CARLOS RICARDO BIFI

ESTATÍSTICA EM UM CURSO DE ADMINISTRAÇÃO DE EMPRESAS

MOBILIZAÇÃO DOS CONCEITOS ESTATÍSTICOS DE BASE

MESTRADO EM EDUCAÇÃO MATEMÁTICA

PUC/SP
2006

CARLOS RICARDO BIFI

ESTATÍSTICA EM UM CURSO DE ADMINISTRAÇÃO DE EMPRESAS

MOBILIZAÇÃO DOS CONCEITOS ESTATÍSTICOS DE BASE

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

PUC/SP

2006

BANCA EXAMINADORA

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

Assinatura: ______________________________ **Local e Data:** ______________

AGRADECIMENTOS

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

Em especial:

- *À Professora Doutora Cileda, orientadora e amiga pelo incentivo, paciência e dedicação;*
- *À minha Esposa Alessandra e minha filha Bárbara pela força e compreensão nos momentos em que estive ausente;*
- *Aos amigos que fiz na pós-graduação pelas discussões e contribuições ao trabalho;*
- *Aos alunos do Curso de Administração de Empresas da USP de Ribeirão Preto que se dispuseram, com boa vontade, a participar desta pesquisa;*
- *À professora Doutora. Silvia Dias Alcântara Machado pelas valorosas contribuições ao trabalho e também como professora na formação para o mestrado;*
- *À professora Doutora Adriana Backx Noronha Viana pelo apoio e amizade, assim como as imensas contribuições durante a fase experimental deste trabalho;*
- *Ao professor Doutor. Saddo Ag. Almouloud pelo apoio, amizade e orientações ao longo de todo o percurso.*

RESUMO

A Estatística tem se destacado ultimamente por sua utilidade em praticamente todas as áreas do conhecimento. Pesquisas e dissertações existentes sobre o assunto sugerem a necessidade de aprofundar os conhecimentos sobre as dificuldades no processo de ensino e aprendizagem dessa disciplina, incluindo-se os aspectos didáticos do tema e os erros que os alunos geralmente cometem após a aprendizagem. Nossa questão principal é investigar se os alunos egressos do componente curricular Estatística do curso de Administração estão capacitados a utilizar e/ou mobilizar, de forma eficaz, as noções estatísticas de base – variabilidade, para resolver problemas práticos dentro da sua área de atuação. Dessa forma, nessa pesquisa, pretendemos verificar o nível de mobilização dos conhecimentos por parte dos alunos do Ensino Superior segundo os termos de A. Robert (1998), que realizou um estudo sobre quatro dimensões de análise dos conteúdos a ensinar no campo da Matemática. A Quarta dimensão, que é alvo do nosso trabalho, trata dos níveis de ajustes em funcionamento dos conhecimentos pelos alunos. Tentaremos diagnosticar qual o nível de conhecimento em que o aluno do Ensino Superior se encontra no conteúdo curricular Estatística quanto a Técnico, Mobilizável e Disponível. Mediante atividades extraídas de situações-problema da realidade do profissional da área de administração, e que exigem conhecimentos de Estatística, pode o aluno fazer a inter-relação entre os conteúdos que estudou e a resolução de problemas práticos da área profissional?

Palavras-chave: Estatística, Análise Exploratória de Dados, Ensino-aprendizagem.

ABSTRACT

Statistics has detach itself lately because it is useful in all knowledge areas. Researches and studies about this subject show the necessity to deepen the knowledge about difficulties in learning – teaching process including didactic aspects of this subject and mistakes made by students after this learning. Our main question is to study if Administration Students, learning statistics, are ready to use, in a competent way, these notions to solve practical problems in their performance area. In this research we intend to verify the mobilization knowledge level by students in college according to studies made by A. Robert (1998) where he researched four analysis dimensions to teach on math. The dimension number four, that is our study goal, studies the levels of mobilization knowledge from students. We will try to diagnose in what knowledge level are students college in statistics studies reffering to tecnique, "mobilizável" and available concepts. Using activities extracted from problem situations in Administration Area and statistics knowledge can students make inter – relations between the students subjects and the practical problems of their professional area?

Keywords: Statistics, Exploratory Analyse, Teaching-learning.

SUMÁRIO

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE GRÁFICOS

LISTA DE FIGURAS

INTRODUÇÃO

A Estatística é uma ciência que tem por objetivo obter, organizar e analisar dados, determinar as correlações que apresentem, tirando delas suas conseqüências para descrição e explicação do que passou e previsão e organização de inferências para tomada de decisões. Sendo assim, no mundo de hoje, a Estatística não é mais disciplina apenas para matemáticos. Ao ler os resultados de uma pesquisa de opinião ou ao receber um relatório de produção ou de vendas de uma empresa, vários conceitos estatísticos estão em jogo, por exemplo, variabilidade e probabilidade. Informações estatísticas sempre estiveram presentes na vida dos cidadãos e, assim como muitas pessoas confiam e as utilizam para nortear suas decisões, outras as olham, desconfiam e/ou “atacam” sua veracidade.

Não é à toa essa desconfiança, principalmente para aqueles que, baseados numa informação, fizeram uma escolha e depois verificaram que as pesquisas estatísticas tiveram seus resultados mal interpretados ou mesmo manipulados. Basta recordar a grande quantidade de equívocos nos prognósticos em eleições passadas, em que institutos de pesquisa “erraram” em diferentes lugares. Segundo Lopes:

> 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 a habilidade de relacionar criticamente os dados apresentados, questionando e ponderando até mesmo sua veracidade. (LOPES, 2004, p. 189).

Dessa forma, não é suficiente que a pessoa desenvolva capacidade de organização e representação de uma coleção de dados apenas; a compreensão de mensagens estatísticas requer a utilização de várias habilidades na interpretação dos dados para obter significado na informação apresentada ao leitor. O que queremos apontar é que a porção escrita de uma mensagem pode ser bastante longa ou até mesmo os dados apresentarem-se por meio de gráficos, utilizando apenas algumas palavras. Em qualquer que seja o caso, a análise realizada nestes dados, deve conter bastante informação ou deve comprovar no que ela está fundada, para permitir que outro ouvinte ou leitor julgue sua racionalidade.

A capacidade dos cidadãos para interpretar uma grande quantidade de dados quantitativos e qualitativos assume, hoje, uma grande importância, ou seja, ser competente em estatística é fundamental para o entendimento das informações que os meios de comunicação social veiculam em todos os níveis de acesso, sejam eles televisivos, escritos ou até mesmo pela rede de internet. Segundo Gal (2002, p. 5) ”..., ser competente em estatística é possuir um conjunto de elementos que contribuem com a habilidade das pessoas para compreender, interpretar, avaliar, e, caso necessário, reagir a mensagens estatísticas”. Esse conjunto de elementos, segundo o autor, é composto de conhecimento de alfabetização, conhecimento estatístico, conhecimento matemático, conhecimento em contexto e questões críticas.

Embora os PCN – Parâmetros Curriculares Nacionais para o Ensino Fundamental e Ensino Médio (1999), que explicitam as habilidades das competências específicas de cada área, em atendimento ao estabelecido na LDB – Lei de Diretrizes e Bases da Educação Nacional, lei n. 9394/96, façam referências diretas à inclusão de elementos de Estatística e Probabilidade na Escola Básica, os alunos estão chegando ao curso superior com pouco ou nenhum contato anterior com esta área do saber. A maioria destes alunos utiliza mecanicamente a Estatística simplesmente como uma ferramenta matemática, usando suas fórmulas e algoritmos sem a devida compreensão do(s) conceito(s) que os justifica(m) e, também; sem a percepção da aplicabilidade desses conceitos na sua área de atuação.

Já nos cursos de graduação em Ciências Humanas, mais especificamente em suas grades curriculares (gestão Administração), enfatiza-se a importância da disciplina Estatística na análise e interpretação de dados, na ajuda de decisões tomadas na prática profissional e também na formação de um cidadão crítico.

Segundo as DCN (Diretrizes Curriculares Nacionais), parecer n. 0146/2002 CES/CNE do curso de graduação em Administração, também a LDB (Lei de Diretrizes e Bases da Educação Nacional) lei n. 4024/61, em seu art. 9°, posteriormente também a Lei de Reforma Universitária, lei n. 5540/68, no art. 26, *estabelecem ao Conselho Federal de Educação a incumbência de fixar os currículos mínimos dos cursos de graduação, válidos para todo o país, os quais foram concebidos com os objetivos elencados no artigo 9º. Os cursos de graduação de Administração devem formar profissionais que revelem, pelo menos, as seguintes competências e habilidades:*

- *reconhecer e definir problemas, equacionar soluções, pensar estrategicamente, introduzir modificações no processo produtivo, atuar preventivamente, transferir e generalizar conhecimentos e exercer, em diferentes graus de complexibilidade, o processo da tomada de decisão; desenvolver expressão e comunicação compatíveis com o exercício profissional, inclusive nos processos de negociação e nas comunicações interpessoais ou intergrupais;*

- *refletir e atuar criticamente sobre a esfera da produção, compreendendo sua posição e função na estrutura produtiva sobre seu controle e gerenciamento;*
- ***desenvolver raciocínio lógico crítico e analítico para operar com valores e formulações matemáticas presentes nas relações formais e causais entre fenômenos produtivos, administrativos e de controle, bem assim expressando-se de modo crítico e criativo diante dos diferentes contextos organizacionais e sociais;***
- *ter iniciativa, criatividade, determinação, vontade política e administrativa, vontade de aprender, abertura às mudanças e consciência da qualidade e das implicações éticas do seu exercício profissional;*
- *desenvolver capacidades de transferir conhecimentos da vida e da experiência cotidianas para ambiente de trabalho e do seu campo de atuação profissional, em diferentes modelos organizacionais, revelando-se profissional adaptável.*

É notório que somente as DCN não garantem o sucesso das competências e habilidades que abordam. É necessário que, durante o contato com a disciplina, o aluno esteja empenhado em dedicar-se aos conteúdos que serão ministrados durante sua formação e, também, que o profissional incumbido de ministrar o curso faça-o de forma clara para atingir os objetivos traçados por esse profissional educador.

Segundo Potter (1995, p. 260), "..., o objetivo mais importante de um curso de Estatística é encorajar os estudantes a serem praticantes deste instrumental. O conhecimento estatístico definitivamente nada significa se ele não se relaciona às questões e problemas reais".

Há indícios de que os estudantes universitários não são encorajados na prática da utilização da Estatística; pelo contrário, suspeita-se que estes mesmos estudantes, ao se depararem com a disciplina estatística, sentem-se incomodados e relacionam as dificuldades que apresentam na Matemática com a disciplina Estatística.

Problemas tais como desinteresse dos alunos, suas dificuldades no trato com o conteúdo e na articulação do conhecimento estatístico oferecido nas aulas, dificuldades com a realidade dos diferentes campos do conhecimento têm se tornado uma questão emergente. Muitas vezes, os alunos não conseguem visualizar como a análise estatística será aplicada na sua futura prática profissional e terminam o curso de graduação sem a instrumentação necessária para a utilização da ferramenta estatística na solução de problemas em sua área de atuação.

Algumas pesquisas como, por exemplo, Batanero (2001), têm revelado que, mesmo para utilização das ferramentas de análise de dados mais simples, os estudantes mostram ter dificuldades. É o caso das representações em gráficos e o manuseio de freqüências relativas, que envolvem as aplicações matemáticas com números racionais e conceitos relacionados às medidas. Percebe-se que esse aluno, durante sua formação, depara-se com situações-problema que são, *a priori*, analisadas pelo professor como "conteúdos ensinados", isto é, parte-se da hipótese de que a maioria dos livros didáticos utilizados nesses cursos são organizados didaticamente para o desenvolver do programa, mas não necessariamente, adequados a realidade profissional que esses alunos irão enfrentar após o término do curso. Diante dessas dificuldades, surgem, por parte dos estudantes, indagações do tipo: *para que servem estes cálculos? Onde, como e quando vou aplicá-los no meu cotidiano profissional?* Muitas vezes, estas perguntas ficam sem respostas, e os exemplos utilizados raramente têm alguma relação com a realidade dos estudantes.

O que pretendemos enfocar é que o profissional formado em Administração, ao analisar um banco de dados, deve ter condições de verificar quais são os invariantes da situação que possibilitem a busca dos modelos e ferramentas estatísticas para transformar dado em informação e, assim, poder inferir para futuras tomadas de decisões.

Partindo desses pressupostos, vale aqui ressaltar que a alfabetização estatística de adultos, citada por Gal (2002), enfatiza a importância de permitir a todas as pessoas trabalharem efetivamente dentro de uma sociedade guiada pela informação. Este estudo preocupa-se com a habilidade das pessoas de agirem como “consumidores de dados” em diversos contextos de vida que, para abreviar, são identificados aqui pela terminologia “contextos de leitura”. Estes contextos emergem, por exemplo, quando as pessoas estão em casa e assistem TV ou lêem um jornal; ou olham para anúncios enquanto fazem compras, quando visitam sites da Internet, quando participam de atividades comunitárias ou assistem a um evento cívico ou político; ou quando lêem matérias e relatórios no trabalho. Segundo esse autor, “... o estudo presente foca na alfabetização estatística uma área de conhecimento crítica, mas freqüentemente negligenciada, que precisa ser direcionada se os adultos, ou os futuros adultos, pretendem se tornar cidadãos e empregados mais informados”.

Segundo Wallman:

> (...) alfabetização estatística é a habilidade para entender e avaliar criticamente resultados estatísticos que fazem parte do cotidiano, juntamente com a habilidade para apreciar as contribuições que o pensamento estatístico pode fazer em decisões públicas e privadas, profissionais e pessoais[1] (WALLMAN, 1993 p. 2).

A Justificativa para este trabalho é reforçar a hipótese de que o ensino da disciplina Estatística nos cursos de graduação precisa ser repensado e direcionado a contextos reais na vida prática profissional dos estudantes em Administração. Pretendemos, com este trabalho, ser útil no campo da pesquisa em Educação Matemática e, reforçar os já existentes na área, contribuindo assim, para uma futura prática didática coerente com a proposta educacional.

---

[1] Wallman (1993) argued that statistical literacy is the ability to understand and critically evaluate statistical results that permeate daily life, coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions.

1. ESTUDOS PRELIMINARES

Neste capítulo abordaremos o ensino da Estatística no Ensino Superior, bem como, sua trajetória no decorrer de algumas décadas. Também faremos uma análise da importância da Matemática no ensino da disciplina de Estatística, abordando os objetos matemáticos envolvidos para os cálculos estatísticos e os obstáculos pedagógicos que a matemática trás para o ensino da disciplina e quais as expectativas dos alunos do curso de Administração frente a disciplina da Matemática e a Estatística.

1.1 A ESTATÍSTICA NO ENSINO SUPERIOR

A palavra Estatística origina-se do latim *statisticum* (referente aos estadistas). Por sua vez, *statisticum*, do latim status (estado). Empregada na forma italiana *statistica* desde 1633 com o sentido de ciência do Estado, só no século XVIII que o termo passou a ser usado com o significado atual (Mirador, 1989). Que a palavra origina-se do latim *status* (estado) não há dúvidas, mas seria Estado (organização social) ou estado (modo de ser)? Esta é uma dúvida que persiste, sem entretanto despertar interesse científico, pois são perfeitamente aceitáveis as duas idéias, uma vez que a Estatística, a princípio, tratava dos negócios de interesse do Estado (organização social) e, de certa forma, era usada para demonstrar em que estado (modo de ser) se encontrava este estado. Hoje, Estatística não é exclusivamente do Estado, sendo muito utilizada por empresas privadas, todavia é muito comum encontrar dois significados diferentes, porém relacionados, da palavra Estatística.

Num sentido mais comum, usa-se esta palavra para designar uma coleção de dados numéricos, mas, por outro lado, a palavra Estatística também é usada referindo-se ao ramo da Matemática Aplicada que se preocupa em analisar um conjunto de dados observados, fornecendo subsídios para sua correta interpretação (Toledo & Ovalle, 1995). A Estatística teve suas origens na história do homem e se desenvolveu paralelamente à evolução da civilização e do conhecimento humano, sendo alimentada, ao longo dos tempos, por contribuições de diversas épocas e regiões do mundo (Milone & Angelini, 1995).

Nos últimos 30 anos, a humanidade presenciou um aumento sem precedentes da aplicação do método estatístico na solução mais diversos problemas. Entretanto a Estatística pode ser considerada milenar, pois, desde a antiguidade, vários povos já registravam o número de habitantes, de nascimentos, de óbitos, faziam estimativas das riquezas individuais e sociais, distribuíam eqüitativamente terras, cobravam impostos e realizavam inquéritos por processos, que, hoje, podem ser chamados de “estatísticas”. Comprovando isso estão os egípcios e os persas, que há aproximadamente 3000 a.C., já coletavam e registraram, em forma de desenhos e hieróglifos, diversos tipos de dados em pirâmides, pedras, etc. Muitos destes dados estatísticos que foram gravados em monumentos egípcios possibilitam, hoje, uma avaliação das condições econômicas do Estado e dos movimentos das populações. Também foram encontrados registros na biblioteca de Assurbanipal, com diversas informações sobre o império Assírio. Na China, foram feitos censos demográficos em 2275 e 2238 a.C., conforme registra o livro sagrado de Confúcio (chouking). Em Roma, também foram feitos os chamados *Census Romanus*, realizados por Sérvio Túlio – 6º Rei de Roma, no ano de 556 a.C.; no ano de 28 a.C., Augusto, primeiro imperador romano, estendeu o recenseamento a todas as províncias romanas (Milone & Angelini, 1995).

Segundo Castro (1975), transcorrido este período inicial, a história da Estatística pode ser dividida em três grandes fases:

- 1ª fase: marcada pela organização de registros sistemáticos de informações e cadastros de interesse do Estado, com finalidades bélicas e/ou tributárias;
- 2ª fase: marcada pela tendência de erigir a Estatística como disciplina autônoma e pelo início de seu desenvolvimento formal. Grandes pensadores e cientistas da época, sobretudo os matemáticos, tiveram importantes participações neste período. Entre os matemáticos, Quételet, com sua "Física social", acabou confirmando a regularidade com que os fatos sociais e demográficos se manifestam, e ainda aproximou muito a Estatística da Matemática. Em 1708, a Universidade de Lena inaugurou um curso de Estatística. Em 1749, o termo "Estatítica" foi generalizado pelo economista e professor alemão Godofred Achenwall, da Universidade de Gottingen. Achenwall também definiu-lhe o objeto e as suas relações com as ciências;
- 3ª fase: é o período que evoluiu até os dias atuais. É marcado pelo aperfeiçoamento técnico e científico, iniciado em 1853 com o 1º Congresso de Estatística. Deste congresso até hoje, o método estatístico vem sendo aprimorado e cada vez mais aplicado aos mais variados campos do conhecimento. A partir de então, as tabelas tornaram-se mais complexas, surgiram representações gráficas cada vez mais modernas, a teoria da probabilidade consolidou a fundamentação teórica da inferência estatística, e a Estatística deixou de ser uma simples catalogação de dados numéricos

coletivos para se tornar o estudo de como chegar a conclusões sobre o todo (população), partindo da observação de partes desse todo (amostra).

Sendo assim, consolidando-se esta última fase, a Estatística se impôs, mercê das inúmeras aplicações e contribuições que trouxe às outras ciências e também nas mais diversas áreas profissionais. Os resultados obtidos com a aplicação dos métodos estatísticos na resolução de problemas dos diversos domínios do conhecimento, aliados à evolução tecnológica dos últimos anos, fizeram com que os conhecimentos estatísticos se tornassem indispensáveis nas grades curriculares, tanto do Ensino Básico como do Ensino Superior. Os próprios êxitos obtidos com a aplicação dos métodos e técnicas estatísticas contribuíram para o desenvolvimento da própria ciência. Segundo Batanero:.

> É indiscutível que o século XX tem sido o século da Estatística, que passou a considerar-se uma das ciências metodológicas fundamentais e base do método científico experimental. O ensino da Estatística, no entanto, ainda se encontra em seu começo, ainda que, como tenha descrito, avança continuamente (BATANERO, 2001 p. 7).

O ensino da Estatística começou por ser feito nas universidades em cursos de pós-graduação e, mais tarde, foi integrado no currículo de algumas licenciaturas. Tratava-se de um ensino de caráter essencialmente informativo, sem grandes preocupações de ordem didática, cujo principal objetivo era formar estatísticos e munir os pesquisadores com novos métodos e técnicas que servissem de ferramenta para o seu trabalho. A integração da Estatística nos currículos de nível superior foi condicionada à perspectiva convencional de que, para se aprender Estatística era necessário possuir sólida formação matemática, que, com o passar do tempo, tornou-se tema fundamental nas pesquisas em Educação Matemática.

> A Estatística constitui hoje em dia técnica de máxima importância no campo da Administração. O conhecimento dos conceitos e métodos da Estatística é imprescindível por parte dos elementos responsáveis pelo Controle de Qualidade, pela Pesquisa Geral e Aplicada, pelo Planejamento Geral, pela Programação da Produção, pela Engenharia de Métodos e Tempos, pela Engenharia de Produtos e por todos os administradores que têm de analisar dados" (MACHLINE et al,1994, p. 119).

Destacando a aplicabilidade do método estatístico no gerenciamento de uma empresa, Werkema (1996, p. 1) destaca que importantes "ferramentas estatísticas podem ser empregadas para uma realização mais eficientes das atividades de gerenciamento de uma empresa".

Conforme Stevenson (1986, p. 3), "o raciocínio estatístico é largamente utilizado pelo governo e na Administração...", afinal, a tomada de decisões sempre esteve presente no cotidiano dos administradores e, muitas vezes, decisões erradas podem gerar conseqüências desastrosas e irremediáveis.

Segundo Chiavenato:

> A tomada de decisões é o núcleo da responsabilidade administrativa. O administrador deve constantemente decidir *o que fazer, quem deve fazer, quando, onde e,* muitas vezes, *como fazer.* Seja ao estabelecer objetivos, alocar recursos ou resolver problemas que surgem pelo caminho, o administrador deve ponderar o efeito da decisão de hoje sobre as oportunidades de amanhã. Decidir é optar ou selecionar dentre várias alternativas de curso de ação aquela que pareça mais adequada (CHIAVENATO, 2000 p. 172).

Para Vlahos:

> O ambiente competitivo em que as empresas operam se está tornando cada vez mais complexo e incerto devido à globalização das atividades e a rápida introdução de novas tecnologias. A maioria das decisões empresariais é um completo conhecimento sobre o futuro (VLAHOS, 1999 p. 143).

Druker (1998, p. 45) diz que tomar decisões "é função básica das empresas e constituem a essência da empresarização". Mas, embora o risco sempre esteja presente no processo de tomada de decisões administrativas, a aplicação do método estatístico pode reduzir o grau de incerteza. Segundo Vlahos (1999, p. 143), “o risco e a incerteza são ingredientes constantes nos negócios. Mas há métodos estatísticos que permitem avaliá-los, e, por conseguinte, superá-los”.

Downing & Clark (1998), referindo-se à aplicação do método estatístico na administração, destacam sua importância na pesquisa de mercado para lançamento de novos produtos, na análise de regressão para determinar relações de causa e efeito, nas inferências populacionais baseadas em amostras e na análise dos resultados de testes de novos produtos.

Um dos primeiros indicadores da expansão da Educação Estatística são os trabalhos realizados pelo *IASE (Associação Internacional para a Instrução Estatística)*, bem como os congressos do *ICOTS (Internacional Conference on the Teaching of Statistics)* e as sessões de educação das conferências *ISI (Internacional Statistical Institute).* Sociedades de Estatística e de Educação também organizaram sessões específicas de Educação Estatística, como, por exemplo, *a ASA (Americam Statistics Association), AERA (American Educational Research Association), Royal Statistical Society*, na Inglaterra, *Sociedade Estatística Japonesa*, S*ociedade Espanhola de Investigação em Educação Matemática*, etc.

As revistas especializadas no ensino da Estatística apontam para uma problemática docente e, em contrapartida, despertam um interesse dos professores em melhorar suas ações nessa prática. Podemos citar a *Teaching Statistics*, que tem 22 anos de existência e que foi desenvolvendo e adquirindo uma identidade e qualidade internacional. Outra revista similar é a *Induzioni y*

Journal of Statistical Education, que é uma revista de Educação Estatística de nível universitário com serviço para associados.

Percebemos que o papel da Estatística no Ensino Superior é o de formar profissionais conscientes e aptos a desempenhar bem essa atividade imprescindível nos dias de hoje.

Apoiado no tripé teoria-metodologia-aplicações, o currículo tem como objetivo desenvolver a aptidão e o raciocínio estatístico do aluno em aulas teóricas e práticas. Nas aulas teóricas, os conteúdos ministrados são voltados para a Estatística Descritiva, que tem o papel principal de suavizar a transição das variáveis aleatórias para a probabilidade, abordando, assim, a Estatística Inferencial. Nas aulas práticas, o futuro profissional aprende a lidar com problemas reais, organizando e analisando dados em diferentes áreas (Estatística Descritiva), entre as quais pesquisa mercadológica, controle de qualidade e otimização de processos industriais. Isso tudo faz com que o profissional, com conhecimento em Estatística em seu Ensino Superior, tenha um melhor direcionamento nas tomadas de suas decisões contribuindo assim, até mesmo no ambiente em que venha trabalhar.

1.2 OBJETO MATEMÁTICO

A definição de Estatística é encontrada de várias formas diferentes, por muitos autores. Segundo Pereira (1997, p. 8) “a Estatística pode ser considerada a tecnologia da ciência, auxiliando a pesquisa desde o seu planejamento até a interpretação dos dados”.

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

O estudo da Estatística divide-se em dois ramos principais: a Estatística descritiva – ramo que trata da organização, do resumo e da apresentação dos dados – e a Estatística Inferencial – ramo que trata de tirar conclusões sobre uma população a partir de uma amostra.

Quando falamos de Estatística, temos que ter em mente que ela se refere a amostra ou população. A amostra é um subconjunto da população no qual é imprescindível que se mantenha a mesma característica da população. A população representa a totalidade da variável em estudo, e, variável estatística é exatamente a característica que está sendo estudada. Exemplos que podemos citar de variável estatística: diâmetros na produção de porcas e parafusos de uma indústria, estaturas de pessoas de uma determinada comunidade, etc.

A variável pode ser apresentada de duas formas: variável qualitativa e variável quantitativa. Na primeira (variável qualitativa), os dados coletados são categorizados, usando-se somente nomes, marcas ou qualidades. Nenhum cálculo matemático pode ser feito nesse formato. Já a variável quantitativa é aquela em que os dados podem ser ordenados e é possível calcular diferenças significativas entre os registros de dados – e, assumem diretamente valores numéricos, definidos em intervalo reais –, podendo ser discreta ou continua. A variável quantitativa é discreta quando for enumerável, ou seja, entre dois valores consecutivos não podemos inserir nenhum outro valor. Por exemplo: números de “livros-caixa” de uma Empresa – não podemos pegar 35,2 livros-caixa para analisar. Já na variável quantitativa contínua, todos os valores podem assumir outro qualquer valor dentro de um intervalo real, isto é, não há como enumerar todos os valores. Por exemplo, estaturas, pesos, entre outras.

Em nosso trabalho serão abordados alguns conceitos estatísticos que envolvem variabilidade e, entre eles, estão as medidas de tendência central, as medidas de posição e as medidas de variação. A medida de tendência central é um valor que representa o centro de um conjunto de dados ordenados de forma crescente ou decrescente. As três medidas de tendência central mais usadas são a média, a mediana e a moda.

Podemos definir média aritmética de um conjunto de dados como a soma dos valores dados dividida pelo número total dos valores. Para encontrar a média de um conjunto de dados, usaremos uma das seguintes fórmulas: para *média amostral:* $\bar{x} = \sum_1^n \frac{x_i}{n}$; para *média populacional:* $\mu = \sum_1^n \frac{x_i}{n}$ (para ambas, Σ denota somatório de um conjunto de dados; x denota variável usada para representar valores individuais dos dados e *n* representa o número de valores em uma amostra).

A mediana de um conjunto de valores é o valor do meio de um conjunto de dados, quando os valores estão dispostos em ordem crescente (ou decrescente). Para o cálculo da mediana para dados agrupados em ordem crescente (ou decrescente), aplicaremos um dos processos abaixo:

- se o número de valores na amostra for ímpar, a mediana é o número localizado exatamente no meio da lista;

- se o número de valores é par, a mediana será a média aritmética dos dois valores centrais.

A moda de um conjunto de dados é o valor que ocorre com maior freqüência. Quando dois valores ocorrem com a mesma freqüência, cada um deles será uma moda, e o conjunto se diz *bimodal*. Se mais de dois valores ocorrem com a mesma freqüência, cada um deles é uma moda, e o conjunto é multimodal. Quando nenhum valor é repetido, o conjunto não tem moda.

Outras medidas que fazem parte deste estudo serão as medidas de posição: o*s quartis*. Os q*uartis* são medidas que dividem aproximadamente o conjunto de dados ordenado em quatro partes iguais: cerca de um quarto $\left(\frac{1}{4}\right)$ dos dados ficam dentro ou abaixo do *primeiro quartil (Q_1)*; cerca da metade dos dados ficam dentro ou abaixo do *segundo quartil Q_2* (o segundo quartil é igual à mediana do conjunto de dados); cerca de três quartos $\left(\frac{3}{4}\right)$dos dados ficam dentro ou abaixo de *terceiro quartil (Q_3).*

As medidas de variação que abordaremos neste trabalho serão as seguintes: *Desvio-padrão* e *Variância.* Começaremos com a amplitude total da amostra, que é a diferença entre o maior e o menor valor da amostra e é facilmente calculada pela expressão: $H = Valor_{\max} - valor_{\min}$. Como medida de variação, a amplitude total tem a vantagem de ser facilmente calculável. Sua desvantagem, porém, é que ela usa somente dois valores do conjunto de dados. Duas medidas de variação que usam todos os valores do conjunto de dados são: a variância e o desvio–padrão. Contudo, antes de falarmos sobre estas medidas, seria coerente falar um pouco do significado do desvio de um conjunto de dados.

Desvio de um conjunto de dados de uma amostra é a diferença entre os valores x_i e a média $\bar{x}$ do conjunto de dados. Usaremos a seguinte fórmula: *desvio* $d = x_i - \bar{x}$.

A Variância de um conjunto de dados é a medida de variação dos valores em relação à média. Para calcular a Variância, recorremos à seguinte fórmula: $s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$.

O Desvio–padrão obtém–se extraindo a raiz quadrada da variância, e definimos por meio da fórmula: $s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$.

Em toda área do conhecimento, profissionais da Educação têm como objetivo principal identificar os pontos difíceis e os erros que surgem durante a aprendizagem. Na Matemática, podemos perceber que existem vários estudos que focam os erros: conceituais ou algorítmicos na disciplina. Na Educação Estatística não é diferente. Nosso trabalho também focará nesse ponto, no qual tentaremos identificar qual(is) a(s) dificuldade(s) e erro(s) que os alunos enfrentam, mesmo depois de terem passado pelo curso de Estatística.

Segundo Batanero (2001), as dificuldades não acontecem de forma aleatória, e, com freqüência, encontramos erros que se repetem regularmente e produzem associações com variáveis próprias das tarefas propostas, dos sujeitos, de circunstâncias presentes ou passadas e até mesmo de concepções dos alunos. Baseados nos conceitos estatísticos que já mencionamos anteriormente, pode-se ainda verificar que estes conceitos estão também apoiados por alguns conceitos matemáticos utilizados como ferramenta para a construção e mobilização dos conceitos estatísticos aqui relacionados.

Por outro lado, alunos do Ensino Superior precisarão estar atentos para alguns procedimentos matemáticos que estarão claramente por trás da criação dos indicadores estatísticos comuns, como porcentagem e média. Ao mesmo tempo, criam-se expectativas quanto à quantidade e nível necessário de conhecimento matemático formal para compreender idéias estatísticas básicas ensinadas ao nível introdutório no Ensino Superior (Moore, 1998).

Estatísticos esclareceram, durante os últimos, anos a natureza de algumas diferenças fundamentais entre Matemática e Estatística, e formularam algumas suposições de trabalho sobre o nível geral da Matemática de que a pessoa precisa para aprender estatística, pelo menos ao nível introdutório no Ensino Superior. Moore (1997) sugeriu que, enquanto os estatísticos fizerem uso "pesado" da Matemática, a aprendizagem em estatística no nível Superior tornar-se-á preocupante. A matemática deveria focar nas idéias estatísticas necessárias, a importância da produção dos dados, e a onipresença de variabilidade.

Entendemos que as derivações matemáticas que estão por trás de idéias principais apresentadas em estatística no Ensino Superior têm sua importância, mas deve ser limitada, uma vez que os computadores automatizaram vários cálculos. Entretanto a intenção não é colocar para os alunos que as principais derivações estatísticas ocorrem por "mágica", sem que estes saibam os conceitos matemáticos envolvidos no processo. Muita ênfase em teoria matemática não é esperada no início da aprendizagem no Ensino Superior; pode romper o desenvolvimento da compreensão intuitiva, necessária às idéias estatísticas fundamentais e aos conceitos que freqüentemente não têm representações matemáticas.

Talvez o conhecimento mais simples esperado dos alunos é a compreensão de que qualquer tentativa de resumir um número grande de observações por variável quantitativa, como, por exemplo, porcentagem, média e probabilidade, requerem alguma aplicação de ferramentas matemáticas e procedimentos. Os alunos precisam ter habilidades numéricas em um nível suficiente para possibilitar a interpretação correta de números usados em relatórios estatísticos. “Estudo sobre números” está crescentemente sendo apontado como uma habilidade essencial para o entendimento apropriado de diversos tipos de números, como “números grandes” e “números pequenos”, inclusive os decimais, frações e percentuais. A compreensão de resultados estatísticos básicos referentes a porcentagens ou médias requer familiaridade intuitiva, até certo ponto formal, sobre os procedimentos matemáticos ou computacionais que estão por trás da geração destas estatísticas. Os alunos devem saber que um estudo baseado em média pode ser influenciado por valores extremos em um conjunto de dados e, conseqüentemente, podem não estar representando a média correta para um conjunto de valores.

As idéias fundamentais da Matemática, embutidas no ensino da Estatística, apresentam-se na forma de “conceitos implícitos”. Podemos citar, como exemplo, os conceitos de proporcionalidade, de porcentagem, de contagem, de somatório, de frações, de produtos notáveis, e ainda conceitos básicos das operações aritméticas. Nestes termos, é praticamente impossível pensar em Estatística sem pensar em Matemática. Os cálculos, as representações, as análises, as variáveis, entre outros tópicos, também estão presentes na Estatística e igualmente no campo da Matemática.

Pensando nas idéias fundamentais do AMT – *Advanced Mathematical Thinking*, como referencial do que ocorre principalmente no estudo da Matemática avançada no nível universitário, três perspectivas foram abordadas por Heid et al (2004) para os estudos das idéias do AMT:

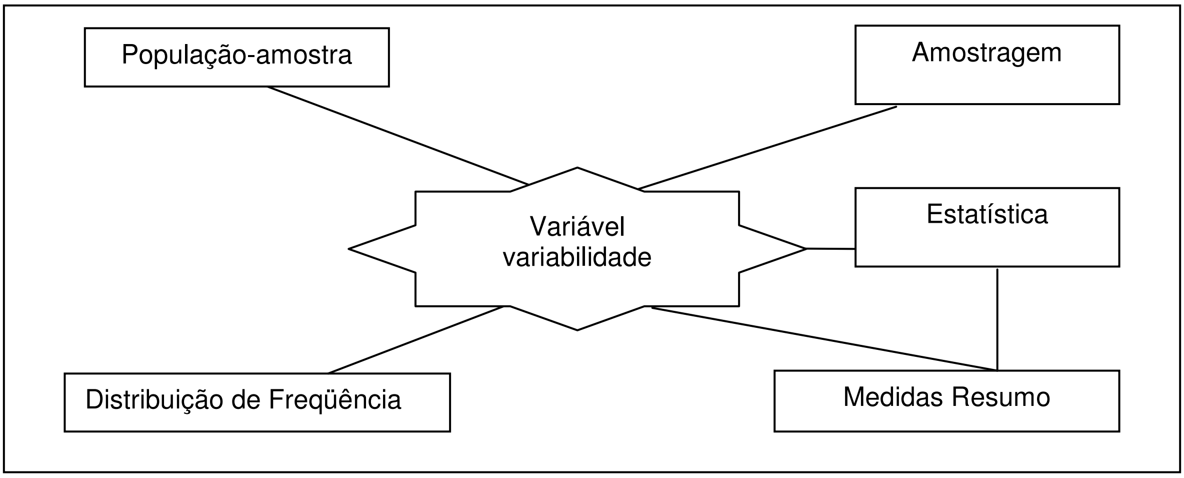
- 1ª – Grupo liderado por Bárbara Edwards, que considera que AMT requer duas condições. A primeira diz sobre raciocínio preciso das idéias matemáticas, e a segunda, que essas idéias matemáticas não são inteiramente acessíveis aos cinco sentidos. A autora enfatiza a "necessidade da fundamentação precisa das idéias matemáticas" e ainda sugere que os professores aprendam "como os estudantes desenvolvem sua capacidade de raciocinar na ausência de exemplos concretos";

- 2ª – Grupo liderado por Chris Rasmussen, que focaliza suas pesquisas em AMT na "Atividade da Matemática Avançada" e concebe a aprendizagem Matemática como participação no fazer Matemática. A concepção do AMT para este segundo grupo está voltada na Matematização Horizontal e Matematização Vertical. A primeira aborda a transformação de um contexto ou problema matemático ou ainda do mundo real de tal forma que permita favorecer uma análise matemática. A segunda, como atividade que se fundamenta ou é construída a partir da Matematização Horizontal.

Vamos nos ater a um exemplo para tentar explicar melhor o que o grupo defende. Imaginemos um operário de obras da construção civil, construindo um cômodo de uma casa, demarcando a área por meio de um retângulo. Para garantir que cada vértice deste retângulo possua ângulos retos, ele se utiliza da seguinte estratégia: partindo de um dos vértices, traça um segmento de reta sobre um dos lados desse retângulo com medida 3 uc (unidade de comprimento). Do outro lado, adjacente ao mesmo vértice, traça outro segmento de reta com 4 uc. Para garantir que o vértice escolhido meça 90º, a diagonal ligada pelos extremos destes dois segmentos deverá medir 5 uc. Esta prática, cuja a fundamentação teórica para justificar tal procedimento, muitas vezes é desconhecida para este operário, é denominada Matematização Horizontal, ou seja, ele a utiliza sem fundamentação, baseado apenas na prática. A justificativa para esta estratégia está fundamentada pelas propriedades do triângulo retângulo de lados, 3, 4 e 5, que garante um ângulo reto. Nesse caso, a fundamentação que dá suporte a esta estratégia é a Matematização Vertical. Seria viável que estas duas vertentes caminhassem paralelamente na construção dos conceitos.

- 3ª – Grupo desenvolvido por Guerson Harel, caracteriza o Pensamento Matemático como Avançado levando-se em conta os obstáculos epistemológicos envolvidos. Harel entende que são necessários grandes esforços e longo período de tempo para se desenvolver o pensamento matemático avançado e que esse trabalho deve iniciar-se nos conteúdos da matemática elementar com estudantes de pouca idade (Harel, 2000 *apud* Heid *et al*, 2004).

As perspectivas abordadas foram de imenso valor para pesquisadores em Educação Matemática, pois elas puderam servir de apoio para entender melhor o raciocínio do estudante, não só do Ensino Superior, mas também de qualquer nível em que ele se encontre. Entre essas perspectivas mostradas, acreditamos que a perspectiva de Chris Rasmussen enquadra-se naquilo em que pensamos, ou seja, não basta ao aluno do Ensino Superior ter a praticidade ou a aplicabilidade a esmo dos cálculos estatísticos, sem uma fundamentação teórica que a justifique. Dessa forma, o campo conceitual, tanto matemático quanto estatístico, envolvido na análise exploratória de dados, é muito vasto e complexo, de onde vem uma justificativa na opção por mostrar apenas os conceitos estatísticos de base envolvidos. Vejam o diagrama:



Quadro 1: conceitos estatísticos de base (Novaes, 2004, p. 63).

Nossa preocupação é verificar os caminhos percorridos pelo aluno na análise exploratória dos dados, ou seja, verificar se ele construiu as relações e esquemas necessários para sua mobilização de forma adequada, podendo assim explicitá-los de forma clara de acordo com o contexto e o problema a ser trabalhado.

2. PROBLEMÁTICA

2.1 O PROBLEMA DE PESQUISA

Como já foi discutido na introdução deste texto, a Estatística tem se destacado cada vez mais no âmbito internacional em todas as áreas do conhecimento. As pesquisas apresentadas por Educadores em Estatística, Batanero (2001), Gal (2002), Novaes (2004), Silva (1999) entre outros, apontam a necessidade de investir na alfabetização estatística para crianças, adolescentes e até mesmo adultos, visando sanar as dificuldades de aprendizado, por parte destes, ao se depararem com esta disciplina.

O objetivo desta pesquisa é verificar como os alunos egressos do componente curricular Estatística no curso de Administração de Empresas mobilizam os conceitos estatísticos (noções estatísticas de base) quando confrontados com problemas que envolvam variabilidade (média e desvio padrão) na análise exploratória de dados.

Focalizamos nosso estudo no problema central:

> Qual o nível de funcionamento dos conceitos, segundo os preceitos de Robert[2], especificamente aqueles ligados ao estudo da variabilidade, dos alunos do Ensino Superior?

Pesquisas apontam que a maioria dos alunos, desde as séries iniciais, não teve quase nenhum tipo de contato com a disciplina Estatística. Segundo

[2] Técnico, Mobilizável ou Disponível. Tais níveis serão apresentados mais detalhadamente na seqüência deste texto.

Vendramini, Chenta e Silva (2004), foram pesquisados 135 alunos de 7ª (46%) e 8ª (54%) séries do ensino fundamental de escolas públicas do interior do Estado de São Paulo, com idades variando de 12 a 17 anos; destes, 55% do gênero masculino. A maioria declarou nunca ter estudado conceitos de estatística (97%), embora tenham afirmado ter estudado tabelas (48%) e gráficos (30%).

Quanto aos conceitos matemáticos, a maioria declarou não ter estudado razão (83%) e taxa (94%). Alguns conceitos matemáticos, que normalmente são estudados antes da 8ª série, não foram citados por uma parcela da amostra: 21,5% afirmaram não ter estudado porcentagem, 29,6% gráficos e 48,1% tabelas. Quanto aos conceitos estatísticos, 84,4% dos alunos declararam não ter estudado média, 97,8% mediana, 98,5% moda, 98,5% medidas de variabilidade.

Os resultados acima citados reforçam nosso trabalho no âmbito de que o aluno do Ensino Superior, ao apresentar possíveis dificuldades na aprendizagem em Estatística, poderá nos levar à suspeita de uma má formação no Ensino Básico. Esta afirmação foi mostrada num estudo sobre o raciocínio estatístico de 325 estudantes universitários, indicando índices de dificuldade elevados para a maioria dos itens de uma prova, que se referia à interpretação de dados apresentados em tabela e de índices de dificuldades mais baixos nos itens referentes à interpretação de dados em gráficos. Uma das explicações para este resultado é o fato de que a tabela apresentada na prova tenha sido de dupla entrada e exigir a comparação das informações nelas contida para se chegar às respostas corretas das questões propostas (Vendramini, Silva & Canale, 2003).

2.2 QUADRO TEÓRICO

Utilizaremos, como referencial teórico em nosso trabalho, Robert (1998). A autora classifica os conteúdos de Matemática a ensinar em quatro dimensões, dispondo-se em etapas o acesso à compreensão das noções matemáticas que nos interessam.

A autora defende que estas etapas ajudam a identificar as possíveis dificuldades que podem surgir por meio de um contexto dado, seja para diagnosticar conhecimentos dos alunos, ou até mesmo propor aos alunos ou estudantes situações-problema que os professores julguem convenientes, para, assim, superar possíveis dificuldades detectadas.

As duas primeiras dimensões que a autora propõe estão relacionadas a outra teoria da Didática da Matemática, como a da Douady (Douady 1985, p. 35 *apud* Carvalho, 2001 p. 7) que, inspirada nos trabalhos de Vigotsky, afirma que um conceito matemático pode ser encarado como *ferramenta*, "quando nosso interesse se foca na sua utilização para resolver problemas" ou como *objeto*, quando o encaramos como "um objeto cultural que faz parte de um corpo científico de conhecimentos".

A terceira dimensão citada, Níveis de Conceitualização, que é associada com a Teoria dos Campos Conceituais (Vergnaud, 1998a), será útil em nosso trabalho e, segundo a autora:

> Trata-se de rotular dentro de um campo de conhecimentos matemáticos (Campo Conceitual) correspondentes a uma organização coerente de uma parte do campo, caracterizada por objetos matemáticos apresentados de certa maneira por teoremas sobre esses objetos, métodos associados a esses teoremas e problemas que os alunos podem resolver com os teoremas do nível considerado, utilizando esses métodos (ROBERT,1998, p.164).

Segundo Vergnaud:

> ...as competências e concepções dos alunos se desenvolvem ao longo do tempo, por meio de experiências com um grande número de situações tanto dentro quanto fora do ambiente escolar. Normalmente, quando é colocada uma nova situação para o aluno, ou seja, um novo domínio, novos dados numéricos ou, até mesmo, novas relações, este usa o conhecimento desenvolvido em sua experiência de situações anteriores e tenta adaptá-lo à nova (VERGNAUD, 1998, p. 173).

A Teoria dos Campos Conceituais considera que existe uma série de fatores que influenciam e interferem na formação e no desenvolvimento dos conceitos e que o conhecimento conceitual deve emergir dentro de situações-problema. O estudo do desenvolvimento de um campo conceitual, segundo esta teoria, requer uma visão por parte do pesquisador de um conceito formado por uma terna de conjuntos (***S, I, R***), no qual:

- ***S*** é um conjunto de situações que tornam o conceito significativo;
- ***I*** é um conjunto de invariantes operatórios (objetos, propriedades e relações) que podem ser mobilizados e usados pelo sujeito para analisar e dominar essas situações;
- ***R*** é um conjunto de representações simbólicas que podem ser usadas para pontuar e representar esses invariantes e, portanto, representar as situações e os procedimentos para lidar com eles.

Na maioria das vezes, professores e pesquisadores da educação têm dificuldades em entender que a compreensão de um conceito, por mais simples que seja, não emerge de um só tipo de situação, assim como uma simples situação sempre envolve mais que um único conceito. Um exemplo típico, em sala de aula no Ensino Superior, envolvendo média, é apresentado no seguinte molde: é apresentado um banco de dados ao aluno, no qual se identifica a variável estatística a ser trabalhada e pede-se para calcular a média. Porém o aluno, que só se depara com esse tipo de situação, poderá ter dificuldade em realizar o mesmo cálculo se os dados forem apresentados de uma outra forma, por exemplo, na forma de tabelas ou até na forma de gráficos. O que queremos mostrar neste exemplo é que o conceito de média precisa ser apresentado de diferentes situações para que o aluno possa construir os significados necessários para a formação e evolução desse conceito.

Sendo assim, os conceitos estatísticos apresentam seus significados a partir de uma variedade de situações, já que cada situação normalmente não pode ser analisada com a ajuda de apenas um conceito. Em outras palavras, nem um só conceito nem uma situação isolada dão conta do processo. Em sintonia com Vergnaud, propomos estudar conceitos estatísticos não como conceitos isolados, mas como conjuntos de conceitos inter-relacionados com conjuntos de situações.

Baseando-nos na teoria de Robert (1998), o que pretendemos identificar é o campo conceitual, no qual os estudantes do Ensino Superior alocam seus conceitos estatísticos em diversas situações de resolução de problemas. A transição desses níveis de conceitualização, segundo a Teoria dos Campos Conceituais, permite-nos distinguir diversos conhecimentos importantes de uma mesma noção, os quais podem ser abordados de diversas formas no decorrer da aprendizagem, por meio da diversidade de situações e de representações.

Na quarta dimensão, *Níveis de mobilização* dos conhecimentos pelos alunos, Robert (1998) descreve três níveis para os conhecimentos adquiridos: nível técnico, mobilizável e disponível. Um conhecimento é caracterizado como *nível técnico*, quando, para resolver um problema, o aluno recorre às indicações isoladas, colocando em jogo aplicações imediatas de teoremas, propriedades, definições, fórmulas, etc.

Trata-se de uma contextualização simples, local e sem adaptações. Por exemplo, se pedirmos para um aluno fazer uma representação gráfica de um conjunto de dados, ele aprendeu como elaborar o gráfico, e pode fazê-lo sem, no entanto, saber como interpretar de forma correta os dados ali representados, pois a interpretação de representações aponta para o estabelecimento de relações complexas, indo além de simples aplicações de fórmulas e procedimentos.

Uma outra situação seria o cálculo da média de um conjunto de dados, no qual o aluno mobiliza o cálculo algorítmico e não percebe a necessidade de associação com outras medidas para a análise da variabilidade desses dados, mostrando, assim, que não relaciona a média com outras medidas, de forma a poder efetivamente analisar os significados possíveis.

Robert (1998) classifica como nível mobilizável o caracterizado por um nível de fazeres em funcionamento mais amplo do que um conhecimento técnico. Por exemplo, a resolução de um problema proposto exige do aluno a adaptação de seus conhecimentos para aplicar o teorema adequado. O aluno precisará aplicar várias vezes o mesmo objeto matemático ou utilizar objetos distintos em etapas sucessivas ou, ainda, articular duas ou mais informações de natureza diferentes. O conhecimento mobilizado põe à prova um “fazer”, que coloca em funcionamento uma justaposição de saberes dentro de um domínio dado, direcionado para uma organização.

Segundo a autora, “Um saber é dito mobilizável, quando é bem identificado, é bem utilizado pelo aluno, mesmo que tenha sido necessária uma adaptação ao contexto particular" (Robert, 1998, p. 166). Percebemos que a autora exalta a adaptação necessária ao conhecimento, no qual esse pode ser sugerido pelo professor ou pelo próprio enunciado do problema, ou seja, o que diferencia conhecimento mobilizável de conhecimento técnico é a possibilidade de estabelecer relações complexas entre o objeto visado a outros e a utilização desse objeto como ferramenta para resolução de problemas complexos. Por exemplo, podemos citar uma atividade que fornece os dados coletados de uma determinada variável e, logo em seguida, pede-se ao aluno uma análise estatística destes dados. Dependendo de como está transcrito o enunciado da atividade, ele pode sugerir um caminho a ser percorrido, ou seja, implicitamente o aluno é conduzido a utilizar cálculos algébricos ou até mesmo raciocínio lógico, para que seja possível, a esse aluno, apresentar resultados favoráveis na resolução do problema.

Por fim, o nível de conhecimento disponível, segundo Robert, corresponde ao saber fazer o que foi proposto sem indicações. Saber procurar por si mesmo, entre os conhecimentos, aquele que pode ser utilizado. Saber dar contra-exemplos, mudar de quadro, sem sugestão, aplicar métodos não previsíveis. No exemplo que se segue, tentaremos ilustrar esses três níveis de funcionamento de um conhecimento:

"O editor-chefe de uma revista pede para um dos membros de sua equipe que investigue os preços da gasolina na cidade B e escreva um artigo sobre eles. Ele reuniu os dados na tabela abaixo. Os dados representam os preços pagos por um galão de gasolina comum em uma amostra aleatória de postos de gasolina da cidade do indivíduo e em outras cidades de tamanho similar".

Preços no varejo para um galão de gasolina comum em 10 postos selecionado ao acaso em quatro cidades

Cidade A	*Cidade B*	*Cidade C*	*Cidade D*
1,69	1,75	1,72	1,79
1,64	1,74	1,71	1,76
1,63	1,72	1,69	1,73
1,68	1,74	1,77	1,72
1,71	1,77	1,71	1,75
1,72	1,78	1,72	1,78
1,68	1,75	1,72	1,71
1,69	1,79	1,76	1,78
1,68	1,77	1,71	1,72
1,63	1,75	1,75	1,74

Tabela 1: Preço de galões de gasolina. (dados fictícios)

- *Determine as medidas-resumo para os preços de gasolina coletados.*
- *Como você pode analisar o comportamento (a variação) de preços em função destas medidas?*

Apresentado esta atividade ao aluno do ensino superior, ele poderá, assim, calcular média, mediana, moda, desvio-padrão, quartis e coeficiente de variação, sem, no entanto, relacioná-los. As medidas obtidas restam estanques. Isso caracteriza um nível técnico de conhecimento sobre estas medidas-resumo. Uma análise dos dados demandaria uma associação entre média, desvio-padrão e coeficiente de variação (variabilidade em torno da média) e entre a mediana e os quartis (variabilidade em torno da mediana), associando, ainda, em cada um dos casos, a medida da amplitude total da amostra. Como o enunciado não sugere essas relações, elas devem ser feitas espontaneamente pelo aluno, o que caracteriza um nível disponível de conhecimento sobre estas medidas. Se o aluno da graduação não associá-las espontaneamente, o professor pode sugerir a associação em vários níveis de perguntas (se você associar essas medidas, quais podem explicar melhor o comportamento dos preços? Ou em um nível mais direto: que medidas podem ser associadas entre si? Qual o significado dessas associações). Note-se que a “dica” dada pelo professor não pode dar ao aluno da graduação a estratégia para resolver o problema proposto, mas apenas indicar caminhos para superação de dificuldades do aluno. Se, com esta indicação do professor, o aluno fizer as associações e interpretações adequadamente, então podemos inferir que ele está em um nível mobilizável de funcionamento dos conhecimentos sobre as medidas-resumo.

A teoria de Robert é comparada por Skemp (1978 *apud* Carvalho & César 2001, p. 7). As autoras abordam o conhecimento como sendo *instrumental e relacional.* O primeiro é quando se denomina uma coleção isolada de regras e algoritmos apreendidos por meio da repetição e da rotina. Quando o conhecimento que o sujeito possui é desse tipo, só se consegue resolver um conjunto limitado de situações, em contextos semelhantes. O segundo é aquele em que o aluno construiu um esquema do conceito que pode ser atualizado sempre que novas situações lhe exijam, ou seja, um conhecimento que consegue mobilizar em face de novas situações.

Segundo as autoras, os objetivos a serem atingidos com o estudo da Estatística deveriam proporcionar aos alunos um conhecimento relacional, e não apenas um conhecimento instrumental dos conceitos. Este fato não é, contudo, um problema que apenas diz respeito à Estatística, como afirmam Sfard e Linchevski (1994a, p. 203): “Professores e pesquisadores queixam-se freqüentemente de que a compreensão que os alunos têm da álgebra é meramente instrumental: as crianças são capazes de ‘avançar nos passos necessários’, mas não são capazes de explicar aquilo que estão por fazer”.

Gal (2002) aborda em seu trabalho que a alfabetização estatística pode ser entendida por alguns como a denotação de conhecimentos mínimos de procedimentos e conceitos estatísticos básicos. Ainda, segundo o autor, o termo alfabetização, quando usado como parte da descrição da capacidade das pessoas para comportamento orientado a metas, é um domínio específico, não só sugestiona um amplo agrupamento de conhecimento efetivo e certas habilidades formais e informais, mas também de convicções desejadas, hábitos de pensamento ou atitudes como também consciência geral e uma perspectiva crítica. Para o autor, o conjunto de procedimentos estatísticos que um indivíduo desenvolve na resolução de um problema pode ser categorizado em cinco bases de conhecimentos inter-relacionados, que são: Alfabetização, Estatística, Matemática, Contexto Global e Contexto Crítico.

Gal (1995, p. 97) discute que é desejável para leitores de relatórios estatísticos saberem que as médias e medianas são modos simples para resumir um conjunto de dados e apresentar o seu “ponto central”; essas médias são afetadas por valores extremos, mais do que as medianas; e isso pode causar a má estruturação do ponto central quando a distribuição ou formação dos dados ou modelos, nos quais eles são calculados, não são representativos da população inteira sob o estudo que estiver sendo realizado. Mais amplamente, é útil para estes alunos estarem atentos ao fato de que tipos diferentes de índices sumários aparentemente simples, como, por exemplo, porcentagem e mediana, podem representar algo diferente, às vezes conflitando visões de um mesmo estudo.

Tratando-se de exibições gráficas ou tabulares, acreditamos que os alunos egressos do Ensino Superior já deveriam saber que os dados podem ser apresentados ou reportados por ambas as representações, visto que estes mesmos alunos já passaram pelo ensino da disciplina, tanto no Ensino Fundamental quanto no Médio. Estas formas de representar os dados coletados, gráficos e tabelas servem para organizar as informações possibilitando novas descoberta e também comparações de tendências a partir dos dados depois apresentados. Neste sentido, espera-se que, primeiramente os alunos da graduação possam realizar uma leitura literal dos dados em tabelas ou gráficos, familiarizem-se com padrões de conversões usados na criação de gráficos e atentem para simples violações em tais conversões.

Também é esperado que os alunos possam fazer, em algum nível, o que Curcio (1987) e Wainer (1992) chamam “leitura entre os dados” e “leitura por trás dos dados”, projeções que podem ser ‘feitas a partir de um dado recebido, e em que se deve observar todos os padrões passados, e não pontos específicos em um gráfico ou tabela (Gal, 1995). Os alunos também devem perceber que gráficos e tabelas podem trazer dados diferentes (e possivelmente conflitantes) de visões sobre um assunto sob investigação. Finalmente, os alunos devem estar atentos para o fato de que gráficos podem ser intencionalmente criados para enganar, realçar ou até mesmo esconder uma tendência especifica ou uma diferença.

Esperamos que a maioria dos alunos tenha um entendimento sobre algum modo típico para resumir dados, usando médias, medianas, porcentagens ou gráficos. Porém, como há diferentes definições usadas para coletar dados – como os processos de amostragem –, e processos aleatórios podem ser envolvidos, os alunos também precisam possuir algum senso de como analisar os dados e as conclusões obtidas, estando atentos a diagnosticar problemas pertinentes a estas situações.

2.3 METODOLOGIA E PROCEDIMENTOS

O campo da Educação Estatística abrange elementos quantitativos e qualitativos, devido a sua natureza dos fenômenos em estudo. A abordagem quantitativa permite ao pesquisador verificar o tipo de conhecimento que os alunos conseguem mobilizar e procurar a existência de padrões na suas respostas. Por outro lado, a abordagem qualitativa possibilita que se compreenda profundamente o papel que as interações sociais podem desempenhar, como elemento facilitador da passagem de um conhecimento técnico para um disponível, tal qual é definido e recomendado pelos atuais componentes curriculares. Essas interações sociais vêm ao encontro de que os alunos carregam consigo na sua experiência educacional, ou seja, cada aluno tem conhecimentos, vivências, sentimentos e expectativas diferentes quando é confrontado com um mesmo tipo de tarefa.

Nosso trabalho visa diagnosticar a mobilização dos conceitos estatísticos de base, sem que influenciemos nas possíveis interações sociais (diálogos entre as participante–duplas) que acontecerão durante a aplicação da atividade, pois é fundamental para o pesquisador respeitar a forma de resolução dos entrevistados, deixando-os livres para utilizar seus próprios raciocínios.

Para responder a nossa questão de pesquisa, será proposta uma situação-problema na forma de uma atividade diagnóstica qualitativa, dividida em três etapas. Esperamos que esta atividade permita diagnosticar qual o nível de funcionamento dos conceitos, segundo os preceitos de Robert (1998) e, especificamente, aqueles ligados ao estudo da variabilidade, por parte dos alunos do Ensino Superior, e, também, permitir identificar o(s) possível(is) erro(s) cometido(s) por estes alunos.

Acreditamos que as situações-problema, quando apresentadas em várias formas de representação, abrem um leque maior de possibilidades de investigar as possíveis dificuldades que os alunos universitários poderão encontrar na resolução de qualquer atividade diagnóstica. Sendo assim, as etapas da atividade se apresentarão em diferentes níveis de complexidade, para que, então, seja possível verificar o nível de mobilização dos conceitos (Robert, 1998) em que estes alunos se encontram. A situação-problema será aplicada para 3 duplas de alunos que já cursaram a disciplina de Estatística, oferecida no curso de Administração, em um tempo estimado de 200 minutos (4horas/aula).

A atividade será entregue para apenas um integrante de cada dupla, para que forcemos cada dupla a dialogar entre si sobre as diversas formas de resolução que ambos podem encontrar, e, assim, poderemos investigar com maior clareza o pensamento estatístico de cada dupla e, por fim, registrar suas possíveis dúvidas para futura análise. Também proporemos o registro do diálogo entre o pesquisador e o aluno durante a resolução, assim como a produção escrita da atividade. As várias formas de procedimentos desenvolvidos por parte das duplas e a forma de encadear o saber estatístico poderão nos permitir analisar o nível de mobilização dos conhecimentos estatísticos de base, baseados nos preceitos de Robert (1998).

Nossa atividade está dividida em três partes com diferentes formas de apresentação. A primeira parte será composta de um banco de dados fictício, no qual constarão idade e renda mensal de quarenta pessoas entrevistadas por uma Empresa de cartões de crédito. Na segunda parte, são apresentadas duas distribuições na forma de tabelas, sendo a primeira sem intervalo de classes e a segunda com intervalo de classes. Na primeira distribuição, relata-se o comportamento de quantidade de carros por número de pessoas, e a segunda distribuição relata o comportamento do tempo no trânsito por número de pessoas.

E, por fim, a terceira parte relata as tabelas da atividade da segunda parte, em forma de gráficos. Nossa idéia em apresentar a atividade em diferentes formas de representação nos ajudará, além do que nos propomos a investigar, é a de identificar se os níveis de dificuldades também se apresentam em diferentes formas. O que queremos dizer é que, se um conjunto de dados apresentados ao aluno, estiver na forma de dados brutos (sem nenhum tipo de organização crescente ou decrescente), as dificuldades serão maiores do que se estes mesmos dados forem apresentados com algum tipo de forma organizacional, por exemplo, na forma de gráficos ou tabelas.

As respostas obtidas nesta atividade serão organizadas, tabuladas por categorias e, posteriormente, analisadas segundo os preceitos de Robert (1998). As categorias que acabo de citar nos ajudarão a identificar qual(is) a(s) possível(is) dificuldade(s) que os alunos apresentarão na resolução da atividade, quais sejam:

- *em relação à média:*

  a. os alunos analisarão a média isoladamente, interpretando corretamente seus resultados;

  b. os alunos analisarão a média isoladamente, porém interpretando seus resultados como valor central da distribuição (concepção errônea, mas muito encontrada);

c. os alunos analisarão o desvio-padrão isoladamente, interpretando corretamente seus resultados como uma medida de dispersão ao redor da média;

d. os alunos relacionarão média e desvio–padrão, interpretando seus valores como indicadores de variabilidade em relação à média;

e. os alunos relacionarão média e desvio–padrão, mas sem interpretar a variabilidade;

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

- *em relação à mediana:*

a. os alunos analisarão a mediana isoladamente, interpretando corretamente seus resultados como valor central de uma distribuição;

b. os alunos analisarão os quartis isoladamente, interpretando corretamente seus resultados como uma medida separatriz;

c. os alunos relacionarão mediana e quartis, interpretando seus valores como indicadores de variabilidade em relação à mediana;

d. os alunos relacionarão mediana e quartis, mas sem interpretar a variabilidade;

Esperamos que, por meio desta atividade, possamos atingir nossos objetivos e assim colaborarmos para que o ensino da disciplina de Estatística seja vista com olhar preocupante, e que as possíveis dificuldades apresentadas pelos alunos não sejam mais encaradas como “rotineiras”, bem como as pesquisas futuras apresentem quadros satisfatórios de aprendizados.

2.4 PESQUISAS QUE ABORDAM O TEMA

Algumas pesquisas voltadas para a(s) dificuldade(s) de aprendizagem da Estatística no curso superior contribuíram e continuam contribuindo para a consolidação da importância de tal estudo. Entre elas, podemos citar Silva (2000), Vendramini (2000), Novaes (2004) e Batanero (2001).

Silva (2000) pesquisou cerca de 643 estudantes das áreas de Ciências Humanas e Exatas de uma universidade particular e apresentou estudos indicando que os alunos apresentavam atitudes negativas em relação à Estatística e desenvolviam ansiedade em relação à 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 uma construção, por parte dos alunos, dos conceitos básicos da disciplina. 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.

A colaboração que Silva (2000) trouxe para o nosso trabalho visa à preocupação na elaboração da atividade que permite diagnosticar os níveis de mobilização por parte dos alunos do Ensino Superior na disciplina Estatística. Tentaremos elaborar uma atividade que seja voltada a aspectos próximos à realidade desses alunos e que desperte neles o interesse de se chegar a uma análise conclusiva da atividade por meio de estratégias próprias. Por outro lado, a hipótese de os alunos do Ensino Superior apresentarem grandes níveis de dificuldade na resolução de problemas, envolvendo cálculos estatísticos, também se tornará fator importante na construção da atividade, já que os resultados apresentados no trabalho de Silva (2000) apontaram atitudes negativas em relação à Estatística. A elaboração de situações-problema, não sendo cuidadosamente construída de uma forma que desperte interesse por parte dos alunos, poderá ser um obstáculo para eles.

Novaes (2004), em seu trabalho, investigou a mobilização dos conceitos estatísticos de base em alunos do curso superior de Turismo. A autora concluiu que os erros cometidos por esses alunos foram os que envolviam processos algébricos; erros nos conceitos de média, moda e mediana; análise inadequada da variabilidade dos dados e dificuldades oriundas de obstáculos epistemológicos e didáticos na resolução de situações-problema. A autora também observou que os erros se repetem e produzem regularidades, sendo que os sujeitos investigados apresentavam, em sua maioria, as mesmas dificuldades. Neste caso, a autora propôs em seu trabalho o ensino dos conteúdos estatísticos com pelo menos uma ordenação, na forma de situações-problema do que se pretende ensinar, de forma a facilitar a construção dos conceitos por parte dos alunos, bem como a sua adequada utilização.

O trabalho de Novaes (2004) é de grande valia para a nossa pesquisa. A investigação da mobilização dos conceitos estatísticos em um curso de Turismo feito pela pesquisadora será nosso referencial para futuras análises e, conseqüentemente, trará à tona nossas suspeitas das dificuldades em curso de Administração. Pretendemos com esse trabalho reforçar a necessidade de uma atenção especial à metodologia do ensino da Estatística, mas, também, proporcionar recursos teóricos para futuras pesquisas na área, que sofre de enorme carência.

De acordo com os trabalhos citados, os alunos, tanto do ciclo básico como do Ensino Superior, associam as dificuldades que tiveram com a Matemática às atuais, com a disciplina Estatística. Quando reconhecem a importância dessa área do saber para a resolução de situações-problema da vida profissional, ficam ansiosos ao sentirem dificuldades, o que os leva a atitudes negativas. Podemos verificar também, nesses trabalhos, a preocupação de se elaborar situações-problema de uma forma organizada, no processo de aprendizagem da Estatística, auxiliando na identificação de possíveis erros por parte dos alunos e compreendendo a origem de tais erros, bem como fazendo um estudo de novas propostas para uma forma correta de abordagem do conteúdo da disciplina.

Batanero (2001) realizou investigações sobre erros e dificuldades na compreensão dos conceitos estatísticos de base. A autora afirmou que grande parte da investigação teórica e experimental, realizada atualmente em Didática da Matemática, mostrou que os alunos produzem respostas erradas ou simplesmente não são capazes de dar nenhuma resposta, quando se pede para realizarem certas tarefas. Nos casos em que não se tratava de falta de atenção dos alunos, os professores acreditavam que a tarefa era considerada de nível elevado de dificuldade para eles, porém esses erros não aconteciam de forma aleatória, imprevisível.

Com freqüência, é possível encontrar regularidades, associações com variáveis próprias das atividades propostas, dos sujeitos ou circunstâncias presentes ou passadas. São inúmeros os motivos que provocam essas regularidades, por isso é importante que o professor reconheça essas estabilidades e regularidades quando forem manifestadas por seus alunos. No caso da probabilidade e da estatística, é importante analisar o raciocínio dos alunos, visto que tratamos com idéias abstratas e não tão ligadas à experiência escolar, como foram os conceitos matemáticos.

Podemos perceber que a natureza da estatística é diferente da cultura determinista da matemática. Os indicadores disso são as controvérsias filosóficas sobre a interpretação e a aplicação de conceitos básicos como os de probabilidade, aleatoriedade, independência ou contraste de hipótese, já que estas controvérsias não são comuns no campo da álgebra ou geometria. Por isso, faz-se necessário o trabalho da Estatística desde as séries iniciais como sugerem os Parâmetros Curriculares Nacionais (PCN) e, quanto mais tarde trabalharmos a formação dos conceitos estatísticos, mais dificuldades os alunos terão para entender a aleatoriedade e a variabilidade contida nos fenômenos do cotidiano, que permita uma leitura correta do mundo.

Batanero (2001) também chamou a atenção para a formação do professor, tanto de Matemática como o de Estatística. No Brasil, quando o tema é abordado em palestras ou congressos, nem sempre é enfocada a formação crítica do profissional ou mesmo do aluno de todos os níveis, e, embora tenhamos excelentes livros-textos que tratam do assunto, a investigação didática está começando a mostrar como alguns erros conceituais e pedagógicos estão presentes com freqüência nestes mesmos livros.

3. PARTE EXPERIMENTAL

Na disciplina Estatística, para alunos do Curso Superior – gestão Administração de Empresas, levanta-se a hipótese de que a maioria deles tenha uma visão parcial da utilidade desta ferramenta, restringindo sua aplicação, ou à organização de dados numéricos de uma amostra, ou a cálculos de média aritmética, desvio–padrão, porcentagens ou à elaboração de gráficos, raramente considerando de maneira global a disciplina e suas aplicações.

Na atividade proposta, os alunos irão analisar uma situação-problema fictícia nos moldes da atividade profissional da área da Administração de Empresas. Conforme especificado anteriormente, o objetivo do nosso trabalho é o de verificar em que nível os alunos estão utilizando os conhecimentos de Estatística adquiridos no curso que freqüentaram, ou seja, quando tais conhecimentos são técnicos, mobilizáveis ou disponíveis para o aluno, segundo os níveis estudados por Robert (1998).

3.1 ATIVIDADE PROPOSTA

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

Tabela 2: Idade e Renda mensal de 40 pessoas (dados fictícios).

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:

Quantidade de carros	*Número de pessoas*
1	10
2	25
3	15
4	5
Total	***55***

Tabela 3: Quantidade de carros/pessoa (dados fictícios)

Tempo no trânsito	*Número de pessoas*
0 ├ 2	10
2 ├ 4	20
4 ├ 6	30
6 ├ 8	15
Total	***75***

Tabela 4: N. de horas no trânsito/pessoa (dados fictício)

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 1: número de carros/pessoa

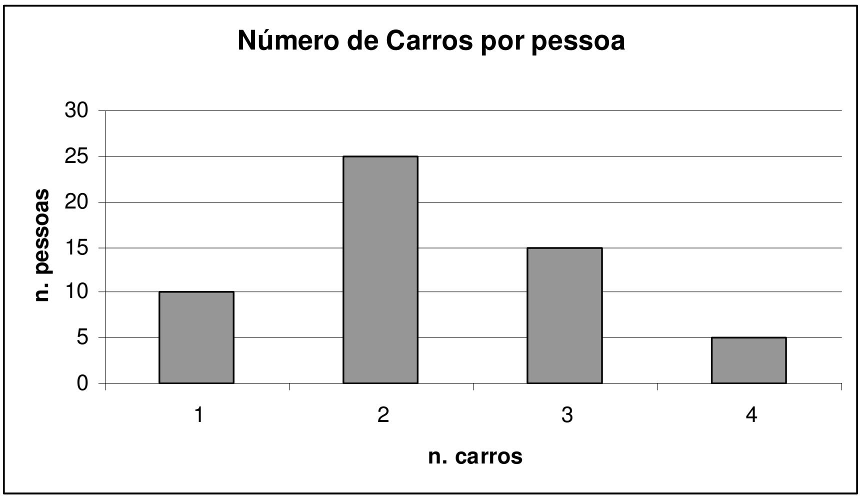
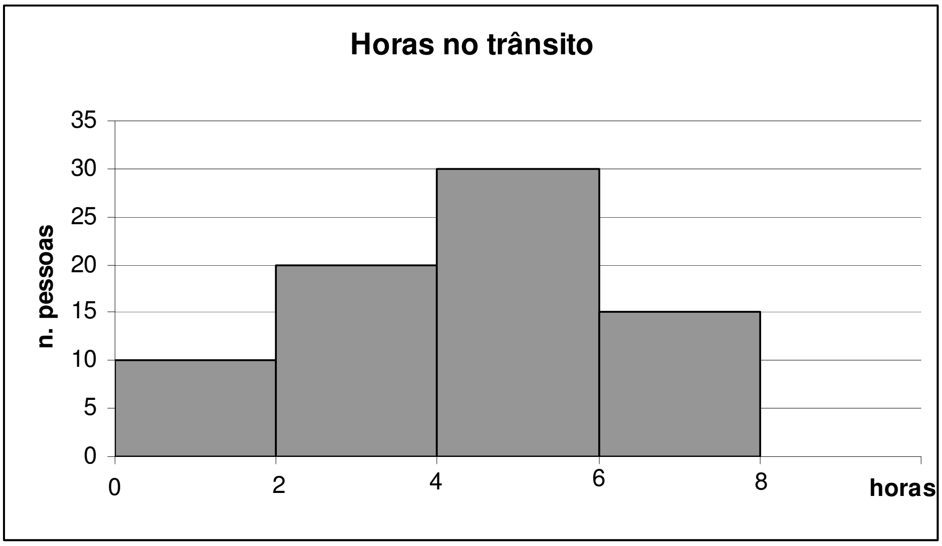


Gráfico 2: horas no trânsito/pessoa.



3.2 ANÁLISE *A PRIORI*

As variáveis didáticas que abordaremos são:

a) tamanho da amostra: é necessário que o aluno saiba exatamente o número de elementos da sua população-amostra; se necessário trabalhar com todos os elementos da amostra ou apenas uma parte dela e, também, o que essa escolha poderá acarretar no resultado final. O tamanho da amostra será o valor que estará envolvido praticamente em todos os cálculos na análise exploratória;

b) amplitude da amostra (diferença entre o maior e o menor valor dos elementos da amostra): tal representação pode sinalizar ao aluno, na primeira parte da atividade, a necessidade de um tratamento da variável como sendo discreta, para otimizar o trabalho de análise deste conjunto de dados;

c) apresentação dos dados: a forma da apresentação dos dados oferecidos sugere tratamento unidimensional das variáveis envolvidas no problema. Na primeira parte da atividade proposta, ela se apresenta de forma bidimensional, deixando ao aluno a escolha da preferência da ordem da variável a tratar;

d) disposição dos alunos: 3 duplas de alunos em diferentes estágios no curso. Na observação de dados coletados, há sempre um objetivo à vista, ou, simplesmente, exploramos os dados para ver o que eles nos revelam. Podemos aplicar muitas técnicas apreendidas num curso de Estatística para esta análise, mas é importante lembrar que, nesta exploração, devemos relacionar três características: (1) natureza ou forma de distribuição; (2) um valor representativo; e (3) uma medida de variação. É imprescindível considerar a distribuição dos dados, porque ela pode afetar não só os métodos estatísticos a serem usados, como também as conclusões a que chegamos.

Assumindo que os alunos na graduação, que já passaram pela disciplina Estatística, entendam por que e como os dados são produzidos, eles precisam estar familiarizados com conceitos básicos e apresentações de dados que são comumente usados para divulgar resultados ao público-alvo. Dois tipos-chaves de conceitos, cuja centralização é notável por muitos são, “por cento” (Parker & Leinhardt, 1995, p. 435) e medidas de tendência central, principalmente a média e a mediana.

Primeiro, os alunos precisam estar atentos à possibilidade de diferentes erros ou preconceitos (em amostragem, medida, conclusão) e manterem uma preocupação saudável quanto à estabilidade e generalidade dos resultados. Segundo, é útil perceber que erros podem ser controlados pelo próprio modelo do estudo, podendo ser calculados e descritos, por exemplo, por meio de declarações de probabilidade. Um conceito mencionado, pelos meios de comunicação, nesta situação, é “margem de erro”. Terceiro, é útil saber que há modos para determinar o significado ou a “veracidade” de uma diferença entre grupos, mas isto requer atenção ao tamanho dos grupos estudados, para a qualidade do processo de amostragem e a possibilidade de que uma amostra seja parcial.

Finalmente, é importante estar atento a que diferenças encontradas ou tendências podem existir, mas não, necessariamente, podem ser grandes ou estáveis o bastante para serem consideradas importantes, ou podem ser causadas por processos de casualidade.

3.2.1 RESOLUÇÃO DA ATIVIDADE

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.

Pediremos aos alunos para realizarem uma análise exploratória dos dados apresentados na **tabela 2**, com o objetivo, conforme já especificado, de diagnosticar em que nível de conceitualização, segundo Robert (1998), os alunos se encontram em relação aos conceitos estatísticos de base, já vistos no curso.

A resolução da atividade será feita por meio de tabela unidimensional, ou seja, trataremos cada variável separadamente.

Variável idade (ROL):

Tabela 5: 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

1ª Estratégia de resolução: Cálculo das medidas a partir do ROL

Para o cálculo da média, usaremos a fórmula: $\overline{x} = \sum_{i}^{n} \frac{x_i}{n}$, neste caso, o valor da média será: $\overline{x} = \frac{1364}{40} = 34{,}1 \ \ anos$;

A variância será denotada pela fórmula: $s^2 = \frac{\sum (x - \overline{x})^2}{n-1}$; logo o valor da variância será $s^2 = \frac{2915{,}6}{39} = 74{,}76$;

O desvio padrão será a raiz quadrada da variância: $\sqrt{s} = 8{,}65$;

O valor mínimo da amostra: 23 anos;

O valor máximo da amostra: 65 anos;

Amplitude da amostra: 42 anos;

Nas medidas separatrizes temos:

O primeiro quartil (Q1): 29 anos;

A mediana (md): 30 anos;

O terceiro quartil (Q3): 40 anos;

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

A análise esperada pelos alunos:

Na amostra, a média é de 34 anos aproximadamente e o desvio–padrão de aproximadamente 8,5 anos, tendo uma amplitude amostral de 42 anos. Percebemos que o coeficiente de variação em torno da média é de aproximadamente 25%, considerada alta. A amplitude em torno da média é de 17 anos, podendo ser representada como mostra a figura 1:

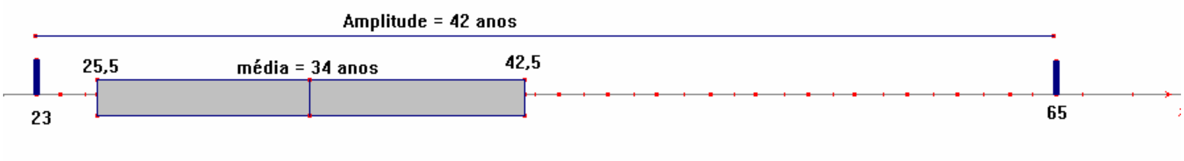


Figura 1: amplitude em torno da média

Percebemos, pela **figura 1**, a existência de uma maior concentração à esquerda de toda amostra em torno da média, ou seja, uma concentração que varia entre 25,5 anos a 42,5 anos em torno de uma média de 34 anos, representando 40% do total da amostra. Significa que, na amostra, de cada 100 pessoas entrevistadas, 40 estão entre 25 e 42 anos de idade. Uma análise geométrica dos dados coletados apresenta uma melhor noção do comportamento da amostra. As concepções erradas sobre média apresentam-se de várias formas pelos alunos do Ensino Superior. Entre elas, está a concepção de que média é atrelada a ponto central, ou seja, 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, p. 87) 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 aluno do Ensino Superior 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 de posição: Quartis.

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

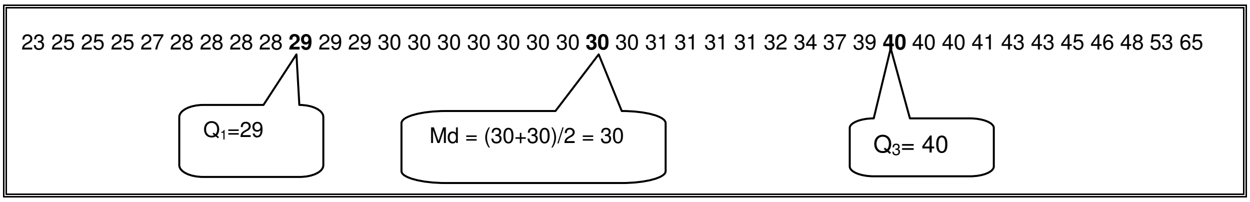


Figura 2: medidas de posição

A variável *idade* tem uma melhor distribuição quando trabalhada as medidas de posição. Logo, para esta análise, a melhor medida que explicaria o comportamento da variável *idade* seria a mediana. Verificaremos isso por meio do Box-plot.

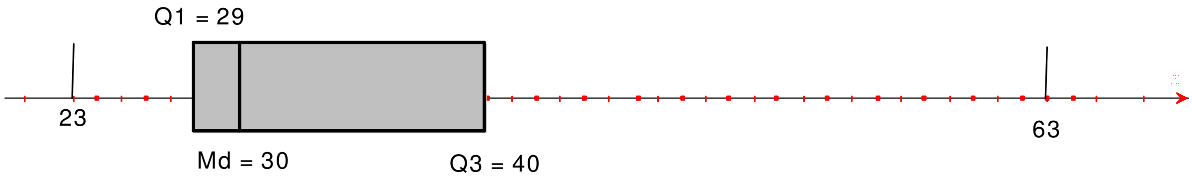


Figura 3: box-plot

Podemos analisar pelo Box-plot[3] (**figura 3**) que, para cada quartil obteremos 25% da amostra. Nestes moldes, verificamos que há uma maior concentração dos dados entre 23 e 30 anos, isto é, 50% da amostra está exatamente entre 23 e 30 anos. Esperamos que o aluno do Ensino Superior tenha essa percepção de que a análise desta variável é mais representativa por meio da mediana. Na análise dos níveis de mobilização citada por Robert (1998), o aluno, em um nível técnico, não relacionará as medidas encontradas e, provavelmente não conseguirá fazer uma análise dos dados por meio das representações geométricas que acabamos de mostrar. Caso aconteça o inverso, ou seja, esta análise relacional entre as medidas ocorrerem e o aluno conseguir fazer uma conclusão coerente sobre os resultados, segundo Robert (1998), este aluno estará em um nível disponível. Por outro lado, se esta análise ocorrer por intervenção do professor, no sentido de mostrar o caminho a ser percorrido e não fornecendo a resposta, este aluno estará em um nível mobilizável (Robert, 1998).

---

[3] Box-plot is a way of summarizing a set of data measured on na interval scale. It is often used in exploratory data analysis. It is a type of graph which is used to show the shape of the distribution, is central value, and variability. The picture produced consists of the most extreme values in the data set (maximum and minimum values) the lower and upper quartiles, and the median. (dfinicion taken from Valerie J. Easton and John H. McColl's Statistics Glossary V1.1)

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 do seguinte recurso: $\cong \sqrt{n}$.

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

Número de classes: $K = \frac{H}{\sqrt{n}}$, (H é a amplitude da amostra e ***n*** é o número de elementos da amostra). Então temos: $K = \frac{42}{\sqrt{40}} \Rightarrow K = \frac{42}{6{,}32} \Rightarrow K = 6{,}64$.

Consideremos 6 classes com intervalo de 7 anos para cada classe

Idade	f_i	Pm = x_i	$X_i.f_i$	F_{ac}	d_i^2 . f_i
23 ├ 30	12	26,5	318	12	867
30 ├ 37	15	33,5	502,5	27	33,75
37 ├ 44	8	40,5	324	35	242
44 ├ 51	3	47,5	142,5	38	468,75
51 ├ 58	1	54,5	54,5	39	380,25
58 ├ 65	1	61,5	61,5	40	702,25
Totais	***40***	***xxxx***	***1403***	***xxx***	***2694***

Tabela 6: Distribuição de freq. com intervalo de classe: Idade

Cálculo das medidas de variação

Calculo da média: $\bar{x} = \sum_{i}^{n} \frac{x_i . f_i}{n} \Rightarrow \bar{x} = \sum_{i}^{40} \frac{1403}{40} = 35{,}075$ anos

Cálculo do desvio–padrão: $s = \sqrt{\frac{\sum (x - \bar{x})^2 . f_i}{n-1}} \Rightarrow s = \sqrt{\frac{2694}{39}} = 8{,}31$ anos.

Na representação geométrica dos cálculos da média e desvio-padrão, podemos observar o seguinte:

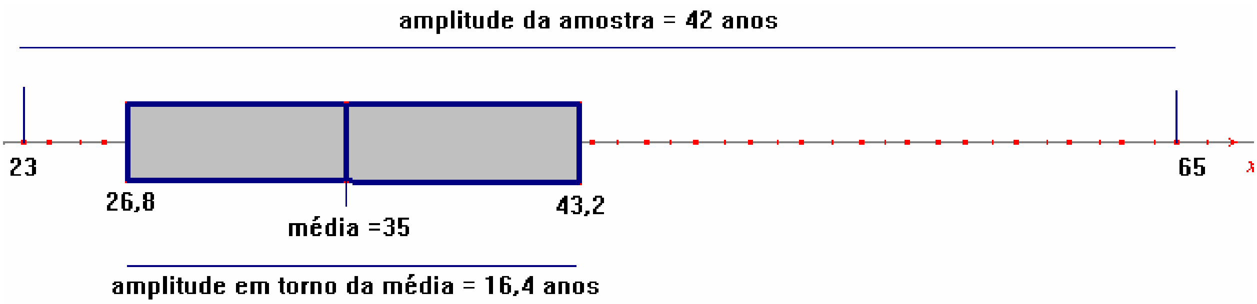


Figura 4: amplitude em torno da média

A representação na **figura 4** aponta para 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 38,6% do total da amostra. Isso significa que na amostra de cada 100 pessoas entrevistadas, podemos esperar que 39 estejam entre 27 e 43 anos de idade. Neste caso, também observamos, por meio da representação geométrica, 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, utilizaremos o método da interpolação linear:

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

$$\begin{array}{ccc} 30 & mediana & 37 \\ \vdots & \vdots & \vdots \\ \frac{12}{40} & 0{,}5 & \frac{27}{40} \end{array} \quad \Rightarrow \quad \frac{med - 30}{0{,}5 - \frac{12}{40}} = \frac{37 - 30}{\frac{27}{40} - \frac{12}{40}} \quad \Rightarrow mediana = 33{,}73\ anos$$

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

$$\begin{array}{ccc} 23 & Quartil\ 1 & 30 \\ \vdots & \vdots & \vdots \\ \frac{0}{40} & 0{,}25 & \frac{12}{40} \end{array} \Rightarrow \frac{Q1-23}{0{,}25-\frac{0}{40}} = \frac{30-23}{\frac{12}{40}-\frac{0}{40}} \Rightarrow Quartil\ 1 = 28{,}83\ anos$$

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

$$\begin{array}{ccc} 37 & Quartil\ 3 & 44 \\ \vdots & \vdots & \vdots \\ \frac{27}{40} & 0{,}75 & \frac{35}{40} \end{array} \Rightarrow \frac{Q3-37}{0{,}75-\frac{27}{40}} = \frac{44-37}{\frac{27}{40}-\frac{35}{40}} \Rightarrow Q3 = 39{,}63\ anos$$

Os dados calculados, expressos no Box-plot, nos trará a representação geométrica da distribuição. Veja:

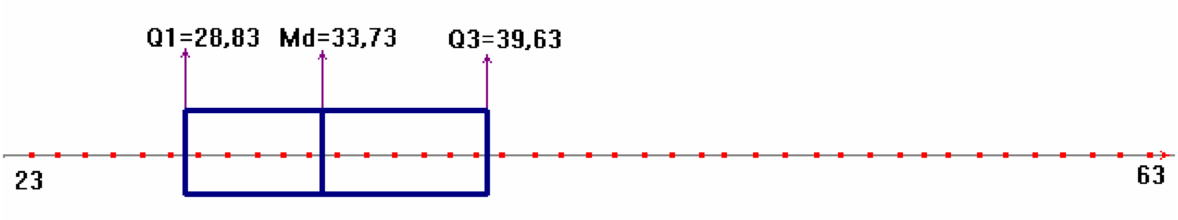


Figura 5: representação dos quartis

Podemos analisar, pelo Box-plot (**figura 5**), que, para cada quartil, obteremos 25% da amostra. Porém 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 será a critério do aluno, mas vale ressaltar que, na nossa análise, *a posteriori*, teremos que diagnosticar se o aluno tem a consciência de que a escolha da segunda estratégia mostra apenas uma estimativa dos resultados da análise. Nestes moldes, percebe-se maior concentração dos dados entre 23 e 33,73 anos (próximos dos 34 anos), isto é, 50% da amostra está exatamente entre 23 e 33,73 anos.

3ª estratégia: por meio de gráfico.

Utilizaremos a tabela da segunda estratégia para elaboração do gráfico, levando em consideração que o Histograma é o tipo de gráfico que mais se utiliza no ensino da análise exploratória de dados.

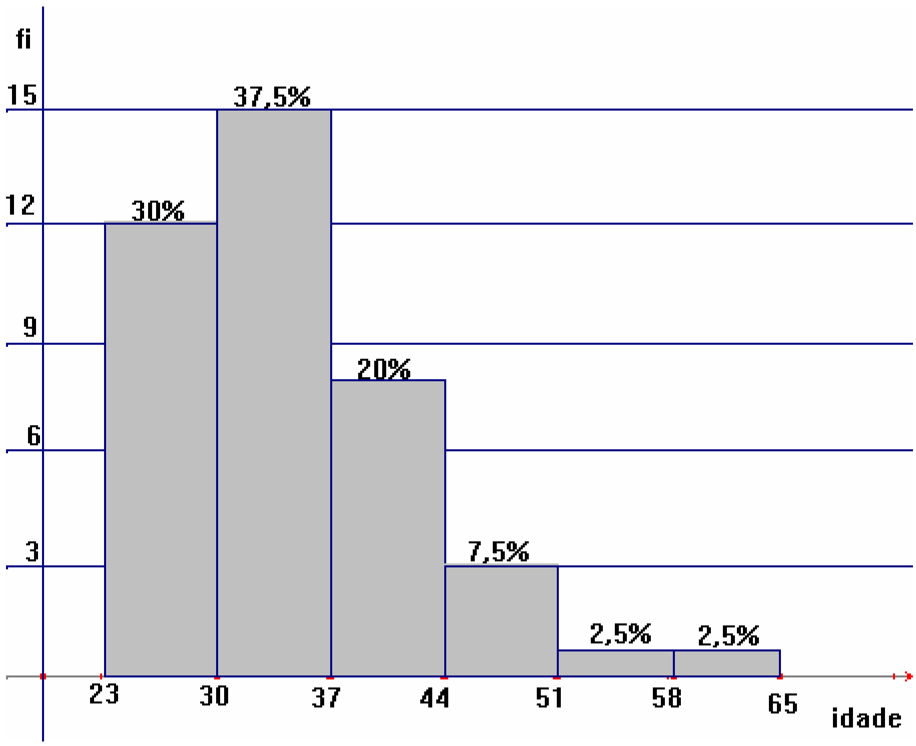


Gráfico 3: Histograma da variável idade

O histograma consiste em retângulos contínuos com base nas faixas de valores da variável e com área igual à freqüência relativa da respectiva faixa. Dessa forma, a altura de cada retângulo é denominada densidade de freqüência ou simplesmente densidade definida pelo quociente da área pela amplitude da faixa. Para a variável *idade*, as densidades de cada faixa podem ser obtidas dividindo-se a coluna f_i da **tabela 6** por 40, que é o total de elementos da amostra. Para facilitar a interpretação, colocamos em cada retângulo o valor percentual.

Usaremos esta terceira estratégia para o cálculo das medidas de variação e separatrizes.

Para o cálculo das medidas de variação, representaremos como mostra a figura a seguir:

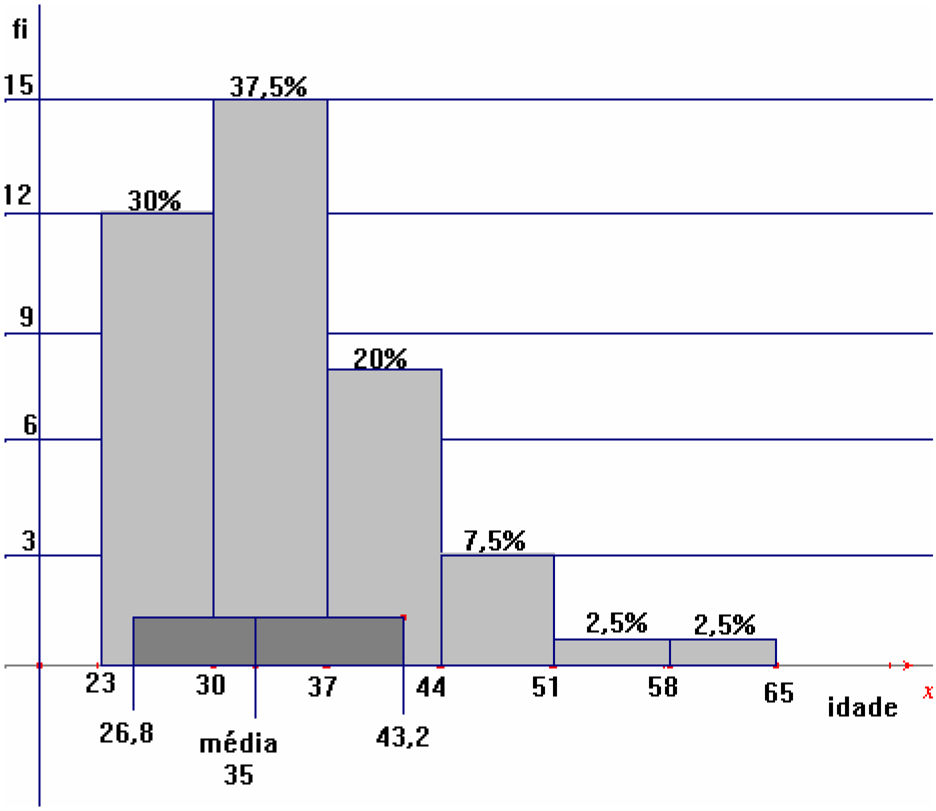


Gráfico 4: Estudo da média por meio do histograma

Este tipo de representação também mostrará a idéia de variação em torno da média, e deixa visivelmente clara a relação da amplitude desta variabilidade com a amplitude da amostra. Os resultados possíveis e esperados são igualmente mostrados nas estratégias anteriores, mais ainda a que aborda a estratégia 2.

Já para os cálculos das medidas separatrizes, mostraremos o processo de como se chegar aos resultados. Calculemos a mediana da variável *idade* por meio do histograma. Inicialmente identificamos o retângulo que deve conter a mediana. Uma simples soma das áreas resulta que a mediana pertence ao intervalo [30;37[, uma vez que até o valor 37 temos acumulado 67,5% (30% +37,5%) das observações dentro dessa faixa, precisamos determinar um retângulo com área igual a 20%, pois é o valor que falta para atingir 50% (30% + **20%** = 50%). Veja a ilustração abaixo, na qual o retângulo procurado está marcado com área mais escura, e os procedimentos finais para o cálculo.

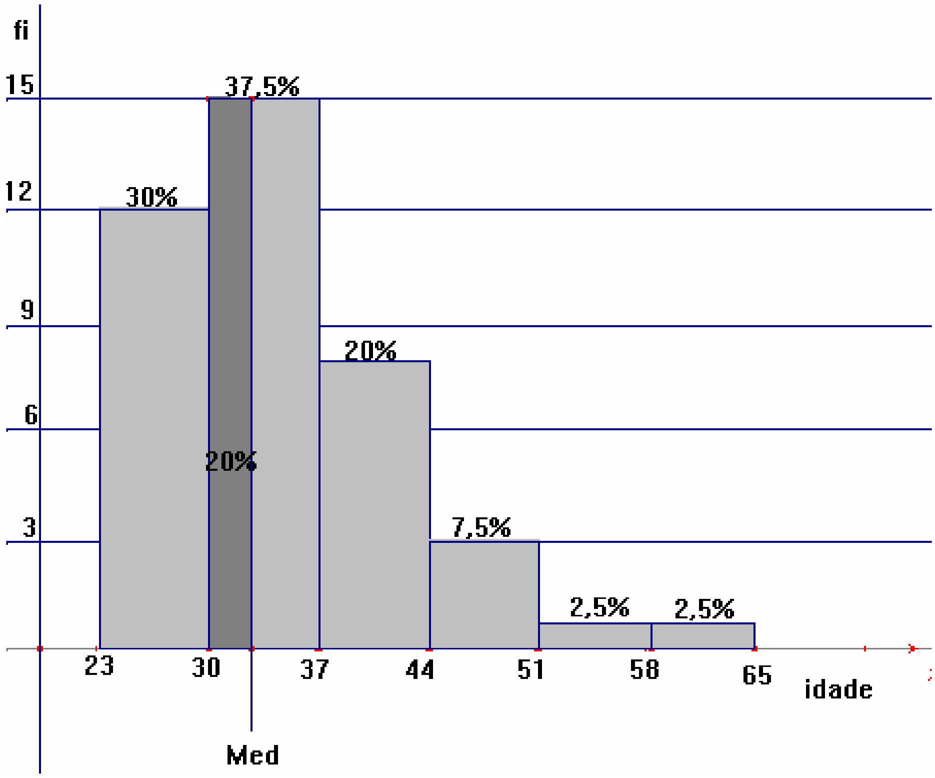


Gráfico 5: Estudo da mediana por meio do histograma

Com uso de proporções, estabelecemos a seguinte igualdade:

$$\frac{med - 30}{0{,}20} = \frac{37 - 30}{0{,}375} \Rightarrow md = 33{,}73$$

Este cálculo pode ser generalizado para situações em que o conjunto de dados é dividido em mais subgrupos. Um caso importante é aquele em que dividimos o conjunto de dados em quatro subgrupos. 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. O cálculo dos valores dos quartis poderá ser realizado de forma semelhante àquela descrita para a mediana, isto é, por meio do histograma. Tomaremos a liberdade de não mostrar os procedimentos para tal cálculo por ser análogo ao anterior e, focalizando a investigação dos níveis de funcionamento, os processos mostrados pela mediana serão suficientes para tal diagnóstico. É importante ressaltar as diversas formas de abordagem de um mesmo assunto para que, em nossa pesquisa, possamos diagnosticar os possíveis erros cometidos por alunos do Ensino Superior nas mais diversas formas de resolução, que poderão surgir na aplicação da atividade.

Variável renda familiar (ROL)

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

Tabela 7: Renda familiar (dados fictícios)

Para o cálculo da média da renda familiar, usaremos a mesma fórmula: $\bar{x} = \sum_{i}^{n} \frac{xi}{n}$. Neste caso, o valor da média da renda familiar será: $\bar{x} = \frac{30137}{40} = 978{,}42\ reais$;

A variância será denotada pela fórmula: $s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$, logo o valor da variância será $s^2 = \frac{7817761{,}426}{39} \cong 200455{,}4212$;

O desvio-padrão será a raiz quadrada da variância: $\sqrt{s} \cong 447{,}72$;

O valor mínimo da amostra: 300 reais;

O valor máximo da amostra: 1800 reais;

Amplitude da amostra: 1200 reais;

Nas medidas separatrizes temos:

O primeiro quartil (Q1): 588,50 reais;

A mediana será (md): 890 reais;

O terceiro quartil (Q3): 1377,50 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 reais aproximadamente e 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%. Podemos analisar geometricamente a variabilidade em torno da média pela figura 6:

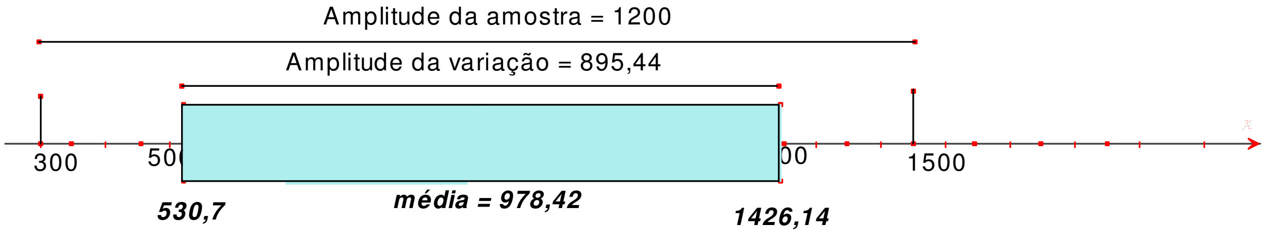


Figura 6: variabilidade em torno da média

Percebemos, pela **figura 6**, a existência de uma concentração uniforme da amostra em torno da média, ou seja, uma concentração em renda mensal que varia entre 530,7 reais e 1426,14 reais em uma média de 978,42 reais, representando 74,62% de toda a amostra. Essa análise geométrica dos dados coletados relata, com maior confiabilidade, o comportamento da amostra e que tipo de distribuição ela apresenta (simétrica ou assimétrica). Neste caso, o aluno poderá definir qual o melhor valor para a análise da variabilidade. Para tanto, vamos fazer um estudo dos quartis, conhecimento este, esperado dos alunos do Ensino Superior.

Utilizamos, para este estudo, o box-plot.

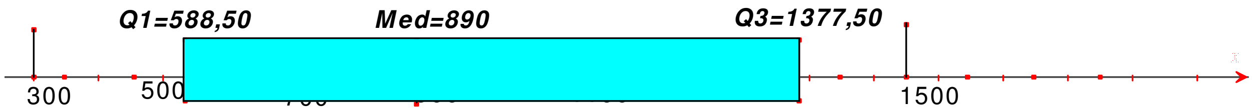


Figura 7: variabilidade em torno da mediana

Como podemos verificar na **figura 7**, o box-plot nos fornece uma análise mais detalhada e com maior precisão do comportamento dos dados coletados, e percebemos que há uma ligeira concentração à esquerda dos dados em relação à mediana. Nota-se uma maior concentração dos valores entre 300,00 reais e 890,00 reais, levando-nos a crer que a mediana é o valor de melhor representa a variabilidade do conjunto. 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 aparente ser um valor representativo. É imprescindível ao aluno do Ensino Superior, ter habilidade e competência de analisar a variabilidade de conjunto de dados, escolhendo bem o valor que melhor representa o conjunto, cercando quase todas as possibilidades, para que suas conclusões futuras sobre o conjunto estudado tenham a confiabilidade que se espera do profissional nas atividades que lhe são atribuídas.

Segunda Parte

Nesta segunda parte, a distribuição apresentar-se-á na forma tabular e o motivo para essa representação é proposital para que possamos diagnosticar se os níveis de mobilização dos estudantes se manifestam por meio de outra forma de representação dos dados. Assim, estaremos diagnosticando o uso pelos alunos de uma diversidade de representações, conforme Vergnaud (1998a). Dessa forma, até por conta da mesma necessidade dos cálculos (média, quartis e desvio-padrão), acreditamos ser mais viável e fácil para o aluno, bastando ele completar a tabela com colunas auxiliares para encontrar os valores pedidos, porém não garantindo o sucesso na atividade.

Tabela 8: distribuição de freqüência - qtde. carros/pessoa

Qtde de carros (xi)	***N. de pessoas (fi)***	$xi.fi$	$di = (xi - \bar{x})$	$di^2.fi$
1	10	10	1 – 2,27 = -1,27	16,129
2	25	50	2 – 2,27 = - 0,27	1,8225
3	15	45	3 – 2,27 = 0,73	7,9935
4	5	20	4 – 2,27 = 1,73	14,9645
Total	***55***	***125***		***40,9095***

Comecemos os cálculos completando a **tabela 3**, fornecida na 2ª parte

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

1) Cálculo da média e desvio-padrão.

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

O desvio padrão será $s = \sqrt{\frac{\sum di^2.fi}{n-1}} \Rightarrow s = \sqrt{\frac{40{,}9095}{54}} = 0{,}87$ carros por pessoa.

A análise esperada pelos alunos:

Nas medidas de variação, temos, na amostra, que a média é de 2 (2,27) carros por pessoa e desvio-padrão de aproximadamente 1 (0,87) carros por pessoa. Tendo uma amplitude amostral de 3 carros, percebemos que o coeficiente de variação em torno da média é de aproximadamente 38,3%. A amplitude em torno da média será aproximadamente de 1 carro. Uma forma de tentar "enxergar" o comportamento da amostra em relação à média seria mostrar uma representação geométrica, que indicaria quais os melhores caminhos a seguir:

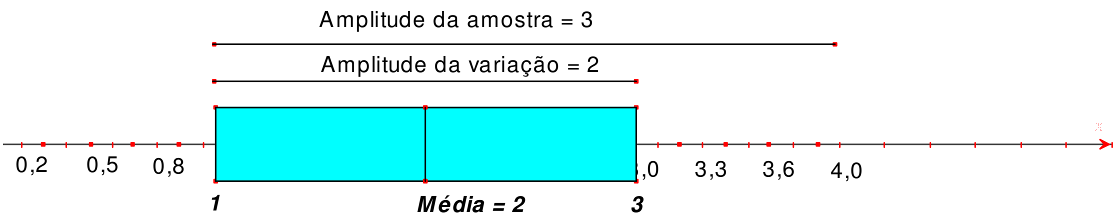


Figura 8: amplitude em relação à média

Por esta disposição, podemos perceber que a dispersão dos dados não é alta, pois a média sendo $\cong 2$ e o desvio-padrão $\cong 1$, comparando com a amplitude total da amostra igual a 3, é fácil perceber que existe uma concentração à esquerda em relação ao conjunto todo e, sendo assim, faz-se necessário um estudo dos quartis.

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

As medidas separatrizes também nos dão uma representatividade do comportamento da amostra, e trarão os dados que parecerão mais adequados para a solução. Para resolver esta questão, utilizamos a seguinte estratégia:

Coloquemos os dados da tabela na forma de rol, dividamos a amostra ao meio e encontraremos a mediana ou Q_2.

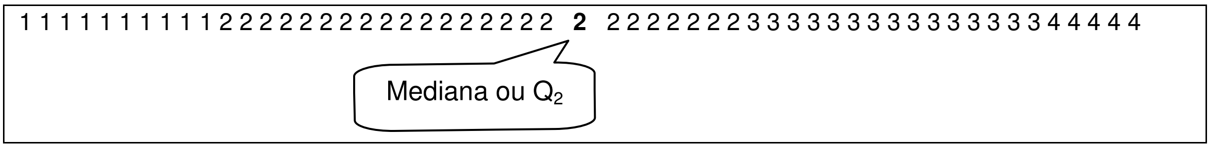


Figura 9: representação geométrica da mediana

Para cada uma das metades (esquerda e direita), dividimos ao meio e encontramos o primeiro quartil Q_1 (valor à esquerda da mediana) e o terceiro quartil Q_3 (valor à direita da mediana).

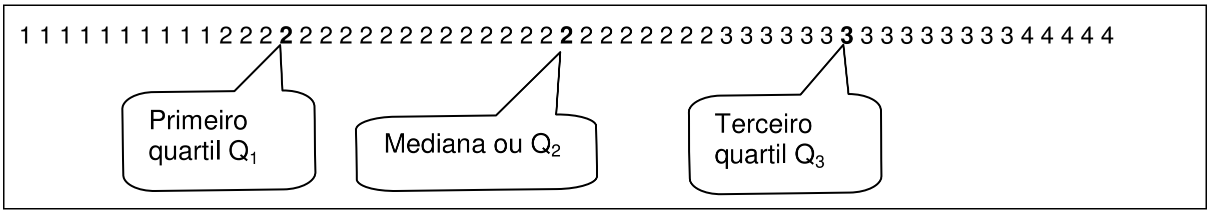


Figura 10: representação geométrica dos quartis

O valor da Mediana é 2; o valor do 1º quartil é 2; e o valor do 3º quartil é 3.

Podemos analisar que cerca da metade dos entrevistados ou 50% utilizam-se de dois carros ou menos, e cerca de três quartos ou 75% dos entrevistados utilizam-se de três carros ou menos. Para a análise no box-plot teríamos:

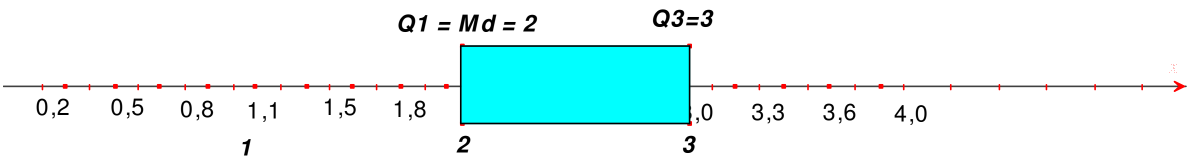


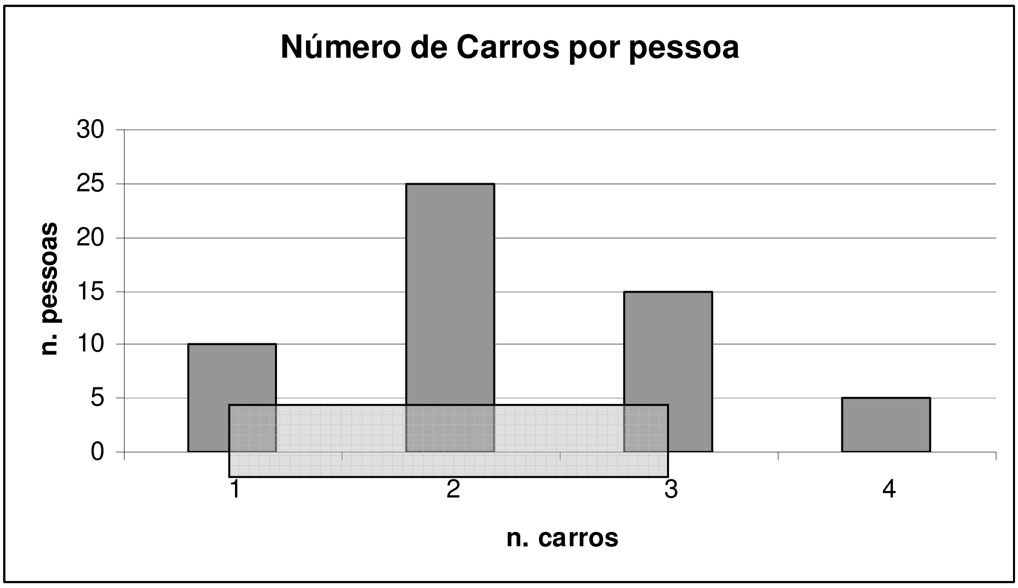
Figura 11: representação da amplitude dos quartis

Podemos verificar claramente que o primeiro quartil e a mediana coincidem, havendo uma concentração dos dados entre 1 a 2 carros. Na análise do box-plot, fica claro que a mediana é o valor mais representativo para o estudo da variabilidade do conjunto de dados.

2ª Estratégia: resolução por meio de gráfico

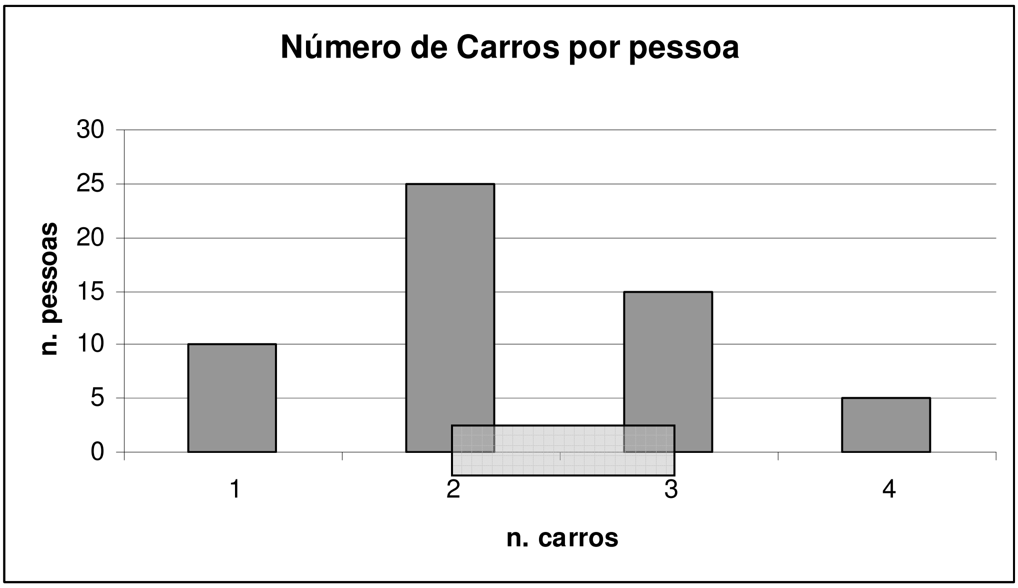
Na construção de gráficos estatísticos, percebemos que os professores do Ensino Superior, às vezes, dedicam pouco tempo para o ensinamento deste tópico. Levantamos a preocupação de que gráficos estatísticos podem ser um facilitador, trazendo informações resumidas das informações, e talvez isso seja o suficiente para o tipo de análise ou, também, poderá ser um complicador, já que, para distribuições com intervalo de classes, muitos dados são perdidos e a análise poderá não mostrar o que realmente se investiga. Por meio do gráfico abaixo tentamos analisar o comportamento do conjunto de dados em estudo.

Gráfico 6: estudo das medidas centrais por meio de gráficos.



Calculados os valores de variação, 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 as medidas separatrizes e verificar qual a melhor escolha para a representação da variabilidade. Vamos investigar a mediana no gráfico abaixo:

Gráfico 7: estudo das medidas separatrizes por meio de gráficos.



Quando mostramos as medidas separatrizes, 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 aluno, ele perceba a necessidade de se atribuir a mediana como melhor valor de representividade.

A tabela 2 da segunda parte apresenta a distribuição com intervalo de classes, e os motivos para tal escolha são os mesmos da atividade da tabela 1.

Comecemos os cálculos, completando a **tabela 4**, fornecida na 2ª parte.

Tabela 9: distribuição de freqüência com intervalo de classe da tabela 4.

Tempo no trânsito	***N. de pessoas (fi)***	$P.médio(xi)$	***xi.fi***	$di = (xi - \overline{x})$	$di^2.fi$
0 ├ 2	10	1	10	1 – 4,33 = - 3,33	110,89
2 ├ 4	20	3	60	3 – 4,33 = -1,33	35,38
4 ├ 6	30	5	150	5 – 4,33 = 0,67	13,47
6 ├ 8	15	7	105	7 – 4,33 = 2,67	106,93
Total	***75***		***325***		***266,67***

1ª estratégia: Cálculo das medidas a partir da tabela

1) Cálculo da média e desvio-padrão.

A média neste caso será $\bar{x} = \frac{\sum xi.fi}{n} \Rightarrow \bar{x} = \frac{325}{75} = 4{,}33$ horas por pessoa;

O desvio padrão será $s = \sqrt{\frac{\sum di^2.fi}{n-1}} \Rightarrow s = \sqrt{\frac{266{,}67}{74}} = 1{,}89$ horas por pessoa.

Podemos verificar o comportamento da variável com mais clareza quando representada geometricamente, e tirar conclusões se a média é um bom valor de representação da variabilidade. Veja:

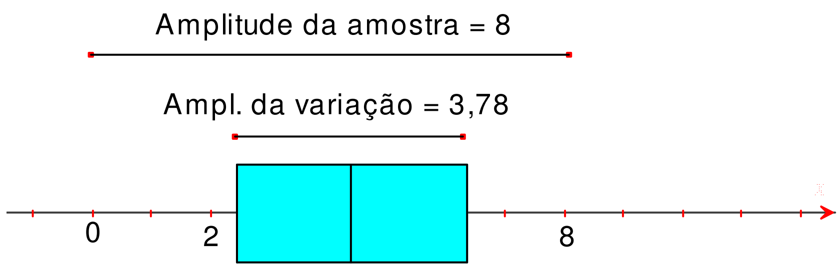


Figura 12: variabilidade em torno da média

A variável em torno da média apresenta uma amplitude de 3,78 horas, representando 47,25 da amplitude total da amostra. Não há segurança em dizermos se a média representa a variabilidade da amostra sem realizar os estudos das separatrizes, e é o que esperamos dos alunos investigados.

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

Como a distribuição está com intervalos de classes, utilizamos as seguintes fórmulas:

Para a mediana: $md = l\inf + \left(\dfrac{\frac{n}{2} - \sum fant}{fi} \right).h$

$l\inf$ = limite inferior à classe da mediana

n = número de elementos da amostra

$\sum fant$ = somatório da freqüência anterior à classe da mediana

fi = freqüência da classe da mediana

h = amplitude da classe

então

$$md = l\inf + \left(\frac{\frac{n}{2} + \sum fant}{fi} \right).h \Rightarrow md = 4 + \left(\frac{\frac{75}{2} - 30}{30} \right).2 \Rightarrow md = 4{,}5$$

Para a Moda: $mo = l\inf + \left(\dfrac{\Delta 1}{\Delta 1 + \Delta 2} \right).h$

$l\inf$ = limite inferior à classe da mediana

h = amplitude da classe

$\Delta 1$ = diferença entre os valores do *fi* da classe modal com a anterior

$\Delta 2$ = diferença entre os valores do *fi* da classe modal com a posterior

então:

$$mo = l\inf + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2} \right).h \Rightarrow mo = 4 + \left(\frac{10}{10 + 15} \right).2 \Rightarrow mo = 4{,}8$$

Para o 1º quartil e o 3º quartil, utilizamos a mesma fórmula da mediana, porém mudamos a fração de *n* para cada item:

1º quartil: $Q1 = l\inf + \left(\dfrac{\dfrac{n}{4} + \sum fant}{fi} \right).h \Rightarrow md = 2 + \left(\dfrac{\dfrac{75}{4} - 10}{20} \right).2 \Rightarrow md = 2,88$

3º quartil: $Q1 = l\inf + \left(\dfrac{\dfrac{3n}{4} + \sum fant}{fi} \right).h \Rightarrow md = 4 + \left(\dfrac{\dfrac{3.75}{4} - 30}{30} \right).2 \Rightarrow md = 5,75$

Análise esperada pelos alunos:

Nas medidas de variação, observamos que a média é de aproximadamente 4 horas por pessoa com desvio-padrão igual a 2 horas por pessoa. A amplitude total é de 8 horas por pessoa, com coeficiente de variação em torno de 50%. Vamos analisar estes dados por meio de uma representação geométrica, ou seja, o box-plot.

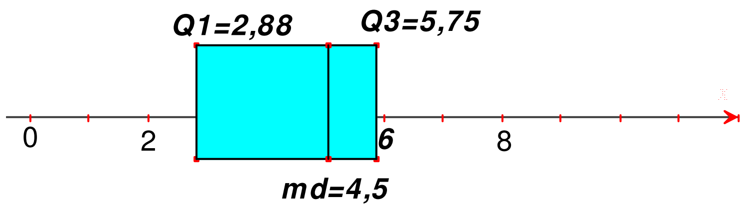


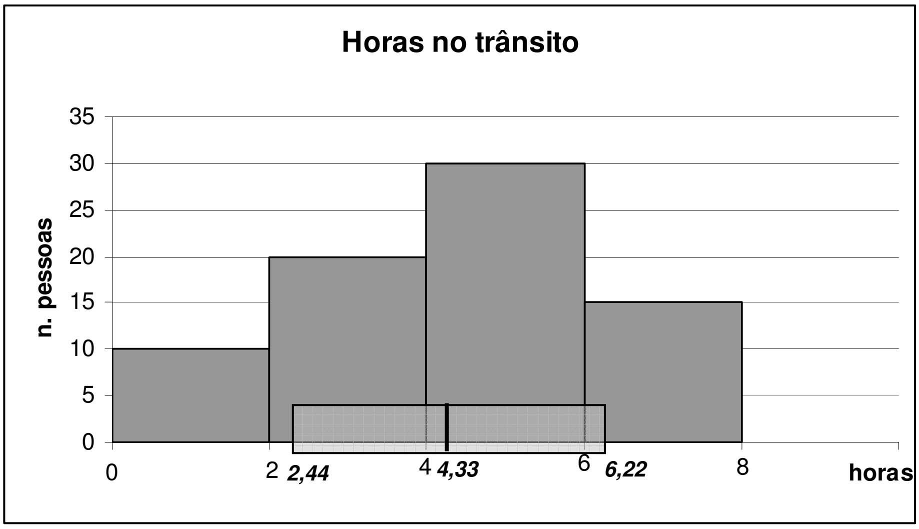
Figura 13: variabilidade em torno da mediana.

Verificamos, mais uma vez, a concentração dos dados (à direita) da mediana, confirmando assim, que a média, mais uma vez, não é um bom valor para análise dos dados, e sim a mediana, por ela explicar melhor o comportamento da amostra.

2ª estratégia: representação gráfica

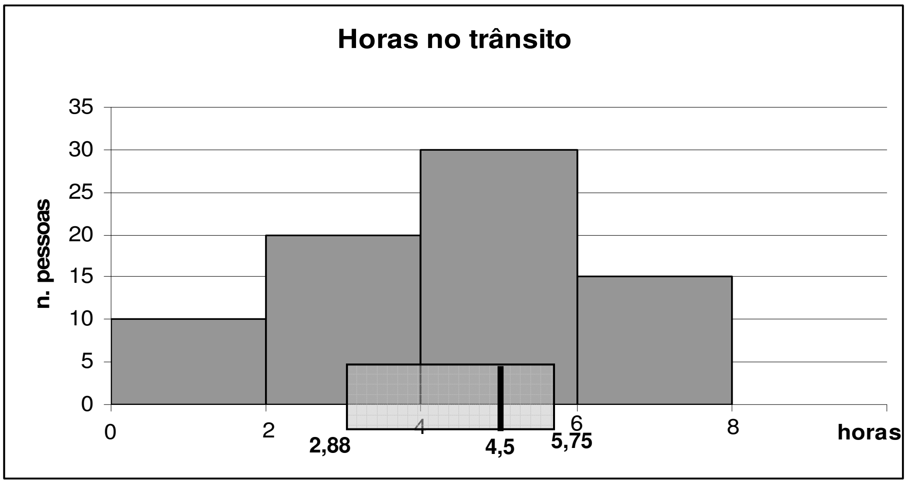
Por meio do gráfico abaixo, tentamos analisar o comportamento do conjunto de dados em estudo (tabela 4). Mostraremos, no gráfico, os intervalos dos desvios em relação à média e, posteriormente, mostraremos os intervalos interquartis. Nessa primeira apresentação mostraremos o intervalo dos desvios em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

Gráfico 8: análise das medidas centrais da tabela 4 por meio de gráficos.



Podemos perceber no gráfico 8 a concentração dos dados em torno da média e, a partir daí, explicitar 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. Nosso próximo passo é calcular a mediana por meio do próprio gráfico acima.

Gráfico 9: análise das medidas separatrizes da tabela 4 por meio de gráficos.



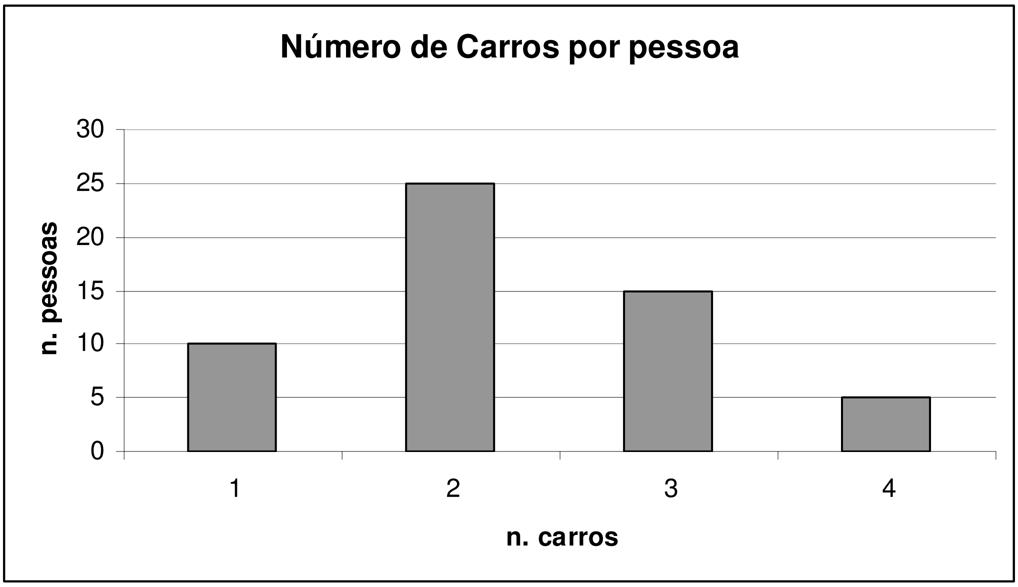
Podemos verificar uma maior concentração à direita em relação à mediana. Neste caso, seria viável e mais seguro analisar o conjunto de dados pela mediana, pois ela é que está representando melhor a amostra dos dados.

Terceira parte

A terceira parte pede para observar os dois gráficos (1 e 2) e, em seguida, 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.

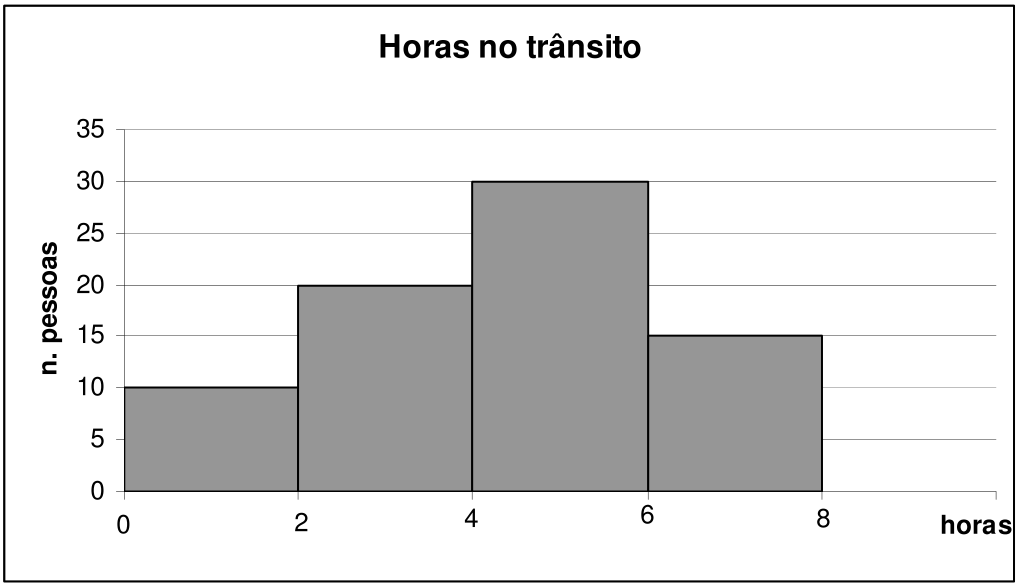
Análise dos Gráficos

Gráfico 10: retomada do gráfico 1.



De acordo com o gráfico, 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 aproximadamente 2 carros por pessoa com desvio- adrão de aproximadamente 1.

Gráfico 11: retomada do gráfico 2.



Nesta terceira e última parte, apresentaremos aos alunos dois gráficos que reforçarão nossa investigação dos níveis de funcionamento dos conceitos estatísticos, pois eles são passagens de estratégias que poderão surgir nas resoluções que se apresentarão pelos alunos. 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.

Podemos resumir estas etapas dizendo que são inúmeros os conhecimentos mínimos necessários para que o aluno do Ensino Superior resolva as atividades propostas por meio das estratégias. Entre eles estão os objetos matemáticos, já citados nesse trabalho. Não podemos deixar de ressaltar que os resultados obtidos por meio destas estratégias estão atrelados a cálculos algébricos. As dificuldades que poderão surgir por conta da Álgebra poderão ser um fator negativo para que nosso trabalho tenha êxito no objetivo traçado.

As atitudes negativas em relação à Matemática na resolução de cálculos estatísticos é fator preponderante na disciplina de Estatística. Silva (2000) apontou, em seu trabalho, que as dificuldades e ansiedades provocadas pela disciplina de Matemática são transferidas para a disciplina Estatística, sendo fator importantíssimo para um desempenho insatisfatório dos alunos. Novaes (2004) também faz menção a que os erros praticados pelos alunos, em sua pesquisa, foram algébricos e de ordem conceitual, e que estes erros apresentavam certa regularidade na maioria dos alunos investigados.

Na resolução por meio de gráficos, também devemos nos preocupar com possíveis erros e falsas interpretações que poderão surgir durante a aplicação da atividade. Os professores de Estatística supõem, às vezes, que a elaboração de gráficos e tabelas é muito simples e dedicam pouco tempo para seu ensinamento Batanero (2001, p.79). 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 aluno calcular as medidas de variação (média, desvio-padrão e coeficiente de variação) e também as medidas separatrizes (quartis), forçando ao aluno uma análise que demandaria 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), porém a variabilidade seria 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 este aluno encontra-se, segundo Robert (1998), a um nível técnico de mobilização.

A situação-problema apresentada nessa estratégia não sugere que o aluno faça associações entre as medidas, porém, em nossa pesquisa, esperamos encontrar alunos que as façam espontaneamente, caracterizando, segundo Robert (1998), um nível disponível de conhecimento. Por outro lado, não havendo essa associação por parte do aluno, e havendo a necessidade de algum tipo de intervenção do professor, sem que este dê a solução do problema, e sim indique o caminho a ser percorrido superando a dificuldade do aluno, classificamos, segundo Robert (1998), em um nível mobilizável.

3.3 ANÁLISE *A POSTERIORI*

Voltaremos ao enunciado da atividade para que o leitor acompanhe melhor esta análise. Nossa atividade **(apêndice I)** foi composta por três etapas: na primeira, apresentamos um banco de dados de variável discreta quantitativa; na segunda, duas tabelas representando uma distribuição de freqüências de variável discreta quantitativa (1ª tabela – quantidade de carros por pessoa) e variável contínua quantitativa (2ª tabela com intervalo de classes – tempo no trânsito por pessoa); e, na terceira, uma distribuição de freqüências representada graficamente. Para as duas primeiras partes da atividade, foi pedido para que os alunos calculassem, nas medidas de dispersão, a média e o desvio-padrão e, para as medidas separatrizes, que calculassem a mediana, moda e quartis. Em seguida, foi solicitado aos alunos que fizessem uma análise dos cálculos e relatassem quais das duas medidas encontradas representava melhor a amostra. Para a última parte da atividade, pediu-se aos alunos 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 aplicação das atividades, os alunos foram organizados em duplas, as quais denominaremos por D1 (dupla 1), D2 (dupla 2) e D3 (dupla 3) na seqüência de nossa análise. Os procedimentos descritos por cada dupla foram analisados de acordo com o descrito no nosso quadro teórico, no qual a educação estatística é retratada como a habilidade para interpretar, avaliar criticamente e, se necessário, discutir sobre: informações estatísticas, argumentos e mensagens (Gal 2002, p. 28) conforme abordado no capítulo que tratou do quadro teórico. Nesse contexto, procuramos identificar os **níveis de mobilização dos conhecimentos pelos alunos** (Robert, 1998), descritos como: técnico, mobilizável e disponível, conforme também apresentado no capítulo do quadro teórico. Buscamos, assim, identificar um patamar não só em um campo de conhecimentos matemáticos, mas também no estatístico, caracterizado e sustentado por objetos matemáticos apresentados, de uma certa maneira, por meio de teoremas matemáticos e associados a diversos quadros, registros de representações e noções intuitivas.

No decorrer das atividades, percebemos que a maioria dos erros cometidos foi de ordem analítica, e não de ordem algorítmica. Isto é, os alunos são capazes de aplicar fórmulas para cálculo das medidas solicitadas, sem, no entanto, conseguir analisar o significado dos resultados obtidos. As análises feitas pelos alunos, tanto na forma escrita quanto na oral, não foram totalmente claras (justificaremos isso logo abaixo), e, sendo assim, tivemos a sensação de que as análises das atividades feitas por esses alunos surgiam pela necessidade de darem respostas. Vale ressaltar aqui uma forte influência do contrato didático vigente: o que se pergunta pelo professor deve sempre ter uma resposta.

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, conforme apresentamos em nosso capítulo, no qual abordamos o quadro teórico. Em nossa análise, categorizaremos os procedimentos de resolução dos alunos, em dois níveis de conhecimento.

Quadro 2: Categorização dos procedimentos

Primeiro nível (Operacional)	Segundo nível (Analítico)
Alfabetização	*Análise global*
Estatística	*Análise crítica*
Matemática	

O primeiro nível, que chamaremos de Operacional, seriam 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. Já que a maioria das mensagens estatísticas é apresentada por textos escritos ou orais, e, sendo assim, requerem que os alunos naveguem por exibições de informações tabulares ou gráficas, a utilização de habilidades específicas da estatística tornase um ponto fundamental para nossa análise. O uso de mensagens estatísticas e matemáticas apresenta várias demandas com respeito à habilidade de alfabetização dos alunos. Por exemplo, os alunos têm que estar atentos ao fato de que os significados de certos termos estatísticos usados na mídia podem ser diferentes de suas acepções coloquiais ou cotidianas. 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.

No que diz respeito à segunda categoria, a 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 promover no aprendiz habilidades e competências necessárias para uma análise, crítica e correta, dos dados. Porém nossa hipótese é de que a maioria das instituições de Ensino Superior não discute como desenvolver essas habilidades e competências nos alunos em curso, preparando-os para uma sociedade que derrama a todo momento, informações estatísticas por diversas formas de comunicação.

Na terceira categoria, Matemática, os alunos precisam estar atentos sobre alguns procedimentos matemáticos que estão claramente por trás dos indicadores estatísticos, como, por exemplo, porcentagem e média. Ao mesmo tempo, expectativas quanto à quantidade e nível necessário de conhecimento matemático formal para compreender idéias estatísticas básicas ensinadas ao nível introdutório no ensino superior têm mudado recentemente (Moore, 1998). 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 sujeito 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 alunos do ensino superior 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.

Porém perguntas sobre o conhecimento matemático de que o aluno precisa ter para entender conceitos mais sofisticados são mais difíceis de serem respondidas e foram parte de um debate entre estatísticos e educadores de matemática (Moore, 1997). Um conhecimento mais profundo dos conceitos relacionados à Matemática, e a própria interpretação do seu exato significado, requer um entendimento mais sólido do que está por trás das idéias estatísticas.

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 alunos do ensino superior, exigindo habilidades em colocar as mensagens dentro de um contexto e, também, o de buscar seu conhecimento no que diz respeito a dados estatísticos. Essa habilidade de interpretar mensagens estatísticas também apóia o processo geral da alfabetização estatística e é um fator crítico para possibilitar o senso de compreensão de qualquer mensagem estatística. Moore (1990) discutiu isso em estatística, na qual o contexto motiva os procedimentos. 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.

Acreditamos que a habilidade dos alunos compreenderem as diferentes formas de apresentações estatísticas depende da informação que eles podem obter da mensagem sobre o que está por trás do estudo que está sendo discutido. Conhecimento de contexto é o principal determinante da familiaridade do aluno com fontes para variação e erro. Se o aluno não está familiarizado com o contexto no qual foram coletados os dados, na forma em que se apresenta o problema a ser resolvido, fica mais difícil compreender e analisar a variabilidade dos dados, interpretar e validar resultados e os modelos utilizados na análise.

Sendo assim, os alunos têm que se manterem atentos e examinar a racionalidade das informações divulgadas nos meios de comunicação escritos ou visuais. Eles devem se preocupar 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.

O que esperamos, de acordo com os níveis atribuídos por nós, é que o aluno do Ensino Superior, que já passou pela disciplina de Estatística, seja, pelo menos, segundo Robert (1998), mobilizável em relação a estes dois níveis que acabamos de apresentar. Seguem agora, as análises das duplas.

3.3.1 ANÁLISE DAS DUPLAS D1, D2 E D3.

Primeira Parte

Foi apresentada aos alunos uma tabela contendo dados brutos relativos à idade e à renda mensal de 40 pessoas. 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.

Na nossa análise, as duplas D1, D2 e D3 não apresentaram dificuldades na manipulação dos cálculos para a primeira parte da atividade. No entanto as duplas não souberam justificar o uso dos cálculos, sugerindo que os valores encontrados seriam auto-explicativos. Essas duplas utilizaram a calculadora como uma ferramenta fundamental para a resolução do problema, sem questionar o procedimento algorítmico. As duplas alegaram que tal procedimento não é relevante, já que a calculadora realiza todos os cálculos necessários e, sendo assim, o que importava era o resultado final. Assim, o procedimento algorítmico com uso de calculadora é perfeitamente dominado por esses alunos, que, no entanto, não conseguem explicar a própria construção destes procedimentos. Isso nos leva a questionar o significado construído pelos alunos sobre estes conceitos. Neste sentido, podemos supor que eles estão em um nível técnico de mobilização dos conceitos abordados.

> Este nível corresponde para nós a dos focos em funcionamento indicados, isolados, colocando em jogo as aplicações imediatas de teoremas, propriedades, definições, fórmulas, etc. Ele contextualiza de maneira simples, sem etapas, sem trabalho preliminar de reconhecimento, sem adaptações. Isto diz respeito principalmente ao funcionamento das ferramentas (inclusive as definições). (ROBERT, 1998 p. 165).

Na resolução final da primeira parte, que envolvia média e desvios, estabelecemos com D1 um debate sobre os valores encontrados e quais seriam seus significados. Percebemos que esta dupla teve extrema dificuldade em responder e fornecer uma análise crítica. Ou seja, a dupla em questão não foi capaz de explicitar qualquer tipo de relatório, oral ou escrito, que poderia justificar os resultados encontrados, como por exemplo, a resolução mostrada no capítulo, que trata da análise *a priori* das atividades. Partiram, assim, para os cálculos das medidas separatrizes.

A estratégia de D1 para os cálculos dos quartis foi a de não se preocupar com fórmulas. Dividiram a amostra em 4 partes iguais, encontrando o 1º, o 2º e o 3 º quartis. Nesta etapa, a D1 mostrou-se bastante prática na manipulação dos dados. Este procedimento é fundamentado no conceito de quartis. Na análise final da primeira parte, D1 chegou à conclusão que a questão 2 da primeira parte, que abordavam os cálculos das medidas separatrizes, seria o melhor valor para representar a amostra. Questionamos o porquê da escolha e obtivemos a seguinte resposta: *chamaremos Pesquisador de P.*

P: Por que vocês escolheram a questão 2 para a análise final?

D1: Por causa do tipo de distribuição.

P: Como assim por causa da distribuição? que tipo é este?

D1: (silêncio..) humm, não sei, mas é por causa da distribuição. Ela tem uma melhor distribuição.

Os protocolos nos permitem inferir que D1 pensava em distribuições simétricas, pois o contrato didático leva a isso, já que os livros didáticos trabalham, na maioria dos seus exemplos, com distribuições simétricas. Acreditamos que a forma de abordagem dos livros didáticos utilizados por alunos do curso superior não favoreceram a superação da dificuldade dos alunos em diferenciar os tipos de distribuições.

Percebemos que esses livros didáticos induzem o aluno a ser somente técnico, segundo Robert (1998). O fato é que se suspeita que as situações-problema, apresentadas na disciplina de Estatística na graduação, não são preparadas para uma mobilização de conhecimentos, ou seja, são situações-problema que apresentam um só tipo de situação, como, por exemplo: “calcule a média, calcule o desvio-padrão”, e deseja-se que essas atividades exijam dos alunos uma análise crítica dos resultados obtidos. Novaes (2004, p. 44), em seu trabalho, escolheu alguns exercícios extraídos de livros didáticos, que foram utilizados pelos seus alunos pesquisados durante o desenvolvimento do componente curricular Estatística. As análises feitas pela autora focaram o(s) diferente(s) tipo(s) de abordagem dos conceitos estatísticos, e observou se esses exercícios estão facilitando a possibilidade de os alunos saírem de um nível técnico, atingirem o nível mobilizável de conhecimento e evoluírem para o nível disponível, nos termos de Robert (1998).

Segundo Novaes (2004, p. 51), não foram encontrados exercícios que dessem ao aluno autonomia para resolver sozinho uma situação-problema, em que possa escolher, entre os conhecimentos, aquele que lhe pode ser mais útil, aplicando-o corretamente para fazer a resolução acertada, e demonstrando, assim, que atingiu o nível disponível de conhecimento. Ainda, segundo a autora, foi observado nos livros uma diversidade de contextos, mas não foi observada uma diversidade de representações, como, por exemplo, pedir ao aluno que calcule a média a partir de um gráfico da distribuição, como foi proposto na nossa atividade. Para a dupla D2, também foi dada a oportunidade de debater sobre os valores encontrados na primeira parte e sobre seus significados. Também a dupla D2 não soube justificar, por meio de uma análise crítica, os resultados encontrados. Apenas justificaram que os resultados eram suficientes. Segundo o relatório feito por D2, tentaram justificar esses cálculos escrevendo o seguinte texto:

D2: "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".

Pela frase de D2, podemos verificar que, para essa dupla, o conceito de média não está totalmente claro. Acreditamos que houve uma confusão envolvendo o conceito de amplitude da amostra com intervalo de confiança, tanto para variável *idade* como para variável *renda*. Pelo que percebemos, a dupla D2 buscou uma análise da variabilidade, tentando mostrar, por meio da amplitude, o comportamento da amostra. Questionado sobre o procedimento, D2 tentou justificar pelo seguinte protocolo:

P: Explique melhor o que vocês quiseram dizer com esta análise.

D2: Queremos mostrar aqui que a amplitude da amostra é alta.

P: E o que isso significa?

D2: oras … Significa que…(olharam-se) temos um intervalo de valores, do maior para o menor, altos.

P: Mas essa análise é o suficiente para alguma conclusão?

D2: Não, temos agora que calcular a mediana e...os quartis e depois analisar novamente.

P: mas do que vocês estão sentindo falta para uma análise, antes desses cálculos, mediana e quartis?

D2: (Olham-se e pensam por alguns minutos)…Não sabemos. É que você pediu aqui para calcular. Acho que, depois de calcular poderemos responder melhor.

Pode-se notar, nessa última frase dos alunos, um forte efeito do contrato didático usual: se uma atividade pede que se façam os cálculos de medidas estatísticas, isso significa que estes cálculos devem servir para alguma coisa. Não consideram o fato de que os cálculos já efetuados podem ser suficientes para a análise solicitada.

Partindo para os cálculos das medidas separatrizes, D2 também não mostrou dificuldades no desenvolvimento dos procedimentos. Assim como no caso da primeira dupla, D2 encontrou 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. O protocolo abaixo mostra a tentativa de justificar os cálculos de D2.

D2: "Essa análise permite a associação das idades com as rendas de uma faixa de clientes. Por exemplo, a renda de 577 do 1º quartil com a idade 29 do 1º quartil. Isso permite identificar segmentos de atuação".

Podemos perceber que D2 tentou fazer uma comparação entre as variáveis, fazendo uma interpretação de variáveis bi-dimensionais, porém não se preocupou 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 de contrato, uma vez que os dados foram apresentados em uma tabela com as duas variáveis. Podemos inferir que esta dupla está no primeiro nível que abordamos no começo do capítulo, pois percebemos que esta dupla possui 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 pela dupla, 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 mobilização, D2 encontra-se em um nível de conhecimento que Robert (1998) classifica como nível técnico.

Para responder à última questão da primeira parte, a D2 respondeu como mostra o protocolo abaixo:

D2: "A 1ª questão é útil no sentido que mostra a grande amplitude de idade e de renda dos clientes, mostrando que é possível dividi-las em faixas, ou segmentos de atuação de marketing, mais precisos, concedendo crédito conforme a faixa de idade e renda e seu perfil. A 2ª questão é importante para identificar esses segmentos. Portanto nós utilizaríamos as duas análises".

Para D2 responder à última questão, necessitou de muito debate entre os integrantes da dupla, percebemos que a dupla D2 precisava dar uma resposta e, quando questionada da reposta dada, a dupla respondeu o seguinte:

P: Como vocês chegaram a esta conclusão?

D2: Bom, pegamos a resposta da questão 1 e a resposta da questão 2, e analisamos com calma e decidimos pela duas questões.

P: Então vocês acham que tanto a questão 1 quanto a questão 2 são medidas que representam bem as variáveis estudadas?

D2: Sim, pois uma não depende muito da outra. Então podemos pegar uma ou outra para explicar.

Podemos perceber que as análises feitas por esses alunos são carregadas de um vocabulário peculiar do curso. Os termos usados para a redação dos resultados não fogem de um contexto próprio à área de Administração, como, por exemplo, "*segmentos de atuação de marketing*", mas, por outro lado, não percebemos uma segurança nas respostas dadas; pelo contrário, D2 não soube concluir com coerência todos os resultados encontrados.

A dupla D3 foi a dupla que demonstrou maior segurança nas respostas das análises. No entanto algumas justificativas apresentadas, também foram efeitos do contrato didático, ou seja, a forte necessidade de apresentar uma resposta, pois o professor havia feito uma pergunta.

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

D3: "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."

Na análise dos diálogos estabelecidos entre os alunos e da sua produção escrita, não podemos identificar uma ligação de idéias ou mesmo de conceitos que justifiquem a afirmação feita. 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, conforme sugerimos em nossa análise *a priori*; coisa que não aconteceu. 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. Ao questionar sobre a frase escrita, tivemos as seguintes respostas:

P: A dupla poderia ser mais clara sobre a análise da questão?

D3: Humm…Tentamos mostrar que a variável "renda mensal" é mais dispersa que a variável "idade". Está errado?

P: Não se preocupem. Não estou aqui para corrigi-los, apenas debater os resultados de vocês, ok? Mas gostaria que vocês explicassem melhor essa idéia de variabilidade que vocês escreveram em relação às variáveis? Ou não tem importância?

D3: Ahhh, acho que não tem muita importância. Apenas foi uma informação que percebemos, mas não vejo isso como coisa que irá servir para relatório…Sei lá.

P: E sobre a média? Quando vocês escreveram "no entanto", parece que este valor salvou a questão? É isso?

D3: Não…É…Acho que a média é um valor que a maioria das pessoas calcula em uma atividade de Estatística. Achamos que, como ela está próxima de um dos extremos, pode indicar algo de maior importância.

P: O que, por exemplo?

D3: Que ela não está dividindo no meio. Esta amostra não é certinha, ela é concentrada em um dos lados, em um dos extremos.

P: Mas, e aí?

D3: (Ficaram três minutos pensando)…Não sabemos.

Percebemos que, nessa dupla, a idéia de simetria está implícita quando afirmam que “esta amostra não é certinha”. Há indícios de que se referiam à distribuição simétrica, porém, na frase “que ela não está dividindo no meio”, pode-se inferir que a dupla confundiu os conceitos de média e mediana, tal como descrito em Batanero (2001) e Novaes (2004), quando comenta que os termos matemáticos que designam esses conceitos têm um significado preciso, embora este nem sempre coincida com o seu significado em linguagem coloquial.

Estas medidas coincidem quando a distribuição é simétrica, o que reforça nossa hipótese de efeito de contrato. Vale aqui ressaltar que as situações-problema, dos livros didáticos, desenvolvidas nos cursos de Estatística, estão voltados a distribuições simétricas, deixando ao aluno poucas opções de diferentes tipos de distribuições, para que ele, possa ter uma melhor noção do comportamento da amostra (Novaes, 2004). Percebe-se que esses tipos de escolha didática, por parte dos livros didáticos, provocam esse tipo de efeito. Essa autora cita os trabalhos de Russel Y Mokros (1991), que classificam em quatro categorias os significados incorretos atribuídos por estudantes à palavra “média”: valor mais freqüente (confusão com o conceito de “moda”), valor razoável (significado coloquial de termo), ponto médio (confusão com o conceito de mediana) e algoritmo (significado restrito em que a média é vista apenas como seu algoritmo de cálculo). Este último significado está nos diálogos das três duplas estudadas.

Percebermos que essa dupla encontra-se no nível operacional da nossa análise, pois os três primeiros níveis são contemplados nos diálogos dessa dupla e, dentro desse nível, inferimos que os alunos encontram-se em um nível de mobilização de técnico para mobilizável, já que os protocolos mostram que os alunos perceberam a concentração em um dos extremos da amostra, sem, no entanto, explorar esse fato. “Um saber é dito mobilizável, quando é bem identificado, é bem utilizado pelo aluno, mesmo que tenha sido necessária uma adaptação ao contexto particular” (Robert, 1998, p. 166). Segundo a autora, a adaptação necessária ao conhecimento pode ser sugerida pelo professor ou pelo próprio enunciado do problema, e isto é o que diferencia conhecimento mobilizável de conhecimento disponível, pelo qual o aluno deve agir de forma autônoma e, sendo assim, poder transformar a informação em uma informação operacional.

Na segunda questão, que tratou das medidas separatrizes, D3 não apresentou dificuldades nos cálculos algorítmicos, porém, na análise critica do resultado encontrado, D3 relatou o seguinte:

> *D3: “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 pela dupla foram os seguintes: para a variável idade, o 1º quartil = 29, o 2º quartil = 30 e 3º quartil = 40; e para a renda mensal, os valores para o 1º, o 2º e o 3º quartis foram, respectivamente, 577, 890, 1385. Infere-se que esta dupla possua a percepção de não simetria, pois, na fala de D3 que diz "*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 o 3º*", a dupla deixa transparecer bem essa idéia, e, quando D3 relata "*Já na variável "renda mensal", os valores são mais próximos, deixando assim a mediana com distâncias semelhantes de ambos os quartis"* inferimos que, nesse protocolo, a dupla está 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 dois bancos de dados representados 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 as duplas investigadas pudessem manipular as informações em diferentes formas de representação, ou seja, conjunto de representações simbólicas utilizadas pelo aluno na manipulação do conceito (Vergnaud, 1995).

Nessa segunda parte, as duplas apresentaram apenas dificuldades para encontrar um método de resolução que justificasse os resultados encontrados, pois os alunos alegaram não ter aprendido a calcular as medidas solicitadas por meio de tabelas com ou sem intervalos de classe. Vale ressaltar que estes alunos sabiam manipular *softwares* estatísticos e lembramos que, na atividade anterior, que partia de um banco de dados não organizado em uma tabela de distribuição de freqüências, os alunos fizeram sem dificuldades os cálculos solicitados.

Ao analisarmos D1, percebemos que o nível de conhecimento dessa dupla, segundo Robert (1998), pode ser classificado como mobilizável. Percebemos que, depois da intervenção do pesquisador, houve sucesso no cálculo dos valores, visto que os alunos transformaram os dados agrupados em dados brutos, podendo, assim, inferir sobre os dados apresentados nos moldes de resolução da primeira parte, e, sendo assim, 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. Poderemos perceber isso no protocolo que segue abaixo:

D1: Professor, não sabemos calcular média e desvio padrão na tabela.

P: Não sabem ou não lembram?

D1: Não aprendemos a calcular esses valores por tabela.

P: Então vamos analisar a primeira tabela. O que vocês estão percebendo?

D1: Nós entendemos que os valores estão compactados, assim...cinco pessoas tem um carro, vinte e cinco pessoas possuem dois carros e assim vai.

P: Perfeito. Então como poderíamos fazer uma comparação com a primeira parte da atividade, veja.

D1: Ahh, então poderemos espalhar os valores e deixá-los como a primeira parte, é isso?

P: Querem tentar? Fazendo isso, vocês são capazes de calcular?

D1: Sim, aí é só jogar na calculadora.

P: Então façam. Mas e a mediana, a moda e os quartis?

D1: Faremos a mesma coisa.

Nessa primeira etapa do protocolo, no que diz respeito à variável discreta, os alunos conseguem fazer a transformação necessária para a resolução do problema proposto, implicando, assim, um nível mobilizável de conhecimento e dentro da categoria operacional, ou seja, o algorítmico. Porém podemos perceber que, na segunda etapa do protocolo, essa mobilização não ocorre. Os alunos tentam manter o mesmo padrão de raciocínio para estabelecer uma estratégia de resolução, não alcançando sucesso. Veja pelo protocolo:

P: E para a segunda tabela, com intervalo de classes, vocês têm idéia de como farão?

D1: Não, aí já não sabemos, porque a primeira linha tem o tempo entre 0 (zero) e 2 (dois). Pelo que lembro, esses valores são contínuos e como vou saber a quantidade que tem entre 0 e 2?

P: Bom, você tem razão, não dá para saber. Mas do que você está sentindo falta? Por exemplo, o que é tem na primeira tabela e não tem na segunda, que ajudaria a calcular os valores?

D1: A variável x_i da fórmula da média, por exemplo. A quantidade de pessoas eu tenho; não tenho aqui a variável x_i.

P: Teria como arranjar essa variável?

D1: Não sei... tem?

P: Bom, e se calculássemos o ponto médio entre os intervalos e tomássemos como variável x_i os valores encontrados (faz no quadro um exemplo prático)

D1: aí dá para resolver. Mas.... podemos fazer assim então?

Nessa segunda etapa dos protocolos, fez-se necessária, enquanto interventor, a produção de um exemplo no quadro negro para introduzir um conhecimento novo aos alunos, para que eles pudessem perceber a similaridade com situações anteriores. Mostrando a variável x_i (ponto médio das classes) para os alunos, foi possível o reconhecimento de contexto por parte deles e, sendo assim, possibilitou a transformação necessária dos dados para resolução do problema. Essa transformação, como podemos perceber, não foi no âmbito de transformar em rol os dados que estavam na forma de intervalos de classe, e sim o procedimento de multiplicar a variável x_i pelo valor da freqüência, método este percebido na primeira parte desta etapa pela dupla. O que sentimos falta foi a não percepção da dupla de que, trabalhando com o ponto médio do intervalo de classe, há perda de precisão. Porém podemos dizer que fica aqui bem clara, a passagem do nível técnico para o mobilizável, segundo Robert (1998). Mas, na seqüência dos protocolos, percebemos que os níveis de dificuldade dos alunos voltam à estaca inicial.

P: Mas como ficaria a mediana, a moda e os quartis?

D1: Aí não há como calcular, pois os valores são contínuos. Aí nunca saberemos qual é o valor que divide essa amostra em partes iguais, por exemplo, pois são valores infinitos.

No final, a dupla não soube justificar qual o valor que explicaria o comportamento da amostra, pois, como a questão tratava de calcular as medidas separatrizes, esta não foi concluída pela dupla. Destacamos o fato de que a dupla nem tentou 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. Por exemplo, 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. D1 poderia tentar dar resposta conclusiva por meio dessa análise visual, ou ,ainda, analisar por meio de uma apreensão perceptiva dos dados representados graficamente.

Nossa hipótese, feita por essa análise, mostrou que, mais uma vez, a necessidade de apresentar cálculos numéricos para justificar resultados foi muito forte para essa dupla. A matemática, que é uma das bases de Gal, manifestou-se com maior intensidade, e foi a ferramenta encontrada para justificar a análise dos dados.

Analisando D2, na mesma etapa da atividade, concluímos que o processo algorítmico não apresentou dificuldades, mesmo a dupla não apresentando os cálculos algébricos que justificassem os resultados encontrados. Não percebemos, até por conta de extrema timidez por parte da dupla, diálogos entre os alunos, que justificassem a análise da atividade feita por ela. O protocolo de D2, que justifica os cálculos encontrados, apresenta-se da seguinte maneira:

D2: "Como conclusão, 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".

Os resultados encontrados foram bem próximos aos que encontramos em nossa análise *a priori*, porém o que podemos perceber é que o texto elaborado por D2 foi exatamente um relato dos valores encontrados, ou seja, o que percebemos é apenas que D2 fez 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, já que D2 precisava, naquele momento, de uma resposta para o professor que acabara de realizar uma pergunta.

O que nos chamou a atenção foi o fato de que a dupla, ao encontrar como resposta a média igual a 2,73 para a primeira tabela, arredondou-a para 2 carros no relatório final. O mesmo não aconteceu com o tempo em horas. O valor encontrado foi a média de 3,67 horas. Assim, fica claro que esses alunos perceberam as variáveis contínua e discreta, porém nos níveis de alfabetização estatística. Gal (2002) 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 do indivíduo). Não é 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. As variáveis quantitativas contínuas têm papel no campo da matemática que é específico e sem alterações, mas, quando trazemos estas mesmas variáveis para a vida real, temos que adaptá-las para que possamos dar sentido lógico ao contexto em que ela está inserida. Inferimos que D2 não se preocupou com essa análise, já que ela apresenta, em sua conclusão, uma média de horas de 3,67.

Para o cálculo das medidas separatrizes, D2 teve o mesmo discurso de D1. Alegaram não ter aprendido esse tipo de cálculo por meio de tabelas. Porém o mesmo diálogo que tivemos com D1 para, quem sabe, induzir a dupla à passagem de níveis de conhecimento, não ocorreu e, sendo assim, ficamos impossibilitados de investigar em qual nível se encontrava a dupla D2. Os cálculos foram realizados e não foram explicitados. A metodologia de resolução e a conclusão da dupla, segue em seu protocolo:

D2: "A utilização da técnica dos quartis apresenta como características a segmentação da amostra, o que nos possibilita concluir que o perfil da amostra é de pessoas com 3 carros e que gastam no mínimo 3 horas no trânsito".

Os resultados mostrados na resolução da atividade somado ao que foi escrito na conclusão apontam que D2 confundiu o conceito de mediana e média. Nossa hipótese é a de que esta dupla usou, como comparativo, os resultados encontrados na média e mediana. Os valores encontrados para a primeira questão foram a média de 2,73 carros por pessoa com tempo médio de 3,67 horas; já os encontrados na mediana foram, respectivamente, para os carros 3 e para as horas 3. No momento em que a dupla relata "...*o que nos possibilita concluir que o perfil da amostra é de pessoas com 3 carros e que gastam no mínimo 3 horas no trânsito",* podemos inferir que há aqui uma confusão entre média e mediana, já que se trata de medidas separatrizes, e não medidas de tendência central. O que esperávamos dessa dupla é exatamente uma análise semelhante ou próxima às que explicitamos na nossa análise *a priori,* no capitulo que tratou do assunto. Podemos aqui também inferir que, por acreditarem que todas as análises feitas pelas três duplas 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).

Analisando D3, na segunda parte da atividade, percebemos a ausência de dificuldades nos cálculos algorítmicos. A dupla também não apresentou os cálculos que justificassem os resultados e utilizou-se somente da calculadora. O protocolo de D3 para a análise da primeira questão foi o seguinte:

D3: "O número de carros tem uma variabilidade menor do que o tempo gasto no trânsito, ou seja, o comportamento das pessoas em relação ao número de carros que compram é mais homogêneo do que o tempo gasto no trânsito".

D3 tentou justificar seus cálculos comparando a variabilidade das variáveis, quantidade de carros e tempo no trânsito, com relação a comportamento das pessoas em adquirir um veículo.

A conclusão sobre as duplas investigadas, referindo-se ao nível operacional, focalizando o processo algébrico nas duas primeiras etapas da atividade, é que as duplas se encontram no nível de conhecimento técnico para mobilizável, segundo Robert (1998). Em se tratando de nível analítico, as duplas não se enquadraram em nenhum dos níveis propostos por Robert, pois nenhum relatório foi apresentado por estas duplas que permitissem uma análise da nossa parte, nas etapas da atividade.

Para a última parte da nossa atividade, não houve possibilidade de diagnosticar os resultados, pois as três duplas estudadas alegaram não ter aprendido a resolução de problemas estatísticos por meio de gráficos. Vejam os resultados por meio da tabela.

Tabela 10: Resultados das análises

<table>
<tr><th colspan="2" rowspan="2">Duplas D1, D2 e D3</th><th rowspan="2"></th><th colspan="2">Preceitos de Id Gal (2002)</th></tr>
<tr><th>Operacional</th><th>Analítico</th></tr>
<tr><td rowspan="2">Ativ. 1</td><td>Média e Desvio</td><td rowspan="6">Preceitos de Aline Robert (1998)</td><td>Disponível</td><td>S/ classificação</td></tr>
<tr><td>Mediana Moda e Quartis</td><td>Disponível</td><td>S/ classificação</td></tr>
<tr><td rowspan="2">Ativ. 2</td><td>Média e Desvio</td><td>Mobiliz p/ 1ª e 2ª tab.</td><td>S/ classificação</td></tr>
<tr><td>Mediana Moda e Quartis</td><td>Mobiliz para 1ª tab.
S/ Cassif. p/ 2ªtab.</td><td>S/ classificação</td></tr>
<tr><td rowspan="2">Ativ. 3</td><td>Média e Desvio</td><td>S/ classificação</td><td>S/ classificação</td></tr>
<tr><td>Mediana Moda e Quartis</td><td>S/ Classificação</td><td>S/ classificação</td></tr>
</table>

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 por alunos, diagnosticadas em pesquisas que abordam o tema, tanto no ensino básico como no ensino superior, tornaram-se molas propulsoras que incentivaram pesquisadores a investigar os possíveis fatores que influenciaram no processo ensino-aprendizagem desses alunos.

Nossa preocupação e incentivo para a pesquisa surgiram diante da minha experiência profissional no nível Médio, quando alunos já formados nesse nível e ingressos no Ensino Superior questionavam-me sobre a aplicabilidade da Estatística na futura vida profissional. Pude observar que, no caso destes alunos, o ensino de Estatística no Ensino Superior está focada 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 do campo da Matemática, mas também mostrar que a Estatística tem suas particularidades, próprias da disciplina, e que é tão importante quanto a 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 aluno egresso no curso superior, mais especificamente no curso de Administração, conceitualiza as idéias básicas de Estatística, envolvendo variabilidade. Tal estudo foi feito à luz dos trabalhos desenvolvidos por dois autores: Aline Robert (1998), que trata do nível de mobilização de conhecimentos pelos alunos (técnico, mobilizável ou disponível), foi nosso pano de fundo sob um ponto de vista didático; e Gal (2002), que trata da Alfabetização Estatística, constituiu nosso quadro teórico Estatístico.

Para atingirmos nosso objetivo e responder à nossa questão de pesquisa, elaboramos e aplicamos atividades com seis alunos do curso de Administração de uma universidade pública no interior do estado de São Paulo, organizados em três duplas, que cursaram o componente curricular Estatística durante a graduação. Duas duplas do segundo semestre, que passaram pelo curso recentemente, e uma dupla do último semestre.

O procedimento utilizado foi uma atividade diagnóstica, audiogravada, que permitiu analisar as produções orais e escritas da atividade proposta aos nossos sujeitos de pesquisa. Tal atividade foi dividida em três etapas: na primeira, apresentamos um banco de dados de variável quantitativa discreta; na segunda, duas tabelas representando uma distribuição de freqüências 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 parte da atividade, foi oferecida a eles uma distribuição de freqüências representada graficamente. Em cada uma das atividades, as questões permaneciam as mesmas: cálculo e interpretação da média e do desvio-padrão, cálculo e interpretação da mediana e dos quartis e, finalmente, decisão de qual o melhor valor resumo para o conjunto de dados estudado. Para a última parte da atividade, pediu-se aos alunos que respondessem à seguinte questão: *"Se você precisasse descrever esses dados para um cliente, como você os analisaria?".*

Dessa forma, buscamos verificar o nível de mobilização dos conhecimentos por parte destes alunos segundo os termos de A. Robert (1998), que realizou um estudo sobre quatro dimensões de análise dos conteúdos a serem ensinados no campo da Matemática e, também, com as cinco bases de Alfabetização em Estatística, segundo Gal (2002), ou seja, buscamos identificar se os alunos egressos do componente curricular Estatística estão capacitados a utilizar/mobilizar, de forma eficaz, as noções estatísticas de base para resolver problemas práticos de sua área de atuação.

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 a mobilização dos conhecimentos dos alunos segundo os preceitos de Robert (1998) quanto a Técnico, Mobilizável e Disponível.

Pudemos perceber que, em nenhuma das duplas investigadas, foi possível identificar invariantes que justificassem possíveis dificuldades dos alunos no âmbito dos cálculos algébricos, ou seja, as duplas não apresentaram dificuldades em calcular as medidas pedidas nas duas primeiras etapas da atividade. Porém, nos diálogos feitos com as duplas, pudemos perceber que nenhuma delas conseguia 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 pelas duplas, levando-nos a inferir que este conhecimento, se existente, permaneceu implícito. Mesmo com questionamentos durante a atividade que visavam proporcionar condições para que as duplas exteriorizassem o significado por eles atribuído aos valores calculados, os alunos permaneciam ligados somente aos valores numéricos, acreditando que estes eram auto-explicativos. Sendo assim, inferimos que as duplas classificam-se em nível técnico no contexto algébrico, segundo Robert (1998).

Podemos resumir, nesta etapa, que as duplas investigadas dominam perfeitamente os cálculos algébricos propostos nas duas primeiras etapas da nossa atividade sem, no entanto, dar significados a estes cálculos. Para as duplas investigadas, o que importa são os resultados, os valores sendo auto-explicativos. Percebemos que o contrato didático surge com muita intensidade, ou seja, os alunos buscaram explicar os valores obtidos apenas por uma leitura dos cálculos apenas porque a tarefa assim o solicitava ou porque o pesquisador os questionava sobre isso. Eles mesmos não sentiam a necessidade dessas interpretações e justificativas.

Das duplas investigadas, duas delas não souberam distinguir a diferença de se trabalhar os conjuntos de dados na forma de rol e na forma tabular. A representação dos dados na forma de tabela foi fator de grande dificuldade para os alunos, sendo que, em nenhuma dupla, foi possível uma completa análise nesse tipo de resolução. Pudemos assim constatar a validade da afirmação feita por Batanero (2001, p. 79), uma vez que estes alunos sabiam trabalhar com conceitos avançados da estatística, tais como análise de cluster, mas sem conseguir explicar os conceitos mais simples e imediatos como média e desvio-padrão para interpretar a variabilidade dos dados.

> Los profesores suponem, as veces, que la elaboración de tablas y gráficos es muy sencilla y dedican poco tiempo a su enseñenza. Sin embargo, elaborar un tabla de frecuencias o un gráfico supone una primera reduccíon estadística, pues se pierden los valores originales de cada uno de los datos individuales pasándose a la distribuicón de frecuencia. Este concepto es ya compljo, al referirse al agregado (población o muestra) y no a los datos particulares, lo que es una característica esencial de muchos conceptos estadísticos. (BATANERO, 2002, p. 79)

Pensamos ser necessário que o profissional educador, ao apresentar um conceito novo, não apresente somente um tipo de representação. O aluno, em contato com algo que para ele é novo, tem que ter a condição de adaptar o que ele já conhece com aquilo que lhe é apresentado como novidade. Ou seja, buscar conhecimentos anteriores, adaptá-los para o novo e dar significado para esse novo conhecimento.

Percebemos, em alguns momentos, que as duplas investigadas realizaram análises equivocadas de alguns conceitos, como, por exemplo, confundir média e mediana. Isso foi diagnosticado pelo fato de as duplas terem atribuído a noção de simetria para qualquer banco de dados. Para elas, toda distribuição é simétrica, assim percebemos que, em toda a atividade, as duplas sentiram a necessidade de modelar a amostra para uma distribuição normal.

É comum encontrarmos esse tipo de equívoco quando educadores em Estatística adotam livros didáticos preparados para o ensino voltado para o contexto escolar da sala de aula. As atividades nos livros didáticos do ensino superior, em sua maioria, são trabalhadas com distribuição simétrica e, assim, os valores das medidas de tendência central e medidas separatrizes são quase que as mesmas, causando assim equívocos nos conceitos de ambas. Faz-se necessário que o profissional educador perceba que tipo de distribuição os alunos estão trabalhando para, assim, colocar situações que pelo menos faça que esses alunos percebam a diferença nas medidas encontradas e, conseqüentemente, terem condições de dizer quais são as medidas que melhor representam a amostra. Isso nos leva a inferir que o trabalho somente com distribuições simétricas leva o aluno a ser somente técnico, segundo Robert (1998), não permitindo ao aluno a oportunidade de questionar sobre outros tipos de distribuição e, então, quem sabe torná-lo um aluno questionador para que passe de um nível mobilizável para disponível.

Durante a aplicação das atividades, percebemos, por parte dos alunos, 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 obrigatoriedade desses valores. Por exemplo, a última atividade que nós apresentamos para as duplas 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 o aluno, de uma forma contextualizada, interdisciplinar e de maneira que o conhecimento esteja disponível sempre quando for necessário. O aluno, na sua formação em nível superior, terá que ter condições de continuar aprendendo e de encontrar respostas para os problemas que aparecerem durante sua trajetória profissional. Para tanto, faz-se necessário que este aluno continue sua vida educacional sempre buscando aperfeiçoar seus conhecimentos.

O caminho que escolhemos para o estudo deste trabalho, juntamente com o quadro teórico que utilizamos, proporcionou-nos novas perspectivas para o ensino de Estatística. Uma das conseqüências deste nosso trabalho é a necessidade da elaboração de uma seqüência didática que permita ao aluno vivenciar todas as fases necessárias para a construção de um conceito e sua mobilização, e o conseqüente estudo dos resultados obtidos quando da aplicação de tal engenharia em um grupo com as mesmas características do grupo pesquisado, ou ampliar a população para outras áreas profissionais.

BIBLIOGRAFIA

BATANERO, C. (2001). *Didáctica de la Estadística, Granada*: Grupo de Investigación en Educación Estadística. ISBN 84-699-4295-6.

BATANERO, C. (En prensa). *Aleatoriedad, Modelización, Simulación. Presentado en las X Jornadas sobre el Aprendizaje y la Enseñanza de las Matemáticas.* Zaragoza, 2001.

BATANERO, C. ;Y GODINO, J. D. *Análisis de Datos Y Su Didática.* Reprografia de la Faculdad de Ciências Univerdidad de Granada, 2001b.

BATANERO, C. *Los Retos de la Cultura Estadistica.* Jornadas Interamericanas de Ensênaza de la Estadistica. Buenos Aires, 2002. Conferencia inaugural.

BROUSSEAU, G. *Les Obstacles épistemologiques et les problémes em Matématiques.* RECHERCHES EM DIDATIQUES DES MATHEMÁTIQUES. La Pensée Sauvage-Editions, v.4.2, pp. 165-198. Grenoble, 1983.

CARVALHO, C. & CÉSAR, M. *Interações entre Pares e Estatísticas: Contributos para o Estudo do Conhecimento Instrumental e Relacional.* QUADRANTE Revista Teórica e de Investigação. Vol 10, N. 1 – 2001.

CASTRO, Lauro S. Viveiros. *Pontos de Estatística.* 17 ed. Rio de Janeiro: Científica, 1975.

CHIAVENATO, Idalberto. *Administração de Empresas: Uma abordagem contigencial:* 3 Ed. São Paulo. Makron Books, 1995.

COUTINHO, C.Q.S. (2001). *Introduction aux situations aléatoires dès le Collège: de la modélisation à la simulation d'expériences de Bernoulli dans l'environnement informatique Cabri-géomètre II.* Tese de doutorado. Université Joseph Fourier. França.

CROSSEN, C. *O Fundo falso das pesquisas: a ciência das verdades torcidas.* Rio de Janeiro: Ed. Revan. (1996)

CURCIO, F.R. (1987). *Comprehension of mathematical relationships expressed in graphs.* Journal for Research in Mathematics Educations, 18, 382-393.

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

DOWNING, Douglas; CLARK, Jeffrey. *Estatística Aplicada. Série essencial. São Paulo: Saraiva, 1998.*

DRUCKER, Peter F. *Administrando em tempos de grandes mudanças 5ª Ed.* São Paulo: Pioneira, 1998.

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

GAL, I (1995) *Statistical tools and Statistical Literacy: The case of the average.* Teaching Statistics, 17(3), 97-99.

GOKHALE, A. A. *Effectiveness of Computer Simulation for Enhancing Higher order Thinking. Journal of Industrial Teacher Education, [S.I.:s.n.], v. 33, p. 36-46. 1996.* Digital Library and Archives. Disponível em: <http://scholar.lib.vt.edu/ejournals/JITE/v33n4/jite-v33n4.gokhale.html>. Acesso em: 20 mai. 2000;

HAND, D. J. (1998). *Breaking Misconceptions - Statistics and its Relationship to Mathematics. The Statistician, 47 (2), 245-250.*

HEID, M. K et al. *Advanced Mathematical Thinking: Implications of various perspectives on advanced mathematical thinking for mathematics education reform*, acessado em 05 de setembro de 2004 em http://www.DefSOnat.psu.edu/stafftmlklmkh2lAMTheid.00f

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

MACHLINE, Claude et al. *Manual de Administração da produção*, 8 Ed. Rio de Janeiro: FGV, 1994, V.2.

MAGINA S. (2001). *Repensando Adição e Subtração – Contribuições da Teoria dos Campos Conceituais* – Ed. PROEM. 2001 – 5 -17.

MILONE, Giuseppe; ANGELINI, Flávio. *Estatística Geral, Vol. 1 e 2*, São Paulo: Atlas, 1995.

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

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

MINISTÉRIO DA EDUCAÇÃO. *Secretaria da Educação Média e Tecnológica. Parâmetros Curriculares Nacionais. Brasília: Ministério da Educação, 1999.*

NOVAES, Diva Valério. *A Mobilização de Conceitos estatísticos: Estudo exploratório com Alunos de um Curso de Tecnologia em Turismo.* Dissertação de Mestrado em Educação Matemática – PUC/SP – 2004.

PEREIRA, B. B. (1997). *Estatística: A tecnologia da Ciência.* Boletim da Associação Brasileira de Estatística, ano XIII, no. 37, 2º Quadrimestre.

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

POTTER, A. M. *Statistics for Sociologists: Teaching Techniques that Work.* Teaching Sociology, v. , 23, july, 1995.

ROBERT, A. *Outis D'analyse dês Contenus Mathématiques á ensiigner au lycée á l'Université.* RECHERCHES EM DIDACTIQUE DÊS MATÉMATIQUES, Vol. 18, nº2, pp. 139-190. 1998

ROITER, K., PETOCZ, P. *Introductory statistics courses-A new way of thinking*. Journal of Statistics Education., v. 4, n. 2, l996.

RUBERG, S. J. e MASON, R. L.. *Increasing public awareness of Statistics as a science and profession starting in high school.* The American Statistician, 42 (3), 167-170. (1988)

SALIBY, E. *Software para Simulação, [S.I.:s.n.].* jun. 2000. Disponível em: <http://www.coppead.ufrj.br/pesquisa/cel/po_psal.htm>. Acesso em: 09 jun. 2000;

SCHANK, R. *Using Simulators to Teach. [S.I.:s.n.],* jul. 2000. Disponível em: <http://www.ils.nwu.edu/~e_for_e/nodes/NODE-125-pg.html>. Acesso em: 20 jul. 2000.

SILVA C., CAZORLA, IRENE, BRITO, FERREIRA M. R. *Concepções e Atitudes em relação à Estatística.* In: Conferência Internacional "Experiências e Perspectivas do Ensino da Estatística: Desafio para o século XXI", 1999. Anais da Conferência Internacional "Experiências e Perspectivas do Ensino da Estatística: Desafio para o século XXI", 1999 p. 18-29, 1999.

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

SFARD, A. & LINCHEVSKI, L. (1994a). *The gains and the pitfalls of reitification: The case of algebra.* Educational Studies in Mathematics, 26, 191-228.

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

SPRENT, P. (1998). *Statistics and Mathematics - Trouble at the Interface?* The Statistician, 47 (2), 239-244.

STENVENSON, Willian J. *Estatística Aplicada a Administração. São Paulo.* Harbra, 1986.

STUART, M. (1995). *Changing the teaching of Statistics.* The Statistician, 44 (1), 45-54.

TOLEDO, Luciano Geraldo; OVALLE, Ivo Izidoro. *Estatística Básica.* 2 Ed. São Paulo: Atlas, 1995.

VENDRAMINI, C. M. M. *Implicações das Atitudes e das Habilidades Matemáticas na Aprendizagem dos Conceitos Estatísticos.* Tese de Doutorado, UNICAMP/SP, 2000.

VENDRAMINI, C.M.M., SILVA, M. C., & CANALE. M. (2003). *Análise de itens de uma prova de raciocínio estatístico.* Artigo submetido à revista Pscicologia em Estudo, Maringá.

VENDRAMINI, C. M. M., CHENTA, V. C., & SILVA, L. S. (2004). *Leitura de dados estatísticos: Um estudo com alunos do ensino fundamental.* Texto não publicado.

VERGNAUD, G. *A Comprehensive Theory of Representation for Mathematics Education.* IMB V17, N2, pp. 167 – 181, 1998.

VLAHOS, Kiriakos. *Minimizando o risco quanto a incerteza. In: Dominando a Administração:* Financial Times: São Paulo: Makron Brooks, 1999.

WAINER, H. (1992). *Understanding Graphs and Tables*. Educational Researcher, 21(I), 14 –23.

WALLMAN, K. K. (1993). *Enhancing Statistical Literacy: Enriching our society.* Journal of the American Statistical Association, 88, 1-8.

WERKEMA, Maria Cristina C. *Como estabelecer conclusões com confiança: entendendo a inferência estatística.* Belo Horizonte: Fundação Christiano Ottoni, Escola de Engenharia – UFMG, 1996.

APÊNDICES

APÊNDICE 1

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	Idade	Renda Mensal	Idade	Renda Mensal	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.

APÊNDICE 2

Segunda parte

De acordo com as tabelas abaixo, Responda às questões:

Números de carros por pessoa entrevistada

Números de carros	***Números de pessoas***
1	10
2	25
3	15
4	5
Total	***55***

Número de horas no trânsito por pessoa

Tempo no trânsito	***Número de pessoas***
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.

APÊNDICE 3

Terceira parte

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

Gráfico 1

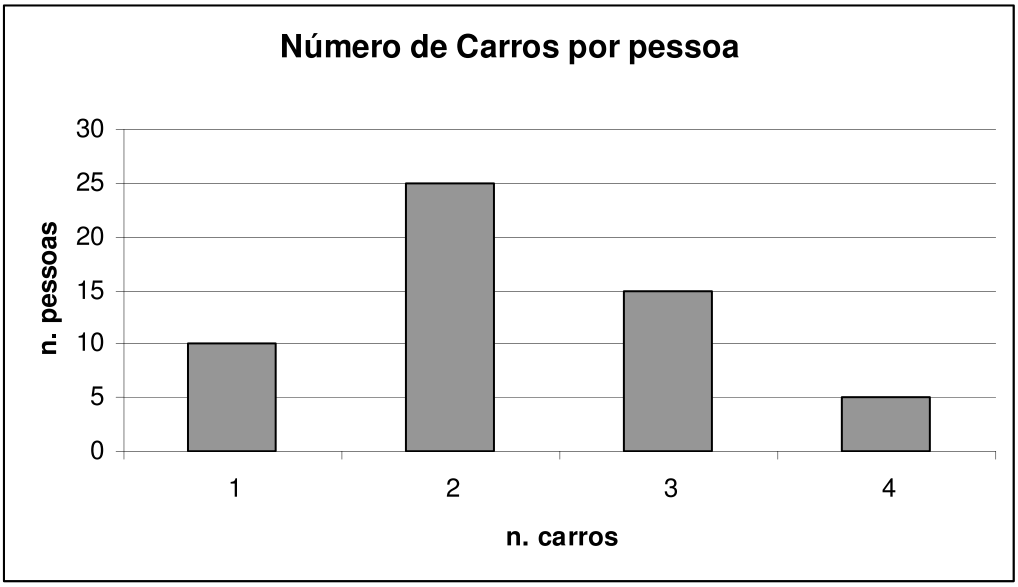


Gráfico 2