

Explorando compreensões sobre amostragem com uma professora utilizando o *software TinkerPlots 2.0*

Martins, Niedja Pereira; Monteiro, Carlos Eduardo Ferreira y Queiroz, Tamires Nogueira
Universidade Federal de Pernambuco

Resumo

Este artigo apresenta aspectos de um estudo de caso que enfocou a compreensão sobre amostra e amostragem por uma professora dos anos iniciais ao utilizar o *software TinkerPlots 2.0*. Buscou-se analisar a compreensão da docente a respeito de três aspectos básicos do conceito de amostra: tamanho, tipo e representatividade, ao utilizar as ferramentas do referido *software*. O processo de coleta de dados da pesquisa envolveu a participação da professora em três etapas: entrevista inicial sobre os conhecimentos a respeito do conceito de amostra e a apresentação do *TinkerPlots 2.0*; realização de atividades de amostragem aleatória simples no *software*; e entrevista final com o objetivo de identificarmos possíveis mudanças na compreensão sobre o conceito. Nos resultados, observamos que a docente modificou sua compreensão sobre amostra, identificou tamanhos e os relacionou adequadamente à representatividade nas amostras. Identificou-se a noção de representatividade ao longo das manipulações dos dados realizadas pela professora. Apesar de que ela reconheceu diferenças na distribuição dos grupos de dados, não estabeleceu relação disso com a representatividade, o tamanho da amostra ou processo de amostragem. Consideramos, então, que a possibilidade de simulação a partir da ferramenta *Sampler* e da manipulação de dados a partir da ferramenta *Plot* pôde propiciar situações de interpretação de dados estatísticos que favoreceram compreensões iniciais sobre amostragem.

Palavras chaves: Educação Estatística; Amostragem; Professores de séries iniciais; Software TinkerPlots.

1. Introdução

O fluxo de informações incitado pelas Tecnologias da Informação e Comunicação estimula mudanças nas linguagens e representações utilizadas no processo de comunicação (LÉVY, 2000). Pode-se afirmar que em virtude disso, o uso frequente de dados estatísticos na sociedade vem aumentando e compreendê-los torna-se uma necessidade para a efetivação da cidadania.

Sendo assim, a Educação Estatística torna-se um importante elemento da Educação Básica por possibilitar o desenvolvimento de uma atitude investigativa, reflexiva e crítica dos estudantes em uma sociedade marcada pelo acúmulo de informações (CAMPOS et al., 2011). De outro modo, a introdução da Estatística na escola pode incorporar em seus processos educativos importantes instrumentos culturais compatíveis a uma sociedade informacional.

Uma maneira que as escolas têm evidenciado para buscar atualizar suas práticas é utilizando computadores no ensino. No campo da Estatística, por exemplo, tem-se a possibilidade de realizar atividades por meio interfaces gráficas, como num *software*, o que contribui para modificar os processos de ensino e aprendizagem (Hawkins, 1997).

Em termos das atividades com amostras, a dificuldade de se trabalhar com uma quantidade grande de dados em situações de amostragem não se torna um grande problema

quando se tem disponíveis instrumentos tecnológicos que permitam a simulação dos dados para a retirada de várias estatísticas.

Refletindo sobre esse conceito, Watson (2004) argumenta que embora o raciocínio sobre ele seja fundamental para legitimar práticas de Estatística, frequentemente a amostra recebe pouca atenção no currículo escolar. A possibilidade de utilizar tecnologias da Informação e Comunicação para desenvolver atividades sobre amostragem, no entanto, pode ser uma alternativa interessante de abordá-lo no ambiente escolar.

Especificamente, alguns *softwares* de análise de dados vem sendo apontados como instrumentos que favorecem o trabalho com dados estatísticos, tal como o *software TinkerPlots 2.0* o qual oferece a possibilidade de simular dados para a construção de diferentes tipos de amostras. Neste artigo, apresentamos um recorte de uma pesquisa que buscou compreender como o conceito de amostra poderia ser desenvolvido por uma professora mediante a utilização do *software TinkerPlots*.

2. Estudos sobre Amostra e Amostragem

A amostra, quando comparada com outros conteúdos estatísticos é pouco abordada nas investigações em Educação Estatística, ao mesmo tempo em que é relevante de ser considerada na escolarização básica (Watson, 2004; Ben-Zvi et al. 2011). Watson (2004) indica que uma possível causa para que professores não realizem atividades de amostragem com seus alunos pode estar relacionada ao fato desse conceito ser componente de um tópico associado à Matemática, mas não exigir rigorosamente a realização de cálculos para serem aprendidos.

Pela importância que tal conceito apresenta no desenvolvimento de conhecimentos ligados a Estatística, alguns documentos internacionais têm preconizado seu ensino nos anos iniciais. Franklin et al. (2005) propõem o Guia para Avaliação e Instrução em Educação Estatística (GAISE), o qual estabelece algumas competências a serem desenvolvidas nos estudantes da escola básica no âmbito internacional, tais como: estabelecer investigações por amostragem; criar amostragem com seleção aleatória; reconhecer se uma amostra é ou não representativa e realizar generalizações a partir de uma amostra.

Para o trabalho com amostras, três aspectos estão intimamente relacionados e apresentam-se enquanto aspectos básicos de serem aprendidos pelas pessoas em atividades de amostragem, são eles: tamanho, representatividade e tipo de amostra.

Bolfarine e Bussab (2005) discutindo sobre a importância da definição do tamanho de uma amostra indicam que a precisão estatística de um processo por amostragem esbarra nas limitações do orçamento da pesquisa e no tamanho da amostra.

Tais preocupações tornam-se importantes, pois há sempre a necessidade de extrair amostras que sejam representativas. Podemos dizer, inclusive, que a representatividade é a condição mais importante em uma investigação quando se pretende generalizar dados. Assim, Bolfarine e Bussab (2005) expõe que uma amostra é dita representativa quando suas características se assemelham ao máximo à população investigada.

O desenvolvimento de atividades sobre amostragem exige a utilização de várias ideias relacionadas. Essas interconexões de diferentes conceitos estatísticos parecem ser fundamentais no aprendizado sobre amostra, por exemplo, Ben-Zvi et al. (2011) cita a ligação da amostra com conceitos tais como intervalo e distribuição, inferência, probabilidade, aleatoriedade e interpretação de gráficos.

Considerando esses aspectos, Prodromou (2010) realizou um estudo que investigou a compreensão de estudantes sobre a relação entre tamanho e representatividade em amostras,

utilizando a versão 1.0 do *software TinkerPlots*. Ela pode perceber, enquanto resultado, que os alunos puderam apresentar não apenas reflexões iniciais sobre tal relação, mas também sobre outros conceitos relacionados, como representatividade, variabilidade e distribuição amostral.

A utilização do *TinkerPlots* para o desenvolvimento de pesquisas sobre amostras também foi objeto de investigação de Kasak e Konold (2010) que desenvolveram um protótipo para a simulação de probabilidades junto a esse *software*.

Nas pesquisas desses autores foi observado que inicialmente os estudantes desse estudo confundiam a ideia de aleatoriedade com acaso. Ao incluir uma ferramenta que permitisse aos sujeitos acompanharem visualmente o desenho de um plano amostral, foi possível perceber um crescimento conceitual dos estudantes em relação à ideia de probabilidade contida na amostra aleatória simples. Nesse sentido, a simulação mostrou ser um elemento importante para a compreensão do conceito de amostra aleatória.

3. Compreendendo o TinkerPlots

O *TinkerPlots* é um *software* projetado para construir e manipular dados estatísticos. Ele possui ferramentas que permitem passar por todas as etapas do processo de tratamento das informações estatísticas, possibilitando aos usuários, explorar dados, representá-los, comunicá-los, formular hipóteses e dentre outros aspectos.

Estudos empíricos apontam para a eficácia do uso do *TinkerPlots* no ensino de conceitos e habilidades estatísticas com crianças a partir dos 9 anos de idade (KONOLD, 2006), pelo fato desse *software* possuir uma interface simples.

O *TinkerPlots* apresenta um menu de ferramentas no idioma inglês e traz ícones de cinco ferramentas básicas: *Cards*, *Table*, *Plot*, *Sampler* e *Text* direcionadas a construção e representação dos dados, conforme a Figura 1:



Figura 1. Ícones do menu de ferramentas do *TinkerPlots 2.0* que é apresentado na tela inicial.

A ferramenta *Cards* serve para registrar a criação de dados da pesquisa a ser realizada no *software*. Ao fazer isso, o usuário poderá inserir atributos (ou variáveis) no *Cards*. A segunda ferramenta do menu do *TinkerPlots* é a *Table*. Ao ativar essa ferramenta acionamos a distribuição dos casos incluídos na ferramenta *Cards* em uma tabela.

Na sequência, podemos acionar a ferramenta *Plot* que permite a visualização dos casos incluídos na ferramenta *Cards* numa subjanela. Ao fazer uso da ferramenta *Plot*, teremos acesso a um segundo menu de ferramentas que permite a manipulação dos dados para a construção de diferentes representações e seleção de diferentes estatísticas.



Figura 2. Menu da ferramenta *Plot* do *TinkerPlots 2.0* que é habilitado ao acionarmos o ícone *Plot* no menu inicial do *software*.

As principais ferramentas disponíveis para a manipulação do *Plot* são: *Separate* (separar), *Order* (ordenar) e *Stack* (empilhar). Essas ferramentas podem ser utilizadas de forma simultânea para construir diferentes representações. Do mesmo modo, é possível fazer uso de outros ícones no menu de ferramentas do *Plot* para identificar valores, estatísticas e parâmetros nos conjunto de dados, como por exemplo as ferramentas *Ref.*, *Averages* e *Counts*.

Na versão 2.0 do *TinkerPlots* foi inserida a ferramenta *Sampler* para o trabalho com probabilidade e amostra. Ela favorece a manipulação de uma grande quantidade de dados sem a necessidade de uma etapa extensa de coleta.

Os componentes básicos da Ferramenta *Sampler* incluem diferentes funções. Os principais componentes da ferramenta *Sampler* são os dispositivos que permitem a inserção das variáveis a serem selecionadas na amostra; o dispositivo *Repeat* que permite a definição do tamanho da amostra a ser extraída e o botão *Run* que permite que os dados sejam retirados do simulador aleatoriamente. A seguir, são mostradas as funções da ferramenta *Sampler* na Figura 3:

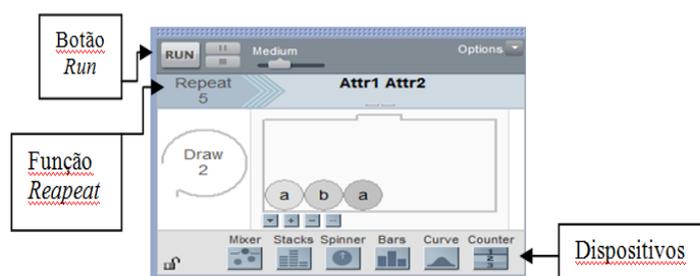


Figura 3. Aspectos da ferramenta *Sampler*.

4. Método

A abordagem metodológica foi composta por dados provenientes de diferentes instrumentos de coleta, tal qual a entrevista semi- estruturada e a videografia. Entrevistamos uma professora que, nesse estudo será chamada de Ana por questões éticas.

A coleta de dados aconteceu em três encontros. No primeiro e no terceiro encontro, a professora participante foi entrevistada com base em questões propostas por Watson (2004). Ainda no primeiro encontro foi realizada uma familiarização da professora com o *software*, na qual buscamos explorar as ferramentas do *TinkerPlots* a partir de um banco de dados que apresentavam características diversas de 24 gatos. No segundo encontro foram submetidas duas atividades sobre amostragem as quais desenvolvidas por meio do *software*.

As duas atividades tiveram como objetivo, propor uma situação na qual fossem estabelecidas relações entre o tamanho de uma amostra aleatória simples e sua representatividade. Para este artigo, fizemos um recorte para apresentar o desenvolvimento de uma das atividades (*FishPopulation*). O exercício apresentava um banco de dados com 625 peixes e informações relativas ao tipo do peixe (geneticamente modificado ou normal), ao tamanho do peixe (em centímetros) e uma identificação numérica em cada peixe. Nessa atividade, a professora foi convidada a extrair casos do simulador *Sampler* para aumentar gradativamente o tamanho da amostra e analisar o impacto do tamanho sobre a representatividade dos dados.

5. Análise dos protocolos das entrevistas

Na atividade “*FishPopulation*” a professora foi conduzida a extrair amostras crescentes de uma mesma população com $N = 625$ e parâmetros inicialmente desconhecidos. A intenção dessa atividade era promover a compreensão de que amostras maiores tendem a representar melhor a população e estimar níveis de confiança informais sobre as suas inferências. Inicialmente a professora apresentou pouca segurança para interpretar os dados em função da amostra conter um tamanho pequeno, como vemos no exemplo abaixo:

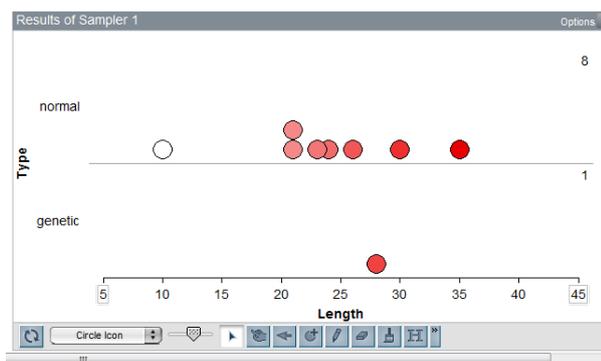


Figura 6: Representação gerada na Atividade 1 com a professora Ana demonstrando uma Amostra Aleatória Simples com $n = 9$.

P.: Olhando para cá, você pode me dizer se os peixes geneticamente modificados são maiores que os peixes normais?

A.: São maiores, não. Se a gente tirar a média... Só tem um.

P.: Só tem um, né? Então, com um só daria pra dizer?

A.: De 600? Dá nada!

P.: O que precisaria?

A.: De 600, pelo menos uns 10 por cento, né?

P.: 10 por cento?

A.: Menos?

P.: O que você acha?

A.: Não sei...

Nessa extrato de entrevista é possível observar que a mesma identifica não poder avaliar se o grupo dos peixes “genéticos” são maiores que os normais em função da disparidade entre um e outro grupo, como também em função da relação entre o tamanho dessa amostra e o tamanho total da população.

Contudo, Ana não manteve essa certeza ao longo da atividade, demonstrando não compreender que para responder corretamente a questão, seria necessário considerar o tamanho dos dois grupos de peixes e não analisar casos isolados. Isso nos parece ser um erro frequente, pois ao interpretar dados de um grupo, algumas pessoas apresentam a estratégia de observar pontos extremos do gráfico para oferecer conclusões a todo conjunto de dados (Asseker, 2011).

Nesse exemplo acima com $n = 9$, a professora optou em utilizar a ferramenta média para identificar qual grupo permanecia com o comprimento maior. Por ser uma medida de tendência central e, as unidades amostrais do grupo de peixes geneticamente modificados estarem em um número inferior às unidades do grupo normal, a média não ajudou a professora a corroborar sua impressão de que os peixes normais estariam em vantagens em relação aos genéticos.

Em função dessa dificuldade, a professora apresentou interpretações que não considerou a possibilidade dos peixes geneticamente modificados serem maiores mesmo quando $n = 28$. Aumentou a amostra para 45 peixes e adicionou a ferramenta *Hat* que pode auxiliar na identificação do aglomerado da representação. A partir dessa função, Ana conseguiu observar que o maior concentrado dos dados se encontrava em relação à escala. Podemos ver esse momento abaixo:

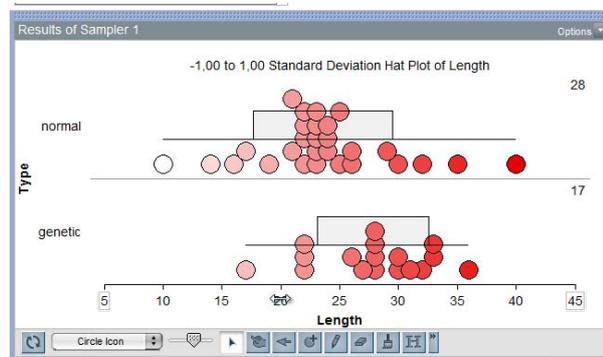


Figura 7. Demonstração do uso da ferramenta *Hat* acionada pela professora Ana para responder o questionamento da Atividade 1 com uma amostra de $n = 45$.

Pesquisadora.: Qual seria o aglomerado que demonstra um tamanho maior em centímetros?

Ana.: Esse (Geneticamente modificados).

P.: E não esse? (Normais).

A.: É.

P.: Então, eu posso dizer que os genéticos estão aglomerados num lugar que indica um tamanho maior na escala do que os normais?

A.: É. Os normais têm muito, mas estão atrás. Entendi.

P.: Então, o que você percebe? Qual é o conjunto que está maior?

A.: Os genéticos.

P.: Com essa quantidade de peixes que você pegou, você poderia dizer que essa ideia que você concluiu se aplicaria aos 625?

A.: Sim.

Nessa Figura 7 pode ser observado que a ferramenta *Hat* seleciona os aglomerados ou concentrações do gráfico permitindo observar de maneira mais clara a tendência de cada grupo. Em seguida, questionamos a professora a respeito da confiança que ela mantinha sobre esta última interpretação e, mais uma vez, a professora oferece um nível de confiança alto para sua inferência.

Somente ao indicarmos que poderíamos inserir uma quantidade maior de elementos ao *Plot* a professora conseguiu indicar explicitamente que sua confiança a respeito da inferência também aumentaria, como mostra o trecho da entrevista seguir:

P.: Você diz que tem 9, numa escala de 0 a 10, de confiança que esse resultado se aplica ali. Se ao invés de colocar mais nove, eu colocasse 50?

A.: Eu ia ter mais confiança ainda.

P.: Por quê?

A.: Porque se é mais, a gente pode ter mais certeza, né?

A professora Ana também fez uso das ferramentas *Average* (Média) e *Ruler* (Régua) sem grandes dificuldades para chegar a resposta final à questão. Mesmo quando adicionarmos todos os 625 casos da população, Ana continuou a considerar as diferenças em relação às medidas as quais ela verificou em diferentes amostras.

6. Considerações finais

A manipulação das amostras por Ana junto ao *software*, como pode ser analisado, aconteceu a partir da mediação da pesquisadora, o que significa dizer que tais resultados não podem ser tomados apenas em função das possibilidades presentes nas ferramentas do *software*.

No decorrer dessa atividade, percebemos que a professora utilizou apenas três ferramentas para generalizar dados de uma amostra à uma população: *Average* (média), *Ruler* (régua) e *Hat* (identifica a tendência dos dados). A professora pode identificar aspectos de variação nos grupos dispostos na amostra. Esse achado nos fez pensar que é possível explorar nos sujeitos a relação entre heterogeneidade/homogeneidade e representatividade em amostras.

Percebemos ainda, que a possibilidade de visualização dos dados através da ferramenta *Plot* foi um elemento que facilitou na observação das diferentes distribuições da amostra. Na medida em que se aumentava o tamanho da amostra, a professora foi estabelecendo respostas mais seguras em relação à população.

Identificamos que o trabalho com amostras crescentes no *Tinkerplots* permitiu a professora compreender que amostras menores tendem a representar de forma menos confiável os dados de uma população. Sabemos, porém, que essa compreensão expressa uma ideia parcial sobre a representatividade em amostras, uma vez que faltam compreensões a respeito do processo de seleção da amostra e da variabilidade dos dados.

Apesar da boa execução da atividade, os dados não oferecem evidências sobre como a professora concebia um fenômeno aleatório e por isso, não podemos concluir que a mesma construiu uma compreensão conceitual sobre a amostragem aleatória simples. Para outros estudos, será interessante desenvolver atividades que explorem a construção de amostras aleatórias simples a fim de que numa investigação como essa, seja possível observar suas hipóteses a respeito dos fenômenos aleatórios.

Referências

- Asseker, A. (2011). *O uso do TinkerPlots para a exploração de dados por professores de escolas rurais*. Dissertação (Mestrado em Educação Matemática e Tecnológica) – Centro de Educação, Universidade Federal de Pernambuco. Recife.
- Ben-Zvi, D. et al. (2011). Children's emergent inferential reasoning about samples in an inquiry-based environment. In: *Proceedings of the Seventh Conference of the European Society For Research In Mathematics Education*, Poland, Rzeszow.
- Bolfarine, H. e Bussab, W. (2005.) *Elementos de amostragem*. São Paulo: Blucher.
- Campos, C. R. et. al. (2011). Educação estatística no contexto da educação crítica. *Bolema* 24, 39, 473-494.
- Franklin, C. et al. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) 14 report*: Alexandria, VA: American Statistical Association,
- Hawkins, A. (1997) Myth-conceptions! In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 1-14). Voorburg, The Netherlands: International Statistical Institute.
- Kazak, S. e Konold, C. (2010) Development of ideas in data and chance through the use of tools provided by computer-based technology. In: *Proceedings of the Eighth International*

Conference on Teaching Statistics. Slovenia: International Statistical Institute, 2010.
Disponível em:

http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_8D2_KAZAK.pdf>.

Konold, C. (2006). Handling complexity in the design of educational *software* tools. In: *Proceedings of the Seventh International Conference on Teaching Statistics*. Salvador.

Lévy, P. *Cibercultura*. São Paulo: Editora 34, 2000. 260.

Prodromou, T. (2011). Students' emerging inferential reasoning about samples and sampling. In: *Anais do 23 biennial conference of the Australian Association of Mathematics Teachers. Mathematics: traditions and [new] practices, 2011*. Disponível em: http://www.merga.net.au/documents/RP_PRODROMOU_MERGA34-AAMT.pdf. Acesso em: 30 ago. 2012.

Watson, J. M. (2004) Developing reasoning about samples. In: Ben-Zvi, D., Garfield, J. (Org.), *The challenge of developing statistical literacy, reasoning and thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishers.