

Análise de dados Biotecnológicos

Saulo Henrique Weber (organizador)

PREFÁCIO

Caro leitor, é com grande entusiasmo e gratidão que compartilho com você este livro, que representa a realização de um sonho que venho cultivando ao longo dos anos. Durante os meus 15 anos de carreira na docência, tenho tido a honra de trabalhar com disciplinas voltadas para a estatística e bioestatística, áreas que sempre me fascinaram pela sua capacidade de extrair significado de dados complexos.

Desde o início da minha jornada como professor, descobri uma paixão por criar materiais de estudo que fossem não apenas informativos, mas também envolventes e contextualizados. Começando com simples resumos para as aulas, expandi esses materiais para apresentações de slides detalhados e, eventualmente, para apostilas mais abrangentes. Ao longo do tempo, percebi a importância de conectar os conceitos teóricos com exemplos práticos, utilizando minhas experiências em pesquisa, que abrangiam desde estudos sobre árvores e peixes até espécies com potencial agronômico.

Minha trajetória como professor de universidade me levou a mergulhar em diversas áreas da ciência animal, explorando temas como genética, fisiologia, nutrição e comportamento. Sempre busquei abordar os temas de forma contextualizada, criando exemplos que ressoassem com as experiências e formações dos meus estudantes.

Recentemente, testemunhei a concretização de um sonho antigo com a criação da disciplina de análise de dados biotecnológicos. Nessa disciplina, finalmente pude proporcionar aos meus alunos a oportunidade de vivenciar todas as etapas de uma pesquisa, desde o planejamento até a interpretação dos resultados, utilizando dados de experimentos que eles próprios realizaram. Essa abordagem imersiva permitiu-lhes compreender não apenas os conceitos teóricos, mas também a importância prática e o rigor exigido na análise de dados em contextos biotecnológicos.

O material inicialmente concebido como uma apostila foi refinado e aprimorado ao longo dos anos, e hoje, com imenso orgulho, transforma-se neste livro que você tem em mãos. Minha esperança é que este livro possa servir como um guia valioso para estudantes e profissionais que buscam compreender os fundamentos e as aplicações da análise de dados biotecnológicos, capacitando-os a enfrentar os desafios dessa área com confiança e habilidade.

Que este livro seja não apenas uma fonte de conhecimento, mas também um catalisador para novas descobertas e conquistas no vasto e fascinante campo da Biotecnologia.

Com os melhores votos, reitero meu apreço e consideração.

Saulo Henrique Weber



1

PLANEJAMENTO DE PESQUISA

Karen Sumire Kubo
Saulo Henrique Weber

1.1 Estatística

A palavra “estatística”, originalmente, referia-se a *status*, que significa “estado”. Um dos primeiros usos dos quais se tem conhecimento de sua aplicação foi para determinar as quantidades de riquezas (alimento) que cada família produzia. O objetivo era a arrecadação de impostos.

Estatística é a ciência que dispõe de processos apropriados para recolher, organizar, classificar, apresentar e interpretar conjuntos de dados. Em outra definição tem-se que ela é uma área da Matemática que relaciona fatos e números de um determinado fenômeno ou população. O objetivo da estatística é extrair o máximo de informações úteis a partir de dados para obter uma melhor compreensão das situações que representam.

Durante seus estudos utilizando o livro *Análise de Dados Biotecnológicos*, você se deparará com termos conhecidos e outros nem tanto. Cada item será desenvolvido plenamente, respeitando um processo sequencial de aprendizagem que culminará na aquisição de conhecimentos que lhe permitirá, quando do surgimento de um problema, planejar, executar e analisar uma pesquisa ou experimento.

1.2 Pesquisa

Uma pesquisa ou experimento é um conjunto de atividades que permitem o estudo de uma situação real ou abstrata. Ela é o processo sistemático de investigação para obter novos conhecimentos ou validar hipóteses. Envolve coleta, análise e interpretação de dados para responder a questões específicas. Sua interpretação varia, indo além do acadêmico para incluir curiosidade, descoberta e inovação, impulsionando avanços em todas as áreas do conhecimento.

1.3 Fases do método estatístico

São etapas do estudo de um problema envolvendo métodos estatísticos determinar objetivos, planejar a experiência que nos vai permitir recolher os dados e extrair o máximo de informações relevantes para o problema em estudo, ou seja, para a população de onde os dados provêm.

As fases do método estatístico podem ser definidas como: definição do problema, planejamento e execução.

1.3.1 Definição do problema

Definir um problema é saber o que se deseja pesquisar. Um problema não surge de uma pesquisa, e sim o contrário. A ordem natural dos acontecimentos é que, a partir de uma situação de dúvida ou indecisão, surge a necessidade da implementação de uma pesquisa.

Antes de iniciar um experimento é preciso saber qual a importância do trabalho, se ainda não foi realizado. Caso existam trabalhos análogos, é fundamental comparar os resultados que foram obtidos em sua pesquisa com os daqueles. A existência de trabalhos similares já publicados não torna desnecessária a realização de uma nova experiência. A repetição de uma pesquisa se faz necessária sempre que se deseja testar novas técnicas, que estas evoluem ou são alteradas. A replicação de um experimento é salutar e pode trazer grandes avanços quando fatores que influenciam na expressão da variável de resposta. Um exemplo clássico é o fator geográfico. Sabe-se que fatores edafoclimáticos influenciam na resposta de uma espécie vegetal. Assim, uma variedade testada e cientificamente eficiente com relação à produção de matéria seca na Região Sul do Brasil pode não ser tão produtiva no Nordeste brasileiro.

Ao se deparar com um problema, deve-se inicialmente entender quais são os fatores relevantes para a sua solução. Isso resulta em duas implicações: o conhecimento prévio e as variáveis a serem mensuradas.

Conhecimento prévio é necessário, pois é preciso conhecer o escopo do problema abordado. Não é possível planejar, implementar e executar uma pesquisa da qual você não tem conhecimento a respeito. Esse conjunto de informações pode ter sido obtido empiricamente (com a prática), por meio de estudos acadêmicos ou leitura de material científico.

Variáveis podem ser dependentes ou independentes. As variáveis são definidas a partir dos dados coletados. O mais importante em relação às variáveis se refere à influência sobre a variável de resposta. Para exemplificar, imagine um estudo sobre a produção de soja. A variável de resposta é a produção de soja, enquanto que as variáveis independentes podem ser solo, clima, cultivares, adubação, entre outros. Fatores que não são inclusos como variáveis são cor dos olhos do pesquisador, número de filhos do dono da fazenda, tipo de bota usada na coleta de dados, entre outros, pois esses não estão relacionados com a produção de soja.

1.3.2 Objetivos

Todo problema surge de uma situação real, seja uma observação empírica ou de questões teóricas. Um ponto fundamental é que os objetivos devem ser estabelecidos obrigatoriamente antes do início. Isso está relacionado com a fidedignidade científica. Você não pode mudar seus objeti-

vos depois de observar os resultados. Esse é mais um motivo do conhecimento prévio do assunto a ser estudado.

É permitido incluir novos objetivos específicos, quando se observa que novas e importantes conclusões podem ser obtidas com os dados coletados. Mas isso não implica em permissão para alterar o escopo da pesquisa após a coleta.

1.3.3 Objetivos gerais

São os principais resultados que se espera obter. Eles delinham metas amplas e abrangentes que se almejam alcançar, sempre alinhado com a pergunta de pesquisa. Eles fornecem uma visão panorâmica do propósito da investigação, indicando as direções principais a seguir. São guias fundamentais que direcionam o planejamento, a execução e a avaliação do estudo. No caso do exemplo anterior, o objetivo geral poderia ser estabelecer a produtividade em função da adubação aplicada.

1.3.4 Objetivos específicos

Os objetivos específicos são resultados secundários, que podem ser obtidos por meio de análises complementares das variáveis mensuradas.



EXEMPLO 1

Considere o exemplo do experimento sobre a produção de soja e que foram coletados dados sobre o solo. Imagine que foi medida, ainda, a altura das plantas. Nesse caso, como objetivo específico poderia ser realizado o ajuste de um modelo de crescimento em altura em função do tempo. Lembre-se de que os objetivos específicos são muito importantes, mas o principal é responder o objetivo geral.

1.3.5 Hipóteses

Hipótese é uma proposição que pode (ou não) ser verdadeira. Pode ser uma suposição, conjectura, possibilidade, chance ou opção. Toda hipótese estatística deve ser uma afirmação lógica. Isso pressupõe um conhecimento prévio sobre o assunto abordado.



EXEMPLO 2

No caso do exemplo da soja, é preciso saber que o solo é um fator que influi na produção de matéria seca, entretanto, não se pode afirmar que quanto mais adubo aplicar, maior será a produção, pois é sabido que a partir de uma determinada concentração, o adubo (seja orgânico ou não) passa a ser tóxico para as plantas.

Uma hipótese estatística pode surgir de diversas fontes, incluindo os itens indicados a seguir.

- **Observações prévias:** ao analisar dados ou observar fenômenos, podem surgir padrões ou tendências que levem a formular uma hipótese sobre a relação entre variáveis.

- **Revisão da literatura:** a análise de estudos anteriores pode sugerir relações entre variáveis ou levantar questões que podem ser testadas a partir de uma hipótese.
- **Teorias existentes:** teorias científicas ou modelos conceituais podem fornecer a base para a formulação de hipóteses a serem testadas empiricamente.
- **Experiência pessoal ou profissional:** a experiência prática em um campo específico pode gerar *insights* que levam à formulação de hipóteses sobre como variáveis se relacionam.
- **Intuição ou criatividade:** às vezes, hipóteses surgem a partir de intuições ou *insights* criativos que os pesquisadores têm sobre o fenômeno em estudo.

Quando uma inferência estatística é realizada, seja por meio de comparação de médias ou frequências, seja pelo estabelecimento de correlações, duas hipóteses são sempre construídas: hipóteses de nulidade e alternativa.

H_0 : hipótese de nulidade. É uma afirmação lógica que estabelece uma condição de igualdade. É sempre essa hipótese que é testada, ou seja, parte-se do princípio que os tratamentos apresentem resultados iguais. Isso garante a honestidade do pesquisador, o que significa que este não estará sendo tendencioso em suas análises.

H_1 : quando se rejeita a hipótese de nulidade, se aceita a hipótese alternativa. Nesse caso, ao aplicar um tratamento, os resultados podem ser: a) um diferente do outro; b) um maior que o outro; c) um menor que o outro. Nesse ponto pode parecer redundante estabelecer essas três possibilidades, mas elas serão importantes quando da aplicação dos testes estatísticos.

Quatro condições devem ser satisfeitas na elaboração de hipóteses estatísticas.

Testável: isso significa que deve ser possível medir ou observar os eventos que se está pesquisando. Portanto, é necessário que o pesquisador possua ou tenha acesso aos equipamentos de mensuração, saiba operá-los ou tenha uma equipe que saiba fazê-lo. Para que seja testável, é fundamental que seja eticamente adequado, portanto, se envolver seres humanos ou animais, é obrigatória a aprovação do projeto em um comitê de pesquisa.

Lógica: a elaboração de hipóteses estatísticas pressupõe conhecimento do pesquisador sobre o assunto pesquisado, para que se possa estabelecer suposições que façam sentido. Fazer sentido em pesquisa significa ter paradigmas (teorias) que sustentem aquela suposição.

Afirmação: toda hipótese estatística, seja nula ou alternativa, deve ser redigida em forma de afirmação. Portanto, não se pode usar termos condicionais. Lembre-se de que essas hipóteses estatísticas serão posteriormente testadas por meio de procedimentos objetivos. Alguns exemplos de palavras subjetivas são “talvez”, “poder” (verbo), entre outras.



EXEMPLO 3

Para entender melhor o que são as hipóteses, vamos voltar ao exemplo anterior (produtividade de soja sujeitas a dois tipos de adubação). Nesse caso, a hipótese de nulidade (H_0) é “A produção de soja será a mesma para os dois tipos de adubação”, enquanto que a alternativa poderia ser “A produção de soja será *diferente* para cada tipo de adubação”. Note que a hipótese alternativa poderia ser “A produção de soja foi *maior* para a adubação A”.

1.4 Planejamento

Uma vez que está estabelecido o objetivo do seu trabalho e quais são as hipóteses a serem testadas, inicia-se a fase de planejamento. Nessa fase é importante ter claro **quais são as conclusões** que você quer tirar do seu trabalho e se estas são relevantes. Para isso, deve-se considerar as etapas a seguir.

- a. Tipo de estatística
 - Descritiva
 - Indutiva ou inferencial
- b. Variáveis envolvidas: quais variáveis estão relacionadas (influenciam) com meu objetivo
- c. População e amostra – censo ou amostragem –, parâmetro e estimativa
- d. Processo de amostragem
- e. Técnica de amostragem
- f. Técnica de coleta
- g. Tratamentos
- h. Coleta
- i. Definições
- j. Análise

1.4.1 Tipo de estatística

Pode ser definida como “profundidade” da pesquisa. Nesse âmbito, o trabalho pode requisitar uma estatística descritiva ou indutiva.

Estatística descritiva: envolve a coleta, organização e descrição dos dados. É o tipo mais simples de estatística, mas tão importante quanto qualquer outra. A estatística descritiva pode ser dividida em dois tipos de estudos, indicados a seguir.

Estudo exploratório: esse tipo de pesquisa será escolhido quando o objetivo for examinar uma investigação pouco estudada. Exemplo: imagine que se deseje testar a viabilidade de uma espécie florestal amazônica em plantios equidistantes no interior do Paraná. Se essa espécie nunca foi testada, não existem informações sobre ela, logo, deve-se fazer um plantio e estabelecer os primeiros dados (se as árvores sobrevivem, se crescem, se a madeira mantém suas características originais desejáveis, entre outros fatores).

Estudo descritivo: é o tipo de pesquisa realizada quando se tem alguma informação sobre o tema em questão. Pode ser também aplicada quando o objetivo é tão somente apresentar a composição de uma população ou amostra. Exemplo: imagine que se queira saber a área basal (área transversal por hectare) em diferentes talhões de *Pinus*. Nesse caso, seriam elaboradas tabelas e gráficos, bem como seriam calculadas a média, desvio padrão, coeficiente de variação e outras estatísticas que favorecessem o entendimento da composição daquela floresta.

Estatística indutiva ou inferencial: refere-se à análise e interpretação. Nesse tipo de estatística, também existem dois tipos de estudos.

Estudos explicativos: são aqueles que vão mais além do que apenas as descrições dos conceitos ou fenômenos; esses visam fazer inferências ou interpretações. Aqui são utilizados

os intervalos de confiança, comparações de médias, medianas, frequências, entre outros. Exemplo 1: pesquisa cujo objetivo fosse determinar o intervalo aceitável de peso de animais para abate em um frigorífico. Exemplo 2: pesquisa cujo objetivo fosse comparar a produção de soja ao se usar dois tipos de adubação.

Estudos correlacionais: podem ser do tipo causal. Esses podem estabelecer uma relação de causa e efeito. Comumente se utiliza a análise de regressão no desenvolvimento desse tipo de estudo, em que se analisa se duas ou mais variáveis estão ou não correlacionadas. Exemplo: pesquisa cujo objetivo fosse compreender a relação entre a produção de leite e a quantidade de ração fornecida às vacas.

Inferência pontual, intervalar e teste de hipóteses

A inferência estatística permite que os cientistas façam generalizações a partir de um conjunto de dados amostrais para a população mais ampla. Existem três abordagens principais na inferência estatística: inferência pontual, intervalar e teste de hipóteses.

Inferência pontual: na inferência pontual, o objetivo é estimar um único valor para um parâmetro desconhecido da população, com base em uma amostra de dados. Por exemplo, em estudos de expressão gênica, os cientistas podem estar interessados na estimativa da média da expressão de um determinado gene em uma população celular.

Intervalos de confiança: os intervalos de confiança são uma extensão da inferência pontual, fornecendo uma faixa de valores dentro da qual o parâmetro populacional é provável que esteja com uma determinada confiança. Esses intervalos são úteis porque reconhecem a incerteza associada à estimativa pontual. Por exemplo, um intervalo de confiança de 95% para a média de expressão gênica indicaria que existe uma probabilidade de 95% de que o verdadeiro valor populacional esteja dentro do intervalo estimado.

Teste de hipóteses: o teste de hipóteses é uma ferramenta estatística que permite aos pesquisadores fazer inferências sobre os parâmetros populacionais com base em evidências amostrais. Nesse contexto, uma hipótese nula é formulada, geralmente indicando que não há efeito ou diferença entre grupos, e é testada em relação a uma hipótese alternativa, que sugere uma diferença significativa. Os resultados do teste de hipóteses fornecem uma indicação da força das evidências contra a hipótese nula, permitindo que os pesquisadores tomem decisões informadas sobre sua rejeição ou aceitação.

1.4.2 Variáveis

Uma variável é uma característica cujos valores podem variar de indivíduo para indivíduo ou de grupo para grupo. Antes de apresentar os tipos de variáveis, cabe ressaltar a diferença entre “dados” e “informação”, uma vez que estes são comumente usados como sinônimos. Dados são os valores coletados a campo (peso de cada pessoa, altura de cada árvore, produção de leite de cada vaca, número de grãos em cada espiga de milho), enquanto que informação é um valor, conjunto de valores ou elemento visual que auxilia no rápido entendimento de um conjunto de dados (média dos pesos dos alunos de uma sala, gráfico da distribuição das alturas de um talhão, mapa com as diferentes produções médias de leite no Brasil). As variáveis podem ser classificadas como qualitativas ou quantitativas.

Variáveis qualitativas: apresentam como resultado uma *qualidade* ou *atributo*. Elas são divididas em dois tipos.

Variáveis qualitativas nominais: são atributos ou nomes. Elas não apresentam ordenação. As estatísticas possíveis são comparação de frequência, tabelas, gráficos. Quando apenas duas possibilidades mutuamente excludentes existem, essa variável é definida como binária. Exemplos: sementes – germina e não germina, espécies, sexo, portador de doença, qual doença.

Variáveis qualitativas ordinais: existe uma ordem nos possíveis resultados. As estatísticas possíveis são comparação de frequência e tendência, tabelas, gráficos, séries. Exemplos: não palatável, palatável e muito palatável; escores corporais, série na universidade, classe social, porte do cão, Famacha, entre outros.

Variáveis quantitativas: apresentam como resultado *medidas* ou *contagens*. As variáveis quantitativas podem ser divididas em discretas e contínuas.

Variáveis quantitativas discretas: só podem assumir certos valores; números inteiros (geralmente) provenientes de contagens. Em um intervalo finito, possui uma quantidade finita de valores. São estatísticas possíveis: comparação de frequência e tendência, tabelas, gráficos, séries, medidas de tendência central e dispersão. Exemplos: número de óvulos por estimulação, número de filhotes por ninhada, quantidade de colônias em uma placa, entre outros.

Variáveis quantitativas contínuas: podem assumir infinitos valores possíveis dentro de um determinado intervalo finito. Em um intervalo finito há uma infinidade de valores possíveis. São estatísticas possíveis: comparação de frequência e tendência, tabelas, gráficos, séries, medidas de tendência central e dispersão. Exemplos: pesos ao nascer de bezerros holandeses, peso ao abate de aves, produção de carne, produção de soja, DAP, H, pressão arterial.

Arredondamento

Durante o processo de análise dos resultados, muitas vezes surgem números que não são inteiros ou com quantidade de casas decimais maior do que aquela obtida com o instrumento de medição. Por exemplo, uma régua escolar comum geralmente tem uma precisão de milímetros. Ao se calcular a média de cinco medições, pode ser que surja um número em milímetros com duas casas decimais. Usar esse número poderia parecer que você fez medições com esse nível de precisão, o que não é verdade. Nesses casos, você deve fazer um arredondamento, conforme a Tabela 1, indicada a seguir.

Tabela 1. Procedimento para arredondamento de números

Condições	Procedimentos	Original	Arredondado
< 5	O último algarismo a permanecer fica inalterado.	53,24	53,2
> 5	Aumenta-se de uma unidade o algarismo a permanecer.	42,87	42,9
		25,08	25,1
		53,99	54,0
= 5	(i) Se ao 5 seguir em qualquer casa um algarismo diferente de zero, aumenta-se uma unidade no algarismo a permanecer.	2,352	2,4
		25,6501	25,7
		76,250002	76,3
= 5	(ii) Se o 5 for o último algarismo ou se ao 5 só seguirem zeros, o último algarismo a ser conservado só será aumentado de uma unidade se for ímpar.	24,75	24,8
		24,65	24,6
		24,7500	24,8
		24,6500	24,6

Fonte: Adaptada de ABNT (2014).

1.4.3 População e amostra

Uma população “S” é um conjunto de indivíduos ou objetos com pelo menos uma característica em comum. Exemplos: seres humanos, animais de uma raça, árvores de uma espécie, alunos de uma universidade, parafusos em uma fábrica, plantas em uma fazenda, entre outros.

Para uma análise estatística, uma amostra “A” é uma parte significativa e representativa da população (Figura 1). Ela pode ser qualquer subconjunto de uma população e pode ser, inclusive, um indivíduo ou todos.

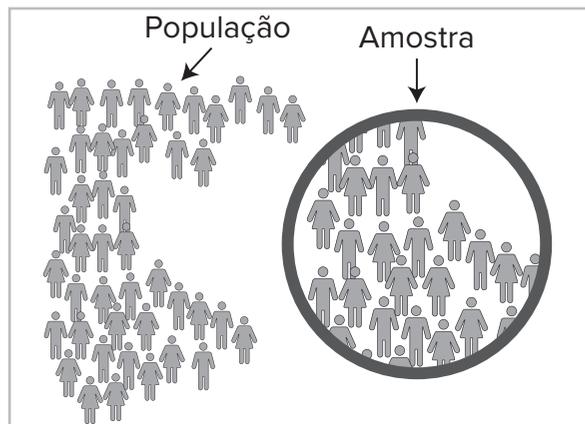


Figura 1. Representação da amostra de uma população

Fonte: Baldresca (2020).

Um **parâmetro** é uma característica numérica referente à população e só pode ser obtido mediante censo.

Estimativa é a característica numérica referente à amostra, que visa *estimar* um parâmetro e é obtida a partir de uma amostragem.

1.4.4 Processo de amostragem

Existem dois processos de amostragem: censo e amostragem. O censo consiste na coleta dos dados de todos os indivíduos da população, enquanto que a estimação consiste na coleta apenas dos dados de uma amostra. É importante que se lembre que uma amostra é composta por um conjunto de unidades amostrais. As diferenças entre censo e amostragem estão apresentadas no quadro a seguir.

Quadro 1. Comparação entre censo e estimação

Censo	Estimação
Caro	Barato
Lento	Rápido
Desatualizado	Atualizado
Nem sempre é viável	Sempre viável
Erro processual de 0%	Erro positivo

Fonte: Elaboração própria.

1.4.5 Tipos de amostragem

Em um processo de amostragem, as amostras podem ser não probabilísticas ou probabilísticas. Em uma amostra não probabilística, você seleciona os itens ou indivíduos sem conhecer suas respectivas probabilidades de seleção. Nesse âmbito existem as amostras por conveniência (pela facilidade ou baixo custo), por blocos, por quotas ou por julgamento. A teoria desenvolvida para amostragens probabilísticas não pode ser aplicada nesse caso.

Em uma amostragem probabilística, os itens são selecionados de acordo com probabilidades conhecidas. Existem quatro tipos de amostragem probabilística, as quais variam em termos de custos, precisão e complexidade: amostragem aleatória simples ao acaso, amostragem sistemática, amostragem aleatória estratificada e amostragem por conglomerados (Figura 2).



EXEMPLO 4

Vamos utilizar um conjunto de dados de um rebanho de gado, brincados de 1 a 1.000, de três raças (Nelore, Jersey e Holandês) para exemplificar cada caso.

Amostragem aleatória simples ao acaso: cada item da população tem a mesma chance de ser selecionado. Cada amostra de tamanho fixo apresenta a mesma chance de seleção em comparação com cada uma das amostras de mesmo tamanho. Ela pode ser com ou sem reposição. Os indivíduos, itens ou áreas devem ser escolhidos randomicamente, ou seja, por sorteio (que pode ser feito com papéis em um saco, em planilhas eletrônicas ou em algum sorteador on-line). Exemplo: seriam sorteados n valores de 1 a 1.000 e os animais cujos brincos correspondessem a esses números seriam amostrados.

Amostra sistemática: nesse caso, a primeira unidade amostral é escolhida aleatoriamente e as demais são determinadas de forma sistemática. Exemplo: sorteia-se o primeiro número (digamos que seja o número 28) e escolhe-se os demais sistematicamente ($28 + 50 = 78; 128; 178; 228; \dots$).

Amostragem estratificada: inicialmente divide-se a população em subconjuntos (estratos) e depois sorteiam-se os indivíduos dentro de cada estrato. Exemplo: os três estratos são as raças. Sorteiam-se n_1 Nelores, n_2 Jerseys e n_3 Holandeses. Caso os estratos tenham diferentes tamanhos, deve-se fazer uma amostragem proporcional. Exemplo: se fossem 100 Nelores, 300 Jerseys e 600 Holandeses, seriam amostrados, por exemplo, 10 Nelores, 30 Jerseys e 60 Holandeses.

Amostragem por conglomerados: você divide sua população em conglomerados em que cada conglomerado seja representativo da população inteira. Então você amostra todos os itens de um ou mais conglomerados. Exemplo: vamos usar um exemplo diferente. Imagine uma pesquisa sobre produção de leite no estado do Paraná. Cada propriedade que produz leite comercialmente seria considerada como um conglomerado. Sorteiam-se algumas delas e amostram-se todas as vacas dessas propriedades.

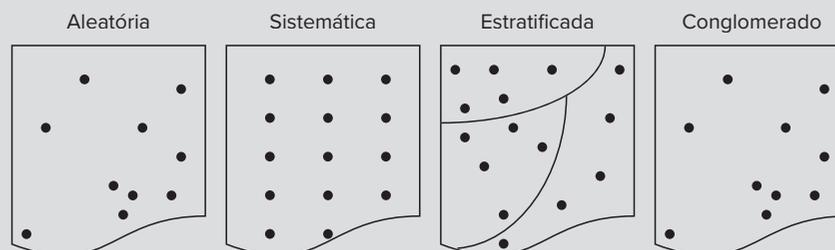


Figura 2. Exemplos de diferentes tipos de amostragem
Fonte: Elaboração própria.

1.4.6 Técnicas de coleta

Após estabelecer o tipo de amostragem e tendo em mente o objetivo do trabalho, estabelece-se a técnica de coleta, que é a forma como serão coletados os dados. Entre as técnicas, existem os tipos indicados a seguir.

Bibliográfica: quando um assunto já foi estudado, pode-se fazer uma pesquisa em materiais científicos disponíveis (livros e revistas científicas). Deve-se tomar muito cuidado com as informações disponíveis na internet que não sejam de institutos de pesquisa ou disponíveis em material certificado. As informações contidas em cartilhas, veiculadas na televisão e em sites como Wikipédia estão sujeitas a opiniões particulares e devem ser avaliadas com cuidado.

Entrevistas com especialistas: pesquisadores ou professores podem deter o conhecimento suficiente para contribuir para a solução de um determinado problema. Exemplo: se as vacas de sua propriedade estão com mastite, você poderia consultar médicos-veterinários para descobrir qual o melhor medicamento para ser aplicado.

Observação sistemática: consiste na observação de um fenômeno natural ou não de forma sistematizada segundo um protocolo ou técnica.

Entrevistas pessoais ou por telefone: deve ser bem sistemática. Antes de iniciar, deve-se treinar os entrevistadores para que todos façam as perguntas na mesma ordem e com a mesma técnica vocal. Uma entrevista pode ser tendenciosa quando o entrevistador coloca suas emoções em sua fala. A neurolinguística deve ser observada cuidadosamente. Quando existem alternativas para o entrevistado responder, essas devem ser dispostas em ordem diferente para não influenciar as respostas. Sabe-se que há uma tendência de as pessoas lerem com mais atenção as primeiras alternativas.

Questionários: conjunto de perguntas a serem lidas e respondidas pelos entrevistados. A elaboração de um questionário deve ser sempre uma ação conjunta entre o especialista da área, um estatístico e uma pessoa com conhecimentos sobre a dinâmica da mente humana.

Levantamento de campo ou laboratório: é o mais comum. No campo ou laboratório fazem-se as medições, mensurações e observações. Exemplos: medir o peso de animais, os DAPs, as alturas dos pés de milho, as produções de soja, a resistência de parafusos, a densidade de madeiras, entre outros.

Estudo de caso: quando não há material experimental suficiente, fazem-se todas as observações possíveis de um único caso. Exemplo: “recentemente” foi observado um caso de síndrome de Brugada por professores da PUCPR.

1.4.7 Tratamento

O tratamento é um tipo de variação imposta pelo pesquisador ou previamente observada. É também chamada de “variação premeditada”. Exemplo: variação na quantidade de ração para observar a *variação* na quantidade de leite produzida.

A variação sistemática é comumente atribuída à natureza do material ou de condições ambientais não controláveis. É importante que se saiba que todos os indivíduos estejam sujeitos a essa variação.

Um exemplo de variação sistemática é a temperatura ao longo do dia. Sabemos que durante o dia a temperatura é mais alta e durante a noite é mais baixa, exceto em casos pontuais e raros.

A variação aleatória é uma heterogeneidade preexistente. Essa também não pode ser controlada, mas deve ser tanto menor quanto se possa selecionar. Exemplo: variação genética. Em um povoamento clonal de eucalipto no interior de São Paulo, submetido à mesma adubação e às mesmas técnicas silviculturais, foram observadas árvores com DAP de 5 cm a 30 cm.

1.4.8 Coleta

A coleta deve ser amplamente planejada. Muitas vezes ela é feita em locais remotos e coisas simples podem inviabilizá-la (por exemplo, a falta de uma caneta ou lápis para anotação). Dentre os itens a serem planejados estão a contratação de pessoal, pagamento, alimentação, prazo de realização, material a ser confeccionado, supervisão, verificação da entrevista, crítica aos dados, entre outros.

1.4.9 Definições

Sempre tendo em mente o objetivo do trabalho, deve-se estabelecer com o máximo de precisão os itens a seguir.

Etapas: sempre que possível, é interessante dividir a pesquisa em etapas, organizando um cronograma de atividades. Exemplo: determinação do objetivo, pesquisa bibliográfica, planejamento, coleta (viagem, acampamento, alimentação, implantação).

Tempo disponível: tempo para a realização de cada etapa.

Recursos financeiros: fonte pagadora; possíveis interessados na pesquisa.

Recursos humanos: quem fará a coleta, quem supervisionará, treinamento de pessoal, digitação dos dados.

Recursos tecnológicos: quais técnicas serão utilizadas, quais equipamentos, construção de equipamentos específicos, armadilhas, baias, entre outros.

Recursos estruturais: laboratórios, salas, casa de vegetação, fazenda, canil, entre outros.

Acessibilidade aos detentores dos dados: a partir de dados que já existem, realizar outras pesquisas, verificar a tecnologia disponível, investigar os dados acessíveis, descobrir se existe alguma questão ética a ser considerada (exemplo: pesquisa sobre aids).

Transporte: verificar qual o meio de transporte, local de encontro, horários, combustível, entre outros.

Material de consumo: botas, calça, máscara, seringa, luva, caneta, prancheta, lápis, borracha, apontador, casqueador, tesoura, tesoura de poda, sacos plásticos, contêineres, facão, óculos de proteção, boné, repelente, filtro solar, focinheira, enforcador, entre outros.

1.4.10 Análises

Planejar previamente quais análises serão realizadas facilita na organização dos dados e posterior interpretação. Deve ser planejado como os dados devem ser processados, o tempo necessário para a digitação e organização, crítica aos dados (conferência, verificação ou análise de consistência), como as informações serão apresentadas (tabelas, gráficos, figuras), medidas estatísticas de cada variável e testes a serem realizados.

1.5 Execução

Após ter planejado, inicia-se a fase de execução. Ela é dividida em preparação, campo e processamento.

Preparação

Construção, pré-teste e reformulações, quando necessárias, do instrumento de pesquisa e do manual de campo; impressão do instrumento e do manual de campo; recrutamento, seleção e treinamento dos pesquisadores de campo; distribuição do trabalho entre a equipe; revisão do veículo de transporte; preparação da alimentação (quando necessário).

Campo

Coleta: obtenção dos dados brutos.

Crítica aos dados: análise qualitativa dos dados. Verificar se foram coletados e digitados corretamente. Excluir possíveis *outliers*.

Processamento e análise

Apuração: digitar e condensar os dados coletados.

Exposição: apresentação em forma de tabelas, figuras e gráficos.

Análise e interpretação dos resultados: momento em que se tiram as conclusões e se fazem as estimativas.



EXEMPLO 5

A Região Norte do Paraná se destaca como uma área significativa para a produção de milho no Brasil. Com seu clima favorável e solos propícios, a região atrai tanto agricultores tradicionais quanto aqueles interessados em tecnologias agrícolas inovadoras, como os cultivos transgênicos.

A bioestatística desempenha um papel crucial na avaliação e comparação de diferentes variedades de culturas, incluindo o milho. Nesse contexto, um estudo comparativo entre milho transgênico e não transgênico ganha relevância, visando não apenas compreender as diferenças em características de crescimento e qualidade das espigas, mas também avaliar questões relacionadas à produtividade e resistência a doenças.

Para conduzir esse estudo, serão utilizadas técnicas estatísticas robustas, empregando medidas de tendência central, dispersão e testes de significância para analisar os dados obtidos. O experimento avaliará a produção de massa seca (g), quantidade de espigas, ocorrência de doenças (quais doenças) e a qualidade das espigas (ruim, boa ou excelente).

Para avaliação da **produção de matéria seca** serão estabelecidas 20 parcelas de 50 m² cada, distribuídas de forma aleatória e igualmente divididas entre os cultivos transgênico e não transgênico. A massa seca será medida ao final do ciclo de crescimento das plantas, proporcionando uma avaliação precisa da biomassa produzida por unidade de área em cada variedade.

Para analisar a **quantidade de espigas** serão selecionadas aleatoriamente 70 plantas por cultivar de milho. A contagem do número de espigas por planta será realizada no momento da colheita.

A **avaliação de doenças** ocorrerá durante o período de crescimento, em que serão registradas quaisquer ocorrências de doenças nas plantações de milho, sejam fúngicas ou bacterianas.

A uniformidade das espigas colhidas definirá a **qualidade das espigas**, que serão classificadas em ruim, boa ou excelente.

A partir desse contexto, faça o que se pede.

- 1. Tratamentos:** quais os tratamentos estão sendo conduzidos nessa pesquisa?
Milho transgênico e não transgênico.
- 2. Variáveis:** identifique as variáveis que estão sendo medidas e classifique-as em qualitativa nominal, qualitativa ordinal, quantitativa discreta ou quantitativa contínua.
 - a. Produção de matéria seca: quantitativa contínua.
 - b. Quantidade de espigas: quantitativa discreta.
 - c. Avaliação de doenças: qualitativa nominal.
 - d. Qualidade das espigas: qualitativa ordinal.
- 3. Coleta:** quais instrumentos são necessários para medir ou avaliar essas variáveis?
 - a. Balança.
 - b. Não há necessidade de uso de equipamentos para essa avaliação.
 - c. Método direto, por exemplo, com auxílio do Manual Básico de Técnicas Fitopatológicas.
 - d. Observação.
- 4. H_0 :** escreva uma hipótese de nulidade.
Uma hipótese poderia ser (existem outras): a produção de matéria seca do milho transgênico é igual à do milho não transgênico.
- 5. H_1 :** escreva uma hipótese alternativa. Lembre-se de que H_0 e H_1 devem ser relacionadas.
A produção de matéria seca do milho transgênico é maior que a do milho não transgênico.
- 6. População:** quem é a população dessa pesquisa?
Variedades de milho transgênico e não transgênico.
- 7. Amostra:** quem é a unidade amostral? Quantas unidades amostrais estão planejadas?
70 plantas por cultivar de milho.
- 8. Censo ou amostragem:** esse experimento será conduzido como censo ou amostragem? Seria possível fazer outro processo de amostragem? Quais seriam as consequências?
Amostragem. Censo é inviável. Impossível em função do tempo, entre outros fatores.
- 9. Será realizada amostragem aleatória, sistemática ou estratificada?**
As três poderiam ser realizadas, mas da maneira como foram apresentadas as informações, o tipo de amostragem mais condizente é aleatória.



1.6 Exercícios

- 1. Objetivo:** estabelecer a contaminação por bactérias de latas de bebidas.
 - a. **Variável:** identifique a variável que está sendo medida e classifique-a em qualitativa nominal, qualitativa ordinal, quantitativa discreta ou quantitativa contínua.
 - b. **Coleta:** quais instrumentos são necessários para medir ou avaliar essa variável?

O livro *Análise de Dados Biotecnológicos* é voltado para professores e estudantes das áreas de Medicina e Ciências da Vida que buscam uma abordagem prática e eficaz no ensino e aprendizado da análise de dados. Para os professores, o livro oferece metodologias inovadoras e estratégias de ensino que promovem uma aprendizagem significativa e duradoura. Já para os estudantes, os conceitos são apresentados de forma clara e organizada, facilitando a compreensão da lógica dos temas e ajudando no desenvolvimento de um sólido mapa mental. Exemplos reais, comentados e resolvidos, trazem concretude aos temas abordados, permitindo ao leitor visualizar aplicações práticas que farão parte de sua futura carreira. Esta obra pode ser adotada como livro-texto para disciplinas como estatística, bioestatística, análise de dados, cálculo diferencial e integral, sendo ideal para cursos de Biotecnologia, Medicina, Biologia, Psicologia, Fisioterapia, Educação Física, Nutrição, Farmácia, entre outros campos das Ciências Biológicas.

