

Centro de Ciências Agrárias e Ambientais da UFBA
Departamento de Engenharia Agrícola

Disciplina: AGR116 – Bioestatística

Professor: Celso Luiz Borges de Oliveira

Assunto: Estatística Descritiva

Tema: Amostragem, Dimensionamento de Amostras, Estimação de Parâmetros

RESUMO E NOTAS DA AULA Nº 03

A maneira de se obter a amostra é tão importante, e existem tantos modos de fazê-lo, que estes procedimentos constituem uma especialidade dentro da Estatística, conhecida como Amostragem. Estes procedimentos podem ser agrupados em dois grandes grupos: as Amostragens Probabilísticas e as Não-Probabilísticas. Ambas têm suas vantagens e desvantagens. A grande vantagem das amostras probabilísticas é medir a precisão da amostra obtida, baseando-se no resultado contido na própria amostra. Tais medidas são bem mais difíceis para os procedimentos das não-probabilísticas.

As principais técnicas de amostragem probabilística são a amostragem casual simples, a sistemática, por meio de conglomerados, a estratificada e a amostragem múltipla. Trataremos com mais ênfase a Amostragem Casual Simples.

As técnicas de amostragem não-probabilísticas incluem os seguintes casos: inacessibilidade a toda a população, a esmo ou sem norma, população formada por material contínuo e amostragens intencionais.

- 1) Amostragem Casual Simples – também chamada de simples ao acaso, aleatória, casual simples, elementar ou randômica; é equivalente a um sorteio lotérico. Nela todos os elementos da população têm igual probabilidade de pertencer à amostra, e todas as possíveis amostras têm também igual probabilidade de ocorrer.

Sendo **N** o número de elementos da população e **n** o número de elementos da amostra, cada elemento da população tem probabilidade n/N de pertencer à amostra. A essa relação n/N denomina-se fração de amostragem ($f = n/N$).

Quando a população é finita ou a amostragem é feita sem reposição esta fração de amostragem, $f \geq 0,05$ (f é maior ou igual a 5%). No caso da população ser infinita ou haver reposição da amostra, o valor de $f < 0,05$ (f é menor que 5%).

Dimensionamento da Amostra

Ao dimensionarmos uma amostra, necessitamos do conhecimento prévio da variância da população e do grau de precisão para a amostragem.

Geralmente, não se tem conhecimento sobre a variabilidade da população e o recomendável é fazer um levantamento prévio, utilizando uma amostra piloto ou pré-amostragem, objetivando ter alguma informação sobre a variância populacional.

O tamanho da amostra n para se estimar uma média \bar{X} pode ser calculado a partir da fórmula:

$$t = \frac{\bar{x} - \bar{X}}{\frac{s}{\sqrt{n_0}}} = \frac{d}{\frac{s}{\sqrt{n_0}}} \therefore n_0 = \frac{t^2 s^2}{d^2} \text{ (para população infinita) em que;}$$

d = grau de precisão ou erro de amostragem desejado;

s^2 = variância da amostra prévia;

t = é o valor da amplitude total estudentizada encontrada na Tabela da Distribuição de t Student com o nível de significância e graus de liberdade ($n-1$).

No caso da população finita faz-se a correção pela fórmula:

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}, \text{ que é o tamanho da amostra.}$$

Passos:

- 1) Preliminarmente a amostra é dimensionada para a população infinita, obtendo-se o tamanho n_0 e, em uma segunda etapa, corrigimos para a população finita obtendo-se n ;
- 2) Se n_0 for inferior a 5% da população, é desnecessário proceder a correção de n_0 ;
- 3) Podemos, sem incorrer em erros grosseiros, para casos de dimensionamento que não se exige muito rigor, tomar para 0,05 (5%), $t = 2$, assim;

$$n_0 = 4s^2/d^2$$

- 4) O valor de d poderá ser fixado ou definido pela fração da estimativa da média, p.ex., $d = 0,10 \bar{x}$ ou $d = 0,05 \bar{x}$;
- 5) A amostra pode também ser dimensionada pelo Coeficiente de Variação, em que;

$$CV = \frac{s}{\bar{x}} \quad e \quad CV_{\bar{x}} = \frac{s(\bar{x})}{\bar{x}}; \quad \text{teremos} \quad n = \frac{(CV_x)^2}{(CV_{\bar{x}})^2}$$

Intervalo de Confiança

As populações são descritas por pelas medidas de certas características chamadas de parâmetros (média - μ , variância - σ^2 , desvio padrão - σ), etc.). As amostras são descritas pelas mesmas características, mas são chamadas de estimadores dos parâmetros ou estatísticas (média - \bar{x} , variância - s^2 , desvio padrão - s , etc.). As estimativas são os valores assumidos pelas grandezas amostrais (ou seja, pelos estimadores). A estimação é o processo pelo qual se obtém informações sobre a população a partir de amostras.

- 1) Estimação por Ponto - consiste em fornecer a melhor estimativa possível para o parâmetro, que será estimado por um valor único, o qual corresponde a um ponto sobre o eixo de variação da variável.

- a) Estimação por ponto da média da população - o melhor estimador de que dispomos para a média da população é a média da amostra \bar{x} , que é um estimador justo de μ , representado por:

$$\bar{x} = \frac{\sum x_i}{n}$$

- b) Estimação por ponto da produção total $\longrightarrow \hat{X} = N \cdot \bar{x}$

- 2) Estimação por Intervalo - é o intervalo que, com probabilidade conhecida, deverá conter o valor real do parâmetro. À probabilidade, designada por $1 - \alpha$, de que o intervalo contenha o valor do parâmetro chamaremos **nível ou grau de confiança** do intervalo (90%, 95%, 99% e 99,9%). À probabilidade α chamamos de **nível de significância** (10%, 5%, 1% e 0,1%). Os valores de α estão contidos na Tabela de t - Amplitude Total Estudentizada.

- a) Estimação por intervalo da média da população infinita ($f < 0,05$)

$$I.C(\bar{X}) : \bar{x} \pm t_{\alpha(n-1)} \cdot s(\bar{x})$$

$$P(\bar{x} - t_{\alpha(n-1)} \cdot s(\bar{x}) < \bar{X} < \bar{x} + t_{\alpha(n-1)} \cdot s(\bar{x})) = 1 - \alpha$$

- b) Estimação por intervalo da média da população finita ($f \geq 0,05$)

$$I.C(\bar{X}) : \bar{x} \pm t_{\alpha(n-1)} \cdot s(\bar{x}) \cdot \sqrt{\frac{N-n}{N-1}}$$

Obs: No caso de população finita, acrescenta-se o **Fator de Correção (F.C.)**.

$$F.C = \sqrt{\frac{N-n}{N-1}}$$

- d) Estimação por intervalo da produção total

$$I.C(X) : N [\bar{x} \pm t_{\alpha(n-1)} \cdot s(\bar{x})]$$

- 3) Erro de Amostragem (ou Grau de Precisão) para a Média - é a diferença entre a média da amostra e a média da população; $d = \bar{x} - \bar{X}$. Quando a média da população não é conhecida – no caso de populações muito grandes ou quando a população é infinita – utiliza-se um valor percentual da média (10%, 5% ou 1%), representa-se o erro de amostragem por, $d = 0,10 \bar{x}$, $d = 0,05 \bar{x}$ ou $d = 0,01 \bar{x}$.

EXERCÍCIO DE APLICAÇÃO DE CONHECIMENTOS

Um canal de 56 ha foi subdividido em áreas de 1 ha. Após o corte dos colmos obteve-se o seguinte resultado, em t/ha.

79	60	79	73	76	86	84	58
84	70	74	77	84	81	74	65
74	54	72	84	87	69	87	92
78	87	67	90	83	72	70	84
74	81	66	77	78	81	62	67
64	82	86	75	64	90	59	93
70	69	73	83	81	71	62	57

- 1) a) Calcular a produção total de colmos do canal.
- b) Calcular a produção média de colmos do canal.

- 2) Usando a amostragem aleatória simples, obtenha uma amostra de tamanho 10 ($n = 10$) – utilize os valores **sublinhados**, e calcule:
 - a) média
 - b) variância
 - c) desvio padrão
 - d) Coeficiente de variação
 - e) desvio padrão da media
 - f) coeficiente de variação da media

- 3) Calcule também;
 - a) a estimativa da média de produção de colmos do canal, por ponto e por intervalo, com 5% de significância. Verificar se a média verdadeira está contida nesse intervalo.
 - b) o erro de amostragem para a média.
 - c) as estimativas do total da produção de colmos, por ponto e por intervalo, com 95% de confiança. Verificar se o valor do total verdadeiro está contido nesse intervalo.
 - d) o erro de amostragem para a produção total de colmos.
 - e) o tamanho para novas amostras, com 95% de confiança, admitindo para o trabalho de amostragem os seguintes graus de precisão:
 - e.1) 10% da média
 - e.2) 5% da média
 - e.3) 1% da média
 - f) Qual a estimativa do erro de amostragem para a amostra prévia ($n = 10$), admitindo-se 5% de significância para esses resultados.

“Um homem nada faria, se para principiar a fazer as coisas esperasse até fazê-las com tal perfeição que ninguém lhes acharia defeito”
Cardeal Newmann

GLOSSÁRIO

Amostragem - processo adequado de obtenção de amostras com garantia de representatividade da população.

Amostragem Probabilística - reúnem todas as técnicas que utilizam mecanismos aleatórios de seleção dos elementos da amostra, atribuindo a cada um deles uma probabilidade, conhecida *a priori*, de pertencer à amostra, isto é, se todos os elementos da população tiverem probabilidade conhecida, e diferente de zero, de pertencer à amostra.

Amostragem Casual Simples - também chamada de amostra simples ao acaso, aleatória, casual, elementar, randômica, e é equivalente a um sorteio lotérico. Nela todos os elementos têm igual probabilidade de pertencer à amostra, e todas as possíveis amostras têm também igual probabilidade de ocorrer.

Amostragem Sistemática - quando os elementos da população se apresentam ordenados e a retirada dos elementos da amostra é feita periodicamente..

Amostragem por meio de conglomerados - consiste em sortear um número suficiente de conglomerados, cujos elementos constituirão a amostra. As unidades de amostragem são os conglomerados e não os elementos individuais da população. É utilizada quando a população apresenta uma subdivisão em pequenos grupos é possível – e muitas vezes conveniente – recorrer a esse tipo de amostragem.

Amostragem Estratificada - consiste em especificar quantos elementos da amostra serão retirados em cada estrato. São considerados três tipos: **uniforme** – quando há igual número de elementos em cada estrato; **proporcional** – o número de elementos existentes no estrato e; **ótima** – toma em cada estrato um número de elementos proporcional ao número de elementos do estrato e também à variação da variável de interesse no estrato, medida pelo seu desvio-padrão.

Amostragem Múltipla - a amostra é retirada em diversas etapas sucessivas e tem por finalidade diminuir o número médio de itens inspecionados em longo prazo, baixando assim o custo de inspeção.

Amostragem Não-Probabilística - são também empregadas em trabalhos estatísticos, por simplicidade ou por impossibilidade de se obterem amostras probabilísticas, como seria desejável.

Amostragem por inacessibilidade a toda a população - em determinadas circunstâncias somos forçados a colher a amostra na parte da população que nos é acessível. Daí surge uma distinção entre **população-objeto** e **população amostrada**. A população-objeto é a que temos em mente ao realizar o trabalho estatístico, mas somente parte dessa população está acessível para a retirada da amostra. Esta parte é a população-amostrada.

Amostragem a esmo ou sem norma - é aquela em que o amostrador, para simplificar o processo, procura ser aleatório sem, no entanto, realizar propriamente o sorteio usando algum dispositivo aleatório confiável.

Amostragem população formada por material contínuo - Nesse caso é impossível realizar a amostragem probabilística, devido à impraticabilidade de um sorteio rigoroso. Se a população for líquida ou gasosa, o que se costuma fazer, com resultado satisfatório, é homogeneizá-la e retirar a amostra a esmo. Às vezes pode ser usado no caso de material sólido.

Amostragem amostragens intencionais - o amostrador deliberadamente escolhe certos elementos para pertencer à amostra, por julgar esses elementos representativos da população. .

BIBLIOGRAFIA CONSULTADA

Barbetta, P. A. *Estatística Aplicada às Ciências Sociais*. Florianópolis. 3ª ed. Editora DA UFSC. 1999.284p.

Costa, J. A. *Notas e Resumos de aulas da disciplina AGR116-Bioestatística*. DEA/AGRUFBA.

Costa, S. F. *Introdução Ilustrada à Estatística*. São Paulo. 3ª ed. Editora Harbra Ltda. 1998. 313p.

Costa Neto, P. L. de O. *Estatística*. Ltda. São Paulo 17ª ed. Editora Edgard Blücher. 1999. 264p.

Regazzi, A. J. *Curso de Iniciação Estatística: Roteiro de Aulas*. Viçosa. DPI/UFV. 1997. 138p.

Santos, J. W. dos; Gheyi, H. R. (Eds.) *Estatística Experimental Aplicada*. Campina grande: Editora Gráfica Marcone Ltda, 2003. 213p. Tópicos de Engenharia Agrícola e Agronômica.

Wonnacott, R. J. e Wonnacott, T. H. *Fundamentos de Estatística*; tradução de Alfredo Alves de Farias. Rio de Janeiro. LTC – Livros Técnicos e Científicos Editora S.A., 1985. 357p.