



CÁLCULO DO TAMANHO AMOSTRAL EM R SHINY

Calculation of the Sample Size in R Shiny

BARASUOL, João B.¹; GELATTI, Gabriel A.²; JOCHIMS, Gabriel S.³;
NASCIMENTO; Antonielli D. do⁴; ANSUJ, Angela P.⁵; FIGUEIRÓ, Michele F.⁶

Resumo: As técnicas de amostragem possibilitam fazer afirmações sobre uma população, baseando-se nos resultados obtidos de amostras representativas. O pacote Shiny do programa estatístico R cria aplicativos para *web* (*web apps*) interativos, que podem ser úteis na resolução de cálculos estatísticos. Este artigo objetiva apresentar o *Sampling*, um *web app* elaborado em R Shiny, que utiliza procedimentos de amostragem para calcular o tamanho amostral para estimar proporção ou média de populações finitas ou infinitas.

Palavras-chave: Amostragem, Tamanho Amostral, Populações Finitas e Infinitas, R Shiny.

Abstract: The sampling techniques allow making assertions about a population, based on the obtained results of representative samples. Shiny package of R statistical software creates interactive web applications (web apps), that can be helpful in resolution of statistical calculations. This article aims at presenting the *Sampling*, a web app developed in R Shiny that uses sampling proceedings to calculate sampling size to estimate proportion or mean of finite or infinite populations.

Keywords: Sampling, Sample Size, Finite and Infinite Populations. R Shiny.

Introdução

A linguagem de programação R é comumente utilizada na análise exploratória de dados, uma vez que ela apresenta um conjunto de bibliotecas ou pacotes que auxiliam na realização de cálculos estatísticos e gráficos. Fazendo-se uso do RStudio, seu ambiente de desenvolvimento integrado, é possível também elaborar relatórios e apresentações dinâmicos, e desenvolver aplicativos para *web*.

¹ Acadêmico do curso de Ciência da Computação, Universidade de Cruz Alta. E-mail: joaobrenobarasuol@hotmail.com.

² Bolsista de IC, acadêmico do curso de Ciência da Computação e estagiário do NEA, Universidade de Cruz Alta, PIBIC/UNICRUZ. E-mail: gabriel.gelatti@sou.unicruz.edu.br.

³ Acadêmico do curso de Ciência da Computação e estagiário do NEA, Universidade de Cruz Alta. E-mail: gabriel.jochims@sou.unicruz.edu.br.

⁴ Acadêmica do curso de Ciência da Computação, Universidade de Cruz Alta. E-mail: antonielli_nascimento@hotmail.com.

⁵ Professora do curso de graduação em Estatística, Universidade Federal de Santa Maria. E-mail: angelaansuj@yahoo.com.

⁶ Professora do curso de graduação em Ciência da Computação e coordenadora do NEA, Universidade de Cruz Alta. E-mail: mfigueiro@unicruz.edu.br.



O pacote Shiny da linguagem R permite a criação de *web apps* interativos, sem que haja a necessidade de conhecimento em linguagens de marcação ou de programação web, tais como HTML, CSS, JavaScript e PHP (BEELEY, 2013).

Em estatística, as técnicas de amostragem, associadas à coleta de dados, garantem a obtenção de amostras da população de interesse. O uso de amostras é recomendado quando a população alvo é muito grande ou o custo da realização da pesquisa é muito alto. Caso contrário, pode-se optar pelo uso do censo. A partir dos resultados encontrados de uma amostra probabilística, é possível fazer inferências estatísticas sobre a população alvo. (BONFARINE; BUSSAB, 2004).

Este artigo objetiva apresentar o *web app* chamado *Sampling* elaborado no pacote Shiny da linguagem R, que calcula o tamanho amostral (n) para estimar uma proporção de população (finita ou infinita), ou uma média de população (finita ou infinita) com variâncias populacionais (conhecidas ou desconhecidas). A amostra a ser selecionada é do tipo probabilística.

A Linguagem R e o Pacote Shiny

O sistema estatístico R é formado por uma linguagem de programação e um ambiente com gráficos, um sistema de depuração, acesso a certas funções do sistema e a capacidade de executar programas armazenados em arquivos de *script* (HORNIK, 2017). A linguagem R é adequada para a realização da análise exploratória de dados, utilizando procedimentos metodológicos da estatística computacional.

Sendo desenvolvido para o uso na área estatística, a linguagem R é capaz de realizar um grande número de procedimentos estatísticos através das bibliotecas e dos pacotes disponíveis. Além disso, ela também possui um amplo conjunto de funções para criação de apresentações de dados através de um ambiente gráfico, através da utilização de bibliotecas e pacotes com propósitos específicos (HORNIK, 2017).

A programação em R acontece no ambiente de desenvolvimento integrado chamado RStudio, onde é possível a criação de *scripts*, acesso ao *console*, visualização dos gráficos e históricos dos dados, suporte referente às bibliotecas e aos pacotes, e entre outras funcionalidades. O RStudio foi criado com o intuito de facilitar a curva de aprendizagem para



novos usuários da linguagem R, assim como, permitir que aqueles mais experientes a utilizem de uma forma mais prática a partir das ferramentas disponíveis no Rstudio (ALLAIRE, 2011).

O pacote Shiny da linguagem R adiciona diversas ferramentas que permitem a criação de *web apps* interativos, sem a necessidade de conhecimentos específicos das linguagens de web, tais como HTML ou PHP. Este pacote é popular pela sua capacidade de construir *web apps* com *layouts* de interface do usuário (IU) simples e interativos, e com respostas rápidas, segundo as interações e necessidades de cada usuário. No final, as estatísticas ou os gráficos são exibidos, sem que haja a necessidade da instalação do sistema estatístico R na máquina do usuário (THE SHINY, 2017).

Uma vez que, a IU do Shiny substitui a necessidade de utilizar os comandos do R, a experiência final torna-se muito mais fácil e agradável para o usuário. Estes *web apps* são capazes de utilizar valores de entrada (*inputs*) para os cálculos estatísticos no R, e passar os resultados em forma de valores de saída (*outputs*) para o usuário. Esta saída pode ser tanto na forma de texto para resultados simples, quanto na representação de tabelas ou gráficos, de acordo com a complexidade dos cálculos ou necessidade do usuário (THE SHINY, 2017).

Procedimentos Metodológicos

O *web app Sampling*, elaborado no pacote Shiny da linguagem R, realiza, quando necessário, o cálculo do tamanho amostral (n) para estimar uma proporção populacional (finita ou infinita), ou uma média populacional (finitas ou infinitas) com variâncias populacionais (conhecidas ou desconhecidas). Considerando que o planejamento adotado na seleção desta amostra de tamanho (n) é do tipo Amostragem Aleatória Simples (AAS). A relevância deste estudo justifica-se pelo fato de que adotar os procedimentos metodológicos corretos no levantamento amostral de pesquisas quantitativas (levantamento ou experimentação), conduz obrigatoriamente a resultados estatísticos mais precisos.

Resultados e Discussões

Como resultado, apresenta-se o desenvolvimento do *web app Sampling*, em R Shiny, que calcula o tamanho amostral. Ele pode ser perfeitamente utilizado em pesquisas



quantitativas, durante a fase de levantamento amostral da população alvo. Foram implementadas funções para os cálculos do tamanho amostral na estimação de uma proporção e de uma média populacionais, tanto nos casos com populações finitas ou infinitas, quanto com variâncias populacionais conhecidas ou desconhecidas.

O *Sampling* foi desenvolvido para ser simples e intuitivo, apresentando uma interface que pode ser facilmente compreendida. As fórmulas matemáticas para o cálculo amostral são apresentadas para o usuário, detalhando cada variável envolvida.

O código foi escrito em linguagem R e se divide em duas partes, IA e Servidor. A interface do usuário contém as funções do Shiny e do compilador de texto LaTeX, utilizado na exibição das expressões matemáticas. Esta interface utiliza-se de painéis condicionais para apresentar apenas o necessário ao cálculo amostral selecionado, como exemplificado na Figura 1.

Figura 1: Trecho da parte IA do código escrito em R.

```
fluidRow(  
  column(width = 5, offset = 0, wellPanel(  
    selectInput("Calculate", "PLEASE SELECT AN OPTION",  
              choice = c("POPULATION PROPORTION", "POPULATION MEAN"),  
              selected = "POPULATION PROPORTION")  
  )),  
  column(width = 4, offset = 0, wellPanel(  
    radioButtons("Size", "POPULATION",  
                choices = c("FINITE", "INFINITE"),  
                selected = "INFINITE")  
  )),  
)
```

Fonte: Elaboração própria pelos autores.

A segunda parte do código é o servidor, onde são armazenados os dados inseridos pelo usuário, fórmula do cálculo amostral é compilada, e cuja resposta final pode ser visualizada na interface do *app web* pelo usuário.

Na interface principal, o usuário deve selecionar se deseja estimar a proporção ou a média populacional. Em seguida, deve-se selecionar se a população alvo é dita finita ou infinita. Se a opção “proporção de uma população infinita” seja escolhida, o erro amostral (e_0), o nível de significância (α) e a proporção amostral (p) de resultados favoráveis devem ser informados. Se população for finita, o tamanho populacional (N) também deverá ser inserido, como mostram as Figuras 2 e 3, respectivamente.



Figura 2: Cálculo do tamanho amostral (estimação de proporção de população finita).

Calculate Sample Size

PLEASE SELECT AN OPTION

POPULATION PROPORTION

POPULATION

FINITE

INFINITE

FORMULA

$$n = \frac{(z_{\alpha/2})^2 \hat{p} (1 - \hat{p}) N}{e_0^2 (N - 1) + (z_{\alpha/2})^2 \hat{p} (1 - \hat{p})}$$

SAMPLE SIZE IS

[1] 212.4892

FINITE POPULATION PROPORTION

Population Size (N)

500

Standard Error (e₀)

0,05

Significance Level (α)

0,05

Proportion of Favorable Results (p̂)

0,4

Fonte: Elaboração própria pelos autores.

Figura 3: Cálculo do tamanho amostral (estimação de proporção de população infinita).

Calculate Sample Size

PLEASE SELECT AN OPTION

POPULATION PROPORTION

POPULATION

FINITE

INFINITE

FORMULA

$$n = \frac{(z_{\alpha/2})^2 \hat{p} (1 - \hat{p})}{e_0^2}$$

SAMPLE SIZE IS

[1] 368.7936

INFINITE POPULATION PROPORTION

Standard Error (e₀)

0,05

Significance Level (α)

0,05

Proportion of favorable results (p̂)

0,4

Fonte: Elaboração própria pelos autores.

No cálculo do tamanho amostral para estimar média populacional, duas opções devem ser selecionadas, se a população é finita ou infinita e se a variância populacional é conhecida ou desconhecida.

Considerando uma população finita com variância populacional conhecida, as incógnitas relacionadas são o erro amostral (e₀), o nível de significância (α), variância populacional (σ²) e o tamanho populacional (N), de acordo com a Figura 4. Caso a variância



populacional seja desconhecida, informam-se também a estimativa da variância amostral (s^2) e os graus de liberdade (v).

De acordo com a Figura 5, as incógnitas para a população infinita, com variância populacional desconhecida, são o erro amostral (e_0), o nível de significância (α), a estimativa da variância amostral (s^2) e os graus de liberdade (v). Com variância populacional (σ^2) conhecida, informa-se seu valor.

Figura 4: Cálculo do tamanho amostral (estimação de média de população finita e variância populacional conhecida).

Calculate Sample Size

PLEASE SELECT AN OPTION

POPULATION MEAN

POPULATION

FINITE
 INFINITE

Variance

KNOWN
 UNKNOWN

FORMULA

$$n = \frac{(z_{\alpha/2})^2 \sigma^2 N}{e_0^2 (N - 1) + (z_{\alpha/2})^2 \sigma^2}$$

SAMPLE SIZE IS

[1] 430.3567

FINITE POPULATION MEAN - KNOWN POPULATION VARIANCE

Population Size (N)

500

Standard Error (e₀)

2

Significance Level (α)

0.07

Population Variance (σ² ≥ 0)

2

Fonte: Elaboração própria pelos autores.

Figura 5: Cálculo do tamanho amostral (estimação de média de população infinita e variância populacional desconhecida).

Calculate Sample Size

PLEASE SELECT AN OPTION

POPULATION MEAN

POPULATION

FINITE
 INFINITE

Variance

KNOWN
 UNKNOWN

FORMULA

$$n = \frac{(t_{\alpha/2, v})^2 s^2}{e_0^2}$$

SAMPLE SIZE IS

[1] 5289.033

INFINITE POPULATION MEAN - UNKNOWN POPULATION VARIANCE

Standard Error (e₀)

2

Freedom Degree (v)

5

Significance Level (α)

0.1

Sample Variance (s² ≥ 0)

2

Fonte: Elaboração própria pelos autores.



Considerações finais

A partir da experiência adquirida na elaboração do *web app Sampling*, que calcula o tamanho amostral de populações finitas ou infinitas, pode-se sugerir que conteúdos abordados apenas teoricamente em sala de aula, podem também ser estudados de maneira mais atrativa, a partir de procedimentos estatísticos e computacionais. Conclui-se que as linguagens de programação, tais como a R, são ferramentas computacionais úteis que auxiliam na construção de aplicativos nas diferentes áreas do conhecimento científico, assim como, como instrumento didático-pedagógico.

Todos estes *web apps* serão disponibilizados na *webpage* do Núcleo de Estatística Aplicada (NEA) da Universidade de Cruz Alta.

REFERÊNCIAS

ALLAIRE, J. J.; RStudio: Integrated Development Environment for R In: The R User Conference, useR! , 2011, Coventry. **Anais...** Coventry : University of Warwick, 2011. p. 14.

HORNIK, Kurt; R FAQ, CRAN-R, 2017. Disponível em: <<https://cran.r-project.org/doc/FAQ/R-FAQ.html>>. Acesso em 03 set. 2018.

THE SHINY. The basic parts of a Shiny app. Shiny, 2017. Disponível em: <<https://shiny.rstudio.com/articles/basics.html>>. Acesso em 03 set. 2018.

BEELEY, Chris. Web Application Development with R using Shiny. New York: Packt open source, 2013.

BOLFARINE, Heleno e BUSSAB, Wilton. Elementos de Amostragem. São Paulo: Blucher, 2004.