



Apresentação

A estatística é uma ciência que consiste em métodos para delinear estudos e analisar dados coletados nos estudos. Os métodos para analisar dados incluem métodos descritivos para resumir dados (estatística descritiva) e métodos inferenciais para fazer previsões (estatística inferencial).

Os dados coletados, quando submetidos a técnicas de análise, proporcionam conhecimento e são fundamentais para a vida social e profissional dos indivíduos, contribuindo para a elaboração de um planejamento estratégico e para a tomada de decisões.

Nesta Unidade de Aprendizagem, você irá aprender os elementos da estatística descritiva, especialmente quanto à distribuição de frequência. Você também estudará a estatística inferencial e os procedimentos para fazer generalizações sobre as características de uma população a partir de informações contidas em amostras.

Bons estudos.

Ao final desta Unidade de Aprendizagem, você deve apresentar os seguintes aprendizados:

- Reconhecer os elementos da estatística descritiva.
- Elaborar distribuições de amostragem e intervalos de confiança.
- Descrever as abordagens de probabilidade.



Desafio

Quando se utiliza somente um valor para inferir um parâmetro de interesse, o procedimento é denominado "estimação pontual". O objetivo é que este único valor numérico esteja próximo do verdadeiro valor do parâmetro.

Considerando o exposto, imagine a seguinte situação:

Carlos está preocupado com seus gastos com alimentação, por isso procurou você para ajudá-lo. Você recomendou que ele anotasse os valores gastos nas últimas 10 compras realizadas no supermercado para fazer a estimativa.



Os valores foram os seguintes:

- ▶ R\$325,52
- ▶ R\$310,87
- ▶ R\$294,96
- ▶ R\$305,72
- ▶ R\$334,66
- ▶ R\$319,88
- ▶ R\$286,66
- ▶ R\$291,11
- ▶ R\$289,31
- ▶ R\$320,15

Com base no seu conhecimento profissional sobre estimação pontual, responda as seguintes questões:

- a) Qual é a estimativa pontual média dos gastos de Carlos com alimentação no supermercado?
- b) Qual é a estimativa pontual da proporção dos gastos de Carlos com alimentação em valores abaixo de R\$300,00?



Infográfico

A análise estatística fornece informações acerca dos dados coletados em uma pesquisa, contribuindo para a tomada de decisões. Mas você sabe quais são os tipos de análise estatística? Você sabe diferenciar os conjuntos de dados?

Neste Infográfico, confira os processos da análise estatística e seus principais conceitos.

ANÁLISE ESTATÍSTICA

A análise estatística envolve dois processos: **estatística descritiva e estatística inferencial**.

Uma análise estatística coleta, resume e apresenta os dados (descreve dados) e busca descobrir padrões e tendências (busca inferência).

Para se obter bons resultados em uma análise estatística, além da aplicação correta de métodos, também é necessário ter clareza dos conceitos utilizados. A seguir, são apresentados alguns desses conceitos (FIGUEIREDO FILHO, 2019):



População – Número total de observações ou de mensurações de interesse.



Amostra – Parte específica da população, um subgrupo do total.



Estatística – Descrição da amostra.



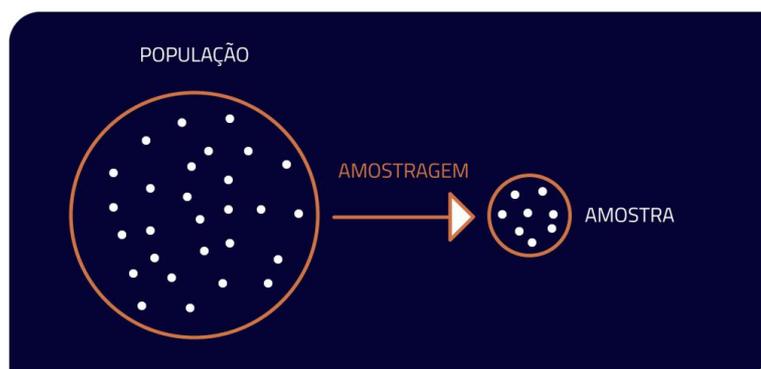
Inferência – Processo de utilização de informações da amostra por inferir sobre as características da população.



Amostragem – Técnica de seleção da amostra.

Conjuntos de dados: população e amostra

Quando se pretende conhecer as características da amostra, utiliza-se a estatística na percepção descritiva, e quando se busca dizer algo sobre a população com base nas informações da amostra, utiliza-se a estatística inferencial. A figura a seguir apresenta a **relação entre população, amostragem e amostra**:



A principal vantagem da amostra é a coleta de informações de forma mais ágil. No entanto, para que uma amostra seja válida, é necessário que ela seja extraída com base em critérios rígidos. **A inferência estatística utiliza o método estatístico em dados amostrais e tira conclusões sobre a população de interesse, descrevendo-a.**



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.



Conteúdo do livro

Os métodos estatísticos dão subsídio para tomada de decisões. Eles consideram todas as causas envolvidas no processo como variável e procuram determinar, no resultado final, que influências cabem a cada uma delas.

A análise estatística é realizada por meio da estatística descritiva, que resume os dados, e da estatística inferencial, que faz previsões baseadas nos dados amostrais. A estatística descritiva busca condensar os dados, usando elementos para descrevê-los, tais como tabelas e diagramas. A inferência estatística tem interesse nos parâmetros desconhecidos da população, buscando relatar a previsão provável da amostra.

No capítulo **Métodos estatísticos aplicados**, base teórica desta Unidade de Aprendizagem, você vai estudar a estatística descritiva e a elaboração de distribuições de frequências. Você vai aprender estatística inferencial e algumas técnicas utilizadas neste método (estimação pontual e intervalo de confiança). Por fim, vai conhecer as principais abordagens de probabilidade (clássica, empírica e subjetiva).

Boa leitura.

INVESTIMENTOS, FINANCIAMENTOS, ESTRATÉGIAS E LUCROS ECONÔMICOS



sagah⁺

Métodos estatísticos aplicados

Flávia Monaco Vieira

OBJETIVOS DE APRENDIZAGEM

- > Reconhecer os elementos da estatística descritiva.
- > Elaborar distribuições de amostragem e intervalos de confiança.
- > Descrever as abordagens de probabilidade.

Introdução

O método estatístico é utilizado para simplificar a análise quantitativa de todas as variantes envolvidas em determinado processo. Basicamente, sua função é usar dados para estudar um fenômeno cujos resultados estão associados a um forte grau de incerteza. Existem essencialmente dois tipos de procedimentos estatísticos: descritivo e inferencial. Enquanto estatística descritiva visa descrever dados, a estatística inferencial preocupa-se com o raciocínio necessário para compreender os dados.

Neste capítulo, você estudará os elementos da estatística descritiva e da estatística inferencial. Além disso, aprenderá a elaborar uma distribuição de frequência, uma estimativa pontual e um intervalo de confiança. Por fim, conhecerá as principais abordagens de probabilidade e seus cálculos.

Elementos da estatística descritiva

O primeiro passo do processo estatístico é realizado por meio da estatística descritiva, utilizada para descrever e compreender os dados. Ainda que os dados estejam geralmente disponíveis apenas para uma amostra, as estatís-

ticas descritivas são úteis também quando os dados estão disponíveis para toda a população, como em um censo.

De acordo com Agresti e Finlay (2012, p. 20) “uma análise estatística é classificada como descritiva se seu propósito principal for descrever dados”. Por sua vez, Silvestre (2007) afirma que o objetivo final da análise estatística é tirar conclusões a partir de dados a fim de enriquecer nosso conhecimento sobre a realidade.

A partir dos métodos de estatística descritiva, é possível organizar, classificar e apresentar os dados mais importantes de um conjunto de características observadas, sendo possível ainda compará-los a outros conjuntos de dados. Para tanto, os dados são reduzidos para facilitar seu entendimento, buscando sempre não distorcer ou perder muita informação (AGRESTI; FINLAY, 2012).

Embora ocorram perdas nessa condensação de dados, uma vez que não se têm as observações originais, essa perda de informação é pequena quando comparada ao ganho obtido com as possibilidades de interpretação oportunizadas pelo resumo dos dados.

A estatística descritiva é composta por elementos que ajudam a organizar os dados. Os elementos descritivos consistem em gráficos, tabelas e medidas de sínteses (porcentagens, índices e médias). Vejamos a seguir detalhes sobre a distribuição de frequência.

Distribuição de frequência

Existem diferentes maneiras de representar os dados, mas sempre é necessário descrevê-los de forma precisa e esclarecedora (BONAFINI, 2014). Desse modo, costuma-se normalmente distribuí-los em classes ou categorias convenientes, ou seja, determinando o número de indivíduos pertencentes a cada classe/categoria.

Esse arranjo tabular dos dados, juntamente com as frequências, corresponde à distribuição de frequências, também chamada de tabela de frequências. Bonafini (2014, p. 78) conceitua a distribuição de frequência como “uma tabela simples que mostra como a variável se dissipa de acordo com divisões por nós estabelecidas”. Seu objetivo é resumir grandes massas de dados brutos em tabelas que listam os valores observados em correspondência com suas repetições verificadas, o que agiliza a análise dos dados (SAMPAIO; ASSUMPÇÃO; FONSECA, 2018).

Podem-se dividir as distribuições de frequência em dois tipos:

- sem intervalos de classe — todos os valores dos dados coletados são apresentados e não há perdas de valores;

- com intervalos de classe — os valores estão representados por faixas de magnitude.

Distribuição de frequência sem intervalos de classe

A construção de uma distribuição de frequência sem intervalos de classe é equivalente à construção de uma tabela simples onde se listam os diferentes valores observados da variável com suas frequências absolutas (F_i).

Em outras palavras, a frequência absoluta corresponde ao número de observações encontradas numa classe ou intervalo específico. Sua equação é assim representada:

$$\sum F_i = n$$

Vejamos a seguir um exemplo de distribuição de frequência sem intervalos de classe.



Exemplo

Suponha que determinada empresa fabricante de peças automotiva queria conhecer a frequência absoluta da sua produção. Para tanto, durante dez dias seguidos foram verificadas as produções diárias. Eis os dados coletados referentes a peças produzidas por dia.

1.120	1.125	1.125	1.123	1.110
1.123	1.120	1.125	1.123	1.125

Eis a distribuição de frequência com base nos dados acima são:

Peças produzidas	Frequência (F_i)
1.110	1
1.120	2
1.123	3
1.125	4
Total	10

Podemos observar que na tabela de distribuição todos os valores da variável (quantidade produzida) foram considerados e, portanto, não houve perda de informação. A frequência foi organizada em ordem crescente e a soma total da frequência ($\sum F_i$) é igual a 10.

Ao ordenarmos os valores de uma variável, fica mais simples analisar seu comportamento. Essa ordenação pode ser crescente ou decrescente.

Como explicado anteriormente, a variável é resumida em frequências absolutas, sendo o total das F_i igual ao número total de observações da variável.

Podemos incluir como colunas complementares numa distribuição de frequências os seguintes itens.

- Frequência percentual ($f_i\%$): representa o percentual de observações pertencente à variável. Nesse caso, a soma da frequência deve ser igual a 100%. Equação da frequência percentual: $f_i\% = (F_i \div n) \times 100$.
- Frequência absoluta acumulada (F_{ai}): representa a soma das frequências absolutas dos valores anteriores. Sua equação é assim definida:
 $F_{ai} = F_1 + F_2 + \dots + F_i$.
- Frequência acumulada relativa ($f_{ai}\%$): representa a porcentagem de vezes que uma frequência apareceu em relação ao todo. A frequência relativa é utilizada para comparar um dado em relação ao todo. A equação é assim determinada: $f_{ai}\% = (F_{ai} \div n) \times 100$.

Uma tabela que contém todas essas frequências é vista como uma distribuição de frequências completa. Vejamos a seguir um exemplo de tabela de frequências completa.



Exemplo

Utilizando a distribuição de frequência elaborada no exemplo anterior, foram incluídas mais três colunas: frequência percentual ($f_i\%$), frequência absoluta acumulada (F_{ai}) e frequência acumulada relativa ($f_{ai}\%$).

Peças produzidas	F_i	$f_i\%$	F_{ai}	$f_{ai}\%$
1.110	1	10%	1	10%
1.120	2	20%	3	30%
1.123	3	30%	6	60%
1.125	4	40%	10	100%
Total	10	100%		

Podemos observar que a frequência absoluta consiste no número de vezes que determinada variável (dado) se repetiu, enquanto a frequência relativa compara a frequência daquele dado em relação ao todo.

Distribuição de frequência com intervalos de classe

Quando os dados são agrupados em classes/categorias, temos aquilo que chamamos de distribuição de frequência com intervalos de classe. Nesse caso, é considerada a frequência em cada classe (quantidade de observações pertencentes). O uso dessa tabela é recomendado para apresentar dados quantitativos contínuos com um número elevado de possíveis valores.

Para construir uma distribuição de frequência com intervalos de classe, é necessário conhecer algumas definições (GUEDES *et al.*, 2005):

- Número de classes (k) — representa o total de classes da variável. Uma classe é uma linha da distribuição de frequências, sendo que o menor valor da classe é denominado limite inferior (l_i) e o maior valor da classe é denominado limite superior (L_i). Cabe ao analista de dados definir o número de classes mais adequado, com base em sua experiência. Vale ressaltar que, se houver muitos intervalos, o resumo não constituirá grande melhoria com relação aos dados brutos; já se houver poucos intervalos, um grande volume de informação se perderá.
- Amplitude das classes (h) — corresponde à divisão entre a amplitude total (AT) e o número de classes (k). A AT é a diferença entre o limite superior da última classe e o limite inferior da primeira classe. O valor só poderá ser arredondado para mais. Desse modo, todas as classes terão a mesma amplitude, o que permitirá a construção de gráficos e o cálculo de medidas descritivas. Equação: $h = AT \div k$.
- Ponto médio (X_i) — corresponde à média aritmética entre o limite superior e o limite inferior. O ponto médio será representado por X_i , onde $X_i = (l_i + L_i) \div 2$, $i = 1, \dots, k$.

Vejamos a seguir um exemplo de distribuição de frequência com intervalos de classe.



Exemplo

Suponha que uma sala de aula tenha 20 alunos e a tabela de frequências referente a suas notas em determinada prova de matemática seja a seguinte:

Notas	F_i	$f_i\%$
0–20	0	0%
20–40	3	7,5%
40–60	9	22,5%
60–80	23	57,5%
80–100	5	12,5%
Total	40	100%

Com base na tabela de distribuição, observamos o seguinte:

- Há 5 classes (k) na tabela de frequência.
- Os intervalos da variável (i) notas 0–20 corresponde à 1ª classe, 20–40 à 2ª classe, 40–60 à 3ª classe, 60–80 à 4ª classe, e 80–100 à 5ª classe.
- Os limites da 1ª classe (notas 0–20) são: limite inferior (l_i) igual a 0 e limite superior (L_i) igual a 20.
- A amplitude das classes (h) é igual a 20, o que corresponde à divisão da amplitude total ($AT = 100$) pelo número de classes ($k = 5$).
- O ponto médio da 1ª classe (X_1) é igual 10 (ou seja, $0 + 20 \div 2$), o ponto médio da 2ª classe (X_2) é 30 ($20 + 40 \div 2$), e assim por diante.

Nas distribuições de frequência, todos os elementos são fixos. O fenômeno especificado se apresenta por meio de gradações, isto é, os dados referentes ao fenômeno que se está representando são reunidos de acordo com sua magnitude.

A seguir estudaremos a estatística inferencial, aplicando as principais técnicas de inferência estatística para estimar os parâmetros populacionais e os intervalos de confiança.

Distribuição de amostragem e os intervalos de confiança

Como vimos anteriormente, a estatística descritiva dispõe de elementos que possibilitam a descrição correta dos dados. Com base nela, podem-se

utilizar algumas técnicas para atingir um grau de segurança que permite ao investigador indicar com grande probabilidade de certeza se o valor é correspondente ao parâmetro populacional ou não (BONAFINI, 2015).

A partir daqui nosso estudo se volta para a estatística inferencial. A inferência estatística se preocupa em tomar decisões sobre determinada população com base apenas nas informações contidas numa amostra aleatória dela.



Saiba mais

Há essencialmente dois tipos de procedimentos em estatística: a estatística descritiva, que tem como objetivo descrever os dados, e a estatística inferencial, que tem por finalidade obter uma afirmação acerca de uma população com base numa amostra. A estatística inferencial se preocupa com o raciocínio necessário para, a partir dos dados, obter conclusões gerais.

Alguns conceitos são importantes no estudo da estatística inferencial:

- estimativa pontual — consiste no uso de dados da amostra (dados amostrais) para estimar valores de parâmetros populacionais desconhecidos;
- confiança — demonstra se o investigador pode “acreditar” num valor calculado;
- teste de hipóteses — utilizado quando é necessário comprovar uma possibilidade que aparenta ser verdadeira.

A **estimativa pontual** utiliza os dados da amostra para chegar a um único número que representa um valor plausível para a característica de interesse. Em outras palavras, é um dado encontrado na amostra e que o investigador quer verificar se é correspondente à população (BONAFINI, 2015).

Pode-se usar como parâmetros a média (μ), a variância (σ^2), o desvio-padrão (σ) e a proporção (p) para apurar a estimativa pontual. A média amostral \bar{x} é a estimativa pontual menos tendenciosa para o parâmetro μ . Para apurar a estimativa pontual da média, basta somar todos os elementos e dividir pela quantidade de unidades (n).

Vejamos a seguir um exemplo de estimativa pontual proporcional.



Exemplo

Um estabelecimento varejista planeja estender seu horário de funcionamento, e para tanto buscou estimar a proporção de clientes que iriam frequentar o estabelecimento durante as horas estendidas.

Foram entrevistados 862 clientes, e 221 deles afirmaram que frequentariam o estabelecimento nesse horário.

A estimativa pontual proporcional é dada pela estatística:

$$p = \frac{\text{número de sucesso na amostra}}{n}$$

$$p = \frac{221}{862}$$

$$p = 0,2564 \times 100 = 25,64\%$$

Com base nessa amostra aleatória, o estabelecimento estimou que aproximadamente 26% dos clientes iriam frequentar o estabelecimento no horário estendido.

Apesar da sua grande utilidade nos estudos estatísticos, o resultado da estimativa pontual depende de qual amostra foi selecionada entre as amostras possíveis. Geralmente, amostras diferentes produzirão estimativas diferentes, devido às diferenças aleatórias de uma amostra para a outra. Por causa disso, dificilmente a estimativa pontual a partir de uma amostra é exatamente igual ao valor real da característica na população.

Desse modo, é necessário indicar de algum modo um intervalo de valores plausíveis para a característica da população, como veremos a seguir.

Intervalo de confiança

Um intervalo de confiança para uma característica da população (parâmetro) é um intervalo de valores plausíveis para esta característica. O nível de confiança c é a probabilidade de que o intervalo estimado contenha o parâmetro populacional (LARSON; FARBER, 2010).

O grau de confiança tem origem na probabilidade associada ao processo de construção do intervalo antes de se obter o resultado amostral. Geralmente, os níveis de confiança são de 90%, 95% ou 99%.

Vejamos a seguir um exemplo simplificado de como funciona o intervalo de confiança.



Exemplo

Suponha que uma empresa deseja encontrar o número médio de acidentes laborais por funcionário ocorridos em sua produção. No entanto, a empresa tem muitos funcionários e, portanto, só tem acesso a uma amostra do todo.

Se a média encontrada na amostra foi de 6 incidentes, com nível de confiança de 95% no intervalo de confiança $[5;7]$, que vai de 5 a 7, pode-se concluir que:

- há 95% de confiança de que o intervalo de confiança [5;7] contém a média populacional verdadeira (aquela que seria encontrada se tivéssemos acesso a todos os funcionários);
 - ao obter uma nova amostra, em 95% das vezes o valor verdadeiro do parâmetro vai estar no intervalo de confiança estabelecido.
-

Um intervalo de confiança é um intervalo de valores utilizado para estimar o verdadeiro valor de parâmetro populacional. Esse parâmetro obscuro é descoberto por meio de um modelo de amostragem calculado a partir dos dados coletados. Um intervalo de confiança sinaliza a margem de incerteza a partir do cálculo realizado. Calcular um intervalo de confiança é uma estratégia que leva em consideração o erro de amostragem.

A seguir, estudaremos as principais abordagens de probabilidade e seus conceitos. Este estudo possibilita ao investigador realizar inferências de fenômenos futuros com base nos dados coletados, isto é, proporciona formas de medir a incerteza dos eventos.

Abordagens de probabilidade

A probabilidade é uma das formas mais conhecidas de estudar situações em que os resultados são variáveis. Ela permite o estudo da incerteza oriunda de fenômenos de caráter aleatório e constitui a base da estatística inferencial (BONAFINI, 2014).

Antes de aprofundarmos o estudo das abordagens de probabilidade, é necessário conhecer alguns de seus conceitos (BONAFINI, 2014):

- **experimento de probabilidade** — corresponde a uma ação pela qual os resultados são obtidos;
- **espaço amostral** — consiste no universo da probabilidade, ou seja, o “todo” do qual as “partes” serão extraídas;
- **resultado** — diz respeito ao que extraímos de maneira individual de cada dado, em que cada resultado composto por uma única variável;
- **evento** — refere-se ao que de fato queremos calcular, representando um subgrupo do espaço amostral.

De acordo com Larson e Farber (2010), o método a ser utilizado para calcular uma probabilidade dependerá do tipo e probabilidade. Ela pode ser escrita como frações, decimais ou porcentagens.

Existem três tipos de probabilidades: clássica, empírica e subjetiva. Em todos os cálculos, usa-se a notação $P(E)$, que se lê: a probabilidade de ocorrência do evento E . Vejamos a seguir a característica de cada tipo de probabilidade.

Probabilidade clássica

A probabilidade clássica é usada quando cada resultado num espaço amostral é igualmente possível de ocorrer. Trata-se da divisão entre o evento que buscamos e o total de eventos presentes no espaço amostral (BONAFINI, 2015; LARSON; FARBER, 2010).

A representação da probabilidade clássica para um evento E é:

$$P(E) = \frac{\text{Número de resultados no evento } E}{\text{Número total de resultados no espaço amostral}}$$

Vejamos a seguir um exemplo envolvendo probabilidade amostral.



Exemplo

Suponha que você tem três canetas de cores diferentes (vermelha, azul e preta) dentro de uma gaveta. Você abre a gaveta e pega uma caneta sem olhar. Qual é a probabilidade de ter pegado a caneta vermelha?

A probabilidade de ocorrência do evento $P(E)$ é $1/3$, ou seja, o número de eventos que nos interessa (1) dividido pelo total de eventos (3).

E se você abrisse a gaveta e pegasse duas canetas sem olhar. Qual seria a probabilidade de você pegar a caneta vermelha?

Agora a probabilidade seria de $2/3$.

De modo geral, a probabilidade clássica calcula as chances de um determinado evento ocorrer.

Probabilidade empírica (frequência relativa)

A probabilidade empírica é usada quando um experimento é repetido várias vezes, o que, por sua vez, cria padrões regulares. A probabilidade empírica pode ser usada mesmo se cada resultado de um evento não tiver a mesma chance de ocorrer (LARSON; FARBER, 2010).

A probabilidade empírica de um evento E é a frequência relativa desse evento, e é assim representada:

$$P(E) = \frac{\text{Frequência do evento } E}{\text{Frequência total}}$$

Conforme aumenta o número de vezes que um experimento de probabilidade é repetido, a probabilidade empírica de um evento aproxima-se de sua probabilidade clássica. Vejamos a seguir um exemplo de probabilidade empírica.



Exemplo

Suponha que determinada empresa esteja organizando o lançamento de um novo produto no mercado. Como preparação, ela conduziu uma pesquisa com 500 possíveis consumidores sobre a qualidade de tal produto. O resultado da pesquisa foi o seguinte: 82 acharam o produto de baixa qualidade, 123 consideraram o produto com qualidade aceitável e 295 descreveram o produto como de alta qualidade.

Com base na pesquisa de qualidade, qual é a probabilidade de um futuro cliente considerar o produto de baixa qualidade?

A probabilidade de ocorrência do evento $P(E)$ é de 16,40%. Acompanhe o cálculo:

$$P(E) = \frac{\text{Frequência do evento } E}{\text{Frequência total}}$$

$$P(E) = \frac{82}{500} = 0,164$$

$$P(E) = 0,164 \times 100 = 16,4\%$$

Probabilidade subjetiva

A probabilidade subjetiva resulta da intuição de suposições fundamentais e estimativas. Normalmente, ela não é realizada por meio de cálculos, já que é oriunda de uma estimativa (BONAFINI, 2014; LARSON; FARBER, 2010).

Vejamos a seguir um exemplo de probabilidade subjetiva.



Exemplo

Imagine que um casal de amigos tenha acabado de noivar e que você foi convidado para a festa de noivado. Durante a festa, várias pessoas especulam sobre quando será a data do casamento.

A noiva responde: provavelmente o casamento será daqui a 1 ano.

E o noivo complementa: com certeza não será este ano, pois estamos juntando o dinheiro.

Observe que a probabilidade do evento ocorrer esse ano é 0%, mas as chances aumentam para o próximo ano. No entanto, essas chances não passam de opiniões.

Como visto, a probabilidade é o estudo das chances numéricas de um evento ocorrer em um experimento aleatório, obtida pela razão entre o evento desejado e os eventos possíveis. A probabilidade é usada para determinar a relação entre uma amostra e uma população. Por sua vez, a probabilidade subjetiva é baseada no julgamento pessoal, na forma de um palpite.

Por fim, podemos concluir que os métodos estatísticos são utilizados para coleta, organização, descrição, análise e interpretação de dados, dando subsídio à tomada de decisões. Por meio dos métodos estatísticos, o investigador estuda um determinado fenômeno. Há dois tipos de procedimentos estatísticos: a estatística descritiva, que busca descrever os dados, e a estatística inferencial, que se empenha em compreender os dados. A partir dos elementos da estatística descritiva, é possível sintetizar uma série de valores, o que facilita sua análise. Já a partir dos métodos de análise inferencial é possível interpretar os dados e tirar conclusões. Esses métodos fundamentam-se nas abordagens de amostragem e probabilidade.

Referências

AGRESTI, A.; FINLAY, B. *Métodos estatísticos para as ciências sociais*. 4. ed. Porto Alegre: Penso, 2012.

BONAFINI, F. C. (org.). *Estatística*. São Paulo: Pearson, 2015.

BONAFINI, F. C. (org.). *Matemática e estatística*. São Paulo: Pearson, 2014.

GUEDES, T. A. et al. Estatística descritiva. In: GUEDES, T. A. et al. *Projeto de ensino: aprender fazendo estatística*. São Paulo: USP, 2005. p. 1-49. Disponível em: https://www.ime.usp.br/~rvicente/Guedes_etal_Estatistica_Descritiva.pdf. Acesso em: 27 set. 2022.

LARSON, R.; FARBER, E. *Estatística aplicada*. 4. ed. São Paulo: Pearson Prentice Hall, 2010.

SAMPAIO, N. A. S.; ASSUMPÇÃO, A. R. P.; FONSECA, B. B. *Estatística descritiva*. Belo Horizonte: Poisson, 2018.

SILVESTRE, A. L. *Análise de dados e estatística descritiva*. Forte da Casa: Escolar, 2007.



Fique atento

Os links para sites da web fornecidos neste capítulo foram todos testados, e seu funcionamento foi comprovado no momento da publicação do material. No entanto, a rede é extremamente dinâmica; suas páginas estão constantemente mudando de local e conteúdo. Assim, os editores declaram não ter qualquer responsabilidade sobre qualidade, precisão ou integridade das informações referidas em tais links.

Conteúdo:

sagah⁺



Dica do professor

A estatística descritiva tem como objetivo resumir os dados e descrevê-los de forma mais simples e clara. Para isso, o pesquisador precisa definir os atributos (variáveis) de interesse e quantificar suas características.

Nesta Dica do Professor, conheça a importância e os usos da estatística descritiva, observando sua utilização em várias áreas do conhecimento, com variáveis distintas (candidatos eleitorais, taxa de desemprego etc.).



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.



Exercícios

- 1) A probabilidade empírica é aquela baseada em observações aleatórias, sem necessariamente apresentar uma base teórica. Ela é utilizada quando se deseja observar com que frequência certos eventos podem ocorrer.

Considere uma empresa que fabrica três modelos diferentes de produtos (A, B e C), e uma amostra de 1.200 produtos vendidos em que tenham sido verificadas as seguintes quantidades de vendas por produto:

Produto A: 531

Produto B: 393

Produto C: 276

Com base nessa amostra, qual a probabilidade de a próxima venda ser do produto A?

- A) 79,37%.
- B) 66,90%.
- C) 55,75%.
- D) 44,25%.
- E) 32,75%.

- 2) A construção de uma distribuição de frequência sem intervalo de classe é equivalente à construção de uma tabela simples, onde se listam os diferentes valores observados da variável.

Com base nessas informações, analise as seguintes afirmativas sobre distribuição de frequências.

I. A frequência absoluta (F_i) é utilizada para comparar um dado em relação ao todo.

II. A frequência percentual ($f_i\%$) representa o percentual de observações pertencente à variável.

III. A frequência absoluta acumulada (F_{ai}) representa o percentual de vezes que uma frequência apareceu em relação ao todo.

Assinale a alternativa que contém a(s) assertiva(s) correta(s).

- A) Apenas I.
- B) Apenas II.
- C) I e II.
- D) I e III.
- E) II e III.

3) Por meio da inferência estatística, um investigador pode atingir um grau de segurança que lhe permite dizer com grande probabilidade de certeza se um valor é correspondente ao parâmetro populacional ou não.

Com base nos conceitos da estatística inferencial, analise as afirmativas a seguir:

I. A estimação pontual consiste no uso de dados da amostra para estimar valores de parâmetros populacionais desconhecidos.

II. A confiança demonstra ao pesquisador se o valor calculado é nulo.

III. O teste de hipóteses é utilizado quando se faz necessário comprovar uma possibilidade que aparenta ser verdadeira.

Assinale a alternativa que contém a(s) assertiva(s) correta(s).

- A) Apenas I.
- B) Apenas II.
- C) I e II.
- D) I e III.
- E) II e III.

4) Um intervalo de confiança para uma característica da população é um intervalo de valores plausíveis para esta característica.

Assinale (V) para verdadeiro e (F) para falso nas seguintes afirmativas sobre intervalo de confiança:

() O intervalo de confiança sinaliza a margem de certeza do cálculo realizado.

() O grau de confiança tem origem na probabilidade associada ao processo de construção do intervalo antes de se obter o resultado amostral.

() Um intervalo de confiança é um intervalo de valores utilizado para estimar o verdadeiro valor de parâmetro populacional.

() Calcular um intervalo de confiança é uma estratégia que leva em consideração o acerto da amostragem.

Assinale a alternativa que apresenta a sequência correta.

A) F - V - V - V.

B) F - V - V - F.

C) V - V - F - F.

D) V - F - V - V.

E) V - F - V - F.

5) A probabilidade permite o estudo da incerteza a partir de um fenômeno de caráter aleatório, constituindo a base da estatística inferencial. Existem basicamente três tipos de probabilidade: a clássica, a empírica e a subjetiva.

Com base nos tipos de probabilidade, analise as afirmativas a seguir:

I. A probabilidade clássica é usada quando cada resultado em um espaço amostral é igualmente possível de ocorrer.

II. A probabilidade empírica é usada quando um experimento é repetido várias vezes, o que, por sua vez, cria padrões regulares.

III. A probabilidade subjetiva é usada quando os dados são muito complexos, sendo seu cálculo feito a partir de um valor conhecido.

Assinale a alternativa que contém a(s) assertiva(s) correta(s).

A) Apenas I.

B) Apenas II.

C) I e II.

D) I e III.

E) II e III.



Na prática

Uma distribuição de frequência é um arranjo de valores que uma ou mais variáveis assumem em uma amostra. Cada entrada na tabela contém a frequência ou o número de ocorrências de valores dentro de um determinado intervalo e, portanto, a tabela resume a distribuição dos valores da amostra.

Neste Na Prática, acompanhe a elaboração de uma distribuição de frequência.

DISTRIBUIÇÃO DE FREQUÊNCIA NOS NEGÓCIOS

A construção de uma distribuição de frequência de um conjunto de dados deve seguir os passos descritos a seguir.

1. Definir o número de classes. Lembrando que não existe uma única fórmula para realizar esse processo, sendo importante considerar a intenção do analista de dados. A recomendação é que sejam definidas de 5 a 20 classes.
2. Identificar a amplitude total.
3. Encontrar os limites de classe inferior e superior para cada uma.

A seguir, acompanhe um caso prático para entender melhor o processo.

Contexto

Márcia é diretora comercial da empresa Aliança Cosméticos Ltda., responsável pela estratégia de vendas e pelo acompanhamento dos resultados obtidos pela equipe. Após o fechamento mensal, Márcia resume o resultado das vendas para apresentar na reunião da alta administração.



A fim de demonstrar os principais resultados, Márcia recorre ao uso de métodos estatísticos, optando pela **construção da tabela de frequência**.

Assim sendo, ela utiliza os dados da quantidade de vendas dos últimos 30 dias (variável aleatória da tabela de frequência).

Coleta e separação dos dados

Os dados obtidos foram os seguintes:

10.151	10.227	10.226	10.401	10.299
10.123	10.256	10.342	10.389	10.337
10.158	10.345	10.527	10.344	10.546
10.320	10.582	10.180	10.273	10.318
10.562	10.435	10.234	10.551	10.441
10.486	10.268	10.333	10.472	10.488

Depois de coletar os dados, Márcia os organizou em ordem crescente e identificou o limite inferior e o limite superior:

Limite inferior	←	10.123	10.342	
		10.151	10.344	
		10.158	10.345	
		10.180	10.389	
		10.226	10.401	
		10.227	10.435	
		10.234	10.441	
		10.256	10.472	
		10.268	10.486	
		10.273	10.488	
		10.299	10.527	
		10.318	10.546	
		10.320	10.551	
		10.333	10.562	
		10.337	10.582	→ Limite superior

Tabela de frequência

Márcia decidiu organizar sua tabela de frequência em 5 classes (k), passando a calcular a amplitude total (AT) e amplitude das classes (h).

$$AT = \text{Limite superior da última classe (-)} - \text{limite inferior da primeira classe}$$

$$AT = 10.582 - 10.123$$

$$AT = 459$$

$$h = AT / K$$

$$h = 459 / 5$$

$$h = 92$$

A tabela de frequência terá 5 colunas:

1. Quantidades de vendas = 5 classes com amplitude total de 12
2. Frequência absoluta (F) = número de observações encontradas na classe
3. Frequência percentual (f%) = $F_i \div \text{total de dados} \times 100$
4. Frequência absoluta acumulada (Fai) = $F_1 + F_2 + \dots + F_i$
5. Frequência acumulada relativa (f%_a) = $F_{ai} \div \text{total de dados} \times 100$

Quantidade de vendas	F _i	f%	F _a	f% _a
10.123 10.215 Classe 1	4	13,33% =(4/30)×100	4 =F ₁	13,33% =(4/30)×100
10.215 10.307 Classe 2	7	23,33% =(7/30)×100	11 =F ₁ + F ₂	36,67% =(11/30)×100
10.307 10.399 Classe 3	8	26,67% =(8/30)×100	19 =F ₁ + F ₂ + F ₃	36,67% =(19/30)×100
10.399 10.491 Classe 4	6	20,00% =(6/30)×100	25 =F ₁ + F ₂ + F ₃ + F ₄	36,67% =(25/30)×100
10.491 10.582 Classe 5	5	16,67% =(5/30)×100	30 =F ₁ + F ₂ + F ₃ + F ₄ + F ₅	100% =(30/30)×100
TOTAL	30	100%		

Análise dos dados

Com a construção, a tabela de frequência torna mais fácil apresentar e analisar os dados. Sendo assim, Márcia tirou as conclusões a seguir.

- ▶ O mínimo de vendas é de 10.123 itens diários.
- ▶ O máximo de vendas é de 10.582 itens diários.
- ▶ A amplitude variável é 459 itens diários.
- ▶ 50% das vendas diárias ficam entre 10.215 e 10.399 itens.



Conclusão

Com as análises apresentadas, foram estabelecidas novas estratégias e metas de vendas.

A distribuição de frequência é uma série estatística que organiza os dados e por meio da qual é possível extrair informações sobre o comportamento desses dados. A organização dos dados evidencia diversos aspectos do fenômeno estudado e contribui para tomada de decisões.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.



Saiba mais

Para ampliar o seu conhecimento a respeito desse assunto, veja abaixo as sugestões do professor:

Métodos estatísticos para as ciências sociais

Por meio da estatística descritiva é possível descrever os dados, tanto utilizando a distribuição da frequência absoluta, da frequência relativa e do ponto médio quanto outros métodos. Além de resumir os dados, a estatística descritiva permite sua apresentação, por meio de tabelas e gráficos. Leia o Capítulo 3, Estatística descritiva, e confira a descrição gráfica de alguns dados.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

Distribuição amostral

A distribuição amostral é a distribuição de probabilidades de uma medida estatística baseada em uma amostra aleatória. Sua importância reside na inferência estatística, tendo em vista a simplificação. Para conhecer mais sobre este conceito e sobre seu uso, assista ao vídeo a seguir.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

O que é intervalo de confiança?

O intervalo de confiança é importante para indicar a margem de incerteza frente a um cálculo efetuado. Ele deve ser utilizado para descrever achados importantes de uma pesquisa. Para compreender sua aplicação, assista ao vídeo a seguir.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.