UNIVERSIDADE DE SÃO PAULO
FACULDADE DE EDUCAÇÃO

Lisbeth Kaiserlian Cordani

O Ensino de Estatística na Universidade e a controvérsia
sobre os fundamentos da inferência

São Paulo
2001

14800

Universidade de São Paulo – Faculdade de Educação
Programa de Pós-Graduação em Educação
Sub-área temática: Ensino de Ciências e Matemática

Tese de doutorado

**O Ensino de Estatística na Universidade e a
controvérsia sobre os fundamentos da inferência**

Lisbeth Kaiserlian Cordani

11 85 055

DEDALUS - Acervo - IME

31000006070

Tese apresentada como requisito
parcial para a obtenção do Título
de Doutora em Educação, sob a
orientação do **prof. Dr. Nilson
José Machado.**

118 5055

São Paulo
Abril de 2001

Offa Lisseth K. Condami

24.7.2001	Q1A276.2 T
	C794e
47842	60c

Para João, Ana e Mabel, que chegaram para iluminar.

Para Umberto, companheiro e cúmplice da celebração da vida.

Agradecimentos

Não caberia em poucas páginas a lista de pessoas que me influenciaram e concorreram para esta trajetória. O seu tamanho é diretamente proporcional aos anos de minha vida. Daí a impossibilidade...

Agradeço a todos aqueles familiares, amigos e professores que, ao se dedicarem à vida com paixão, foram capazes de me emocionar.

Agradeço, em particular,

ao *Nilson José Machado*, orientador e amigo, sempre disposto a mostrar os vários caminhos que poderiam me levar a ver *a luz no fim do túnel*;

aos meus amigos do IME USP, clássicos e/ou Bayesianos, companheiros de tantas jornadas, que contribuíram para o desenvolvimento de muitas das idéias aqui expostas;

aos meus colegas de percurso na Faculdade de Educação USP;

aos meus filhos muito queridos, *Marina e Renato*, por terem compreendido minha indisponibilidade em tantos momentos.

Resumo

A maioria dos cursos universitários brasileiros tem, em seu currículo, uma disciplina básica obrigatória de elementos de probabilidade e estatística. Além dos procedimentos de natureza descritiva, associados a análise de dados, fazem parte da ementa dessas disciplinas procedimentos inferenciais, geralmente apresentados dentro da teoria clássica (ou frequentista) de Neyman-Pearson. Não é costume nesta disciplina nem discutir aspectos epistemológicos ligados à inferência estatística e nem incluir a apresentação da escola Bayesiana, como uma possível alternativa.

Sabidamente, tal disciplina é um entrave na vida escolar, tanto do aluno como do professor. Do aluno, porque este se depara, em boa parte das vezes, com um oferecimento mecânico da disciplina, sem motivação de natureza aplicada e sem vínculo aparente com sua realidade próxima curricular. Do professor, porque encontra geralmente alunos, além de despreparados com relação aos conceitos primários de incerteza e variabilidade, também com predisposição negativa, devido ao tabu associado à disciplina.

Com o intuito de discutir a necessidade do oferecimento das primeiras noções inferenciais nessa disciplina, bem como responder a pergunta **“qual a inferência que deve ser ensinada numa disciplina básica de um curso universitário?”** buscamos caracterizar, ao longo do trabalho, as relações da estatística com: criação científica em geral e racionalismo e empirismo em particular; a existência ou não de um método científico; o objetivismo e o subjetivismo; os paradigmas das escolas clássica e Bayesiana; aprendizagem e cognição. Foram analisadas e comparadas as abordagens inferenciais feitas segundo cada escola, bem como apresentados alguns exemplos.

A sugestão deste trabalho é de que o [programa de uma primeira disciplina inclua os aspectos epistemológicos ligados à inferência, bem como a apresentação do tópico inferência estatística segundo as duas abordagens: clássica e Bayesiana. Isto eliminaria, pelo menos nos primeiros contatos do aluno com a área, a proposta de rompimento com a escola clássica preconizada por muitos adeptos da escola Bayesiana, bem como a proposta de resistência (manutenção do “*status quo*”), defendida por muitos elementos da escola clássica.] Na verdade a proposta preconiza a coexistência entre as duas escolas numa apresentação de curso básico, pois entendemos que o dever do professor é mostrar o estado da arte da área a seus alunos, deixando a opção (se isto fizer sentido) para uma etapa futura, seja acadêmica ou profissional.

Abstract

In general, most of the undergraduate courses in Brazil offer a basic discipline in probability and statistics. Beyond the descriptive procedures, associated with data analysis, these courses present to the students some inferential techniques, usually linked to the classical (frequentist) Neyman-Pearson school. It is not common to present the inferential aspects from the Bayesian point of view. Everybody knows that both student and teacher have problems with this basic discipline. The student because he receives, in general, a mechanical course, without motivation, with no links to their other disciplines and the teacher because he (or she) usually teaches to very naïve students concerning concepts like uncertainty and variability. Added to that, students seem to have some fear towards the discipline (taboo).

In order to discuss the first inferential notions presented in this discipline, and to answer the question **“which inference should we teach into the basic discipline of statistics to undergraduate students?”** we have tried, in this work, to characterise the relationship between statistics and the following aspects: scientific creation in general and empirism and rationalism in particular; the existence or not of a scientific method; the objectivism and the subjectivism; the paradigms associated to the classical and the Bayesian schools; learning and some cognitive aspects. We have made some comparison between the inferential approaches, and some examples have been presented.

This work suggests that the first program of a basic discipline of probability and statistics should include some epistemological inferential aspects as well as the introduction of inferential statistics by means of both approaches: classical and Bayesian. This action will prevent, at least at the first contact, the members of the Bayesian school from proposing the rupture with the classical school, and also the members of the classical one from maintaining the *status quo*. In fact, the proposal is of a coexistence of both schools in a first level, because we think it is a teacher's duty to show the state of art to his (or her) students, giving the possibility of option (if necessary) for a following step.

Résumé

La plupart des courses universitaires au Brésil a, dans son curriculum, une discipline élémentaire, obligatoire, de probabilité et statistique. Au delà des procédures descriptives, associées à l'analyse de données, les procédures inferentielles font aussi partie intégrante de la discipline, et appartiennent à la théorie classique (ou fréquentiste) de Neyman-Pearson. Il n'est pas habituel de présenter cette dernière partie sous l'angle de l'école Bayésienne. On y sait que cette discipline pose des problèmes aux élèves aussi bien qu'aux professeurs. De la part des élèves, parce qu'ils reçoivent une discipline mécanique, sans motivation appliquée, et sans aucune liaison avec leur quotidien curriculaire. Quant au professeur, il reçoit les élèves sans aucune connaissance sur les notions d'incertitude et de variabilité, et au même temps demotivés *a priori*, car il y a un tabou associé à cette discipline.

Pour discuter la nécessité d'offrir les notions d'inférence dans cette discipline, et pour répondre à la question **“que est que c'est l'inférence qu'on doit enseigner dans un cours élémentaire de statistique à l'université?”** nous avons essayé de caractériser, pendant ce travail, les relations de la statistique avec: la création scientifique en général et l'empirisme et le rationalisme, en particulier; l'existence ou pas d'une méthode scientifique; l'objectivisme et le subjectivisme; les paradigmes de l'école classique et de l'école Bayésienne; l'apprentissage et la cognition.

Ce travail suggère que le programme pour une première présentation de la discipline doit inclure les aspects épistémologiques liés à l'inférence, comme l'offre de l'inférence statistique selon les deux écoles: classique et Bayésienne. Cette choix supprimerait, au moins au premier contact, la rupture avec l'école classique, défendue par beaucoup de membres de l'école Bayésienne, au même temps que la manutention du *status quo*, défendu pour les membres de l'école classique. En réalité, la thèse propose la coexistence entre les deux écoles dans une présentation élémentaire, en entendant que le devoir du professeur est lequel de montrer l'état d'art de la statistique à ses élèves – l'option (si nécessaire) sera laissée au futur.

Índice

Capítulo 1 – Apresentação do tema 1

- 1.1 Introdução 1
 - 1.1.1 Desenvolvimento atual da estatística e sua utilidade 2
 - 1.1.2 Estatística e informação 4
- 1.2 A estatística no currículo pré-universitário 6
- 1.3 O ensino da Estatística na Universidade 10
 - 1.3.1 Formação dos professores 11
 - 1.3.2 O curso introdutório 13
 - 1.3.3 Escolas de inferência 16
 - 1.3.4 Dificuldades dos alunos 19
- 1.4 Objetivos, destinatários e estrutura do trabalho 22

Capítulo 2 – Estatística e método 25

- 2.1 Introdução 25
- 2.2 Domínio da incerteza 25
- 2.3 Criação científica 27
 - 2.3.1 Empirismo e racionalismo 28
 - 2.3.2 Método ou métodos 31
- 2.4 Conclusão 35

Capítulo 3 – A importância da estatística 37

- 3.1 Introdução 37
- 3.2 Objetivismo e subjetivismo 40
- 3.3 O uso da estatística 42
- 3.4 As escolas clássica e Bayesiana como paradigmas 45
- 3.5 Conclusão 48

Capítulo 4 – A lógica da inferência 49

- 4.1 Introdução 49
- 4.2 Inferência 50
- 4.3 Inferência estatística 54
- 4.4 Raciocínios condicionais 56
- 4.5 As inferências de Fisher, Neyman-Pearson e Bayesiana 57
- 4.6 Conclusão 61

Capítulo 5, – Alguns elementos para a disciplina introdutória 63

- 5.1 Introdução 63
 - 5.1.2 Probabilidade e suas definições 63
 - 5.1.3 Estatística e suas definições 67
- 5.2 Probabilidade e o teorema de Bayes 70
- 5.3 Conclusão 77

Capítulo 6 – Técnicas de inferência estatística 79

- 6.1 Introdução 79
- 6.2 Teste de hipóteses 80
 - 6.2.1 Exemplo com variável contínua 83
 - 6.2.1.1 Teste de significância de Fisher 84
 - 6.2.1.2 Teste de hipóteses de Neyman – Pearson 85
 - 6.2.1.3 Teste Bayesiano 90
 - 6.2.2 Exemplo com variável discreta 94
 - 6.2.2.1 Teste de significância de Fisher 94
 - 6.2.2.2 Teste de hipóteses de Neyman – Pearson 96
 - 6.2.2.3 Teste Bayesiano 97
 - 6.2.3 Comparação dos três procedimentos 99
- 6.3 Estimação 100
 - 6.3.1 Estimação pontual 100
 - 6.3.2 Estimação por intervalo 105
- 6.4 Conclusão 104

Capítulo 7 – A estatística e o aluno 109

- 7.1 Introdução 109
- 7.2 Estatística e aprendizagem 111
- 7.3 Estatística e cognição 113
 - 7.3.1 Modelos mentais 114
 - 7.3.2 Intuições e suas distorções 117
 - 7.3.3 Raciocínio lógico e a tarefa de seleção de Wason 120
- 7.4 Conclusão 124

Capítulo 8 – Considerações finais 126

- 8.1 Introdução 126
- 8.2 Alguns contrastes entre as escolas clássica e Bayesiana 127
- 8.3 Proposta de programa 130
- 8.4 Últimas observações 132

Referências bibliográficas 137

“Uma coleção de dados não nos proporciona informação científica, da mesma forma que uma mina não fornece artigos de metal.” (Cohen)☹

“Um paradigma pode ao mesmo tempo elucidar e cegar, revelar e ocultar. É no seu seio que se esconde o problema - chave do jogo da verdade e do erro.” (Morin)☺

Capítulo 1 Apresentação do tema

1.1 Introdução

O século XX assistiu ao desenvolvimento de uma das ferramentas mais utilizadas hoje em dia pelos cientistas, analistas econômico-sociais, profissionais liberais, jornalistas, dentre outros: a Estatística, que trata dados de maneira descritiva e/ou inferencial, qualquer que seja a definição que possamos encontrar para isto. Não há como negar a necessidade de um conhecimento de abordagem quantitativa em todas essas áreas, principalmente com o acúmulo de informações disponíveis, as quais só serão bem digeridas se objetivamente apresentadas e interpretadas e a Estatística tem tido um papel relevante nessas análises. Além disso, de acordo com Besson (1995), as diferentes sociedades têm necessidade de relatos de seu futuro, que ensejem projetos, esperanças, temores e limites. Disso tudo decorre a necessidade científica, social e mesmo política do ensino da Estatística, disciplina que está explicitamente presente em quase todos os cursos de nível superior no Brasil. Essa presença suscita imediatamente questões ligadas não só ao ensino propriamente dito, como também à interpretação dos resultados obtidos através de análises de dados, dada a expectativa de se encontrar a “verdade” com este instrumental.

Houve época, antes do século XIX, em que a Estatística era identificada como certeza científica. Em parte podemos dizer que mesmo no século XX, com a busca da objetividade, esperava-se que a Estatística pudesse, mecanicamente, produzir um certo grau de certeza a uma dada hipótese. Isto foi encarado, segundo Gigerenzer (1992), como o sonho da mecanização do conhecimento e a Estatística estaria sendo usada a serviço da objetividade. Esta expectativa foi modificada, e atualmente espera-se uma posição mais cuidadosa face aos resultados encontrados, permitindo que outros componentes ajudem na tomada de decisão, como propõe Barnett(1973):

“Estatística é o estudo de como informação deveria ser empregada para reflexão

e ação em uma situação prática envolvendo incerteza”.

☉ 1.1.1 Desenvolvimento atual da Estatística e sua utilidade

No mundo acadêmico atual, praticamente todas as carreiras (exatas, biológicas ou humanas) têm em seu currículo uma disciplina introdutória de Estatística, prova de que o raciocínio estatístico e/ou probabilístico é considerado fundamental em quase todas as áreas do conhecimento. Os objetivos do conhecimento básico dessa disciplina incluem:

- destacar a importância da coleta de dados, discutindo qual a informação realmente relevante para a resolução de problemas;
- discutir a importância de uma análise consistente dos dados;
- divulgar a linguagem e o raciocínio estatístico/probabilístico para tomada de decisão face à incerteza;
- fazer compreender a noção de variabilidade, inerente a qualquer característica a ser pesquisada;
- fornecer elementos básicos para compreensão e crítica de artigos publicados em revistas especializadas ou até mesmo na mídia, para o grande público;
- desenvolver no aluno o espírito crítico em sua futura atividade profissional;
- fornecer elementos para a busca continuada da melhoria de processos;
- mostrar ao aluno a existência de limitações nas interpretações de análises em estatística. ☉

Geralmente os resultados de pesquisas são publicados com base em coleta e análise estatística de dados. Assim é necessário conhecer conceitos básicos de Estatística, para que o profissional possa ler um trabalho científico e refletir sobre a adequação da solução ou mesmo dialogar com um estatístico para realizar uma análise de dados. Numa atividade ligada à administração, por exemplo, a Estatística servirá para equacionar os problemas da área e verificar se o processo é ágil e eficiente, enquanto que na de um profissional liberal será útil para o entendimento das últimas publicações de sua especialidade, sabendo dar a devida interpretação aos resultados. Além disso, o cidadão

comum é envolvido diariamente por notícias que contêm “Estatísticas” sendo portanto importante o conhecimento básico da área para não ficar refém da interpretação dos dados, podendo interpretar as pesquisas, tanto no campo científico formal quanto no de divulgação, bem como as notícias do dia-a-dia, que envolvam observações numéricas ou categorizadas e /ou tomada de decisão. Numa situação ideal de educação continuada, podemos aqui considerar como aprendizes, não somente os estudantes propriamente ditos, mas também os operários, os funcionários, os acadêmicos, os cidadãos de modo geral; todos podem, de alguma forma, ser considerados *pesquisadores* e é nesse sentido que esta palavra será usada daqui para a frente.

④ A necessidade crescente de formação docente na área de Estatística contribuiu para que o ISI (International Statistical Institute) organizasse, a partir de 1982, uma série de conferências internacionais sobre o ensino da Estatística (*ICOTS - International Conferences on Teaching Statistics*). A partir da criação do IASE (International Association for Statistical Education), em 1991, como uma seção do ISI, todas as atividades de ensino em Estatística foram aí concentradas, favorecendo tanto o intercâmbio de experiências como o estabelecimento de metas e desafios para o futuro do ensino da Estatística. Uma amostra dessa preocupação, incluindo cada vez mais uma interação da Estatística com a sociedade, foram os temas da última reunião que o IASE organizou em Singapura, em 1998:

- *Educação Estatística no nível escolar (pré-universitário)*
- *Educação Estatística no nível pós-secundário*
- *Educação Estatística no local de trabalho*
- *Educação Estatística e a sociedade global*
- *Uma perspectiva internacional de educação Estatística*
- *Pesquisa em ensino da Estatística*
- *O papel da tecnologia no ensino da Estatística*
- *Outros desenvolvimentos em educação Estatística* ⑤

Voltando ao ponto do interesse acadêmico, muitas carreiras precisam que seus pesquisadores tenham um conhecimento mais profundo da área de Estatística, e a procura tem aumentado ao longo dos anos. De fato, se examinarmos os periódicos específicos às ciências aplicadas, veremos que a maioria das publicações contém análise Estatística. De acordo com Scheaffer, presidente da ASA (American Statistical Association), em matéria publicada no ISI Newsletter (2001, v.25(1)),

“ Ser estatístico nos dias de hoje é muito empolgante! Dados estão por toda a parte. Tecnologia para produzir e analisar dados avança mais rapidamente do que a maioria de nós alcança compreender (talvez mais rápido do que desejamos). Lideranças em negócios, indústria, governo e educação começam a perceber (embora não tão rapidamente como desejamos) que a Estatística e os estatísticos podem dar contribuições importantes à sociedade.”

1.1.2 Estatística e informação

Ensinar e aprender, práticas do cotidiano escolar, são atividades diretamente ligadas ao conhecimento ou à construção de significados, seja por parte do aluno ou do professor, da escola ou da sociedade. e seu uso consciencioso auxilia a extração de informações pertinentes, que poderão vir a integrar o corpo de conhecimentos associados a determinada área. De acordo com Machado (1995), para melhor compreender as obrigações a serem desempenhadas pela escola, não nos devemos centrar somente no par informação/formação mas sim na inter-relação entre dados, informação e conhecimento. Uma vez que conhecimento pode ser construído através de dados e informação e como a Estatística dispõe de ferramentas para tratar dados coletados em diversas situações é costume dizer que a Estatística e a Probabilidade são ferramentas adequadas e imprescindíveis para o avanço do mesmo (conhecimento).

Estamos na era da tecnologia de informação, que colabora com mudanças de impacto nos aspectos políticos, sociais e econômicos da sociedade globalizada. Estas mudanças afetam também a coleta e análise de dados e isto deveria ser levado em conta no ensino da Probabilidade e Estatística desde o curso introdutório. Atualmente os diversos setores das grandes organizações empresariais estão integrados de forma a fornecer informações para seus bancos de dados, preparados para armazenar imensas quantidades de características dos clientes, por exemplo, colhidas a velocidades há pouco tempo inimagináveis. Os profissionais dessas empresas farão parte de equipes interdisciplinares que, para enfrentar esta nova realidade, quase certamente conterão estatísticos. Também profissionais de outras áreas já estão assumindo o papel de analistas, o que mostra a importância para essas pessoas de um conhecimento pelo menos básico de Estatística.

Essa forma atual de coleta de informações deu ensejo ao aparecimento de uma nova área dentro da análise de dados, chamada "*Data Mining*", que nada mais é do que um conjunto de procedimentos que auxiliam a coleta e análise de grandes conjuntos de dados de uma organização, com o objetivo de extrair informações, reconhecendo padrões e usando-os para tomada de decisão. Os domínios em que esta área está se desenvolvendo são de natureza bem diversa, como por exemplo Administração e Medicina. É claro que o primeiro escrutínio tem que ser feito quanto à qualidade dos dados, pois de nada adianta uma grande quantidade de dados errados, ou mesmo coletados segundo uma perspectiva equivocada. Muitas vezes a qualidade é ignorada a favor da quantidade e os analistas devem estar atentos a isto. O novo nome sugere uma garimpagem de dados, coletados aos milhares, com objetivos de classificar variáveis e descobrir relações entre elas, fazer previsões, etc. A organização dos dados exige uma padronização que facilita a operação de auxílio à tomada de decisão mediante potentes recursos computacionais. A disciplina introdutória de Estatística, da forma como é desenvolvida atualmente, não está preparada para colocar em discussão este conceito. Uma reforma nesta disciplina, que atenda às expectativas do novo século, dependerá fundamentalmente do preparo dos professores.

1.2 A Estatística no currículo pré-universitário

☉ Com relação ao ensino da Estatística, até há muito pouco tempo somente aparecia a necessidade de aprendizagem dos conceitos estatísticos/probabilísticos nos diferentes cursos universitários (exceção feita ao extinto Curso Normal, de formação para o magistério primário, de 1ª a 4ª série, que corresponderia ao atual curso médio, onde era ensinada exclusivamente a Estatística descritiva). No entanto já se fazia sentir, há tempos, a necessidade de um currículo pré-universitário que contemplasse de maneira explícita e definitiva a área de Probabilidade e Estatística. A Lei de Diretrizes e Bases da Educação Nacional (Lei 9394 de 20/12/1996) deu ensejo a uma mudança radical nos programas curriculares e favoreceu a criação, pelo Ministério de Educação (MEC), de uma série de documentos, denominados Parâmetros Curriculares Nacionais (PCN), com os primeiros volumes destinados ao ensino fundamental (de 1ª a 8ª) publicados em 1998 e os volumes seguintes dedicados ao ensino médio, em 1999.

Nos PCN's é destacada a importância de se incluir, já no ensino fundamental, elementos de Estatística e Probabilidade, uma vez que está sendo detectada uma demanda social por este tipo de conhecimento. Nesse documento foi colocada a seguinte afirmação:

"Estar alfabetizado neste final de século, supõe saber ler e interpretar dados apresentados de maneira organizada e construir representações para formular e resolver problemas que impliquem o recolhimento de dados e a análise de informações. Esta característica da vida contemporânea traz ao currículo de Matemática uma demanda em abordar elementos de Estatística, Combinatória, Probabilidade, desde os ciclos iniciais." ☉

Esta vinculação (da Estatística com a matemática) no ensino pré-universitário está sempre presente aos conteúdos programáticos associados à disciplina matemática. Tem sido constante, nos meios ligados à docência, a indagação tanto no âmbito nacional quanto no internacional se esta vinculação é desejável e significativa, tanto que alguns

países (Portugal, por exemplo) já oferecem a disciplina Estatística, separada da Matemática, no nível médio de ensino. Isto é assunto para uma discussão que não faz parte deste trabalho. Voltemos então aos PCN's.

Permeando todos os ciclos da Educação é possível detectar a preocupação constante de que os alunos sejam capazes de (dentre outros aspectos):

- utilizar diferentes fontes de informação e recursos tecnológicos para adquirir e construir conhecimento;
- questionar a realidade, formulando problemas e tratando de resolvê-los, utilizando para isto o pensamento lógico, a criatividade, a intuição, a capacidade de análise crítica, selecionando procedimentos e verificando sua adequação.

Essas capacidades poderiam ser exploradas no ensino fundamental através da aprendizagem de ferramentas Estatísticas, que poderão ser vistas tanto isoladamente dentro (no caso) da disciplina Matemática, ou então como uma atividade de interação com outras disciplinas, através dos chamados *temas transversais* (termos usados nos PCN's).

Percebe-se (nos PCN's) a preocupação com o chamado tratamento da informação, que inclui coleta, organização e descrição de dados, interpretação e identificação de comportamentos aleatórios. Recomenda-se que seja contemplada uma abordagem que permita ao aluno descrever, analisar, avaliar e tomar decisões, itens que são privilegiados numa análise Estatística. A ênfase sugerida depende do ciclo em que o aluno se encontra (cada dois anos do ensino fundamental corresponde a um ciclo; como são oito anos, tem-se quatro ciclos), com dificuldades crescentes ao longo dos ciclos, tomando o cuidado de não privilegiar o caminho das fórmulas e definições. Os computadores são reconhecidos neste contexto pois são facilitadores de ações que permitem o tratamento de dados reais, quase sempre em quantidade expressiva, necessitando de habilidades para condensá-los e interpretá-los.

A recomendação dos parâmetros para o ensino médio dá maior flexibilidade às escolas, que deverão estabelecer suas propostas pedagógicas definindo a proporção das áreas na grade curricular, bem como conteúdos a serem nela incluídos. Qualquer que seja a orientação, é preciso que o aluno esteja preparado para compreender o caráter aleatório dos fenômenos naturais e/ou sociais, bem como ser capaz de retirar e interpretar resultados amostrais, eventualmente usando cálculo de Probabilidades.

Também o ENEM (Exame Nacional do Ensino Médio) , organizado pelo MEC, apresenta em seu documento básico, publicado em março de 2000, cinco competências esperadas do aluno que conclui o ensino médio, duas das quais podem ser entendidas como, pelo menos parcialmente, ligadas à aprendizagem de conceitos estatísticos e/ou probabilísticos:

- selecionar, organizar, relacionar, interpretar dados e informações representados de diferentes formas, para tomar decisões e enfrentar situações-problema;
- relacionar informações, representadas em diferentes formas com conhecimentos disponíveis em situações concretas, para construir argumentação consistente.

Quanto às vinte e uma habilidades a serem avaliadas pelo ENEM, dez delas estão relacionadas de alguma forma a uma abordagem estatístico/probabilística , que inclui o cálculo de medidas descritivas e gráficos, distribuições de variáveis, cálculo de Probabilidades e até construção de índices. Neste último tópico há sugestão de discutir o significado e a importância de tais construções, trabalhando, por exemplo, com o IDH (Índice de Desenvolvimento Humano), elaborado pelas Nações Unidas, calculado levando em conta indicadores de **educação, saúde e renda**, que mede a qualidade de vida das pessoas em 174 países. No cálculo divulgado no ano 2000 com a atual metodologia, o BRASIL ocupava a 74ª posição. Um estudo deste indicador está de acordo com a 12ª das habilidades listadas e pode ser abordado dentro de um contexto de análise Estatística :

“Analisar fatores socioeconômicos e ambientais associados ao desenvolvimento, às condições de vida e saúde da população humana, por meio da interpretação de diferentes indicadores.”

Uma abordagem dos principais conceitos ligados à área de Probabilidade e Estatística no ensino pré-universitário facilitaria a vida diária bem como o trabalho dos indivíduos, uma vez que informações quantitativas estão cada vez mais disponíveis. Também os alunos ingressantes no terceiro grau teriam uma outra formação, não sendo tão *ingênuos* no tema. Isto favoreceria o oferecimento de um curso em outro nível de maturidade em relação ao que pode ser hoje oferecido.

Como esta não é ainda a realidade brasileira de ensino pré-universitário (e isto também acontece em muitos países ditos desenvolvidos) e apesar de considerarmos um avanço esta proposta contida nos PCN's, faremos as seguintes recomendações:

- menos ênfase na vinculação de Probabilidade ao cálculo combinatório (pois corre-se o risco de que o aluno associe Probabilidade definitivamente a arranjos, combinações, etc.);
- menos prestígio à média aritmética como medida representativa de um conjunto de dados, dando espaço para medidas mais robustas como a mediana, por exemplo;
- a inclusão, indispensável, de medidas de variabilidade, trabalhando inicialmente com a amplitude e posteriormente com o desvio padrão.

A reforma preconizada pelo MEC (e eventualmente melhorada) demandará, a curto e médio prazos, professores qualificados, com concepções pedagógicas modernas, capazes de refletir sobre concepções de ensino-aprendizagem de modo amplo, a fim de que as pessoas em geral e futuros ingressantes do terceiro grau, em particular, tenham noções prévias sobre Probabilidade e Estatística, que favoreçam sua atuação como cidadãos, profissionais e alunos de nível universitário. Esta formação prévia propiciará que o programa inicial da disciplina em um curso superior seja mais denso e apropriado à área de concentração do aluno, em relação ao que é oferecido hoje.

Atualmente, dado que esta reforma ainda não atingiu os ingressantes do ensino superior, estes apresentam um analfabetismo crônico na área de Estatística e Probabilidade, e é destes alunos (os ingressantes) que passaremos a falar daqui por diante.

1.3 Ensino da Estatística na Universidade

Como a Estatística faz parte do elenco básico de disciplinas em praticamente todos os cursos universitários, podemos nos perguntar como ela é encarada pelos alunos e como é o seu desempenho. Sabe-se que historicamente os alunos não se mostram motivados pela Estatística, quer seja pela dificuldade alegada do seu conteúdo, quer seja pelo pouco uso que dela fazem as disciplinas que a seguem na graduação em algumas das unidades de origem (a retomada na pós-graduação, depois de anos de afastamento, é geralmente dolorosa para o aluno!). Muitas vezes, mesmo em cursos que têm a disciplina de Estatística como obrigatória, as demais disciplinas não levam em conta a exposição dos alunos às noções elementares da Estatística e apenas são apresentadas as conclusões advindas de estudos experimentais, sem que se faça menção explícita ao uso de técnicas quantitativas, não sendo oferecida ao aluno a oportunidade de discussão dos resultados. É como se o modelo descrito gerasse os dados observados e, neste caso, a característica de incerteza não estaria mais presente. Dada esta situação, a existência da disciplina de Estatística em tal curso parece mais estranha ainda ao aluno, que já não nutre, desde o início, nenhuma simpatia especial por ela.

É preciso considerar que os alunos aqui mencionados estão na faixa entre 18 e 21 anos, e que portanto têm uma bagagem sócio-cultural. As aulas tradicionais de Estatística não costumam levar em conta esse histórico, e geralmente os professores nem se importam em verificar como funciona a intuição desses alunos e nem em tentar decifrar quais são os modelos mentais por eles utilizados. Espera-se que o aluno use para essas aulas um compartimento vazio de seu cérebro e que o encha com os conceitos “corretos”.

Essa tradição dificulta a espontaneidade do aluno, que nunca se manifesta nas aulas, temendo não ser capaz de fazer comentários pertinentes.

Os alunos ingressantes no curso superior geralmente vêm com conhecimentos insuficientes (ou até mesmo inexistentes) de temas como, por exemplo, *Probabilidade, variabilidade, raciocínio aleatório*, fato este agravado por problemas de entendimento de linguagem (o qual afeta não só o desempenho em Estatística, como também em todas as outras áreas). Além disso boa parte dos alunos tem uma visão determinística do mundo, culpa em parte da visão difundida pela maioria das disciplinas, nos ensinamentos fundamental e médio. Todos esses componentes, incluindo a perspectiva muitas vezes equivocada do professor, aliados a um assunto que tem complexidade conceitual, servem para dificultar o trajeto discente, tornando essa disciplina uma das responsáveis pelo alto índice de retenção de alunos, principalmente nas universidades públicas.

1.3.1 Formação dos professores

Para atender à demanda dos inúmeros cursos universitários que colocam a disciplina de Estatística em seu curriculum é necessário dispor de professores qualificados bem como de instalações compatíveis com a modernização tecnológica. É necessário mudar a concepção de boa parte dos professores, que insistem numa orientação exclusivamente computacional para a disciplina, com cálculos muitas vezes tediosos, para uma base mais experimental, ligada ao contexto cultural do aluno bem como à sua área de concentração do curso. É preciso construir diretrizes que auxiliem o treinamento de professores, mas é importante estar ciente das dificuldades inerentes a esta tarefa, principalmente nesse campo desconhecido que é a educação Estatística.

A idéia geral que permeia a nova LDB é de flexibilização das ementas dos cursos superiores e, dado que a Estatística é uma disciplina geralmente responsável pelos altos índices de repetência nos primeiros anos universitários, será que vamos assistir a uma diminuição da carga a ela associada? Creio que cabe a nós, professores, desmitificar esta

disciplina, oferecendo um conteúdo atraente e motivador, atualizado e compatível com o nível de conhecimento dos alunos, por meio de práticas pedagógicas que comprometam o aluno com o aprendizado, oferecendo ainda, como optativa, uma disciplina com mais formalidade para os que têm interesse e/ou vocação para este tipo de raciocínio.

O que se verifica, no entanto, é que para os iniciados no assunto (e particularmente os professores) o desenvolvimento tradicionalmente proposto para a ementa da disciplina parece ser muito natural, com uma progressão de conhecimentos “coerente”. Mas o que podemos dizer do ponto de vista didático-pedagógico? Por que a reprovação é alta? Por que a evasão é grande? Sabemos que essa disciplina tem sido, por anos e anos, o nó górdio dos cursos de graduação, principalmente nas áreas de humanas e biológicas, e assistimos muitas vezes a diferentes unidades universitárias se rendendo à pressão dos alunos para a retirada da obrigatoriedade da mesma. Mesmo na área de exatas reconhecemos que os alunos, que normalmente não têm barreiras a métodos quantitativos, aprendem a resolver os *problemas-tipo*, sem no entanto perceberem o cerne do raciocínio; na verdade eles aplicam uma Estatística instrumental. O que se percebe é que os professores se sentem confortáveis por ensinarem o que lhes foi ensinado, e como não há prestígio acadêmico por ser um bom professor, esta atividade é relegada a segundo plano: o avanço na carreira acadêmica é que importa, e o prestígio quase sempre é atrelado ao número de artigos científicos publicados. Este é o quadro nas universidades que fazem pesquisa. Nas demais, como os professores, na maioria das vezes, dão um número excessivo de aulas, o quadro favorece menos ainda os aspectos didático-pedagógicos desejáveis para um resultado compatível com um ensino de bom nível. O campo da educação Estatística é complexo e pouco desenvolvido. É necessário um conjunto de iniciativas para consolidar a área e melhorar o correspondente conhecimento do ensino e aprendizagem. A 6ª Conferência Internacional do Ensino da Estatística (ICOTS) a ser realizada em 2002, destinou uma sessão especial para apresentação de comunicações sobre o tema “Treinamento de professores”, o que mostra a preocupação internacional com o tema.

1.3.2 O curso introdutório

Dada a influência das escolas anglo-saxônicas na formação dos primeiros professores brasileiros com doutorado em Probabilidade e Estatística, as ementas sugeridas para a disciplina introdutória nas áreas aplicadas, os chamados *cursos de serviço*, foram também por elas fortemente influenciadas. A parte inicial apresenta, além da Estatística descritiva, elementos de Probabilidade (espaço amostral, definição e axiomas relacionados a Probabilidades), conceitos de variável aleatória e distribuição de Probabilidade, bem como de distribuições amostrais. A análise inferencial, que é a parte final, descrita na grande maioria dos textos básicos nacionais e internacionais, apresenta procedimentos da teoria clássica de inferência (estimação e teste de hipóteses). Este foi (e ainda é) o aspecto geral da ementa utilizada na grande maioria dos cursos oferecidos aos alunos ingressantes em praticamente todos os cursos universitários (tanto no Brasil quanto no exterior)

Podemos identificar dois gargalos nessa disciplina introdutória de Estatística: um deles refere-se ao cálculo básico de **Probabilidades** e o outro à metodologia de **inferência Estatística**, material ao qual o aluno é exposto geralmente pela primeira vez em sua vida. Embora estejam de certa forma ligadas, cada uma destas abordagens resolve um tipo de problema.

Na teoria de Probabilidades, por exemplo, é dada uma informação sobre uma população e quer-se saber qual a Probabilidade de se encontrar este ou aquele resultado num experimento ou numa pesquisa observacional. (Se uma moeda é *não viciada*, ou seja, se a Probabilidade de sair o resultado cara em um lançamento for igual a $\frac{1}{2}$, qual a Probabilidade de se obter quatro caras em quatro lances independentes dessa moeda? Através de um raciocínio puramente dedutivo, obtido através do uso conveniente de regras de Probabilidade, chega-se à resposta.

Quanto à inferência Estatística, o caminho é *contrário*, no sentido de que é oferecido um resultado amostral através de dados coletados e quer-se *caracterizar a*

população à qual eles pertencem, com respeito a determinados parâmetros. Se o lançamento de uma moeda quatro vezes, de modo independente, forneceu como resultado quatro caras, a moeda é não viciada? Aqui vemos que o interesse é saber se a moeda pertence a uma determinada população com Probabilidade de sucesso $\frac{1}{2}$, mas a resposta não é tão fácil como foi no exemplo dado no parágrafo anterior. A inferência Estatística pode então, pela sua natureza, ser considerada como um aspecto do problema de indução, um problema que continua a desafiar soluções, ao qual voltaremos oportunamente.

Há muitas escolhas disponíveis na literatura para a abordagem da análise inferencial, mas vamos nos ater, neste trabalho, a duas delas, que têm concepções fundamentalmente diferentes: a **teoria Clássica** e a **teoria Bayesiana**. A primeira pode ser encarada como paradigma nas disciplinas básicas, em que elementos de inferência são apresentados através de um conjunto de regras e técnicas, que o aluno “aprende” (?) de maneira instrumental, sem nenhuma discussão dos fundamentos a ela subjacentes. Um dos esteios desta teoria é a objetividade, no sentido da neutralidade do observador, que não incorpora suas opiniões à análise, chegando às Probabilidades através de uma abordagem frequentista. A segunda, pode ser vista como um contraponto à perspectiva clássica, pois defende o uso da subjetividade, levando em conta a opinião do observador no estabelecimento das chamadas Probabilidades *a priori*.

Numa disciplina introdutória de Estatística, é preciso muito esforço para introduzir, numa primeira fase, o raciocínio probabilístico, incluindo aí a noção de Probabilidade condicional e, numa segunda etapa, os princípios subjacentes às técnicas de inferência (apresentadas geralmente sob o ponto de vista clássico), com um resultado nem sempre positivo. Muitos professores dessa disciplina fazem, naturalmente, perguntas como estas: qual o mecanismo de aprendizado que está por trás da teoria que estamos ensinando? Como funciona a intuição do aluno? Como aproveitar o conhecimento *tácito* do aluno? A dificuldade está na linguagem? Há deficiências no raciocínio lógico? Uma discussão prévia sobre procedimentos científicos ajudaria o desenvolvimento da disciplina? Como aspectos psicológicos relacionados ao aprendizado relacionam-se com o problema? E muitas outras.

Os textos apresentados no nível básico dão a aparência de que a Estatística é mera tecnologia e que uma vez aplicados os modelos “corretos” a resposta “certa” será obtida. Geralmente não há menção da existência de várias correntes filosóficas que respaldam a lógica e a inferência na Estatística, nem do desacordo entre elas ou mesmo dentro delas. O material segue usualmente a chamada linha clássica ou frequentista, desenvolvida por Neyman e Pearson, que será discutida posteriormente. Há muita controvérsia entre as duas teorias (clássica e Bayesiana) e uma das mais contundentes críticas aos textos que usam a abordagem clássica pode ser vista em Howson e Urbach (1993), adeptos da escola Bayesiana, que afirmam:

“Embora sejam profundamente falaciosos, os métodos de inferência clássica são defendidos em centenas de livros, textos obrigatórios em milhares de instituições de educação superior, onde centenas de milhares de estudantes são obrigados a estudar”.

Algumas iniciativas isoladas para desenvolver um curso introdutório de Estatística Bayesiana começaram a ser feitas na década de 70, com um texto de Blackwell (1969), professor radicado na Universidade de Berkeley, California (EUA), que introduziu o tema como uma abordagem introdutória para uma primeira disciplina de Probabilidade e Estatística (segundo a entrevista dada pelo autor a DeGroot em 1986, ele mesmo só usou o próprio livro durante 5 anos!). Nessa época esse texto foi utilizado no Instituto de Matemática e Estatística da USP, embora também por pouco tempo, para cursos básicos, onde nem sempre se chegava à discussão do raciocínio Bayesiano para inferência, presente nos últimos capítulos.

Apesar de algumas tentativas terem surgido há mais de 30 anos, ainda é escasso o desenvolvimento da escola Bayesiana. Na mesma entrevista acima referida, Blackwell afirma que se a Estatística Bayesiana crescer será menos pela influência dos estatísticos e mais pela influências dos pesquisadores de outras áreas, que se sentem mais à vontade com a nova perspectiva. A literatura básica disponível é insuficiente e há controvérsias

entre vários educadores em Estatística sobre a conveniência ou não de fazer tal abordagem. Mesmo nos Estados Unidos onde se pode dizer que há atualmente uma massa crítica maior de adeptos desta teoria do que em décadas passadas, não há mais do que uma dúzia de cursos básicos em graduação que enfocam tal teoria (Berry, D. The American Statistician, 51(3), 1997).

Ainda no campo de controvérsia, também encontramos professores céticos quanto ao ensino da inferência Bayesiana, como Moore (1997a) que afirma:

"...os Bayesianos ainda aparecem em clara minoria, não tendo conseguido persuadir a eclética maioria dos estatísticos a respeito de suas concepções. Devemos hesitar em reestruturar nosso ensino aos iniciantes da maneira por eles recomendada."

1.3.3 Escolas de inferência

Situando cronologicamente no século XX, sir R. Fisher, em meados da década de 20 na Inglaterra (Rothamsted), e J. Neyman no fim da década de 30 nos Estados Unidos (Berkeley), foram dois nomes de peso no desenvolvimento da inferência Estatística clássica, para a qual deram muitas contribuições que foram usadas, ampliadas, polemizadas ou criticadas reiteradas vezes por cientistas e/ou usuários de todo o mundo. Foi também nesse mesmo século que a Estatística Bayesiana se desenvolveu, apesar ser baseada num teorema enunciado no século XVIII, pelo Rev. Bayes,

Todas as abordagens inferenciais utilizam a noção de Probabilidade, mas definir o que é Probabilidade e discutir como ela é conectada com a Estatística tem levado muitos pesquisadores a discussões intermináveis que, segundo Savage (1972),

"...não se via desde a Torre de Babel."

A própria definição de Estatística pode também levar a controvérsias e voltaremos a este assunto no capítulo 4. Nesta seção comentaremos sucintamente as abordagens inferenciais do ponto de vista clássico e do ponto de vista Bayesiano, tema que será retomado nos próximos capítulos.

A teoria inferencial tradicionalmente apresentada aos alunos segue a chamada linha clássica, como já comentado aqui, e supõe um modelo probabilístico, geralmente sob a forma de uma família paramétrica de distribuições. Supõe ainda que os dados são provenientes de observações de uma variável aleatória definida num espaço amostral, variável esta que tem sua distribuição conhecida a menos, por exemplo, de um parâmetro. Os métodos inferenciais são construídos num curso básico geralmente para que se tenha a informação sobre o parâmetro desconhecido, a fim de que a população (distribuição) seja *completamente* conhecida.

Esses métodos são conhecidos como *estimação e teste de hipóteses*, sendo que o primeiro apresenta para o parâmetro desconhecido uma estimativa pontual e uma intervalar (intervalo de confiança), esta última com um coeficiente de confiança estipulado *a priori*. O segundo método, o teste de hipóteses, estabelece duas hipóteses complementares (Hipótese nula e Hipótese alternativa) e constrói a chamada região crítica com um nível de significância pré-estabelecido: se o valor da chamada Estatística do teste pertencer à região crítica, a hipótese nula será rejeitada; caso contrário não o será (decisão essa sempre acompanhada de uma probabilidade de erro, associado ao nível de significância estabelecido).

No caso de teste de hipóteses, há diversas maneiras de discutir a construção dessas hipóteses, mas há uma tradição em se ensinar que o que um pesquisador deseja *provar* deve ser colocado na hipótese alternativa. Assim, o estabelecimento de hipóteses é feito pelo método indireto, e o aluno “aprende” a estabelecer essa regra: se o objetivo é testar se um método de ensino é superior a outro, coloca-se na hipótese nula que eles são iguais, e na hipótese alternativa que eles são diferentes (aqui seria uma abordagem relativa a parâmetros de duas populações).

De imediato pode-se argumentar que esta “regra” nem sempre condiz com a realidade, como por exemplo no caso recente da implantação no Brasil de medicamentos genéricos, onde a realização de um ensaio clínico tem como objetivo demonstrar que o princípio ativo da droga genérica é compatível com a da comercial (estudo de *bioequivalência*, onde o resultado positivo é alcançado quando não se rejeita a hipótese nula!). É possível invocar também, para a construção de tais hipóteses, a magnitude dos possíveis tipos de erro (erros de primeira e segunda espécies) que podem ser cometidos numa conclusão do teste e, nesse caso, as hipóteses seriam construídas de modo a minimizar o erro considerado mais importante. No entanto, para o aluno iniciante, a “regra” acima é o processo mais comumente ensinado.

A teoria Bayesiana fornece critérios diferentes daqueles usados na Estatística clássica para tirar conclusões. Na verdade a teoria Bayesiana pode ser usada tanto numa situação probabilística (dedução) como numa situação Estatística (indução), ou seja, fornece tratamento unificado para os dois tipos de situação. O uso da técnica para raciocínio dedutivo (cálculo probabilístico) é feito através do chamado teorema de Bayes e nesse caso não há nenhuma discussão em torno de seu uso (na verdade, ele está em praticamente todas as ementas que incluem o cálculo de Probabilidades, inclusive naquelas em que a opção pela inferência é a clássica). O uso para inferência Estatística é que é controvertido, uma vez que contém um elemento subjetivo que é o estabelecimento da distribuição *a priori* de Probabilidades, *a qual deve ser estabelecida antes do experimento ser realizado*, e definida subjetivamente (e com critério) pelo pesquisador.

Na teoria Bayesiana as inferências são também associadas a hipóteses, e os resultados são expressos em termos probabilísticos, através das chamadas Probabilidades *a posteriori* (que atualizam as Probabilidades *a priori*, com base nas evidências amostrais). Como em algumas situações (principalmente naquelas em que a distribuição *a priori* é não informativa) o resultado numérico tanto para a teoria clássica como para a Bayesiana é praticamente o mesmo, houve um relaxamento na preocupação

epistemológica com a teoria clássica, a qual continuou exercendo influência em praticamente todas as pesquisas quantitativas, com alguns pesquisadores aplicados relutando em aceitar análises Bayesianas, temendo não serem aceitos pela comunidade acadêmica, usuária dos testes de hipóteses clássicos. Além disso, como a teoria Bayesiana é altamente dependente de forte suporte computacional para os modelos mais complexos, este foi um fator determinante na escolha da teoria clássica. Com o avanço das ferramentas computacionais, este argumento tende a se tornar irrelevante.

Assim, uma das diferenças entre as escolas clássica e Bayesiana é que esta última incorpora informações *a priori* às evidências amostrais observadas pelo pesquisador com relação ao seu problema, enquanto que para a clássica isto não pode ser feito; na verdade, para a inferência clássica é relevante tanto a informação fornecida pelos dados da amostra, quanto o processo gerador da mesma. Isto equivale, na teoria Bayesiana a associar ao parâmetro uma distribuição de Probabilidade (distribuição *a priori*); no caso clássico, ao parâmetro não pode ser associada qualquer distribuição de Probabilidade, pois ele é considerado como uma entidade fixa.

Talvez não faça sentido procurar diferenças e semelhanças, porque a inferência Bayesiana viria a constituir um novo *paradigma* (Lindley, 1990). Como preconizava Kuhn (1975), este deveria se impor através de uma *revolução* (ou até que os seus críticos, geralmente de uma geração anterior, *deixem de existir*). Pela controvérsia na literatura sobre os fundamentos subjacentes à metodologia clássica, principalmente no aspecto dos testes de hipótese, diversos profissionais defendem as vantagens de substituí-la pela inferência Bayesiana.

1.3.4 Dificuldades dos alunos

O que se assiste, em grande parte dos cursos universitários, é uma aversão generalizada dos alunos pela disciplina de Estatística, geralmente oferecida nos

primeiros semestres da graduação, e que apresenta um alto índice de reprovação. Como já dissemos, os grandes vilões são o cálculo de Probabilidades e a inferência Estatística.

Com relação ao cálculo de Probabilidades, é oportuno destacar alguns problemas de natureza cognitiva, descritos em Lecoutre (1988,1998). Isso pode sugerir caminhos a serem considerados pelos professores dessa disciplina. Voltaremos a esse aspecto no capítulo final.

No que concerne ao ensino de noções de inferência, é preciso refletir se um desconhecimento total (por parte da grande maioria dos alunos ingressantes) dos procedimentos científicos em geral, bem como dos mecanismos de dedução e indução, em particular, não limitam a compreensão dos temas apresentados na disciplina. O que significaria para os alunos estabelecer hipóteses? Como já foi comentado, em sua grande maioria eles apresentam raciocínio determinístico, o que se constitui numa barreira para o desenvolvimento do pensamento estatístico/probabilístico, produto das informações apresentadas sem crítica e/ou discussão durante as etapas do ensino fundamental e médio e mesmo em outras disciplinas no próprio curso universitário: se dizemos numa aula na universidade que uma máquina apresenta 10% de peças com falha, muitos acham que se as nove primeiras vierem sem falha a 10^a, *com certeza*, virá com a falha!

Mesmo quase inexistente nos cursos básicos (e mesmo em outros mais avançados) a abordagem inferencial Bayesiana, é apresentada, pelos seus adeptos, como a mais adequada ao senso comum, por ser de raciocínio e interpretação diretos. Isto facilitaria, segundo eles, a compreensão dos alunos e os professores não teriam que recorrer à construção do edifício da inferência clássica, o qual, segundo dizem, está alicerçado em *bases duvidosas*.

Segundo Batanero (2000), uma possível causa de confusão dos alunos com relação ao ensino dos testes de hipóteses, em inferência clássica, pode estar na estrutura formal dos mesmos, que é similar àquela da prova por contradição (*modus tollens*) e, segundo Anderson (1995),

"o sucesso no mundo real depende muito mais do conhecimento conceitual disponível do que da capacidade de aplicar corretamente o raciocínio tipo modus tollens".

Também nem sempre fica claro ao aluno que antes do estabelecimento de uma hipótese Estatística existem uma hipótese substantiva, uma hipótese de pesquisa e uma hipótese experimental, e que é sobre essa última que poderão ser tiradas algumas conclusões, com certa margem de erro. Na verdade, espera-se do teste de hipótese mais do que ele pode oferecer e aos alunos não é dada nem a chance de discutir o procedimento. Tampouco se discutem os aspectos éticos ligados ao experimento, o valor teórico e prático da hipótese na qual se está interessado, e aspectos ligados ao custo do mesmo.

Segundo Popper (1975), o teste de uma hipótese Estatística é uma operação de caráter dedutivo: métodos estatísticos são essencialmente hipotético-dedutivos e operam por eliminação das hipóteses inadequadas. Ele afirma ainda que através de recursos a inferências puramente dedutivas (como o *modus tollens*, da lógica tradicional), já comentado acima, é possível concluir acerca da falsidade de enunciados universais a partir da "verdade" de enunciados singulares e que a indução é um mito. Oakes (1986) aproveita a posição de Popper para então perguntar se também a inferência Estatística é um mito. Ele (Oakes) acha razoável responder afirmativamente, uma vez que a inferência Estatística pode ser considerada como uma das facetas do problema não resolvido da indução.

Na verdade, os alunos têm que ter em mente que procedimentos inferenciais não nos dão um veredicto, mas podem ser encarados como ferramentas para que o pesquisador tire suas conclusões (pelo menos temporariamente). O problema é o uso indiscriminado de técnicas Estatísticas, às quais muitas vezes se dá uma interpretação indevida. O que tem que ser deixado claro, é que a Estatística dispõe de ferramentas de ajuda para o pesquisador de uma determinada área, mas que não deve substituir o próprio pesquisador e sua faculdade de pensar sobre a própria área de investigação.

1.4 Objetivos, destinatários e estrutura do trabalho

Desenvolvemos neste trabalho um roteiro complementar para dar significado à disciplina introdutória de Estatística em cursos universitários. Neste capítulo 1, introduzimos a necessidade de uma cultura Estatística por parte de futuros profissionais de qualquer área do conhecimento, bem como nossa inquietação pela falta de alfabetização em conceitos aleatórios de nossos ingressantes universitários, saudando com esperança a iniciativa dos PCN's em colocar uma cultura estatístico/probabilística no ensino pré-universitário. Além disso delineamos o ensino da Estatística básica na Universidade, introduzindo a discussão sobre as duas escolas de inferência, ou seja, a escola clássica e a escola Bayesiana. No capítulo 2 serão apresentadas algumas noções ligadas ao pensamento científico e ao método de fazer ciência, particularizando para o uso da Estatística; as escolas ligadas ao empirismo e ao racionalismo e suas ligações com as análises Estatísticas de pesquisadores também são apresentadas. O capítulo 3 destaca o papel da Estatística na investigação científica, formalizada a partir do século XX através de pesquisadores da escola anglo-saxônica. O paradigma da teoria clássica e a "candidata" a paradigma, a escola Bayesiana, são apresentados em seus aspectos objetivos e subjetivos, respectivamente. Como o cerne do trabalho diz respeito à inferência apresentada aos alunos numa disciplina básica de Estatística, o capítulo 4 contemplou inicialmente a inferência científica descrevendo os processos dedutivo e indutivo bem como alguns aspectos dos raciocínios condicionais do tipo *modus ponens* e *modus tollens* e sua ligação com a inferência Estatística de Fisher, de Neyman Pearson e a Bayesiana. Definir Probabilidade e Estatística e estabelecer a relação entre elas tem levado pesquisadores a muitas discussões e no capítulo 5 são apresentadas algumas definições para que se tenha uma idéia da variedade com que aparecem para os alunos e/ou interessados no assunto. Como o teorema de Bayes é a principal ferramenta do raciocínio Bayesiano, ainda neste capítulo ele é apresentado sob o ponto de vista probabilístico, através de dois exemplos: o primeiro refere-se a um experimento de bolas em urnas e o segundo a um experimento na área de saúde (meningite). No capítulo 6 apresentamos o que poderíamos chamar de inferência comparada, com exemplos de testes de hipóteses sendo resolvidos pelas técnicas de Fisher, Neyman Pearson e

Bayesiana, mostrando que é possível discutir as três abordagens, sob o ponto de vista técnico, somente com um conhecimento elementar de Probabilidade e Estatística, geralmente oferecido nas disciplinas introdutórias antes do tópico inferência. Ainda nesse capítulo é apresentado um exemplo de estimação. Atualmente as diferentes formas de raciocínio e modelos mentais, estudados em psicologia cognitiva, começam a interessar os pesquisadores em educação em Estatística, e o capítulo 7 enfatiza as relações da Estatística com a aprendizagem, com a cognição e com os modelos mentais; são também descritos alguns vieses apresentados pelos alunos ligados a tópicos da disciplina introdutória de Probabilidade e Estatística, bem como a tarefa de seleção de Wason, ligada a raciocínios condicionais. Concluímos o trabalho no capítulo 8, apresentando alguns contrastes entre as escolas clássica e Bayesiana, com uma sugestão de programa para uma disciplina introdutória de Probabilidade e Estatística, que contempla a coexistência entre essas duas escolas de inferência.

Ao preparar este trabalho tivemos em mente, em primeiro lugar, o professor de Estatística de um curso superior, que está ministrando a disciplina introdutória de Probabilidade e Estatística, na grande parte dos chamados "cursos de serviço". Esses professores dificilmente têm oportunidade de discutir com seus alunos aspectos importantes, que dão contexto a um tema tão complexo como a inferência Estatística. Em segundo lugar, o usuário de Estatística, advindo de outra área, interessado em conhecer alguns aspectos importantes que permeiam suas análises ou as apresentadas em artigos de sua área de concentração. A intenção é que este material sirva de reflexão e apoio para esses profissionais, e que, no caso daqueles de educação, os estimule a perseguir o objetivo de transformar a disciplina numa atividade desafiadora e inquietante para os alunos, ao mesmo tempo que interessante e útil.

Inferência Estatística não é um assunto simples mas não deve ser retirada da ementa de uma disciplina introdutória, como sugerem alguns professores. Ela deve ser ministrada não somente sob o aspecto técnico, mas também sob o ponto de vista epistemológico. Além disso deve ser mostrado o estado da arte do tema, discutindo os pontos fortes e as fraquezas das duas principais teorias que são a teoria clássica e a teoria

Bayesiana. Outras perspectivas de escolha para o ensino da inferência podem ser consideradas como fazendo parte desta bifurcação (como por exemplo, inferência não paramétrica).

A sugestão final é a coexistência do ensino das duas abordagens num curso introdutório, pois, como já há tempo sugeria Plackett (1966),

".....Minha visão é a de que as diferenças de opinião deveriam ser pacientemente exploradas, e que os pontos concordantes deveriam ser enfatizados. Estamos tratando com idéias que são mais profundas do que lógica e matemática e o ponto importante é que não devemos partir necessariamente das mesmas suposições mas sim obter, se possível, essencialmente as mesmas conclusões dadas as evidências."

Capítulo 2 Estatística e Método

2.1 Introdução

O conhecimento científico ligado às ciências empíricas é geralmente estruturado levando em conta o raciocínio estocástico, uma vez que depende fortemente do estudo das relações entre variáveis, as quais são intrínsecas à área de Estatística. Fazer ciência, para alguns pesquisadores, significa seguir um conjunto de procedimentos, aos quais quase sempre a Estatística está diretamente ligada e o nome *método* é muitas vezes invocado para esse conjunto. Não é difícil ouvir nos meios acadêmicos, principalmente de professores de Estatística desgostosos com o baixo aproveitamento dos alunos, que antes de ensinar Estatística ao corpo discente de diferentes áreas seria preciso que os mesmos tivessem noções básicas do que vem a ser *método científico*. Mas o que vem a ser, de fato, o método científico? Existe o *método científico*? Neste capítulo vamos descrever, ainda que superficialmente, o desenrolar do pensamento científico e sua ligação com a Estatística, buscando ligar essas idéias com o que se entende por *método*.

2.2 Domínio da incerteza

Desde tempos imemoriais o Homem busca o conhecimento de sua origem e de seu destino. Os séculos XVII - XIX assistiram a grandes transformações que levaram tanto à Revolução Industrial como à chamada Revolução Científica, acompanhando o desenvolvimento do espírito crítico do ser humano. Com o avanço do conhecimento científico, que ocupou espaço de outros saberes, como religião, arte, poesia, etc., acreditou-se que era uma questão de tempo chegar à *verdade* sobre a Humanidade e esta não era somente uma crença do homem comum, não afeito ao pensamento científico. Até mesmo Laplace (1749-1827), astrônomo, físico e matemático francês, cujo nome está definitivamente ligado ao cálculo de probabilidades, principalmente devido ao texto *Ensayo filosófico sobre las probabilidades*, tendo igualmente contribuído para o

desenvolvimento inicial da Estatística Matemática, acreditava que, em determinado momento, poderia haver uma inteligência para quem nada seria incerto, uma vez que nossa incerteza se devia a problemas decorrentes de atrasos tecnológicos que, ao serem sanados, conduzir-nos-ia à “*verdade*”. Este determinismo de Laplace foi um exemplo da crença: “*basta medir para dominar, basta explicar para compreender*”.

O século XIX terminou mas ainda continuou a expectativa de que o domínio sobre as questões da natureza era iminente, pois acreditava-se que as hipóteses científicas, uma vez confirmadas pela experiência, se tornariam *verdades* absolutas (portanto inquestionáveis). No entanto, já na primeira metade do século XX, algumas teorias supostamente verdadeiras tiveram que ser abandonadas, com novas teorias rompendo com as anteriores, sem necessariamente se constituírem na palavra final. Algumas delas podem ser consideradas como responsáveis por esta ruptura e dentre elas citamos a teoria da relatividade de Einstein (1879-1955), o princípio da incerteza de Heisenberg (1901-1976), o princípio da inconsistência de Gödel (1906-1978), todas elas desenvolvidas nas três primeiras décadas do século XX.

A busca da precisão teve como resultado a imprecisão e parece já não existir a crença de que a Ciência a qualquer momento vai fornecer todas as respostas. Atualmente, cientistas associados a um grande projeto internacional, liderado pelos EUA, denominado Genoma Humano, dedicado a decifrar o código genético, estão convencidos (pelo menos por enquanto!) de que nem tudo na natureza humana pode ser explicado pelo DNA, o que colocou em xeque muitos conceitos ligados à área de genética (já *consagrados*!). Podemos dizer que entramos no século XXI sem nenhuma certeza de estarmos caminhando em direção à *verdade*. Na opinião de alguns filósofos, é possível que estejamos cada vez mais distantes dela e o que faria sentido seria refletir sobre a realidade: o que passa a ser importante é a imagem do mundo que pode estar contida no espírito humano.

Esse estado de coisas sugere a reaproximação entre pensamento científico e pensamento filosófico, uma vez que a Ciência, não sendo capaz de conquistar a precisão,

encontrou limites na explicação do Universo, mostrando que não há certezas definitivas. Parece haver consenso de que a já comentada hegemonia do conhecimento científico sobre as áreas artísticas e/ou filosóficas, apesar de ter apresentado sucesso não desprezível em áreas como a Medicina, por exemplo, não se traduziu numa melhora significativa do bem estar do ser humano como um todo. Citando Santos (1989):

"... é alienante concluir que a acumulação de tanto conhecimento sobre o mundo se tenha traduzido em tão pouca sabedoria do mundo, do homem consigo próprio, com os outros, com a natureza".

É preciso compreender a prática científica para além da consciência ingênua ou oficial dos cientistas e das instituições de ciência (Santos, 1989), e evidentemente o raciocínio estocástico não pode ficar à margem dessa compreensão.

2.3 Criação científica

O ser humano continua a se dedicar a estudos tanto teóricos como aplicados, experimentais e/ou observacionais, sendo que a Ciência tem desempenhado, em muitos casos, o papel de uma âncora em terra supostamente firme. Ciência (do latim, *scientia*, conhecimento, arte, habilidade, de acordo com Larousse) é considerada como sendo um conjunto organizado de conhecimentos a respeito de uma certa área, e é costume chamar o procedimento de aquisição desse conhecimento de *método científico* (expressão ambígua, segundo Bunge, 1980). Os últimos séculos assistiram ao desenvolvimento da chamada ciência ativa, em oposição à ciência contemplativa da Antiguidade. Conhecimento ficou, pouco a pouco, sendo reconhecido como conhecimento científico, associado à produção tecnológica e o termo ciência foi então reservado para áreas em que a idéia de *progresso* (sempre sob o aspecto tecnológico) tem um sentido óbvio (o que evidentemente não é o caso nem da Arte e nem da Filosofia).

Fazer ciência significa, dentre outras coisas, sair do senso comum, buscando resultados consistentes e fecundos. Mas como se faz ciência? Será que existe um *método*, a partir do qual, os resultados são certamente obtidos?

A comunidade científica esteve, por muito tempo, à procura de um conjunto de procedimentos que garantissem a descoberta de quais afirmações seriam *verdadeiras*, e portanto significativas, e quais as falsas, sujeitas a serem desconsideradas. Apesar de a busca por um *método* científico, com um conjunto de regras pré-estabelecidas a serem seguidas, ter tornado mais rigorosa cada afirmação proveniente de investigação científica, o tempo mostrou que a esperança de encontrar tal conjunto foi se tornando cada vez menor.

A palavra *método* pode encerrar em si mesma uma discussão pois deriva etimologicamente do grego *methodos* que significa caminho para chegar a um fim, sendo assim definida em Larousse (1998) “(filos.) marcha racional do espírito para chegar ao conhecimento ou à demonstração de uma *verdade*”. Esta definição permitiria acreditar que é possível a existência de um *método* no qual podemos confiar para atingir o conhecimento pleno a respeito de determinado assunto. É claro que sempre seguimos um caminho para atingir uma meta; no entanto não existe um modelo de caminho para atingir todas as metas.

2.3.1 Empirismo e Racionalismo

Uma das primeiras sistematizações *metodológicas* de que se tem notícia foi feita por Bacon (século XVII) que pretendeu esgotar toda a possibilidade de conhecimento através do *método* de acumular observações, caracterizando o início formal do **empirismo**, uma posição filosófica que adota o princípio de que todo o conhecimento deriva de uma base empírica e que tem como lema uma frase de inspiração Aristotélica:

“Nada está no intelecto que não tenha passado pelos sentidos”.

A grande contribuição de Bacon foi a defesa do *método* experimental contra a ciência clássica teórica e especulativa, defesa essa baseada na concepção de um pensamento crítico sobre o progresso da ciência e da tecnologia. Este *método* não tinha teorias *a priori* e preconizava que a *verdade* iria aparecer pela simples justaposição de dados, colhidos criteriosamente. Ou seja, de acordo com a teoria de Bacon, bastava um conjunto de dados para se obter uma informação de cunho científico e com isso fazer a ciência avançar.

Descartes, contemporâneo de Bacon, desenvolve o *método racionalista*, em que a razão é o ponto central a partir do qual o conhecimento é criado – essa perspectiva leva em conta que o homem traz dentro de si a possibilidade do conhecimento, com a racionalidade fazendo parte da condição humana. A partir de algumas premissas básicas, a respeito da própria gênese do fenômeno, o cientista deve deduzir o comportamento do mesmo e só então testar o modelo experimentalmente. Essas duas escolas são consideradas como *paradigmas* da primeira fase da filosofia moderna (até o século XIX).

No século XX a tradição empirista teve continuidade no chamado positivismo lógico, que incorporava as idéias positivistas de Comte ao formalismo da lógica matemática, cujos seguidores fundaram o Círculo de Viena em 1929. Esse movimento caracterizou-se pela ênfase na necessidade de um rigor metodológico para a investigação científica e pela idéia de que as teorias, para serem consideradas científicas, teriam que se submeter a um critério de verificação. Com o advento da 2ª Guerra Mundial, o grupo se dispersou e houve um enfraquecimento da confiança exagerada na força explicativa das teorias científicas. Mesmo assim pode-se dizer que ainda hoje o positivismo exerce sua influência em ambientes de pesquisa.

A pesquisa, em seu processo de investigação, dispondo cada vez mais de dados coletados, cada vez mais faz uso do raciocínio estocástico (probabilístico/estatístico). Grosso modo, sem entrar nas ramificações e transformações que as escolas empirista e

racionalista enfrentaram, temos notado que os pesquisadores que utilizam análise Estatística em seus trabalhos são prioritariamente de dois tipos:

* **empirista**, que coleta dados por vocação e/ou tradição e deseja que a Estatística dê sentido aos seus dados;

* **racionalista**, que a partir do fenômeno de interesse constrói um modelo teórico a partir da gênese do próprio fenômeno e que usa as técnicas em Estatística para *comprovação* de sua teoria.

Em muitos anos de prática de auxílio a pesquisadores acadêmicos de diferentes áreas, encontramos muito mais aqueles identificados com o primeiro tipo, empirista, que usam a Estatística no início da análise, do que do segundo tipo, racionalista, que vão à gênese do fenômeno, para então testá-lo empiricamente, fazendo uso da Estatística na etapa final da análise.

Esse comportamento se deve, em primeiro lugar, ao ambiente típico em que as pesquisas se desenvolvem, na maior parte das vezes em laboratórios, em que os novatos se juntam a grupos de pesquisa já formados por pesquisadores mais experientes, continuando as linhas já iniciadas, com as respectivas coletas de dados (quando for o caso), nem sempre sabendo bem porque este é o problema a ser enfrentado pelo grupo, qual o paradigma atual nessa parte da ciência que estão desenvolvendo, quais são as propostas de ponta emergentes, etc. Isso acaba num círculo vicioso, em que os novatos de agora se tornarão os experientes de depois, que por sua vez orientarão os novatos e assim por diante. Nesse caso a reflexão nem sempre é um ingrediente cotidiano e há casos em que valores espúrios em relação à massa de dados coletada e/ou conhecida são deixados de lado sem maiores referências, mesmo que possam se constituir em novidades no ramo. As ciências biológicas e afins, com seus experimentos, e mais recentemente as ciências sociais, com seus estudos na maioria observacionais, são áreas em que encontramos exemplos desse comportamento.

Em segundo lugar, há a dificuldade de formação em modelagem para que o segundo tipo (racionalista) possa ser mais utilizado. Neste caso, um conhecimento matemático é praticamente indispensável, e este não é o tópico preferido da maioria dos pesquisadores das áreas ditas não exatas, e o trabalho interdisciplinar não é ainda o cotidiano da maioria dos ambientes de pesquisa. Uma vez conhecido bem o fenômeno quanto à sua gênese e, com conhecimentos básicos de análise matemática, é possível em muitos casos descrever um modelo teórico para o mesmo, estudar suas propriedades, seus casos limite, suas aplicações, para somente então recorrer ao método experimental, com coleta de dados, para eventualmente validar tal modelo. Neste caso podemos dizer que o uso da Estatística é feito *a posteriori*. As ciências físicas, pela proximidade com o raciocínio matemático, utilizam bastante este tipo de procedimento.

2.3.2 Método ou métodos

Qualificar a palavra *método* com o termo científico – *método científico* – pressupõe

um desejo de normatizar a maneira de fazer ciência: seguindo certos procedimentos pré-estabelecidos chegar-se-ia ao conhecimento (uma clara influência da escola positivista). O termo é muitas vezes invocado mas raramente definido. Alguns autores apresentaram ementas curriculares do que seria uma disciplina de metodologia científica. Segundo Gale (1979), por exemplo, a visão moderna de *método científico*, presente por várias décadas em disciplinas de metodologia ministradas em escolas norte-americanas, inclui: a) montagem do experimento; b) observação de ocorrências no experimento; c) estabelecimento de uma hipótese inicial para descrever a ocorrência; d) construção de novos experimentos para testar a hipótese; e) correção ou modificação da hipótese à luz dos resultados do último experimento. Esta é também a sequência sugerida por Beveridge (1981), que discute algumas dificuldades lógicas e práticas dos itens apresentados. É possível identificar nesses passos a influência de Bacon e da teoria empirista, bem como

a necessidade de ferramentas para o tratamento dos valores observados, pois em algumas das etapas descritas é possível perceber a necessidade da abordagem Estatística.

Na literatura Estatística encontramos, na segunda metade do século XX, autores consagrados na área de Estatística fazendo alusão ao *método científico*, para definir ou caracterizar a Estatística. Good (1989), que defende os princípios da escola Bayesiana, caracteriza *método científico* em termos de 14 itens, e considera Estatística como sendo parte do *método*, fazendo uso de todas as características. A definição dada por Kendall e Stuart (1969), representantes da escola clássica de inferência, é a seguinte:

“ Estatística é o ramo do método científico que trata com observações obtidas por contagem ou mensuração de características de populações de fenômenos naturais ”.

Apesar das citações acima conterem o termo *método científico*, elas também não acenam com uma definição para o mesmo, mas ajudam a manter a indefinição em torno de sua existência bem como incentivam a recomendação de que se deva ensinar aos alunos procedimentos básicos de *metodologia científica* – principalmente aos que farão uso do raciocínio estocástico para tomada de decisão em pesquisas empíricas. No fim do capítulo retomaremos essa discussão.

Inúmeros são os caminhos que a mente humana percorre na pesquisa científica. Na sua obra “A Criação Científica”, Moles (1971) descreve alguns *métodos*, por ele chamados de *métodos heurísticos*, que descrevem os processos mentais e/ou comportamentais que caracterizam os primeiros passos dessa criação científica. Para ele, o funcionamento do espírito humano que dará origem à criação científica (ciência em vias de se fazer) não é necessariamente lógico, embora a ciência acabada siga os cânones da lógica formal. Tantas maneiras diferentes de deixar vir à tona o pensamento científico, como as descritas por Moles, de fato não podem ser confinadas numa única estratégia metodológica. O caminho científico não está pronto e é preciso enveredar por

uma rede de procedimentos heurísticos, seguindo muitas vezes a orientação vaga que usamos ao percorrer labirintos. Lembrando A. Machado (1979):

"Caminante no hay camino, el camino se hace al andar".

Ainda de acordo com Moles, a *ciência do provável*, que utiliza a Estatística como ferramenta e na qual o homem tem um papel abstrato (isto é, não aparece), está se transformando na *ciência do percebido*, em que passa a ser importante o modo pelo qual as pessoas fazem ciência, incluindo aí aspectos psicológicos, políticos, sociais e econômicos. Podemos relacionar a oposição entre ciência do provável e ciência do percebido com a oposição entre análise clássica e análise Bayesiana, nessa ordem. Voltaremos a este tópico no próximo capítulo.

De qualquer forma é preciso ter em mente que a ciência a ser desenvolvida dependerá menos do pesquisador individual e mais da política das instituições de pesquisa e dos órgãos financiadores da mesma, tributária da filosofia presente nos correspondentes planejamentos. Assim

"Uma filosofia idealista sugerirá como modelo da ciência, a torre de marfim; uma filosofia empirista inspirará o fomento da pesquisa empírica sem orientação teórica; uma filosofia pragmatista inspirará desprezo pela investigação básica, e assim por diante. Somente uma epistemologia equilibrada poderá inspirar uma política equilibrada da ciência, uma política que fomente o desenvolvimento integral e ininterrupto tanto da investigação básica como da pesquisa aplicada."
(Bunge, 1980).

A política utilizada terá, portanto, grandes implicações no desenvolvimento científico e na maneira como a ciência será considerada, tanto pelos meios acadêmicos quanto pela sociedade.

Como pensar na existência de um *método científico* se, continuamente, a ciência está derrubando algo como comprovado (pelo *método científico*) por algo novo, também comprovado (pelo *método científico*)? De fato, concordamos com Bunge,

“...a expressão método científico é enganosa pois pode induzir a crer que consiste num conjunto de regras exaustivas e infalíveis que qualquer um pode manejar para inventar idéias e pô-las à prova.

.....

Método Científico não é nada mais do que a maneira de fazer boa ciência, natural ou social, pura ou aplicada, formal ou factual. Posto que o método científico é a maneira de conduzir investigações científicas, não pode ser apreendido separadamente destas últimas. Vai-se dominando o método – e talvez também modificando-o – à medida que se faz investigação original”.

Isto parece ser o que Mayo (1996) chama de regras metodológicas: procedimentos para conduzir investigações confiáveis, evitando erros fatais. Tais procedimentos, em sua opinião, incluem aleatorização, experimentos controlados, conveniência de evitar hipóteses circunstanciais, etc. As regras citadas por Gale poderiam também fazer parte deste contexto, ao invés de caracterizarem, como se pretende, o *método científico*. Na verdade poderíamos descrever vários conjuntos de regras de procedimentos que também seriam sujeitos a controvérsias, diretamente ligados à área ou objeto de estudo, à escola de análise a ser utilizada, ou ainda ao estilo do pesquisador.

Segundo a chamada epistemologia anarquista de Feyerabend (1975), para quem reduzir a Ciência a um *método* é o mesmo que destruí-la, a Ciência não tem nenhuma característica especial que a faça mais digna de crença do que um mito qualquer. Esta foi uma frase muito polêmica, suavizada posteriormente pelo autor, mas que poderia ser equivalente ao que disse Bunge (1980):

“A criação original, ao contrário das tarefas rotineiras, não parece ser regulamentável”.

Mais polêmico, Earman (1996) faz objeções ao que ele chama de “caçada” para encontrar uma *metodologia* da ciência, e diz concordar com Feyerabend, não por anarquismo mas, dentre outras coisas, porque tais cientistas perdem tempo em tentar demarcar o limite entre a ciência e não ciência. O que é necessário, segundo Earman, é

“ um pouco de Bayesianismo e bastante reflexão tranqüila.”

2.4 Conclusão

Em nossa opinião, de fato não existe o *método científico*, universal, que molde todas as formas de conhecimento, capaz de provar que as teorias científicas ou são *verdadeiras* ou são *falsas*, mas sim métodos, procedimentos ou regras de acordo com o pesquisador e o ramo ao qual ele se dedica, num processo de evolução em direção a metas, pré-traçadas ou não. Dada a natureza estocástica da maioria dos fenômenos estudados pelos pesquisadores, o raciocínio e as técnicas probabilísticas/Estatísticas estarão quase sempre presentes em alguma dessas etapas a serem cumpridas, mas não devem ser usadas cegamente, a técnica pela técnica, como se a resposta brotasse diretamente das análises. Uma atitude menos pragmática no uso da Estatística, por parte dos “fazedores” de ciência, certamente seria benéfica.

Sendo assim, não vemos sentido no oferecimento de uma disciplina de metodologia da ciência como pré-requisito à disciplina introdutória de Estatística, embora achemos que discussões como as apresentadas aqui poderiam ser material de algumas aulas da disciplina básica de probabilidade e Estatística num curso superior. Para facilitar a compreensão dos alunos pelos métodos estatísticos seria ideal vinculá-los à prática científica propriamente dita que ocorre ou no seio da instituição ou em outros locais. O que de fato reforçaria a necessidade da Estatística em cada área seria uma disciplina básica integrada com as disciplinas de natureza quantitativa da própria área de

concentração do aluno – a interdisciplinaridade viria favorecer a contextualização e conceitos e técnicas poderiam ser mais bem discutidos e analisados.

Nos capítulos seguintes, sempre que aparecer a palavra método, ela estará sendo utilizada em seu sentido mais amplo, ou seja, simplesmente como sinônimo de regra ou procedimento.

Capítulo 3 A importância da Estatística

3.1 Introdução

As Estatísticas e probabilidades surgem de modo mais visível apenas no século XVII e seu desenvolvimento acompanha a expansão capitalista do hemisfério norte. Índices, taxas, porcentagens, médias, tendências, curvas, mapas mercantis, medidores de audiências, histogramas e “tortas” agregam massas heterogêneas, resumem uma multiplicidade de histórias singulares, expõem o perfil global de uma pluralidade dispersa de acontecimentos (Lévy, 1999). O verbete *Estatística* foi introduzido no século XVIII, com origem na palavra latina *status* (Estado), e serviu inicialmente a objetivos ligados à organização político-social, como o fornecimento de dados ao sistema de poder vigente.

No século XIX, o impulso das Estatísticas sociais se baseou na sua capacidade de estabelecer leis sobre fatos da sociedade, que poderiam justificar medidas políticas: criminalidade, saúde, educação, etc. (Besson, 1995). Ainda no final do século XIX foram feitas algumas tentativas no campo da amostragem, notadamente nos Estados Unidos da América, para sondagem de opinião em época de eleições, sem ainda uma preocupação explícita com a representatividade da amostra. Conforme descrito em Driesbeke (1997) também os russos, nessa época, desenvolveram algumas técnicas de amostragem para conhecimento da população e território, que somente foram divulgadas na Europa em meados do século XX.

A partir do início do século XX teve início o desenvolvimento formal de uma teoria Estatística, fortemente embasada em raciocínio probabilístico e ferramental matemático. A partir do Círculo de Viena (1929), a Estatística começa a ter um papel de destaque na investigação científica, com o raciocínio e técnicas da escola clássica proliferando em todas as ciências factuais, primeiro com as Ciências Naturais e mais tarde com as Ciências Sociais, trazendo uma esperança de objetividade com as

conclusões obtidas. Como já foi visto no capítulo anterior, nessa época havia entre os intelectuais grande influência do positivismo lógico, que aliava a tradição empirista ao formalismo matemático. Para essa corrente de pensamento uma afirmação só teria sentido se fosse ou provada analiticamente ou verificada empiricamente. A inferência Estatística apresentou-se então, através da escola clássica, como uma ferramenta para preencher este critério de verificação: os testes de hipóteses foram considerados como a resposta probabilística para confirmação de hipóteses (embora esta não fosse inicialmente a idéia defendida por Fisher). Popper (1902-1994), muitas vezes identificado com o positivismo lógico, aproximou-se do Círculo de Viena mas depois afastou-se, criticando a lógica indutiva e o princípio da verificabilidade: para ele a *indução* não era garantia de sucesso na busca do conhecimento, simplesmente porque é um procedimento falho e somente os procedimentos que usam a *dedução* são logicamente válidos. O desenvolvimento dos métodos estatísticos e dos campos de aplicação, com as correspondentes interpretações nem sempre corretas, fez a Estatística, na opinião de Efron (1998), passar de *uma pequena área obscura para uma grande área obscura*.

As áreas notadamente experimentais, como as Ciências Biológicas, estão à frente cronologicamente no uso do raciocínio probabilístico/estatístico principalmente através da Estatística clássica, constituindo-se em muitos casos a própria gênese de muitas das técnicas comumente usadas hoje em dia. O Centro de Estatística Aplicada do Instituto de Matemática e Estatística da Universidade de São Paulo (CEA IME USP) tem atendido pesquisadores de várias áreas do conhecimento ao longo dos últimos vinte anos e, apesar de contar com a participação cada vez maior de várias áreas da universidade, um levantamento mostra que, nos últimos cinco anos, pelo menos 60% das análises estatísticas foram feitas para pesquisadores das áreas de Ciências Biológicas e afins.

A tradição do uso do raciocínio estatístico nas ciências biológicas começou com *sir* R. Fisher (1890-1962), um cientista inglês da primeira metade do século XX, considerado por Efron (1998) a figura mais importante da Estatística do século. Ele trabalhava na estação experimental de agronomia de Rothamsted, na Inglaterra, e desenvolveu parte de sua teoria em resposta aos problemas de aplicação a que era

submetido. Os aspectos inferenciais de sua teoria foram muito criticados e seus contemporâneos, J. Neyman (1894 -1981) e E. Pearson (1895-1980, filho do famoso pesquisador inglês Karl Pearson), apresentaram uma teoria com alguma semelhança à de Fisher, mas com outro encaminhamento com o qual se propunham a resolver as críticas apresentadas.

A teoria Bayesiana teve origem a partir de um resultado atribuído ao pastor inglês T. Bayes (1702-1761), cujo teorema, que leva seu nome, foi publicado postumamente em 1763. A inferência Bayesiana utiliza a opinião *a priori* do pesquisador que será devidamente atualizada pelos dados através da aplicação do teorema de Bayes. Este teorema em si não causa nenhum tipo de controvérsia ao ser usado no âmbito probabilístico. No entanto o seu uso para efeitos inferenciais, que considera uma distribuição *a priori*, representando a *opinião* do pesquisador, é colocado em dúvida pelos adeptos da escola clássica.

Fisher e a dupla Neyman - Pearson, embora pertencentes à escola frequentista, praticamente romperam relações, numa atitude semelhante à discórdia atual entre clássicos e Bayesianos. Em um ponto porém os três concordavam: iriam desenvolver uma teoria *perfeitamente objetiva*, descartando os chamados métodos Bayesianos e seu *subjetivismo*.

Uma das restrições ao uso do método Bayesiano, é que pessoas com experiências diferentes e opiniões iniciais diferentes terão distribuições *a priori* diferentes, que podem conduzir a resultados finais muito distintos. A crítica, baseada no subjetivismo que as opiniões iniciais podem trazer ao problema, é uma herança do positivismo, que renegava o conhecimento *a priori*, limitando-se a estabelecer relações entre os fenômenos observados. Essa possível divergência de conclusões não acontece com o método clássico: duas pessoas, mesmo com opiniões iniciais diferentes, chegarão ao mesmo resultado de análise, uma vez que suas opiniões *a priori* não são incorporadas na análise. O procedimento Bayesiano é considerado subjetivo e o clássico objetivo. Para a escola clássica, a subjetividade é uma das principais falhas da escola Bayesiana. Esta,

por sua vez, acredita que a objetividade da escola clássica é pretensiosa e inútil. Vamos analisar a seguir esses dois termos e sua relação com a oposição entre as escolas clássica e Bayesiana.

3.2 Objetivismo e subjetivismo

Da mesma forma que, sob o ponto de vista semântico, *subjetivo* e *objetivo* são dois adjetivos que se definem por oposição entre si, boa parte da comunidade Estatística vê uma oposição entre as mencionadas escolas ou mesmo participa de tal oposição. O subjetivismo da escola Bayesiana é encarado muitas vezes pejorativamente pelos seus opositores, membros da escola clássica e defensores do objetivismo, pelo receio de que tudo seja permitido.

Na verdade, a postura Bayesiana mais parece se aproximar da *ciência do percebido* de Moles (1971), mencionada no capítulo anterior: nesse contexto passa a ter importância o pensamento, a observação pessoal e até o instinto. Como já vimos, para Moles a *ciência do provável*, na qual o homem tem um papel abstrato, está deixando lugar para a *ciência do percebido*. O pretenso método *objetivo* de observação da teoria clássica poderia ser equiparado à ciência do provável, e o citado movimento desta em direção à ciência do percebido poderia ser pensado como o movimento de procurar na Estatística Bayesiana uma alternativa para a Estatística clássica. Esta última não leva em conta as dificuldades da isenção total, embora devesse fazê-lo, uma vez que os dados são coletados por um observador:

“Lembremo-nos que nós não observamos a natureza ela mesma, mas a natureza submetida ao nosso método de investigação”. (Watzlawick, 1988).

Enquanto o termo *subjetivo* diz respeito somente ao pensamento humano, a palavra *objetivo* (por oposição) diz respeito somente ao objeto a ser examinado, com referências exteriores ao espírito. O estatístico clássico, sendo (ou parecendo) objetivo,

não se deixa influenciar pelos seus pensamentos, e espera que os seus resultados transmitam a *verdade* sobre o seu experimento. O estatístico Bayesiano, ao ser subjetivo, interpreta a sua população de interesse *a priori* (ou seja, antes de incorporar os dados observados) e tira as suas conclusões *a posteriori* (ao incorporar os resultados observados à distribuição *a priori*).

Lakoff e Johnson (1980), em um texto em que discutem como os conceitos metafóricos estão diretamente ligados à linguagem, pensamento e ação em nossa vida cotidiana, apresentam os “mitos” do objetivismo e do subjetivismo, que serão aproveitados aqui para fazer um paralelo com as atitudes dos adeptos de uma escola contra a outra.

Clássico contra Bayesiano

- faz parte da condição humana ter erros de percepção – portanto não podemos confiar no juízo subjetivo dos indivíduos ao pretenderem estabelecer distribuições *a priori*;
- os métodos proporcionados pela Estatística clássica são baseados em metodologia universalmente válida, acima de qualquer subjetividade;
- somente o conhecimento objetivo é *verdadeiro* e nos permite uma visão imparcial do mundo;
- a subjetividade é irracional e portanto indesejável na argumentação científica.

Bayesiano contra clássico

- as intuições dos pesquisadores são os melhores guias para as ações e não colocar as distribuições *a priori* seria abdicar desta condição;

- a objetividade pode ser inútil em certos casos – trata, por exemplo, com valores médios associados a grandes conjuntos de informações e não dá a devida atenção para valores associados a um indivíduo em particular;
- a objetividade pode ser injusta porque ignora os aspectos mais relevantes de nossa experiência particular em favor do universal e impessoal.

Os pontos colocados acima podem ser considerados como uma visão simplificada das razões do antagonismo entre os dois grupos, mas podem ser encarados como pontos de partida para um confronto. Lakoff e Johnson apresentam o que eles chamam de posição alternativa ao objetivismo e ao subjetivismo, denominada através do neologismo “experencialismo”, que satisfaz os interesses de um e de outro mas sem acreditar nem na verdade absoluta nem em sua necessidade (objetivismo) e também sem compartilhar a idéia de que a imaginação está livre de limitações (subjetivismo). Essa posição de síntese é adotada pelos mencionados autores, que consideram que a verdade é função de nosso sistema conceitual, e que pessoas com sistemas diferentes entendem o mundo de modo diferente. Tal visão está mais de acordo com o ponto de vista Bayesiano, pois parece estimular a colocação de distribuições *a priori*, que retratem o conhecimento anterior do pesquisador.

3.3 O uso da Estatística

A Estatística exerce vários papéis interligados, na pesquisa, no ensino, na comunidade. No campo científico podemos considerar como sendo duplo o seu papel: de um lado como uma ciência, que se desenvolve como uma disciplina da Matemática, como a Estatística matemática, os processos estocásticos, os aspectos lógicos associados à tomada de decisão, etc. e, de outro, fazendo parte de um procedimento inerente às diversas ciências factuais para as quais ela é invocada. As ciências factuais precisam

testar experimentalmente suas hipóteses substantivas, diferentemente das ciências formais (lógica, matemática) que usam principalmente raciocínios dedutivos para demonstrar resultados. Segundo Batanero (2000) as hipóteses substantivas nas ciências factuais não podem ser provadas diretamente, pois em geral se referem a entidades teóricas. São então transformadas em hipóteses de pesquisa, as quais também nem sempre podem ser confirmadas diretamente, de onde decorre a necessidade da construção de um instrumento para que se teste uma hipótese experimental, momento este em que a Estatística ocupa um lugar de destaque.

Tanto na corrente do racionalismo como na do empirismo discutidas no capítulo anterior, o que é mais comum sob o ponto de vista de análise Estatística é a procura por uma **análise inferencial clássica**. Isto se dá por várias razões que incluem: exigência destas técnicas pelos *referees* das revistas científicas, disponibilidade das técnicas na literatura existente, disponibilidade de aplicativos estatísticos para as análises, hábito do pesquisador e/ou orientador. Tais análises vão desde um teste de comparação entre médias de dois grupos (supostamente trivial), cuja técnica tradicional (dadas certas premissas) seria um teste de hipóteses usando a distribuição “*t*” de Student, até uma abordagem multivariada com ferramental de apoio forte em álgebra linear, como por exemplo uma análise de regressão multivariada.

Com essa descrição não estamos querendo fixar dois extremos da abordagem Estatística clássica: simplesmente foi dado um exemplo de técnica considerada *simples* versus uma técnica considerada mais *sofisticada*. Em grande parte dessas análises, o pesquisador está em busca de um resultado *significativo* (que equivale, por exemplo, na linguagem técnica, a rejeitar a hipótese – chamada de hipótese nula – de que os grupos apresentam a mesma resposta frente a dois tratamentos), traduzido pelo chamado nível descritivo p (“*p-value*”). É através desse valor que o pesquisador decide se os grupos podem ser considerados como diferentes (com a correspondente rejeição da hipótese nula) ou se a diferença observada entre eles pode ser atribuída ao acaso (sendo que neste último caso o resultado é dito não significativo e a hipótese nula não é rejeitada).

Por sua vez, a escola Bayesiana está sendo paulatinamente difundida nos trabalhos de aplicação, mas ainda de uma maneira incipiente. Em primeiro lugar, porque grande parte dos textos básicos não lhe dão espaço; em segundo, porque ela depende fortemente de recursos computacionais e, por último, porque as comunidades, que usam tradicionalmente as técnicas clássicas, preferem continuar a fazê-lo. Esta abordagem também procura resolver os problemas descritos nos parágrafos anteriores e a técnica reside em avaliar como os dados mudam a impressão que o pesquisador tem *a priori* sobre um dado fenômeno. Ela permite que, além de se poder incorporar na análise informações anteriores, os resultados sejam revistos enquanto o experimento está sendo realizado e isto pode representar uma economia de custos ou mesmo uma possibilidade de reverter uma situação durante um ensaio, se os resultados, obtidos em tempo real, mostrarem que isso é mais conveniente.

Uma das grandes polêmicas entre as duas escolas diz respeito à coleta de dados. A Estatística clássica desenvolveu a área de planejamento experimental para orientar como os dados devem ser coletados, a fim de preservar a coleta objetiva, enquanto que a escola Bayesiana parte do pressuposto de que toda a informação de nosso interesse está contida nos dados, não importando o modo como eles foram retirados: o tipo de análise a ser utilizado no caso clássico depende fundamentalmente do procedimento utilizado na coleta de dados, enquanto que na análise Bayesiana isto não vem ao caso, e os dados são analisados independentemente do método utilizado para a coleta, desde que o pesquisador o considere coerente. Blackwell (1986), que como já foi dito no capítulo 1 é um estatístico da linha Bayesiana, diz que se surpreendeu com o avanço dessa teoria, mas acha que ela deixa a desejar do ponto de vista de planejamento de experimentos. Ele sugere a aleatorização não como uma idéia subjacente à teoria Bayesiana, o que não é, mas como uma proteção contra possível imperfeição humana.

A técnica clássica é menos flexível sob os dois aspectos recém mencionados: o pesquisador deve evitar incorporar suas opiniões na análise para não trazer vícios nos resultados e deve cuidar para que o tamanho inicial da amostra ou o planejamento experimental seja cumprido, sem interromper o experimento. É portanto muito fácil de

ver duas características opostas entre as duas escolas; a clássica, procurando ser objetiva, não se deixando influenciar por opiniões e a Bayesiana, ao contrário, incorporando a subjetividade na estratégia de análise. Muitos pesquisadores, mesmo não discordando da filosofia Bayesiana, preferem continuar a pensar da forma tradicional e aplicar o método clássico. Um exemplo dessa postura é o comentário feito pelo bioestatístico L. Kessler, da *Food and Drug Administration* (FDA) dos Estados Unidos, em artigo da revista *Science* (1999, v. 286, p. 1460-1464):

“Nós não vamos mudar o paradigma estatístico da noite para o dia”.

3.4 As escolas Clássica e Bayesiana como paradigmas

O termo *paradigma* foi utilizado algumas vezes neste texto e como se ajusta bem à controvérsia entre as duas escolas vamos explorá-lo um pouco. A forma latina é idêntica à da língua portuguesa e significa modelo, norma, padrão. Kuhn (1975) o caracteriza como a existência de um campo de pesquisa que apresenta duas características:

- concentração em torno de si de um grupo duradouro de partidários;
- comportamento suficientemente aberto para abrigar toda a espécie de problemas relativos à área.

Horgan (1998) diz que paradigma é um conjunto de procedimentos ou idéias que instruem os cientistas, implicitamente, sobre aquilo em que devem acreditar. Segundo Gale (1979) um paradigma é aquilo que é ensinado nos textos didáticos, nas escolas fundamental e média e até mesmo nas universidades.

Com essas colocações, e sem entrar nas discussões polêmicas sobre o termo, podemos dizer que o paradigma “atual” é a teoria de inferência clássica, uma vez que, tanto nos ambientes acadêmicos, como nas aulas de Estatística em diferentes países do

mundo, predominam largamente os adeptos dessa escola. Também a literatura disponível, tanto os livros-texto como os artigos de publicação especializada privilegiam esta abordagem: uma visita a livrarias científicas ou a bibliotecas acadêmicas, tanto no Brasil quanto no exterior, mostra esta realidade. Este *status quo* estabelece um círculo vicioso.

A Estatística não é um ramo acabado do conhecimento e embora grande parte do seu progresso tenha sido no campo tecnológico, a preocupação epistemológica tem ocupado cada vez mais a comunidade acadêmica, em especial em virtude do desenvolvimento da escola Bayesiana. Um número ainda pequeno de centros com ênfase estritamente Bayesiana está conseguindo chamar a atenção para a área, inclusive com simpósios específicos, situação que pode ser encarada como uma tentativa de mudança de paradigma. De acordo com Lindley (1990), a Estatística Bayesiana não é uma ramo da Estatística e sim uma nova maneira de olhar a Estatística:

“ Ele (o paradigma Bayesiano) é um novo paradigma e a substituição de um (escola clássica) pelo outro requer uma verdadeira revolução científica, no sentido dado por Kuhn”.

Como adepto da escola clássica, Lehman, um dos mais famosos teóricos da Estatística clássica no século XX, apresenta seus comentários no próprio artigo de Lindley acima citado, dizendo entre outras coisas que

“ Nós podemos desfrutar da beleza da utopia de Lindley desde que nós não a levemos muito em consideração”.

Sob o ponto de vista de pesquisa acadêmica, apesar da maioria dos trabalhos apresentados em congressos de Estatística privilegiar a abordagem clássica (com exceção evidente das Escolas Bayesianas, que são reuniões específicas, já comentadas acima), vemos emergir, pouco a pouco grupos que aderem parcial ou totalmente ao novo paradigma que é a inferência Bayesiana. As facilidades computacionais impulsionaram o desenvolvimento destas técnicas, tributárias quase sempre de algoritmos iterativos.

Apesar disso, ainda assim há uma notável predominância da Estatística clássica, principalmente quando essa disciplina é oferecida como o chamado *curso de serviço*, oferecido às áreas de aplicação.

Não é de se estranhar o comentário de Blackwell (*in* DeGroot, 1986), a respeito de um eventual crescimento da área Bayesiana, ele que é um de seus renomados adeptos:

“ se isto acontecer será menos pela influência dos estatísticos e mais pela influência de outras áreas que se sentem mais à vontade com as premissas Bayesianas do que com as clássicas ”.

Um novo paradigma deve ser capaz de solucionar problemas não resolvidos (ou mal resolvidos) pelo paradigma anterior, bem como de preservar a já adquirida capacidade de resolver problemas. Apesar de Kuhn (1975) afirmar que dois paradigmas diferentes não podem conviver numa mesma comunidade, o que temos assistido é que cada vez mais os departamentos de Estatística têm abrigado simultaneamente defensores das duas escolas, mesmo que sob tensão em alguns casos, e ainda que com superioridade numérica para os adeptos da escola clássica, na maioria das vezes. É possível encontrar pesquisadores que transitam por ambas as áreas, menos num esforço filosófico de aproximação entre elas e mais numa possibilidade de comunicar seus resultados de pesquisa usando técnicas diferentes para efeitos de publicação. Não há em geral uma preocupação epistemológica explícita, pois segundo Chalmers (1982) o cientista é treinado para ser eficiente e nem sempre está consciente da natureza do paradigma no qual trabalha. Lacey (1998) não vê pela qual paradigmas incompatíveis e igualmente fecundos não possam coexistir, ainda que sob tensão, em diferentes cenários internacionais.

3.5 Conclusão

É nítida a influência do positivismo lógico na estrutura dos métodos desenvolvidos pela Estatística clássica, onde uma argumentação científica somente seria considerada válida dentro de um contexto objetivo, uma vez que a para seus adeptos a subjetividade é irracional. Por sua vez, a Estatística Bayesiana foi desenvolvida numa perspectiva de incorporação de subjetivismo, onde intuições e opiniões podem e devem ser incorporadas à análise. Segundo essa visão, a objetividade pode ser injusta, por ser impessoal.

O paradigma atual é, sem dúvida, a Estatística clássica e apesar de muitas tentativas terem sido feitas para pelo menos deixar mais visível para a comunidade a existência de métodos alternativos, podemos considerar que fatores como a inércia dos pesquisadores, o temor do uso da subjetividade de modo incoerente, e a já comentada influência do positivismo lógico, ainda presente em muitos ambientes de pesquisa, não favorecem essa possibilidade.

Apesar desses problemas, não se pode perder de vista que o nosso propósito é discutir o conteúdo da disciplina de probabilidade e Estatística oferecida ao aluno de ensino superior, principalmente no que diz respeito ao aspecto inferencial, bem como as dificuldades dos alunos nesse aprendizado. Mesmo que ainda de maneira incipiente, alguns segmentos da própria comunidade acadêmica estão progredindo numa direção de coexistência entre as duas escolas, e não seria pedagógico nem mesmo honesto ocultar dos alunos este fato, mesmo que sejam de cursos básicos. A solução seria introduzir no programa da disciplina uma discussão sobre paradigmas, mostrando os pontos conflitantes, dando menos valor à técnica pela técnica. Apresentar um material contextualizado e adequado ao desenvolvimento cognitivo dos alunos parece uma postura adequada.

Capítulo 4 A lógica da inferência

4.1 Introdução

Como amplamente discutido nos capítulos anteriores, a Estatística é utilizada como ferramenta em diversas áreas das ciências factuais, analisando dados provenientes tanto de experimentos controlados (geralmente nas ciências naturais) quanto de estudos observacionais (geralmente nas ciências sociais). Essas observações empíricas são usadas para embasar as teorias científicas, submetidas a avaliação através de hipóteses. Mesmo estando cientes de que as teorias não podem ser comprovadas de modo definitivo, este embasamento se fortalece à medida em que encontremos evidências para passar de uma situação particular para uma situação geral, via os chamados *procedimentos inferenciais*. O que se espera é que os métodos forneçam subsídios para que se faça um julgamento da população a partir de uma amostra. Há várias soluções propostas na literatura. Apesar de serem diferentes sob os pontos de vista epistemológico e metodológico, buscam encontrar a melhor estratégia para dar respostas aos cientistas face à incerteza.

Numa disciplina introdutória, a inferência Estatística é oferecida aos alunos como um processo de passagem do particular (amostra) para o geral (população). Aos alunos é oferecida uma inferência *instrumental*, geralmente sem aprofundamentos de natureza epistemológica, sem nenhum questionamento sobre: o que é inferência, que tipo de inferência podemos ou queremos fazer, o que significa sair do particular para o geral, como a filosofia encara o problema, se o problema é de natureza lógica ou cognitiva, se há mais de uma maneira de resolver o problema, etc. Enfim, essas idéias não são discutidas, deixando muitas vezes a impressão (para os alunos) de que tal passagem é muito simples, até mesmo óbvia. Sabemos que isto não é verdadeiro, pois há muita sofisticação conceitual envolvida, podendo mesmo ser que nessa apresentação *simplista* resida a dificuldade e/ou resistência dos alunos no aprendizado dessa disciplina.

Aliás, muitos dos alunos, e também boa parte da população, acreditam que as teorias científicas *confirmadas* não são mais passíveis de questionamento – a Estatística é muitas vezes, equivocadamente, invocada como responsável por essa *confirmação*, quando o natural seria encarar um resultado positivo como indicativo e não como conclusivo. Barnett (1973), um estatístico preocupado com a inferência comparada, insistia na exposição do aluno à variedade de atitudes filosóficas e conceituais que proliferam na teoria e prática da Estatística. Em nossa opinião, concordano com Barnett, esta discussão deve ser apresentada já numa disciplina introdutória, mesmo que para isso algumas das técnicas clássicas, habitualmente ensinadas, sejam suprimidas (por exemplo, *todas* as maneiras de testar as diferenças entre duas médias – variâncias iguais, diferentes, conhecidas, etc.).

Neste capítulo vamos apresentar algumas idéias sobre *inferência* de modo geral e *inferência estatística* de modo particular, bem como uma descrição dos raciocínios condicionais no âmbito da lógica.

4.2 Inferência

"Inferência é um processo pelo qual se chega a uma proposição, afirmada na base de uma ou mais proposições aceitas como ponto de partida do processo." (Copi, 1968)

No processo de inferência é possível destacar dois tipos de raciocínio: dedutivo e indutivo. O primeiro deles, o dedutivo, é chamado de raciocínio lógico por excelência, pois a partir de premissas verdadeiras chega-se a uma conclusão verdadeira. Neste caso as premissas contêm a conclusão. O segundo, o raciocínio indutivo, fornece uma conclusão mais geral do que as contidas nas premissas, conclusão essa que pode ser falsa mesmo que as premissas sejam verdadeiras. Então a inferência dedutiva é sempre conclusiva, enquanto que a inferência indutiva é sempre inconclusiva. Estamos aqui tratando da chamada indução incompleta, onde não estão incluídos os casos triviais de

simples enumeração (indução simples) e nem o caso de indução matemática, a qual, na verdade, é um raciocínio rigorosamente demonstrativo.

Há muita controvérsia sob o ponto de vista epistemológico acerca da indução. A procura de enunciados que possam ser deduzidos de acordo com as regras da lógica formal é um problema muito discutido em filosofia, desde a Antigüidade, e interessa aos filósofos, aos cientistas e ao ser humano em geral em busca de conhecimento. Mesmo a utilização dos termos indução e inferência pode ser diversa, de acordo com a época, o autor, e ainda com a escola filosófica em questão.

Muitos filósofos afirmam que somente a inferência dedutiva é válida pois é passível de uma justificação lógica. Hume (século XVIII) parece ter sido o primeiro filósofo a tratar de modo sistemático do problema da indução, apesar de seu interesse estar centrado nos casos de inferência causal, sobre o qual ele discorria sem fazer uso ou mesmo referência ao emprego técnico de probabilidades. Ele argumentava que as conclusões tiradas a partir da indução não deviam necessariamente ser refutadas, mas careciam de justificativa lógica ou racional, embora pudessem ser úteis para a sobrevivência da Humanidade. Laplace incluía a indução como uma das maneiras de se chegar à verdade, embora deixasse claro que ela não era suficiente para se estabelecer conclusões com rigor.

Peirce, matemático e filósofo do século XIX, antecipou muitos dos pontos que posteriormente seriam defendidos pelos adeptos da teoria de inferência clássica. Com relação à indução, ele comentou:

“A indução é a verificação experimental de uma teoria. Sua justificativa está em que, embora a conclusão da investigação em um estágio qualquer possa ser mais ou menos longe da verdade, mesmo assim a aplicação ulterior do mesmo método deve mitigar o erro.....Trata-se de um simples cálculo de probabilidade objetiva....Os lógicos muitas vezes afirmam que a conclusão indutiva se aproxima da verdade devido à uniformidade da natureza”.

Segundo Paulos(1998), Peirce apresentou uma justificativa pragmática da indução, dizendo que

"Pode ser que indução não funcione, porém se algo tiver que funcionar, será a própria indução."

Popper (1975), por sua vez, rejeitava a indução e propunha que experiências deviam ser levadas a efeito para submeter a teoria a testes rigorosos, que pudessem refutá-la: é possível refutar hipóteses mas jamais prová-las, pois a coerência nem sempre fortalece a verdade mas a incoerência, sem dúvida, estabelece a falsidade. Tanto mais refutável uma teoria for, maior será o seu conteúdo afirmativo. A proposta de Popper constituiu-se no chamado método hipotético-dedutivo: a partir de uma hipótese inicial, raciocina-se que se ela for verdadeira algumas situações devem acontecer; planeja-se então um experimento destinado a refutar tal hipótese; se os resultados forem incompatíveis com as situações esperadas, a hipótese é rejeitada, caso contrário, a hipótese é corroborada (mas não provada).

Com ou sem a influência de Popper, o método que foi proposto por Fisher é do tipo hipotético-dedutivo, embora ele se considerasse no caminho da inferência indutiva. A postura de Fisher deve ter contribuído para que Da Costa (1993) sugerisse que o método hipotético-dedutivo fosse incluído na classificação de inferência indutiva, dizendo que todas as inferências indutivas se referem a este método.

Tanto Chalmers (1982) como Beveridge (1981) argumentam que não é possível justificar logicamente o problema da indução, face às dificuldades, reconhecidas pelos filósofos, de toda a tentativa de justificação deste tipo de raciocínio, o qual se baseia ou na analogia ou na generalização. No entanto é um fato que todas as informações das ciências factuais vêm do passado e que tentamos, a partir delas e de alguma forma, prever o futuro. Uma ajuda parcial, nesse sentido, viria com o *Princípio da Uniformidade da Natureza*, que apesar de ser responsável por parte de nosso comportamento cotidiano,

também é controverso: “o futuro será como o passado” ou “efeitos semelhantes devem ter causas semelhantes”. Segundo Hume, não há nenhuma justificativa para essa suposição, pois para ele não há garantia da veracidade das proposições enunciadas pela ciências experimentais.

Este é um problema que interessa tanto a filósofos como a estatísticos e probabilistas. Segundo Boudot (1972), como a filosofia clássica não forneceu meios para resolver o problema da indução, houve uma caminhada em direção ao cálculo de probabilidades, sem afirmar, no entanto que isto possa se constituir numa lógica da indução:

“ Enquanto os lógicos se esforçavam em justificar a passagem das conseqüências aos princípios, chamada de indução, para uma certeza aos princípios, os homens da ciência se contentavam em determinar as hipóteses mais plausíveis. ”

Lambert e Brittan (1972) distinguem os conceitos de *plausabilidade* e generalização Estatística, caracterizando o primeiro como sendo probabilidade indutiva. Por *plausabilidade* entendem uma relação entre enunciados: medida do grau de apoio que certos enunciados conferem a outros. Acreditam que a *generalização estatística*, por sua vez, pode levar a um argumento do tipo indutivo quando há interesse em explicar acontecimentos particulares; mas que se o interesse for explicar acontecimentos de massa, a generalização Estatística está mais para deduzir conclusões gerais do que para induzir acontecimentos particulares.

Para Lindley (1990), a lógica formal nos capacita a deduzir da hipótese H as conseqüências x e y, mas não nos ajuda a provar H ou deduzir que y pode se seguir a x.

“ Inferência científica é essencialmente a utilização de dados observados no passado para prever dados futuros ainda não observados. ”

.....

O problema fundamental de inferência Estatística é poder, a partir de um conjunto x de observações, expressar opinião sobre um outro conjunto y , ainda não observado". Lindley (1990).

A busca pela possibilidade de se extrair mais conhecimento acerca de uma população pela evidência de resultados verificados empiricamente não ficou somente no campo da filosofia, e o próximo tópico vai analisar a inferência sob o ponto de vista da Estatística.

4.3 Inferência Estatística

A justificativa para a abordagem inferencial na disciplina de Estatística é a impossibilidade que temos (quase sempre) de trabalhar com a população de interesse, por diversas razões, que incluem: dificuldade de acesso físico a toda a população (caso de doenças raras), dificuldade de acesso a toda a população sob o ponto de vista econômico (caso de pesquisa eleitoral), além de várias outras. O censo brasileiro de 2000 se propôs a trabalhar com toda a população do país em algumas questões básicas, como por exemplo sexo, cor, idade e fez uma amostragem sistemática de 10% (ou 25% em alguns casos) da população para outros quesitos. Na grande maioria das vezes a pesquisa não é feita do tipo censitário, e o usual é recorrer a amostras da população de interesse: pretende-se que seja possível tirar conclusões sobre a população a partir de uma análise da amostra. Isto significa que temos que preparar nossos alunos para uma passagem da amostra para a população, ou, em outras palavras, do *particular para o geral*.

Usualmente o termo **inferência científica** se converte em **inferência estatística** quando a relação entre o estado da natureza desconhecido por nós e aquilo que observamos se traduz em termos probabilísticos, ou seja, a inferência estatística pressupõe uma formulação probabilística para a passagem da amostra para a população. Disso decorre também o termo muitas vezes utilizado de *inferência indutiva*, que permitiria ir além das constatações experimentais. Isto parece razoável, quando pensamos

na característica inconclusiva oferecida pela indução: considerar as teorias mais prováveis ou menos prováveis, em função dos resultados empíricos obtidos em experimentos ou em estudos observacionais.

Somente no século XX é que teve início o estudo formal da inferência Estatística. Em muitos textos encontra-se a afirmação de que ela está ligada ao raciocínio indutivo, mas muitos dos adeptos da teoria Bayesiana reivindicam esta prerrogativa para o seu método de inferência, alegando que o método clássico não utiliza um raciocínio indutivo. Good (1989), um brilhante estatístico da linha Bayesiana, entende por indução científica a mudança da probabilidade de uma hipótese, à luz da evidência, e completa que o problema é *parcialmente* resolvido por meio do teorema de Bayes.

Na inferência Estatística Bayesiana é feita uma generalização do raciocínio indutivo presente no teorema de Bayes: neste ponto raciocínio indutivo está sendo usado no sentido dado por Anderson (1995):

“ Termo usado para descrever processos através dos quais se chega a conclusões prováveis ao invés de certas”.

No teorema está presente uma quantidade que, embora seja uma afirmação probabilística (priori), supõe-se de valor conhecido – é o que podemos chamar de inferência probabilística (ou como raciocínio indutivo de Anderson). Na inferência Estatística Bayesiana esse valor, como não é conhecido, é atribuído pelo analista, daí o elemento subjetivo, causa de grande parte da controvérsia. Como as evidências podem ser usadas para revisar as probabilidades *a priori*, o método é encarado por muitos como sendo do tipo indutivo (ou quase, segundo Good). Por sua vez, os estatísticos clássicos (Fisher, Neyman e Pearson) desdenham a subjetividade do método Bayesiano e acreditam ter encontrado uma metodologia objetiva para a inferência Estatística.

Fisher acreditava ser possível obter inferências válidas a partir dos resultados de experimentos. Para ele, inferência indutiva era o único processo sob o qual os novos

conhecimentos poderiam aparecer. Ele admitia um certo grau de incerteza para as conclusões advindas da indução, mas achava que esse grau de incerteza poderia ser explicado através de uma expressão rigorosa (a justificativa estaria no cálculo das probabilidades). Isto significa que Fisher acreditava ser possível fazer uma inferência rigorosa mesmo a partir da incerteza. Aparentemente influenciado pelas idéias de Popper, Fisher propôs um método inferencial – teste de significância – que seria usado para *falsear* uma hipótese (chamada por ele de hipótese nula), usando o procedimento da lógica condicional, conhecido como raciocínio condicional *modus tollens* (argumento por contradição).

Neyman não advogava a indução para o seu método inferencial e, mais cauteloso, preferia chamar a sua técnica de *conduta indutiva*, um processo através do qual o pesquisador decide por uma ação a tomar. Por outro lado, o método desenvolvido por Neyman e Pearson guarda semelhanças com o método de Fisher, como será discutido posteriormente.

Antes de retomar a discussão sobre as escolas de inferência e sua controvérsia vamos apresentar as idéias básicas do raciocínio condicional, do ponto de vista da lógica formal.

4.4 Raciocínios condicionais

O raciocínio condicional na lógica clássica se apresenta sob duas formas:

1 – *modus ponens* –

- se acontecer P então acontecerá Q (premissa condicional)
- P aconteceu (premissa categórica)

∴ Q também aconteceu

Notação: Se P então Q P ∴ Q

2 – *modus tollens* –

■ se acontecer P então acontecerá Q (premissa condicional)

■ Q não aconteceu (premissa categórica)

∴ P também não aconteceu

Notação: Se P então Q $\sim Q$ ∴ $\sim P$

onde P e Q são, por convenção, denominados, respectivamente, de antecedente e conseqüente. No primeiro caso, a premissa categórica afirma a proposição antecedente da premissa condicional (P) e a conclusão afirma a conseqüente (Q). Diz-se que este argumento está no modo afirmativo ou *modus ponens* (*ponere* – forma latina de afirmar). No segundo caso, a premissa categórica nega a proposição conseqüente da premissa condicional ($\sim Q$) e a conclusão nega a sua antecedente ($\sim P$). Diz-se que este segundo argumento está no modo negativo ou *modus tollens* (*tollere* é a forma latina de negar). Essas são as duas formas consideradas válidas dentro da lógica formal, sendo que as demais possibilidades são consideradas falácias associadas ao raciocínio lógico, algumas delas acontecendo às vezes e outras sendo simplesmente impossíveis. Como exemplo, se afirmarmos o conseqüente (Q) é possível tanto P acontecer (P) como não acontecer ($\sim P$). Se colocarmos ambos os raciocínios condicionais válidos sob a forma de probabilidades, teríamos: para *modus ponens* $P(Q | P) = 1$ e para *modus tollens*, $P(\sim P | \sim Q) = 1$.

As duas formas válidas de raciocínio condicional discutidas acima são consideradas dedutivas, pois não há elemento de incerteza nas conclusões (as probabilidades descritas são associadas ao que se chama de *evento certo*!). Vamos voltar aos raciocínios ligados às escolas de inferência Estatística.

4.5 As inferências de Fisher, Neyman - Pearson e Bayesiana

Voltando a Fisher, poderíamos resumir o uso feito por ele do raciocínio *modus tollens*, para os testes de significância, da seguinte maneira:

Se a hipótese H for verdadeira (P) então um comportamento amostral de certo tipo é pouco provável (Q). A evidência amostral não se comportou da forma esperada (não Q) então a hipótese H deve ser refutada (não P).

A aplicação do raciocínio não permite dizer que a hipótese é pouco provável pois o procedimento não permite estabelecer probabilidades para hipóteses. Exemplificando o raciocínio Fisheriano: se uma nova vacina contra a AIDS tiver de fato poder de inoculação igual a 40% (P) então a proporção amostral de inoculados fora de certo intervalo é pouco provável (Q); a proporção (observada) de inoculados não pertence ao determinado intervalo (não Q) e portanto rejeita-se que a nova vacina tenha poder de inoculação igual a 40% (não P).

A conclusão acima é verdadeira? É válido extrapolar para a população de interesse uma conclusão obtida empiricamente através de uma amostra? Na verdade, levando em conta que as implicações são mais brandas do que na lógica formal, no sentido da plausibilidade dos resultados, Fisher, para dar credibilidade a essa passagem, associa ao resultado uma probabilidade, que será relacionada ao nível de significância do teste, ao qual voltaremos no próximo capítulo, onde também será discutido em que momento está associada a atribuição desta probabilidade.

Fisher ainda propôs, em seu texto, vários esquemas de experimentação, baseados em estruturas lógicas, que produziriam caminhos para a obtenção de inferências *completamente* válidas. Com isso ele acreditava ter superado a dificuldade teórica associada a inferências indutivas. Suas idéias causaram grande impacto em várias ciências experimentais, notadamente na psicologia, que passou do estudo de caso para o estudo de grupos de pessoas.

Assim como Fisher, também Neyman e Pearson estabeleceram suas ferramentas inferenciais sem fazer uso das probabilidades *a priori*. Eles evitavam o termo indução, dizendo que o que existia não era raciocínio indutivo e sim comportamento indutivo (o que os levou a enfocar o teste de hipóteses como um procedimento de decisão,

associando certos riscos ao processo). Em sua teoria, além da hipótese nula proposta por Fisher, existe a hipótese alternativa (se a hipótese nula é rejeitada, diz-se que há indícios a favor da hipótese alternativa). Neste procedimento também é incluído um novo erro (além daquele associado ao nível de significância de Fisher), que é o chamado erro do tipo II, que consiste em não rejeitar uma hipótese dado que ela é falsa, cujo cálculo se baseia em valores do parâmetro sob a hipótese alternativa (no capítulo 6 voltaremos ao tema). Vemos que é possível dizer que a teoria de Neyman – Pearson está fortemente ligada à de Fisher (hipótese nula, nível de significância, decisão de rejeitar ou não a hipótese testada).

Sob o ponto de vista do ensino, a teoria dos testes de hipóteses de Neyman-Pearson, a partir dos anos 50, juntou-se à teoria de Fisher, produzindo, na opinião de Gigerenzer (1992) uma lógica híbrida da inferência Estatística, que se constitui no que estamos aqui denominando de teoria clássica. Esta abordagem causou um grande impacto nos programas acadêmicos, livros-texto, bem como na prática experimental, pois o que se via (e ainda se vê) é a aplicação mecânica de certas técnicas inferenciais, como único suporte para tomada de decisão de aceitar ou rejeitar teorias, uma receita para produzir conhecimento. Em muitos casos, esse procedimento era tido como o próprio *método científico*!

Com relação às idéias de Bayes, cujo método das probabilidades inversas foi durante muito tempo o único candidato a resolver o problema da indução, houve um renascimento em meados do século XX, que não foi suficiente para destronar a influência da escola clássica e um dos empecilhos foi certamente a dificuldade computacional. Interessante notar que no campo da psicologia, que assistiu a uma revolução cognitiva nos anos 60, os pesquisadores nunca usavam as técnicas Bayesianas, muito criticadas, para análise de seus experimentos, mas supunham que muitos indivíduos usam o raciocínio Bayesiano para pensar racionalmente. Hoje em dia, segundo Lecoutre (2000), a APA (Associação de Psicologia Americana) sugere abrir aos pesquisadores a possibilidade de apresentar seus resultados para publicação usando técnicas alternativas às análises clássicas, como por exemplo os métodos Bayesianos.

O raciocínio usado nesses problemas, que envolve o uso do teorema de Bayes, é puramente objetivo, se conhecermos de modo explícito o valor da probabilidade *a priori*. Apesar de ser um tópico com o qual os alunos sentem dificuldade, pelos motivos relacionados à confusão com probabilidades condicionais, é passível de ser trabalhado para a introdução da filosofia Bayesiana, uma vez que os elementos necessários para a compreensão fazem parte do programa básico da disciplina, e envolvem somente o cálculo básico de probabilidades. O uso deste procedimento probabilístico não é o que causa controvérsia na literatura sobre procedimentos inferenciais. Como nem sempre a *priori* pode ser colocada de modo inequívoco, o uso da técnica para o raciocínio estatístico é que é discutível uma vez que contém um elemento subjetivo: a distribuição *a priori* de probabilidades, a qual deve ser estabelecida antes do experimento ser realizado.

Nos raciocínios condicionais já comentados, considerava-se a afirmação do antecedente dada a afirmação do conseqüente como uma falácia, uma vez que isto não se verifica com probabilidade um (evento certo). No entanto é possível usar este silogismo do ponto de vista probabilístico, como um argumento indutivo, através do teorema de Bayes:

Se a hipótese H for verdadeira (P) então a evidência amostral tem que se comportar de uma certa forma (Q). A evidência amostral se comportou da forma esperada (Q), então a hipótese H é verdadeira (P).

O procedimento Bayesiano inclui de início a distribuição *a priori* para o parâmetro envolvido (ou para a hipótese H), em seguida a evidência amostral, que se reflete na construção de uma função (*verossimilhança*, calculada com base no parâmetro fixado). Depois, a construção da distribuição *a posteriori* (distribuição condicional do parâmetro de interesse dada a evidência amostral – ou da hipótese H dada a evidência) sobre a qual as conclusões serão tomadas. Na verdade este procedimento permite que se revise a probabilidade inicial da hipótese H (com base nos dados) e o que se tem como resultado é uma probabilidade novamente associada à hipótese (mesmo condicionada).

No caso de Fisher (ou de Neyman-Pearson) não é obtida uma probabilidade para a hipótese de interesse. Fisher, embora achando que Bayes era merecedor de um lugar de destaque na história da ciência, foi muito crítico ao dizer que, em aplicações práticas, a teoria Bayesiana não tinha utilidade (Efron, 1998).

4.6 Conclusão

Esta discussão, apesar de incompleta, procurou mostrar diferentes vertentes de inferência que poderiam ser discutidas com os alunos, para mostrar que não há consenso na área. Qualquer que seja o método inferencial proposto, será que estamos passando do particular para o geral? Este seria o caminho natural a percorrer, uma vez que o pesquisador, tendo a amostra em mãos, deseja uma técnica que o faça *passar* para a população ou seja, do particular para o geral. O método clássico é construtivo, baseado num referencial que é o espaço amostral, e é hipotético-dedutivo, indo do geral (população) para o particular (amostra) e devolve a conclusão para a população: daí a cautela de Neyman em chamar o processo de conduta indutiva ao invés de método de indução. Segundo Rivadulla (1991), a abordagem de Neyman-Pearson não é, e nem pretende ser, uma teoria da qual aprendemos de um modo indutivo-probabilístico a partir da experiência. Por outro lado, a escola Bayesiana que para muitos de seus adeptos é considerada a única indutiva, ainda continua fazendo discussões epistemológicas a esse respeito.

Podemos fazer diferentes generalizações a partir de um conjunto de fatos: novamente o fim do caminho é um julgamento subjetivo. De qualquer forma toda a lógica do raciocínio estatístico depende da teoria subjacente de probabilidade defendida pelo pesquisador bem como da escola inferencial advogada por ele. Tanto uma como outra estão sujeitas a controvérsias, o que pode causar dificuldades para a obtenção de conclusões. Conhecer o estado da arte da área, certamente ajudará tanto novatos quanto mais experientes a ficarem mais atentos e precavidos.

Segundo Howson (1993), os estatísticos Fisher, Neyman e Pearson, inspirados na obra de Popper, tinham a ambição de criar um padrão de mérito científico objetivo, convincente e não probabilístico. Suas teorias conseguiram se firmar na área e os procedimentos que eles recomendavam para planejamento de experimentos e análise de dados tornaram-se padrões de exatidão para muitos cientistas. Ainda segundo Howson, não é merecida a grande influência exercida pela Estatística clássica, pois seus métodos não contêm a propalada objetividade e até mesmo costumam violá-la. Na verdade é impossível não usar um pouco de subjetivismo na aplicação dos métodos estatísticos, quer sejam do tipo clássico quer do Bayesiano.

Segundo Box (1983), o ser humano precisa ser capaz de *deduzir* quais os fatos que poderiam ser esperados como realizações de um modelo particular e, mais difícil ainda, *induzir* quais modelos estão de acordo com fatos particulares. A nosso ver a inferência Estatística, apesar de mais restritiva do que a inferência científica, pode desempenhar um papel importante nessas tarefas, uma vez que auxilia a tirar conclusões sobre o não observado, a partir de dados observados. Isto evidentemente não elimina a controvérsia no seio da própria inferência Estatística, retratada por essa frase bastante pessimista (Oakes, 1986),

"Quanto mais se pesquisa na área de fundamentos em inferência, mais parece um poço sem fundo."

transmitindo a idéia de que nenhuma das abordagens para a inferência Estatística é inteiramente satisfatória.

Capítulo 5 Alguns elementos para a disciplina introdutória

5.1 Introdução

Sempre justificamos aos nossos alunos que a inclusão do tópico de probabilidade na ementa de uma disciplina básica se deve à necessidade, principalmente, da quantificação do risco que está associado a qualquer tomada de decisão (aqui o termo decisão foi usado de modo informal, não ligado necessariamente à Teoria da Decisão). O raciocínio probabilístico é de importância vital em processos inferenciais, tanto clássicos quanto Bayesianos; entretanto essa é uma área em que os alunos geralmente não se sentem à vontade e em que os professores convivem com o “fracasso” nas soluções de problemas. Algumas das dificuldades dos alunos serão analisadas do ponto de vista *cognitivo* em algumas seções deste e dos próximos capítulos.

Em uma disciplina que normalmente se intitula Introdução à Probabilidade e Estatística o mínimo que se espera é uma definição dos termos que aparecem em seu título. Como já dissemos no início do trabalho, definir o que é probabilidade e discutir como ela é conectada com a estatística tem levado muitos pesquisadores a discussões intermináveis.

5.1.1 Probabilidade e suas definições

Segundo Hartigan (1983), existem três enfoques para a teoria de probabilidades: a) teoria lógica (Laplace, Keynes, Jeffreys); b) teoria empírica (Von Mises, Kolmogorov) e c) teoria subjetiva (De Finetti, Good).

a) Teoria Lógica

- Laplace (1814): **Probabilidade** de um evento é a razão entre o número de casos favoráveis e o número de casos (igualmente) possíveis. Para que esta teoria se aplique é necessário que se tenha eventos equiprováveis, cuja definição depende da própria definição de probabilidade. Portanto estamos diante de uma tautologia.
- Keynes (1921): **Probabilidade** é uma opinião racional sobre uma proposição baseada no conhecimento de outra proposição. Dificilmente é possível estabelecer um valor numérico para cada opinião individual, mas é possível comparar pares de opiniões. No entanto Keynes não oferece regras para o cômputo das probabilidades, nem para as comparações entre elas.
- Jeffreys (1939): Mesmo conceito de Keynes, porém apresentando uma maneira de construir **probabilidades “a priori”** para diferentes problemas estatísticos.

b) **Teoria Empírica**

- Von Mises (1928): **Probabilidade** de um evento é tomada como o limite da frequência relativa, considerando repetições infinitas do experimento, sob as mesmas condições (daí o nome frequentista que esta teoria também possui). No entanto, observação empírica é finita e esta teoria seria aplicável àquelas sequências finitas cujo comportamento se aproximasse da idealização da sequência infinita. Para Feller (1968) a probabilidade não se refere a opiniões mas sim aos resultados possíveis de um experimento conceitual.

“É necessário conservar em mente a interpretação intuitiva do conceito de probabilidade...Imaginando-se o experimento repetido um grande número de vezes, espera-se que um evento, cuja probabilidade seja 0,6, ocorra aproximadamente 60 vezes em 100 tentativas”.

- Kolmogorov (1933): formalizou **probabilidade** como uma medida, interpretando probabilidade.

c) **Teoria Subjetiva** - (probabilidades são sempre associadas a indivíduos)

- De Finetti (1930): **Probabilidade** atribuída a um evento por um indivíduo é revelada através do quanto ele está disposto a apostar naquele evento. Várias objeções são discutidas na literatura, desde aquelas que tratam de quão prática seria esta definição até aquelas que discutem se ela seria útil. Segundo Lindley (1985)

“...probabilidade não tem existência fora do indivíduo, como o comprimento de uma mesa. Probabilidade exprime a relação entre um indivíduo e o mundo que ele está contemplando”.

- Good (1950) considera o grau de opinião (**probabilidade**) de um indivíduo acerca de uma proposição como uma noção primitiva, sem definição precisa.

Dizendo que os fundamentos são as partes mais controvertidas de muitas ciências, se não de todas, Savage (1972) procurou distinguir três classes de interpretação de probabilidade, que guardam alguma semelhança com algumas já descritas neste capítulo:

- **objetivista** – a evidência da concordância entre o comportamento de um evento repetitivo e o conceito matemático de repetições independentes e com mesma probabilidade será obtida tão somente pela observação de algumas repetições do evento, e não de outras fontes, quaisquer que elas sejam;
- **personalista** – a probabilidade mede a confiança que determinado indivíduo tem na verdade de uma proposição;

- **necessária** – a probabilidade mede o quanto um conjunto de proposições (fora da opinião pessoal) confirma a verdade de outro, no sentido da lógica formal.

À parte dessas teorias, o aluno em particular ou mesmo o cidadão em geral interessado em definir de um modo mais formal a palavra probabilidade, poderia consultar um dicionário de uso corrente (em português e em inglês). Como encontraria o verbete **probabilidade**? Seria possível encontrar várias definições, a maioria delas ligadas ao senso comum e a situações do cotidiano. Para os nossos propósitos, escolhemos alguns exemplos, onde a ligação com a área de *matemática* era explícita. Começamos pelos de língua portuguesa:

“Probabilidade. S. f. ... 3. Mat. Número positivo e menor que a unidade, que se associa a um evento aleatório, e que se mede pela frequência relativa da sua ocorrência, numa longa sucessão de eventos”. (Ferreira, 1988 – Aurélio).

“Probabilidade. S. f. ...4. Mat. Razão entre o número de casos favoráveis pelo número de casos possíveis para um evento qualquer, onde os casos possíveis têm, supostamente, o mesmo grau de possibilidade”. (Larousse, 1998).

Na língua inglesa lê-se, por exemplo,

“ Probability, n.math. The ratio of the number of chances by which an event may happen, to the number by which it may both happen and fail”. (Thatcher, 1970 – Dicionário Webster).¹

“ Probability, n. 4. The mathematically calculated chance that something will happen”. (Longman, 1995).²

¹ razão entre o número de chances em que um evento pode ocorrer e o número de vezes em que o mesmo evento pode ou não ocorrer.

² a chance calculada matematicamente de que alguma coisa acontecerá.

Como se vê a orientação no dicionário Aurélio vai na direção frequentista de Von Mises e na enciclopédia Larousse a abordagem é baseada na definição clássica de Laplace. Nos exemplos de língua inglesa, o dicionário Webster também utiliza a definição de Laplace e o dicionário Longman apresenta uma visão mais ligada ao cálculo matemático, quase axiomática.

Então, definir probabilidade, que é apresentada aos alunos e consta nos textos básicos como algo indiscutível, não é uma tarefa tão fácil como se dá a entender numa disciplina introdutória. É importante que professor e aluno discutam os conceitos, e que estabeleçam as bases da disciplina levando em conta a sua diversidade.

5.1.2 Estatística e suas definições

A mesma motivação da seção anterior nos leva a apresentar algumas definições associadas ao termo **estatística**. Barnett (1973), que como já dissemos esteve sempre interessado no estudo da inferência comparada, apresenta diferentes definições as quais são encontradas nos livros-texto, de acordo com o propósito a que foram escritos: desde coleta e apresentação de dados (estatística descritiva) até métodos para análise e interpretação de dados (estatística inferencial). Como um exemplo do primeiro uso, ele utiliza a seguinte definição:

“Estatística é uma ciência que se refere à experimentação bem como à coleta, descrição e análise de dados...Métodos estatísticos são ferramentas para análise de dados.”

Como exemplo de definição ligada ao segundo propósito, citou Kendall e Stuart (1969):

“ Estatística é o ramo do método científico que trata com observações obtidas por contagem ou mensuração de características de populações de fenômenos naturais ”.

No entanto é possível encontrar definições de Estatística que refletem a corrente filosófica que permeará o texto que lhe segue. Destacaremos duas delas, também apresentadas em Barnett, sendo a primeira devida a Hoel (1962):

“ Métodos estatísticos são aqueles essencialmente desenvolvidos para tratar com dados que foram obtidos através de uma operação repetitiva”;

e a segunda devida a Lindley (1965):

“ Estatística é o estudo de como os graus de opinião são alterados pelos dados”.

Barnett então resolve dar ele próprio uma definição que seja suficientemente geral para acomodar as diferenças filosóficas e conceituais existentes e ao mesmo tempo bastante específica para delinear os componentes básicos de *qualquer* teoria estatística:

“Estatística é o estudo de como informação deveria ser empregada para reflexão e ação em uma situação prática envolvendo incerteza”.

Uma interpretação personalista é oferecida por Savage (1972), onde decidimos traduzir *vagueness* por incerteza, por analogia à palavra usada acima por Barnett (o próprio Savage ao comentar que não discutiria este termo, já estaria indicando a dificuldade que poderíamos encontrar em sua interpretação e/ou tradução):

“Estatística pode, talvez, ser definida como a arte de tratar com a incerteza e com diferenças interpessoais em situações de decisão”.

Também para o verbete **estatística** fizemos uma comparação entre as formas apresentadas nos diferentes dicionários, com os mesmos utilizados para a probabilidade, com o seguinte resultado:

“ Estatística. S. f. 1. Parte da matemática em que se investigam os processos de obtenção, organização e análise de dados, sobre uma população..., e os métodos de tirar conclusões e fazer ilações ou predições com base nesses dados”. (Ferreira, 1988 – Aurélio) .

“ Estatística. S. f. 2. Conjunto dos métodos que têm por objetivos a coleta, o tratamento e a interpretação dos dados observados, relativos a um grupo de indivíduos ou de unidades”. (Larousse, 1998).

“ Statistics, n. A collection of facts which admit of numerical statement and of arrangement in tables, especially facts illustrating the physical, social, moral, intellectual, political, industrial, and economic condition of communities or classes of men”. (Thatcher, 1970 – Webster).³

“ Statistics, n. 1b) The science of dealing with and explaining a collection of numbers which represent facts or measurements”. (Longman, 1995).⁴

Nos parágrafos precedentes verifica-se como é necessária uma visão ampla das várias posições dos pesquisadores no assunto, inclusive para dar ao aluno a opção de discussão para construir sua própria visão. A maneira de trabalhar a informação obtida em situações de incerteza reflete a escola inferencial à qual a pessoa está ligada. Se o referencial se basear em prováveis repetições de determinadas situações experimentais e em raciocínio frequentista, por exemplo, estaremos nos encaminhando para a estrutura exigida pela escola clássica; por outro lado, se o pesquisador tiver um comportamento subjetivo quanto à probabilidade *a priori* de determinada hipótese, por exemplo, seu encaminhamento estará ligado à escola Bayesiana.

³ uma coleção de fatos que admitem uma afirmação numérica e uma organização em tabelas, especialmente fatos que ilustram condições físicas, sociais, morais, intelectuais, políticas, industriais e econômicas de comunidades ou classes de seres humanos.

⁴ a ciência de tratar e explicar uma coleção de números que representam fatos ou medidas.

O dicionário Webster apresenta uma definição que só contempla os aspectos descritivos da estatística, enquanto que os demais incluem uma parte relativa à interpretação ou explicação, que podem estar ligados a uma eventual atitude inferencial, sem no entanto discutir os processos de interpretação. Nas definições de Hoel e de Kendall e Stuart é possível perceber um desejo de encaminhamento clássico, enquanto que os demais, sentindo provavelmente a controvérsia envolvida, preferem, na definição, apresentar um pensamento que abrigue várias correntes (mesmo que sua prática seja melhor definida posteriormente).

5.2 Probabilidade e o teorema de Bayes

A teoria Bayesiana, como o próprio nome indica, é baseada no teorema de Bayes, o qual é demonstrado formalmente (em algumas situações de sala de aula), ou então é apresentado somente sob a forma de um problema mais elaborado, usando probabilidades. Pela maneira como é apresentado, é costume referir-se a ele como o teorema das probabilidades *inversas* ou o teorema das *causas*. O uso do resultado causa polêmica quando é realizado num contexto de inferência estatística. O seu uso, do ponto de vista probabilístico, não causa divergências e é este que vamos abordar aqui, através de um exemplo onde eventualmente poderá ser considerado algum elemento de natureza cognitiva.

Serão apresentados a seguir dois exemplos análogos, colocados em contextos diferentes, comumente apresentados aos alunos no capítulo de probabilidade, numa disciplina introdutória de estatística. O primeiro deles refere-se a um experimento de bolas em urnas (preferido pelos *professores*) e o segundo a um experimento ligado à área de saúde (preferido pelos *alunos*).

O exemplo a ser tratado em primeiro lugar corresponde a um experimento que consiste em retirar uma bola de uma urna escolhida ao acaso dentre duas urnas A e B, que apresentam a seguinte composição:

Urna A - 3 bolas verdes e 2 azuis

Urna B - 4 bolas verdes e 1 azul.

O experimento consiste de duas fases, sendo a primeira sortear uma urna (por exemplo através de um lance de uma moeda) e a segunda, retirar aleatoriamente uma bola da urna sorteada. Quatro perguntas são habitualmente feitas:

- 1) se a urna A foi a sorteada, qual a probabilidade de ser retirada uma bola azul?
- 2) se a urna B foi a sorteada, qual a probabilidade de ser retirada uma bola azul?
- 3) qual a probabilidade de ser retirada uma bola azul?
- 4) se for sorteada uma bola azul, qual a probabilidade de ter sido sorteada a urna A?

Na primeira e segunda perguntas, dada a causa (uma determinada urna foi sorteada), pergunta-se a probabilidade do efeito retirar aleatoriamente uma bola de determinada cor). Na quarta pergunta, dado o efeito, quer-se saber a causa. Na terceira situação, a pergunta não especifica nada sobre urnas e portanto todas as possibilidades têm que ser analisadas e aglutinadas.

Temos então três perguntas condicionais (1ª, 2ª e 4ª), mas as duas primeiras têm como resposta um caminho direto de cálculo enquanto que para a 4ª pergunta deveríamos propor a resposta, fazendo um caminho “inverso”, cujo cálculo é indireto. Para a terceira, como dissemos, temos que verificar todas as causas (sorteio das urnas) que levam ao efeito desejado (bola azul).

Supondo que a seleção inicial da urna seja feita através do lançamento de uma moeda não viciada, e levando em conta a composição das urnas estabelecida acima, o diagrama de árvore toma a seguinte forma:

Nesse diagrama, as probabilidades de cada caminho completo (ou seja, de cada possibilidade de obtenção de um resultado final do experimento) são encontradas através de regras básicas de probabilidade. Por exemplo: probabilidade de ser sorteada a urna A e de ter sido (em seguida) sorteada a bola azul é escrita como $P(\text{urna A e azul})$ e seu valor é obtido através do produto do primeiro ramo (probabilidade incondicional ou probabilidade a *priori* = $1/2$) pelo outro ramo (probabilidade de certa cor **dada** a urna sorteada) = $2/5$ sendo que esta última é uma probabilidade do tipo condicional. Na notação usual temos $P(\text{urna A e azul}) = P(\text{urna A}) \cdot P(\text{bola azul} \mid \text{urna A}) = 1/2 \cdot 2/5 = 1/5$.

Para a resposta à pergunta 1) basta considerar **um** ramo da árvore, o ramo **azul** que se segue ao ramo **urna A** ($\text{bola azul} \mid \text{urna A}$), cuja probabilidade é obtida pelo quociente entre o número de bolas azuis na urna A e o número total de bolas na urna A: $P(\text{bola azul} \mid \text{urna A}) = 2/5$. Analogamente obtém-se para a pergunta 2) a resposta: $P(\text{bola azul} \mid \text{urna B}) = 1/5$.

Quanto à pergunta 3), é preciso considerar que há dois caminhos a serem percorridos, ou seja, há duas maneiras para se obter uma bola azul: ou pelo caminho da

urna A ou pelo caminho da urna B (os eventos são mutuamente exclusivos e a probabilidade é obtida pela soma das probabilidades de cada evento). Assim,

$$\begin{aligned} P(\text{azul}) &= P(\text{urna A e azul}) + P(\text{urna B e azul}) = \\ &P(\text{urna A}) \cdot P(\text{azul} \mid \text{urna A}) + P(\text{urna B}) \cdot P(\text{azul} \mid \text{urna B}) = \\ &(1/2) \cdot (2/5) + (1/2) \cdot (1/5) = 3/10. \end{aligned}$$

Então, a probabilidade de ser sorteada uma bola azul é 0,30 (aqui poderia ser motivadora uma discussão de situação de jogo). Como os alunos geralmente apresentam o viés da equiprobabilidade (Lecoutre, 1988) seria interessante modificar o mecanismo inicial da escolha das urnas de forma a ter probabilidades *a priori* diferentes de 1/2, com a correspondente alteração do resultado final.

A descrição acima está contida em muitos textos de noções de estatística e pode parecer de fácil aprendizado, mas os alunos costumam ter problemas para fazer tanto o produto como a adição propostos acima. O diagrama de árvore, que é um modelo intuitivo, torna o problema visível pictoricamente e facilita o aprendizado instrumental: o aluno “aprende” que pode multiplicar entre si as probabilidades dos ramos que formam um caminho plausível, para obter a probabilidade desse caminho. No entanto esse produto, *autorizado* pela própria definição de probabilidade condicional, de onde se deduz a probabilidade conjunta, entra no conjunto de relações abstratas que o aluno tem dificuldade em aprender.

Para responder à pergunta 4), no entanto, nem sempre o caminho é tão suave. O roteiro para chegar a esta resposta não é tão explícito como os anteriores, pois, de acordo com a construção, o diagrama de árvore vai na direção “dada a *causa*” (urna sorteada) “qual o *efeito*” (cor sorteada), enquanto que a pergunta 4) vai na direção contrária, ou seja, “dado o *efeito*” (bola azul sorteada) “qual a *causa*” que o produziu (urna sorteada)? Em termos probabilísticos, a notação é a seguinte:

$$P(\text{urna A} \mid \text{bola azul sorteada}) = ?$$

Nesse caso, podemos aplicar diretamente a definição de probabilidade condicional:

$$\begin{aligned} P(\text{urna A} \mid \text{bola azul}) &= P(\text{urna A e bola azul}) / P(\text{azul}) = \\ &= P(\text{urna A}) \cdot P(\text{bola azul} \mid \text{urna A}) / P(\text{bola azul}) = \\ &= (1/2) \cdot (2/5) / (3/10) = 2/3, \end{aligned}$$

cuja interpretação é: se o experimento resultar numa bola azul, será mais provável que a urna A tenha sido a sorteada (mesmo porque, $P(\text{urna B} \mid \text{azul}) = 1 - 2/3 = 1/3$). O raciocínio utilizado para responder à pergunta 4) corresponde à aplicação do teorema de Bayes, cujo caminho é:

- a probabilidade *a priori* de ser sorteada a urna A é de $1/2$;
- dadas as evidências (bola azul), isto foi incorporado ao cálculo, produzindo o que se chama de probabilidade *a posteriori* da urna A, dado que a bola sorteada foi a azul;
- a probabilidade *a priori* associada à urna A que era $1/2$ foi “atualizada” e seu valor, agora chamado de probabilidade *a posteriori* é $2/3$;
- concluindo: a informação de que a bola sorteada foi azul aumenta a probabilidade da urna A ter sido a sorteada $\rightarrow$ de $1/2$ (probabilidade *a priori* da urna A ser sorteada) para $2/3$ (que é a probabilidade *a posteriori* da urna A ter sido sorteada, dada a informação de que saiu uma bola azul).

De qualquer forma, é bom ressaltar que o aluno tem muita dificuldade em apresentar a resposta correta de um problema como esse, principalmente pela dificuldade que ele tem em reconhecer a diferença entre $P(A \mid B)$ e $P(B \mid A)$ (no exemplo, $P(\text{urna A} \mid \text{bola azul})$ e $P(\text{bola azul} \mid \text{urna A})$, apesar do encaminhamento ter sido feito de forma a que todas as probabilidades de que ele necessita sejam calculadas, passo a passo. Já comentamos sobre o fato do assunto *bolas em urnas* não ser motivador para o aluno, mas é costume começar desta forma, não só porque o professor está mais acostumado, como já o dissemos, mas pela simplicidade em trabalhar com os elementos como bolas,

urnas, moedas, etc., que favorecem inclusive certos exercícios de simulação em classe. Cabe ao professor transferir esse esquema para um contexto mais próximo do cotidiano do aluno e é precisamente isto que faremos a seguir.

Considere-se, por exemplo, que a probabilidade que um indivíduo de uma certa população tem de contrair meningite seja de 0,005 (0,5%) – bem baixa! Existe um teste disponível para detectar a doença, mas que não apresenta acerto em 100% dos casos. Vamos supor que dado que a pessoa tem meningite, o teste dá positivo (acusa que o paciente tem meningite) em 95% dos casos, e também, dado que a pessoa não tem meningite, o teste dá negativo (acusa que o paciente não tem meningite) em 90% dos casos. Este contexto já nos oferece três probabilidades, sendo duas delas do tipo condicional, a saber:

$$P(\text{meningite}) = 0,005 \text{ (incondicional)}$$

$$P(\text{teste positivo} \mid \text{meningite}) = 0,95 \text{ (condicional)}$$

$$P(\text{teste negativo} \mid \text{não meningite}) = 0,90 \text{ (condicional)}.$$

Se um aluno da classe suspeitar que tem meningite, fizer o teste e se o resultado for positivo, será normal que ele se desesperar? Na maioria das vezes as pessoas intuitivamente se desesperam, porque acham que têm 95% de chance de estar com a doença. Isto porque há uma confusão entre as probabilidades condicionais

$$P(\text{teste positivo} \mid \text{meningite}) = 0,95$$

e

$$P(\text{ter meningite} \mid \text{teste positivo}) = ?$$

Esta última é a probabilidade que o aluno deveria calcular antes de se desesperar. Esta nova probabilidade, $P(\text{meningite} \mid \text{teste positivo})$, é a que passaremos a calcular (e é semelhante à pergunta 4 do exemplo anterior, de bolas em urnas).

É possível repetir o mesmo esquema de árvore do problema das urnas, substituindo **ter meningite e não ter meningite** por urna A e urna B, respectivamente, e **teste positivo** por bola azul e **negativo** por bola verde, com as correspondentes probabilidades do novo enunciado.

Isto posto vem:

$P(\text{meningite} \mid \text{teste positivo}) =$

$P(\text{meningite}) \cdot P(\text{teste positivo} \mid \text{meningite}) / P(\text{teste positivo}).$

Para o cálculo do denominador é preciso levar em conta todas as possibilidades do teste dar positivo (equivalente ao que fizemos no exemplo anterior, quando foi calculada a $P(\text{bola azul})$). Assim,

$$P(\text{teste positivo}) = P(\text{meningite}) \cdot P(\text{teste positivo} \mid \text{meningite}) + \\ P(\text{não meningite}) \cdot P(\text{teste positivo} \mid \text{não meningite}).$$

Com os dados do problema,

$$P(\text{teste positivo}) = 0,005 \cdot 0,95 + 0,995 \cdot 0,10 = 0,104,$$

o que produz

$$P(\text{meningite} \mid \text{teste positivo}) = (0,005 \cdot 0,95 / 0,104) = 0,046,$$

resposta ao nosso problema, ou seja, a probabilidade do aluno ter meningite, dado que o teste deu positivo, é 4,6%, que embora possa deixar o aluno preocupado, evidentemente não é para se desesperar como se fosse 95%!

Então, antes de se submeter ao exame, a probabilidade *a priori* de ele ter meningite era de 0,005 (0,5%), correspondente à incidência da doença na população. Uma vez colhida a informação de que o teste deu positivo, a probabilidade *a posteriori* de ele ter meningite cresceu quase dez vezes (4,6%), mas ainda é pequena. Por quê? Devido à baixa incidência da doença na população, que é de 0,5%. A probabilidade de ele não ter meningite, dado que o teste deu positivo, é complementar da anterior, e vale 95,4%: aí

o indivíduo percebe quão grande é a chance de ele não ter a doença, mesmo sabendo que o teste deu positivo. O que o aluno, ou mesmo o cidadão comum, confunde são idéias associadas às probabilidades condicionais $P(\text{teste positivo} \mid \text{meningite})=0,95$ e $P(\text{ter meningite} \mid \text{teste positivo})=0,046$, intuindo (falsamente) que essa é uma relação comutativa! É claro que dois valores tão diferentes entre si, como o são 0,95 com 0,046, não seriam confundidos pelo senso comum, mas também o senso comum não tem arraigado o raciocínio “inverso” do cálculo de probabilidades e o contexto formal afugenta boa parte dos alunos. Na verdade parece que as pessoas (e os alunos em particular) muitas vezes ignoram a probabilidade *a priori*, não levando em conta, como nesse caso por exemplo, a baixa incidência da doença na população.

5.3 Conclusão

Neste capítulo procuramos descrever diferentes maneiras de definir probabilidade e estatística, tanto pelos pesquisadores reconhecidos da área (mesmo que com o viés anglo-saxão) como por dicionários a serem consultados pelo cidadão. Não devemos ocultar dos alunos a diversidade de opções encontradas, o que já é um bom começo para introduzi-los, aos poucos, aos problemas epistemológicos da área.

No exemplo da meningite consideramos a aplicação do teorema de Bayes, sob o ponto de vista probabilístico, supondo que a probabilidade *a priori* seja perfeitamente conhecida (neste caso, a incidência da doença na população foi considerado como conhecida, embora isto possa ser discutível). Vimos então que a probabilidade *a priori* de ter meningite, $P(\text{meningite})$, foi atualizada pelos dados $\{P(\text{teste positivo} \mid \text{meningite})\}$, produzindo a probabilidade *a posteriori* $\{P(\text{meningite} \mid \text{teste positivo})\}$. Isto significa que a técnica Bayesiana calcula a probabilidade *a posteriori* usando a probabilidade *a priori* e as evidências (dados). Resumindo, tanto no problema das urnas quanto neste da meningite, tem-se:

$$posteriori \approx f(priori, \text{dados}).$$

Exemplos com as técnicas inferenciais clássica e Bayesiana serão apresentados no próximo capítulo.

Capítulo 6 Técnicas de Inferência Estatística

6.1 Introdução

Este capítulo não tem como objetivo comparar a estatística clássica com a Bayesiana do ponto de vista formal, e nem de decidir por uma ou por outra em áreas de aplicação que necessitam de uma análise quantitativa do tipo inferencial. Nosso intuito aqui é descrever duas técnicas inferenciais alternativas entre si e verificar as dificuldades na abordagem a partir de uma ou de outra linha, pois a preocupação é com as disciplinas básicas de estatística ensinadas nos primeiros anos do curso superior, na maioria das vezes para alunos que não sabem ainda o que é pesquisa e que não estão motivados para estes aspectos. Estamos pois analisando as duas escolas, clássica e Bayesiana, no atendimento àquela parte do programa da disciplina introdutória de probabilidade e estatística que trata da inferência

A inferência estatística apresentada aos alunos pode ser considerada de duas maneiras: ou sob o ponto de vista de estimação ou como um problema de teste de hipóteses. No primeiro caso, com abuso de linguagem, poderíamos dizer que é mais uma *especulação* inferencial a respeito do comportamento de alguma característica de uma população, enquanto que, no segundo caso, o pesquisador tem motivos para estabelecer hipóteses, associadas a valores de interesse para os parâmetros em questão, hipóteses essas que serão postas à prova.

Um tratamento mais geral, que não é apresentado normalmente aos alunos em uma disciplina introdutória de estatística, diz respeito diretamente à análise da decisão, com a atribuição do conceito de *utilidade* (tributário do conceito de *valor*) às possíveis consequências de tomadas de decisão: a sugestão é escolher a decisão que maximize a utilidade esperada (ver Lindley, 1985). Tanto na teoria de Neyman-Pearson como na Bayesiana é possível discutir os métodos inferenciais sob o ponto de vista da *teoria da*

decisão, mas isto seria sugerido para uma segunda disciplina de estatística, o que não é objeto de discussão neste trabalho.

É importante frisar que neste capítulo serão usados muitos conceitos que normalmente são vistos como pré-requisitos para a construção dos testes de hipóteses ou dos métodos de estimação. Esses tópicos serão considerados conhecidos (no sentido didático, ou seja, já teriam sido apresentados e discutidos com os alunos) para o desenvolvimento dos exemplos.

6.2 Teste de Hipóteses

O contexto em que é iniciado o assunto da inferência estatística numa disciplina introdutória diz respeito a problemas variados, como já comentamos anteriormente, como uma possível associação entre efeitos de drogas medicamentosas ou o nível de satisfação de funcionários em empresas. Alguns outros exemplos poderiam ser citados:

- decidir se a proporção de eleitores que se declaram a favor de um determinado candidato atingirá o limite necessário para que ele vença no primeiro turno;
- verificar se a média de salário de uma parcela da população está acima do salário histórico (padronizado) desta mesma parcela;
- verificar se, para crianças em fase de alfabetização, o método de ensino A é superior ao método de ensino B;
- analisar se a taxa associada à inflação brasileira ao longo do ano xxxx foi igual à do ano yyyy, etc.

Em todos estes casos, não temos acesso à população (por um ou outro motivo) e temos que recorrer a amostras. Para que se possa usar a amostra para tirar conclusões a cerca da população, são utilizadas técnicas da chamada inferência estatística, cujos aspectos de natureza lógica foram abordados no capítulo 4. De acordo com Batanero (2000), muitas

vezes o que está em jogo é uma hipótese substantiva, de natureza teórica (ou hipótese primitiva, na linguagem de Mayo, 1996), que dá origem a uma hipótese de pesquisa a qual, por sua vez, se transforma numa hipótese experimental ou observacional: esta última será colocada à prova, através da chamada hipótese estatística. É óbvio, então, que a hipótese estatística é mais restritiva do que uma hipótese científica. A passagem da conclusão obtida a partir de um procedimento experimental para o conhecimento científico é um assunto controvertido, e estamos tão somente apresentando uma abordagem para a parte inicial desse longo caminho.

A lógica do raciocínio inferencial em estatística está diretamente ligada com os conceitos de probabilidade e o método usado para obter conclusões: aceitar ou rejeitar uma hipótese, atribuir probabilidades a hipóteses, tomar uma decisão entre possíveis ações, etc. Com relação aos princípios básicos inferenciais apresentados numa disciplina introdutória de estatística, vamos analisar como os temas habitualmente são propostos aos alunos.

No exemplo referente à pesquisa eleitoral, em que uma amostra dos eleitores seria selecionada (de acordo com um esquema de amostragem previamente acordado), é geralmente posta à prova a hipótese de que a proporção de eleitores em determinado candidato é 50% (equivalente a 0,50). Neste caso, a hipótese nula seria $H_0 : p = 0,50$, onde p é a proporção de eleitores na população que votariam neste candidato.

No exemplo dos salários, a hipótese a ser testada poderia ser $H: \mu = \text{R\$ } 180,00$, onde μ é a média de salário da população em estudo. No caso da comparação dos métodos de ensino, a hipótese a ser testada seria $H: \mu_T = \mu_M$ (onde μ_T é a média da população a ser submetida ao esquema tradicional e μ_M é a média do grupo submetido ao esquema modificado). Todas as hipóteses acima colocadas são as chamadas *hipóteses estatísticas* que, como já comentamos, são uma forma simplificada da hipótese inicial do pesquisador, que é a hipótese primitiva ou substantiva. Para o último exemplo, de comparação de métodos de ensino, é possível simular um possível pensamento do

pesquisador sobre o seu problema, até chegar à hipótese estatística, que seria objeto de nosso estudo:

- hipótese substantiva (mecanismo não observável diretamente):

incluir elementos da cultura infantil no aprendizado facilita a alfabetização;

- hipótese de pesquisa (situação observável) :

aplicar a um grupo de crianças um método de alfabetização que inclua elementos da cultura local acelera o processo de alfabetização;

- hipótese experimental (especificação para pesquisa):

o resultado de um teste de alfabetização para um grupo de crianças submetido ao método modificado é superior ao do teste de outro grupo de crianças submetido ao método tradicional, que não incorpora elementos da cultura local.

- hipótese estatística:

$H : \mu_T = \mu_M$ (onde μ_T é a média da população a ser submetida ao esquema tradicional e μ_M é a média do grupo submetido ao esquema modificado).

Nos dois primeiros exemplos (eleitoral e salarial) há somente uma população envolvida, enquanto que no último caso são duas as populações envolvidas. Repete-se o problema de não termos acesso a toda a população (ou populações) e teremos que recorrer a amostras para que possamos testar nossas hipóteses. Em cada exemplo dado, a hipótese que será posta à prova (hipótese nula) não é aquela que o pesquisador quer verificar, mas sim a que diz que o que ele deseja verificar não acontece. Isto na verdade segue a tradição iniciada por Fisher, para quem o único propósito de se levar a cabo um experimento é dar às observações a oportunidade de refutar a hipótese nula.

Em seguida serão analisados dois problemas, sendo o primeiro com variável (aleatória) contínua e o segundo com variável (aleatória) discreta, através dos métodos inferenciais de Fisher, de Neyman-Pearson e Bayesiano. A escolha da variável contínua em primeiro lugar deve-se ao fato de que a maioria dos textos básicos de estatística ou das ementas das disciplinas introdutórias usa o caso contínuo para familiarizar o aluno com os conceitos iniciais de inferência. Poderíamos somente utilizar os dois últimos métodos mencionados, uma vez que o procedimento de Neyman-Pearson pode ser considerado como tendo englobado o de Fisher, apesar da disputa pessoal entre os pesquisadores. No entanto, apesar de o teste de significância de Fisher não ser incluído explicitamente nos textos e/ou disciplinas básicas de estatística, achamos interessante colocá-lo aqui, como referência.

6.2.1 Exemplo com variável contínua

Um pesquisador acredita que tem um método de ensino que acelera a reação de crianças de dez anos a uma determinada tarefa de prontidão motora, diminuindo o tempo de execução da mesma. Poderíamos considerar que a hipótese substantiva seja a de que existem métodos que aceleram a reação a determinada tarefa, numa dada faixa etária. O pesquisador então passa por vários estágios até chegar a uma hipótese estatística, ligada à eficiência que o seu método tem em reduzir o tempo de reação de crianças de dez anos, por exemplo, a determinada tarefa. Para isso ele realiza um experimento com um certo número de crianças e marca o tempo (em minutos) de execução da tarefa de cada criança antes e depois da aplicação do método. Para cada criança teríamos então duas medidas de tempo, cuja diferença poderia ser usada para os propósitos do pesquisador. Assim, numa situação extrema, em que todos os tempos registrados antes e depois fossem iguais para cada criança, todas as diferenças seriam nulas, o que evidentemente não daria suporte à teoria do pesquisador.

Parece então natural que o estudo se concentre nos valores das diferenças observadas para cada criança e chamaremos de D a variável (aleatória) que representa

tais diferenças. Vamos, por questão de simplicidade, supor que D tenha distribuição normal com parâmetros θ (média desconhecida) e 1 (desvio padrão conhecido: suposição pouco realista, mas que facilita a presente discussão). Uma hipótese estatística poderia se referir à característica θ (média da população das diferenças), a qual, se fosse nula, como já o dissemos, não traria evidências para suportar a hipótese do pesquisador.

6.2.1.1 Teste de significância de Fisher

Segundo a proposta inicial de Fisher, a hipótese estatística a ser posta à prova seria justamente aquela contrária à tese do pesquisador, ou seja, será posta à prova a hipótese $H_0 : \theta = 0$, cujo significado é de que seria nula a média populacional das diferenças.

Vamos supor que o pesquisador tenha feito observações em dez crianças e que as diferenças (em minutos) anotadas para cada criança entre o tempo de reação antes e depois da aplicação do método foram

1,2 0,8 -0,4 1,7 -0,5 -1,1 1,7 2,0 1,5
2,5.

Nosso interesse está centrado na média populacional (θ) e a teoria estatística sugere que a média amostral ($\bar{D}$) seja usada como representante (estimador) de θ . Além disso, pode ser demonstrado que $\bar{D}$, que também é uma variável (aleatória), tem distribuição normal com média θ e desvio padrão $(1/\sqrt{n})$, desde que tenha sido obtida através de amostras aleatórias de tamanho n de diferenças D , cuja distribuição seja normal, com média θ e desvio padrão 1 . A notação usual neste caso é $\bar{D} \sim N(\theta; 0,32)$, pois $n=10$. A estimativa $\bar{d}$ do estimador $\bar{D}$, ou seja, a média amostral observada foi, neste caso, $\bar{d} = 0,94$.

Supondo a hipótese nula verdadeira, podemos escrever

$(\bar{D} | H_0) \sim N(0, 0,32)$, e o valor obtido como média amostral (0,94) deve ser cotejado com os demais valores desta distribuição. Verifica-se que ele se distancia da média proposta pela hipótese nula por quase três desvios padrão (2,94) o que produziria uma probabilidade igual a 0,0016 à sua direita, na curva normal de $\bar{D} | H_0$ (valores disponíveis em tabelas ou em aplicativos computacionais). Este valor, 0,0016, é chamado de *nível descritivo* do teste e simbolizado por p (“*p-value*”).

Nessas circunstâncias, com uma probabilidade tão pequena, e usando o raciocínio de Fisher, diremos que ou a hipótese nula não está correta (e deveria ser refutada) ou então ocorreu um valor de média amostral extremamente raro (e nesse caso a hipótese nula não poderia ser refutada). Fisher sugere o que ele chama de procedimento indutivo, ou seja usar a amostra para tirar conclusões sobre a população de onde vêm os dados, estabelecendo um limite arbitrário para a probabilidade (nível de significância) de ocorrência do evento, sendo que probabilidades inferiores a este limite levariam à refutação da hipótese nula. Segundo Fisher, este limite é arbitrário e subjetivo, mas por sua própria influência vem sendo usado historicamente o valor de 5% (às vezes, numa prática mais rigorosa é proposto o valor de 1%).

No exemplo acima, como $0,0016 < 0,05$, (isto é, $p < 0,05$), há indícios para que a hipótese nula seja refutada. Nesta decisão está embutida uma possibilidade de erro, de 0,16%, correspondente a refutar a hipótese indevidamente, pois essa é justamente a probabilidade (p) de observarmos um valor de média amostral superior a 0,94 na distribuição $\bar{D} | H_0$. Este procedimento parece ser de pouca utilidade, se a conclusão parar no próprio valor do nível descritivo p . Daí porque o próprio Fisher, ao ser criticado por isso, dizia que o teste de significância somente deve ser usado como uma ferramenta heurística para sugerir outros caminhos de investigação.

6.2.1.2 Teste de hipóteses de Neyman-Pearson

O teste de hipóteses proposto por Neyman e Pearson pressupõe um par de hipóteses: a hipótese nula H_0 e a hipótese alternativa H_a , exaustivas e mutuamente exclusivas, que, no exemplo do tempo de reação, teriam a seguinte forma genérica: $H_0 : \theta = 0$ e $H_a : \theta > 0$, onde a hipótese nula afirma que o tempo médio de reação não diminui com o método e a hipótese alternativa afirma que o tempo médio de reação diminui com o método. Isto equivale a dividir o espaço paramétrico em duas partes, a primeira que representa a hipótese nula (a ser testada) e a segunda, que representa a hipótese alternativa (caminho a ser tomado no caso de rejeição da hipótese nula). A construção da hipótese alternativa, neste exemplo, é do tipo *monocaudal* ($\theta > 0$) em vez de *bicaudal* ($\theta \neq 0$), pois, se o pesquisador acredita que o método dele diminui o tempo de reação, não lhe interessa verificar se o referido tempo piora e, na verdade, valores do parâmetro inferiores a 0 podem ser englobados na hipótese nula. Neste caso, somente valores de média amostral “muito” à direita da distribuição (em sua cauda) levarão à rejeição da hipótese nula, os demais levando à não rejeição. As duas hipóteses, nula e alternativa, se referem ao valor do parâmetro θ (média populacional), que será estimado pela média amostral $\bar{D}$ a qual, como já dito no exemplo anterior, tem distribuição normal $N(\theta; 1/\sqrt{n})$.

Enquanto que o teste de Fisher analisa tão somente a chamada hipótese nula, nesta abordagem aparece uma outra hipótese, a hipótese alternativa, e são definidos dois tipos de erro: o *erro do tipo I*, que seria rejeitar a hipótese nula quando ela é verdadeira e o *erro do tipo II*, que corresponde a não rejeitar a hipótese nula quando ela for falsa (como se vê, são erros com estrutura condicional). A notação usual para as probabilidades associadas a estes erros é, respectivamente, $P(\text{erro I}) = \alpha$ e $P(\text{erro II}) = \beta$. O teste *ideal*, segundo o critério de Neyman-Pearson, é aquele que não só produz uma pequena probabilidade de erro do tipo I (isto é, uma pequena probabilidade de rejeitar falsamente a hipótese nula H_0) mas que também produza o máximo *poder discriminatório* entre a hipótese nula e possíveis alternativas ($\text{poder} = 1 - \beta$). Simplificando o problema, vamos considerar um valor específico para a hipótese alternativa H_a , em vez de deixá-la assumir

valores num intervalo, o que isto significa que a hipótese alternativa será *simples*, tal como a hipótese nula.

Voltando ao exemplo, vamos escolher arbitrariamente a seguinte hipótese alternativa $H_a : \theta = 1$, o que significa que a média das diferenças vale 1 minuto, ou seja, o método diminui, em 1 minuto, o tempo de execução da tarefa). Temos então um teste de hipótese nula simples contra hipótese alternativa simples:

$$H_0 : \theta = 0 \quad \text{vs} \quad H_a : \theta = 1.$$

Este procedimento é mais completo do que o de Fisher, no sentido de que a existência de uma hipótese alternativa, que não aparecia nos testes de significância, fornece mais elementos para tirar conclusões e, ao mesmo tempo, favorece a verificação das limitações do próprio procedimento, inclusive com o cálculo de outros erros associados às possíveis conclusões. O próprio estabelecimento das hipóteses não segue a idéia de Fisher, para quem o interesse na especificação de uma hipótese estaria diretamente ligado à possibilidade de sua refutação. Na teoria de Neyman-Pearson, a escolha entre a hipótese nula e a hipótese alternativa está diretamente ligada à importância dos erros de tipo I e tipo II associados ao problema, sob o ponto de vista da área de aplicação. A sugestão seria controlar o erro mais importante (sob a ótica do pesquisador) através do nível de significância (que é a probabilidade do erro do tipo I), situação que indicaria a hipótese nula do problema.

Faz parte da técnica de Neyman-Pearson a fixação (*a priori*) do nível de significância do teste, que é a probabilidade do erro do tipo I, α , que por influência de Fisher, como já dissemos, costuma ser colocado como 5% ou 1%. Da fixação do valor de α decorre a construção da chamada *região crítica de tamanho α* , que é constituída pelos intervalos extremos da distribuição normal que levam à rejeição da hipótese nula H_0 .

No exemplo em questão, como a hipótese alternativa apresenta um valor superior ao da hipótese nula, a região crítica será monocaudal, significando que somente valores *muito* à direita da distribuição (em sua cauda) levarão à rejeição da hipótese nula. O valor de α é que calibra este limite. Com a escolha (arbitrária) de $\alpha = 0,05$ (5%), a região crítica é constituída pelo intervalo da distribuição $\bar{D} \mid H_0 \sim N(0, 0,32)$ a partir do valor que deixa a probabilidade α à sua direita.

A esta altura do curso, assume-se que os alunos já tenham completo domínio sobre o cálculo de áreas associadas à distribuição normal, o que facilita a construção da região crítica (RC) deste exemplo. Ela será, então, representada pelo $R.C. = \{ \bar{d} \mid \bar{d} > 0,52 \}$, pois 0,52 é o ponto na distribuição $\bar{D} \mid H_0 (N(0, 0,32))$ que deixa uma área de 5% à sua direita e a concentração da probabilidade do erro do tipo I é toda à direita dado que a hipótese nula somente será rejeitada no caso de um possível valor médio de diferenças ser suficientemente maior do que zero.

Para a distribuição em questão, tolera-se (para não se refutar H_0) um valor de média até 0,52, a partir do qual o procedimento indica a rejeição da hipótese nula H_0 , em favor da hipótese alternativa H_a . Em outras palavras, se o valor a ser observado para a média da amostra obtida a partir do experimento estiver na região crítica, rejeita-se H_0 e, até “segunda ordem”, tudo se passa como se a hipótese alternativa fosse verdadeira. Assim rejeitar a hipótese nula indica como caminho a hipótese alternativa (a média populacional θ é igual a 1), significando que o método do pesquisador diminui em um minuto, em média, o tempo de reação das crianças de dez anos em determinada tarefa. Por outro lado, se o valor da média observada não pertencer à região crítica, não se rejeita H_0 .

Com as suposições feitas até agora, podemos também calcular a probabilidade de erro do tipo II (β), uma vez que a hipótese alternativa é do tipo simples (nos demais casos, ou seja, com a hipótese alternativa composta, constrói-se uma função). Neste caso particular, é possível calcular (β) e o correspondente poder do teste, já definido por (1-

β). O erro do tipo II corresponde a não rejeitar a hipótese nula, sendo ela falsa (ou seja, condicionado ao fato de ela ser falsa). A probabilidade do erro II, portanto, equivale ao cálculo da área da curva normal à esquerda de 0,52, que é o ponto que limita a região crítica, só que agora numa curva normal com a média proposta pela hipótese alternativa, ou seja, $\theta = 1$ (pois estamos considerando a hipótese nula como sendo falsa e então a alternativa seria a verdadeira!). A área em questão é calculada a partir da distribuição $\bar{D} | H_a (N(1, 0,32))$, pois mantivemos o mesmo desvio padrão anterior. Neste caso, o cálculo de β , $P(\bar{D} < 0,52 | H_a) = 0,07$ (7%), o que produz um poder (probabilidade de rejeitar a hipótese nula sendo ela falsa) de 0,93 (93%).

Este tipo de teste discrimina bem entre as duas alternativas propostas. Valores do parâmetro menores do que a unidade para a alternativa H_a produziriam probabilidades associadas ao erro do tipo II maiores e, correspondentemente, menor poder de discriminação entre elas.

Até aqui não entrou o experimento propriamente dito pois somente foram feitas considerações teóricas sobre o valor que a média amostral deveria ter para que fosse rejeitada a hipótese nula, sob determinadas suposições iniciais, como por exemplo, normalidade e desvio padrão conhecido. Isto mostra como este procedimento vai do desconhecido (população) para o conhecido (amostra), embora tenha como objetivo o contrário, isto é, a passagem do particular (amostra) para o geral (população). Uma vez construída a região crítica, serão utilizados os valores obtidos no experimento, que resultaram no valor para a média amostral $\bar{d} = 0,94$: como $0,94 > 0,52$, há evidências que permitem refutar a hipótese nula e portanto tudo se passa como se a hipótese alternativa fosse a “verdadeira”, corroborando com a hipótese do pesquisador de que seu método diminui o tempo de reação a determinada tarefa por crianças de dez anos (neste caso, tudo se passa como se o tempo de reação individual de cada criança fosse diminuído em um minuto, em média, uma vez que a hipótese alternativa apresenta somente um valor, a unidade, e não um intervalo).

A conclusão de rejeitar a hipótese nula é equivalente à obtida no teste de significância de Fisher, mas os elementos usados não são exatamente os mesmos. Além disso, no teste de Neyman-Pearson há a hipótese alternativa a ser considerada, com o correspondente erro do tipo II. A interpretação probabilística que se dá ao nível de significância α é a seguinte: numa grande sequência de amostras independentes de tamanho n de uma população normal de diferenças D , com média $\theta=0$, e desvio padrão 1, o teste construído vai rejeitar, indevidamente, que $\theta=0$ em 5% ($= \alpha$) das vezes. Esta interpretação é, portanto, do tipo frequentista, e o procedimento Neyman-Pearson não permite que se atribua probabilidades às hipóteses consideradas.

6.2.1.3 Teste Bayesiano

Na proposta Bayesiana, continuamos a usar o mesmo exemplo do tempo de reação, com as hipóteses

$$H_0 : \theta = 0 \quad \text{vs} \quad H_a : \theta = 1,$$

onde, como antes, estamos considerando as hipóteses exaustivas e mutuamente exclusivas. Antes de realizar o experimento, é preciso atribuir probabilidades à característica θ , que até agora era um valor fixo, um parâmetro, uma característica da população, e que no contexto Bayesiano toma o papel de uma variável (aleatória). Se não houver motivos para se ter preferência ou por $\theta = 0$ ou por $\theta = 1$, pode-se atribuir a mesma probabilidade para cada hipótese considerada, isto é, $\frac{1}{2}$ ($P(\theta = 0) = P(\theta = 1) = \frac{1}{2}$). Estas probabilidades são conhecidas como probabilidades *a priori* das hipóteses nula e alternativa, respectivamente. Outros valores poderiam ter sido atribuídos ao invés de $\frac{1}{2}$, como será visto oportunamente.

A partir de agora serão utilizados alguns conceitos de probabilidade, principalmente probabilidade condicional, com a qual os alunos deverão estar

familiarizados antes de começar este tópico. A probabilidade de que $\theta = 0$ (H_0), dado que observamos $\bar{D}$, é a chamada probabilidade *a posteriori* de H_0 e pode ser escrita da seguinte forma, usando somente a definição de probabilidade condicional:

$$\begin{aligned} P(H_0 | \bar{D}) &= P(H_0 \cap \bar{D}) / P(\bar{D}) = \\ &= P(H_0) P(\bar{D} | H_0) / P(\bar{D}) = \\ &= P(H_0) P(\bar{D} | H_0) / \{P(H_0) P(\bar{D} | H_0) + P(H_a) P(\bar{D} | H_a)\}. \end{aligned} \quad (1)$$

Analogamente, para a probabilidade *a posteriori* de H_a , tem-se:

$$\begin{aligned} P(H_a | \bar{D}) &= P(H_a \cap \bar{D}) / P(\bar{D}) = \\ &= P(H_a) P(\bar{D} | H_a) / P(\bar{D}) = \\ &= P(H_a) P(\bar{D} | H_a) / \{P(H_0) P(\bar{D} | H_0) + P(H_a) P(\bar{D} | H_a)\}. \end{aligned} \quad (2)$$

Fazendo o quociente entre (1) e (2) vem:

$$\frac{P(H_0 | \bar{D})}{P(H_a | \bar{D})} = \frac{P(H_0)}{P(H_a)} \frac{P(\bar{D} | H_0)}{P(\bar{D} | H_a)} \quad (3)$$

1

2

3

O quociente 1 é a razão das probabilidades *a posteriori* das hipóteses que estão sendo testadas, ou seja, das probabilidades de cada hipótese, dada a evidência amostral $\bar{D}$; o quociente 2 é a razão das probabilidades *a priori* das mesmas hipóteses e o quociente 3 é considerado como a razão entre as *verossimilhanças*.

Aqui vale a pena uma pausa para alguns comentários sobre *verossimilhança*. Os principais elementos utilizados na inferência Bayesiana são as distribuições *a priori* e *a posteriori*, associadas ao parâmetro de interesse (o qual, neste exemplo é a média das diferenças dos tempos de reação, simbolizada por θ). Então a distribuição *a priori* de θ é

$P(\theta)$ e a distribuição *a posteriori* é $P(\theta \mid \text{dados})$, ambas se referindo a probabilidades associadas a θ , e para as quais, logicamente, valem as leis das probabilidades. A aplicação do teorema de Bayes permite escrever $P(\theta \mid \text{dados}) = P(\theta) P(\text{dados} \mid \theta) / P(\text{dados})$, e a quantidade $P(\text{dados} \mid \theta)$, muitas vezes anotada como $L(\theta)$, recebe o nome de verossimilhança. Apesar de não ser formalmente uma probabilidade, pois os seus valores não somam 1, muitos autores a consideram como tal (padronizando os valores): daí o uso da mesma notação $P(\cdot)$. Verossimilhança pode ser considerada como uma medida de relativa plausibilidade (Oakes, 1986) e a área sob a curva por ela delimitada não tem qualquer significado (na verdade é uma função de θ , para uma configuração fixada dos dados).

Neste exemplo, estamos usando as distribuições fazendo referência explícita às hipóteses associadas ao parâmetro de interesse θ , o que é equivalente a fazê-lo através do próprio parâmetro.

Voltando à resolução do problema, o quociente $\underline{1}$ será usado para decidir entre a hipótese H_0 e a hipótese H_a . Genericamente ele é comparado com uma constante C , que leva em conta conceitos como *utilidade e função de perda*. O caso mais comum, e pode-se dizer mais intuitivo, é aquele em que não se prioriza de início nenhuma das hipóteses em termos de valor ou utilidade, e então a escolha se faz através de $C=1$. Isto é, se o quociente $\underline{1}$ for maior do que a unidade, diríamos que as evidências suportam a hipótese nula H_0 . Caso contrário, diríamos que suportam a hipótese alternativa H_a .

Considerando o exemplo do tempo de reação, com as suposições feitas de normalidade e variância unitária, podemos escrever:

distribuição *a priori* : $P(H) = P(H) = \frac{1}{2}$ (escolha da distribuição uniforme)

verossimilhanças: $(\bar{D} \mid H_0) \sim N(0, 0,32)$

$(\bar{D} \mid H_a) \sim N(1, 0,32)$.

Aplicando os valores correspondentes na expressão (3), através da função de verossimilhança associada à distribuição normal, vem:

$$\begin{aligned}\frac{P(H_0 | \bar{D})}{P(H_a | \bar{D})} &= \frac{\frac{1}{2} \cdot \exp(-\frac{n}{2} \bar{D}^2)}{\frac{1}{2} \cdot \exp(-\frac{n}{2} (\bar{D}-1)^2)} \\ &= \exp(-(n/2) (\bar{D}^2 - (\bar{D}-1)^2)) \\ &= \exp(-(n/2) (2\bar{D} - 1))\end{aligned}$$

Os dados fornecidos pelo problema são: $n=10$ e $\bar{D} = 0,94$. Daí vem:

$$\begin{aligned}\frac{P(H_0 | \bar{D})}{P(H_a | \bar{D})} &= \boxed{\exp(-(n/2) (2\bar{D} - 1))} \\ &= \exp(-5 (2 \cdot 0,94 - 1)) = \exp(-4,4) = 0,012\end{aligned}$$

O quociente entre as probabilidades *a posteriori* sendo muito pequeno, significa que a probabilidade *a posteriori* da hipótese alternativa, dada a evidência amostral, é muito maior do que a probabilidade da hipótese nula, sob a mesma evidência. Dessa forma, rejeita-se H_0 em favor de H_a , o que significa que há evidências de que o método do pesquisador diminui em um minuto, em média, o tempo de reação das crianças para a realização da tarefa. Neste exemplo, em que a razão entre as distribuições *a priori* é igual a um, a razão entre as distribuições *a posteriori* é a própria razão entre verossimilhanças.

Resumindo, para a decisão entre H_0 e H_a neste caso, é suficiente escolher aquela hipótese cuja probabilidade *a posteriori* seja maior.

Neste exemplo, a distribuição *a priori* escolhida foi do tipo uniforme, e é sob essas condições que os resultados obtidos com a abordagem Bayesiana mais se aproximam daqueles obtidos com a teoria clássica. Entretanto, se houvesse motivos consistentes para o pesquisador supor que a distribuição uniforme não seria realista para o problema e que o melhor seria colocar $P(H_0) = 1/3$ e $P(H_a) = 2/3$, o valor do quociente final seria $(0,012/2) = 0,006$, aumentando mais ainda a chance de se rejeitar a hipótese nula. Isto ocorreu porque ela já começou com “desvantagem” no estabelecimento da probabilidade *a priori*, considerando sempre que o pesquisador tivesse motivos para isso. Por outro lado, se fosse colocado para a probabilidade *a priori* o valor $P(H_0) = 5/6$ e $P(H_a) = 1/6$, o quociente final seria $(5 \cdot 0,012) = 0,06$, o que para alguns pesquisadores poderia sugerir a não rejeição de H_0 .

6.2.2 Exemplo com variável discreta

Considere um estudo destinado a saber se a proporção de fumantes em uma certa população excede 25%. Para isso foi retirada da população uma amostra aleatória de 100 pessoas e foi observado que o número de fumantes era 35 (informação amostral). Vamos novamente abordar o problema sob três pontos de vista: teste de significância de Fisher, teste de hipóteses de Neyman-Pearson e procedimento Bayesiano.

6.2.2.1 Teste de significância de Fisher

Segundo a proposta inicial de Fisher, vamos por à prova a hipótese nula $H_0 : p = 25\%$ (0,25). Uma hipótese equivalente a esta, e que é mais conveniente para efeitos didáticos, pois trabalha diretamente com a distribuição binomial, extensivamente trabalhada na disciplina introdutória antes deste ponto, é $H_0 : np = 25$, ou seja, se de fato a proporção de fumantes for 25%, então espera-se que em 100 pessoas desta população haja, em média, 25 fumantes. Assim pode-se trabalhar com a variável (aleatória) F , que

representa o número de fumantes em 100 indivíduos (número de *sucessos*), com probabilidade de *sucesso* (ser fumante) igual a p . Se os ensaios forem independentes, podemos dizer que F segue uma distribuição binomial, com parâmetros p (probabilidade de *sucesso*) e $n=100$ (número de ensaios).

A distribuição de probabilidade binomial é bem conhecida e podemos calcular as probabilidades a ela associadas através da expressão $\{P(F=k) = C_{n,k} p^k (1-p)^{n-k}\}$, ou procurá-las em tabelas, ou ainda usar algum aplicativo estatístico para esse fim. Se a hipótese nula for verdadeira o número esperado de fumantes dentre os 100 pesquisados vale $0,25 \times 100=25$.

Para testar a hipótese proposta teremos que relacionar a amostra com o valor do parâmetro estipulado pela hipótese, da seguinte forma: se a hipótese nula for verdadeira, qual a probabilidade de observarmos um valor no mínimo tão extremo como o que foi de fato observado? Esta probabilidade, cuja notação usual é $P(F \geq 35 | H_0)$, pode ser calculada, usando para a distribuição de F os parâmetros $p = 25\%$ ($0,25$) e $n=100$. Temos então que $P(F \geq 35) = 0,0164$ (este é o valor no nível descritivo, p). Nessas circunstâncias, com uma probabilidade tão pequena, e usando o raciocínio de Fisher, diremos que ou ocorreu um número de sucessos extremamente raro (e a hipótese não poderia ser refutada) ou então a hipótese nula não está correta e poderia ser refutada.

Como já comentado no caso contínuo, Fisher sugere o que ele chama de procedimento indutivo. Nesse exemplo, como $0,0164 < 0,05$, há indícios de que a hipótese nula deva ser refutada. No caso, há uma possibilidade de erro, de 1,64%, de refutar a hipótese indevidamente, pois essa é justamente a probabilidade de observarmos no mínimo 35 fumantes numa amostra de 100, com probabilidade de sucesso (ser fumante) igual a 25%. Não podemos no entanto concluir (e esse é um dos erros que muitos alunos tendem a cometer) que a probabilidade da hipótese nula seja igual a 0,0164, pois não se trata de uma probabilidade sobre uma afirmação.

6.2.2.2 Teste de hipóteses de Neyman-Pearson

Repetindo o procedimento de Neyman-Pearson utilizado para variável contínua, temos o seguinte par de hipóteses para o problema dos fumantes:

$$H_0 : p = 0,25 \text{ e } H_a : p > 0,25.$$

Usando o processo anterior, essas hipóteses poderiam ser reescritas como

$$H_0 : np = 25 \text{ e } H_a : np > 25.$$

A construção da hipótese alternativa *monocaudal* ($np > 25$) em vez de *bicaudal* ($np \neq 25$) deve-se ao fato do interesse do estudo estar centrado em saber se a proporção excede 25% e não simplesmente se difere de 25%, quando então faria sentido a hipótese bicaudal.

Tais hipóteses, como no caso anterior, também se referem a valores de parâmetros, que são características populacionais, e aqui o procedimento também faz uso da distribuição binomial da variável (aleatória) F . Como já descrito no caso contínuo, faz parte do procedimento Neyman-Pearson a criação de uma região crítica de tamanho α , que é a probabilidade de erro do tipo I. No caso de uma distribuição binomial $B(100, p)$, para $\alpha = 5\%$, a região crítica, formada pelos pontos que levam à rejeição da hipótese nula H_0 , é encontrada através das probabilidades associadas à distribuição binomial $B(100, 0,25)$, e pode ser escrita da seguinte forma: $R.C. = \{33, 34, 35, \dots, 100\}$. Na verdade, por problemas de continuidade, esta região crítica não tem um valor probabilístico exato de 5% mas aproxima-se deste. Assim, se o número de fumantes na amostra estiver na região crítica, há evidências para se refutar a hipótese nula e, caso contrário, a hipótese nula não será refutada. Neste caso, como o número de fumantes na amostra foi igual a 35, valor que pertence à região crítica, a decisão será a de rejeitar a hipótese nula, ou seja, haveria evidências de que a proporção de fumantes na população excede 25%.

Essa última afirmação está associada a um erro de aproximadamente 5% , isto é, se repetirmos um número muito grande de vezes este experimento numa população em que a verdadeira proporção de fumantes seja de fato 25%, refutaremos erroneamente H_0 em aproximadamente 5% das vezes. Se a hipótese alternativa, ao invés de ser composta ($np > 25$) fosse simples, por exemplo $H_a: np=40$ (o que corresponde a uma proporção de sucessos igual a 0,40), o cálculo do erro do tipo II seria muito fácil, bastando para isso calcular $P(F < 33 \mid p=0,40)$, cujo resultado é 0,0615 ($\approx 6\%$). Isto significa que, se repetirmos um número muito grande de vezes este experimento, numa população em que a verdadeira proporção de fumantes é 40%, não refutaremos H_0 quando ela for falsa em aproximadamente 6% das vezes. No caso descrito os dois erros estão bem equilibrados, mas devemos levar em conta que foi estabelecido arbitrariamente o valor do parâmetro para a hipótese alternativa, deixando para o problema somente a escolha entre os dois valores de p (25% ou 40%).

Esta não é a situação usual, mas do ponto de vista didático é conveniente usá-la para reforço dos conceitos dos dois tipos de erro e suas interpretações, bem como dar uma certa ênfase no conceito de poder do teste. A situação mais geral indica a necessidade de construir uma função com diferentes valores de p da hipótese alternativa, para que a conclusão tenha um respaldo maior.

6.2.2.3 Teste Bayesiano

Vamos resolver agora o mesmo problema sob o ponto de vista Bayesiano. A estrutura que será utilizada envolve as duas hipóteses simples $H_0: p = 0,25$ e $H_a: p = 0,40$, já descritas anteriormente e que são equivalentes a, respectivamente, $H_0: np=25$ e $H_a: np=40$. O procedimento de Bayes tem como primeiro passo a atribuição de probabilidades *a priori* para as hipóteses H_0 e H_a . Na ausência de preferências, vamos usar a mesma estratégia do caso contínuo, atribuindo-se a mesma probabilidade para

ambas as hipóteses, ou seja $P(H_0) = P(H_a) = 1/2$. Estas probabilidades são as probabilidades *a priori* das hipóteses nula e alternativa, respectivamente.

Retomando a estrutura de probabilidades condicionais, já vista para variáveis contínuas, tem-se para as distribuições *a posteriori*, uma vez observado o valor de F:

$$\begin{aligned} P(H_0 | F) &= P(H_0 \cap F) / P(F) = \\ &= P(H_0) P(F | H_0) / P(F) = \\ &= P(H_0) P(F | H_0) / \{P(H_0) P(F | H_0) + P(H_a) P(F | H_a)\} \end{aligned} \quad (3)$$

e

$$\begin{aligned} P(H_a | F) &= P(H_a \cap F) / P(F) = \\ &= P(H_a) P(F | H_a) / P(F) = \\ &= P(H_a) P(F | H_a) / \{P(H_0) P(F | H_0) + P(H_a) P(F | H_a)\}, \end{aligned} \quad (4)$$

cujo quociente é análogo ao obtido na seção 6.2.1.3:

$$\frac{P(H_0 | F)}{P(H_a | F)} = \frac{P(H_0)}{P(H_a)} \frac{P(F | H_0)}{P(F | H_a)} \quad (5)$$

$$\boxed{1} \quad \boxed{2} \quad \boxed{3}$$

O procedimento será o mesmo usado no caso contínuo, onde 1 é a razão das probabilidades *a posteriori* das hipóteses que estão sendo testadas (ou seja, das probabilidades de cada hipótese, dada a evidência amostral F), o quociente 2 é a razão das probabilidades *a priori* das mesmas hipóteses e o quociente 3 é considerado como a razão entre as verossimilhanças. Se o quociente entre as probabilidades *a posteriori* for maior do que a unidade, diz-se que a evidência suporta a hipótese nula H_0 . Caso contrário, a evidência suporta a hipótese alternativa H_a .

No caso presente, em que o valor observado para F foi de 35 fumantes, para uma amostra com $n=100$, podemos reescrever a relação (5) como

$$\frac{1}{2} C_{100,35} 0,25^{35} \cdot 0,75^{65} / \frac{1}{2} C_{100,35} 0,40^{35} \cdot 0,60^{65} = 0,14.$$

Este quociente, sendo bem inferior à unidade, mostra que a probabilidade *a posteriori* de H_a $\{P(H_a \text{ dada a evidência amostral } F)\}$ é bem superior àquela de H_o (dada a mesma evidência): a sugestão é refutar a hipótese nula H_o em favor da hipótese alternativa H_a . Este resultado vai na mesma direção do que os anteriores, como no caso contínuo.

Se para as distribuições *a priori* tivéssemos $P(H_o)=1/3$ e $P(H_a) = 2/3$, a relação (5) teria como valor 0,07, o que mostraria tendência ainda maior para a hipótese alternativa, dado que já inicialmente deu-se mais peso a ela. No entanto se fosse dado como *priori* para a hipótese nula o valor 0,6 (e, conseqüentemente, 0,4 para a alternativa), o quociente (5) teria o valor de 0,84, mostrando que, mesmo que a preferência *a priori* tenha sido pelo valor $p=0,25$, a probabilidade *a posteriori* para $p=0,40$ revelou-se mais adequada, face às evidências amostrais.

6.2.3 Comparação entre os três procedimentos

Para o exemplo proposto, as três abordagens vão na direção de rejeitar a hipótese nula, cada uma usando os próprios argumentos, com interpretações próprias. Na técnica de Fisher, no entanto, como a alternativa não é sequer considerada, o raciocínio parece truncado. Não é sem motivo que os dois procedimentos (Fisher e Neyman-Pearson) acabaram fluindo na literatura para uma teoria híbrida, que mescla nível de significância do teste com nível descritivo. A teoria Bayesiana, utilizando *priori* uniforme, também acena na mesma direção anterior. Por outro lado, dependendo do estabelecimento da *priori*, feito de modo subjetivo e respeitando o princípio da coerência (ver Gelman, 1995), o resultado poderá não concordar com o das outras técnicas.

Os dados coletados, tanto no caso contínuo como no caso discreto, trabalhados sob as três abordagens, indicam pela rejeição da hipótese nula: no caso contínuo há evidências de que o método proposto diminui o tempo de reação a determinada tarefa e, no caso discreto, há evidências de que a proporção de fumantes seja maior do que 25%.

6.3 Estimação

Os aspectos discutidos na seção anterior dizem respeito à parte da inferência relativa aos testes de hipóteses. Nos dois exemplos propostos havia uma atribuição de valor ao parâmetro considerado e isto era posto à prova. No entanto, há circunstâncias em que o desejo é situar o valor do parâmetro, ao invés de testar valores para o mesmo. Esta situação é chamada de **estimação** e é também uma técnica inferencial e, grosso modo, podemos dizer que ela se divide em **estimação pontual** e **estimação por intervalo**. Esta não é uma nomenclatura muito adequada para grande parte dos Bayesianos, mas vamos mantê-la para facilitar a unidade na discussão. Para que o capítulo não fique muito extenso, será apresentado um exemplo de estimação “pontual” Bayesiana, com uma discussão sucinta sobre a abordagem pontual clássica. Também com relação à estimação por intervalo será feita uma apresentação sucinta do tema.

6.3.1 Estimação pontual

O exemplo a seguir foi adaptado de Albert (1996): o técnico de uma equipe de futebol deseja conhecer a proporção (p) de acertos a gol, do mais novo contratado do time, em batidas do limite da área, e sem barreira. Vamos definir **acerto** como *sucesso* e **erro** como *fracasso* (linguagem comum em *ensaios de Bernoulli*); p será a probabilidade de acerto (*sucesso*) do jogador em questão.

Para estimar p na abordagem Bayesiana, vamos analisar os seguintes passos:

- Definir o conhecimento *a priori* sobre p

Podemos considerar que p possa assumir qualquer valor de uma distribuição contínua $[0,1]$ ou que assume valores numa distribuição discreta de probabilidades. Em qualquer dos casos esta distribuição é chamada de distribuição *a priori*, pois ela representa a informação sobre p antes da coleta dos dados. Neste exemplo vamos considerar a segunda possibilidade, isto é, vamos atribuir a p valores de uma variável discreta, valores esses escolhidos segundo o que é largamente conhecido na área esportiva, na opinião do técnico do time. Ele então informa que uma proporção $p=0,40$ define o jogador como fraco; $p=0,50$, como regular; $p=0,75$ como bom e $p=0,90$ como excepcional.

Como antes de conhecer a habilidade do jogador o valor de p é desconhecido, sendo justamente o valor que se deseja estimar, a distribuição *a priori* de p pode ser definida inicialmente com os quatro valores acima mencionados. O termo distribuição significa aqui o conjunto dos valores de p e suas respectivas probabilidades de ocorrência. Para completar a distribuição pode ser usado um critério não informativo (atribuir probabilidades iguais para todos os valores, no caso $\frac{1}{4}$ para cada um) ou um critério baseado na experiência (ou informação) prévia do técnico. Aqui está um diferencial importante com relação à teoria clássica, já visto no caso de teste de hipóteses: nela, o valor p , desconhecido, é fixo, não admitindo uma distribuição de probabilidades. Na abordagem Bayesiana, o parâmetro desconhecido é considerado uma variável (aleatória) e a incerteza sobre ele é expressa em termos probabilísticos. Vamos adotar aqui, primeiramente, o critério não informativo, isto é, probabilidades iguais para todos os valores de p (distribuição uniforme discreta):

Valores de p	0,40	0,50	0,75	0,90
Probabilidades	0,25	0,25	0,25	0,25

- Verificar quantos sucessos foram observados nos n ensaios

Considere-se que o número observado de sucessos tenha sido 12 em 20 cobranças, ou seja, o atleta acertou 12 chutes a gol em 20 cobranças independentes de falta. Então a verossimilhança associada pode ser escrita como $p^{12} (1-p)^8$.

- Calcular a distribuição *a posteriori* de p

Para o cálculo da distribuição *a posteriori* vamos aplicar o teorema de Bayes, a exemplo do que foi feito anteriormente, segundo a relação

$$(posteriori) \approx (priori) \cdot (verossimilhança),$$

onde o símbolo $\approx$ representa proporcionalidade. Neste caso,

$$\begin{aligned} P(p=0,40 \mid \text{foram observados 12 sucessos}) &= \\ P(p=0,40) \cdot P(12 \text{ sucessos} \mid p=0,40) / P(12 \text{ sucessos}) &= \\ 0,25 \cdot 0,40^{12} 0,60^8 / (0,25 \cdot (\sum p^{12} (1-p)^8)) &= 0,1637, \end{aligned}$$

para todos os valores de p apresentados em (6).

Esta probabilidade *a posteriori* (16,37%) relativa ao valor de $p = 0,40$, é menor do que aquela colocada *a priori* (0,25), mostrando que a evidência amostral diminuiu a chance do valor $p = 0,40$. Em outras palavras, com o resultado de 12 acertos, o jogador teria menos chance de ser considerado “fraco” na opinião do técnico. Os demais valores para a distribuição *a posteriori* encontram-se na tabela a seguir.

p	priori	posteriori
0,40	0,25	0,1637
0,50	0,25	0,5537
0,75	0,25	0,2808
0,90	0,25	0,0018

<p>média da distribuição <i>a priori</i> : 0,6375</p> <p>média da distribuição <i>a posteriori</i> = 0,5546</p> <p>(ambas obtidas através do produto dos valores de p pelas respectivas probabilidades)</p>
--

A priori, qualquer resultado, dentre os quatro apresentados para **p**, seria viável, uma vez que todas as probabilidades *a priori* são iguais a 0,25. *A priori*, o valor esperado para **p** (a média da distribuição *a priori*) é igual a 0,6375, que poderia ser interpretado como uma expectativa do treinador de que seu novo jogador seja pelo menos um pouco mais do que regular (se não, para que contratá-lo?). *A posteriori*, uma vez conhecido o resultado amostral de 12 acertos, o valor mais provável para **p** é 0,50 (pois a maior probabilidade *a posteriori*, 0,5537, é obtida para **p**=0,50). A probabilidade de que ele seja excepcional, dado o resultado de 12 sucessos, é 0,0018, bem inferior à probabilidade *a priori* dele ser excepcional, que é de 0,25. É claro que a resposta depende da *priori* considerada e supõe-se que ela seja sempre baseada na experiência do pesquisador (aqui do técnico), e que atenda às regras da coerência (ver Gelman et al. , 1995).

As conclusões apresentadas no parágrafo anterior foram baseadas numa discussão dos valores da tabela da distribuição *a posteriori* de **p**. É usual, também, usar para estimativa de **p** uma medida de posição associada à distribuição *a posteriori*, como por exemplo a média (poderia também ser a moda ou a mediana). Nesse caso então, uma estimativa Bayesiana para **p** poderia ser a média da distribuição *a posteriori*, que é igual a 0,5546, mostrando que a expectativa do técnico com o novo jogador diminuiu depois das evidências (antes era de 0,6375!). Após o teste o jogador contratado parece estar mais para “regular” do que para “bom” , e caberá ao técnico decidir pela sua contratação ou não.

Uma distribuição *a priori* diferente da que foi colocada acima poderia levar a resultados diferentes. Vamos supor que o técnico tenha mais dados sobre a nova contratação e que então a sua distribuição *a priori* seja mais informativa do que a anterior, a saber:

p	<i>priori</i>	<i>posteriori</i>
0,40	0,10	0,0681
0,50	0,15	0,3457
0,75	0,50	0,5844
0,90	0,25	0,0018

<p>média da distribuição <i>a priori</i> : 0,715</p> <p>média da distribuição <i>a posteriori</i> = 0,640</p> <p>(ambas obtidas através do produto dos valores de p pelas respectivas probabilidades)</p>
--

Neste caso, com a distribuição *a priori* modificada em função de informações prévias do técnico, o valor de **p** mais provável inicialmente seria **p=0,75**, ou seja, um jogador “bom”. O valor esperado da distribuição *a priori* é 0,715, o que mostra maior confiança inicial do técnico no novo jogador do que no problema com a distribuição *a priori* uniforme. Neste caso, o valor (*a posteriori*) mais provável para **p** continua sendo 0,75, mas o valor esperado da distribuição *a posteriori*, que pode ser utilizado como estimativa para **p**, diminuiu para 0,640, sugerindo ao técnico que o jogador está mais perto de ser “regular” do que ele supunha inicialmente. Muitas vezes este valor esperado da distribuição *a posteriori* é usado como a probabilidade de sucesso numa ocasião futura.

Como se vê, a distribuição *a posteriori* mudou de um exemplo para outro, acompanhando as mudanças da distribuição *a priori*. Por outro lado, a verossimilhança, que depende somente dos dados, não se alterou nos dois exemplos. Pode-se dizer que a maior confiança *a priori* do técnico no jogador “puxou” a distribuição *a posteriori* para

$p=0,75$, diferentemente do que aconteceu quando o técnico não tinha opinião inicial sobre o jogador.

Este mesmo problema pode ser resolvido pelo método clássico, envolvendo estimação pontual: como em 20 tentativas o jogador acertou 12, a estimativa de p é igual a $(12/20) = 0,60$. Este resultado não utiliza informações ou opiniões prévias do técnico. O número é obtido tão somente a partir da evidência amostral e o resultado é mais próximo ao Bayesiano com distribuição *a priori* não informativa. Por outro lado, a interpretação é diferente. De qualquer forma, neste problema de estimação, vemos que o método Bayesiano tem uma abordagem mais abrangente, podendo quantificar probabilidades *a posteriori* para diferentes valores da probabilidade de sucesso p . No entanto, o fato de o resultado se modificar de acordo com a distribuição *a priori* pode ser visto tanto como qualidade, pelos adeptos da escola Bayesiana, quanto como defeito, pelos seus críticos contumazes.

6.3.2 Estimação por intervalo

A teoria clássica oferece a possibilidade de calcular uma estimativa intervalar para a probabilidade de sucesso p , através do *intervalo de confiança*, que pode ser vista como a justaposição de uma margem de erro associada à estimativa de p . A afirmação associada ao intervalo é feita sobre o próprio intervalo e não sobre o parâmetro p , que seria a entidade de interesse. Normalmente esta é uma desavença entre as escolas clássica e Bayesiana, pois é muito difícil para um aluno e até mesmo para um pesquisador deixar de dar a conclusão como se fosse ligada à probabilidade do parâmetro estar no intervalo dado: isto traria mais “conforto” os interessados na análise, mas a teoria usada não autoriza essa interpretação!

O método clássico usual para este exemplo envolve a aproximação da distribuição binomial pela normal, mas não vamos entrar em detalhes. A interpretação do

intervalo de confiança (com coeficiente γ) para p é a seguinte: se o experimento for repetido um número muito grande de vezes em condições idênticas, o intervalo encontrado conterá o verdadeiro valor do parâmetro numa proporção igual a $100\gamma\%$. Como dissemos, esta afirmação refere-se ao intervalo e não ao parâmetro. Já a abordagem Bayesiana oferece um *intervalo de credibilidade* para o parâmetro p , com base em sua distribuição *a posteriori*. Este intervalo pode produzir um valor numérico análogo àquele produzido na análise clássica, mas a interpretação é diferente, provavelmente mais próxima da linguagem usual do aluno (ou do pesquisador), uma vez que a atribuição de probabilidade é feita sobre o parâmetro.

6.4 Conclusão

Os métodos de inferência clássica se baseiam na teoria freqüentista de Von Mises, e trabalham com um conjunto de referência, que é o espaço amostral. O uso da teoria freqüentista se deve ao fato de que a conclusão está baseada em todas as possíveis amostras que poderiam ter sido observadas, mas não o foram, e por repetições, sob as mesmas condições, do processo de amostragem ou de experimentação, cujo comportamento pode ser entendido a partir do raciocínio com freqüências. Daí a relevância do processo de obtenção dos dados para esta teoria.

A teoria Bayesiana, por sua vez, atribui às hipóteses iniciais uma probabilidade *a priori* que não é derivada nem de repetições experimentais e nem do processo de obtenção dos dados. Na verdade é a expressão da crença do pesquisador, e não necessita da consideração sobre repetições hipotéticas do experimento. O que se pede é que, na atribuição de probabilidades, seja usado o princípio da coerência, o que dá suporte ao procedimento. A repetição do uso do próprio teorema de Bayes para rever as probabilidades iniciais é usada como sugestão no sentido de evitar eventuais inconsistências de opinião do pesquisador. A própria teoria de planejamento experimental clássico, que sugere a aleatorização das unidades experimentais, não é utilizada

formalmente na teoria Bayesiana, pois a mesma parte do princípio que toda a informação sobre a amostra está contida nos dados, e não na maneira como eles foram gerados.

Do ponto de vista da precisão, a teoria frequentista escolhe a região de rejeição da hipótese *a priori*, bem como o nível de significância do teste, o que sugere um procedimento de precisão pré-experimental. O nível descritivo aqui é usado espuriamente, pois não se enquadra na filosofia Neyman-Pearson e sim na de Fisher. Daí dizer-se que hoje em dia usa-se uma teoria híbrida para o teste usual da escola clássica. No caso Bayesiano, a inferência é feita através da distribuição *a posteriori*, a qual é condicionada aos dados obtidos, e a precisão é pós-experimental. Pretendendo fazer o mesmo caminho, ou seja, ir do particular para o geral, as escolas usam dois procedimentos diferentes: a clássica vai da população para a amostra e a Bayesiana da amostra para a população. Daí alguns Bayesianos dizerem que o único método indutivo em estatística é o seu. Um comentário de Lindley(1985) pode suportar essa opinião:

“Para uso científico, seja F uma lei geral e E um caso especial – através do teorema de Bayes passamos de uma afirmação probabilística sobre $E|F$, que geralmente é uma tarefa fácil, para uma afirmação (também probabilística) sobre $F|E$ – este teorema explica como fazer afirmações gerais com base em casos especiais”.

No teste de hipóteses clássico, a hipótese a ser testada é verdadeira ou falsa e o método não prevê probabilidade associada a ela. O que está disponível é a probabilidade condicional $P(\text{rejeitar hipótese} \mid \text{hipótese é verdadeira})$, que é um cálculo derivado da distribuição amostral, e não a probabilidade de a hipótese ser verdadeira nem a probabilidade de a hipótese ser verdadeira dado que foi rejeitada. Portanto, as únicas probabilidades disponíveis na teoria clássica se relacionam aos erros de tipo I e II, não se constituindo em grau de confirmação. Na teoria Bayesiana, são atribuídas probabilidades *a priori* às hipóteses, que são revistas à luz das evidências, produzindo as probabilidades *a posteriori*. Este método permite então calcular as probabilidades das hipóteses, que para muitos é o que todo pesquisador quer saber.

Uma das divergências cruciais entre as duas teorias é a caracterização do parâmetro sobre o qual se fará a inferência. Para o clássico, trata-se de uma característica desconhecida e fixa da população, que será estimada através da amostra. Para o Bayesiano, parâmetro, além de ser uma característica desconhecida da população, envolve uma incerteza, que precisa ser modelada por alguma distribuição de probabilidade, resultando então no conceito de que parâmetro, para esta teoria, é uma variável aleatória. Daí nasce a distribuição *a priori*, causa de grande parte da discussão entre as duas escolas. Apesar de ser possível o uso de diferentes distribuições *a priori*, acredita-se que as evidências amostrais acabam por controlar a distribuição *a posteriori*, fazendo com que, a longo prazo, as opiniões converjam. Nessa construção subjetiva reside a força mas também a fraqueza do método Bayesiano!

Nos exemplos apresentados, entendemos que é possível oferecer aos alunos de um curso introdutório as três abordagens inferenciais, tendo como pré-requisito apenas tópicos que já fazem parte do conteúdo de uma disciplina dessa natureza.

Capítulo 7 A estatística e o aluno

7.1 Introdução

Investigação científica é um processo em constante desenvolvimento e, a todo momento, novas pesquisas trazem à luz resultados muitas vezes surpreendentes, que derrubam todo um edifício construído supostamente em terra firme, erguendo um outro no lugar. Parafraseando o cançãoeiro popular, *“o resultado é verdadeiro enquanto dura”*. Muitos cientistas, apesar de conhecerem esta dinâmica, parecem acreditar numa força objetiva e definitiva que uma análise quantitativa conferiria aos dados, e muitas vezes coletam dados maciçamente, os quais ficam então à procura de uma hipótese e/ou de uma análise. O mito da análise estatística definitiva é passado geralmente aos alunos de cursos universitários, a quem geralmente é oferecido um curso de natureza determinística, como se os resultados das análises efetuadas não pudessem mais ser questionados.

De acordo com os padrões atuais, em quase todos os cursos universitários é oferecida uma disciplina básica de estatística, cuja ementa atual segue a tradição americana dos anos 60 (século XX), época em que foi muito valorizada a busca de uma base científica e tecnológica do conhecimento. Hoje podemos reconhecer outras necessidades para esta inclusão e, dentre elas, a mais imediata seria o consumo das estatísticas disponíveis nos meios de comunicação pelo próprio cidadão, que deveria estar preparado para o aproveitamento inteligente e independente das informações quantitativas. Além da ciência e da tecnologia, os alunos dedicados a profissões ligadas aos negócios, à indústria e ao governo, deveriam também ter sua formação contemplada com as ferramentas e com o raciocínio probabilístico/estatístico.

Os alunos de um curso superior, geralmente sem a formação inicial prévia na área de probabilidade e estatística, quase sempre desconhecem a razão pela qual a disciplina de estatística está em seu currículo, uma vez que, não tendo preocupações científicas e/ou

tecnológicas e por estarem no início da sua formação (a disciplina de estatística é geralmente oferecida no início da grade curricular), não vêem como essa disciplina possa vir algum dia a lhes ser de utilidade. Isto poderia ser mudado logo na apresentação da disciplina, desde que fosse prática usual integrá-la com outras disciplinas do curso. Entretanto, como geralmente os cursos são divididos por disciplinas “não-comunicantes”, o aluno não tem a oportunidade dessa visão global da aprendizagem como um processo. Além disso, como o material apresentado está centrado grande parte em exemplos com jogadas de dados, moedas ou sorteios de bolas em urnas, mais simples para o professor e mais disponíveis nos textos, embora sem qualquer apelo para os alunos, a distância entre estes e a disciplina vai se tornando intransponível em grande parte dos casos, e é responsável, em boa parte, pela repetência, desistência e evasão.

Começar pelos exemplos mais fáceis para compreender a técnica pela técnica pode acarretar aprendizagem mais rápida, mas o professor deve incluir também exemplos motivadores, reais e atuais, e mostrar o caminho da transferência de conceitos e/ou técnicas, nem sempre tão evidentes quanto se supõe no ensino tradicional. Os estudantes quase sempre precisam da orientação do professor para fazer identificações corretas: nem o conhecimento abstrato e nem o instrumental são suficientes, havendo necessidade da introdução de conexões que façam sentido para o aluno. Segundo Bastien (in Morin, 2000),

“a evolução cognitiva não caminha para o estabelecimento de conhecimentos cada vez mais abstratos, mas, ao contrário, para sua contextualização”.

Até aqui este trabalho se dispôs a relatar aspectos mais ligados à epistemologia, no que concerne à controvérsia entre as escolas clássica e Bayesiana, sugerindo, por meio da sequência dos capítulos, um caminho lógico a ser seguido por aqueles que se iniciam no tema e na polêmica dos métodos inferenciais. Pretendemos acrescentar, neste capítulo, alguns comentários ligados à parceria aluno/professor, em várias dimensões, procurando esclarecer algumas dificuldades normalmente apresentadas pelos alunos, responsáveis,

em grande parte, pelo mito que se criou em relação à disciplina introdutória de probabilidade e estatística nos cursos superiores.

7.2 Estatística e aprendizagem

Partindo do princípio de que o conhecimento dos professores seria facilmente transferido aos alunos, como se as mentes destes últimos fossem semelhantes a “baldes vazios”, que seriam pouco a pouco completados, as disciplinas de estatística foram montadas inicialmente com muita ênfase na memorização, e com um nível teórico incompatível com uma primeira abordagem ao assunto. O conhecimento, que em sua gênese é um processo carregado de incerteza, tornou-se, no ensino universitário, um produto pronto, sem levar em conta os processos de sua aquisição, perdendo grande parte da sua dinâmica e significado para o aluno.

Tal “filosofia” de aprendizagem discutida no parágrafo anterior já está se tornando obsoleta e, na opinião de Machado (1995), a idéia de aquisição de conhecimento foi substituída, primeiro pela idéia de cadeia (elos entrelaçados), numa concepção cartesiana do ensino do mais simples ao mais complexo e, agora, a partir da última década, pela idéia de rede:

“... a idéia de conhecimento como rede de significados parece ser o novo paradigma para a concepção de conhecimento.”

Essa visão é reforçada pelo comentário de Box (1983), um estatístico aplicado da escola norte-americana, para quem o cérebro pode ser separado em dois hemisférios e, na maioria das pessoas, o esquerdo é prioritariamente responsável pelas atividades da linguagem e da lógica, e o direito pelas imagens, padrões e processos *indutivos*, isso tudo realizado através de milhões de conexões neurais (ou neuronais, como defende Carvalho, 2001). Imagina-se que haja cooperação entre os dois hemisférios, e a habilidade em

analisar dados, prerrogativa do hemisfério direito, seria transferida para o conhecimento no hemisfério esquerdo, trazendo mais elementos para a descoberta científica.

Aprender é um processo ativo, resultado de várias interações, como por exemplo a do professor / aluno, mas também aluno / aluno, aluno / sociedade, aluno / meios de comunicação, etc., e a concepção de rede, com vários nós, de onde partem feixes de relações parece estar, de fato, mais de acordo com o que se pensa atualmente acerca do comportamento da mente humana e suas interações com o mundo. A aprendizagem portanto se beneficia do trabalho em grupo, da equipe integrada, da interação de competências, pois o conjunto de todas as redes individuais poderia ser pensado como uma grande rede, cujos nós, virtuais, fariam a integração de todas as mentes entre si.

Devlin(1999) comenta que a interação é um elemento importante na aquisição do conhecimento e que alguns profissionais adquirem 2/3 de seu conhecimento através da interação direta com outras pessoas e 1/3 através de meios de comunicação, incluindo aí material escrito. Essas relações, que fazem parte do cotidiano escolar, não podem estar desvinculadas das questões relacionadas ao conhecimento e à aprendizagem

No entanto, essa nova compreensão do processo pedagógico não foi acompanhada de imediato por uma mudança nos programas e métodos didáticos relacionados ao programa de estatística básica, por vários motivos, que incluem: milhares de livros-texto estão disponíveis em bibliotecas e livrarias de todo o mundo e outros estão sendo produzidos da forma tradicional; há muito pouco incentivo, sob o ponto de vista de produção acadêmica, para que professores universitários, mesmo que sejam competentes didaticamente, escrevam seus textos (principalmente no Brasil); credibilidade acadêmica em muitos lugares é sinônimo de aulas difíceis, obscuras e mesmo inacessíveis; os professores tendem a perpetuar o ensino que eles mesmo tiveram, pois isso demanda menos esforço; não há oportunidades para que os docentes universitários se preparem para atuar como educadores, pois isto não lhes garante *status* profissional.

Além disso a ênfase exagerada que se dá para publicações de pesquisa acadêmica, no campo de atuação específica dos professores, muitas vezes inibe o comprometimento dos mesmos com questões do ensino, e quem perde com isso é a educação como um todo e o aluno em particular.

Não basta apenas ensinar bem. É preciso também que o aluno aprenda e isto acontecerá somente quando o objeto de aprendizagem fizer sentido para este aluno. A compreensão dos conceitos em uma disciplina de estatística seria muito facilitada se os professores também estivessem preocupados com a aprendizagem, a cada momento da ação didática, que seria certamente favorecida pela inclusão de projetos ligados à área de interesse profissional do curso. Entendemos que essa preocupação com a aprendizagem está diretamente ligada à discussão dos aspectos fundamentais associados às diferentes escolas de inferência estatística numa disciplina introdutória.

7.3 Estatística e cognição

Conhecer como pensam os indivíduos e como se desenvolve a inteligência humana, são os objetivos de uma área que cresceu na segunda metade do século XX: a *psicologia cognitiva*. Até então havia muita especulação sobre o tema, considerado mais de natureza filosófica do que científica. Entendemos que tanto ensino em geral como ensino de elementos de probabilidade e estatística em particular não podem desconsiderar essa área, ligada diretamente ao estado de compreensão dos alunos. Muitas vezes os professores estão preocupados somente com os aspectos didáticos ligados à apresentação de uma disciplina: como transmitir conhecimento? Esta é uma preocupação legítima, mas que, sozinha, tende a aumentar a assimetria existente entre o aluno e o professor. Morin (2000) afirma:

"A comunicação não garante a compreensão. A informação, se for bem transmitida e compreendida, traz inteligibilidade, condição primeira necessária, mas não suficiente, para compreensão."

A psicologia cognitiva procura entender como se processa a informação na mente humana, como se dá a compreensão, o raciocínio. O raciocínio condicional aparece com muita frequência na disciplina de probabilidade e estatística e é considerado como um dos responsáveis pelas dificuldades apresentadas pelos alunos. Especificamente, no tópico de probabilidades, ele é invocado para o estudo de probabilidades condicionais (incluindo o teorema de Bayes) e independência, continuando no tópico de estatística inferencial, para considerar probabilidades associadas às observações dada a hipótese (estatística clássica) ou probabilidade da hipótese dadas as observações (estatística Bayesiana). A controvérsia entre as escolas clássica e Bayesiana pode ser pensada nesse contexto: o raciocínio natural do ser humano segue a perspectiva Bayesiana? Ou esta é muito mais sofisticada do que os elementos usuais utilizados no raciocínio condicional? Qual o tipo de inferência que deve ser ensinado em cursos introdutórios? As técnicas estatísticas estão à nossa disposição para que tiremos conclusões sobre populações a partir de amostras, ou seja, conclusões do particular para o geral. Que tipo de conclusão pode ser tirada a partir de tais técnicas, tanto em uma como em outra escola?

Um dos problemas crônicos na área do ensino dessa disciplina é que, por muito tempo, somente o aspecto lógico era levado em conta e havia uma discriminação automática em relação aos alunos que não apresentavam, de imediato, a resposta *lógica* e *certa* a determinada questão proposta pelo professor. Com o aparecimento da ciência cognitiva (ocorrido em 1976, segundo Anderson, 1995), uma área que procura integrar psicologia, linguística, lógica, inteligência artificial, para *conhecer como se conhece*, outros fatores deverão ser contemplados na relação professor/ aluno/disciplina de forma a compreender quais deles podem influenciar a recepção de informações e a sua compreensão pelos alunos. Assim, as sub-áreas da ciência cognitiva, bem como suas inter-relações, devem ser levadas em conta no processo educacional.

7.3.1 Modelos mentais

De acordo com Birnbaum (1990), os pesquisadores da área cognitiva ficam surpresos como o fato dos homens terem sido racionais o bastante para inventarem a teoria da probabilidade mas não racionais o suficiente para usá-la no raciocínio diário. Segundo Anderson (1995) os indivíduos geralmente usam *modelos mentais* que não são necessariamente modelos formais: eles constroem um modelo específico de uma situação e determinam o que é verdadeiro para aquela específica situação. Ele cita inclusive um exemplo recente: um dos prováveis erros no desastre de Chernobyl foi o fato de os engenheiros nem pensarem, por muitas horas, na possibilidade de que o reator tivesse vazamentos! Modelos mentais são às vezes montados em cima de heurísticas, que são estratégias do pensamento que aceitam um certo risco de erro, mesmo que tenham a eficiência como objetivo.

Entendemos que os professores deveriam levar em conta não só as respostas que os alunos oferecem às questões mas também os modelos mentais utilizados para esse fim. Nem sempre o modelo espontâneo leva à resposta correta e o aluno deveria estar ciente disso e estar preparado para reformulá-lo.

Normalmente os professores não agem desta forma com respeito aos alunos, o que certamente prejudica a sua compreensão sobre os mecanismos de raciocínio por eles apresentados, em particular no campo da probabilidade e estatística. A literatura em psicologia cognitiva apresenta diversos estudos que ligam o funcionamento cognitivo a situações de incerteza, o que é certamente uma contribuição importante para o ensino. Alguns deles mostram que modelos combinatórios ou modelos lógicos são raramente utilizados de modo espontâneo, embora estejam disponíveis na mente da maioria dos indivíduos. A esse respeito, Lecoutre (1992) afirma ter ativado os modelos corretos dos indivíduos através de estratégias pedagógicas, como por exemplo encobrindo o aspecto probabilístico associado ao problema e dirigindo a atenção do aluno para uma construção ou até mesmo para um raciocínio lógico.

Uma pergunta diretamente ligada ao tema de nosso trabalho e que os interessados na ligação entre a psicologia cognitiva e a teoria de probabilidade procuram responder é

se os indivíduos se comportam espontaneamente como Bayesianos em questões de inferência. Segundo Lecoutre (1990) parece que na maioria dos casos os indivíduos divergem do comportamento Bayesiano ideal: são mais conservativos ao darem um peso indevido às probabilidades iniciais, com respeito às novas informações. Entretanto Anderson (1995) afirma que, embora os julgamentos probabilísticos conscientes dos indivíduos não correspondam ao teorema de Bayes, seu comportamento indica que eles o utilizam inconscientemente. Isto ocorreria porque, em sua opinião, os sujeitos nem sempre são capazes de tornar explícito o seu comportamento implícito.

Uma discussão profícua sobre a maneira de os indivíduos atribuírem probabilidades a diferentes situações começou com o artigo de Tversky e Kahneman (1974), citado inúmeras vezes em artigos ligados à interface entre cognição e probabilidade. Eles mostraram que as pessoas se baseiam em princípios heurísticos que reduzem a complexa atribuição de probabilidades a operações de julgamento mais simples e concluíram que

“essas heurísticas são muito úteis, mas às vezes nos levam a erros graves e sistemáticos.”

Algumas idéias errôneas sobre probabilidade, pré-concebidas, fortemente arraigadas, muitas vezes frutos de falhas da própria intuição, ou produto ainda da intuição primária, são muitas vezes difíceis de serem reformuladas. Conceitos errados, se não corrigidos a tempo, acompanham o indivíduo por toda a vida, influenciando negativamente até mesmo suas tomadas de decisão no campo profissional. Até os professores podem cair na armadilha de sua própria intuição, podendo influenciar negativamente os alunos. Nesse sentido nunca é demais lembrar uma frase de Shaughnessy (1992):

“É preciso primeiro resolver o problema das concepções erradas dos professores antes de esperar que eles sejam competentes para evitar as concepções erradas dos alunos”.

7.3.2 Intuições e suas distorções

Segundo Fischbein (1994), *intuição* é um tipo especial de cognição, caracterizada pela auto-evidência e pela tempestividade. Isto porque ela aparece no indivíduo de maneira subjetiva, sem necessidade de prova ou de apoio empírico. As intuições ou se desenvolvem nos indivíduos como efeito de sua própria experiência de vida, independentemente de instrução (intuições primárias), ou são criadas por algum mecanismo ligado a instrução (novas intuições ou intuições secundárias). Elas representam poderosos componentes de qualquer atividade mental produtiva, cujo impacto deve ser levado em conta em todo processo educativo. Muitas vezes os alunos apresentam dificuldades de aprendizagem devido a estratégias intuitivas inadequadas, e cabe ao professor ajudar o aluno para a sua conscientização e correspondente correção.

O artigo de Tversky e Kahneman, mencionado na seção anterior, descreve algumas distorções ou vieses advindos de heurísticas inadequadas e Lecoutre e Fischbein (1998) ampliam o conjunto dessas distorções, na expectativa de melhor conhecimento do terreno intuitivo dos indivíduos, dos modelos cognitivos disponíveis e

"assim definir um modo privilegiado de ensino, apoiando-se mais em representações probabilísticas intuitivas do que em pressupostos teóricos cujo caráter abstrato nem sempre garante a disponibilidade."

O professor de probabilidade e estatística tem que estar consciente de um fato que inibe a espontaneidade dos alunos em situações em que tem que ser invocado o raciocínio probabilístico: os erros cometidos são embaraçosos, porque discriminatórios, e parecem ser inevitáveis. Segundo Kahneman e Tversky (1982)

“erros de raciocínio são muitas vezes desconcertantes – ou porque a solução não encontrada aparece a seguir como sendo óbvia ou porque o erro cometido é atraente embora a resposta seja sabidamente equivocada.”

A conscientização dos vieses que serão discutidos a seguir certamente trará benefícios para a relação professor/ aluno, no contexto de uma disciplina introdutória de probabilidade e estatística. Conhecer um pouco mais sobre modelos mentais e representações intuitivas dos alunos ligadas aos conceitos do raciocínio probabilístico/estatístico é essencial para o desenvolvimento pleno de uma atividade didático-pedagógica.

Vamos a seguir descrever alguns desses vieses que podem encontrar-se diretamente ligados a tópicos de uma disciplina introdutória de estatística:

- **viés da representatividade –**

refere-se à dominância da informação individual sobre o raciocínio probabilístico – por exemplo, suponha-se que o ambiente em que foi proposto o problema da meningite (capítulo 5) estivesse psicologicamente influenciado pela divulgação da existência de casos de meningite; se um aluno, suspeitando que tem meningite, fizer o teste (cuja probabilidade de detectar meningite dado que o sujeito tem meningite é de 95%) e se o resultado der positivo, seria natural ele se desesperar? A resposta, neste contexto, é afirmativa: o aluno se desespera porque vai desconsiderar a taxa de incidência na população (que o exemplo descreve como sendo 0,5%) e vai associar – erroneamente – a probabilidade de 95% como sendo sua chance de estar com meningite. Em outras palavras, priorizou uma informação individual que, do ponto de vista do problema, era não informativa, em detrimento da taxa basal da doença;

- **viés da ilusão do jogador –**

este viés tem duas vertentes em que podemos reconhecer o comportamento dos alunos :

- uma sequência de três caras em três lançamentos independentes de uma moeda honesta praticamente confere (erroneamente) ao próximo lançamento da mesma moeda o resultado coroa (com certeza!) – aqui o jogador se ilude pensando que uma alternância de resultados é mais representativa do acaso;
- uma sequência de três caras em três lançamentos independentes de uma moeda honesta praticamente confere (erroneamente) ao próximo lançamento da mesma moeda o resultado cara (com certeza!) – aqui o jogador se ilude acreditando na “estabilidade” do aleatório;
- **viés da equiprobabilidade –**
são comuns situações de sala de aula em que a resposta é 50% quando a pergunta diz respeito à probabilidade de uma entre duas situações: probabilidade de eu morrer hoje? probabilidade de uma peça ser fabricada com defeito? probabilidade de tirar um par de 6 ou um 5 e um 6 em dois lançamentos independentes de um dado? A heurística (enganosa) utilizada é de que os eventos de natureza aleatória são (sempre!) equiprováveis;
- **viés da conjunção –**
em certas situações o aluno é levado (erroneamente) a supor que a probabilidade do evento composto ($A \text{ e } B$) é maior do que a probabilidade de cada um dos eventos que o compõem. Lecoutre e Fischbein (1998) realizaram um experimento onde verificaram que este viés diminui com a idade, mas alunos universitários franceses do primeiro ano de psicologia, que fizeram parte do estudo citado ($n=385$), apresentaram uma proporção de erros nessa direção da ordem de 65%. Segundo Birnbaum (1990), em algumas situações a conjunção e significa adição (exemplo: mulheres e crianças nos salva-vidas) – como a adição é maior do que as suas partes, a falácia cometida por alguns indivíduos estaria explicada;
- **fenômeno Falk –**

este é um viés diretamente ligado às probabilidades condicionais e está ligado a situações em que é dado $P(A|B)$ e se deseja obter $P(B|A)$. No problema de bolas em urnas descrito no capítulo 5, encontramos uma ilustração: para o cálculo da $P(\text{sorteio de uma bola azul} | \text{urna A foi selecionada})$ vai-se na direção do experimento realizado – sortear uma urna e depois uma bola; no entanto, a probabilidade requerida na 4ª pergunta ($P(\text{urna A ter sido selecionada} | \text{bola azul foi selecionada})$) do mesmo exercício, vai na direção contrária à da realização do experimento, o que pode ser sintetizado pela frase de Lecoutre e Fischbein:

“inverte o raciocínio causa – efeito ou o eixo passado-futuro, o que vai de encontro à intuição.”

Neste caso há uma ativação do modelo mental de irreversibilidade do tempo que não se aplica ao problema proposto, e o aluno não vê saída para a resposta.

7.3.3 Raciocínio lógico e a tarefa de seleção de Wason

Embora os alunos possam ser treinados a usarem regras formais de lógica na sala de aula, não é essa a maneira com a qual eles pensam no cotidiano; por isso disciplinas em que o procedimento usado faz parte do senso comum são mais fáceis de serem compreendidas por uma grande quantidade de estudantes, enquanto as mais formais, que requerem uma abstração e um aprendizado específicos têm mais dificuldade. Interessante notar que um exercício onde se requer o uso de *modus tollens* numa sala de aula pode ser resolvido de modo incorreto pelos alunos, mas provavelmente uma situação contextualizada na vida real tem mais chance de sucesso. De acordo com Anderson,

“Raciocínio humano e tomada de decisão são bastante robustos no contexto de problemas da vida real”.

Sternberg (2000), a propósito da tarefa de seleção de Wason, que descreveremos abaixo, mostra que a maioria das pessoas se mostra muito à vontade com o raciocínio *modus ponens*, o mesmo não acontecendo com o *modus tollens*, pois neste último tentam negar o antecedente (P) para testar a proposição condicional, quando o que deveriam fazer seria negar o conseqüente.

A tarefa de Wason consiste na apresentação de quatro cartões sobre uma mesa e uma afirmação condicional a eles associada; os indivíduos devem indicar quais as condições de verificação da veracidade da proposição. Os quatro cartões apresentam letra de um lado e número de outro e vamos simbolizar as faces expostas da seguinte forma:

J 3 A 6.

A proposição apresentada foi

“ Se a carta tiver uma consoante (P) então o número na outra face será par (Q)”,

o que equivale à premissa condicional : Se P então Q (ou $P \rightarrow Q$).

A tarefa consiste em pedir que o indivíduo vire o menor número possível de cartas para verificar a veracidade da proposição. A primeira tentativa de verificação mais comum das pessoas é através do uso do *modus ponens*: As cartas com consoante serão viradas para verificar se contêm um número par. A tarefa só apresenta uma carta com consoante (J) a qual, se apresentar a face par do outro lado, confirmará a afirmação. No entanto, se for ímpar, a afirmação condicional de início não será verdadeira e não será necessária mais nenhuma verificação. Isto mostra que a carta **J obrigatoriamente** terá que ser testada.

Se o número no verso da carta J for par, então a carta com o número 3 também precisa ser examinada **obrigatoriamente**, pois ela representa a negação do conseqüente (não par) e portanto logicamente deveria apresentar do outro lado uma não consoante,

para que a afirmação inicial seja considerada verdadeira. Este último raciocínio é do tipo *modus tollens*: (Se P então Q : $\sim Q \therefore \sim P$).

Portanto somente duas cartas teriam que ser examinadas. A terceira carta não seria necessário investigar pois, sendo uma vogal ($\sim P$), nada é dito na afirmação inicial sobre o que devemos encontrar no verso. Raciocínio análogo serve para a carta de número 6. Negação do antecedente ou afirmação do conseqüente não são informações necessárias para verificar a adequação da afirmação. Segundo descrito em Sternberg, a maioria das pessoas sugere a investigação da última carta (com o número par), que consiste na verificação pela afirmação do conseqüente, o que poderia produzir uma falácia. Em verdade, não importa o que tenha na outra face a carta com o número 6. Qualquer que seja o seu conteúdo, não refutaria a proposição, pois se for consoante, ela é verdadeira e se for vogal não há elementos para refutá-la, pois esta se refere apenas às cartas que têm uma consoante. A verificação necessária e suficiente para a investigação se resume à observação de duas cartas: J (P) e 3 (não Q).

Propusemos esta tarefa para um grupo de 47 alunos do segundo ano de administração de um curso universitário, e os resultados mostraram somente 9% de acertos (J e 3) e, dentre outras combinações, houve 47% de sugestões para verificar a carta 6. De fato, o índice de acertos foi baixo, já esperado de acordo com a descrição de Wason, embora ele tivesse obtido quase o dobro de acertos para esta situação (ainda assim, baixo).

Repetimos o problema dentro de um contexto menos abstrato do que o das cartas e mais familiar ao aluno, adaptando o exemplo de venda de bebida alcoólica, descrito em Anderson (1995). Sugerimos que os alunos se imaginassem fiscais adentrando um bar para verificar se a seguinte legislação estava sendo cumprida:

“ Se uma pessoa estiver bebendo cerveja (P) então ela deverá ter mais do que 18 anos (Q)”,

e que encontraram quatro mesas ocupadas. Uma delas com um velho e uma garrafa sem rótulo, outra com um adolescente e uma garrafa também sem rótulo, a terceira com uma pessoa de costas e um refrigerante e a última com uma pessoa de costas e uma cerveja. Esta é uma situação de simulação, que faz parte do cotidiano do aluno; as mesas significam os cartões, e as informações que cada mesa tem são, na ordem,

velho	adolescente	refrigerante	cerveja.
-------	-------------	--------------	----------

A pergunta feita aos alunos foi: qual o número mínimo de mesas a serem investigadas (e quais são) para a verificar se a lei está sendo cumprida? Pelos mesmos motivos descritos no exemplo abstrato das cartas, a verificação necessária e suficiente se dará investigando a mesa com a **cerveja** (P) e a mesa com o **adolescente** ($\sim Q$). Nesta situação o índice de acertos aumentou consideravelmente. De fato, o problema foi proposto ao conjunto de alunos anterior (com um acréscimo de 6 pessoas) e o resultado mostrou uma porcentagem de acertos (77%) relativamente próxima aos 87% obtido por Wason e nitidamente superior aos 9% anteriores, mostrando a importância do contexto no raciocínio dos alunos (os dois exercícios foram propostos com intervalo de uma semana e não foi feito nenhum comentário do primeiro antes de apresentar o segundo).

Machado(1990) apresenta os quatro tipos de formulação da tarefa de Wason, a saber: a) abstrato; b) concreto arbitrário; c) concreto com simulação; d) concreto realista. As duas formulações descritas em nossa execução desta tarefa com os alunos podem ser consideradas como sendo, respectivamente, formulação abstrata e formulação concreta com simulação. Esta última, contextualizada, apresentou um resultado muito superior à situação puramente abstrata (na qual predominou o raciocínio não lógico por parte dos alunos), o que indica que os indivíduos têm implicitamente o raciocínio lógico, mas que só o explicitam mediante um contexto menos abstrato u que lhe pareça mais familiar. Machado examina as diferentes dimensões do concreto associadas à tarefa de Wason e classifica a formulação c) segundo a dimensão do concreto *conteúdo de significações*, que tem um papel preponderante na atuação dos indivíduos, o que justificaria a superioridade de acertos nessa situação.

Pode haver também um confundimento da razão pela qual o número de acertos foi sensivelmente maior no segundo exemplo. Ou o problema contextualizado, por si só, já dá ao indivíduo maior chance de chegar à resposta correta, ou a contextualização que inclui a busca por uma violação legal (Anderson, 1995), é um estímulo maior para raciocinar na direção correta.

Além da existência da dificuldade do uso do raciocínio *modus tollens* no raciocínio lógico, como vimos, há também um problema em sua generalização para a inferência estatística clássica. Segundo Batanero (2000) existe um viés conhecido como a ilusão da prova probabilística por contradição – *modus tollens* – descrito como a crença errônea de que a hipótese nula tornou-se *improvável* ao ser rejeitada. Já vimos que, apesar do interesse do pesquisador em atribuir probabilidades à sua hipótese, o raciocínio frequentista (ou clássico) não oferece essa possibilidade.

7.4 Conclusão

O raciocínio ou pensamento probabilístico-estatístico faz parte do cotidiano das pessoas, na maior parte das vezes de maneira informal: *acaso*, *chance*, mais *provável*, menos *provável*, *previsão* de resultados, *revisão* de nossas *estimativas* anteriores, etc. Em muitos casos podemos perceber que o pensamento está numa direção intuitiva ou mesmo *tácita* (sei responder mas não sei explicar). A dimensão tácita varia de indivíduo para indivíduo, e isto é percebido imediatamente numa sala de aula pelo professor mais atento: há os alunos cuja resposta às questões propostas é imediata e correta; há aqueles em que a resposta é imediata, embora não alinhada com o conteúdo em questão e, por fim, há os que, mesmo eventualmente interessados, não têm qualquer resposta, pois ou não conseguem perceber o contexto, e então o que está sendo proposto não lhes faz nenhum sentido, ou então têm uma certa inibição perante um possível erro de raciocínio.

De acordo com Devlin (2000), nós, seres humanos, funcionamos pobremente quando se trata de extrair informações de dados estatísticos. A estatística espontânea, segundo Besson (1995), é tão errônea quanto a sociologia espontânea. As distorções ocasionadas por raciocínios heurísticos e/ou intuições errôneas podem ser responsáveis por isso, trazendo também problemas para o desempenho dos alunos em uma disciplina introdutória de probabilidade e estatística.

O que se espera de um aluno que tenha cursado a disciplina de estatística é que ele, com o conhecimento de sua área de atuação, estabeleça suas teorias, colete dados e esteja preparado para analisá-los descritivamente. Além disso, seria desejável que ele possa extrair o que há de latente nos dados, a fim de produzir informação, que juntamente com um contexto e uma metodologia estatística possa auxiliar no movimento contínuo em busca do conhecimento. No ambiente de trabalho futuro em que ele irá atuar, geralmente isso não é feito de maneira isolada, mas sim com uma equipe, da qual fará parte (idealmente) um profissional diretamente ligado à estatística com o qual o novo membro irá interagir. Preparar e motivar o futuro profissional é missão também do professor de estatística, que deve, através de propostas didático-pedagógicas atraentes, conseguir o engajamento do aluno, corrigindo os vieses que ele possa apresentar ao usar indevidamente raciocínios de natureza heurística.

Capítulo 8 Considerações finais

8.1 Introdução

Existem três aspectos sob os quais podem ser consideradas as controvérsias sobre os testes de hipóteses: na estatística teórica (inferências de Fisher, Neyman-Pearson e Bayesiana), em aplicações da estatística, inclusive nas sociedades e veículos de publicação científica e na área de ensino (Ito, P.K., *in* IASE Review, novembro de 1999). A preocupação deste trabalho foi abordar a primeira vertente da controvérsia no que se refere à sua ligação com o último, ou seja, com o ensino. Procurou-se aqui incluir diversos aspectos ligados a uma compreensão inicial do problema, com o intuito de apresentar uma proposta conciliatória para a apresentação do tema inferência estatística a alunos ingressantes no terceiro grau.

Não tendo ainda inquietações nem científicas nem tecnológicas a responder, e confundindo a estatística com a matemática, os alunos têm aversão essencial a esta disciplina e cabe aos professores mudar este panorama. Ter consciência das controvérsias inerentes ao próprio tema da inferência bem como dos problemas associados aos aspectos cognitivos ligados ao raciocínio da incerteza ajudará, de maneira definitiva, a melhorar as atividades ligadas tanto ao ensino como à aprendizagem, com a correspondente desmitificação da estatística.

Ambas as abordagens, clássica e Bayesiana, trabalham com raciocínio probabilístico formal, e como vimos no capítulo anterior, os alunos tendem a apresentar falhas devidas a raciocínios heurísticos incorretos ou intuições falhas. Segundo Fischbein (1994), sem uma direta intervenção de uma informação empírica, muitos indivíduos não são capazes de, espontaneamente, confiar somente em seus esquemas lógicos para obter conclusões formais corretas. Esperar que somente uma parcela mínima de alunos entenda o tema para valorizá-la é uma atitude discriminatória e anti-ética, que não deveria ter lugar num ambiente de parceria entre professor e aluno. A valorização deve ser feita

através do oferecimento de atividades paralelas ao curso, como por exemplo, bolsas de iniciação científica, projetos temáticos, projetos integrados, estágios em laboratórios, etc. Entendemos que a sala de aula deve ser o lugar em que se privilegia a dignidade do aluno, com suas facilidades e dificuldades. De acordo com Anderson (1995), os indivíduos, embora possam ser treinados de acordo com regras formais em aulas de lógica e de estatística, não é desta forma que eles raciocinam no dia-a-dia: mesmo que eles não exibam qualquer conhecimento explícito e cometam erros quando instados a fazerem julgamentos explícitos, demonstram possuir um conhecimento implícito dos princípios Bayesianos.

8.2 Alguns contrastes entre as escolas clássica e Bayesiana

Como já vimos, uma das principais distinções entre a abordagem clássica (Neyman - Pearson) e a Bayesiana é que esta última faz uso de conhecimento *a priori*, subjetivo, sobre o parâmetro de interesse e aquela não. Muitos professores do ensino superior, acostumados a lecionar inferência pelo paradigma usual da escola clássica, têm dificuldade com a suposta noção de que o parâmetro, uma quantidade desconhecida e fixa para a escola clássica, “virou” uma variável aleatória na escola Bayesiana e eles não aceitam a possibilidade de abordar o tema para os alunos, com medo de confundi-los. O que acontece, em verdade, é que os Bayesianos modelam o seu conhecimento sobre o parâmetro através de uma distribuição de probabilidade *a priori*, e o aleatório está nessa perspectiva. Isso não era explicitado por eles, talvez propositadamente, e por muito tempo foi mantida a mística em torno do parâmetro se tornar, *como num passe de mágica*, uma variável aleatória! Por seu lado, a teoria clássica também faz suposições iniciais sobre o modelo e *controla* a maneira de obter os dados, através da amostragem ou do planejamento de experimentos.

A escola clássica tira suas conclusões usando a probabilidade da evidência amostral dada a hipótese $\{ P(X | H) \}$ e se essa probabilidade é baixa não se pode inferir que a probabilidade de H seja baixa. A teoria que nos fornece a probabilidade

diretamente sobre a hipótese é a fornecida pela escola Bayesiana, que tira suas conclusões usando a probabilidade da hipótese dadas as observações ($P(H | X)$). A conclusão que a clássica permite sobre a hipótese é sua rejeição ou não rejeição. Muitas vezes os resultados numéricos de ambas as análises coincidem (situação vista no capítulo 6), mas a interpretação é diferente, sendo fundamental que esta seja dada levando em conta a opção inferencial. Um dos erros mais comuns, que é preciso evitar, é uma conclusão Bayesiana para uma técnica clássica, como se esta última atribuisse probabilidades para as hipóteses.

Efron (1998) destaca algumas linhas de contraste entre as escolas Bayesiana e clássica no sentido filosófico e também psicológico. A elas **adicionamos** os pares, dinâmica/estática e subjetiva/objetiva:

Sintético / Analítico

O método Bayesiano é sintético pois faz combinações de todos os elementos necessários à análise para sintetizá-los na distribuição *a posteriori*.

O método clássico é analítico, pois decompõe todo o problema e o resolve parte a parte, dentro de uma perspectiva cartesiana.

Individual / Universal

O método Bayesiano é individual, pois seus resultados são baseados em opiniões individuais, e frequentemente referentes a uma previsão futura sobre um indivíduo.

O método clássico é universal, pretendendo dar respostas ao mundo científico, por meio de características globais da população.

Coerência / Optimalidade

O método Bayesiano procura coerência, uma propriedade desejável numa tomada de decisão, que se espera consistente.

O método clássico procura optimalidade de seus procedimentos, ou seja está em busca de regras ótimas.

Otimista / Pessimista

O método Bayesiano é otimista e uma frase que poderia caracterizá-lo psicologicamente seria: “*Se alguma coisa tiver que dar certo, dará*”. É considerado mais agressivo.

O método clássico é pessimista e a frase que a ele se poderia associar é “*Se alguma coisa tiver que dar errado, dará*”, sendo considerado como um método defensivo.

Dinâmica / Estática

O método Bayesiano é dinâmico no sentido de que o método sempre permite revisar suas probabilidades à luz dos dados obtidos.

O método clássico é estático (no sentido de rígido) e não faz revisões de suas conclusões, pois usa o método hipotético dedutivo, “engessado pelo espaço amostral”.

Subjetivo / Objetivo

Para o método Bayesiano a subjetividade é mais justa porque não ignora os aspectos mais relevantes da experiência particular.

Para o método clássico somente o conhecimento objetivo é verdadeiro e nos permite uma visão imparcial do mundo.

Os contrastes apresentados delimitam bem a postura de cada escola e cada item poderia ser estudado exaustivamente. Além desses contrastes, vimos também que há muitos pontos divergentes entre as duas abordagens. No interior das comunidades de pesquisa há uma separação nítida entre os adeptos de cada escola, apesar de existirem estatísticos clássicos que usam métodos Bayesianos em suas pesquisas de maneira pragmática, para efeito de comparação de resultados. Por outro lado há muitos adeptos da escola Bayesiana que ensinam a disciplina básica de probabilidade e estatística somente

sob o prisma da escola clássica, quer seja pela tradição quer seja pela ementa imposta pela unidade de ensino.

Um retrato recente dessa polêmica pode ser resumido na frase de Kass (*in* Efron, 1998):

“Apesar de minha preferência pelo pensamento Bayesiano, não vejo como o próximo século possa ser exclusivamente Bayesiano. Sinto que apesar do contraste entre frequentista e Bayesiano ser estritamente um problema lógico, pode ser pensado metaforicamente como um problema de linguagem - são duas linguagens alternativas que são usadas para trabalhar com a incerteza, com oradores fluentes de um lado podendo compreender qualquer fenômeno descrito por igualmente fluentes oradores do outro, apesar de não haver boa tradução para certas frases.”

8.3 Proposta de programa

Há muitos obstáculos a serem transpostos ao se oferecer aos alunos ingressantes na universidade noções básicas de probabilidade e estatística. Tais obstáculos podem ser ou de natureza cognitiva, como os vieses apontados no capítulo anterior, ou de natureza didática, como dificuldades de apresentação do tema pelos professores ou ainda de natureza epistemológica, relacionados aos fundamentos inerentes à disciplina, neste caso com ênfase aos aspectos ligados à inferência. Após a discussão de vários desses obstáculos em capítulos anteriores, será apresentada a seguir uma proposta para a disciplina introdutória

Habitualmente, a orientação inicial da disciplina se alterna entre começar com probabilidade ou com análise descritiva de dados. Sob um ponto de vista rígido e formal, nada há a objetar com relação ao início com probabilidade - mas se queremos de fato tornar a matéria palatável ao aluno, certamente o início com uma análise de dados é mais próximo ao seu cotidiano pois os resultados descritivos são imediatamente visíveis.

Dados reais, como por exemplo os coletados pelo IBGE, ou de áreas como finanças ou pesquisa de mercado, estão disponíveis em vários meios de comunicação e são facilmente transferíveis para um laboratório didático. Além disso, alguns periódicos de circulação diária também podem fornecer material abundante, inclusive com definições no âmbito da estatística, que até há bem pouco tempo não eram oferecidas nesse tipo de publicação.

Depois dessa motivação inicial, pode ser introduzido o tópico de probabilidade, que servirá como base para o raciocínio inferencial e que deve ser apresentado de um modo atraente, sem atrelá-lo compulsoriamente às técnicas de contagem, as quais poderão ser sugeridas oportunamente, para alunos especialmente interessados. Uma abordagem com tabelas de dupla entrada ou com diagrama de árvore para um número pequeno de ensaios, incluindo simulação em sala de aula, favorece a aprendizagem neste caso, pois diminui o peso da dimensão abstrata inerente ao tema. A seguir, a disciplina apresentaria o conceito de variável aleatória, bem como algumas distribuições (incluindo distribuições amostrais) para então introduzir a inferência estatística propriamente dita. E neste ponto cabe a pergunta: como iniciar o tópico ligado à inferência em disciplinas de natureza introdutória? Qual a escola inferencial que deve ser ensinada àqueles que tomam contato pela primeira vez com o assunto?

De início, sugerimos que a apresentação do tópico inferência não seja técnica, como ocorre na maioria dos cursos, mas que tenha um caráter epistemológico, numa orientação semelhante àquela dada nos capítulos 2 e 4, modificada convenientemente de acordo com os propósitos de cada curso. Em seguida, deve ser iniciada a abordagem inferencial propriamente dita, contemplando as escolas de Neyman-Pearson e Bayesiana, cujos tópicos seriam discutidos passo a passo, à medida que os exemplos propostos fossem sendo resolvidos. A disciplina terminaria com uma análise crítica da literatura disponível em artigos científicos ou tecnológicos inerentes à área (ou áreas) para as quais o curso é desenvolvido. Uma experiência da autora numa disciplina introdutória, mostrou uma divisão imediata entre os alunos, ao ser apresentado um exemplo de inferência numa e noutra abordagem: os mais objetivos declarando-se a favor da concepção clássica,

condenando o uso de opiniões *a priori*, e os demais francamente adeptos da escola Bayesiana, não podendo entender como prescindir da opinião do pesquisador!

A proposta apresentada favorece a discussão dos fundamentos em detrimento de aspectos técnicos usualmente privilegiados em cursos desta natureza, como por exemplo várias situações de testes de hipóteses e/ou estimação; nela os aspectos críticos de cada escola podem ser enfatizados e discutidos em sala de aula, contribuindo para a formação dos alunos, desenvolvimento de aptidões para uma real aprendizagem da Estatística .

8.4 Últimas observações

Apesar de já termos feito, neste capítulo, uma proposta de programa básico para a disciplina introdutória, gostaríamos de retomar a pergunta

qual a inferência a ser ensinada em cursos iniciais?

para respondê-la sob outra ótica, que vai ao encontro de nossa própria proposta. As respostas que oferecemos são as seguintes:

1 - Pelo critério de disponibilidade de livros-texto, seria a clássica, uma vez que as ementas tradicionalmente apresentam somente a inferência clássica. A menção da perspectiva Bayesiana se faz unicamente no capítulo de probabilidade, como aplicação de probabilidade condicional.

2 - Pelo critério de disponibilidade de professores, seria a clássica, pois a grande maioria destes tem esta formação, e muitos nem estão com a preocupação de povoar a mente com outras possibilidades.

3 - Pelo critério de disponibilidade computacional, seria a clássica, pois a oferta de aplicativos estatísticos é praticamente toda feita na perspectiva clássica, pela já comentada dificuldade numérica dos métodos mais sofisticados. Existem algumas tentativas isoladas de oferecer um ou outro texto introdutório de inferência Bayesiana,

aproveitando os aplicativos existentes e oferecendo ao leitor um disquete com “macros” apropriadas a cada caso estudado.

4 - Pelo critério do número de análises em artigos publicados nas diferentes áreas aplicadas, seria a clássica, dada a grande representatividade nos artigos publicados.

5 - Pelo critério de conforto do pesquisador (ou do aluno), seria a clássica, pois o comprometimento termina depois da coleta dos dados, deixando à estatística a tarefa de analisar e tirar conclusões. O conforto aqui diz respeito à segurança de que duas pessoas com os mesmos dados chegam às mesmas conclusões (usando as mesmas premissas), sem se comprometerem com o caráter subjetivo das distribuições *a priori*.

6 - Pelo critério do engajamento, seria a Bayesiana, que pressupõe um envolvimento maior do pesquisador ou do aluno, desde o estabelecimento da distribuição *a priori* até a discussão sobre a distribuição *a posteriori*, e a eventual revisão, num processo dinâmico e contínuo. Sob estes aspectos podemos considerar a escola clássica como sendo muito rígida, enquanto que a Bayesiana é mais flexível.

7 - Pelo critério de atender à pergunta do pesquisador, ou seja, fornecer qual a probabilidade da sua hipótese, seria a Bayesiana, uma vez que a clássica não permite uma resposta desse tipo (apesar de muitas pessoas interpretarem o resultado clássico desse modo!).

8 - Pelo critério de facilidade de compreensão do aluno, diríamos que ambas têm certas dificuldades: a clássica pelos problemas filosóficos que enfrenta, principalmente ligados ao raciocínio por contradição, e por não ser intuitiva, não fornecendo as probabilidades das hipóteses; a Bayesiana, apesar de mais intuitiva em seu encaminhamento, por ser fortemente dependente de probabilidades condicionais (tópico de difícil entendimento pelos alunos iniciantes) bem como de razoável conhecimento e traquejo com distribuições de probabilidades.

9 - Pelo critério de compromisso com todas as fontes produtoras do conhecimento, deveríamos ensinar num curso introdutório as duas vertentes, por todos os motivos expostos acima, usando somente exemplos simples, para caracterizar cada concepção. Isto nos faria (a nós professores) mais responsáveis perante os alunos, no sentido de lhes oferecer as ferramentas disponíveis para que eles iniciem sua vida profissional escolhendo essa ou aquela orientação, de acordo com sua filosofia, inclinação, área de atuação, etc. Inclui-se aí a perspectiva do *ecumenismo*, em que as duas escolas poderiam coexistir.

Quantificando as respostas acima, alguém poderia argumentar a favor da escola clássica, uma vez que nos quesitos propostos ela seria a escolhida na maioria das situações. Entretanto, isto se dá justamente pelo fato de a estatística clássica ser o paradigma atual: rompimentos ocorrem raramente, principalmente no mundo acadêmico que é (paradoxalmente) fortemente conservador. Além disso o peso que sugerimos dar à última resposta supera em muito a soma dos pesos de todas as outras, o que favorece proposta que já havíamos apresentado na seção 8.3 ou seja, da coexistência das duas abordagens num curso introdutório.

Esta proposta que envolve a apresentação das duas escolas ao aluno é *neutra*, no sentido de não colocar diante do aluno qualquer partido tomado: apresentamos o estado da arte da área, como faria, por exemplo, um professor de ciência política, ao apresentar várias correntes, capitalismo, socialismo, etc., pois elas coexistem enquanto formas de ação política, e o professor não tem que se posicionar perante o aluno por uma ou outra mesmo que ele seja militante em uma delas. De modo análogo, entendemos que a estatística Bayesiana não é um “ramo” da estatística geral, mas sim algo a parte, que poderia ser encarado, como já foi dito, como um novo paradigma.

Como último argumento a favor da coexistência, lembramos o que já foi comentado neste trabalho a respeito da fragmentação do ensino. Hierarquicamente, assistimos à fragmentação do saber (o científico e o humanístico), à fragmentação das áreas em cursos universitários, a fragmentação dos cursos em disciplinas. Fragmentar o

universo da disciplina de estatística (em vez de mostrar seu “estado da arte”) não nos parece a melhor opção. Fragmentação talvez tenha favorecido alguns avanços em tecnologia mas, como disse Morin (2000),

“produziu nova cegueira para os problemas globais, fundamentais e complexos, e esta cegueira gerou inúmeros erros e ilusões, a começar por parte dos cientistas, técnicos e especialistas”.

Achamos que a proposta deste trabalho, retratada pela reta simbolizada pela coexistência, ajudará os professores de Estatística a favorecerem que os alunos possam apreender “o que está tecido junto”!

* com a escola clássica

** à escola Bayesiana

A ação de trazer para o campo pedagógico a tensão entre as escolas clássica e Bayesiana, vem ao encontro dos objetivos da Educação, entendida como um processo em construção, um movimento que prepara o espírito para a crítica. As contradições que

possam advir dessa prática da coexistência objetivam estender a prática pedagógica a uma dimensão que extrapole os objetivos internos à escola, projetando as expectativas da formação do aluno para uma sociedade em constante movimento. Explicitar o compromisso da convivência com o conflito, através dos caminhos oferecidos pela Educação, pode ser uma das poucas propostas viáveis.

I

Referências bibliográficas

- Albert, J. H. (1996). Bayesian computation using Minitab. Duxbury Press. Belmont.
- Anderson, J.R. (1995). Cognitive psychology and its implications. Freeman. New York.
- Barnett, V. D. (1973). Comparative statistical inference. Wiley. Chichester.
- Batanero, C. (2000). *Controversies around the role of statistical tests in experimental research*. Mathematical Thinking and Learning 2(1&2).
- Berry, D. (1997). *Teaching elementary Bayesian statistics with real applications in science*. The American Statistician, 51(3) (p. 241-246).
- Besson, J.L. (1995). A ilusão das estatísticas. Editora UNESP. São Paulo.
- Beveridge, W.I.B. (1981). Sementes da descoberta científica. EDUSP. São Paulo.
- Birnbaum, M.H., Anderson, C.J. and Hynan, L.G. (1990). *Theories of bias in probability judgment*. In: Cognitive Biases (p. 477-498). Caverni, J-P. et al. Eds. Elsevier Science Publishers B.V.
- Blackwell, D. (1969). Basic statistics. McGraw-Hill Book Co. New York.
- Bombassaro, L.C. (1992). As fronteiras da epistemologia. Vozes. São Paulo.
- Botter, D. ; Paula, G.; Leite, J.G. e Cordani, L.K. (1996). Noções de estatística. IME USP. São Paulo.
- Boudot, M. (1972). Logique inductive et probabilité. Librairie Armand Colin. Paris.
- Box G.E.P , Leonard, T. and Wu C-F., eds. (1983). Scientific inference, data analysis, and robustness. Academic Press, Inc. New York.
- Bunge, M. (1980) Ciência e Desenvolvimento. EDUSP. São Paulo.
- Bunge, M. (1980). Epistemologia. EDUSP. São Paulo.
- Carvalho, L.A.V. (2001). Datamining: a mineração de dados no marketing, medicina, economia, engenharia e administração. Editora Érica. São Paulo.
- Chalmers, A F (1982). O que é a ciência afinal? Ed. Brasiliense. São Paulo.

- Cohen, M. R. (1945,1992). *Introducción a la lógica*. 2ª ed. Fondo de Cultura Económica, México.
- Copi, I. M. (1968) . *Introdução à Lógica*. 2ª ed. Editora Mestre Jou. São Paulo.
- Cox, D.R. 1997). *The Current Position of Statistics: A Personal View*. International Statistical Review, v. 65 (3) (p. 261-290).
- Da Costa, N. C. A. (1993). *Lógica indutiva e probabilidade*. Ed. Hucitec e EDUSP. São Paulo.
- DeGroot, M.H. (1986). *A conversation with David Blackwell*. Statistical Science, vol 1(1) (p. 40-53).
- Devlin, K. (1991). *Logic and information*. Cambridge University Press. Cambridge.
- Devlin, K. (1999). *Infosense - Turning information into knowledge*. W.H.Freeman and Company.
- Droesbeke, J.J. et Tassi, P. (1997). *Histoire de la Statistique*. Presses Universitaires de France. Paris.
- Earman, J. (1996) *Bayes or Bust?* MIT Press. Cambridge.
- Efron B. (1998). *R. A. FISHER in the 21st Century*. Statistical Science, v. 13(2) (p. 95-122).
- ENEM (2000). *Exame Nacional do Ensino Médio*. MEC. Brasília.
- Feller, W. (1968). *An introduction to probability theory and its applications*. Wiley.
- Ferreira, A.B.H. (1988). *Dicionário Aurélio Básico da Língua Portuguesa*. Ed. Nova Fronteira S/A.
- Feyerabend (1975) *Against Method*. New Left Books. London.
- Fischbein, E. (1994). *Intuition in science and mathematics*. D. Reidel Publ. Co. Boston.
- Fisher, R. A . (1953). *The design of experiments*. Oliver and Boyd. London.
- Frawley, W. (1997). *Vygotsky e a ciência cognitiva*. Artmed Ed. Porto Alegre.
- Gal, I. , and Garfield, J.B. eds. (1997). *The assessment challenge in Statistics Education*. IOS Press. Amsterdam.
- Gale, G. (1979). *Theory of science*. McGraw Hill Inc. New York.

- Gelman, A. , Carlin, J.B. , Stern, H.S. and Rubin, D.B. (1995). *Bayesian Data Analysis*. Chapman & Hall. London.
- Gigerenzer, G. et al. (1989). *The empire of chance*. Cambridge University Press. Cambridge.
- Gigerenzer, G. (1992). In: *Companion Encyclopedia of the history and philosophy of the mathematical sciences*. Grattan-Guinness ed. Routledge Inc.
- Good, I.J. (1985). *Scientific Method and Statistics*. In *Encyclopedia of Statistical Sciences*, v. 8 (p. 291-304). Wiley. New York.
- Good, I. J. (1989) *The interface between statistics and the philosophy of science*. In: *Logic, Methodology and Philosophy of Science*, VIII. Fenstad, J. et al. eds. Elsevier Science Publishers. Amsterdam.
- Hartigan, J. (1983). *Bayes theory*. Springer. New York.
- Hegenberg L. (1976). *Etapas da investigação científica*. EDUSP. São Paulo.
- Hoel, P. G. (1962). *Introdução à Estatística Matemática*. Editora Atlas Ltda. São Paulo.
- Horgan, J. (1998). *O fim da ciência*. Companhia das Letras. São Paulo.
- Howson, C. and Urbach, P. (1993). *Scientific reasoning: the Bayesian approach*. 2^a. ed. Open Court. Chicago.
- Hume, D. (1999). In: *Os Pensadores*. Editora Nova Cultural Ltda. São Paulo.
- Jungermann, H. and Thüring, M. (1993). *Causal knowledge and the expression of uncertainty*. In: *The cognitive Psychology of Knowledge* (p. 53-73). Strube G. and Wender K.F. Eds. Elsevier Science Publishers B.V.
- Kahneman, D. and Tversky, A. (1982). *On the study of statistical intuitions*. In: *Judgment under uncertainty heuristics and biases* (p. 493-508). Kahneman, D, Slovic, P. and Tversky, A. Eds. Cambridge University Press. Cambridge.
- Kant, I. (1999). *Vida e Obra (Os Pensadores)*. Editora Nova Cultural Ltda. São Paulo.
- Kendall, M.G. and Stuart, A. (1969). *The Advanced Theory of Statistics*, v. 1. 3^a ed. Charles Griffin and Co. London.
- Kneebone, G.T. (1963). *Mathematical logic and the foundations of mathematics*. Van Nostrand.

- Kuhn, T. S. (1975). A estrutura das revoluções científicas. Editora Perspectiva. São Paulo.
- Lacey, H. (1998). Valores e atividade científica. Discurso Editorial. São Paulo.
- Lakoff, G. e Johnson, M. (1980). Metaforas de la vida cotidiana. Catedra.
- Lambert, K. e Brittan Jr, G. G. (1972). Introdução à filosofia da ciência. Ed. Cultrix.
- Laplace, P.S. (1985). Ensayo filosófico sobre las probabilidades. Alianza Editorial, Madrid.
- Larousse Cultural, (1998). Grande Enciclopédia . Nova Cultural Ltda. São Paulo.
- Lecoutre, B. and Poitevineau, J. (2000). *Aller au-delà des tests de signification traditionnels: vers de nouvelles normes de publication*. L' année psychologique, 100 (p. 683-713).
- Lecoutre, M.P. and Durand, J-L. (1988). *Jugements probabilistes et modèles cognitifs: étude d' une situation aléatoire*. Educational Studies in Mathematics, 19 (p.357-368).
- Lecoutre, M.P. , Durand, J-L. and Cordier, J. (1990). *A study of two biases in probabilistic judgements: representativeness and equiprobability*. In: Cognitive Biases (p. 563-575). Elsevier Science Publishers B.V.
- Lehrer, K. (1990). Theory of knowledge. Westview Press. Colorado.
- Lévy P. (1999). A inteligência coletiva . Edições Loyola. São Paulo.
- Lindley, D.V. (1965). Introduction to Probability and Statistics: from a Bayesian viewpoint. Cambridge University Press. Cambridge.
- Lindley, D.V. (1985). Making decisions. J. Wiley and Sons. London.
- Lindley, D.V. (1990). *The 1988 Wald Memorial Lectures: The present position in Bayesian Statistics*. Statistical Science, v. 5 (1) (p. 44-89).
- Longman Group Ltd. (1995). Dictionary of Contemporary English. Longman Group Ltd. Suffolk.
- Lopes, A. E. C. O. (1988). A estatística e sua história: uma contribuição para o ensino da estatística aplicada à educação. Dissertação de mestrado. PUC. São Paulo.
- Machado, A. (1979). Poesías Completas. 11ª ed. Espasa-Calpe, S.A. Madrid.
- Machado, N.J. (1990). Matemática e língua materna. 4ª ed. Cortez Editora. São Paulo.

- Machado, N.J. (1995). Epistemologia e didática. 3ª edição. Cortez Editora. São Paulo.
- Machado, N.J. (1997). Cidadania e educação. Escrituras Editora. São Paulo.
- Marcondes, D. (1997). Iniciação à história da filosofia. Jorge Zahar Editor.
- Mayo, D.G. (1996). Error and the growth of experimental knowledge. The University of Chicago Press. Chicago.
- Moles, A. (1998). A criação científica. Editora Perspectiva. São Paulo.
- Moles, A (1990, 1995) As ciências do impreciso. Civilização Brasileira. São Paulo.
- Moore, D.S. (1997a). *Bayes for beginners? Some reasons to hesitate*. The American Statistician, v. 51(3) (p. 254-261).
- Moore, D.S. (1997b). *New pedagogy and new content: the case of statistics*. International Statistical Review, 65 (2) (p. 123-165).
- Morin, E. (2000). Os Sete Saberes necessários à Educação do Futuro. Cortez Editora. São Paulo.
- Neyman, J. ; Pearson, E.S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. Part I. Biometrika, XX ^ (p. 175-240).
- Oakes, M. (1986). Statistical Inference. John Wiley and Sons. New York.
- Paulos, J. A. (1998). Once upon a number. The Penguin Press.
- PCN (1998). Parâmetros Curriculares Nacionais. Ensino Fundamental. MEC. Brasília.
- PCN (1999). Parâmetros Curriculares Nacionais. Ensino Médio. MEC. Brasília.
- Peirce, C.S. (1999). Semiótica. Editora Perspectiva. 3ª ed. São Paulo.
- Plackett, R.L. (1966). *Current trends in statistical inference*. J.R.S.S. A, 129 (p. 249-267).
- Polya, G. (1968). Patterns of plausible inference. University Press. Princeton.
- Popper, K. (1975). A lógica da pesquisa científica. Cultrix & EDUSP. São Paulo.
- Prosch, H. (1986). Michael Polanyi: a critical exposition. Suny Series in Cultural Perspectives.

- Reisberg, D. (1997). *Cognition*. W W Norton and Co. New York.
- Rivadulla, A. (1991). *Probabilidad e inferencia científica*. Anthropos. Barcelona.
- Salmon, W.C. (1967). *The foundations of scientific inference*. U. of Pittsburgh Press. Pittsburgh.
- Santos, B. S. (1989). *Introdução a uma ciência pós-moderna*. Ed. Graal Ltda. Rio de Janeiro.
- Savage, L.J. (1972). *The foundations of statistics*. Dover Publications, Inc. New York.
- Shaughnessy, J. M. (1992). *Research on Probability and Statistics: Reflections and directions*. In: Grouws, D.A. (ed.) *Handbook of research on mathematics teaching and learning*. Macmillan. New York.
- Sternberg, R.J. (2000). *Psicologia cognitiva*. Artmed. Porto Alegre.
- Stigler, S.M. (1999). *The foundations of statistics at Stanford*. *The American Statistician*, vol 53(3) (p. 263-266).
- Thatcher, V. S. ed. (1970). *The New Webster Encyclopedic Dictionary of the English Language*. Consolidated Book Publishers. Chicago
- Tversky, A. and Kahneman, D. (1974). *Judgment under uncertainty: heuristics and biases*. *Science*, v. 185 (p. 1124-1131).
- Vallecillos J., A . (1996). *Inferencia estadística y enseñanza: un análisis didáctico del contraste de hipótesis estadísticas*. Editorial Comares. Granada.
- Watzlawick, P. (1988). *L'invention de la réalité*. Editorial.
- Watzlawick, P. e Krieg, P. eds. (1995). *O olhar do observador*. Editorial Psy II. São Paulo.
- Wright, G. and Ayton, P (1990). *Biases in probabilistic judgment – a historical perspective*. In: *Cognitive Biases*. Caverni, J-P. et al. Eds. Elsevier Science Publishers B.V.

DOAÇÃO

Data: 23/07/04

De: Profª Lisbeth

BIBLIOTECA "CARLOS BENJAMIN LYRA"
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
— UNIVERSIDADE DE SÃO PAULO —
Cidade Universitária
C.P. 68.281 - AG. Cidade de São Paulo
05315-970 - SÃO PAULO - BRASIL
Tel: 818.6174; 818.6109; 818.6269 Fax: 818.5036
e-mail: bib@ime.usp.br