

MAT02025 - Amostragem 1

A estimativa do tamanho da amostra

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2021



Um exemplo hipotético

Um exemplo hipotético

- ▶ No planejamento de uma pesquisa por amostragem, sempre chega-se a uma etapa em que deve ser tomada uma decisão sobre o **tamanho da amostra**.
- ▶ **A decisão é importante.**
 - ▶ Uma amostra muito grande implica em desperdício de recursos;
 - ▶ Uma amostra muito pequena diminui a utilidade dos resultados.
- ▶ A decisão nem sempre pode ser feita de forma satisfatória; frequentemente não possuímos informações suficientes para ter certeza de que nossa escolha do tamanho da amostra é a melhor.
- ▶ A **teoria da amostragem** fornece uma estrutura para pensar de forma inteligente sobre o problema.

Um exemplo hipotético

- ▶ Um exemplo hipotético mostra as etapas envolvidas para se chegar a uma solução.
- ▶ Um antropólogo se prepara para estudar os habitantes de alguma ilha.
- ▶ Entre outras coisas, ele deseja estimar a porcentagem de habitantes pertencentes ao grupo sanguíneo O.
- ▶ Foi obtida a cooperação necessária, para que seja possível obter uma **amostra aleatória simples**.
- ▶ **Qual deve ser o tamanho da amostra?**

Um exemplo hipotético

- ▶ Esta pergunta não pode ser respondida sem que antes se responda outra:
 - ▶ Com que precisão o antropólogo deseja saber a porcentagem de pessoas com grupo sanguíneo O?
- ▶ Em resposta, ele afirma que ficará satisfeito se a porcentagem estiver correta em $\pm 5\%$, no sentido de que, se a amostra mostrar que 43% têm grupo sanguíneo O, a porcentagem para toda a ilha certamente ficará entre 38 e 48.

Um exemplo hipotético

- ▶ Para evitar mal-entendidos, pode ser aconselhável apontar ao antropólogo que não podemos garantir absolutamente a precisão dentro de 5%, exceto se examinarmos todos indivíduos da ilha.
- ▶ Por maior que seja o valor de n , há uma chance de uma amostra “ **muito infeliz**” (*azarada*) tenha uma margem de erro superior aos 5% desejados.
- ▶ O antropólogo responde friamente que está ciente disso, que está disposto a ter uma chance de 1 em 20 de obter uma amostra infeliz e que tudo o que ele pede é o valor de n em vez de uma aula sobre estatística (!).

Um exemplo hipotético

- ▶ Agora podemos fazer uma estimativa aproximada de n .
- ▶ Para simplificar as coisas, a cpf é ignorada e a porcentagem amostral p é considerada normalmente distribuída.
 - ▶ Se essas suposições são razoáveis, pode ser verificado quando o n inicial é conhecido.

Um exemplo hipotético

- ▶ Em termos técnicos, p deve estar dentro dos limites de $(P \pm 5)$, exceto para uma possibilidade em 20.
- ▶ Uma vez que p é assumido normalmente distribuído em torno de P , ele estará no intervalo $(P \pm 2\sigma_p)$, exceto por uma possibilidade em vinte.
- ▶ Além disso,

$$\sigma_p \doteq \sqrt{PQ/n}.$$

- ▶ Portanto, podemos escrever que

$$2\sqrt{PQ/n} = 5 \quad \text{ou} \quad n = 4PQ/25.$$

Um exemplo hipotético

- ▶ Nesse ponto, surge uma dificuldade comum a todos os problemas de estimativa do tamanho da amostra.
- ▶ Uma fórmula para n foi obtida, mas n **depende de alguma propriedade da população** que será amostrada.
- ▶ Neste caso, a propriedade é a quantidade P que gostaríamos de medir.
- ▶ Portanto, perguntamos ao antropólogo se ele pode nos dar alguma ideia do provável valor de P .
 - ▶ Ele responde que a partir de dados anteriores sobre outros grupos étnicos, e de suas especulações sobre a história étnico-racial desta ilha, ele ficará surpreso se P estiver fora a faixa de 30 a 60%.

Um exemplo hipotético

- ▶ Esta informação é suficiente para fornecer uma resposta útil.
- ▶ Para qualquer valor de P entre 30 e 60, o produto PQ está entre 2100 e um máximo de 2500 em $P = 50$.
- ▶ O n correspondente está entre 336 e 400.
 - ▶ Por segurança, 400 é considerado como a estimativa inicial de n .

Um exemplo hipotético

- ▶ As suposições feitas nesta análise podem agora ser reexaminadas.
- ▶ Com $n = 400$ e P entre 30 e 60, a distribuição de p deve ser próxima da normal.
- ▶ Se a cpf é necessário depende do número de pessoas na ilha.
 - ▶ Se a população exceder 8000, a fração de amostragem é inferior a 5% e nenhum ajuste para cpf é necessário.

A análise do problema

A análise do problema

As principais etapas envolvidas na escolha do tamanho da amostra são as seguintes.

1. Deve haver alguma declaração sobre o que se espera da amostra.
 - ▶ Essa declaração pode ser em termos de limites de erro desejados, como no exemplo anterior, ou em termos de alguma decisão a ser tomada ou ação a ser tomada quando os resultados da amostra forem conhecidos.
 - ▶ **A responsabilidade** pelo enquadramento da declaração recai principalmente sobre **as pessoas que desejam usar os resultados da pesquisa, embora frequentemente precisem de orientação para expressar seus desejos em termos numéricos.**

A análise do problema

2. Alguma equação que conecte n com a precisão desejada da amostra deve ser encontrada.
 - ▶ **A equação irá variar de acordo com o conteúdo da declaração de precisão e com o tipo de amostragem que é contemplado.**
 - ▶ Uma das vantagens da amostragem probabilística é que ela permite que essa equação seja construída.
3. Esta equação conterá, como parâmetros, certas propriedades desconhecidas da população.
 - ▶ **Devem ser estimados para dar resultados específicos (concretos).**

A análise do problema

4. Muitas vezes acontece que os dados devem ser publicados para certas subdivisões principais da população e que os limites de erro desejados são estabelecidos para cada subdivisão.
 - ▶ **Um cálculo separado é feito para n em cada subdivisão, e o n total é encontrado pela soma.**
5. Mais de um item ou característica geralmente é medido em uma pesquisa por amostragem: às vezes, o número de itens é grande.
 - ▶ Se um grau desejado de precisão é prescrito para cada item, os cálculos levam a uma série de valores conflitantes de n , um para cada item.
 - ▶ **Algum método deve ser encontrado para reconciliar esses valores.**

A análise do problema

6. O valor escolhido de n deve ser avaliado para ver se é consistente com os recursos disponíveis para colher a amostra.
- ▶ Isso exige uma estimativa do custo, trabalho, tempo e materiais necessários para obter o tamanho proposto da amostra.
 - ▶ **Às vezes fica claro que n terá que ser drasticamente reduzido.**
 - ▶ Uma difícil decisão deve então ser tomada - se prosseguir com um tamanho de amostra muito menor, reduzindo assim a precisão, ou abandonar os esforços até que mais recursos possam ser encontrados.

A especificação da precisão

A especificação da precisão

- ▶ A declaração de precisão desejada pode ser feita fornecendo a quantidade de erro que estamos dispostos a tolerar nas estimativas de amostra.
- ▶ Este montante é determinado, da melhor maneira possível, em função dos usos que serão dados os resultados das amostras.
- ▶ Às vezes é difícil decidir quanto erro deve ser tolerado, especialmente quando os resultados têm vários usos diferentes.

A especificação da precisão

- ▶ Suponha que perguntamos ao antropólogo por que ele desejava que a porcentagem com grupo sanguíneo **O** fosse correta para 5% em vez de, digamos, 4 ou 6%.
- ▶ Ele pode responder que os dados do grupo sanguíneo devem ser usados principalmente para classificação étnico-racial.
- ▶ Ele suspeita fortemente que os insulares pertencem a um tipo étnico-racial com P de cerca de 35% ou a um com P de cerca de 50%.
- ▶ Um limite de erro de 5% na estimativa parecia-lhe pequeno o suficiente para permitir a classificação em um desses tipos.
 - ▶ Ele, entretanto, não teria objeções violentas aos limites de erro de 4 ou 6%.

A especificação da precisão

- ▶ Assim, a escolha de um limite de 5% de erro pelo antropólogo foi até certo ponto arbitrária.
- ▶ A esse respeito, o exemplo é típico da maneira pela qual um limite de erro é frequentemente decidido.
- ▶ Na verdade, o antropólogo tinha mais certeza do que queria do que muitos outros cientistas e administradores terão.
- ▶ Quando a questão do grau desejado de precisão é levantada pela primeira vez, tais pessoas podem confessar que nunca pensaram a respeito e não têm ideia da resposta.
- ▶ Entretanto, depois de discutido o assunto, eles podem frequentemente indicar pelo menos aproximadamente o tamanho de um limite de erro que lhes parece razoável.

Para casa

- ▶ Revisar os tópicos discutidos nesta aula.
- ▶ Suponha que você irá realizar um levantamento por amostragem, utilizando amostragem aleatória simples, para estimar a porcentagem da presença de uma certa característica (do seu interesse; do tipo dicotômica) em uma população alvo.
 - ▶ Defina a característica de seu interesse e as populações alvo e de estudo.
 - ▶ Defina uma margem de erro; justifique a especificação deste valor para os limites de erro.
 - ▶ Defina a declaração de confiança das futuras estimativas.
 - ▶ Estime o tamanho da amostra para este problema.
 - ▶ Elabore conclusões com respeito ao resultado, suposições necessárias e a viabilidade de se executar este levantamento por amostragem.
 - ▶ Compartilhe os seus achados no Fórum Geral do Moodle.

Próxima aula

- ▶ Tamanho de amostra para proporções e médias.

Por hoje é só!

Bons estudos!

