

ESTATÍSTICA E PROBABILIDADE

Extraído do Guia do PNL D 2012

Um dos conhecimentos mais utilizados hoje em dia é a estatística, que descreve os dados observados em pesquisas ou experimentos em quase todas as atividades humanas e desenvolve metodologias para a tomada de decisão na presença da incerteza. É cada vez mais relevante, para todo cidadão, interpretar criticamente resultados de pesquisas estatísticas. Para isso, é importante que situações que envolvam dados da realidade física ou social – que precisam ser coletados, selecionados, organizados, apresentados e interpretados – façam parte da formação básica de nossos alunos. É também importante saber fazer inferências, com base em informações qualitativas ou dados numéricos.

Diversos estudos na área educacional propõem abordagens para a formação estatística dos alunos com atividades que exigem um maior envolvimento deles no planejamento de pesquisas, construção de questões, definição adequada de população e amostra para cada tipo de pesquisa, coleta e organização de dados, distribuições de frequência, medidas de tendência central e de dispersão. No entanto, são poucas as obras didáticas nas quais são discutidas de maneira satisfatória as etapas de planejamento de uma pesquisa estatística.

Gráficos e tabelas são muito frequentes nos livros didáticos para o ensino básico. Além disso, algumas delas reservam capítulos específicos para o estudo mais detalhado desses tipos de representação. Em contrapartida, são poucas as coleções que exploram aspectos importantes da estatística, associados à análise dos gráficos: a população pesquisada (se é uma amostragem ou uma pesquisa censitária); a opção por apresentar frequência absoluta ou relativa e suas consequências; a escolha de escalas adequadas para cada eixo; as variáveis que estão sendo relacionadas em um mesmo gráfico; a necessidade ou não do uso de legenda; entre outros. São, ainda, raras as atividades que incentivem a análise crítica de uma representação usada na mídia ou em divulgação científica de pesquisas. Não se propõe sua comparação com outras formas de representação, e tampouco se reflete se há indução a interpretações equivocadas. Ao contrário, incluem-se gráficos e tabelas com inadequações, como é o caso de muitos pictogramas utilizados na imprensa escrita, sem qualquer advertência para que o aluno detecte erros ou falta de informação. As falhas, nesses casos, muitas vezes dificultam o acesso ao conteúdo apresentado.

Algumas coleções ainda trazem uma abordagem da estatística por meio de exemplos fictícios, com foco em nomenclatura e em procedimentos de cálculo desprovidos de validação e de interpretação. O cálculo de medidas descritivas deveria ser analisado à luz do raciocínio estatístico e não meramente por meio dos resultados numéricos. Aprender

técnicas de cálculo sem ser capaz de interpretar seus resultados é enfadonho e desnecessário.

Os recursos computacionais, como planilhas eletrônicas, calculadoras simples ou científicas, também permanecem pouco explorados nas coleções aprovadas. Em apenas uma delas esse trabalho é efetivamente valorizado. Como sabemos, uma das vantagens do uso das tecnologias atuais de informação e comunicação é que elas possibilitam, em um tempo cada vez menor, estender em muito o número de dados que podem ser trabalhados nos experimentos. Acima de tudo, isso abre espaço para que possamos investir na busca do significado e na interpretação dos dados obtidos e das medidas estatísticas associadas a ele, que são fundamentais para um efetivo trabalho técnico ou científico. Contudo, não podemos esquecer que é indispensável, também, compreender e saber justificar os procedimentos de cálculo e as fórmulas que os definem, para que se possa exercer o controle dos cálculos e dos algoritmos realizados pelos artefatos tecnológicos.

Em alguns textos, uma omissão verificada na seleção de conteúdos da estatística é a classificação de variáveis – quantitativas (ou numéricas) e qualitativas (ou categorizadas). Essa classificação é fundamental na definição do tipo de organização e apresentação dos dados (os tipos de gráficos adequados) e em decisões a respeito da análise a ser adotada. Disso decorre uma imprecisão encontrada em algumas obras ao chamar gráfico de colunas de histograma, ou vice-versa. O histograma é um gráfico construído a partir de uma tabela com dados de uma variável quantitativa. Neste caso, os valores assumidos são agrupados em intervalos para os quais se constroem as colunas verticais e contíguas (sem espaço entre elas) com alturas proporcionais às frequências de ocorrência de cada intervalo. No que se refere às variáveis quantitativas, um dos conceitos considerados fundamentais na estatística é o de variabilidade, e a medida mais simples para introduzir o conceito é a amplitude (diferença entre o valor máximo e o valor mínimo), raramente mencionada nas obras. Mesmo no estudo da variância, do desvio padrão ou do desvio médio absoluto, a interpretação associada ao conceito de variabilidade não é valorizada. Assim, esse estudo costuma ser reduzido a técnicas operatórias, com pouca discussão de seus significados para a compreensão dos dados. Alguns textos, que buscam avançar desnecessariamente nos conteúdos da área de estatística, incorrem em imprecisões ao fazerem uma abordagem inicial da inferência estatística. Um das inadequações verificadas ocorre no ajuste de histogramas por uma curva normal, o que nem sempre faz sentido. A própria inclusão, no ensino básico, do estudo da distribuição normal exige maior conhecimento estatístico.

No estudo das probabilidades, podem ser identificados pontos positivos em algumas obras. Nelas, observamos maior cuidado na abordagem dos conceitos básicos no campo das probabilidades e preocupação em associá-los a problemas reais e sugestivos. A noção

de incerteza de ocorrência de alguns eventos é igualmente tratada de forma satisfatória, em oposição ao caráter determinístico de outros eventos.

A despeito desses aspectos elogiáveis, há limitações na maioria das obras, no campo das probabilidades. Uma delas é que, tanto na introdução dos conteúdos, quanto em alguns problemas propostos, há contextualizações inadequadas ou artificiais.

Outra limitação vem de uma tradição arraigada de se anteceder o estudo das probabilidades (e da estatística) por um longo e fragmentado capítulo de análise combinatória. É certo que a contagem de possibilidades é uma ferramenta muito útil para resolvermos problemas de probabilidades, quando se utiliza sua definição clássica. No entanto, estender demais a preparação em análise combinatória induz o aluno a pensar que, sem todo aquele arsenal, não é possível compreender probabilidades (nem estatística).

Em relação às noções básicas de probabilidade, observamos que, em algumas obras, não é apresentada, de maneira apropriada, a noção de independência probabilística entre dois eventos definidos em um mesmo espaço amostral. Nesse caso, é conveniente, antes de abordar independência, estudar o conceito de probabilidade condicional, o que nem sempre é feito nos livros aprovados.

No ensino básico, adota-se quase sempre a chamada definição clássica de probabilidade de ocorrência de um evento num determinado experimento aleatório, simplificada: *“o quociente do número de casos favoráveis ao evento pelo número de casos possíveis no experimento”*. Tal definição tem como premissa fundamental que os eventos possíveis tenham a mesma probabilidade (chance) de ocorrência, noutros termos, que sejam equiprováveis¹. No entanto, tal suposição nem sempre é explicitada nos textos didáticos. Por exemplo, em experimentos com dados ou moedas, é necessário explicitar a suposição de que eles são ou não “honestos”.

¹ Uma das críticas dessa definição aponta a sua circularidade, pois estaríamos utilizando a noção de *“mesma probabilidade”* para definir probabilidade. Essa crítica é contornada de duas maneiras, em geral. Uma é basear a suposição de equiprobabilidade em propriedades de simetria e homogeneidade presentes nos objetos envolvidos no experimento. Por exemplo, no lançamento de um dado simétrico e homogêneo (“honesto”) é razoável supor que todas as faces tenham a mesma chance de ficar voltadas para cima. A outra maneira é tomar a premissa de equiprobabilidade como um dos pontos de partida para construir um modelo probabilístico do fenômeno em questão.