

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
PUC/SP**

MÁRCIA VIEIRA

**ANÁLISE EXPLORATÓRIA DE DADOS:
Uma abordagem com alunos do Ensino Médio**

MESTRADO EM EDUCAÇÃO MATEMÁTICA

São Paulo

2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
PUC/SP**

MÁRCIA VIEIRA

**ANÁLISE EXPLORATÓRIA DE DADOS:
Uma abordagem com alunos do Ensino Médio**

*Dissertação de mestrado, apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência parcial para obtenção do título de **MESTRE EM EDUCAÇÃO MATEMÁTICA**, sob a orientação da Profa. Dr(a). Cileda Q. S. Coutinho.*

São Paulo

2008

BANCA EXAMINADORA

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta Dissertação por processos de fotocopiadoras ou eletrônicos.

Assinatura: _____ **Local e Data:** _____

AGRADECIMENTOS

Agradeço a todos que contribuíram direta ou indiretamente para a concretização deste trabalho.

À Deus, que me proporcionou saúde e força em todos os momentos.

À Professora Dra. Cileda Q. S. Coutinho, pela sua orientação e compreensão em todos os momentos, e pelas idéias que tanto contribuíram na realização desse trabalho.

Aos membros da Banca Examinadora, Prof. Dr. Saddo Ag Almouloud e Prof. Dra. Érica Valéria Alves, pelas valiosas contribuições e sugestões para essa pesquisa.

Aos professores da PUC-SP, e a todos os colegas com os quais eu tive a oportunidade de conviver durante o curso de mestrado.

À família, namorado e amigos, pela cooperação e compreensão durante todos os momentos.

Aos colegas e aos alunos da E. E. Prof. Oscavo de Paula e Silva, que contribuíram com a parte experimental deste trabalho.

Por fim, à Secretaria de Educação do Estado de São Paulo, pela concessão da bolsa de estudos.

RESUMO

O presente trabalho tem como objetivo estudar as interações entre aluno e um ambiente de estatística dinâmica, que neste trabalho será o software Fathom, segundo a abordagem da Análise Exploratória de Dados. Discutimos quais os conceitos e quais os procedimentos necessários, visando à construção de uma análise crítica de um conjunto de dados, favorecida pelo dinamismo do ambiente computacional, que será uma ferramenta para facilitar a mobilização de diferentes tipos de registros de representações semióticas deste conjunto. Como referenciais teóricos, consideramos os níveis propostos por Curcio (1989, 2001) para analisar a compreensão gráfica mobilizada pelos alunos em situação de resolução de problemas propostos em contexto estatístico, e na teoria dos Registros de Representação Semiótica, de Duval (1994). Buscamos assim estabelecer uma leitura desta teoria, amplamente utilizada em pesquisas na área da Educação Matemática relativamente a conceitos geométricos e algébricos, dessa vez para a representação dos conceitos estatísticos. Buscamos especialmente estudar os tipos de apreensões de uma figura, no caso, os tipos de apreensões de um gráfico ou tabela estatística. Para tanto, elaboramos uma seqüência didática de atividades desenvolvidas com o uso do software, com base nos pressupostos da Engenharia Didática (ARTIGUE, 1988). Antes de iniciar o trabalho com as atividades da seqüência didática, os alunos realizaram um teste diagnóstico preparado por nós, em que pudemos identificar suas principais dificuldades em relação aos conceitos estatísticos. O desenvolvimento da seqüência didática mostrou que as interações com o ambiente informatizado e com os grupos, nas articulações dos diferentes tipos de representação, contribuíram com a compreensão de conceitos como a média aritmética e a mediana, e também com a análise e interpretação de gráficos de colunas e de pontos (Dot-Plot). No entanto, estas variáveis ainda foram insuficientes na compreensão de medidas como os quartis, e do gráfico Box-Plot.

Palavras-chave: Análise Exploratória de Dados; Registros de Representação Semiótica; Estatística; Fathom.

ABSTRACT

The object of the present paper is to study the interactions between the student and the environment of dynamic statistics, which, in this paper, will be the software Fathom, according to the approach of the Exploratory Data Analysis. We have discussed which concepts and procedures are necessary, aiming to the construction of critical analysis of a group of data, favored by the dynamism of the computational environment, which will be a tool to facilitate the mobilization of different types of registers of semiotic representations of this set. As theoretical reference, we have considered the levels proposed by Curcio (1989, 2001) to analyze the graphic comprehension mobilized by students in a situation of problem solving proposed in statistical context, and in the theory of Register of Semiotic Representation, by Duval (1994). We have tried, this way, to establish an understanding of this theory, largely used in researches in the Mathematics Education area relatively to geometric and algebraic concepts, this time to a representation of the statistics concept. We have especially tried to study the kinds of understanding of a picture, in this case, the kinds of understanding of statistical graphics or tables. In order to do so, we have elaborated a didactics sequence of activities developed with the use of the software, with bases on the Didactics Engineering (ARTIGUE, 1988). Before starting working with the activities of didactic sequence, we have realized a diagnostic test with the students, in which we were able to identify their main difficulties concerning statistical concepts. The development of the didactics sequence has shown that the interaction with the computerized environment and with the groups, in the articulations of different kinds of representation, have contributed to the comprehension of concepts as the arithmetic mean and median, and also with the analysis and interpretation of graphics of columns and dots (Dot-Plot). However, these variables are still insufficient to the comprehension of measures as quarts and the Box-Plot graphic.

Key-words: Exploratory data analysis; Register of Semiotic Representation; Statistics; Fathom

LISTA DE FIGURAS

Figura 1: Exemplo de manipulação de gráficos.....	14
Figura 2: Exemplo Leitura dos Dados, de Curcio (1989, p.4).....	21
Figura 3: Exemplo Leitura entre os dados, de Curcio (1989, p.4).....	22
Figura 4: Exemplo Leitura além dos dados, de Curcio (1989, p.4).....	23
Figura 5: Gráfico de setores para a variável série.....	25
Figura 6: Exemplo de “tratamento” em álgebra (Moretti, 2002, p.153).....	26
Figura 7: Mudança da escala do gráfico da variável idade.....	28
Figura 8: Gráfico no plano cartesiano da parábola de equação $y=2x^2$	28
Figura 9: Gráfico de setores para a variável gênero.....	29
Figura 10: Gráfico da função.....	30
Figura 11: Gráfico de colunas para a variável gênero.....	31
Figura 12: Gráfico de setores para a variável gênero.....	32
Figura 13: Triângulo.....	33
Figura 14: Gráfico de colunas da variável série.....	33
Figura 15: Apreensão discursiva de uma figura.....	34
Figura 16: Apreensão Discursiva (Box-Plot).....	34
Figura 17: Construção de um triângulo equilátero.....	35
Figura 18: Apreensão seqüencial de uma figura em estatística.....	36
Figura 19: Modificação do paralelogramo ABCD.....	37
Figura 20: Exemplo de apreensão operatória em estatística.....	37
Figura 21: Estrutura do Pensamento Estatístico (SILVA, 2007).....	44
Figura 22: Tela do Fathom – Inserindo um banco de dados.....	52
Figura 23: Tela do Fathom – Gráficos de variáveis qualitativas.....	53
Figura 24: Tela do Fathom – Gráficos de variáveis quantitativas.....	54
Figura 25: Tela do Fathom – Representação gráfica de medidas-resumo.....	55
Figura 26: Tela do Fathom.....	56
Figura 27: Exemplo de um gráfico de setores da variável gênero de um grupo de estudantes.....	60
Figura 28: Gráfico de barras e gráfico de colunas para a variável gênero.....	60
Figura 29: Exemplo de um gráfico histograma da variável altura de um grupo de estudantes.....	61
Figura 30: Ramo-e-folhas para a variável salários.....	62
Figura 31: Exemplo de um gráfico Dot-Plot para a variável idade de um grupo de estudantes.....	62
Figura 32: Exemplo de um gráfico Box Plot da variável altura de um grupo de estudantes.....	63
Figura 33: Representação dos quartis da variável altura no Histograma e no Box-Plot.....	66
Figura 34: Protocolo grupo 1.....	74
Figura 35: Protocolo grupo 1.....	75
Figura 36: Protocolo grupo 1.....	76
Figura 37: Protocolo grupo 1.....	76
Figura 38: Protocolo grupo 2.....	76
Figura 39: Protocolo grupo 3.....	76
Figura 40: Protocolo grupo 2.....	77
Figura 41: Protocolo grupo 3.....	77
Figura 42: Protocolo grupo 1.....	78
Figura 43: Protocolo grupo 2.....	78
Figura 44: Protocolo grupo 3.....	79
Figura 45: Protocolo grupo 1.....	79
Figura 46: Protocolo grupo 2.....	81
Figura 47: Protocolo grupo 3.....	82
Figura 48: Protocolo grupo 1.....	82
Figura 49: Dot-plot para a variável idade.....	85
Figura 50: Possibilidades de gráficos para a variável idade.....	85
Figura 51: Protocolo grupo 2.....	87
Figura 52: Protocolo grupo 3.....	88
Figura 53: Protocolo grupo 1.....	88
Figura 54: Protocolo grupo 2.....	89
Figura 55: Protocolo grupo 1.....	89
Figura 56: Tabelas de distribuição de freqüência da variável gênero no Fathom.....	94

Figura 57: Tabelas de distribuição de frequência da variável idade, no Fathom.....	95
Figura 58: Média aritmética da variável idade, no Fathom.....	95
Figura 59: Tabela para as variáveis gênero e escolaridade no Fathom.....	96
Figura 60: Protocolo grupo 1.....	97
Figura 61: Protocolo grupo 3.....	98
Figura 62: Protocolo grupo 1 (Fathom).....	99
Figura 63: Protocolo grupo 3.....	100
Figura 64: Tabela construída pelos alunos no Fathom.....	101
Figura 65: Gráfico para a variável gênero no Fathom.....	102
Figura 66: Ribbon Chart – Variável Gênero.....	102
Figura 67: Gráficos para as variáveis escolaridade e cargo no Fathom.....	103
Figura 68: Gráficos Histograma e Dot Plot para a variável idade no Fathom.....	104
Figura 69: Tela do Fathom, relacionando diferentes tipos de registros de representação de um conjunto de dados.....	105
Figura 70: Protocolo grupo 3.....	107
Figura 71: Protocolo grupo 3.....	108
Figura 72: Protocolo grupo 1.....	108
Figura 73: Protocolo grupo 1.....	109
Figura 74: Protocolo grupo 3.....	109
Figura 75: Média aritmética para as variáveis idade e salário no Fathom.....	110
Figura 76: Representação gráfica da média aritmética da variável salário, no Fathom.....	112
Figura 77: Média aritmética e mediana da variável idade no Fathom.....	113
Figura 78: Representação gráfica da média aritmética e da mediana da variável salário, no Fathom.....	114
Figura 79: Representação gráfica da média aritmética, da mediana e dos quartis da variável salário, no Fathom.....	115
Figura 80: Gráfico Box-plot da variável salário, no Fathom.....	116
Figura 81: Visualização do histograma e Box-plot (apreensão seqüencial da figura).....	117
Figura 82: Protocolo grupo 1.....	124
Figura 83: Protocolo grupo 3.....	125
Figura 84: Protocolo grupo 2.....	125
Figura 85: Protocolo grupo 2.....	126
Figura 86: Protocolo grupo 3.....	126
Figura 87: Protocolo grupo 1.....	127
Figura 88: Protocolo grupo 3.....	127
Figura 89: Protocolo grupo 1.....	127
Figura 90: Protocolo grupo 1.....	128
Figura 91: Protocolo grupo 2.....	128
Figura 92: Protocolo grupo 3.....	128
Figura 93: Protocolo grupo 3.....	130
Figura 94: Protocolo grupo 2.....	130
Figura 95: Protocolo grupo 1.....	130
Figura 96: Gráfico da variável idade, construído com lápis e papel.....	136
Figura 97: Cálculo da média e mediana.....	136
Figura 98: Parte do relatório do grupo 1.....	137
Figura 99: Visualização simultânea dos gráficos construídos pelo grupo 1, para o gênero feminino.....	139
Figura 100: Visualização simultânea dos gráficos construídos pelo grupo 1, para o gênero masculino.....	139
Figura 101: Gráfico da variável idade e cálculo da média aritmética (lápis e papel).....	140
Figura 102: Gráfico de colunas para a variável “sexo”, construído pelos alunos.....	142
Figura 103: Tabela de distribuição de frequências para a variável idade, e cálculo da média e mediana.....	143
Figura 104: Gráfico para a variável peso, construído pelos alunos no Fathom.....	144

LISTA DE TABELAS

TABELA 1: Distribuição de 40 estudantes de ensino médio, segundo a série de escolaridade.....	25
TABELA 2: Escolaridade dos Funcionários.....	27
TABELA 3: Distribuição dos funcionários, segundo sua escolaridade.....	27
TABELA 4: Distribuição dos estudantes de ensino médio, por gênero.....	29
TABELA 5: Distribuição dos estudantes de ensino médio, por gênero	31
TABELA 6: Distribuição dos estudantes de ensino médio, por gênero.....	58
TABELA 7: Distribuição dos estudantes, segundo suas alturas.....	64
TABELA 8: Distribuição dos gastos do orçamento da União, por setor.....	81
TABELA 9: Distribuição das idades.....	84

LISTA DE QUADROS

QUADRO 1: Relações hipotéticas entre as Teorias de Duval e Curcio.....	38
--	----

SUMÁRIO

Introdução.....	11
Capítulo I – Problemática.....	13
1.1 Justificativa.....	13
1.2 Objetivos e Questão de Pesquisa.....	16
1.3 Metodologia de Pesquisa.....	17
1.3.1 A Engenharia Didática como metodologia.....	17
Capítulo II – Fundamentação Teórica.....	20
2.1 Os Níveis de Compreensão Gráfica, de Frances R. Curcio.....	20
2.2 Os Registros de Representação Semiótica, de Raymond Duval.....	23
2.3 Relações entre os níveis de compreensão gráfica de Curcio, e os tipos de apreensões, de Duval	37
Capítulo III – Estudos Preliminares.....	40
3.1 A Análise Exploratória de Dados	40
3.2 Pensamento Estatístico.....	42
3.3 Revisão Bibliográfica.....	45
3.4 O Software Fathom.....	51
3.5 Objeto de Estudo.....	56
3.6 Parâmetros Curriculares Nacionais.....	67
Capítulo IV: Seqüência Didática.....	69
4.1 Descrição da Parte Experimental da pesquisa.....	69
4.2 Teste-Diagnóstico: Análises e Resultados.....	70
4.2.1 Análises e resultados da primeira questão.....	71
4.2.2 Análises e resultados da segunda questão.....	80
4.2.3 Análises e resultados da terceira questão.....	83
4.2.4 Análises e resultados da quarta questão.....	83
4.3 Seqüência Didática: Análise a priori e Análise a posteriori	90
4.3.1 Análise a Priori e Análise a Posteriori da Parte A da seqüência didática.....	90
4.3.2 Análise a Priori e Análise a Posteriori da Parte B da seqüência didática.....	131
Considerações Finais.....	147
Referências.....	153
ANEXOS	156
TESTE DIAGNÓSTICO.....	156
SEQUÊNCIA DIDÁTICA.....	159
Telas do Fathom – Parte A das atividades.....	173
Telas do Fathom – Grupo 1.....	173
Telas do Fathom – Grupo 2.....	174
Telas do Fathom – Grupo 3.....	175
Relatórios dos Grupos (Parte B das atividades).....	176
Relatório Grupo 1	176
Relatório Grupo 2	180
Relatório Grupo 3	182

Introdução

Neste trabalho discutimos o uso de um ambiente computacional para a introdução à Análise Exploratória de Dados¹, que tem sua porta de entrada pela Estatística Descritiva, a qual, atualmente, faz parte do currículo a ser desenvolvido tanto no Ensino Fundamental como no Médio.

Nosso objetivo é promover a discussão de conceitos e procedimentos didáticos necessários para uma análise crítica de um conjunto de dados, favorecida pelo dinamismo de um ambiente computacional no momento de utilizar representações múltiplas desse conjunto para melhor percepção e apreensão das informações ali constantes.

Este texto foi organizado em quatro capítulos. No primeiro, apresentamos justificativa e relevância deste estudo, explicitando os objetivos e a questão de pesquisa que pretendemos responder com nossa investigação. Apresentamos também a metodologia de pesquisa que adotamos para o desenvolvimento de tal estudo.

No segundo capítulo, apresentamos os pressupostos teóricos que fundamentam o presente trabalho. Além dos Níveis de compreensão gráfica, de Curcio (1989, 2001), apresentamos também a teoria dos Registros de Representação Semiótica, de Duval (1994, 1998, 2003, 2004), discutindo como aplicar esta teoria (que visa principalmente à geometria e à álgebra) no estudo dos tipos de apreensões de uma representação semiótica de um objeto estatístico.

No terceiro capítulo, apresentamos as leituras realizadas na área de educação estatística, que foram utilizadas como base para o desenvolvimento deste trabalho. As contribuições dos trabalhos dos diversos autores que analisamos não se restringiram somente aos resultados de suas pesquisas, mas nos auxiliaram na apresentação da Análise Exploratória de Dados, do Pensamento Estatístico, do software Fathom, e do objeto de estudo em questão. Apresentamos também a abordagem dos Parâmetros Curriculares Nacionais, referente ao ensino da estatística no Ensino Médio.

¹ Filosofia nos estudos estatísticos, introduzida por Tukey (1977, apud BATANERO, 1991), que será definida no item 3.1 deste trabalho.

No quarto capítulo, apresentamos uma síntese da seqüência didática que foi desenvolvida com os alunos sujeitos desta pesquisa, a análise a priori desta seqüência, os resultados de sua aplicação e a análise a posteriori. Neste mesmo capítulo, consta também o teste diagnóstico que foi aplicado com os sujeitos da pesquisa, no intuito de verificar os conhecimentos prévios destes alunos em relação aos conceitos estatísticos que pretendíamos abordar ou que serviriam de base para a abordagem de novos conceitos, e relatamos os resultados deste teste.

Após o quarto capítulo, apresentamos as considerações finais do trabalho e as possíveis perspectivas para trabalhos futuros nesta temática.

Capítulo I – Problemática

1.1 Justificativa

Atualmente, e a cada dia mais em nosso cotidiano, temos o contato com situações que envolvem uma grande quantidade de informações. Na maioria das vezes, estas são expostas na mídia na forma de gráficos e tabelas. É só ligar a televisão ou abrir uma revista, que nos deparamos com alguma informação expressa nessa forma, como pesquisas eleitorais, andamento do campeonato de futebol, aumento ou queda de itens de consumo ou de taxa de juros. Na escola, os alunos também têm acesso a esse tipo de informação diariamente através de livros didáticos, para-didáticos, jornais, revistas. No trabalho, o cidadão tem contato com gráficos de produção, despesas, lucros, etc.

Este avanço crescente na exposição de informações em gráficos e tabelas deve-se, em grande parte, à popularização dos computadores, que tornam mais rápido e fácil sua construção, além de permitirem uma infinidade de opções na sua escolha.

Porém nem sempre essa facilidade no manuseio de um software garante que as informações serão expressas na forma correta, pois, por desconhecimento de alguns conceitos básicos de estatística, uma pessoa pode construir um gráfico no computador que não as expresse da forma mais adequada. Além disso, existem ainda casos em que se manipulam os dados para “maquiar” algum tipo de informação num gráfico, ou seja, a representação gráfica pode ser vulnerável a manipulações tendenciosas, que podem mostrar informações imprecisas. Por exemplo, a distorção de um gráfico pode ser feita pelo simples “alongamento” de um de seus eixos, tirando a proporcionalidade ideal para a interpretação dos dados ali representados. Mostramos um exemplo ilustrado por Wainer (1992, apud CAZORLA 2002), de um gráfico (figura 1a, CAZORLA 2002, p.4) publicado em uma revista, cuja idéia é mostrar que, enquanto o gasto com educação feito pelo governo americano aumentou rapidamente na última década, o desempenho dos estudantes (medido com as médias das notas da avaliação SAT, Scholastic Aptitude Tests), não correspondeu a este investimento. Porém, manipulando-se

separadamente as escalas verticais (figura 1b, CAZORLA, 2002, p.4), o gráfico pode mostrar o oposto disto.

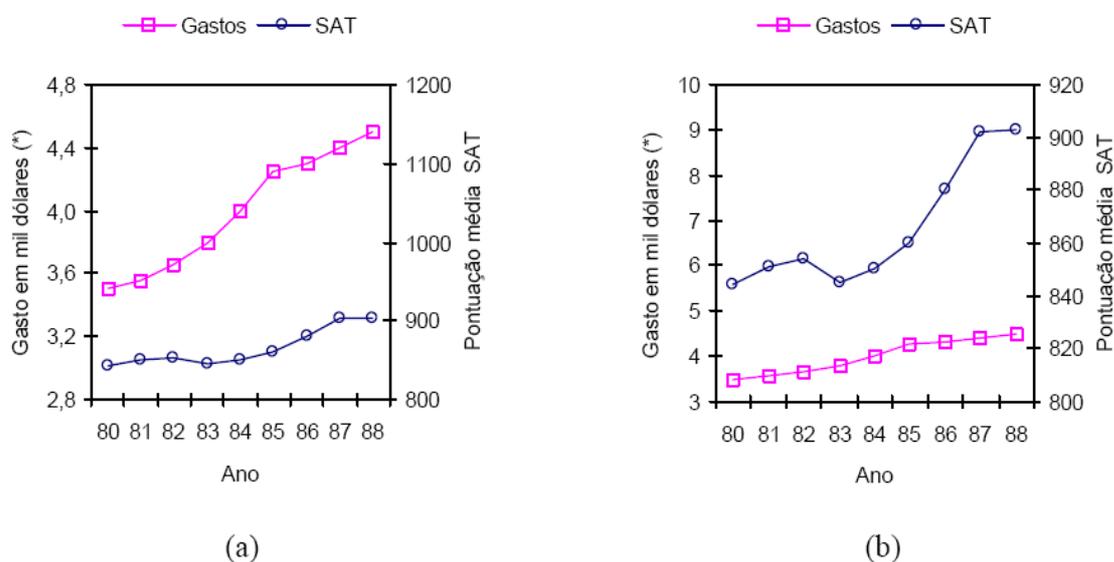


Figura 1: Exemplo de manipulação de gráficos

Alguns autores, como Cazorla (2002) e Godino (1995), ressaltam a importância que a estatística tem alcançado atualmente, tanto na cultura básica quanto na formação profissional do cidadão. Isso se deve à abundância de informação que ele se confronta em seu trabalho diário, e, para estes autores, a alfabetização em estatística torna-se cada vez mais necessária, a fim de evitar o uso indiscriminado das técnicas estatísticas, que podem levar a conclusões erradas.

Batanero (2002) ressalta que “para uma cidadania plena, o pensamento estatístico² é tão necessário quanto a capacidade de ler e escrever”.

Para Godino (1995), este rápido desenvolvimento da estatística e sua difusão deve-se principalmente à influência dos computadores, pelo modo em que os dados podem ser coletados e processados. Este mesmo autor ainda considera que a capacitação estatística hoje deve incluir o conhecimento de formas de processar dados em programas estatísticos, sendo que o computador deve ser visto não só um recurso de cálculo, mas também como um importante recurso didático.

² Estratégias mentais utilizadas pelo indivíduo para tomar decisão em um ciclo investigativo (Wild, e Pfannkuch, 1999).

Alguns dos objetivos didáticos no uso de computadores no ensino da estatística apontados por Godino (1995, p.2) são:

- Capacitar o aluno a: coletar, organizar, depurar, reunir, representar e analisar sistemas de dados de complexidade acessível;
- Nos sistemas de dados, ter ao menos três componentes: a descrição das variáveis, os valores das variáveis e os resumos estatísticos de cada variável, e sobre cada um destes componentes realizar operações e transformações internas e externas, podendo estabelecer relações entre estes componentes (dependência, implicação, similaridade);
- Um ponto de partida da estatística deveria ser o trabalho com sistemas de dados reais (resultados desportivos de suas equipes favoritas, preços de guloseimas da escola, meios de transporte usados para ir à escola), de modo que poderiam ver que construir um sistema de dados próprio e analisá-lo não é o mesmo que resolver um problema de cálculo rotineiro de um livro didático;
- Este sistema de dados poderia ser analisado com a ajuda do computador. O manuseio de um programa computacional é um objetivo importante na atualidade. Um problema tradicional em estatística é a defasagem na compreensão dos conceitos e na aplicação dos cálculos (a existência de programas computacionais de fácil manuseio pode ajudar nestes cálculos);
- Conscientizar o aluno de que esta maior facilidade no emprego de procedimentos estatísticos implica o perigo do uso inadequado da estatística;
- Conscientizar os alunos de que um mesmo problema estatístico pode ser resolvido por diferentes procedimentos, e nem todos os procedimentos adaptam-se bem a todos os problemas.

Godino (1995) ainda ressalta que o uso do computador por si só não resolve o problema do ensino, mas pode proporcionar ao aluno um ambiente que lhe permita generalizar, observar, refletir e interagir sobre fenômenos que dificilmente poderiam se explorar sem uma ferramenta informática, utilizando um núcleo de conceitos e procedimentos matemáticos inter-relacionados, mediante sistemas dinâmicos e de múltiplas representações.

Dentro deste enfoque, nesta pesquisa trabalhamos a Análise Exploratória de Dados com alunos de Ensino Médio, utilizando o software de estatística dinâmica Fathom, com o objetivo de desenvolver a capacidade de analisar e interpretar as informações expressas em gráficos e tabelas.

Optamos por trabalhar com alunos de Ensino Médio por ser a fase final da escolaridade básica, preparo do jovem para o ingresso no mercado de trabalho e na universidade, e também pela contribuição à pesquisa científica, já que são poucos os estudos na área de educação estatística neste nível de escolaridade, como ressalta Batanero (1994, p.2), *“muitos estudos se concentram em crianças pequenas ou em estudantes de universidade, sendo escassas as investigações com estudantes de 11 a 16 anos”*. Apesar de a autora ter feito essa afirmação em 1994, hoje esse cenário ainda não é muito diferente, conforme se pode perceber na revisão bibliográfica.

1.2 Objetivos e Questão de Pesquisa

O objetivo de nossa pesquisa é, a partir do desenvolvimento de uma seqüência didática, estudar as interações entre o aluno e um ambiente de estatística dinâmica, introduzindo a análise exploratória de dados para o nível do segundo ano do Ensino Médio, permitindo que o sujeito desenvolva a capacidade de organizar, analisar e interpretar os dados. Esta seqüência didática será desenvolvida com uso de um ambiente de estatística dinâmica, que é o software Fathom.

O ambiente computacional será uma ferramenta para facilitar a mobilização simultânea de diferentes tipos de registros de representações semióticas de um conjunto de dados, como as tabelas, os gráficos e as medidas-resumo, permitindo assim a obtenção do maior número de informações.

Neste contexto, levantamos as seguintes questões de pesquisa, que pretendemos responder durante o desenvolvimento deste trabalho:

- *Em que medida as articulações entre os diferentes tipos de registros de representações semióticas podem constituir o acesso à compreensão em estatística?*
- *O uso de um software de estatística dinâmica pode gerar/potencializar a necessidade de diversos registros de representações semióticas, e permitir que o sujeito desenvolva a capacidade de organizar, analisar e interpretar dados?*

- *Quais tipos de articulações entre os registros de representações semióticas de um conjunto de dados podem ser identificados, a partir do uso do ambiente computacional, visando à análise crítica dos dados?*

1.3 Metodologia de Pesquisa

Neste trabalho, adotaremos como metodologia de pesquisa os pressupostos da engenharia didática, pois é o método que consideramos mais adequado para coletar os dados que permitirão responder nossa questão de pesquisa.

1.3.1 A Engenharia Didática como metodologia

A noção de engenharia didática emergiu na Educação Matemática no início da década de 1980, com o objetivo de estilizar uma forma do trabalho didático. Artigue (1988, p.193) faz uma analogia entre o trabalho do pesquisador e do engenheiro:

...aquela que era comparável ao trabalho do engenheiro que, para realizar um projeto preciso, se apóia nos conhecimentos científicos do seu domínio, aceita submeter-se a um controle do tipo científico, mas, ao mesmo tempo, se encontra obrigado a trabalhar sobre objetos muito mais complexos do que os objetos depurados da ciência, e, portanto, a estudar de uma forma prática, com todos os meios ao seu alcance, problemas de que a ciência não quer ou ainda não é capaz de se encarregar...

A engenharia didática é inserida no quadro teórico da Educação Matemática, com a finalidade de fornecer ferramentas para analisar fenômenos didáticos observados no desenvolvimento de situações didáticas. Artigue (1988, p.196) relata algumas das características gerais dessa metodologia:

...A engenharia didática, vista como metodologia de investigação, caracteriza-se primeiramente por um esquema experimental baseado em realizações didáticas na sala de aula, isto é, na

concepção, na realização, na observação e na análise de seqüências de ensino...

O processo experimental da engenharia didática é composto por quatro diferentes fases, que apresentamos a seguir, já explicitando como serão desenvolvidas em nosso trabalho:

1. **Análises prévias:** Nesta fase, são realizadas as análises preliminares de considerações sobre o quadro teórico didático geral e sobre a análise da epistemologia dos conteúdos abordados, do ensino atual e seus efeitos, das dificuldades e obstáculos, que contribuem para evolução dos alunos. Aqui apresentaremos os principais resultados de pesquisa na área de educação estatística e a abordagem da análise exploratória de dados, definiremos o objeto de estudo a ser pesquisado, e falaremos também sobre o que tratam os Parâmetros Curriculares Nacionais a respeito do ensino da estatística. Destacamos que não faremos uma análise epistemológica do tema, mas apenas a análise didática. Identifica-se assim, já neste primeiro momento, uma perspectiva de pesquisa: uma análise epistemológica dos conceitos de base da Estatística Descritiva, assim como uma análise da evolução do ensino desses conceitos, identificando dessa forma as várias fases na evolução da transposição didática.
2. **Análise a priori:** Na fase da concepção e análise a priori das situações didáticas, o pesquisador, com base nas análises prévias, escolhe certo número de variáveis do sistema, que supõe serem variáveis pertinentes ao problema estudado, sendo retomadas e aprofundadas durante o desenvolvimento da pesquisa. É a escolha das variáveis que vai dar forma às situações didáticas construídas para o desenvolvimento da fase experimental. Nesta fase, faremos a análise a priori da seqüência didática elaborada, pautados em nossa fundamentação teórica e nas análises preliminares.

3. **Experimentação:** Ocorre na realização da engenharia com uma determinada população de alunos, tendo início no momento em que se dá o contato pesquisador – professor – observador com os alunos sujeitos da investigação. Nessa fase, devem-se respeitar ao máximo as escolhas feitas nas análises a priori, para evitar o fracasso da engenharia; no entanto, passada cada sessão, tudo pode ser revisto e corrigido a partir de uma breve análise a posteriori. Cada nova mudança deve passar novamente por uma análise teórica ou a priori. Nesta etapa, aplicaremos a seqüência didática com alunos do segundo ano do Ensino Médio, de uma escola estadual de Santo André (explicitaremos melhor as características dos sujeitos no item 4.1 deste trabalho).

4. **Análise a posteriori e validação:** A análise a posteriori se apóia no conjunto dos dados colhidos durante a experimentação e nas produções dos alunos. Nesse momento ocorre o tratamento e a coleta dos dados coletados na fase de experimentação. É no confronto das duas análises, a priori e a posteriori, que se apóia essencialmente a validação das hipóteses envolvidas na investigação (validação interna, que é a característica básica de uma engenharia didática). É nesta fase que os fatos observados são transformados em fenômenos didáticos pela sua teorização. Aqui, analisaremos os dados coletados durante a aplicação da seqüência didática com os alunos, confrontando estes dados com a análise a priori, sempre pautados em nossa fundamentação teórica.

Capítulo II – Fundamentação Teórica

Como referenciais teóricos que ajudarão a modelar os fenômenos didáticos identificados, utilizaremos os Registros de Representação Semiótica (DUVAL, 1994, 1998, 2003, 2004) para analisar a articulação de seu uso na representação e manipulação dos objetos estatísticos construídos pelo Fathom. Embora Duval tenha proposto sua Teoria visando principalmente à geometria e à álgebra, buscaremos, em nossa pesquisa, estabelecer um paralelo para a representação dos conceitos estatísticos.

Utilizaremos também os Níveis de Compreensão Gráfica, de Frances R. Curcio (1989, 2001), para analisar a compreensão gráfica mobilizada por alunos em situação de resolução de problemas propostos em contexto estatístico.

2.1 Os Níveis de Compreensão Gráfica, de Frances R. Curcio

Os resultados de pesquisa de Curcio (1989, 2001) indicam que os estudantes de Ensino Fundamental e Médio deveriam ser envolvidos em coleta de dados reais ligados a contextos de seu cotidiano para construir seus próprios gráficos, o que poderia estimulá-los a levantar hipóteses sobre os dados e verbalizar as relações e padrões observados.

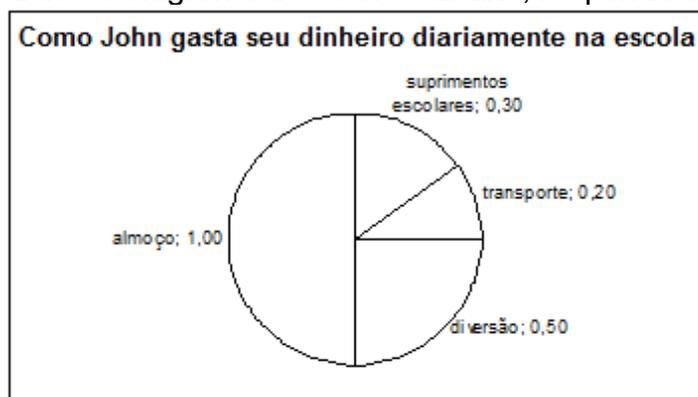
Para Curcio (1989, p.1), os gráficos são meios de comunicar e classificar dados, permitindo a comparação e exibindo relações. Embora a leitura literal dos dados apresentados seja uma habilidade importante na compreensão desse tipo de representação, o seu potencial máximo será atingido quando o leitor for capaz de interpretar e generalizar os dados.

A pesquisa desse autor sugere que existem três níveis distintos para a compreensão gráfica. No que diz respeito à forma utilizada, os três níveis são: leitura dos dados; leitura entre os dados; leitura além dos dados.

- **Leitura dos dados:** este nível de compreensão requer uma leitura literal do gráfico. O leitor simplesmente “levanta” os fatos explicitamente

apresentados ou a informação encontrada nas etiquetas dos eixos, diretamente do gráfico. Neste nível, não existe interpretação, e este tipo de compreensão é uma tarefa que requer poucos recursos cognitivos.

Exemplo: Dado o diagrama de setores abaixo, responda o que se pede.



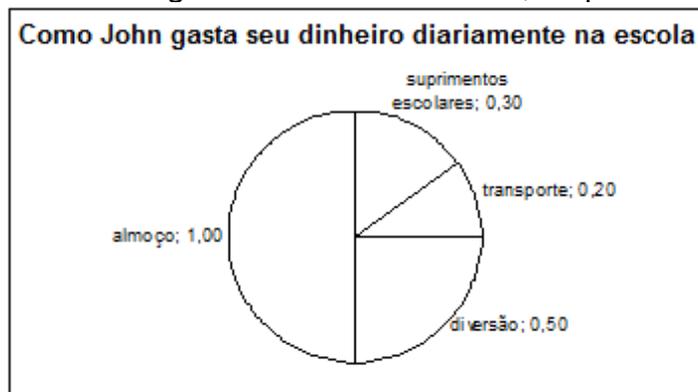
- 1) O que o gráfico conta para você?
 - a) Como John gasta seu dinheiro nos finais de semana
 - b) Como John gasta seu dinheiro durante uma semana completa
 - c) Como John gasta seu dinheiro durante as férias
 - d) Como John gasta seu dinheiro durante um dia na escola.

- 2) Quanto John gasta em suprimentos escolares em um dia?
 - a) R\$0,20
 - b) R\$0,30
 - c) R\$0,50
 - d) R\$1,00

Figura 2: Exemplo Leitura dos Dados, de Curcio (1989, p.4)

- **Leitura entre os dados:** este nível de compreensão exige a interpretação e integração do dado no gráfico. Requer uma habilidade de comparar quantidades (por exemplo, maior que, mais alto ou mais baixo) e o uso de outros conceitos matemáticos e habilidades (por exemplo, as operações de adição, subtração, multiplicação e divisão) que permitem ao leitor combinar e integrar dados e identificar as relações matemáticas expressas no gráfico. A leitura entre os dados requer, “ao menos, um degrau de inferência lógica ou pragmática, necessário para passar da questão à resposta e ambos, questão e resposta, são derivados do texto” (Pearson and Johnson 1978, p. 161, apud Curcio, 1989).

Exemplo: Dado o diagrama de setores abaixo, responda o que se pede.



- 1) Quanto John necessita para pagar seu almoço em 5 dias de escola?

a) R\$1,00	c) R\$4,00
b) R\$2,00	d) R\$5,00

- 2) Qual é o total de dinheiro que John precisa para 5 dias na escola?

a) R\$2,00	c) R\$10,00
b) R\$7,00	d) R\$14,00

Figura 4: Exemplo Leitura além dos dados, de Curcio (1989, p.4)

2.2 Os Registros de Representação Semiótica, de Raymond Duval

Raymond Duval, filósofo e psicólogo francês, desenvolveu no Instituto de Pesquisa em Educação Matemática de Estrasburgo, na França, importantes estudos relativos à Psicologia Cognitiva. Sua teoria sobre registros de representação semiótica tem se mostrado importante instrumento de pesquisa no estudo da complexidade da aprendizagem de matemática.

Para Duval (2003, p.11), um dos principais objetivos do ensino da matemática é contribuir para o desenvolvimento geral das capacidades de raciocínio, de análise e de visualização dos alunos. Este autor destaca que a atividade cognitiva requerida pela matemática envolve duas características importantes:

- **A importância primordial das representações semióticas,**

pelo fato de que as possibilidades de tratamento matemático dependem do sistema de representação utilizado, e também pelo fato de que os objetos matemáticos não são objetos diretamente perceptíveis ou observáveis com a ajuda de instrumentos (o acesso a estes objetos está ligado à utilização de um sistema de representação que os permite designar);

- **A grande variedade de representações semióticas utilizadas em matemática**, como sistemas de numeração, figuras geométricas, escritas algébricas e formais, representações gráficas e tabulares e em língua natural.

Assim como os objetos matemáticos, os objetos estatísticos não são facilmente acessíveis ou perceptíveis: o acesso a esses objetos passa necessariamente por representações semióticas, e não se deve jamais confundir um objeto com sua representação. Para Duval (2004, p. 31), uma representação pode funcionar verdadeiramente como representação, e permitir o acesso ao objeto de estudo representado somente quando se cumprem duas condições: “que se disponha de ao menos dois sistemas semióticos diferentes para produzir a representação de um objeto, de uma situação, de um processo, e que espontaneamente possa se converter de um sistema semiótico para outro as representações produzidas”.

Desse modo, a compreensão em matemática implica a capacidade de transitar entre mais de um registro de representação semiótica. Neste sentido, Duval (2003) afirma que “a originalidade da atividade matemática está na mobilização simultânea de ao menos dois registros de representação, ou na possibilidade de trocar a todo o momento de registro de representação”. Pode-se afirmar o mesmo quanto à atividade estatística.

Dependendo da situação, em uma resolução de problema, um registro pode aparecer de forma privilegiada, mas deve existir sempre a possibilidade de passar de um registro ao outro. Por exemplo, em uma análise de um conjunto de dados, deve-se poder passar do banco de dados para uma distribuição em sua forma tabular ou em sua forma gráfica, e ainda para sua representação por meio de medidas-resumo. Assim, a resolução de problemas estatísticos passa não só pela transformação de um registro a outro, como também pelo uso simultâneo desses diversos registros para a obtenção do maior número de informações, permitindo a análise crítica dos dados, segundo os princípios da Análise

Exploratória de Dados.

Segundo Duval et. al. (1999), “o conteúdo de uma representação depende mais do registro de representação do que do objeto representado”. Isso implica que passar de um objeto para outro não é somente mudar o modo de tratamento, é também explicar as propriedades ou os diferentes aspectos de um mesmo objeto. Duas representações de um mesmo objeto, produzidas em dois registros diferentes, podem não ter o mesmo conteúdo, ou seja, formas diferentes de representar um mesmo objeto de estudo podem utilizar propriedades ou aspectos diferentes, que podem permitir um olhar diferenciado sobre os dados. Por isso, a compreensão matemática está ligada ao fato de dispor de ao menos dois tipos de representações diferentes para não confundir o conteúdo de uma representação com o objeto representado.

Por exemplo, podemos representar uma distribuição de frequências para certa variável de diferentes formas, como uma tabela ou um gráfico.

TABELA 1 – DISTRIBUIÇÃO DE 40 ESTUDANTES DE ENSINO MÉDIO, SEGUNDO A SÉRIE DE ESCOLARIDADE.

SÉRIE	Nº ALUNOS
Primeira	10
Segunda	15
Terceira	15
TOTAL	40

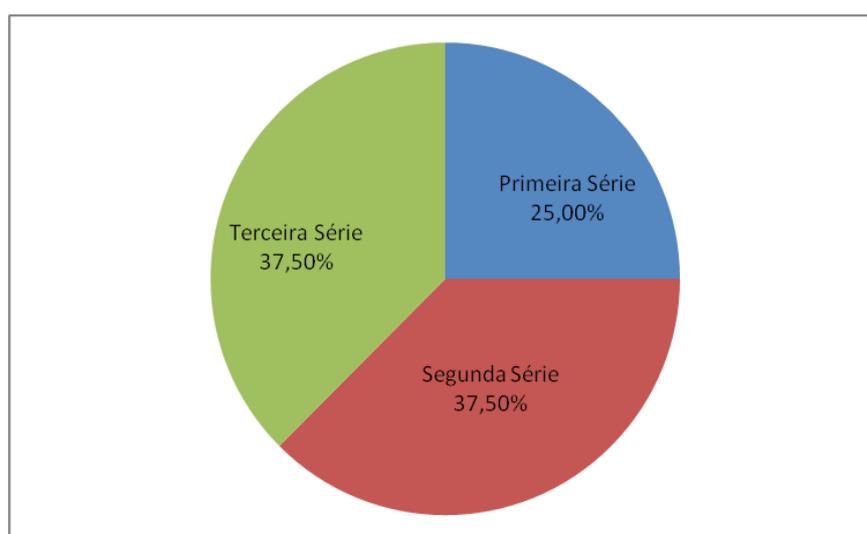


Figura 5: Gráfico de setores para a variável série

Para Duval (2003, p.22), “é a articulação dos registros que constitui uma

condição de acesso à compreensão em matemática, e não o inverso, ou seja, o enclausuramento de cada registro” e, em nosso trabalho, acreditamos que esta articulação entre os registros também constitui o acesso à compreensão da estatística.

Para análise do funcionamento cognitivo da compreensão, Duval considera dois tipos diferentes de transformação entre registros de representações semióticas: os tratamentos e as conversões.

O tratamento é a transformação da representação efetuada dentro de um mesmo registro (corresponde a procedimentos de justificação, produção de respostas). Por exemplo, Moretti (2003, p.153), em seu artigo sobre translação como recurso no esboço de curvas, mostra que a parábola resultante da expressão $y = 2x^2 - 8x + 10$ pode ser vista como o deslocamento da parábola $y = 2x^2$, da seguinte forma:

$$\begin{aligned}
 y = 2x^2 - 8x - 10 &\Leftrightarrow y = 2(x^2 - 4x - 5) \Leftrightarrow \\
 &\Leftrightarrow y = 2((x-2)^2 - 4 - 5) \Leftrightarrow \\
 &\Leftrightarrow y = 2((x-2)^2 - 9) \Leftrightarrow \\
 &\Leftrightarrow y = 2(x-2)^2 - 18 \Leftrightarrow \\
 &\Leftrightarrow y + 18 = 2(x-2)^2
 \end{aligned}$$

Figura 6: Exemplo de “tratamento” em álgebra (Moretti, 2002, p.153)

Pensando em Estatística, poderíamos considerar como tratamento a passagem da enumeração em série de rol para uma distribuição de freqüências ou, no caso de variáveis qualitativas ou quantitativas discretas, a mudança de escala em um gráfico. Para exemplificar melhor um tratamento, podemos pensar na situação apresentada a seguir.

Um questionário foi aplicado aos dez funcionários do setor de contabilidade de uma empresa, fornecendo os dados apresentados na tabela:

Tabela 2: Escolaridade dos funcionários

Funcionário	Curso
Carlos	Superior
João	Superior
José	Médio
Cinta	Médio
Pedro	Médio
Maria	Médio
Caio	Médio
Gabriela	Médio
Roberto	Fundamental
César	Fundamental

Podemos realizar um tratamento na tabela (na perspectiva de Duval), construindo a distribuição de freqüências, uma vez que o objeto continua a ser representado por um mesmo tipo de registro, ou seja, a tabela.

Tabela 3: Distribuição dos funcionários, segundo sua escolaridade

Curso	Nº Funcionários
Fundamental	2
Médio	6
Superior	2
Total	10

Um outro exemplo de tratamento em estatística, como citamos anteriormente, poderia ser a mudança de escalas em um gráfico, provocada pela mudança da amplitude das classes consideradas, assim como pela mudança de escala no eixo das freqüências, que indica o número de funcionários, como mostramos na figura (7).

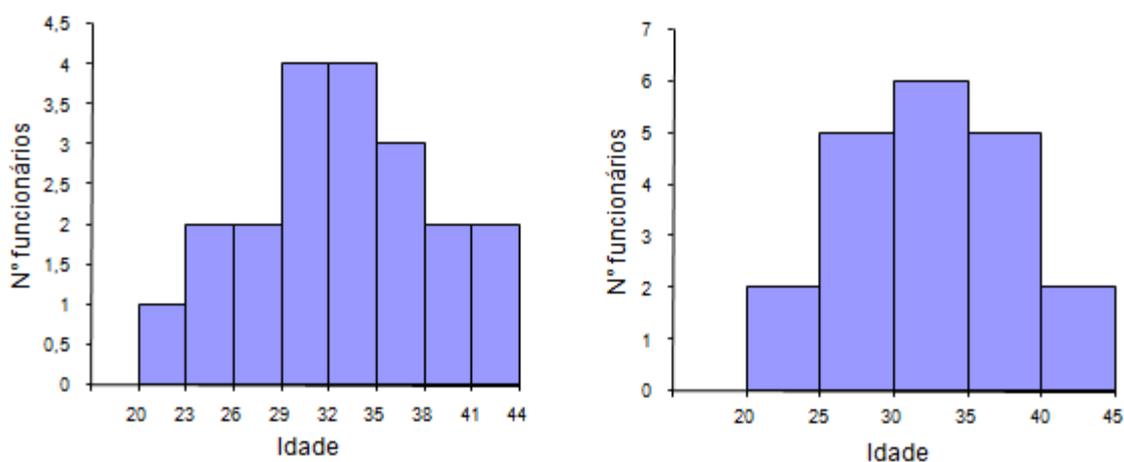


Figura 7: Mudança da escala do gráfico da variável idade

A conversão é a transformação da representação que ocorre entre registros, mudando o sistema de representação, mas conservando a referência aos mesmos objetos, ou seja, conservando o objeto matemático da representação inicial. Por exemplo, ocorre a conversão na passagem da parábola de expressão $y = 2x^2$ para o seu gráfico cartesiano (MORETTI, 2003, p.155).

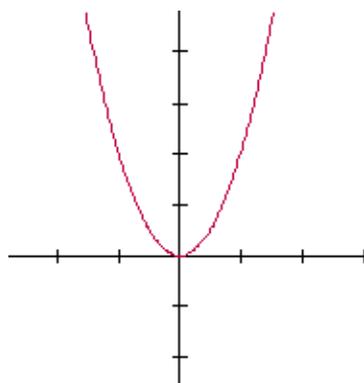


Figura 8: Gráfico no plano cartesiano da parábola de equação $y=2x^2$

Em estatística, podemos pensar em conversão como a passagem dos dados de uma tabela para um gráfico. Por exemplo, a tabela abaixo mostra a distribuição de freqüências para a variável gênero, de uma pesquisa realizada com alunos do Ensino Médio:

**TABELA 4 – DISTRIBUIÇÃO DOS ESTUDANTES DE ENSINO MÉDIO,
POR GÊNERO**

GENÉRO	Nº ALUNOS
Feminino	15
Masculino	25
TOTAL	40

Podemos efetuar a conversão de registros, representando essa mesma distribuição na forma gráfica:

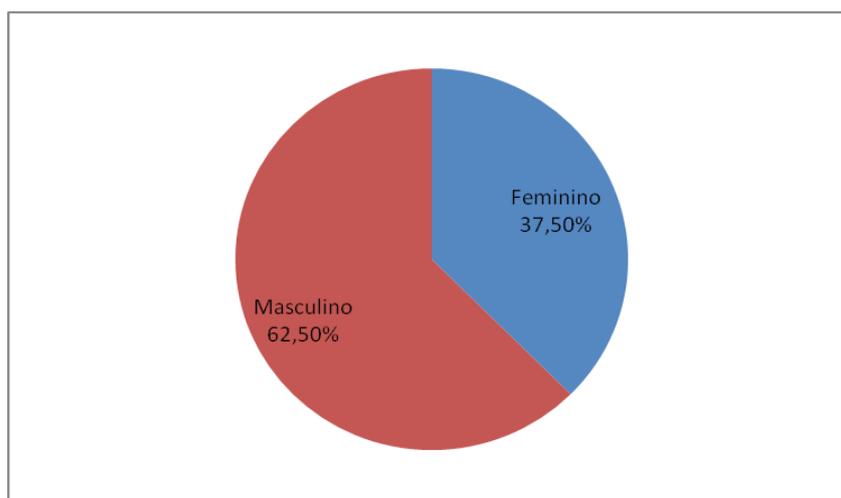


Figura 9: Gráfico de setores para a variável gênero

Duval (1996, p.3) relata que as operações de conversão são cognitivamente mais complexas do que as de tratamento, e, conseqüentemente, são mais difíceis de adquirir, ou construir.

A respeito das atividades de conversão, são observáveis dois tipos de fenômenos:

1. As variações de congruência e não congruência. A congruência ocorre quando a representação final transparece na representação inicial, e a conversão está próxima de uma situação de simples codificação. A não-congruência ocorre quando a representação final não transparece absolutamente na representação inicial;
2. A heterogeneidade nos dois sentidos de conversão. A importância no sentido da conversão deve-se pelo fato de que nem sempre a conversão efetua-se quando se inverte os registros inicial e final, e a

inversão do sentido da conversão pode levar a contrastes de acerto, pois o estudante pode não reconhecer a situação apresentada na representação que ele próprio tinha escolhido anteriormente.

Por exemplo, consideremos uma função representada na forma algébrica por $y = 3x + 2$. Para passar do registro algébrico para o registro gráfico desta função, o estudante pode obter alguns pontos por substituição na expressão da função, que são localizados em um sistema cartesiano para que se possa obter o esboço do gráfico.

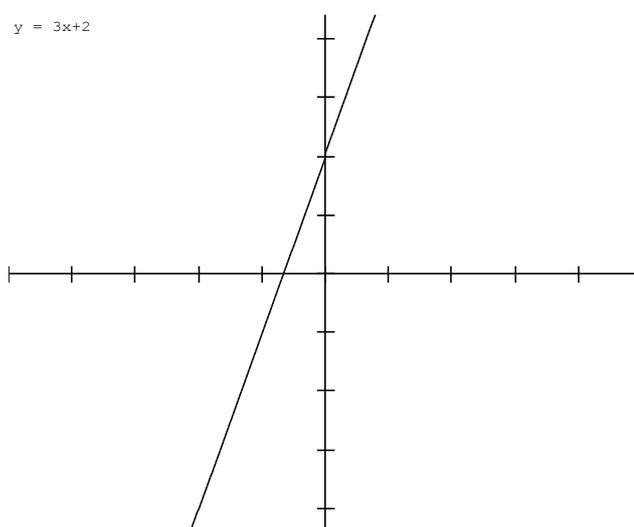


Figura 10: Gráfico da função $y = 3x + 2$

No entanto, se alterarmos o sentido da conversão, ou seja, se apresentarmos ao estudante o esboço do gráfico de uma função no plano cartesiano e solicitarmos que ele obtenha a expressão algébrica que representa tal função, o estudante poderá ter uma dificuldade maior para relacionar os pontos do gráfico com uma expressão algébrica que os generalize.

Neste sentido, dizemos então que, no primeiro caso, na passagem do registro algébrico para o gráfico, há congruência, pois a conversão dos registros pode ser efetuada de forma direta pelo estudante. Já, ao inverter o sentido da conversão, que seria a passagem do registro gráfico para o registro algébrico, não haverá congruência, pois a conversão dos registros não é realizada de forma direta.

Podemos pensar de modo parecido em estatística. Para passar de uma

tabela de distribuição de freqüências de uma variável estatística para o seu gráfico, podemos dizer que existe congruência, pois a passagem é feita de forma direta.

Porém, na passagem do gráfico para a tabela, pode haver ou não congruência, dependendo do tipo de gráfico.

Quando a leitura do gráfico é direta, há congruência (por exemplo, a passagem de um gráfico de colunas para a sua tabela de distribuição de freqüências).

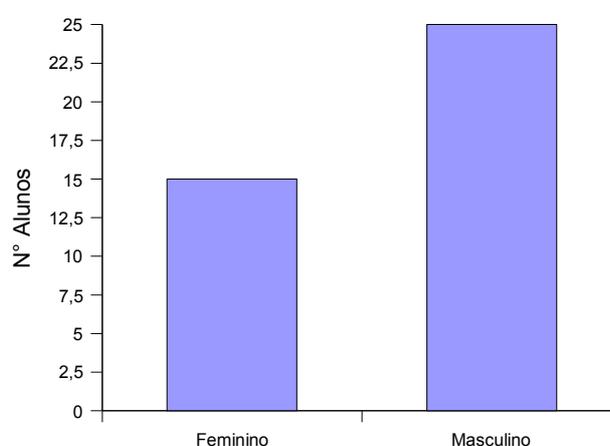


Figura 11: Gráfico de colunas para a variável gênero

TABELA 5 – DISTRIBUIÇÃO DOS ESTUDANTES DE ENSINO MÉDIO, POR GÊNERO

GENÉRO	Nº ALUNOS
Feminino	15
Masculino	25
TOTAL	40

Já para alguns outros tipos de gráficos, a passagem para a tabela não é feita de forma direta, não havendo congruência. Por exemplo, para passar de um gráfico de setores para a tabela de distribuição de freqüências, é necessário calcular a porcentagem de cada setor pelo ângulo, o que não é realizado de uma forma direta (ou mesmo de um gráfico de freqüências acumuladas).

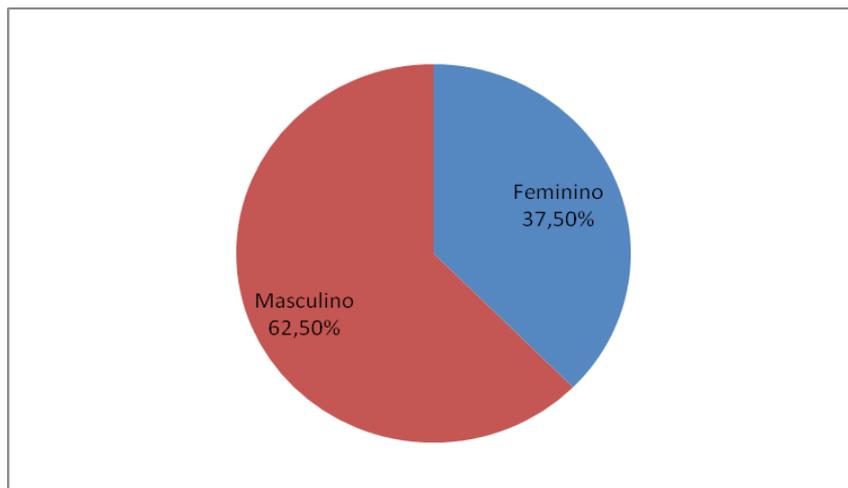


Figura 12: Gráfico de setores para a variável gênero

A representação figural é considerada de grande importância para o estudo da matemática, pois, na resolução de um problema, uma figura pode mostrar mais facilmente a idéia de solução do problema, sendo mais acessível ao aluno do que um texto, por exemplo. Porém um desenho pode ser visto de várias maneiras por diferentes indivíduos. Duval (1994, p.123) afirma que há várias maneiras de apreender uma figura em um contexto geométrico, destacando quatro tipos de diferentes apreensões: a apreensão perceptiva, a apreensão discursiva, a apreensão seqüencial e a apreensão operatória. Embora Duval tenha desenvolvido estes estudos em um contexto de geometria, acreditamos que a representação figural é muito importante na compreensão de conceitos estatísticos, e buscaremos compreender as diferentes apreensões no contexto de resolução de problemas que envolvem a análise de um conjunto de dados.

A **apreensão perceptiva** de uma figura permite identificar ou reconhecer imediatamente uma forma ou um objeto matemático (em geometria), no plano e no espaço. Segundo Almouloud (2003, p.127), a apreensão perceptiva é a *interpretação das formas da figura em uma situação geométrica*. Duval (1994, p.124) diz que a apreensão perceptiva tem a função epistemológica de identificação dos objetos em duas ou três dimensões, e isso se faz pelos tratamentos cognitivos efetuados automaticamente e, portanto, inconscientemente.

Por exemplo, temos o seguinte objeto matemático representado no plano (figura 13). A apreensão perceptiva da figura permite identificar que ela representa

um triângulo.

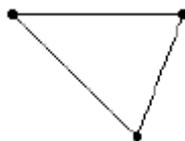


Figura 13: Triângulo

Em estatística, podemos estabelecer como exemplo que a apreensão perceptiva do gráfico abaixo (figura 14) permite identificar que esta figura representa um gráfico de colunas da variável série, e que cada coluna representa a quantidade de alunos de determinada série. Assim, essa apreensão permite fazer a leitura direta dos eixos.

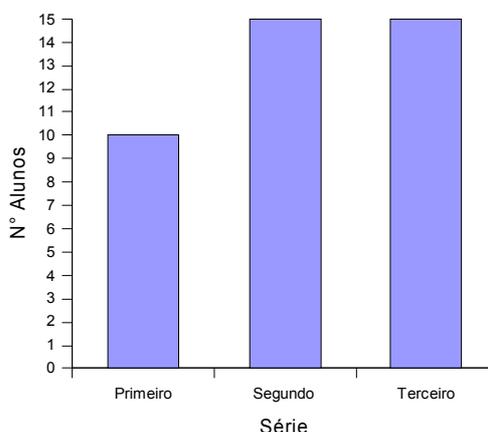


Figura 14: Gráfico de colunas da variável série

Vale ressaltar que, do ponto de vista de recursos cognitivos dispendidos, esta situação requer mais do indivíduo do que a primeira (apreensão perceptiva do triângulo). Isso se deve à numerosidade de variantes e atributos deste exemplo, que são o tipo de variável, a quantidade de valores da variável, a leitura dos eixos; o que nos faz levantar a hipótese de que a apreensão perceptiva em relação à estatística seria mais complexa do que na geometria, mas permite a leitura simples de um gráfico (leitura dos dados, de acordo com Curcio, 1989). Para os outros tipos de apreensões, levantamos a mesma hipótese: em estatística, são mais complexas.

Na **apreensão discursiva**, são explicitadas propriedades matemáticas da figura (além daquelas propriedades que já foram indicadas por uma legenda ou

pelas hipóteses). Segundo Almouloud (2003, p. 127), a apreensão discursiva é a *interpretação dos elementos da figura geométrica, privilegiando a articulação dos enunciados, levando em consideração a rede semântica de propriedades do objeto*. Duval (1994, p.124) diz que estas explicitações são de natureza dedutiva, e sua função epistemológica é de demonstração (como no exemplo abaixo).

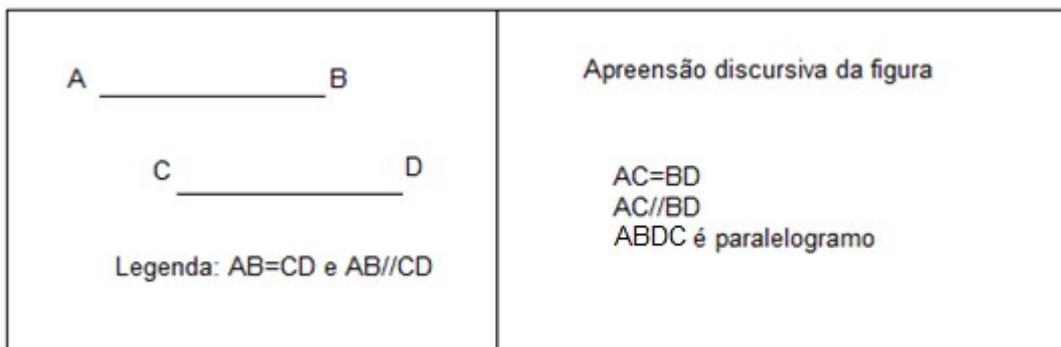


Figura 15: Apreensão discursiva de uma figura

Em estatística, podemos estabelecer como exemplo que a apreensão discursiva de um gráfico Box-plot seria a explicitação das propriedades do gráfico, obtidas através do estudo da variação dos dados. Por exemplo, pela visualização de um Box-plot, podemos identificar algumas propriedades, localizando as medidas pela posição que ela ocupa.

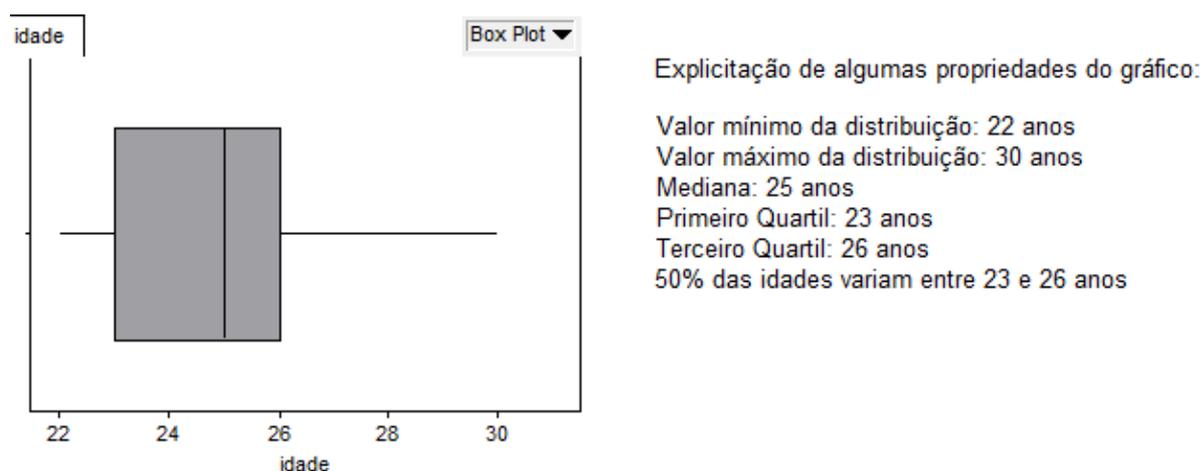


Figura 16: Apreensão Discursiva (Box-Plot)

A **apreensão seqüencial** de uma figura refere-se à ordem da construção da mesma, que depende não somente das propriedades matemáticas da figura à construir, mas também de técnicas e dos instrumentos utilizados (régua e

compasso, por exemplo). Segundo Almouloud (2003, p.127), este tipo de apreensão é *solicitado nas tarefas de descrição com objetivo de reproduzir uma figura*. Para Duval (1994, p.126), a função da apreensão seqüencial de uma figura é de modelo.

Por exemplo, a apreensão seqüencial de um triângulo eqüilátero seria a seqüência de passos para a construção deste triângulo.

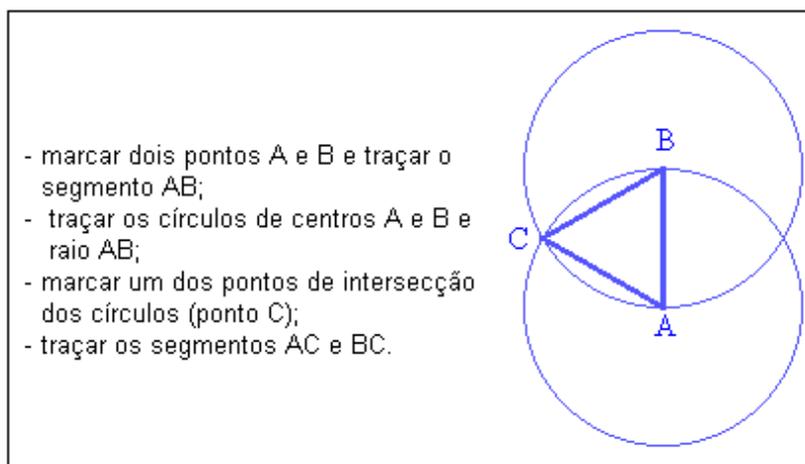


Figura 17: Construção de um triângulo eqüilátero.

Em estatística, podemos estabelecer como exemplo que a apreensão seqüencial de um gráfico de colunas seria a seqüência de passos para a construção do gráfico Box Plot, ou seja, a passagem do gráfico de colunas para um polígono de freqüência acumulada, e depois para o Box-plot, ou então a passagem de um gráfico de colunas ou histograma diretamente para o Box-plot, pela determinação algébrica da mediana e dos quartis. Na figura, mostramos a passagem de um histograma para um polígono de freqüências acumuladas e depois para o Box-plot.

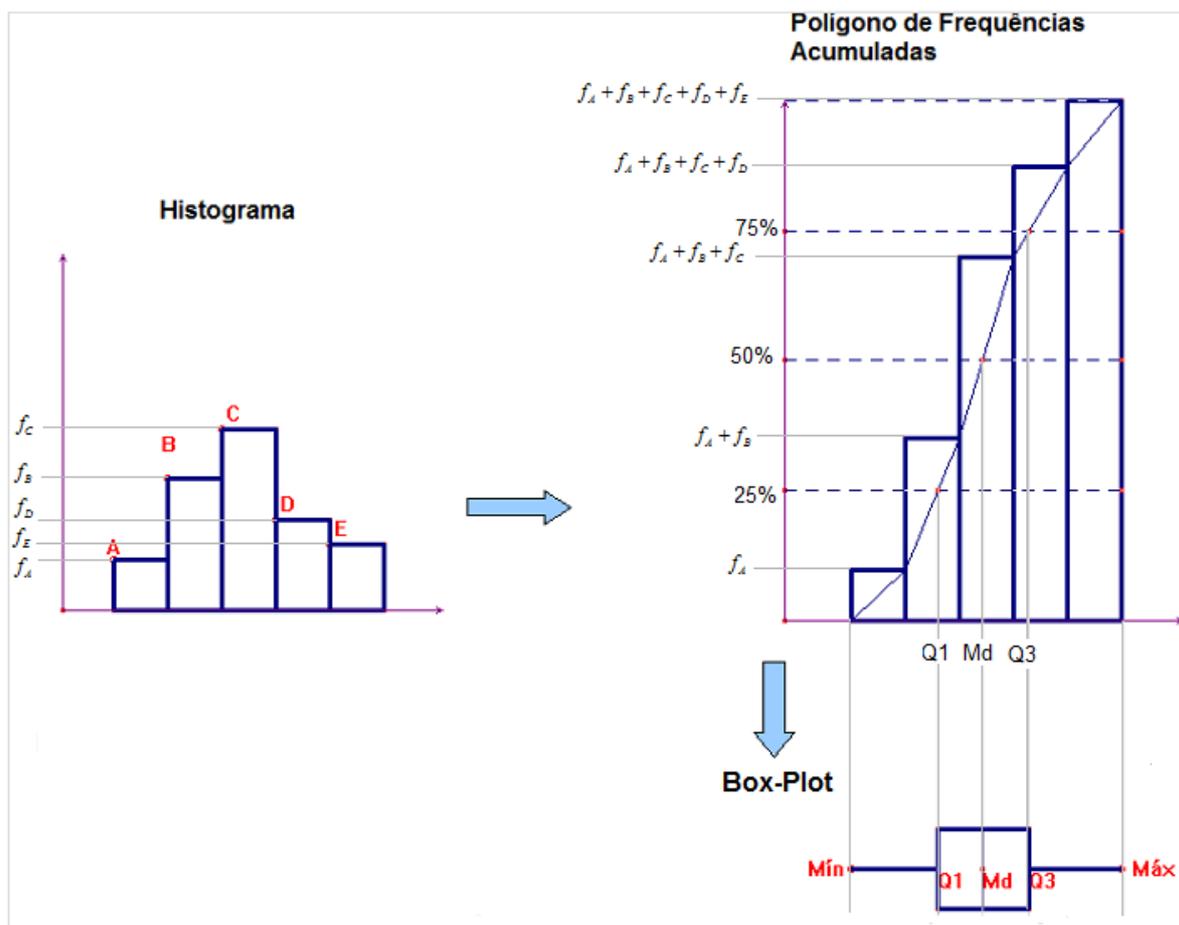


Figura 18: Apreensão seqüencial de uma figura em estatística.

Na realidade, aqui temos duas vezes a apreensão seqüencial: a primeira, para construir o polígono de freqüências acumuladas; a segunda, que parte do polígono para a construção do Box-plot, pela determinação da intersecção do polígono com as linhas indicativas de 25%, 50% e 75% das freqüências acumuladas, respectivamente.

A **apreensão operatória** corresponde à modificação (transformação) de uma figura dada em outras figuras possíveis. Segundo Almouloud (2003, p.127), a apreensão operatória *está centrada nas modificações possíveis de uma figura e na reorganização perceptiva que essas modificações sugerem*. Para Duval (1994, p.126), sua função é de exploração heurística, pois freqüentemente a figura geométrica é transformada em outras figuras para nos mostrar uma “idéia” da solução de um problema ou de uma demonstração.

Por exemplo, para obter a área de um paralelogramo ABCD, podemos modificar esta figura em um retângulo ABXY (como na figura 19), o que seria um

tipo de apreensão operatória.

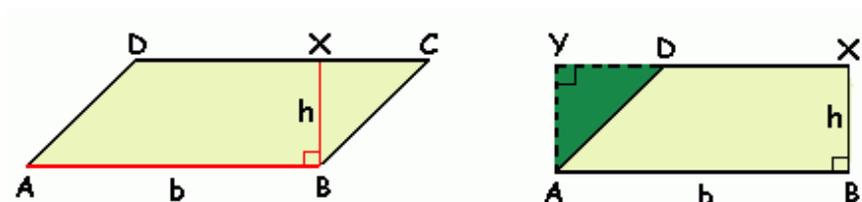


Figura 19: Modificação do paralelogramo ABCD

Em Estatística, podemos estabelecer como exemplo de apreensão operatória, a passagem de um histograma para um polígono de freqüências acumuladas.

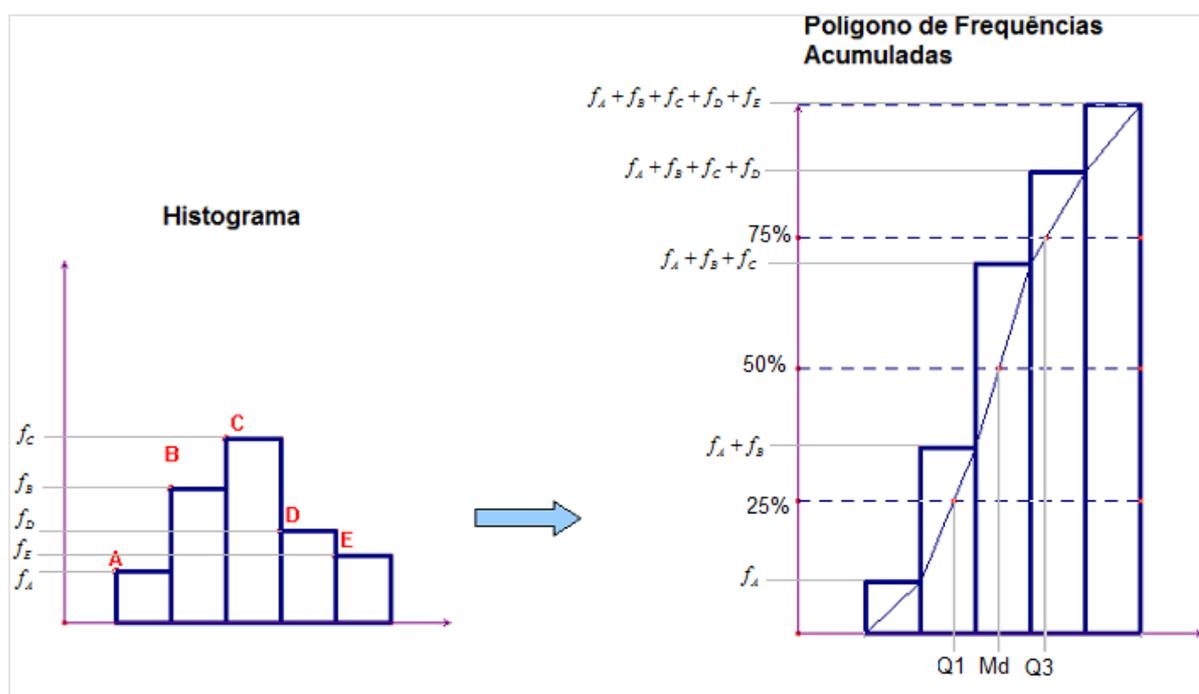


Figura 20: Exemplo de apreensão operatória em estatística

2.3 Relações entre os níveis de compreensão gráfica de Curcio, e os tipos de apreensões, de Duval

Acreditamos que os trabalhos de Duval e de Curcio adequam-se muito bem ao enfoque da análise exploratória de dados, pois ressaltam a importância de diferentes tipos de representação de um mesmo objeto para uma

aprendizagem mais significativa.

Assim, nosso objetivo é tentar estabelecer um paralelo entre os tipos de apreensões de uma figura, de Duval, e os Níveis de compreensão gráfica, de Curcio.

No quadro abaixo, fazemos uma hipótese de como relacionar estas duas teorias.

Quadro 1: Relações hipotéticas entre as Teorias de Duval e Curcio

Tipos de apreensões possíveis de uma figura	Níveis de Compreensão Gráfica
Apreensão Perceptiva	Nível 1: Leitura dos Dados
Apreensão Discursiva	Nível 1: Leitura dos Dados + Nível 2: Leitura entre os Dados
Apreensão Operatória Apreensão Seqüencial	Nível 1: Leitura dos Dados + Nível 2: Leitura entre os Dados + Nível 3: Leitura além dos Dados

A apreensão perceptiva, de Duval, requer a identificação do objeto matemático, que se faz pelos tratamentos cognitivos efetuados automaticamente. Acreditamos que este tipo de apreensão está presente no “nível 1” de compreensão gráfica de Curcio, “leitura dos dados”, pois o sujeito apenas levanta dados e informações explícitas no gráfico. Ele visualiza a figura e levanta os dados apresentados explicitamente nela.

Na apreensão discursiva de Duval, são explicitadas propriedades matemáticas da figura, é a interpretação dos elementos da figura. Em estatística, cremos que para este tipo de apreensão, o sujeito deve ler os dados e ler entre os dados, estando no “nível 2” de compreensão gráfica segundo Curcio, pois, ao explicitar as propriedades do gráfico, o sujeito também interpreta e identifica relações nele.

A apreensão operatória de Duval é a apreensão que corresponde à modificação de uma figura em outras possíveis. Em estatística, acreditamos que este tipo de apreensão exige a leitura dos dados, a leitura entre os dados, e a leitura além dos dados, portanto o sujeito estaria no “nível 3” de compreensão

gráfica de Curcio, pois, ao modificar um gráfico o sujeito vai além dos dados e das informações explícitas no mesmo.

A apreensão seqüencial refere-se à ordem da construção de uma figura, e é solicitado nas tarefas de descrição, com objetivo de reproduzir uma figura. Em estatística, acreditamos que este tipo de apreensão exige a leitura dos dados, a leitura entre os dados e a leitura além dos dados, pois a seqüência de passos exigida para a construção de gráficos requer que o sujeito reflita além dos dados e das informações disponíveis.

Vale ressaltar que estas relações são apenas algumas hipóteses que estamos levantando em nosso trabalho. A equivalência entre estas teorias implica a realização de um estudo empírico, que necessita de ferramentas de validação estatística, que não é o foco deste trabalho, mas que seria uma perspectiva para trabalhos futuros, em que se possam aprofundar estes estudos. Neste trabalho, buscaremos apenas construir situações didáticas que envolvam os tipos de apreensões de uma figura e os níveis de compreensão gráfica, buscando identificar as condições de mobilização simultânea.

Capítulo III – Estudos Preliminares

3.1 A Análise Exploratória de Dados

O desenvolvimento tecnológico dos últimos anos vem permitindo cada vez mais o aumento das possibilidades de construção de representações gráficas e tabulares de uma distribuição de freqüências, assim como de cálculo das medidas-resumo de uma distribuição, possibilitando o surgimento de uma nova filosofia nos estudos estatísticos, introduzida por Tukey (1977, apud BATANERO 1991): a *Análise Exploratória de Dados*.

Segundo Biehler (1989, p.191), a abordagem da Análise Exploratória de Dados representa um desafio nos pontos de vista, atitudes e valores mais tradicionais, que influem nos currículos e nos enfoques do ensino de estatística.

Ainda segundo este autor, uma idéia subjacente da análise exploratória de dados é que, variando a representação e empregando representações múltiplas de dados, tem-se um meio de desenvolver novos conhecimentos e idéias.

Nosso objetivo neste trabalho é desenvolver o estudo da Análise Exploratória de Dados com alunos do segundo ano do ensino médio. Estudos recentes (BATANERO, 1991; BIEHLER, 2003; GODINO, 1995) investigam a utilização desse enfoque em sala de aula com estudantes de diferentes níveis de escolaridade. Batanero (1991, p.2) indica algumas características dele, que o fazem um tema apropriado de estudo no Ensino Médio:

- **Possibilidade de gerar situações de aprendizagem referentes a temas de interesse para o aluno.** O usual é trabalhar sobre um projeto em que se possam coletar os dados mediante experimentos ou na realização de uma pesquisa, ou retirar dados da internet ou de um anuário estatístico. Isto pode motivar os estudantes, que podem ver a utilidade da estatística na investigação;
- **Forte apoio em representações gráficas:** Uma idéia fundamental da análise exploratória de dados é que o uso de representações múltiplas de dados se converte em um meio de desenvolver novos conhecimentos e perspectivas, e isto coincide com a importância que se dá às representações gráficas nos novos parâmetros curriculares;
- **Não necessita de uma teoria matemática complexa:** Como a análise de dados não propõe que estes se distribuam

segundo uma lei de probabilidade clássica, utiliza somente noções matemáticas muito elementares e procedimentos gráficos fáceis de realizar, e assim qualquer aluno pode fazer uma análise de dados.

Neste sentido, o trabalho em ambiente computacional revela-se fundamental para que o aluno possa efetivamente preocupar-se com as interpretações, com os conceitos envolvidos, sem perder-se em cálculos e construções de pouco significado cognitivo para ele. Concordamos com Batanero, Estepa e Godino (1991), quando se afirma que o trabalho com computador pode facilitar, tornando mais rápido o trabalho de cálculo e representação gráfica, já que o interesse concentra-se na comparação de todas as representações possíveis sobre os mesmos dados.

Godino (1995, p.4) ressalta que um ponto importante da Análise Exploratória de Dados é que não se trata de um conjunto de métodos, e sim de uma filosofia de aplicação da estatística, que consiste no estudo dos dados a partir de todas as perspectivas e com todas as ferramentas possíveis, com propósito de extrair o máximo de informação, gerando novas hipóteses, no sentido de conjecturar sobre as observações de que dispomos. Godino (1995, p.5) ainda cita alguns procedimentos que os alunos poderiam aplicar usando um programa computacional:

- Estudo univariante: estudo das distribuições de freqüências e resumos estatísticos de cada uma das variáveis de todo o conjunto de dados. Para isso, pode-se utilizar programas que permitam calcular estatísticas elementares (medidas resumo, dispersão) e tabelas e gráficos de distribuição de freqüência (histograma, polígono de freqüência, box-plot). A representação gráfica de uma distribuição acompanhada de suas medidas-resumo pode ajudar a decidir, para cada variável, qual a medida mais adequada como base de comparação. Ao dispor de computadores, é possível uma atitude mais crítica e analítica dos dados coletados, podendo comparar as características dessa distribuição com outras disponíveis;
- Comparação de sub-populações: Estabelecer análises estatísticas para diferentes subamostras (meninos e meninas, por exemplo), comparando as estatísticas e as distribuições de freqüências;
- Estudo da associação estatística: O estudo de tabelas de contingência, das variáveis agrupadas em intervalos, permite comparar as distribuições condicionais e a diferença entre

freqüências observadas e esperadas, e estudar os conceitos de associação e independência entre variáveis estatísticas.

Em nosso trabalho, as contribuições do ambiente informatizado para o desenvolvimento de habilidades necessárias para a análise exploratória de dados seriam as possibilidades de diversos tipos de representações que o software proporciona, no sentido de que, com o máximo de representação, pode-se obter o máximo de informação.

3.2 Pensamento Estatístico

Wild e Pfannkuch (1998, 1999) escreveram sobre o “Pensamento Estatístico”, tomando como ponto de partida a concepção de que o objetivo fundamental da investigação estatística é a aprendizagem na esfera do contexto do problema e que o “Pensamento Estatístico” é a interação entre o estatístico e o contextual. Ainda segundo estes autores, o Pensamento Estatístico é utilizado em contraste com a aplicação mecânica das técnicas estatísticas.

Para Snee (1990, apud WILD e PFANNKUCH, 1998, p.1), o Pensamento Estatístico se define como:

“processos de pensamento que reconhecem a presença da variação em torno de tudo o que nos rodeia e presente em qualquer coisa que fazemos, todo o trabalho é uma série de processos interconectados, e a identificação, a caracterização, a quantificação, o controle e a redução da variação proporcionam oportunidades para melhorias”.

Wild e Pfannkuch (1998, p. 3) ressaltam que o Pensamento Estatístico consiste em processos genéricos do pensamento, os quais deveriam ter lugar onde a metodologia estatística encontre um problema do mundo real.

Campos (2007), considera que o Pensamento Estatístico ocorre quando os modelos matemáticos são associados à natureza contextual do problema em questão, ou seja, quando surge a identificação da situação analisada e faz-se uma escolha adequada das ferramentas estatísticas necessárias para sua descrição e interpretação.

Wild e Pfannkuch (1999) apresentam uma estrutura de Pensamento

Estatístico baseada em quatro dimensões: ciclo investigativo; tipos de pensamento; ciclo interrogativo; disposições.

O ciclo investigativo refere-se a forma de que se atua e de que se pensa durante o curso de uma investigação estatística (WILD E PFANNKUCH, 1999, p. 4). Foi adaptado do modelo PPDAC (*Problem, plan, data, analysis, conclusions*), que, segundo estes autores, tem o objetivo de resolver um problema real, geralmente com a intenção de mudar um *sistema* para melhorar algo.

A segunda dimensão, “tipos de pensamento”, refere-se ao tipo de pensamento fundamental e ao tipo de pensamento geral. O pensamento fundamental é o reconhecimento da necessidade de dados, a transnumeração (formas de mudar representações de dados e aspectos de um sistema para melhorar a compreensão desse sistema), a consideração da variação, o uso de modelos estatísticos e a integração da estatística com o contexto. O pensamento geral refere-se ao planejamento do ciclo investigativo: o pensamento estratégico, a modelização e a aplicação de técnicas.

O ciclo interrogativo é um processo geral de pensamento de uso constante em resolução de problemas estatísticos. O *pensador* gera possibilidades (a partir do contexto, dos dados e do conhecimento estatístico), busca informação e idéias para, posteriormente, interpretar o resultado estatístico (ler – traduzir – resumir – comparar – conectar).

Para Wild e Pfannkuch (1999, p. 12), as disposições podem ser entendidas como o compromisso do pensador com o problema: *curiosidade* (querer investigar mais); *imaginação* (procurar enxergar o problema sob diferentes pontos de vista); *cepticismo* (questionar as conclusões); *lógica* (detectar quando uma idéia surge de outra); *abertura* (considerar novas idéias) e *perseverança*.

Silva (2007, p.30), entende o Pensamento Estatístico como as estratégias mentais utilizadas pelo indivíduo para tomar decisão em toda a etapa de um ciclo investigativo, e sugere uma interpretação do Pensamento Estatístico baseado em Wild e Pfannkuch (1999), no esquema a seguir:

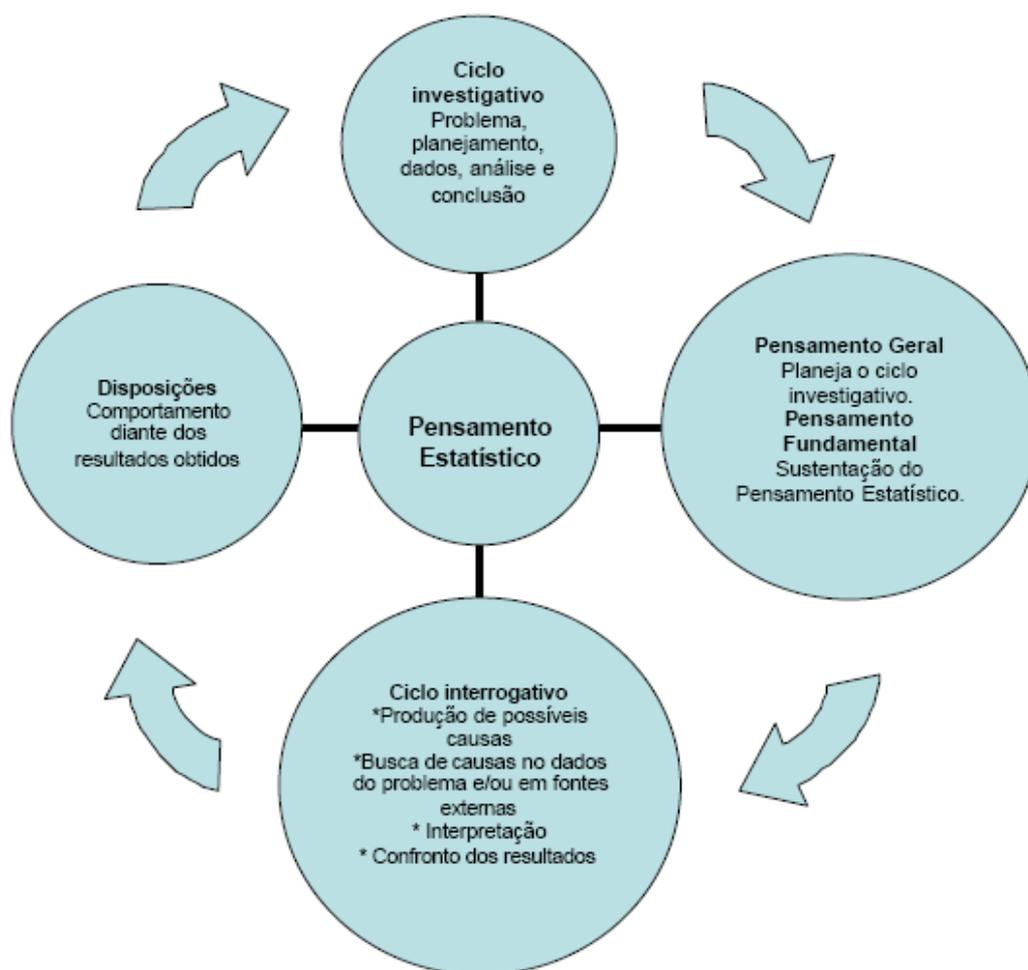


Figura 21: Estrutura do Pensamento Estatístico (SILVA, 2007)

Para Campos (2007, p. 39), uma característica do Pensamento Estatístico é prover a habilidade de enxergar o processo de maneira global, com suas interações e seus porquês, entender suas diversas relações e o significado de suas variações, explorar os dados além do que os textos prescrevem e gerar questões não previstas inicialmente. Assim, o pensador estatístico é capaz de ir além do que lhe é ensinado, questionando espontaneamente e investigando os resultados acerca dos dados envolvidos num contexto específico.

Outros autores têm abordado aspectos que buscam caracterizar o Pensamento Estatístico, mas vamos limitar nosso universo aos que foram aqui citados.

3.3 Revisão Bibliográfica

Durante o desenvolvimento deste trabalho, analisamos vários resultados de pesquisas na área do ensino e aprendizagem de estatística, como Cazorla (2002), Godino (1995), Batanero, Godino, Green, Holmes, e Vallecillos (1994), Baker, Biehler e Konold (2004), Silva (2007).

Cazorla (2002), em sua tese “A relação entre a habilidade viso-pictórica e o domínio de conceitos estatísticos na leitura de gráficos”, investigou os fatores que interferem na leitura de gráficos estatísticos. Para a autora, o fundamental para uma boa interpretação de gráficos é o desenvolvimento do Pensamento Estatístico, que se refere não só ao domínio de conceitos e procedimentos estatísticos, mas também à capacidade de utilizar de forma adequada as ferramentas estatísticas na solução de problemas, contextualizando as informações e indo além delas, sendo capaz de entender a essência do saber e inferir prováveis tendências.

Uma das teorias que a autora estudou para contribuir na fundamentação de seu trabalho foi a “Teoria das habilidades matemáticas”, de Krutetskii (1976). Para Krutetskii (1976, apud CAZORLA, 2002, p. 130), habilidades seriam especificidades das pessoas, que permitem a realização de determinada atividade ou tarefa com sucesso (referindo-se às esferas sensorial, cognitiva e motora, incluindo aspectos como percepção, atenção, memória, imaginação e pensamento), sendo que a prontidão ou a capacidade de realizar com sucesso uma atividade ou tarefa depende da habilidade do indivíduo. Cazorla (2002, p.130) afirma que são ainda necessárias outras condições psicológicas, tais como a atitude frente à atividade, os traços de personalidade, o estado mental, bem como os conhecimentos, destrezas e hábitos. Então, para Krutetskii (1976, apud CAZORLA, 2002, p.132), a habilidade matemática é *uma síntese única de propriedades, uma qualidade integral da mente, incluindo diversos aspectos mentais, e desenvolvida durante a atividade matemática.*

Cazorla (2002, p.132) descreve que, segundo Krutetskii (1976), existem três estágios básicos na atividade mental durante a solução de problemas matemáticos:

- **Obtenção da informação matemática:** habilidade para compreender a estrutura formal do problema;
- **Processamento da informação matemática,** em que vários componentes são evidenciados, podendo ou não interagir entre si;
- **Retenção da informação matemática:** esse estágio admite a existência de uma memória matemática (memória generalizada para relações matemáticas, tipos característicos, esquemas de argumentos e provas, métodos de solução dos problemas e tipos de abordagens utilizadas).

Cazorla (2002, p.132) afirma que a cada um desses estágios corresponde a uma ou várias habilidades matemáticas, constituídas por componentes, sendo que esses componentes relacionam-se segundo um modelo estrutural hierárquico, em que cada fator corresponde a um dos estágios básicos da atividade mental durante a solução de problemas matemáticos, além de um componente geral, que é identificado como **componente sintético**.

Para Krutetskii (1976, apud CAZORLA 2002, p.134), o segundo estágio, o do processamento de informações, é composto por:

- habilidade para **pensar logicamente** na área das relações espaciais e quantitativas, e habilidade para pensar em símbolos matemáticos;
- habilidade para **generalizar** de forma abrangente e rápida os conteúdos matemáticos, as relações e as operações;
- habilidade para **resumir** os processos matemáticos e os sistemas correspondentes de operações, além da habilidade para pensar através de estruturas curtas e eficientes;
- **flexibilidade** dos processos mentais na atividade matemática;
- **elegância**, que seria a inclinação para a clareza, simplicidade, economia e racionalidade da solução;
- habilidade para uma rápida e livre reconstrução do processo mental (**reversibilidade** dos processos mentais no raciocínio matemático).

Segundo Cazorla (2002, p. 134), o **componente geral sintético** (NEUMANN, 1995, apud CAZORLA, 2002) refere-se ao tipo de mente ou pensamento matemático, que para Krutetskii (1976, apud CAZORLA, 2002) era a qualidade matemática da mente, e que foi classificada em três grupos:

- Mente de tipo **analítica:** caracteriza-se pela predominância dos componentes lógico-verbais (pessoas que possuem

este tipo de mente operam com facilidade esquemas abstratos, sem suporte visual para perceber objetos ou modelos durante a solução de problemas; mesmo no caso do problema ser enunciado através de figuras ou desenhos, estes tendem a representar o enunciado em termos lógico-verbais);

- Mente do tipo **geométrico (viso-pictórica)**: caracteriza-se pela predominância dos componentes viso-figurativos (pessoas que possuem este tipo de mente precisam representar os problemas através de esquemas e figuras, necessitam criar suportes visuais na solução de problemas, visualizar objetos ou diagramas, e são marcadas por um forte desenvolvimento de conceitos espaciais;
- Mente do tipo **harmônico**: caracteriza-se pelo equilíbrio no domínio dos esquemas abstratos e das representações visuais (pessoas desse tipo, dependendo do problema, buscam a melhor forma de solucioná-lo, podendo ou não se valer de representações geométricas). Apresenta dois subtipos: a) Subtipo **harmônico abstrato**: pode usar suportes visuais na solução de problemas, mas prefere não fazê-lo; b) Subtipo **harmônico pictorial**: pode usar suportes visuais na solução de problemas e prefere fazê-lo.

Cazorla (2002) realizou sua pesquisa com 814 estudantes de graduação matriculados em várias turmas da disciplina de Estatística em uma universidade na Bahia, durante os anos de 1999 e 2000. Em relação à habilidade viso-pictórica (de KRUTETSKII, 1976), a avaliação foi realizada através de uma prova matemática, composta por três partes: figural, de geometria plana e de geometria espacial. A autora considerou o desempenho médio na prova (a média de acertos foi de 43,1%) como um nível fraco para usuários de estatística de nível universitário. Segundo a autora, os sujeitos conseguiram extrair informação ao nível elementar, extrair tendências quando explícitas e fazer projeções quando existia um padrão regular, porém apresentaram dificuldade no domínio da linguagem gráfica, na percepção de padrões mais complexos, e não conseguiram perceber a necessidade da utilização de conceitos mais elaborados para uma análise mais consistente. Ainda segundo a autora, foi encontrada uma relação significativa, linear e positiva entre o desempenho na prova matemática e na prova de gráficos, o que implica que quanto maior a habilidade viso-pictórica, maior o nível da leitura de gráficos. A autora acredita que o papel da instrução poderia e deveria elevar essas habilidades a patamares maiores, percebendo que

ainda há muito trabalho a ser feito na busca do desenvolvimento do Pensamento Estatístico.

Batareno, Godino, Green, Holmes e Vallecillos (1994) relatam os principais erros e dificuldades na compreensão dos conceitos estatísticos elementares. Segundo os autores (p.2), uma dessas dificuldades, é que algumas concepções, que permitem resolver um conjunto de tarefas, mostram-se limitadas e inapropriadas quando se aplicam a casos mais gerais, e que o sujeito mostra uma resistência a sua substituição. Outra dificuldade relatada no artigo deve-se a uma falta de conhecimento básico necessário para uma compreensão correta de um conceito ou procedimento dado.

No que se refere à representação gráfica, estes autores analisaram os principais erros, levando em consideração os níveis de compreensão gráfica de Curcio (1989), e encontraram as principais dificuldades aparecem no segundo e no terceiro nível (leitura entre os dados e leitura além dos dados). Os autores citam o trabalho de Li y Shen (1992, apud BATANERO, GODINO, GREEN, HOLMES e VALLECILLOS,1994), que relata exemplos de escolha incorreta do tipo de gráfico em projetos estatísticos realizados por estudantes de educação secundária (alguns alunos utilizaram polígonos de frequências com variáveis qualitativas). Para o autor, este problema agrava-se na disponibilidade de um software para a representação gráfica, e no desconhecimento do modo correto em que se deve ser empregado por parte dos alunos. Li y Shen (1992, apud BATANERO, GODINO, GREEN, HOLMES e VALLECILLOS,1994), ainda citam uma lista de erros de caráter técnico, como omitir as escalas dos eixos, não especificar a origem das coordenadas, não proporcionar divisões suficientes nas escalas dos eixos, comparar quantidades heterogêneas em um mesmo gráfico.

Baker, Biehler e Konold (2004), escreveram um artigo levantando o questionamento de que se estudantes de até 14 anos devem aprender a construir o gráfico Box-plot. Neste artigo, os autores realizaram uma pesquisa com alguns estudantes, e utilizaram o software Fathom para construir os gráficos após a coleta dos dados. Baker, Biehler e Konold (2004) consideram que o Box-plot é muito útil visualmente, para comparar distribuições, pois fornece uma visão compacta de onde os dados estão centrados e como eles são distribuídos na variável. Segundo os autores, em vários países este gráfico tornou-se parte do

currículo de análise de dados, porém a faixa etária em que se trabalha com ele difere-se consideravelmente entre estes países. Nos Estados Unidos, por exemplo, o Box-plot é trabalhado com alunos a partir de 12 anos. Em países como Austrália, Bélgica e Holanda, com alunos entre 15-16 anos.

Apesar das vantagens visuais no trabalho com o Box-plot, os autores citam alguns desafios que os estudantes podem enfrentar no trabalho com este tipo de gráfico (p.2):

- O Box-plot geralmente não permite perceber casos individuais³;
- A mediana não é tão intuitiva aos estudantes como os pesquisadores esperavam;
- A divisão dos dados em quartis é de difícil compreensão.

Segundo Baker, Biehler e Konold (2004), o Box-plot é rico conceitualmente, mas para compreendê-lo é necessário ter conhecimento sobre primeiro quartil, mediana e terceiro quartil, sendo necessário entender que a mediana é utilizada como uma medida central de distribuição.

Os autores concluem que o ensino do Box-plot deve ser iniciado com alunos acima da faixa etária de 14 anos, e sempre que for trabalhado este assunto, deve-se dedicar tempo suficiente para instrução, para poder facilitar a compreensão da mediana e dos quartis.

Entre as principais dificuldades (p. 9), os autores destacaram que:

- Os alunos tendem a ver os dados como casos individuais (como preparação para aprendizagem do Box-plot, pode-se utilizar uma combinação do Dot Plot e quartis);
- O Box-plot não exibe densidades ou frequências, em contraste com o que os alunos estão acostumados (para diminuir essa dificuldade, pode-se utilizar um software que exiba o Dot-plot e o Box-plot ao mesmo tempo);
- Embora a mediana seja relativamente fácil de aprender como um procedimento para obtenção de um valor médio, é mais difícil reconhecê-la como medida central de um grupo;
- Existe uma variedade de definições de quartis e, ao utilizar um software, o professor deve estar ciente da definição que o software utiliza⁴.

Silva (2007), em sua tese “Pensamento Estatístico e Raciocínio sobre variação: um estudo com professores de matemática”, realizou uma pesquisa com

³ Exceto os outliers e extremos.

⁴ Para este tipo de problema, podemos considerar o trabalho de Graham (2006), que diz que um cálculo “aproximativo” é suficiente, uma vez que é uma estimativa do parâmetro.

professores (sendo nove professores da escola básica e dois alunos do curso de graduação em matemática), com o objetivo de verificar o raciocínio sobre variação e variabilidade nas etapas do ciclo investigativo do Pensamento Estatístico elaboradas por Wild e Pfannkuch (1998, 1999).

Em seu trabalho, Silva (2007) fez inicialmente um diagnóstico com os professores sujeitos de sua pesquisa, com o objetivo de verificar como eles atribuíam significado à estatística e como os conceitos relacionados à variabilidade faziam parte dessa significação. Os resultados deste diagnóstico permitiram identificar pontos frágeis na significação de estatística para estes professores, como a ausência de raciocínio sobre variação (os professores apenas verbalizaram o desvio padrão, sem levar em consideração sua aplicação como ferramenta de pesquisa ou como conteúdo a ser ensinado). Foi observado que os professores utilizavam apenas a distribuição de frequências e sua respectiva representação gráfica para analisar os resultados de uma pesquisa, o que indicou a não abordagem do conceito de variação em suas aulas.

Durante a realização das atividades da pesquisa-ação (metodologia adotada pela autora em seu trabalho), foi possível notar que os professores utilizaram quase completamente o ciclo investigativo, utilizaram parcialmente os tipos de pensamento e o ciclo interrogativo, e apenas um professor apresentou disposição sobre os resultados, ou seja, transitou por todas as dimensões do Pensamento Estatístico. A autora ainda relata que a experiência com a elaboração de uma pesquisa, desde a definição de seus objetivos até a coleta e montagem do banco de dados (atividade baseada em PONTE, 2005, que também utilizaremos em nosso trabalho), permitiu um avanço no desenvolvimento do Pensamento Estatístico dos professores, que já transitavam em três das quatro dimensões de sua estrutura. Porém, o desenvolvimento do Pensamento Estatístico não implicou diretamente em um nível mais avançado do raciocínio de variação. Ainda segundo a autora, a discussão sobre as medidas de tendência central, permitiu observar a interpretação equivocada de média como maioria, que foi um fato impeditivo para a percepção da necessidade de uma medida de variação. A pronúncia “maior variação” pode ser interpretada como variação entre as observações, e raramente será entendida como variação em torno da média ou de qualquer medida de tendência central. No geral, concluiu-se que, durante o

desenvolvimento da pesquisa-ação existia raciocínio sobre variação, mas este não pode ser considerado como raciocínio completo de variação.

Um dos principais resultados da pesquisa desta autora foi a pouca familiaridade dos professores com a apreensão da variação dos dados. Pode-se inferir o mesmo para alunos.

3.4 O Software *Fathom*

O *Fathom* é um software de estatística dinâmica, desenvolvido especificamente para o ensino, utilizando as opiniões de alguns especialistas em educação estatística (FINZER, 2000 apud SÁNCHEZ, 2002).

Algumas das características essenciais deste software (SÁNCHEZ, 2002, p.3) são:

- Possibilidade de mover objetos;
- Possibilidade da representação e da coordenação simultânea de dados em tabelas e gráficos;
- Possibilidade de representar em gráficos e tabelas as características mais importantes de um conjunto de dados;
- Possibilidade de acesso imediato a uma série de funções, que fornecem uma ampla flexibilidade para o trabalho estatístico.

Nosso primeiro contato com este software foi através dos trabalhos de Rolf Biehler (2003) e de Avila (2005) e a partir destes trabalhos surgiu o interesse em trabalhar com o software por ser de fácil manuseio e possibilitar o acesso rápido a diferentes tipos de representações, como tabelas, gráficos e medidas resumo, características que vêm ao encontro do enfoque da Análise Exploratória de Dados utilizado neste trabalho.

Biehler (2003), em um estudo sobre a utilização de softwares na análise de dados, faz uma breve comparação entre o software Excel e o *Fathom*, explicando que este último oferece essencialmente mais possibilidades de experimentações com gráficos e métodos de simulação para análises estatísticas.

Vale ressaltar que o software não é livre⁵, e seu idioma é o inglês, pontos

⁵ Informações técnicas e financeiras sobre o software estão disponíveis no site www.keypress.com (acesso em maio/2008)

que dificultam o trabalho com ele em escolas públicas brasileiras. Para realizar esta pesquisa, utilizamos a licença disponível na Pontifícia Universidade Católica de São Paulo.

Iniciando o software, o usuário clica no ícone “New Collection” para inserir um novo banco de dados, e depois no ícone “New case table” para inserir os dados na tabela (os dados podem ser digitados diretamente na tabela, ou também podem ser copiados de tabelas prontas em outros aplicativos, como os softwares Word ou Excel, desde que estas tabelas sejam copiadas no bloco de notas do Windows, ou seja, desde que tenham formato “txt”; nesse caso, deve-se utilizar a ferramenta “Import from file” no menu “File”).

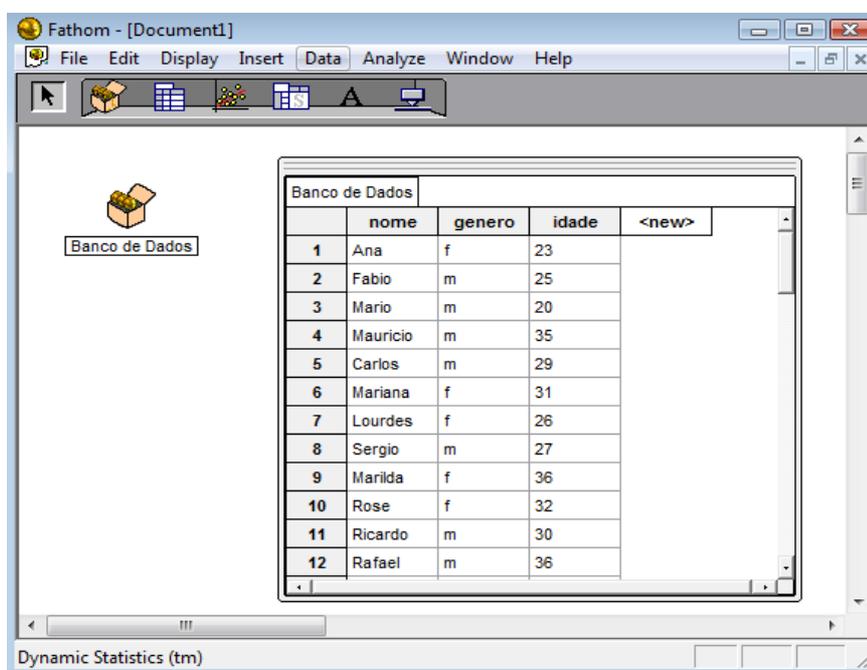


Figura 22: Tela do Fathom – Inserindo um banco de dados

Ao inserir o banco de dados, o usuário passa a ter acesso a diferentes representações, já que pode construir gráficos e tabelas à partir deste banco de dados. Por exemplo, para construir o gráfico da variável gênero, basta clicar no ícone “New Graph” e abrir a janela na qual o novo gráfico será construído. Em seguida, basta clicar na coluna da variável e arrastar até o gráfico.

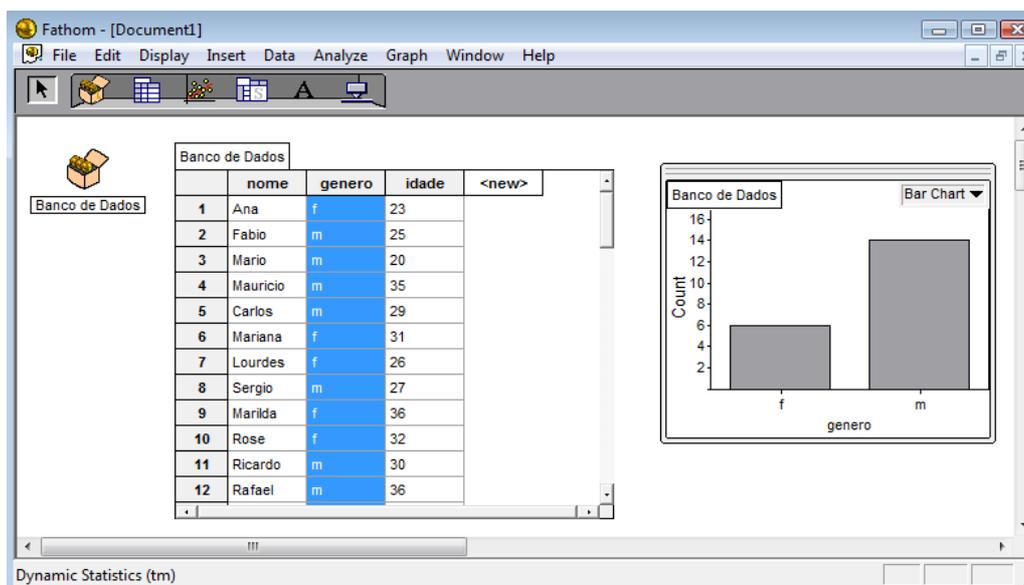


Figura 23: Tela do Fathom – Gráficos de variáveis qualitativas

Existem vários tipos de gráficos disponíveis, que o usuário pode mudar apenas clicando na seta no canto superior direito do gráfico. Observe alguns dos diferentes tipos de gráficos disponíveis para a variável idade:

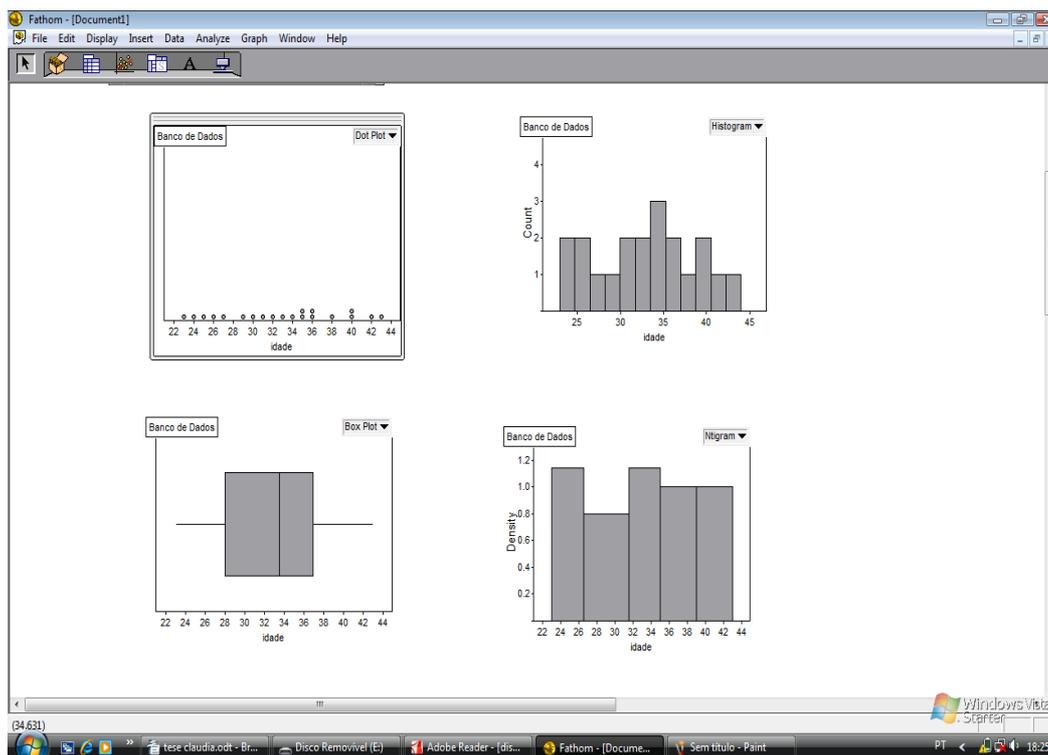


Figura 24: Tela do Fathom – Gráficos de variáveis quantitativas

Vale lembrar que nem todos os tipos de gráficos estão disponíveis no Fathom. Por exemplo, o software não faz o diagrama circular (gráfico de setores) ou o gráfico de barras horizontais.

O software possui recursos que permitem calcular as medidas-resumo de uma distribuição de freqüências e visualizá-las no gráfico. Para isso, é necessário inicialmente selecionar o gráfico. Em seguida, basta clicar no menu “Graph”, que será aberta uma janela com várias opções. Para a média aritmética, clica-se em “Plot Value”, depois em “Functions”, na seqüência em “Statistical”, e finalmente em “mean”. Para a mediana, basta seguir o mesmo procedimento, e no final clicar em “median”. Este mesmo menu dá acesso a uma infinidade de outras funções.

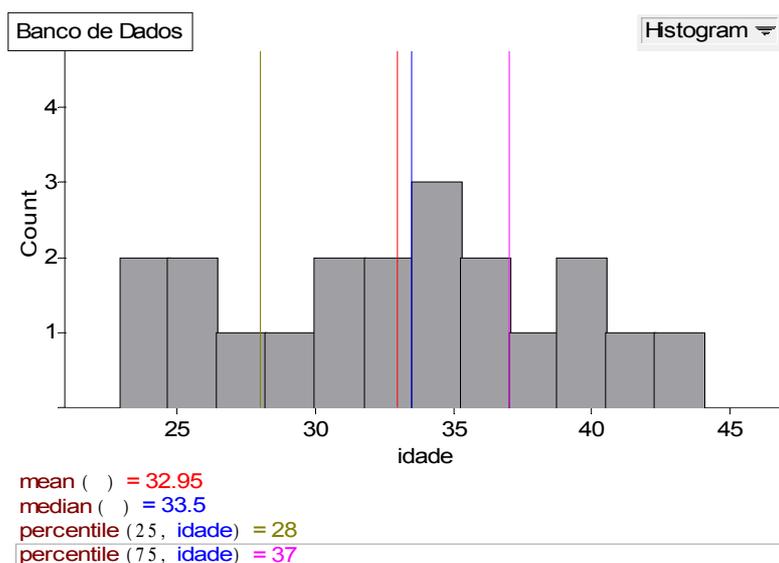


Figura 25: Tela do Fathom – Representação gráfica de medidas-resumo

Uma ferramenta interessante do software é a possibilidade de selecionar um determinado atributo e visualizar tudo o que está relacionado a ele. Por exemplo, ao clicar na coluna referente ao gênero feminino no gráfico relativo a essa variável, o software destaca todos os dados relacionados a este gênero nos outros gráficos e também no banco de dados.

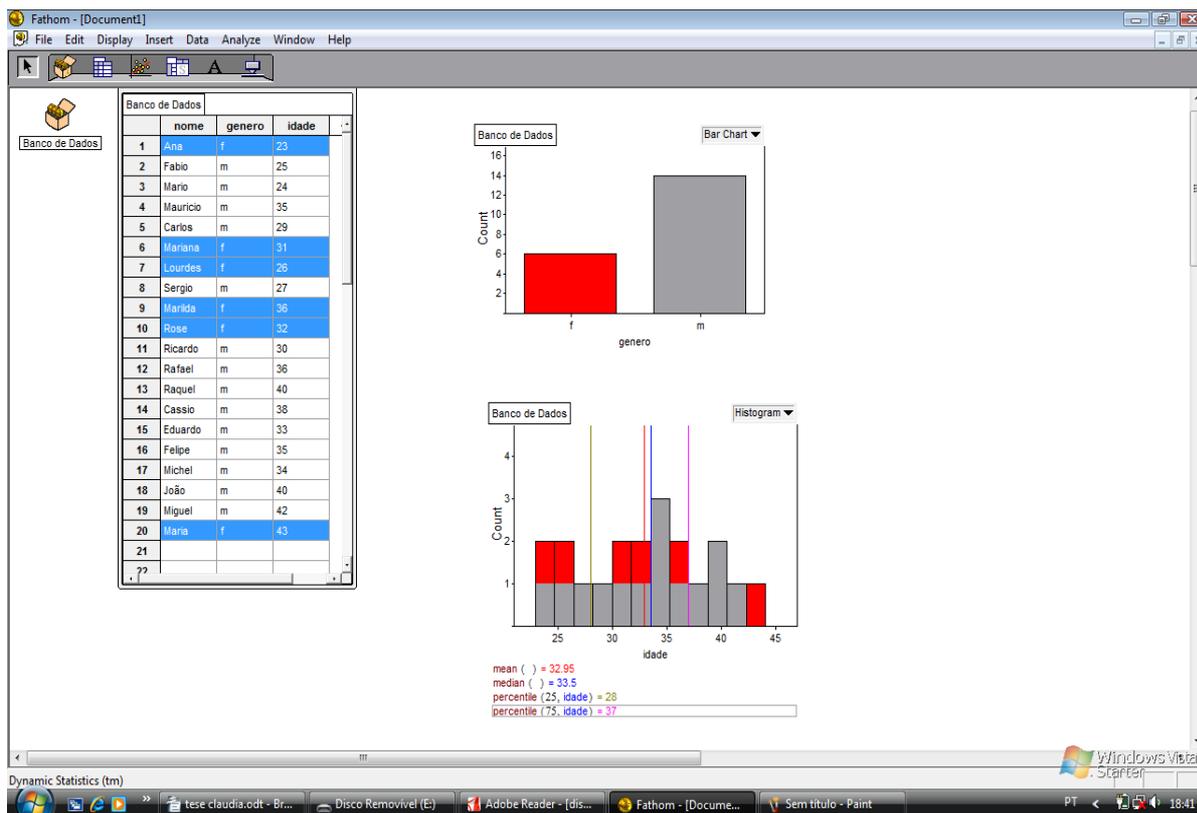


Figura 26: Tela do Fathom

Esta possibilidade de utilização de representações simultâneas proporcionada pelo Fathom é o foco principal da utilização deste software em nosso trabalho.

Destacamos, neste capítulo, algumas funções do software Fathom, mas lembramos que existem inúmeras outras funções para o trabalho com estatística e probabilidade que são de grande potencial para o ensino e pesquisa na área de educação estatística. Algumas dessas funções serão mais detalhadas em nossa seqüência didática.

3.5 Objeto de Estudo

Batanero (2001) em um estudo sobre história da estatística, descreve que as suas origens são muito antigas, e que se tem encontrado provas de coletas de dados sobre população, bens e produção em civilizações como a China (aproximadamente 1000 anos a.C.), Suméria e Egito. Segundo a autora, inclusive

na Bíblia, no livro de Números, aparecem referências ao recrutamento dos israelitas em idade de serviço militar. Porém só recentemente, a estatística tem adquirido a categoria de ciência. Aos poucos foram criadas sociedades e oficinas estatísticas para organizar a coleta de dados estatísticos (a primeira delas foi na França, em 1800). Como conseqüência, foi possível comparar as estatísticas de cada país em relação aos demais, para determinar os fatores determinantes do crescimento econômico, e começaram os congressos internacionais, com o objetivo de homogeneizar os métodos utilizados (o primeiro foi organizado por Quetelet em Bruxelas, 1853). Em 1885, nasceu o Instituto Internacional de Estatística (ISI), cuja finalidade específica é conseguir uniformidade nos métodos de coletas de dados e abstração de resultados e enviar aos governos o uso correto da estatística na solução de problemas políticos e sociais.

Para Cabriá (1994, apud BATANERO 2001, p. 9):

A estatística estuda o comportamento dos fenômenos chamados coletivos. Está caracterizada por uma informação sobre um coletivo ou universo, o que constitui seu objeto material; um modo próprio de raciocínio, o método estatístico, o que constitui seu objeto formal e certas previsões sobre o futuro, o que implica um ambiente de incerteza, que constituem seu objeto ou causa final.

Magalhães e Lima (2005, p. 1) entendem a Estatística como:

...um conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados em qualquer área do conhecimento.

Baseando-se no posicionamento destes autores (CABRIÁ, MAGALHÃES e LIMA), podemos então definir a estatística como o *“estudo do comportamento dos fenômenos de um determinado grupo, a partir de um conjunto de técnicas que permitem organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados nas mais diversas áreas do conhecimento”*.

Batanero (2001), ainda reforça a importância que a estatística tem alcançado, tanto como cultura básica, como no trabalho profissional e na investigação, devido à quantidade de informações com que o cidadão depara-se diariamente, já que na maioria das vezes estas informações vêm expressas na

forma de tabelas e gráficos estatísticos, sendo necessário um conhecimento básico desta ciência para sua correta interpretação.

A estatística pode ser dividida em três áreas: *Estatística Descritiva*; *Probabilidade*; *Inferência Estatística*. A *estatística descritiva* pode ser definida como um conjunto de técnicas destinadas a descrever e resumir os dados, a fim de que possamos tirar conclusões a respeito de características de interesse. A *probabilidade* pode ser pensada como a teoria matemática utilizada para se estudar a incerteza oriunda de fenômenos de caráter aleatório. A *inferência estatística* é o estudo de técnicas que possibilitam a extrapolação a um grande conjunto de dados, de informações e de conclusões obtidas a partir de subconjunto de valores, usualmente de dimensão muito menor.

Para “tratar” os dados de um experimento estatístico com o objetivo de extrair informações de características importantes, faz-se uso de distribuição de freqüências em forma de tabelas e gráficos (lembrando que tal procedimento deve levar em conta a natureza dos dados).

A tabela de distribuição de freqüências contém informações resumidas para a variável. Esta tabela conterá as categorias, ou seja, os valores assumidos pela variável e o número de observações (que são denominadas freqüências absolutas). Pode-se também acrescentar nesta tabela as informações sobre a freqüência relativa, que é o quociente entre a freqüência absoluta e a quantidade total de observações da variável, e pode ser expresso em forma decimal ou porcentagem. Por exemplo, a tabela a seguir mostra a distribuição de freqüências do gênero de um grupo de estudantes.

TABELA 6 – DISTRIBUIÇÃO DOS ESTUDANTES DE ENSINO MÉDIO, POR GÊNERO

GÊNERO	N° ALUNOS	
	N°	%
Feminino	25	0,45
Masculino	30	0,55
TOTAL	55	1

A organização dos dados em tabelas de distribuição de freqüência proporciona um meio eficaz de estudo do comportamento de características de interesse. Mas, muitas vezes, a informação contida nas tabelas pode ser mais

facilmente visualizada por gráficos, devido ao baixo custo cognitivo da apreensão perceptiva.

O gráfico estatístico é um tipo de representação cujo objetivo é produzir uma impressão mais rápida e viva do fenômeno em estudo. Segundo Magalhães e Lima (2005, p.11), os meios de comunicação apresentam diariamente os mais variados tipos de gráficos para auxiliar na apresentação de informações, assim como os órgãos públicos e empresas utilizam cada vez mais os gráficos em documentos internos e relatórios de atividades desempenho.

Magalhães e Lima (2005) ressaltam que a utilização de recursos visuais na criação de gráficos deve ser feita cuidadosamente, pois um gráfico desproporcional em suas medidas pode dar falsa impressão de desempenho e conduzir a conclusões equivocadas, e que o uso de divulgação ética e criteriosa de dados devem ser pré-requisitos indispensáveis e inegociáveis.

Segundo Chambers (1983 apud BUSSAB e MORETTIN, 2002, p.3), os gráficos podem ser utilizados para diversos fins:

- Buscar padrões e relações;
- Confirmar (ou não) certas expectativas que se tinham sobre os dados;
- Descobrir novos fenômenos;
- Confirmar (ou não) suposições feitas sobre os procedimentos estatísticos estudados;
- Apresentar resultados de modo mais rápido e fácil.

Existe uma grande diversidade de tipos de gráficos que podem ser utilizados, e com o avanço tecnológico, a construção de gráficos em programas computacionais tem sido cada vez mais simplificada. Para Costa (1992), os melhores gráficos são aqueles que primam pela simplicidade e clareza. Os gráficos mais usuais para variáveis qualitativas e discretas são os gráficos de barras, colunas e de setores. No caso de variáveis agrupadas, os mais usuais são o histograma e o polígono de frequências. Apresentamos na seqüência alguns tipos de gráficos mais utilizados.

O **gráfico de setores** (ou diagrama circular, ou disco), consiste em repartir um disco em setores circulares correspondentes às porcentagens de cada valor assumido pela variável em questão. Este tipo de gráfico adapta-se muito bem às variáveis qualitativas. É empregado sempre que desejamos ressaltar a participação do dado no total (MAGALHÃES e LIMA, 2005, p.13).

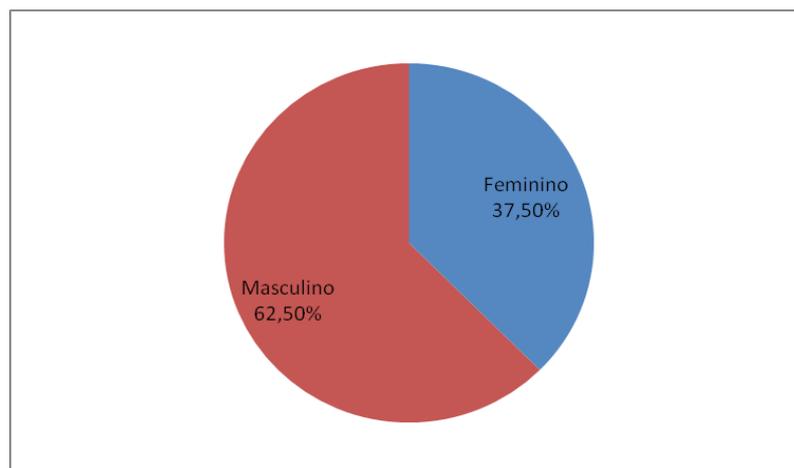


Figura 27: Exemplo de um gráfico de setores da variável gênero de um grupo de estudantes.

O **gráfico de colunas** utiliza o plano cartesiano com os valores da variável no eixo das abscissas e as freqüências no eixo ordenadas; e para cada valor da variável, desenha-se uma barra com altura correspondendo à sua freqüência. O **gráfico de barras** utiliza o plano cartesiano com os valores da variável no eixo das ordenadas e as freqüências no eixo das abscissas, e para cada valor da variável desenha-se uma barra com comprimento, correspondendo à sua freqüência. Estes tipos de gráficos adaptam-se melhor às variáveis discretas ou qualitativas (MAGALHÃES e LIMA, 2005, p.13).

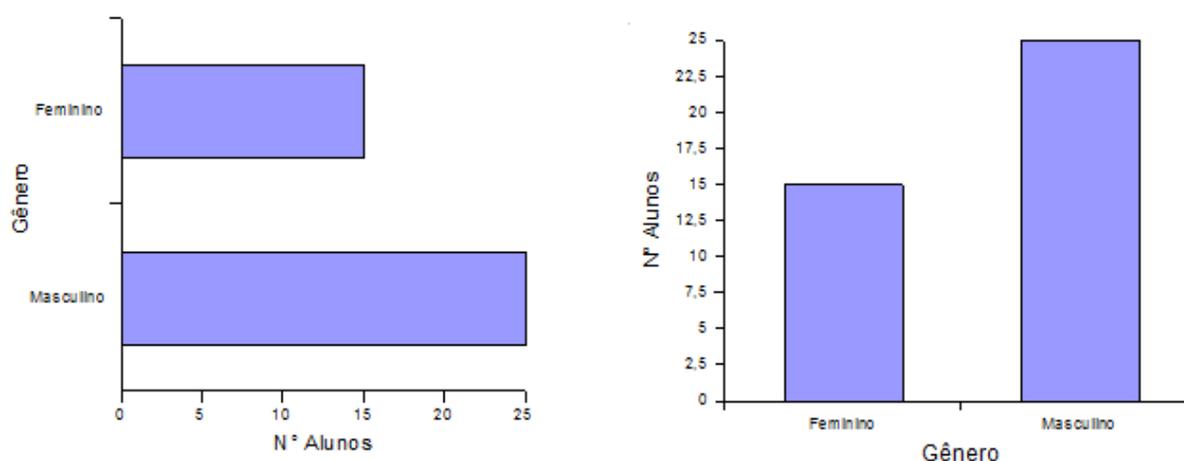


Figura 28: Gráfico de barras e gráfico de colunas para a variável gênero

O **histograma** consiste em retângulos justapostos com base nas faixas de valores da variável e com área igual à freqüência relativa da respectiva faixa

(MAGALHÃES e LIMA, 2005, p.14).

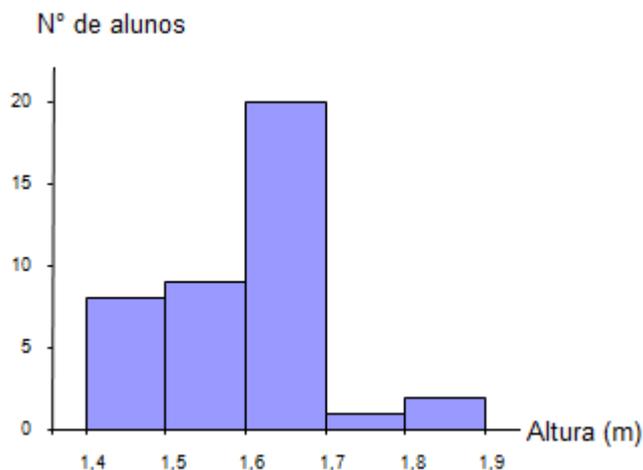


Figura 29: Exemplo de um gráfico histograma da variável altura de um grupo de estudantes.

Este tipo de gráfico pode ser utilizado no cálculo da mediana, já que a área de um histograma é proporcional à soma das frequências.

O **ramo e folhas** é uma representação gráfica alternativa para resumir um conjunto de valores, que, segundo Bussab e Morettin (2002, p. 20), tem o objetivo de se obter uma idéia da forma de sua distribuição, com a vantagem de não perder (ou perder pouca) informação sobre os dados em si. Ainda segundo estes autores, não existe uma regra fixa para a construção deste tipo de gráfico, mas a idéia básica é dividir cada observação em duas partes: a primeira (o ramo) é colocada à esquerda de uma linha vertical, e a segunda (a folha) é colocada à direita. Uma das principais vantagens deste tipo de gráfico é que ele permite estudar a distribuição dos dados, entre outros, e é acessível desde as primeiras séries de escolaridade, pois basta que o aluno conheça o valor posicional no sistema de numeração decimal.

Por exemplo, na figura 30 (BUSSAB e MORETTIN, 2002), foi construído o ramo-e-folhas dos salários de 36 empregados de uma companhia. Para os salários 4,00 e 4,56, o 4 é o ramo e 00 e 56 são as folhas, respectivamente.

4	00 56
5	25 73
6	26 66 86
7	39 44 59
8	12 46 74 95
9	13 35 77 80
10	53 76
11	06 59
12	00 79
13	23 60 85
14	69 71
15	99
16	22 61
17	26
18	75
19	40
20	
21	
22	
23	30

Figura 30: Ramo-e-folhas para a variável salários.

Nesse gráfico, os espaços em branco significam a ausência de folhas.

O **Dot-Plot** é um gráfico de pontos, que representa cada observação obtida em uma escala horizontal, permitindo visualizar a representação dos dados ao longo desta.

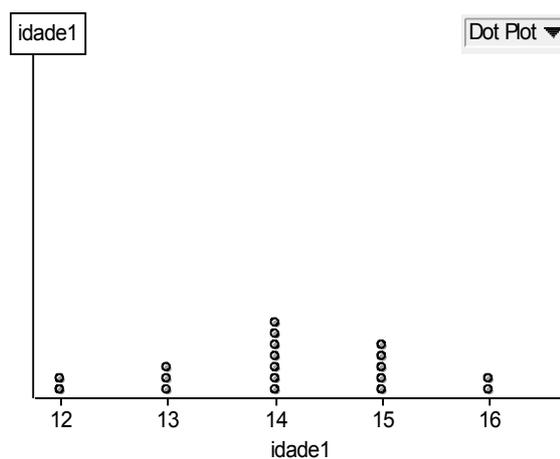


Figura 31: Exemplo de um gráfico Dot-Plot para a variável idade de um grupo de estudantes.

O **Box Plot** (ou gráfico de cinco pontos) é uma representação gráfica

envolvendo os quartis, a mediana, o valor máximo e o valor mínimo. Para construir o box-plot, definimos uma caixa com o nível superior dado pelo terceiro quartil e o nível inferior pelo primeiro quartil (ou o nível direito dado pelo terceiro quartil e o nível esquerdo pelo primeiro quartil). A mediana é representada por um traço no interior da caixa. Segmentos de reta são colocados da caixa até os valores máximo e mínimo (MAGALHÃES e LIMA, 2005, p.17).

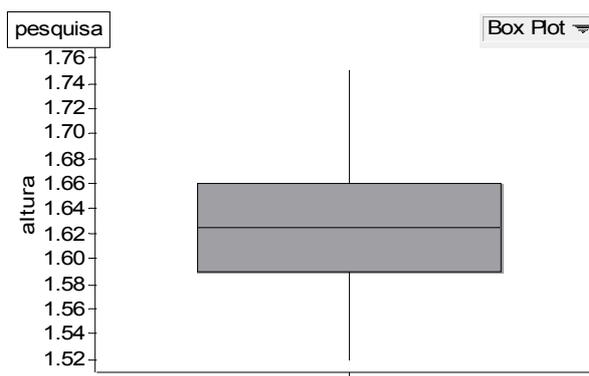


Figura 32: Exemplo de um gráfico Box Plot da variável altura de um grupo de estudantes.

Este tipo de gráfico é uma representação muito útil no sentido de informar a variabilidade e simetria dos dados, e também para detectar diferenças nos comportamentos de grupos de variáveis.

Para representar um conjunto de dados, utiliza-se também medidas que resumem os dados e descrevem suas características, permitindo assim a análise da variação desses dados, que são as **medidas-resumo**. Estas medidas também constituem representações semióticas do conjunto de dados em questão.

A **média aritmética** é o ponto de equilíbrio dos valores da distribuição. É calculada como a soma das observações dividida pelo número de observações (BUSSAB e MORETTIN, 2002). Vale lembrar que a média é um valor representativo de um conjunto de dados, não precisando ser necessariamente um dos valores observados. Ela ainda leva em conta todos os valores observados, podendo ser, portanto, influenciada por valores discrepantes.

Por exemplo, vamos considerar a tabela a seguir com as alturas de um grupo de estudantes.

**TABELA 7: DISTRIBUIÇÃO DOS ESTUDANTES,
SEGUNDO SUAS ALTURAS (em metros)**

Altura	Nº de alunos
1,52	1
1,54	1
1,55	1
1,56	1
1,58	3
1,59	4
1,6	4
1,62	5
1,63	4
1,64	1
1,65	4
1,66	2
1,68	2
1,69	1
1,7	1
1,71	2
1,73	1
1,75	2
Total	40

Calculando a média, temos:

$$Média = \frac{1,52 + 1,54 + 1,55 + 1,56 + 1,58 \times 3 + 1,59 \times 4 + 1,60 \times 4 + 1,62 \times 5 + 1,63 \times 4 + 1,64 + 1,65 \times 4 + 1,66 \times 2 + 1,68 \times 2 + 1,69 + 1,70 + 1,71 \times 2 + 1,73 + 1,75 \times 2}{40}$$

$$Média = 1,63125m$$

Considerando este exemplo, concluímos então que a altura média destes estudantes é 1,63125m, ou seja, as alturas observadas se distribuem em torno de 1,63125m. Porém, lembramos este valor não nos informa sobre a variação dos dados, ou seja, sobre o grau de concentração ou dispersão em torno da média (variabilidade). Para isso, deve-se associar outras medidas, tais como a amplitude e/ou o desvio padrão.

A **amplitude**, referente a certa variável, é definida como a diferença entre o maior e o menor valor do conjunto de dados (MAGALHÃES e LIMA, 2005, p.103).

Por exemplo, a amplitude das alturas da tabela 6 é: 1,75m – 1,52m =

0,23m.

O **desvio em relação à média** é a diferença entre cada valor observado e a média. A soma desses desvios sempre dá zero, logo a média deles também é zero. Dessa maneira, essa medida não seria informativa sobre a dispersão das observações. Portanto, para obter o desvio médio, utilizamos o valor absoluto dessas diferenças, que é definido como a média dos valores absolutos em relação à média (MAGALHÃES e LIMA, 2005, p.104).

Para o cálculo da **variância**, consideramos a mesma média dos desvios, elevando ao quadrado cada uma de suas parcelas (para eliminar a possibilidade dessa soma ser nula). O denominador é o número de observações subtraído de uma unidade.

O **desvio-padrão** é definido como a raiz quadrada do valor encontrado na variância.

A **mediana** é o valor que ocupa a posição central dos dados ordenados (MAGALHÃES e LIMA, 2005, p.94), ou seja, é o valor que é maior ou igual a 50% dos dados ordenados e menor ou igual a 50% dos dados. É uma medida de posição, que pode ser utilizada para uma melhor visualização da variação de um conjunto de dados quando associada a outras medidas, tais como quartis, uma vez que não é influenciada por valores extremos.

Quando a quantidade de dados de um conjunto é ímpar, a mediana é um valor do próprio conjunto, ou seja, é o valor central dos dados ordenados. Quando a quantidade de termos for par, a mediana é a média aritmética dos dois valores centrais.

Por exemplo, considerando os dados da tabela 6, para obter a mediana, temos:

Valor da 20ª posição: 1,62m

Valor da 21ª posição: 1,63m

$$Mediana = \frac{1,62 + 1,63}{2} = 1,625$$

Os **quartis** são separatrizes que dividem a área de uma distribuição de

freqüências em regiões de áreas iguais e múltiplos de $\frac{1}{4}$ da área total (MORETTIN, 2000, p.16).

O **Primeiro Quartil** é o valor (separatriz) que divide a distribuição em duas partes, tal que 25% dos valores sejam menores que ele, e 75% dos valores sejam maiores que ele (MORETTIN, 2000, p.16). Por exemplo, considerando a tabela 6, o primeiro quartil seria o valor da altura entre a 10^o e a 11^o posição, que no caso é 1,59m.

O **Segundo Quartil** coincide com a mediana da distribuição. (MORETTIN, 2000, p.16).

O **Terceiro Quartil** é o valor que divide a distribuição em duas partes, tal que 75% dos valores sejam menores que ele, e 25% dos valores sejam maiores que ele (MORETTIN, 2000, p.16). Por exemplo, no caso da tabela 6, o terceiro quartil seria o valor da altura entre a 30^o e a 31^o posição, que é 1,66m.

Os quartis podem ser representados na forma gráfica, como nos exemplos a seguir, localizando-os no histograma e no box-plot:

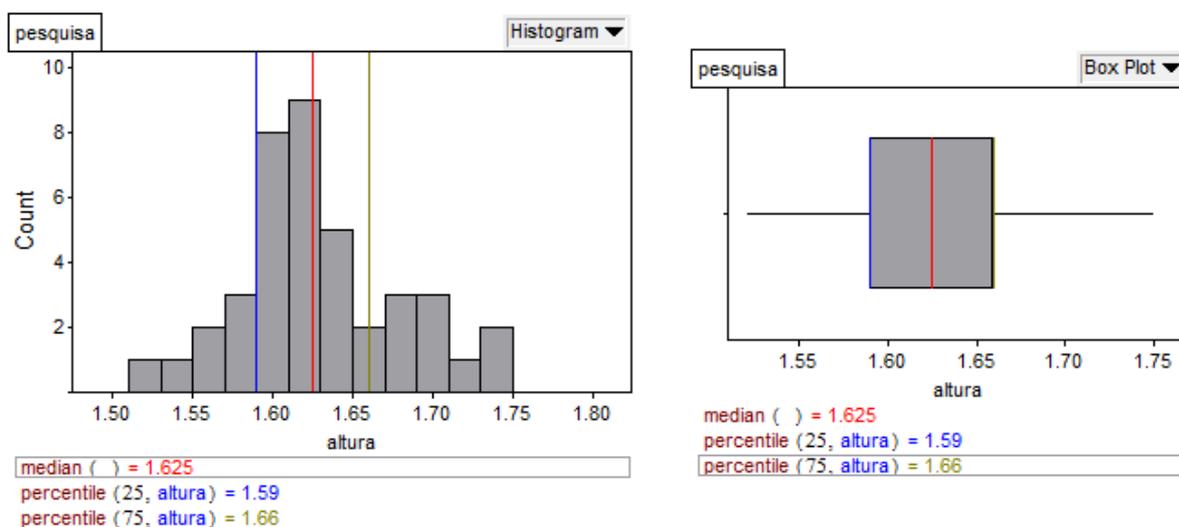


Figura 33: Representação dos quartis da variável altura no Histograma e no Box-Plot

A **moda** é o valor mais freqüente de uma distribuição de freqüências, ou seja, é o valor da variável que corresponde ao valor máximo na distribuição de freqüência (MORETTIN, 2000).

Por exemplo, considerando a tabela 6, a moda é o valor que mais se repete, ou seja, a altura 1,62m.

3.6 Parâmetros Curriculares Nacionais

Os Parâmetros Curriculares Nacionais para o Ensino Médio indicam algumas competências que a escola deve desenvolver com o objetivo de preparar o aluno para um aprendizado permanente. Algumas dessas competências relacionadas com o tema deste trabalho (BRASIL, 2002, p. 111) são:

- Ler e interpretar dados ou informações apresentados em diferentes linguagens e representações, como tabelas, gráficos, esquemas, diagramas, árvore de possibilidades, fórmulas, equações ou representações geométricas;
- Traduzir uma situação dada em determinada linguagem para outra, por exemplo, transformar situações dadas em linguagem discursiva em esquemas, tabelas, gráficos, entre outras;
- Selecionar diferentes formas para representar um dado ou conjunto de dados e informações, reconhecendo as vantagens e limites de cada uma delas.

Essas competências relacionadas nos Parâmetros Curriculares Nacionais (PCN) vêm ao encontro do nosso trabalho ao enfatizar a leitura e interpretação de informações em diferentes linguagens e representações. Quanto ao segundo item, que fala em traduzir as situações dadas para outras linguagens, acreditamos que está de acordo com nosso referencial teórico, que seriam, no caso, as transformações de representações semióticas, de Duval. O terceiro item, que trata do reconhecimento das vantagens e limites de cada tipo de representação, também se relaciona com as atividades de nossa seqüência didática.

Segundo os PCN, o ensino da matemática está dividido em três temas: Álgebra, Números e Funções; Geometria e Medidas; Análise de Dados (BRASIL, 2002, p. 117). O ensino de estatística é discutido no terceiro tema, "Análise de dados". Este tema estruturador permite o desenvolvimento de várias competências relativas à contextualização sócio-cultural, como a análise de situações reais presentes no mundo contemporâneo e a articulação com diversas áreas do conhecimento, contribuindo para compreensão e uso de representações gráficas, identificação de regularidades, interpretação e uso de modelos matemáticos e conhecimento de formas específicas de raciocinar em matemática. Nesse contexto, o computador ganha importância como instrumento que permite

a abordagem de problemas com dados reais ao mesmo tempo em que o aluno pode ter a oportunidade de se familiarizar com as máquinas e os softwares. Acreditamos também que essas competências descritas nos PCN, estão de acordo com o enfoque dado na Análise Exploratória de Dados.

Os conteúdos e habilidades propostos para as unidades temáticas a serem desenvolvidos nesse tema (BRASIL, 2002, p.124) seriam:

Estatística: descrição de dados; representações gráficas; análise de dados: médias, moda e mediana, variância e desvio padrão.

- Identificar formas adequadas para descrever e representar dados numéricos e informações de natureza social, econômica, política, científico - tecnológica ou abstrata;
- Ler e interpretar dados e informações de caráter estatístico apresentados em diferentes linguagens e representações, na mídia ou em outros textos e meios de comunicação.
- Obter médias e avaliar desvios de conjuntos de dados ou informações de diferentes naturezas.
- Compreender e emitir juízos sobre informações estatísticas de natureza social, econômica, política ou científica apresentadas em textos, notícias, propagandas, censos, pesquisas e outros meios.

Ainda segundo os Parâmetros Curriculares Nacionais (BRASIL, 2002, p.130), um elemento importante da comunicação em matemática é a multiplicidade de formas textuais a que os alunos devem ser expostos. Gráficos, tabelas, esquemas, desenhos, fórmulas, textos jornalísticos, mapas são, na escola e fora dela, as diferentes linguagens e representações que o aluno deve compreender para argumentar e posicionar-se frente a novas informações.

Essa abordagem proposta pelos Parâmetros Curriculares Nacionais explora os mesmos fundamentos que acreditamos de grande importância para a aprendizagem significativa dos conceitos estatísticos e que estão em acordo com nosso referencial teórico.

No trabalho em ambiente computacional, o software Fathom é uma ferramenta que pode contribuir para a compreensão e interpretação dos conceitos estatísticos, nesse mesmo enfoque que tratam os PCN, por permitir acesso rápido a diferentes tipos de representação, possibilitando a coordenação simultânea de dados em tabelas e gráficos, e o acesso a diversas funções que permitem uma maior flexibilidade em uma análise de dados.

Capítulo IV: Seqüência Didática

4.1 Descrição da Parte Experimental da pesquisa

A parte experimental de nossa pesquisa é composta por um teste-diagnóstico e pela seqüência didática realizada com o uso do software Fathom.

O objetivo do teste-diagnóstico é verificar os conhecimentos prévios destes alunos em relação aos conceitos estatísticos. O teste foi composto por quatro questões e teve duração de duas aulas de 45 minutos.

A seqüência didática é composta por duas partes, A e B, com duração prevista de seis aulas de 45 minutos cada, sendo três aulas em cada dia.

Na parte A da seqüência, foram realizadas as “*atividades de familiarização*”, em que o aluno entra em contato com o software Fathom e desenvolve as atividades propostas a partir de um banco de dados que ele insere no software.

Na parte B da seqüência, o aluno é solicitado a preparar um relatório composto por tabelas, gráficos e textos explicativos, visando a determinar as principais características de um determinado grupo de estudantes.

A experimentação desta pesquisa foi realizada entre os meses de maio e junho de 2008, com alunos do segundo ano do Ensino Médio do período noturno, de uma escola estadual localizada no município de Santo André, em São Paulo.

Para selecionar os alunos sujeitos da pesquisa, conversamos com duas turmas do segundo ano do Ensino Médio desta escola, informando como seriam realizadas as atividades referentes à pesquisa, e convidando os alunos que estivessem interessados em participar como voluntários. Tínhamos interesse em selecionar oito alunos, organizados em quatro duplas. No entanto dez alunos se prontificaram à participar, e organizaram-se em grupos, da seguinte forma:

- Grupo 1: Duas alunas e um aluno.
- Grupo 2: Uma aluna e um aluno.
- Grupo 3: Duas alunas e um aluno.
- Grupo 4: Uma aluna e um aluno.

Os quatros grupos descritos realizaram o teste diagnóstico no mês de maio de 2008.

Marcamos o desenvolvimento das atividades da seqüência didática para a semana seguinte. Porém houve um pequeno atraso neste início, devido ao não funcionamento do laboratório de informática da escola.

Optamos então em aplicar as atividades utilizando três computadores pessoais (sendo um da pesquisadora, e os outros cedidos por professores da escola). Sendo assim, instalamos o software Fathom (licença emprestada pela PUC/SP) nestes computadores apenas para realizar a seqüência didática, e desinstalamos o software ao final dela.

O início da primeira parte (parte A) da seqüência didática foi realizado então em junho de 2008. Os alunos de um dos grupos (grupo 4) não compareceram à escola nesta semana (acreditamos que devido ao período de greve dos professores das escolas estaduais), não realizando as atividades da seqüência didática. Portanto continuamos a pesquisa somente com os três primeiros grupos, ou seja, com oito alunos, que continuaram até o fim das atividades.

Os alunos desenvolveram as atividades da parte A e, no final da aula, foi-lhes entregue a parte B, para que os alunos pudessem iniciar a coleta de dados. Então, no final do mês de junho de 2008, os alunos já com os dados coletados, desenvolveram a parte B da seqüência didática. Todas as seções foram áudio-gravadas e acompanhadas por observador.

4.2 Teste-Diagnóstico: Análises e Resultados

O objetivo do teste-diagnóstico aplicado antes de iniciar a seqüência didática foi levantar os conhecimentos prévios dos alunos em relação aos conceitos estatísticos. Identificando qual o conhecimento do aluno em relação à estatística, poderíamos, ao final da aplicação da seqüência didática, apontar quais foram os avanços, significativos ou não em relação à aprendizagem destes conceitos, alcançados durante o desenvolvimento deste trabalho. Ressaltamos que este teste teve apenas o caráter de verificação, e que, portanto, não foi

analisado de acordo com nossos pressupostos teóricos.

O teste foi composto por quatro questões, nas quais pretendíamos verificar se, e como os alunos:

- Realizam a leitura de dados em tabelas e gráficos;
- Identificam e diferenciam as variáveis estatísticas de uma pesquisa;
- Conseguem converter uma distribuição de frequências representada em tabela para o gráfico, e em gráfico para a tabela;
- Conhecem, e sabem calcular as medidas-resumo de uma distribuição de frequências (média aritmética, mediana, moda, quartis).

Para resolver os problemas propostos neste teste, eram necessários alguns conhecimentos prévios, como: leitura, análise e construção de tabelas e gráficos; operações básicas, como adição, subtração, multiplicação, divisão e porcentagem; cálculo de medidas resumo, como a média aritmética.

Nesta situação, havia algumas variáveis didáticas em jogo. O trabalho em grupo tinha o objetivo de provocar discussões significativas sobre os conceitos e os procedimentos envolvidos nos problemas propostos entre os alunos sujeitos da pesquisa. Os problemas propostos tinham como finalidade identificar se os alunos conseguiam construir, ler e interpretar tabelas e gráficos e obter os valores de medidas resumos, além de converter dados expressos em porcentagem, converter os dados da tabela para o gráfico, do gráfico para a tabela e do rol de dados para o gráfico.

4.2.1 Análises e resultados da primeira questão.

1) Em certa empresa, a distribuição do número de funcionários (segundo seu gênero), para os diversos cargos e seus respectivos salários é a seguinte:

Cargo	Gênero	Nº de funcionários	Salário (em R\$)
Gerente	Masculino	1	3500
Auxiliar Administrativo	Masculino	3	820
Auxiliar Administrativo	Feminino	5	820
Vendedores	Masculino	12	540
Vendedores	Feminino	10	540
Serventes	Masculino	3	400
Serventes	Feminino	3	400

- a) Quantos funcionários têm a empresa?
- b) Quantos funcionários são do gênero feminino?
- c) Qual a porcentagem de funcionários do gênero masculino?
- d) Quais são as variáveis desta pesquisa?
- e) Você considera que as variáveis da pesquisa são de mesmo tipo ou são de tipos diferentes? Justifique.
- f) Qual a média salarial dos funcionários desta empresa?
- g) Se colocarmos esses funcionários em ordem crescente de salários, qual seria a posição do funcionário que dividiria essa fila em dois grupos como o mesmo número de funcionários. Escreva o cargo e o salário deste funcionário.
- h) Qual o salário que você considera que melhor representa o salário pago por esta empresa? Por quê?
- i) Construa um gráfico que represente os cargos dos funcionários desta empresa.
- j) Construa um gráfico que represente os gêneros dos funcionários desta empresa.
- k) Construa um gráfico que represente os salários dos funcionários desta empresa.

Análise da Primeira Questão:

Nesta primeira questão, o aluno inicialmente deveria analisar os dados da tabela para responder as questões.

Nos itens (a) e (b), o aluno levantaria os dados explícitos na tabela, e somaria as quantidades. Como resultados, esperávamos os seguintes valores:

- a) 37 funcionários

b) 18 funcionárias

Já no item (c), além de levantar os dados da tabela, o aluno deveria convertê-los em porcentagem, identificando o total e a parte relativa a esse todo que está considerando, obtendo como resultado:

c) Aproximadamente 51,3% dos funcionários são do gênero masculino.

Nos itens (d) e (e), pretendíamos verificar se o aluno conhecia e conseguia diferenciar as variáveis estatísticas de uma pesquisa, categorizando-as quanto ao tipo (qualitativa ou quantitativa). Esperávamos como respostas:

a) As variáveis são: cargo, gênero, n° de funcionários e salário.

b) São de tipos diferentes. Existem variáveis qualitativas e variáveis quantitativas

No item (f), pretendíamos verificar se o aluno conseguia calcular a média aritmética para a variável salário, e no item (g), se o aluno tinha a noção intuitiva de mediana. Já no item (h), pretendíamos verificar qual valor o aluno utilizaria para representar o salário desta empresa. Como resultados, esperávamos:

$$f) \mu = \frac{350 + 8 \times 820 + 22 \times 540 + 6 \times 400}{37} = 657,83^6$$

g) Ordenando os dados da distribuição temos:

400 – 400 – 400 – 400 – 400 – 400 – 540 – 540 – 540 – 540 – 540 – 540 – 540 –
540 – 540 – 540 – 540 – 540 – 540 – 540 – 540 – 540 – 540 – 540 – 540 –
540 – 540 – 820 – 820 – 820 – 820 – 820 – 820 – 820 – 820 – 820 – 3500

O termo central desta distribuição é o 19º termo (termo que divide o conjunto de dados em dois subconjuntos com 18 elementos cada). O salário é de R\$ 540,00 e o cargo é de vendedor.

h) O aluno poderia utilizar algumas das medidas-resumo para representar este salário, que poderia ser a média aritmética, cujo valor é de R\$657,83, ou a mediana, cujo o valor é de R\$540,00, ou ainda a moda, cujo valor também é de R\$540,00, justificando a escolha de tal medida.

Nos itens (i), (j), (k), pretendíamos verificar se o aluno conseguia construir

⁶ Como trata-se do total de funcionários, a média é designada por μ .

gráficos estatísticos para variáveis qualitativas e quantitativas, transferindo os dados explicitados na tabela para o gráfico, realizando uma conversão de registro de representação (na perspectiva de Duval, 2003) da tabela para o gráfico. Os alunos poderiam construir gráficos de colunas, de barras, ou de setores.

Resultados da Primeira Questão:

Os alunos receberam as atividades referentes ao teste diagnóstico e resolveram no tempo previsto de duas aulas.

As primeiras questões (itens a, b, c) solicitavam apenas o levantamento de dados explícitos na tabela, e foram respondidas corretamente, mesmo quando se tratava de representar os dados na forma de porcentagem, em que foram efetuados tratamentos dos dados (na perspectiva de Duval, 2003 – Teoria de Registros de Representação Semiótica).

No item (d), que questionava quais eram as variáveis estatísticas consideradas na pesquisa, dois grupos (grupo 2 e grupo 3) responderam corretamente, porém, com base nas observações das discussões dos grupos, constatamos que os alunos não conheciam o termo “variável estatística”, e responderam corretamente possivelmente influenciados pela palavra “característica” mencionada na questão. O outro grupo (grupo 1) não conseguiu responder corretamente, tentando estabelecer alguma relação entre os dados da tabela, como mostramos no protocolo a seguir:

d) Quais são as características (variáveis estatísticas) consideradas nesta pesquisa?
A porcentagem de homens é maior do que a porcentagem de mulheres

Figura 34: Protocolo grupo 1

No item (e), que questionava se as variáveis eram de mesmo tipo ou diferentes, dois grupos (grupos 2 e 3) não responderam, e um grupo (grupo 1) respondeu incorretamente, conforme protocolo a seguir:

e) Você considera que as variáveis da pesquisa são de mesmo tipo ou são de tipos diferentes? Justifique.

São diferentes porque dependendo do cargo que se exercido o salário é diferente.

Figura 35: Protocolo grupo 1

Nestes dois itens (d, e), verificamos que os alunos tiveram dificuldades em identificar e diferenciar as variáveis de uma pesquisa, possivelmente porque não associavam o termo “variável estatística” às características relacionadas em uma pesquisa, ou seja, desconheciam o significado deste termo no contexto da estatística.

No item (f), que solicitava a média aritmética, um grupo (grupo 1) resolveu corretamente. Os outros dois grupos (grupo 2 e 3) forneceram o resultado incorreto. Na atividade, não constava como os alunos fizeram este cálculo, mas com base nas observações dos grupos, verificamos que os alunos calcularam a média do seguinte modo:

$$\frac{3500 + 820 + 820 + 540 + 540 + 400 + 400}{7} = 1002,85$$

Podemos verificar que estes dois grupos, ao calcular a média, não levaram em conta a frequência em cada salário. Com base nestes resultados, podemos concluir que os alunos conhecem o algoritmo da média aritmética simples, porém não conseguem identificar a necessidade do uso do algoritmo de média ponderada. Batanero (2000, p.7) cita que este tipo de erro é freqüente, pois “se aplica o algoritmo de forma mecânica e sem compreender seu significado”.

No item (g), pretendíamos verificar se o aluno tinha uma noção intuitiva de mediana. Apenas o grupo 1 resolveu e conseguiu achar o valor correto. Os outros dois grupos não responderam.

representasse os cargos desta empresa. Os três grupos construíram um gráfico de pontos para uma variável qualitativa, ou seja, associaram o gráfico solicitado ao gráfico de uma função, que estavam acostumados a fazer nas aulas de matemática. A seguir, mostramos gráficos construídos pelos grupos:

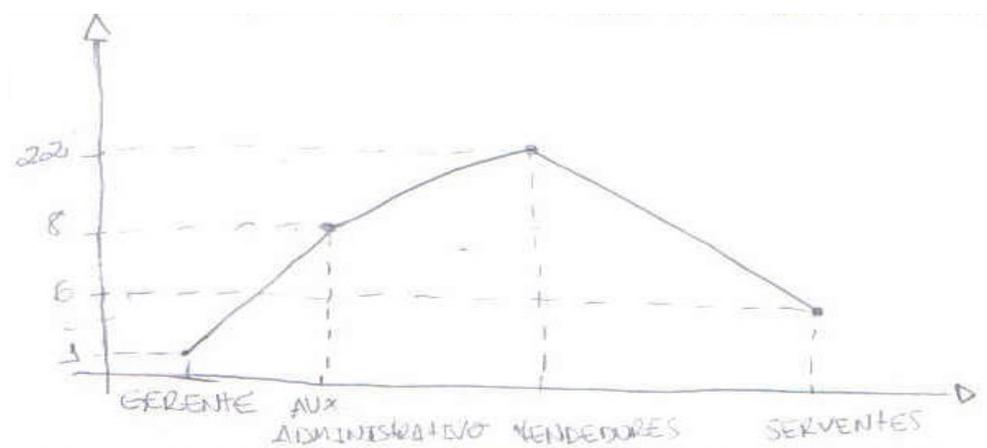


Figura 40: Protocolo grupo 2

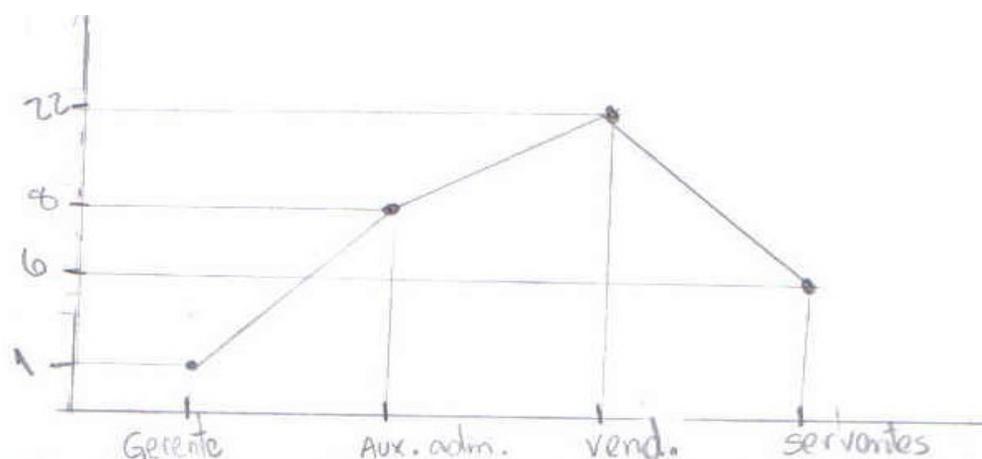


Figura 41: Protocolo grupo 3

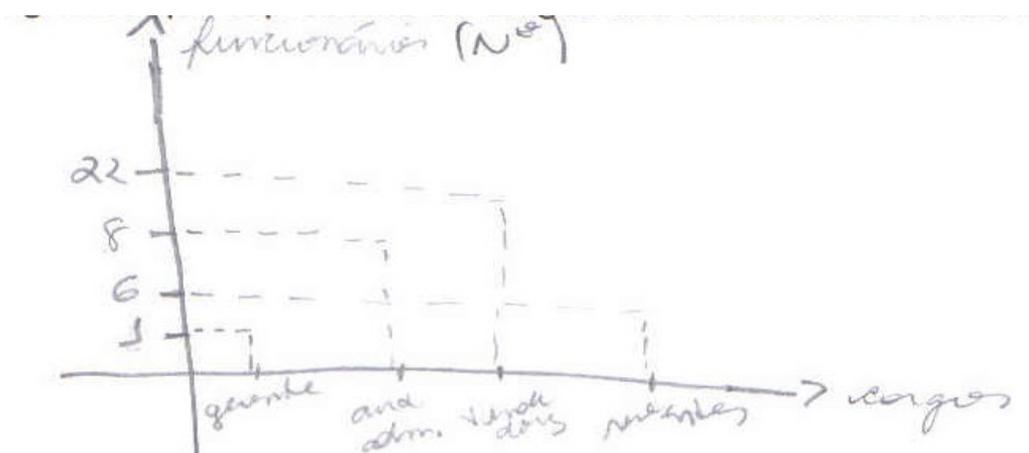


Figura 42: Protocolo grupo 1

No item (j), solicitamos um gráfico que representasse o gênero dos funcionários desta empresa. Um grupo (grupo 2) construiu um gráfico de setores para esta variável, porém os outros grupos continuaram a associar o gráfico solicitado ao gráfico de uma função, conforme verificamos nos protocolos seguintes.

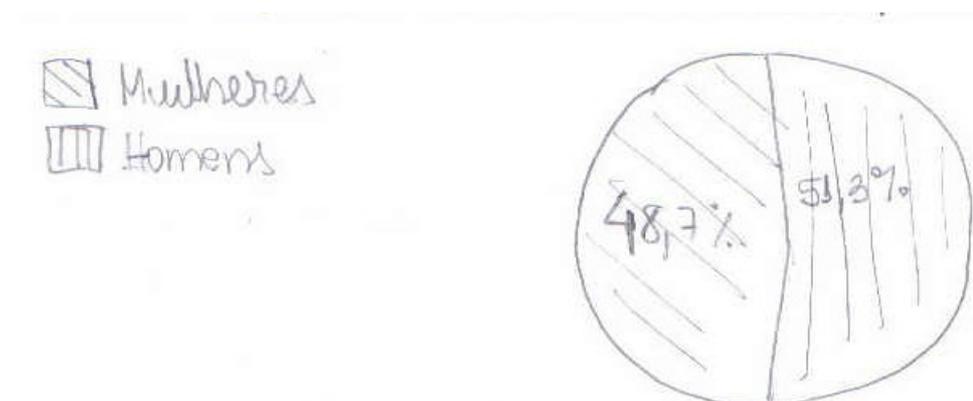


Figura 43: Protocolo grupo 2

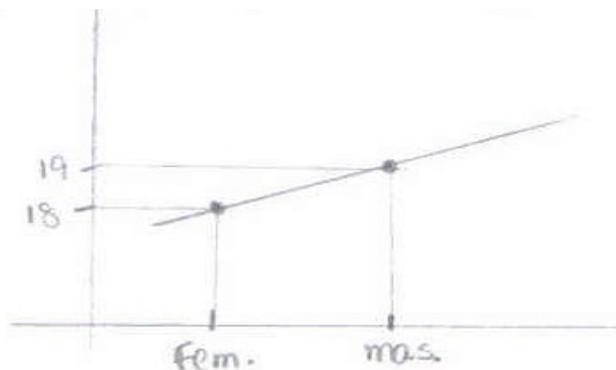


Figura 44: Protocolo grupo 3

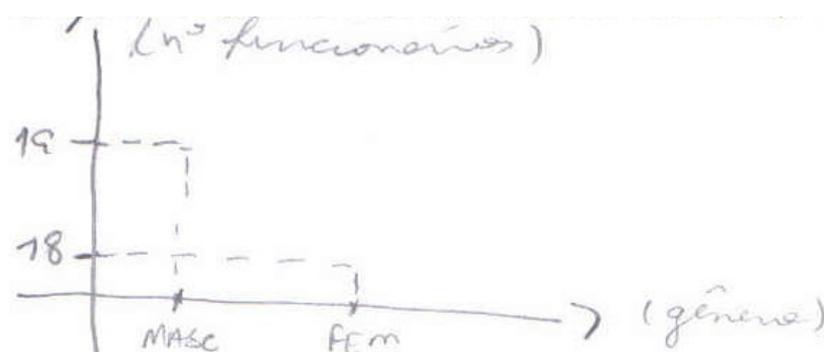


Figura 45: Protocolo grupo 1

No item (k), em que solicitamos a construção de um gráfico que representasse o salário, novamente os grupos 1 e 3 construíram gráficos de pontos, e o grupo 2 construiu um gráfico de colunas.

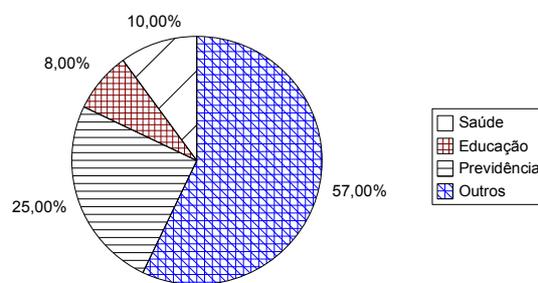
Em relação à construção de gráficos estatísticos, verificamos que, apesar dos alunos conhecerem este tipo de gráfico, eles apresentam certa tendência a construir gráficos ligando pontos, ou seja, associando o termo “gráfico” ao gráfico de uma função, possivelmente influenciados pelo ensino do conceito de função no primeiro ano do Ensino Médio. O trabalho de Li y Shen (1992, apud BATANERO, 2001), mostra exemplos de construção incorreta de gráficos em projetos estatísticos realizados por estudantes de ensino médio, como a utilização de gráficos de pontos para variáveis qualitativas e a omissão das escalas nos eixos, entre outros.

4.2.2 Análises e resultados da segunda questão.

Análise da Segunda Questão:

Na *segunda* questão, que consta um gráfico de setores, informa-se como são divididos os gastos do orçamento da União.

2) O gráfico abaixo, publicado pela revista *Veja* (1999), mostra como são divididos os 188 bilhões de reais do orçamento da União, entre os setores de saúde, educação, previdência e outros.



Construa uma tabela que mostre os valores em reais gastos em cada um dos setores.

Nesta questão, os alunos deveriam analisar o gráfico, coletando as informações expostas na forma de porcentagem, calcular o valor em reais dos gastos em cada setor e construir uma tabela que informe estes valores. Os cálculos deveriam ser realizados manualmente, pois não houve o uso de calculadoras durante as atividades. Optamos por não utilizar a calculadora, pois achamos que não havia necessidade neste caso.

Pretendíamos verificar se o aluno conseguia realizar a passagem do gráfico para a tabela, realizando novamente uma conversão de registros de representação (Duval, 2003), desta vez do figural para o numérico.

Como solução, esperávamos que os alunos construíssem uma tabela do seguinte modo:

TABELA 8: DISTRIBUIÇÃO DOS GASTOS DO ORÇAMENTO DA UNIÃO, POR SETOR

Setor	Gastos (em bilhões de reais)
Saúde	18,8
Educação	15,04
Previdência	47
Outros	107,16
Total	188

Fonte: Revista Veja (1999)

O montante a ser trabalhado pelos alunos, expresso em bilhões de reais, poderia se revelar como um complicador, mas resolvemos manter o valor devido ao nível de escolaridade dos alunos e para manter a veracidade do contexto.

Resultados da Segunda Questão:

Analisando os resultados do teste, observamos que, nesta questão, apenas um grupo (grupo 1) teve dificuldade para calcular os valores, enquanto os outros dois grupos calcularam corretamente. Quanto à apresentação da tabela, os grupos 2 e 3, esqueceram-se de acrescentar o título e a fonte dos dados, enquanto que o grupo 1 também omitiu estes itens, além dos títulos das colunas e das unidades em que estão sendo informados os dados (no caso, em bilhões de reais).

SETORES	R\$ (em bilhões)
EDUCAÇÃO	15,04
PREVIDÊNCIA	47
SAÚDE	18,8
OUTROS	107,16

Figura 46: Protocolo grupo 2

Setores	Billões em R\$
Saúde	18,8
Educação	15,04
Previdência	47
Outros	107,16

Figura 47: Protocolo grupo 3

150,00 → educação
 188,00 - saúde
 470,00 - previdência
 1071,600 - outros

Figura 48: Protocolo grupo 1

Apesar de tabela ser uma representação usualmente presente na mídia, observa-se que os alunos envolvidos nessa atividade não estão habituados a construí-las. Pode-se assim inferir que a aprendizagem do uso (construção) de tabelas não foi efetivado, mesmo que os alunos tenham tido contato com esse tipo de representação em séries anteriores de sua escolaridade.

4.2.3 Análises e resultados da terceira questão.

Análise da Terceira Questão:

3) Para analisar um conjunto de dados, podemos utilizar medidas que resumem os dados e descrevem suas características, permitindo assim a análise da variação destes dados. Abaixo, indique quais destas medidas que você conhece:

Média Aritmética

Mediana

Moda

Quartis

Na terceira questão, solicitamos que o aluno indicasse quais medidas estatísticas ele conhecia. Pretendíamos, com isso, identificar quais são as medidas-resumo que o aluno diz conhecer (ou ao menos já ouviu falar). Observe-se aqui que não serão identificados casos do aluno saber calcular ou determinar a medida, sem contudo associá-la ao seu nome, à sua designação.

Resultados da Terceira Questão:

Os grupos 2 e 3 não assinalaram nenhuma das medidas. O grupo 1, assinalou “média aritmética” e “mediana”. Estas respostas estão em acordo com os resultados obtidos nos itens (f) e (g) da primeira questão, em que somente o grupo 1 apresentou resultados corretos para a média e a mediana.

4.2.4 Análises e resultados da quarta questão

Análise da Quarta Questão:

Na quarta questão, é exposto o rol com as idades de algumas pessoas que prestaram um determinado concurso, para que o aluno organizasse e analisasse conforme necessidade para responder as questões propostas.

4) As idades das quarenta pessoas que prestaram um concurso de seleção para uma determinada instituição são as seguintes

18, 18, 20, 22, 21, 18, 19, 20, 22, 18,
 19, 20, 21, 18, 20, 19, 20, 19, 22, 19,
 20, 21, 20, 18, 19, 18, 22, 21, 20, 19,
 20, 18, 21, 18, 18, 19, 20, 19, 18, 18

- a) Construa um gráfico de barras que represente as idades destas pessoas.
- b) Quantas pessoas têm 20 ou menos de 20 anos? _____
- c) Qual idade você considera mais apropriada para representar esta distribuição? Justifique. _____
- d) Tente relacionar cada um dos valores das idades desta distribuição com as respectivas medidas:
- I. Média Aritmética () 18 anos
- II. Mediana () 19,5 anos
- III. Moda () 19 anos
- IV. Primeiro Quartil () 20 anos
- V. Terceiro Quartil () 18 anos

No item (a), o aluno deveria organizar os dados, e depois construir um gráfico de barras para esta distribuição. O aluno podia inicialmente ordenar todos os dados, ou então construir uma tabela com a distribuição das idades:

TABELA 9: DISTRIBUIÇÃO DAS IDADES

Idade (em anos)	Nº de pessoas
18	12
19	9
20	10
21	5
22	4
Total	40

Outro procedimento possível seria construir um esquema do tipo “Dot-plot”, para após construir o gráfico de colunas solicitado.

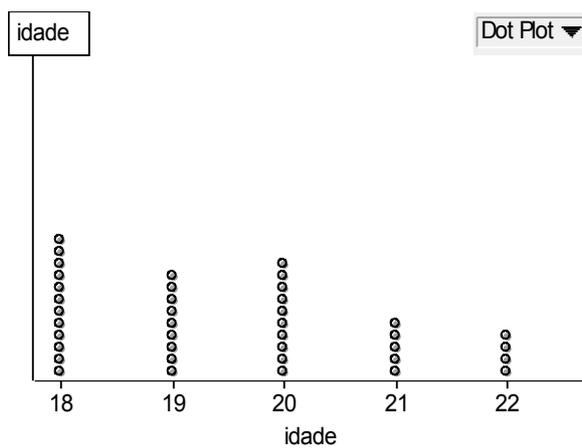


Figura 49: Dot-plot para a variável idade

Com os dados organizados, o aluno podia construir um gráfico de barras, de colunas ou de setores para estas idades.

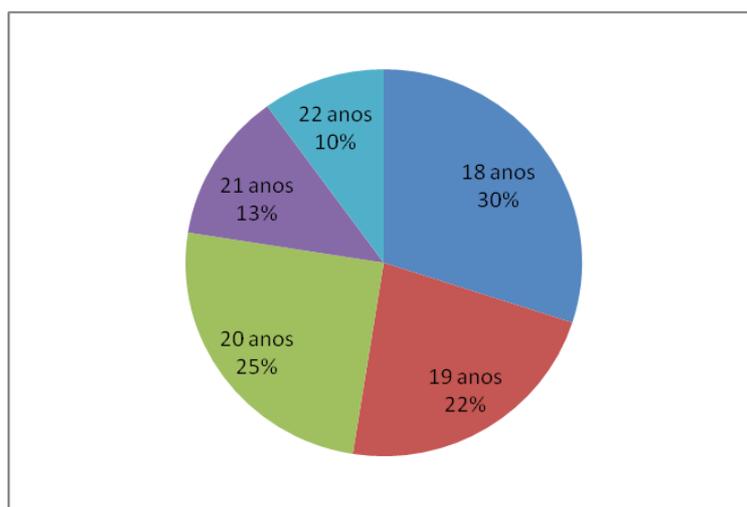
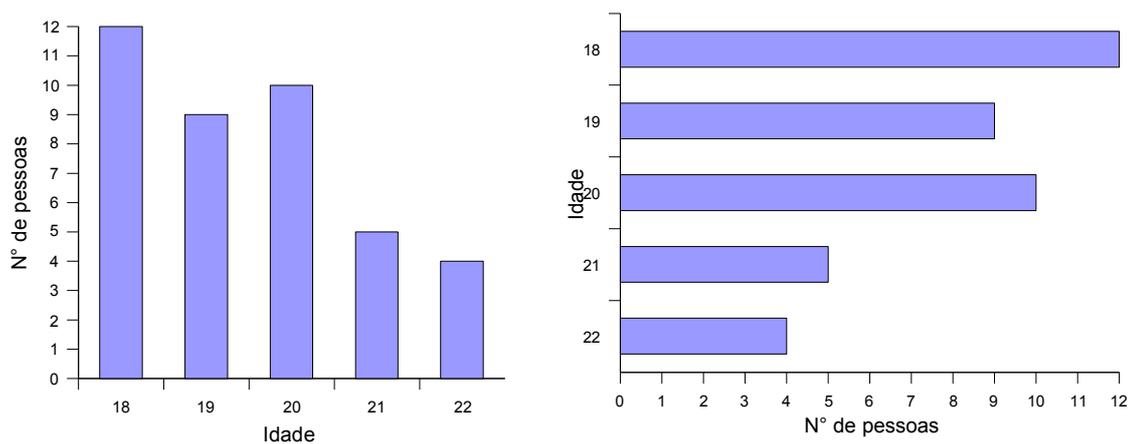


Figura 50: Possibilidades de gráficos para a variável idade

No item (b), o aluno analisaria o gráfico, ou a tabela com os dados

organizados, ou ainda o rol dos dados, para verificar a quantidade de pessoas com 20 anos ou menos. Como resultado, esperávamos que os alunos verificassem que são 31 pessoas, por contagem ou somando as freqüências para as idades de 18, 19 e 20 anos.

No item (c), pretendíamos verificar qual a idade que o aluno consideraria mais representativa para esta distribuição, e como ele justificaria esta escolha. Neste item, existem várias possibilidades de respostas. Os alunos poderiam escolher a média aritmética, que geralmente é o valor mais utilizado para este tipo de questão. Os alunos também poderiam escolher outras medidas, como a mediana e a moda, sempre justificando a sua escolha.

No item (d), novamente pretendíamos verificar quais medidas estatísticas o aluno conhecia, tentando relacionar os valores com as respectivas medidas. Os resultados seriam:

I) Média aritmética:

$$\mu = \frac{12 \times 18 + 9 \times 19 + 10 \times 20 + 5 \times 21 + 4 \times 22}{40} = 19,5 \text{ anos}$$

II) Mediana: Média aritmética entre o 20° e o 21° elemento

$$\left. \begin{array}{l} 20^\circ \text{ elemento no rol : } 19 \\ 21^\circ \text{ elemento no rol : } 19 \end{array} \right\} \Rightarrow Md = \frac{19 + 19}{2} = 19 \text{ anos}$$

III) Moda: Valor mais freqüente

$$\text{Maior freqüência : } 12 \Rightarrow Mo = 18 \text{ anos}$$

IV) Primeiro Quartil: Média aritmética entre o 10° e o 11° elemento

$$\left. \begin{array}{l} 10^\circ \text{ elemento no rol : } 18 \\ 11^\circ \text{ elemento no rol : } 18 \end{array} \right\} \Rightarrow Q_1 = \frac{18 + 18}{2} = 18 \text{ anos}$$

V) Terceiro Quartil: Média aritmética entre o 30° e o 31° elemento

$$\left. \begin{array}{l} 30^\circ \text{ elemento no rol : } 20 \\ 31^\circ \text{ elemento no rol : } 20 \end{array} \right\} \Rightarrow Q_3 = \frac{20 + 20}{2} = 20 \text{ anos}$$

De onde a solução procurada seria:

- | | |
|----------------------|-----------------|
| I. Média Aritmética | (III) 18 anos |
| II. Mediana | (I) 19,5 anos |
| III. Moda | (II) 19 anos |
| IV. Primeiro Quartil | (V) 20 anos |
| V. Terceiro Quartil | (IV) 18 anos |

Resultados da Quarta Questão:

No primeiro item, o aluno deveria organizar os dados para construir um gráfico de barras que representasse as idades informadas. Dois grupos organizaram inicialmente os dados em uma tabela, e depois construíram o gráfico de colunas. Outro grupo, o 2, apenas realizou a contagem dos dados, e depois construiu o gráfico de colunas. Mostramos os gráficos construídos a seguir.

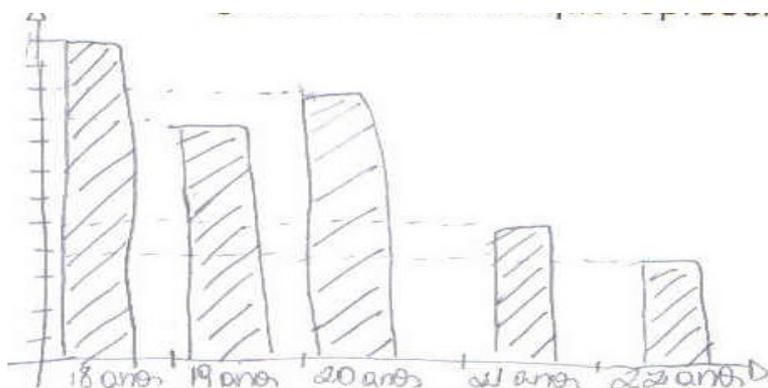


Figura 51: Protocolo grupo 2

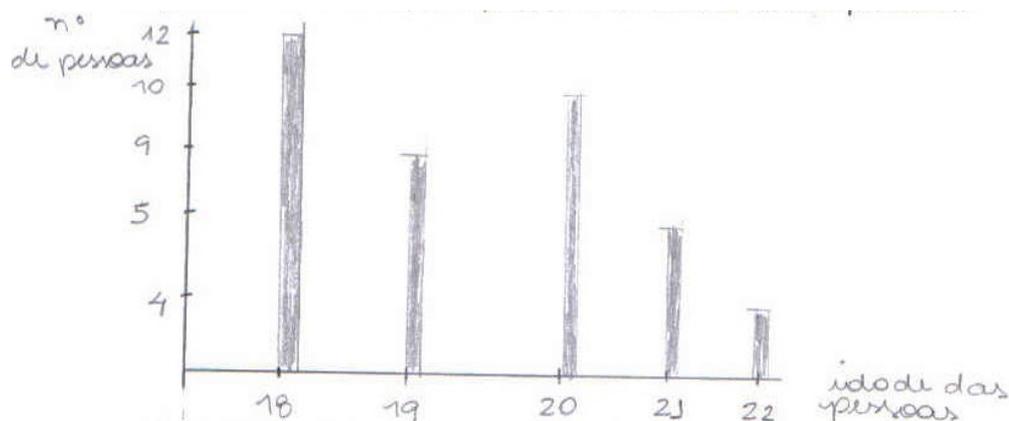


Figura 52: Protocolo grupo 3

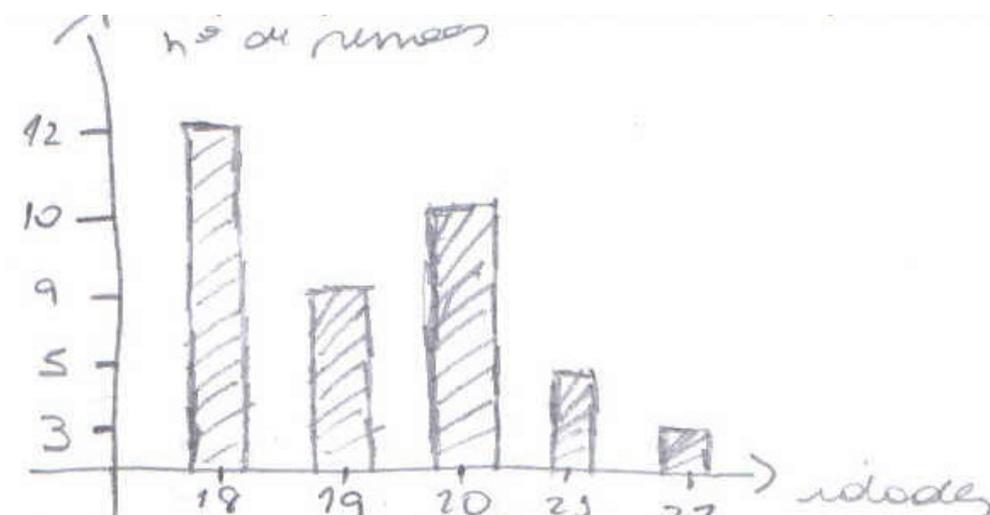


Figura 53: Protocolo grupo 1

Verificamos que desta vez os alunos construíram enfim gráficos de colunas para esta variável, e não o gráfico de função com que estavam acostumados, pois foi solicitado no enunciado da questão este tipo de gráfico. No entanto a construção não foi precisa (feita sem uso de régua graduada).

No item (c) da quarta questão, pretendíamos verificar qual a idade que o aluno consideraria mais representativa para esta distribuição, e como ele justificaria esta escolha. Dois grupos (grupos 2 e 3) escolheram a idade 18, justificando que existe um maior número de pessoas com essa idade, utilizando novamente o conceito de moda de forma intuitiva. O grupo 1 escolheu a idade 22,

ou seja, o valor máximo desta distribuição, conforme protocolos a seguir.

c) Qual idade você considera mais apropriada para representar esta distribuição? Justifique. 18 anos porque o número de pessoas que representam esta distribuição com esta idade é maior.

Figura 54: Protocolo grupo 2

c) Qual idade você considera mais apropriada para representar esta distribuição? Justifique. 22, porque o valor de pesos com esta idade está presente em todos os outros nomes de meses de outros idades.

Figura 55: Protocolo grupo 1

Resultados Gerais do Teste-Diagnóstico

Na análise geral do teste-diagnóstico, supomos que os alunos tiveram pouco contato com a estatística durante sua vida escolar, ou a aprendizagem não foi efetiva (pode ter sido apenas local, não resistindo ao tempo), desconhecendo o que é uma variável estatística, o que são medidas-resumo e como calculá-las. Em relação aos gráficos, os alunos não apresentaram dificuldades na sua leitura, porém apresentaram dificuldades na sua construção.

Pudemos observar que os alunos conseguem ler dados em gráficos e tabelas, mesmo quando estão expressos em unidades diferentes, como a porcentagem. Porém os alunos apresentaram dificuldades em representar estes dados, principalmente na forma gráfica, com a tendência de construir gráfico de pontos na maioria dos casos (talvez influenciados pelo ensino do conceito de função nas aulas de matemática do primeiro ano do Ensino Médio).

Os alunos também apresentaram dificuldades no cálculo da média aritmética, principalmente nos casos que não se trata da média aritmética simples, erro que Batanero (2000, p.7) afirma acontecer, pois “se aplica o algoritmo de forma mecânica e sem compreender seu significado”. Quanto às outras medidas-resumo, como a mediana e a moda, verificamos que os alunos

pouco as conhecem, apesar de usarem intuitivamente a moda para escolher o valor representativo de uma distribuição, por considerarem o valor mais freqüente.

4.3 Seqüência Didática: Análise a priori e Análise a posteriori

4.3.1 Análise a Priori e Análise a Posteriori da Parte A da seqüência didática

A parte A das atividades da seqüência didática foi construída para que o aluno tivesse seu primeiro contato com o software Fathom, com o objetivo de apropriar-se das ferramentas oferecidas no ambiente, além de aprofundar/revisar alguns conceitos estatísticos, permitindo ao sujeito a construção de relações.

Esperava-se que durante o desenvolvimento das atividades, o aluno pudesse:

- Aprender a construir um banco de dados no Fathom;
- Entender o significado da expressão “variável estatística”, e diferenciar variável qualitativa de quantitativa;
- Construir e interpretar a tabela de distribuição de freqüências de uma variável;
- Construir e analisar os diferentes tipos de gráfico de distribuição de freqüências de uma variável;
- Utilizar os recursos do software para identificar relações entre gráficos e tabelas;
- Utilizar os recursos do software para obter a média e a mediana de uma distribuição de freqüências, relacionando essas medidas com a amplitude total dos dados;
- Utilizar os recursos do software para localizar a média e a mediana no gráfico, relacionando essas medidas com a amplitude total dos dados;
- Interpretar o significado destas medidas, relacionando-as com a amplitude total dos dados para melhor apreender a variação deles;
- Utilizar os recursos do software para determinar os desvios de uma variável, interpretando o sentido destes em relação à média;

- Relacionar tabelas e gráficos para responder questões de interpretação dos dados;
- Interpretar o significado de cada tipo de representação e julgar qual o tipo de representação é mais adequada para cada situação.

Nesta situação, as variáveis didáticas em jogo eram:

- a) O ambiente informatizado, que teve por objetivo proporcionar uma ampla visualização do conjunto de dados pelo uso simultâneo de diferentes tipos de representação;
- b) O trabalho em grupo, que teve o objetivo de provocar discussões significativas sobre os conceitos e os procedimentos envolvidos nos problemas propostos entre os alunos sujeitos da pesquisa, assim como tornar explícitos os conhecimentos mobilizados na resolução dos problemas;
- c) O enfoque experimental escolhidos para a organização das atividades propostas, que teve o objetivo de familiarizar os alunos com o software e com os diferentes tipos de representação de uma distribuição de freqüências.

Para desenvolver as atividades propostas, eram necessários poucos conhecimentos prévios. Entre eles, o conhecimento básico em informática, para manuseio no software, além de conhecimentos de operações matemáticas básicas e habilidades para análise e interpretação de tabelas e gráficos. Vale lembrar que uma das características da Análise Exploratória de Dados, como cita Batanero (1991), é a utilização apenas de noções elementares de matemática, não necessitando de uma teoria matemática complexa, permitindo a qualquer aluno realizar a análise de dados.

Na seqüência didática, foram propostas atividades para que os alunos pudessem conhecer as ferramentas básicas do software, como construção de tabelas e gráficos, e cálculos de medidas-resumo, apropriando-se delas para utilização na atividade posterior. Relatamos estas atividades a seguir, com suas análises a priori e a posteriori.

Análise a Priori da Primeira Fase da Parte A das atividades

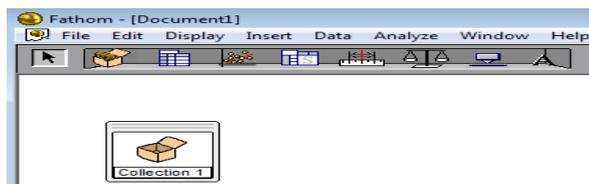
Iniciando a seqüência didática, o aluno observa um conjunto de dados em que constam nome, gênero, idade, cargo, escolaridade e salário de 20 funcionários de uma empresa. A partir deste conjunto de dados, foram propostas inicialmente as seguintes questões:

- Quais as características dos sujeitos observados que estão sendo alvo desta pesquisa? _____
- Se formos classificar estas características, quais critérios poderiam ser usados? Por quê? _____

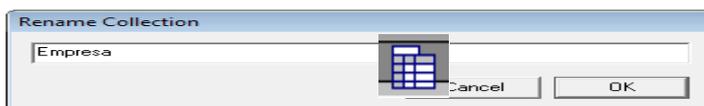
Para o primeiro item, esperávamos que o aluno respondesse que foram pesquisadas características como gênero, idade, cargo, escolaridade e salário. No segundo item, esperávamos que o aluno classificasse as características como numéricas e não numéricas, relacionando como variáveis quantitativas e qualitativas. Nesta faixa de escolaridade, esperava-se que eles já tivessem algum conhecimento básico sobre estatística, conforme orientam os Parâmetros Curriculares Nacionais, as Orientações Curriculares para o Ensino Médio e também conforme os conteúdos propostos nos livros didáticos para a Escola Básica. Apesar disso, na análise do teste-diagnóstico verificamos que estes alunos não conheciam variáveis estatísticas, então podíamos também esperar que os alunos encontrassem dificuldades para responder estas questões, ou até não as respondessem.

Nesta primeira fase das atividades, iniciamos o trabalho com as tabelas.

- 1) Inicialmente, construiremos o banco de dados da empresa no software Fathom. Para isso, siga os procedimentos abaixo.
 - a) Clique no ícone para iniciar o Fathom.
 - b) Clique em “**new collection**”, e arraste até a tela.



c) Clique duas vezes em “**collection1**”; aparecerá a caixa “rename collection”, onde você pode inserir um nome para o seu banco de dados.



d) Agora, clique em “**new case table**”, e arraste até a tela.

e) Insira os dados da pesquisa na tabela do Fathom. Para isso, clique em “**New**” e escreva o título da coluna. Depois, tecle “enter”, para continuar inserindo os dados nas linhas. (*Obs.: não utilize acentuação, pontuação ou espaçamentos*).

2) Agora, construiremos uma tabela de distribuição de freqüências para a variável “**gênero**”.

a) Clique em “**new summary table**” ;

b) No banco de dados do Fathom, selecione a coluna “**gênero**” e arraste-a até a nova tabela.

3) Seguindo o mesmo procedimento, construa também as tabelas de distribuição de freqüências para as variáveis “cargo” e “escolaridade”.

Nesta fase das atividades, o aluno iria conhecer as ferramentas disponíveis para montar um banco de dados no Fathom e também para a construção de tabelas de distribuição de freqüências, para variáveis qualitativas. O aluno poderia perceber que a tabela resultante do exercício (2) ficaria com dois tipos de apresentações diferentes, dependendo da direção da “seta”, que são mostrados na figura abaixo.

pesquisa	Summary Table		
↓	genero		Row Summary
	f	m	
	7	13	20

S1 = count ()

pesquisa	Summary Table	
↓	→	
genero	f	7
	m	13
	Column Summary	20

S1 = count ()

Figura 56: Tabelas de distribuição de freqüência da variável gênero no Fathom

Em seguida, no quarto item, o aluno é solicitado a responder sobre os tipos de variáveis.

4) Você considera que a variável “gênero” e a variável “idade” são de mesmo tipo? Por quê? _____

Nesta questão, esperávamos que o aluno identificasse a variável “gênero” como qualitativa e a variável “idade” como quantitativa. Porém o aluno poderia apresentar dificuldades para responder este tipo de questão, conforme verificado no teste-diagnóstico. As interações com o ambiente, com as situações propostas nas atividades e com os integrantes do grupo devem atuar no sentido de minimizar estas dificuldades.

No quinto item, foram passadas as instruções de como construir as tabelas de distribuição de freqüências para variáveis quantitativas, que requerem um tratamento diferente no software, pois, ao tentar construir a tabela da mesma forma que a qualitativa, o aluno obterá a média aritmética, ao invés das freqüências para cada valor assumido pela variável.

5) Construa uma tabela de distribuição de freqüências para a variável “idade”.

Para isso, clique em “**new summary table**” , selecione a coluna da variável “**idade**” no banco de dados e arraste-a até a nova tabela.

Obs.: para tabelas com dados numéricos (ou seja, variáveis quantitativas), mantenha apertada a tecla **shift** enquanto arrasta a variável até a tabela.

6) Construa a tabela de distribuição de freqüências para a variável “salário”

Esta tabela também poderia ser construída em dois formatos diferentes,

dependendo da posição da “seta”. Os resultados possíveis são mostrados na figura abaixo.

pesquisa	Summary Table																
↓	idade													Row Summary			
	23	25	26	27	29	30	31	32	33	34	35	36	38	40	42	43	20
	1	1	1	1	1	2	1	1	1	1	2	2	1	2	1	1	20

S1 = count ()

pesquisa	Summary Table	
↓		
idade	23	1
	25	1
	26	1
	27	1
	29	1
	30	2
	31	1
	32	1
	33	1
	34	1
	35	2
	36	2
	38	1
40	2	
42	1	
43	1	
Column Summary		20

S1 = count ()

Figura 57: Tabelas de distribuição de freqüência da variável idade, no Fathom

Se o aluno não mantivesse apertada a tecla shift enquanto arrastava a variável, como resultado ele obterá uma tabela com a média aritmética, e não com a distribuição das freqüências, como mostramos na figura a seguir.

pesquisa	Summary Table
↓	idade
	33.25

S1 = mean ()

Figura 58: Média aritmética da variável idade, no Fathom

Em seguida, indicamos os procedimentos para construir tabelas no Fathom combinando duas variáveis, a variável gênero e a variável escolaridade, o que proporciona outro tipo de visualização dos dados, ou seja, outro tipo de apreensão desse registro de representação semiótica.

7) Podemos também construir tabelas combinando algumas variáveis. Por exemplo, vamos construir uma tabela que mostre o nível de escolaridade dos funcionários de gênero masculino e feminino.

a) Clique em “**new summary table**” para construir a tabela;

b) Clique na variável “**escolaridade**” e arraste até o lado **direito** da tabela:

pesquisa		Summary Table			
		escolaridade			Row Summary
		fundamental	medio	superior	
		8	7	5	20

S1 = count ()

c) Agora, clique na variável “**gênero**” e arraste até o lado de **baixo** da tabela.

Como resultado, o aluno obterá a seguinte tabela:

pesquisa		Summary Table			
		escolaridade			Row Summary
		fundamental	medio	superior	
genero	f	3	2	2	7
	m	5	5	3	13
Column Summary		8	7	5	20

S1 = count ()

Figura 59: Tabela para as variáveis gênero e escolaridade no Fathom

Desta forma, foi criado um novo registro de representação semiótica, uma tabela que informa o nível de escolaridade dos funcionários da empresa, por gênero. Vamos admitir, nesse trabalho, que o processo de amalgamar duas tabelas simples, transformando-as em uma única tabela de contingência, proporciona uma diferente percepção dos dados, sem mudar seu conteúdo. Introduce-se, dessa forma, uma nova informação à precedente, relativa às subcategorias geradas pela consideração simultânea das duas variáveis. Realiza-se assim um “tratamento”, pois manteve-se o registro numérico, mudando apenas a disposição dos dados (de forma análoga à fatoração de uma expressão algébrica). Levantamos a hipótese, em nosso trabalho, de que este tipo de reorganização caracteriza uma “apreensão operatória” da tabela, na perspectiva de Duval (1994), e que tal apreensão requer a “leitura entre os dados” da tabela, fazendo-se aqui uma analogia entre os níveis de leitura de gráficos propostos por Curcio (1989, 2001) e a leitura de tabelas.

Análise a Posteriori da Primeira Fase da Parte A das atividades

Antes do início das atividades da parte A da seqüência didática, mostramos aos grupos os resultados do teste diagnóstico, comentando os principais erros cometidos, principalmente em relação aos gráficos, e também procuramos brevemente explicar o que eram as variáveis estatísticas e como diferenciá-las.

Então as atividades foram entregues ao grupo. Explicamos que eles deveriam ler as atividades e desenvolvê-las com o uso do computador e, surgindo qualquer dúvida em relação ao funcionamento do software, poderiam solicitar nossa ajuda, já que estes alunos utilizavam o Fathom pela primeira vez.

Os três grupos iniciaram a leitura dos enunciados. Após observarem o conjunto de dados, foi levantada a primeira questão referente às variáveis estatísticas envolvidas. O primeiro item, sobre quais eram as características pesquisadas, foi respondido corretamente pelos três grupos. Já o segundo item, sobre como poderíamos classificar estas características, gerou dúvidas nos alunos. Os três grupos discutiram bastante sobre a questão.

O grupo 1 inicia o seguinte diálogo:

- Eu acho que é gênero, cargo e escolaridade.
- E idade?
- Acho que idade não, pois idade pode ter várias.
- E salário, não?
- Ah, não sei, acho que não.

Após essa discussão, responderam da seguinte forma:

- Se formos classificar estas características, quais critérios poderiam ser usados?
Por quê? gênero, ~~idade~~, escolaridade, ~~salário~~ porque
um qualitativo e o outro é quantitativo, dando para
diferenciá-los.

Figura 60: Protocolo grupo 1

O terceiro grupo também discute bastante sobre qual característica usar, e responde:

- Se formos classificar estas características, quais critérios poderiam ser usados? Por quê? Carregue Escolaridade Porque de fato
mediram e que realmente mudam.

Figura 61: Protocolo grupo 3

O grupo 2 não respondeu a questão. Analisando estas respostas, verificamos que os alunos procuraram escolher uma variável que representasse o conjunto de todos os dados, ou que apresentasse alguma diferença no banco de dados, não procurando um critério geral que classificasse as variáveis, levando em conta a grandeza envolvida no que se está observando. Observa-se que o grupo 1, inclusive, não conseguiu apreender a diferença entre variável qualitativa e variável quantitativa, classificando a escolaridade como sendo quantitativa, ao invés de qualitativa ordinal. Podemos supor que os alunos, de uma forma geral, não compreenderam o enunciado e, portanto, não conseguiram identificar o problema proposto: identificar a categoria para cada uma das variáveis identificadas no banco de dados.

Com base nestes resultados, percebemos que não houve avanços em relação à identificação e diferenciação de variáveis estatísticas, conceitos que esperávamos que alunos deste nível de escolaridade já conhecessem, conforme nossa análise a priori. O diálogo dos grupos evidencia um conhecimento insuficiente ou mal-construído.

No primeiro item das atividades, os alunos deveriam construir o banco de dados no Fathom, seguindo os procedimentos discriminados na ficha entregue a eles. Todos os grupos conseguiram inserir o banco de dados e não tiveram dificuldades no manuseio com o software.

No segundo e no terceiro item, os alunos deveriam construir com o software tabelas de distribuição de freqüências para variáveis qualitativas, o que todos os grupos conseguiram fazer sem dificuldades. Ao construir a tabela de distribuição de freqüências para a variável “gênero”, aparecem os valores das freqüências para cada gênero, assim como o total de elementos observados (Row Summary), conforme a figura seguinte.

Empresa	Summary Table		Row Summary
	Genero		
	f	m	
	7	13	20

S1 = count ()

Figura 62: Protocolo grupo 1 (Fathom)

Uma integrante do grupo 1 pergunta, e a colega do mesmo grupo responde:

- O que é o “7” e o “13”?
- É que tem 7 pessoas de sexo feminino e 13 pessoas de sexo masculino, e o 20 é o total de pessoas.

Pelo diálogo, percebemos que a primeira aluna ainda não estava conseguindo interpretar as tabelas, enquanto a segunda aluna demonstra ter alcançado a apreensão perceptiva da figura, ou seja, analisa adequadamente as informações contidas nessa representação, descrevendo as etapas de construção de uma distribuição de frequências, estando no nível da “leitura dos dados”, em analogia aos níveis de Curcio (1989, 2001) para leitura de gráficos. Estas interações entre os integrantes do grupo contribuem no sentido da atribuição e explicitação de significado construído às representações pelos alunos. Depois do auxílio da colega de grupo, a primeira aluna também consegue realizar a apreensão perceptiva da figura, evoluindo para o nível⁷ da “leitura dos dados”.

No quarto item, o aluno deveria responder uma questão sobre tipos de variáveis, em que esperávamos que identificassem a variável “gênero” como qualitativa e a variável “idade” como quantitativa. O grupo 1 respondeu corretamente, apesar de ter errado nessa categorização anteriormente, conforme protocolo apresentado na Figura 60; o grupo 2 novamente não respondeu e o grupo 3 respondeu da seguinte forma:

⁷ Desse ponto do texto em diante vamos nos referir aos níveis de leitura de tabelas sempre em analogia aos níveis de leitura gráfica propostos por Curcio (1989, 2001), sem explicitar essa associação a cada momento, evitando assim a repetição no texto. Dessa forma, assumimos como uma premissa de trabalho essa possibilidade de associação, deixando a comprovação da mesma como uma perspectiva de pesquisa a ser realizada.

4) Você considera que a variável “gênero” e a variável “idade” são de mesmo tipo? Por quê? não pois o gênero se limita e a idade é mais variável

Figura 63: Protocolo grupo 3

Pelo protocolo, verificamos que o terceiro grupo identificou que a variável idade “é mais variável”, ou seja, assume uma quantidade maior de valores, enquanto a variável gênero “se limita”, ou seja, apresenta somente duas opções. Porém não identificou essas diferenças como qualitativa e quantitativa. Em nossa análise a priori, fizemos a suposição de que as interações com o ambiente, com as situações propostas nas atividades e com os integrantes dos grupos atuariam no sentido de minimizar as dificuldades em relação ao conceito de variável, assim como a explicação dada por nós no início do encontro, quando foi comentada a atividade diagnóstica. Observando estes resultados, percebemos que os alunos começam a visualizar as diferenças entre as variáveis a partir da variação dos dados observados (apreensão da variabilidade), mas estes avanços ainda não são suficientes para uma categorização formal em termos de quantitativa ou qualitativa. Podemos também inferir que os alunos não sentiram a necessidade de tal categorização para a resolução do problema proposto.

No quinto e no sexto item da atividade, o aluno deveria construir a tabela de distribuição de frequências para variáveis quantitativas, e no oitavo item, o aluno deveria construir tabelas combinando duas variáveis, “gênero” e “escolaridade”.

Os alunos não apresentaram dificuldades no manuseio do software, e conseguiram construir todas as tabelas. Comentaram que acharam “*muito prático construir este tipo de tabela no software*”. Porém, foram pouco receptivos às questões em que tinham que escrever ou justificar algo, e sempre perguntavam “*precisa responder mesmo?*” Ou seja, os alunos gostavam de manusear o software, mas não queriam responder por escrito às perguntas, principalmente quando se tratava de justificativas. No entanto, estas justificativas, ou mesmo os primeiros passos de uma análise de dados apareciam naturalmente do diálogo dos alunos.

Enquanto construía as tabelas, os alunos comentavam o que estavam visualizando nelas: “*Olha, aqui mostra que 2 mulheres têm nível superior, enquanto 3 homens têm nível superior*”, de que podemos inferir um nível de

“leitura entre os dados”.

Empresa		Summary Table			Row Summary
		escolaridade			
		fundamental	medio	superior	
Genero	f	3	2	2	7
	m	5	5	3	13
Column Summary		8	7	5	20

S1 = count ()

Figura 64: Tabela construída pelos alunos no Fathom

Desta maneira, realizavam a “apreensão operatória” dos dados da tabela. Pelos dados por nós observados durante a resolução das atividades pelos alunos, podemos inferir que estas interações com o software e as discussões em grupo tenham contribuído para a compreensão dos dados visualizados e para que atingissem esse nível de apreensão dos registros tabulares de representação semiótica de uma distribuição de freqüências.

Análise a Priori da Segunda Fase da Parte A das atividades

Na segunda fase da parte A, trabalhamos com a representação gráfica, que é outro tipo de representação de uma distribuição de freqüências. Foi apresentado o seguinte enunciado aos alunos:

1) Nesta fase, utilizaremos o software Fathom para construir os **gráficos**, que constituem um outro tipo de representação de uma Distribuição de Freqüências.

- a) Clique no ícone  para construir o gráfico, e arraste até a tela;
- b) Clique na coluna da variável “**gênero**” para selecioná-la, e a arraste até a área do gráfico, sobre a caixa “**Drop an attribute here**”.

Este gráfico é chamado “diagrama de colunas” e é, normalmente, utilizado para representar ou variáveis qualitativas ou variáveis quantitativas discretas. Ele coloca em evidência a comparação entre os valores observados.

Iniciando a atividade, os alunos deveriam construir um gráfico de colunas de uma variável qualitativa no Fathom, a variável “gênero”. O enunciado da atividade explica detalhadamente como construir um diagrama de colunas (“bar

chart” no software, que tem seu menu em inglês) e seguindo as instruções dadas, os alunos deveriam obter como resultado o seguinte gráfico:

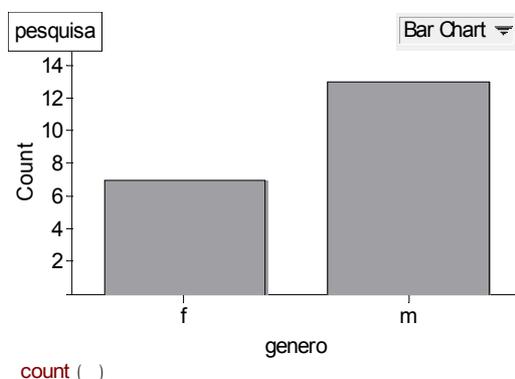


Figura 65: Gráfico para a variável gênero no Fathom

Depois de construído o gráfico, os alunos utilizariam um recurso do software que permite a escolha de diferentes tipos de gráficos.

2) Clique no ícone “bar chart” escolha o outro tipo de gráfico. O que você observa? É o mesmo tipo de comparação que ele representa? _____

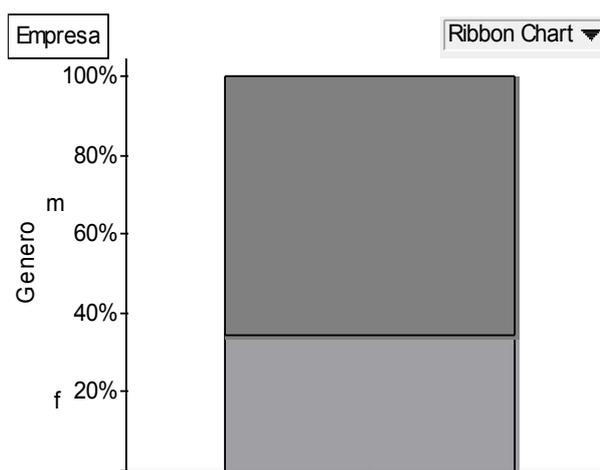


Figura 66: Ribbon Chart – Variável Gênero

Foi solicitado aos alunos que comparassem os diferentes tipos de gráfico, com o objetivo de que percebessem as diferenças entre os mesmos, para que pudessem escolher o mais adequado para cada tipo de situação ou identificar se

cada tipo traria ou não diferentes informações sobre os dados ali representados. Os alunos deveriam verificar que o segundo gráfico (Ribbon Chart) representava percentualmente os dados, podendo ser comparado ao gráfico de setores, já trabalhado no Ensino Fundamental, conforme orientações dos Parâmetros Curriculares Nacionais, Orientações Curriculares para o Ensino Médio e conforme apresentado pelos livros didáticos em geral.

Logo depois, os alunos seguem os mesmos procedimentos para construir os gráficos para as variáveis “cargo” e “escolaridade”.

3) Construa os gráficos para as variáveis “cargo” e “escolaridade”.

Como resultado para este item, os alunos deveriam obter os seguintes gráficos:

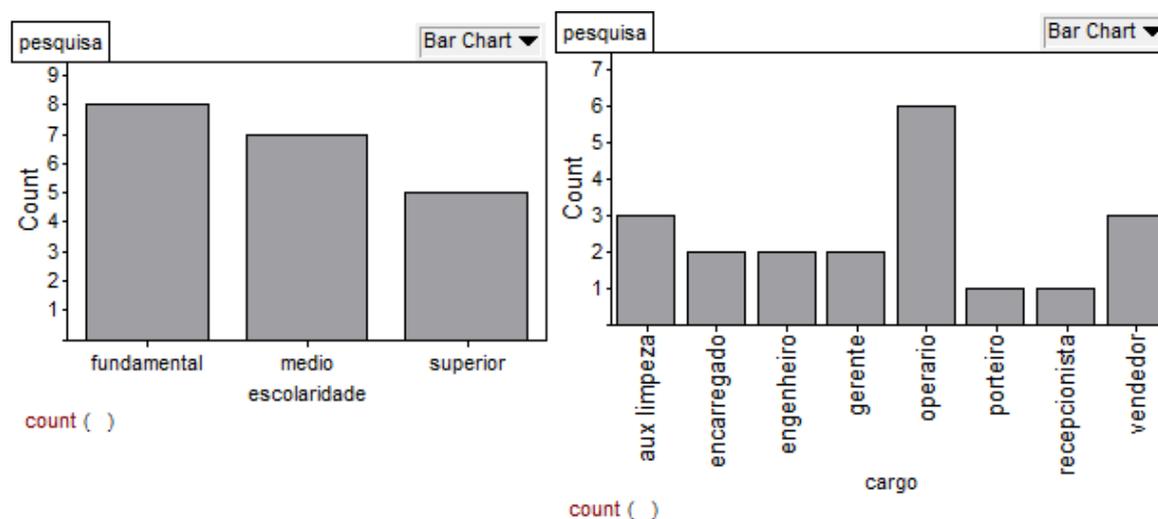


Figura 67: Gráficos para as variáveis escolaridade e cargo no Fathom

Continuando nesta fase da atividade, os alunos construiriam gráficos para variáveis quantitativas.

4) Construa agora um histograma para a variável idade. (Escolha o Histogram no menu do FATHOM).

Inicialmente, seria construído o histograma para a variável idade. Em seguida, solicitamos que os alunos utilizassem um recurso do software e mudassem o gráfico construído no item anterior para o gráfico Dot-plot, comparando os dois tipos de gráficos e explicando as variações observadas em cada um. Esperávamos que os alunos explicassem qual tipo de visualização cada gráfico permite, qual a diferença entre as escalas dos gráficos e qual tipo de gráfico eles julgavam representar melhor os dados.

5) Construa um gráfico de pontos para a variável “idade” (Escolha Dot Plot). Explique a diferença entre os dois gráficos. Como você explicaria a variação dessa variável pela observação dos dois gráficos? _____

Como resultado, deveriam ser obtidos os seguintes gráficos, representando a distribuição dos dados:

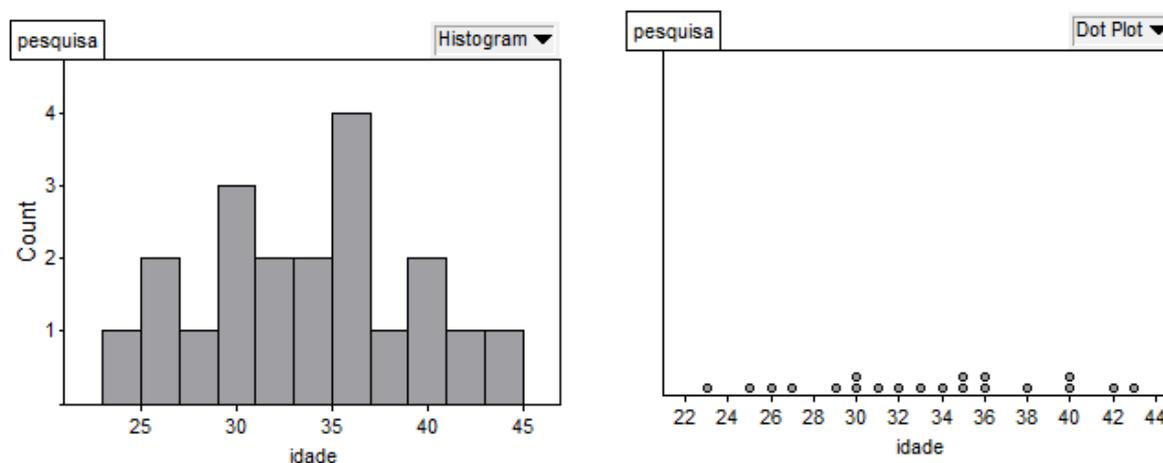


Figura 68: Gráficos Histograma e Dot Plot para a variável idade no Fathom

Esta mudança no tipo de gráfico caracteriza um tratamento, pois o conteúdo representado não foi modificado, apenas transformado de representação em classes (distribuição de frequências para variável contínua) para representação individual de cada ponto observado, conservando o objeto matemático da representação inicial. Esta transformação foi feita inteiramente

pelo software, mas poderia ser analisada e interpretada pelo aluno.

Finalizando esta fase da seqüência didática, no sexto item, o aluno teria que utilizar um recurso do software que permite relacionar os gráficos de diferentes variáveis com o banco de dados. Este recurso possibilita uma nova visualização dos dados, integrando os diversos registros de representações semióticas construídos.

6) Agora, no gráfico da variável gênero, clique sobre a barra que indica o gênero masculino. O que você observa no banco de dados e nos outros gráficos?



Figura 69: Tela do Fathom, relacionando diferentes tipos de registros de representação de um conjunto de dados.

Utilizando este recurso de visualização simultânea dos vários registros de representação semiótica na tela, o aluno pode perceber e interpretar a conversão e o tratamento de registros, da tabela de dados brutos para os gráficos, de diferentes variáveis da pesquisa, fazendo uso da multiplicidade de informações que essas transformações trazem para a análise. Este recurso pode ser bastante útil na interpretação do conjunto de dados e na construção de relações pelo

aluno. Este enfoque facilita a passagem para o nível de “leitura além dos dados” (CURCIO, 1989) e está em acordo com os princípios da Análise Exploratória de Dados, pois o aluno faz uso simultâneo de múltiplas representações para extrair o máximo de informações possíveis.

Análise a Posteriori da Segunda Fase da Parte A das atividades

Ao iniciar a segunda fase das atividades, os alunos já conseguiram rapidamente construir o gráfico do primeiro item e, novamente, não apresentaram dificuldades em relação ao manuseio do software. Alguns integrantes do grupo citam que foi fácil construir o gráfico com esse software e, falam que, se fossem construir o mesmo gráfico no Excel, teriam mais dificuldades. Um aluno fala que *“já tentou construir gráficos no Excel na empresa em que trabalha e teve mais trabalho”⁸*.

No segundo item, foi solicitado aos alunos que mudassem o tipo de gráfico do primeiro item e comparassem os resultados obtidos. Os três grupos responderam de forma parecida, que os dois gráficos não apresentam o mesmo tipo de comparação, pois o segundo gráfico apresenta os dados em porcentagem, como mostramos no protocolo de um dos grupos abaixo. Apesar de verificarem que este gráfico apresentava os dados em porcentagem, nenhum dos alunos associou o “ribbon chart” com o diagrama de setores, o que caracterizaria uma apreensão operatória com transformação mereológica⁹ da figura analisada (passagem do retângulo subdividido segundo proporções determinadas pelas frequências relativas – dados percentuais – para o círculo subdividido segundo os mesmos critérios).

⁸ Os alunos sujeitos dessa pesquisa tinham entre 16 e 17 anos, estudavam no período noturno, e alguns deles já trabalhavam.

⁹ Modificação Mereológica: A figura pode separar-se em partes que são subfiguras da figura dada, fracionando-se e reagrupando-se, isto é, uma relação da parte e do todo (ALMOULOU, 2003, p.127).

2) Clique no ícone "bar chart" escolha o outro tipo de gráfico. O que você observa? É o mesmo tipo de comparação que ele representa? Ele representa a
comparação de forma diferente. Porém o "Ribbon chart"
representa em porcentagem.

Figura 70: Protocolo grupo 3

O terceiro item solicitou a construção de outros gráficos para variáveis qualitativas. Os alunos construíram os gráficos sem dificuldades. Os grupos utilizaram-se dos recursos do software e escolheram o outro tipo de gráfico disponível para esse tipo de variável, mas acabaram voltando para o gráfico de colunas, pois achavam mais simples para visualizar os dados.

No quarto item, os alunos construíram o gráfico Histograma para a variável idade e, no quinto item, modificaram para o gráfico Dot-Plot. Solicitamos que os alunos explicassem a variação dos dados pela observação dos dois gráficos. Esperávamos que explicassem qual tipo de visualização cada gráfico permitia, qual a diferença entre as escalas dos gráficos e qual tipo de gráfico ele julgava representar melhor os dados.

O grupo 2 não respondeu este item. O grupo 3 respondeu que no Dot-Plot, cada pessoa representa um ponto, porém não conseguiu explicar bem as diferenças entre os dois gráficos, justificando que o histograma mostra a "quantidade por idade", não atentando para o fato deste gráfico ter dividido as idades em classes (mostramos a resposta deste grupo no protocolo a seguir). No entanto essa resposta nos permite inferir que o grupo percebe o tipo de tratamento realizado no registro de representação semiótica: a mudança de um agrupamento em classes para a representação de cada dado individualmente na reta numerada (eixo do gráfico).

5) Construa um gráfico de pontos para a variável "idade" (Escolha Dot Plot). Explique a diferença entre os dois gráficos. Como você explicaria a variação dessa variável pela observação dos dois gráficos? Histograma é quantidade por idade e no dot plot é idade por quantidade. Ao invés de termos colunas até a quantidade de pessoas, temos por pessoa (cada pessoa represente um ponto).

Figura 71: Protocolo grupo 3

Já o grupo 1 observou as diferenças nas escalas dos gráficos, explicando que o *"Dot Plot varia de dois em dois e o Histograma de cinco em cinco"*. Este grupo percebe um dos tratamentos efetuados nos gráficos, que foi a mudança de escala. Todavia nenhum dos grupos produz outras explicações quanto à variabilidade dos dados nos gráficos.

5) Construa um gráfico de pontos para a variável "idade" (Escolha Dot Plot). Explique a diferença entre os dois gráficos. Como você explicaria a variação dessa variável pela observação dos dois gráficos? O dot plot varia-se de dois em dois, já o Histograma varia-se de cinco em cinco.

Figura 72: Protocolo grupo 1

No sexto item, solicitamos que os alunos utilizassem um recurso do software que permite a visualização simultânea das diversas representações, relacionando os gráficos de diferentes variáveis, e explicassem o que observavam no banco de dados e nos outros gráficos.

O grupo 2 novamente não respondeu a questão.

O grupo 1 percebeu que este recurso mostra os dados referentes ao gênero masculino em todos os gráficos das demais variáveis pertencentes ao banco de dados.

6) Agora, no gráfico da variável gênero, clique sobre a barra que indica o gênero masculino. O que você observa no banco de dados e nos outros gráficos?

Os dicoramas, mostram que o programa indica todas as características do gênero masculino, ou seja, gênero; idade; cargo; escolaridade e salário.

Figura 73: Protocolo grupo 1

O grupo 3 relata todas as observações dos diferentes tipos de gráficos, explicando qual a relação entre as variáveis, conforme mostramos no protocolo a seguir.

6) Agora, no gráfico da variável gênero, clique sobre a barra que indica o gênero masculino. O que você observa no banco de dados e nos outros gráficos?

A maioria das empresas contratam mais homens do que mulheres. A grande parte tem o ensino médio completo, e a maioria dos homens operários e engenheiros. E a grande maioria tem entre 35 e 40 anos.

Figura 74: Protocolo grupo 3

Observando estes resultados, percebemos que o primeiro grupo conseguiu verificar que tal recurso permite visualizar simultaneamente os vários registros de representação semiótica. Entretanto o terceiro grupo já analisa os dados, construindo relações entre os mesmos, fazendo uso da multiplicidade de informações que puderam ser visualizadas e apresentando assim uma apreensão operatória desses diversos registros figurais e uma leitura além dos dados, segundo os níveis de Curcio (1989, 2001).

Análise a Priori da Terceira Fase da Parte A das atividades

Na terceira fase da parte A, trabalhamos com as medidas-resumo de uma distribuição de freqüências.

Iniciamos com a média aritmética, que é uma medida já conhecida pelos alunos desde o início da escolaridade, ainda que não dentro de um contexto de

estudo da estatística: os alunos tiram média das notas obtidas nas diversas disciplinas, têm contato com o termo na mídia, etc. Como verificamos no teste-diagnóstico, os alunos conhecem média aritmética, mas apresentaram dificuldades no seu cálculo, principalmente quando não se trata de média aritmética simples, pois não compreendem o seu significado.

1) Construa uma tabela de distribuição de freqüências para a variável “idade”, como você fez na primeira fase, mas agora sem apertar a tecla shift (somente arraste a variável até a tabela).

Observe que apareceu somente um valor na tabela. Este é o valor da “média aritmética” para a variável “idade” (no Fathom, quando construímos tabelas de variáveis quantitativas, se você **não** mantiver apertada a tecla shift, a tabela mostra a média aritmética).

A média aritmética é uma medida que resume e representa um conjunto de dados em um único valor. A média de um conjunto é o valor ao redor do qual os valores observados se distribuem, e calculada como a soma dos valores dividida pelo número total de observações. Podemos fazer uma analogia entre a média de um conjunto de dados e o fiel de uma balança em equilíbrio.

2) Seguindo o mesmo procedimento, obtenha o valor da média aritmética para a variável “salário”.

Ao terminar o primeiro e o segundo item da atividade, o aluno obtém como resultado os valores da média aritmética em uma tabela.

pesquisa	Summary Table
↓	⇒ idade
	33.25
S1 = mean ()	

pesquisa	Summary Table
↓	⇒ salario
	1242.5
S1 = mean ()	

Figura 75: Média aritmética para as variáveis idade e salário no Fathom

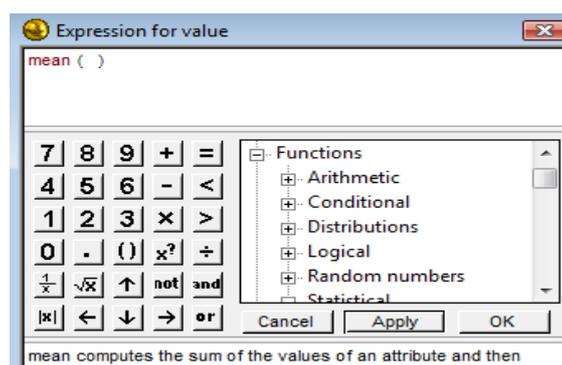
O software calcula este valor rapidamente, e o aluno não visualiza como o cálculo foi efetuado, apenas o resultado final. Dessa forma, o software não

contribui muito com o desenvolvimento do significado desta medida para o aluno pela compreensão do algoritmo. Nesse caso, o significado deveria ser construído pela efetiva aplicação do conceito na análise dos dados. Por exemplo, qual o significado do valor 33,25 obtido como média das idades observadas?

A atividade apresenta uma definição de média aritmética e como ela pode ser calculada, no intuito de contribuir para a compreensão deste significado. Outro recurso do software que pode contribuir com a compreensão do significado de média aritmética é a possibilidade de representar este valor no gráfico, proporcionando uma diferente apreensão dessa medida a partir da sua representação gráfica, podendo facilitar a atribuição do significado desta medida pela associação da média com a amplitude total do conjunto de dados. Dessa forma, o aluno pode, pelo menos, ter uma apreensão perceptiva dessa medida e de sua relação com a amplitude total.

3) Iremos agora representar o valor da média no gráfico.

- a) Construa um gráfico (histogram) para a variável “salário”;
- b) Clique em “**Graph**”, “**Plot Value**”;
- c) No menu que se abrirá, clique em “**Functions**”, “**Statistical**”, “**One Attribute**”.
- d) Clique (duas vezes) em “**Mean** ”.



Seguindo estes procedimentos, o aluno obtém como resultado o seguinte registro de representação semiótica da média do conjunto de dados (linha vermelha no gráfico):

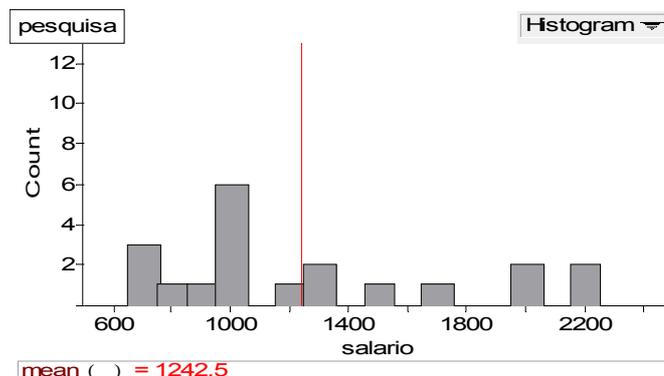


Figura 76: Representação gráfica da média aritmética da variável salário, no Fathom

Assim, neste tipo de representação, o aluno pode visualizar como os demais valores distribuem-se em relação à média (em relação ao posicionamento da linha vermelha), relacionando-a com a amplitude total dos dados, sem aspectos que poderiam se constituir em “distratores”, como o cálculo manual da média (independentemente do tamanho do conjunto a ser considerado) e sua localização na reta numerada.

Desta forma, foi realizada uma conversão de registros (DUVAL, 2003), pois inicialmente a média foi representada em um registro numérico, que depois foi transformado em uma representação gráfica, ou seja, foi modificado o registro de representação, mas conservado o objeto matemático em questão.

A atividade propõe questões que visam a possibilitar que o aluno perceba que a média sozinha não é suficiente para analisar um conjunto de dados e que existem outros valores para essa análise, complementando a informação. Em nossa seqüência, optamos por não trabalhar diretamente com o desvio-padrão, e ao invés disso, passamos para o estudo da mediana e dos quartis, cognitivamente menos complexos do que a variância e o desvio-padrão, e que permitiam a introdução de mais um registro de representação semiótica: o Box-plot.

“A média sozinha é suficiente para que se possa analisar um conjunto de dados? Pense, por exemplo: Qual o intervalo de variação dos valores observados? Qual é a posição da média dentro deste intervalo?”

Nesta questão inserida no meio da atividade, não esperávamos respostas precisas do aluno. Pretendíamos apenas provocar uma reflexão e discussão entre os grupos sobre a variabilidade dos dados.

No quarto item, o aluno obtém o valor da mediana da variável idade na mesma tabela utilizada para a média, o que possibilita a verificação das diferenças entre as duas medidas.

4) Na mesma tabela da média das idades (construída no item 1), vamos mostrar também a mediana.

a) No Fathom, clique na tabela acima, depois clique em **“Summary”**, **“Add Formula”**;

b) No menu, clique em **“Functions”**, **“Statistical”**, **“One Attribute”**. Clique (duas vezes) em **“median”** e **“Ok”**.

Você terá o valor que ocupa a posição central nas idades ordenadas.

Seguindo estes procedimentos, o aluno obtém como resultado o valor da mediana na mesma tabela da média aritmética.

pesquisa	Summary Table
↓	⇒ idade
	33.25
	33.5

S1 = mean ()
S2 = median ()

Figura 77: Média aritmética e mediana da variável idade no Fathom

No quinto item, o aluno representa também a mediana no mesmo gráfico em que foi representada a média.

5) Representaremos a mediana no gráfico (pode ser no mesmo gráfico do item 3).

a) Clique no gráfico da variável salário que você construiu a média;

b) Clique em **“Graph”**, **“Plot Value”**;

c) No menu que se abrirá, clique em **“Functions”**, **“Statistical”**, **“One Attribute”**, **“Median (duplo clique)”**.

6) Como os demais valores representados distribuem-se em relação à média? E em relação à mediana? _____

O resultado obtido resulta na seguinte representação gráfica (linha azul para representar a localização da mediana, linha vermelha para representar a localização da média, observando que a representação numérica fornecida também pelo software aparece nas mesmas cores do que as respectivas linhas):

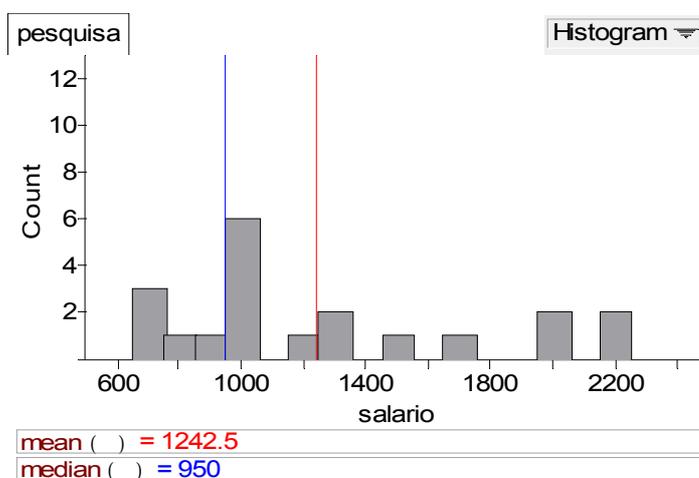


Figura 78: Representação gráfica da média aritmética e da mediana da variável salário, no Fathom

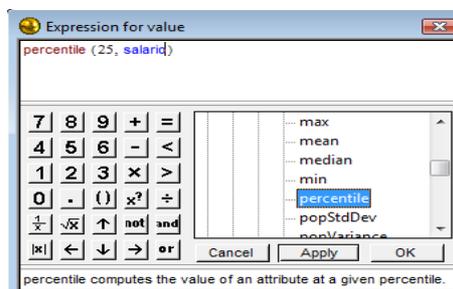
Analisando este gráfico, o aluno poderia verificar como os demais valores do conjunto de dados distribuem-se em relação a essas duas medidas para, no sexto item, poder estudar/analisar esses valores em relação à amplitude total do conjunto de dados.

Continuando a seqüência, são passados os procedimentos para representar os quartis no mesmo gráfico.

7) Para representar os quartis, utilizaremos ainda o mesmo histograma da variável salário.

- Selecione o gráfico e clique em **“Graph”, “Plot value”**;
- No menu, clique em **“Functions”, “Statistical”, “One Attribute”, “Percentile”** (duas vezes)
- Na caixa de funções aparecerá **“percentile ()”**. Dentro dos parênteses,

escreva **(25, salario)**, e clique em **OK**. Você terá representado no gráfico o primeiro quartil;



d) Repita o mesmo procedimento, e na caixa da função percentil, escreva **(75, salario)**, e clique em **OK**. Será representado o terceiro quartil.

No sétimo item, o aluno deveria representar os quartis no mesmo gráfico da variável salário. Este item pode causar maior dificuldade para os alunos, pois requer a utilização de uma função do software, inserindo uma fórmula que calcule os quartis e, como verificamos no teste diagnóstico, os alunos não conhecem ainda esse conceito.

O resultado obtido é apresentado da seguinte forma no gráfico (linhas de mesma cor que a representação numérica apresentada abaixo do gráfico):

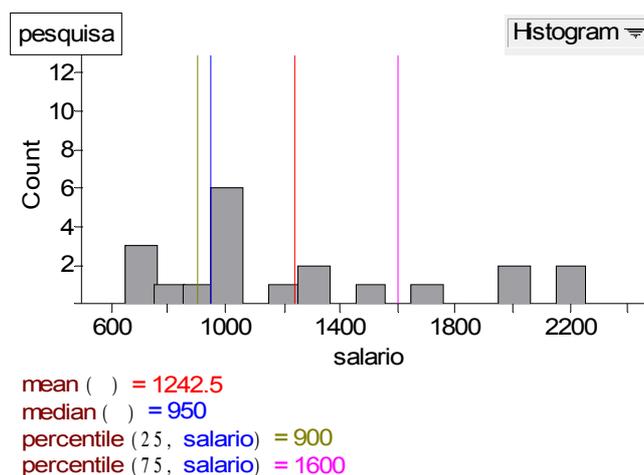


Figura 79: Representação gráfica da média aritmética, da mediana e dos quartis da variável salário, no Fathom

Ainda que o software use cores para diferenciar as linhas, acreditamos que

seria didaticamente mais interessante se cada linha recebesse também uma “etiqueta”, identificando-a.

Logo após a construção deste histograma da variável salário com a representação das medidas, o aluno deveria construir outro tipo de gráfico que também representa estas medidas, que seria o Box-plot.

8) Construa um novo gráfico para a variável “salário” e, no canto superior direito do gráfico, clique na seta e selecione **Box Plot**, para representar este gráfico.

O resultado dessa manipulação seria o seguinte gráfico:

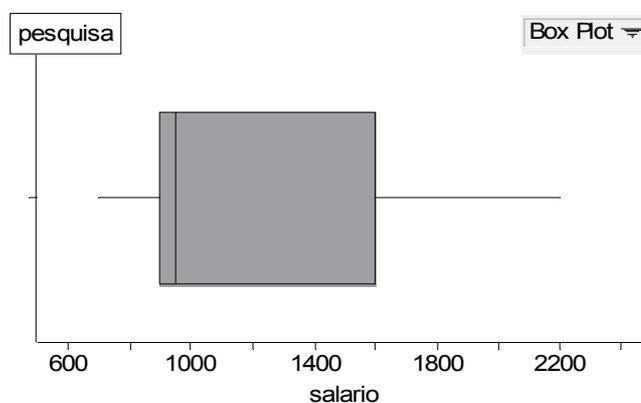


Figura 80: Gráfico Box-plot da variável salário, no Fathom

Nesta fase, era esperado que o aluno comparasse os dois gráficos, relacionando a mediana, os quartis e os valores máximo e mínimo da distribuição. Consideramos que a seqüência de passos necessária para a passagem do Histograma para o Box-plot caracteriza uma apreensão seqüencial, na perspectiva de Duval (1994), conforme apresentado na página 35, e, para realizar tal apreensão, é necessário que o aluno atinja o nível de leitura “além dos dados”.

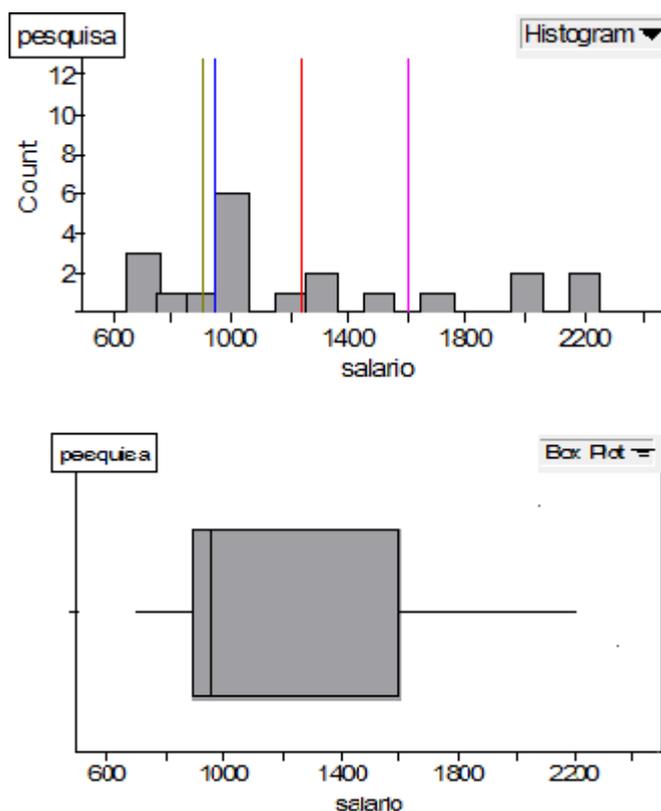


Figura 81: Visualização do histograma e Box-plot (apreensão seqüencial da figura)

Nos itens (9) e (10), solicitamos que os alunos comparassem os dois gráficos construídos, evidenciando os tipos de informações que conseguissem extrair de cada um deles, e escolhessem um deles para estudar este conjunto de dados, justificando sua escolha. Observe-se que, na filosofia da Análise Exploratória de dados, tal escolha não seria necessária, uma vez que é a mobilização conjunta de múltiplas representações que permite levantar a maior quantidade de informações possível dos dados analisados. Tal escolha fez-se por motivos didáticos: evitar uma possível dificuldade na manipulação conjunta de dois objetos, sendo que um deles foi introduzido nessa própria atividade.

9) Em relação aos últimos gráficos utilizados para representar os quartis, responda:

Qual tipo de gráfico você prefere? _____ Justifique _____

10) Compare os dois tipos de gráfico, dizendo as principais diferenças entre eles e quais informações são mais evidentes em um ou em outro. _____

No item (11), o aluno deveria inserir uma nova coluna no banco de dados, para calcular os desvios em relação à média aritmética. Para calcular esses desvios, o aluno utilizaria uma função do software, inserindo uma fórmula que calcula a diferença entre cada valor da variável e a sua média aritmética. Nesse momento, sim, começaria o trabalho para a introdução da idéia de desvios em torno da média. Não trabalhamos o desvio padrão, pois este conceito geralmente é trabalhado no terceiro ano do Ensino Médio (nesta pesquisa estamos trabalhando com alunos do segundo ano do Ensino Médio) e, como verificamos no teste-diagnóstico, os conhecimentos dos alunos em relação à estatística eram mínimos, portanto seriam muitos conceitos novos aos alunos.

11) Veremos agora como inserir uma nova coluna no banco de dados inicial. Vamos assim considerar a distância entre cada valor observado para a variável “idade” e o valor obtido para a média. Chamaremos cada uma dessas distâncias de “desvio”. Podemos chamar esta nova coluna de “idadedesvios”.

- a) Na tabela, clique em **<new>** e insira o título para a coluna;
- b) Para calcular os desvios, clique na coluna “**idadedesvios**”;
- c) No menu, clique em “**Edit**”, “**Edit Formula**”;



d) Na caixa, escreva “**idade-mean(idade)**” e clique em **OK** (será calculado para cada valor da variável, a diferença entre o valor observado e a média aritmética).

12) Observe o resultado na tabela. O que significa o resultado “positivo” nesse cálculo? E o “negativo”? _____

Depois de inserir esta nova coluna com os desvios, foi solicitado que o aluno interpretasse estes resultados, comparando a média com os desvios e com a amplitude total.

Esperamos que após o desenvolvimento destas atividades o aluno consiga diferenciar os diferentes tipos de representação de uma distribuição de freqüências, e consiga integrar e coordenar as representações mais adequadas para diferentes situações. Duval afirma que é a articulação dos registros que constitui uma condição de acesso à compreensão em matemática. Dessa forma, acreditamos que esta articulação entre os diferentes registros contribui para a compreensão dos conceitos estatísticos que estamos trabalhando nestas atividades.

No final da terceira fase das atividades, os alunos respondem um questionário que visa à interpretação dos dados. Para responder a estas questões, o aluno deveria observar os diferentes tipos de representações construídos no Fathom, evoluindo da “leitura dos dados”, para os níveis da “leitura entre os dados” e da “leitura além dos dados”, nos termos de Curcio (1989), assim como evoluindo nos tipos de apreensão de uma figura, nos termos de Duval (1994).

- a) Quantos funcionários responderam à pesquisa? _____
- b) Do total de entrevistados, qual é a quantidade de funcionários do gênero feminino? _____ E do masculino? _____

Nos itens (a) e (b), o aluno realiza a “leitura dos dados”, levantando os explícitos na tabela de distribuição de freqüências para a variável “gênero”, o que requer uma “apreensão perceptiva” dessa representação. Como resultados, o aluno deveria obter os seguintes valores:

- a) 20 funcionários
- b) 7 (feminino) e 13 (masculino)

- c) Em relação ao nível de escolaridade, quantas mulheres têm nível superior? ____ E quantos homens? ____

No item (c), é necessário que o aluno esteja no nível de leitura entre os dados, pois precisa comparar os efetivos de cada gênero, feminino e masculino. Para tanto, o aluno pode utilizar a tabela de distribuição de freqüências que

combina os dois tipos de variáveis (gênero e escolaridade), realizando a sua “apreensão operatória”, ou então pode verificar no banco de dados, linha a linha (que pode ser mais demorado e pode levar mais facilmente a enganos de contagem), realizando apenas uma “apreensão perceptiva” dos registros de representação utilizados. Os valores esperados como respostas eram: 2 mulheres e 3 homens.

d) Entre as idades observadas, qual a idade máxima e a idade mínima?

No item (d), o aluno observa a idade mínima e máxima dos funcionários, o que pode ser feito através de qualquer uma das representações. Na utilização da tabela com os dados brutos, o aluno procura as idades mínimas e máximas. Na utilização da tabela de distribuição de frequências, o aluno visualiza este valor explicitamente, uma vez que os valores aparecem de forma ordenada, realizando apenas a leitura dos dados. Em ambos os casos, o aluno realiza a “apreensão perceptiva” do registro tabular (do registro numérico apresentado em forma tabular). Na utilização do gráfico, o aluno também consegue visualizar a idade mínima e máxima, realizando uma “apreensão discursiva” da figura em um nível de leitura entre os dados, porque basta uma leitura do eixo no qual a escala das idades é representada para se fazer uma inferência sobre o valor mínimo ou máximo pelo estudo da escala representada (ver gráfico ilustrado na figura 68, página 103). Como resultado, verifica-se que a idade máxima seria 43 anos e a mínima seria 23 anos.

e) Qual é o menor salário nesta empresa? _____ E o salário máximo? _____
E a média dos salários? _____

No início do item (e), o aluno observa o salário mínimo e o salário máximo em qualquer uma das representações desta variável. A média dos salários pode ser identificada graficamente, ou então pode ser efetuado seu cálculo no Fathom novamente. Neste item, o aluno realiza a leitura entre os dados e a “apreensão discursiva” da figura. Os resultados obtidos seriam:

Menor salário: R\$ 700,00

Maior salário: R\$ 2200,00

Média dos salários: R\$1242,50

f) Quais são os cargos que têm o salário acima da média nesta empresa? _____

g) Quantos funcionários recebem salário abaixo da média nesta empresa? _____

No item (f), o aluno, já conhecendo a média dos salários, verifica nas tabelas quais cargos têm salário acima deste valor, e no item (g), verifica quais cargos têm salários abaixo deste valor, realizando a leitura entre os dados. Para facilitar esta leitura, o aluno pode utilizar o software e construir uma tabela de distribuição de freqüências combinando as duas variáveis – cargo e salário, realizando a “apreensão operatória” da tabela. Os resultados obtidos seriam:

f) Engenheiro, vendedor e gerente.

g) 12 funcionários.

h) Quantos funcionários do gênero masculino recebem salário acima da média na empresa? ___ E do gênero feminino? _____

i) Quantos funcionários do gênero feminino recebem mais que R\$1900,00? ___ E do gênero masculino? _____

j) Quantos funcionários do gênero feminino recebem menos que R\$900,00? _____ E do gênero masculino? _____

Nos itens (h), (i) e (j), o aluno realiza a leitura entre os dados, relacionando as variáveis. Para tanto, vários recursos do software podem ser utilizados, como a construção de tabelas que combinem estas variáveis (apreensão operatória), ou a verificação no banco de dados (apreensão perceptiva), ou então o uso simultâneo de diversas representações (apreensão operatória). Os valores que esperávamos como resultados seriam:

h) Masculino: 5; Feminino: 3

i) Feminino: 1; Masculino: 3

j) Feminino: 4; Masculino: 1

k) Qual valor você utilizaria para representar o salário desta empresa?

No item (k), o aluno escolhe somente um valor que represente o salário desta empresa, justificando sua escolha, realizando a leitura além dos dados. Como resultado, o aluno pode escolher o valor da média aritmética, ou da mediana, ou ainda o valor mais freqüente, utilizando intuitivamente o conceito de moda, já que esta foi a medida mais utilizada no teste diagnóstico. Neste item, o aluno efetua a “apreensão discursiva”.

Análise a Posteriori da Terceira Fase da Parte A das atividades

Os alunos iniciaram a terceira fase das atividades utilizando uma ferramenta do software para obter a média aritmética para a variável “idade” e para a variável “salário”. Eles acharam “prático” o uso desta ferramenta para o cálculo da média.

No terceiro item, a média foi representada no gráfico. Neste item, os alunos também não apresentaram dificuldade em relação ao manuseio com o software. Inicialmente, acharam difícil a visualização deste valor no gráfico, mas depois conseguiram perceber o intervalo de variação no qual este valor se situava, conforme mostramos no trecho de um diálogo a seguir.

- O que é isso mesmo?
- Olha, essa linha vermelha é a média. Está entre R\$ 1200,00 e R\$ 1400,00.
- Ah, é mesmo....olha lá o cálculo, deu 1242,50.
- É... então aqui (apontam a linha da média) deve ser o R\$ 1242,50.
- É, está mesmo próximo deste valor mesmo.

Neste trecho, percebemos que as interações entre os integrantes do grupo e o software auxiliaram na visualização do valor da média aritmética no gráfico e na sua compreensão como medida indicativa de variação em torno de um valor. Os alunos puderam visualizar o resultado da “conversão” do valor da média aritmética para sua representação gráfica.

No próximo item, os alunos obtêm a mediana para a variável idade na

mesma tabela da média. Eles observaram e comentaram que “*é quase o mesmo valor*”.

No quinto item, os alunos representaram a mediana para a variável “salário” no gráfico. Os alunos perceberam que para esta variável o valor da mediana não é tão próximo da média quanto no item anterior, e comentaram que, neste caso, “*o valor da mediana está próximo de R\$ 1000,00*”. Para verificar o valor correto, um dos grupos fez o cálculo da mediana na tabela para a variável “salário”. Logo depois, perceberam que este valor aparecia abaixo do gráfico. Neste item, verificamos que o aluno visualiza a medida na forma gráfica e efetua uma conversão de registro para confirmar o seu resultado. Desta forma, o aluno está realizando uma apreensão discursiva desta figura, explicitando algumas propriedades da figura.

Em seguida, foi perguntado como os demais valores se distribuem em relação à média e à mediana. Os grupos discutem bastante para responder este item. Relatamos abaixo um trecho de uma das discussões entre os grupos.

- Olha, a média é R\$1242,50, está vendo? É essa linha vermelha aqui (aponta a linha no gráfico)... então estes valores aqui (aponta o gráfico) estão dentro da média, entendeu?... E a mediana, está aqui (aponta a linha da mediana no gráfico), está vendo? Está abaixo do valor da média... entendeu?
- Tá... entendi... mas como vou escrever isso aqui?
- Então, fala assim.... estes valores aqui (apontando o gráfico) estão dentro da média...
- Dentro da média? É, acho que é... não, espera, não estão acima da média? Olha, a linha vermelha está em cima da média, os valores depois da linha estão acima da média... e a mediana está na linha azul.
- Ah, então até aqui é a mediana? (aponta os valores desde o início do gráfico até a linha azul).
- Não... a linha azul é o ponto da mediana... , o que vem antes está abaixo da mediana.
- Ah, mas você entendeu o que é a mediana?
- Eu entendi que a média, pega todos, soma, e divide pela quantidade de pessoas... mas a mediana eu não entendi direito.

No diálogo, percebemos que os alunos utilizaram várias vezes o termo “dentro da média”. Enquanto falavam este termo, eles apontavam a coluna no gráfico em que estava a linha que representava a média, parecendo entender que todos os valores daquela coluna eram a média. Ou seja, para esses alunos, a média não é um valor representativo de um conjunto de dados, mas sim um

intervalo que contém alguns valores desse conjunto. Após alguns diálogos entre os integrantes do grupo, parecem perceber que a linha representa a média aritmética e que os valores localizados depois da linha estão acima da média e antes da linha estão abaixo da média. Nesta etapa, o grupo realiza uma apreensão discursiva da figura.

As dúvidas em relação à mediana ainda permanecem e, neste momento, os alunos deste grupo voltam a ler a definição da mediana nas folhas das atividades. Um dos alunos fala: *“É o valor do meio... mas, sei lá... ainda não entendi direito...”*

Os alunos ainda ficaram confusos em relação ao conceito de mediana, nos chamaram para perguntar *“o que era a mediana mesmo?”*.

Neste momento, percebemos que os três grupos ainda estavam confusos em relação ao conceito de mediana, e chamamos a atenção de todos para a discussão de um breve exemplo. Perguntamos a idade de todos os alunos ali presentes, e as escrevemos na lousa em ordem crescente: 15 – 15 – 16 – 16 – 16 – 16 – 17 – 18. Em seguida, discutimos com os alunos as definições da folha de atividades, explicando que inicialmente ordenam-se os dados, para depois verificar qual o termo que ocupa a posição central. Como neste caso são oito termos, dois termos ocupam a posição central, então é necessário obter a média aritmética entre estes dois termos. Os alunos concluem que o valor da mediana neste caso é “16”. Um dos alunos aproveita para utilizar este mesmo exemplo e calcular a média aritmética, nos perguntando se o seu cálculo estava correto. Solicitamos que o aluno desenvolvesse o cálculo na lousa, e perguntamos aos outros alunos se estava correto. Todos concordaram que sim.

Continuando o diálogo no mesmo grupo (grupo 1):

Ah, entendeu agora? Vamos responder então...

Entender, eu entendi... mas como vou explicar isso?

As respostas formuladas pelos grupos foram as seguintes:

6) Como os demais valores representados se distribuem em relação à média? E em relação à mediana? *Em relação à média eles estão acima do valor, e em relação a mediana eles estão dentro do valor da mesma.*

Figura 82: Protocolo grupo 1

6) Como os demais valores representados se distribuem em relação à média? E em relação à mediana? A maioria das pessoas ganham menos de R\$ 1242. A parte mediana, ganham mais que R\$ 950.

Figura 83: Protocolo grupo 3

6) Como os demais valores representados se distribuem em relação à média? E em relação à mediana? A soma dos valores dividido pela quant mediana é o valor que ocupa a posição central ou seja, 50%.

Figura 84: Protocolo grupo 2

Baseando-se nas discussões que foram realizadas, percebemos que os alunos começam a compreender as medidas, mas ainda não conseguiam explicar seus significados, ou seja, não conseguiam expressá-los na linguagem natural (que também constitui um registro de representação semiótica para Duval), mas apenas o algoritmo para sua determinação. Dessa forma, o software parece ter ajudado na sua visualização, mas não na construção de um significado para o conceito.

No sétimo item, os alunos determinam o primeiro e o terceiro quartil. Novamente surgem as dúvidas, e após ler e reler a atividade, um dos alunos falou: “Nossa, o que é quartil?” Novamente, chamamos a atenção de todos e usamos o mesmo exemplo das idades para explicar para os alunos esse conceito. Neste momento, eles dizem ter entendido o seu significado.

Continuando as atividades, os alunos observaram os valores o primeiro quartil e do terceiro quartil no gráfico da variável “salário”. Os alunos do grupo 1, olhando o gráfico e as medidas, comentaram:

- Ah... acho que, se for assim então, a mediana deve ser o segundo quartil.
- É mesmo, porque a mediana está aqui (aponta o gráfico), esse é o primeiro quartil, esse é o terceiro... então a mediana deve ser o segundo.

Neste diálogo, percebemos que os alunos não conhecem quartil, mas

começam a compreender seu significado a partir da discussão e da visualização no gráfico, conseguindo relacionar o segundo quartil com a mediana. Neste caso, poderíamos dizer que estes alunos realizaram uma apreensão discursiva da figura, explicitando a linha da mediana como segundo quartil, lendo entre os dados.

Em seguida, a atividade solicita a construção do Box-plot para representar os quartis. Novamente, um dos alunos falou: “Nossa, o que é isso agora... nunca ouvi falar em box-plot...”

Os grupos acharam essa representação confusa, conforme mostramos neste trecho de um diálogo:

- Você entendeu esse Box-plot?
- Ah, eu entendi que ele separa por quartil... mas é difícil de ver. E que é uma caixa...
- É, achei confuso, o outro (referindo-se ao histograma) é bem mais fácil.

Quando solicitamos que os alunos respondessem qual tipo de gráfico eles preferiam, todos responderam que era o histograma, justificando que o histograma “é mais fácil de entender”, conforme mostramos nos protocolos a seguir.

9) Em relação aos últimos gráficos utilizados para representar os quartis, responda:

Qual tipo de gráfico você prefere? Histograma Justifique Porque é bem mais fácil de entender

Figura 85: Protocolo grupo 2

9) Em relação aos últimos gráficos utilizados para representar os quartis, responda:

Qual tipo de gráfico você prefere? Histograma Justifique Porque é de mais fácil compreensão

Figura 86: Protocolo grupo 3

9) Em relação aos últimos gráficos utilizados para representar os quartis, responda:

Qual tipo de gráfico você prefere? Gráfico de barras Justifique pois mostra de modo mais simples e objetivo as informações passadas.

Figura 87: Protocolo grupo 1

Baker, Biehler e Konold (2004), em um estudo sobre o ensino do Box Plot, relatam que este tipo de gráfico é rico conceitualmente, mas para compreendê-lo é necessário ter conhecimento sobre quartis. Estes autores ainda relatam, em seus resultados de pesquisa, que a mediana não é tão intuitiva aos estudantes como os pesquisadores esperavam, e que a divisão dos dados em quartis é de difícil compreensão. Podemos constatar que os alunos sujeitos de nossa pesquisa encontraram as mesmas dificuldades em relação ao Box-plot, por não terem conhecimentos sobre os quartis, e também porque o conceito de mediana não é tão intuitivo quanto pensávamos, e o ambiente informatizado não foi suficiente para a compreensão do Box-plot como representação semiótica da distribuição.

Ao solicitarmos que os alunos comparassem os dois tipos de gráficos, explicando as principais diferenças, e quais informações são mais evidentes em cada, eles citaram que no histograma as informações são mais claras e que, para ler o Box-plot “*tem que ter um certo conhecimento*”, ou seja, o conhecimento sobre quartis, que Baker, Biehler e Konold (2004) citaram. Em seguida, mostramos os protocolos desta questão.

10) Compare os dois tipos de gráfico, dizendo as principais diferenças entre eles e quais informações são mais evidentes em um ou em outro. No histograma é mais fácil de ler, Box Plot tem que ter um certo conhecimento

Box Plot → Valor mínimo e máximo e a mediana em ambos

Figura 88: Protocolo grupo 3

10) Compare os dois tipos de gráfico, dizendo as principais diferenças entre eles e quais informações são mais evidentes em um ou em outro. Histograma, pois as informações estão mais claras e de forma mais objetiva de ser entendida e compreendida.

Figura 89: Protocolo grupo 1

Continuando as atividades, trabalhamos com os desvios em relação à média. Os alunos construíram uma nova coluna no banco de dados e calcularam estes desvios. Em seguida, solicitamos que os alunos observassem os resultados dos desvios, explicando o significado do resultado positivo e do resultado negativo nesse cálculo. No geral, os alunos explicaram que o resultado negativo mostrava os valores que estavam abaixo da média, e o resultado positivo mostrava os valores que estavam acima da média, conforme mostramos nos protocolos dos grupos a seguir.

12) Observe o resultado na tabela. O que significa o resultado "positivo" nesse cálculo? E o "negativo"? negativo -> está abaixo da média.
positivo -> está acima da média.

Figura 90: Protocolo grupo 1

12) Observe o resultado na tabela. O que significa o resultado "positivo" nesse cálculo? E o "negativo"? Personas com 34 anos ou mais, o resultado deu
positivo. E personas com 33 anos ou menos, deu negativo.

Figura 91: Protocolo grupo 2

12) Observe o resultado na tabela. O que significa o resultado "positivo" nesse cálculo? E o "negativo"? A idade é maior que a média. (positivo)
A idade é menor que a média (negativo)

Figura 92: Protocolo grupo 3

Terminando esta etapa das atividades, os alunos responderam ao questionário.

Todos os grupos responderam corretamente os itens (a), (b), (c), (d), (e) do questionário. Para responder estes itens, os grupos utilizaram as tabelas de distribuição de frequências das variáveis "gênero", "idade", "escolaridade" e "salário", já que todas estavam disponíveis, realizando a apreensão perceptiva (itens a, b), o que requer apenas a leitura dos dados, a apreensão operatória (item c), e a apreensão discursiva (itens d, e), que requer a leitura entre os dados.

No item (f), sobre quais cargos têm salário acima da média, os grupos 2 e 3 responderam da forma esperada, e o grupo 1 esqueceu um dos cargos. Para responder este item, os grupos 2 e 3 consultaram no banco de dados a coluna com os valores dos desvios do salário, observando quais eram os cargos com desvios positivos, identificando os cargos de gerente, engenheiro e vendedor pela leitura dos dados, realizando a “apreensão perceptiva”. O grupo 1 verificou primeiro qual era a média aritmética dos salários, e depois consultou no banco de dados, na coluna de salários, quais eram os cargos com salários acima deste valor, identificando os cargos de gerente e engenheiro, não observando o cargo de vendedor, cometendo erro previsto em nossa análise a priori.

No item (g), somente o grupo 3 respondeu como esperávamos. Os grupos 1 e 2 responderam quais os cargos que recebiam salário acima da média. Ambos os grupos erraram na interpretação do problema, trocando a palavra “quantos” por “quais” (talvez possam ter sido influenciados pelo item anterior).

No item (h), os grupos 2 e 3 responderam corretamente, e o grupo 1 novamente esqueceu alguns dados. Para responder este item, os grupos 2 e 3 olharam no banco de dados a coluna dos desvios dos salários, identificando os desvios positivos para o gênero masculino. O grupo 1 respondeu baseando-se no item (f), ou seja, como já tinham verificado no item (f) o salário acima da média para os cargos de gerente e engenheiro, apenas verificaram na tabela de distribuição de freqüências a quantidade de funcionários nestes cargos.

Os itens (i) e (j) foram respondidos corretamente pelos três grupos, que fizeram a leitura entre os dados. Para responder estes itens, os grupos 2 e 3 consultaram o banco de dados, realizando a “apreensão perceptiva”, e o grupo 1 consultou as tabelas de distribuição de freqüências, realizando a “apreensão operatória”. Neste momento, o grupo 1 percebeu que o cargo de encarregado também recebe salário acima da média, mas considerou esta informação apenas para responder estas duas questões, e não se lembrou de voltar para corrigir as questões anteriores.

No item (k), que exigia a leitura além dos dados para a escolha de um valor representativo para o salário desta distribuição, verificamos que apenas um grupo escolheu uma medida-resumo para representar este salário, e os outros grupos utilizaram valores aleatórios, com justificativas que não explicavam o real

significado deste valor para o grupo. Cada grupo considerou um valor diferente, conforme mostramos nos protocolos a seguir.

k) Qual valor você utilizaria para representar o salário desta empresa? R\$ 950
 Justifique: Porque é a média de todos os salários.

Figura 93: Protocolo grupo 3

k) Qual valor você utilizaria para representar o salário desta empresa? 1242,5
 Justifique: A média de salário e boa para os serviços prestados

Figura 94: Protocolo grupo 2

k) Qual valor você utilizaria para representar o salário desta empresa? R\$ 700,00
 Justifique: Por este valor se encaixa no salários de todos

Figura 95: Protocolo grupo 1

O grupo 3 considerou o valor de R\$950,00, justificando que este era o valor da média. Porém o valor correto da média era de R\$1242,50. O grupo 2 considerou o valor da média aritmética, de R\$1242,50, justificando que esta média era “boa para os serviços prestados”. O grupo 1 considerou o valor de R\$700,00, pois achavam que este valor se “*encaixava no salários de todos os funcionários*”. Em nossa análise a priori, esperávamos que os alunos utilizassem uma das medidas-resumo para este item. Porém, baseando-se nestes resultados, percebemos que somente um dos grupos utilizou o valor da média aritmética para representar o salário da empresa, e os outros dois grupos continuaram utilizando valores aleatórios, da mesma forma que fizeram no teste diagnóstico. Pode-se inferir que a visualização das medidas não foi suficiente para que os alunos associassem as medidas com representações semióticas do conjunto de dados.

Com base nos resultados deste questionário, percebemos que os alunos conseguiram responder corretamente a maioria das questões de análise de dados, inferindo então que estes alunos estão no nível da leitura entre os dados, segundo os termos de Curcio (1989). Para coletar as informações que permitiriam responder as questões, os alunos geralmente escolhiam como representação as tabelas, pois localizavam mais rapidamente informações solicitadas, fazendo

pouco uso de outras ferramentas do software.

4.3.2 Análise a Priori e Análise a Posteriori da Parte B da seqüência didática

Análise a Priori da Parte B das atividades

Na parte B das atividades, o aluno deveria fazer uso dos procedimentos que conheceu na atividade de familiarização (parte A), para preparar um relatório composto por tabelas, gráficos, medidas-resumo e textos explicativos, que visasse a informar as principais características de um grupo de estudantes de sua escola.

O objetivo desta fase era que o aluno, após o desenvolvimento da parte A, pudesse escolher variáveis de pesquisa que considerasse mais adequadas para traçar o perfil dos estudantes de sua escola, coletar estes dados e construir seu banco de dados no Fathom, organizando-os de forma adequada, construindo os diversos registros de representação semiótica para levantar o maior número de informações possíveis sobre os dados, escrevendo breves textos que explicassem as características observadas. O software atuaria como ferramenta para construção das diversas representações consideradas.

Esperava-se que, durante o desenvolvimento das atividades, o aluno pudesse:

- Escolher as variáveis de pesquisa que julgasse mais conveniente para caracterizar o perfil de sua turma;
- Coletar e organizar os dados de forma adequada;
- Escolher os tipos de representações que julgasse mais adequado para representar cada variável;
- Redigir um texto que descrevesse as características de sua pesquisa.

Vale lembrar que essas etapas propostas para as atividades de nossa seqüência didática estão em acordo com os objetivos didáticos no uso de computadores no ensino da estatística, apontados por Godino (1995), que ressalta que um dos objetivos do ensino de estatística é capacitar o aluno a coletar, organizar, depurar, reunir, representar e analisar os dados.

Nessa atividade, o aluno estaria envolvido em uma coleta de dados de colegas de sua escola, que são dados reais ligados ao cotidiano do aluno, o que, para Curcio (1989), pode estimular o aluno a levantar questionamentos sobre os dados e contribuir para a compreensão das relações expressas nos gráficos. Godino (1995) também ressalta a importância do trabalho com sistemas de dados reais. Batanero (2001) cita que projetos em que se possam coletar os dados motiva os alunos, que podem ver a utilidade da estatística.

Nesta etapa das atividades, havia algumas variáveis didáticas em jogo:

- O ambiente informatizado, que atuava no sentido de proporcionar o uso simultâneo de várias representações;
- A coleta de dados e o trabalho com dados reais e do cotidiano do aluno, que tinha o objetivo de motivá-lo, podendo ver a utilidade da estatística. Esta possibilidade de gerar situações de aprendizagem referentes a temas de interesse do aluno é uma das características da Análise Exploratória de Dados, citada por Batanero (1991);
- O trabalho em grupo, que teve o objetivo de levantar questionamentos sobre o conjunto de dados e provocar discussões significativas sobre as informações coletadas, e sobre as diferentes possibilidades de representar estas informações, entre os alunos sujeitos da pesquisa.

Os conhecimentos prévios necessários para o desenvolvimento desta etapa das atividades eram a habilidade no manuseio no software Fathom, que o aluno deveria ter desenvolvido durante a atividade de familiarização, e a habilidade de analisar e representar os dados em diferentes registros, como tabelas, gráficos e medidas resumo, que também foram trabalhadas durante as atividades de familiarização.

Durante o desenvolvimento desta atividade, o aluno, sujeito desta pesquisa, poderia transitar pelas quatro dimensões da estrutura do Pensamento Estatístico (WILD e PFANNKUCH, 1999), podendo assim desenvolver esse tipo de pensamento:

- No ciclo investigativo, o sujeito planeja como coletar, organizar e analisar os dados;
- Na segunda dimensão, dos “tipos de pensamento”, o sujeito reconhece a importância dos dados, realiza a transnumeração (mudando a

representação dos dados para melhor compreensão dos mesmos), considera a variação dos dados e utiliza modelos estatísticos para organizar e analisar os dados;

- No ciclo interrogativo, o sujeito busca informações e idéias, para, posteriormente, interpretar o resultado estatístico, lendo, resumindo e comparando os dados;
- Na quarta dimensão, o sujeito pode apresentar disposições para querer investigar mais, questionar conclusões, considerar novas idéias.

Para desenvolver as atividades, os alunos inicialmente deveriam discutir sobre quais variáveis entrariam na caracterização deste grupo de estudantes, como achavam que poderia ser o perfil destes estudantes, e como iriam traçar esse perfil.

Suponha que você queira comunicar a um aluno de outro estado ou outro país, como são os estudantes de sua escola. Para isso, você irá preparar um relatório que mostre resumidamente as características principais destes estudantes.

1) Preparação das questões de investigação: discuta, com seu colega de dupla, sobre:

a) Quais dados (físicos, sociais, culturais...) devem entrar na caracterização do aluno típico?

b) Será necessário traçar um perfil para as meninas e outro para os meninos? Por quê?

Depois dessa discussão, os alunos deveriam preparar folhas de registros e coletar os dados na sua escola.

2) Preparação da coleta de dados:

- Prepare folhas de registro para os dados que serão pesquisados;
- Escolha as variáveis da pesquisa;
- Peça aos seus colegas de escola que respondam as perguntas da pesquisa.

Após coletar e organizar os dados, os alunos deveriam escolher os tipos de

representações que julgassem mais apropriadas para representar cada variável.

3) Organização e representação dos dados:

- Organize seus dados adequadamente (como você aprendeu nas atividades de familiarização);
- Para cada tipo de variável da sua pesquisa, construa os tipos de representações (gráficos, tabelas) que você considerar mais apropriado para representar tal variável. Quando considerar conveniente, utilize também as medidas-resumo;
- Escolha uma das variáveis para fazer as representações somente utilizando lápis e papel e, para as demais, utilize o software Fathom;
- Acompanhando estas representações estatísticas, redija pequenos textos que servirão como relatórios, descrevendo os estudantes de sua escola pelo conjunto de dados que você coletou.

Os resultados desta etapa seriam utilizados para verificarmos os tipos de articulações entre as representações deste conjunto de dados que foram identificadas pelos alunos. Como nesta etapa os alunos ficaram livres para escolher o tipo de representação que julgassem convenientes, podíamos esperar as seguintes possibilidades de resultados:

- Os alunos poderiam ser influenciados pela atividade de familiarização, desenvolvendo a parte B da mesma forma;
- Os alunos poderiam utilizar uma quantidade menor de recursos, excluindo os conceitos que achassem mais complexos;
- Os alunos poderiam ir além da atividade de familiarização, descobrindo novos recursos e possibilidades do software.

Solicitamos ao aluno que escolhesse uma das variáveis e a representasse utilizando somente lápis e papel para verificarmos se houve avanços em relação aos conceitos estatísticos trabalhados nas atividades da parte A.

Na última parte da atividade, o aluno deveria redigir um texto explicativo, descrevendo os estudantes de sua escola pelos dados coletados. A análise dos resultados desta parte deveria mostrar o tipo de interpretação que o aluno fez deste conjunto de dados. Assim poderíamos analisar as articulações mobilizadas

pelo aluno entre os diferentes tipos de representação, os tipos de apreensões (Duval, 2003) destas representações que foram mobilizadas e o nível de compreensão gráfica (Curcio 1989, 2001), relacionando estes resultados com o quadro teórico de nossa pesquisa, visando a responder nossa questão de pesquisa.

Análise a Posteriori da Parte B das atividades

Após o término da terceira fase da parte A das atividades, entregamos aos alunos a atividade B, para que eles pudessem discutir, preparar e realizar a coleta de dados.

Os três grupos decidiram preparar a folha de registros comum para os três, para após a coleta de dados, escolherem uma amostra deste conjunto para desenvolver a atividade.

Então, prepararam folhas de registros contendo as seguintes variáveis: idade, altura, peso, esporte preferido, série, tipo de música preferido, matéria preferida, tempo gasto para chegar à escola, profissão que pretende seguir. Depois, pediram que duas turmas da sétima série e duas turmas da oitava série do período da manhã respondessem a pesquisa. No total, cento e onze alunos responderam, utilizando quatro folhas de registros.

Após a realização desta coleta de dados, cada grupo realizou a parte B das atividades no computador, em horários diferentes, mas sob nossa orientação e supervisão. Cada grupo escolheu as variáveis com que iria trabalhar, e como definiria a amostra utilizada.

Resultados do primeiro grupo

Os alunos do primeiro grupo discutiram sobre como organizariam os dados coletados na pesquisa e decidiram utilizar as seguintes variáveis para montar seu banco de dados: idade, esporte preferido, tipo de música, tempo gasto para chegar à escola, matéria preferida, futura profissão, sexo. O grupo decidiu utilizar um total de 35 elementos do banco de dados coletado para montar o banco de dados a ser utilizado na atividade (não utilizaram o banco de dados completo,

ao teste-diagnóstico, pois o grupo conseguiu representar graficamente a variável, conseguiu localizar a mediana ordenando os dados, e conseguiu calcular a média aritmética desta distribuição.

No Fathom, este grupo preferiu utilizar a representação gráfica, construindo um gráfico de colunas que representasse cada variável. O grupo não redigiu um relatório final, mas após a representação gráfica de cada variável, escreveu uma frase descrevendo esta representação. No gráfico da variável “tempo gasto para chegar à escola”, o grupo mostrou também o valor da média aritmética no gráfico.

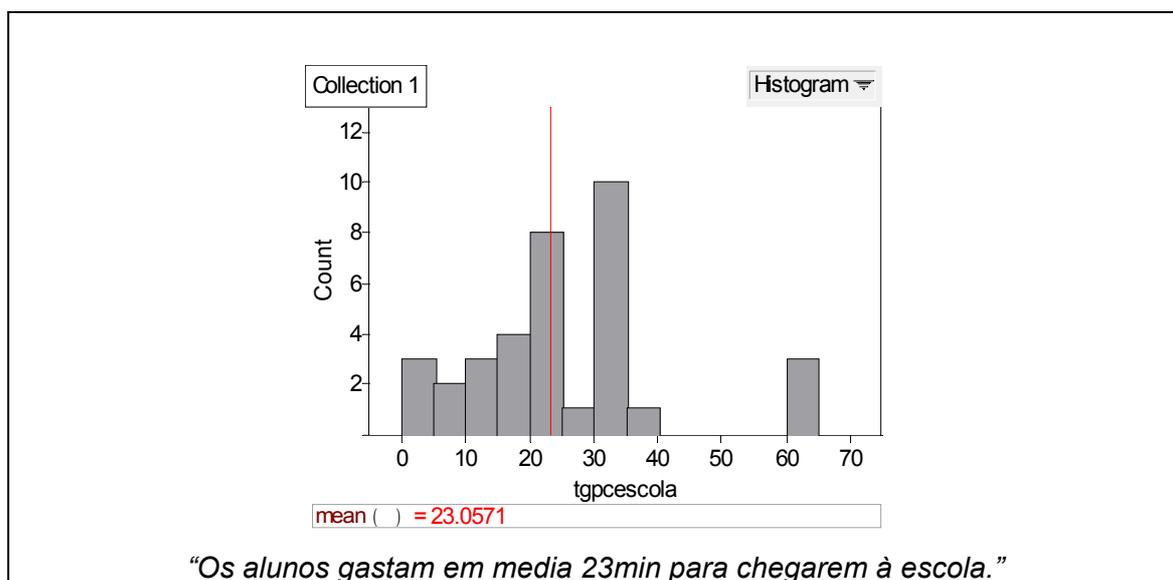


Figura 98: Parte do relatório do grupo 1

O grupo tentou calcular a média para as outras variáveis, mas após algumas tentativas em alguns dos gráficos, os alunos perceberam que não era possível.

- Porque não aparece “plot value” em nenhum?
- Não sei... ah... acho que é porque não é número.
- Ah, é verdade, acho que é isso mesmo.

Após algumas tentativas, os alunos intuíram que só era possível obter as medidas resumos para variáveis numéricas, ou seja, quantitativas, e evidenciamos o papel do ambiente computacional como elemento de validação: o fato de não permitir o cálculo das medidas de variáveis não numéricas.

Como o grupo não redigiu um texto final para o relatório, e escreveu apenas uma frase para cada representação, reproduzimos todas as frases do relatório a seguir.

- Os alunos entrevistados estão na faixa etária de 12 a 15 anos, sendo que a maioria possui 13 anos de idade e a minoria 12 anos.
- Os alunos gastam em média 23min para chegarem a escola.
- A maioria dos entrevistados são do sexo masculino.
- O esporte mais praticado é o futebol.
- Os estilos de músicas mais ouvidos são rock e black.
- As matérias preferidas são inglês e matemática.
- A maioria não sabe qual será sua profissão futura.

Podemos verificar que, para a idade, os alunos analisaram como os dados variavam, utilizando o valor mínimo e o valor máximo da distribuição, e a moda, explicando que a idade variava de 12 a 15 anos, e a maior parte dos alunos tinha 13 anos. Para chegar a esta conclusão, os alunos utilizaram apenas o gráfico. Em seguida, para analisar o tempo gasto para chegar à escola, os alunos utilizaram a média aritmética desta distribuição, que era 23 minutos. Neste momento, os alunos comentam sobre o tempo que cada um gastava para chegar à escola, comparando com o valor obtido na média, e chegaram a conclusões de que o tempo que eles gastavam realmente era muito próximo no valor obtido como média na pesquisa. Este tipo de comparação é interessante para o aluno, facilitando a compreensão do significado do valor obtido, vendo utilidade na sua pesquisa, como Godino (1995) cita que, deste modo, os alunos *poderiam ver que construir um sistema de dados próprio e analisá-lo não é o mesmo que resolver um problema de cálculo rotineiro de um livro didático*. Continuando a análise, percebemos que os alunos tendem a utilizar a característica mais freqüente para analisar os dados, utilizando muito a palavra “maioria” ou “mais”, ou seja, realizam a “apreensão perceptiva” e a “leitura dos dados” do gráfico, verificando os valores mais freqüentes nessa representação.

Após construir um gráfico para cada variável, o grupo utilizou o recurso de visualização simultânea dos gráficos no software Fathom, relacionando as variáveis para o gênero feminino e para o gênero masculino.



“A maioria das meninas tem 14 anos, praticam handebol, gostam de psy, preferem a matéria de português e querem seguir a profissão de psicóloga”.

Figura 99: Visualização simultânea dos gráficos construídos pelo grupo 1, para o gênero feminino



“Os meninos praticam mais futebol, gostam de inglês, a maioria possui 13 anos, querem seguir a profissão de atleta”.

Figura 100: Visualização simultânea dos gráficos construídos pelo grupo 1, para o gênero masculino

Na visualização simultânea dos gráficos, este grupo procurou estabelecer

uma relação entre todas as variáveis para o gênero feminino, e depois fez o mesmo para o gênero masculino. Podemos dizer que deste modo, o grupo realizou uma apreensão operatória dos gráficos, reorganizando as informações obtidas para cada gênero.

Resultados do segundo grupo

O segundo grupo, após discussão, decidiu utilizar em seu banco de dados as seguintes variáveis: idade, altura, esporte preferido, futura profissão e tipo de música preferido, escolhendo de forma aleatória um total de 30 amostras. Após as escolhas realizadas, o grupo digitou os dados no banco de dados do Fathom.

Para a representação a lápis e papel, o grupo escolheu a variável idade, construindo um gráfico de barras para representar esta variável, e também calculando o valor da média aritmética.

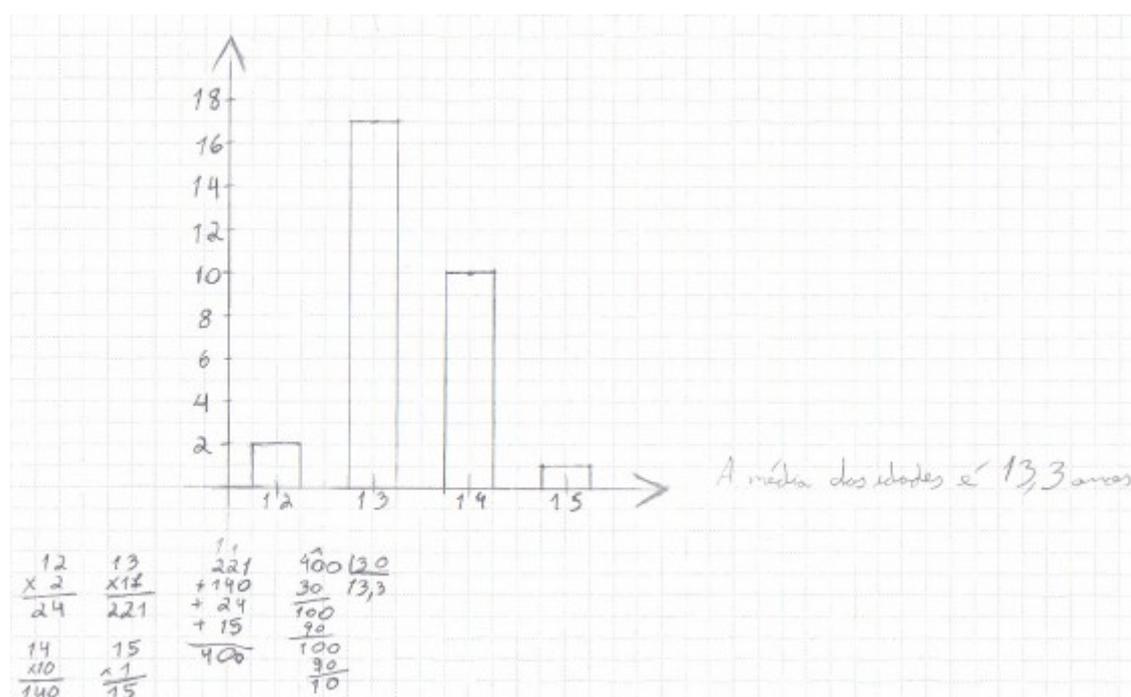


Figura 101: Gráfico da variável idade e cálculo da média aritmética (lápiz e papel)

Este grupo também apresentou avanços, conseguindo construir o gráfico de colunas e calcular a média aritmética, mas limitaram-se a esta medida.

O segundo grupo também escolheu a representação gráfica para montar seu relatório, construindo um gráfico para cada variável, e escrevendo logo após o gráfico uma frase com sua análise. O grupo construiu gráficos de pontos (Dot-

plot), para as variáveis idade e altura, e gráfico de barras para as outras variáveis, ou seja, como estes tipos de gráficos são escolhidos automaticamente pelo software, o grupo não optou por mudar para outros tipos de gráficos. Este grupo não utilizou mais nenhum outro tipo de representação, e nenhum outro recurso do software (utilizou o mínimo dos recursos disponíveis pelo software).

O grupo também não escreveu um texto final, mas somente uma frase para cada representação, conforme mostramos a seguir.

- A maioria dos alunos pesquisados possuem 13 anos de idade.
- A maioria dos alunos entrevistados possuem entre 1.65 e 1.70m.
- Grande parte dos entrevistados têm como esporte favorito o Futebol.
- Grande parte dos entrevistados têm como gosto musical o Rock.
- A maioria alunos entrevistados querem seguir a carreira de engenheiro.

Analisando as frases, percebemos que estes alunos também utilizam muito o termo “maioria”, limitando-se à “apreensão perceptiva” à “leitura dos dados” do gráfico e dos valores que são mais freqüentes.

Resultados do terceiro grupo

O terceiro grupo utilizou as seguintes variáveis em seu banco de dados: nome, sexo, idade, peso, estilo musical, esporte preferido, matéria preferida, escolhendo 30 elementos do banco de dados de forma aleatória, para depois iniciarem a digitação dos dados no Fathom.

Para representar a lápis e papel, este grupo escolheu a variável “sexo”, construindo o gráfico de colunas para esta variável.

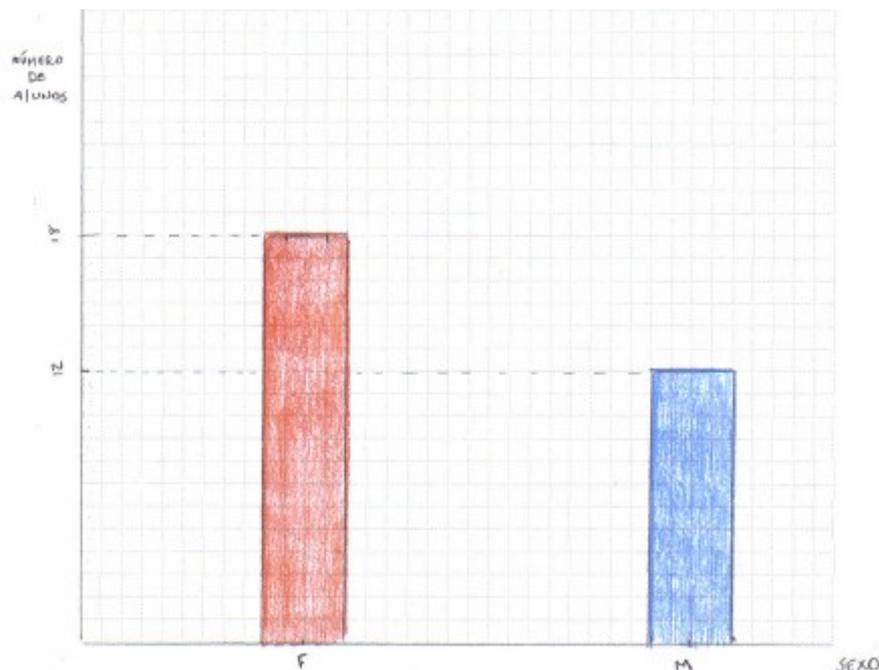


Figura 102: Gráfico de colunas para a variável “sexo”, construído pelos alunos

Os alunos queriam calcular as medidas resumo para esta variável, mas verificaram que:

- Não dá pra calcular a média, pois a variável não é numérica.
- É mesmo... e agora?
- Vamos calcular a média de outra então...

Então o grupo calculou a média aritmética e a mediana para a variável “idade”. Os alunos iniciaram o cálculo destas medidas, organizando os dados em uma tabela de distribuição de frequências.

Idade	Nº DE PESSOAS
12	1
13	17
14	12

MÉDIA

$$12 \cdot 1 = 12$$

$$13 \cdot 17 = 221$$

$$14 \cdot 12 = 168$$

$$12 + 221 + 168 = 401 \div 30 = 13.3$$

Mediana

São 30 alunos, a metade de 30 é 15; essa metade está na idade de 13 anos.

Figura 103: Tabela de distribuição de freqüências para a variável idade, e cálculo da média e mediana.

Analisando estes resultados, podemos verificar que houve avanços, pois o grupo representou uma variável qualitativa na forma gráfica e conseguiu também obter duas medidas-resumo para uma variável quantitativa, a média aritmética e a mediana, tarefa que este grupo não conseguiu desenvolver no teste-diagnóstico.

No Fathom, os alunos deste grupo também preferiram a representação gráfica. Para cada variável, os alunos construíram um gráfico de colunas que a representasse e, para variáveis quantitativas, representaram também a média aritmética no gráfico, como mostramos um exemplo a seguir.

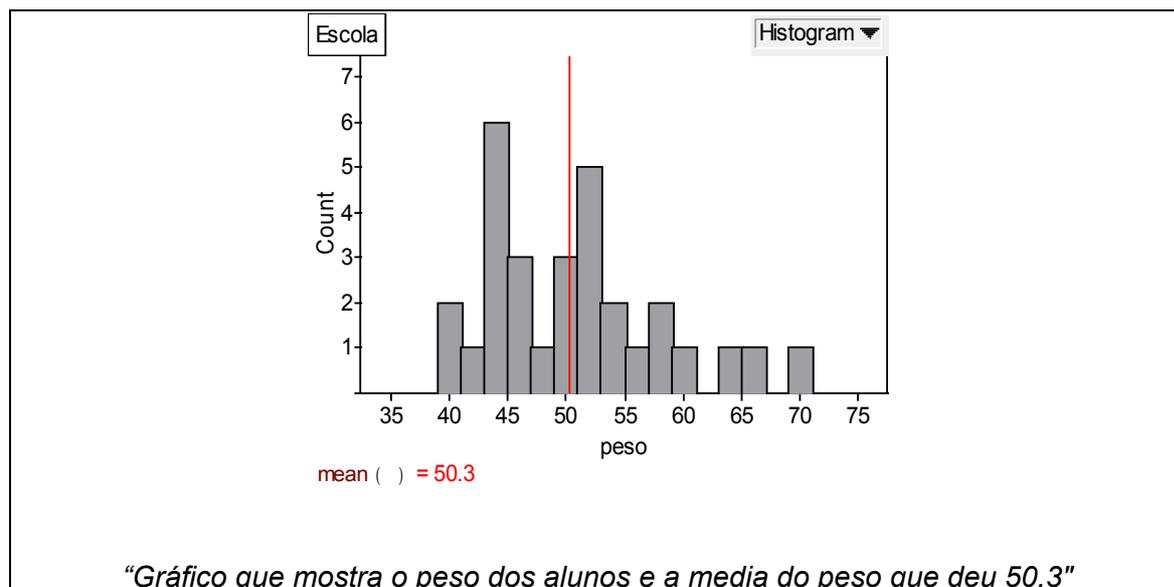


Figura 104: Gráfico para a variável peso, construído pelos alunos no Fathom

Para as variáveis quantitativas, os alunos também construíram tabelas para o cálculo da média aritmética, mas estas não foram utilizadas no relatório final. Eles usaram este valor para conferir os resultados do gráfico, verificando que era realmente o mesmo. Neste relatório, este grupo também não redigiu um texto final. Após cada representação, o grupo escreveu uma frase que a descrevesse. Mostramos, a seguir, estas frases:

- A maioria dos estudantes da oitava série da escola Oscavo tem preferência para o Rock e Black.
- Gráfico de idade com a média de idade com cerca de 13.3 anos.
- Gráfico que mostra o peso dos alunos e a media do peso que deu 50,3 kg.
- Da quantidade de alunos que gostam do esporte, a maioria é de Handball.
- A Matéria preferida dos alunos é Matemática.
- A maioria dos alunos são do sexo feminino.

Analisando as frases dos alunos, percebemos que este grupo também utiliza a palavra “maioria” ao escolher o valor representativo de uma distribuição, realizando uma “apreensão perceptiva” da figura. Para as variáveis quantitativas, os alunos utilizaram a média aritmética da distribuição, realizando uma “apreensão discursiva” da figura. Este grupo também costumava comparar os resultados obtidos na sua pesquisa com suas características pessoais, concluindo que estes resultados geralmente estavam próximos do que esperavam. Também

compararam os resultados obtidos na sua pesquisa com os obtidos pelos outros grupos, identificando resultados parecidos.

Resultados Gerais da Parte B

Após a análise dos resultados obtidos pelos grupos, percebemos que houve avanços em relação às atividades desenvolvidas no teste diagnóstico. Em relação aos gráficos, a tendência em construir o gráfico de pontos diminuiu, e todos os grupos construíram gráficos de barras para representar uma variável estatística. Em relação à média aritmética, os alunos também evoluíram, e todos os grupos calcularam (manualmente) a média aritmética da distribuição, levando em consideração a frequência dos dados, ou seja, calcularam a média ponderada. Em relação à mediana, dois grupos utilizaram esta medida, e pelas dúvidas levantadas durante o desenvolvimento da parte A da seqüência didática, percebemos que os alunos começam a compreender o significado desta medida. Em relação aos quartis, percebemos que ainda existem dificuldades relacionadas a estas medidas, já que elas não foram nem citadas pelos alunos na parte B, e estas dificuldades refletiram na compreensão do Box-Plot, ou seja, as atividades propostas não foram suficientes para que os alunos pudessem construir o conceito de forma significativa.

Para representar os dados, os alunos preferem os gráficos, e para analisar estas representações, os alunos realizam a “apreensão perceptiva” deles, identificando os dados que são mais evidentes, com a palavra “maioria”. No caso das variáveis quantitativas, em alguns casos, os alunos representaram a média aritmética no gráfico, realizando a “apreensão discursiva” da figura, estando no nível da “leitura entre os dados”.

Os alunos não utilizaram muitos recursos disponíveis pelo software. A utilização destes recursos poderia ter contribuído com uma análise mais rica, propiciando diferentes visualizações dos dados e explorando o máximo de informações, de acordo com os objetivos da análise exploratória de dados.

Os alunos foram pouco receptivos às atividades em que era necessário produzir justificativas e, conseqüentemente, não elaboraram um texto explicativo na parte B, como foi solicitado. Observa-se que, na parte A, não foi solicitado ao

aluno um relatório, mas ele deveria responder questões que o ajudariam a resolver o problema proposto. Dessa forma, a redação de um texto era uma atividade ainda desconhecida e, por isso, problemática, constituindo provavelmente uma quebra do contrato didático usual em uma aula de matemática: a produção de textos não é uma atividade para essa disciplina.

Em relação ao Pensamento Estatístico, verificamos que os grupos transitaram por três das quatro dimensões de sua estrutura (elaborada por WILD e PFANNKUCH, 1999): no planejamento do ciclo investigativo; nos tipos de pensamento, realizando a transnumeração, organizando e analisando os dados; no ciclo interrogativo, interpretando os resultados e comparando os dados. Contudo, não apresentaram as disposições para investigar mais ou questionar os resultados obtidos, mostrando assim que o pensamento estatístico ainda não está desenvolvido nesses alunos.

Considerações Finais

O objetivo deste trabalho foi estudar as interações entre o aluno e um ambiente de estatística dinâmica, no caso, o software Fathom, segundo a abordagem da Análise Exploratória de Dados.

Segundo Batanero (1991), uma idéia fundamental da Análise Exploratória de Dados é que o uso de múltiplas representações de um conjunto de dados propicia o desenvolvimento de novos conhecimentos e idéias.

Duval (2003) afirma que a originalidade da atividade matemática está na mobilização simultânea de ao menos dois registros de representação, ou na possibilidade de trocar, a todo o momento, de registro de representação. Assim, a compreensão em matemática implica a capacidade de transitar entre mais de um registro de representação semiótica.

Considerando a importância da multiplicidade de registros de representação para o desenvolvimento da Análise Exploratória de Dados, buscamos em nosso trabalho estabelecer uma leitura da teoria de Duval para a representação dos conceitos estatísticos, principalmente no que diz respeito aos tipos de apreensões de uma figura (DUVAL, 1994). Para tanto, estudamos também os Níveis de Compreensão Gráfica, de Curcio (1989), buscando identificar as condições de mobilização simultânea dessas duas teorias.

Vale ressaltar que a utilização da teoria dos Registros de Representação Semiótica, de Duval, na área de Educação Estatística, é relativamente recente, e por isso algumas questões em nosso trabalho poderão ficar em aberto, já que nossa pesquisa é uma das primeiras a trabalhar com tal abordagem. É importante também ressaltar que as relações que buscamos estabelecer entre essas teorias são apenas hipotéticas, pois a equivalência de tal relação implica a realização de um estudo empírico, que não foi o foco deste trabalho.

Ainda considerando a importância da multiplicidade de registros de representação, escolhemos trabalhar com o software Fathom por possibilitar o acesso rápido aos diferentes tipos de representações de um conjunto de dados. Para tanto, elaboramos uma seqüência didática de atividades desenvolvidas com o uso do software, com base nos pressupostos da Engenharia Didática

(ARTIGUE, 1988), metodologia de pesquisa que julgamos ser a mais adequada para coletar os dados que permitirão responder nossa questão de pesquisa.

Antes de iniciar o trabalho com as atividades da seqüência didática, aplicamos um teste-diagnóstico nos alunos, sujeitos de pesquisa, no intuito de levantar os conhecimentos prévios destes alunos em relação aos conceitos estatísticos.

Analisando os resultados gerais deste, verificamos que os alunos conseguiam ler dados em gráficos e tabelas, porém apresentavam dificuldades em representar os dados, ou seja, apresentavam dificuldades na construção de gráficos e tabelas, com a tendência de construir gráfico de pontos na maioria dos casos (talvez influenciados pelo ensino do conceito de função do primeiro ano do Ensino Médio). Tal dificuldade também foi apontada em outros trabalhos da nossa revisão bibliográfica, como os exemplos citados por Li y Shen (1992, apud BATANERO, 2001), que relata erros na construção de gráficos por estudantes de Ensino Médio, como a utilização de gráfico de pontos para variáveis qualitativas.

Em relação à média aritmética, os alunos apresentavam dificuldades no seu cálculo, principalmente nos casos que não se tratava da média aritmética simples, dificuldade também citada por Batanero (2000), que afirma que, tal erro acontece, pois aplica-se o algoritmo de forma mecânica e sem compreender seu significado. Os alunos utilizavam o conceito de moda intuitivamente, para escolher o valor representativo de uma distribuição. E em relação as outras medidas-resumo, verificamos que os alunos não as conheciam.

Iniciamos os trabalhos da seqüência didática com as atividades de familiarização com o Fathom, para que os alunos pudessem conhecer algumas das ferramentas disponíveis pelo software. Os alunos foram bem receptivos ao trabalho neste ambiente, e se adaptaram rápido ao uso das ferramentas trabalhadas.

Enquanto familiarizavam-se com o software, os conceitos estatísticos iam sendo trabalhados ou revisitados. As interações com o ambiente e com os integrantes do grupo provocavam discussões que contribuía com a construção de novos conhecimentos. Os alunos tinham a oportunidade de visualizar tabelas e diferentes gráficos de uma mesma distribuição de freqüências, e ainda visualizar medidas-resumo nas tabelas e nos gráficos.

Durante o desenvolvimento desta parte da seqüência didática, percebemos que as interações com o ambiente informatizado e as interações entre os grupos contribuíram com a compreensão de conceitos como a média aritmética e a mediana, que foi muito discutida, e também com a análise e interpretação de gráficos de colunas (Histograma) e de pontos (Dot-Plot). No entanto estas variáveis ainda foram insuficientes na compreensão de medidas como os quartis, e do gráfico Box-Plot.

O estudo de Baker, Biehler e Konold (2004) apresenta dificuldades parecidas no trabalho com o gráfico Box-Plot. Os autores citam que tais dificuldades acontecem, pois, para compreender este tipo de gráfico, é necessário ter conhecimentos sobre quartis, e a divisão dos dados em quartis é de difícil compreensão, e a mediana não é tão intuitiva aos estudantes como os pesquisadores esperavam. Por isso é necessário organizar atividades específicas e em maior número para facilitar a compreensão da mediana e dos quartis.

Então, acreditamos que, em nosso trabalho, como os alunos tiveram contato com tantos conceitos novos, seria necessário mais tempo para que pudessem compreender os significados dos quartis e, por conseguinte, a construção do gráfico Box-Plot.

Ao final da parte A das atividades, os alunos responderam as questões de interpretação, em que poderiam mobilizar as diferentes representações para obter as informações necessárias. Neste momento, verificamos que os alunos preferiam consultar estas informações nas tabelas, fazendo pouco uso dos gráficos, evidenciando a pouca familiaridade no uso desse tipo de registro de representação semiótica de uma distribuição de freqüências.

Nas atividades da parte B da seqüência didática, os alunos foram envolvidos em atividades que trabalhavam desde a coleta dos dados, até a representação e análise deles.

Nesta etapa, verificamos que os grupos transitaram por três das quatro dimensões da estrutura do Pensamento Estatístico (WILD e PFANNKUCH, 1998). No ciclo investigativo, os alunos definiram o que pesquisar e como coletar, organizar e analisar estes dados. Na segunda dimensão, os alunos escolheram qual tipo de representação dos dados proporcionava uma melhor compreensão deles, organizando e analisando os dados. No ciclo interrogativo, os alunos

interpretaram o resultado de sua pesquisa, resumindo e comparando os dados. Porém não apresentaram disposições para investigar mais ou questionar os dados, ou seja, o pensamento estatístico ainda não está desenvolvido nesses alunos.

Todos os grupos preferiram utilizar gráficos de colunas para representar os dados. Dois grupos utilizaram também a representação gráfica das medidas-resumo em seu relatório, indicando o aumento na familiaridade no uso desse registro de representação semiótica quando comparado à primeira fase da atividade.

Na representação com lápis e papel, verificamos que houve avanços na construção de gráficos em relação ao teste diagnóstico. Acreditamos que o ambiente informatizado pode ter contribuído com estes avanços, na visualização de diferentes tipos de gráficos. Ainda neste item, verificamos avanços também em relação ao cálculo da média aritmética ponderada, e na obtenção da mediana. Acreditamos que estes avanços foram propiciados pelas interações com o ambiente e com os grupos.

Em relação à análise dos dados, os alunos não redigiram um relatório como foi solicitado, mas escreveram frases soltas para resumir cada distribuição. Ao levarmos em consideração que estes alunos nunca tinham trabalhado com este tipo de atividade, que exigia a redação de uma análise estatística, podemos afirmar que os resultados obtidos foram significativos, já que os alunos começaram a escrever uma pequena análise.

Ainda em relação a esta atividade, verificamos que os alunos utilizaram poucos recursos do software. Somente em alguns casos analisavam a variação dos dados, e utilizaram por várias vezes a palavra “maioria”, escolhendo os valores mais freqüentes para representar uma distribuição.

Neste momento, podemos retomar as questões que motivaram esta pesquisa:

- *Em que medida as articulações entre os diferentes tipos de registros de representações semióticas podem constituir o acesso à compreensão em estatística?*
- *O uso de um software de estatística dinâmica pode gerar/potencializar a necessidade de diversos registros de representações semióticas, e permitir*

que o sujeito desenvolva a capacidade de organizar, analisar e interpretar dados?

- *Quais tipos de articulações entre os registros de representações semióticas de um conjunto de dados podem ser identificados, a partir do uso do ambiente computacional, visando à análise crítica dos dados?*

Considerando as reflexões apresentadas nas análises e nesta etapa do nosso trabalho, e baseando-se no que pudemos observar na evolução conceitual dos alunos durante a seqüência didática que trabalhamos, podemos responder a essas questões.

Quanto à primeira delas, pudemos perceber a importância da articulação dos diferentes tipos de registros de representação semiótica, como entre tabelas e gráficos, representação numérica e gráfica de uma medida como a média ou a mediana, para a compreensão destes conceitos pelos alunos.

Quanto à segunda questão, pudemos perceber que a influência do ambiente computacional foi fundamental no sentido de possibilitar a visualização dos diferentes tipos de registros de representação semiótica de uma distribuição de freqüências, principalmente no caso da representação gráfica de medidas como a média e a mediana, contribuindo para a compreensão destes conceitos. Apesar de os alunos ainda não terem compreendido o significado de outros conceitos, como os quartis e a construção do Box-Plot, acreditamos que o uso desse ambiente possa contribuir com a compreensão de tais conceitos, desde que seja desenvolvida uma seqüência de atividades específica, com maior tempo de duração. Fazemos também a hipótese sobre a necessidade de uma associação explícita entre o Box-plot e o Dot-plot, permitindo a visualização das relações que se pode estabelecer na análise da distribuição dos dados. Tal associação permitiria também a melhor atribuição de significados aos quartis.

Quanto à terceira questão, percebemos, durante nossas análises, que a utilização do ambiente computacional possibilitou aos alunos a articulação de diferentes representações de um mesmo objeto, além de permitir a visualização simultânea destes vários registros, propiciando a obtenção de um maior número de informações sobre os dados. No entanto os alunos nem sempre faziam uso dessa multiplicidade de informações, e na análise geral das atividades,

constatamos que os alunos preferiam as tabelas para ler os dados, ou procurar informações (parte A das atividades), enquanto que para representar os dados, preferiam os gráficos (parte B das atividades). Uma primeira questão que se apresenta como perspectiva a ser investigada mais profundamente diz respeito ao tipo de mobilização de conhecimentos e ao tipo de mente matemática dos alunos, nos termos de Krutestki (1976, apud CAZORLA, 2002), que explicariam tais escolhas.

As reflexões apresentadas nesse estudo também sugerem outras questões. Uma delas seria a verificação da equivalência entre as teorias de Duval (1994) sobre os tipos de apreensões de uma figura, e de Curcio (1989) sobre os Níveis de Compreensão Gráfica, verificação que pode ser realizada através de estudos empíricos, necessitando de ferramentas de validação estatística.

Uma das limitações deste trabalho diz respeito ao desenvolvimento da compreensão do conceito de quartis e da construção do Box-Plot. Uma terceira questão a ser investigada em trabalhos futuros é se o desenvolvimento de uma seqüência didática com uso do software Fathom simultaneamente ao uso de ambiente não computacional, que dedicasse um maior tempo para o estudo destes conceitos, possibilitaria a sua compreensão.

Referências

ALMOULOU, S. A. Registros de representação semiótica e compreensão de conceitos geométricos. In: MACHADO, S. D. A. (org.). **Aprendizagem em Matemática: Registros de representação semiótica**. Campinas, SP: Papyrus, 2003, p.125-147.

ARTIGUE, M. Engenharia Didática. In: BRUN, J. **Didáctica das Matemáticas**. Lisboa: Instituto Piaget, 1996, p. 193-218.

AVILA, R. (org.). **Paquete Didáctico de Estadística y Probabilidad I**. Universidade Nacional Autónoma de México, 2005.

BAKER, A, BIEHLER, R., KONOLD, C. **Should young students learn about box plots?** Curricular Development in Statistics Education. Sweden, 2004, p. 163-173. Disponível em http://www.stat.auckland.ac.nz/~iase/publications/rt04/4.2/Bakker_etal.pdf (acesso em 01/04/08, às 17:55)

BATANERO, C. **Didáctica de la Estadística**. Universidad de Granada. Espanha, 2001: <http://www.ugr.es/~batanero>.

BATANERO, C. **Significado y comprensión de las medidas de posición central**. Universidad de Granada, 2000, UNO 25, 41-58.

BATANERO, C. ESTEPA, A. GODINO, J. **Análisis exploratório de datos: sus posibilidades em la enseñanza secundaria**. Granada, 1991. Suma, n.9, p.25-31.

BATANERO, C., GODINO, J. D., GREEN, D. R., HOLMES, P., VALLECILLOS, A. **Errores y dificultades em la comprensión de los conceptos estadísticos elementales**. *International Journal of Mathematics Education in Science and Technology*, 1994, p.527-547.

BIEHLER, R. KOMBRINK, K. **Muffins: Statistik mit komplexen Datensätzen – Freizeitgestaltung und Mediennutzung von Jugendlichen**. Universität Kassel: 2003. Erscheint in *Stochastik in der Schule*, 23, p. 11-26. <http://www.mathematik.uin-kassel.de/didaktik/mathelehren.htm>

BIEHLER, R. Perspectivas educativas del análisis de datos. In: MORRIS, R. **Estudios em educación matemática: La enseñanza de la estadística**. Paris: UNESCO, 1989. (Traducción: Ocar Dodera)

BRASIL, Ministério da Educação. Secretaria de Educação Média e Tecnologia. **PCN+ Ensino Médio: Orientações Educacionais complementares aos Parâmetros Curriculares Nacionais**. Brasília: Ministério da Educação, 2002.

BUSSAB, W.O., MORETTIN, P. A. **Estatística Básica**. São Paulo: Editora

Saraiva, 2002.

CAMPOS, C. R. **A Educação Estatística: Uma investigação acerca dos aspectos relevantes à didática da estatística em cursos de graduação**. Rio Claro: UNESP, 2007. Doutorado. Orientadora: Maria Lucia Lorenzetti Wodewotzki

CAZORLA, I. M. **A relação entre a habilidade viso-pictórica e o domínio de conceitos estatísticos na leitura de gráficos**. Campinas: FE-Unicamp, 2002. Doutorado. Orientadora: Márcia Regina F. de Brito.

CAZORLA, I. M., SANTANA, E. R. S. **Tratamento da informação para o ensino fundamental e médio**. Itabuna: Via Literarum, 2006.

COSTA, S. F. **Introdução Ilustrada à Estatística**. Editora Harbra. São Paulo, 1992.

CURCIO, F. R. **Developing Graph Comprehension**. National Council of Teachers of Mathematics, New York, 1989.

_____. **Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications**. Journal for Research in Mathematics Education, Vol. 32, N. 2, (Mar., 2001), p. 124-158.

DUVAL, R. **Les différents fonctionnements possibles d'une figure dans une démarche géométrique**. Repères, n.17, p. 121-138, 1994.

_____. **Conversion et articulation des représentations analogiques**. Séminaire IUFM. Nord Pas-de-Calais, 1999.

_____. Registros de representação semiótica e funcionamento cognitivo da compreensão em matemática. In: MACHADO, S. D. A. (org.). **Aprendizagem em Matemática: Registros de representação semiótica**. Campinas, SP: Papirus, 2003, p.11-33.

_____. **Semiosis e pensamiento humano. Registros semióticos y aprendizajes intelectuales**. Universidad del Valle, Grupo de Educación Matemática, 2004. Traducion: Myrian Vega Restrepo, 1999.

GODINO, J. D. **¿Qué aportan los ordenadores al aprendizaje y la enseñanza de la estadística?** UNO, 5, 1995, p. 45-56.

GRAHAM, A. **Developing Thinking in Statistics**. Open Unviersity. Paul Chapman Educational, 2006

MAGALHÃES, M. N.; LIMA, A. C. P. **Noções de Probabilidade e Estatística**. São Paulo, SP: EDUSP, 2005.

MORETTI, M. T. A translação como recurso no esboço de curvas por meio da

interpretação global de propriedades figurais. In: MACHADO, S. D. A. (org.). **Aprendizagem em Matemática: Registros de representação semiótica**. Campinas, SP: Papyrus, 2002, p.149-160.

MORETTIN, L. G. **Estatística Básica, Inferência**. São Paulo: Makron Books, 2000.

PONTE, J. P., BROCARD, J., OLIVEIRA, H. **Investigações matemáticas na sala de aula**. Belo Horizonte, MG: Autêntica, 2005.

SANCHÉZ, E. S **Teachers' beliefs about usefulness of simulation with the educational software fathom for developing probability concepts in statistics classroom**. ICOTS 6, 2002.

SILVA, C. B. **Pensamento estatístico e raciocínio sobre variação: um estudo com professores de matemática**. São Paulo: PUC/SP, 2007. Doutorado. Orientadora: Prof. Dr. Cileda de Queiroz e Silva Coutinho.

WILD, C. J., PFANNKUCH, M. **What is Statistical Thinking?** Proceedings of the Fifth International Conference on Teaching of Statistics. Singapura: 1998, vol.1, p. 333-339.

_____. **Statistical Thinking in Empirical Enquiry**. International Statistical Review. Auckland: 1999, n. 67, p.223-265.

ANEXOS

TESTE DIAGNÓSTICO

1) Em certa empresa, a distribuição do número de funcionários (segundo seu gênero), para os diversos cargos e seus respectivos salários é a seguinte:

Cargo	Gênero	Nº de funcionários	Salário (em R\$)
Gerente	Masculino	1	3500
Auxiliar Administrativo	Masculino	3	820
Auxiliar Administrativo	Feminino	5	820
Vendedores	Masculino	12	540
Vendedores	Feminino	10	540
Serventes	Masculino	3	400
Serventes	Feminino	3	400

Responda:

- a) Quantos funcionários tem a empresa? _____
- b) Quantos funcionários são do gênero feminino? _____
- c) Qual a porcentagem de funcionários do gênero masculino? _____
- d) Quais são as variáveis desta pesquisa? _____

- e) Você considera que as variáveis da pesquisa são de mesmo tipo ou são de tipos diferentes? Justifique.

- f) Qual a média salarial dos funcionários desta empresa? _____
- g) Se colocarmos esses funcionários em ordem crescente de salários, qual seria a posição do funcionário que dividiria essa fila em dois grupos como o mesmo número de funcionários. Escreva o cargo e o salário deste funcionário.

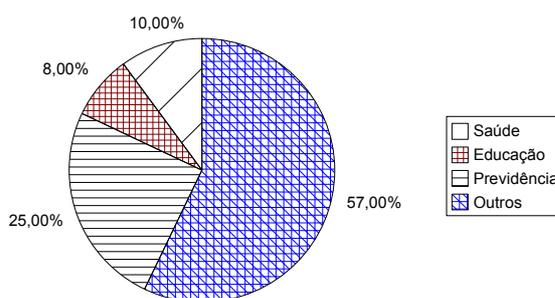
- h) Qual o salário que você considera que melhor representa o salário pago por esta empresa? Por quê?

i) Construa um gráfico que represente os cargos dos funcionários desta empresa.

j) Construa um gráfico que represente os gêneros dos funcionários desta empresa.

k) Construa um gráfico que represente os salários dos funcionários desta empresa.

2) O gráfico abaixo, publicado pela revista Veja (1999), mostra como são divididos os 188 bilhões de reais do orçamento da União, entre os setores de saúde, educação, previdência e outros.



Construa uma tabela que mostre os valores em reais gastos em cada um dos setores.

3) Para analisar um conjunto de dados, podemos utilizar medidas que resumem

os dados e descrevem suas características, permitindo assim a análise da variação destes dados. Abaixo, indique quais destas medidas que você conhece:

() Média Aritmética

() Mediana

() Moda

() Quartis

4) As idades das quarenta pessoas que prestaram um concurso de seleção para uma determinada instituição são as seguintes

18, 18, 20, 22, 21, 18, 19, 20, 22, 18,
19, 20, 21, 18, 20, 19, 20, 19, 22, 19,
20, 21, 20, 18, 19, 18, 22, 21, 20, 19,
20, 18, 21, 18, 18, 19, 20, 19, 18, 18

a) Construa um gráfico de barras que represente as idades destas pessoas.

b) Quantas pessoas têm 20 ou menos de 20 anos? _____

c) Qual idade você considera mais apropriada para representar esta distribuição? Justifique.

d) Tente relacionar cada um dos valores das idades desta distribuição com as respectivas medidas:

I. Média Aritmética

() 18 anos

II. Mediana

() 19,5 anos

III. Moda

() 19 anos

IV. Primeiro Quartil

() 20 anos

V. Terceiro Quartil

() 18 anos

SEQUÊNCIA DIDÁTICA

Parte A - Familiarização com o software

Realizaremos essas atividades com o **software de estatística dinâmica Fathom**. Nesta fase, nosso objetivo é que você conheça e familiarize-se com o software, revendo (ou conhecendo) alguns conceitos estatísticos.

Observe o seguinte conjunto de dados, de uma pesquisa realizada com 20 funcionários de uma empresa:

Nome	Gênero	Idade	Cargo	Escolaridade	Salário
Ana	f	23	recepcionista	médio	850
Fabio	m	25	operário	fundamental	950
Mario	m	30	operário	fundamental	950
Mauricio	m	35	gerente	superior	2000
Carlos	m	29	engenheiro	superior	2200
Mariana	f	31	gerente	superior	2000
Lourdes	f	26	vendedor	médio	1300
Sergio	m	27	operário	médio	950
Marilda	f	36	aux limpeza	fundamental	700
Rose	f	32	aux limpeza	fundamental	700
Ricardo	m	30	vendedor	médio	1700
Rafael	m	36	engenheiro	superior	2200
Raquel	f	40	encarregado	superior	1300
Cassio	m	38	operario	médio	950
Eduardo	m	33	operário	fundamental	950
Felipe	m	35	encarregado	médio	1200
Michel	m	34	vendedor	médio	1500
João	m	40	operário	fundamental	950
Miguel	m	42	porteiro	fundamental	800
Maria	f	43	aux limpeza	fundamental	700

A partir deste conjunto de dados podemos levantar inicialmente as seguintes questões:

- Quais as características dos sujeitos observados que estão sendo alvo desta pesquisa? _____
- Se formos classificar estas características, quais critérios poderiam ser usados? Por quê? _____

Obs.: Vale lembrar que as características observadas constituem o que

chamamos de “variável estatística”.

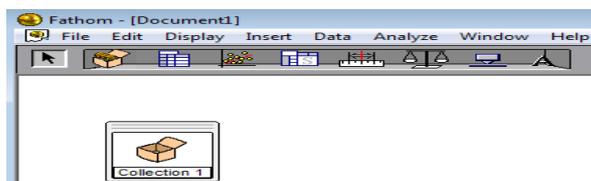
Primeira Fase: Elaboração de tabelas no Fathom

1) Inicialmente, construiremos o banco de dados da empresa no software Fathom.

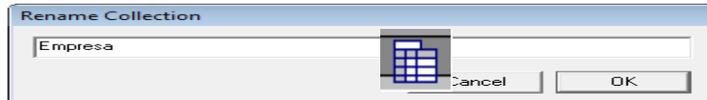
Para isso, siga os procedimentos abaixo.

a) Clique no ícone para iniciar o Fathom.

b) Clique em “**new collection**”, e arraste até a tela.



c) Clique duas vezes em “**collection1**”; aparecerá a caixa “rename collection”, onde você pode inserir um nome para o seu banco de dados.



d) Agora, clique em “**new case table**”, e arraste até a tela.

e) Insira os dados da pesquisa na tabela do Fathom. Para isso, clique em “**New**” e escreva o título da coluna. Depois, tecele “enter”, para continuar inserindo os dados nas linhas. (*Obs.: não utilize acentuação, pontuação ou espaçamentos*).

Empresa			
	nome	genero	<new>
1	Ana		
2	Fabio		
3	Mario		

2) Agora, construiremos uma tabela de distribuição de freqüências para a variável “**gênero**”.

a) Clique em “**new summary table**” ;

b) No banco de dados do Fathom, selecione a coluna “**gênero**” e arraste-a até a nova tabela.

pesquisa	Summary Table			
↓	⇒	gênero		Row Summary
		f	m	
		7	13	20

S1 = count ()

3) Seguindo o mesmo procedimento, construa também as tabelas de distribuição de freqüências para as variáveis “cargo” e “escolaridade”.

4) Você considera que a variável “gênero” e a variável “idade” são de mesmo tipo? Por quê? _____

5) Construa uma tabela de distribuição de freqüências para a variável “idade”.

Para isso, clique em “**new summary table**” , selecione a coluna da variável “**idade**” no banco de dados e arraste-a até a nova tabela.

Obs.: para tabelas com dados numéricos (ou seja, variáveis quantitativas), mantenha apertada a tecla **shift** enquanto arrasta a variável até a tabela.

6) Construa uma tabela de distribuição de freqüências para a variável “salário”.

7) Podemos também construir tabelas combinando algumas variáveis. Por exemplo, vamos construir uma tabela que mostre o nível de escolaridade dos funcionários de gênero masculino e feminino.

a) Clique em “**new summary table**” para construir a tabela;

b) Clique na variável “**escolaridade**” e arraste até o lado **direito** da tabela:

pesquisa	Summary Table				
↓	⇒	escolaridade			Row Summary
		fundamental	medio	superior	
		8	7	5	20

S1 = count ()

c) Agora, clique na variável “**gênero**” e arraste até o lado de **baixo** da tabela.

pesquisa		Summary Table			
		escolaridade			Row Summary
		fundamental	medio	superior	
gênero	f	3	2	2	7
	m	5	5	3	13
Column Summary		8	7	5	20

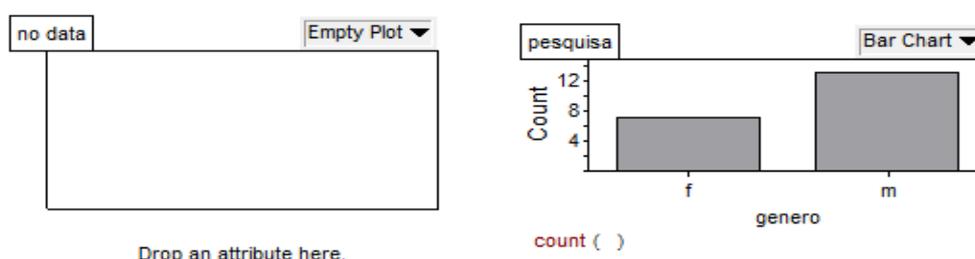
S1 = count ()

8) Construa uma tabela que mostre os cargos dos funcionários de gênero masculino e feminino.

Segunda fase: Construção de gráficos no Fathom

1) Nesta fase, utilizaremos o software Fathom para construir os **gráficos**, que constituem um outro tipo de representação de uma Distribuição de Frequências.

- a) Clique no ícone  para construir o gráfico, e arraste até a tela;
- b) Clique na coluna da variável “**gênero**” para selecioná-la, e a arraste até a área do gráfico, sobre a caixa “**Drop an attribute here**”.



Este gráfico é chamado “diagrama de colunas”, e é normalmente utilizado para representar variáveis qualitativas ou mesmo variáveis quantitativas discretas. Ele coloca em evidência a comparação entre os valores observados.

2) Clique no ícone “bar chart” escolha o outro tipo de gráfico. O que você observa? É o mesmo tipo de comparação que ele representa?

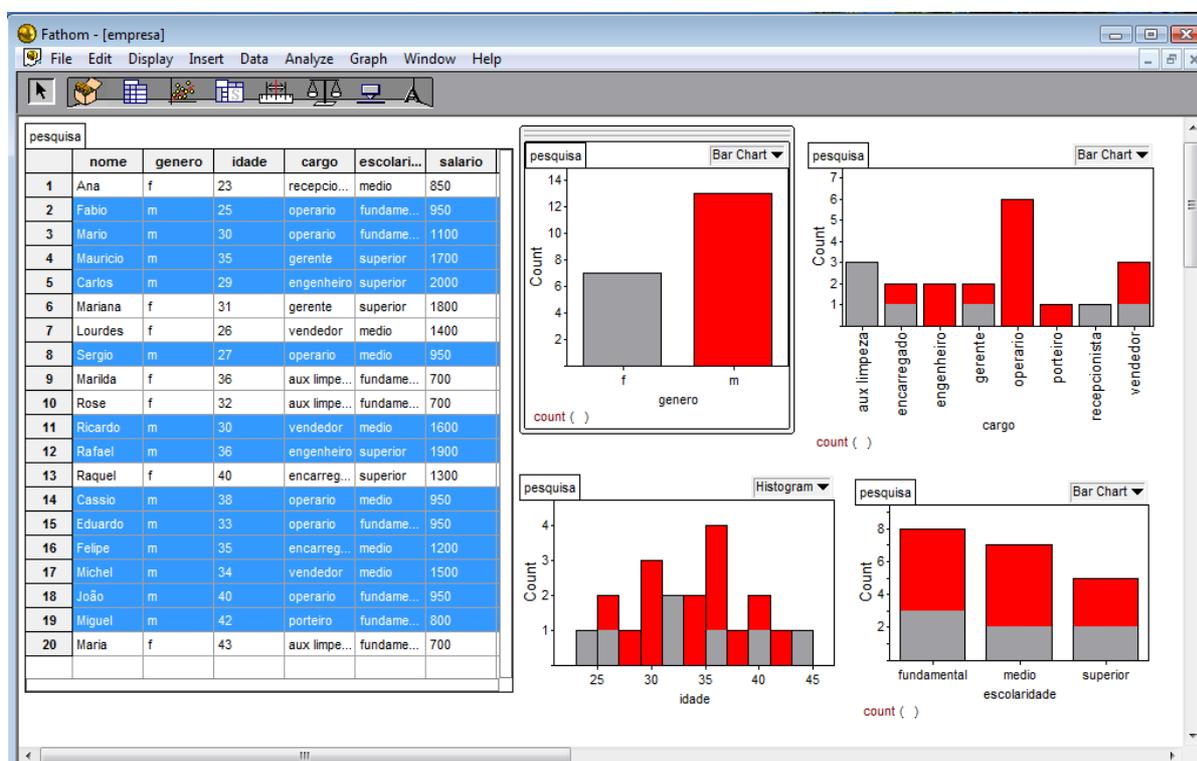
3) Construa os gráficos para as variáveis “cargo” e “escolaridade”.

4) Construa agora um histograma para a variável idade. (Escolha o Histogram).

5) Construa um gráfico de pontos para a variável “idade” (Escolha Dot Plot). Explique a diferença entre os dois gráficos. Como você explicaria a variação dessa variável pela observação dos dois gráficos? _____

Obs.: A variável “idade” é uma variável quantitativa contínua, mas pode ser tratada como discreta, por ser apresentada em número de anos completos (uma contagem). O software Fathom, por definição, trata toda variável numérica como contínua. Então, para construir o gráfico Histograma, o software faz uma “correção de continuidade”, ou seja, soma e subtrai meia unidade ao valor da variável, construindo assim um intervalo e propiciando a percepção da continuidade. Sendo assim, para variáveis discretas, é mais adequado utilizar o gráfico Dot Plot (neste tipo de gráfico, cada ponto representa um valor observado).

6) Agora, no gráfico da variável gênero, clique sobre a barra que indica o gênero masculino. O que você observa no banco de dados e nos outros gráficos?



Terceira fase: Medidas-Resumo, Tabelas e Gráficos no Fathom

1) Construa uma tabela de distribuição de frequências para a variável “idade”, como você fez na primeira fase, mas agora sem apertar a tecla shift (somente arraste a variável até a tabela).

pesquisa	Summary Table
↓	⇒ idade
	33.25

S1 = mean ()

Observe que apareceu somente um valor na tabela. Este é o valor da “média aritmética” para a variável “idade” (no Fathom, quando construímos tabelas de variáveis quantitativas, se você **não** mantiver apertada a tecla shift, a tabela mostra a média aritmética).

A média aritmética é uma medida que resume e representa um conjunto de dados em um único valor. A média de um conjunto é o valor ao redor do qual os valores observados se distribuem, e calculada como a soma dos valores dividida pelo número total de observações. Podemos fazer uma analogia entre a média de um conjunto de dados e o fiel de uma balança em equilíbrio.

2) Seguindo o mesmo procedimento, obtenha o valor da média aritmética para a variável “salário”

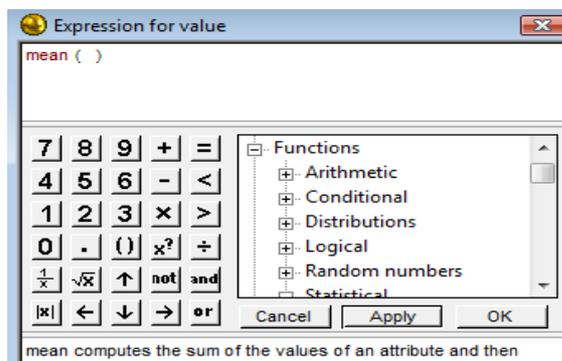
3) Iremos agora representar o valor da média no gráfico.

a) Construa um gráfico (histogram) para a variável “salário”;

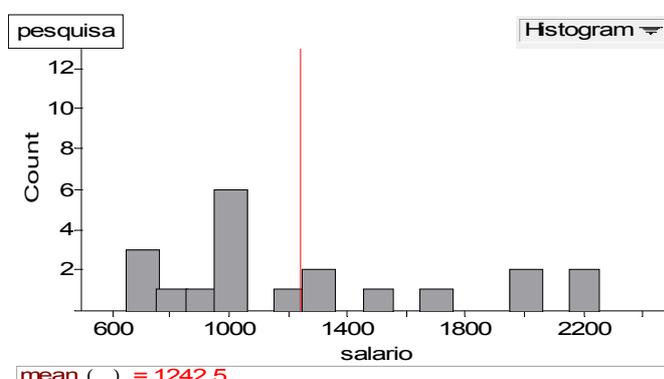
b) Clique em “**Graph**”, “**Plot Value**”;

c) No menu que se abrirá, clique em “**Functions**”, “**Statistical**”, “**One Attribute**”.

d) Clique (duas vezes) em “**Mean** ”.



Veja o gráfico:



“A média sozinha é suficiente para que se possa analisar um conjunto de dados? Pense, por exemplo: Qual o intervalo de variação dos valores observados? Qual é a posição da média dentro deste intervalo?”

Uma outra medida que podemos utilizar para sintetizar e representar um conjunto de dados é a mediana. A mediana é o valor que ocupa a posição central dos dados ordenados, ou seja, é o valor que é maior ou igual a 50% dos dados ordenados e menor ou igual a 50% dos dados observados. Podemos fazer uma analogia entre a mediana e o ponto médio de um segmento.

4) Na mesma tabela da média das idades (construída no item 1), vamos mostrar também a mediana.

a) No Fathom, clique na tabela acima, depois clique em “Summary”, “Add

Formula”;

b) No menu, clique em **“Functions”, “Statistical”, “One Attribute”**. Clique (duas vezes) em **“median”** e **“Ok”**.

Você terá o valor que ocupa a posição central nas idades ordenadas.

pesquisa	Summary Table
↓	idade
	33.25
	33.5

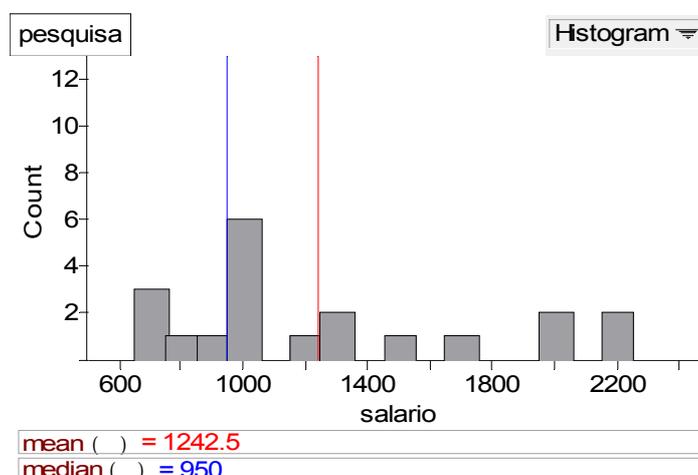
S1 = mean ()
S2 = median ()

5) Representaremos a mediana no gráfico (pode ser no mesmo gráfico do item 3).

a) Clique no gráfico da variável salário que você construiu a média;

b) Clique em **“Graph”, “Plot Value”**;

c) No menu que se abrirá, clique em **“Functions”, “Statistical”, “One Attribute”, “Median (duplo clique)”**.



6) Como os demais valores representados se distribuem em relação à média? E em relação à mediana? _____

Vamos determinar, além da mediana, dois valores tais que 25% das observações ordenadas serão menores ou iguais a um deles e 75% serão menores ou iguais

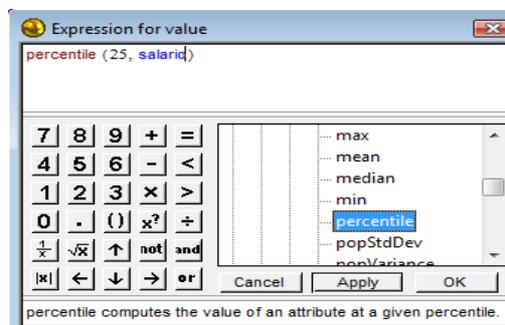
ao outro. Tais valores são denominados, respectivamente, **primeiro quartil** e **terceiro quartil**. Note que a mediana representa o segundo quartil.

7) Para representar os quartis, utilizaremos ainda o mesmo histograma da variável salário.

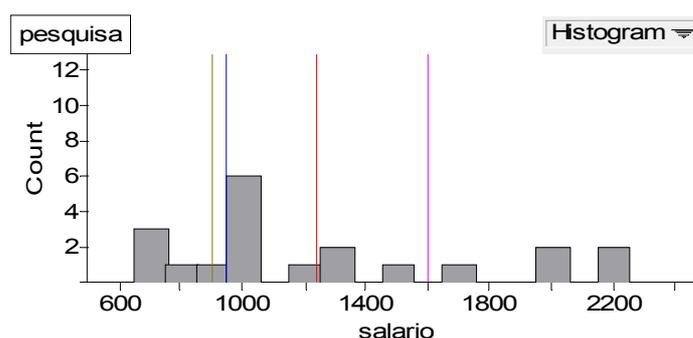
a) Selecione o gráfico e clique em “**Graph**”, “**Plot value**”;

b) No menu, clique em “**Functions**”, “**Statistical**”, “**One Attribute**”, “**Percentile**” (duas vezes)

c) Na caixa de funções aparecerá “**percentile ()**”. Dentro dos parênteses, escreva **(25, salario)**, e clique em **OK**. Você terá representado no gráfico o primeiro quartil;



d) Repita o mesmo procedimento, e na caixa da função percentile, escreva **(75, salario)**, e clique em **OK**. Será representado o terceiro quartil.

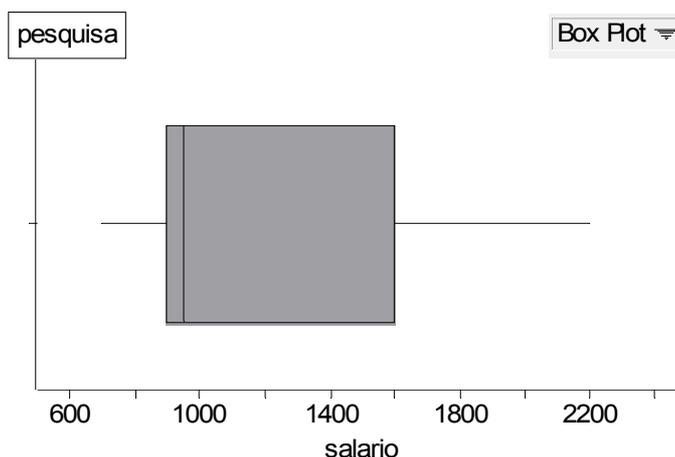


mean () = 1242.5
 median () = 950
 percentile (25, salario) = 900
 percentile (75, salario) = 1600

Uma representação gráfica envolvendo quartis, mediana, valor máximo e valor mínimo é o “**box-plot**”. No **box-plot**, define-se uma “caixa” com o limite esquerdo dado pelo primeiro quartil e o limite direito pelo terceiro quartil. A

mediana é representada por um traço no interior da caixa, e segmentos de reta são colocados da caixa até os valores máximo e mínimo.

8) Construa um novo gráfico para a variável salário, e no canto superior direito do gráfico, clique na seta e selecione **Box Plot**, para representar este gráfico.



9) Em relação aos últimos gráficos utilizados para representar os quartis, responda:

Qual tipo de gráfico você prefere? _____ Justifique _____

10) Compare os dois tipos de gráfico, dizendo as principais diferenças entre eles e quais informações são mais evidentes em um ou em outro.

11) Veremos agora como inserir uma nova coluna no banco de dados inicial. Vamos assim considerar a distância entre cada valor observado para a variável “idade” e o valor obtido para a média. Chamaremos cada uma dessas distâncias de “desvio”. Podemos chamar esta nova coluna de “idadedesvios”.

- Na tabela, clique em **<new>** e insira o título para a coluna;
- Para calcular os desvios, clique na coluna **“idadedesvios”**;
- No menu, clique em **“Edit”, “Edit Formula”**;



d) Na caixa, escreva “**idade–mean(idade)**” e clique em **OK** (será calculado para cada valor da variável, a diferença entre o valor observado e a média aritmética).

pesquisa							
	nome	genero	idade	cargo	escolari...	salario	idadede..
1	a	f	23	recepcao...	medio	850	-10.25
2	io	m	25	operario	fundame...	950	-8.25
3	io	m	30	operario	fundame...	1100	-3.25
4	uricio	m	35	gerente	superior	1700	1.75
5	los	m	29	engenheiro	superior	2000	-4.25
6	iana	f	31	gerente	superior	1800	-2.25
7	irdes	f	26	vendedor	medio	1400	-7.25
8	gio	m	27	operario	medio	950	-6.25
9	ilda	f	36	aux limpe...	fundame...	700	2.75

12) Observe o resultado na tabela. O que significa o resultado “positivo” nesse cálculo? E o “negativo”? _____

13) Vamos inserir os desvios na tabela de distribuição de freqüência da variável idade (que foi construída na primeira fase da atividade).

Clique na coluna **idadedesvios**, e a arraste até o canto esquerdo da tabela.

pesquisa		Summary Table														Row Summary			
		idade																	
		23	25	26	27	29	30	31	32	33	34	35	36	38	40	42	43		
	idadedesvio	1	1	1	1	2	1	1	1	1	2	2	1	2	1	1	1	20	
		-10.25	-8.25	-7.25	-6.25	-4.25	-3.25	-2.25	-1.25	-0.25	0.75	1.75	2.75	4.75	6.75	8.75	9.75	0	

S1 = count ()
S2 = mean ()

14) Construa também uma coluna para os desvios da variável salário, e depois insira na tabela de distribuição de freqüências do salário.

Baseando-se nas atividades desenvolvidas até o momento, responda:

- a) Quantos funcionários responderam à pesquisa? _____
- b) Do total de entrevistados, qual é a quantidade de funcionários do gênero feminino? _____ E do masculino? _____
- c) Em relação ao nível de escolaridade, quantas mulheres têm nível superior? _____ E quantos homens? _____
- d) Entre as idades observadas, qual a idade máxima e a idade mínima?

- e) Qual é o menor salário nesta empresa? _____ E o salário máximo? _____ E a média dos salários? _____
- f) Quais são os cargos que têm o salário acima da média nesta empresa? _____
- g) Quantos funcionários recebem salário abaixo da média nesta empresa? _____
- h) Quantos funcionários do gênero masculino recebem salário acima da média na empresa? _____ E do gênero feminino? _____
- i) Quantos funcionários do gênero feminino recebem mais que R\$ 1900,00? _____ E do gênero masculino? _____
- j) Quantos funcionários do gênero feminino recebem menos que R\$ 900,00? _____ E do gênero masculino? _____
- k) Qual valor você utilizaria para representar o salário desta empresa? _____ Justifique: _____

Parte B – Descrevendo os estudantes de sua escola (Adaptado de PONTE, 2005).

Suponha que você queira comunicar a um aluno de outro estado ou outro país, como são os estudantes de sua escola. Para isso, você irá preparar um relatório que mostre resumidamente as características principais destes estudantes.

1) Preparação das questões de investigação: discuta, com seu colega de dupla, sobre:

a) Quais dados (físicos, sociais, culturais...) devem entrar na caracterização do aluno típico?

b) Será necessário traçar um perfil para as meninas e outro para os meninos? Por quê?

2) Preparação da coleta de dados:

- Prepare folhas de registro para os dados que serão pesquisados;
- Escolha as variáveis da pesquisa;
- Peça aos seus colegas de escola que respondam as perguntas da pesquisa.

3) Organização e representação dos dados:

- Organize seus dados adequadamente (como você aprendeu nas atividades de familiarização);
- Para cada tipo de variável da sua pesquisa, construa os tipos de representações (gráficos, tabelas) que você considerar mais apropriado para representar tal variável. Quando considerar conveniente, utilize também as medidas-resumo;
- Escolha uma das variáveis para fazer as representações somente utilizando lápis e papel e, para as demais, utilize o software Fathom;
- Acompanhando estas representações estatísticas, redija pequenos textos que servirão como relatórios, descrevendo

Telas do Fathom – Parte A das atividades

Telas do Fathom – Grupo 1

Fathom - [carol, michelle]

File Edit Display Insert Data Analyze Window Help

	nome	genero	idade	cargo	escolaridade	salario
1	Ana	f	23	recepcao...	medio	850
2	Fabio	m	25	operario	fundamental	950
3	mario	m	30	operario	fundamental	950
4	mauricio	m	35	gerente	superior	2000
5	carlos	m	29	engenheiro	superior	2200
6	mariana	f	31	gerente	superior	2000
7	lourdes	f	26	vendedor	medio	1300
8	sergio	m	27	operario	medio	950
9	marida	f	36	aux limpe...	fundamental	700
10	rose	f	32	aux limpe...	fundamental	700
11	ricardo	m	30	vendedor	medio	1700
12	rafael	m	36	engenheiro	superior	2200
13	raquel	f	40	encarreg...	superior	1300
14	cassio	m	38	operario	medio	950
15	eduardo	m	33	operario	fundamental	950
16	felipe	m	35	encarreg...	medio	1200
17	michel	m	34	vendedor	medio	1500
18	joao	m	40	operario	fundamental	950
19	miguel	m	42	porteiro	fundamental	800
20	maria	f	43	aux limpe...	fundamental	700

Summary Table

		cargo								Row Summary
		aux limpeza	encarregado	engenheiro	gerente	operario	porteiro	receptionista	vendedor	
genero	f	3	1	0	1	0	0	1	1	7
genero	m	0	1	2	1	6	1	0	2	13
Column Summary	3	2	2	2	6	1	1	3	20	

S1 = count ()

Summary Table

		escolaridade			Row Summary
		fundamental	medio	superior	
genero	f	3	2	2	7
genero	m	5	5	3	13
Column Summary	8	7	5	20	

S1 = count ()

Summary Table

		genero		Row Summary
		f	m	
idade	33	25		
idade	33.5			
Column Summary	7	13	20	

S1 = mean ()
S2 = median ()

Summary Table

		genero		Row Summary
		f	m	
salario	1242.5			
Column Summary	7	13	20	

S1 = mean ()

Summary Table

		escolaridade			Row Summary
		fundamental	medio	superior	
idade	8	7	5	20	
Column Summary	8	7	5	20	

S1 = count ()

Summary Table

		cargo								Row Summary
		aux limpeza	encarregado	engenheiro	gerente	operario	porteiro	receptionista	vendedor	
idade	3	2	2	2	6	1	1	3	20	
Column Summary	3	2	2	2	6	1	1	3	20	

S1 = count ()

Summary Table

		escolaridade			Row Summary
		fundamental	medio	superior	
idade	8	7	5	20	
Column Summary	8	7	5	20	

S1 = count ()

Dynamic Statistics (tm)

Windows Vista Starter

Fathom - [carol, michelle]

File Edit Display Insert Data Analyze Window Help

Summary Table

		idade												Row Summary				
		23	25	26	27	29	30	31	32	33	34	35	36	38	40	42	43	
idadedesvios	1	1	1	1	1	2	1	1	1	1	2	2	1	2	1	1	1	20
idadedesvios	-10.25	-8.25	-7.25	-6.25	-4.25	-3.25	-2.25	-1.25	-0.25	0.75	1.75	2.75	4.75	6.75	8.75	9.75	0	
Column Summary	1	1	1	1	2	1	1	1	1	2	2	1	2	1	1	1	20	

Summary Table

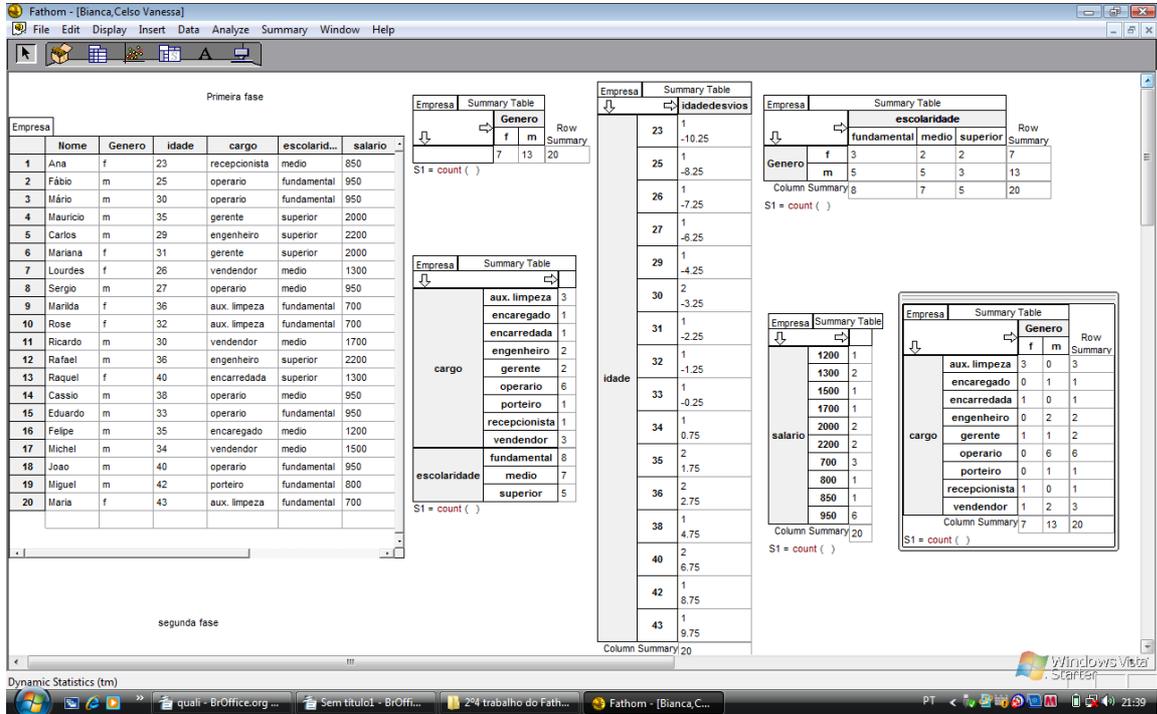
		salario												Row Summary
		1200	1300	1500	1700	2000	2200	700	800	850	950			
salariodesvios	1	2	1	1	2	2	3	1	1	1	6			
salariodesvios	-42.5	57.5	257.5	457.5	757.5	957.5	-542.5	-442.5	-392.5	-292.5				
Column Summary	1	2	1	1	2	2	3	1	1	1	6			

S1 = count ()

Dynamic Statistics (tm)

Windows Vista Starter

Telas do Fathom – Grupo 2



Telas do Fathom – Grupo 3

The screenshot shows the Fathom software interface with several data tables and summary statistics. The main table is a 'Cargo' table with columns for different job roles and a 'Row Summary' column. Other tables show summary statistics for 'Genero', 'Escolaridade', and 'Idade'. A large table on the right lists individual data points for 14 people, including their names, genders, ages, job titles, education levels, salaries, and IDs.

Nome	Genero	Idade	Cargo	Escolari...	salario	idi...
1 Ana	F	23	Recepção	medio	850	-10
2 Fabio	M	25	Operario	fundame...	950	-8
3 Mario	M	30	Operario	fundame...	950	-3
4 Mauricio	M	35	Gerente	Superior	2000	1.7
5 Carlos	M	29	Engenheiro	Superior	2200	-4
6 Mariana	F	31	Gerente	Superior	2000	-2
7 Lourdes	f	26	vendedor	medio	1300	-7
8 Sergio	m	27	operario	medio	950	-6
9 Marilda	f	36	aux. limp...	fundame...	700	2.7
10 Rose	f	32	aux. limp...	fundame...	700	-1
11 Ricardo	m	30	Vendedor	medio	1700	-3
12 Rafael	m	36	engenheiro	superior	2200	2.7
13 Raquel	f	40	encarreg...	superior	1300	6.7
14 Cassio	m	38	operario	medio	950	4.7
15 Eduardo	m	33	operario	fundame...	950	0

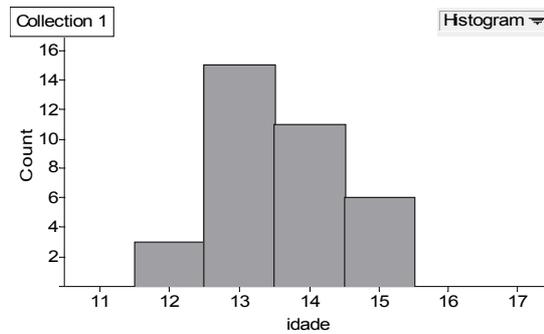
The screenshot shows the Fathom software interface with several statistical plots for the 'salario' variable. The plots include a histogram, a bar chart, a box plot, and a dot plot. The histogram shows the distribution of salaries, with a peak around 1000. The bar chart shows the count for each salary value. The box plot shows the median, quartiles, and range of the salary data. The dot plot shows individual data points for the salary variable.

Summary statistics for 'salario':

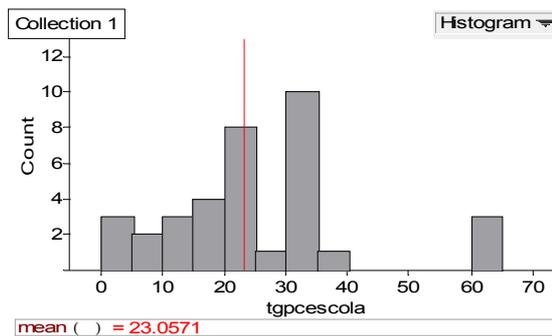
- mean () = 1242.5
- median () = 950
- percentile (25, salario) = 950
- percentile (75, salario) = 1600

Relatórios dos Grupos (Parte B das atividades)

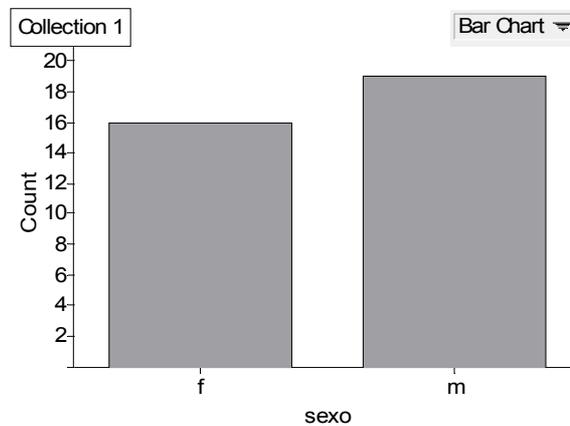
Relatório Grupo 1



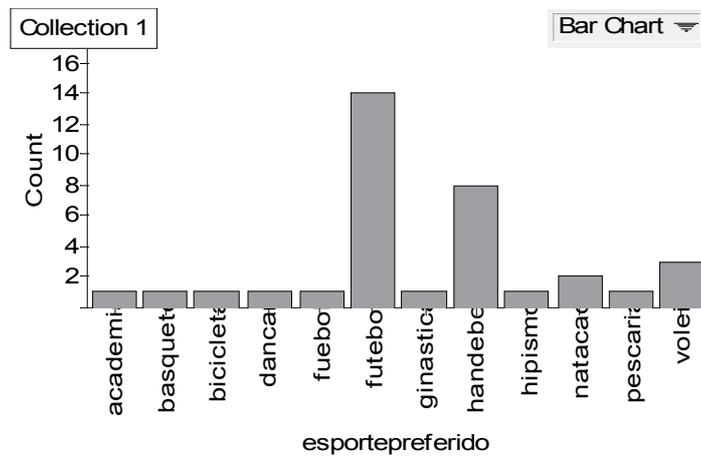
Os alunos entrevistados estão na faixa etária de 12 a 15 anos, sendo que a maioria possui 13 anos de idade e a minoria 12 anos.



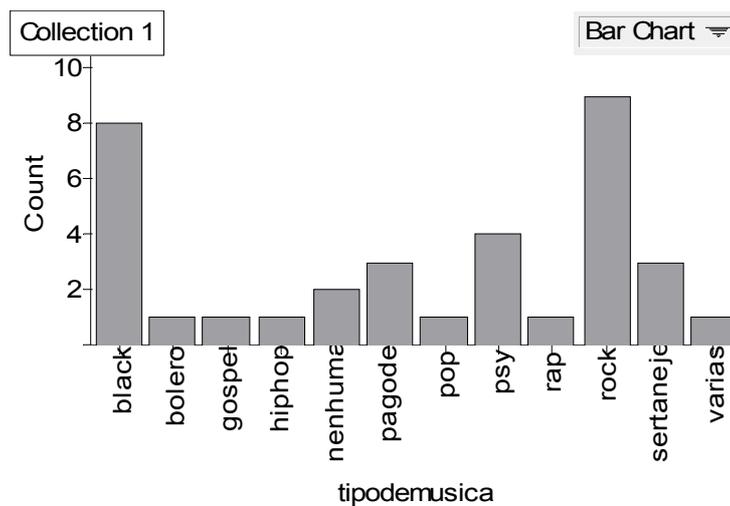
Os alunos gastam em media 23min para chegarem a escola.



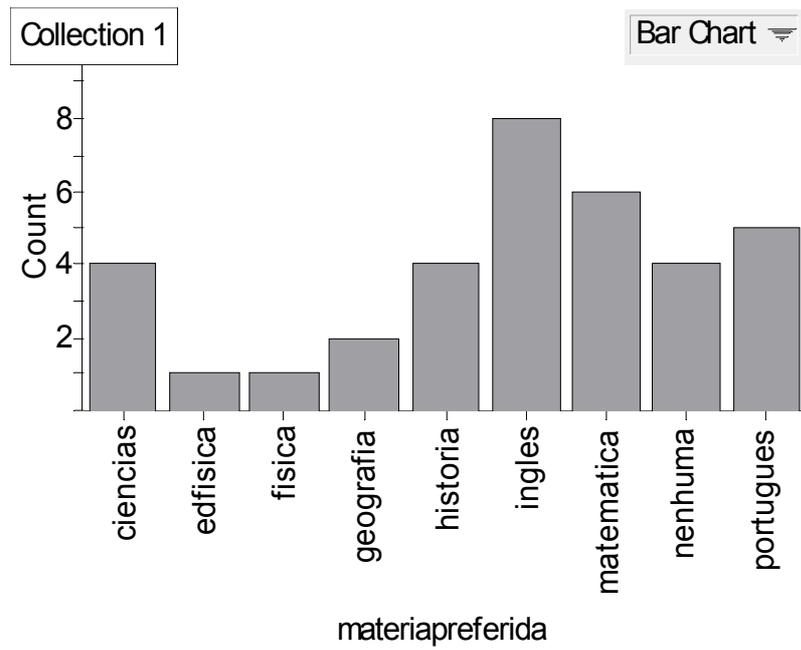
A maioria dos entrevistados são do sexo masculino.



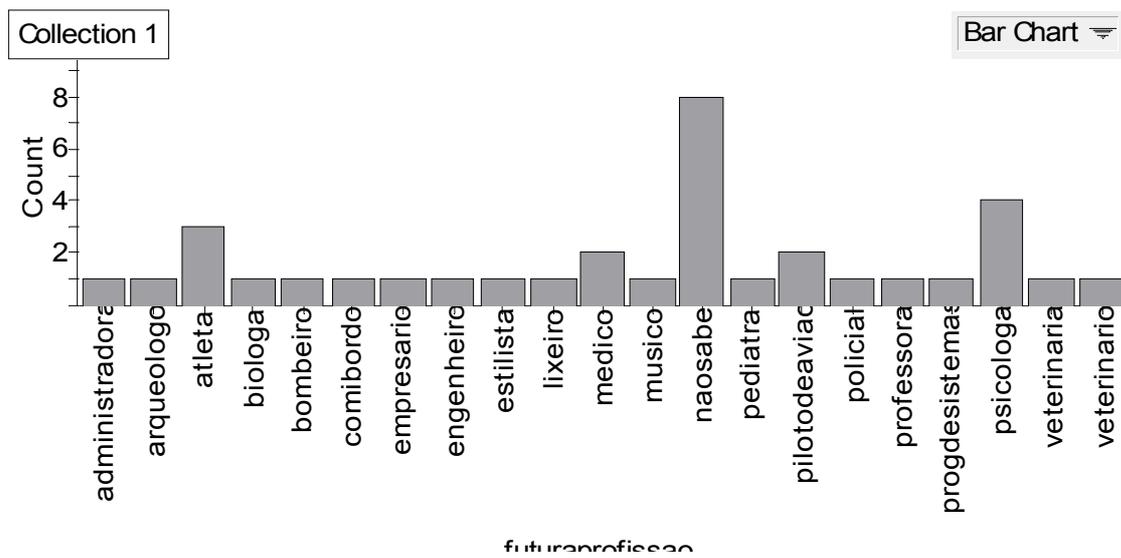
O esporte mais praticado e o futebol.



Os estilos de musicas mais ouvidos são rock e black.



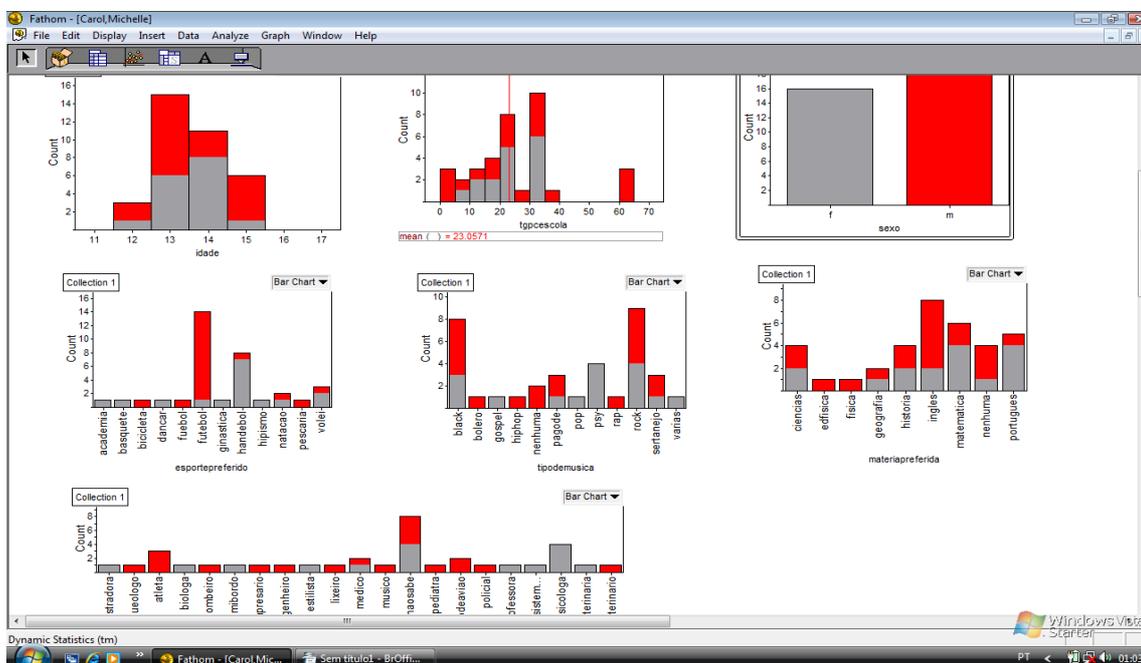
As matérias preferidas são inglês e matemática.



A maioria não sabe qual será sua profissão futura.

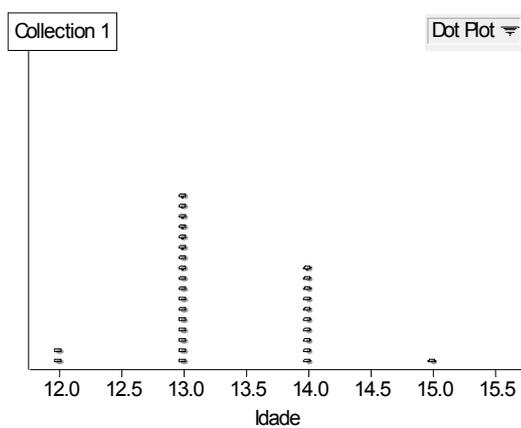


A maioria das meninas tem 14 anos, praticam handebol, gostam de psy, preferem a matéria de português e querem seguir a profissão de psicóloga.

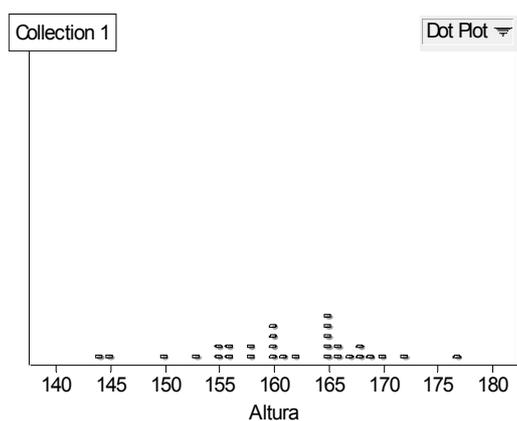


Os meninos praticam mais futebol, gostam de inglês, a maioria possui 13 anos, querem seguir a profissão de atleta.

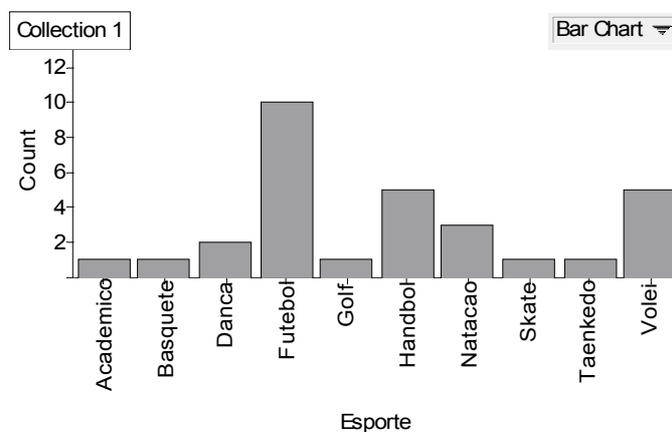
Relatório Grupo 2



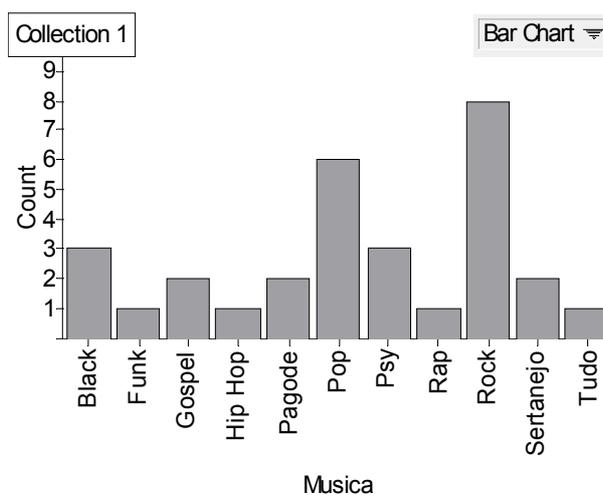
A maioria dos alunos pesquisados possuem 13 anos de idade.



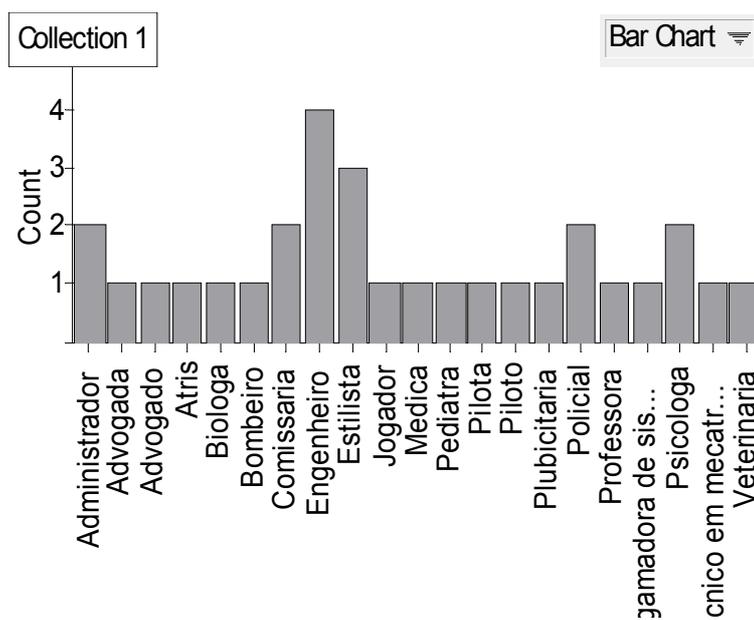
A maioria dos alunos entrevistados possuem entre 1.65 e 1.70 m.



Grande parte dos entrevistados têm como esporte favorito o Futebol.

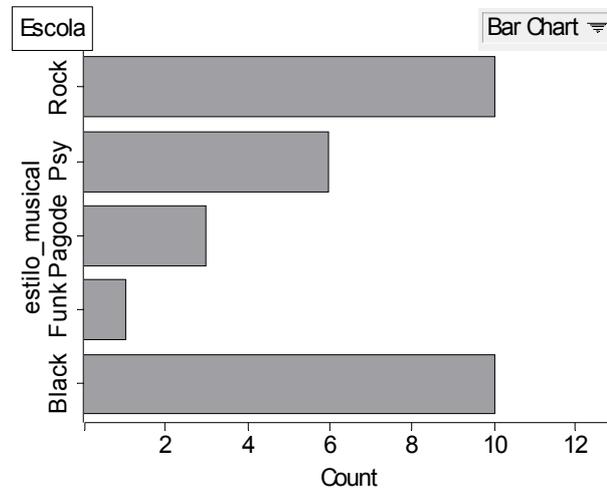


Grande parte dos entrevistados têm como gosto musical o Rock.



A maioria alunos entrevistados querem seguir a carreira como engenheiros.

Relatório Grupo 3



A maioria dos estudantes da oitava série da escola Oscavo tem preferência para o Rock e Black.

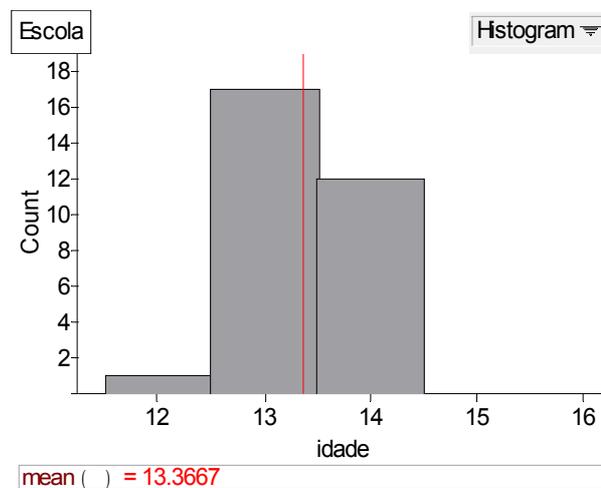


Gráfico de idade com a media de idade com cerca de 13.3

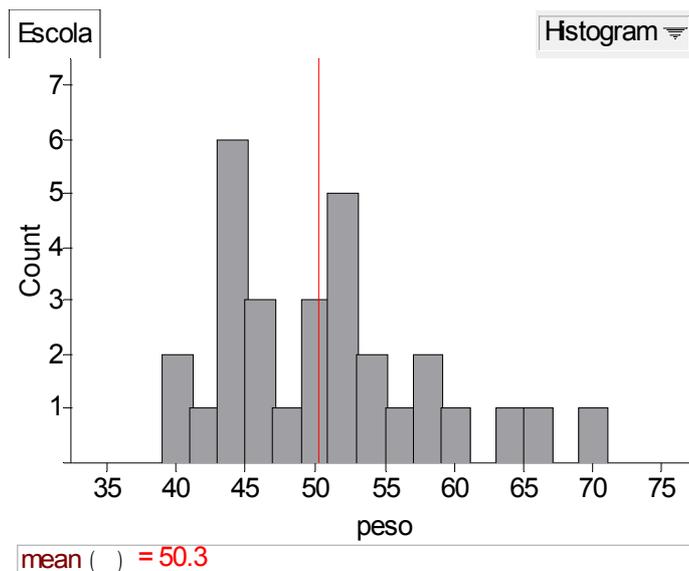
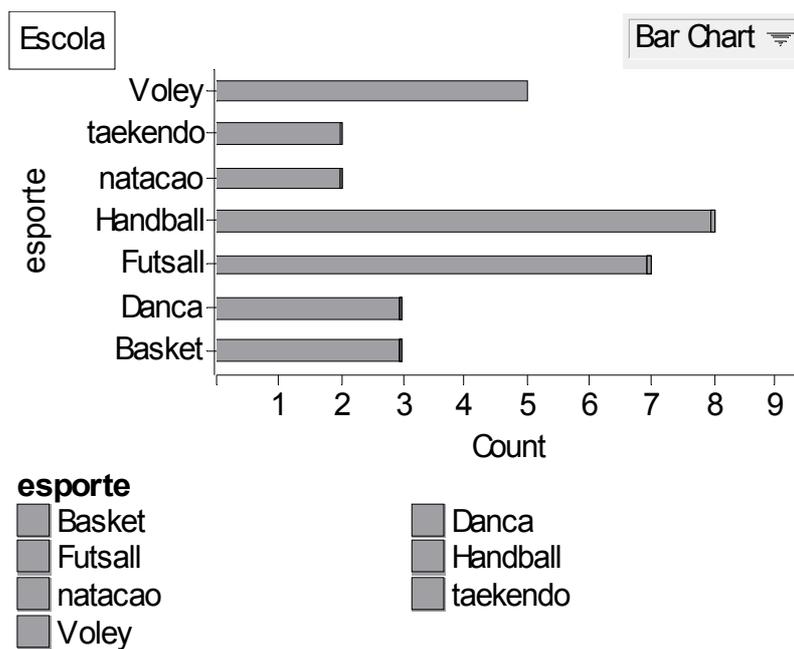
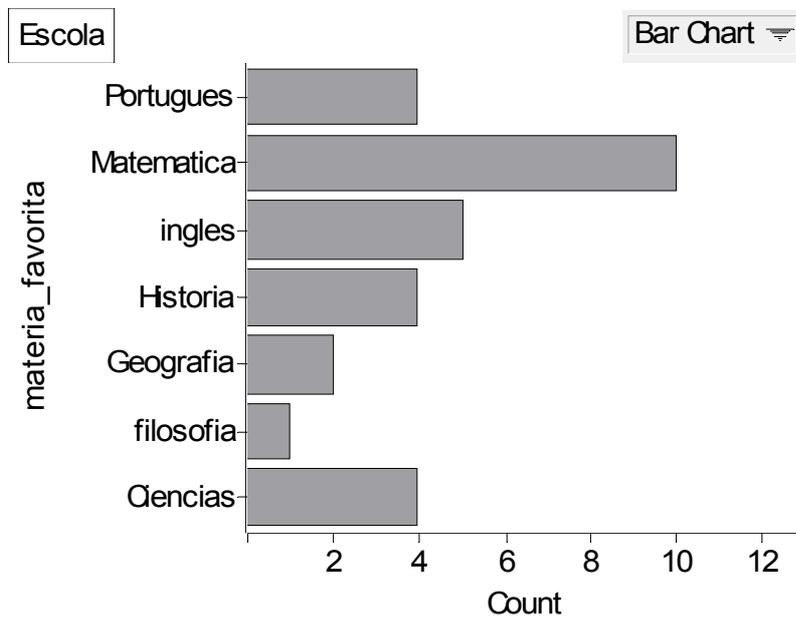


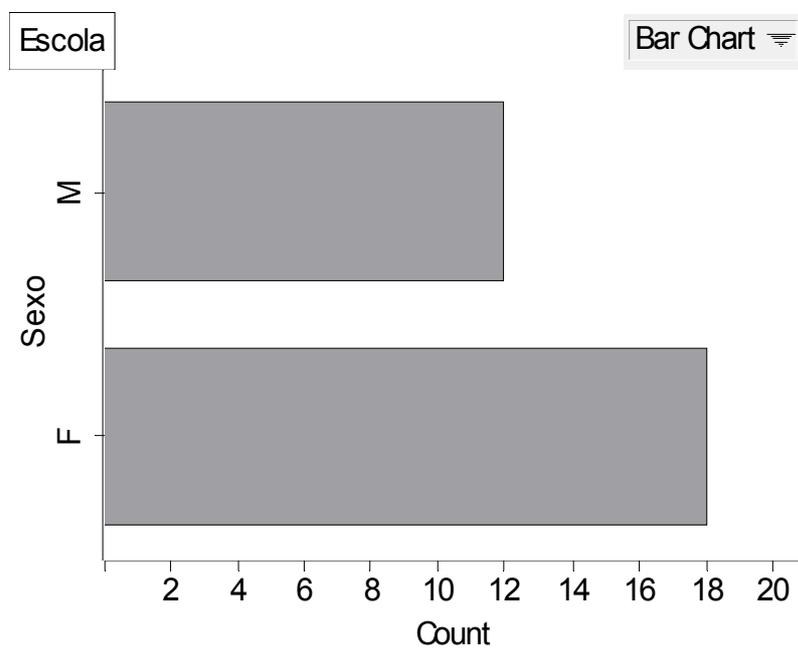
Gráfico que mostra o peso dos alunos e a media do peso que deu 50,3



A quantidade de alunos que gostam do esporte, a maioria é de Handball.



A Matéria preferida dos alunos é Matemática



A maioria dos alunos são do sexo feminino.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)