Semana Nacional de Ciência e Tecnologia

ICET-UFAM/2017

MINICURSO:
COMO CALCULAR UMA AMOSTRA

PROFESSOR: HIDELBRANDO FERREIRA RODRIGUES

Semana Nacional de Ciência e Tecnologia

ICET-UFAM/2017

ESTATÍSTICO: UFAM/1996

MESTRE EM ECONOMIA (DESENVOLVIMENTO REGIONAL) – UFAM/2004

CONCLUINDO DOUTORADO EM ENGENHARIA DE PRODUÇÃO - UFSC

1. Considerações Iniciais

1.1 MOTIVAÇÃO INICIAL

1.2 MOTIVAÇÃO / PRÉ-REQUISITO

1.3 TOMADA DE DECISÃO

1.4 COMO A ESTATÍSTICA PODE AJUDAR NA TOMADA DE DECISÃO

1.1 Motivação Inicial

- Afinal de contas, do que se trata exatamente o fato de necessitarmos realizar um levantamento amostral?
- O que está por trás da realização de uma pesquisa?
- Por que eu preciso saber o tamanho de uma amostra quando resolvo realizar um trabalho científico?

1.2 Motivação / Pré-requisito

- Alunos de TCC costumam me procurar para saber qual o tamanho da amostra para realizarem suas pesquisas.

- Normalmente me perguntam se um determinado número de amostras é suficiente para suas pesquisas.

1.3 A tomada de decisão

Processo

Identificação de um problema/oportunidade

Seleção de uma linha de ação para resolvê-lo

1.3 Tomada de Decisão

Palpite

Estatística

A tomada de decisão

Vários fatores afetam a tomada de decisão:

- Tempo disponível para a tomada de decisão
- A importância da decisão
- *O ambiente*
- *Certeza ou incerteza e risco*
- Agentes decisores
- Conflito de interesses

Podemos classificar a tomada de decisão de diversas formas

Nível hierárquico na empresa

Tipo de informação disponível

Número de decisores

Podemos classificar a tomada de decisão de diversas formas

Nível hierárquico na empresa

- Estratégico
- Gerencial
- Operacional

Podemos classificar a tomada de decisão de diversas formas

Tipo de informação disponível

- Estruturada
- Semiestruturada
- Não estruturada

Podemos classificar a tomada de decisão de diversas formas

Número de decisores

Decisão individual

Decisão em grupo

1.4. Como a estatística pode ajudar na tomada de decisão?

A ESTATÍSTICA ESTÁ INTERESSADA NOS MÉTODOS CIENTÍFICOS PARA COLETA, ORGANIZAÇÃO, RESUMO, APRESENTAÇÃO E ANÁLISE DOS DADOS, BEM COMO NA OBTENÇÃO DE CONCLUSÕES VÁLIDAS E NA TOMADA DE DECISÕES BASEADAS NAS ANÁLISES

(SPIEGEL, 2009).

2. A Pesquisa Estatística

2.1 CLASSIFICAÇÃO DA PESQUISA ESTATÍSTICA

2.2.FASES DA PESQUISA ESTATÍSTICA

2.3. EXPLORANDO CONCEITOS DA ESTATÍSTICA

2.4 TIPOS DE VARIÁVEIS

2.5 CONFIABILIDADE

2.6 SOFTWARE ESTATÍSTICO

2.1 Classificação da Pesquisa Estatística

Alguns autores classificam a pesquisa estatística em pesquisa de levantamento e pesquisa por experimento.

[Crespo (2002) e Reis e Lino (2013)]

2.1 Classificação da Pesquisa Estatística

- Na pesquisa de levantamento são observadas as características ou fenômenos presentes na população, com mínima intervenção por parte do pesquisador.

- Nesse tipo de pesquisa, não é possível evidenciar relações de causa e efeito, apenas se medem as variáveis em estudo, sendo possível somente afirmar relações entre elas.

2.1 Classificação da Pesquisa Estatística

- Na pesquisa experimental, o pesquisador tem um controle sobre as condições de pesquisa, conseguindo eliminar quase todas as causas de variação mediante um planejamento do experimento.
- Dessa forma, a pesquisa experimental é o único tipo de pesquisa que permite selecionar as variáveis capazes de influenciar a característica em estudo e definir as formas de controle e de observação dos efeitos que essas variáveis produzem nessa característica.

2.1 Classificação da Pesquisa Estatística

- Alguns autores, entre os quais Crespo (2002) e Reis e Lino (2013), classificam a pesquisa estatística em pesquisa de levantamento e pesquisa por experimento.
- Na pesquisa de levantamento são observadas as características ou fenômenos presentes na população, com mínima intervenção por parte do pesquisador. Nesse tipo de pesquisa, não é possível evidenciar relações de causa e efeito, apenas se medem as variáveis em estudo, sendo possível somente afirmar relações entre elas.

2.2.Fases da pesquisa estatística

a) definição do problema: determinação do que se quer pesquisar;

b) planejamento: definição dos procedimentos necessários para o desenvolvimento da pesquisa;

c) coleta de dados: obtenção, reunião e registro sistemático de dados, podendo ser direta ou indireta;

d) apuração dos dados: resumo dos dados, mediante a sua contagem ou agrupamento;

e) apresentação dos dados: pode ser por meio de tabelas, gráficos ou medidas;

f) análise e interpretação dos dados: conclusões que auxiliam na resolução do problema.

2.3. Explorando conceitos da estatística

População

- População é o conjunto de todos os itens, objetos ou pessoas, enfim, entes que se pretende analisar – razão pela qual devem apresentar pelo menos uma propriedade em comum.
- Uma população pode ser formada por pessoas, famílias, estabelecimentos comerciais ou industriais, contas-correntes, peças de uma linha de produção ou qualquer outro tipo de elemento.

2.3 Explorando conceitos da estatística

Amostra

- Quando se analisa uma parte dessa população, tem-se uma "amostra", que é subconjunto do todo.
- A amostra é definida previamente e obtida com a consideração de alguns critérios, a fim de que seja significativa (quanto ao número de elementos) e se mostre representativa, apresentando as mesmas propriedades da população.

2.3 Explorando conceitos da estatística

As melhores técnicas de amostragem são aquelas que usam a ideia de aleatoriedade

2.3 Explorando conceitos da estatística

- Ao se realizar um levantamento de dados, deve ser definido se este será realizado em toda a população (“censo”) ou em uma amostra dessa população (“levantamento por amostragem”).

- De maneira geral, opta-se pelo levantamento por amostragem por ser mais rápido e mais econômico.

2.3 Explorando conceitos da estatística

- É bastante usual representar as observações de uma característica presente em toda a população por uma medida denominada de "parâmetro", obtida a partir de um conjunto de observações de todo o grupo, ou seja, a população.

- Os mais utilizados para representar a população são: a média (μ), a variância (σ^2) e a proporção de sucesso (π).

- Tratando-se de uma amostra, tem-se a "estatística", que é uma medida obtida para um subconjunto de observações ou amostra.

- As estatísticas mais usuais para representar dados amostrais são as mesmas usadas para representar populações: a média ($\bar{x}$), a variância (s^2) e a proporção de sucesso (p), sendo representadas, entretanto, por símbolos distintos por apresentarem propriedades diferentes.

2.3 Explorando conceitos da estatística

Um parâmetro é uma constante, enquanto uma estatística é uma variável aleatória

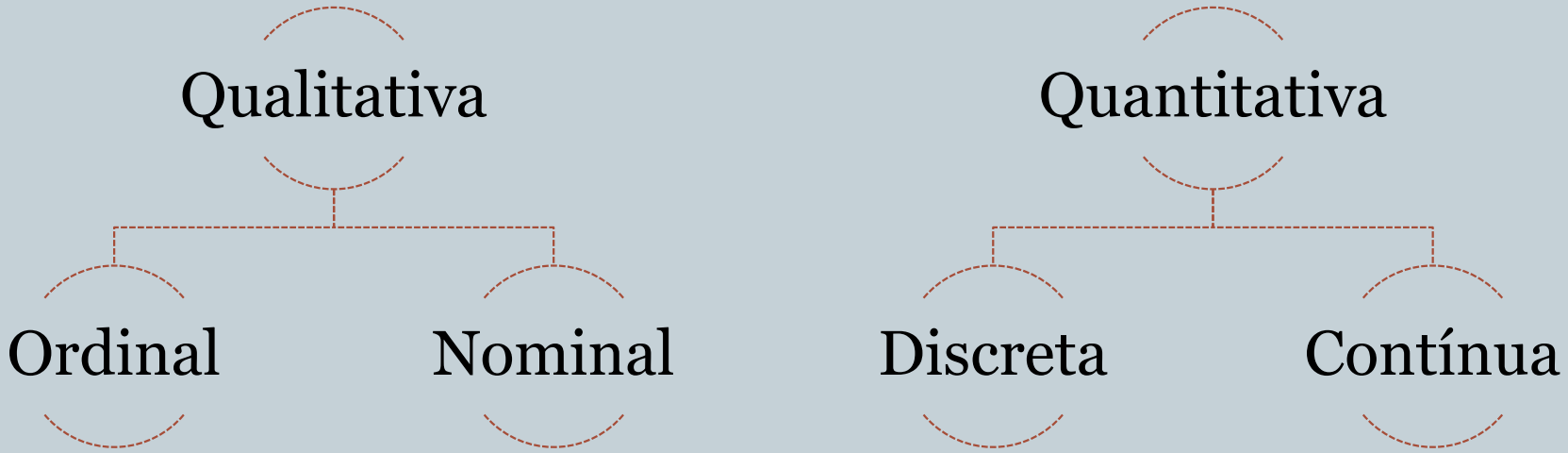
2.4 Tipos de variáveis

- Variáveis são características estudadas em uma população ou amostra que podem assumir diferentes valores numéricos ou categóricos.

- Dados ou observações são as informações inerentes às variáveis que caracterizam os elementos que constituem a população ou a amostra em estudo.

- Por exemplo: quilometragem é uma variável; 530 km é um dado.

2.4 Tipos de variáveis



2.5 Confiabilidade dos dados

- Ao resumir um conjunto de dados, deve-se observar se são consistentes, ou seja, verificar se os resultados estão de acordo com o esperado, ou se alguma mudança relativamente forte pode ter ocorrido.
- Muitas vezes, tem-se uma expectativa em relação aos dados, porém os resultados e a análise se apresentam diferentes do esperado.
- Tal fato de forma alguma invalida o estudo, mas muito provavelmente contradirá uma ou mais hipóteses iniciais.

2.5 Confiabilidade dos dados

- Durante a coleta de dados, deve-se sempre ter em mente que estes dependem do método utilizado.
- Muitas vezes, especialmente na área da Engenharia, os dados são coletados por meio de um equipamento de medição, que é um dos elementos que constituem um sistema de medição.
- O pesquisador deve saber como usá-lo corretamente, além de saber como analisar e interpretar os resultados.
- Entretanto, devido às diversas fontes de variação que afetam um sistema de medição, leituras repetidas sobre a mesma peça, por exemplo, podem não produzir um mesmo e idêntico resultado.
- Dessa forma, os efeitos das distintas fontes de variação sobre o sistema de medição devem ser avaliados.

Confiabilidade dos dados

- Uma das ferramentas utilizadas com essa finalidade são os estudos de “repetibilidade” e “reprodutibilidade” (R&R).

- Nos estudos de R&R, o instrumento de medição é utilizado para medir, repetidas vezes, a grandeza de interesse.

2.5 Confiabilidade dos dados

- Conforme o VIM (2012), podemos definir a repetibilidade como: condição de medição num conjunto de condições, as quais incluem o mesmo procedimento de medição, os mesmos operadores, o mesmo sistema de medição, as mesmas condições de operação e o mesmo local, assim como medições repetidas no mesmo objeto ou em objetos similares durante um curto período de tempo.

2.5 Confiabilidade dos dados

- Já a reprodutibilidade é definida como: condição de medição num conjunto de condições, as quais incluem diferentes locais, diferentes operadores, diferentes sistemas de medição e medições repetidas no mesmo objeto ou em objetos similares (VIM, 2012).

2.5 Confiabilidade dos dados

- A repetibilidade estima a variabilidade dos resultados obtidos por um mesmo operador, usando o mesmo equipamento de medição segundo um mesmo procedimento.
- Pode-se interpretar a repetibilidade como a menor variabilidade que pode ser entregue por um sistema de medição.
- Uma repetibilidade ruim pode estar relacionada à manutenção falha ou à falta de robustez do equipamento de medição, local inadequado para a realização da medição, ou, ainda, a uma variação dentro da própria amostra.

2. 5 Confiabilidade dos dados

- Por outro lado, a reprodutibilidade avalia, caso diferentes operadores sejam envolvidos no estudo, a variabilidade das médias das medições feitas por operador.
- Tem-se, assim, uma indicação da robustez do sistema de medição, do grau de treinamento dos operadores, da uniformidade dos métodos empregados pelos operadores (AMERICAN SOCIETY FOR QUALITY CONTROL, 2010).

2.5 Confiabilidade dos dados

Em suma, o pesquisador deve estar bastante atento aos seus dados e, em especial, a como foram coletados

2. 6 Software estatístico

- O *software* estatístico é de suma importância tanto no meio acadêmico como no empresarial, quer pela sua facilidade de utilização, quer pela eficiência no tratamento de grandes conjuntos de dados.
- Atualmente, há uma concordância entre os educadores de que as disciplinas de estatística devem ser auxiliadas por algum tipo de tecnologia para reduzir a necessidade de execução de cálculos manuais e fornecer aos alunos o acesso a conjuntos de dados de casos práticos e situações reais (BECKER, 1996; MALTBY, 2001; SCHUYTEN; THAS, 2007; VERZANI, 2008; GOULD, 2010).
- Entretanto, o uso de um *software* estatístico, em processos de aprendizagem, deve ser fundamentado por certo conhecimento das técnicas estatísticas envolvidas ou conduzido por quem possui esses conhecimentos (ALVES; CUNHA, 2013).

2. 6 Software estatístico

- Existem diversos *softwares* estatísticos, alguns de uso gratuito.
- Entre os gratuitos destacam-se: INSTAT, Biostat, R e SEstatNet.
- Entre os pagos destacam-se: Minitab, SAS, SPSS e Statistica.

2. 6 Software estatístico

- O INSTAT é um pacote estatístico geral. É simples o suficiente para ser útil no ensino da estatística, como também pode auxiliar a pesquisa que requer uma análise de dados. Tem sido amplamente utilizado no Reino Unido e em outros lugares por uma série de empresas, institutos de pesquisa, escolas, faculdades e universidades. Também tem sido utilizado como apoio em cursos de estatística e cursos relacionados a agricultura, saúde e climatologia (INSTAT, 2013).

2. 6 Software estatístico

- O Biostat é um *software* desenvolvido por professores da Universidade Federal do Pará e possui vários pacotes estatísticos de fácil aplicabilidade para iniciantes. De acordo com Ayres et al. (2007), autores do *software*, seu uso é bastante facilitado, principalmente pelo *help* do programa, que, na verdade, é um livro de apoio. Nele podem ser encontrados diversos testes estatísticos, dimensionamento de amostras e tipos de experimento. Ainda, para facilitar sua aprendizagem e consequente utilização, cada procedimento abordado em cada pacote vem acompanhado de um exemplo.

2. 6 Software estatístico

- O Minitab é um *software* estatístico que se caracteriza pela simplicidade e *interface* amigável.

- Conforme Alves e Cunha (2013), esse *software* oferece um grande número de procedimentos para análise estatística de dados, permitindo obter estatísticas descritivas, simulações e distribuições, inferência estatística, análise da variância, regressão, análise de dados categóricos, testes não paramétricos, análise de séries temporais, ferramentas de planejamento de experimentos e ferramentas de controle de qualidade, bem como vários tipos de gráficos.

2. 6 Software estatístico

- O SAS (*Statistical Analysis System*) é um sistema integrado de aplicações para o processamento e a análise estatística de dados. De acordo com Alves e Cunha (2013), o programa trabalha com quatro formas básicas sobre os dados: acessar, manipular, analisar e apresentar. O núcleo de todo o sistema é o SAS *base*, o qual permite criar tabelas e proceder à manipulação dos dados. Um de seus pontos fortes é o fornecimento de ferramentas para simulação.

2. 6 Software estatístico

- O SPSS (Statistical Package for the Social Sciences) é um software para análise estatística de dados.
- Possui uma interface muito amigável, que se utiliza de menus, janelas e caixas de diálogo e permite realizar cálculos complexos e visualizar seus resultados de forma simples. Inicialmente, o SPSS foi desenvolvido para tratar dados da área de ciências sociais, mas atualmente é muito utilizado nos meios acadêmico e empresarial (ALVES; CUNHA, 2013).
- É comercializado em módulos, e o mais simples, o módulo básico, permite executar tabulações, cruzamentos, gráficos, cálculo de medidas, testes de hipóteses paramétricos e não paramétricos, regressão e correlação, entre outros.

2. 6 Software estatístico

- O software Statistica é desenvolvido pela Statsoft.
- Conforme Alves e Cunha (2013), é um aplicativo que inclui estatísticas descritivas, correlações, testes t e outros testes para as diferenças entre grupos, tabelas de frequências, cruzamentos, métodos de regressão múltipla, métodos não paramétricos, rotinas de Anova/Manova, módulos de ajustamento das distribuições e um vasto conjunto de ferramentas para gráficos.
- Ainda contempla módulos adicionais compostos por modelos lineares/não lineares avançados, técnicas exploratórias multivariadas, gráficos para controle de qualidade, índices de capacidade de processo e planejamento de experimentos, dentre outros (ALVES; CUNHA, 2013).
- Esse software se destaca pela qualidade dos gráficos construídos.

3. Revelando o Poder dos Dados

"VOCÊ NÃO PODE CORRIGIR COM A ANÁLISE O QUE VOCÊ ESTRAGOU NO PLANEJAMENTO."

3.1 Como responder às seguintes perguntas:

- Há um “gene de explosão”?
- A marcação de pinguins para identificação os machuca?
- Os humanos, subconscientemente, liberam sinais químicos (feromônios)?
- Qual proporção das pessoas que lavam as mãos depois de usar um banheiro público?
- Se os pais pudessem voltar atrás no tempo, ainda escolheriam ter filhos?

3.1 Como responder às seguintes perguntas:

- Por que aranhas adolescentes se envolvem em jogos preliminares?
- Qual percentual de professores de faculdade se consideram professores “acima da média”?
- O fato de os jogadores de um time baterem várias vezes nas mãos uns dos outros ajuda o time a vencer?
- O que é melhor para o pico do desempenho: um pequeno e suave aquecimento ou um longo e intenso aquecimento?
- Para os homens, a cor vermelha aumenta a atratividade das mulheres?

3.1 Como responder às seguintes perguntas:

- Os moradores de cidades são mais propensos que os moradores do campo a ter distúrbios de humor e ansiedade?
- Há verdade em se dizer “beleza adormecida”?
- As clínicas de bronzeamento orientam mal seus clientes em relação aos perigos do bronzeamento?
- Qual percentual dos jovens adultos nos Estados Unidos que voltam a morar com os pais?
- O aumento do volume da música em um bar faz com que as pessoas bebam mais cerveja?

3.2 Prêmio IgNoble

- 1991 - **Química**: Jacques Benveniste, produtivo e dedicado correspondente da revista *Nature*, pela persistência na descoberta de que a água é um líquido inteligente e por demonstrar que ela é capaz de se lembrar de acontecimentos muito depois de todos os vestígios desses acontecimentos terem desaparecido.

3.2 Prêmio IgNoble

- 1992 - **Medicina**: F. Kanda, E. Yagi, M. Fukuda, K. Nakajima, T. Ohta, e O. Nakata do *Shisedo Research Center* em Yokohama, por seu estudo pioneiro "*Elucidação dos Componentes Químicos Responsáveis pelo Chulé do Pé*" (*Elucidation of Chemical Compounds Responsible for Foot Malodour*), especialmente pela conclusão de que as pessoas que pensam que têm chulé, têm, e as que pensam que não têm, não têm.

3.2 Prêmio IgNoble

- **Psicologia**: John Edward Mack da Escola de Medicina de Harvard e David M. Jacobs da Universidade Temple, pela sua conclusão de que as pessoas que acreditam terem sido raptadas por extraterrestres provavelmente o foram—e especialmente pela sua conclusão "o objectivo da abdução é a produção de filhos".

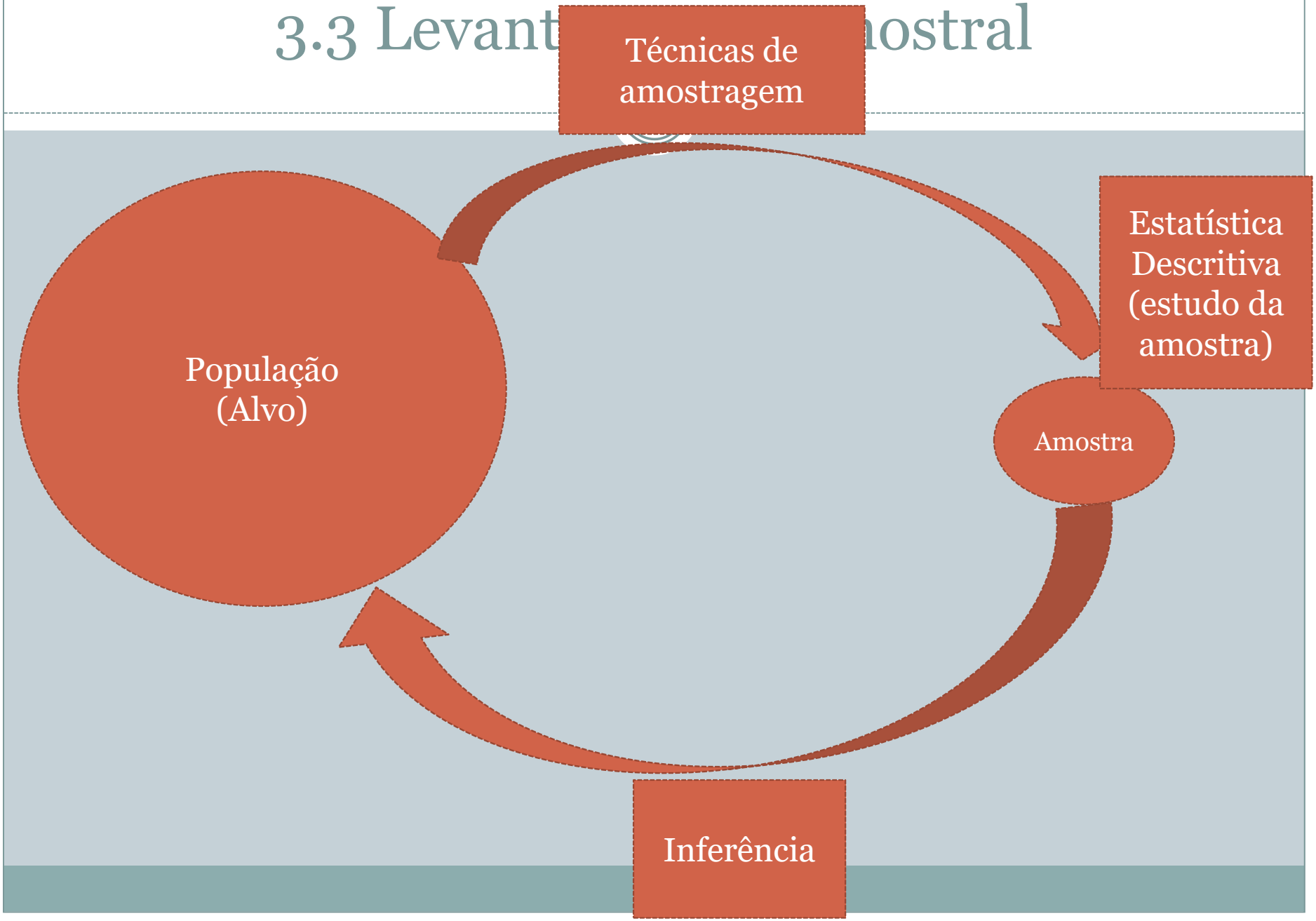
3.2 Prêmio IgNoble

- 2017 - **Física**: Marc-Antoine Fardin, por utilizar a dinâmica dos fluidos para provar sua tese "*Pode um Gato ser Ambos, um Sólido e um Líquido?*".
- 2017 - **Anatomia**: James Heathcote, por seu estudo de pesquisa médica "*Porque os Velhos têm Orelhas Grandes?*".
- 2017 - **Dinâmica dos fluidos**: Jiwon Han, por seu estudo da dinâmica de um líquido, para saber o que acontece quando uma pessoa caminha para trás enquanto carrega uma xícara de café.

3.2 Prêmio IgNobel

- 2017 - **Nutrição**: Fernanda Ito, Enrico Bernard e Rodrigo Torres, pelo primeiro relatório científico "*Sangue Humano na Dieta do Morcego-Vampiro-de-Perna-Peluda*".
- 2017 - **Obstetrícia**: Marisa López-Teijón, Álex García-Faura, Alberto Prats-Galino e Luis Pallarés Aniorte, por demonstrarem em seu estudo "*Expressão Facial do Feto em Resposta à Emissão Intravaginal de Música*", que um feto humano em desenvolvimento reage mais fortemente à música que é tocada eletromecanicamente dentro da vagina da mãe do que quando é tocada eletromecanicamente na sua barriga.

3.3 Levant ostral



4. Plano de Amostragem

4. Vamos praticar

1. Vamos definir nossa população alvo =>Participantes do minicurso “Como calcular uma amostra”

1. Objetivos:

   1º. Levantar informações socioeconômicas dos participantes;

   2º. Levantar informações: Obter informações antropométricas dos participantes do minicurso “Como calcular uma amostra”

Plano de Amostragem

- Na prática da pesquisa em geral, o tamanho da amostra parece sintetizar todas as questões relacionadas ao processo de amostragem.

- E, às vezes, esse aspecto ainda é traduzido pelas clássicas questões: Que porcentagem da população deve ser observada? 5 ou 10 % será significante?

- Elaborar um plano de amostragem significa construir soluções não apenas para determinar o número de unidades a serem observadas, mas também para determinar as condições concretas para operacionalizar os planos de observação e de análise de dados definidos em etapas anteriores previstas no processo de levantamento.

Plano de Amostragem

O plano de amostragem deverá, portanto, atender as seguintes exigências:

a) Orientar-se pelos objetivos e hipóteses da pesquisa, ajustando-os aos recursos e à realidade que será efetivamente observada.

b) Servir como base necessária para o cumprimento da inferência estatística, ou seja, possibilitar a relação dos resultados amostrais (estimativa) com os valores populacionais desconhecidos, alvos da pesquisa.

Plano de Amostragem

O plano de amostragem deverá, portanto, atender as seguintes exigências:

c) Conter instruções claras, completas e factíveis que, além de facilitar as tarefas operacionais envolvidas, possam, defender seu suporte teórico probabilístico.

d) Ser eficiente, isto é, viabilizar a coleta, o processamento e a análise de dados dentro dos níveis de precisão especificados e do menor custo possível.

DETERMINAÇÃO DO TAMANHO DE UMA AMOSTRA COM BASE NA ESTIMATIVA DA MÉDIA POPULACIONAL

- Os valores de confiança mais utilizados e os valores de Z correspondentes podem ser encontrados na Tabela 1:

Tabela 1 – Valores críticos associados ao grau de confiança na amostra

Grau de Confiança	α	Valor Crítico $Z_{\alpha/2}$
90%	0,10	1,645
95%	0,05	1,96
99%	0,01	2,575

DETERMINAÇÃO DO TAMANHO DE UMA AMOSTRA COM BASE NA ESTIMATIVA DA MÉDIA POPULACIONAL

EXEMPLO 1

Um economista deseja estimar a renda média para o primeiro ano de trabalho de um bacharel em direito. Quantos valores de renda devem ser tomados, se o economista deseja ter 95% de confiança em que a média amostral esteja a menos de R$500,00 da verdadeira média populacional? Suponha que saibamos, por um estudo prévio, que para tais rendas, σ = R$6250,00.

$$n = \left(\frac{Z_{\alpha/2}}{E} \right)^2 = \left(\frac{1{,}96 \cdot 6250}{500} \right)^2 = 600{,}25 = 601 \quad (Arredondado\ para\ cima)$$

EXPERIÊNCIA: *Baseado nos dados do EXEMPLO 1, utilize uma margem de erro maior, como R$1.000,00 e determine qual seria o tamanho da amostra necessário nesta situação.*