

RICARDO CARDOSO

**O Professor de Matemática e a análise exploratória
de dados no Ensino Médio**

**MESTRADO PROFISSIONAL EM ENSINO DE
MATEMÁTICA**

**PUC/SP
2007**

RICARDO CARDOSO

**O Professor de Matemática e a análise exploratória no
Ensino Médio**

Dissertação apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência parcial para obtenção do título de **MESTRE PROFISSIONAL EM ENSINO DE MATEMÁTICA**, sob a orientação da **Professora Doutora Cileda de Queiroz e Silva Coutinho**.

**PUC/SP
SÃO PAULO
2007**

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:** _____

DEDICATÓRIA

À minha mãe Izail (em memória)

Ao meu pai, Antonio,

A minha amada esposa, pelo compreensão e apoio

AGRADECIMENTO

A todos que de alguma maneira contribuíram para a realização deste trabalho.

Em especial:

- *A DEUS pela onipresença;*
- *À Professora Doutora **Cileda**, orientadora e amiga pelo incentivo, paciência e dedicação;*
- *Aos Amigos **Pedro Bigattão, Ricardo Sérgio Vasques, Sérgio Alves, , Carlos Biffi, Jonas Borsetti** pelo apoio, incentivo e companheirismo.*
- *Aos demais colegas pelas discussões e contribuições ao trabalho;*
- *Aos professores de Matemática Ensino Médio que se dispuseram, com boa vontade, a participar desta pesquisa;*
- *Ao professor Doutor. **Saddo Ag. Almouloud** pelo apoio, amizade e orientações ao longo de todo o percurso.*
- *A Secretaria da Educação por oferecer a oportunidade da execução desta pesquisa.*

RESUMO

A Estatística tem se destacado ultimamente por sua utilidade em praticamente todas as áreas do conhecimento humano. Pesquisas e dissertações existentes sobre o assunto sugerem a necessidade de aprofundar os conhecimentos sobre as dificuldades no processo de ensino da disciplina. Nossa questão principal é investigar se os Professores da Rede Pública desenvolvem o ensino da Estatística Descritiva no ensino médio, e se estão aptos a utilizar, de forma eficaz, as noções estatísticas de base – Organização de dados, Medidas de Tendência Central, Separatrizes e Dispersão, para resolver problemas práticos de seu cotidiano. O intuito dessa pesquisa é verificar o nível de mobilização dos conhecimentos por parte dos Professores do Ensino Médio. Segundo LINS (2004, p.54) “sabemos que persiste a impressão geral – não documentada de forma sistemática por pesquisas – de que a formação matemática do licenciado, em boa parte similar a do futuro bacharel, não contribui de modo substancial para a formação daquele futuro profissional, a não ser ao reforçar as rotinas de aulas expositivas”. Tentaremos diagnosticar qual o nível de conhecimento em que o Professor do Ensino Médio se encontra no conteúdo curricular de Estatística. Baseando-se no questionário de BIFI (2006, p. 54), verificar se o professor de matemática do ensino médio é capaz de calcular, justificar e relacionar as medidas descritas. As respostas das atividades serão analisadas com o auxílio do software C.H.I.C.

Palavras-chave: Estatística, Organização dos Dados, Análise Exploratória dos Dados.

ABSTRACT

Statistics has stand out lately due to its utility in almost every area of human knowledge. Existing research and essay on the subject suggest the need for deepening knowledge on the difficulties with teaching process of this discipline. Our main issue is to investigate whether Brazilian public school teachers develop the Descriptive Statistics in high school and whether they are able to use, efficiently, the basic statistics concepts – Data Organization, Measures of Central Tendency, Separatrixes and Dispersion to solve daily practical problems. The aim of this research is to verify the level of knowledge mobilization by high-school teachers. According to LINS (2004, p.54) “we know that the general impression – not registered systematically by researches – that the licenciante’s mathematical formation, in great part similar to the future graduate, does not contribute substantially to the formation of this future professional, unless to reinforce the expositive classes routines”. We will try to diagnose in which level of knowledge the high-school teacher is in school curriculum of Statistics. Based on BIFI’s questionnaire (2006, p.54), we will verify whether the high-school Mathematics teacher is able to calculate, justify and relate the described measures. The answers of the activities will be analyzed through C.H.I.C. software.

Key-words: Statistics, Data Organization, Exploratory analysis of data.

SUMÁRIO

INTRODUÇÃO	14
1 PROBLEMÁTICA	17
1.1 INTRODUÇÃO	17
1.2 O PROBLEMA DE PESQUISA	17
1.3 METODOLOGIA E PROCEDIMENTOS.....	18
1.4 QUADRO TEÓRICO	21
2 ESTUDOS PRELIMINARES	30
2.1 PESQUISAS QUE ABORDAM O TEMA.....	30
2.2 NOSSO OBJETO: OS CONCEITOS ESTATÍSTICOS DE BASE	33
2.3 CONCEITOS ESTATÍSTICOS ELEMENTARES	36
3 PARTE EXPERIMENTAL	43
3.1 ATIVIDADE PROPOSTA	44
3.2 ANÁLISE A PRIORI.....	47
3.3 ANÁLISE A POSTERIORI	70
3.3.1 <i>Análise Hierárquica de Similaridade</i>	71
3.3.2 <i>Análise Apresentada Pelos Professores</i>	80
CONSIDERAÇÕES FINAIS	88
REFERÊNCIA BIBLIOGRAFICA	92
ANEXO I	94
APÊNDICE A	98
APÊNDICE B	99
APÊNDICE C	100
APÊNDICE D	103
APÊNDICE E	105

LISTA DE GRÁFICOS

GRÁFICO 1: Nº DE ALUNOS X TEMPO DE ESTUDO	38
GRÁFICO 2: NÚMERO DE CARROS/PESSOA.....	46
GRÁFICO 3: HORAS NO TRÂNSITO/PESSOA.....	46
GRÁFICO 4: HISTOGRAMA DA VARIÁVEL IDADE	53
GRÁFICO 5: ESTUDO DA MÉDIA POR MEIO DO HISTOGRAMA	54
GRÁFICO 6: ESTUDO DA MEDIANA POR MEIO DO HISTOGRAMA	55
GRÁFICO 7: HISTOGRAMA DA DISTRIBUIÇÃO DE RENDA COM OS LIMITES DE REFERÊNCIA E UM DESVIO PADRÃO	57
GRÁFICO 8: ESTUDO DAS MEDIDAS CENTRAIS POR MEIO DE GRÁFICOS.....	62
GRÁFICO 9: ESTUDO DAS MEDIDAS SEPARATRIZES POR MEIO DE GRÁFICOS.	63
GRÁFICO 10: ANÁLISE DAS MEDIDAS CENTRAIS DA TABELA 4 POR MEIO DE GRÁFICOS.....	66
GRÁFICO 11: ANÁLISE DAS MEDIDAS SEPARATRIZES DA TABELA 4 POR MEIO DE GRÁFICOS.	66
GRÁFICO 12: RETOMADA DO GRÁFICO 1.	67
GRÁFICO 13: RETOMADA DO GRÁFICO 2.	68

LISTA DE FIGURAS

FIGURA 1. ESQUEMA 1 DOS NÍVEIS DE LETRAMENTO ESTATÍSTICO	28
FIGURA 2: BOX – PLOT 1	39
FIGURA 3: DOT-PLOT	49
FIGURA 4: REPRESENTAÇÃO DOS QUARTIS E MEDIANA	52
FIGURA 5: VARIABILIDADE EM TORNO DA MEDIANA.....	57
FIGURA 6: REPRESENTAÇÃO DA MEDIANA A PARTIR DA SÉRIE DO ROL.....	60
FIGURA 7: REPRESENTAÇÃO DOS QUARTIS A PARTIR DA SÉRIE DO ROL.....	60
FIGURA 8: REPRESENTAÇÃO DA QUANTIDADE DE CARRO/PESSOA.....	61
FIGURA 9: VARIABILIDADE EM TORNO DA MEDIANA.....	65

LISTA DE TABELAS

TABELA 1: IDADE E RENDA MENSAL DE 40 PESSOAS (DADOS FICTÍCIOS).....	44
TABELA 2: QUANTIDADE DE CARROS/PESSOA (DADOS FICTÍCIOS)	45
TABELA 3: N. DE HORAS NO TRÂNSITO/PESSOA (DADOS FICTÍCIO).....	45
TABELA 4: DADOS ORDENADOS DA VARIÁVEL IDADE	47
TABELA 5: CÁLCULOS PARA ENCONTRAR A MÉDIA E VARIÂNCIA DE DADOS AGRUPADOS.....	51
TABELA 6: RENDA FAMILIAR (DADOS FICTÍCIOS)	55
TABELA 7: DISTRIBUIÇÃO DE FREQUÊNCIA - QTDE. CARROS/PESSOA	59
TABELA 8: DISTRIBUIÇÃO DE FREQUÊNCIA COM INTERVALO DE CLASSE DO NÚMERO DE HORAS NO TRÂNSITO POR PESSOA.....	63

LISTA DE QUADROS

QUADRO 1: CATEGORIZAÇÃO DOS PROCEDIMENTOS.....	22
QUADRO 2 : NÍVEIS DE LETRAMENTO E PENSAMENTO ESTATÍSTICO (ASA)	29

INTRODUÇÃO

A Estatística é uma ciência cujo campo de aplicação estende-se a muitas áreas do conhecimento humano, e tem por objetivo organizar e analisar dados, possibilitando e fundamentando a tomada de decisão. Dados estatísticos sempre fizeram parte da vida das pessoas, assim como muitas confiam, acreditam e utilizam, outras refutam sua veracidade.

Segundo Lopes (2004, p.198)

Não basta ao cidadão entender as porcentagens expostas em índices estatísticos como o crescimento populacional, taxas de inflação, desemprego, etc. O cidadão precisa muitas vezes realizar análise minuciosa dos dados, o que requer uma habilidade de relacionar ou criticar os dados apresentados, e questionando e ponderando até mesmo sua veracidade.

O que queremos apontar é que em qualquer que seja a situação, para uma análise crítica de dados, as informações devem ser suficientes, para que o leitor possa julgar sua fidedignidade, e mesmo levantar suas próprias conjecturas. A capacidade das pessoas interpretar uma grande quantidade de dados, assume, hoje em dia, um papel fundamental para análise e entendimento das informações em que os meios de comunicação veiculam diariamente.

De acordo com Gal (2002, p.2-3), o autor acredita que:

a) capacidade de interpretar e avaliar criticamente a informação estatística os argumentos apoiados em dados e os fenômenos estocásticos que pessoas podem encontrar em diversos contextos, incluindo os meios de comunicação, mas não limitando-se a eles. b) capacidade para discutir ou comunicar suas opiniões a respeito das informações estatísticas quando são relevantes.

De acordo com os Parâmetros Curriculares Nacionais do Ensino Médio PCN+ (1999), o Ensino de Estatística, está inserido no Tema 3 Análise de dados, portanto aponta para necessidade de uma alfabetização estatística do cidadão, e sugerem:

A análise de dados tem sido essencial em problemas sociais e econômicos, como nas estatísticas relacionadas a saúde, populações, transportes, orçamentos e questões de mercado. Propõe-se que constitua o terceiro eixo ou tema estruturador do ensino, e tem como objetos de estudo os conjuntos finitos de dados, que podem ser numéricos ou informações qualitativas, o que dá origem a

procedimentos bem distintos daqueles dos demais temas, pela maneira como são feitas as quantificações, usando-se processos de contagem combinatórios, freqüências e medidas estatísticas e probabilidades. Este tema pode ser organizado em três unidades temáticas: Estatística, Contagem e Probabilidade.

A Estatística e a Probabilidade devem ser vistas, então, como um conjunto de idéias e procedimentos que permitem aplicar a Matemática em questões do mundo real, mais especialmente aquelas provenientes de outras áreas. Devem ser vistas também como formas de a Matemática quantificar e interpretar conjuntos de dados ou informações que não podem ser quantificados direta ou exatamente. Cabe à Estatística, por exemplo, analisar a intenção de voto em uma eleição ou o possível êxito do lançamento de um produto no mercado, antes da eleição em si e da fabricação do produto. Isso é feito através da pesquisa estatística, que envolve amostras, levantamento de dados e análise das informações obtidas.(PCN+, 1999, p.126)

Os PCN+ (1999) sugerem que a estatística pode contribuir muito para o exercício da cidadania, mas para isso é necessário desenvolver uma prática pedagógica em que se desenvolvam a criatividade, iniciativa pessoal e autonomia, pois dessa forma é que formaremos cidadãos críticos.

Em vista disso, o objetivo desta pesquisa é verificar de que forma os professores de Matemática do Ensino Médio, mobilizam os conhecimentos estatísticos quando confrontados com problemas que envolvem Medidas de Tendência Central, Variabilidade e representação Gráfica de um conjunto de dados. Esse trabalho é parte de um trabalho maior, composto por esta pesquisa e as pesquisas de Pereira (2007) e Vasques (2007).

PEREIRA, Sérgio Alves: Um estudo a respeito do professor de matemática e a implementação de uma seqüência didática para a abordagem da estatística no ensino médio.

VASQUES, Ricardo Sergio Braga: Mobilização dos Conceitos Estatísticos - um estudo diagnóstico desses conceitos, envolvendo variabilidade, com os alunos do ensino médio.

Para tanto, consideramos as seguintes questões: o professor, ao trabalhar um conjunto de dados, utiliza os princípios da análise exploratória? Que elementos dessa análise ele mobiliza na resolução de problemas?

Buscando responder a estas questões, nossa pesquisa será apresentada da seguinte forma:

No primeiro capítulo, fazemos uma breve apresentação do problema de pesquisa, apresentando um quadro teórico para nos fundamentarmos nos níveis de alfabetização, em que o professor de Matemática do Ensino Médio se encontra.

No segundo capítulo, algumas pesquisas que abordam o tema de nossa pesquisa, e os objetos matemáticos utilizados.

No terceiro capítulo, foi proposta uma atividade diagnóstica constituída de três partes, sendo a primeira um banco de dados, a segunda tabelas com ou sem intervalos de classe e a terceira gráficos.

Os dados foram analisados com o auxílio do software C.H.I.C. (Classificação Hierárquica Implicativa e Coesitiva), cujo objetivo é a análise de associação entre as variáveis, relacionando sujeitos a variáveis fornecendo um índice de qualidade dessa associação.

1 PROBLEMÁTICA

1.1 Introdução

Acreditamos que este estudo traga contribuições à formação de professores, quando se discute elementos relacionados à aprendizagem de Estatística no Ensino Médio.

A Estatística vem apresentando um rápido desenvolvimento como ciência, tendo em vista sua produção científica que se amplia com a participação de investigadores de áreas distintas, os quais não se restringem a educadores matemáticos e estatísticos.

1.2 O Problema de Pesquisa

Sabemos que a estatística, tem se destacado cada vez mais em todas as áreas do conhecimento. As pesquisas apresentadas por Educadores em Estatística, como Batanero, Estepa, Godino (1991), Wild e Pfannkuch (1998), Morais (2006), Silva (1999), Lopes (2002) entre outros, apontam a necessidade de investir na alfabetização estatística para crianças, adolescentes e até mesmo adultos, visando sanar as dificuldades, por parte destes, ao se depararem com dados ao serem analisados. Neste sentido, o papel dos professores é fundamental nesse processo.

O objetivo desta pesquisa é verificar de que forma os professores de Matemática do Ensino Médio, mobilizam os conhecimentos estatísticos quando confrontados com problemas que envolvem Medidas de Tendência Central, Variabilidade e representação Gráfica de um conjunto de dados.

Focalizamos nosso estudo no problema central:

De que forma os professores de Matemática do Ensino Médio, de escolas públicas de São Paulo em exercício, mobilizam os conhecimentos estatísticos quando confrontados com problemas que envolvam Medidas de Tendência Central, Variabilidade e representação Gráfica, e qual o nível de alfabetização estatística destes professores?

Segundo Shamos (1995, apud Wild, 1999) a alfabetização estatística tem uma estrutura composta por três níveis:

- O primeiro, considerado básico, é a alfabetização cultural, referindo-se às pessoas que compreendem termos básicos usados comumente nos meios de comunicação diante de assuntos relacionados à ciência.

- O segundo, chamado de alfabetização funcional, relativo à capacidade do sujeito conversar, ler, e escrever informações utilizando termos científicos coerentes.

- O terceiro, alfabetização científica relativo aos conhecimentos científicos de esquemas conceituais primordiais ou de teorias que fundamentem a ciência aliada à compreensão dos processos científicos e investigativos mobilizados na resolução de situações-problema. Nessa fase, o indivíduo age com autonomia e segurança nas escolhas de métodos e representações estatísticas, como também na capacidade de analisar os dados considerando-se a variabilidade existente.

1.3 Metodologia e Procedimentos

Para respondermos nossa questão de pesquisa, aplicamos um questionário para 29 professores de matemática do ensino médio que estão atualmente em exercício, em escolas públicas do Estado de São Paulo (SP), particularmente na região sul da grande São Paulo. A escolha desses professores não obedeceu a nenhum critério de amostragem aleatória, uma vez que buscamos a participação voluntária dos docentes, com os quais já tínhamos estabelecido vínculos profissionais. Dessa forma pensamos minimizar o

constrangimento e a sensação de estar sendo avaliado ao solicitarmos que respondessem à atividade proposta.

Vale ressaltar que os professores, eram convidados a resolver um problema, parte da nossa atividade diagnóstica, dividida em três etapas que deveriam ser resolvidas, individualmente, podendo ou não consultar seu livro didático, caso fosse necessário relembrar alguma fórmula. Com um tempo estimado de 150 minutos (3horas/aula), ao término, todas as atividades foram recolhidas pelo pesquisador. Esta atividade diagnóstica constituída com base nas atividades propostas por Bifi (2006), nos permitirá diagnosticar os níveis de alfabetização: científico, funcional e cultural de acordo com Shamos (1995, apud Wild, 1999)

Para analisar essa atividade diagnóstica, contamos com o auxílio do *software* Classificação Hierárquica, Implicativa e Coesitiva –C.H.I.C.– que possibilita análise hierárquica de similaridade e/ou coesitiva dos dados coletados, evidenciando as inter-relações encontradas entre as respostas dadas por cada professor .

Acreditamos que as situações-problemas, quando apresentadas em várias formas de representação, abrem um leque maior de possibilidades de investigar as possíveis dificuldades que os professores do ensino médio poderão encontrar durante a resolução. Dessa forma, as etapas da atividade se apresentarão em diferentes níveis de complexidade, para que seja possível verificar o nível de alfabetização Estatística em que o professor se encontra.

O instrumento acima citado (anexo 1) é composto por três partes distintas: a primeira será composta por um banco de dados fictício, no qual constarão a idade e renda mensal de quarenta pessoas supostamente entrevistadas por uma empresa de cartões de crédito. Na segunda parte, são apresentados duas tabelas de distribuição de freqüências, sendo a primeira sem intervalo de classes e a segunda com intervalo de classes. Na primeira distribuição é representado o comportamento da variável “quantidade de carros por pessoa”, na segunda o comportamento do tempo no trânsito por um número de pessoas.

A terceira parte representa a distribuição da segunda parte de forma gráfica.

O intuito das três formas de representação é identificar os níveis de dificuldade relacionados às diferentes formas de apresentação de uma situação-problema. Em outras palavras, fazemos a hipótese de que se os dados brutos forem apresentados ao professor sem nenhuma organização (crescente ou decrescente) as dificuldades serão maiores, do que estes mesmos dados apresentados em forma de gráficos e tabelas.

As categorias que nos ajudarão a identificar quais as possíveis dificuldades que os professores apresentarão na resolução da atividade, foram construídas a partir dos critérios identificados por Bifi (2006). Porém, para este trabalho com professores, pensamos em uma adaptação em função de seus conhecimentos e do que ele julga importante do seu trabalho com seus alunos. (saber a ensinar, nos termos da transposição didática de Chevallard, (1991).

- em relação a média

a) os professores analisarão a média, interpretando corretamente seus resultados;

b) os professores analisarão a média, porém interpretando seus resultados como um ponto de equilíbrio da distribuição;

c) os professores analisarão o desvio padrão, interpretando corretamente seus resultados como medida de dispersão, fazendo associação com média;

d) os professores relacionarão média e desvio padrão, interpretando seus valores como indicadores de variabilidade em relação à média;

e) os professores relacionarão a média e o desvio padrão, mas sem interpretar a variabilidade;

f) outras categorias que podem emergir na análise dos resultados.

- em relação à mediana

a) os professores analisarão a mediana, interpretando corretamente seus resultados como valor central de uma distribuição, de forma que existam tantos valores superiores ou iguais como inferiores ou iguais a ela;

b) os professores analisarão os quartis, interpretando corretamente seus resultados como uma medida separatriz;

c) os professores analisarão os quartis, que é uma medida que permite dividir a distribuição em quatro partes iguais, como indicadores de variabilidade em relação a mediana;

d) os professores relacionarão mediana e quartis, interpretando as medidas de variabilidade;

Esperamos que, por meio desta atividade, possamos atingir nossos objetivos e assim colaborarmos para o ensino e a aprendizagem da estatística, identificando pontos importantes que devem ser objetos de uma formação continuada, de uma reflexão conjunta pelos professores do Ensino Médio.

Acreditamos que a partir de um estudo diagnóstico como o que estamos propondo, possam ser planejadas parcerias que permitam ao professor rever, melhorar, buscar formação sobre o ensino e a aprendizagem da Estatística no Ensino Médio.

1.4 Quadro Teórico

Shamos (1995, apud Moraes, 2006), utiliza uma estrutura composta por três níveis para categorizar a alfabetização estatística dos sujeitos. O primeiro, considerado básico, é a alfabetização *cultural*, referindo-se às pessoas que compreendem termos básicos usados comumente nos meios de comunicação diante de assuntos relacionados à ciência. O segundo, chamado de alfabetização *funcional*, relativo à capacidade do sujeito de conversar, ler, e escrever informações utilizando termos científicos coerentes. O terceiro e último, é a alfabetização *científica* relativo aos conhecimentos científicos de esquemas conceituais primordiais ou de teorias que fundamentem a ciência aliada à compreensão dos processos científicos e investigativos mobilizados na resolução

de situações-problema. Nessa fase, o indivíduo age com autonomia e segurança nas escolhas de métodos e representações estatísticas, como também na capacidade de analisar dados considerando-se a variabilidade existente.

Conforme Gal (2002), o conjunto de procedimentos estatísticos que um indivíduo desenvolve na resolução de um problema pode ser categorizado em cinco bases de conhecimento inter-relacionadas, que são: Alfabetização, Estatística, Matemática, Contexto Global e Contexto Crítico. Em nossa análise, categorizaremos os procedimentos de resolução dos professores, em dois níveis de conhecimento.

QUADRO 1: CATEGORIZAÇÃO DOS PROCEDIMENTOS

Primeiro nível (Operacional)	Segundo nível (Analítico)
<i>Alfabetização</i>	<i>Análise global</i>
<i>Estatística</i>	<i>Análise crítica</i>
<i>Matemática</i>	

Identificamos como nível operacional, as três primeiras bases de conhecimento citadas por Gal (2002), ou seja: Alfabetização, Estatística e Matemática.

A alfabetização abre espaço para rever as bases de conhecimentos citadas por Gal (2002), necessária no aprendizado em Estatística. Fazendo com que os professores busquem por exibições de informações tabulares ou gráficas, uma vez que as estatísticas são apresentadas por textos escritos ou orais, a utilização de habilidades específicas da estatística torna-se um ponto fundamental para nossa análise. O uso de textos estatísticos apresenta várias demandas com respeito à habilidade de alfabetização dos professores. Por exemplo, os significados de certos termos estatísticos usados na mídia podem ser diferentes de suas acepções coloquiais ou cotidiana. As mensagens veiculadas podem usar termos técnicos de um modo apropriadamente profissional, mas também podem conter jargão estatístico, que é ambíguo ou errôneo. Podemos perceber que

alguns jornais e outros canais de mídia tendem a empregar convenções em resultados estatísticos, como recorrer a “erro de amostragem” (ou “margem de erro”), ao discutir resultados de votações, mas sem explicar o significado dos termos usados.

Segundo Triola (1998 p.10) *Um erro amostral é a diferença entre um resultado amostral e o verdadeiro resultado populacional; tais erros resultam de flutuações amostrais.* “E margem de erro, é a diferença máxima provável entre a média amostral observada e a verdadeira média populacional.

No que diz respeito à segunda categoria, Estatística, ela é condição prévia e óbvia para compreender e interpretar mensagens estatísticas, juntamente com os conceitos e procedimentos matemáticos relacionados ao assunto em estudo. O contato com a Estatística tem como objetivo identificar habilidades e competências necessárias para uma análise, crítica e correta, dos dados.

Na terceira categoria, Matemática, os professores precisam estar atentos sobre alguns procedimentos matemáticos que estão claramente por trás dos indicadores estatísticos, como, por exemplo, porcentagem e média.

Percebe-se que, com a ajuda da tecnologia informatizada, os cálculos matemáticos estão deixando de ser ferramenta de uso cognitivamente custoso, uma vez que é feita pelos *softwares* disponíveis no mercado, cabendo ao professor a análise crítica de seus resultados em um contexto estatístico.

É igualmente importante a preocupação com a compreensão de resultados estatísticos referentes a porcentagens ou médias, o que requer não apenas familiaridade intuitiva, mas até certo ponto formal sobre os procedimentos matemáticos envolvidos. Por outro lado, é desejável que os professores do ensino médio conheçam os aspectos conceituais da média, e não somente o aspecto algorítmico, para compreenderem melhor que a média pode ser influenciada por valores extremos em um conjunto de dados e, conseqüentemente, pode não ser um bom representante para um conjunto de valores (notas de aula).

No segundo nível, os analíticos, abordaremos as duas últimas bases do conhecimento citado por Gal (2002), que seriam os contextos globais e críticos. Nessa etapa, essas categorias estão apropriadas à interpretação de mensagens estatísticas pelos professores do ensino médio, exigindo habilidades em responder dentro de um contexto e, também, o de buscar seu conhecimento no que diz respeito a dados estatísticos. Essa habilidade de interpretar valores estatísticos também apóia o processo geral da alfabetização estatística e é um fator crítico para possibilitar o senso de compreensão de qualquer dado estatístico. (Nota de aula) “Os dados estatísticos deveriam ser vistos como números com um sentido, e, conseqüentemente, esta é a fonte de significado e a base para interpretação dos resultados obtidos.”

As habilidades dos professores em compreender as diferentes formas de apresentações estatísticas dependem da informação que eles podem obter da mensagem sobre o que está por trás do estudo que está sendo discutido, se o professor não está familiarizado com o contexto no qual foram coletados os dados, fica mais difícil compreender e analisar a variabilidade dos dados, interpretar e validar resultados e os modelos utilizados na análise.

Sendo assim, os professores devem preocupar-se com a validade das mensagens transmitidas, a natureza e credibilidade da evidência que está por trás das informações ou conclusões apresentadas, e refletirem sobre possíveis alternativas de interpretação para as conclusões mostradas por estes meios de comunicação.

Acreditamos que o professor do ensino médio, já tenha cursado a disciplina Estatística, esteja, pelo menos, segundo Shamos (1995, apud Moraes, 2006), no nível funcional em relação a estes dois níveis que acabamos de apresentar.

Acreditamos na necessidade de desenvolver habilidades estatísticas que tornem um indivíduo capaz de agir consciente e criticamente diante das questões da atualidade.

Neste sentido, discutiremos algumas idéias que julgamos importantes para o desenvolvimento destas habilidades. Abordamos inicialmente a definição de pensamento estatístico, adotada por alguns pesquisadores e que vem ao encontro do que pensamos ser base para tal trabalho. Para Snee¹(1990, p.118, apud Wild e Pfannkuch, 1999) o pensamento estatístico é definido como:

[...] processos mentais que reconhecem a variação como algo que nos rodeia e que está sempre presente em tudo o que fazemos. Todo o trabalho é uma série de processos interconectados de forma que identificando, caracterizando, quantificando, controlando e reduzindo a variação, proporcionamos oportunidades de crescimento.

Nessa definição, este autor enfatiza a importância da noção de variabilidade no pensamento estatístico. Segundo ele, esses processos são explícitos e vão além da resolução de um problema particular, envolvendo raciocínios que partem de um dado procedimento podendo até chegar a uma ou mais conclusões. O pensamento estatístico possibilita visualizar o processo como um todo com interação, justificação e compreensão do relacionamento dos dados, além da mensuração da variação presente. Assim, é possível explorar os dados para além do que está prescrito no texto, de modo que gere novos questionamentos e tomadas de decisões conscientes e críticas.

Na busca de definir o pensamento estatístico, a American Statistical Association – ASA – propôs elementos constituintes deste, apoiados nas idéias de (Moore 1990, apud Wild, 1999) considerando:

1. a onipresença da variação do processo;
2. a necessidade dos dados relativos ao processo;
3. a produção dos dados baseados na variação existente;

¹ SNEE (1990, p.118, apud Pfannkuch e Wild, 1999) defined statistical thinking as “thought processes, which recognize that variation is all around us and present in everything we do, all work reducing variation provide opportunities for improvement”.

4. a quantificação da variação;

5. a explanação da variação.

Tendo em vista a importância desses elementos na formação do pensamento estatístico a ASA, a Mathematical Association of America – MAA – Joint Committee on Undergraduate Statistics considerou como elementos essenciais para análise de dados:

1. a necessidade dos dados;

2. a importância da produção dos dados;

3. a onipresença da variabilidade;

4. a medição e modelagem da variação existente.

Os pesquisadores Wild e Pfannkuch (1999) investigam os processos complexos do pensamento envolvidos na resolução de problemas da realidade, compreendendo a estatística como possibilidade real para obter uma melhoria nas soluções desses problemas. Como vemos nas palavras destes autores:

[...] nosso interesse é desenvolver uma estrutura para os modelos do pensamento envolvidos na resolução de problemas, nas estratégias para esta resolução, bem como a integração de elementos estatísticos presentes na mesma (Wild E Pfannkuch, 1999, p.3, apud Morais, 2006, p.24).

Acreditamos que a formação do pensamento estatístico enfatiza a necessidade e a produção dos dados, como também o estudo da variação por eles apresentada. No entanto, a capacidade de leitura e interpretação de dados, organizados em tabelas e gráficos, desenvolve habilidades condizentes com um nível de alfabetização estatístico. Nossa compreensão sobre alfabetização estatístico baseia-se nas concepções de Gal (2002), Wallman, (1993, *apud* Gal, 2002), e Shamos (1995, apud Wild, 1999). Adotamos o termo alfabetização como tradução do termo inglês “literacy”, de acordo com a definição de alfabetização funcional apresentada nos relatórios do INAF. No relatório de 2002, encontramos:

A definição sobre o que é analfabetismo funcional vem, ao longo das últimas décadas, sofrendo revisões significativas, como reflexo das próprias mudanças sociais. Em 1958, a UNESCO definia como

alfabetizada uma pessoa que capaz de ler e escrever um enunciado simples, relacionado a sua vida diária. Vinte anos depois, a UNESCO sugeriu a adoção dos conceitos de analfabetismo e alfabetismo funcional. É considerada alfabetizada funcional a pessoa capaz de utilizar a leitura/escrita para fazer frente as demandas de seu contexto social e usar essas habilidades para continuar aprendendo e se desenvolvendo ao longo da vida. (INAF, 2002, p.5)

Gal (2002) acredita que a alfabetização estatística subentende um conhecimento mínimo de conceitos e de procedimentos estatísticos básicos. Wallan² (1993, *apud* Gal, 2002), assegura que ela consiste na habilidade de compreensão e avaliação crítica dos resultados estatísticos vivenciados em nosso cotidiano, aliados à habilidade de apreciar suas contribuições nas tomadas de decisões públicas, confidenciais e pessoais.

Shamos (1995, *apud* Wild, 1999), por sua vez, utiliza uma estrutura composta por três níveis. O primeiro, considerado básico, é o *cultural*, referindo-se às pessoas que compreendem termos básicos usados comumente nos meios de comunicação diante de assuntos relacionados à ciência. O segundo, chamado de *funcional*, relativo à capacidade do sujeito de conversar, ler, e escrever informações utilizando termos científicos coerentes. O terceiro e último, o *científico* relativo aos conhecimentos científicos de esquemas conceituais primordiais ou de teorias que fundamentem a ciência aliada à compreensão dos processos científicos e investigativos mobilizados na resolução de situações-problema. Nessa fase, o indivíduo age com autonomia e segurança nas escolhas de métodos e representações estatísticas, como também na capacidade de analisar dados considerando-se a variabilidade existente. O figura 1, apresentado em Morais (2006), e que adotaremos nesse trabalho, ilustra os três níveis citados.

² WALLAN, K.K (1993, Enhancing Statistical literacy: Enriching our society. Journal of the American Statistical Association, 88 p.1-8apud Gal, 2002)

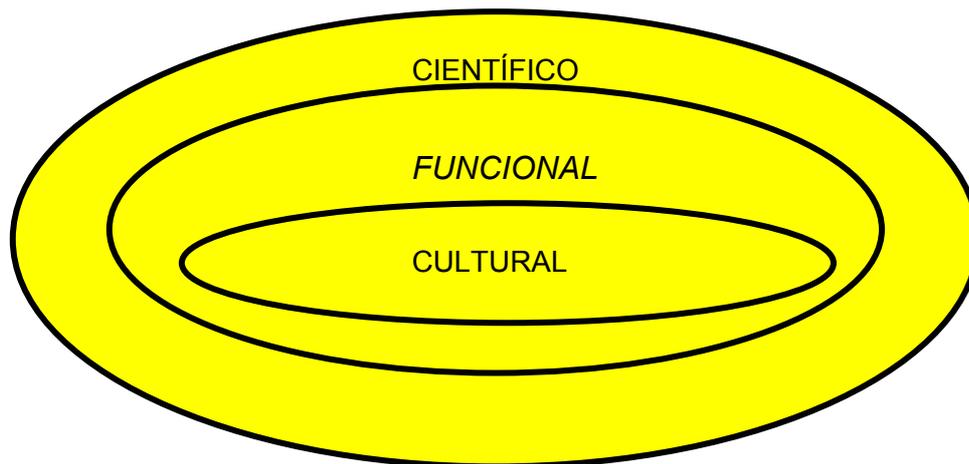


FIGURA 1. ESQUEMA 1 DOS NÍVEIS DE LETRAMENTO ESTATÍSTICO

Fonte: Morais 2006. p.25

Essas concepções, refletem nosso entendimento por alfabetização estatística, quando somos capazes de ler e reconhecer informações que estejam representadas em tabelas e/ou gráficos, estamos no nível *cultural*. Porém, se além de ler, interpretamos informações contidas em dados que estão representados em tabelas e/ou gráficos, ou mesmo, organizá-los nessas representações, identificando e considerando a variação na análise dos mesmos, estamos no nível *funcional*. E, se além das habilidades mencionadas anteriormente, ainda somos capazes de fazer inferências e previsões sobre as informações contidas nos diversos registros, analisando e considerando a variabilidade existente, estamos no nível *científico*. Observamos assim que, de acordo com o esquema apresentado acima, partimos do mais geral para o mais específico, sendo que a cada nível compreende os anteriores: desenvolvemos primeiramente o nível cultural, apenas quando este está estável no sujeito, passamos ao nível seguinte, o funcional, para finalmente atingir o mais completo, nível científico.

Com o quadro 2 relacionamos os níveis de alfabetização estatística de (Shamos 1995, apud Wild, 1999) com a definição de pensamento Estatístico segundo a American Statistical Association – ASA

Níveis de Letramento Estatístico	Pensamento Estatístico (ASA)
Cultural	1. a onipresença da variação do processo
Funcional	2. a produção dos dados baseados na variação existente
Científico	3. a quantificação da variação 4. a explanação da variação

QUADRO 2 : NÍVEIS DE LETRAMENTO E PENSAMENTO ESTATÍSTICO (ASA)

De acordo com o quadro acima, os dados levantados durante a pesquisa serão analisados sob essa perspectiva.

2 ESTUDOS PRELIMINARES

Neste capítulo abordaremos quais conceitos de Estatística o professor de Matemática do Ensino Médio deveria trabalhar para a alfabetização estatística dos seus alunos.

Segundo os PCN+ (1999), *“Os conteúdos e habilidades propostos para as unidades temáticas a serem desenvolvidas nesse tema seriam:*

1. 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 estatísticos 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.(p.127)

A capacidade de leitura e interpretação de dados, organizados em tabelas e gráficos, desenvolve habilidades condizentes com um nível de alfabetização estatístico, onde que a formação do pensamento estatístico enfatiza a necessidade e a produção dos dados, como também o estudo da variação por eles apresentada.

2.1 Pesquisas que Abordam o Tema

Algumas pesquisas voltadas para a(s) dificuldade(s) de aprendizagem da Estatística, contribuíram e continuam contribuindo para a consolidação da importância de tal estudo. Entre elas, podemos citar, particularmente, os trabalhos de Silva (2000), Morais (2006) e Batanero (2001), nos quais fundamentamos nossa pesquisa.

Silva (2000) pesquisou cerca de 643 estudantes das áreas de Ciências Humanas e Exatas de uma universidade particular e indica que os alunos apresentavam atitudes negativas em relação à Estatística e desenvolviam ansiedade em relação a esta disciplina. A autora mostrou ainda que existe correlação positiva e significativa entre as atitudes dos alunos em relação à Matemática e à nota final da disciplina Estatística, isto é, as atitudes negativas com relação à Matemática são transferidas para a Estatística. A autora sugeriu que os alunos precisavam desenvolver atitudes positivas em relação à Estatística, como condição para obter melhores resultados.

Podemos interpretar essa afirmação, levantando a hipótese da necessidade de uma abordagem de situações didáticas adequadas na disciplina Estatística, voltadas para a construção, por parte dos alunos, de seus conceitos básicos. A autora, baseada em Moore (1997), relatou que o ensino de Estatística deveria apresentar problemas com dados reais, concentrando-se em aspectos que não necessitavam de memorização, mas sim de interpretação, estratégias para uma exploração efetiva de dados, com um diagnóstico básico preliminar para a inferência. De acordo com esses resultados, buscamos construir nosso instrumento diagnóstico (a ser apresentado mais adiante neste texto) contemplando tal tipo de atividade.

A pesquisa de Moraes (2006) buscou responder à seguinte questão: Que relações podem ser estabelecidas entre as concepções docentes sobre a sua prática no ensino da estatística e as formas de apresentação desses conteúdos nos livros didáticos?

As concepções a que se refere à autora, estão relacionadas ao desencontro entre as orientações dadas nos livros didáticos e os PCN (1998) para o ensino fundamental. Partindo da hipótese de que os professores tratam as representações tabulares e gráficas dos dados simplesmente como registro, não destacando os conceitos matemáticos e estatísticos mobilizados em cada um deles.

De acordo com a autora os resultados evidenciam que no grupo pesquisado, os professores, com aproximadamente 15 anos de carreira, têm uma visão reducionista da estatística, identificando apenas o conceito estatístico de

média aritmética e o conhecimento matemático de números e decimais nas tarefas propostas. A autora Morais (2006) desenvolve seu trabalho aplicando um instrumento diagnóstico que contém duas situações-problemas para serem discutidas pelos professores entrevistados. A primeira situação-problema contemplava dados relativos à variável quantitativa discreta e, a segunda, contínua. Ambas têm questões abertas e distintas, que exploram a resolução do professor em nível compatível com a resolução de seu aluno. Em cada situação-problema, o docente identifica os conceitos estatísticos e matemáticos presentes na resolução proposta, bem como indica os conceitos já conhecidos pelos alunos e, até mesmo, aqueles cuja abordagem acreditam ser necessária ao nível de escolaridade dos alunos.

Assim, de acordo com estes resultados, buscamos organizar as atividades que compunham nosso instrumento diagnóstico de forma a poder identificar o nível de alfabetização estatística dos professores entrevistados.

Batanero³ (2001) também chamou a atenção para a formação do professor, tanto de Matemática como de Estatística, afirmando que:

A estatística está sendo facilitada a um número crescente de pessoas, provocando uma grande demanda na formação básica desta matéria, que tem sido solicitada nos níveis universitários, aos professores de matemática.

Os novos currículos de educação primária e secundária incluem de forma generalizada recomendações sobre o ensino da estatística. BATANERO(2001, p.6)

Acreditamos que o ensino da Estatística, nos possibilita fazer conjecturas, formular hipóteses, estabelecer relações, que são processos necessários para a resolução de problemas. Para isso, talvez seja necessária uma prática pedagógica que promova a investigação e a análise exploratórias de dados, possibilitando aos professores de Matemática do Ensino Médio, em exercício,

³. [...]Todo ello ha facilitado el uso de la estadística a un numero creciente de personas, provocando una gran demanda de formación básica en esta materia, que ha sido encomendada, en los niveles no universitarios, a los profesores de matemáticas.

Los nuevos currículos de educación primaria y secundaria incluyen en forma generalizada recomendaciones sobre la enseñanza de la estadística.

tomarem consciência dos conceitos estatísticos permitindo-lhes fazer observações e tirar conclusões.

2.2 Nosso Objeto: Os conceitos estatísticos de base

A definição de Estatística é encontrada de formas diferentes, por muitos autores. Segundo Triola (1999, p.2) *“a Estatística é uma coleção de métodos para planejar experimentos, obter dados e organizá-los, resumi-los, analisá-los, interpretá-los e deles extrair conclusões”*.

Para este autor, a Estatística, além de ser uma técnica de coleta e apresentação de dados (análise exploratória e descrição, gráficos e tabelas), compreende também modelagem (probabilidade e processos estocásticos), análise indutiva (inferência: testes e estimação) e previsão e controle (verificação).

Para Vieira (2003, p.1) *“a Estatística não é, porém, simples coleção de números, embora as medidas ou observações na forma numérica sejam sua base. No sentido acadêmico, Estatística é a ciência que trata da coleta, apuração, classificação e análise dos dados”*

Em sua essência, a Estatística é a ciência que representa e interpreta dados, sejam eles numéricos ou não. Pode-se dizer que seu objetivo é o de apresentar informações sobre dados em análise para que se tenha maior compreensão dos fatos.

A Estatística trabalha com dados, dessa forma são necessários elementos básicos para a construção do raciocínio estatístico.

Variável estatística, é a característica que vai ser observada, medida ou contada nos elementos da população ou amostra e que pode variar, ou seja, assumir valor diferente de elemento para elemento e pode ser qualitativa ou quantitativa.

Entre as variáveis quantitativas, distinguem-se as variáveis discretas e as contínuas, as discretas são aquelas, que por sua natureza, assumem valores resultantes de uma contagem. Em uma linguagem matemática, pode-se dizer que uma variável discreta é aquela cujo conjunto de valores tem uma relação biunívoca com o conjunto dos números naturais, ou seja, entre dois valores

consecutivos não se pode inserir nenhum outro valor. Assim, por exemplo, número de pessoas de uma família, número de empregados de uma empresa, e mesmo as notas de uma avaliação são consideradas variáveis discretas, porém, ao medirmos magnitudes tais como peso, altura, temperatura, etc estamos trabalhando com variáveis contínuas, ou seja, entre dois valores consecutivos sempre podemos inserir um novo valor (e por conseqüência, não se pode estabelecer uma relação biunívoca com um conjunto enumerável)

Após a coleta e a construção de em um banco de dados, iniciaremos a Análise Exploratória de Dados. Para Godino, Batanero e Estepa, a análise exploratória dos dados é uma filosofia que consiste:

No estudo dos dados a partir de todas as perspectivas e com todas as ferramentas possíveis, incluindo as já existentes. O propósito é extrair toda a informação possível, gerar novas hipóteses no sentido de construir conjecturas sobre as observações que dispomos. (GODINO, BATANERO; ESTEPA; 1991, p.2)⁴

Esta etapa é fundamental, pois uma análise detalhada permite ao pesquisador familiarizar-se com os dados, organizá-los e sintetizá-los de forma a obter as informações necessárias do conjunto de dados para responder as questões que estão sendo investigadas. Tradicionalmente, a análise exploratória dos dados limitava-se a calcular algumas medidas de posição e variabilidade. No final da década de 70, John Tukey criou uma nova corrente de análise utilizando principalmente técnicas visuais, descrevendo quase sem utilizar cálculos, alguma forma de regularidade ou padrão nos dados, em oposição aos resumos numéricos. Nessa etapa, trabalham-se com tabelas, gráficos ou esquemas e medidas resumos que descrevem a tendência dos dados, quantifiquem e descrevam a sua variabilidade, permitam a detecção de estruturas interessantes (modelos) e valores atípicos no banco de dados.

C. Batanero, A. Estepa y J. D. Godino. Análisis Exploratorio de Datos: sus Posibilidades en la Enseñanza Secundaria. *Suma*, n. 9, 1991: 25-31. Disponível em: <<http://www.ugr.es/~batanero>>. Acesso em: 14 dez. 2006.

⁴Esta filosofía consiste en el estudio de los datos desde todas las perspectiva, y con todas las herramientas posibles, incluso las ya existentes. El propósito es extraer cuanta información sea posible, generar hipótesis nuevas, en el sentido de conjeturar sobre las observaciones de las que disponemos.

Segundo Wild e Pfannkuch (1998, p.3)⁵ "São considerados quatro elementos fundamentais do pensamento estatístico. Levar em conta a variação; a construção e o pensamento a partir de modelos; ter algum conhecimento de base, tanto no domínio estatístico como no domínio do contexto e sínteses e integração."(tradução nossa)

Segundo Machado (2000 p.103) "*As noções, os conceitos realmente relevantes sempre terminam por transbordar as fronteiras disciplinares. Certamente é possível tratá-los de modo disciplinar, mas sempre a custa de um empobrecimento de seu significado.*"

Para o desenvolvimento do raciocínio estatístico, adotaremos como conceitos elementares aqueles enunciados por Gal (2002) como os cinco blocos do conhecimento estatístico de base: o conhecimento da causa da necessidade dos dados e de como estes podem ser produzidos; a familiaridade com termos básicos e idéias relacionadas com a estatística descritiva, familiaridade com termos básicos e idéias relacionadas com representações gráficas e tabulares, a compreensão de noções básicas de probabilidade e, finalmente, o conhecimento do alcance das conclusões estatísticas ou inferências.

Desta forma, do fato de visarmos o nível de escolaridade relativo ao Ensino Médio, tomaremos o significado de número e os conhecimentos básicos da Estatística Descritiva como ponto de partida para a construção do raciocínio estatístico: organização e representação de um conjunto de dados, cálculo ou determinação de medidas-resumo e percepção da variabilidade.

⁵ Consideramos los siguientes cuatro elementos de "pensamiento estadístico". Tomar en cuenta la variación; la construcción a partir de modelos; tener algun conocimiento base, tanto en el dominio estadístico como en el dominio del contexto; y síntesis o integración.

2.3 Conceitos Estatísticos Elementares

Consideramos, no projeto maior, como conceitos estatísticos elementares, aqueles que estão na base do desenvolvimento do raciocínio estatístico e, por consequência, da alfabetização estatística.

Adotaremos como conceitos elementares aqueles enunciados por Gal (2002) os do conhecimento estatístico de base: o reconhecimento da necessidade dos dados e de como estes podem ser produzidos; a familiaridade com termos básicos e idéias relacionadas com a estatística descritiva, familiaridade com termos básicos e idéias relacionadas com representações gráficas e tabulares.

Organização e representação de um conjunto de dados

Da nossa opção por este trabalho, o significado dos conhecimentos básicos da Estatística Descritiva, optamos por apresentar estes conceitos de forma sempre contextualizada, em um primeiro contato do aluno com esses elementos, visando o início do desenvolvimento do raciocínio estatístico.

Apresentaremos assim as idéias básicas relacionadas a alguns destes conceitos.

A primeira delas diz respeito à diferença entre população e amostra.

Segundo Barbeta (2003), uma população é o conjunto de elementos que formam o universo do nosso estudo e que é possível de ser observado. Uma amostra é qualquer subconjunto finito da população.

Uma vez definida população e amostra, vamos definir o que será observado.

Variável estatística: é a característica da população que se quer observar. Esta característica pode ser qualitativa (nominal ou ordinal) ou quantitativa (discreta ou contínua). Uma variável qualitativa nominal descreve uma qualidade, sem, no entanto, estabelecer níveis de hierarquia. Citamos, entre outras música predileta, esporte preferido, cor dos olhos. Uma variável qualitativa ordinal descreve uma qualidade, mas identificando níveis hierárquicos. Citemos por exemplo nível de escolaridade, classe sócio-econômica ou qualquer tipo de

opinião expressa em tipo de escalas. As variáveis quantitativas são expressas por números. As quantitativas discretas são aquelas cujo conjunto admite uma relação biunívoca com o conjunto dos números naturais (ou seja, é um conjunto enumerável). Para Barbeta (2003), são as variáveis que “*só assumem valores que podem ser listados*”. As variáveis quantitativas contínuas são aquelas que, contrariamente às variáveis discretas, podem assumir qualquer valor em um intervalo real.

Uma vez coletados os dados (preferencialmente em pesquisa pelos alunos e com coleta a partir de uma problematização), passa-se a sua representação e organização, seja em forma de tabelas de distribuição de freqüências ou em forma de gráfico. Uma distribuição de freqüências pode ser uma função empírica, na qual cada valor da variável observada é associado a sua freqüência.

Cada representação nos dá uma visão sobre diferentes aspectos, e portanto, dessas representações, faz com que possamos desenvolver uma análise exploratória de dados satisfatória. Para Batanero (2001), uma idéia fundamental da análise exploratória de dados é o uso de vários tipos de representações múltiplas de dados e se converte em um meio para desenvolver novos conhecimentos e perspectivas. Por exemplo, passar de listagem de números a uma representação gráfica, com o histograma ou um Box-Plot que possibilite a comparação de várias amostras.

O histograma, construído no sistema de eixos ortogonais, é uma representação gráfica em forma de colunas justapostas. Para a sua construção deve ser observado o seguinte:

- a) A área total do histograma deve ser proporcional à freqüência total, e as áreas parciais (área de cada retângulo), proporcionais às freqüências das classes.
- b) As bases dos retângulos são proporcionais à amplitude do intervalo de classe. Assim, amplitudes iguais levam à bases de mesma medida, enquanto que amplitudes distintas devem ser representadas por medidas distintas.

c) Caso as amplitudes dos intervalos de classes sejam unitárias, a altura de cada retângulo terá como correspondente suas respectivas freqüências.

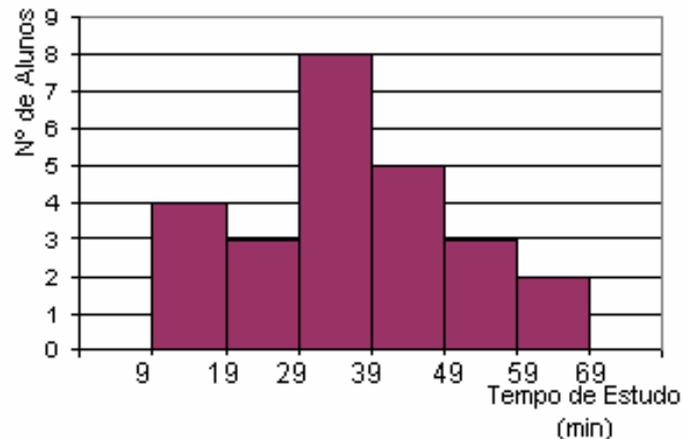


GRÁFICO1: Nº DE ALUNOS X TEMPO DE ESTUDO

Segundo Barbeta (2003), uma maneira de apresentar aspectos relevantes de uma distribuição de freqüência é através do chamado diagrama de caixas (Box-plot) ou desenho esquemático. Sua construção se inicia pela determinação do primeiro quartil, mediana e terceiro quartil e é preciso também observar o limite inferior e o superior.

No exemplo da distribuição representada pelo histograma do gráfico 1, teremos:

$$Q_1 = 26,5$$

$$Q_2 = Md = 35,87$$

$$Q_3 = 46,5$$

$$Min = 9$$

$$Máx = 69$$

Passando esses dados para um eixo orientado, temos o Box-plot apresentado na figura 2.

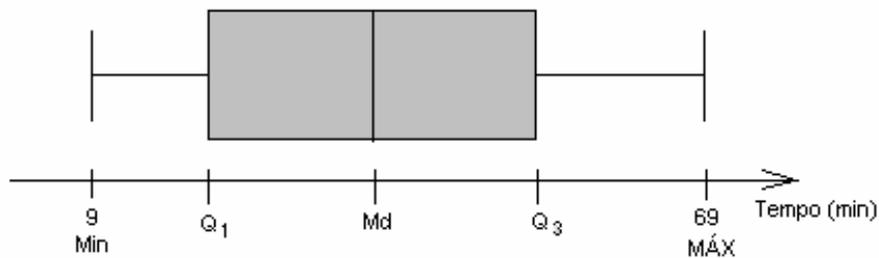


FIGURA 2: BOX – PLOT 1

Este tipo de gráfico permite visualizar o conjunto de dados organizado em quatro grupos com 25% dos elementos em cada um destes grupos.

Determinação e interpretação das medidas-resumo de um conjunto de dados

Uma distribuição de frequência (ou mesmo um conjunto de dados apresentado simplesmente em uma série de ROL) pode ser interpretada com auxílio das medidas de tendência central em associação com medidas de dispersão (o que permite a apreensão da variabilidade). Um exemplo usual no ensino é a amplitude total, que é definida como sendo a diferença entre o maior e o menor dos valores assumidos por uma variável quantitativa. Por exemplo, observando o tempo mínimo e o tempo máximo de estudo dos alunos, teríamos ($69 - 9 = 60$), ou seja a amplitude total é de 60 minutos.

Uma outra medida muito utilizada, a média aritmética, é uma medida que representa um conjunto de dados indicando o ponto ao redor do qual os demais valores se concentram. Sem dúvida é a mais popular entre todas as medidas. Seu cálculo remete à divisão em partes iguais do todo entre seus componentes. Assim, seu algoritmo consiste em somar todos os valores da variável e dividir pelo número de dados.

$$\begin{aligned} \text{Média em Matemática} &= \frac{\text{total da soma das notas}}{\text{N}^\circ \text{ de notas consideradas}} = \\ &= \frac{6,5 + 4,5 + 5,0 + 7,0 + 8,5 + 9,0 + 3,5 + 4,0 + 6,5 + 7,5}{10} = \frac{62}{10} = 6,2 \end{aligned}$$

Ou seja, a média das notas dos dez alunos desse grupo é de 6,2 pontos, o que significa que as notas do grupo variam em torno do valor 6,2, sem

entretanto indicar o grau de dispersão dos mesmos também em relação ao valor 6,2.

No entanto, perde-se muito do seu significado se não pudermos identificar o grau dessa concentração de valores ao redor da média.

O desvio-padrão é a medida que melhor indica essa concentração. Seu valor é determinado pelo cálculo da raiz quadrada da variância, que por sua vez, é determinada pelo cálculo da média dos quadrados da diferença entre cada valor observado e a média. Ou seja

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

ou em sua forma mais operacional:

$$\sigma^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{N}}{N}$$

Tratamos aqui da variância populacional. Para o caso da variância amostral:

$$s^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}$$

ou

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{N}}{N-1}$$

Já a moda é o valor que se repete com maior frequência. Para a variável idade, o valor que mais aparece é 14 anos, logo a moda é 14 anos. Sua interpretação é bastante intuitiva e de baixo custo cognitivo. A moda é uma medida que pode ser determinada para qualquer tipo de variável, seja qualitativa

ou quantitativa e pode ser trabalhada com os alunos mesmo nas séries iniciais de escolaridade.

Uma outra medida-resumo é a mediana, que divide em duas partes iguais (50%) um conjunto de dados ordenado. É o termo central na distribuição. Para determinar a mediana de um conjunto de dados não organizado em uma tabela de distribuição de freqüências, tal como no exemplo (idade), procederemos da seguinte forma:

1º passo: Ordenar os dados

13, 13, 13, 14, 14, 14, 14, 14, 15, 15

2º passo: Buscar o valor tal que 50% dos dados observados são maiores ou iguais a ele, e simultaneamente, 50% dos dados observados são menores ou iguais a ele. No nosso exemplo, como temos um número par de observações (10 observações), este valor será a média aritmética entre os termos que ocupam a 5ª e a 6ª posições.

$$\text{Md} = \frac{14 + 14}{2} = 14$$

Interpretando a mediana: como a mediana divide o conjunto em duas partes iguais (50%), conclui-se de forma bastante intuitiva que 50% dos alunos têm menos de 14 anos e 50% dos alunos têm mais de 14 anos.

Os quartis dividem um conjunto de dados ordenados em quatro partes iguais. Assim, devem-se ter três quartis: Q_1 , Q_2 e Q_3 , em que Q_2 coincide exatamente com a mediana. De um ponto de vista cognitivo, pode-se propor ao aluno um estudo sobre a localização dos quartis a partir de uma analogia com a determinação da Mediana, sem recursos a regras e fórmulas encontradas nos livros didáticos. De acordo com o exemplo Idade apresentado acima vamos determinar o primeiro quartil e terceiro quartil:

1º passo: Ordenar os dados

13, 13, 13, 14, 14, 14, 14, 14, 15, 15

2º passo: Buscar o valor tal que 25% dos dados observados são maiores ou iguais a ele, e simultaneamente, 75% dos dados observados são menores ou

iguais a ele. No nosso exemplo, como temos um número par de observações (10 observações), este valor será a média aritmética entre os termos que ocupam a 2ª e a 3ª posições.

$$Q_1 = \frac{13+13}{2} = 13$$

3º passo: Buscar o valor tal que 75% dos dados observados são maiores ou iguais a ele, e simultaneamente, 25% dos dados observados são menores ou iguais a ele. No nosso exemplo, como temos um número par de observações (10 observações), este valor será a média aritmética entre os termos que ocupam a 7ª e a 8ª posições.

$$Q_3 = \frac{14+14}{2} = 14$$

Interpretando os quartis: como os quartis divide o conjunto em quatro partes iguais (25%), conclui-se de forma bastante intuitiva que 25% dos alunos têm menos de 13 anos e 75% dos alunos têm mais de 13 anos. E de forma análoga 75% dos alunos tem menos de 14 anos e 25% dos alunos tem mais de 14 anos

3 PARTE EXPERIMENTAL

Acreditamos que a formação do pensamento estatístico enfatiza a necessidade e a produção dos dados, como também o estudo da variação por eles apresentada. No entanto, a capacidade de leitura e interpretação de dados, organizados em tabelas e gráficos, desenvolve habilidades condizentes com um nível de alfabetização estatístico, conforme apresentamos anteriormente em nosso capítulo 1.4.

A partir de nossa questão de pesquisa, construímos a hipótese de que os professores de Matemática do Ensino Médio, tenham uma visão pelo menos parcial da Estatística Descritiva e suas aplicações, embora não desenvolvam esse tema com seu alunos por insegurança nesses conhecimentos.

Na atividade que propusemos aos professores participantes de nosso trabalho, solicitamos a eles que analisassem uma situação-problema hipotética baseada em Bifi (2006), onde o objetivo do trabalho era verificar em que nível de alfabetização estatística, o professor de matemática do ensino médio se encontra.

3.1 ATIVIDADE PROPOSTA

Esta atividade foi proposta por Bifi (2006, p.54) e é composta por três partes.

Primeira parte

Uma Empresa de cartões de crédito solicitou uma análise do banco de dados abaixo, construído a partir das respostas a um questionário que buscava levantar a idade e a renda mensal de 40 pessoas.

TABELA 1: IDADE E RENDA MENSAL DE 40 PESSOAS (DADOS FICTÍCIOS).

Idade	Renda Mensal						
30	1.180	28	1.420	37	387	40	387
28	490	46	630	29	1.600	25	1.600
28	1.200	30	1.000	43	1.770	30	1.770
40	540	31	760	43	1.770	45	1.770
29	860	23	1.000	31	1.200	31	1.200
31	850	29	700	30	1.200	65	1.200
30	500	27	400	30	400	53	400
32	1.600	48	380	30	1.400	25	1.400
41	700	30	1.800	30	1.400	34	1.400
39	1.420	40	554	28	800	25	800

Fonte: Bifi (2006, p.54)

Questões:

- 1) Encontre, nas variáveis *idade* e *renda mensal*, a média e o desvio-padrão. Como você analisaria esses resultados?
- 2) Encontre, nas variáveis, *idade* e *renda mensal*, a mediana, o 1º quartil e o 3º quartil. Como você analisaria esses resultados?
- 3) Se você precisasse explicar o *comportamento* da variável *idade* para um cliente, você usaria o item (1) ou o item (2)? Explique por quê.

Segunda parte

De acordo com as tabelas abaixo, responda às questões:

TABELA 2: QUANTIDADE DE CARROS/PESSOA (DADOS FICTÍCIOS)

<i>Quantidade de carros</i>	<i>Número de pessoas</i>
1	10
2	25
3	15
4	5
Total	55

Fonte: Bifi (2006, p.55)

TABELA 3: N. DE HORAS NO TRÂNSITO/PESSOA (DADOS FICTÍCIO)

<i>Tempo no trânsito</i>	<i>Número de pessoas</i>
0 2	10
2 4	20
4 6	30
6 8	15
Total	75

Fonte: Bifi (2006, p.55)

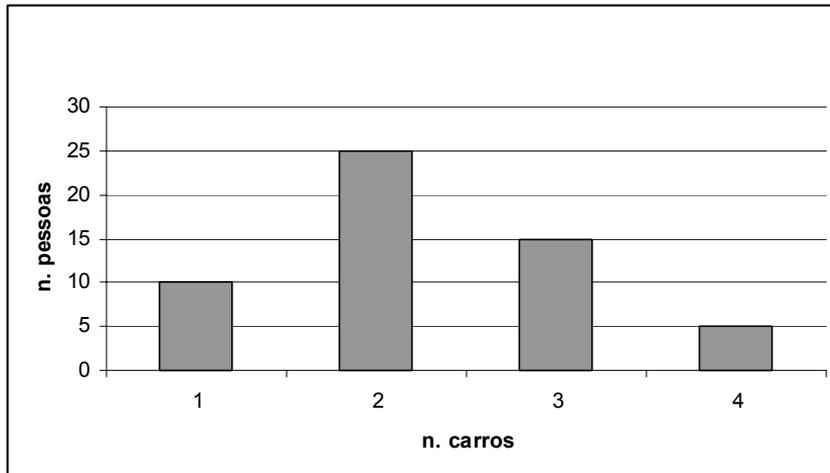
Questões

- 1) Determine, nas tabelas acima, a média e o desvio-padrão. Como você analisaria esses resultados?
- 2) Determine, nas tabelas acima, a mediana, o 1º quartil e o 3º quartil. Como você analisaria esses resultados?
- 3) Se você precisasse descrever os dados *Quantidade de carros* e *Tempo no trânsito* para um cliente, você usaria o item (1) ou o item (2)? Explique por quê.

Terceira parte

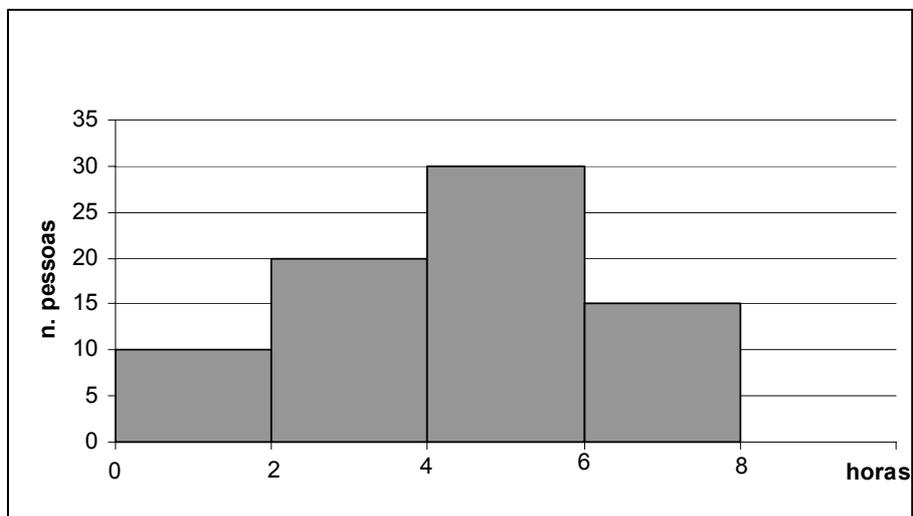
Observe os gráficos abaixo. Responda a pergunta: “Se você precisasse descrever esses dados para um cliente, como você os analisaria?”.

GRÁFICO 2: NÚMERO DE CARROS/PESSOA



Fonte: Bifi (2006, p.55)

GRÁFICO 3: HORAS NO TRÂNSITO/PESSOA.



Fonte: Bifi (2006, p.55)

3.2 ANÁLISE A PRIORI

Primeira parte

De acordo com o enunciado:

Uma Empresa de cartões de crédito solicitou uma análise do banco de dados abaixo, construído a partir das respostas a um questionário que buscava levantar a idade e a renda mensal de 40 pessoas.

A resolução da atividade será feita por meio de tabela unidimensional, a partir do rol de cada uma das variáveis, idade e renda mensal, tratando-as assim isoladamente. Ou seja, espera-se que os professores coloquem os valores observados em ordem crescente para iniciar o tratamento dos dados.

Variável idade (ROL):

TABELA 4: DADOS ORDENADOS DA VARIÁVEL IDADE.

23	25	25	25	27
28	28	28	28	29
29	29	30	30	30
30	30	30	30	30
30	31	31	31	31
32	34	37	39	40
40	40	41	43	43
45	46	48	53	65

Fonte: Bifi (2006, p.59)

A partir desse rol, o professor deverá calcular as medidas-resumo solicitadas no enunciado dos itens do enunciado: média e desvio-padrão, mediana e quartis, para em seguida proceder a análise dos valores calculados em função do contexto proposto pela situação-problema.

Aplicando o algoritmo para o cálculo da média, tem-se:

$$\bar{x} = \sum_i^n \frac{x_i}{n}, \text{ ou seja } \bar{x} = \frac{1364}{40} \cong 34,1 \text{ anos. Isto é, as idades observadas}$$

variam em torno de 34,1 anos, com valor mínimo igual a 23 e valor máximo igual a 65, variando assim com uma amplitude de 42 anos.

Para determinar o grau de concentração ao redor da média, o professor deve determinar o desvio-padrão. Assim, começa pela determinação da variância por meio do seu algoritmo: $s^2 = \frac{\sum(x-\bar{x})^2}{n}$, ou seja $s^2 = \frac{2934}{40} = 73,35$

O desvio padrão é a raiz quadrada da variância: $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} = \sqrt{\frac{2934}{40}} \cong 8,69$ anos. Esse valor deverá ser analisado em função do valor da média, para que se possa compreender a concentração (ou dispersão) dos dados ao redor dessa média.

Observando a amostra, com média de 34,1 anos aproximadamente e o desvio-padrão de aproximadamente 8,69 anos, ou seja, (34,1 – 8,69 ; 34,1 + 8,69) que representam os valores acima e abaixo da média, dessa forma podemos perceber a existência de uma maior concentração em torno da média, ou seja, uma concentração que varia entre 25,3 anos a 42,7 anos em torno de uma média de 34,1 anos, representando 72,5% do total da amostra, e essa porcentagem, é calculada através da razão entre o número de sujeitos entre 25,3 anos 42,7 anos e o total de participantes, que representa 40 sujeitos. Portanto, de cada 100 pessoas entrevistadas, 72 estão entre 25,3 e 42,7 anos de idade. Uma análise gráfica dos dados coletados apresenta uma melhor noção do comportamento da amostra. As concepções sobre média apresentam-se de várias formas pelos professores. Entre estas, está a concepção de que média é atrelada a ponto central, ou seja, um ponto médio, e que não há por parte deles, a preocupação de uma análise, *a priori*, do tipo de distribuição que estão manipulando, se são simétricas ou assimétricas. Batanero (2001) observa que a média tende a situar o centro dos dados da distribuição, propriedade que é certa para distribuições simétricas. Quando a distribuição é muito assimétrica, a média é desprezada e a moda e a mediana seriam os valores mais representativos dos dados. É necessário que o professor faça essa distinção do tipo de distribuição (simétrica ou assimétrica), para, assim, realizar uma análise correta dos dados coletados com a escolha correta do valor mais representativo para o estudo da variabilidade. Por exemplo, os quartis.

É o que pretendemos mostrar nos próximos passos.

Análise esperada envolve as medidas separatrizes: Quartis.

Colocando os dados em ordem crescente, encontramos assim o primeiro quartil, a mediana e o terceiro quartil. Vejamos:

23 25 25 25 27 28 28 28 28 29 29 29 30 30 30 30 30 30 30 30 30 31 31 31 31 32 34 37 39 40 40 40 41 43 43 45 46 48 53 65

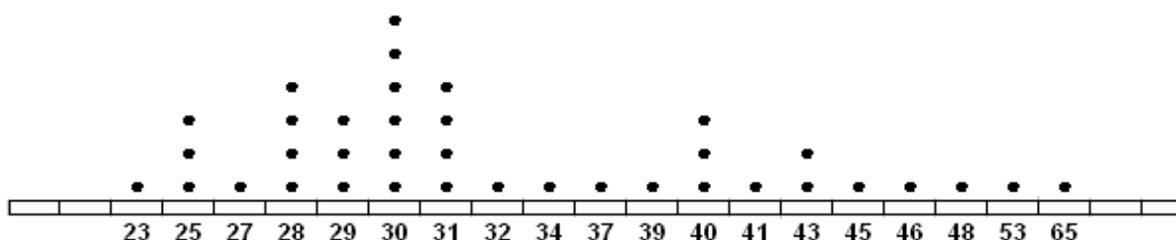


FIGURA 3: DOT-PLOT

A mediana é a medida que melhor representa a variável idade nesse rol, pois podemos analisar que, para cada quartil limitamos 25% da amostra. Nestes moldes, verificamos que, 50% da amostra está exatamente entre 23 e 30 anos. Esperamos que o Professor do Ensino Médio observe que a análise desta variável é mais representativa por meio da mediana. Na análise dos níveis de letramento citada por Shamos (1995, apud Moraes, 2006), o professor, encontra-se em nível cultural, não relacionará as medidas encontradas e, provavelmente, não conseguirá fazer uma análise dos dados por meio das representações gráficas que acabamos de mostrar. Caso aconteça o inverso, ou seja, esta análise relacional entre as medidas ocorrerem e o professor conseguir fazer uma conclusão coerente sobre os resultados, segundo Shamos (1995, apud Moraes, 2006), este professor estará em um nível funcional. Por outro lado, se o professor é capaz de fazer inferências e previsões sobre as informações contidas nos diversos registros, analisando e considerando a variabilidade existente, estará em um nível científico.

Para o cálculo do 1º e 3º quartis, para dados ordenados, será utilizado o seguinte algoritmo: $PQ_i = \frac{in}{4}$, onde $i = \{1,2,3\}$, que representa o quartil procurado e PQ_i indica a posição do quartil, e a essa posição está relacionado, o elemento que corresponde ao conjunto ordenado de valores.

O primeiro quartil (Q1): $PQ_i = \frac{in}{4} \Rightarrow PQ_1 = \frac{1 \times 40}{4} = 10$. Portanto o primeiro quartil será o décimo elemento do conjunto ordenado de valores, que corresponde a 29 anos.

O terceiro quartil (Q3): $PQ_i = \frac{in}{4} \Rightarrow PQ_3 = \frac{3 \times 40}{4} = 30$. Portanto o terceiro quartil será o trigésimo elemento do conjunto ordenado de valores, que corresponde a 40 anos.

O Cálculo da mediana (md) para dados ordenados será utilizado o seguinte algoritmo: $Pmd = \frac{n}{2}$, onde Pmd indica a posição da mediana, e essa posição está relacionada, o elemento que corresponde ao conjunto ordenado de valores.

Portanto a mediana corresponde a: $Pmd = \frac{40}{2} = 20$. Portanto o valor da mediana corresponde a vigésima posição do conjunto ordenado de valores, que representa 30 anos.

A moda, o elemento que aparece com mais freqüência dentro dessa amostra será, 30 anos.

2ª Estratégia de resolução: Cálculo das medidas por meio de uma tabela de distribuição de freqüência

De acordo com a tabela de distribuição de freqüência com intervalo de classes, calcularemos as medidas de variação e separatrizes:

Para obter o número de classes, utilizaremos a fórmula de Sturges:
 $k \cong 1 + 3,3 \cdot \log n$.

Sabendo que a amplitude da amostra é de 42 anos, usaremos a seguinte regra:

Número de classes: $h \cong \frac{H}{k}$, (sendo H a amplitude da amostra é k e corresponde ao número de classes). Em nosso exemplo: $h = \frac{42}{6,36} = 6,61$, consideremos o inteiro, desse resultado, temos 6 classes com intervalo de 7 anos para cada classe.

TABELA 5: CÁLCULOS PARA ENCONTRAR A MÉDIA E VARIÂNCIA DE DADOS AGRUPADOS

Idade	f_i	$P_m = x_i$	$X_i \cdot f_i$	F_{ac}	$X_i^2 \cdot f_i$
23 30	12	26,5	318	12	8427
30 37	15	33,5	502,5	27	16833,75
37 44	8	40,5	324	35	13122
44 51	3	47,5	142,5	38	6768,75
51 58	1	54,5	54,5	39	2970,25
58 65	1	61,5	61,5	40	3782,25
Totais	40	-	1403	-	51904

Fonte: Bifi (2006 p.63)

Cálculo das medidas de variação

$$\text{Cálculo da média: } \bar{x} = \frac{\sum_i^n x_i \cdot f_i}{n} \Rightarrow \bar{x} = \frac{1403}{40} \cong 35,075 \text{ anos}$$

$$\text{Cálculo do desvio-padrão: } s = \sqrt{\frac{\sum x^2 \cdot fi - \frac{(\sum x \cdot fi)^2}{n}}{n-1}} \Rightarrow s = \sqrt{\frac{51904 - \frac{(1403)^2}{40}}{40-1}} = 8,31$$

anos.

Devemos observar que as estatísticas não são necessariamente iguais, quando se estuda dados bruto e dados agrupados. Ao se agrupar dados e calcular as estatísticas há perda de precisão, pois os dados originais são substituídos por estimativas do intervalo de classe, nesse caso pelo ponto médio do intervalo, por isso a variância diminui. Analisando os cálculos da média e desvio-padrão, podemos observar uma não-simetria na distribuição dos dados, com uma concentração que varia entre 26,8 a 43,2 anos em torno de uma média de 35 anos, representando 38,09% do total da amostra. Isso significa que na amostra de cada 100 pessoas entrevistadas, podemos esperar que 38 estejam entre 27 e 43 anos de idade. Neste caso, também observamos, que a distribuição é assimétrica e, sendo assim, pode-se sugerir que a melhor análise poderia ser feita pelas medidas separatrizes, e não pela associação entre média e desvio-padrão.

Cálculo das medidas separatrizes:

Para os cálculos dos quartis e mediana, utilizaremos o método de interpolação linear, conforme (Bonjorno,1992, p. 315 – 316).

Identificada a classe mediana (2ª classe), temos a seguinte proporção:

$$\begin{array}{ccc} 30 & \text{md} & 37 \\ \frac{12}{40} & \frac{1}{2} & \frac{27}{40} \end{array} \Rightarrow \frac{\text{md} - 30}{\frac{1}{2} - \frac{12}{40}} = \frac{37 - 30}{\frac{27}{40} - \frac{12}{40}} \Rightarrow \text{md} = 33,73 \text{ anos}$$

Identificada a classe do 1º quartil (1ª classe), temos a proporção que se segue:

$$\begin{array}{ccc} 23 & Q_1 & 30 \\ \frac{0}{40} & \frac{1}{4} & \frac{12}{40} \end{array} \Rightarrow \frac{Q_1 - 23}{\frac{1}{4} - \frac{0}{40}} = \frac{30 - 23}{\frac{12}{40} - \frac{0}{40}} \Rightarrow Q_1 = 25,60 \text{ anos}$$

Identificada a classe do 3º quartil (3ª classe), temos a proporção que se segue:

$$\begin{array}{ccc} 37 & Q_3 & 44 \\ \frac{27}{40} & \frac{3}{4} & \frac{35}{40} \end{array} \Rightarrow \frac{Q_3 - 37}{\frac{3}{4} - \frac{27}{40}} = \frac{44 - 37}{\frac{27}{40} - \frac{35}{40}} \Rightarrow Q_3 = 39,63 \text{ anos}$$

Os dados calculados, expressos no Box-plot, nos trará a representação gráfica da distribuição da tabela 5.

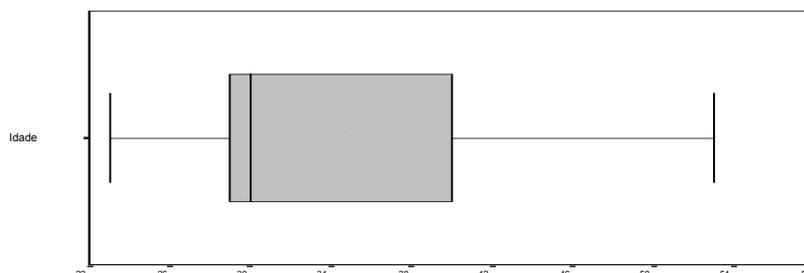


FIGURA 4: REPRESENTAÇÃO DOS QUARTIS E MEDIANA

Podemos analisar, utilizando cinco medidas estatísticas: valor mínimo, valor máximo, mediana, primeiro e terceiro quartil. Estas medidas podem ser melhor visualizadas com o uso do box-plot. Este conjunto de medidas oferece a idéia de posição, dispersão e dados discrepantes (**figura 4**). Verificamos que os resultados obtidos nos mostram uma estimativa do comportamento da amostra, que se diferencia da representação dos mesmos dados quando trabalhados na primeira estratégia. A escolha das estratégias na análise em questão fica a critério do professor, mas vale ressaltar que, em nossa análise *a posteriori*, teremos que

diagnosticar se o professor tem a consciência de que a escolha da segunda estratégia mostra apenas uma estimativa dos resultados da análise. Percebe-se maior concentração dos dados entre 23 anos e 33,73 anos (próximos dos 34 anos), isto é, 50% da amostra está exatamente entre 23 e 33,73 anos, é uma distribuição assimétrica, e há uma maior dispersão entre a mediana e o terceiro quartil e podemos observar que há uma cauda mais curta do lado direito mostrando que existe uma maior concentração entre o valor mínimo e o primeiro quartil e a cauda superior há uma maior dispersão entre o terceiro quartil e o valor máximo.

3ª estratégia: por meio de gráfico.

Utilizaremos a tabela 5 para elaboração do gráfico, levando em consideração que o Histograma é o tipo de gráfico mais usual utilizado na Escola Básica.

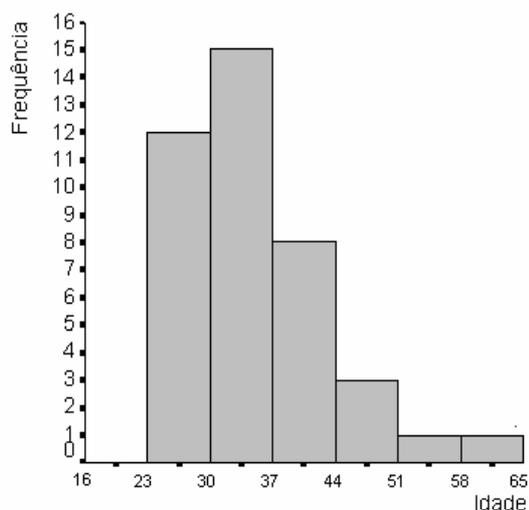


GRÁFICO 4: HISTOGRAMA DA VARIÁVEL IDADE

O histograma consiste em retângulos justapostos, de forma que a área de cada retângulo seja proporcional à frequência da classe que ela representa. Usaremos esta terceira estratégia para o estudo conjunto das medidas de variação e separatrizes.

Para o estudo das medidas de variação, representaremos como mostra o gráfico a seguir:

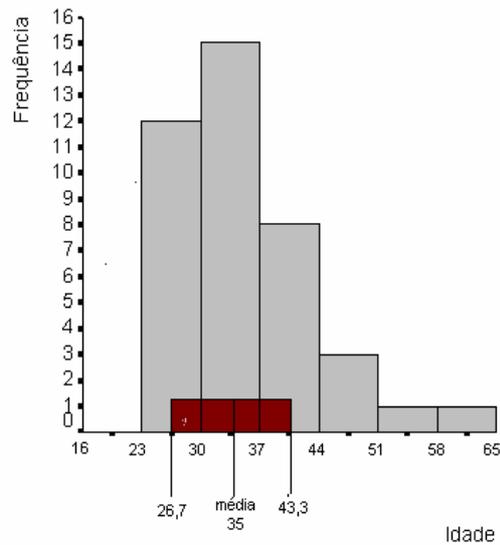


GRÁFICO 5: ESTUDO DA MÉDIA POR MEIO DO HISTOGRAMA

Com este tipo de representação é possível observar a idéia de variação em torno da média, deixando clara a relação da amplitude desta variabilidade com a amplitude da amostra. Os resultados apresentados no gráfico, média 35 anos e os valores inferior e superior à média foram obtidos conforme os cálculos das medidas de variação, apresentada na página 43.

Já para o estudo das medidas separatrizes, calculamos a mediana da variável *idade* por meio do histograma. A mediana será, neste caso, o valor da variável pelo qual metade ou cinquenta por cento da freqüência total ($N/2$) fica situada abaixo ou acima dele. Para isto pode-se usar a interpolação simples para obter uma boa aproximação desse valor, uma vez que conhecemos os limites (inferior e superior) e a freqüência mediana, conforme esquema:

$$\left\langle \begin{array}{l} \text{limite inferior (li)} \\ \text{Freqüência acumulada até li} \end{array} \right. \quad \left\langle \begin{array}{l} Md \\ \text{Freqüência da Md} \end{array} \right. \quad \left\langle \begin{array}{l} \text{Limite superior (Li)} \\ \text{Freqüência acumulada até o Li} \end{array} \right.$$

Desta forma:

$$\begin{array}{ccc} 30 & md & 37 \\ \frac{12}{40} & \frac{20}{40} & \frac{(12+15)}{40} \end{array}$$

Observe que o uso de representação fracionária dos números evita a propagação do erro de aproximação.

Assim:

$$\frac{md - 30}{\frac{20}{40} - \frac{12}{40}} = \frac{37 - 30}{\frac{27}{40} - \frac{12}{40}}$$

De onde conclui-se que a mediana é igual 33,73 anos.

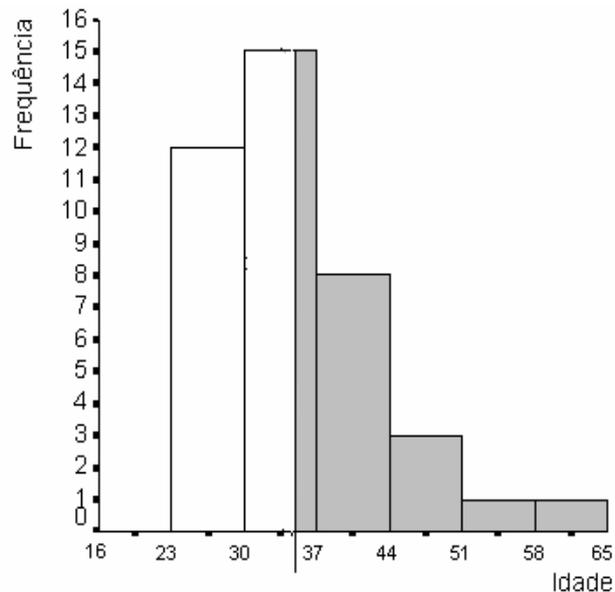


GRÁFICO 6: ESTUDO DA MEDIANA POR MEIO DO HISTOGRAMA

Para determinação dos quartis podemos usar estratégia análoga. Para tanto, deveremos determinar, além da mediana, dois valores tais que 25% das observações ordenadas estarão abaixo de um deles e 75% estarão abaixo do outro. Estes valores são o primeiro quartil e terceiro quartil de acordo com a tabela abaixo temos.

Variável renda familiar (ROL)

TABELA 6: RENDA FAMILIAR (DADOS FICTÍCIOS)

300	380	387	400	400
406	490	500	540	554
600	630	700	700	760
770	800	850	860	890
890	1000	1000	1160	1180
1200	1200	1200	1340	1370
1400	1400	1420	1420	1500
1600	1600	1770	1770	1800

Fonte: Bifi (2006, p.69)

Para o cálculo da média da renda familiar, usaremos a mesma fórmula:

$$\bar{x} = \sum_i^n \frac{x_i}{n}. \text{ Neste caso, o valor da média da renda familiar será: } \bar{x} = \frac{30137}{40} \cong 978,42 .$$

A variância será denotada pela fórmula: $s^2 = \frac{\sum (x - \bar{x})^2}{n}$, logo o valor da variância será: $s^2 = \frac{7817801,77}{40} = 195445,04$

O desvio-padrão será a raiz quadrada da variância: $\sqrt{s} \cong 442,09$.

O valor mínimo da amostra: 300 reais.

O valor máximo da amostra: 1800 reais.

Amplitude da amostra: 1500 reais.

Utilizando os processos mencionadas anteriormente, nas medidas separatrizes temos:

O primeiro quartil (Q1): 565,50 reais.

A mediana será (md): 890 reais.

O terceiro quartil (Q3): 1392,50 reais.

A moda, o elemento que aparece com mais freqüência dentro dessa amostra, será: 1200 reais.

Temos, na amostra, que a média salarial é de 978,42 reais aproximadamente e o desvio-padrão de aproximadamente 448 reais. Tendo uma amplitude amostral de 1500 reais, percebemos que o coeficiente de variação em torno da média é de aproximadamente 46%. Analisando a variabilidade em torno da média de acordo com o histograma temos:

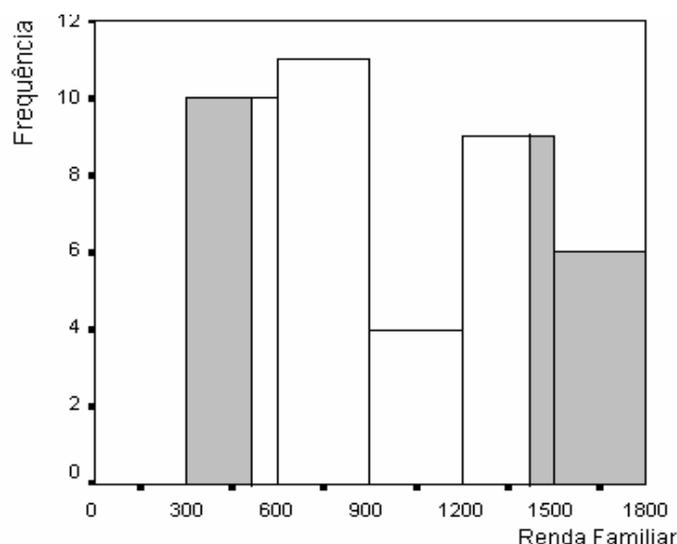


GRÁFICO 7: HISTOGRAMA DA DISTRIBUIÇÃO DE RENDA COM OS LIMITES DE REFERÊNCIA E UM DESVIO PADRÃO

Analisando o gráfico 7, podemos observar a existência de uma concentração uniforme da amostra em torno da média, ou seja, uma concentração, em renda mensal que varia entre $978,4 \pm 447,72$, que é o desvio padrão, representando 65% de toda a amostra. Essa análise gráfica nos permite, com maior segurança, verificar o comportamento da amostra e que tipo de distribuição ela apresenta (simétrica ou assimétrica). Neste caso, o professor poderá definir qual o melhor valor para a análise da variabilidade e assim, ajudar seus alunos na aquisição de autonomia para esta tomada de decisões. Para tanto, vamos fazer um estudo dos quartis, conhecimento este, esperado dos professores do ensino médio.

Utilizamos, para este estudo, o box-plot, apresentado na figura 5.

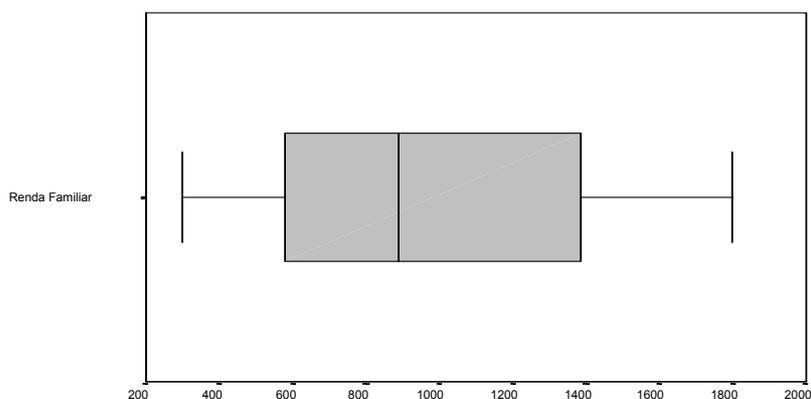


FIGURA 5: VARIABILIDADE EM TORNO DA MEDIANA

Como podemos verificar na **figura 5**, o box-plot nos fornece a oportunidade para uma análise mais detalhada e com maior precisão do comportamento dos dados coletados, e pela facilidade oferecida pela visualização da distribuição percebemos que há uma ligeira concentração à esquerda dos dados em relação à mediana. Nota-se uma maior concentração dos valores entre o menor valor e a mediana, ou seja, entre R\$300,00 e R\$890,00. O que pretendemos mostrar é que, ao analisar a variabilidade de um conjunto de dados, não podemos simplesmente nos ater a um só tipo de medida, por exemplo, a média, mesmo que ela seja um valor representativo. É imprescindível que o professor do ensino médio, tenha as habilidades e competências de analisar a variabilidade de conjunto de dados, escolhendo, dessa forma o valor que melhor representa o conjunto. Observando o histograma e o box-plot conjuntamente, podemos observar que na realidade, as medidas-resumo fornecem informações complementares sobre o conjunto de dados.

Segunda Parte

Nesta segunda parte, a distribuição apresentar-se-á na forma tabular, para que possamos diagnosticar se os níveis de mobilização dos conhecimentos estatísticos dos professores se alteram pelo uso de outra forma de representação dos dados. Conforme Bifi (2006), adaptado a esta pesquisa

[...]dessa forma, até por conta da mesma necessidade dos cálculos (média, quartis e desvio-padrão), acreditamos ser mais viável e fácil para o professor, bastando ele completar a tabela com colunas auxiliares para encontrar os valores pedidos, porém não garantindo o sucesso na atividade". (Bifi, 2006, p.72)

Questões

- 1) Determine para tabela 7, a média e o desvio-padrão. Como você analisaria esses resultados?
- 2) Determine para tabela 7 a mediana, o 1º quartil e o 3º quartil. Como você analisaria esses resultados?
- 3) Se você precisasse descrever os dados "Número de carros" e "Tempo no trânsito" para um cliente, você usaria o item (1) ou o item (2)? Explique por quê.

TABELA 7: DISTRIBUIÇÃO DE FREQUÊNCIA - QTDE. CARROS/PESSOA

<i>Qtde de carros (xi)</i>	<i>N. de pessoas (fi)</i>	<i>Fri(%)</i>	<i>xi.fi</i>	<i>xi.².fi</i>
1	10	18,18	10	10
2	25	45,45	50	100
3	15	27,27	45	135
4	5	9,1	20	80
Total	55	100	125	325

Fonte: Bifi (2006, p.72)

Começemos os cálculos completando a **tabela 7**, fornecida na 2ª parte

1ª estratégia: Cálculo das medidas a partir da distribuição de frequência

1) Questão 1: Cálculo da média e desvio-padrão.

A média neste caso será $\bar{x} = \frac{\sum xi.fi}{n} \Rightarrow \bar{x} = \frac{125}{55} \cong 2,27.$

O desvio padrão será $s = \sqrt{\frac{\sum x^2.fi - \frac{(\sum x.fi)^2}{n}}{n-1}} \Rightarrow s = \sqrt{\frac{325 - \frac{(125)^2}{55}}{54}} = 0,87.$

Análise dos dados, que esperamos que o professor desenvolva.

Nas medidas de variação, temos, na amostra, que a média é de 2,27 carros por pessoa e desvio-padrão de aproximadamente 0,87 carros por pessoa. Tendo uma amplitude amostral de 3 carros, (4 – 1) percebemos que o coeficiente de variação em torno da média é de aproximadamente 38,3%. A amplitude em torno da média será 0,87 carro.

Ou seja, podemos perceber que a dispersão dos dados não é alta, pois a média sendo 2,27 e o desvio-padrão 0,87, observando o intervalo entre a média e o desvio padrão, ou seja, 2,27 – 0,87 e 2,27 + 0,87, representando uma variação entre 1,40 e 3,14 carros, temos uma ocorrência de 40 carros, em um total de 55 carros, ou seja, representa aproximadamente que 72% dos valores estão no intervalo da média menos o desvio padrão e a média mais o desvio padrão ($\bar{x} - s$ e $\bar{x} + s$).

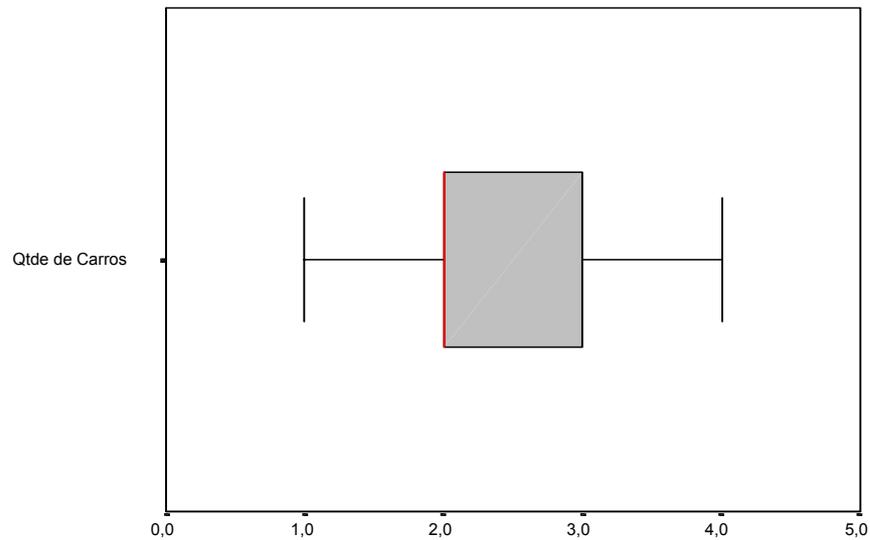


FIGURA 8: REPRESENTAÇÃO DA QUANTIDADE DE CARRO/PESSOA

Podemos verificar claramente que o primeiro quartil e a mediana coincidem, havendo uma concentração dos dados entre 1 e 2 carros. Analisando o box-plot, podemos observar que a mediana é um valor representativo para o estudo da variabilidade do conjunto de dados, completando a informação obtida pelo estudo da média e desvio-padrão

2ª Estratégia: resolução por meio de gráfico

Levantamos a hipótese de que gráficos estatísticos podem ser um facilitador, na aprendizagem dos conceitos de base da Estatística Descritiva, trazendo informações resumidas das informações, e talvez isso seja o suficiente para o tipo de análise. Por outro lado, poderá ser um complicador, já que, para distribuições com intervalo de classes, há perda de informação e a análise poderá não mostrar o que realmente se investiga. Cabe ao professor as escolhas para o desenvolvimento das atividades em sala de aula. Por meio do gráfico 8 tentamos analisar o comportamento do conjunto de dados em estudo.

Número de Carros por pessoa

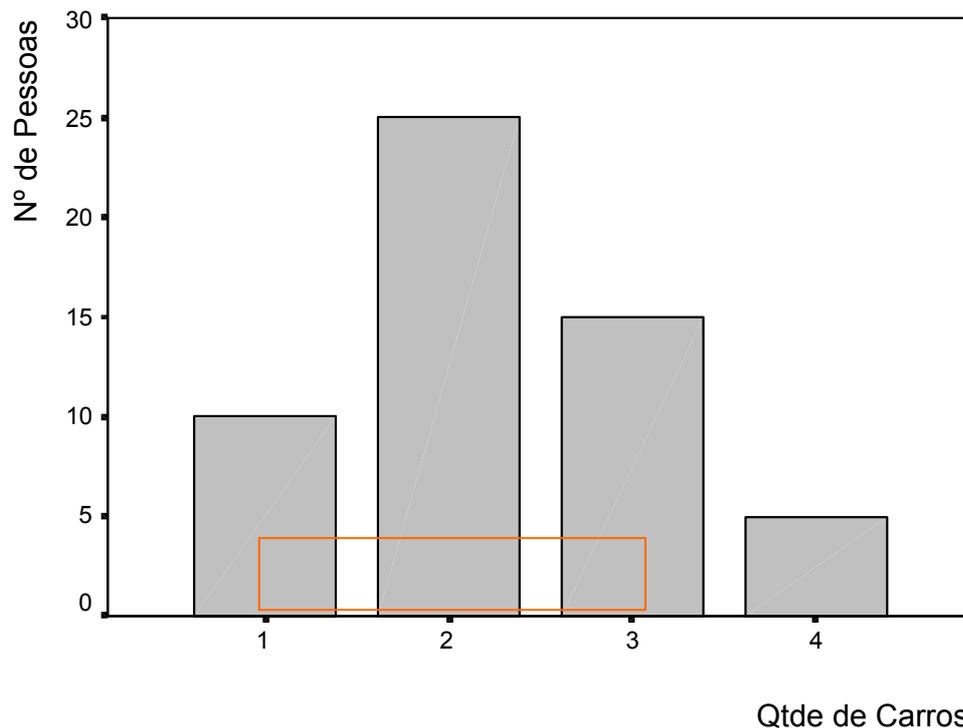


GRÁFICO 8: ESTUDO DAS MEDIDAS CENTRAIS POR MEIO DE GRÁFICOS.

Calculados os valores, podemos representá-los no gráfico e ter uma noção clara do comportamento da variável e, por outro lado, também é fácil perceber que existem valores que estão distantes do intervalo da variação em torno da média. Nesse caso, percebemos a necessidade de recorrermos novamente às medidas separatrizes e verificar qual a melhor escolha para a representação da variabilidade. Vamos analisar a mediana no gráfico 9.

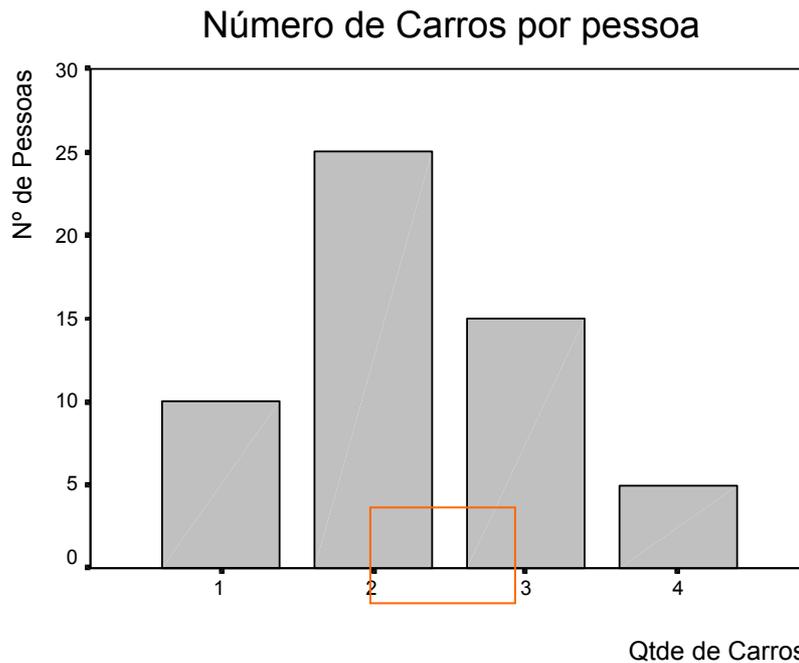


GRÁFICO 9: ESTUDO DAS MEDIDAS SEPARATRIZES POR MEIO DE GRÁFICOS.

Quando mostramos as medidas separatrizes (p.60), percebemos que o primeiro quartil e a mediana coincidem e, neste caso, também verificamos uma maior concentração dos valores nos primeiros 50% da amostra. Esperamos que, nessa análise, feita pelo professor, ele perceba a importância do uso conjunto das medidas-resumo.

Começamos os cálculos, completando a **tabela 3**, fornecida na 2ª parte do apêndice.

TABELA 8: DISTRIBUIÇÃO DE FREQUÊNCIA COM INTERVALO DE CLASSE DO NÚMERO DE HORAS NO TRÂNSITO POR PESSOA.

<i>Tempo no trânsito</i>	<i>N. de pessoas (fi)</i>	<i>P.médio(xi)</i>	<i>xi.fi</i>	<i>xi².fi</i>	<i>Fri(%)</i>
0 2	10	1	10	10	13,33
2 4	20	3	60	180	26,67
4 6	30	5	150	750	40
6 8	15	7	105	735	20
Total	75		325	1665	100

Fonte: Bifi (2006, p.76)

1ª estratégia: Cálculo das medidas a partir da tabela

1) Cálculo da média e desvio-padrão.

A média neste caso será $\bar{x} = \frac{\sum x_i \cdot f_i}{n} \Rightarrow \bar{x} = \frac{325}{75} \cong 4,33$ horas por pessoa;

O desvio padrão será $s = \sqrt{\frac{\sum x^2 \cdot f_i - \frac{(\sum x \cdot f_i)^2}{n}}{n-1}} \Rightarrow s = \sqrt{\frac{1665 - \frac{(325)^2}{75}}{74}} = 1,86$ horas por pessoa.

A variável em torno da média apresenta uma amplitude de 3,72 horas, ou seja $\bar{x} \pm s$, representando aproximadamente 66,7% da amplitude total da amostra. Portanto, não podemos afirmar se a média representa a variabilidade da amostra sem realizar o estudo do 1º quartil, mediana e 3º quartil, que esperamos que os professores investiguem, que medida utilizar para a interpretação dos dados

2) Cálculo da mediana, 1º quartil e 3º quartil

Como a distribuição está com intervalos de classes, utilizaremos o método de interpolação; já apresentado anteriormente neste texto.

Identificada a classe mediana (3º classe), temos a seguinte proporção:

$$\begin{array}{ccc} 4 & Md & 6 \\ \frac{30}{75} & \frac{1}{2} & \frac{60}{75} \end{array} \Rightarrow \frac{Md - 4}{\frac{1}{2} - \frac{30}{75}} = \frac{6 - 4}{\frac{60}{75} - \frac{30}{75}} \Rightarrow Md = 4,5 \text{ anos}$$

Identificada a classe do 1º quartil (2ª classe), temos a proporção que se segue:

$$\begin{array}{ccc} 2 & Q_1 & 4 \\ \frac{10}{75} & \frac{1}{4} & \frac{30}{75} \end{array} \Rightarrow \frac{Q_1 - 2}{\frac{1}{4} - \frac{10}{75}} = \frac{4 - 2}{\frac{30}{75} - \frac{10}{75}} \Rightarrow Q_1 = 2,88 \text{ anos}$$

Identificada a classe do 3º quartil (3ª classe), temos a proporção que se segue:

$$\begin{array}{ccc} 4 & Q_3 & 6 \\ \frac{30}{75} & \frac{3}{4} & \frac{60}{75} \end{array} \Rightarrow \frac{Q_3 - 4}{\frac{3}{4} - \frac{30}{75}} = \frac{6 - 4}{\frac{60}{75} - \frac{30}{75}} \Rightarrow Q_3 = 5,75 \text{ anos}$$

Análise dos dados, que esperamos que o professor desenvolva:

Nas medidas de variação, observamos que a média é de aproximadamente 4,33 horas por pessoa com desvio-padrão igual a 1,86 horas por pessoa. A amplitude total é de 8 horas por pessoa, com coeficiente de

variação em torno de 50%. Analisaremos estes dados por meio de uma representação geométrica, ou seja, o box-plot., apresentada na figura 9.

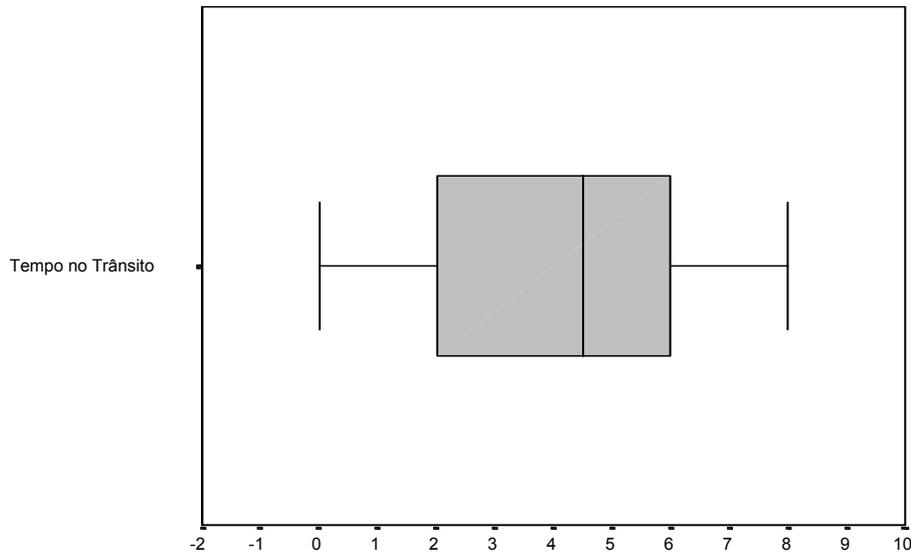


FIGURA 9: VARIABILIDADE EM TORNO DA MEDIANA.

Verificamos, de acordo com a figura 9, mais uma vez, a concentração dos dados (à direita) da mediana, confirmando assim, que a média, mais uma vez, não é um bom valor único para análise dos dados, e sim a mediana, por ela explicar melhor o comportamento da amostra (conjunto: Md e \bar{x}).

2ª estratégia: representação gráfica

Por meio do gráfico 10, p.66, tentamos analisar o comportamento do conjunto de dados em estudo (tabela 4). Mostraremos, no gráfico, os intervalos dos desvios em relação à média e, posteriormente, mostraremos os intervalos interquartis. Nessa primeira apresentação mostraremos o intervalo dos desvios em relação à média.

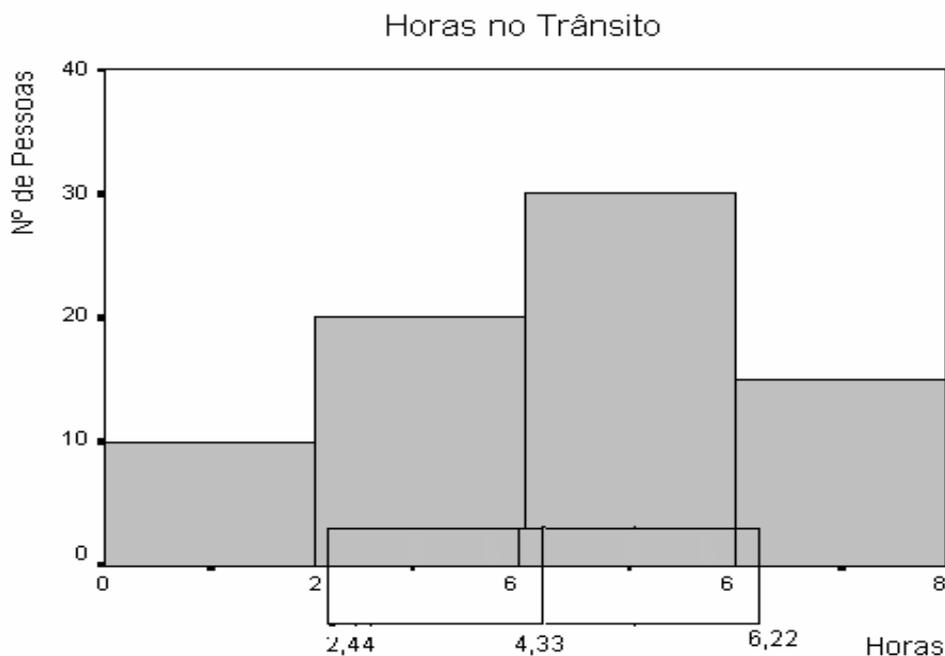


GRÁFICO 10: ANÁLISE DAS MEDIDAS CENTRAIS DA TABELA 4 POR MEIO DE GRÁFICOS.

Podemos perceber, no gráfico 10, a concentração dos dados em torno da média e, a partir daí, tirar algumas conclusões pertinentes ao conjunto de dados em estudo. Por exemplo, que a média está na classe de 4 a 6, que os desvios apontam para no mínimo 2 horas e no máximo 6 horas de tempo no trânsito. Cálculo da mediana, baseando-se no gráfico acima.

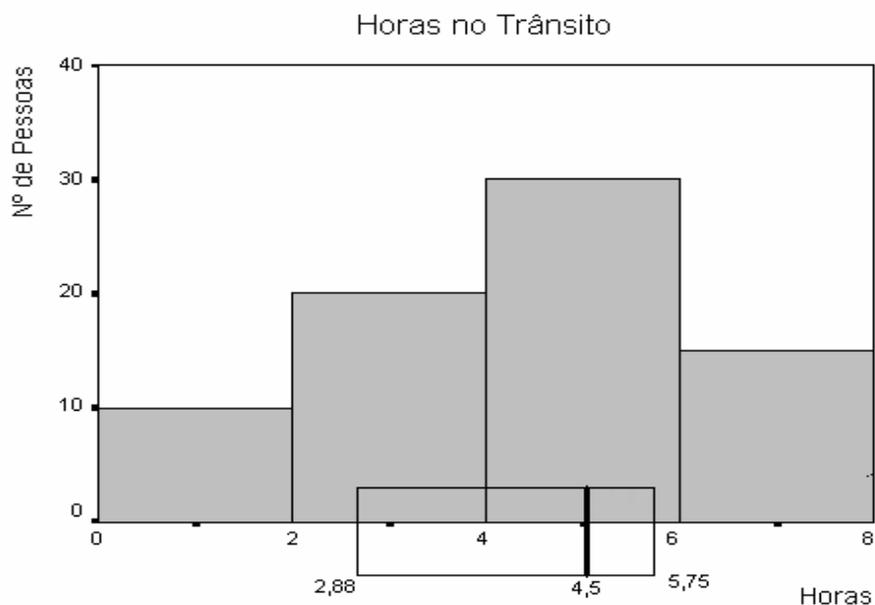


GRÁFICO 11: ANÁLISE DAS MEDIDAS SEPARATRIZES DA TABELA 4 POR MEIO DE GRÁFICOS.

Podemos verificar, no gráfico 11, uma maior concentração à direita, ou seja 25% está entre Q_1 e a Md contra 25% entre a Md e Q_3 , dessa forma podemos observar uma assimetria a direita. Portanto, é prudente analisarmos o

conjunto de dados pela mediana, pois ela é que está representando melhor a amostra dos dados.

Terceira parte

Observando os gráficos (1 e 2), pergunta-se: “Se você precisasse descrever esses dados para um cliente, como você os analisaria?” Também enfocamos a necessidade de um outro tipo de representação de dados para que possamos diagnosticar as dificuldades que poderão surgir neste tipo de representação.

Análise dos Gráficos

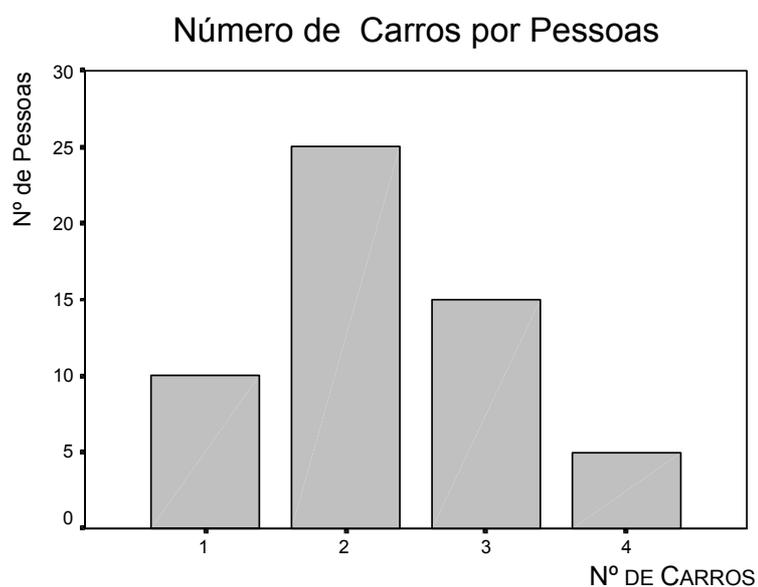


GRÁFICO 12: RETOMADA DO GRÁFICO 1.

De acordo com o gráfico 12, foram entrevistadas 55 pessoas. Destas pessoas, 10 possuem apenas 1 (um) carro, 25 possuem 2 (dois) carros, 15 possuem 3 (três) carros e 5 (cinco) possuem 4 carros. A média de carros por pessoa é de 2,27 carros, com desvio-padrão de 1,29.

A partir do gráfico 12 pode-se, determinar as medidas resumo, completando assim a análise tal como foi mostrado na parte 2.

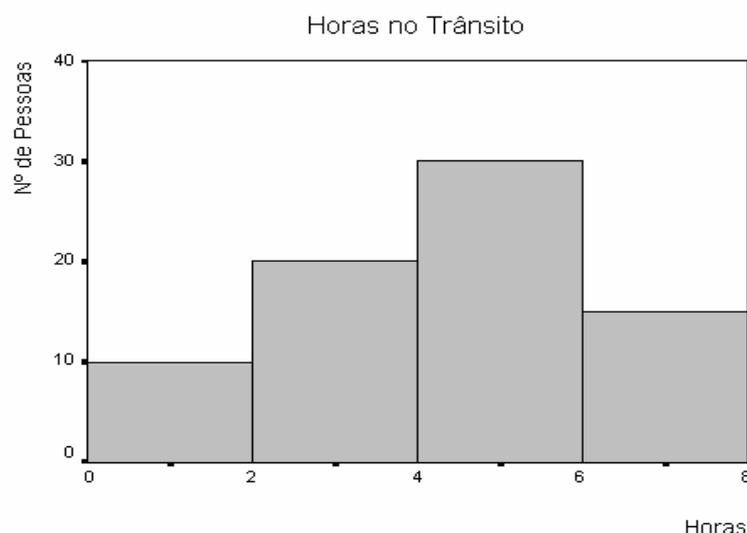


GRÁFICO 13: RETOMADA DO GRÁFICO 2.

Apresentaremos aos professores, os gráficos 12 e 13, que reforçarão nossa investigação sobre os níveis de funcionamento dos conceitos estatísticos, pois eles podem favorecer estratégias mais intuitivas, que poderão surgir nas resoluções que se apresentarão pelos professores. Esta terceira etapa se faz necessária se, por acaso, nas etapas passadas, não surgirem estratégias deste tipo. Porém vale ressaltar que a resolução desta terceira etapa já foi abordada em nossa análise na parte 2.

Não podemos deixar de observar que os resultados obtidos por meio destas estratégias não estão atrelados necessariamente a cálculos algébricos. Em procedimentos mais intuitivos, o gráfico permite uma apreensão visual da variação dos dados, permitindo ao sujeito (professor e ou aluno) a descrição simples daquilo que ele observa, sem uso das medidas-resumos. Por exemplo, pode perceber que a amplitude total, a forma da distribuição (simétrica ou assimétrica), a região de concentração dos dados. Na resolução por meio de gráficos, também devemos nos preocupar com possíveis erros e falsas interpretações que poderão surgir durante a aplicação da atividade. Segundo Batanero (2001), os professores de Matemática do Ensino Médio supõem, às vezes, que a elaboração de gráficos e tabelas é muito simples e dedicam pouco tempo para trabalhar estes tópicos. Curcio (1989) estudou que, a compreensão e as relações matemáticas expressas nos gráficos possuem os seguintes fatores:

- 1) conhecimento prévio do tema que se refere o gráfico;

- 2) conhecimento prévio do conteúdo matemático do gráfico, isto é, os conceitos numéricos, relações e operações contidas nos mesmos;
- 3) conhecimento prévio do tipo de gráfico implantado (barras, pictogramas, etc.);

Independentemente das dificuldades que poderão surgir, as estratégias mostradas no nosso trabalho, nas três partes da atividade, permitirão ao professor calcular as medidas de variação (média, desvio-padrão e coeficiente de variação) e também as medidas separatrizes (quartis e mediana), fazendo com que o professor faça uma associação entre essas medidas de variação, levando-o, assim, a um estudo da variabilidade em torno da média. Ou, ainda, uma associação entre as medidas separatrizes (mediana e quartis), e da variabilidade em torno da mediana. Os dois casos, no entanto, estão associados com a amplitude total da amostra. A não associação entre estas medidas nos leva a crer que o professor encontra-se, segundo Shamos (1995, apud Morais, 2006), em um nível cultural de alfabetização estatística, levando a uma mobilização dos conhecimentos em nível técnico.

A situação-problema apresentada no apêndice, não sugere explicitamente que o professor faça associações entre as medidas. Em nossa pesquisa, esperamos encontrar professores que as façam espontaneamente para responder ao item (3), caracterizando, segundo Shamos (1995, apud Morais, 2006), um nível funcional de alfabetização.

Para a nossa análise consideramos:

a) tamanho da amostra: é necessário que o professor saiba identificar exatamente o número de elementos da sua população-amostra; e identificar se é possível trabalhar com toda população ou parte dela, observando em que essa escolha poderá afetar o resultado final. O tamanho da amostra será o valor que estará envolvido praticamente em todos os cálculos da análise exploratória dos dados;

b) amplitude da amostra (diferença entre o maior e o menor valor dos elementos da amostra): tal representação pode sinalizar ao professor, uma medida rápida da variação dos dados, pois a amplitude depende apenas de dois valores: o maior e o menor de um conjunto de dados.

Segundo Robert (1998, apud Bifi, 2006) um conhecimento é mobilizado no nível técnico quando é feita simplesmente a substituição em fórmulas e expressões, sem adaptações deste conhecimento.

Assumindo que os professores já passaram pela disciplina Estatística, entendem porque e como os dados são produzidos, mas há necessidade de familiarização dos conceitos básicos e apresentações de dados que são utilizados para a divulgação do público-alvo.

Segundo Parker e Leinhardt, (1995) dois conceitos são muito comuns, “por cento” e medidas de tendência central, principalmente média e mediana, ou seja, são valores que comumente aparecem porcentagem e média é amplamente divulgada, tanto na mídia como em livros didáticos desde as séries iniciais.

3.3 ANÁLISE A POSTERIORI

Voltaremos ao enunciado da nossa atividade para que o leitor acompanhe melhor esta análise. Nosso instrumento (**anexo I**) foi organizado três etapas: na primeira, apresentamos um banco de dados de variável qualitativa discreta; na segunda, duas tabelas representando uma distribuição de frequências de variável quantitativa discreta (1ª tabela – quantidade de carros por pessoa) e variável quantitativa contínua (2ª tabela com intervalo de classes – tempo no trânsito por pessoa); e, na terceira, uma distribuição de frequências representada graficamente número de carros por pessoa. Para as duas primeiras partes da atividade, foi pedido para que os professores calculassem a média e o desvio-padrão, assim como que calculassem a mediana e quartis. Em seguida, foi solicitado aos professores que fizessem uma análise dos cálculos e relatassem quais dos conjuntos de medidas encontradas representava melhor a amostra. Para a última parte da atividade, pediu-se aos professores que analisassem os dois gráficos apresentados e que respondessem a seguinte questão: “Se você precisasse descrever esses dados para um cliente, como você os analisaria”.

Para a aplicação do instrumento assim composto, os professores resolveram individualmente os problemas propostos. Em um primeiro momento, buscaremos estudar a associação entre as variáveis identificadas nas estratégias de resolução apresentadas pelos professores, pertinentes em relação aos nossos

objetivos, com auxílio do software C.H.I.C. (Classificação Hierárquica Implicativa e Coesitiva).

Em um segundo momento, estudaremos a estratégia desenvolvida por cada professor de acordo com o descrito no nosso quadro teórico, no qual a alfabetização estatística é retratada como a habilidade para interpretar, avaliar criticamente e, se necessário, discutir sobre: informações estatísticas, argumentos e mensagens (Gal 2002). Nesse contexto, procuramos identificar na análise individual de cada uma das atividades entregues pelos professores entrevistados, os **níveis de alfabetização** (Shamos,1995, apud Moraes, 2006), categorizados como: cultural, funcional e científico.

3.3.1 Análise Hierárquica de Similaridade

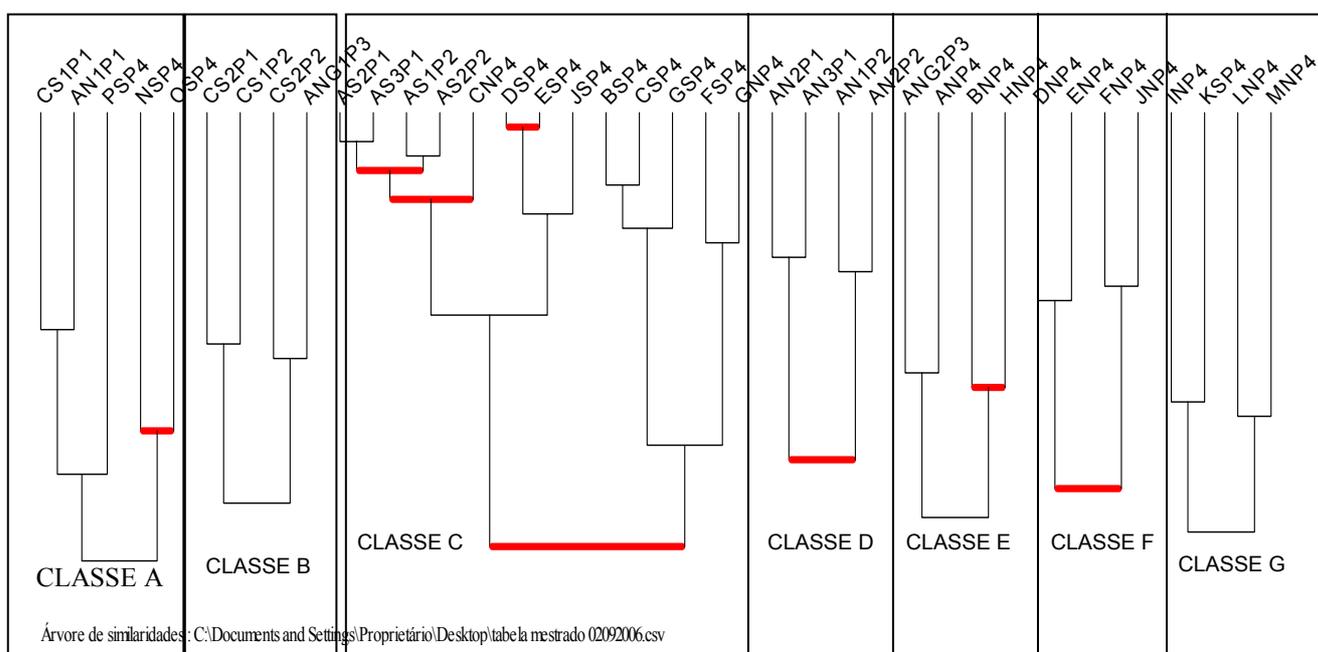
Para analisar esse questionário, anexo I, contamos com o auxílio do *software* Classificação Hierárquica, Implicativa e Coesitiva –C.H.I.C.– que possibilita análise hierárquica de similaridade e/ou coesitiva dos dados qualitativos coletados, evidenciando as inter-relações encontradas entre as respostas dadas por cada professor. Para utilizarmos o programa C.H.I.C., inicialmente, codificamos todas as respostas dadas pelos professores no instrumento diagnóstico.

Depois de estabelecidas as variáveis principais, (Apêndice E) e as variáveis suplementares (Apêndice F), ambas devidamente codificadas, elaboramos uma planilha eletrônica (Excel) explicitando os códigos atribuídos anteriormente a cada variável, em cada resposta dada por cada professor. Nessa tabela, cada linha identifica o professor e cada coluna a codificação recebida, por questão. Cada código dessa planilha tinha duas possibilidades de ocorrência, 0 ou 1, as quais atribuímos 1 para a presença do atributo ou 0 para a ausência do mesmo (variáveis binárias e, portanto, dicotômicas). Assim, a seguinte configuração para o professor, para o caso do professor estar na categoria TM3 s: TM1 s = 0; TM2 s = 0; TM3 s = 1; TM4 s = 0 e TM5 s = 0. Após a codificação dos dados, na planilha do Excel, foi submetido ao *software* C.H.I.C., solicitando, primeiramente, uma classificação das variáveis principais e das suplementares por meio de uma análise hierárquica de similaridade que possibilita uma visualização de semelhanças de comportamento mapeadas em níveis de uma

árvore hierárquica. As variáveis são organizadas e analisadas segundo agrupamentos que esse *software* disponibiliza.

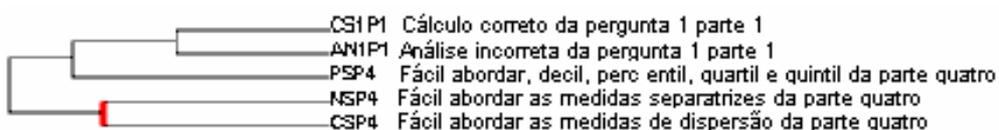
Faremos uma análise de baseada na árvore de similaridade construída pelo C.H.I.C., estabelecendo agrupamentos significativos entre as variáveis principais. Estudaremos os agrupamentos formados em relação ao nível de similaridade calculado (que indica a probabilidade de que os sujeitos tenham o mesmo comportamento em relação às variáveis associadas). Analisaremos também a contribuição de cada variável suplementar (variáveis de caracterização dos sujeitos) para o agrupamento identificado.

Ver codificação nos apêndices E e F.



Este gráfico nos permite observar a presença de sete classes, representadas acima. Apenas para facilitar a compreensão dessa nossa análise, buscaremos reproduzir e analisar cada uma das classes individualmente.

Classe A

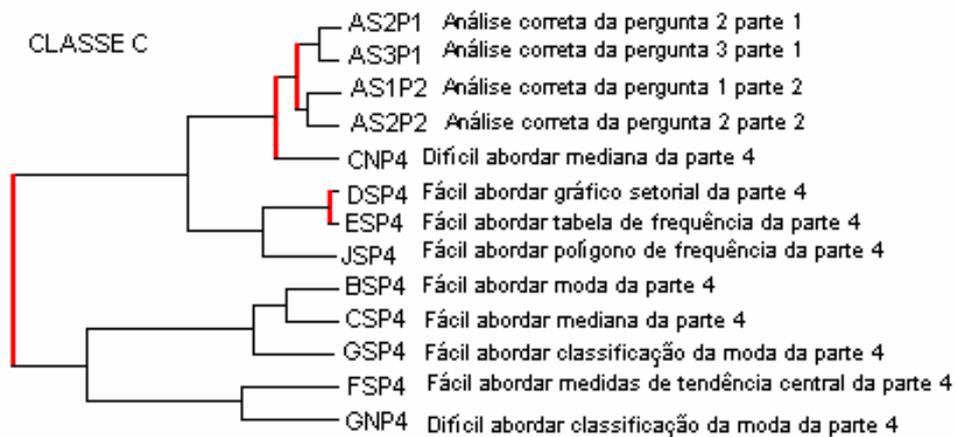


A Classe A apresenta um nó significativo (grau de associação entre as variáveis) das variáveis NSP4 (Fácil abordar as medidas separatrizes da parte quatro) e OSP4 (Fácil abordar medidas de dispersão da parte quatro), com um grau de similaridade igual a 0,5. O grau de similaridade igual a 0,5 significa a probabilidade dos professores que responderam NSP4 e OSP4 simultaneamente⁶. Ainda nessa classe, podemos observar o agrupamento CS1P1 (Cálculo correto da pergunta 1 parte 1⁷) com AN1P1 (Análise incorreta da pergunta 1 parte 1⁸), com grau de similaridade igual a 0,5. Comparando a afirmação do professor (conhece o objeto) com o que ele fez (colocou corretamente mas não analisou), podemos inferir a concepção de estatística desse professor como tecnicista. Uma outra sub-classe é observada quando o grupo (CS1P1, AN1P1) é associado à variável PSP4, (O professor afirma ser fácil abordar, decil, percentil, quartil e quintil) com grau de similaridade igual a 0,25 (baixo), ou seja aumenta a probabilidade de não associação destas variáveis. Quanto à contribuição e à tipicidade, relativa à Classe A, a variável suplementar que mais contribui para ambas foi TM4 (19 a 30 anos de magistério, de acordo com as fases propostas por de HUBERMAN,(1989, apud Bolívar, 1990, p.53), que corresponde a um risco de 0,245.

⁶ Lembramos que a parte 4 consta de uma lista de termos que o professor deve assinalar caso afirme conhecer. Ou seja, é uma declaração do professor sobre o que ele conhece em conceitos estatísticos. Para melhor análise, esta declaração deve ser confrontada com o que o professor fez efetivamente nas parte 1,2 e 3 do instrumento

⁷ Determinar a média e o desvio padrão

⁸ Analisar os resultados da média e do desvio-padrão,



A classe C apresenta quatro nós significativos (grau de associação entre as variáveis). No outro bloco, identificado pela comprovação da análise correta de resultados de cálculo das partes 1 e 2, temos:

- os professores que fizeram a análise correta da pergunta 2, (mediana e quartis), parte 1 também o fizeram para a pergunta e, parte 1 (explicar corretamente o comportamento da variável, com grau de similaridade de 0,999999.

- Os professores que fizeram análise correta da pergunta 1 parte 2 (média e desvio-padrão) também o fizeram para a pergunta 2, parte 2 (mediana e quartis) com similaridade igual a 0,999999.

- os dois grupos acima têm mesmo comportamento com índice de similaridade igual a 0,999997.

- os professores deste novo agrupamento também afirmam ser difícil abordar a mediana com similaridade 0,989419

- os professores que responderam de fácil abordagem os gráficos de setores da parte 4, também responderam ser fácil abordar tabela de frequência da parte 4, com grau de similaridade de 1.

- os professores deste novo agrupamento também afirmam ser fácil abordar polígono de frequência da parte 4, com similaridade de 0,920425.

- os professores que afirmam ser fácil abordar moda da parte 4, também o fizeram com relação à abordagem da mediana da parte 4, com similaridade de 0,999804.

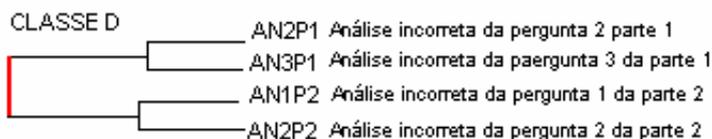
- os professores deste agrupamento também afirmam ser fácil abordar classificação da moda da parte 4, com similaridade de 0,839073.

- os professores que admitem ser fácil abordar medidas de tendência central da parte 4, também o fizeram com a questão de difícil abordagem da moda da parte 4, com similaridade de 0,786422.

- os professores deste agrupamento acreditam ser fácil abordar moda, mediana e classificação da moda, com similaridade de 0,301605 que indica uma dessemelhança entre as variáveis.

- os professores deste grupo, também afirmam a dificuldade de análise correta das perguntas 2 e 3 da parte 1 e da pergunta 1 e 2 da parte 2 e mediana da parte 4, porém a facilidade em abordar gráfico de setor, tabela de frequência, polígono de frequência, com similaridade de 0,536924.

Referindo-se à contribuição e à tipicidade, da Classe C, a variável suplementar que mais contribuiu para o sub grupo ((AS2P1 AS3P1),(AS1P2 AS2P2)) CNP4) foi TM3 (7 a 18 anos de magistério, de acordo com as fases de HUBERMAN (1989), apud Bolívar p.53) , que corresponde a um risco de 0,343. Entretanto, o subgrupo (AS2P1 AS3P1) (AS1P2 AS2P2)) CNP4) ((DSP4 ESP4) JSP4)), representado pela variável suplementar TM4 (19 a 30 anos de magistério, de acordo com a fases de HUBERMAN (1989, apud Bolívar, 1990, p.53), correspondendo a um risco de 0,245, representando a confiabilidade entre as respostas.



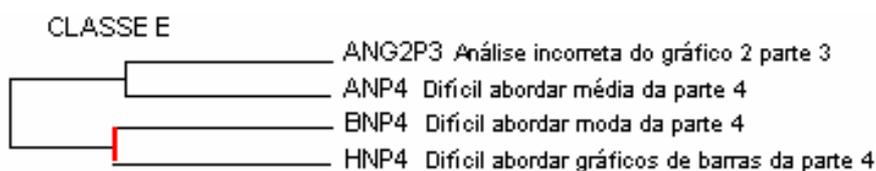
A classe D apresenta um nó significativo (grau de associação entre as variáveis). Neste bloco, identificado pela comprovação da análise incorreta das perguntas da parte 1 e 2 temos:

- os professores que fizeram análise incorreta da pergunta 2 parte 1 (mediana, primeiro e terceiro quartil), também o fizeram para a pergunta 3 parte 1 (comportamento da variável idade), com similaridade de 0,711226.

- os professores que fizeram análise incorreta 1 da parte 2 (média e desvio-padrão), também o fizeram para a pergunta 2 da parte 2 (mediana, primeiro e terceiro quartil), com similaridade igual a 0,711266.

- os dois grupos tem o mesmo comportamento com índice de similaridade igual a 0,255934, que é baixo, ocorrendo uma dessemelhança entre as variáveis, ou seja os professores não relacionam as variáveis entre si, simplesmente aceitam seus resultados.

Referindo-se à contribuição e à tipicidade, referente à Classe D, a variável suplementar que mais contribuiu para ambas foi TM4 (19 a 30 anos de magistério, de acordo com as fases de HUBERMAN (1989, apud Bolívar, 1990 p.53), que corresponde a um risco de 0,417.



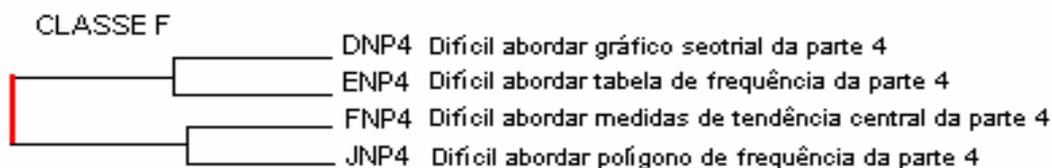
A classe E apresenta um nó significativo (grau de associação entre as variáveis) Neste bloco, identificado pela comprovação da análise incorreta do gráfico da parte e dificuldade de abordagem da média, moda e gráficos de barra da parte 4 temos:

- os professores que fizeram análise incorreta do gráfico 2 parte 3 (horas no trânsito), também tem dificuldade de abordar média na parte 4, com similaridade igual a 0,5.

- os professores que tem dificuldade em abordar moda na parte 4, também tem dificuldade em abordar gráfico de barras na parte 4, com similaridade igual 0,5.

- os dois grupos acima tem o mesmo comportamento com índice de similaridade igual a 0,0625, indicando uma dessemelhança entre as variáveis.

Referindo-se à contribuição e à tipicidade, referente à Classe E, a variável suplementar que mais contribuiu foi TM3 (7 a 18 anos de magistério, de acordo com as fases de HUBERMAN (1989), apud Bolívar, 1990, p. 53), que corresponde a um risco de 0,594, representando a confiabilidade entre as respostas



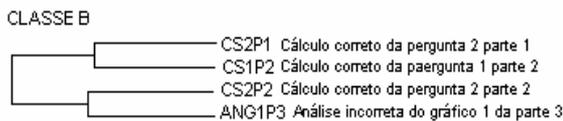
A classe F apresenta um nó significativo (grau de associação entre as variáveis) Neste bloco, identificado pela dificuldade de abordar gráfico setorial, tabela de frequência, medidas de tendência central e polígono de frequência, temos:

- os professores em gráfico setorial da parte 4, também tem dificuldade em abordar tabela de frequência, com similaridade de 0,573658.

- os professores que tem dificuldade em abordar medidas de tendência central e polígono de frequência na parte 4, com grau de similaridade igual a 0,625799.

- os dois grupos acima tem o mesmo comportamento, com índice de similaridade igual a 0,102173, onde há uma dessemelhança entre as variáveis.

Referindo-se à contribuição e à tipicidade, referente à Classe E, a variável suplementar que mais contribuiu foi TM1 (3 a 5 anos de magistério, de acordo com as fases de HUBERMAN (1989, apud Bolívar, 1990, p. 53), que corresponde a um risco de 0,621. Esse risco representa a confiabilidade entre as respostas DNP4, ENP4 (Difícil abordar gráfico setorial da parte 4), (Difícil abordar tabela de frequência da parte 4), e FNP4, JNP4, (Difícil abordar medidas de tendência central da parte 4), (Difícil abordar polígono de frequência da parte 4), ou seja, observamos, que os professores tem dificuldades em abordar esses temas, pelo fato de não analisarem de forma correta as representações gráficas.

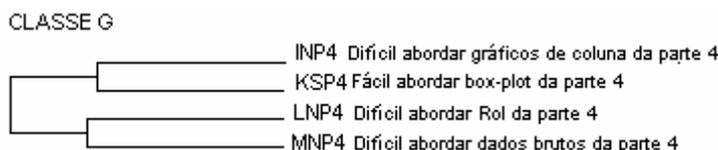


A classe B não apresenta um nó significativo (grau de associação entre as variáveis). Neste bloco, identificado pelo cálculo correto das perguntas 1 e 2 das partes 1 e 2 e análise incorreta do gráfico 1 da parte 3 (número de carros por pessoa), temos:

- os professores que calcularam corretamente a pergunta 2 parte 1 (renda mensal, mediana e primeiro quartil), também calcularam corretamente a pergunta 1 parte 2 (média e desvio padrão), com similaridade igual a 0,5

- os professores que calcularam corretamente a pergunta 2 parte 2 (mediana, primeiro e terceiro quartil), fizeram análise incorreta do gráfico 1 parte 3 (número de carros por pessoa), tem similaridade igual a 0,5

- os dois grupos acima tem o mesmo comportamento, com índice de similaridade de 0,0625, que indica uma dessemelhança entre as variáveis.



A classe G não apresenta um nó significativo (grau de associação entre as variáveis). Neste bloco, identificado por abordar gráfico de colunas, Box-plot, Rol e dados brutos da parte 4, temos:

- os professores que têm dificuldades em abordar gráficos de coluna da parte 4, afirmam não ter dificuldades em abordar Box-plot, com similaridade de 0,5.

- os professores que têm dificuldades em abordar Rol, também têm dificuldades em abordar dados brutos, com similaridade igual a 0,5.

- os dois grupos acima, têm o mesmo comportamento, com índice de similaridade igual a 0,0625, indica dessemelhança entre as variáveis.

Ainda com o auxílio do software C.H.I.C., utilizamos como variável suplementar o tempo de magistério, um modelo proposto por Huberman (1989), que segundo este autor, as fases da carreira dos professores estão centradas fundamentalmente nos anos de experiência docente e não na idade, conforme quadro abaixo.

Anos de Carreira	Fases/Temas
1 – 3	Entrada na Carreira Sobreviver e descobrir
4 – 6	Estabilização. Consolidação de Um repertório pedagógico
7 – 18	Diversificação, “Ativismo → Redelineamento
19 – 30	Serenidade, Distanciamento afetivo Conservadorismo
31 – 4	Ruptura (serena ou amarga)

Quadro 3: Anos de carreira e suas fases

No transcorrer das atividades observamos que a maioria dos erros cometidos foi de ordem analítica, e não de ordem algorítmica. Portanto, os professores são capazes de aplicar fórmulas para cálculo das medidas solicitadas, sem, no entanto, conseguir analisar o significado dos resultados obtidos. As análises feitas pelos professores, tanto na forma escrita quanto na oral, não foram totalmente claras (justificaremos isso logo abaixo). Podemos observar que as análises feitas por esses professores surgiam pela necessidade de justificar uma resposta, pela demanda do pesquisador a análise não foi espontânea, levando-nos a observar que os participantes dessa pesquisa não possuem uma alfabetização estatística de base.

3.3.2 Análise Apresentada Pelos Professores

Primeira Parte

Foi apresentada aos professores uma tabela contendo dados brutos relativos à idade e à renda mensal de 40 pessoas. Bifi (2006). As questões colocadas foram:

- 1) Encontre, nas variáveis *idade* e *renda mensal*, a média e o desvio-padrão. Como você analisaria esses resultados?
- 2) Encontre, nas variáveis *idade* e *renda mensal*, a mediana, o 1º quartil e o 3º quartil. Como você analisaria esses resultados?
- 3) Se você precisasse explicar o *comportamento* da variável *idade* para um cliente, você usaria o item (1) ou o item (2)? Explique por quê.

Em nossa análise, pudemos observar que os professores não apresentaram dificuldades na manipulação dos cálculos para a primeira parte da atividade. Porém, analisando as respostas das atividades, observamos que os professores não souberam justificar o uso dos cálculos. Alguns professores utilizaram, além da calculadora como uma ferramenta fundamental para a resolução do problema, o livro didático, porém, sem questionar o procedimento algorítmico. Os professores alegam que tal procedimento não é relevante, alegando que o importante é o resultado final. Assim, o procedimento algorítmico com uso de calculadora é perfeitamente dominado pelos professores, que, no entanto, não conseguem explicar a própria construção destes procedimentos ou mesmo interpretar os resultados. Neste sentido, podemos supor que eles estejam em um nível cultural de alfabetização, segundo Shamos (1995 apud Morais, 2006) e quanto a mobilização dos conceitos abordados, a contextualizando é feita de maneira simples, sem etapas, sem trabalho preliminar de reconhecimento, sem adaptações (o que corresponde ao nível técnico de mobilização dos conceitos de acordo com Robert (1998, apud Bifi 2006).

Ao final da primeira parte, que envolvia média e desvio-padrão, estabelecemos com os professores um debate sobre os valores encontrados e quais seriam seus significados. Ficou clara a extrema dificuldade em responder e fornecer uma análise crítica. Ou seja, os professores não conseguem explicitar qualquer tipo de relatório, oral ou escrito, que poderia justificar os resultados encontrados partindo, assim, para os cálculos das medidas separatrizes.

Para os cálculos dos quartis, a maior parte dos professores utilizou fórmulas, e apenas cinco professores ordenaram e dividiram a amostra em 4 partes iguais, encontrando o 1º, o 2º e o 3º quartis. Dessa forma, esses cinco professores, utilizaram estratégia de contagem e agrupamento, em um nível bastante intuitivo para o conceito, mas que faz mais apelo ao significado do que ao procedimento na manipulação dos dados. Este procedimento é fundamentado no conceito de quartis. Questionando os professores do porquê do cálculo dos quartis utilizando a ordenação dos dados, me foi respondido que é por analogia à geometria plana e por terem aproximadamente o mesmo número de sujeitos o que mostra que a apreensão do conceito não é suficiente para a construção de significados, embora a estratégia usada tenha sido conceitual.

Esses resultados nos remetem a pensar que os professores acreditam que as distribuições sejam simétricas, já que os livros didáticos trabalham, na maioria dos seus exemplos, com distribuições simétricas. Acreditamos que a forma de abordagem dos livros didáticos utilizados pelos professores do ensino médio não favorece a superação da dificuldade dos professores e alunos em diferenciar os tipos de distribuições.

Observamos que esses livros didáticos, agora adotados em todo o ensino médio oficial, induzem professor e aluno a ter um nível somente cultural, segundo Shamos (1995 apud Moraes, 2006). As pesquisas que tratam da análise de livros indicam que as situações-problema, apresentadas nos livros didáticos, não são preparadas para uma mobilização de conhecimentos, ou seja, são situações-problema que apresentam um só tipo de resolução, como, por exemplo: “calcule a média, calcule a mediana, calcule o desvio-padrão”, sem que essas atividades exijam dos alunos uma análise crítica dos resultados obtidos. Para os professores foram dadas oportunidades para discutirem os valores encontrados na primeira parte do questionário, e sobre seus significados. Os professores

justificaram que os resultados eram suficientes. Alguns acreditam que esse tipo de análise, associada à média e desvio-padrão, serve para mostrar, basicamente, que a amplitude das idades é alta, tal como a renda. Só isso. “de pouca utilidade *prática*”.

Pelo que foi justificado, podemos observar que o conceito de média não está totalmente claro. Acreditamos que os conceitos de amplitude da amostra, tanto para variável *idade* como para variável *renda*, tenha gerado uma confusão. Pelo que percebemos, os professores buscam uma análise da variabilidade, tentando mostrar, por meio da amplitude, o comportamento da amostra.

Partindo para os cálculos das medidas separatrizes, os professores não mostraram dificuldades no desenvolvimento dos procedimentos, porém há uma extrema dificuldade em redigir o texto que justificasse tais cálculos. Pelo texto redigido, podemos perceber que eles tentam criar uma explicação para os valores simplesmente por demanda do pesquisador, uma vez que, para eles, os próprios valores seriam auto-explicativos.

Percebemos que alguns professores tentam fazer uma comparação entre as variáveis, fazendo uma interpretação de variável bi-dimensional, porém não se preocuparam em verificar que o valor das medidas separatrizes são medidas unidimensionais e, sendo assim, não cabe uma análise comparativa. Tal comparação pode ser um efeito da forma de apresentação do banco de dados. Podemos observar que este grupo de professores está no primeiro nível que abordamos no começo do capítulo, pois percebemos que estes professores apresentam uma alfabetização em estatística, já que identifica o texto que está sendo lido, conhece os cálculos a serem abordados da Estatística e da Matemática, até porque, nessa segunda questão, não há cálculos matemáticos que exijam alto conhecimento da ciência em questão. Porém a análise feita pelos professores, indicou uma dificuldade em redigir um texto que mostre uma análise crítica correta dos resultados, conforme mostramos na nossa análise *a priori*. Quanto aos níveis de alfabetização, esses professores encontram-se em um nível de conhecimento que Shamos (1995, apud Morais (2006)) classifica como nível cultural. Percebemos, que a maior segurança nas respostas dadas, são aquelas que foram utilizadas os algoritmos, porém não souberam relacionar com coerência todos os resultados encontrados.

No entanto, algumas justificativas apresentadas, também foram efeitos do contrato didático, ou seja, a forte necessidade de apresentar uma resposta, pois o há uma pergunta a ser respondida.

Para a primeira questão, depois dos cálculos de média e desvio-padrão, há uma tentativa de comparar o comportamento das duas variáveis, conforme podemos perceber na afirmação feita por eles:

“A variabilidade da variável ‘renda mensal’ é superior à da variável ‘idade’. No entanto a média da variável ‘idade’ está mais próxima de um dos extremos da amostra.”

De acordo com a redação dos professores, não identificamos uma ligação de idéias ou mesmo de conceitos que justifiquem essa afirmação. Quando escreveram que a variabilidade da variável “renda mensal” é superior à da “idade”, espera-se uma justificativa do porquê desta observação, e o que ela acarretaria na análise crítica, para a variação dos dados em torno das respectivas médias conforme sugerimos em nossa análise *a priori*; fato que não ocorre. Ao escrever que a média da variável “idade” está próxima de um dos extremos da amostra, eles não se referem à não simetria na distribuição dos dados, o que poderia levá-los a uma análise mais global dos resultados, conforme o segundo nível que estabelecemos para o pensamento estatístico.

Estas medidas coincidem quando a distribuição é simétrica, o que reforça nossa hipótese de efeito de contrato na forma de apresentação dos problemas de Estatística. Vale aqui ressaltar que as situações-problemas dos livros didáticos estão voltadas, em sua maioria, a distribuições simétricas, deixando ao professor e ao aluno poucas opções de diferentes tipos de distribuições, para que ele possa ter uma melhor noção do comportamento da amostra (Novaes, 2004)

Percebermos que há professores que evoluem na entrevista, passando de um nível cultural para funcional, pois percebem a concentração em um dos extremos da amostra, sem, no entanto, explorar esse fato. “Um saber é dito *funcional*, quando somos capazes além de ler, interpretarmos informações contidas em dados que estão representados em tabelas e ou gráficos, ou mesmo

organizá-los nessas representações, identificando e considerando a variação na análise” (Shamos, 1995,apud Morais, 2006, p.24).

Na segunda questão, que tratou das medidas separatrizes, os professores não apresentaram dificuldades nos cálculos algorítmicos, porém, na análise crítica do resultado encontrado, os professores relatam que:

A diferença observada entre os valores dos quartis, para a variável idade, é bastante desigual, estando mais próximos aos valores do 1º e 2º quartis do que do 3º. Estes dados servem para comprovar a predominância de casos com baixa idade. Já na variável “renda média”, os valores são mais próximos deixando assim a mediana com distâncias semelhantes de ambos os quartis.

Os valores encontrados foram os seguintes: para a variável idade, o 1º quartil igual 29, o 2º quartil igual 30 e 3º quartil igual 40; e para a renda mensal, os valores para o 1º, o 2º e o 3º quartis foram, respectivamente, 566, 890, 1392. Pode-se observar, que alguns professores possuem uma percepção de não simetria, pois, em sua fala. *“Observamos que a variável idade, é bastante desigual, estando mais próximos aos valores do 1º e 2º quartis do que o 3º”, “Porém a variável, “renda mensal”, os valores são mais próximos, deixando assim a mediana com distâncias semelhantes de ambos os quartis”* acreditamos que os professores estão querendo dizer que a distribuição da renda é mais simétrica do que a variável idade.

Segunda parte da atividade

Analisaremos a segunda parte da atividade, que é apresentada por meio de duas distribuições de frequências representadas por tabelas (anexo I), sendo que a primeira tabela está representada sem intervalos de classe e a segunda tabela com intervalo de classes. As questões feitas, nessa segunda parte, são análogas às da primeira, e o objetivo de apresentar os dados na forma de tabelas foi para que os professores investigados pudessem manipular as informações em diferentes formas de representação.

Nessa segunda parte, os professores apresentaram apenas dificuldades para encontrar um método de resolução que justificasse os resultados encontrados, pois alguns professores alegaram não ter encontrado uma forma de calcular as medidas solicitadas por meio de tabelas com ou sem intervalos de classe, tal como observado com alunos de um curso de administração, por Bifi (2006), e a estratégia de resolução, também foi observada por este autor.

Pois, depois da intervenção do pesquisador, houve sucesso no cálculo dos valores, visto que os professores transformaram os dados agrupados em dados brutos, podendo, assim, inferir sobre os dados apresentados nos moldes de resolução da primeira parte, realizando com sucesso todos os cálculos pedidos. Queremos lembrar que toda esta análise encontra-se dentro do nível operacional de nossa categorização, que trata, dentre outros fatores, do contexto matemático. Percebemos que o nível de conhecimento desses professores, na segunda parte da atividade, segundo Shamos (1995, apud Morais, 2006), pode ser classificado como *cultural*.

Nessa primeira etapa, no que diz respeito à variável discreta, os professores conseguem fazer a transformação necessária para a resolução do problema proposto, implicando, assim, um nível funcional de conhecimento e dentro da categoria operacional, ou seja, o algorítmico. O mesmo ocorre na segunda etapa, os professores mantêm o mesmo padrão de raciocínio para estabelecer uma estratégia de resolução, ou seja, determinam o ponto médio de cada classe, justificando que cada classe está em unidade intervalar, necessitando do cálculo do ponto médio de cada classe, alcançando dessa forma, os valores das média, mediana e quartis, com o auxílio de algoritmo.

Observamos que os professores, não sabem justificar o que explicaria o comportamento da amostra. Destacamos o fato de os professores não tentarem analisar os dados da tabela utilizando as separatrizes, ainda que superficialmente. Segundo Gal (2002), em suas bases de alfabetização na Estatística, não há regras ou critérios para uma análise crítica de dados estatísticos.

Segundo Bifi (2006, p.84),

“o autor não defende que suas bases de alfabetização sigam uma ordem e que estas devam necessariamente passar pelos cálculos estatísticos ou matemáticos. Ao analisar uma tabela ou até mesmo um gráfico estatístico, pode o leitor usar seu senso crítico e intuitivo, e perceber, por uma análise visual, o que esses dados podem estar dizendo.”

Os professores poderiam tentar dar resposta conclusiva por meio dessa análise visual, ou ,ainda, analisar por meio de uma apreensão perceptiva dos dados representados graficamente.

Porém a conclusão comum dos professores foi: “podemos observar que a variabilidade da amostra é elevada em ambas. Nota-se que o perfil das amostras é de pessoas com 2 carros ou mais, e que passam 3,67 horas para mais no trânsito”. Eles associam, assim, variabilidade com o valor modal, o que é uma estratégia errônea.

Os resultados encontrados foram bem próximos aos que apresentamos em nossa análise *a priori*, porém o que podemos perceber é que o texto elaborado pelos professores foi exatamente um relato dos valores encontrados, ou seja, o que percebemos é apenas uma transcrição da linguagem matemática encontrada nos resultados para a linguagem coloquial, sem, contudo, fazer uma análise do significado desses valores. Assim, o efeito do contrato didático usual está novamente presente, os professores precisavam, naquele momento, de uma resposta à pergunta feita.

O que nos chamou a atenção foi o fato de que os professores, ao encontrarem como resposta a média igual a 2,73 para a primeira tabela, arredondou-a para 2 carros no relatório final, arredondamento que não deveria ter ocorrido pelo fato da média ser o valor ao redor do qual os valores assumidos pela variável estudada se distribuem. O mesmo não aconteceu com o tempo em horas. O valor encontrado foi a média de 3,67 horas. Assim, fica claro que esses professores perceberam as variáveis contínua e discreta, porém nos níveis mais elementares de alfabetização estatística. Segundo Gal (2002) ele faz referência não apenas ao que se pode escrever sobre relatórios estatísticos, mas, também, se o contexto está coerente com o que foi calculado e com o que está sendo redigido (alfabetização funcional). Não basta apenas transcrever o que se

encontra como valores estatísticos, sem fazer uma análise crítica. Devemos, além de relatar o que foi encontrado, ter uma análise crítica dos resultados e, assim, percebe-se o que foi encontrado como resultado numérico faz algum sentido no relatório conclusivo. Observamos que os professores não se preocuparam com essa análise, já que ela apresenta, em sua conclusão, uma média de horas de 3,67, fazendo-nos inferir um nível cultural de alfabetização.

O que esperávamos dos professores era exatamente uma análise semelhante ou próxima às que explicitamos em nossa análise *a priori*, no capítulo que tratou do assunto. Podemos aqui também inferir que, por acreditarem que todas as análises feitas pelos professores tratavam de distribuições simétricas e, sendo assim, admitindo normalidade na distribuição, sem uma justificativa para tal pensamento, a média, a mediana e a moda poderiam ter sempre o mesmo valor (novamente efeito de contrato: as opções didáticas dos livros e da maioria dos professores têm sido pela apresentação quase exclusiva de exemplos que se limitam às distribuições simétricas).

Para a última parte da nossa atividade, os professores fizeram uma análise separada de cada gráfico, ou seja, fizeram uma leitura dos valores apresentados nos gráfico, sem, contudo, os analisar. Dessa forma conseguiram visualizar graficamente as medidas de tendência central e as medidas separatrizes, já calculadas, mais uma vez os professores se vêem na obrigatoriedade de oferecer uma resposta, sem contudo relacioná-la. (nível cultural)

CONSIDERAÇÕES FINAIS

Sabemos que a Estatística tem se destacado há várias décadas e vem sendo um dos principais temas de pesquisa em Educação Matemática. As dificuldades apresentadas e diagnosticadas em pesquisas que abordam o tema, tanto no ensino básico como no ensino médio, tornaram-se molas propulsoras que incentivam pesquisadores a investigar os possíveis fatores que influenciam no processo de ensino e aprendizagem.

Nossa preocupação e incentivo para a pesquisa surgiram diante da minha experiência profissional no nível Médio, quando alunos desse nível, de escolas públicas, questionavam-me sobre a aplicabilidade da Estatística na futura vida profissional. Pude observar que, no caso destes alunos, o ensino de Estatística está focado somente na prática de cálculos, no exercitar de algoritmos, sem qualquer preocupação com a atribuição de significados aos conceitos explorados, de forma a permitir aos alunos uma análise crítica de seus resultados.

A relevância deste trabalho, além de estar atrelada à necessidade de buscar novas metodologias que auxiliem profissionais e pesquisadores da Educação Matemática em sua interface com a Educação Estatística é, ainda, um alerta de que o ensino da disciplina Estatística não pode ser encarado como somente um ramo da Matemática, mas também mostrar que a Estatística tem particularidades, próprias da disciplina, e que é tão importante quanto à própria Matemática no processo de formação do cidadão de qualquer área profissional, tornando-o alfabetizado nesse componente.

Antes de apresentar os resultados finais do nosso trabalho, faremos um resumo de como foi conduzida a nossa pesquisa. A Estrutura do trabalho foi apresentada em nosso capítulo de Introdução. O objetivo deste trabalho foi apresentado em nossa problemática, e visava estudar como o professor de Matemática do Ensino Médio, conceitualizam as idéias básicas de Estatística, envolvendo variabilidade. Tal estudo foi feito à luz dos trabalhos desenvolvidos por dois autores: Shamos (1995, apud Morais, 2006), que trata dos níveis de letramento (científico, funcional e cultural), foi nosso pano de fundo sob um ponto de vista didático; e Gal (2002), que trata da Alfabetização Estatística, constituiu nosso quadro teórico Estatístico.

No intuito de responder nossas questões, ou seja, de que forma os professores de Matemática do Ensino Médio, de escolas públicas de São Paulo em exercício, mobilizam os conhecimentos estatísticos quando confrontados com problemas que envolvem Medidas de Tendência Central, Variabilidade e representação Gráfica.

Aplicamos um questionário para 29 professores de matemática do ensino médio que estão atualmente em exercício, em escolas públicas do Estado de São Paulo (SP). A escolha desses professores não obedeceu a uma amostragem aleatória, uma vez que buscamos a participação voluntária dos docentes, com os quais já tínhamos estabelecido vínculos profissionais, dessa forma há uma liberdade na busca do conhecimento individual de cada professor com relação a análise exploratória de dados, sem que haja constrangimento na resolução da atividades propostos. Esse primeiro contexto permitiu estender o convite a outros profissionais que lecionam em outras escolas.

Vale ressaltar que os professores convidados recebiam uma situação problema na forma de atividade diagnóstica, dividida em três etapas para serem resolvidas, individualmente, podendo ou não consultar seu livro didático, caso seja necessário relembrar alguma fórmula. Com um tempo estimado de 150 minutos (3horas/aula), ao término, todas as atividades foram recolhidas pelo pesquisador. Esta atividade diagnóstica baseada em Bifi (2006) nos permitiu diagnosticar os níveis de letramento: científico, funcional e cultural de acordo com Shamos (1995, apud Morais, 2006)

Para que pudéssemos analisar os resultados da pesquisa, dividimos as bases elencadas por Gal (2002) em dois níveis. O primeiro focou a Alfabetização, a Estatística e a Matemática, e o segundo focou a análise crítica e global. E, sendo assim, dentro desses níveis, investigamos os níveis de letramento: científico, funcional e cultural de acordo com Shamos (1995, apud Morais, 2006)

Pudemos perceber que, não foi possível identificar invariantes que justificassem possíveis dificuldades dos professores no âmbito dos cálculos algébricos, ou seja, os professores não apresentaram dificuldades em calcular as medidas pedidas nas duas primeiras etapas da atividade. Porém, pudemos perceber que os professores não conseguiam justificar ou dar significado aos cálculos que foram feitos, apesar de estarem todos corretos. Os conceitos mobilizados nos cálculos não tiveram seu significado explicitado ou mesmo

justificado pelos professores, levando-nos a inferir que este conhecimento, se existente, permaneceu implícito. Mesmo com questionamentos durante a atividade que visavam proporcionar condições para que os professores exteriorizassem o significado por eles atribuído aos valores calculados, os professores permaneceram ligados somente aos valores numéricos, acreditando que estes eram auto-explicativos. Sendo assim, inferimos que os professores classificam-se em nível, segundo Shamos (1995, apud Morais, 2006), cultural.

Percebemos, em alguns momentos, que os professores investigados realizaram análises equivocadas de alguns conceitos, como, por exemplo, confundir média e mediana. Isso foi diagnosticado pelo fato dos professores atribuírem, para qualquer banco de dados, a noção de simetria. Para eles, toda distribuição é simétrica. Assim percebemos que, em toda a atividade, os professores sentiram a necessidade de modelar a amostra para uma distribuição simétrica.

É comum encontrarmos esse tipo de equívoco quando professores adotam livros didáticos, nos quais em sua maioria, são trabalhadas apenas com distribuição simétrica e, assim, os valores das medidas de tendência central e medidas separatrizes coincidem. Faz-se necessário que o professor do ensino médio perceba que tipo de distribuição está sendo trabalhado para, então perceberem a diferença nas medidas encontradas e, conseqüentemente, dizer quais são as medidas que melhor representam a amostra. Isso nos leva a inferir que o trabalho somente com distribuições simétricas leva o professor a atingir o nível cultural, segundo Shamos (1995, apud Morais, 2006), o que não permite a esse professor a oportunidade de questionar sobre outros tipos de distribuição e, então, quem sabe torná-lo um questionador para que passe a um nível cultural.

Durante a aplicação das atividades, percebemos, por parte dos professores, a obrigatoriedade de, em primeiro lugar, calcular-se a média e o desvio-padrão, sem se dar conta de verificar, até por uma análise superficial da amostra, se realmente haveria a necessidade desses valores. Por exemplo, a última atividade que apresentamos para os professores não exigia iniciar os cálculos pelas medidas de tendência central, e sim que fizessem uma análise visual do gráfico e, depois, tentassem explicar o comportamento dos dados informados pelos dois gráficos. Chegamos a inferir, nesse caso, que a resolução

de atividades de estatística passa obrigatoriamente por algum processo algébrico. Fica evidente, nesse caso, a necessidade de que a alfabetização estatística deva ser contemplada na sua totalidade nas propostas de Gal (2002): a análise crítica e global dos dados coletados. Entendemos que esta análise não precisa ser necessariamente depois dos dados codificados, mas sim uma análise *a priori* da codificação dos dados, procurando um melhor caminho a seguir.

Nossa idéia foi, dessa forma, a de propor uma reflexão sobre o ensino de Estatística para que este trabalho tenha sentido para professores de Matemática do Ensino Médio, de escolas públicas de São Paulo em exercício, ter condições de continuar aprendendo e de encontrar respostas para os problemas que aparecerem durante sua trajetória profissional.

O caminho que escolhemos para o estudo neste trabalho, juntamente com o quadro teórico que utilizamos, nos proporcionaram novas perspectivas para o ensino de Estatística, uma das conseqüências é a necessidade de uma seqüência didática que permita o professor vivenciar todas as fases necessárias para a construção de um conhecimento estatístico, e conseqüentemente estudo dos resultados obtidos de uma situação-problema, ou seja, análise exploratória dos dados.

REFERÊNCIAS

BARBETA, A. P Estatística Aplicada as Ciências Sociais – 5ª Edição – Editora UFSC 2003.

BATANERO, C Didáctica de la Matemática – Departamento de Didáctica de la Matemática - Universidade de Granada. 2001.

BATANERO, C. ESTEPA, A; GODINO, J. D. *Sus Possibilidades en la Enseñaza Secundária.* (suma nº 9, 1991)

BIFI, Ricardo C. *Estatística em um Curso de Administração de Empresas: Mobilização dos Conceitos Estatísticos de Base. Dissertação de Mestrado em Educação Matemática – PUC/SP - 2006.*

BOLIVAR, A. (org). Profissão Professor: O Itinerário... Bauru, EDUSC. 2002
Carmen Batanero y Juan D. Godino, 2001 ANÁLISIS DE DATOS Y SU DIDÁCTICA Departamento de Didáctica de la Matemática

CURCIO, F. R (1989) *Developing graph comprehension.* Reston. VA: N.C.T.M.

GAL, I (2002) *Adult's Statistical Literacy: Meanings, Components, Responsibilities* - Appeared in: *Internacional Statistical Review*, 2002, 70 (1), 1-25.

GIOVANNI, J.R., & BONJORNO, J. R. 2º Grau Matemática 3 – Editora – FTD.

BOLIVAR, A. (org): *Fases da Carreira dos Professores* HUBERMAN (1990),

LINS, Rômulo C. (coord) *A formação Pedagógica nas disciplinas matemáticas e a formação matemática nas disciplinas pedagógicas, em cursos de licenciatura em Matemática. VII EPEM, São Paulo: SBEM – SP, p. 48 – 50, 2004*

LOPES, Celi A. E; *A Probabilidade e a Estatística no Currículo de Matemática no Ensino Fundamental Brasileiro*, 1998.

LOPES, C. (2004): *Literária Estatística e INAF 2002.* In Fonseca M.C (org) *Letramento no Brasil – habilidades matemáticas.* Ed Global. P.187-197

LOPES, C.A. Espasadin : *O Conhecimento Profissional dos Professores e suas Relações com Estatística e Probabilidade na Educação Infantil. Tese de Doutorado UNICAMP – 2006.*

MACHADO, Nilson J. *Educação Projetos e valores.* São Paulo: Escrituras Editoras 2000 (Coleção Ensaio Transversais)

MIRADOR, *Enciclopédia Internacional*, 8V. São Paulo: Britânia, 1989.

MOORE (1997). (apud Wild, 1999) *New Pedagogy and new content: the case of statistics*. *Internacional Statistical Review*, 65(2), 123-137.

MORAIS, TULA, M. R. *Um Estudo sobre o Pensamento Estatístico: Componentes e Habilidades*. Dissertação de Mestrado em Educação Matemática – PUC/SP – 2006.

PARKER, M., & LEINHARDT, G. (1995). *Percent: A privileged proportion*. *Review of Educational Research*, 65(4), 421-481.

VIEIRA, S. *Bioestatística: tópicos avançados*. 1ª Edição. Editora Campos – 2003.

ISBN 85-352-0967-0

SILVA, C. B. *Atitudes em relação à Estatística: Um Estudo com alunos de graduação*. Dissertação de Mestrado. UNICAMP/SP, 2000

SNEE, R. D. (1988). (apud Wild e Pfannkuch, 1999) *Mathematics is only one tool that Statistician use*. *The College Mathematics Journal*, 19, 30-32.

TOLEDO, Luciano G; OVALLE, Isidoro I. *Estatística Básica 2ª Edição*. Editora Atlas. 1995. ISBN 85-224-1791-1

TRIOLA, Mario F. *Introdução à Estatística, 7ª edição, Livros Técnicos e Científicos*. 1999. ISBN 85-216-1154-4.

WILD, C; PFANNKUCH, M. (1998) *Que és el ensamiento Estadístico*. Publicado no *Jornal Hipóteses Alternativa*. Vol. 56 – nº 2 – Septiembre – 2005.

Anexo I

Instrumento aplicado a partir de adaptações feitas ao utilizado por Bifi (2006)

Tempo de Magistério _____

Primeira parte

Uma Empresa de cartões de crédito solicitou uma análise do banco de dados abaixo, construído a partir das respostas a um questionário que buscava levantar a idade e a renda mensal de 40 pessoas.

Idade	Renda Mensal						
23	300	29	600	30	890	40	1400
25	380	29	630	31	1000	40	1400
25	387	30	700	31	1000	41	1420
25	400	30	700	31	1160	43	1420
27	400	30	760	31	1180	43	1500
28	400	30	770	32	1200	45	1600
28	490	30	800	34	1200	46	1600
28	500	30	850	37	1200	48	1770
28	540	30	860	39	1340	53	1770
29	554	30	890	40	1370	65	1800

Questões:

1) Encontre das variáveis, idade e renda mensal, a média e o desvio-padrão. Como você analisaria esses resultados?

2) Encontre das variáveis, idade e renda mensal, a mediana, o 1º quartil e o 3º quartil. Como você analisaria esses resultados?

3) Se você precisasse explicar o “comportamento” da variável “idade” para um cliente, você usaria o item (1) ou o item (2)? Explique por quê.

Segunda parte

De acordo com as tabelas abaixo, Responda às questões:

Números de carros por pessoa entrevistada

<i>Números de carros</i>	<i>Números de pessoas</i>
1	10
2	25
3	15
4	5
Total	55

Número de horas no trânsito por pessoa

<i>Tempo no trânsito</i>	<i>Número de pessoas</i>
0 2	10
2 4	20
4 6	30
6 8	15
Total	75

Questões

1) Determine nas tabelas acima, a média e o desvio-padrão. Como você analisaria esses resultados?

2) Determine nas tabelas acima a mediana, o 1º quartil e o 3º quartil. Como você analisaria esses resultados?

3) Se você precisasse descrever os dados “Número de carros” e “Tempo no trânsito” para um cliente, você usaria o item (1) ou o item (2)? Explique por quê.

Terceira parte

Observe os gráficos abaixo. Responda a pergunta: “Se você precisasse descrever esses dados para um cliente, como você analisaria?”.

Gráfico 1

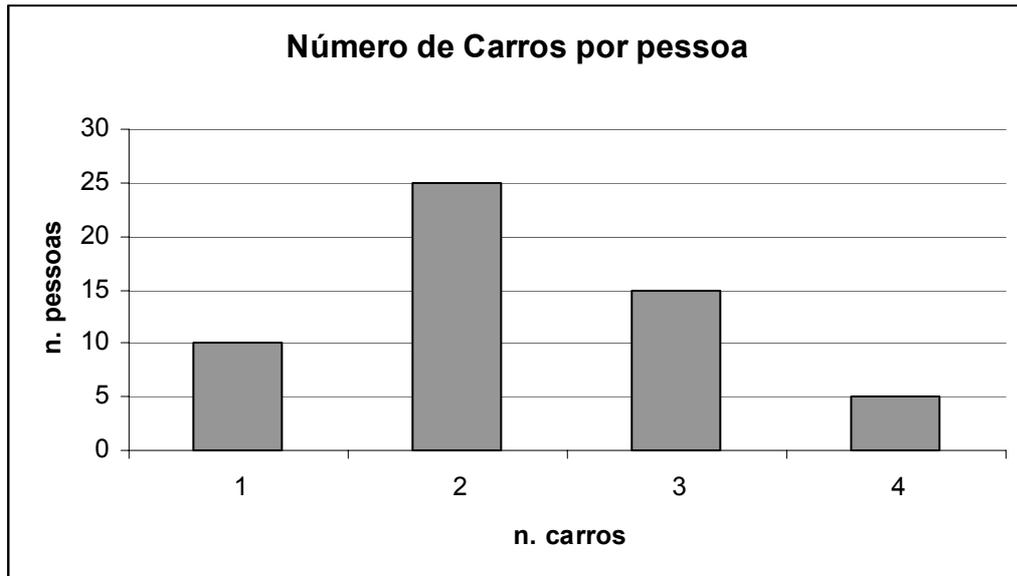
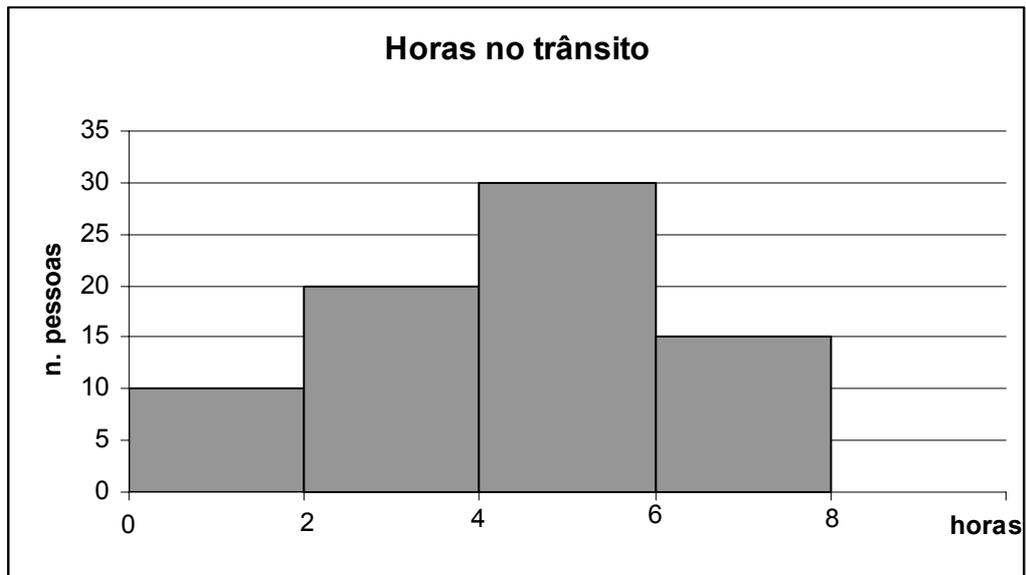


Gráfico 2



Parte 4

Assinale os itens mais difíceis de serem abordados com os alunos.

- a) () Média
- b) () Moda
- c) () Mediana
- d) () Gráfico Setorial
- e) () Tabela de Freqüência
- f) () Medidas de Tendência Central
- g) () Classificação da Moda: Amodal, Unimodal, Bimodal etc.
- h) () Gráfico de Barras
- i) () Gráfico de Colunas
- j) () Polígono de Freqüência
- k) () Box Plot
- l) () Rol
- m) () Dados Brutos
- n) () Medidas Separatrizes
- o) () Medidas de dispersão
- p) () Decil, percentil, quartil e quintil

Assinale os itens mais fáceis de serem abordados com os alunos

- a) () Média
- b) () Moda
- c) () Mediana
- d) () Gráfico Setorial
- e) () Tabela de Freqüência
- f) () Medidas de Tendência Central
- g) () Classificação da Moda: Amodal, Unimodal, Bimodal etc.
- h) () Gráfico de Barras
- i) () Gráfico de Colunas
- j) () Polígono de Freqüência
- k) () Box Plot
- l) () Rol
- m) () Dados Brutos
- n) () Medidas Separatrizes
- o) () Medidas de dispersão
- p) () Decil, percentil, quartil e quintil

Apêndice B

Índice de Similaridade

Classificação ao nível: 1 : (DSP4 ESP4) similaridade : 1
Classificação ao nível: 2 : (AS2P1 AS3P1) similaridade : 0.999999
Classificação ao nível: 3 : (AS1P2 AS2P2) similaridade : 0.999999
Classificação ao nível: 4 : ((AS2P1 AS3P1) (AS1P2 AS2P2)) similaridade : 0.999997
Classificação ao nível: 5 : (BSP4 CSP4) similaridade : 0.999804
Classificação ao nível: 6 : (((AS2P1 AS3P1) (AS1P2 AS2P2)) CNP4) similaridade : 0.989419
Classificação ao nível: 7 : ((DSP4 ESP4) JSP4) similaridade : 0.920425
Classificação ao nível: 8 : ((BSP4 CSP4) GSP4) similaridade : 0.839073
Classificação ao nível: 9 : (FSP4 GNP4) similaridade : 0.786422
Classificação ao nível: 10 : (AN2P1 AN3P1) similaridade : 0.711266
Classificação ao nível: 11 : (AN1P2 AN2P2) similaridade : 0.711266
Classificação ao nível: 12 : (FNP4 JNP4) similaridade : 0.625799
Classificação ao nível: 13 : (DNP4 ENP4) similaridade : 0.573658
Classificação ao nível: 14 : (((AS2P1 AS3P1) (AS1P2 AS2P2)) CNP4) ((DSP4 ESP4) JSP4))
similaridade : 0.536924
Classificação ao nível: 15 : (CS1P1 AN1P1) similaridade : 0.5
Classificação ao nível: 16 : (CS2P1 CS1P2) similaridade : 0.5
Classificação ao nível: 17 : (CS2P2 ANG1P3) similaridade : 0.5
Classificação ao nível: 18 : (ANG2P3 ANP4) similaridade : 0.5
Classificação ao nível: 19 : (BNP4 HNP4) similaridade : 0.5
Classificação ao nível: 20 : (INP4 KSP4) similaridade : 0.5
Classificação ao nível: 21 : (LNP4 MNP4) similaridade : 0.5
Classificação ao nível: 22 : (NSP4 OSP4) similaridade : 0.5
Classificação ao nível: 23 : (((BSP4 CSP4) GSP4) (FSP4 GNP4)) similaridade : 0.301635
Classificação ao nível: 24 : ((AN2P1 AN3P1) (AN1P2 AN2P2)) similaridade : 0.255934
Classificação ao nível: 25 : ((CS1P1 AN1P1) PSP4) similaridade : 0.25
Classificação ao nível: 26 : ((DNP4 ENP4) (FNP4 JNP4)) similaridade : 0.102173
Classificação ao nível: 27 : ((CS2P1 CS1P2) (CS2P2 ANG1P3)) similaridade : 0.0625
Classificação ao nível: 28 : ((ANG2P3 ANP4) (BNP4 HNP4)) similaridade : 0.0625
Classificação ao nível: 29 : ((INP4 KSP4) (LNP4 MNP4)) similaridade : 0.0625

Classificação ao nível: 30 : (((((AS2P1 AS3P1) (AS1P2 AS2P2)) CNP4) ((DSP4 ESP4) JSP4))
(((BSP4 CSP4) GSP4) (FSP4 GNP4))) similaridade : 0.029922

Classificação ao nível: 31 : (((CS1P1 AN1P1) PSP4) (NSP4 OSP4)) similaridade : 0.015625

O nó mais significativo está no nível : 1

Nós significativos

ao nível : 1

ao nível : 4

ao nível : 6

ao nível : 19

ao nível : 22

ao nível : 24

ao nível : 26

ao nível : 30

Apêndice C

Tipicalidade

Tipicalidade à classe : DSP4,ESP4 (1)

A variável típica a esta classe é TM4 com um risco de : 0.24

Tipicalidade à classe : AS2P1,AS3P1 (2)

A variável típica a esta classe é TM3 com um risco de : 0.34

Tipicalidade à classe : AS1P2,AS2P2 (3)

A variável típica a esta classe é TM3 com um risco de : 0.34

Tipicalidade à classe : AS2P1,AS3P1,AS1P2,AS2P2 (2,3,4)

A variável típica a esta classe é TM3 com um risco de : 0.343

Tipicalidade à classe : BSP4,CSP4 (5)

A variável típica a esta classe é TM4 com um risco de : 0.24

Tipicalidade à classe : AS2P1,AS3P1,AS1P2,AS2P2,CNP4 (2,3,4,6)

A variável típica a esta classe é TM3 com um risco de : 0.34

Tipicalidade à classe : DSP4,ESP4,JSP4 (1,7)

A variável típica a esta classe é TM4 com um risco de : 0.245

Tipicalidade à classe : BSP4,CSP4,GSP4 (5,8)

A variável típica a esta classe é TM4 com um risco de : 0.245

Tipicalidade à classe : FSP4,GNP4 (9)

A variável típica a esta classe é TM3 com um risco de : 0.401

Tipicalidade à classe : AN2P1,AN3P1 (10)

A variável típica a esta classe é TM4 com um risco de : 0.41

Tipicalidade à classe : AN1P2,AN2P2 (11)

A variável típica a esta classe é TM4 com um risco de : 0.417

Tipicalidade à classe : FNP4,JNP4 (12)

A variável típica a esta classe é TM1 com um risco de : 0.621

Tipicalidade à classe : DNP4,ENP4 (13)

A variável típica a esta classe é TM1 com um risco de : 1

Tipicalidade à classe : AS2P1,AS3P1,AS1P2,AS2P2,CNP4,DSP4,ESP4,JSP4 (1,2,3,4,6,7,14)

A variável típica a esta classe é TM4 com um risco de : 0.245

Tipicalidade à classe : CS1P1,AN1P1 (15)

A variável típica a esta classe é TM4 com um risco de : 0.755

Tipicalidade à classe : CS2P1,CS1P2 (16)

A variável típica a esta classe é TM4 com um risco de : 0.755

Tipicalidade à classe : CS2P2,ANG1P3 (17)

A variável típica a esta classe é TM4 com um risco de : 0.755

Tipicalidade à classe : ANG2P3,ANP4 (18)

A variável típica a esta classe é TM4 com um risco de : 0.755

Tipicalidade à classe : BNP4,HNP4 (19)

A variável típica a esta classe é TM4 com um risco de : 0.75

Tipicalidade à classe : INP4,KSP4 (20)

A variável típica a esta classe é TM4 com um risco de : 0.755

Tipicalidade à classe : LNP4,MNP4 (21)

A variável típica a esta classe é TM4 com um risco de : 0.755

Tipicalidade à classe : NSP4,OSP4 (22)

A variável típica a esta classe é TM4 com um risco de : 0.755

Tipicalidade à classe : BSP4,CSP4,GSP4,FSP4,GNP4 (5,8,9,23)

A variável típica a esta classe é TM4 com um risco de : 0.245

Tipicalidade à classe : AN2P1,AN3P1,AN1P2,AN2P2 (10,11,24)

A variável típica a esta classe é TM4 com um risco de : 0.417

Tipicalidade à classe : CS1P1,AN1P1,PSP4 (15,25)

A variável típica a esta classe é TM4 com um risco de : 0.755

Tipicalidade à classe : DNP4,ENP4,FNP4,JNP4 (12,13,26)

A variável típica a esta classe é TM1 com um risco de : 0.621

Tipicalidade à classe : CS2P1,CS1P2,CS2P2,ANG1P3 (16,17,27)

A variável típica a esta classe é TM4 com um risco de : 0.755

Tipicalidade à classe : ANG2P3,ANP4,BNP4,HNP4 (18,19,28)

A variável típica a esta classe é TM4 com um risco de : 0.755

Tipicalidade à classe : INP4,KSP4,LNP4,MNP4 (20,21,29)

A variável típica a esta classe é TM4 com um risco de : 0.755

Tipicalidade à classe :

AS2P1,AS3P1,AS1P2,AS2P2,CNP4,DSP4,ESP4,JSP4,BSP4,CSP4,GSP4,FSP4,GNP4 (1,2,3,4,5,6,7,8,9,14,23,30)

A variável típica a esta classe é TM4 com um risco de : 0.245

Tipicalidade à classe : CS1P1,AN1P1,PSP4,NSP4,OSP4 (15,22,25,31)

A variável típica a esta classe é TM4 com um risco de : 0.755

Apêndice D

Contribuição

Contribuição à classe : DSP4,ESP4 (1)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.24

Contribuição à classe : AS2P1,AS3P1 (2)

A variável que contribui mais a esta classe é TM3 com um risco de : 0.343

Contribuição à classe : AS1P2,AS2P2 (3)

A variável que contribui mais a esta classe é TM3 com um risco de : 0.343

Contribuição à classe : AS2P1,AS3P1,AS1P2,AS2P2 (2,3,4)

A variável que contribui mais a esta classe é TM3 com um risco de : 0.343

Contribuição à classe : BSP4,CSP4 (5)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.245

Contribuição à classe : AS2P1,AS3P1,AS1P2,AS2P2,CNP4 (2,3,4,6)

Contribuição à classe : DSP4,ESP4,JSP4 (1,7)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.245

Contribuição à classe : BSP4,CSP4,GSP4 (5,8)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.24

Contribuição à classe : FSP4,GNP4 (9)

A variável que contribui mais a esta classe é TM3 com um risco de : 0.401

Contribuição à classe : AN2P1,AN3P1 (10)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.417

Contribuição à classe : AN1P2,AN2P2 (11)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.417

Contribuição à classe : FNP4,JNP4 (12)

A variável que contribui mais a esta classe é TM1 com um risco de : 0.621

Contribuição à classe : DNP4,ENP4 (13)

A variável que contribui mais a esta classe é TM3 com um risco de : 0.496

Contribuição à classe : AS2P1,AS3P1,AS1P2,AS2P2,CNP4,DSP4,ESP4,JSP4 (1,2,3,4,6,7,14)

A variável que contribui mais a esta classe é TM3 com um risco de : 0.534

Contribuição à classe : CS1P1,AN1P1 (15)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.755

Contribuição à classe : CS2P1,CS1P2 (16)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.755

Contribuição à classe : CS2P2,ANG1P3 (17)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.755

Contribuição à classe : ANG2P3,ANP4 (18)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.755

Contribuição à classe : BNP4,HNP4 (19)

A variável que contribui mais a esta classe é TM3 com um risco de : 0.594

Contribuição à classe : INP4,KSP4 (20)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.755

Contribuição à classe : LNP4,MNP4 (21)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.755

Contribuição à classe : NSP4,OSP4 (22)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.755

Contribuição à classe : BSP4,CSP4,GSP4,FSP4,GNP4 (5,8,9,23)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.435

Contribuição à classe : AN2P1,AN3P1,AN1P2,AN2P2 (10,11,24)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.417

Contribuição à classe : CS1P1,AN1P1,PSP4 (15,25)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.755

Contribuição à classe : DNP4,ENP4,FNP4,JNP4 (12,13,26)

A variável que contribui mais a esta classe é TM1 com um risco de : 0.621

Contribuição à classe : CS2P1,CS1P2,CS2P2,ANG1P3 (16,17,27)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.755

Contribuição à classe : ANG2P3,ANP4,BNP4,HNP4 (18,19,28)

A variável que contribui mais a esta classe é TM3 com um risco de : 0.594

Contribuição à classe : INP4,KSP4,LNP4,MNP4 (20,21,29)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.755

Contribuição à classe :

AS2P1,AS3P1,AS1P2,AS2P2,CNP4,DSP4,ESP4,JSP4,BSP4,CSP4,GSP4,FSP4,GNP4
(1,2,3,4,5,6,7,8,9,14,23,30)

A variável que contribui mais a esta classe é TM4 com um risco de : 0.517

Contribuição à classe : CS1P1,AN1P1,PSP4,NSP4,OSP4 (15,22,25,31)

A variável TM5 contribui a esta classe com um risco de : 1

A variável que contribui mais a esta classe é TM4 com um risco de : 0.755

Apêndice E

Codificação da Variáveis

CS1P1 = CÁLCULO CORRETO DA PERGUNTA 1 PARTE 1

CN1P1 = CÁLCULO INCORRETO DA PERGUNTA 1 PARTE 1

AS1P1 = ANÁLISE CORRETA DA PERGUNTA 1 PARTE 1

AN1P1 = ANÁLISE INCORRETA DA PERGUNTA 1 PARTE 1

AS3P1 = ANÁLISE CORRETA DA PERGUNTA 3 PARTE 1

AN3P1 = ANÁLISE INCORRETA DA PERGUNTA 3 PARTE 1

CS2P1 = CÁLCULO CORRETO DA PERGUNTA 2 PARTE 1

CN2P1 = CÁLCULO INCORRETO DA PERGUNTA 2 DA PARTE

1

AS2P1 = ANÁLISE CORRETA DA PERGUNTA 2 PARTE 1

AN2P1 = ANÁLISE INCORRETA DA PERGUNTA 2 PARTE 1

AS3P1 = ANÁLISE CORRETA DA PERGUNTA 3 PARTE 1

AN3P1 = ANÁLISE INCORRETA DA PERGUNTA 3 PARTE 1

CS1P2 = CÁLCULO CORRETO DA PERGUNTA 1 PARTE 2

CN1P2 = CÁLCULO INCORRETO DA PERGUNDA 2 PARTE 2

AS1P2 = ANÁLISE CORRETA DA PERGUNTA 1 DA PARTE 2

AN1P2 = ANÁLISE INCORRETA DA PERGUNTA 1 PARTE2

CS2P2 = CÁLCULO CORRETO DA PERGUNTA 2 PARTE 2

CN2P2 = CÁLCULO INCORRETO DAPERGUNTA 2 PARTE2

AS2P2 = ANÁLISE CORRETA DA PERGUNTA 2 PARTE 2

AN2P2 = ANÁLISE INCORRETA DAPERGUNTA 2 PARTE 2

ASG1P3 = ANÁLISE CORRETA DO GRÁFICO 1 PARTE 3

ANG1P3 = ANÁLISE INCORRETA DO GRÁFICO 1 DA PARTE 3

ASG2P3 = ANÁLISE CORRETA DO GRÁFICO 2 DA PARTE 3

ANG2P3 = ANÁLISE INCORRETA DO GRÁFICO 2 DA PARTE 3

ASP4 = FÁCIL ABORDAR MÉDIA DA PARTE 4

ANP4 = DIFÍCIL ABORDAR MÉDIA DA PARTE 4

BSP4 = FÁCIL ABORDAR MODA PARTE 4

BNP4 = DIFÍCIL ABORDAR MODA PARTE 4

CSP4 = FÁCIL ABORDAR MEDIANA PARTE 4

CNP4 = DIFÍCIL ABORDAR MEDIANA PARTE 4

DSP4 = FÁCIL ABORDAR GRÁFICO SETORIAL DA PARTE 4

DNP4 = DIFÍCIL ABORDAR GRÁFICO SETORIAL DA PARTE 4

ESP4 = FÁCIL ABORDAR TABELA DE FREQUÊNCIA DA PARTE 4

ENP4 = DIFÍCIL ABORDAR TABELA DE FREQUÊNCIA DA PARTE 4

FSP4 = FÁCIL ABORDAR MEDIDAS DE TENDÊNCIA CENTRAL DA PARTE 4

FNP4 = DIFÍCIL AOBDRAR MEDIDAS DE TENDÊNCIA CENTRAL DA PARTE 4

GSP4 = FÁCIL ABORDAR CLASSIFICAÇÃO DA MODA DA PARTE 4
GNP4 = DIFÍCIL ABORDAR CLASSIFICAÇÃO DA MODA DA PARTE 4
HSP4 = FÁCIL ABORDAR GRÁFICO DE BARRAS DA PARTE 4
HNP4 = DIFÍCIL ABORDAR GRÁFICO DE BARRAS DA PARTE 4
ISP4 = FÁCIL ABORDAR GRÁFICO DE COLUNAS DA PARTE 4
INP4 = DIFÍCIL ABORDAR GRÁFICO DE COLUNAS DA PARTE 4
JSP4 = FÁCIL ABORDAR POLÍGONO DE FREQUÊNCIA DA PARTE 4
JNP4 = DIFÍCIL ABORDAR POLÍGONO DE FREQUÊNCIA DA PARTE 4
KSP4 = FÁCIL ABORDAR BOX-PLOT DA PARTE 4
KNP4 = DIFÍCIL ABORDAR BOX-PLOT DA PARTE 4
LSP4 = FÁCIL ABORDAR ROL DA PARTE 4
LNP4 = DIFÍCIL ABORDAR ROL DA PARTE 4
MSP4 = FÁCIL ABORDAR DADOS BRUTOS DA PARTE 4
MNP4 = DIFÍCIL ABORDAR DADOS BRUTOS DA PARTE 4
NSP4 = FÁCIL ABORDAR MEDIDAS SEPARATRIZES DA PARTE 4
NNP4 = DIFÍCIL ABORDAR MEDIDAS SEPARATRIZES DA PARTE 4
OSP4 = FÁCIL ABORDAR MEDIDAS DE DISPERSÃO DA PARTE 4
ONP4 = DIFÍCIL ABORDAR MEDIDAS DE DISPERSÃO DA PARTE 4
PSP4 = FÁCIL ABORDAR DECIL, PERCENTIL, QUARTIL E QUINTIL DA PARTE 4
PNP4 = DIFÍCIL ABORDAR DECIL, PERCENTIL, QUARTIL, E QUINTIL DA PARTE 4

FASES DA CARREIRA EM ANOS

TM1 S = 3 - 5
TM2 S = 4 - 6
TM3 S = 7 - 18
TM4 S = 19 - 30
TM5 S = 31 - 40