência: Definição de Agrupamento .....	46
5.4.5	Seqüência: Definição de Parâmetros .....	46
5.4.6	Seqüência: Teste de Distribuição .....	47
5.4.7	Seqüência: Visualização de Gráfico.....	47
5.4.8	Seqüência: Execução de Teste de Adequação.....	48



<b>5.5</b>	<b>Diagrama de Classes .....</b>	<b>49</b>
5.5.1	Diagrama de Classes: Visão Geral .....	49
5.5.2	Diagrama de Classes: Objeto Distribuição.....	50
5.5.3	Diagrama de Classes: Objeto Dados .....	51
<b>5.6</b>	<b>Testes .....</b>	<b>51</b>
5.6.1	Testando o módulo CP-AvaDisPro .....	52
5.6.2	Testando o AvaDisPro.....	53
<b>6</b>	<b>Demonstração do AvaDisPro .....</b>	<b>56</b>
6.1	Definindo o Arquivo de Dados .....	56
6.2	Abrindo o Arquivo de Dados .....	57
6.3	Verificando Parâmetros.....	58
6.4	Testando uma Distribuição de Probabilidade .....	59
6.5	Testando todas Distribuições com $\chi^2$ ou K-S.....	60
6.6	Executando os Testes $\chi^2$ e K-S Automaticamente .....	61
6.7	O Módulo CP-AvaDisPro .....	62
<b>7</b>	<b>Avaliação do AvaDisPro .....</b>	<b>63</b>
7.1	Avaliação de Aspectos Estatísticos .....	63
7.2	Avaliação de Aspectos da Engenharia de Software .....	64
7.3	Avaliação de Aspectos Gerais .....	65
7.4	Análise de Resultados .....	66
<b>8</b>	<b>Considerações Finais.....</b>	<b>67</b>
8.1	Trabalhos Futuros.....	67
	<b>Referências Bibliográficas .....</b>	<b>69</b>
	<b>Anexo A Especializando uma classe do tipo Distribuição .....</b>	<b>71</b>
	<b>Anexo B Utilizando o AvaDisPro no GNU/Linux .....</b>	<b>75</b>
	<b>Anexo C Programando o AvaDisPro em Delphi e Kylix .....</b>	<b>77</b>
	<b>Anexo D Tabela do teste <math>\chi^2</math> .....</b>	<b>78</b>
	<b>Anexo E Tabela do teste K-S .....</b>	<b>79</b>
	<b>Anexo F Questionário de Avaliação do AvaDisPro.....</b>	<b>80</b>

## Lista de Figuras

Figura 1.1 – Sítio do AvaDisPro .....	16
Figura 2.1 – Tela Principal do <i>Input Analyzer</i> .....	19
Figura 2.2 – A ferramenta DataPlot.....	20
Figura 3.1 – FDP da Distribuição Exponencial [NIS 2002, p. 410].....	28
Figura 3.2 – FDP da Distribuição Gama [NIS 2002, p. 439] .....	29
Figura 3.3 – FDP da Distribuição Normal [NIS 2002, p. 377].....	31
Figura 3.4 – FDP da Distribuição Uniforme [NIS 2002, p. 384] .....	32
Figura 3.5 – FMP da Distribuição Binomial [NIS 2002, p. 482] .....	34
Figura 3.6 – FMP da Distribuição Poisson [NIS 2002, p. 485].....	36
Figura 4.1 – FPA da Normal versus FDA da distribuição empírica [NIS 2002, p. 323] .....	38
Figura 5.1 – Diagrama de Caso de Uso .....	42
Figura 5.2 – Diagrama de Seqüência para Escolha de Tipo de Dado .....	44
Figura 5.3 – Diagrama de Seqüência para Leitura de Dados Brutos.....	45
Figura 5.4 – Diagrama de Seqüência para Leitura de Dados Agrupados.....	45
Figura 5.5 – Diagrama de Seqüência para Definição de Agrupamento .....	46
Figura 5.6 – Diagrama de Seqüência para Definição de Parâmetros .....	46
Figura 5.7 – Diagrama de Seqüência para Teste de Distribuição.....	47
Figura 5.8 – Diagrama de Seqüência para Visualização de Gráfico .....	47
Figura 5.9 – Diagrama de Seqüência para Execução de Teste de Adequação .....	48
Figura 5.10 – Diagrama de Classes: Visão Geral.....	49
Figura 5.11 – Diagrama de Classes: Objeto Distribuição .....	50
Figura 5.12 – Diagrama de Classes: Objeto Dados.....	51
Figura 6.1 – Arquivo de Dados do Usuário.....	56
Figura 6.2 – Tela de Abertura de Dados do AvaDisPro.....	57
Figura 6.3 – Definindo parâmetros da Distribuição Normal.....	58
Figura 6.4 – Resultado dos testes $\chi^2$ e K-S com a Distribuição Normal .....	59
Figura 6.5 – Relatório dos testes $\chi^2$ e K-S para a Normal.....	60
Figura 6.6 – Resultado do teste $\chi^2$ para todas distribuições contínuas.....	61
Figura 6.7 – Opção para executar automaticamente os testes $\chi^2$ e K-S .....	61
Figura 6.8 – Opção para executar o módulo CP-AvaDisPro.....	62
Figura 6.9 – Tela do módulo CP-AvaDisPro .....	62

Figura 7.1 – Avaliação de usuários sobre a acurácia dos testes de adequação.....	63
Figura 7.2 – Avaliação de usuários sobre o desempenho do mecanismo de agrupamento .....	64
Figura 7.3 – Avaliação de usuários sobre a estabilidade em tempo de execução .....	65
Figura 7.4 – Avaliação de usuários sobre o grau de importância do AvaDisPro .....	65

## Lista de Tabelas

Tabela 2.1 – Resumo dos softwares avaliados .....	18
Tabela 2.2 – Resultado da avaliação de métricas de hardware e software .....	22
Tabela 2.3 – Resultado da avaliação de métricas de estatística.....	23
Tabela 2.4 – Resultado final da avaliação de softwares .....	23
Tabela 2.5 – Distribuições disponíveis pelos softwares .....	24
Tabela 3.1 – Propriedades da Distribuição Exponencial.....	28
Tabela 3.2 – Propriedades da distribuição Gama .....	29
Tabela 3.3 – Propriedades da distribuição Log-Logistic.....	30
Tabela 3.4 – Propriedades da distribuição Normal.....	30
Tabela 3.5 – Propriedades da distribuição Triangular.....	31
Tabela 3.6 – Propriedades da distribuição Uniforme .....	32
Tabela 3.7 – Propriedades da distribuição Weibull.....	32
Tabela 3.8 – Propriedades da distribuição Binomial.....	33
Tabela 3.9 – Propriedades da distribuição Binomial Negativa .....	34
Tabela 3.10 – Propriedades da distribuição Geométrica .....	35
Tabela 3.11 – Propriedades da distribuição Poisson .....	35
Tabela 3.12 – Propriedades da distribuição Uniforme Discreta.....	36
Tabela 5.1 – Ferramentas e Metodologias de Desenvolvimento.....	41
Tabela 5.2 – Resultado do teste da FPA.....	52
Tabela 5.3 – Conjunto de Dados Brutos.....	53
Tabela 5.4 – Parâmetros gerados pelas ferramentas.....	53
Tabela 5.5 – Resultado do teste $\chi^2$ .....	54
Tabela 5.6 – Resultado do teste K-S.....	54
Tabela 5.7 – Resultados dos testes K-S e $\chi^2$ .....	55
Tabela 7.1 – Resumo da avaliação de usuários .....	66

# 1 Introdução

Este capítulo apresenta uma introdução ao desenvolvimento da ferramenta AvaDisPro (Avaliador de Distribuições de Probabilidade). Discute-se a motivação e o contexto do seu desenvolvimento. Também são mostrados os objetivos desta monografia, além da sua estrutura e do sítio no qual está disponível a ferramenta.

## 1.1 Motivação

A estatística é uma ciência derivada da matemática cuja utilização dá-se em diversas áreas do conhecimento humano [LEV 1998]. Em especial, a área de simulação de sistemas depende da estatística no sentido de requerer – quando de modelagem estocástica – que cada variável modelada esteja acompanhada de sua respectiva distribuição de probabilidade. Uma distribuição representa o comportamento de uma variável através de sua função de probabilidade [HAY 1996], e pode ser de dois tipos: discreta, a qual representa valores discretos; e contínua, que representa valores contínuos.

Para a área de simulação, bem como para qualquer outra área que utilize distribuições de probabilidade, é de importância vital que o conjunto de dados modelado seja corretamente identificado com a distribuição mais apropriada à sua utilização. Para isso se faz necessário aplicar algum teste de hipótese [SPI 1972], como, por exemplo, o Qui-Quadrado ( $\chi^2$ ) ou o Kolmogorov-Smirnov (K-S), para se certificar de que realmente a distribuição escolhida não oferece restrições ao uso do conjunto de dados selecionado, sob risco de comprometer a confiabilidade estatística da operação<sup>1</sup>.

Existem algumas ferramentas computacionais comerciais, tais como o SPSS, o Statistica e o Arena, que resolvem este e outros problemas estatísticos (ver capítulo 2). O seu uso, especialmente em âmbito acadêmico, esbarra em três dificuldades: custo, utilização e licença de uso. O custo destes pacotes é muitas vezes inacessível para universidades e também para estudantes. Como resultado, o estudante termina por utilizar nenhuma ferramenta de apoio em seu processo de aprendizagem. Já com relação à utilização, estes softwares são pacotes comerciais (que não são apropriados ao uso por um estudante, dada sua natureza pouco didática) e geralmente estão em língua inglesa, o que é outro fator de desestímulo ao seu uso. Por fim, estes softwares têm uma licença de uso proprietária que é restritiva, a qual não permite, por exemplo, que a comunidade acadêmica possa estudar o seu código-fonte e realizar melhorias.

---

<sup>1</sup> Ao longo de todo o texto utiliza-se expressões como: "a distribuição de probabilidade foi aceita", ou ainda "o conjunto de dados segue uma distribuição" etc. Sabe-se que uma distribuição é escolhida por falta de evidências em contrário, isto é, por sua não rejeição em algum teste de adequação. No entanto, para facilitar a leitura do texto, eventualmente empregou-se o uso destas expressões.

Existem alguns softwares sob licença *freeware* (gratuita), como o DataPlot, o OpenStat, o StatCalc e o ViSta, os quais – apesar de não apresentarem restrições com custo – apresentam problemas com relação à licença de uso, já que este tipo de distribuição também não permite modificação do código-fonte do sistema. Dessa forma, caso erros sejam detectados em alguma parte do software, à exceção do seu autor, ninguém poderia corrigi-los. Ferramentas genéricas, como o Microsoft Excel, suportam apenas parcialmente o processo estatístico. Esta, inclusive, apresenta problemas de acurácia em determinadas operações [BER 1999], o que desaconselha o seu uso para fins nos quais um rigor estatístico maior seja importante.

Com o crescente uso da computação na sociedade, faz-se mister a existência de ferramentas computacionais acessíveis ao estudante de disciplinas de estatística, ou ainda a pesquisadores, seja em relação a custo, utilização e licença de uso.

## 1.2 Objetivos

O objetivo principal deste trabalho é o desenvolvimento de uma ferramenta computacional de apoio ao processo estatístico de escolha de distribuições de probabilidade. Tal software deve responder às seguintes questões: dado um conjunto de dados qualquer, existe alguma distribuição que melhor o represente? E, caso exista, qual é essa distribuição? E quais são os valores de seus parâmetros?

Como objetivos secundários, pode-se citar os seguintes:

- a) Oferecer interface em língua portuguesa: devido à escassez de softwares desenvolvidos em língua portuguesa na área estatística, optou-se por desenvolvê-lo nessa língua, de modo que essa carência seja – ao menos parcialmente – resolvida.
- b) Ser direcionado ao público acadêmico: o AvaDisPro é um software didático sem pretensões comerciais, cujo uso far-se-á em disciplinas nas quais o ensino de probabilidade se faça presente, ou ainda na área de pesquisa. Dessa forma, o esforço de desenvolvimento está direcionado em mostrar os passos para a resolução do problema ao qual se propõe a resolver, e não em simplesmente apresentar uma resposta final.
- c) Ter uma licença de uso flexível: a idéia é utilizar uma licença de uso baseada na GPL (*General Public License*) [FREa 2003], que é a licença padrão do movimento GNU (*GNU is Not Unix!*) [FREb 2003]. Utilizou-se uma variação dessa licença, de modo a restringir o uso do AvaDisPro para fins comerciais, permitindo, no entanto, o uso, a cópia e a modificação para fins de ensino ou pesquisa.
- d) Ser portátil: o desenvolvimento da ferramenta é realizado utilizando a biblioteca CLX (*Component Library for Cross Platform*), que é uma biblioteca intercambiável entre os ambientes Windows e GNU/Linux [DIA 2001], [SON 2001].

- e) Oferecer documentação apropriada: dada à natureza didática do software e também à sua licença de uso, a documentação da ferramenta deve ser rica e suficiente de modo que tanto usuários quanto desenvolvedores possam usar ou modificar o AvaDisPro com um esforço mínimo.
- f) Possuir uma arquitetura extensível: o AvaDisPro foi projetado para ser extensível, de modo que o desenvolvimento futuro seja facilitado. Para isso, por exemplo, utilizou-se uma arquitetura que separa classes de interface de classes de negócio [SCH 2001], além do uso de técnicas de programação sob o modelo orientado a objetos [CAN 2000], [CAR 2001], como herança e polimorfismo (ver capítulo 5 e anexo A).

## 1.3 Estrutura

O presente trabalho está estruturado nos seguintes capítulos a seguir:

- Análise de Ferramentas: o capítulo 2 descreve a avaliação de alguns softwares estatísticos sob algumas métricas de dois grupos: métricas para software e hardware e métricas para estatística. Também se apresenta os critérios de avaliação e os resultados da análise, mostrando que os objetivos propostos para o AvaDisPro não são atendidos por estas ferramentas de maneira integral.
- Distribuições de Probabilidade: o capítulo 3 descreve as distribuições implementadas pelo AvaDisPro, que são: Exponencial, Log-Logistic, Gama, Triangular, Uniforme e Weibull, entre as contínuas; e Binomial, Binomial Negativa, Geométrica, Poisson e Uniforme Discreta, entre as discretas. São mostradas as suas propriedades, bem como uma descrição da aplicabilidade de cada distribuição.
- Testes de Adequação: o capítulo 4 descreve os testes implementados pelo AvaDisPro, que são o  $\chi^2$  e o K-S. Tais testes visam a determinar se um conjunto de dados segue ou não alguma distribuição, de acordo com a falta de evidências em contrário.
- Projeto do AvaDisPro: o capítulo 5 apresenta o projeto do AvaDisPro, incluindo a sua modelagem seguindo a linguagem UML (*Unified Modeling Language*). A arquitetura de desenvolvimento é discutida, e o resultado de alguns testes é apresentado.
- Demonstração do AvaDisPro: o capítulo 6 fornece uma demonstração das potencialidades do software com base em um exemplo. As suas funcionalidades são descritas passo-a-passo, de modo que se possa mostrar os recursos do AvaDisPro.
- Avaliação do AvaDisPro: o capítulo 7 descreve o resultado de uma pesquisa de avaliação do AvaDisPro. Ao todo 12 questões foram abordadas em forma de questionário a alunos de terceiro grau que necessitavam utilizar distribuições de probabilidade no contexto da área de simulação e modelagem de sistemas.

E também de acordo com os seguintes anexos:

- Especializando uma classe do tipo Distribuição: mostra um tutorial de como inserir uma nova classe do tipo distribuição no AvaDisPro.
- Utilizando o AvaDisPro no GNU/Linux: roteiro para utilização do AvaDisPro.
- Programando o AvaDisPro em Delphi e Kylix: apresenta algumas diretivas de programação nos ambientes Windows e GNU/Linux.
- Tabela do Teste  $\chi^2$ : mostra os valores críticos usados no AvaDisPro para o teste  $\chi^2$ .
- Tabela do Teste K-S: mostra os valores críticos usados no AvaDisPro para o teste K-S.
- Questionário de Avaliação do AvaDisPro: apresenta as perguntas enviadas aos alunos para avaliação do AvaDisPro.

## 1.4 Sítio do AvaDisPro

O AvaDisPro está disponível para cópia no sítio <http://www.avadispro.cjb.net> (figura 1.1). Nesse local também se encontra o código-fonte do projeto e a presente monografia. O acesso é livre para quem deseja utilizar e modificar a ferramenta, de acordo com a sua licença de uso.



Figura 1.1 – Sítio do AvaDisPro



## 2 Análise de Ferramentas

Este capítulo mostra uma análise de alguns softwares estatísticos que auxiliam no processo de escolha de distribuições de probabilidade. Testou-se ao todo oito ferramentas, que são de dois tipos: comerciais (Arena, Excel, SPSS e Statistica) e não-comerciais (DataPlot, OpenStat, StatCalc e ViSta).

Apresentam-se as métricas segundo as quais os softwares foram avaliados, os critérios de avaliação, uma descrição de cada ferramenta, além dos resultados obtidos, que mostram a carência de softwares que atentam aos objetivos propostos para o AvaDisPro (ver seção 1.2).

### 2.1 Métricas

Para a avaliação de softwares utilizou-se dois grupos de métricas, que são: métricas para hardware e software e métricas para estatística. A seguir apresentam-se as métricas para o primeiro grupo:

- a) Licença de Uso: diz respeito ao nível de liberdade que o software oferece ao usuário.
- b) Plataforma: refere-se à quantidade de plataformas nas quais o software está disponível.
- c) Usabilidade: diz respeito a uma série de conceitos que visam a tornar melhor o uso do software por seus utilizadores. Nessa categoria inclui-se os seguintes itens [PRE 1995]: facilidade de uso (quantidades de opções para chegar num resultado, uso de interface gráfica etc.), flexibilidade e documentação (manuais, ajuda on-line etc.).
- d) Didática: refere-se à capacidade do software ser auto-explicativo em seu uso.
- e) Extensibilidade: diz respeito ao nível de personalização permitido ao usuário. Por exemplo, se é possível incluir novas características no software. Em geral, é necessário modificar o código-fonte para uma possibilidade maior de adaptação.

Já com relação à estatística, utilizou-se as seguintes métricas:

- a) Entrada de Dados: refere-se aos tipos de formatação para entrada de dados. Por exemplo, se o software lê dados brutos e consegue classificá-los apropriadamente.
- b) Saída de Dados: refere-se à riqueza de informações que o software apresenta após realizar seu processamento, como a apresentação de histogramas e medidas estatísticas.
- c) Distribuições de Probabilidade: refere-se à quantidade de distribuições suportadas.
- d) Testes Disponíveis: se o software valida os dados aplicando os testes  $\chi^2$  e K-S.
- e) Automação de Resultados: refere-se à capacidade do software de automaticamente mostrar os resultados de suas avaliações com um mínimo de esforço ao usuário.

## 2.2 Critérios

Avaliou-se cada software de acordo com a seguinte escala numérica: 0 (a característica não é satisfeita) até 1 (satisfeita integralmente). Todas as dez métricas têm peso 1 e a totalização horizontal gera o resultado individual por software por conjunto de métricas. Já a média da totalização vertical fornece a média aritmética por métrica. A fórmula a seguir é utilizada para se obter a média aritmética final (ver tabela 2.4):

$$m = \frac{\sum_{i=1}^j n_i}{j}$$

Onde  $j$  é o total de métricas utilizadas,  $n$  é a nota por métrica e  $m$  é a média final por software.

## 2.3 Softwares Avaliados

Esta seção apresenta uma descrição dos softwares avaliados. A tabela 2.1 mostra um resumo sobre estas ferramentas, incluindo a versão testada, o autor e o seu endereço web.

Tabela 2.1 – Resumo dos softwares avaliados

Software	Versão	Autor	Web
Arena	3.5	System Modeling Corp.	–
DataPlot	–	NIST	<a href="http://www.itl.nist.gov/div898/software/dataplot">http://www.itl.nist.gov/div898/software/dataplot</a>
Excel	10.0	Microsoft Inc.	<a href="http://www.microsoft.com">http://www.microsoft.com</a>
OpenStat	3.1	prof. Bill Miller	<a href="http://openstat.homestead.com/OpenStatMain.html">http://openstat.homestead.com/OpenStatMain.html</a>
SPSS	11.0	SPPS Inc.	<a href="http://www.spss.com">http://www.spss.com</a>
StatCalc	1.1	prof. K. Krishnamoorthy	<a href="http://www.ucs.louisiana.edu/~kxk4695/StatCalc.htm">http://www.ucs.louisiana.edu/~kxk4695/StatCalc.htm</a>
Statistica	6.0	StatSoft Inc.	<a href="http://www.statsoftinc.com">http://www.statsoftinc.com</a>
ViSta	6.4	prof. Forrest W. Young	<a href="http://www.visualstats.org">http://www.visualstats.org</a>

### 2.3.1 Arena – Módulo Input Analyzer

Software de simulação desenvolvido pela *System Modeling Corporation* sob licença comercial. Possui um módulo independente de avaliação de distribuições chamado *Input Analyzer*. Está disponível para a plataforma PC (*Personal Computer*) e sistema operacional Windows 95 (ou superior).

Esta ferramenta lê dados em formato texto e os agrupa automaticamente em classes, além de permitir a personalização do histograma gerado (figura 2.1). Também é possível alterar os parâmetros de algumas distribuições, como a Gama ou a Weibull, por exemplo.

Os testes disponíveis são os  $\chi^2$  e o K-S, e a ferramenta suporta ao todo 12 distribuições de probabilidade, não havendo possibilidade de inserção de novas distribuições.

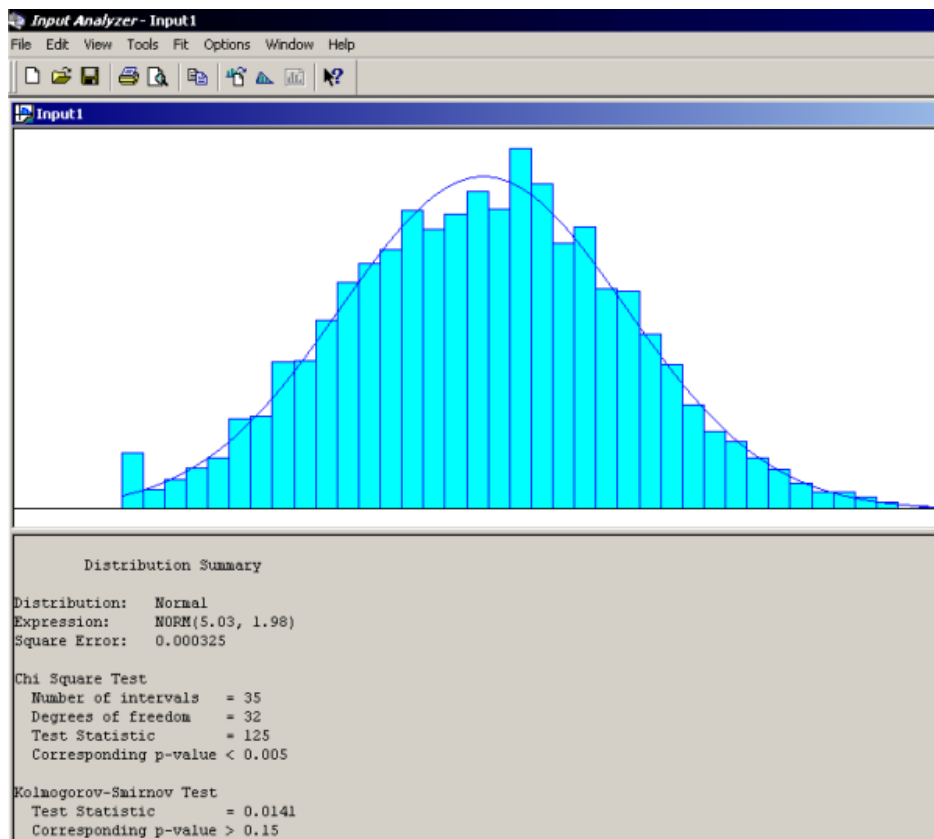


Figura 2.1 – Tela Principal do *Input Analyzer*

A ferramenta é simples de usar e rápida. A interface é intuitiva, o que ajuda o usuário a rapidamente escolher as opções de seu interesse. Com relação à documentação, o Arena possui ajuda de uso operacional, não oferecendo qualquer ajuda em termos estatísticos.

### 2.3.2 DataPlot

Ferramenta de domínio público desenvolvida pelo NIST (*National Institute of Standards and Technology*), disponível para vários ambientes (Windows, Unix, GNU/Linux etc.), tem por objetivo ser uma ferramenta para modelagem, análise e otimização de processos científicos.

O DataPlot suporta 22 distribuições e os testes  $\chi^2$  e K-S. Possui opções para geração de gráficos e estimação de parâmetros.

A documentação do software é completa, fornecendo além da ajuda operacional, um rico material sobre estatística, incluindo um livro eletrônico [NIS 2002]. A sua operação, no entanto, é confusa e o software não permite que todas suas opções sejam realizadas com uso de interface gráfica, tendo a interface por linha de comando de ser utilizada. A figura 2.2 mostra uma tela do DataPlot para cálculo da distribuição Normal.

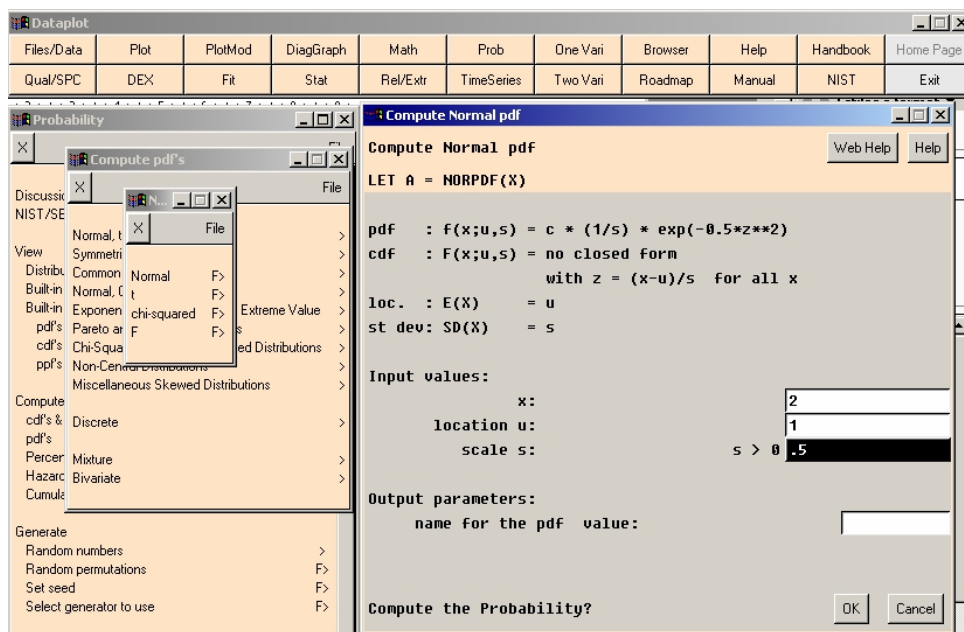


Figura 2.2 – A ferramenta DataPlot

### 2.3.3 Excel

Planilha eletrônica desenvolvida pela Microsoft sob licença comercial. Disponível para as plataformas PC e Macintosh, apresenta versões em vários idiomas (incluindo o português). Trata-se duma ferramenta genérica, que possui algumas funções estatísticas.

O Excel suporta 13 distribuições de probabilidade, assim como o teste  $\chi^2$ . Também lê dados brutos em vários formatos, como HTML, ASCII ou DBF. A ajuda operacional do programa é completa, bem como a ajuda estatística para cada função.

A operação do Excel é simples e existem recursos que visam a facilitar o trabalho do usuário, como, por exemplo, o uso de assistentes para elaboração de fórmulas.

Por ser uma ferramenta genérica, o Excel não possui um formato adequado para validar os dados do usuário, tendo este que calcular manualmente as funções de seu interesse. Além disso, existem problemas de acurácia no resultado de algumas funções [BER 1999], o que desaconselha o seu uso para fins nos quais um rigor estatístico seja importante.

### 2.3.4 OpenStat

Software sob licença *freeware*. Disponível para a plataforma Windows e hardware PC, suporta apenas 3 distribuições de probabilidade e não aplica testes de adequação.

O seu uso é confuso (não há padronização na interface e as opções não apresentam coerência entre si) e vários erros ocorreram nos testes realizados. A ajuda operacional do software, entretanto, é completa, bem como a ajuda para os processos estatísticos implementados.

### 2.3.5 SPSS

Tradicional pacote comercial estatístico produzido pela SPSS Inc. Está disponível para o ambiente Windows e hardware PC.

O SPSS é um pacote completo, tendo opções de comunicação com banco de dados, ou ainda a possibilidade de se trabalhar com várias variáveis ao mesmo tempo. Suporta 19 distribuições de probabilidade, além de realizar os testes  $\chi^2$  e K-S.

Com uma documentação rica e completa, o software torna-se mais amigável ao usuário, já que devido às diversas opções existentes nem sempre é fácil encontrar a opção de uso apropriada. O SPSS oferece também alguns tutoriais sobre o seu uso e também sobre a aplicabilidade estatística em diversas áreas do conhecimento humano.

### 2.3.6 StatCalc

Software de avaliação de distribuições de probabilidade sob licença *freeware*. Está disponível para plataforma PC e sistema operacional Windows 95 (ou superior).

O StatCalc realiza o cálculo da função de probabilidade e também da função de probabilidade acumulada de 17 distribuições, mas não implementa testes de adequação.

O software é de fácil uso, oferecendo opções de ajuda completas e úteis, apesar da interface estar prejudicada em função de vários elementos estarem presentes numa mesma tela, o que dificulta a sua visualização. Para cada fórmula utilizada, o StatCalc oferece um conjunto adequado de informações de suporte e ajuda.

### 2.3.7 Statistica

Ferramenta genérica desenvolvida pela StatSoft, sob licença de uso comercial. Disponível para as plataformas PC e Macintosh (sistemas operacionais Windows 95 e MacOS).

O Statistica é uma ferramenta ampla que possui alguns recursos, como a comunicação com banco de dados, o uso de técnicas de *Data Mining* e várias opções de visualização de dados. Suporta 13 distribuições de probabilidade através do módulo *Probability Calculator*, que permite calcular os parâmetros e probabilidades de um conjunto de dados. Também possui um módulo para aplicação dos testes de adequação  $\chi^2$  e K-S.

O Statistica é fácil de operar, tendo uma ajuda operacional completa. Com relação à ajuda estatística, também o software mostra-se completo e, além de apresentar os conceitos básicos de cada distribuição, possui módulos nos quais ensina ao usuário diversos conceitos estatísticos, relacionando-os com o uso da ferramenta.

### 2.3.8 ViSta

ViSta é o acrônimo de *The Visual Statistics System*, sob licença *freeware* (porém, com permissão de modificação do código-fonte sob autorização de seu autor). Disponível para os sistemas operacionais Windows, Unix e MacOS, além das plataformas de hardware PC e Macintosh. Possui versões em língua inglesa, espanhola e francesa.

O ViSta suporta somente distribuições contínuas, sendo 10 no total, bem como teste  $\chi^2$ . O software não lê dados brutos, porém possui um módulo para sua edição, bem como importação de dados externos em formato proprietário.

A operação do software é confusa, com várias telas sobrepondo-se entre si. Os resultados apresentados não possuem uma padronização clara. A documentação do ViSta limita-se à ajuda operacional, tendo poucas informações sobre o seu significado estatístico.

## 2.4 Resultados

Os resultados da avaliação dos softwares encontram-se nas tabelas a seguir. A tabela 2.2 apresenta as notas de cada software com relação às métricas de software e hardware. Já a tabela 2.3 apresenta as notas das métricas com relação à estatística.

Tabela 2.2 – Resultado da avaliação de métricas de hardware e software

Métrica / Software	Licença de Uso <sup>2</sup>	Plataforma <sup>3</sup>	Usabilidade	Didática	Extensibilidade	Média / Software
Arena	0,0	0,0	1,0	0,5	0,0	<b>0,3</b>
DataPlot	0,5	1,0	0,0	0,5	0,0	<b>0,4</b>
Excel	0,0	0,5	1,0	0,5	0,0	<b>0,4</b>
OpenStat	0,5	0,0	0,0	0,0	0,0	<b>0,1</b>
SPSS	0,0	0,5	1,0	0,5	0,0	<b>0,4</b>
StatCalc	0,5	0,0	0,5	0,5	0,0	<b>0,3</b>
Statistica	0,0	0,5	1,0	0,5	0,0	<b>0,4</b>
ViSta	1,0	1,0	0,0	0,0	0,0	<b>0,4</b>
Média	<b>0,3125</b>	<b>0,4375</b>	<b>0,5625</b>	<b>0,3750</b>	<b>0,0000</b>	-

Os resultados mostram que a extensibilidade é um objetivo que não é atingido por nenhuma ferramenta avaliada, isto é, os softwares não estão preparados – nem foram projetados – para serem adaptados por terceiros. A licença de uso apresentou média de 31,25%, e didática de 37,50%, o que mostra a carência de softwares que levem em consideração esses fatores, de acordo com os objetivos propostos para o AvaDisPro (ver seção 1.2).

<sup>2</sup> 0: Comercial; 0,5: *freeware*; 1: GPL ou semelhante.

<sup>3</sup> 0: Suporte a uma plataforma; 0,5: suporte a duas; 1: suporte a três ou mais.

Tabela 2.3 – Resultado da avaliação de métricas de estatística

Métrica / Software	Entrada de Dados <sup>4</sup>	Saída de Dados <sup>5</sup>	Distribuições <sup>6</sup>	Testes de Adequação <sup>7</sup>	Automação Resultados	Média / Software
Arena	0,6	0,6	0,6	1,0	1,0	<b>0,76</b>
DataPlot	0,3	0,6	1,0	1,0	0,0	<b>0,58</b>
Excel	1,0	1,0	0,6	0,5	0,0	<b>0,62</b>
OpenStat	0,6	0,3	0,0	0,0	0,0	<b>0,18</b>
SPSS	1,0	1,0	1,0	1,0	0,0	<b>0,80</b>
StatCalc	0,0	0,0	1,0	0,0	0,0	<b>0,20</b>
Statistica	1,0	1,0	0,6	1,0	0,0	<b>0,72</b>
ViSta	0,0	0,3	0,3	0,5	0,0	<b>0,22</b>
Média	<b>0,5625</b>	<b>0,6000</b>	<b>0,6375</b>	<b>0,6250</b>	<b>0,1250</b>	-

O desempenho das ferramentas foi melhor nas métricas relacionadas à estatística. O único item que ficou com uma média baixa é automação de resultados, com 12,5%. Apenas o software Arena apresenta opção para automaticamente testar todas distribuições disponíveis e selecionar a mais apropriada ao conjunto de dados.

A tabela 2.4 mostra o resultado final da avaliação de softwares. As 4 ferramentas sob licença comercial tiveram o melhor desempenho, sendo o SPSS e o Statistica as que ficaram com melhor média, com 60% e 56%, respectivamente. Tais resultados evidenciam a carências de softwares da área, de acordo com as métricas propostas nesta avaliação.

Tabela 2.4 – Resultado final da avaliação de softwares

Métrica / Software	Métricas Software	Métricas Estatística	Média / Software	Média / Software (%)
SPSS	0,4	0,80	<b>0,6000</b>	<b>60,00</b>
Statistica	0,4	0,72	<b>0,5600</b>	<b>56,00</b>
Arena	0,3	0,76	<b>0,5300</b>	<b>53,00</b>
Excel	0,4	0,62	<b>0,5100</b>	<b>51,00</b>
DataPlot	0,4	0,58	<b>0,4900</b>	<b>49,00</b>
ViSta	0,4	0,22	<b>0,3100</b>	<b>31,00</b>
StatCalc	0,3	0,20	<b>0,2500</b>	<b>25,00</b>
OpenStat	0,1	0,18	<b>0,1400</b>	<b>14,00</b>
Média	<b>0,3375</b>	<b>0,5100</b>	<b>0,4238</b>	<b>42,38</b>

<sup>4</sup> 0: não lê dados externos; 0,3: lê um formato; 0,6: lê dois formatos; 1: lê três ou mais.

<sup>5</sup> 0: mostra só o resultado final; 0,3: mostra um tipo de saída; 0,6: mostra dois tipos; 1: mostra três ou mais.

<sup>6</sup> 0: suporta de 1 a 5 distribuições; 0,3: de 6 a 10; 0,6: de 11 a 15; 1: 16 ou mais.

<sup>7</sup> 0: não suporta testes de adequação; 0,5: suporta um teste; 1,0: suporta dois ou mais.

A tabela 2.5 apresenta a relação das distribuições de probabilidade implementadas por cada software. A totalização vertical representa a quantidade de distribuições que cada software suporta e a totalização horizontal a quantidade de vezes que cada distribuição está disponível, num total de 28 distribuições.

Tabela 2.5 – Distribuições disponíveis pelos softwares

Software / Distribuição	Arena	Data-Plot	Excel	Open Stat	SPSS	Stat-Calc	Statística	ViSta	$\Sigma$
Anglit		✓							1
Beta	✓	✓	✓		✓	✓	✓	✓	7
Binomial		✓	✓		✓	✓		✓	5
Binomial Negativa		✓	✓		✓	✓		✓	5
Cauchy		✓			✓	✓	✓		4
$\chi^2$		✓	✓	✓	✓	✓	✓	✓	7
Empírica	✓								1
Erlang	✓								1
Exponencial	✓	✓	✓		✓			✓	5
F			✓	✓	✓	✓	✓		5
Hipergeométrica		✓	✓		✓	✓			4
Gama	✓	✓	✓		✓	✓	✓	✓	7
Geométrica		✓			✓	✓			3
Johnson	✓								1
Laplace					✓	✓	✓		3
Log-Normal	✓	✓	✓		✓	✓	✓		6
Logistic		✓			✓	✓	✓		4
Normal	✓	✓	✓	✓	✓	✓	✓	✓	8
Pareto		✓			✓	✓	✓		4
Poisson	✓	✓	✓		✓	✓		✓	6
Rayleigh							✓		1
Semi-circular		✓							1
t Estudante		✓	✓		✓	✓	✓	✓	6
Triangular	✓	✓							2
Tukey-lambda		✓							1
Weibull	✓	✓	✓		✓	✓	✓		6
Uniforme	✓	✓			✓			✓	4
Uniforme Discreta		✓							1
$\Sigma$	12	22	13	3	19	17	13	10	—



### 3 Distribuições de Probabilidade

Este capítulo fornece uma visão geral das distribuições de probabilidade implementadas pelo AvaDisPro. Ao todo são utilizadas 12 distribuições, sendo sete contínuas (Exponencial, Gamma, Log-Logistic, Normal, Triangular, Uniforme e Weibull) e cinco discretas (Binomial, Binomial Negativa, Geométrica, Poisson e Uniforme Discreta). O critério para escolha dessas distribuições baseia-se em importância e utilização, de acordo com a bibliografia consultada [HAY 1996], [LAW 2000], [NIS 2002] e [McL 2000].

Existem ainda outras distribuições de probabilidade (por exemplo: Beta, LogNormal, Cauchy, Pearson, Johnson etc.) ou ainda variações entre distribuições [McL 2000], [LEE 1994] e [SAB 2000], as quais não são implementadas pelo AvaDisPro. Devido ao caráter extensível e aberto da ferramenta, no entanto, essa lacuna pode ser preenchida futuramente pela comunidade acadêmica.

A seguir apresenta-se uma descrição das distribuições do AvaDisPro, incluindo as expressões matemáticas que foram utilizadas na sua implementação. Também se conceitualiza as funções densidade (ou massa) e cumulativa, bem como alguns métodos para estimação de parâmetros.

#### 3.1 Conceito de Distribuição Discreta

Uma função de probabilidade discreta,  $p(x)$ , é uma função que satisfaz as seguintes propriedades [NIS 2002]:

- a) A probabilidade que  $x$  assuma um valor específico é  $p(x)$ :  $P[X = x] = p(x) = p_x$
- b)  $p(x)$  é não-negativo para todos números reais  $x$ .
- c) A soma de  $p(x)$  de todas as possibilidades de  $x$  é 1, isto é:  $\sum_j p_j = 1$   
onde  $j$  representa os valores possíveis que  $x$  pode assumir e  $p_j$  é a probabilidade em  $x_j$ .

Como consequência das propriedades  $b$  e  $c$  tem-se que:  $0 \leq p(x) \leq 1$ .

De acordo com essa definição, tem-se que uma função de probabilidade discreta é uma função que assume números discretos [NIS 2002]. Assim, a função que permite valores negativos ou valores maiores que um não é uma função de probabilidade.

### 3.2 Conceito de Distribuição Contínua

Uma função de probabilidade contínua,  $f(x)$ , é uma função que satisfaz as seguintes propriedades [NIS 2002]:

- a) A probabilidade de  $x$  estar em dois pontos  $a$  e  $b$  é:  $p[a \leq x \leq b] = \int_a^b f(x)dx$
- b) A probabilidade é não-negativa para todos números reais  $x$ .
- c) A integral da função de probabilidade é um, isto é:  $\int_{-\infty}^{\infty} f(x)dx = 1$

Com base nessa definição, tem-se que desde que a função de probabilidade é definida para um número infinito de termos num intervalo contínuo, então a probabilidade de um ponto específico é sempre zero [NIS 2002]. Assim, probabilidades são mensuradas em intervalos.

### 3.3 Função Densidade (ou Massa)

Para uma função contínua, a função densidade de probabilidade (FDP) é a probabilidade que uma variável tenha o valor  $x$  [NIS 2002], e a probabilidade de  $x$  ser igual a determinado valor é zero. Desde que as distribuições contínuas iniciem em um ponto zero, então se pode expressar a FDP como a integração entre dois pontos [NIS 2002]:

$$\int_a^b f(x)dx = p[a \leq X \leq b]$$

Já para funções discretas, a função massa de probabilidade (FMP) é a probabilidade que uma variável tenha o valor de  $x$  e é dada por [NIS 2002]:

$$f(x) = p[X = x]$$

### 3.4 Função Cumulativa

Para ambas funções discreta e contínua, a função de probabilidade acumulada (FPA) é a probabilidade que uma variável tenha um valor que seja igual ou menor que  $x$  [NIS 2002], isto é:

$$F(x) = \Pr[X \leq x] = \alpha$$

A definição matemática para distribuições contínuas é dada por [NIS 2002]:

$$F(x) = \int_{-\infty}^x f(u)du$$

Já para distribuições discretas [NIS 2002]:

$$F(x) = \sum_{i=0}^x f(i)$$

### 3.5 Parametrização

As distribuições de probabilidade, sejam elas discretas ou contínuas, podem ser divididas em dois grupos, de acordo com suas propriedades [McL 2000]. No primeiro grupo, tem-se as distribuições cujos parâmetros são estimados diretamente de uma amostra, através de medidas como média ( $\mu$ ) ou desvio padrão ( $\sigma$ ). Já no segundo grupo, têm-se distribuições cujos parâmetros precisam ser estimados com uso de algum método específico, isto é, não se sabe os seus valores a priori, e para descobri-los é preciso aplicar algum método de estimação.

Para estimar os parâmetros de cada uma dessas distribuições existem alguns métodos, entre os mais conhecidos pode-se citar [NIS 2002]:

- a) Método dos Momentos (MOM): método baseado na obtenção de parâmetros de acordo com os momentos da amostra [NIS 2002]. Tem como vantagem a simplicidade de ser obtido quando puder ser aplicado (o que nem sempre ocorre com todas distribuições) e como principal desvantagem o fato de não produzir propriedades otimizadas, como ocorre, por exemplo, com o método *Maximum Likelihood Estimation* e o dos Mínimos Quadrados [NIS 2002].
- b) *Maximum Likelihood Estimation* (MLE): utiliza-se de uma expressão matemática conhecida como função *likelihood*. Tal método consiste em escolher o valor do parâmetro que maximiza a probabilidade de se obter a amostra observada, de acordo com uma função de probabilidade [MUR 1990]. Tem como vantagem o fornecimento de um esquema consistente de parametrização. Como desvantagem, o uso de equações matemáticas não-triviais e a obtenção de propriedades otimizadas pode não ocorrer para amostras pequena [NIS 2002].
- c) Mínimos Quadrados: método não linear que fornece uma alternativa ao MLE. Tem como vantagem obter razoáveis estimativas quando o MLE não puder ser aplicado e como principal desvantagem não produzir propriedades tão otimizadas quanto o MLE [NIS 2002].

Não é simples a escolha de um método em particular que forneça parâmetros apropriados para cada distribuição [HAY 1996]. Não existe um método único que produza tais parâmetros para todas distribuições, tendo o contexto dos dados analisados como principal referencial para a escolha do método mais apropriado [LAW 2000].

Assim, para a estimação de parâmetros das distribuições implementadas pelo AvaDisPro, utilizou-se uma mescla destes métodos, de acordo com a viabilidade de implementação de cada método.

## 3.6 Distribuições Contínuas

As distribuições de probabilidade contínuas a seguir representam o conjunto de distribuições implementadas pelo AvaDisPro, de acordo com as propriedades de cada distribuição.

### 3.6.1 Exponencial

Distribuição utilizada para modelar intervalos de tempo de chegada de clientes em um sistema, cuja chegada ocorre com uma determinada taxa constante [LAW 2000]. Também pode ser utilizada para modelar o intervalo de tempo até a falha de uma peça de um equipamento [HAY 1996]. A tabela 3.1 apresenta as propriedades da Exponencial implementadas pelo AvaDisPro e a figura 3.1 o seu gráfico<sup>8</sup> da FDP.

Tabela 3.1 – Propriedades da Distribuição Exponencial

Exponencial – Propriedades [LAW 2000]	
FDP	$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x \geq 0$
FPA	$F(x) = 1 - e^{-x/\beta}, \quad x \geq 0$
Parâmetros	$\beta > 0$
Constantes	$e$ (logaritmo de Euler)
Intervalo	$[0, \infty)$
Estimação (MLE)	$\beta = \bar{X}(n)$

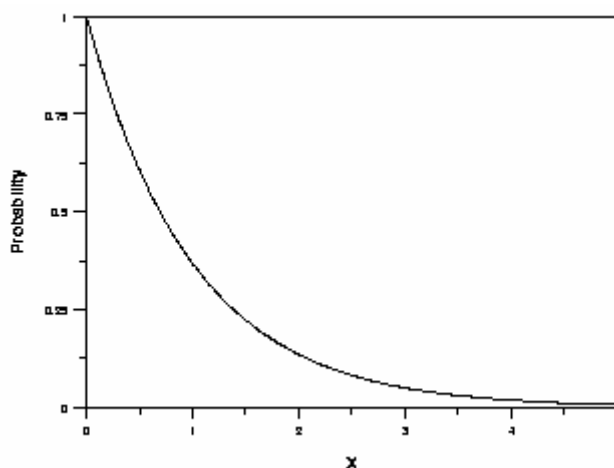


Figura 3.1 – FDP da Distribuição Exponencial [NIS 2002, p. 410]

<sup>8</sup> O rótulo da ordenada não é probabilidade, e sim função densidade de probabilidade. Isso se aplica a todas distribuições contínuas. Neste trabalho, os gráficos da Exponencial, Gama, Normal e Uniforme.

### 3.6.2 Gama

Distribuição utilizada para modelar o tempo para completar alguma tarefa, isto é, serviços ao cliente ou manutenção de máquinas [LAW 2000]. A tabela 3.2 apresenta as propriedades da Gama e a figura 3.2 a sua FDP.

Tabela 3.2 – Propriedades da distribuição Gama

Gama – Propriedades [LAW 2000]	
FDP	$f(x) = \frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)}, \quad x > 0$
FPA	$F(x) = \frac{\Gamma_x(\alpha)}{\Gamma(\alpha)}, \quad x > 0$
Função Gama	$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$
Função Gama (Incompleta)	$\Gamma_x(z) = \int_0^x t^{z-1} e^{-t} dt$
Parâmetros	$\alpha > 0 \quad \beta > 0$
Intervalo	$[0, \infty)$
Estimação (MOM) [NIS 2002]	$\alpha = \left(\frac{\mu}{\sigma}\right)^2$ $\beta = \frac{\sigma^2}{\mu}$

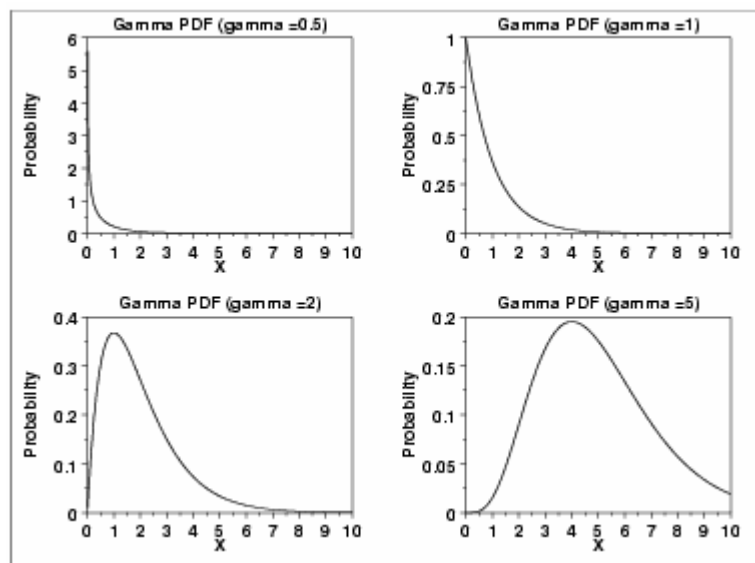


Figura 3.2 – FDP da Distribuição Gama [NIS 2002, p. 439]

### 3.6.3 Log-Logistic

Distribuição utilizada para modelar o tempo para completar alguma tarefa [LAW 2000]. A tabela 3.3 apresenta as propriedades da distribuição Log-Logistic.

Tabela 3.3 – Propriedades da distribuição Log-Logistic

Log-Logistic – Propriedades [LAW 2000]	
FDP	$f(x) = \frac{\alpha(x/\beta)^{\alpha-1}}{\beta[1+(x/\beta)^\alpha]^2}, \quad x > 0$
FPA	$F(x) = \frac{1}{1+(x/\beta)^{-\alpha}}, \quad x > 0$
Parâmetros	$\alpha > 0 \quad \beta > 0$
Constantes	$\pi = 3,141593...$
Intervalo	$[0, \infty)$
Estimação (MLE)	$\alpha = \left  \operatorname{cosec}\left(\frac{\pi}{\sigma}\right) \right  \quad \beta = \frac{1}{3}(\pi\alpha)^2$

### 3.6.4 Normal

Também conhecida como distribuição de Gauss, é utilizada na modelagem de erros de diversos tipos, como, por exemplo, no ponto de impacto de uma bomba ou na quantidade que são a soma de números grandes de outras quantidades (utilizando o teorema do limite central) [LAW 2000]. É considerada a distribuição mais importante [NIS 2002], servindo de base para outras distribuições. A tabela 3.4 apresenta as propriedades da Normal e a figura 3.3 apresenta a sua FDP.

Tabela 3.4 – Propriedades da distribuição Normal

Normal – Propriedades [LAW 2000]	
FDP	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}, \quad x \text{ é um número real.}$
FPA	$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad x \text{ é um número real.}$
Parâmetros	$\mu \in (-\infty, \infty) \quad \sigma > 0$
Intervalo	$(-\infty, \infty)$
Estimação (MLE)	$\mu = \bar{X}(x) \quad \sigma = \left[ \frac{n-1}{n} S^2(n) \right]^{1/2}$

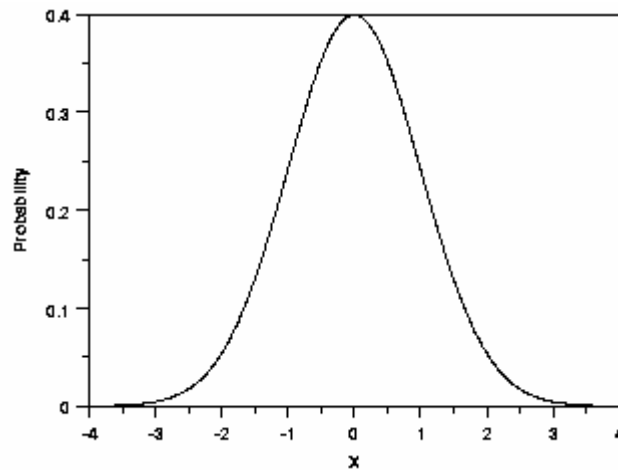


Figura 3.3 – FDP da Distribuição Normal [NIS 2002, p. 377]

### 3.6.5 Triangular

Distribuição utilizada como modelo bruto na ausência de dados [NIS 2002]. A tabela 3.5 apresenta as propriedades da Triangular.

Tabela 3.5 – Propriedades da distribuição Triangular

Triangular – Propriedades [LAW 2000]	
FDP	$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)}, & a \leq x \leq c \\ \frac{2(b-x)}{(b-a)(b-c)}, & c < x \leq b \end{cases}$
FPA	$F(x) = \begin{cases} \frac{(x-a)^2}{(b-a)(c-a)}, & a \leq x \leq c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)}, & c < x \leq b \end{cases}$
Parâmetros	$a < c < b$
Intervalo	$[a, b]$
Estimação (MOM)	$a = \min_{1 \leq i \leq n} X_i \quad b = \max_{1 \leq i \leq n} X_i$ $c = \left  \bar{X} * 3 - \min_{1 \leq i \leq n} X_i - \max_{1 \leq i \leq n} X_i \right $

### 3.6.6 Uniforme

Considerada a distribuição contínua mais simples [HAY 1996], tendo em sua FDP uma probabilidade horizontal, é utilizada na validação de números randômicos [LAW 2000]. A tabela 3.6 apresenta as propriedades da Uniforme e a figura 3.4 mostra a sua FDP.

Tabela 3.6 – Propriedades da distribuição Uniforme

Uniforme – Propriedades [LAW 2000]	
FDP	$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$
FPA	$F(x) = \frac{x-a}{b-a}, \quad a \leq x \leq b$
Parâmetros	$a < b$
Intervalo	$[a, b]$
Estimação (MLE)	$a = \min_{1 \leq i \leq n} X_i$ $b = \max_{1 \leq i \leq n} X_i$

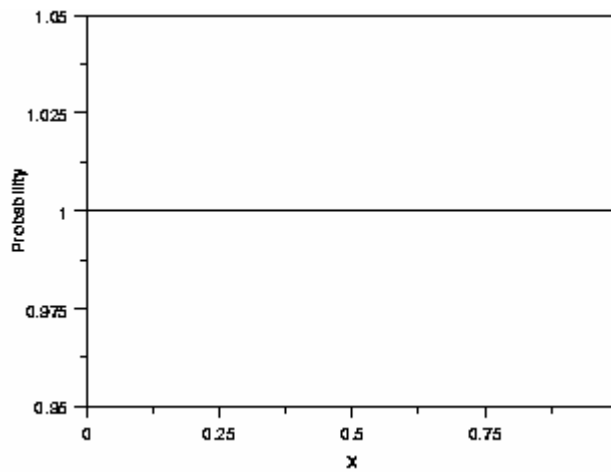


Figura 3.4 – FDP da Distribuição Uniforme [NIS 2002, p. 384]

### 3.6.7 Weibull

A distribuição Weibull é utilizada para modelar o tempo para se completar alguma tarefa e o tempo entre falhas de uma peça [LAW 2000]. A tabela 3.7 apresenta as suas propriedades.

Tabela 3.7 – Propriedades da distribuição Weibull

Weibull – Propriedades [LAW 2000]	
FDP	$f(x) = \alpha \beta^{-\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha}, \quad x > 0$
FPA	$F(x) = 1 - e^{-(x/\beta)^\alpha}, \quad x > 0$
Parâmetros	$\alpha > 0 \quad \beta > 0$
Intervalo	$[0, \infty)$
Estimação (MOM) [MOH 2000]	$\alpha = \frac{1}{n} \sum_{i=1}^n X_i \quad \beta = \frac{\sigma^2}{\mu}$



### 3.7 Distribuições Discretas

As distribuições de probabilidade discretas a seguir representam o conjunto de distribuições implementado pelo AvaDisPro, de acordo com as propriedades de cada distribuição.

#### 3.7.1 Binomial

Distribuição proposta por Bernoulli (1713), é utilizada para modelar o número de sucessos num experimento  $t$  independente de Bernoulli com probabilidade  $p$  de sucesso em cada experimento [LAW 2000]. Também é utilizada para verificar o número de itens defeituosos num lote de tamanho  $t$ . Possui aplicações em controle de qualidade, seguros e outros problemas industriais. [HAY 1996]. A tabela 3.8 apresenta as propriedades da Binomial e a figura 3.5 um gráfico para algumas FMP dessa distribuição.

Tabela 3.8 – Propriedades da distribuição Binomial

Binomial – Propriedades [LAW 2000]	
FMP	$p(x) = \binom{t}{x} p^x (1-p)^{t-x}, \quad x \in \{0, 1, \dots, t\}$
FPA	$F(x) = \sum_{i=0}^{\lfloor x \rfloor} \binom{t}{i} p^i (1-p)^{t-i}, \quad 0 \leq x \leq t$
Coeficiente Binomial	$\binom{t}{x} = \frac{t!}{x!(t-x)!}$
Parâmetros	$p \in (0, 1)$ $t$ é um inteiro positivo
Intervalo	$\{0, 1, \dots, t\}$
Estimação (MLE)	$t = n \quad p = \frac{\bar{X}(n)}{t}$

#### 3.7.2 Binomial Negativa

Também conhecida como Pascal, a distribuição Binomial Negativa é utilizada para modelar o número de falhas antes de sucessos  $s$  num experimento independente de Bernoulli com probabilidade  $p$  de sucesso em cada experimento [LAW 2000]. Também é utilizada para verificar o número de itens bons inspecionados antes de se encontrar um item defeituoso  $s$  [LAW 2000]. A tabela 3.9 apresenta as propriedades as suas propriedades.

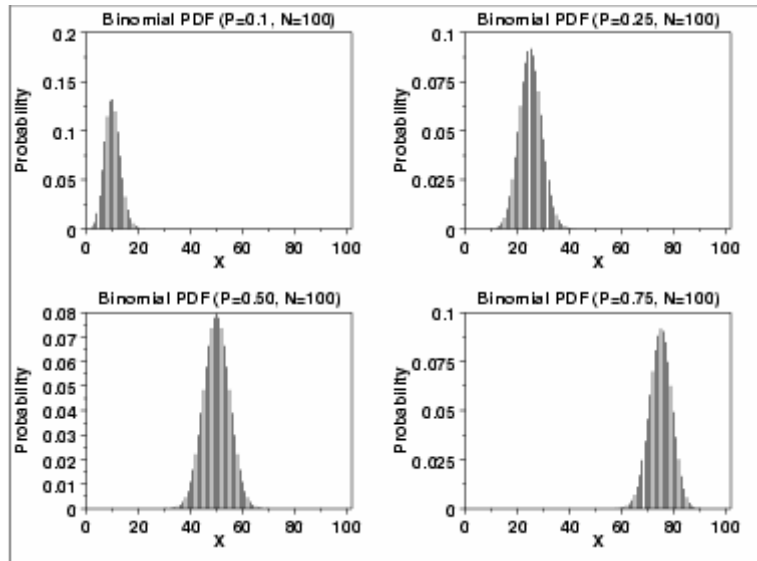


Figura 3.5 – FMP da Distribuição Binomial [NIS 2002, p. 482]

Tabela 3.9 – Propriedades da distribuição Binomial Negativa

Binomial Negativa – Propriedades [LAW 2000]	
FMP	$p(x) = \binom{s+x-1}{x} p^s (1-p)^x, \quad x \in \{0,1,\dots\}$
FPA	$F(x) = \sum_{i=0}^{\lfloor x \rfloor} \binom{s+i-1}{i} p^s (1-p)^i, \quad x \geq 0$
Parâmetros	$p \in (0,1)$ $s$ é um inteiro positivo
Intervalo	$\{0,1,\dots\}$
Estimação (MLE)	$s = n$ $p = \frac{s}{\overline{X}(n) + s}$

### 3.7.3 Geométrica

Distribuição utilizada para modelar o número de falhas antes do primeiro sucesso numa sequência dum experimento independente de Bernoulli com probabilidade  $p$  de sucesso em cada experimento [LAW 2000]. Também conhecida como distribuição Pascal (quando  $s = 1$ ), é utilizada para verificar o número de itens inspecionados antes de se encontrar o primeiro item defeituoso [LAW 2000]. A tabela 3.10 apresenta as propriedades da distribuição Geométrica.

Tabela 3.10 – Propriedades da distribuição Geométrica

Geométrica – Propriedades [LAW 2000]	
FMP	$p(x) = p(1-p)^x, \quad x \in \{0,1,\dots\}$
FPA	$F(x) = 1 - (1-p)^{\lfloor x \rfloor + 1}, \quad x \geq 0$
Parâmetros	$p \in (0,1)$
Intervalo	$\{0,1,\dots\}$
Estimação (MLE)	$p = \frac{1}{\overline{X}(n)+1}$

### 3.7.4 Poisson

Distribuição proposta pelo matemático francês Simeon Denis Poisson (1781 – 1840) [HAY 1996]. É utilizada para definir uma variável aleatória que conta o número de eventos ocorridos dentro de certos limites. Por exemplo, um pesquisador pode estar interessado no número de itens defeituosos de uma peça, no número de partículas radioativas emitidas por uma substância ou ainda no número de chamadas telefônicas recebidas por um operador dentro de um determinado limite [HAY 1996]. A tabela 3.11 mostra as propriedades da Poisson e a figura 3.6 um gráfico para algumas FMP dessa distribuição.

Tabela 3.11 – Propriedades da distribuição Poisson

Poisson – Propriedades [LAW 2000]	
FMP	$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x \in \{0,1,\dots\}$
FPA	$F(x) = e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!}, \quad x \geq 0$
Parâmetros	$\lambda > 0$
Intervalo	$\{0,1,\dots\}$
Estimação (MLE)	$\lambda = \overline{X}(n)$

### 3.7.5 Uniforme Discreta

Distribuição utilizada para modelar a ocorrência aleatória com diversas possibilidades de resultado, cada um sendo distribuído igualmente [LAW 2000]. É, em geral, utilizada como um primeiro modelo para uma quantidade de inteiros variando de  $i$  até  $j$  [HAY 1996]. A tabela 3.12 mostra as propriedades da distribuição Uniforme Discreta.

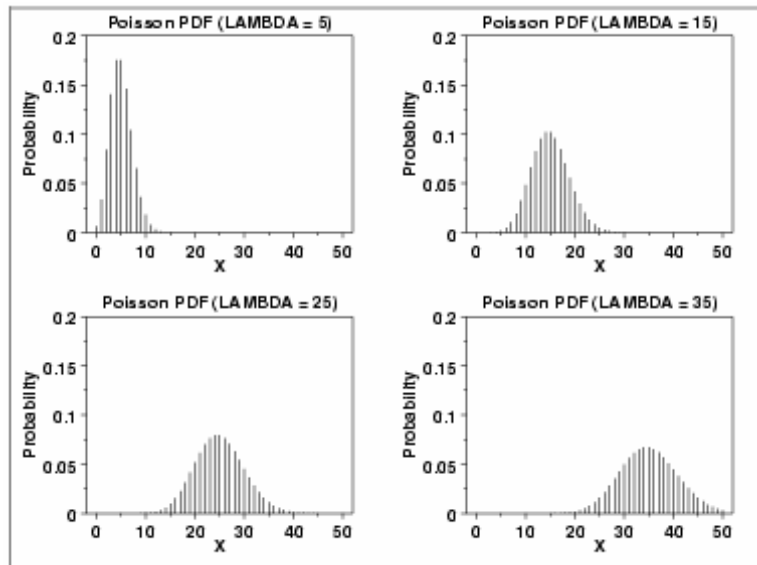


Figura 3.6 – FMP da Distribuição Poisson [NIS 2002, p. 485]

Tabela 3.12 – Propriedades da distribuição Uniforme Discreta

Uniforme Discreta – Propriedades [LAW 2000]	
FMP	$p(x) = \frac{1}{j-i+1}, \quad x \in \{i, i+1, \dots, j\}$
FPA	$F(x) = \frac{\lfloor x \rfloor - i + 1}{j - i + 1}, \quad i \leq x \leq j$
Parâmetros	$i \leq j$ $i$ e $j$ são inteiros
Intervalo	$\{i, i+1, \dots, j\}$
Estimação (MLE)	$i = \min_{1 \leq k \leq n} X_k$ $j = \max_{1 \leq k \leq n} X_k$

## 4 Testes de Adequação

Neste capítulo apresenta-se os testes implementados pelo AvaDisPro para verificar se um conjunto de dados qualquer se adapta a uma determinada distribuição de probabilidade. Para tal, escolheu-se dois testes: o Qui-Quadrado ( $\chi^2$ ) e Kolmogorov-Smirnov (K-S), sendo o primeiro melhor indicado quando uma amostra tem tamanho maior a 30 elementos [SPI 1972].

O critério para escolha destes dois testes baseia-se na utilidade (atendem os requisitos propostos pelo sistema, isto é, testar uma amostra com distribuições de probabilidade) e na utilização, já que são empregados em boa parte dos softwares estatísticos (ver capítulo 2).

Como o AvaDisPro é uma ferramenta extensível e aberta, futuras atualizações de outros testes poderão ser realizadas futuramente.

### 4.1 Qui-Quadrado ( $\chi^2$ )

O teste  $\chi^2$  é usado para testar se uma amostra de dados pode ser representada por uma distribuição específica [McL 2000] e é definido em forma de hipótese [NIS 2002]:

$H_0$ : Os dados seguem uma distribuição específica.

$H_1$ : Os dados não seguem uma distribuição específica.

Para efeitos de teste, os dados são divididos dentro de  $k$  classes e a estatística do teste é definida como [NIS 2002]:

$$\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$$

onde  $o_i$  é a frequência observada para cada classe  $i$  e  $e_i$  é a frequência esperada para cada classe  $i$ . A frequência esperada é calculada como [NIS 2002]:

$$E_i = N(F(Y_u) - F(Y_l))$$

onde  $F$  é a função de probabilidade acumulada para a distribuição que está sendo testada,  $Y_u$  é o limite superior da classe  $i$ ,  $Y_l$  é o limite inferior para a classe  $i$  e  $N$  é o tamanho da amostra.

#### 4.1.1 Região Crítica

A estatística do teste segue aproximadamente a distribuição  $\chi^2$  com  $(k - c)$  graus de liberdade, onde  $k$  é o número de células não-vazias e  $c$  é o número de parâmetros da distribuição + 1 [NIS 2002]. Por exemplo, para aplicar o teste  $\chi^2$  usando a distribuição normal,  $c = 3$  (considerando dois parâmetros,  $\mu$  e  $\sigma$ ).

Assim, a hipótese que a variável testada segue alguma distribuição é rejeitada se:

$$\chi^2 > \chi^2_{(\alpha, k-c)}$$

onde  $\chi^2_{(\alpha, k-c)}$  é o resultado da função  $\chi^2$  com  $k - c$  graus de liberdade e nível de significância de  $\alpha$  [NIS 2002].

O teste  $\chi^2$  pode ser aplicado em qualquer distribuição de probabilidade (contínua ou discreta), mas se recomenda que o seu resultado seja ignorado caso o tamanho da amostra seja menor que 30 elementos, já que nesse caso há uma tendência de aceitação de  $H_0$  em qualquer situação [SPI 1972]. Outra restrição do  $\chi^2$  refere-se às frequências esperadas: quando  $e_i < 5$  não se considera apropriado o seu uso [SPI 1972].

## 4.2 Kolmogorov-Smirnov (K-S)

O teste Kolmogorov-Smirnov (K-S) é baseado na Função de Distribuição Empírica (FDE) [NIS 2002]. Dado  $n$  pontos de dados ordenados  $Y_1, Y_2, \dots, Y_N$ , a FDE é definida como:

$$E_N = \frac{n(i)}{N}$$

onde  $n(i)$  é o número de pontos menor que  $Y_i$ , os quais são ordenados do menor ao maior valor.

A figura 4.1 representa a FDE com a função cumulativa da distribuição normal para 100 números aleatórios [NIS 2002]. O teste K-S é baseado na distância máxima entre estas duas curvas.

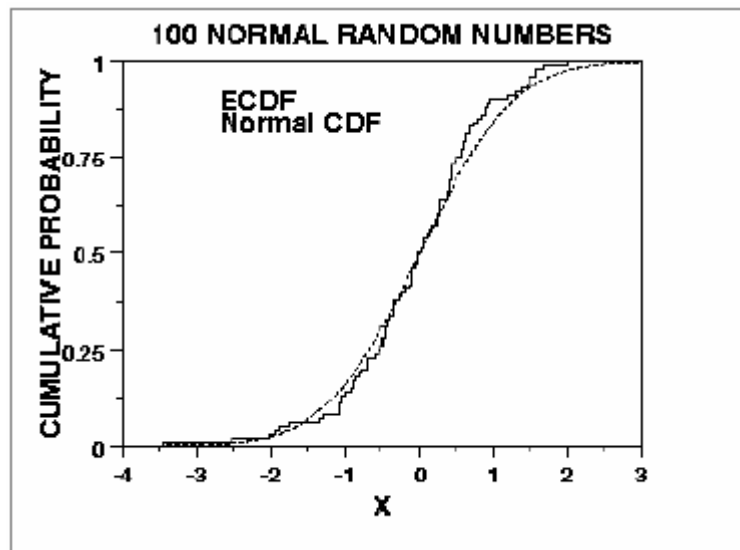


Figura 4.1 – FPA da Normal versus FDA da distribuição empírica [NIS 2002, p. 323]

O teste K-S é definido como [NIS 2002]:

$H_0$ : Os dados seguem uma distribuição específica.

$H_a$ : Os dados não seguem uma distribuição específica.

A estatística do teste é definida como [NIS 2002]:

$$D = \max_{1 \leq i \leq N} \left| F(Y_i) - \frac{i}{N} \right|$$

onde  $F$  é a função cumulativa da distribuição que se está a testar<sup>9</sup>.

A hipótese é rejeitada se a estatística do teste,  $D$ , é maior que o valor crítico obtido da tabela K-S<sup>10</sup>. Caso contrário, a hipótese é aceita.

---

<sup>9</sup> Alguns autores consideram que o teste K-S deve ser aplicado somente em distribuições contínuas, como é o caso de Nist/Sematech [NIS 2002]. Entretanto, outros autores, como Law [LAW 2000] e Spiegel [SPI 1972] não fazem essa separação, razão pela qual o AvaDisPro implementa o teste K-S para ambos tipos de dados (contínuo e discreto).

<sup>10</sup> Segundo Nist/Sematech [NIS 2002], existem algumas variações de valores de tabelas K-S na literatura, as quais consideram diferentes escalas para o teste K-S e regiões críticas. O AvaDisPro implementa a tabela (ver o anexo E) disponível em *Probability And Statistics: For Engineers And Scientists*, de Hayter [HAY 1996].

## 5 Projeto do AvaDisPro

Este capítulo apresenta o projeto da ferramenta AvaDisPro. A arquitetura do software é discutida, sendo proposta uma solução de desenvolvimento em duas camadas. Também são apresentados os diagramas de caso de uso, seqüência e classe, que seguem a linguagem UML, mais precisamente a MRDS (Metodologia para Rápido Desenvolvimento de Sistemas), proposta por Schmitz e Silveira [SCH 2001], a qual representa uma metodologia interessante para o desenvolvimento de sistemas pequenos. Discute-se também os testes realizados na ferramenta, e os seus resultados são apresentados.

Dessa forma, as informações deste capítulo devem representar um importante subsídio para o desenvolvimento futuro do AvaDisPro.

### 5.1 Arquitetura

Definir um modelo de desenvolvimento é uma etapa importante na construção de sistemas, de modo que se possa afirmar que o software é construído com padrões previamente estabelecidos [PRE 1995]. A UML é a linguagem de modelagem padrão atualmente [FOW 2000] e algumas metodologias foram desenvolvidas para o seu uso, como, por exemplo, o RUP (*Rational Unified Process*) [IBM 2003]. O RUP é uma metodologia extensa que exige, na maior parte dos casos, que um subconjunto seja extraído de acordo com a necessidade do projeto.

Uma ferramenta cujo objetivo é auxiliar no projeto de pequenos sistemas é o FastCASE [SCH 2001]. Trata-se de um software gratuito que implementa a MRDS, a qual foi empregada pelas seguintes razões:

- a) Utiliza um subconjunto da UML com base em quatro diagramas: caso de uso, classe, seqüência e estados. Tal redução é importante para o desenvolvimento de pequenos sistemas.
- b) Introduz uma metodologia de desenvolvimento na qual as classes do sistema são divididas em classes de negócio, controle e domínio [SCH 2000]. Tal arquitetura tende a melhorar o processo de desenvolvimento, de modo que uma futura manutenção seja facilitada, já que se separou convenientemente o negócio do sistema da sua interface e de seu controle [SCH 2001].

Para o processo de desenvolvimento do AvaDisPro, utilizou-se como base a MRDS, à exceção de alguns detalhes próprios da natureza do problema, como, por exemplo, a não utilização de classes de domínio, já que o AvaDisPro não trabalha com base de dados. Assim, o AvaDisPro apresenta uma arquitetura em duas camadas, isto é, classes de interface e classes de negócio.



## 5.2 Ferramentas Utilizadas

O desenvolvimento do AvaDisPro foi realizado utilizando-se as seguintes ferramentas, de acordo com a tabela 5.1:

Tabela 5.1 – Ferramentas e Metodologias de Desenvolvimento

Ferramentas e Metodologias de Desenvolvimento	
Linguagem de Programação	Object Pascal
Ambiente de Programação	Delphi 6 / Kylix 2
Biblioteca de Desenvolvimento	CLX <sup>11</sup>
Linguagem de Modelagem	UML
Metodologia de Desenvolvimento	Orientação a Objetos com MRDS
Ambiente de Modelagem	Rational Rose 2000 <sup>12</sup>

## 5.3 Diagrama de Caso de Uso

A seguir apresentam-se os casos de uso implementados, bem como suas descrições. A figura 5.1 mostra o diagrama de caso de uso para o AvaDisPro. Esse diagrama apresenta um único ator, que é o usuário do sistema, que interage em 8 situações.

### 5.3.1 Caso de Uso: Escolha de Tipo de Dados

O usuário pode escolher o tipo de dado que deseja utilizar no sistema, que são dois: contínuo, que representa variáveis contínuas, e discreto, que representa variáveis discretas. Ao escolher um tipo de dado, as distribuições de probabilidade para aquele conjunto tornam-se disponíveis. Quando um arquivo de dados já estiver aberto, o sistema automaticamente deve fazer a alternância entre as distribuições disponíveis, de acordo com o tipo de dado selecionado.

---

<sup>11</sup> Biblioteca compatível com os ambientes GNU/Linux e MS-Windows.

<sup>12</sup> Apesar das tentativas em se utilizar alguma ferramenta gratuita e/ou livre para modelagem do sistema, não foi possível encontrar nenhum software que satisfizesse os requisitos mínimos de desenho e documentação para o AvaDisPro. As seguintes ferramentas foram testadas: ArgoUML 0.12, dia 0.9, FastCASE 1.1a, UMLet 1.8 e Violet 0.12.

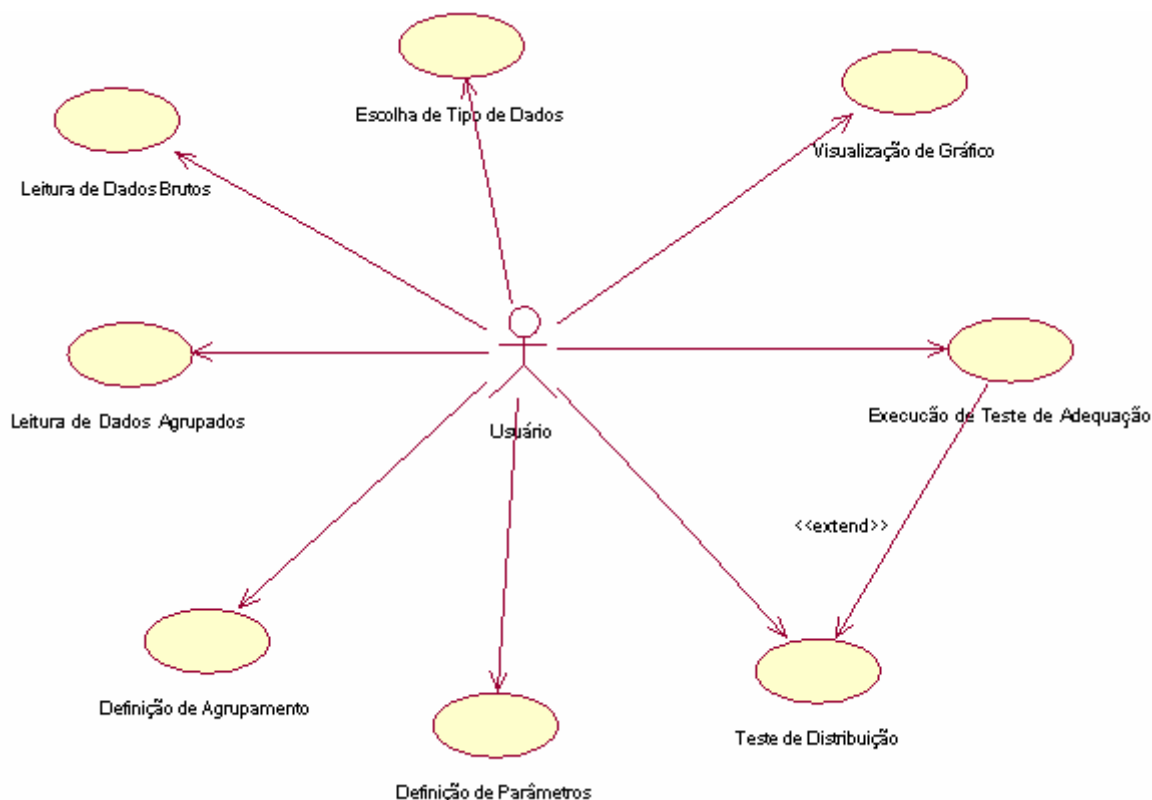


Figura 5.1 – Diagrama de Caso de Uso

### 5.3.2 Caso de Uso: Leitura de Dados Brutos

Um arquivo de dados do tipo texto representa dados brutos. O usuário já deve ter editado seu arquivo previamente antes de selecioná-lo pelo AvaDisPro. O dados devem ser simplesmente digitados – em qualquer disposição – em um arquivo texto. Após selecionar a opção para leitura de dados, o AvaDisPro automaticamente executa os seguintes procedimentos: insere em uma tabela todos os valores lidos; gera um resumo com informações estatísticas e agrupa os dados em classes apropriadas de acordo com o número de intervalos definido pelo usuário.

### 5.3.3 Caso de Uso: Leitura de Dados Agrupados

Um arquivo de dados do tipo texto representa dados agrupados. O usuário já deve ter editado seu arquivo previamente antes de selecioná-lo pelo AvaDisPro. O formato do arquivo deve ser: os dados devem estar divididos em classes, uma por linha, como por exemplo: 20 40 10, que significa: 20 é o limite inferior da classe, 40 representa o limite superior e 10 é a quantidade de elementos naquele intervalo. Após selecionar a opção para leitura de dados, o AvaDisPro verifica se os dados do arquivo estão corretos. Caso estejam, o usuário poderá executar suas opções. Caso contrário, um aviso é emitido ao usuário sobre o erro na formação do seu arquivo.

#### 5.3.4 Caso de Uso: Definição de Agrupamento

Após a abertura do arquivo de entrada pelo usuário, se o tipo de dado for bruto é possível alterar três parâmetros do agrupamento em classes: valor inicial do intervalo, valor final do intervalo e total de classes. Depois de alguma alteração, o AvaDisPro reagrupa as classes de modo a refletir a nova realidade. As duas primeiras opções permitem testar um subconjunto do conjunto de dados, já que se pode convenientemente selecionar faixas de dados dentro do próprio conjunto de dados. Caso o tipo de dado já esteja agrupado, não é possível modificar a definição de agrupamento. Nesse caso, o próprio arquivo deve ser modificado.

#### 5.3.5 Caso de Uso: Definição de Parâmetros

Após abrir o seu arquivo de dados, o usuário pode selecionar alguma distribuição de probabilidade e alterar o(s) seu(s) parâmetro(s). Os parâmetros iniciais são calculados pelo AvaDisPro. Se o usuário alterar um parâmetro de maneira incorreta, isto é, que viole alguma regra da distribuição escolhida, o AvaDisPro quando for realizar algum teste de adequação informará um erro, indicando qual distribuição o contém. As modificações de parâmetros são válidas somente quando o arquivo de dados do usuário estiver aberto. Em outra futura reabertura de arquivo, os parâmetros iniciais são os mesmos calculados pelo AvaDisPro.

#### 5.3.6 Caso de Uso: Teste de Distribuição

Uma vez aberto o arquivo de dados, o usuário pode escolher uma distribuição de probabilidade para realizar os testes de adequação  $\chi^2$  e K-S. Ao selecionar a opção para teste, o AvaDisPro mostra na tela a tabela desses testes e também mostra se o conjunto de dados segue ou não aquela distribuição. Também é disponibilizada uma opção para visualizar os valores do teste  $\chi^2$  aos níveis de significância 1, 5, 10, 25 e 50%, e também valores do teste K-S (1, 5, 20%).

#### 5.3.7 Caso de Uso: Visualização de Gráfico

O usuário seleciona a opção de visualização de gráfico após estar com algum dado carregado em memória. Se o tipo de dado for bruto e o usuário modificar algum parâmetro (valor inicial ou final do intervalo ou total de classes) o gráfico é automaticamente atualizado. A exibição do gráfico mostra no eixo  $x$  os intervalos de cada classe, e no eixo  $y$  a quantidade de elementos daquele intervalo.

### 5.3.8 Caso de Uso: Execução de Teste de Adequação

O usuário, tendo o seu arquivo de dados aberto, pode escolher executar o teste  $\chi^2$  ou o teste K-S. A execução de um desses testes é aplicada para todas distribuições de probabilidade do tipo de dado selecionado. Ao final do processo, é possível visualizar os resultados do teste para cada distribuição, bem como seus parâmetros e se esta foi ou não rejeitada. Caso algum parâmetro de alguma distribuição esteja incorreto, o sistema pára sua execução e o usuário é informado sobre qual distribuição incorre em erro.

## 5.4 Diagramas de Seqüência

Um diagrama de seqüência, às vezes também denominado diagrama de interação, trilha de mensagem ou ainda trilha de evento, representa os objetos colaborando entre si seguidos de uma narrativa de caso de uso [FUR 1998]. Cada caso de uso pode dar origem a um ou mais diagramas de seqüência. A seguir apresenta-se esses diagramas para os casos de uso supracitados.

### 5.4.1 Seqüência: Escolha de Tipo de Dado

A figura 5.2 mostra o diagrama de seqüência para o caso de uso Escolha de Tipo de Dado.

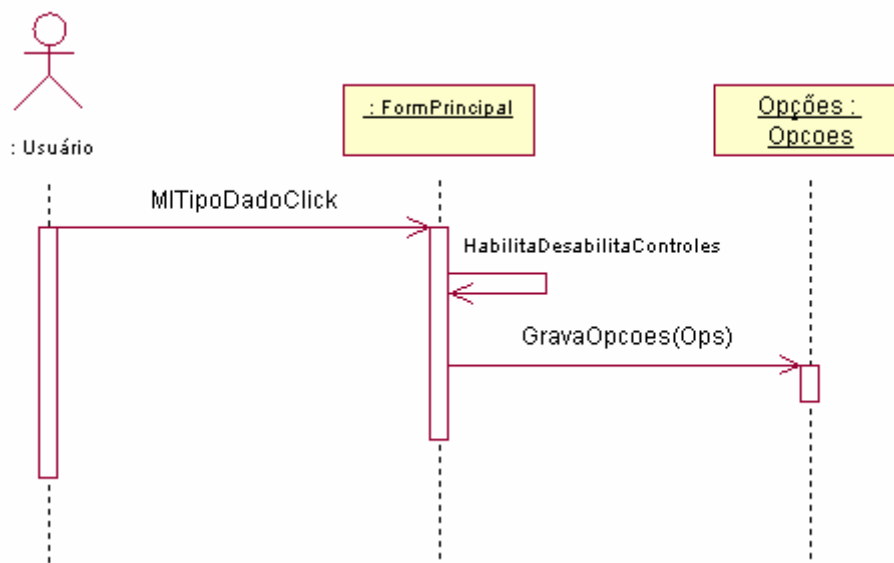


Figura 5.2 – Diagrama de Seqüência para Escolha de Tipo de Dado

### 5.4.2 Sequência: Leitura de Dados Brutos

A figura 5.3 mostra o diagrama de sequência para o caso de uso Leitura de Dados Brutos.

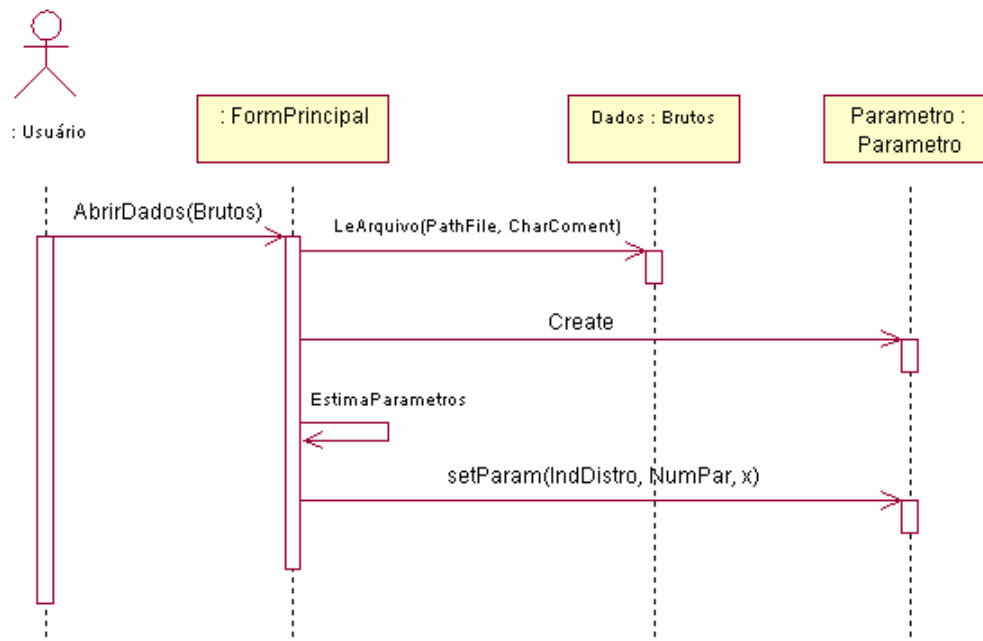


Figura 5.3 – Diagrama de Sequência para Leitura de Dados Brutos

### 5.4.3 Sequência: Leitura de Dados Agrupados

A figura 5.4 mostra o diagrama de sequência para o caso de uso Leitura de Dados Agrupados.

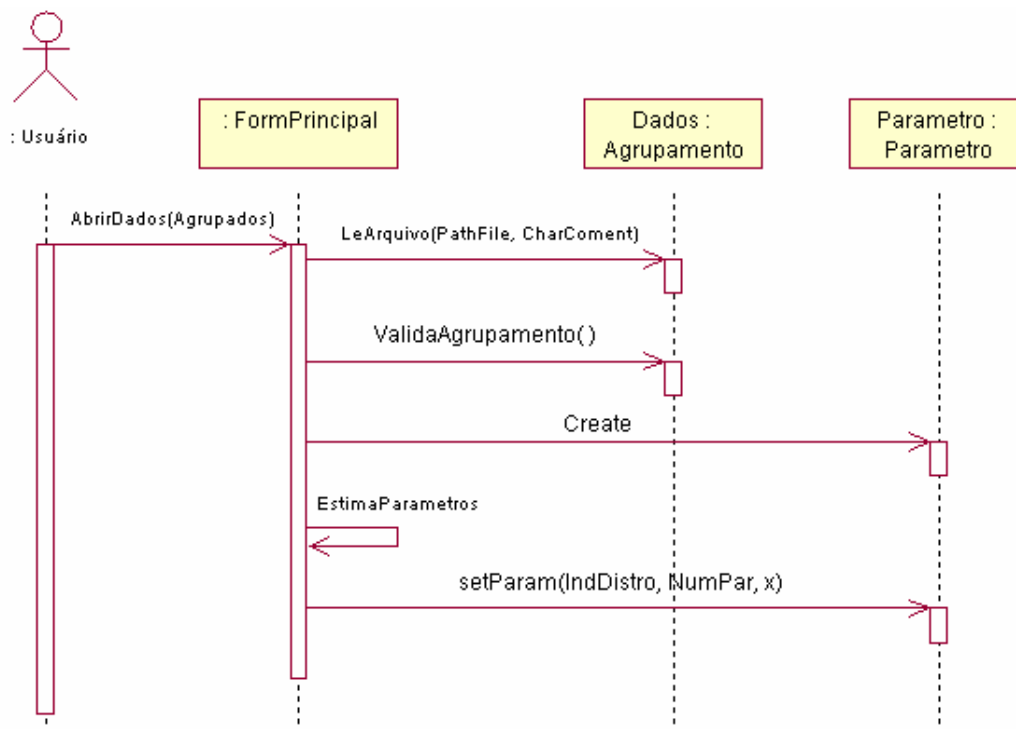


Figura 5.4 – Diagrama de Sequência para Leitura de Dados Agrupados

#### 5.4.4 Sequência: Definição de Agrupamento

A figura 5.5 mostra o diagrama de sequência para o caso de uso Definição de Agrupamento.

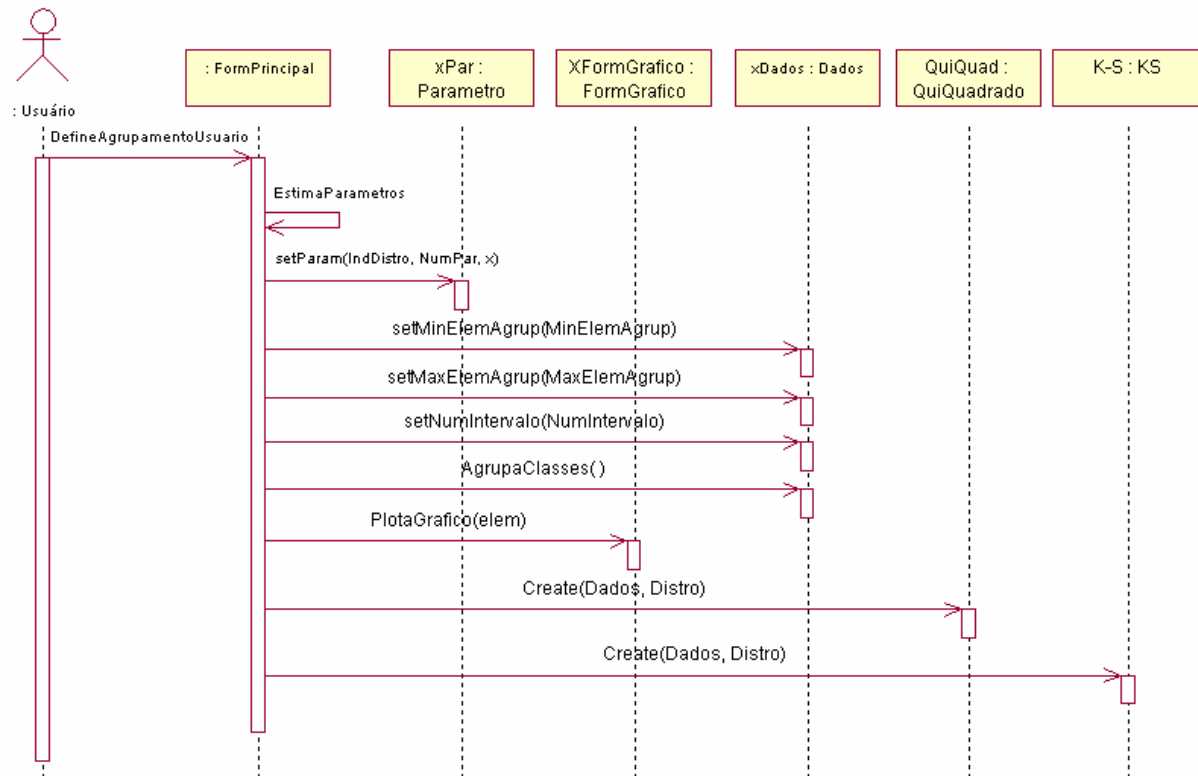


Figura 5.5 – Diagrama de Sequência para Definição de Agrupamento

#### 5.4.5 Sequência: Definição de Parâmetros

A figura 5.6 mostra o diagrama de sequência para o caso de uso Definição de Parâmetros.

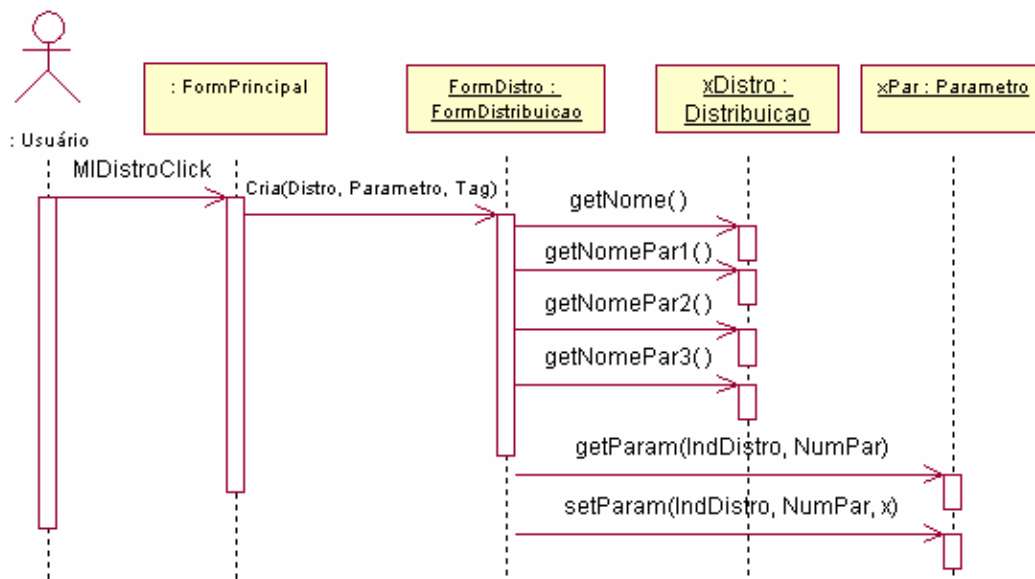


Figura 5.6 – Diagrama de Sequência para Definição de Parâmetros

#### 5.4.6 Sequência: Teste de Distribuição

A figura 5.7 mostra o diagrama de sequência para o caso de uso Teste de Distribuição.

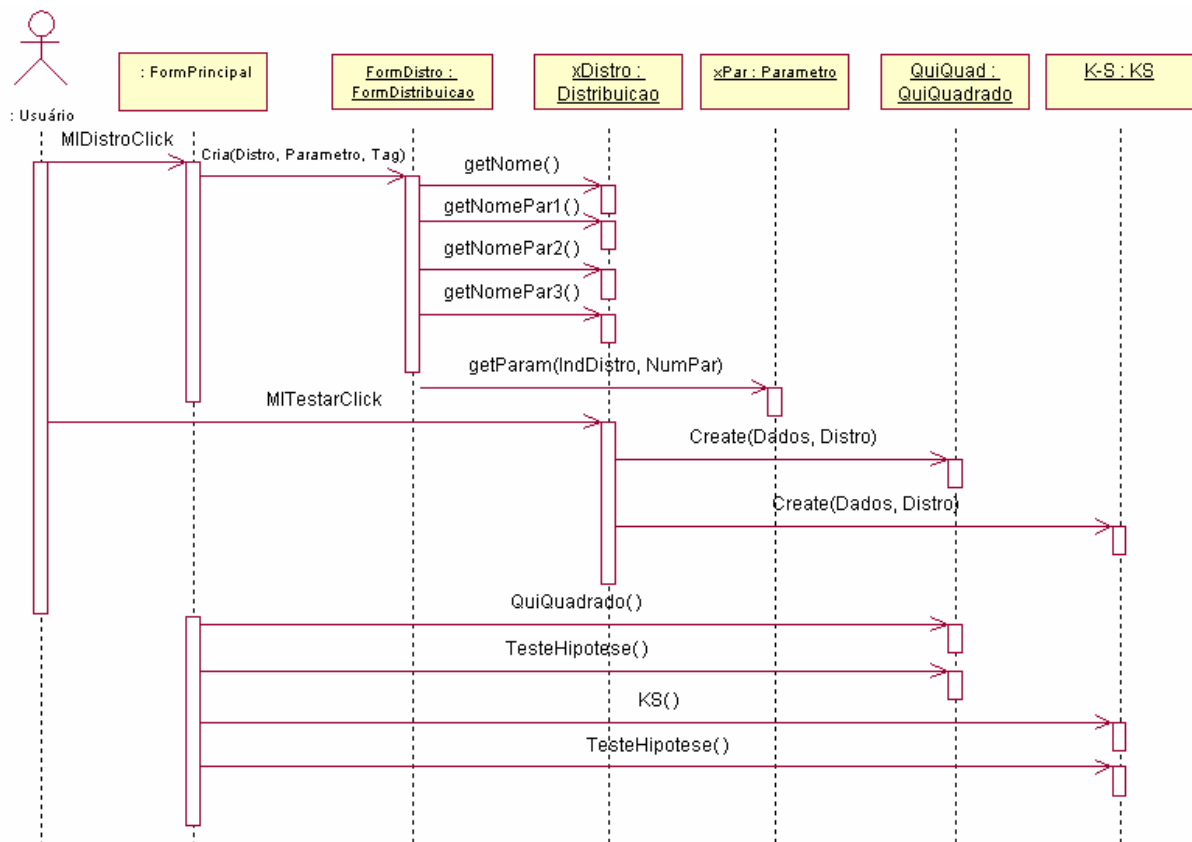


Figura 5.7 – Diagrama de Sequência para Teste de Distribuição

#### 5.4.7 Sequência: Visualização de Gráfico

A figura 5.8 mostra o diagrama de sequência para o caso de uso Visualização de Gráfico.

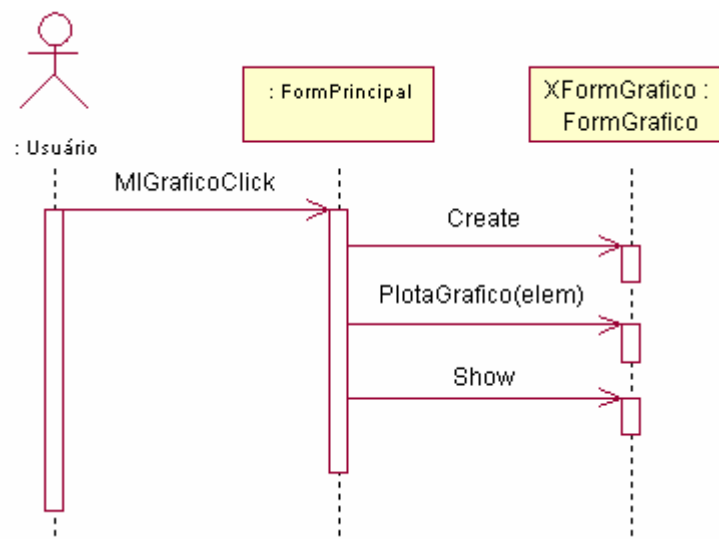


Figura 5.8 – Diagrama de Sequência para Visualização de Gráfico

#### 5.4.8 Seqüência: Execução de Teste de Adequação

A figura 5.9 mostra o diagrama de seqüência para o caso de uso Execução de Teste de Adequação.

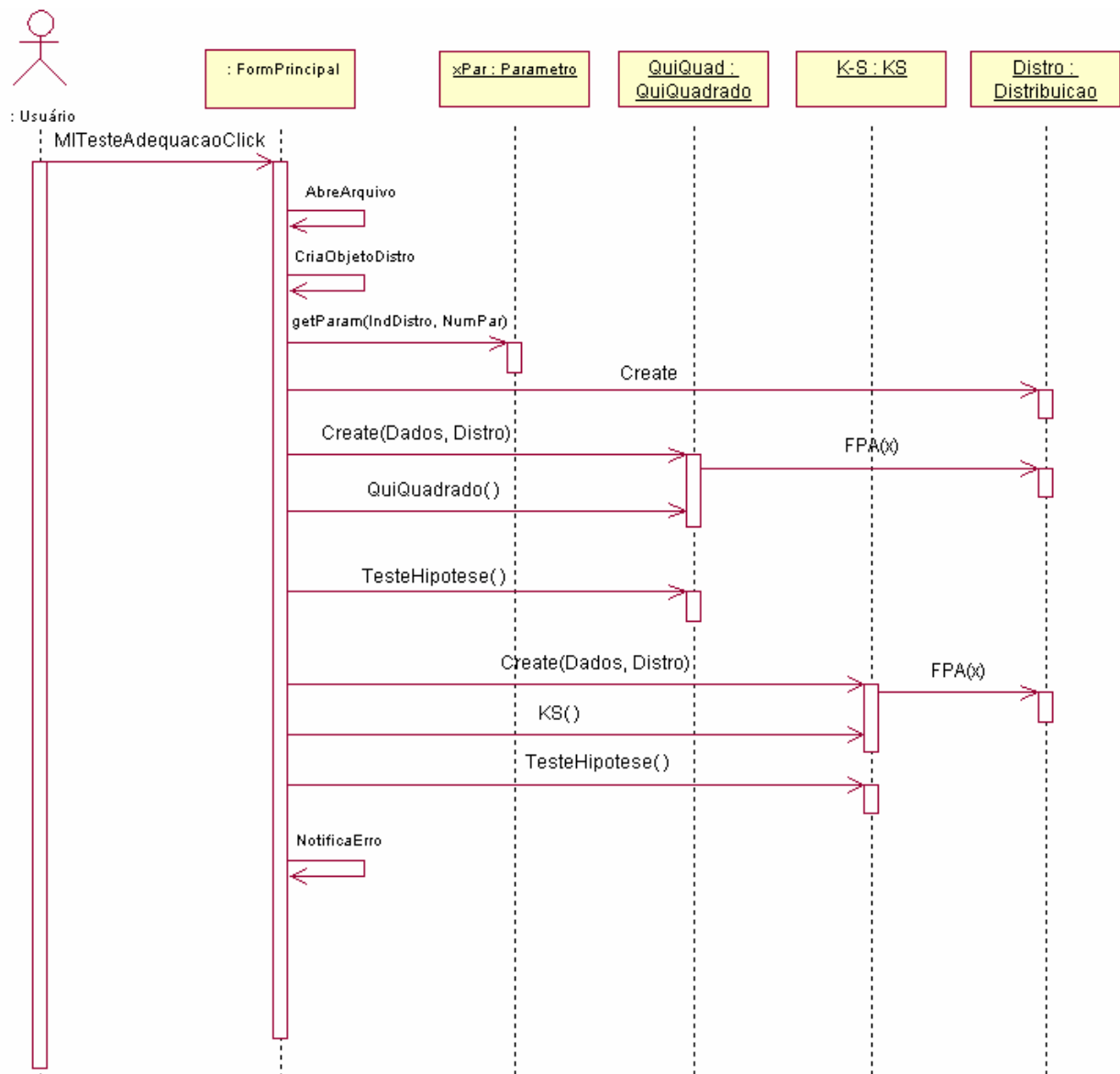


Figura 5.9 – Diagrama de Seqüência para Execução de Teste de Adequação



## 5.5 Diagrama de Classes

A seguir apresenta-se o diagrama de classes do AvaDisPro. Para uma melhor leitura, três visões foram criadas: a geral, que engloba uma visão a partir da classe principal; a do objeto *distribuicao*, que contém todas as classes de distribuições; e a do objeto *dados*, que mostra a especialização dos dados em brutos e agrupados. A soma desses três diagramas compõe o modelo de classes do AvaDisPro.

### 5.5.1 Diagrama de Classes: Visão Geral

O diagrama da figura 5.10 mostra os relacionamentos a partir da classe principal, *FormPrincipal*. As classes do tipo interface estão identificadas com `<<Interface>>`, e seus métodos e atributos estão suprimidos, já que não se tratam de classes de negócio.

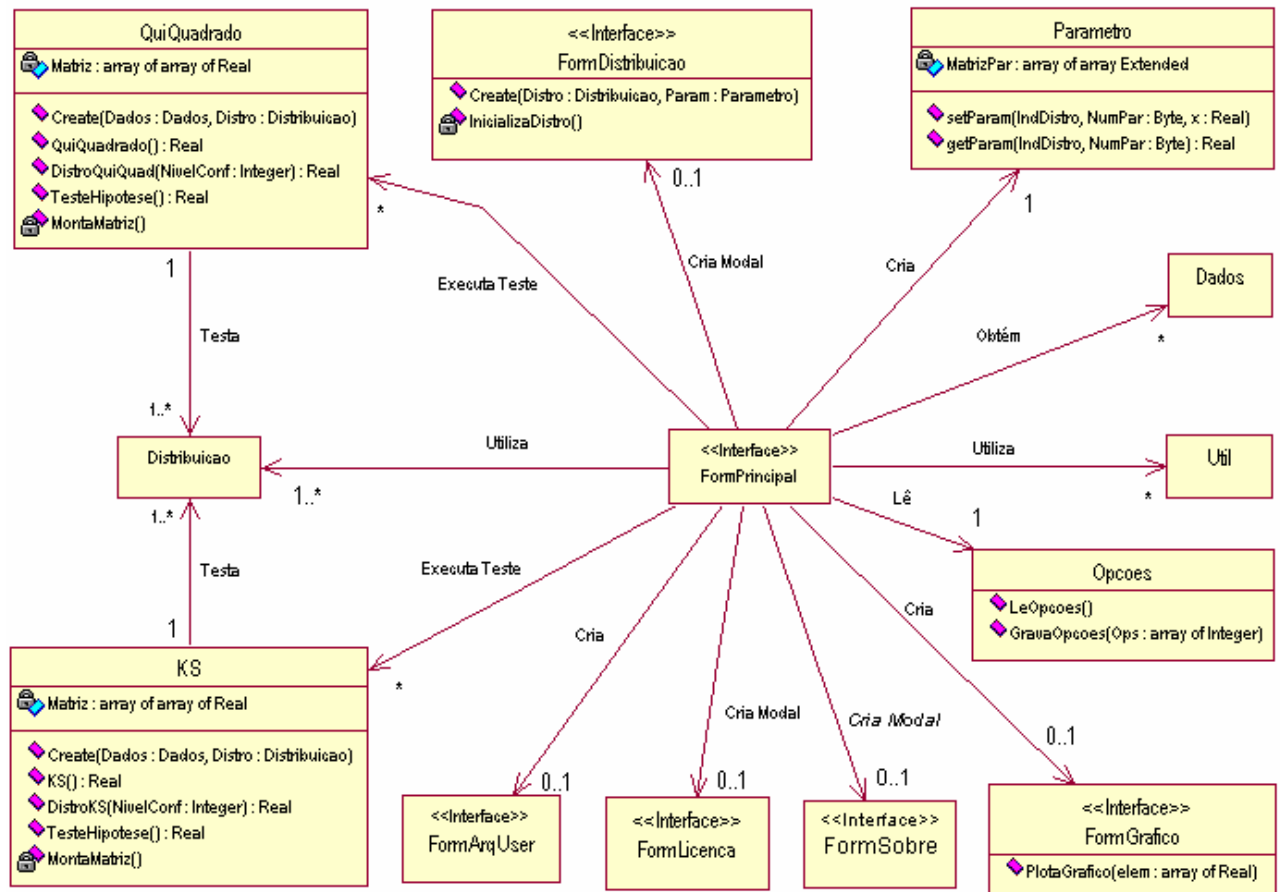


Figura 5.10 – Diagrama de Classes: Visão Geral

### 5.5.2 Diagrama de Classes: Objeto Distribuição

O objeto *distribuicao* (figura 5.11) é o núcleo central do AvaDisPro. A partir dele, é possível incluir novas distribuições no sistema. Para tal, utilizou-se a classe abstrata *Distribuicao*, que é especializada por duas classes – também abstratas – *Continua* e *Discreta*. Para, então, inserir novas distribuições no AvaDisPro, basta especializar a classe apropriada, de acordo com o tipo da distribuição. Tal mecanismo visa a conferir extensibilidade ao software [FUR 1998], uma vez que somente um objeto é criado em o todo sistema, o qual pode representar todas as distribuições implementadas pelo AvaDisPro com uso de uma única variável.

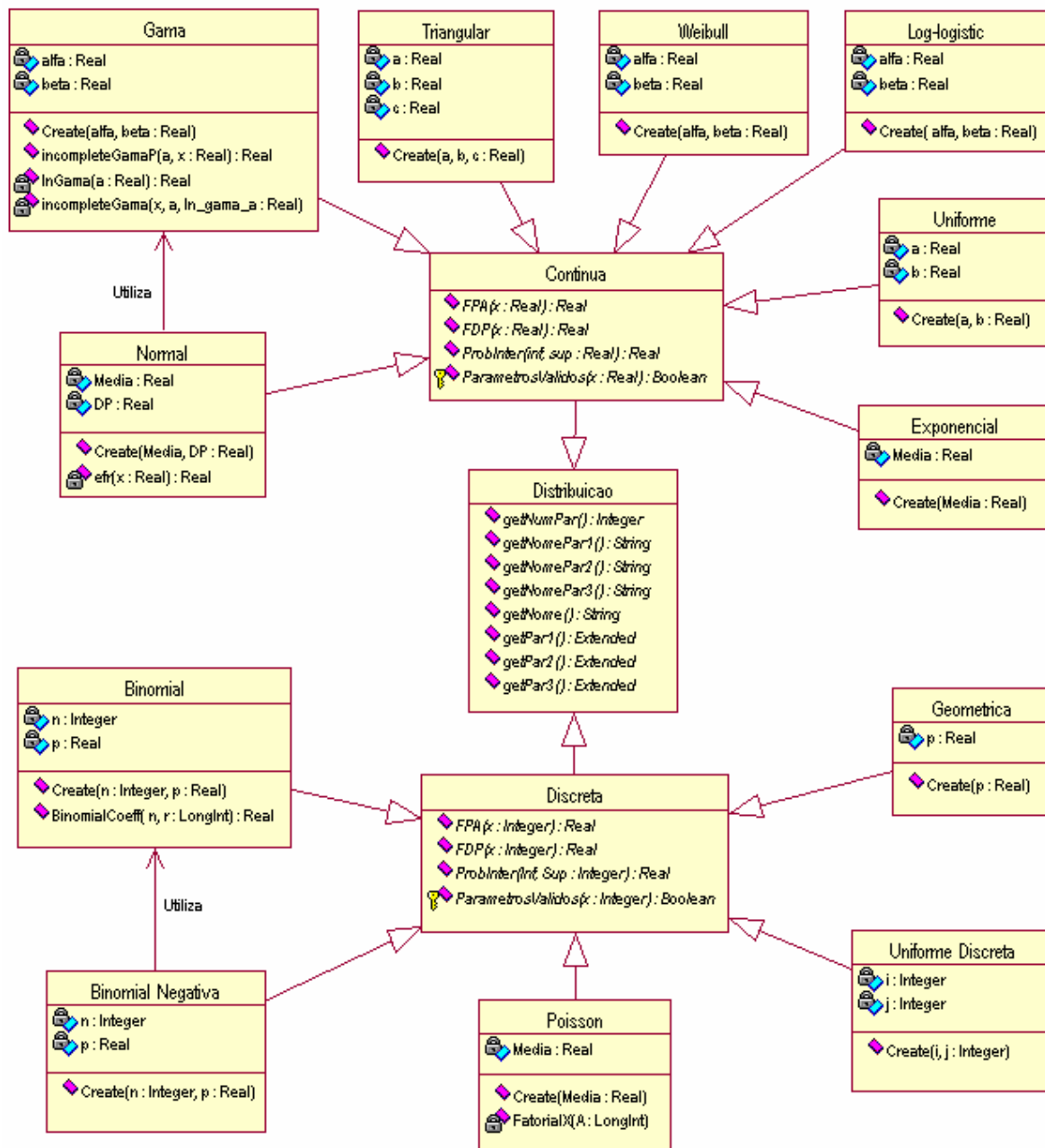


Figura 5.11 – Diagrama de Classes: Objeto Distribuição

### 5.5.3 Diagrama de Classes: Objeto Dados

A figura 5.12 mostra o diagrama de classes para o objeto *Dados*. Também aqui a idéia de extensibilidade está presente, já que cada tipo de dado pode ser estendido para suportar a leitura de arquivos dos mais diversos formatos possíveis. De início, o AvaDisPro lê dados de dois tipos: brutos ou agrupados. O primeiro tipo representa dados sem qualquer formatação, já no segundo os dados são divididos em classes definidas pelo usuário.

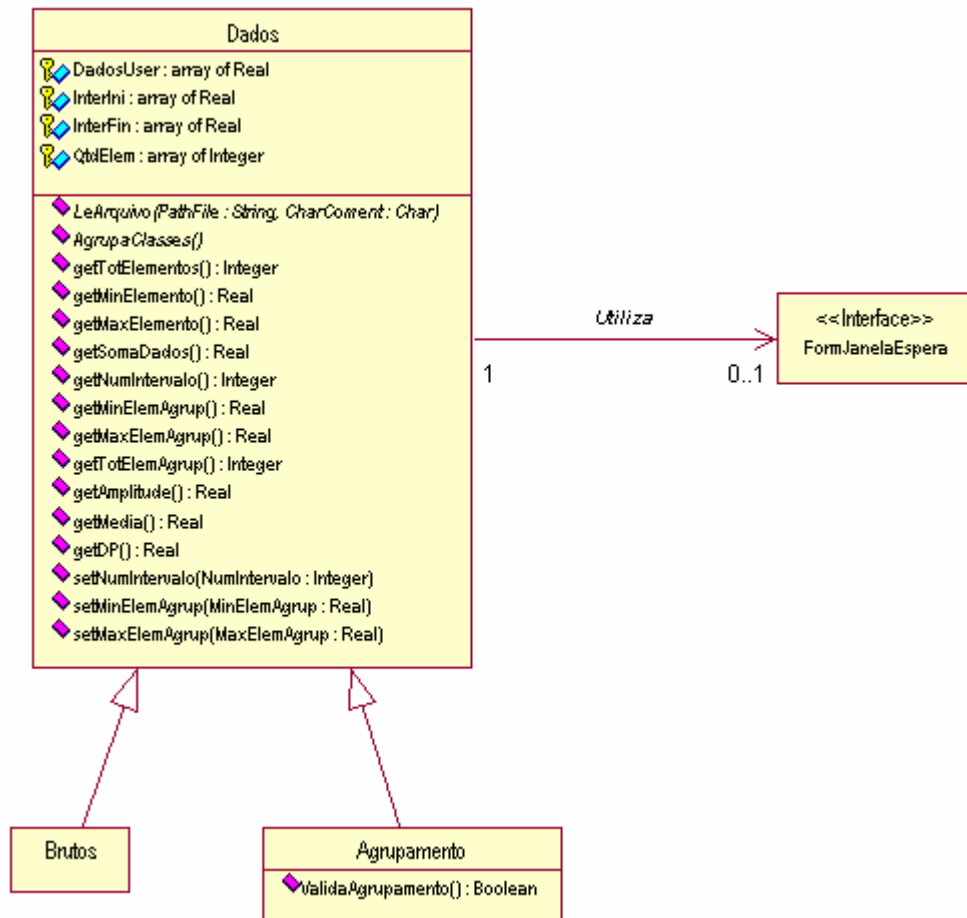


Figura 5.12 – Diagrama de Classes: Objeto Dados

## 5.6 Testes

Os testes realizados no AvaDisPro procuram determinar a correta implementação dos algoritmos utilizados. Para tal, utilizou-se o método de comparação com outras ferramentas para validar o software (também se utilizou a avaliação de usuários para verificar possíveis erros no AvaDisPro: ver capítulo 7).

Tal método, embora não garanta formalmente a correta validação dos algoritmos, dá uma margem de segurança razoável, posto que a comparação com softwares já em uso no mercado

permite a conclusão que, se os resultados apresentados forem os mesmos, então existe uma grande possibilidade da implementação também estar correta, partindo do princípio que tais softwares estão corretamente implementados.

### 5.6.1 Testando o módulo CP-AvaDisPro

Testou-se a função cumulativa de algumas distribuições, visto que essa função é utilizada nos testes  $\chi^2$  e K-S. A tabela 5.2 apresenta os resultados dos testes.

Tabela 5.2 – Resultado do teste da FPA

Distribuição	P (X ≤ x)	AvaDisPro	SPSS	MS-Excel	Statística
<b>Uniforme</b> a = 2 b = 7	3	0,20000000	0,20000000	0,20000000	Não Disponível
	4	0,40000000	0,40000000	0,40000000	
	5	0,60000000	0,60000000	0,60000000	
<b>Exponencial</b> $\mu = 1,5$	1	0,48658288	0,48658288	0,48658288	0,486583
	2	0,73640286	0,73640286	0,73640286	0,736403
	3	0,86466472	0,86466472	0,86466472	0,864665
<b>Normal</b> $\mu = 4,32$ $\sigma = 2,87$	3,5	0,38754848	0,38754848	0,38754854	0,387548
	5,5	0,65951866	0,65951866	0,65951862	0,659519
	7,5	0,86607210	0,86607210	0,86607206	0,866072
<b>Gama</b> $\alpha = 3,5$ $\beta = 5,8$	2	0,00158603	0,00158603	0,00158603	Disponível sem parâmetro $\beta$
	5	0,02647786	0,02647786	0,02647786	
	9	0,12472971	0,12472971	0,12472971	
<b>Weibull</b> $\alpha = 2$ $\beta = 8$	2	0,06058694	0,06058694	0,06058694	0,060587
	5	0,32336615	0,32336615	0,32336615	0,323366
	10	0,79038861	0,79038861	0,79038861	0,790389
<b>Binomial</b> n = 15 p = 0,7	6	0,01524253	0,01524253	0,01524253	Não Disponível
	8	0,13114257	0,13114257	0,13114257	
	10	0,48450894	0,48450894	0,48450894	
<b>Geométrica</b> p = 0,2	1	0,36000000	0,36000000	0,36000000	Não Disponível
	3	0,59040000	0,59040000	0,59040000	
	5	0,73785600	0,73785600	0,73785600	
<b>Poisson</b> $\mu = 4,2$	3	0,39540337	0,39540337	0,39540337	Não Disponível
	5	0,75314289	0,75314289	0,75314289	
	7	0,93605666	0,93605666	0,93605666	

Os softwares escolhidos, dentre os que foram avaliados conforme o capítulo 2, são o Statistica, o Microsoft Excel e o SPSS, os quais são ferramentas comerciais largamente utilizadas.

Tais resultados mostram que, à exceção da distribuição Normal, todos os demais resultados são iguais usando uma precisão de 8 dígitos. Aliás, este teste pôde comprovar os problemas relatados por Berry [BER 1999], que o Microsoft Excel apresenta problemas de acurácia em algumas funções, como o caso da Normal a partir da sétima casa decimal de acordo com a tabela 5.2.

### 5.6.2 Testando o AvaDisPro

Testou-se os resultados produzidos pelo AvaDisPro com outras três ferramentas: Arena, Crystal Ball e XLStat<sup>13</sup>. Escolheu-se esses softwares por serem os únicos (dentro dos testados conforme o capítulo 2) que automatizam os resultados finais, desde a leitura de dados brutos até a interpretação dos resultados dos testes de adequação. A tabela 5.3 mostra o conjunto de dados utilizado para a comparação entre ferramentas.

Tabela 5.3 – Conjunto de Dados Brutos

Conjunto de Dados				
5,92967	6,35459	5,45298	2,31701	5,60527
4,02595	5,30993	5,28506	3,03203	3,57468
4,34990	7,63119	7,32429	4,76752	5,12594
5,85373	3,22086	4,93933	8,21779	4,07427
6,87463	7,40743	3,98845	3,80545	2,16358
2,95762	3,91789	2,60964	5,89833	2,01489
5,12584	2,06911	5,03726	6,45160	3,48472
4,76267	6,03881	7,48211	6,50704	3,60315
3,83762	3,12301	6,06527	3,26955	2,67221
7,15297	5,90531	6,00990	5,78106	5,80670

Escolheu-se as distribuições de probabilidade implementadas pelo AvaDisPro e que tivessem ao menos dois softwares que a implementassem. Por isso, a distribuição log-logistic não foi testada. A tabela 5.4 mostra a estimação de parâmetros realizada por cada ferramenta.

Tabela 5.4 – Parâmetros gerados pelas ferramentas

Distribuição / Ferramenta	Parâmetro	AvaDisPro	Arena	Crystal Ball	XLStat
<b>Exponencial</b>	$\mu$	4,97	2,88	0,20	0,21
<b>Gama</b>	$\alpha$	0,55	1,78	-4,31	Não Disponível
	$\beta$	9,00	1,62	0,29	
<b>Normal</b>	$\mu$	4,97	4,88	4,88	4,86
	$\sigma$	1,66	1,62	1,64	2,73
<b>Triangular</b>	a	2,01	2,00	1,51	Não Disponível
	b	8,22	8,84	9,00	
	c	4,67	5,42	3,84	
<b>Uniforme</b>	a	2,01	2,00	1,89	2,02
	b	8,22	8,84	8,34	8,22
<b>Weibull</b>	$\alpha$	4,88	3,15	1,41	3,37
	$\beta$	0,55	1,58	3,92	5,43

<sup>13</sup> As ferramentas Crystal Ball (<http://www.crystalball.com>) e XLStat (<http://www.xlstat.com>) não foram avaliadas (capítulo 2), porque funcionam somente no MS-Excel (em forma de *plug-in*), não caracterizando – assim – serem softwares autônomos e independentes.

Conforme a tabela 5.4, os parâmetros apresentados por cada ferramenta mostram uma diferença significativa entre seus valores. Por exemplo, a distribuição gama teve valores alfa estimados em 0,55 (AvaDisPro), 1,78 (Arena) e -4,31 (Crystal Ball), o que demonstra que métodos diferentes podem produzir valores diferentes, de acordo com o contexto dos dados<sup>14</sup>.

Tabela 5.5 – Resultado do teste  $\chi^2$

Distribuição / Ferramenta	AvaDisPro	Arena	Crystal Ball	XLStat
<b>Exponencial</b>	49,091	23,2	65,92	39,48
<b>Gama</b>	86,436	12,50	9,76	-
<b>Normal</b>	2,952	4,98	9,40	5,667
<b>Triangular</b>	12,212	5,95	8,68	-
<b>Uniforme</b>	2,600	14,00	13,00	2,122
<b>Weibull</b>	32,000	7,12	7,96	5,991

Tabela 5.6 – Resultado do teste K-S

Distribuição / Ferramenta	AvaDisPro	Arena	Crystal Ball	XLStat
<b>Exponencial</b>	0,281	0,196	0,3325	0,319
<b>Gama</b>	0,368	0,116	0,1014	-
<b>Normal</b>	0,059	0,091	0,0892	0,092
<b>Triangular</b>	0,107	0,216	0,1155	-
<b>Uniforme</b>	0,080	0,206	0,1530	0,143
<b>Weibull</b>	0,800	0,135	0,1004	0,090

As tabelas 5.5 e 5.6 apresentam o resultado do cálculo dos testes  $\chi^2$  e K-S. Todas as ferramentas utilizaram cinco classes, com amplitude definida por cada software. Dessa forma, procurou-se assegurar a correta execução destes testes, já que o teste  $\chi^2$ , por exemplo, apresenta sensibilidade ao número de classes utilizadas [SPI 1972], gerando valores diferentes de acordo com o seu número.

O AvaDisPro obteve valores não tão divergentes nas distribuições exponencial, normal, triangular e uniforme<sup>15</sup>, ficando as distribuições gama e weibull com valores discrepantes se comparadas às outras ferramentas<sup>16</sup>.

<sup>14</sup> As ferramentas Arena e Crystal Ball não informam o método utilizado para estimação de parâmetros.

<sup>15</sup> Este teste não tem por objetivo determinar qual ferramenta é melhor (até porque mais testes seriam necessários, dependendo do contexto analisado), mas sim simplesmente comparar uma saída de dados do AvaDisPro com outros softwares.

<sup>16</sup> Não é possível comparar a implementação destas programas, pois se tratam de softwares proprietários.

A tabela 5.7 mostra o resultado final para ambos testes  $\chi^2$  e K-S. O AvaDisPro não rejeitou duas distribuições (normal e uniforme) com o teste  $\chi^2$  e, além dessas, a triangular com o K-S.

Tabela 5.7 – Resultados dos testes K-S e  $\chi^2$

Distribuição / Ferramenta	Teste de Adequação	AvaDisPro	Arena	Crystal Ball	XLStat
<b>Exponencial</b>	K-S	Rejeição	Rejeição	Rejeição	Rejeição
	$\chi^2$	Rejeição	Rejeição	Rejeição	Rejeição
<b>Gama</b>	K-S	Rejeição	Rejeição	$\alpha = 0,05$	Não disponível
	$\chi^2$	Rejeição	Rejeição	Rejeição	
<b>Normal</b>	K-S	$\alpha = 0,2$	$\alpha = 0,05$	Rejeição	$\alpha = 0,05$
	$\chi^2$	$\alpha = 0,1$	$\alpha = 0,05$	Rejeição	$\alpha = 0,05$
<b>Triangular</b>	K-S	$\alpha = 0,2$	$\alpha = 0,05$	$\alpha = 0,05$	Não disponível
	$\chi^2$	Rejeição	$\alpha = 0,05$	$\alpha = 0,05$	
<b>Uniforme</b>	K-S	$\alpha = 0,2$	Rejeição	Rejeição	$\alpha = 0,05$
	$\chi^2$	$\alpha = 0,25$	Rejeição	Rejeição	$\alpha = 0,05$
<b>Weibull</b>	K-S	Rejeição	$\alpha = 0,05$	$\alpha = 0,05$	$\alpha = 0,05$
	$\chi^2$	Rejeição	$\alpha = 0,05$	$\alpha = 0,05$	$\alpha = 0,05$

Os níveis de significância nos quais a hipótese testada não é rejeitada só são diferentes de 5% no AvaDisPro, pois as outras ferramentas sempre mostram a não rejeição nesse nível. É importante saber em qual nível a distribuição não foi rejeitada, já que – quanto maior o valor de  $\alpha$  – menor é o nível de rejeição [NIS 2002].

## 6 Demonstração do AvaDisPro

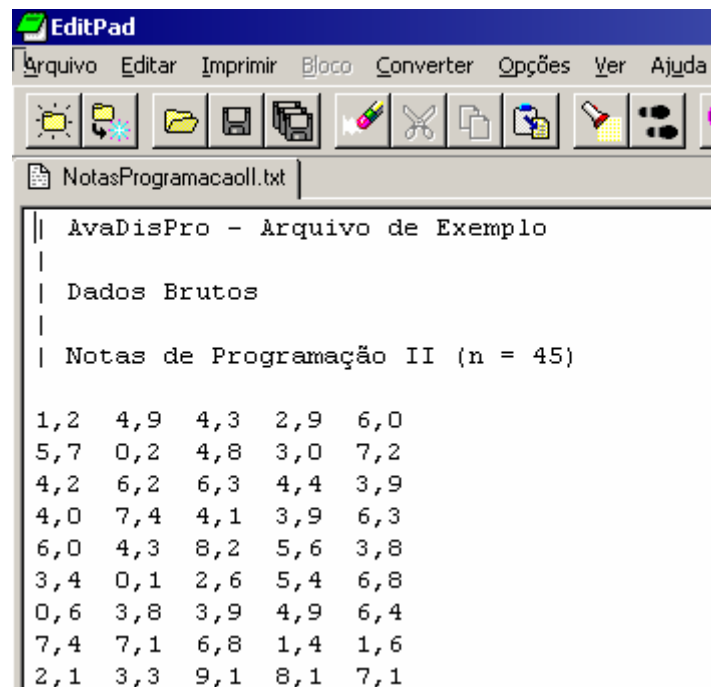
Este capítulo mostra as funcionalidades do AvaDisPro através de um exemplo. Utilizou-se como variável as notas de uma disciplina de graduação, com uma população de 45 elementos. A seguir apresenta-se os passos para verificar se essa variável segue alguma distribuição de probabilidade e, caso siga, qual é esta distribuição e quais são seus parâmetros.

### 6.1 Definindo o Arquivo de Dados

O primeiro passo para utilizar o AvaDisPro é organizar os dados de origem, que podem ser<sup>17</sup>:

- Dados Brutos: são dados sem qualquer tipo de formatação.
- Dados Agrupados: são dados pré-agrupados em classes definidas pelo usuário. A estrutura do arquivo segue o seguinte formato: limite inferior da classe, limite superior da classe e quantidade de elementos do intervalo para cada linha do arquivo. Por exemplo, 15 25 8, significa que 15 é o limite inferior, 25 é o limite superior e 8 é a quantidade de elementos nesse intervalo. Para efeitos de arredondamento, o AvaDisPro considera o intervalo inferior como fechado (neste exemplo, de 15 até 24,99...).

Neste exemplo, utilizou-se dados brutos, conforme mostra a figura 6.1. A disposição dos elementos é flexível, podendo seguir qualquer ordem ou posicionamento.



```
|| AvaDisPro - Arquivo de Exemplo
|
| Dados Brutos
|
| Notas de Programação II (n = 45)
|
1,2  4,9  4,3  2,9  6,0
5,7  0,2  4,8  3,0  7,2
4,2  6,2  6,3  4,4  3,9
4,0  7,4  4,1  3,9  6,3
6,0  4,3  8,2  5,6  3,8
3,4  0,1  2,6  5,4  6,8
0,6  3,8  3,9  4,9  6,4
7,4  7,1  6,8  1,4  1,6
2,1  3,3  9,1  8,1  7,1
```

Figura 6.1 – Arquivo de Dados do Usuário

<sup>17</sup> Devido à extensibilidade do AvaDisPro, é possível especializar classes para ler arquivos de diversos formatos.



## 6.2 Abrindo o Arquivo de Dados

Uma vez definido o arquivo de dados, o usuário está apto a utilizar o AvaDisPro para realizar os testes de adequação. Para tal, deve selecionar o menu *Arquivo*, opção *Abrir Dados Brutos*. A figura 6.2 mostra a tela principal do software após a abertura do arquivo.

The screenshot shows the AvaDisPro 1.0 software interface. The title bar indicates the file path: E:\Delphi6\Projects\AvaDisPro\Dados\NotasProgramacaoII.txt. The menu bar includes Arquivo, Ver, Opções, Distribuições, Testes, and Ajuda.

**Conjunto de Dados**

Ordem	Valor
1	1,20
2	4,90
3	4,30
4	2,90
5	6,00
6	5,70
7	0,20
8	4,80
9	3,00
10	7,20
11	4,20
12	6,20
13	6,30
14	4,40
15	3,90
16	4,00
17	7,40
18	4,10
19	3,90

**Resumo**

Medida	Valor
Elementos	45
Mínimo	0,10
Máximo	9,10
Soma Total	210,70
Média Aritmética	4,68
Desvio-Padrão	2,22

**Agrupamento**

Vl. Inicial	Vl. Final	Qtd.	%
0,10	1,90	6	13,33
1,90	3,70	6	13,33
3,70	5,50	15	33,33
5,50	7,30	13	28,89
7,30	9,10	5	11,11

**Intervalos (Personalizar)**

Valor Inicial: 0,10  
 Valor Final: 9,10  
 Amplitude: 1,80  
 Total: 5

Buttons: Aplicar, Restaurar

**Resumo (Agrupamento)**

Total Elementos: 45  
 Média Aritmética: 4,80  
 Desvio Padrão: 2,14

**Qui-Quadrado**

Id	ei %	ei	oi %	oi

**K-S**

Id	ei	Fo(x)	Sn(x)	D

**Resultado**

Buttons: Detalhes, Limpar

Figura 6.2 – Tela de Abertura de Dados do AvaDisPro

Ao abrir o arquivo, O AvaDisPro automaticamente gera um resumo dos dados lidos, calculando algumas medidas, como o tamanho, o valor mínimo, o valor máximo, a soma total, a média aritmética e o desvio-padrão da amostra.

Além disso, o AvaDisPro automaticamente agrupa os dados em classes, considerando a seguinte fórmula para cálculo da amplitude:

$$amp = \frac{Max - Min}{TotInt}$$

onde *Max* e *Min* são, respectivamente, os elementos de maior e menor valor e *TotInt* é o total de intervalos.

É possível personalizar a forma de agrupamento através das seguintes opções:

- Aumentar o valor do limite inferior ou superior, podendo com isso passar a utilizar um subconjunto de dados.
- Definir o total de intervalos, podendo ajustar um maior ou menor número de intervalos, já que testes como  $\chi^2$  ou K-S são sensíveis à formação de classes [McL 2000].

Também é possível determinar como o AvaDisPro irá tratar os dados, isto é, se como discretos ou se como contínuos.

### 6.3 Verificando Parâmetros

Após abrir o arquivo de dados, o usuário pode verificar os parâmetros iniciais de cada distribuição de probabilidade. Para tal, seleciona-se o menu *Distribuições*, opção *Contínuas* (no caso dos dados serem contínuos) e escolhe-se a distribuição escolhida. A figura 6.3 mostra os parâmetros da seleção da Distribuição Normal.

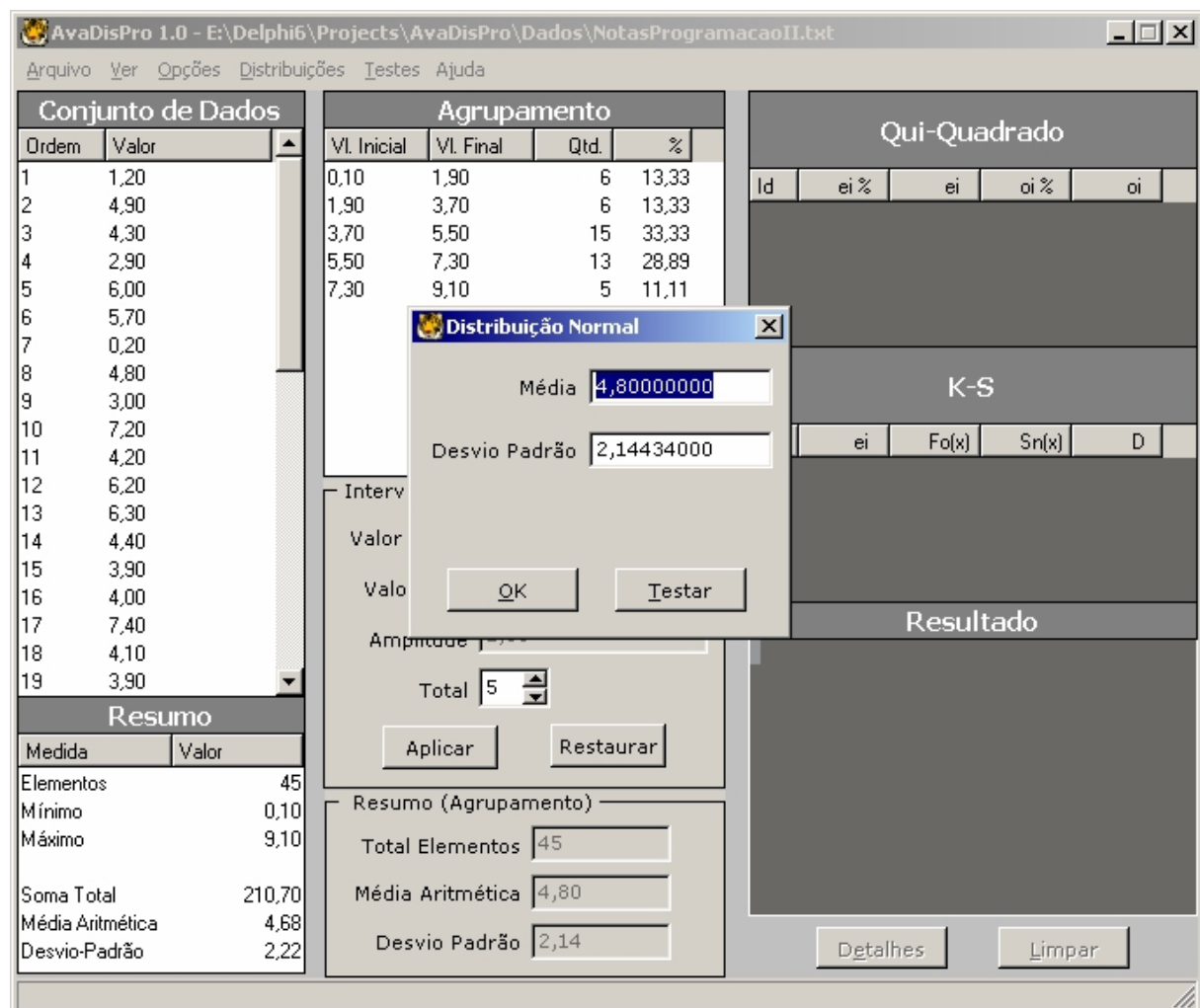


Figura 6.3 – Definindo parâmetros da Distribuição Normal

## 6.4 Testando uma Distribuição de Probabilidade

O AvaDisPro permite testar uma distribuição em particular. Para tal, deve-se seguir os mesmos passos descritos na seção 6.3 e selecionar a opção *Testar*. Automaticamente o AvaDisPro monta as tabelas dos testes  $\chi^2$  e K-S, além de mostrar o valor de cada teste e se a hipótese foi ou não rejeitada, conforme a figura 6.4.

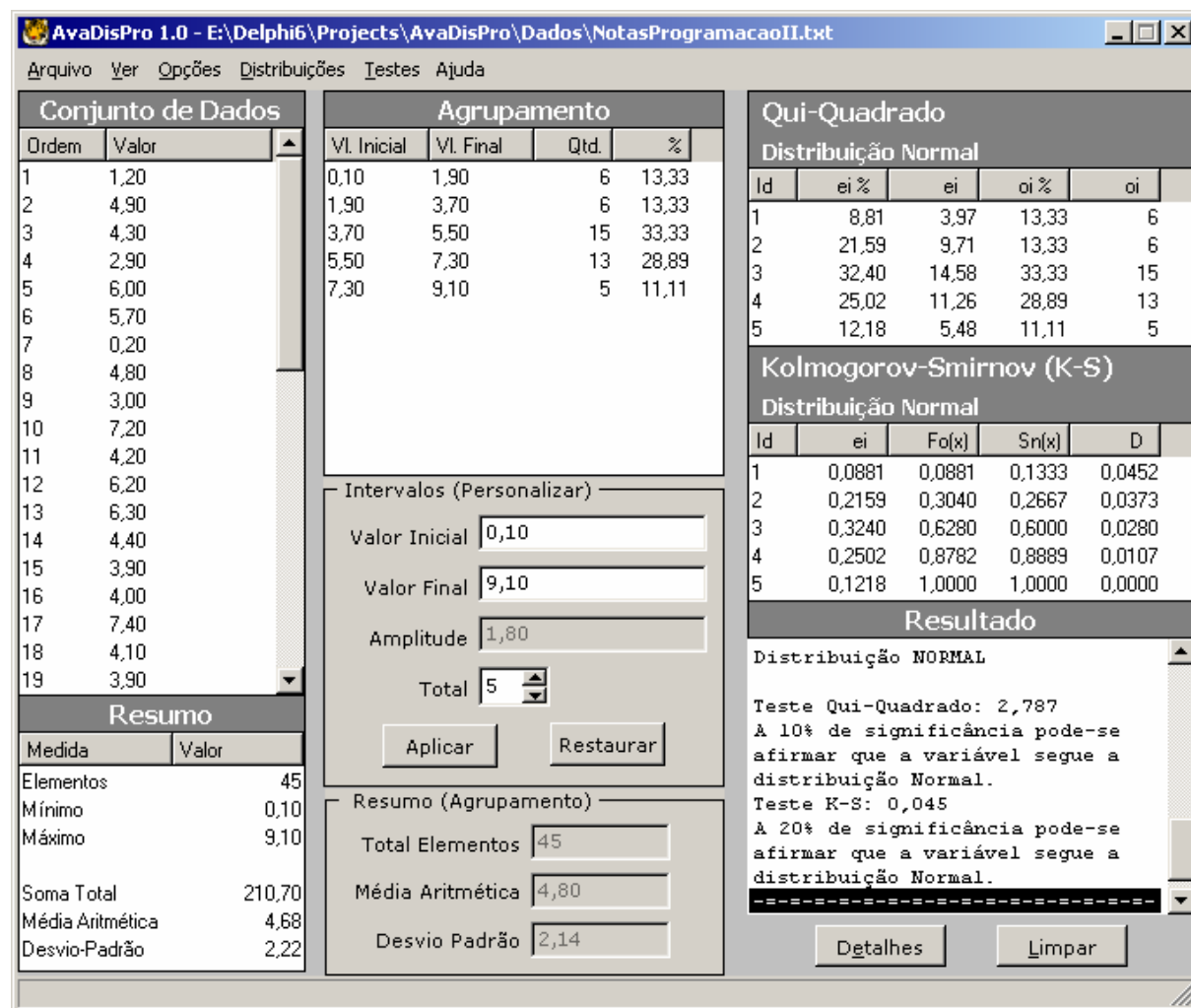


Figura 6.4 – Resultado dos testes  $\chi^2$  e K-S com a Distribuição Normal

Neste ponto, ao utilizar a opção *Detalhes*, o AvaDisPro mostra um relatório contendo as informações dos testes realizados para a distribuição escolhida. Um arquivo é gerado sob o nome *output.txt*, e é gravado no diretório de instalação do AvaDisPro, para posterior consulta. A figura 6.5 mostra esses resultados, que incluem também alguns níveis de significância para cada teste, podendo verificar-se qual foi o nível de rejeição da variável testada.

Desta forma, o AvaDisPro mostra todos os passos e cálculos realizados, e não simplesmente o resultado final do processamento. Assim, procura-se afirmar o caráter didático do software, ao contrário de ferramentas cujo objetivo é puramente comercial (ver capítulo 2).

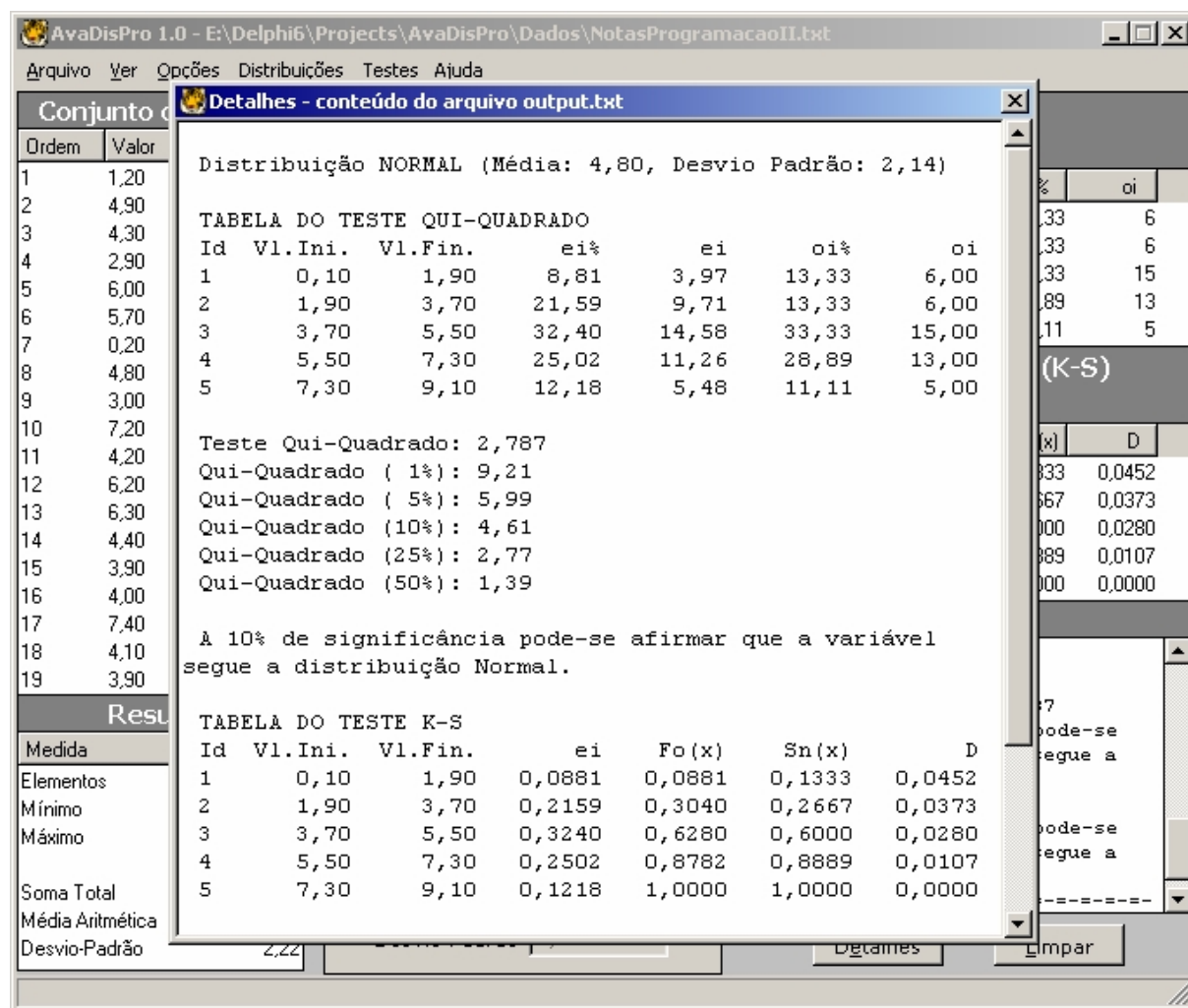


Figura 6.5 – Relatório dos testes  $\chi^2$  e K-S para a Normal

## 6.5 Testando todas Distribuições com $\chi^2$ ou K-S

Uma maneira mais rápida para se verificar se a variável segue uma ou mais distribuições é utilizar algum teste de adequação para todas distribuições contínuas ou discretas. Seleciona-se o menu *Testes*, e escolhe-se a opção *Qui-Quadrado* ou *K-S*. O AvaDisPro automaticamente realiza o cálculo para todas as distribuições. A figura 6.6 mostra o resultado dessa operação.

No exemplo utilizado neste capítulo, duas distribuições não foram rejeitadas: a Normal (10% de significância) e a Triangular (1%). Todas as demais distribuições foram rejeitadas, e o resultado do teste é mostrado, bem como os parâmetros utilizados.

Desta forma, busca-se atingir uma maior facilidade de uso, já que com um esforço pequeno do usuário é possível chegar ao resultado final e verificar as hipóteses testadas.

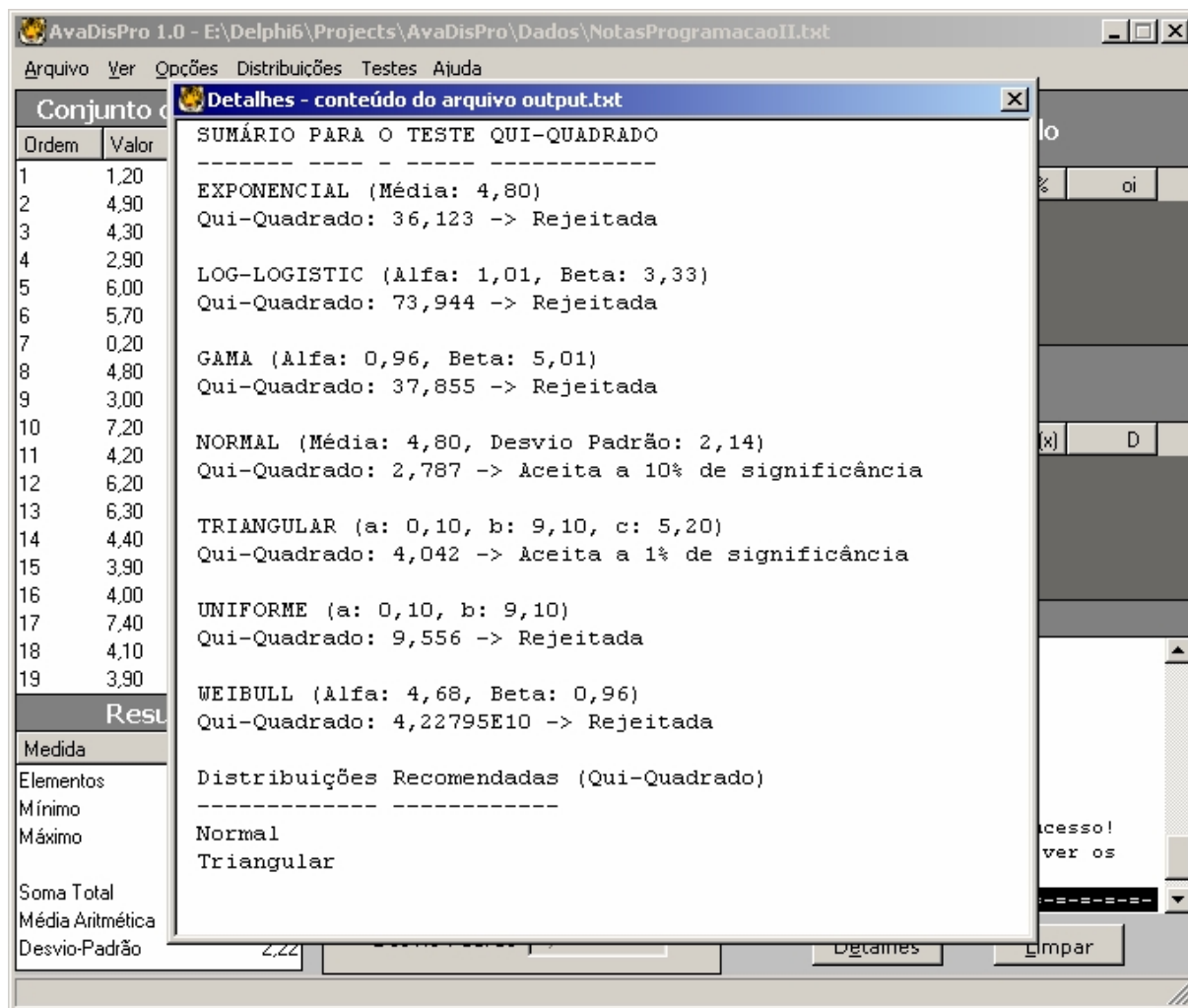


Figura 6.6 – Resultado do teste  $\chi^2$  para todas distribuições contínuas

## 6.6 Executando os Testes $\chi^2$ e K-S Automaticamente

Também é possível executar ambos testes  $\chi^2$  e K-S ao abrir o arquivo de dados. Para tal, seleciona-se o menu *Opções*, opção *Executar Automaticamente* e em seguida o teste a executar (*Qui-Quadrado* ou *K-S*). A figura 6.7 mostra esta opção.

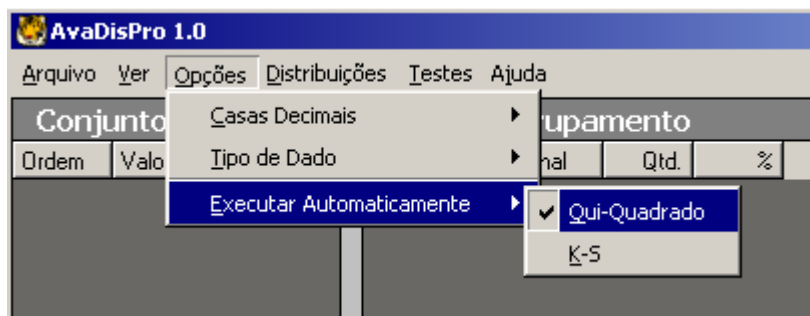


Figura 6.7 – Opção para executar automaticamente os testes  $\chi^2$  e K-S

## 6.7 O Módulo CP-AvaDisPro

O módulo CP-AvaDisPro é um programa que complementa o AvaDisPro, à medida que permite calcular as funções densidade e cumulativa de cada distribuição implementada pelo software. A figura 6.8 mostra como ativar o CP-AvaDisPro através do AvaDisPro.

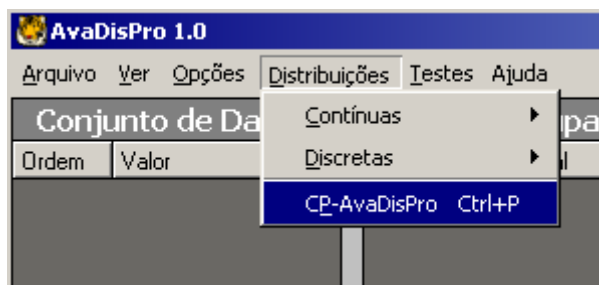


Figura 6.8 – Opção para executar o módulo CP-AvaDisPro

Neste exemplo, utilizou-se como cálculo a probabilidade de  $X$  no ponto 5,5, conforme a figura 6.9. O software calcula a FDP e a FPA, nesse caso para a distribuição Normal. É possível também definir a precisão de dados (de 1 até 14 dígitos), bem como o resultado dessas funções em percentual.

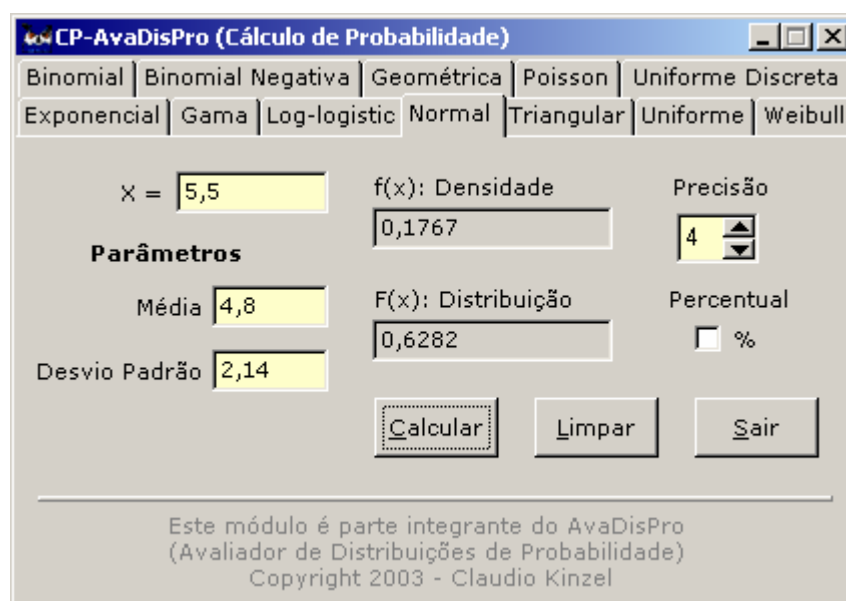


Figura 6.9 – Tela do módulo CP-AvaDisPro

O módulo CP-AvaDisPro é um complemento ao AvaDisPro, uma vez que realiza o processo inverso: calcular probabilidades individuais, ao contrário da execução dos testes de adequação, que calculam todas as probabilidades dos intervalos selecionados. Dessa forma, deve funcionar como mecanismo de apoio ao processo de aprendizagem e teste das próprias distribuições implementadas pelo AvaDisPro.

## 7 Avaliação do AvaDisPro

Este capítulo apresenta o resultado de uma avaliação da ferramenta AvaDisPro, na qual 15 estudantes<sup>18</sup> de 3º grau responderam a um questionário com 12 questões (ver anexo F), e os resultados são descritos a seguir.

### 7.1 Avaliação de Aspectos Estatísticos

As questões de 1 a 6 procuram avaliar os seguintes aspectos da ferramenta AvaDisPro: testes de adequação, quantidade de distribuições, importação de dados brutos, mecanismo de agrupamento e estimação de parâmetros.

A primeira questão visava a avaliar a opinião dos usuários com relação aos testes de adequação presentes no AvaDisPro. Mais precisamente, se eles eram suficientes ou se haveria espaço para outro tipo de teste. Todos os alunos responderam que os testes  $\chi^2$  e K-S são suficientes. Já a segunda questão indagava a respeito do leque de distribuições de probabilidade oferecidas pela ferramenta. Para 14 alunos, ou 93,33%, as distribuições implementadas são suficientes e 6,66%, ou 1 aluno, afirmou que sentiu falta da distribuição Erlang.

A partir da terceira questão (inclusive) as respostas buscam avaliar algum aspecto particular do AvaDisPro, sendo objetivas numa escala de 0 (não apresenta a característica) até 10 (apresenta integralmente a característica).

A terceira questão avalia a acurácia dos testes  $\chi^2$  e K-S. O índice de aprovação entre os avaliados foi de 80,67%, conforme a figura 7.1.

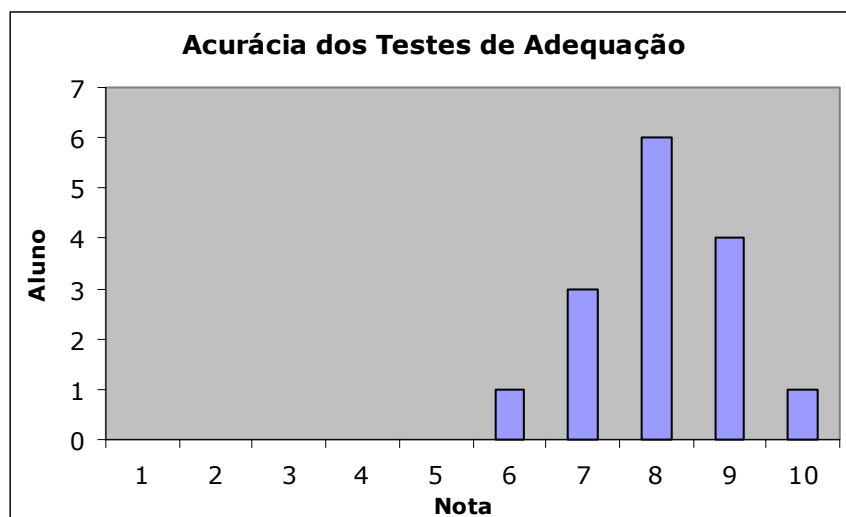


Figura 7.1 – Avaliação de usuários sobre a acurácia dos testes de adequação

<sup>18</sup> Os estudantes em questão são da disciplina Simulação e Modelagem de Sistemas, do curso de Informática da Universidade do Vale do Rio dos Sinos, edição 2003/2.

A quarta questão indagava sobre a facilidade de importação de dados brutos. Para 74% o AvaDisPro apresenta facilidade e flexibilidade nesse item. A quinta questão versava sobre o desempenho do mecanismo de agrupamento de dados, isto é, a organização dos dados em classes. O índice desta questão ficou em 85,33%, conforme a figura 7.2.

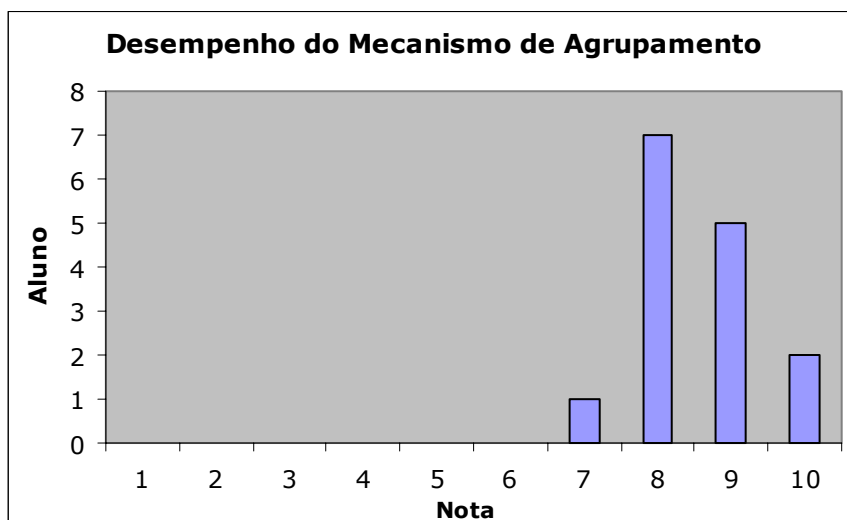


Figura 7.2 – Avaliação de usuários sobre o desempenho do mecanismo de agrupamento

Finalmente, a questão 6 buscava avaliar a qualidade da estimação de parâmetros. Para 80% dos alunos o AvaDisPro estima correta ou apropriadamente os parâmetros das distribuições implementadas pela ferramenta.

## 7.2 Avaliação de Aspectos da Engenharia de Software

As questões de número 7 a 9 avaliam três aspectos de engenharia de software importantes para qualquer projeto [PRE 1995]: documentação, facilidade de uso e estabilidade. A questão 7 avalia a qualidade de documentação do AvaDisPro e 74% foi o percentual de satisfação dos usuários. Já a questão 8 versava sobre a facilidade de uso (uso da interface, documentação etc.) e para 80,67% dos usuários o AvaDisPro pode ser considerado fácil de usar.

A figura 7.3 apresenta um gráfico para a questão 9, que avaliava a estabilidade da ferramenta. Para 86,67% o AvaDisPro está muito estável, o que reflete a importância do uso de uma arquitetura bem definida na construção de sistemas.

Nos três itens avaliados na categoria de aspectos da engenharia de software o AvaDisPro obteve média aritmética de 8,04, que é um resultado interessante em se tratando de uma ferramenta sem pretensões comerciais e recém em sua primeira versão.



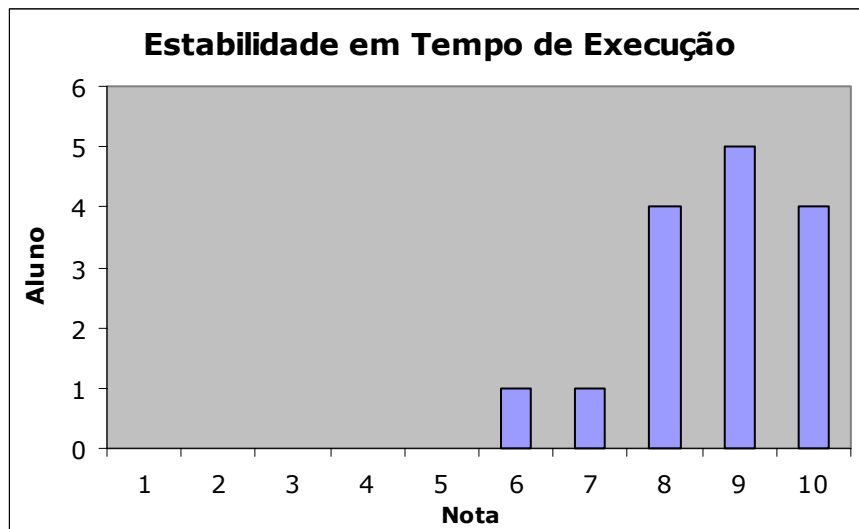


Figura 7.3 – Avaliação de usuários sobre a estabilidade em tempo de execução

### 7.3 Avaliação de Aspectos Gerais

As questões de número 10 e 11 buscam avaliar dois itens em relação ao projeto do AvaDisPro: a sua importância em termos acadêmicos e de inovação e o seu uso como complemento educacional em disciplinas como Simulação de Sistemas e Estatística.

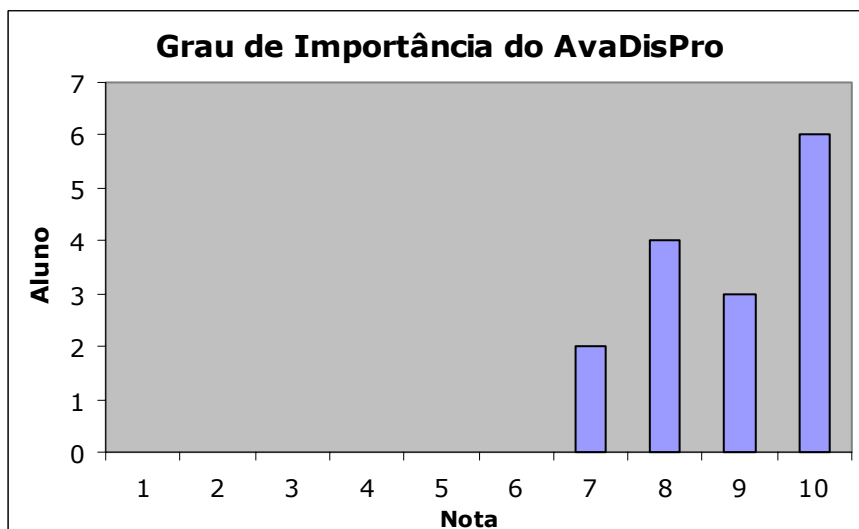


Figura 7.4 – Avaliação de usuários sobre o grau de importância do AvaDisPro

No primeiro caso, 88,67% foi o índice de importância da ferramenta na avaliação de usuários (figura 7.4), e 82% foi o percentual no quesito do uso do AvaDisPro como complemento educacional.

## 7.4 Análise de Resultados

A avaliação realizada com usuários é um instrumento importante para verificar a ferramenta AvaDisPro e fornecer importantes subsídios em termos de melhorias futuras no software. A tabela 7.1 apresenta um resumo desta avaliação, mostrando as médias de cada questão, e também os subtotais por categoria.

Tabela 7.1 – Resumo da avaliação de usuários

Número Questão	Objetivo	Média Aritmética	Média Aritmética %
3	Acurácia dos Testes de Adequação	8,07	80,67
4	Facilidade de Importação de Dados Brutos	7,40	74,00
5	Desempenho do Mecanismo de Agrupamento	8,53	85,33
6	Qualidade da Estimação de Parâmetros	8,00	80,00
<b>Média Categoria Estatística</b>		<b>8,00</b>	<b>80,00</b>
7	Qualidade da Documentação	7,40	74,00
8	Facilidade de Uso	8,07	80,67
9	Estabilidade em Tempo de Execução	8,67	86,67
<b>Média Categoria Engenharia de Software</b>		<b>8,04</b>	<b>80,44</b>
10	Grau de Importância do AvaDisPro	8,87	88,67
11	Uso do AvaDisPro como Complemento Educacional	8,20	82,00
<b>Média Categoria Aspectos Gerais</b>		<b>8,53</b>	<b>85,33</b>
<b>Média Global</b>		<b>8,13</b>	<b>81,33</b>

As questões relativas à categoria estatística obtiveram média 8,0, ao passo que as questões de engenharia de software tiveram média de 8,04, com destaque para a questão 9 (estabilidade em tempo de execução) com 8,67. Já em aspectos gerais a média ficou em 8,53, mostrando que – na opinião dos alunos – o projeto do AvaDisPro é importante e seu uso como complemento educacional é válido.

Além das questões objetivas, a questão 12 perguntava ao usuário se este tinha algum comentário ou sugestão a fazer. Alguns alunos emitiram opiniões a respeito do software e outros sugeriram melhorias diversas. Dessa forma, esta avaliação representou um importante mecanismo de avaliação do AvaDisPro e os resultados obtidos foram considerados satisfatórios.

## 8 Considerações Finais

Apresentou-se neste trabalho o projeto da ferramenta AvaDisPro, software que busca responder as seguintes questões: dado um conjunto de dados qualquer, existe alguma distribuição que melhor o represente<sup>19</sup>? E, caso exista, qual é essa distribuição? E quais são os valores de seus parâmetros?

Para averiguar a carência de softwares que respondam a essas perguntas, e também que tenham características como licença de uso flexível, possibilidade de ampliação futura, entre outras, avaliou-se 8 ferramentas computacionais, sendo que nenhuma delas atingiu os requisitos propostos para o AvaDisPro, caracterizando – dessa forma – a necessidade deste projeto.

Também se mostrou as distribuições de probabilidade implementadas, suas características e funções (cumulativa e densidade), bem como sua aplicabilidade. Os testes de adequação implementados pelo AvaDisPro ( $\chi^2$  e K-S) também foram descritos.

O projeto da ferramenta foi apresentado em detalhes, de modo que possa representar um importante subsídio para o desenvolvimento futuro do AvaDisPro. Os diagramas de caso de uso, seqüência e classe, que constituem o cerne da modelagem UML, foram discutidos e apresentados. O software também foi descrito em forma de tutorial, de modo que suas principais características fossem demonstradas.

A avaliação de potenciais usuários foi relatada, de modo a apurar alguns aspectos importantes do software, como facilidade de uso, importância do projeto, nível de documentação etc. Tal pesquisa representou um importante *feedback* por parte de utilizadores do AvaDisPro.

Desta forma, este trabalho mostrou a ferramenta AvaDisPro, software que procura preencher uma lacuna importante na área da probabilidade, especialmente no meio acadêmico. Dada a sua natureza dinâmica, entretanto, este trabalho não termina aqui. Ele pode e deve ser continuado futuramente, por qualquer pessoa interessada na área, desde que os conceitos básicos da licença de uso do AvaDisPro sejam respeitados.

### 8.1 Trabalhos Futuros

Os pontos a seguir são algumas sugestões de possíveis melhorias na ferramenta AvaDisPro e servem como base para quem desejar estendê-la futuramente:

- Adicionar outras distribuições de probabilidade no AvaDisPro (tais como: Beta, Cauchy, Erlang, Hipergeométrica, Logistic etc.), de modo que um leque maior de opções esteja presente (para uma relação de distribuições, ver tabela 2.5);

---

<sup>19</sup> Considerando o conjunto de distribuições implementado pelo AvaDisPro.

- Adicionar outros formatos para leitura de dados, como por exemplo: ler dados de planilhas como StarOffice ou MS-Excel, ou ainda, sob o contexto de simulação de sistemas, ler dados de tempos entre chegadas;
- Melhorar o gráfico gerado pelo software, incluindo também a curva de cada distribuição testada no histograma gerado;
- Adicionar outros métodos para estimação de parâmetros, podendo oferecer ao usuário a opção de aplicar os testes de adequação de acordo com cada método implementado;
- Traduzir o AvaDisPro para outras línguas, de modo que a sua utilidade seja ampliada. Nesse sentido, também é interessante construir um mecanismo que reconheça a língua utilizada de acordo com a opção desejada, e a tradução possa ser copiada de um arquivo texto, por exemplo;
- Incluir opções de relatórios no AvaDisPro, incluindo informações sobre a amostra testada, os resultados dos testes de adequação e os gráficos gerados.
- Oferecer a opção de salvar o estado do sistema, de acordo com as opções do usuário.

A implementação destas funcionalidades torna o AvaDisPro uma ferramenta mais completa, de modo que a sua utilidade possa ser ampliada.

## Referências Bibliográficas

- [BER 1999] BERRY, Wilson. McCULLOUGH, B.D. On the Accuracy of Statistical Procedures in Microsoft Excel 97. **Elsevier Science B. V.**, Washington, n. 31, p. 27-37, 1999.
- [CAN 2000] CANTÙ, Marco. **Dominando o Delphi 5**. São Paulo: Makron Books, 2000.
- [CAR 2001] CARVALHO, Façal F. **Delphi 6: Programação Orientada a Objetos**. São Paulo: Érica, 2001.
- [DIA 2001] DIAS, Adílson S. **Kylix: a Ferramenta Linux para Programadores**. Rio de Janeiro: Ciência Moderna, 2001.
- [FOW 2000] FOWLER, Martin. SCOTT, Kendall. **UML Essencial: um Breve Guia para a Linguagem-padrão de Modelagem de Objetos**. Porto Alegre: Bookman, 2000.
- [FREa 2003] Free Software Foundation. **GNU General Public License**. On-line. Disponível em: <http://www.gnu.org/copyleft/gpl.html>. Acesso em: 10/05/2003.
- [FREb 2003] Free Software Foundation. **GNU's Not Unix!** On-line. Disponível em: <http://www.gnu.org>. Acesso em: 10/05/2003.
- [FUR 1998] FURLAN, José D. **Modelagem de Objetos através da UML**. São Paulo: Makron Books, 1998.
- [HAY 1996] HAYTER, Anthony J. **Probability And Statistics: For Engineers And Scientists**. Boston: Pws, 1996.
- [IBM 2003] IBM. **Rational Unified Process**. On-line. Disponível em: <http://www-3.ibm.com/software/awdtools/rup>. Acesso em: 19/09/2003.
- [LAW 2000] LAW, Averill M. KELTON, W. David. **Simulation Modeling and Analysis**. Estados Unidos: McGraw-Hill, 2000.
- [LEE 1994] LEEMIS, Lawrence M. **Input Modeling**. Winter Simulation Conference, Estados Unidos, 1994.
- [LEV 1998] LEVINE, David M. BERENSON, Mark L. STEPHAN, David. **Estatística: Teoria e Aplicação usando o Microsoft Excel em Português**. Rio de Janeiro: Livros Técnicos e Científicos, 1998.
- [McL 2000] McLAUGHLIN, Michael P. **Regress+ Apêndice A: A Compendium of Common Probability Distributions ver. 2.3**. User's Guide. 2000.

- [MOH 2000] MOHAMMAD, Al-Fawzan A. **Methods for Estimating the Parameters of the Weibull Distribution**. On-line. Disponível em:  
<http://interstat.stat.vt.edu/InterStat/ARTICLES/2000/articles/O00001.pdf>. Acesso em: 06/09/2003.
- [MUR 1990] MURTEIRA, Bento J. F. **Probabilidade e Estatística**. Vol. II. Portugal: McGraw-Hill, 1990.
- [NIS 2002] NIST/SEMATECH. **e-Handbook of Statistical Methods**, On-line. Disponível em: <http://www.itl.nist.gov/div898/handbook>. Acesso em: 11/08/2003.
- [PRE 1995] PRESSMAN, Roger S. **Engenharia de Software**. São Paulo: Makron Books, 1995.
- [SAB 2000] SABUNCUOGLU, Ihsan. YILMAZ, Anil. Input Data Analysis Using Neural-Networks. **Simulation**, vol. 1, p. 128-137, mar. 2000.
- [SCH 2001] SCHMITZ, Eber. SILVEIRA, Denis. **MRDS: Uma Metodologia de Desenvolvimento de Software em Empresas de Pequeno e Médio Porte**. II Conferência da Associação Portuguesa de Sistemas de Informação, Portugal, Novembro de 2001. Disponível em: <http://equipe.nce.ufrj.br/eber/fast2000/MRDS-Revisado.pdf>. Acesso em: 02/10/2002.
- [SCH 2000] SCHMITZ, Eber. SILVEIRA, Denis. **Uma Metodologia Rápida para o Desenvolvimento Visual de Software Orientado a Objetos**. VIII Congreso Iberoamericano de Educación Superior en Computación - CIESC-CLEI2000, Mexico, Setembro de 2000. Disponível em:  
<http://equipe.nce.ufrj.br/eber/fast2000/UmaMetodologiaRapida.pdf>. Acesso em: 02/10/2002.
- [SON 2001] SONNINO, Bruno. **Kylix: Delphi para Linux**. São Paulo: Makron, 2001.
- [SPI 1972] SPIEGEL, Murray R. **Estatística**. São Paulo: McGraw-Hill do Brasil, 1972.

## Anexo A Especializando uma classe do tipo Distribuição

Neste anexo é apresentado um roteiro de como inserir uma nova distribuição de probabilidade no AvaDisPro. Todo o esforço em termos de modelagem foi realizado de modo que a extensibilidade do software ofereça ao desenvolvedor uma facilidade de execução ao implementar novas funcionalidades no AvaDisPro.

Para inserir uma nova distribuição, dois passos devem ser seguidos:

- Modificação da classe de interface: modifica-se o leiaute do AvaDisPro através de sua interface gráfica (figura A.1).
- Criação da classe de negócio: cria-se uma nova classe de uma distribuição, contínua ou discreta, implementando-se todos métodos da superclasse (que, no caso, é abstrata).

No primeiro passo, os seguintes procedimentos devem ser realizados:

- Inserir um item de menu (componente *TMainMenu*) com o nome da nova distribuição, na categoria apropriada (Contínuas ou Discretas).
- Definir o evento *OnClick* com o método *MIDistroClick*.
- Atualizar o método *CriaObjetoDistro*, para criar o objeto da nova distribuição no momento apropriado.
- Atualizar o método *EstimaParametros*, o qual implementa os parâmetros iniciais de cada distribuição. Para tal, utiliza-se o método *setParam*, da classe *Parametro*, para definir parâmetros iniciais da nova distribuição.

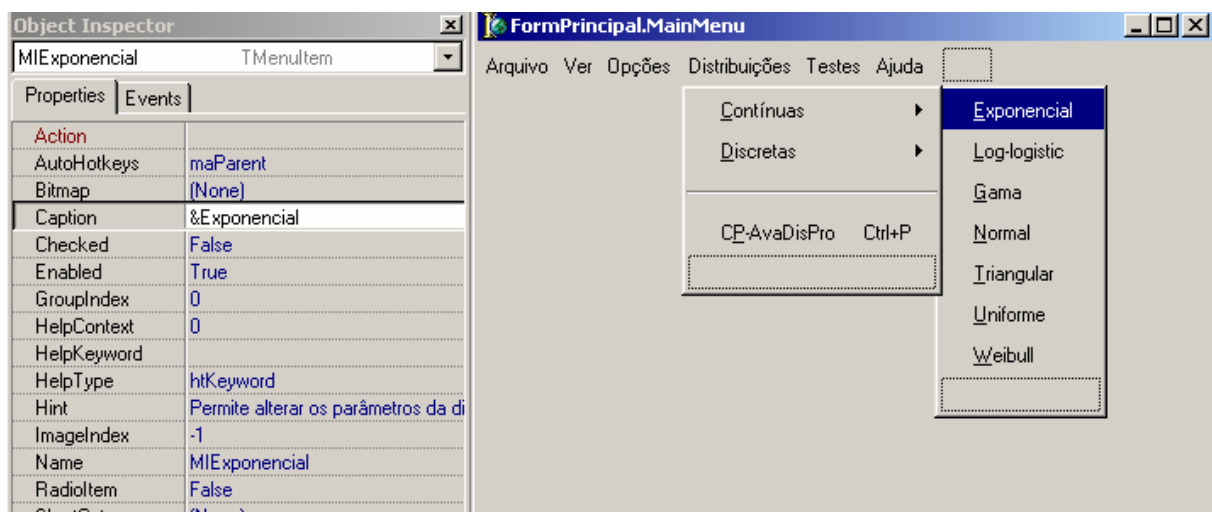


Figura A.1: Inserindo uma nova distribuição no AvaDisPro

Já para a criação de uma classe do tipo distribuição, primeiro mostra-se no quadro A.1 a implementação das classes abstratas *TDistribuicao*, *TContinua* e *TDiscreta*.

```

{ ===== }
{                                     Definição da classe abstrata Distribuicao                                     }
{ ===== }
type
  TDistribuicao = class (TObject)
  public
    { Retorna a quantidade de parâmetros necessários para a distribuição }
    function getNumPar: Byte; virtual; abstract;

    { Retorna o nome dos parâmetros de uma DP }
    function getNomePar1: String; virtual; abstract;
    function getNomePar2: String; virtual; abstract;
    function getNomePar3: String; virtual; abstract;

    { Retorna os parâmetros de uma DP }
    function getPar1: Extended; virtual; abstract;
    function getPar2: Extended; virtual; abstract;
    function getPar3: Extended; virtual; abstract;

    { Retorna o nome da distribuição }
    function getNome: String; virtual; abstract;
  end;
{ ===== }
{                                     Definição da classe abstrata Distribuicao Continua                                     }
{ ===== }
type
  TContinua = class(TDistribuicao)
  public
    { Calcula a FPA (Função de Probabilidade Acumulada) de uma DP Continua }
    function FPA (x: Extended): Extended; virtual; abstract;

    { Calcula a FDP (Função Densidade de Probabilidade) de uma DP Continua }
    function FDP (x: Extended): Extended; virtual; abstract;

    { Calcula a probabilidade de um intervalo; dado um limite inferior e um
      limite superior, calcula a diferença de probabilidades entre eles }
    function ProbInter (_inf, _sup: Extended): Extended; virtual; abstract;
  protected
    { Método utilizado para validar os parâmetros de uma DP Contínua }
    function ParametrosValidos (x: Extended): Boolean; virtual; abstract;
  end;
{ ===== }
{                                     Definição da classe abstrata Distribuicao Discreta                                     }
{ ===== }
type
  TDiscreta = class(TDistribuicao)
  public
    { Calcula a FPA (Função de Probabilidade Acumulada) de uma DP Discreta }
    function FPA (x: Integer): Extended; virtual; abstract;

    { Calcula a FMP (Função Massa de Probabilidade) de uma DP Discreta }
    function FDP (x: Integer): Extended; virtual; abstract;

    { Calcula a probabilidade de um intervalo; dado um limite inferior e um
      limite superior, calcula a diferença de probabilidades entre eles }
    function ProbInter (_inf, _sup: Integer): Extended; virtual; abstract;
  protected
    { Método utilizado para validar os parâmetros de uma DP Discreta }
    function ParametrosValidos (x: Integer): Boolean; virtual; abstract;
  end;

```

Quadro A.1: Definição das classes TDistribuicao, TContinua e TDiscreta



Para criar uma nova distribuição, implementa-se a nova classe de acordo com o seu tipo (contínuo ou discreto). No exemplo a seguir, especializou-se a distribuição contínua exponencial, conforme o quadro A.2:

```
unit Exponencial;

interface

uses
  Math, SysUtils, Distribuicao;

type
  TExponencial = class (TContinua)
  public
    constructor Create (_media: Extended);
    function FPA (x: Extended): Extended; override;
    function FDP (x: Extended): Extended; override;
    function ProbInter (inf, sup: Extended): Extended; override;
    function getNumPar: Byte; override;
    function getNomePar1: String; override;
    function getNomePar2: String; override;
    function getNomePar3: String; override;
    function getPar1: Extended; override;
    function getPar2: Extended; override;
    function getPar3: Extended; override;
    function getNome: String; override;
  protected
    function ParametrosValidos(x: Extended): Boolean; override;
  private
    media: Extended;
  end;

implementation

{ ===== }
{               Implementação da classe TExponencial               }
{ ===== }
constructor TExponencial.Create (_media: Extended);
begin
  media := _media;
end;

function TExponencial.FPA (x: Extended): Extended;
begin
  if ParametrosValidos (x) then
    Result := 1 - Math.Power (e, (-1 / media * x))
  else
    Result := 0;
end;

function TExponencial.FDP (x: Extended): Extended;
begin
  if ParametrosValidos (x) then
    Result := ( 1 / media ) * Math.Power (e, (-x / media))
  else
    Result := 0;
end;

function TExponencial.ProbInter(inf, sup: Extended): Extended;
begin
  Result := FPA(sup) - FPA(inf);
end;
```

```

function TExponencial.getNumPar: Byte;
begin
    Result := 1;
end;

function TExponencial.getNomePar1: String;
begin
    Result := 'Média';
end;

function TExponencial.getNomePar2: String;
begin
    Result := '';
end;

function TExponencial.getNomePar3: String;
begin
    Result := '';
end;

function TExponencial.getPar1: Extended;
begin
    Result := Media;
end;

function TExponencial.getPar2: Extended;
begin
    Result := 0;
end;

function TExponencial.getPar3: Extended;
begin
    Result := 0;
end;

function TExponencial.getNome: String;
begin
    Result := 'Exponencial';
end;

function TExponencial.ParametrosValidos (x: Extended): Boolean;
begin
    if x <= 0 then
        raise Exception.Create('Exponencial: Probabilidade deve ser maior que zero')
    else if media <= 0 then
        raise Exception.Create('Exponencial: Parâmetro média deve ser maior que zero')
    else
        Result := True;
    end;
end.

```

Quadro A.2: implementação da classe TExponencial

Implementando uma nova distribuição com os métodos acima permite ao AvaDisPro reconhecê-la e passar a utilizá-la. Seguindo o esquema proposto neste anexo, pode-se incluir novas distribuições de probabilidade no AvaDisPro de maneira rápida e simples, tornando-o mais completo e útil para um maior número de pessoas.

## Anexo B Utilizando o AvaDisPro no GNU/Linux

Neste anexo é mostrado como instalar e utilizar o AvaDisPro num ambiente GNU/Linux. Para tal, utilizou-se a distribuição Conectiva, versão 9. Além dos binários AvaDisPro e CalculoDistro, é necessário copiar a biblioteca libqtintf-6.5-qt2.3.so para dentro do disco rígido.

Supondo que seja criado um diretório em /home/usuario/avadispro e nele inseridos estes arquivos, ao tentar digitar o comando ./AvaDisPro ocorrerá um erro semelhante a este:

error while loading shared libraries: libqtintf-6.5-qt2.3.so: cannot open shared object file.

Isso ocorre porque é preciso informar o caminho desta biblioteca para o sistema. Assim, é preciso modificar o arquivo de configuração /etc/ld.so.conf e inserir um novo diretório no caminho (*path*), por exemplo, /home/usuario/avadispro. Depois disso, é preciso rodar o comando ldconfig para atualizar as configurações de sistema. Outra solução é copiar o libqtintf-6.5-qt2.3.so para dentro de um diretório que já esteja no *path*.

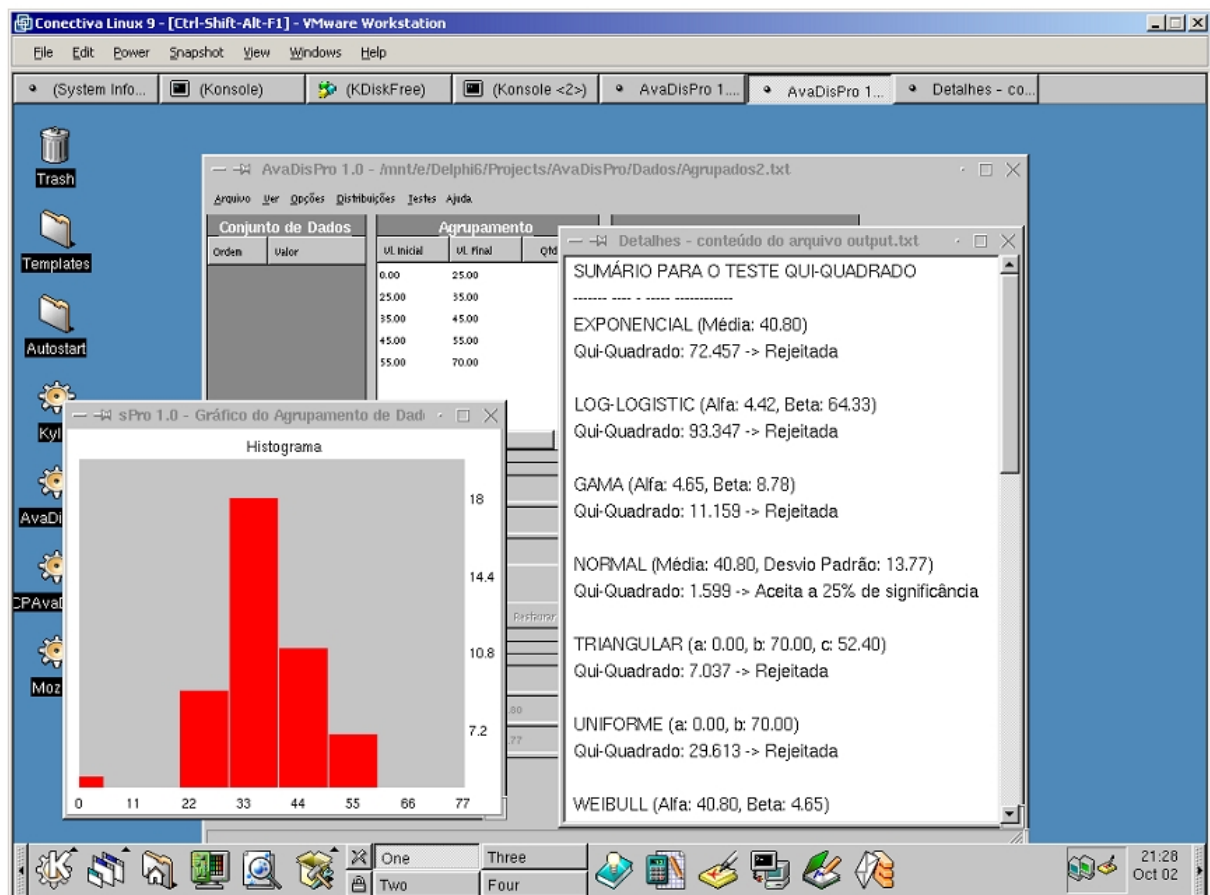


Figura B.1: Execução do AvaDisPro no GNU/Linux

Também é possível que, ao executar o AvaDisPro, ocorra um erro envolvendo outro arquivo que não o supracitado. Neste caso, pode-se utilizar o comando `ldd AvaDisPro` para verificar todos os arquivos necessários para a execução do AvaDisPro. Caso alguns deles não estejam presentes ou desatualizados, é necessário atualizar o sistema.

## Anexo C Programando o AvaDisPro em Delphi e Kylix

Neste anexo são apresentadas algumas diretivas para programar o AvaDisPro nos ambientes Windows e GNU/Linux. Apesar da portabilidade da biblioteca CLX, algumas restrições devem ser levadas em consideração, de acordo com o sistema operacional que se utiliza.

Os passos a seguir mostram alguns cuidados que se deve ter ao programar um sistema portátil usando a CLX:

- a) Componentes Visuais: o AvaDisPro utiliza componentes padrões da CLX (*TLabel*, *TEdit*, *TListView*, *TMainMenu*, *TButton* etc.). Apesar disso, notou-se algumas distorções visuais em alguns componentes, razão pela qual se recomenda verificar todas opções em ambos ambientes para se certificar que o resultado esperado é apropriado. Também algumas propriedades não funcionam corretamente em ambos ambientes, como, por exemplo, o valor inicial de um *TSpinEdit* no GNU/Linux.
- b) Componentes de Terceiros: o AvaDisPro utiliza somente um componente de terceiros, que é o *PlotPanel*, para desenhar gráficos. Tal componente está sob licença *free* e apresenta compatibilidade entre Windows e GNU/Linux. No entanto, é preciso ter certo cuidado ao utilizar outros componentes, de modo que: (i) a compatibilidade seja garantida e (ii) a licença de uso do componente seja adequada à do AvaDisPro.
- c) Diretivas de Pré-Processamento: quando for necessário programar partes específicas do código para determinada plataforma, são necessárias certas diretivas, conforme quadro C.1 a seguir:

```
program AvaDisPro;

{$IFDEF MSWINDOWS}
uses
  QForms,
  FrmPrincipal in 'Source\FrmPrincipal.pas' {FormPrincipal},
  FrmJanelaEspera in 'Source\FrmJanelaEspera.pas' {FormJanelaEspera},
  ...
{$ENDIF}
{$IFDEF Linux}
uses
  QForms,
  FrmPrincipal in 'Source/FrmPrincipal.pas' {FormPrincipal},
  FrmJanelaEspera in 'Source/FrmJanelaEspera.pas' {FormJanelaEspera},
  ...
{$ENDIF}
```

Quadro C.1: usando diretivas de pré-processamento

- d) Fontes: é preciso verificar a compatibilidade de fontes entre ambos ambientes. Se se quiser utilizar alguma fonte em especial, pode-se utilizar as diretivas de pré-processamento. Nesse caso, porém, talvez o esforço seja demasiado e não valha o trabalho braçal de especificar pontos específicos em todo o código-fonte.

## Anexo D Tabela do teste $\chi^2$

Tabela  $\chi^2$ : NIST/SEMATECH, *e-Handbook of Statistical Methods* [NIS 2002]

graus de liberdade	$\alpha = 0.50$	$\alpha = 0.25$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
1	0.45	1.32	2.71	3.84	6.63
2	1.39	2.77	4.61	5.99	9.21
3	2.37	4.11	6.25	7.81	11.3
4	3.36	5.39	7.78	9.49	13.3
5	4.35	6.63	9.24	11.1	15.1
6	5.35	7.84	10.6	12.6	16.8
7	6.35	9.04	12.0	14.1	18.5
8	7.34	10.2	13.4	15.5	20.1
9	8.34	11.4	14.7	16.9	21.7
10	9.34	12.5	16.0	18.3	23.2
11	10.3	13.7	17.3	19.7	24.7
12	11.3	14.8	18.5	21.0	26.2
13	12.3	16.0	19.8	22.4	27.7
14	13.3	17.1	21.1	23.7	29.1
15	14.3	18.2	22.3	25.0	30.6
16	15.3	19.4	23.5	26.3	32.0
17	16.3	20.5	24.8	27.6	33.4
18	17.3	21.6	26.0	28.9	34.8
19	18.3	22.7	27.2	30.1	36.2
20	19.3	23.8	28.4	31.4	37.6
21	20.3	24.9	29.6	32.7	38.9
22	21.3	26.0	30.8	33.9	40.5
23	22.3	27.1	32.0	35.2	41.6
24	23.3	28.2	33.1	36.4	43.0
25	24.3	29.3	34.4	37.7	44.3
26	25.3	30.4	35.6	38.9	45.6
27	26.3	31.5	36.7	40.1	47.0
28	27.3	32.6	37.9	41.3	48.3
29	28.3	33.7	39.1	42.6	49.6
30	29.3	34.8	40.3	43.8	50.9

## Anexo E Tabela do teste K-S

Tabela K-S: Probability And Statistics: For Engineers And Scientists, de Hayter [HAY 1996]

n	$\alpha = 0.2$	$\alpha = 0.05$	$\alpha = 0.01$
1	0.900	0.975	0.995
2	0.684	0.842	0.929
3	0.565	0.708	0.829
4	0.493	0.624	0.734
5	0.447	0.563	0.669
6	0.410	0.519	0.617
7	0.381	0.483	0.576
8	0.358	0.454	0.542
9	0.339	0.430	0.513
10	0.323	0.409	0.489
11	0.308	0.391	0.468
12	0.296	0.375	0.449
13	0.285	0.361	0.432
14	0.275	0.349	0.418
15	0.266	0.338	0.404
16	0.258	0.327	0.392
17	0.250	0.318	0.381
18	0.244	0.309	0.371
19	0.237	0.301	0.361
20	0.232	0.294	0.352
21	0.226	0.287	0.344
22	0.221	0.281	0.337
23	0.216	0.275	0.330
24	0.212	0.269	0.323
25	0.208	0.264	0.317
26	0.204	0.259	0.311
27	0.200	0.254	0.305
28	0.197	0.250	0.300
29	0.193	0.246	0.295
30	0.190	0.242	0.290
Aproximação para $n > 30$	$\frac{1.07}{\sqrt{n}}$	$\frac{1.38}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

## Anexo F Questionário de Avaliação do AvaDisPro

<p>1) Todos os testes de adequação necessários para uma correta validação dos dados estão presentes no AvaDisPro? (SIM   NÃO)</p> <p>(No caso da responder ser NAO, qual seria o teste/método específico?)</p> <p>Resposta: SIM (   )</p> <p>NAO (   ) Justificativa (no caso de 'NAO'):</p>
<p>2) Em relação ao leque de distribuições oferecido, este é suficiente ou é sentida a falta de alguma(s) distribuição(ões) em particular? (caso afirmativo, qual(is)?)</p> <p>RESPOSTA: (   ) SIM, é suficiente</p> <p>(   ) NAO, faltam distribuições (CITAR: ...)</p>
<p>3) Avalie os resultados dos testes de adequação seguindo a escala (0: ruins/errados/imprecisos até 10: ótimos/acurados).</p> <p>RESPOSTA: (0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)</p> <p>(   ) (   ) (   ) (   ) (   ) (   ) (   ) (   ) (   ) (   ) (   )</p>
<p>4) Avalie a facilidade e flexibilidade de importação dos dados brutos pelo software (0:pouca facilidade até 10: muita facilidade).</p> <p>RESPOSTA: (0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)</p> <p>(   ) (   ) (   ) (   ) (   ) (   ) (   ) (   ) (   ) (   ) (   )</p>
<p>5) Avalie o desempenho do mecanismo de agrupamento dos dados feito pelo software (0: agrupamento inadequados até 10: agrupamentos adequados).</p> <p>RESPOSTA: (0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)</p> <p>(   ) (   ) (   ) (   ) (   ) (   ) (   ) (   ) (   ) (   ) (   )</p>
<p>6) Em relação à estimação dos parâmetros de distribuições, você diria que esta é (0: ruim até 10: ótima)</p> <p>RESPOSTA: (0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)</p> <p>(   ) (   ) (   ) (   ) (   ) (   ) (   ) (   ) (   ) (   ) (   )</p>



<p>7) Em termos de documentação, você diria que o software esta (0: mal documentado até 10: possui ótima documentação).</p> <p>RESPOSTA: (0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)</p> <p>( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( )</p>
<p>8) Em termos de facilidade de uso (incluido-se aí o projeto da interface de uso, do mecanismo de ajuda etc.) você diria que o software é (0: difícil de usar até 10: fácil de usar).</p> <p>RESPOSTA: (0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)</p> <p>( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( )</p>
<p>9) Em termos de estabilidade em tempo de execução, você diria que o software está (0: muito instável até 10: muito estável).</p> <p>RESPOSTA: (0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)</p> <p>( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( )</p>
<p>10) Qual o nível de importância que você dá para este trabalho (considerando interesse acadêmico / relevância do assunto / nível de inovação etc.)?</p> <p>(0: nenhuma importância até 10: muito importante)</p> <p>RESPOSTA: (0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)</p> <p>( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( )</p>
<p>11) Você acredita que o AvaDisPro pode ser utilizado como complemento ao ensino de distribuições de probabilidade e testes de adequação, em disciplinas como Estatística ou outras afim?</p> <p>(0: nunca deve ser utilizado até 10: sempre deve ser utilizado)</p> <p>RESPOSTA: (0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)</p> <p>( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( )</p>
<p>12) Quais sugestões você teria em termos de alterações a serem feitas no software e em termos de remoção/inclusão de funcionalidades no mesmo?</p> <p>RESPOSTA:</p>