




Enap

Monitoramento da biodiversidade: gestão, análise e síntese dos dados

Módulo

3

Noções de estatística
no âmbito do Programa
Monitora I





Fundação Escola Nacional de Administração Pública

Presidente

Diogo Godinho Ramos Costa

Diretor de Educação Continuada

Paulo Marques

Coordenador-Geral de Educação a Distância

Carlos Eduardo dos Santos

Conteudista/s

Jumara M. Souza (conteudista ICMBio, 2020)

Equipe responsável:

Ana Paula Medeiros Araújo (produção gráfica, 2020)

Bruna W. F. Miranda (CGGP/ICMBio, 2020)

Guilherme Telles (implementação Articulate, 2020)

Juliana Bermudez (revisão textual, 2020)

Kamila S. N. Oliveira (pedagoga ICMBio, 2020)

Lavínia Cavalcanti Martini Teixeira dos Santos (coordenadora, 2020)

Michelli Lopes (implementação Moodle, 2020)

Priscila Campos Pereira (coordenadora, 2020)

Rosana L. S. Siqueira (CGGP/ICMBio, 2020)

Sheila Rodrigues de Freitas (coordenação web, 2020)

Tathiana C. de Souza (coordenadora ComobOMOB/ICMBio, 2020)

Ugo José B. Bezerra (coordenador substituto ComobOMOB/ICMBio, 2020)

Vanessa Mubarak Albim (diagramação, 2020)

Curso produzido em Brasília 2020.

Desenvolvimento do curso realizado no âmbito do acordo de Cooperação Técnica FUB / CDT / Laboratório Latitude e Enap.



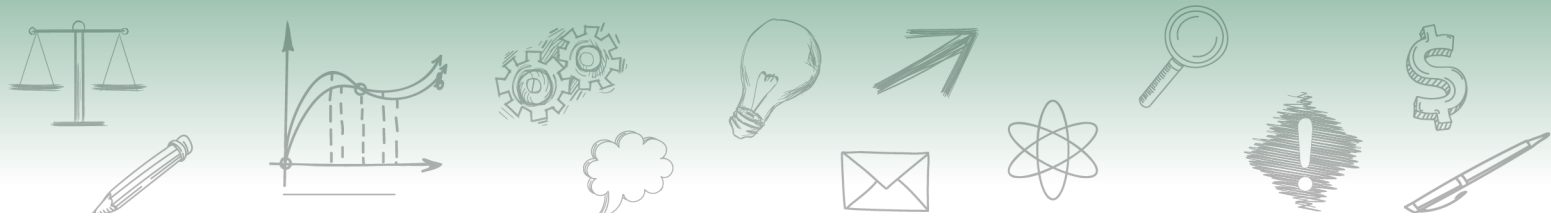
**Escola Nacional de
Administração Pública**

Enap, 2020

Enap Escola Nacional de Administração Pública

Diretoria de Educação Continuada

SAIS - Área 2-A - 70610-900 — Brasília, DF



Sumário

Unidade 1 - O pensamento científico no monitoramento da biodiversidade..... 5

1.1 Contextualização	5
1.2 Método científico	6
1.3 Etapas de construção no método científico	7
1.4 Pesquisa versus monitoramento	9

Unidade 2 - Introdução aos principais conceitos e abordagens estatísticas 11

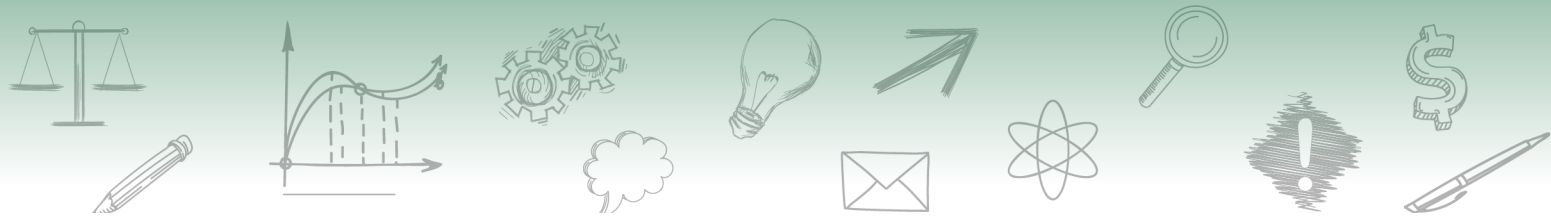
2.1 Contextualização	11
2.2 O que é estatística?	12
2.3 Principais abordagens estatísticas	14
2.4 Variáveis	15
2.5 Aleatorização	18
2.6 Software estatísticos	23

Unidade 3 - Noções sobre análise descritiva no Monitora: medidas de tendência central ou de posição 25

3.1 Contextualização	25
3.2 Média aritmética	26
3.3 Mediana	29
3.4 Moda	32
3.5 Comparando as medidas: média, mediana e moda	33

Referências..... 37





Módulo 3

Noções de estatística no âmbito do Programa Monitora I

Unidade 1 - O pensamento científico no monitoramento da biodiversidade

Ao final desta unidade, você deverá ser capaz de citar as etapas do método científico no Programa Monitora.

1.1 Contextualização



O Dicionário Online de Português define ciência como o “conhecimento profundo sobre alguma coisa” e a “reunião de saberes organizados obtidos por observação, pesquisa ou pela demonstração de certos acontecimentos, fatos ou fenômenos, sendo sistematizados por métodos ou de maneira racional: as normas da ciência”. A ciência também pode ser definida como uma rede de estudiosos, projetos, artigos e ideias. Essa rede é complexa, auto-organizada e está em constante evolução.

E quem faz ciência? Aposto que você pensou: “Essa é fácil. É o cientista!” Acredite: eu, você e qualquer pessoa pode fazer ciência. E sim, aquele que faz ciência é de certa forma um estudioso, um cientista.

Talvez a pergunta que devemos fazer é: como se faz ciência? Agora, você deve estar pensando em algo extremamente complexo, talvez em um laboratório com muitos aparelhos de última geração, com homens e mulheres de jalecos brancos e expressões loucas. Mas nem sempre é assim. Algumas das maiores descobertas da ciência surgiram por meio de observações e experimentações simples, conduzidas por pessoas curiosas, e é assim até hoje.

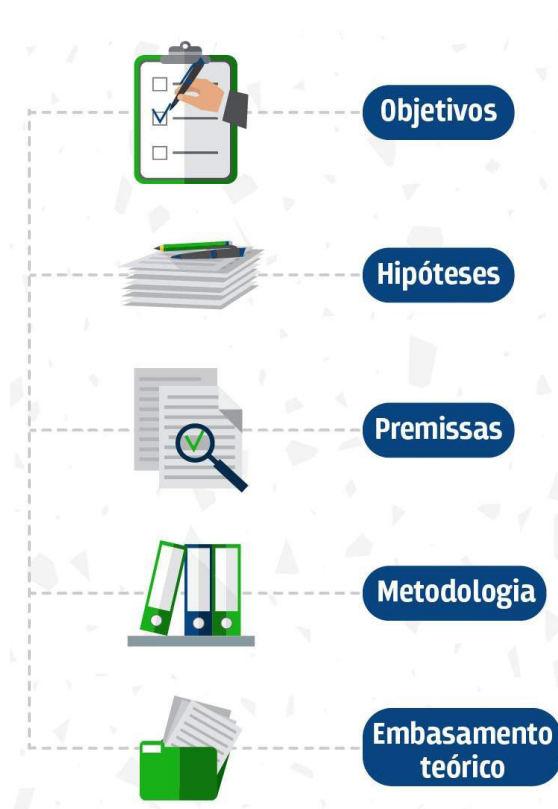
Você já deve ter ouvido ou lido alguma história que envolvia maçãs, pipas, banheiras, coroas, ilhas, bicos de pássaros e até cortiça, relacionadas com algumas das maiores descobertas da ciência. O que todas elas têm em comum é a presença de pessoas curiosas e de uma estrutura racional (um método) empregada para resolver determinado problema ou questão.



De acordo com a segunda definição de ciência apresentada no início desta unidade, existe uma padronização e uma lógica que estruturam o pensamento científico. Essa estrutura de pensamento independe da área de estudo, seja uma abordagem voltada para questões filosóficas ou aplicadas às ciências sociais, exatas, tecnológicas, biológicas, ambientais ou qualquer outro ramo da ciência. E é essa estruturação do método científico que vamos apresentar agora.

1.2 Método científico

Sempre que abordamos ciência e pesquisa, é comum depararmos com termos como objetivo, hipóteses, premissas, metodologia e embasamento teórico. Entender cada um deles ajuda a compreender melhor como o pensamento científico é estruturado.



- **Objetivos:**

São afirmações que mostram o motivo da racionalização, da pesquisa ou do monitoramento. Exemplo: gerar informação qualificada para a avaliação continuada da efetividade das UCs federais e do Sistema Nacional de Unidades de Conservação.

- **Hipóteses:**

São formulações provisórias que deverão ser testadas, confrontadas e postas a prova.

Exemplo: os indicadores monitorados não sofrem alterações significativas ao longo dos anos.

- **Premissas:**

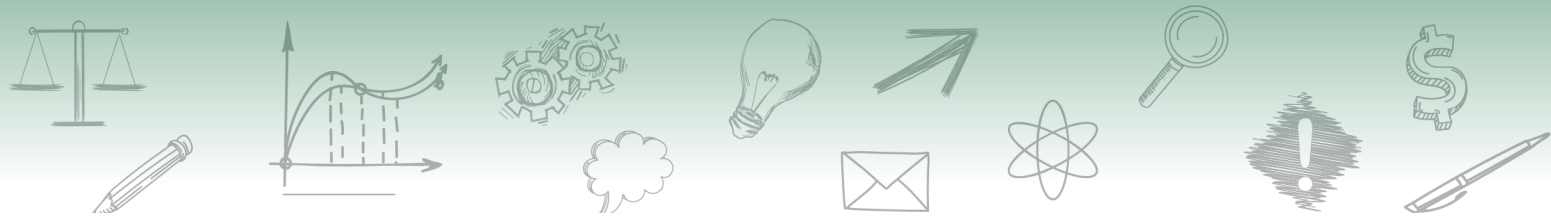
Representam a base da argumentação, são evidências ou condições necessárias para que o raciocínio lógico seja coerente.

Exemplo: as três premissas que baseiam a formulação da hipótese “os indicadores

monitorados não sofrem alterações significativas ao longo dos anos” são: (I) todos os dados do Monitora são coletados em UCs; (II) as UCs representam áreas protegidas; e (III) os dados do monitoramento são coletados em áreas de referência dentro dessas UCs.

- **Metodologia:**

Funciona como um passo a passo para a obtenção dos objetivos propostos. A metodologia diz respeito a questões do tipo “Quando?”, “Onde?”, “Como?” ou “Quanto?”. Todos os protocolos que integram o Monitora apresentam um conjunto de procedimentos a serem seguidos durante a execução do monitoramento.



Exemplo: as fichas de automonitoramento da pesca, subprograma Aquático Continental, devem ser preenchidas durante sete dias contínuos para cada um dos quatro ciclos hidrológicos (seca, enchente, cheia e vazante).

- **Embasamento teórico:**

É o conjunto de conhecimentos científicos que dá suporte às ideias, problemáticas e questões relacionadas a determinado alvo de pesquisa ou monitoramento. Ele fortalece o processo de argumentação, sendo fundamental ao longo da construção da informação e do conhecimento.

1.3 Etapas de construção no método científico

Podemos pensar em alguns passos que guiam a construção do método científico. Embora a abordagem que iremos apresentar aqui seja bastante difundida, ela não é única. Você poderá perceber que muito do que apresentaremos a partir de agora está inserido nas suas tomadas de decisões diárias.



1º passo: observando e questionando

Tudo começa com uma pergunta ou uma observação que gera um questionamento. Indagações simples como: “Por que isso é assim?”, “Como isso acontece?”, “O que acontece se trocar X por Y numa receita?”, ou perguntas mais elaboradas do tipo: “Como X se comporta ao longo do tempo?”, “Como X interage com Y?”, “Existe uma relação de causa e efeito entre X e Y?”.



2º passo: formulando hipóteses

Com o questionamento feito, é hora de elaborar algumas hipóteses. Elas têm relação direta com os questionamentos e podem ser simples ou mais elaboradas. Representam uma resposta ou explicação do fenômeno considerado na pergunta. As hipóteses devem fazer sentido e, normalmente, precisam se apoiar em premissas.

Exemplo:

Pergunta/questionamento: A estratégia de unidades de conservação é eficiente na conservação da biodiversidade?

Hipótese: As UCs representam áreas preservadas capazes de proteger a biodiversidade.
Premissas:

(I) As áreas protegidas do Brasil recebem pouca influência das ações antrópicas comparadas com as demais áreas.

(II) As UCs fazem parte das áreas protegidas do Brasil.

3º passo: testando as hipóteses

Agora que já formulamos algumas hipóteses, é hora de testá-las ou confrontá-las. Para isso, é preciso desenhar uma estratégia adequada, ou seja, formular ou adequar uma metodologia capaz de responder de forma satisfatória os questionamentos e aplicá-la. A metodologia correta garante que os dados sejam coletados no formato, periodicidade e qualidade necessários.

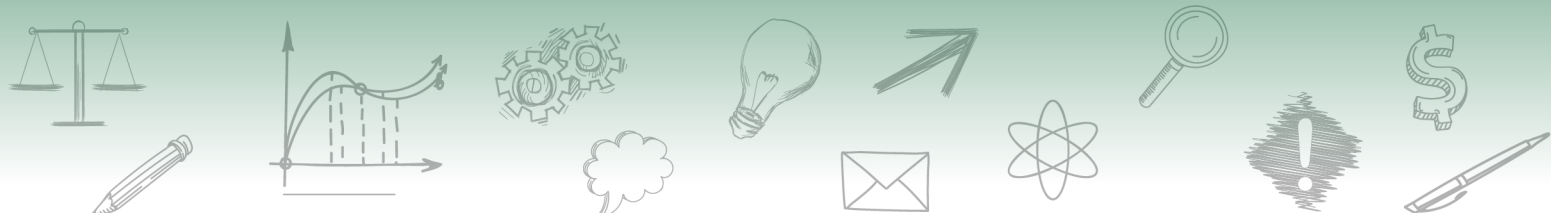
No Monitora, os protocolos foram pensados para otimizar as respostas dos questionamentos e para atender os objetivos propostos. Um exemplo disso são as periodicidades sazonais para coleta dos conjuntos de dados em cada um dos protocolos.

Os dados do componente Igarapé, o qual integra o subprograma Aquático Continental, devem ser coletados sempre no período de vazante e no horário das 10 às 14h. Essas restrições maximizam a captura (peixes) e a visualização de odonatas. Dados coletados fora desse padrão podem comprometer a interpretação dos resultados.

Independente da metodologia aplicada para responder uma questão, tenha em mente que sempre vão existir vantagens e limitações associadas a qualquer uma delas. Nenhuma abordagem é 100% eficiente e eficaz. Existem diferentes formas de explorar uma mesma questão.

4º passo

Por fim, vamos interpretar e analisar os dados coletados. O resultado final deve ser uma conclusão lógica sobre o questionamento inicial, esteja ele de acordo ou não com o que era esperado. Se todo o processo é conduzido de forma satisfatória, não existem resultados errados.



Ao longo do processo de aplicação do método científico, é comum que o questionamento inicial se desdobre em outros questionamentos. Não se decepcione caso a conclusão ou a resposta obtida não for a que você esperava. É muito provável que, para responder a pergunta inicial de forma satisfatória, seja necessário promover mudanças ou adaptações em algumas das etapas. Muitas vezes, essas necessidades só ficam evidentes ao final do processo.

Como dissemos no início deste tópico, não existe uma única maneira de construir um raciocínio científico. Tenha em mente que a proposta exposta aqui representa uma síntese de muitas formas para conduzir uma abordagem científica.

1.4 Pesquisa versus monitoramento

Pesquisa e monitoramento são duas abordagens científicas distintas. Vamos conversar sobre cada uma e entender em qual delas o Programa Monitora se enquadra.

Pesquisa científica

Podemos definir uma pesquisa científica como uma abordagem que tem início, meio e fim, e busca responder questões específicas. Conforme Volpato (2015), as pesquisas podem ser divididas em dois tipos: descritivas e associativas.

- **Pesquisa descritiva**
Busca descrever uma ou mais variáveis. Exemplo: um trabalho que busque descrever uma nova espécie de planta ou molécula.
- **Pesquisa associativa**
Contempla a relação entre as variáveis e pode ser de dois tipos:
 - ✓ **Sem interferência entre as variáveis:** a associação é caracterizada pela ausência de interferência entre as variáveis. Exemplo: a floração de uma determinada espécie de bromélia não influencia o período de floração de uma espécie de gramínea.
 - ✓ **Com interferência entre as variáveis:** uma mudança em uma variável afetará a outra. Exemplo: alterações em uma floresta gera mudanças na proporção de tribos de borboletas frugívoras.

Monitoramento

O monitoramento consiste no acompanhamento de algo ou alguém. No caso do Monitora, acompanhamos o meio ambiente, de forma sistemática e contínua, por longo prazo. O objetivo é identificar padrões e variações, além de gerar dados que podem ser usados para diversos fins.

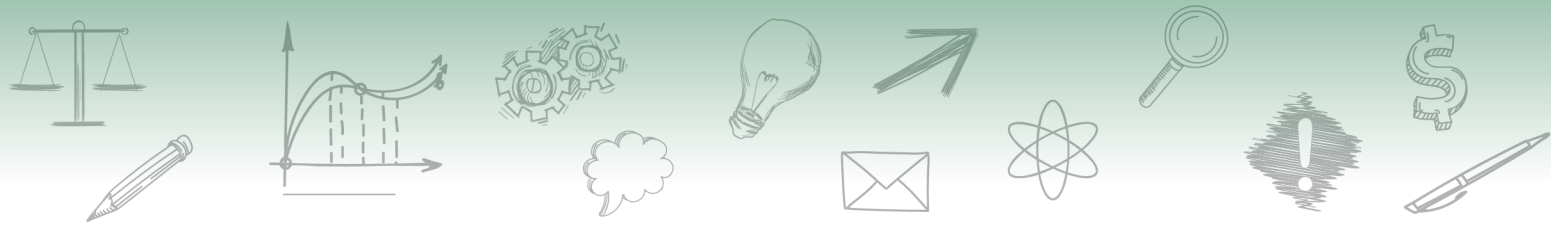
Pode ser categorizado em dois grupos, a depender de quem o conduz. Nesse sentido, há grupos de monitoramento conduzidos por cientistas e grupos de monitoramento participativos.



- **Realizados por cientistas**
Os realizados por cientistas utilizam métodos mais complexos e produzem resultados robustos. No entanto, os resultados não são facilmente absorvidos pela gestão.
- **Com participação da sociedade**
Por outro lado, aqueles realizados com participação da sociedade (como é o caso do Monitora), utilizam métodos mais simples e os resultados tendem a ser absorvidos mais rapidamente pela gestão.

Também é possível categorizar os monitoramentos ambientais em relação a seus objetivos. Nesse caso, existem monitoramentos científicos, de alerta e de manejo.





- **Científico**
O monitoramento científico busca entender a dinâmica dos sistemas biológicos.
- **De alerta**
O de alerta objetiva produzir avisos sobre alterações na biodiversidade.
- **De manejo**
E o de manejo busca produzir informações para subsidiar decisões

O Monitora é um programa de monitoramento ambiental que busca aproximar os pontos positivos da pesquisa e do monitoramento em si. Além de contar com a participação social, os métodos e as formas de gestão de dados foram pensados para reduzir o tempo das análises e facilitar a divulgação dos resultados.

O programa também está em constante processo de aprimoramento para otimizar esforços e desenhos amostrais, nos quais as pesquisas para monitoramento são fundamentais. Assim, o Monitora busca atender a diversas demandas, sem ficar restrito a um modelo único de monitoramento.

Unidade 2 - Introdução aos principais conceitos e abordagens estatísticas

Ao final desta unidade, você deverá ser capaz de descrever os principais conceitos, abordagens e ferramentas estatísticas utilizados no Monitora.

2.1 Contextualização

Nesta unidade, aprenderemos algumas das abordagens relacionadas à estatística descritiva e como podemos aplicar isso aos dados coletados pelo Monitora.



Você identifica a aplicação da estatística no seu dia a dia?

Como você a percebe?

Qual é a aplicação da estatística dentro do monitoramento?



SAIBA MAIS 🔍

Lembre-se que este é um curso introdutório e, para saber mais sobre os assuntos de estatística tratados, você pode acessar o material disponibilizado na biblioteca e consultar os seguintes livros-texto:

- CALLEGARI-JACQUES, S. M. Bioestatística: princípios e aplicações. Porto Alegre: Artmed, 2003.
- GOTELLI, N. J.; ELLISON, A. M. Princípios de Estatística em Ecologia. Porto Alegre: Artmed, 2011.

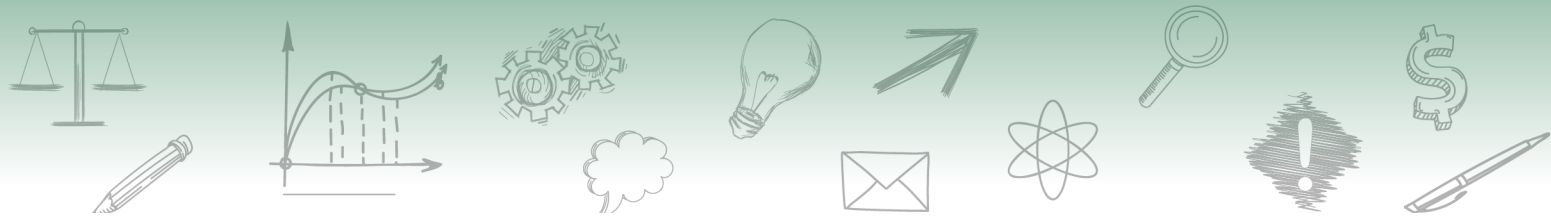
Agora, vamos começar a falar sobre estatística.

2.2 O que é estatística?

Podemos pensar na estatística como uma abordagem que conduz a uma síntese numérica. Essa síntese evidencia o que há de significativo ou de generalizado para um conjunto de dados.

DESTAQUE

Desse modo, a estatística é uma forma de gerar informação e conhecimento quantificável e comparável, a partir de um conjunto de dados.



A estatística envolve:

- Métodos relacionados com o planejamento e a condução do estudo.
- Descrição dos dados coletados.
- Tomada de decisões.
- Predições ou inferências sobre fenômenos relacionados aos dados.

Na sua forma básica, a estatística não busca o excepcional, mas o que é geral. O poder de inferência está no conjunto dos dados, ou seja, observar um dado isolado pode não dizer muito, pois é o conjunto de dados que possui a informação.

SAIBA MAIS

O primeiro levantamento estatístico remonta à 3050 a.C., no Egito, tendo como objetivo informar o Estado sobre recursos humanos e econômicos. Quer conhecer um pouco mais sobre a história da estatística? O livro de David Salsburg, chamado *Uma Senhora Toma Chá*, narra suavemente parte da história da estatística e um pouco da vida de alguns dos nomes que revolucionaram a análise de dados.



No Monitora, coletamos um grande volume de dados. São muitas UCs reunindo diferentes dados para responder perguntas diversas, sejam elas comuns a todo o sistema ou de interesse local.



Entretanto, como responder essa gama de perguntas a partir desse grande conjunto de dados? Como usar os dados coletados para produzir informação e conhecimento de forma quantificável? Como usar tal informação a fim de identificar tendências ou padrões relevantes para o sistema ou a realidade da UC?

As respostas para essas perguntas serão tratadas a seguir.

2.3 Principais abordagens estatísticas

As abordagens estatísticas podem ser divididas em dois grandes grupos: estatística descritiva e estatística inferencial. As características que definem cada uma delas são:

- **Estatística descritiva**

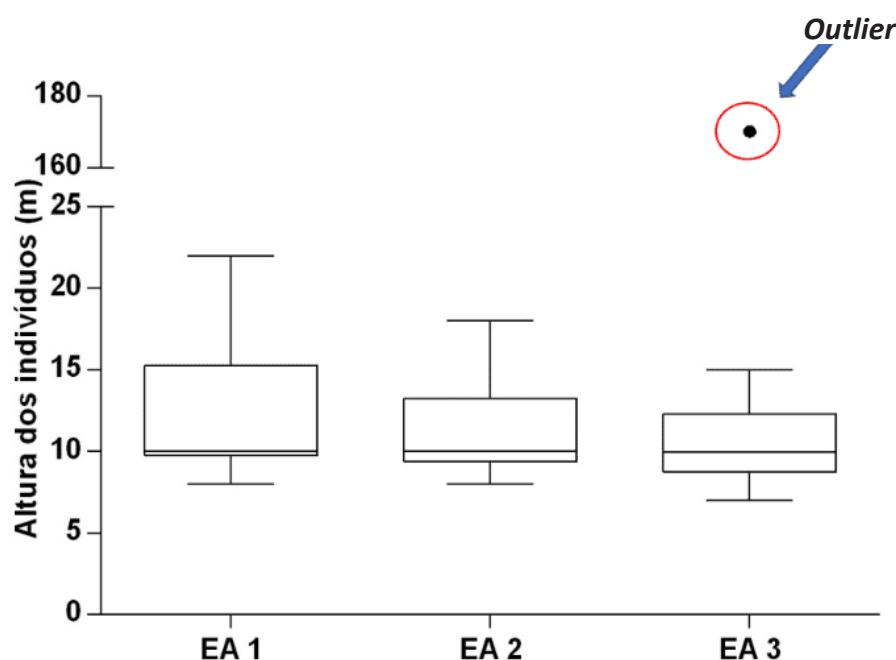
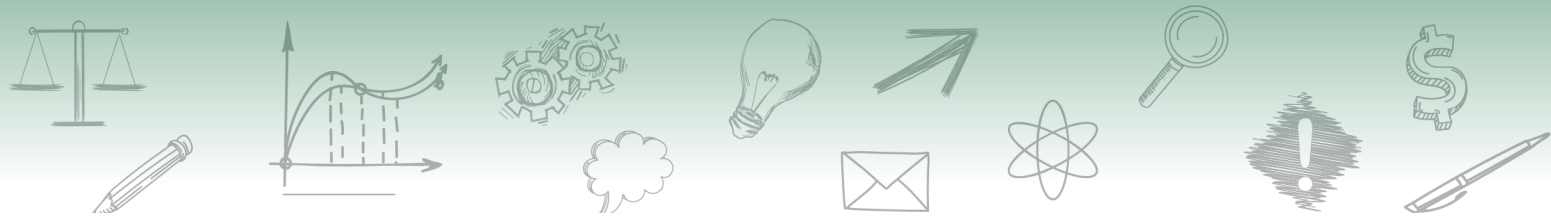
Na estatística descritiva, o objetivo principal é descrever e analisar um conjunto de dados, sem se preocupar em formular conclusões assertivas ou inferências sobre ele. Além disso, por meio da análise descritiva, é possível identificar inconsistências ou potenciais erros.

Um exemplo prático da aplicação dessas análises é a identificação de valores que destoam da tendência geral do conjunto, os chamados *outliers*, representados por um ponto fora do padrão, como no gráfico a seguir.

No gráfico, é possível perceber a tendência da altura média das plantas e acompanhar sua variação em cada uma das estações amostrais (EAs) monitoradas. Além disso, esse tipo de gráfico permite uma identificação rápida dos valores máximos e mínimos do indicador (nesse caso, a altura) em cada uma das EAs, assim como dos *outliers*.

As análises descritivas geralmente respondem questões acerca dos próprios dados, por exemplo: “qual é a melhor forma de resumir os aspectos essenciais dos dados?” ou “os dados apresentam regularidades, tendências, ciclos, padrões?”.

A estatística descritiva também possibilita expor, de forma simplificada, a informação contida em um grande volume de dados. Essa, inclusive, é uma das necessidades do Monitora.



- **Estatística Inferencial**

A estatística inferencial é focada nas generalizações feitas a partir de amostras aleatórias dos dados coletados de uma população.

Dentro do contexto estatístico, podemos definir o termo população como todo o conjunto de elementos que se quer estudar, enquanto o termo amostra pode ser definido como um subconjunto da população.

É coletando dados de parcelas reduzidas das populações que os estudos científicos conseguem inferir informações sobre as populações de interesse. Isso é válido também para o Monitora.

Imagine ter que medir todas as árvores de uma UC ou anotar medidas de tamanho e peso de todos os peixes das áreas de entorno e do interior da UC para podermos monitorar ou pesquisar esses indicadores. Seria necessário um esforço físico e logístico praticamente impossível de ser realizado. E isso certamente inviabilizaria o monitoramento ou a pesquisa. Entretanto, ao utilizar a estatística de inferência, é possível coletar dados de uma amostra representativa da população e, a partir deles, conduzir generalizações sobre o todo.

2.4 Variáveis

Variável é a característica que será medida e observada nos indivíduos amostrados, podendo ser representada por dados numéricos ou não. As variáveis podem ser organizadas em duas categorias, quantitativas ou qualitativas, e cada uma dessas categorias é subdividida em outras.



DESTAQUE

De forma simplificada, as variáveis qualitativas quase sempre são representadas por palavras ou letras e relacionadas a uma característica ou a uma qualidade. Por outro lado, as variáveis quantitativas quase sempre são números.

No entanto, nem sempre a variável apresentada como número será quantitativa. Em algumas fichas de coleta utilizadas no Monitora, por exemplo, existem campos para anotar a coordenada geográfica do ponto de amostragem ou o número da unidade amostral. Embora esses dados sejam numéricos, eles são considerados variáveis qualitativas, pois estão relacionados a características do local.

Vamos detalhar agora as características das variáveis e as suas subdivisões.

VARIÁVEL QUANTITATIVA

Nessa categoria, a característica observada assume valores numéricos. Geralmente, esses valores são originados de medições (como o número de avistamentos), cálculos (como o cálculo do tamanho do caranguejo a partir da medida do diâmetro da galeria) ou enumerações.

- **VARIÁVEL QUANTITATIVA DISCRETA**

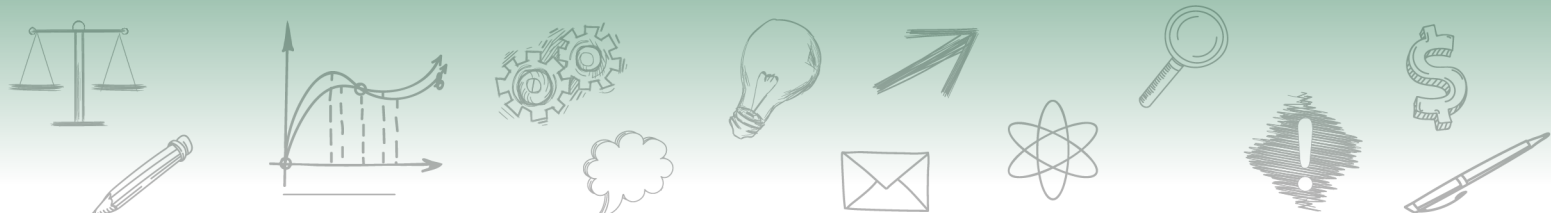
Assume valores dentro de um conjunto de números especificados. Exemplo: imagine que durante uma coleta do protocolo de mamíferos e aves, componente Florestal, ao percorrer a trilha, o monitor anota: 3 inhambus, 5 pacas e 8 macacos de cheiro. Os números 3, 5 e 8 representam dados quantitativos discretos. Note que, como estamos falando de seres vivos, não há possibilidade de termos valores como 1,5 paca ou 3,8 macacos. As variáveis quantitativas discretas são representadas por números inteiros.

- **VARIÁVEL QUANTITATIVA CONTÍNUA**

Assume valores em um intervalo contínuo de números. Em geral, esse tipo de variável é proveniente de medições. Exemplo: durante as medições de largura ou profundidade dos igarapés, componente Igarapé, podemos obter valores como: 0,2 cm, 35 cm e 4,2 m. Os números não inteiros (fracionais ou decimais) representam as variáveis contínuas.

VARIÁVEL QUALITATIVA (OU CATEGÓRICA)

Representa uma qualidade ou atributo associados ao que estamos monitorando. Geralmente, é representada por dados não numéricos. No monitoramento, alguns exemplos são: condições do ambiente (sol/ensolarado, nublado), data, local da coleta (Igarapé 2, Estação Amostral 1), tribos de borboletas frugívoras e morfotipos de peixe. Entretanto, é preciso ter atenção! Existem variáveis numéricas que são classificadas como qualitativas, como os números atribuídos às estações ou unidades amostrais presentes nos protocolos.



- **VARÁVEL QUALITATIVA NOMINAL**

Para essa variável, não existe nenhuma ordenação, ou seja, a ordem de apresentação dos dados não influencia na interpretação. Exemplo: na tabela a seguir, a primeira coluna traz a variável “tribo”. A ordem como as tribos aparecem não influencia na interpretação dos dados. Em outras palavras, não há uma hierarquia dentro das tribos.

Tribos	Número de indivíduos
Satyrini	138
Brassolini	69
Morfini	37
Epicaliini	21
Anaeini	15
Preponini	6
Coeini	5
Ageroniini	3
Haeterini	3
Epiphilini	1
Melanitini	1

- **VARIÁVEL QUALITATIVA ORDINAL**

Para essa variável, existe uma ordem em seus resultados. Os dados apresentam uma hierarquia de disposição, uma escala gradativa (muito/pouco) ou temporal (meses do ano). Exemplo: na tabela a seguir, são apresentados números de borboletas da tribo Satyrini amostrados ao longo dos anos. A coluna “Ano” representa uma variável qualitativa ordinal, pois existe uma ordem a ser seguida, nesse caso, uma ordem temporal que afeta a interpretação de todo o conjunto de dados.

Ano	Tribos	Número de indivíduos
Ano I	Satyrini	138
Ano II	Satyrini	125
Ano III	Satyrini	120
Ano IV	Satyrini	90
Ano VI	Satyrini	88
Ano VII	Satyrini	102
Ano VIII	Satyrini	82
Ano IX	Satyrini	93
Ano X	Satyrini	75
Ano XI	Satyrini	62
Ano XII	Satyrini	66



2.5 Aleatorização

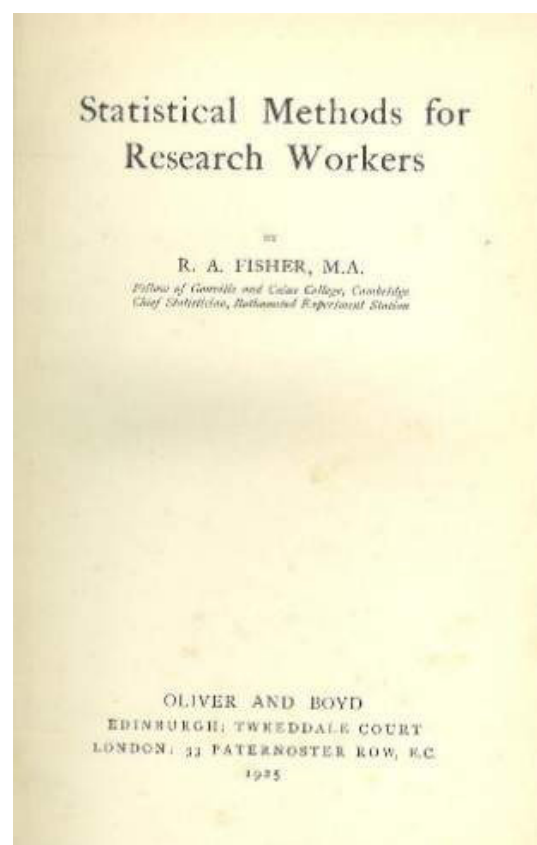


Nada se pode provar sem um desenho experimental aleatório.

Ronald A. Fisher, 1925



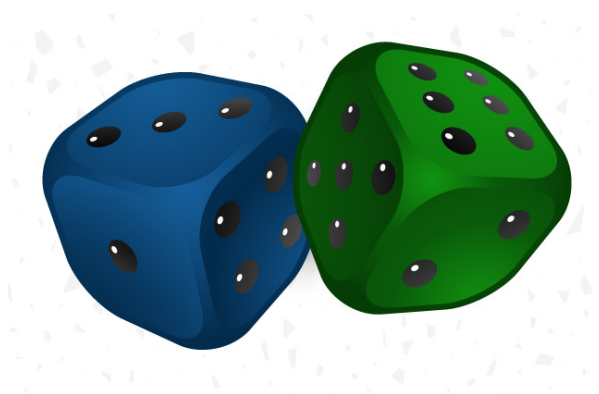
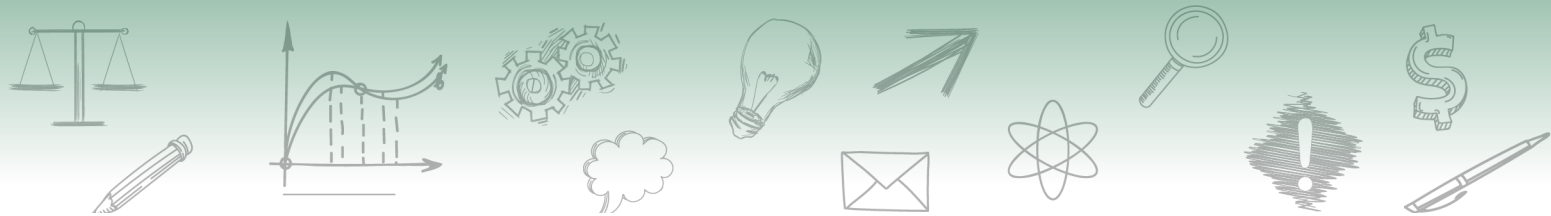
Essa frase está presente em uma das publicações consideradas um marco na história da ciência: *Statistical Methods for Research Workers* (Métodos estatísticos para pesquisadores), livro escrito em 1925 por Ronald A. Fisher.



Ronald Fisher e a capa de seu livro, publicado em 1925.

Um dos principais pontos defendidos por Fisher durante sua carreira foi a aplicação de um bom desenho amostral aleatório na condução de uma abordagem científica.

De maneira simples, escolher de forma aleatória, ao acaso, significa que qualquer elemento de uma população ou um conjunto de elementos disponíveis tem a mesma chance de ser escolhido.



Pode parecer algo simples, mas tentar, conscientemente, escolher algo de forma aleatória não é trivial, pois todo o nosso processo de escolha ou de tomada de decisão é enviesado. É o chamado viés de escolha, ou seja, quando uma escolha é feita de forma influenciada ou guiada. Isso ocorre porque, mesmo de forma inconsciente, nosso cérebro processa as escolhas baseado em um conjunto de informações disponíveis que acabam influenciando nossa escolha. É precisamente para evitar essa influência que Fisher defende a importância da escolha aleatória.

É importante observar, no entanto, que enviesar o processo de tomada de decisão ou de escolha nem sempre é algo ruim, pois também pode nos ajudar a manter um padrão coerente nas nossas ações. Por exemplo, no Monitora, a escolha das áreas elegíveis para implementação das estações amostrais (EAs) e unidades amostrais (UAs) deve sempre seguir as orientações presentes no protocolo. Essas orientações são responsáveis por guiar e padronizar as ações.

DESTAQUE

Áreas elegíveis ou de referência para o Monitora são regiões da UC que atendem aos requisitos para implantação das EAs e UAs, locais onde serão coletados os dados do monitoramento. Desse modo, a definição dessas áreas elegíveis não é aleatória, já que é feita de forma padronizada, de acordo com critérios específicos.

Entretanto, no programa, outras escolhas e tomadas de decisão precisam ser isentas de viés, ou seja, precisam ser feitas de forma aleatória. Um exemplo é a seleção dos locais, dentro das áreas elegíveis, onde serão implementadas as EAs ou UAs, ou seja, onde haverá monitoramento. Isso porque, nesses casos, a presença de viés pode influenciar, direta ou indiretamente, nos resultados obtidos.


Acompanhe a explicação de como definir aleatoriamente, por meio de sorteio, as áreas para implantação das EAs do componente Florestal.



Passo 1

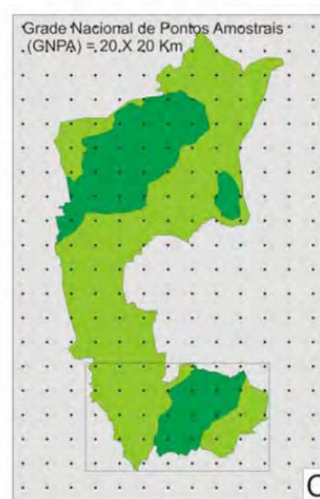
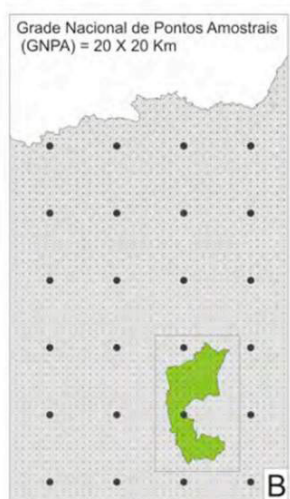


LEGENDA:

 Unidade de Conservação


Definir a unidade de conservação que receberá as estações amostrais.


Passo 2

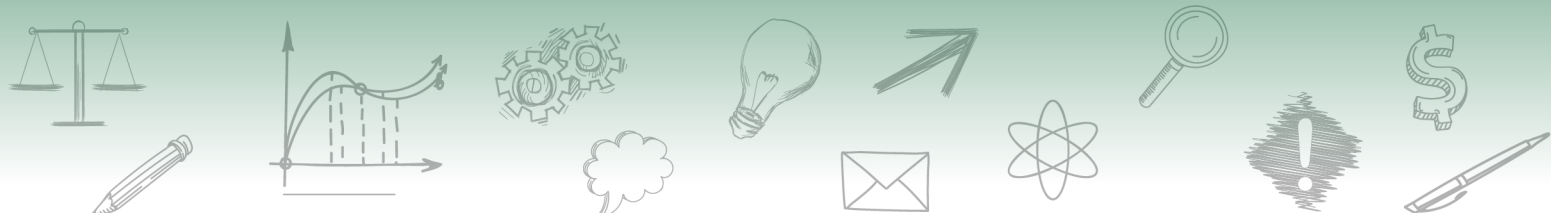


LEGENDA:

 Unidade de Conservação

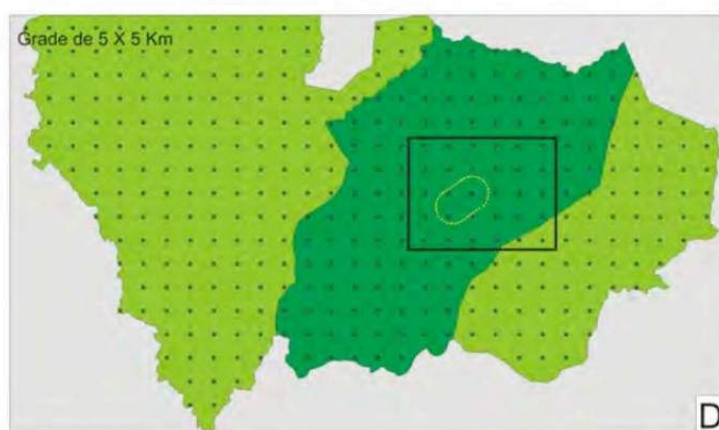
 Áreas elegíveis ou em condição de referência

 Áreas não elegíveis ou em condição indeterminada



Aplicar sobre o mapa da UC a Grade Nacional de Pontos Amostrais (GNPA) padrão, conjunto de pontos pequenos na figura e distantes 20 Km entre si, estabelecidos pelo Serviço Florestal Brasileiro (SFB).

Passo 3

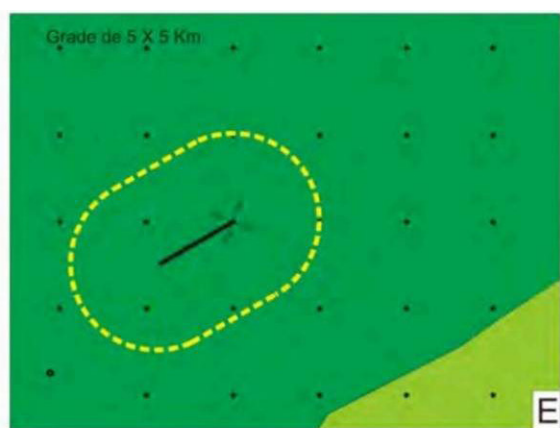


LEGENDA:

- Unidade de Conservação
- Limite da área de inelegibilidade no entorno da EA
- Áreas elegíveis ou em condição de referência
- Áreas não elegíveis ou em condição indeterminada

Determinar as áreas elegíveis ou de referência nos limites da UC. Lembre-se que essas áreas são aptas a receber as estações amostrais.

Passo 4



LEGENDA:

- Unidade de Conservação
- Representação da Estação de Amostragem
- Limite da área de inelegibilidade no entorno da EA
- Áreas elegíveis ou em condição de referência
- Áreas não elegíveis ou em condição indeterminada



Adaptar a grade do SFB, diminuindo a distância entre os pontos da GNPA padrão de 20 Km para 5 Km. Então, numerar esses pontos e sorteá-los.

DICA

Use um saquinho contendo pedacinhos de papéis com os números dos pontos para o sorteio.

No componente Florestal, as EAs devem manter uma distância entre si de no mínimo 5 Km. Por isso, considera-se tal distância em volta das EAs uma área inelegível (tracejado amarelo na figura). Se o ponto sorteado estiver localizado nessa área inelegível, ele deve ser desconsiderado e um novo sorteio deve ser feito.

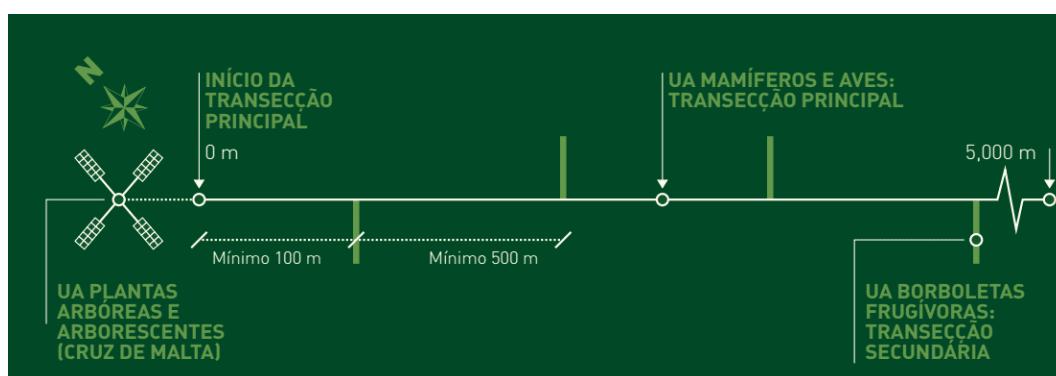
Passo 5

Implantar a estação amostral.

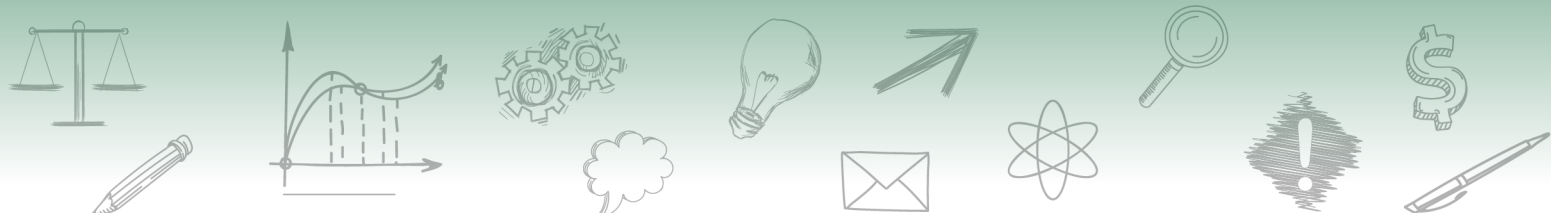
SAIBA MAIS

As estações amostrais são compostas por um conjunto de unidades amostrais. No componente Florestal, as EAs incluem as UAs dos alvos globais e possuem o seguinte formato:

Para saber mais sobre as características das EAs e UAs do componente Florestal, além de detalhes da sua implantação, sugerimos acessar o *Roteiro Metodológico de Aplicação* na biblioteca virtual.



Você pode estar pensando que o seu protocolo não aborda nenhum desses aspectos mencionados. Mas você tem certeza disso? Confira o seu protocolo e procure uma recomendação sobre a escolha dos locais para implementação das EAs ou das UAs, ou se não há uma orientação sobre os locais que serão amostrados. Essa escolha é um passo importante na aplicação das



ações de monitoramento e devemos executá-la da melhor forma possível. Lembre-se que, no Monitora, temos diversos alvos e nem sempre o protocolo exige estações amostrais, como é o caso automonitoramento da pesca.

2.6 Software estatísticos

Durante muitos anos, as análises estatísticas eram feitas sem auxílio de computadores. Essa prática, a de realizar os cálculos manualmente, pode ser muito instrutiva, uma vez que ajuda a conhecer as equações por trás das análises e entender a lógica contida em cada uma delas. Entretanto, fazer os cálculos a mão pode ser bem demorado e mais passível de erro, principalmente para grandes volumes de dados, como é o caso do Monitora.

SAIBA MAIS 🔍

Durante muito tempo, os cálculos utilizados pela *National Aeronautics and Space Administration* (NASA) foram feitos a mão, por pessoas conhecidas como calculadoras humanas. Um pouco da história de algumas delas é contada no livro *Estrelas Além do Tempo*, o qual foi adaptado para o cinema.



Embora nosso cérebro seja um prodígio em processar dados e informações, nosso desempenho com grande volume de dados, na média geral, não é nada se comparado com os computadores modernos. Por esse motivo, vamos conhecer um pouco sobre algumas das ferramentas computacionais utilizadas para análises estatísticas.

Hoje, o mercado dos softwares estatísticos é dividido em dois grandes grupos: os softwares de licença gratuita e os pagos.

The R Project for Statistical Computing é um exemplo de software de licença gratuita. Ele é uma linguagem e um ambiente desenvolvido para executar estatística computacional e gráfica.



Entretanto, seus desenvolvedores evitam pensar nele como um software estatístico, justamente devido a sua flexibilidade funcional.

DESTAQUE

Além de ser gratuito e ter uma ampla gama de funcionalidades, o R opera por meio de linhas de comando, digitadas no console para que as ações ocorram. Para muitas pessoas, essa é sua maior limitação, pois é preciso conhecer os códigos para utilizá-lo.

Isso não é um problema na maioria dos softwares estatísticos pagos, pois eles têm uma interface mais intuitiva e botões de comando para execução das ações. Assim como o R, os softwares estatísticos pagos possuem muitas funcionalidades e alguns são exclusivos para a criação de gráficos.

A desvantagem é que o uso é liberado somente por meio da compra da licença de utilização. Os preços e o modelo da compra, anual ou vitalícia, podem variar dependendo do produto.

DESTAQUE

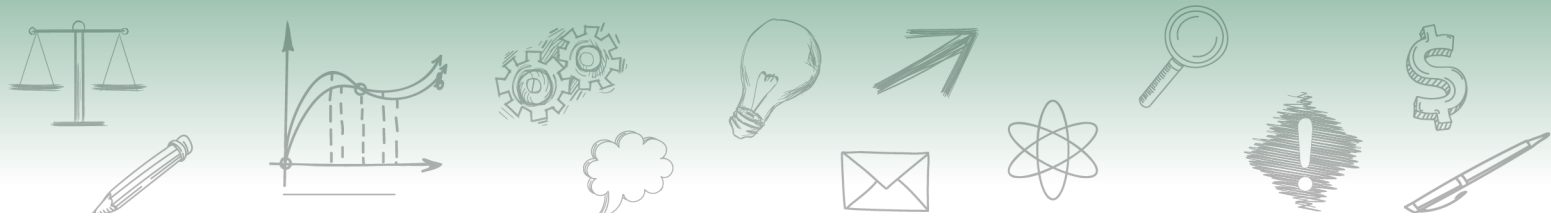
E qual deles é o melhor para usarmos? A escolha vai depender de vários fatores. O principal é observar se o software estatístico que você vai utilizar atende suas necessidades.

Nem sempre será necessário usar um software complexo na realização de cálculos e condução de algumas das análises usadas pelo Monitora, sendo possível utilizar editores de planilhas eletrônicas, como o Excel, para isso. Entretanto, tenha em mente que essa não é a funcionalidade à qual os editores de planilha se destinam. O ideal é utilizar ferramentas desenvolvidas especificamente para análise de dados, sejam elas gratuitas ou pagas.

Independente da ferramenta escolhida, é importante considerar dois pontos principais:

- **Conheça bem as funcionalidades oferecidas e o formato de entrada do dado requerido pelo software**

Tenha atenção ao fato de que softwares diferentes têm aparências e regras distintas. Entender o tutorial do software antes de utilizá-lo ajuda a compreender questões como o formato e a maneira correta de entrada dos dados na plataforma. Existem livros-texto focados em ferramentas específicas (alguns estão citados na nossa lista de referências). Uma busca rápida na internet pode indicar materiais de apoio para essas e outras ferramentas estatísticas. Mas, assim como todo conteúdo disponível na internet, tenha cuidado com a fonte! Dê preferência para sites e publicações disponibilizadas por instituições de pesquisa e de ensino superior, como faculdades e universidades, ou para materiais de apoio (help ou FAQs) disponibilizados na página da empresa ou grupo responsável pela ferramenta utilizada.



- **Use a ferramenta**

Como qualquer atividade que desempenhamos, analisar os dados do Programa de Monitoramento torna-se mais fácil com a prática e a repetição. O mesmo serve para o uso da ferramenta estatística. Simule dados e análises para diferentes situações. Isso fará com que você se familiarize com a ferramenta e tornará mais simples trabalhar os dados do monitoramento.

Diferentes softwares podem gerar resultados diversos para a mesma análise, apesar de utilizarem o mesmo conjunto de dados. Mas não se assuste, pois, na maioria dos casos, essa diferença não é significativa. Ela pode acontecer devido à diferença entre algoritmos e fórmulas utilizados pelos programas.

Unidade 3 - Noções sobre análise descritiva no Monitora: medidas de tendência central ou de posição

Ao final desta unidade, você deverá ser capaz de aplicar as medidas de tendência central aos dados coletados pelo Programa Monitora.

3.1 Contextualização

A exploração e a análise descritiva dos dados representam a fase inicial do processo de construção da informação e do conhecimento. Nessa etapa, são organizados, resumidos e descritos alguns aspectos do conjunto de dados de forma a identificar padrões ou inconsistências. Embora os dados coletados pelo nosso programa apresentem um padrão esperado, pensado para responder determinadas questões, é importante conduzir uma análise exploratória e conhecer bem seu conjunto de dados.

DESTAQUE

Antes de iniciar qualquer análise, observe se há algum dado muito discrepante, como valores extremos ou com formatos diferentes do esperado, ou algum campo vazio. Tenha atenção, pois essas inconsistências podem prejudicar sua análise.

Muitas das questões de interesse local serão respondidas e compreendidas ao longo da etapa de exploração e por meio de análises descritivas. Isso não quer dizer que essas questões sejam mais simples ou menos relevantes, e sim que não é preciso realizar uma análise de inferência com o conjunto de dados para respondê-las.

As principais ferramentas da análise descritiva são as representações visuais dos dados, como gráficos e tabelas, e as medidas utilizadas para síntese numérica da informação, como médias, índices, porcentagens, erros e desvios.



Podemos dividir as análises descritivas em dois grupos: medidas de tendência central (posição ou locação) e medidas de dispersão ou de variabilidade.

Nesta unidade, abordaremos as medidas de tendência central. Elas correspondem a valores que sintetizam os dados, mediante um valor com potencial para representar o conjunto.

Existem três medidas com esse potencial: a média, a mediana e a moda. Vamos conhecê-las!

3.2 Média aritmética

A média aritmética talvez seja a medida de tendência mais conhecida, pois o cálculo da média é simples. De forma resumida, basta somar todos os dados e dividir o valor pelo número de elementos do conjunto, conforme a seguinte fórmula:

$$\bar{X} = \frac{\text{Soma de todos os valores coletados}}{\text{Tamanho da amostra (número de elementos)}}$$

Acompanhe os seguintes exemplos:

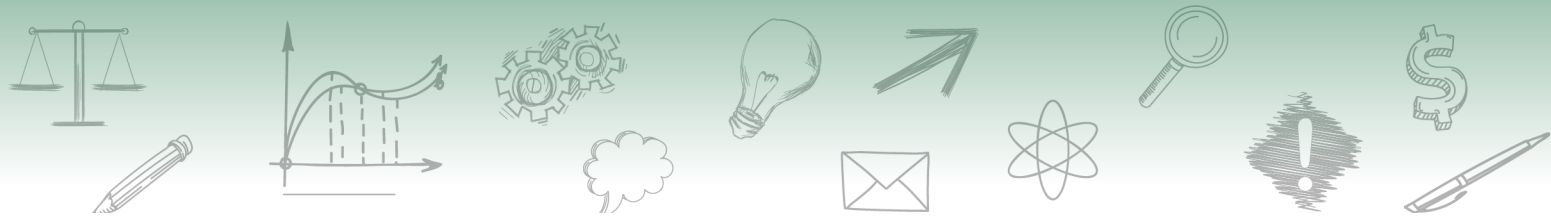
Exemplo 1

Na tabela a seguir, estão representados dados hipotéticos para o alvo caranguejo-uçá, subprograma Marinho Costeiro, Programa Monitora. A variável, destacada em vermelho, é o diâmetro da galeria (DG).

Dados coletados no protocolo do alvo caranguejo-uçá, componente Manguezal, subprograma Marinho Costeiro

UC	DATA	EQUIPE	MANGUEZAL	DG (cm)
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,2
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,3
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,3
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,7
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,3
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	3,0
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,9
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,2

Ao todo, temos oito medições de galerias diferentes. Vamos ordená-las do menor para o maior valor (2,2; 2,2; 2,3; 2,3; 2,3; 2,7; 2,9; e 3,0) e, então, calcular a média aritmética:



$$\bar{X} = \frac{2,2 + 2,2 + 2,3 + 2,3 + 2,3 + 2,7 + 2,9 + 3,0}{8} = 2,5$$

IMPORTANTE !

O diâmetro médio das galerias amostradas no Manguezal 1, na referida data e UC, é de 2,5 cm. Ou seja, se precisarmos de um único valor para resumir o diâmetro das galerias dessa UC, esse valor é 2,5 cm.

Exemplo 2

Agora, analise os dados da tabela a seguir, na qual acrescentamos dados do diâmetro da galeria para diferentes anos, os quais estão destacados por cores distintas.

Dados coletados pelo protocolo do alvo caranguejo-uçá, componente Manguezal, subprograma Marinho Costeiro

UC	DATA	EQUIPE	MANGUEZAL	DG (cm)
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,2
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,3
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,3
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,7
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,3
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	3,0
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,9
ESEC X	Ano 1	Hugo, José e Luiz	MG_1	2,2
ESEC X	Ano 2	Hugo, José e Maria	MG_1	2,6
ESEC X	Ano 2	Hugo, José e Maria	MG_1	2,0
ESEC X	Ano 2	Hugo, José e Maria	MG_1	2,2
ESEC X	Ano 2	Hugo, José e Maria	MG_1	2,2
ESEC X	Ano 2	Hugo, José e Maria	MG_1	2,2
ESEC X	Ano 2	Hugo, José e Maria	MG_1	2,3
ESEC X	Ano 2	Hugo, José e Maria	MG_1	2,3
ESEC X	Ano 2	Hugo, José e Maria	MG_1	2,3
ESEC X	Ano 3	Maria, José e Lúcia	MG_1	2,3
ESEC X	Ano 3	Maria, José e Lúcia	MG_1	2,2
ESEC X	Ano 3	Maria, José e Lúcia	MG_1	2,3
ESEC X	Ano 3	Maria, José e Lúcia	MG_1	1,9



ESEC X	Ano 3	Maria, José e Lúcia	MG_1	1,9
ESEC X	Ano 3	Maria, José e Lúcia	MG_1	2,1
ESEC X	Ano 3	Maria, José e Lúcia	MG_1	2,2
ESEC X	Ano 3	Maria, José e Lúcia	MG_1	1,9

Como podemos perceber, há um conjunto de dados mais complexo. Note que são referentes à mesma UC e ao mesmo local (Manguezal 1), mas a anos diferentes (Ano 1, Ano 2 e Ano 3). Considerando esse conjunto, vamos calcular a média para cada ano.

Ano 1

$$\bar{X} = \frac{2,2 + 2,2 + 2,3 + 2,3 + 2,3 + 2,7 + 2,9 + 3,0}{8} = 2,5$$

Ano 2

$$\bar{X} = \frac{2,0 + 2,2 + 2,2 + 2,2 + 2,3 + 2,3 + 2,3 + 2,6}{8} = 2,3$$

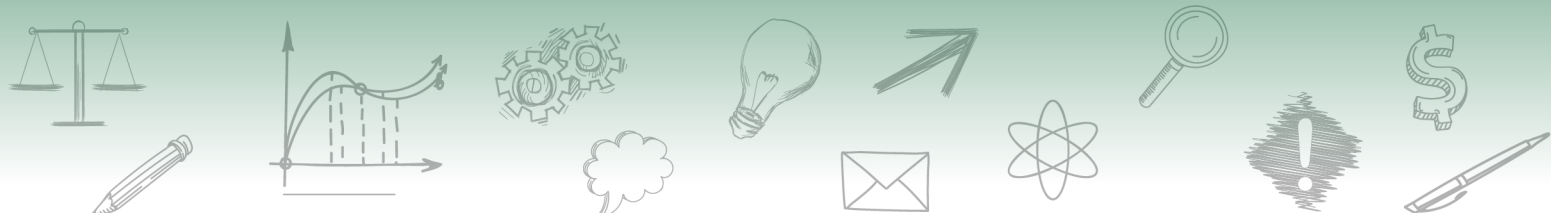
Ano 3

$$\bar{X} = \frac{1,9 + 1,9 + 1,9 + 2,1 + 2,2 + 2,2 + 2,3 + 2,3}{8} = 2,1$$

No ano 1, temos um valor de DG médio igual a 2,5; no ano 2, de 2,3; e, no ano 3, de 2,1. Esse conjunto de dados aponta um indício de redução no diâmetro das galerias com o passar dos anos, o que, possivelmente, está relacionado a uma diminuição no tamanho dos caranguejos da região.

IMPORTANTE !

Valores de média, como os expostos aqui, tornam-se mais precisos com o aumento do esforço de coleta, ou seja, quando aumentamos o número de amostras. Isso acontece porque, ao fazer isso, chegamos mais perto do valor real daquele indicador, isso é, o valor que de fato representa melhor toda a população.



SAIBA MAIS

Acabamos de conhecer a média aritmética simples. Porém, existem outros tipos de média com aplicações específicas que não trataremos aqui. Se quiser saber mais, você pode pesquisar sobre as médias aritméticas ponderada, geométrica e harmônica.

3.3 Mediana

A mediana pode ser definida como sendo o valor do meio de um conjunto de dados, quando dispostos em ordem crescente. Basicamente, basta distribuir os dados em ordem crescente, de forma que metade dos valores fique acima do valor central e a outra metade abaixo dele.

Existem algumas orientações que ajudam a definir o valor da mediana para um conjunto de dados. Vamos a elas:

DESTAQUE

- Quando o número total de elementos que compõem o conjunto de dados for ímpar, existirá apenas um valor do meio, que representa a mediana.
- Quando o número total de elementos que compõem o conjunto de dados for par, haverá dois valores do meio e a mediana será o resultado da média aritmética simples entre os dois elementos.

Vamos a um exemplo prático:

Na tabela a seguir, constam dados hipotéticos do alvo habitat, componente Igarapé, subprograma Aquático Continental.

Profundidade do igarapé em diferentes trechos, dados hipotéticos do alvo habitat, componente Igarapé, subprograma Aquático Continental

UC	DATA	EQUIPE	UA	TRECHO	PROFUNDIDADE (m)
ESEC X	Ano 1	Maria, José, Lucia	IG_1	0 m	0,5
ESEC X	Ano 1	Maria, José, Lucia	IG_1	0 m	0,9
ESEC X	Ano 1	Maria, José, Lucia	IG_1	0 m	1,1
ESEC X	Ano 1	Maria, José, Lucia	IG_1	0 m	1,1
ESEC X	Ano 1	Maria, José, Lucia	IG_1	0 m	0,4
ESEC X	Ano 1	Maria, José, Lucia	IG_1	25 m	0,5
ESEC X	Ano 1	Maria, José, Lucia	IG_1	25 m	0,7



ESEC X	Ano 1	Maria, José, Lucia	IG_1	25 m	0,9
ESEC X	Ano 1	Maria, José, Lucia	IG_1	25 m	0,8
ESEC X	Ano 1	Maria, José, Lucia	IG_1	25 m	0,5
ESEC X	Ano 1	Maria, José, Lucia	IG_1	50 m	0,2
ESEC X	Ano 1	Maria, José, Lucia	IG_1	50 m	0,8
ESEC X	Ano 1	Maria, José, Lucia	IG_1	50 m	1,3
ESEC X	Ano 1	Maria, José, Lucia	IG_1	50 m	0,8
ESEC X	Ano 1	Maria, José, Lucia	IG_1	50 m	0,6
ESEC X	Ano 1	Maria, José, Lucia	IG_1	75 m	0,5
ESEC X	Ano 1	Maria, José, Lucia	IG_1	75 m	0,9
ESEC X	Ano 1	Maria, José, Lucia	IG_1	75 m	1,1
ESEC X	Ano 1	Maria, José, Lucia	IG_1	75 m	0,6
ESEC X	Ano 1	Maria, José, Lucia	IG_1	100 m	0,3
ESEC X	Ano 1	Maria, José, Lucia	IG_1	100 m	0,8
ESEC X	Ano 1	Maria, José, Lucia	IG_1	100 m	0,7
ESEC X	Ano 1	Maria, José, Lucia	IG_1	100 m	0,6

Observe que os dados da tabela são referentes a um único evento de coleta e foram apurados na mesma unidade amostral e no mesmo ano. Perceba que cada unidade é subdividida em cinco trechos (0, 25, 50, 75 e 100 m) e neles são tomadas cinco medidas de profundidade. Para fins didáticos, os trechos 75 e 100 m do nosso exemplo terão apenas quatro medidas. Assim, ao todo, temos 23 medições de profundidades ao longo das cinco subdivisões do igarapé. São elas:

0,2; 0,4; 0,5; 0,5; 0,5; 0,5; 0,5; 0,6; 0,6; 0,6; 0,7; 0,8; 0,8; 0,8; 0,8; 0,8; 0,9; 0,9; 0,9; 1,1; 1,1; 1,1; 1,3

Com isso, a mediana para o conjunto de valores é 0,8.

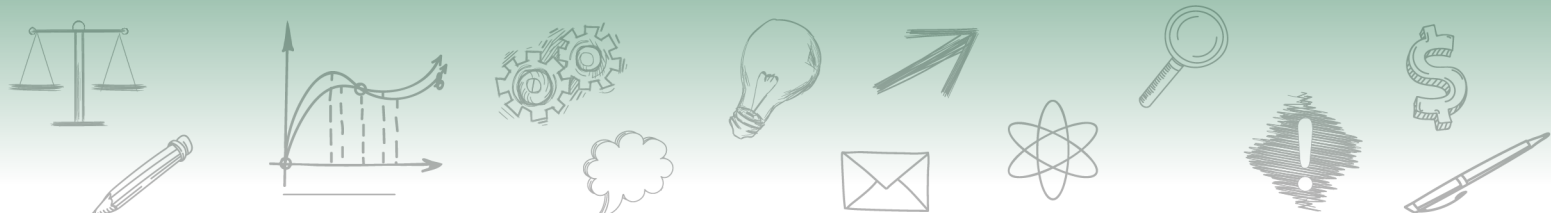
IMPORTANTE !

Dica: uma maneira rápida de encontrar o valor central em um conjunto de elementos de tamanho ímpar é somando 1 ao número total de elementos e dividindo esse valor por 2. O resultado dará a posição do elemento central ou mediana. No nosso exemplo, 23 é o número total de elementos, portanto: $(23+1)/2 = 12$. A décima segunda posição é onde está a mediana.

Agora, vamos calcular a mediana para cada um dos trechos.

Trecho 0 m: 0,4; 0,5; **0,9**; 1,1; 1,1.

A mediana corresponde ao elemento central do conjunto: 0,9.



Trecho 25 m: 0,5; 0,5; **0,7**; 0,8; 0,9.

A mediana corresponde ao elemento central do conjunto: 0,7.

Trecho 50 m: 0,2; 0,6; **0,8**; 0,8; 1,3.

A mediana corresponde ao elemento central do conjunto: 0,8.

Trecho 75 m: 0,5; 0,6; 0,9; 1,1.

Como temos 4 elementos, que é um número par, devemos calcular a média entre os dois elementos centrais. Logo, a mediana é 0,8, conforme o seguinte cálculo:

$$\bar{X} = \frac{0,6 + 0,9}{2} = \mathbf{0,8}$$

Trecho 100 m: 0,3; 0,6; 0,7; 0,8.

Também é um conjunto com número par de elementos. Logo, a mediana é 0,7, conforme o seguinte cálculo:

$$\bar{X} = \frac{0,6 + 0,7}{2} = \mathbf{0,7}$$

Observe que todos os valores de mediana obtidos, tanto para o conjunto total de dados como para cada um dos cinco trechos da UA, são valores próximos. Isso pode ser um indicativo de que esse igarapé apresenta uma estrutura homogênea.

DESTAQUE

A mediana, como medida de tendência central, geralmente é mais intuitiva que a média, uma vez que representa, de fato, o centro ou o meio do conjunto de dados. Observe que, ao conhecer a mediana, os valores extremos, mínimo e máximo, ficam evidentes.

Vamos usar o seguinte conjunto de valores:

7,86; 8,20; 8,60; 9,15; 9,20; 9,50; 9,90; 10,3; **10,3**; 10,6; 12,2; 12,6; 14,1; 15,3; 170.

Ao buscarmos a mediana, conseguimos perceber os valores mínimos e máximos. Observe como o valor máximo, 170, destoa dos demais.



IMPORTANTE !

Identificar valores extremos é um bom exercício para verificar potenciais equívocos que possam ter ocorrido na hora de anotar ou digitar os dados.

3.4 Moda

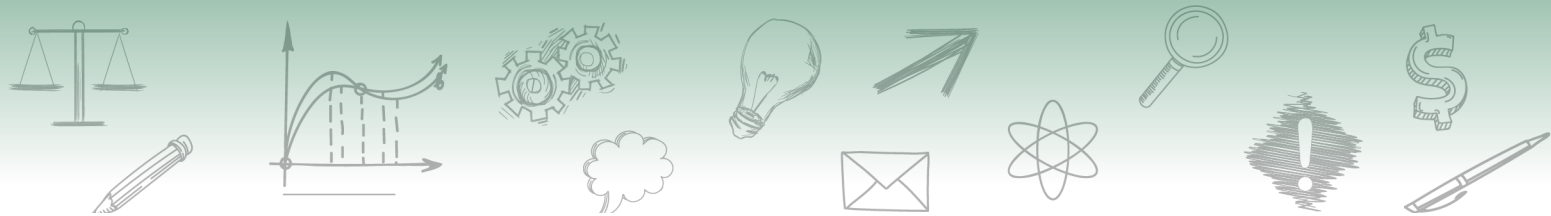
A moda é definida como a medição que ocorre mais frequentemente em um conjunto de dados. Essa medida de tendência é caracterizada pelo valor, ou faixa de valores, típico da variável. Podemos representar o que é típico por meio do valor que se repete mais vezes no conjunto.

DESTAQUE

A moda é a única das medidas de tendência central que faz sentido no caso de variáveis qualitativas. Assim, a que aparecer com maior frequência é chamada de categoria modal e, por isso, é representada pelo pico em um gráfico de frequência.

Número de plantas presentes em cada uma das classes de circunferência à altura do peito (CAP) definidas a partir de dados coletados do alvo plantas, componente Florestal, subprograma Terrestre

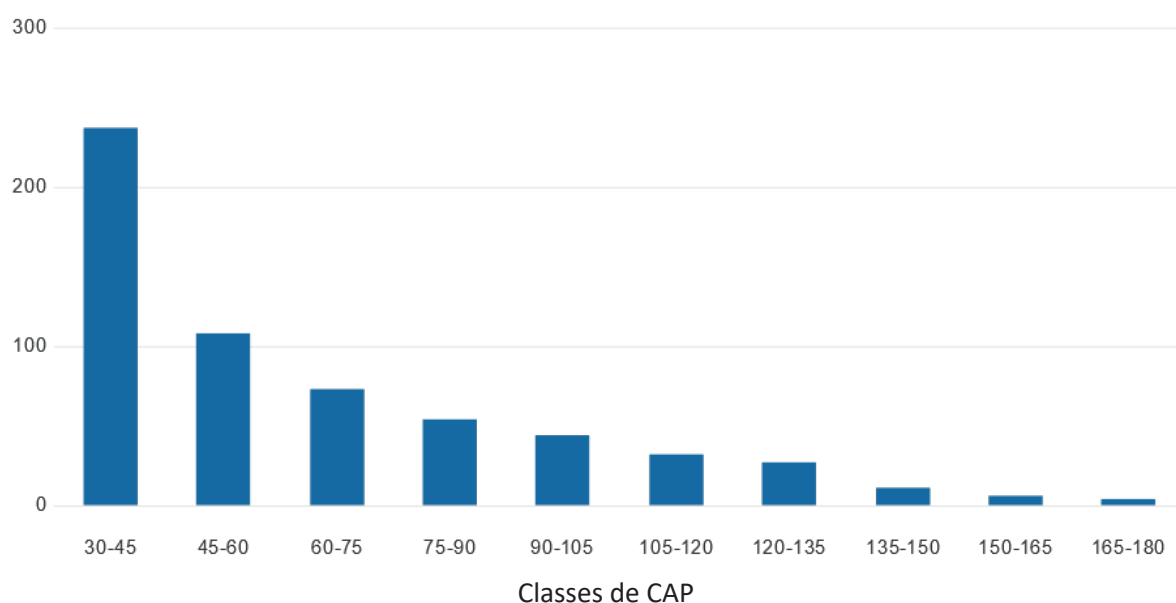
UC	DATA	EQUIPE	UA	CLASSES CAP (cm)	NÚMERO DE INDIVÍDUOS
ESEC X	Ano 1	Hugo, José e Luiz	1	30-45	237
ESEC X	Ano 1	Hugo, José e Luiz	1	45-60	108
ESEC X	Ano 1	Hugo, José e Luiz	1	60-75	73
ESEC X	Ano 1	Hugo, José e Luiz	1	75-90	54
ESEC X	Ano 1	Hugo, José e Luiz	1	90-105	44
ESEC X	Ano 1	Hugo, José e Luiz	1	105-120	32
ESEC X	Ano 1	Hugo, José e Luiz	1	120-135	27
ESEC X	Ano 1	Hugo, José e Luiz	1	135-150	11
ESEC X	Ano 1	Hugo, José e Luiz	1	150-165	6
ESEC X	Ano 1	Hugo, José e Luiz	1	165-180	4



A tabela apresenta dados hipotéticos do protocolo de plantas, componente Florestal, subprograma Terrestre, referentes a um mesmo evento de coleta.

A seguir, os mesmos dados são apresentados em um gráfico de frequência.

Número de plantas presentes por classes de circunferência à altura do peito (CAP) definidas a partir de dados coletados do alvo plantas, componente Florestal, subprograma Terrestre



Podemos perceber que a classe com os indivíduos cujo tamanho de circunferência à altura do peito variou entre 30 e 45 cm foi a mais numerosa e, no gráfico, representa o pico de frequência. Logo, ela é a nossa classe modal, pois representa a variação de CAP típica das árvores amostradas.

3.5 Comparando as medidas: média, mediana e moda

Vamos comparar as três medidas que acabamos de estudar.

DESTAQUE

A média é a medida de posição mais usada, principalmente devido à facilidade de tratamento estatístico e suas propriedades. No entanto, ela é muito influenciada pela presença de valores extremos: valores muito grandes ou muito pequenos em relação aos demais.



Como a média usa todos os valores coletados em seu cálculo, a presença de valores extremos pode tender o resultado em direção a eles, deslocando também a representação do centro.

Vamos voltar ao último conjunto de valores utilizado:

7,86; 8,20; 8,60; 9,15; 9,20; 9,50; 9,90; **10,3**; 10,3; 10,6; 12,2; 12,6; 14,1; 15,3; **170**.



Zona central

O resultado da média para esse conjunto de valores é igual a 21,18, conforme o seguinte cálculo:

$$X = \frac{7,86 + 8,20 + 8,60 + 9,15 + 9,20 + 9,50 + 9,90 + 10,3 + 10,3 + 10,6 + 12,2 + 12,6 + 14,1 + 15,3 + 170}{15} = 21,18$$

Podemos perceber como esse valor médio está deslocado da zona central dos valores, que oscilam entre 9,50 e 10,6. Isso ocorre devido a presença de um valor extremo: 170. Agora, vamos calcular a média desconsiderando esse valor:

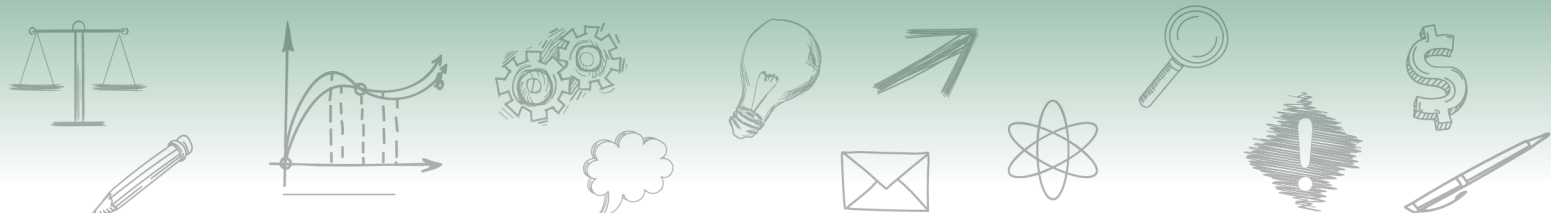
$$X = \frac{7,86 + 8,20 + 8,60 + 9,15 + 9,20 + 9,50 + 9,90 + 10,3 + 10,3 + 10,6 + 12,2 + 12,6 + 14,1 + 15,3}{14} = 9,82$$

O resultado é igual a 9,82. Um valor muito mais próximo da faixa central de valores do nosso conjunto, não é mesmo?

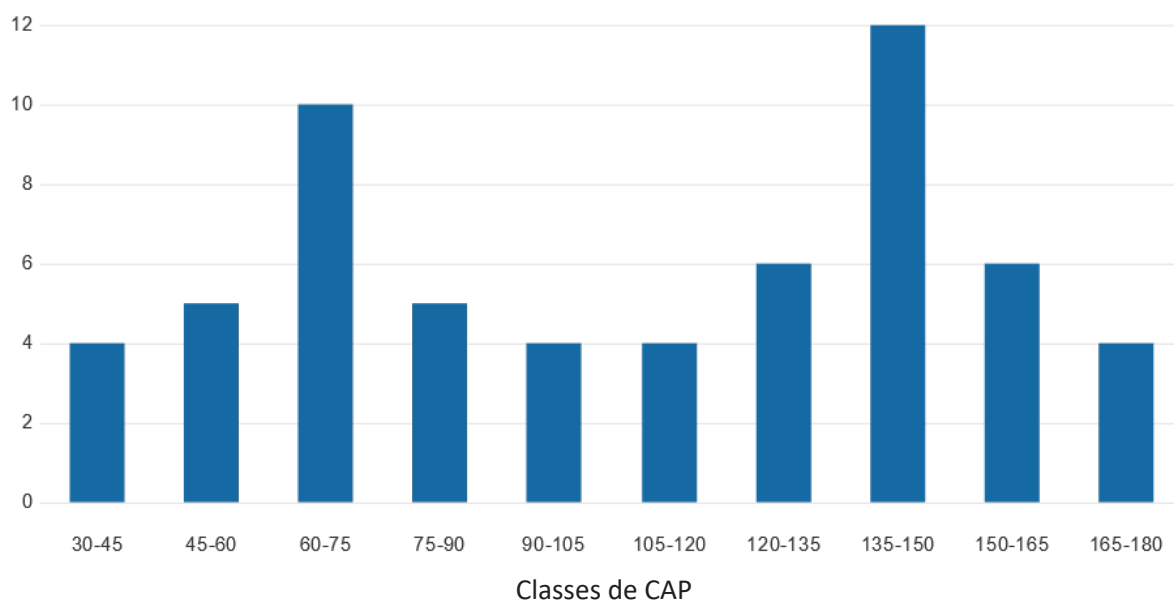
DESTAQUE

A mediana, por sua vez, não é tão influenciada por valores extremos, pois utiliza a ordem dos elementos, e não seus valores, para o cálculo. Assim, se um elemento do conjunto de dados tem o seu valor alterado, por um erro de anotação ou digitação, por exemplo, mas sua ordem continua a mesma, a mediana não sofre influência nenhuma. Entretanto, ela baseia-se em apenas um número de toda a amostra e não é possível associá-la a uma medida de variância.

A moda possui algumas vantagens sobre a média e a mediana, especialmente quando os dados possuem distribuição com dois (bimodais) ou mais (multimodais ou polimodais) picos de frequência. Acompanhe no gráfico a seguir:



Frequência de indivíduos arbóreos em classes de circunferência à altura do peito (CAP) em determinada UC



IMPORTANTE !

No gráfico, há claramente dois picos de frequência: a classe 60-75 e a classe 135-150. Logo, a distribuição mostrada é bimodal. Nem a média nem a mediana resumiriam bem a informação contida nesse conjunto de dados. Por isso, a melhor alternativa é usar a moda.

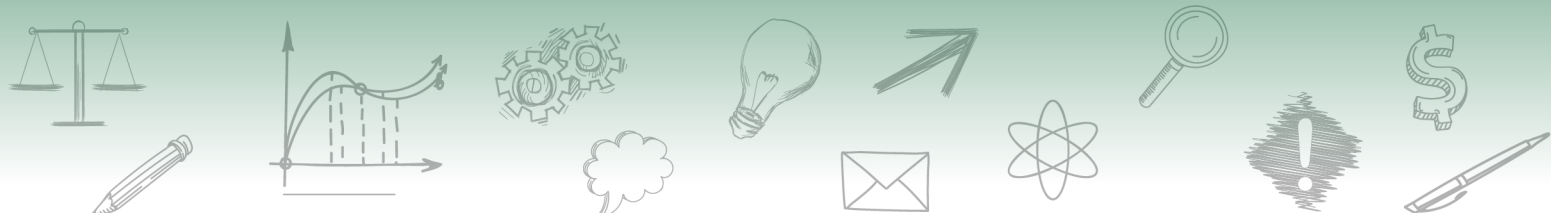
Essas duas classes modais são as que representam da melhor forma os valores típicos do conjunto de dados. Além disso, a moda deixa nítida a existência de dois grandes grupos de plantas presentes na UC.

A figura a seguir resume alguns dos principais pontos positivos e limitações da média, mediana e moda.



Resumo esquemático dos pontos positivos e negativos da utilização da média, mediana e moda.

Você pode estar se perguntando: “mas qual dessas medidas de tendência central devo usar para os meus dados?”. A resposta é todas. Devemos sempre explorar todas as medidas de tendência central, entretanto, é bom ter atenção às diferentes informações que cada uma delas pode fornecer sobre o conjunto de dados.



Referências

Unidade 1

CIÊNCIA. In: DICIONÁRIO online de português. Porto: 7Graus, 2020. Disponível em: <https://www.dicio.com.br/ciencia/>. Acesso em: 10 jun. 2019.

CONSTANTINO, P. A. L.; CRUZ, A. T. **Monitoreo de la Biodiversidad en América Latina**: panorama y recomendaciones para estructurar una iniciativa. Brasília, GKNoronha, 2016.

DANIELSEN, F.; BURGESS, N. D.; JENSEN, P. M.; PIRHOFFER-WALZL, K. Environmental monitoring: the scale and speed of implementation varies according to the degree of people involvement. **Journal of Applied Ecology**, London, v. 47, n. 6, p. 1166-1168, Dec. 2010.

FORTUNATO, S.; BERGSTROM, C. T.; BÖRNER, K.; EVANS, J. A.; HELBING, D.; MILOJEVIĆ, S.; PETERSEN, A. M.; RADICCHI, F.; SINATRA, R.; UZZI, B.; VESPIGNANI, A.; WALTMAN, L.; WANG, D.; BARABÁSI, A. L. Science of Science. **Science**, v. 359, n. 6379, p. 1-7, 2 Mar. 2018.

RIBEIRO, K. T. (org.). **Estratégia do Programa Nacional de Monitoramento da Biodiversidade**: Programa Monitora: estrutura, articulações e perspectivas. 1. ed. Brasília: ICMBio, 2018.

VOLPATO, G. L. O método lógico para redação científica. **RECIIS**, Rio de Janeiro, v. 9, n. 1, p. 1-14, jan./mar. 2015.

Unidade 2

CALLEGARI-JACQUES, S.M. **Bioestatística**: princípios e aplicações. Porto Alegre: Artmed, 2003.

FISHER, R. A. **Statistical Methods for Research Workers**. London: Oliver and Boyd, 1925.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**: data mining, inference and prediction. 2nd ed. New York: Springer, 2017.

GOTELLI, N. J.; ELLISON, A. M. **Princípios de Estatística em Ecologia**. Porto Alegre: Artmed, 2011.

GUEDES, T. A.; ACORSI, C. R. L.; MARTINS, A. B. T.; JANEIRO, V. **Estatística descritiva**. Projeto de ensino: aprender fazendo estatística. Maringá: Universidade Federal de Maringá, 2005. Disponível em: http://www.each.usp.br/rvicente/Guedes_et al_Estatistica_Descritiva.pdf. Acesso em: 20 abr. 2019.

REIS, E. A.; REIS, I. A. **Análise Descritiva de Dados**. 1. ed. Belo Horizonte: UFMG, 2002. Disponível em: <http://www.est.ufmg.br/portal/arquivos/rts/rte0202.pdf>. Acesso em: 20 jun. 2019.



Unidade 3

GUEDES, T. A.; ACORSI, C. R. L.; MARTINS, A. B. T.; JANEIRO, V. **Estatística descritiva**. Projeto de ensino: aprender fazendo estatística. Maringá: Universidade Federal de Maringá, 2005. Disponível em: <http://www.each.usp.br/rvicente/Guedes_et al _Estatistica_Descritiva.pdf>. Acesso em: 20 abr. 2019.

REIS, E. A.; REIS, I. A. **Análise Descritiva de Dados**. 1. ed. Belo Horizonte: UFMG, 2002. Disponível em: <http://www.est.ufmg.br/portal/arquivos/rts/rte0202.pdf>. Acesso em: 20 jun. 2019.