

ANÁLISE EXPLORATÓRIA DE DADOS: UM ESTUDO DIAGNÓSTICO SOBRE CONCEPÇÕES DE PROFESSORES

COUTINHO¹, Cileda de Queiroz e Silva – PUC-SP – cileda@pucsp.br

MIGUEL², Maria Inez Rodrigues – PUC-SP – minez@pucsp.br

GT: Educação Matemática / n.19

Agência Financiadora: Sem Financiamento

Introdução

A Análise Exploratória de Dados é o pano de fundo do presente trabalho. Com as mudanças curriculares ocorridas na última década em diversos países, tais como E.U.A. (National Council of Teachers of Mathematics), Inglaterra (Curriculum Nacional de Inglaterra y Gales) e Espanha (Diseño Curricular Base), assim como no Brasil (Parâmetros Curriculares Nacionais), que trouxeram a Estatística e a Probabilidade para a Escola Básica, vieram, também, alterações no enfoque de apresentação destas disciplinas. Tais mudanças apontam para uma passagem de um enfoque tecnicista, no qual o aspecto matemático é predominante e os modelos são pré-estabelecidos, para um analítico, no qual se busca construir modelos a partir do estudo de dados observados

A análise exploratória de dados é uma filosofia que consiste

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

Nesta perspectiva, busca-se identificar quais conceitos e idéias constituem a base para a aplicação e implementação dessa filosofia, focando o trabalho com a Estatística na Escola Básica. Neste sentido, deve-se considerar a utilidade, para os alunos, dos conceitos e procedimentos que serão então desenvolvidos. Para estes autores, o trabalho com a filosofia da Análise Exploratória de Dados, precisa:

- possibilitar a geração de situações de aprendizagem contextualizadas em temas que sejam de interesse para o aluno;
- lançar mão de um forte apoio às representações gráficas (que facilitam a percepção da variabilidade no conjunto de dados observados);
- empregar, preferencialmente, as estatísticas de ordem (que aportam maior facilidade na atribuição de significado pelo aluno da Escola Básica);
- utilizar diferentes escalas de categorização das variáveis para o estudo dos dados observados;

¹ Programa de Estudos Pós-graduados em Educação Matemática

² Programa de Estudos Pós-graduados em Educação Matemática

- não utilizar (pois não é necessário) ferramentas matemáticas complexas.

Desta forma, para o trabalho nessa filosofia, é necessário, como afirmado acima, que o aluno possa perceber a problematização envolvida na coleta das informações (que ele mesmo pode fazer). Este é o cerne da atribuição de significado para os resultados obtidos nos cálculos estatísticos visando à análise, pois, somente dessa forma, o aluno poderá conceber o correto significado aos valores encontrados.

Para Gal (2002), as informações estatísticas podem ser representadas, basicamente, em três formatos: língua natural, símbolos ou números e tabelas ou gráficos; usualmente, utilizam-se combinações de mais de uma destas formas. Vê-se, assim, o reforço da necessidade do contexto escolhido para o nível dos alunos, uma vez que precisam analisar e interpretar tais representações adequadamente. Este autor aponta que outros pesquisadores mostram que documentos com mensagens matemáticas, ou estatísticas, apresentam o uso de termos específicos que podem ser interpretados erroneamente em seu significado coloquial. Alerta, ainda, que as mensagens podem conter jargões estatísticos, o que pode provocar ambigüidade. A partir dessas colocações, pode-se inferir a necessidade de um trabalho apurado, na introdução da Análise Exploratória como componente curricular da Escola Básica, visando não apenas o uso adequado de tais termos, mas, também, a preparação do sujeito para uma leitura adequada, nos diversos contextos da vida particular ou profissional.

Para que se implemente tal componente curricular, Gal (2002) aponta cinco elementos que são assumidos como conhecimentos estatísticos de base:

- conhecimento da razão da necessidade dos dados e como estes podem ser produzidos;
- familiaridade com termos e idéias básicos da estatística descritiva;
- familiaridade com termos e idéias básicos da representação gráfica e tabular;
- compreensão de noções básicas de probabilidade;
- conhecimento de métodos de obtenção de inferências ou conclusões estatísticas.

Neste contexto, é fácil perceber a importância do papel do professor na organização e gestão das atividades que permitam aos alunos a construção dos conceitos

e das ferramentas necessárias à Análise Exploratória de Dados. Neste sentido, o objetivo do presente estudo é identificar concepções³ docentes sobre tal tema.

Aspectos teóricos e metodológicos

Para atingir o objetivo pretendido, anunciado acima, foi elaborado um instrumento diagnóstico na forma de um questionário, que foi tratado com a ajuda do software CHIC (Classificação Hierárquica, Implicativa e Coesitiva), particularmente a análise de similaridades e coesitiva.

Trinta e três professores de Matemática da Escola Básica, sendo 25 deles alunos de um mestrado profissional da cidade de São Paulo e os demais participantes de um projeto de pesquisa que inclui formação continuada na mesma instituição.

O questionário (apêndice A) foi organizado em três partes; a primeira delas visa construir o perfil dos respondentes, a segunda busca identificar o grau de concordância que os professores atribuem aos itens propostos por Gal (2002) em relação à alfabetização estatística e a terceira envolve análise didática e matemática de situações-problema sobre o tema.

A análise das informações colhidas com o grupo de professores foi amparada por um *software*, denominado CHIC (Classificação Hierárquica, Implicativa e Coesitiva). Esse *software* foi desenvolvido por um grupo de pesquisadores coordenados por Régis Gras, no início dos anos 1990, com o intuito de sintetizar e estruturar as respostas dos professores para obter uma tipologia de comportamentos, por meio de um tratamento multidimensional de dados estatísticos.

O *software* CHIC tem por funções essenciais extrair de um conjunto de dados, cruzando sujeitos e variáveis (ou atributos), regras de associação entre variáveis, fornecer um índice de qualidade de associação e de representar uma estruturação das variáveis obtida por meio destas regras. (COUTURIER; BODIN; GRAS, p.1 do texto de ajuda do software).

Segundo Almouloud (2005), quando se trata de um exemplo de pesquisa em educação,

[...] na análise qualitativa das informações, no intuito de tomar decisões que se apóiam em uma certa estabilidade e pertinência de respostas, o pesquisador, muitas vezes, recorre às análises estatísticas de dados multidimensionais. Essas análises permitem:

³ Segundo Thompson (1992, p.232, apud Azcárate, 1996, p.41), uma concepção é “uma estrutura mental de caráter geral, que inclui crenças, conceitos, significados, regras, imagens mentais e preferências, conscientes ou inconscientes”.

- sintetizar e estruturar os dados multidimensionais a fim de identificar as variáveis estatísticas (e/ou didáticas), os fatores em jogo, suas relações, sua hierarquia, etc.
- evidenciar a dinâmica dos comportamentos de alunos ou professores em situação de resolução de problemas. (ALMOULOU, 2005)

Desta forma, considerando-se o conjunto V de atributos (ou caracteres) e o conjunto E de sujeitos que possuem ou não estes atributos, deve-se utilizar ferramentas que permitam, simultaneamente:

- tratar diferentes tipos de variáveis (binárias, modais, frequenciais, intervalares);
- quantificar a significação dos valores atribuídos à qualidade das associações, consistência da(s) regra(s) que as determinaram, da ordenação das classes resultantes, a tipicidade e contribuição de sujeitos ou categorias de sujeitos à constituição dessas classes;
- representar, por um gráfico, tendo fixado um intervalo de confiança, caminho de regras e por uma hierarquia de regras sobre regras;
- suprimir, acrescentar variáveis.

O CHIC foi desenvolvido com estes objetivos, e pode trabalhar com variáveis binárias, frequenciais e intervalares. Nosso estudo limita-se ao uso das binárias e a associação entre elas pode ser feita por grau de similaridade e/ou de implicação.

[...] a análise hierárquica de similaridade permite constituir, por meio de um critério, partições cada vez mais finas sobre um conjunto de variáveis estatísticas. Tais partições são construídas de modo ascendente em uma árvore permitindo estudar e interpretar, em termos de tipologia e semelhança (dessemelhança), classes de variáveis. Na análise implicativa dos dados, chega-se a estruturas implicativas no sentido de que uma atitude *a* tem como consequência, ou não, uma atitude *b* ($a \rightarrow b$). (ALMOULOU, 2005, p.3).

Nessa análise, se consideram os conjuntos A (sujeitos com a característica “a”) e B (sujeitos com a característica “b”) contidos em um conjunto E (conjunto dos *n* sujeitos pesquisados), sendo n_a e n_b o número de sujeitos em cada conjunto, respectivamente.

O critério de similaridade exprime-se da seguinte maneira nos casos das variáveis binárias: duas variáveis têm comportamentos semelhantes, quando o número *k*, dos sujeitos que os verificam simultaneamente (elementos de $A \cap B$), é suficientemente grande quando comparado, por um lado, ao valor obtido no caso da ausência de relação entre “a” e “b” e, por outro lado, ao número de elementos n_a e n_b .

Medimos esta semelhança pela probabilidade de “que k seja superior ao número aleatório esperado nesta situação”. Para isso, considera-se um conjunto Y, tomado aleatoriamente entre os elementos do conjunto das partes de E, tal que Y e B tenham a mesma cardinalidade ($n_b = n_y$), como representado na Figura 1.

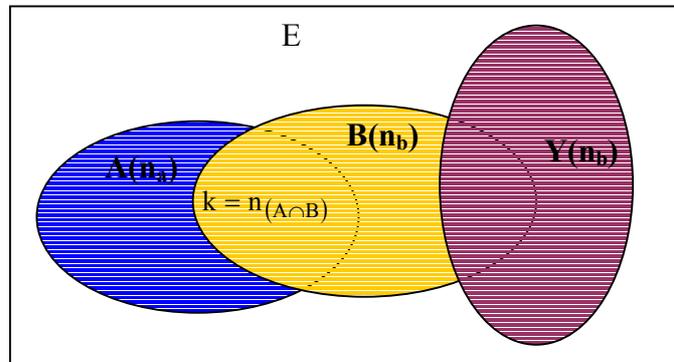


Figura 1. Conjunto Y - aleatório e tal que $n_y = n_b$

O número de elementos da intersecção, $A \cap Y$, designado por s_a é, para Lerman (1970), uma variável aleatória (S_a) sob a hipótese de ausência de relação a priori entre A e Y. Nestas condições, a similaridade entre “a” e “b” cresce à medida que o valor observado k for maior que o valor esperado de S_a ; o índice inicial de similaridade será dado pela probabilidade $P(S_a < k)$.

Considera-se, a seguir, um conjunto X, escolhido aleatoriamente entre os elementos do conjunto das partes de E que tenham a mesma cardinalidade de A ($n_x = n_a$). Da mesma forma, se definimos s_b como sendo o número de elementos da intersecção entre X e B, tem-se que S_b é uma variável aleatória com a mesma distribuição de S_a . Segundo Almouloud (1992),

[...] mostra-se que, para n suficientemente grande, a variável aleatória

$$S(a, b) = \frac{S_a - \mu}{\sigma} = \frac{S_b - \mu}{\sigma} \text{ segue uma lei normal padrão, onde}$$

$$\mu = \frac{n_a \cdot n_b}{n} \text{ e } \sigma^2 = \frac{n_a \cdot n_a^- \cdot n_b \cdot n_b^-}{n^2 \cdot (n-1)} \text{ [...] considerando que as}$$

associações positivas entre “a” e “b”, que servem para caracterizar a proximidade entre “a” e “b” (eventos raros) o índice poissoniano é preferível ao binomial ou hipergeométrico, aos olhos das análises

$$\text{didáticas e é dado por: } s_p = \frac{n(A \cap b) - \frac{n_a \cdot n_b}{n}}{\sqrt{\frac{n_a \cdot n_b}{n}}}. \text{ (ALMOULOU,}$$

1992, p.8-9).

O índice de similaridade entre variáveis serve, na seqüência do estudo, para definir um índice entre duas classes de variáveis, segundo esse mesmo princípio de comparação entre a observação e o que deve ser considerado devido ao acaso.

Assim, como em todos os métodos de classificação, procura-se constituir, em um conjunto V das variáveis, partições cada vez menos finas, construídas de maneira ascendente. Essas partições encaixadas são representadas por uma árvore obtida a partir de um critério de similaridade entre variáveis. A similaridade é definida a partir do cruzamento do conjunto V das variáveis com um conjunto E de sujeitos (ou de objetos). Este tipo de análise permite ao usuário estudar e depois interpretar, em termos de tipologia e de semelhança (ou não semelhança) decrescente, classes de variáveis, constituídas significativamente a certos níveis (da árvore) e se opondo a outros nestes mesmos níveis. Assim, ao dizermos que as variáveis V_1 e V_2 , por exemplo, são agrupadas pelo critério de similaridade, estamos afirmando que os sujeitos que respondem V_1 têm comportamento similar aos que respondem V_2 . Pode-se calcular o índice de similaridade entre as variáveis, que é a probabilidade de que efetivamente os dois grupos tenham comportamento similar, ou identificar quais as características típicas do grupo no qual essa similaridade foi identificada.

Dessa forma, para construir uma árvore de similaridades, reúnem-se em uma classe de primeiro nível as duas variáveis que se parecem mais, no sentido do índice de similaridade, depois duas outras variáveis ou uma variável e a classe já formada e, sucessivamente, outras variáveis ou classes de variáveis.

Análise dos resultados

Em um primeiro momento na análise dos dados, busca-se traçar a caracterização dos professores que responderam o questionário, quanto ao gênero, idade e tempo que leciona matemática. Os demais itens da parte A do instrumento não foram discriminantes e por isso não serão abordados.

Assim, entre os 33 respondentes, 17 são do sexo feminino e 16 do sexo masculino, com as idades distribuídas conforme Tabela 1.

Tabela 1. Idade dos 33 professores que responderam ao questionário

Idade (anos completos)	Número de professores	
	N_i	%
21 a 28	5	15,15

29 a 33	7	21,21
34 a 40	8	24,24
41 a 50	6	18,18
Mais de 50	7	21,21
TOTAL	33	100,00

As faixas etárias discriminadas no questionário foram determinadas a partir de Sikes et al (1985), citado em Bolívar (2002) e constam na Tabela 2.

Tabela 2: Categorias de vida profissional segundo a idade do professor

Idades (anos completos)	Características
21 a 28	A entrada na profissão
29 a 33	Transposição/crise dos 30
34 a 40	Estabilização e compromisso
41 a 50	Crise na metade da carreira
Mais de 50	Recapitulação da carreira

Quanto ao tempo que leciona ou lecionou matemática, os resultados estão apresentados na Tabela 3.

As faixas discriminadas para o tempo de magistério foram determinadas a partir de Huberman (1990) e estão sintetizadas na Tabela 4.

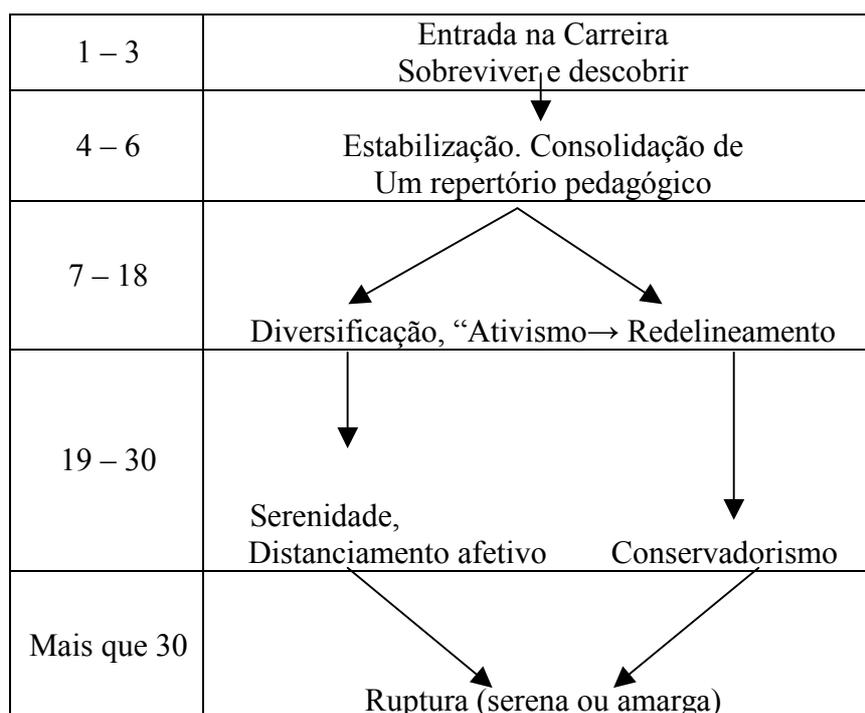
Passando para a parte B do instrumento (Apêndice A), a Tabela 5 apresenta as frequências observadas para cada uma das possibilidades de respostas.

Tabela 3. Tempo de magistério dos professores que responderam ao questionário

Tempo de Magistério (anos completos)	Número de professores	
	N _i	%
Até 6	7	21,21
De 7 a 18	20	60,61
De 19 a 40	6	18,18
TOTAL	33	100,00

Tabela 4: Categorias de vida profissional, segundo o tempo de magistério

Anos de Carreira	Fases/Temas
------------------	-------------



Fonte: Huberman (1990), apud Bolivar (2002)

Pode-se perceber, pela leitura da Tabela 5 que, embora a maioria dos professores concorde plenamente com muitos itens, que retratam etapas da formação do pensamento estatístico, tem-se vários professores que não apresentam a mesma opinião. Uma análise aprofundada das relações entre as opiniões destes professores e os itens de perfil fará parte da seqüência deste texto, com auxílio do software CHIC.

Os dados levantados na parte B do questionário foram devidamente codificados e tabulados para tratamento com o CHIC. A opção para a análise de similaridades foi a teoria clássica (usual para populações menores do que 1000) e a distribuição binomial, obtendo-se a árvore apresentada na Figura 2.

Tabela 5: Frequência do grau de concordância das afirmações da parte B do questionário

Para planejar o desenvolvimento de aulas, (em qualquer nível de ensino) que tenham por objetivo trabalhar conceitos da estatística descritiva, o professor deve:	F r1	F r2	F r3	Total
q1b possibilitar a geração de situações de aprendizagem contextualizadas em temas que sejam de interesse do aluno;	1	8	24	33
q2b lançar mão de representações gráficas, que, entre outros aspectos, podem facilitar a percepção da variabilidade em um conjunto de dados observados;	8	6	19	33
q3b empregar, preferencialmente, as estatísticas de ordem ⁴ , que aportam maior facilidade na atribuição de significado pelo aluno;	5	14	13	33

⁴ Estatísticas de ordem: mediana, quartis, decis, percentis.

q4b não utilizar (pois não é necessário) fórmulas que exijam manipulações matemáticas complexas.	7	22	4	33
q5b considerar a necessidade dos alunos fazerem a coleta dos dados e como estes dados podem ser produzidos;	1	6	26	33
q6b considerar a necessidade de familiaridade com termos e idéias básicos da estatística descritiva, como média, moda, mediana, quartis, variância, desvio médio e desvio padrão;	2	7	23	33
q7b considerar a necessidade de familiaridade com termos e idéias básicos da representação gráfica e tabular, como frequência absoluta e relativa, frequência acumulada, gráfico de barras, colunas, de setor, histograma, ogiva, etc.;	1	12	19	33
q8b considerar a compreensão de noções básicas de probabilidade;	1	11	20	33
q9b considerar o estudo descritivo para a obtenção de inferências ou conclusões estatísticas.	3	14	14	33

Nota: (r1) corresponde a discordância total; (r2) concordância parcial e (r3) concordância plena.

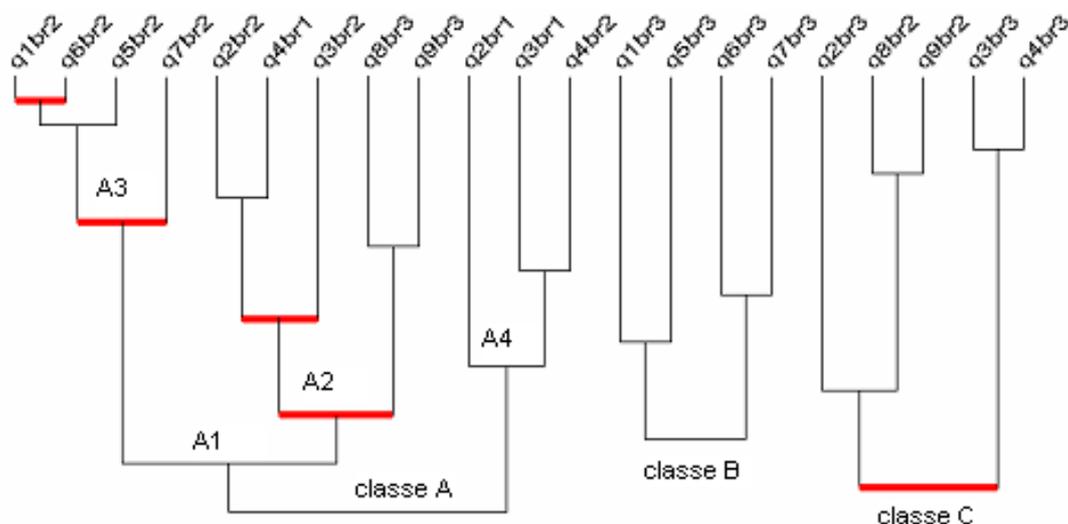
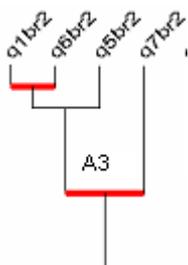


Figura 1. Árvore de Similaridades

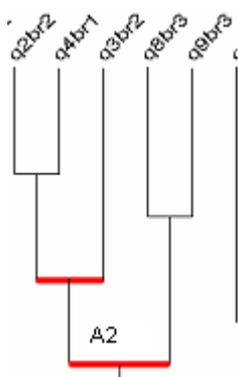
do

os agrupamentos feitos pelo *software*. A primeira delas é formada, ainda, por três grandes subclasses, A1 (composta por A2 e A3) e A4. O início da análise refere-se à subclasse com maior índice de similaridade, A3, que apresenta dois nós significativos. Definimos como um nó significativo “àquelas associações significativas de uma árvore de similaridades, ou seja, as que correspondem a uma classificação com a melhor compatibilidade em relação aos valores obtidos e à qualidade dos valores das similaridades”. (COUTURIER; BODIN; GRAS, p.1 da ajuda do software).



Analisando primeiramente a subclasse A3, observa-se como característica deste grupo a concordância parcial com itens que discutem a aprendizagem dos conceitos de base da estatística descritiva, tais como: a geração de situações de aprendizagem contextualizadas em temas que sejam de interesse do aluno (q1br2), considerar a necessidade de familiaridade com termos e idéias básicos da estatística descritiva (q6br2), que deve considerar a necessidade de os alunos fazerem a coleta de dados e refletirem em como estes dados podem ser produzidos (q5br2), e a necessidade da familiaridade com termos e idéias básicos da representação tabular (q7br2). Os professores com mais de 50 anos de idade são o que mais contribuíram para a identificação dos agrupamentos até aqui obtidos, com risco de 0,204 e, como tal são definidos no CHIC como variável típica do agrupamento. Entende-se como risco a probabilidade de se cometer um erro na afirmação feita.

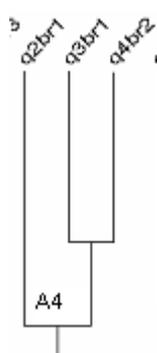
Observa-se assim que estes professores podem, efetivamente, ser favoráveis ao trabalho com a estatística em suas aulas dentro de princípios que podem encaminhar para a filosofia da Análise Exploratória de Dados e permitindo ao aluno ocupar uma posição efetivamente de aprendiz ao invés de reproduzir técnicas apresentadas previamente.



A subclasse A2 é formada por quatro sub-agrupamentos, que se caracterizam por uma visão mais intuitiva na abordagem dos conceitos de base da estatística descritiva. Desta forma, estes professores concordam parcialmente que o professor deve lançar mão de representações gráficas para facilitar a percepção da variabilidade do conjunto de dados observados (q2br2), concordam com a não necessidade do uso de fórmulas que exijam manipulações matemáticas complexas (q4br1) e concordam parcialmente que se devem empregar, preferencialmente, as estatísticas de ordem para maior facilidade na atribuição de significados pelo aluno. Por outro lado, eles discordam totalmente que se deve considerar, tanto a compreensão de noções básicas de probabilidade (q8br3), e quanto ao estudo descritivo para a obtenção de inferências ou conclusões estatísticas (q9br3). Observe-se que os dois aspectos de discordância estão ligados à inferência estatística, reforçando a característica de uma visão intuitiva dentro desta classe. A variável típica é aquela composta por professores que cursaram o Ensino Médio em cursos profissionalizantes, magistério ou supletivo, com risco de 0,128.

Observa-se assim que este grupo de professores que mostram uma visão mais intuitiva da Estatística Descritiva também tem condições de abordar a Estatística em sala de aula com uma visão da Análise Exploratória de Dados. No entanto, sabe-se que para uma aprendizagem efetiva a visão intuitiva não é suficiente e que, portanto, é preciso apresentar aos alunos o conceito em suas várias representações e vários enfoques, não fugindo aos aspectos formais necessários, como por exemplo, as noções básicas de probabilidade. Estas noções podem ser abordadas desde as séries iniciais da escolaridade, conforme resultados de pesquisas e orientações dos Parâmetros Curriculares Nacionais para o Ensino Fundamental.

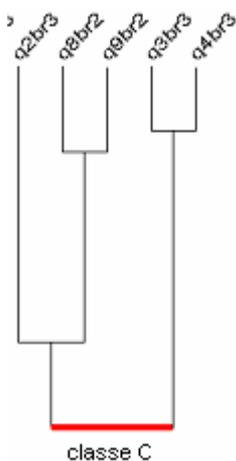
As subclasses A2 e A3, já descritas, se associam formando a subclasse A1, com similaridade 0,273. Baixos índices de similaridade podem ser vistos também como índices de dissimilaridade, ou seja, a não semelhança entre os grupos A2 e A3. A variável típica de A1 refere-se a professores com até seis anos de magistério de matemática, com risco de 0,118. Neste caso, a análise feita indica que os professores com menos tempo de magistério em matemática não possuem comportamentos semelhantes quanto aos conceitos de base da Estatística Descritiva e à abordagem intuitiva destes conceitos.



Passemos à subclasse A4. Os professores que discordam completamente que se devem empregar preferencialmente as estatísticas de ordem (q3br1), concordam parcialmente que não se devem utilizar fórmulas com alto grau de complexidade matemática (q4br2) e discordam totalmente que se deve lançar mão de representações gráficas para facilitar percepção da variabilidade (q2br1). A variável típica dessa subclasse A4 corresponde a professores com mais que 50 anos de idade, com risco de 0,194. Percebe-se nesses agrupamentos uma atitude quase negativa quanto à abordagem da Estatística, evidenciada pela negação no uso de estatísticas de ordem e das representações gráficas que, conforme resultados de pesquisas na área, tem baixo custo cognitivo devido à sua pouca complexidade. Podemos levantar a hipótese de que a idade destes professores é determinante desta postura, pois em sua formação inicial provavelmente tiveram seu contato com a Estatística em uma abordagem excessivamente tecnicista, usual no ensino da época.

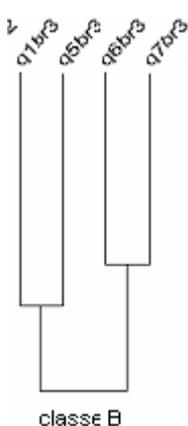
A subclasse A4 se associa com a subclasse A1, formando a classe A, com índice 0,026, indicando efetivamente um grau de dissimilaridade entre elas. A variável típica

dessa dissimilaridade é o grupo de professores com até seis anos de magistério de matemática, com risco de 0,118.



A classe C é caracterizada pela concordância total ou parcial com termos fundamentais para a Análise Exploratória de dados. Assim, concordam completamente tanto com o fato de que se devem empregar preferencialmente as estatísticas de ordem (q3br3) como com a não utilização de fórmulas com muita complexidade matemática na manipulação (q4br3). Concordam parcialmente com o fato de que se deve considerar a compreensão das noções básicas de probabilidade (q8br2), concordam parcialmente com o fato de que se deve considerar o estudo descritivo para a obtenção de inferências ou conclusões estatísticas (q9br2) e concordam plenamente que se deve lançar mão das representações gráficas para facilitar a percepção da variabilidade dos dados. Os dois grupos se relacionam formando a classe C, com similaridade igual a 0,182 (dissimilaridade). A variável típica dessa dissimilaridade é formada pelos professores com idades entre 41 e 50 anos, com risco de 0,169.

Percebe-se nessa classe um agrupamento de professores que mesmo tendo sido formados em uma época de enfoque tecnicista para a abordagem da estatística, possuem opiniões favoráveis à adoção da filosofia da Análise Exploratória de Dados, evidenciando um comportamento coerente com a crise na metade da carreira (41 a 50 anos) identificada por Sikes et al (1985, apud Bolivar, 2002).



A última classe a ser analisada, classe B, é caracterizada pela concordância plena com itens da Análise Exploratória de dados: os professores que concordam plenamente que se deve considerar a familiaridade com termos e idéias da estatística descritiva (q6br3); que se deve considerar a familiaridade com termos e idéias básicos da representação gráfica e tabular (q7br3); que se deve possibilitar a geração de situações de aprendizagem contextualizadas em temas que sejam de interesse dos alunos (q1br3) e que se deve considerar a necessidade dos alunos fazerem a coleta dos dados, considerando como estes podem ser produzidos (q5br3). A variável típica a essa classe é formada por professores do gênero feminino, com risco 0,091.

Observa-se assim que as professoras que responderam ao questionário têm uma posição bastante definida quanto à abordagem de conceitos estatísticos em sala de aula, abordagem essa que favorece a adoção da filosofia da Análise Exploratória de Dados.

Em relação ao diagnóstico correspondente à Parte C do questionário, decidimos por sua não inclusão no banco de dados a ser tratado pelo CHIC, pela ausência de diferenças significativas que pudessem sugerir relações de interesse para nossa análise. Assim sendo, no que se segue, apresentamos uma síntese do que foi explicitado pelos professores participantes da pesquisa e nossa reflexão a respeito.

Dois sujeitos, sendo um professor do Ensino Médio e outro do Ensino Fundamental II, não responderam essa parte do questionário, indicando um completo desconhecimento do conteúdo abordado. Diferentemente, outros oito, dentre os quais um deles é professor do Ensino Superior, explicitaram que nunca ministraram aulas sobre o tema e que, portanto, não teriam condições de opinar sobre sua organização didática ou matemática.

Dentre os 23 respondentes, foi praticamente unânime a necessidade de inclusão da frequência relativa, representada em porcentagem que, além de familiar aos estudantes, possibilita a construção do gráfico de setores. Em relação a este último, citado pela grande maioria dos professores, foram mencionadas as dificuldades em relação à determinação dos ângulos e manuseio de régua e transferidor. Este tipo de dificuldade pode indicar que, de um ponto de vista didático, este gráfico só é abordado para leitura dos dados e não para construção, conforme apresentação na maioria dos livros didáticos hoje adotados. Vale lembrar que mesmo essa leitura é, quase sempre, desprovida de análise crítica.

Porcentagem é um tema de estudo obrigatório, que tem como aplicação a construção de gráficos de setor, que permite trabalhar medida de ângulo e o uso do transferidor. Porém, todo esse estudo não se refere ao interesse específico da estatística, que deve vir a seguir, como por exemplo, se existe valor da variável predominantemente, valores equivalentes, visualização dos possíveis valores, etc. As respostas nos questionários indicam que esses professores podem estar considerando que a finalidade da análise dos dados é construir o gráfico de setores, talvez, porque eles tenham conhecimento e domínio apenas sobre os conteúdos citados para sua elaboração.

A leitura dos questionários permite afirmar que o conhecimento a respeito da Estatística Descritiva limita-se a coleta de dados e respectiva construção de tabelas (com

freqüência absoluta, relativa e acumulada) e gráficos (barra, coluna e setor). Aqueles que se referiram às medidas de tendência e dispersão foram unânimes em afirmar que ensinam a aplicar fórmulas e que a dificuldade dos alunos está nos cálculos. Alguns chegaram a citar que o trabalho poderia ser facilitado com o uso de calculadoras e computador, mas que, dessa forma, os conceitos ficam sem significado (termo utilizado pelos professores). Observa-se aqui uma tendência para a abordagem tecnicista, o que é coerente com os resultados identificados pelo uso do CHIC.

Nenhuma das respostas apresentou a solução matemática das atividades propostas. Apenas uma limitou-se ao cálculo da média, variância e desvio padrão, sendo os dois últimos com valores absurdos, incompatíveis com os dados. Outros indicaram como fórmulas para o cálculo da média, para a atividade 2, $\sum \frac{N_i}{4}$ e para a atividade 1, $\frac{\sum f_i}{16}$. A opção em utilizar, no questionário, N_i para representar a freqüência foi fundamentada em obras reconhecidas como referência na área; mesmo que essa notação possa não ser usual nos livros didáticos, a simples observação das tabelas apresentadas e o conhecimento do significado do conceito de freqüência não deveriam causar equívocos. Questiona-se, portanto, a leitura que os professores fizeram das informações das tabelas apresentadas.

A ausência da solução matemática pode ser interpretada como sendo apenas uma omissão, mas os comentários registrados indicam que o motivo é o desconhecimento total do conteúdo. Alguns explicitaram que não ensinam mediana e quartis porque é muito difícil para os alunos compreenderem as fórmulas e outros porque não sabem esse conteúdo. Note-se que, entre os 23 respondentes, apenas cinco professores discordam totalmente com o uso de estatísticas de ordem, enquanto os demais se dividem, quase igualmente, entre a concordância total e a parcial deste item. Levantamos assim uma questão: como concordar em ensinar algo que não dominamos?

Nossa reflexão a esse respeito é que há carência de material didático com opções de ensino que abordem a determinação da mediana e dos quartis sem o uso de fórmulas, o que pode favorecer o seu ensino e aprendizagem. Ainda mais, a forma que os livros abordam tais conceitos, não é clara o suficiente a ponto de um professor responder na atividade 2, que a mediana é 13,5, desconsiderando completamente as freqüências.

Nenhum professor comentou a apresentação das duas atividades que tinham o objetivo didático de desenvolver os conceitos inicialmente com “poucos” elementos

para depois extrapolar para “muitos”, nem tampouco, a importância e diferença das três variáveis da atividade 1 e dos quatro exemplos da atividade 2, com valores distintos das variáveis didáticas: valor da variável e número de elementos.

Várias interpretações podem ser dadas para justificar essas ausências. Uma delas é o desconhecimento das medidas de tendência central e dispersão; outra pode ser a inexperiência em preparar seu próprio material didático, como uma seqüência de ensino para desenvolver os conteúdos em suas aulas; acrescenta-se a isso a falta de tempo pelo acúmulo de aulas e o hábito em seguir o livro didático sem questionamento.

Silva (2005), Rossini (2006) e Silva (2007) desenvolveram experimentos em que os professores foram organizados em grupos para elaborarem seqüências de ensino sobre frações, funções e estatística, respectivamente. No primeiro caso, a pesquisadora acabou reorganizando as atividades propostas pelos professores a fim de obter uma seqüência didática viável de ser aplicada aos alunos; já no segundo, os mesmos professores foram capazes de elaborar e aplicar uma seqüência sobre funções, depois de um trabalho de exaustivas tentativas, no entanto, no terceiro caso, o mesmo grupo não conseguiu elaborar as atividades de estatística, alegando que o tempo de formação não foi suficiente para que se sentissem seguros em relação ao conteúdo a ser abordado.

Essas pesquisas e as respostas dos questionários mostram a carência, nos cursos de graduação, de um trabalho na formação inicial, relacionado à análise crítica dos livros didáticos e elaboração de seqüências de ensino, mas que, com insistência e amparo é uma habilidade e competência que pode ser desenvolvida, não só em relação à análise de dados, mas em qualquer tema.

Considerações finais

Pode-se perceber: um discurso docente predominantemente favorável aos itens que compõem a filosofia da Análise Exploratória de Dados. No entanto, as respostas ao bloco C do questionário e as práticas observadas usualmente nos dias de hoje mostram que este tema não tem sido trabalhado em sala de aula. Como exemplo, os resultados obtidos em Moraes (2006), Bigattão (2007) e Cardoso (2007) convergem para o diagnóstico de um não conhecimento dos conceitos de base da Estatística Descritiva que permitiriam um trabalho eficaz com a Análise Exploratória de Dados.

Referências

ALMOULOU, S.A. *L'ordinateur, outil d'aide à l'apprentissage de la démonstration et de traitement de données didactiques*. 1992. 294f. Tese (Doutorado em Educação Matemática)-Université de Rennes I, Rennes, 1992.

- ALMOULOU, S. A. *L'analyse statistique de données multidimensionnelles: outil révélateur des conceptions d'enseignants en formation*. In: ENCONTRO CHIC, 2005
- AZCÁRATE, P. G. *Estudio de las Concepciones disciplinares de futuros Profesores de Primaria en torno a las nociones de Aleatoriedad y Probabilidad*. Granada: Comares, 1996
- BATANERO C.; ESTEPA A.; GODINO J.D. Análisis exploratorio de datos: sus posibilidades en la enseñanza secundaria. *Suma*, 9, 25-31. 1991.
- BIGATTÃO, P.A. Jr. *Concepção do professor de matemática sobre o ensino da estocástica*. 2007. Dissertação (Mestrado profissional em Educação Matemática)- Pontifícia Universidade Católica de São Paulo, 2007.
- BOLIVAR, A. (org). *Profissão professor: o itinerário profissional e a construção da escola*. 53-59. Bauru: EDUSC, 2002.
- CARDOSO, R. *O professor de Matemática e a Análise Exploratória de Dados no Ensino Médio*. 2007. Dissertação (Mestrado profissional em Educação Matemática)- Pontifícia Universidade Católica de São Paulo, 2007
- COUTURIER; BODIN; GRAS, R.. *Material de apoio para o uso do software CHIC*.
- GAL, I. Adults' statistical literacy: meanings, components, responsibilities. *International Statistical Review*, Netherlands: ISI, v. 70, n.1, p. 1-50, abr. 2002.
- MORAIS, T.M.R. *Um estudo sobre o pensamento estatístico: componentes e habilidades*. 2006. 134f. Dissertação (Mestrado em Educação Matemática)- Pontifícia Universidade Católica de São Paulo, 2006.
- ROSSINI, R. *Saberes docentes sobre o tema função: uma investigação das praxeologias*. 2006. 382f. Tese (Doutorado em Educação Matemática)- Pontifícia Universidade Católica de São Paulo, 2006.
- SILVA, C.B. *Pensamento estatístico e raciocínio sobre variação: um estudo com professores de matemática*. 2007. 354 f. Tese (Doutorado em Educação Matemática)- Pontifícia Universidade Católica de São Paulo, 2007.
- SILVA, M.J.F. *Investigando saberes de professores do ensino fundamental com enfoque em números fracionários para a quinta série*. 2005. 301f. Tese (Doutorado em Educação Matemática)-Pontifícia Universidade Católica de São Paulo, 2005.

quartis, variância, desvio médio e desvio padrão;			
○ Considerar a necessidade de familiaridade com termos e idéias básicos da representação gráfica e tabular, como frequência absoluta e relativa, frequência acumulada, gráfico de barras, colunas, de setor, histograma, ogiva, etc.;	1	2	3
○ Considerar a compreensão de noções básicas de probabilidade;	1	2	3
○ Considerar o estudo descritivo para a obtenção de inferências ou conclusões estatísticas.	1	2	3

Parte C

Em cada atividade responda os itens abaixo:

Como você trabalharia com seus alunos a análise destes dados, em relação a:

- tabela de distribuição de frequências e gráficos;
- média e desvio padrão;
- mediana e quartis

Descreva com detalhes como você, professor, resolveria essa atividade. Descreva também como você a trabalharia com seus alunos. Quais as possíveis dificuldades dos alunos nesse trabalho? Quais as possíveis dificuldades que um professor teria para trabalhar esse problema?

Atividade 1. Dezesseis alunos foram pesquisados sobre as variáveis: (X) número de irmãos e (Y) número de pessoas que moram na residência, e quinze, sobre a variável (Z) número de cômodos na residência, obtendo-se os dados brutos, cujas organizações em Série de Rol estão expostas nas Tabelas 1, 2 e 3.

Tabela 6. Série de Rol dos valores da variável (X) Número de irmãos

0	0	1	1	1	1	1	1	1	2	2	2	2	2	3	3	3
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Tabela 7. Série de Rol dos valores da variável (Y) Número de pessoas na residência

3	4	4	4	4	5	5	5	5	5	5	6	6	6	7	7
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Tabela 8. Série de Rol dos valores da variável (Z) Número de cômodos na residência

3	3	3	3	3	3	3	4	4	4	4	4	4	5	5	5
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Atividade 2. Na Tabela a seguir, os dados do exemplo 1 correspondem às idades (em número de anos) de 480 alunos da Unidade A de uma escola de idiomas; do exemplo 2, dos 481 alunos da Unidade B dessa escola; o exemplo 3, dos 482 alunos da Unidade C; e, finalmente, o exemplo 4, dos 483 alunos da Unidade D.

Distribuições de frequências

Exemplo 1			Exemplo 2			Exemplo 3			Exemplo 4		
X_i	N_i	F_i	Y_i	N_i	F_i	Z_i	N_i	F_i	W_i	N_i	F_i
12	86	86	12	97	97	12	113	113	12	34	34
13	95	181	13	132	229	13	124	237	13	200	234
14	179	360	14	132	361	14	132	369	14	202	436
15	120	480	15	120	481	15	113	482	15	47	483
total	N= 480		total	N= 481		total	N= 482		total	N= 483	

Fonte: Dados fictícios.

Notas: N_i representa a frequência absoluta e F_i , a frequência acumulada.