Curso: Marketing
Professor: Ms. José João de Melo (JOTA)

ESTATÍSTICA

No mundo atual, a empresa é uma das sustentações da Economia dos Povos. A direção de uma empresa exige de seu administrador a importante tarefa de tomar decisões, e o conhecimento e o uso da Estatística facilitarão o trabalho de organizar, dirigir e controlar a empresa.

Através da coleta de dados e do recenseamento de opiniões podemos conhecer a realidade geográfica e social, os recursos naturais, humanos e financeiros disponíveis, as expectativas da comunidade em relação à empresa, e a partir disso estabelecer suas metas com maiores possibilidades de serem alcançadas.

Nas múltiplas atividades de hoje, o homem lança mão de técnicas e processos estatísticos que só com o estudo desta disciplina pode-se evitar o erro das generalizações apressadas a respeito de tabelas e gráficos apresentados em jornais, revistas e televisão, erros estes cometidos frequentemente por quem conhece de forma superficial a Estatística.

A **Estatística** é o ramo da Matemática Aplicada desenvolvida para a coleta, a classificação, a apresentação, a análise e a interpretação de dados quantitativos e a utilização desses dados para a tomada de decisões. A estatística possui duas funções ou campos bem amplos. A primeira função é descritiva e a segunda é indutiva.

A **Estatística Descritiva** pode ser interpretada como uma função cujo objetivo é a observação de fenômenos de mesma natureza, a coleta de dados numéricos referentes a esses fenômenos, a organização e a classificação desses dados observados e a sua apresentação através de gráficos e tabelas, além do cálculo de coeficientes que permitem descrever resumidamente os fenômenos estudados.

A **Estatística Indutiva** ou inferência estatística refere-se a um processo de generalização, a partir de resultados particulares. Consiste em obter e generalizar conclusões para um todo com base em uma parte estudada. O processo de generalização está associado a uma margem de incerteza, em razão desta margem se torna importante o estudo de Probabilidade.

População e amostra

Dois conceitos muito utilizados em Estatística são população ou universo estatístico e amostra.

População é o conjunto da totalidade dos indivíduos sobre o qual se faz uma inferência e **amostra** é uma parte selecionada da totalidade de observações abrangidas pela população, através da qual se faz um juízo ou inferência sobre as características da população. Mas, para as inferências serem corretas, é necessário garantir que a amostra seja representativa da população, isto é, a amostra deve possuir as mesmas características básicas da população, no que diz respeito ao fenômeno que desejamos pesquisar. Para que a amostra seja representativa é necessário o uso de algumas técnicas de amostragem, descrevemos a seguir três das principais técnicas de amostragem:

- Amostragem casual ou aleatória simples: Esse tipo de amostragem é equivalente a um sorteio lotérico. Na prática a amostragem aleatória simples pode ser realizada numerando-se a população de 1 a n e sorteando-se, a seguir, por meio de um dispositivo aleatório qualquer.

- **Amostragem proporcional estratificada:** Muitas vezes a população se divide em subpopulações, denominados **estratos**. Como é provável que a variável em estudo apresente, de estrato em estrato, um comportamento heterogêneo e, dentro de cada estrato, um comportamento homogêneo, convém que o sorteio dos elementos da amostra leve em consideração tais estratos.

- **Amostragem sistemática:** Quando os elementos da população já se acham ordenados, não há necessidade de construir os sistemas de referência. São exemplos os prontuários de vendas de um produto, as casa de uma rua, as linhas de produção etc. Nesses casos, a seleção dos elementos que constituirão a amostra pode ser feita por um sistema imposto pelo pesquisador. A esse tipo de amostragem denominamos de amostragem sistemática.

Variáveis

Num estudo estatístico o observador poderá anotar ou medir a intensidade efetiva de um caráter variável em cada um dos objetos ou pessoas observadas. Pode, por exemplo, registrar a idade das pessoas ao morrer, a estatura, o peso, a quantidade de produtos vendidos em uma empresa etc. Esse caráter variável pode ser chamado simplesmente de variáveis.

As variáveis podem ser **qualitativas** ou **quantitativas.** As variáveis qualitativas se referem a um atributo, normalmente, não numérico da população, pode ser, por exemplo: profissão, sexo, raça etc. Enquanto que as variáveis quantitativas apresentam como possível resultado números resultantes de uma contagem ou mensuração, pode ser, por exemplo: número de funcionários de uma empresa, altura dos jogadores de um time de futebol, renda etc.

As variáveis qualitativas podem ser:

- nominal: quando não existe ordenação dos possíveis resultados. Exemplos: sexo, profissão etc.

- ordinal: quando existe certa ordem nos possíveis resultados. Exemplos: escolaridade, nível sócio econômico etc.

As variáveis quantitativas podem ser:

- discretas: quando o resultado é expresso por um número inteiro, o qual é obtido mediante a contagem. Exemplos: número de peças produzidas por uma máquina, número de pessoas da família.

- contínuas: quando o resultado é um número real, que em geral é resultante de uma mensuração. Exemplos: massa, altura, idade etc.

Exercícios

1) A direção de um parque contratou uma equipe de pesquisadores para coletar algumas informações sobre os seus freqüentadores. Os cem entrevistados responderam às seguintes questões: sexo, idade, número de vezes por semana que vão ao parque, período da visita (manhã, tarde, começo da noite), tempo de permanência e quantia gasta nas dependências do parque. Classifique cada uma das variáveis desse estudo em qualitativa ou quantitativa discreta ou quantitativa continua.

2) Classifique as variáveis em qualitativas, quantitativas discretas ou quantitativas contínuas.

a) cor dos cabelos

b) número de filhos

c) número de livros de uma biblioteca

Fases do método estatístico

Existem diversas etapas ou operações que devem ser desenvolvidas para se chegar aos resultados finais de um estudo estatístico completo. Essas etapas são denominadas fases do trabalho estatístico e são do âmbito da Estatística Descritiva. As principais fases do método estatístico são as seguintes:

- Definição do problema: saber exatamente aquilo que se pretende pesquisar o que corresponde a definir corretamente o problema.

- Planejamento: consiste em determinar o procedimento necessário para resolver o problema e, em especial, como levantar as informações sobre o assunto objeto do estudo.

- Coleta de dados: consiste na obtenção, reunião e registro sistemático de dados, com um objetivo determinado.

- Apuração dos dados: consiste em resumir os dados através de sua contagem e agrupamento.

- Apresentação dos dados: essa fase pode ser tabular (apresentação numérica dos dados) ou gráfica (apresentação geométrica)

- Análise e interpretação dos dados: essa fase é a mais importante e também a mais delicada. Essa etapa consiste em tirar conclusões que auxiliem o pesquisador a resolver seu problema. A análise dos dados estatísticos está ligada ao cálculo de medidas, cuja finalidade principal é descrever o fenômeno. Sendo assim, o conjunto de dados a ser analisado pode ser expresso por números resumos que evidenciam as características particulares desse conjunto.

Arredondamento de dados

A estatística trabalha com números e muitas vezes esses números não são inteiros, em função disso é necessário o arredondamento dos mesmos. Esse arredondamento deve obedecer alguns critérios recomendados pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

Esses critérios ou regras são:

1) Arredondamento por falta: Quando o primeiro dígito, aquele situado mais à esquerda entre os que irão ser eliminados, for menor que cinco não deverá ser alterado o dígito remanescente.

2) Arredondamento por excesso: Quando o primeiro dígito após aquele que será arredondado for maior que cinco ou igual a cinco seguido por dígitos maiores que zero, o dígito remanescente será acrescido de uma unidade.

3) Arredondamento de dígitos seguidos do cinco: Quando o dígito mais à esquerda dos que serão eliminados for um cinco ou um cinco seguido somente de zeros, o último dígito remanescente, se for par, não se alterará, se for ímpar será aumentado de uma unidade.

Exemplos

1) Arredonde cada um dos números abaixo, conforme a precisão pedida:

1.1) Para o décimo mais próximo:

a) 20,32 =	b) 25, 7384=
c) 45,19=	d) 38, 650000001 =
e) 7,45 =	f) 120, 4500 =
g) 29, 99 =	h) 499, 9701 =

1.2) Para o centésimo mais próximo:

a) 42,723 =	b) 123,8549 =
c) 129,65 =	d) 2,9519 =
e) 38,995 =	f) 34, 465 =

1.3) Para a unidade mais próxima:

a) 10,16 =	b) 98,981 =
c) 69,65 =	d) 37,29 =
e) 39,5 =	f) 44,59 =

Compensação

Em muitos casos, mesmo fazendo corretamente o arredondamento, ocorre uma pequena diferença para a apresentação dos resultados e, se torna necessário que tal diferença desapareça, o que é possível pela prática que denominamos de compensação. Praticamente usamos descarregar a diferença na(s) maior(es) parcela(s). Convém, ainda, observar que, se a maior parcela é igual ao dobro de qualquer outra parcela (ou maior que esse dobro), descarregamos a diferença (maior que uma unidade) apenas na maior parcela.

Exemplos:

1) Arredonde para o centésimo mais próximo e compense, se necessário:

0,060 + 0,119 + 0,233 + 0,313 + 0,164 + 0,091 + 0,030 = 1,000.

2) Arredonde para a unidade mais próxima e compense, se necessário:

4,0 + 7,6 + 12,4 + 27,4 + 11,4 + 8,0 = 70,8.

Exercícios

1) Arredonde cada um dos números abaixo, conforme a precisão pedida:
1.1) Para o décimo mais próximo:

a) 23,41 =	b) 234, 784=
c) 45,09 =	d) 38, 850002 =
e) 78,85 =	f) 120, 4500 =
g) 129, 98 =	h) 199, 97 =

1.2) Para o centésimo mais próximo:

a) 46,727 =	b) 123,842 =
c) 123,65 =	d) 299,951 =
e) 38,255 =	f) 37, 485 =

1.3) Para a unidade mais próxima:

a) 16,6 =	b) 49,98 =
c) 67,65 =	d) 68,2 =
e) 129,5 =	f) 49,49 =

2) Arredonde para a unidade mais próxima e compense, se necessário:
33,4 %+ 23,4% + 20,0% + 3,2% + 11,1% + 8,9% = 100%.

Proporção, razão e porcentagem

Apesar de as noções de razão, proporção e porcentagem estarem em texto de Matemática ou Matemática Financeira, será útil em estatística discutirmos estes termos.

Sejam dois números reais a e b, com b ≠ 0. Chama-se **razão** entre a e b o quociente a : b ou $\frac{a}{b}$.

A igualdade entre duas razões recebe o nome de **proporção**. A proporção de indivíduos em uma dada categoria é definida através do quociente entre o número de indivíduos pertencentes a essa categoria e o número total de indivíduos considerados, devendo as categorias ser mutuamente exclusivas.

A razão cujo denominador é 100 recebe o nome de **porcentagem.** Devido a sua importância elas costumam ser representadas por um símbolo especial (%), que substitui o denominador 100. Nesse caso, as razões centesimais também recebem o nome de taxa de porcentagem.

Do ponto de vista estatístico a razão, a proporção e a porcentagem podem ser consideradas como medidas que permitem estabelecer comparações entre os diversos grupos.

Exemplos

1) Uma máquina fabrica 900 produtos por dia, dos quais desejamos ter uma amostra de 5,5 % desses produtos. Que procedimentos devemos usar para a escolha desses produtos baseado numa amostragem sistemática?

2) Uma empresa possui 2500 clientes em 8 cidades distintas: A, B, C, D, E, F, G e H. Sendo 350 na cidade A, 320 na B, 300 na C, 290 na D, 340 na E, 320 na F, 310 na G e 270 na H. Se essa empresa pretende fazer uma pesquisa sobre a satisfação de seus clientes, então:
a) qual o tipo de amostragem mais indicado?
b) obtenha uma amostra de 200 clientes.

3) Uma amostra de 45 elementos será retirada de uma população ordenada formada por 2430 indivíduos. Na ordenação geral, qual dos elementos seguintes seria escolhido para pertencer à amostra, sabendo-se que o elemento de ordem 1350 a ela pertence?

a) 725º b) 1000º c) 1638º d) 1620 e) 1700

Exercícios

1) O diretor de uma empresa na qual trabalham 280 homens e 320 mulheres deseja conhecer as condições de vida extra-empresa de seus funcionários e não dispondo de tempo para entrevistar todas as famílias, resolveu fazer um levantamento com 5% de amostragem. Obtenha para esse diretor, o número de pessoas de cada sexo dessa amostra.

2) Numa cidade com 156000 eleitores, há três candidatos a prefeito. Foi feito um levantamento com amostragem de 4% da população e constatou-se que o candidato A tem 30% da preferência do eleitorado, o candidato B tem 40% e o candidato C tem 10% da preferência do eleitorado. Calcule:
a) quantas pessoas estavam na amostra feita?
b) quantos eleitores disseram votar em A? Em B? Em C?

3) Uma população encontra-se dividida em três estratos, com tamanhos, respectivamente, $e_1 = 40$, $e_2 = 100$ e $e_3 = 60$. Sabendo-se que, ao ser realizada uma amostragem proporcional estratificada, nove elementos da amostra pertenciam ao estrato e_3, determine o número total de elementos da amostra.

4) Numa pesquisa com 1500 moradores de uma cidade, constatou-se que 4,5% não têm conta em banco. Responda:
a) que quantidade de pessoas corresponde a essa porcentagem?
b) quantas pessoas dessa amostra da cidade têm conta em banco?

5) Uma amostra de 32 elementos será retirada de uma população ordenada formada por 2432 indivíduos. Na ordenação geral, qual dos elementos seguintes seria escolhido para pertencer à amostra, sabendo-se que o elemento de ordem 1420 a ela pertence?
a) 725º b) 1120º c) 1648º d) 1701 e) 1700

Somatório

A Matemática fornece ainda outra noção de grande utilidade para a estatística: o somatório. Este operador facilita a indicação, a formulação de medidas e algumas operações algébricas desenvolvidas pela Estatística. Por esse motivo vamos estudar este assunto como um apêndice à Estatística.

Utiliza-se a letra grega sigma maiúsculo: $\sum$. Por exemplo, a soma dos números do conjunto X = {2, 4, 7, 9, 10, 15} pode ser indicada por: $\sum_{i=1}^{6} x_i = 2+4+7+9+10+15$. A Expressão deve ser lida assim: somatório de x_i, com i variando de 1 a 6.

Operar com os somatórios requer o conhecimento de algumas propriedades, regras e definições. Aqui chamaremos simplesmente de propriedades:

P_1) O somatório de uma constante é igual ao produto do número de termos pela constante.

P_2) O somatório de um produto de uma constante por uma variável que depende do somatório é igual ao produto da constante pelo somatório da variável.

P_3) Vale para o somatório a propriedade distributiva em relação à adição algébrica.

P_4) O quadrado da soma é diferente da soma dos quadrados.

P_5) O produto de duas somas é diferente da soma dos produtos.

DISTRIBUIÇÃO DE FREQÜÊNCIA

- Dados não agrupados em classes

Após a coleta, os dados originais ainda não se encontram pronto para a análise, por não estarem numericamente organizados. Esses dados são chamados de **dados brutos**. Precisamos, então começar a organizar esses dados obedecendo a uma determinada ordem. Essa primeira ordenação recebe o nome de rol.

Rol é uma lista em que os valores estão dispostos em uma determinada ordem, crescente ou decrescente. Essa primeira organização dos dados proporciona algumas vantagens em relação à forma original, poderemos notar, por exemplo, que existem valores que se repetem. Essa repetição sugere que se condensem todos os resultados em uma tabela de freqüência, estabelecendo a correspondência entre o valor individual e o número de vezes que ele foi observado.

O número de observações recebe o nome de freqüência. A tabela de freqüência permite a apresentação esteticamente mais vantajosa dos dados, facilitando ainda a verificação do comportamento do fenômeno. Além disso, é possível a obtenção de estatísticas com menos cálculo.

As tabelas de valores podem ser construídas com valores não agrupados em classe, ou seja, uma tabela onde os valores de cada variável aparecem individualmente. Esse tipo de apresentação é utilizado para representar uma variável discreta.

Muitas vezes, mesmo com o risco de se perder algum detalhe, há vantagem em resumir os dados em uma distribuição de freqüência com dados agrupados em classes. Esse tipo de agrupamento será conveniente principalmente quando a variável do estudo for contínua.

As tabelas de distribuição de freqüência apresentam alguns elementos e termos próprios que é necessário conhecer. Destacamos estes termos a seguir:

- Freqüência absoluta simples: É o número de observações correspondente a essa classe ou a esse valor. A freqüência absoluta simples é simbolizada por f_j.

- Amplitude total: É a diferença entre o maior e o menor valor observado da variável em estudo. A amplitude total é indicada por A_t.

- Classe: É cada um dos valores em que se subdivide a amplitude total do conjunto de valores observados da variável.

- Limites de classes: São seus valores extremos.

- Freqüência simples relativa: Representa a proporção de observações de um valor individual ou de uma classe, em relação ao número total de observações. Para calcular a freqüência relativa basta dividir a freqüência absoluta da classe pelo número total de observações. A freqüência relativa é indicada por fr_j.

- Freqüências relativas percentuais: É a freqüência que se obtém multiplicando-se a freqüência relativa simples por 100. Indicamos as freqüências relativas percentuais por fr_j (%). Observa-se, ainda, que a soma das freqüências relativas percentuais deve ser igual a 100%.

- Freqüência absoluta acumulada: É a soma da freqüência simples absoluta dessa classe ou desse valor com as freqüências simples absolutas das classes ou dos valores anteriores. Sempre que queremos saber quantos valores ocorrem até uma determinada classe recorre-se à freqüência acumulada. Indicamos a freqüência acumulada por $F_{j.}$

- Freqüência relativa acumulada: É a soma da freqüência simples relativa dessa classe ou desse valor com as freqüências relativas simples das classes ou dos valores anteriores. Indicamos a freqüência relativa acumulada por Fr_j.

- Freqüência relativa acumulada percentual: É a freqüência que se obtém multiplicando-se a freqüência relativa acumulada por 100. Indicamos a freqüência relativa acumulada por Fr_j (%).

Exemplos

1) Uma empresa fabricante de instrumentos de precisão está interessada em saber o número de aparelhos defeituosos rejeitados pelo controle de qualidade. As estatísticas, fornecidas por essa seção, referem-se ao período de 2001 a 2004.

Número mensal de aparelhos defeituosos

Ano/mês	J	F	M	A	M	J	J	A	S	O	N	D
2001	6	2	5	6	0	8	7	6	3	4	5	8
2002	10	9	7	6	3	4	6	4	5	4	0	1
2003	3	6	7	9	3	1	4	6	5	3	5	4
2004	7	2	5	8	6	4	2	5	1	6	5	2

Construa uma tabela de freqüências de valores não agrupados.

2) A tabela abaixo se refere a uma pesquisa, realizada com 200 sócios de um clube, a respeito do esporte preferido:

Esporte	f_j	fr_j	fr_j (%)
Futebol	108		
Vôlei		0,210	
Basquete			
Natação	12		
Outros			8,5
Total			

Complete os espaços da tabela.

Exercícios

1) Considere a seguinte distribuição correspondente aos diferentes preços de um determinado produto em vinte lojas pesquisadas.

Preços (R$)	Número de lojas (fj)
10	2
11	5
12	6
13	6
14	1
Total	20

a) Quantas lojas apresentaram um preço de R$ 12,00?
b) Construa uma tabela de freqüências simples relativas, freqüências absolutas acumuladas, freqüências simples relativas percentuais e freqüência relativas acumuladas percentuais.
c) Quantas lojas apresentaram um preço de até R$ 12,00?
d) Qual a porcentagem de lojas com preço maior que R$ 11,00 e menor que R$14,00?

2) Um produto é vendido por apenas três empresas, em determinado ano, para um total de 180.000 unidades vendidas, a empresa A vendeu 72.000, a empresa B vendeu 48.000 e a empresa C 60.000. Determine a distribuição percentual das vendas desse produto.

3) A distribuição abaixo indica o número de acidentes ocorridos com 70 motoristas de uma empresa de ônibus.

Nº de acidentes	0	1	2	3	4	5	6	7
Nº de motoristas	20	10	16	9	6	5	3	1

Determine:

a) o número de motoristas que não sofreram nenhum acidente.
b) o número de motoristas que sofreram pelo menos 4 acidentes.
c) o número de motoristas que sofreram menos de 3 acidentes.
d) o número de motoristas que sofreram no mínimo 3 e no máximo 5 acidentes.
e) a porcentagem dos motoristas que sofreram no máximo 2 acidentes.

4) Com o objetivo de obter um perfil de seu público nos finais de semana, o proprietário de um teatro contratou dois pesquisadores para coletar dados referentes a sua clientela. Os pesquisadores escolheram seis objetos de estudo: sexo, idade, nível de escolaridade, estado civil, renda mensal e meio de transporte para chegar ao teatro. Num final de semana foram entrevistados 20 freqüentadores desse teatro. Os resultados estão na tabela seguinte:

Sexo	Idade	Nível de escolaridade*	Estado civil	Transporte	Renda mensal (em salários mínimos)
Masculino	28	EM	Casado	Carro	11,8
Masculino	38	EM	Casado	Carro	13,9
Feminino	24	ES	Solteira	Carro	12,4
Masculino	43	EM	Casado	Carro	19,5
Feminino	32	ES	Separada	Ônibus	12,1
Feminino	19	EM	Solteira	A pé	5,0
Masculino	22	ES	Solteiro	Ônibus	8,9
Masculino	25	EM	Solteiro	Ônibus	13,3
Masculino	41	ES	Casado	A pé	14,7
Feminino	40	EF	Solteira	Carro	16,6
Feminino	35	ES	Solteira	Carro	9,3
Masculino	29	EF	Casado	Carro	11,6
Masculino	31	EF	Separado	Carro	10,2
Feminino	36	ES	Solteira	Carro	16,0
Feminino	48	EM	Casada	Carro	18,8
Feminino	23	EM	Casada	A pé	15,4
Masculino	27	ES	Solteiro	A pé	10,7
Masculino	26	ES	Solteiro	Ônibus	8,2
Masculino	29	ES	Separado	Ônibus	12,5
Masculino	30	EF	Casado	Carro	7,6

* EF: ensino fundamental completo; EM: ensino médio completo; ES: ensino superior completo.

a) Construa uma tabela de freqüências para a variável sexo. (f_j, fr_j(%).
b) Construa uma tabela de freqüências para a variável nível de escolaridade.(f_j, F_j, fr_j(%), Fr_j(%)).
c) Construa uma tabela de freqüências para a variável estado civil.(f_j, fr_j(%)).
d) Construa uma tabela de freqüências para a variável transporte.(f_j, fr_j(%)).

- Dados agrupados em classes

Para a distribuição de freqüência com dados agrupados é importante que se tenha um número adequado de classes. Para determinar o número de classes há diversos métodos. A regra de Sturges estabelece que o número de classes seja calculado da seguinte maneira: k = 1 + 3,3.log n, onde k é o número de classes e n é o número total de observações.

O inconveniente da regra de Sturges é propor um número demasiado de classes para um número pequeno de observações e poucas classes para um número de observações grandes. Turman L. Kelley sugere os seguintes números de classes com base no número total de observações para efeito de representação gráfica:

N	5	10	25	50	100	200	500	100
k	2	4	6	8	10	12	15	15

Apesar de termos esses e outros critérios para a determinação do número de classes, o analista deverá ter em mente que a escolha dependerá da natureza dos dados e da unidade de medida que eles forem expressos.

Para a construção de uma tabela de freqüência é importante, tanto quanto possível, que seus pontos médios coincidam com a concentração dos valores reais.

Para a construção de uma tabela de freqüências com dados agrupados em classe é conveniente adotar-se um roteiro. O roteiro proposto consta das seguintes etapas:

I - Listar os dados brutos (transformados ou não em rol)

II - Encontrar a amplitude total do conjunto de valores observados (A_t)

III – Escolher o número de classes (k).

IV – Determinar a amplitude do intervalo de classe. (Sugere-se que a amplitude do intervalo seja A_t/k, porém convém arredondar adequadamente o número correspondente à divisão).

V – Determinar os limites das classes, escolhendo-se, preferencialmente, números inteiros.

VI – Construir a tabela de freqüências.

Exemplos

1) Os dados seguintes representam 20 observações relativas à temperatura, em ^{0}C, de determinados estados do Brasil.

14	22	29	10	10	21	27	16	34	15
29	20	28	16	12	11	11	20	13	28

a) Determinar o número de classes pela regra de Sturges.

b) Construir a tabela de freqüência absoluta simples, determinar as freqüências absolutas acumuladas, as freqüências relativas simples, as freqüências relativas percentuais, as freqüências relativas acumuladas e as freqüências relativas acumuladas percentuais.

2) A tabela abaixo apresenta os salários pagos a 500 funcionários de uma empresa.

N^o de salários mínimos	N^o de operários (f_j)	fr_j(%)	F_j	Fr_j (%)
0 ├ 2	200			
2 ├ 4	150			
4 ├ 6	50			
6 ├ 8	75			
8 ├ 10	25			
Total	500			

a) Complete essa tabela com as respectivas freqüências.

b) Quantos operários ganham menos de 2 salários mínimos?

c) Quantos operários ganham até 6 salários mínimos exclusive?

d) Qual a porcentagem de operários com salário maior ou igual a 6 e menor que 8 salários mínimos?

e) Qual a porcentagem de operários com salário inferior a 4 salários mínimos?

Exercícios

1) No quadro a seguir estão registradas as massas, em quilogramas, de 50 pessoas que freqüentam uma academia de ginástica.

72	81	57	64	87	90	69	77	73	74
80	96	55	88	92	47	60	68	80	58
77	76	59	83	81	90	68	65	74	57
91	97	86	82	73	64	69	71	88	94
77	72	81	49	75	52	50	63	70	91

Elabore um quadro de distribuição de freqüências absolutas e relativas percentuais. Se for conveniente, utilize a regra de Sturges para a escolha do número de classes.

2) Com os dados do exercício 4 da secção anterior, construa uma tabela de freqüências para as variáveis:
a) Idade. (f_j, F_j, fr_j(%), Fr_j(%)).
b) Renda mensal.(f_j, F_j, fr_j(%), Fr_j(%)).

REPRESENTAÇÃO GRÁFICA

O gráfico estatístico é uma forma de apresentação dos dados estatísticos, cujo objetivo é o de produzir, no investigador ou no público, uma impressão mais rápida do fenômeno em estudo. Sendo assim a representação gráfica se torna um importante complemento da representação tabular. Os gráficos permitem a visualização imediata dos valores observados, além de proporcionarem uma idéia preliminar mais satisfatória da concentração ou da dispersão dos valores.

Os títulos dos gráficos devem ser escritos em letras maiúsculas e responderem às perguntas: O que? Quando? Onde?.

Existem diversos tipos de gráficos, os principais destacamos a seguir.

- Gráfico em barras (horizontais)

Os gráficos em barras têm por finalidade comparar grandezas por meio de retângulos de mesma largura e comprimentos proporcionais às grandezas envolvidas. Para a construção destes gráficos devemos observar algumas orientações:
1ª) As barras devem ser desenhadas observando sua ordem de grandeza para facilitar a leitura e a análise dos valores.
2ª) Normalmente, a ordem é decrescente, a barra superior representa o maior valor. Categorias gerais que costumam vir com inscrições do tipo “outros”, “demais” etc deverão ser representadas na barra inferior, mesmo que o seu comprimento exceda o de alguma outra.
3ª) O gráfico deverá conter uma linha zero claramente definida e uma escala de quantidades ininterrupta, caso contrário a leitura e a interpretação do gráfico poderão ficar distorcidas.

- Gráfico em colunas (verticais)

Os gráficos em colunas têm a mesma finalidade dos gráficos em barras. Utilizamos os gráficos em colunas quando as legendas forem breves, caso contrário, é preferível o gráfico em barras. Os gráficos em colunas prestam-se em especial à representação e análise de dados relacionados com séries temporais, como, por exemplo, as vendas de um produto em períodos sucessivos. Sendo assim, as colunas devem ficar dispostas em ordem cronológica.

Muitas vezes, costuma-se colocar no topo ou no interior de cada coluna o valor correspondente à sua altura. Esse procedimento permite eliminar a escala.

- Gráfico em linhas

Os gráficos em linhas são usados para a representação de séries temporais quando estas séries cobrirem um grande número de períodos de tempo, nestes casos as representações dos valores através de colunas podem conduzir a uma excessiva concentração de dados.

Os gráficos em linhas também são bastante utilizados na identificação de tendências de aumento ou diminuição dos valores numéricos de uma dada informação. Assim, usamos com freqüência esse tipo de representação em análises tais como lucros de empresas, incidência de moléstias, índices de crescimento populacional, taxa de mortalidade infantil, índices de custo de vida etc.

Os gráficos em linhas facilitam a comparação entre duas ou mais grandezas e por esse motivo são usados quando temos o objetivo de comparar o comportamento de duas ou mais variáveis em estudo, como por exemplo, os valores de importação e exportação de uma empresa em um determinado período.

- Gráficos pictóricos

Os gráficos pictóricos ou pictogramas são construídos a partir de figuras representativas da intensidade ou da modalidade do fenômeno em estudo. Esses gráficos são freqüentes em jornais e revistas, tendo como principal vantagem o fato de despertarem a atenção do público leigo. Existem algumas regras para a construção desses gráficos:

1ª) Os símbolos devem ser auto-explicativos.

2ª) As diferentes quantidades devem expressar-se mediante maior ou menor número de símbolos, e não mediante um aumento ou diminuição do tamanho do símbolo.

3ª) Os gráficos devem proporcionar uma visão geral do fenômeno, e não privilegiar detalhes minuciosos.

4ª) Os pictogramas estabelecem comparações gerais, devendo ser evitados, conseqüentemente, para interpretar afirmações ou dados isolados.

- Gráficos em setores (pizza)

Os gráficos em setores são usados para representar valores absolutos ou porcentagens. Recomenda-se que o gráfico em setores seja empregado quando há no máximo sete dados, se fizermos, por exemplo, um levantamento sobre as vendas de uma empresa nas cinco regiões do Brasil recomenda-se o uso do gráfico em setores, já se fizermos um levantamento das vendas desta empresa nos diversos estados brasileiros não se recomenda o uso desse tipo de gráfico, pois o círculo fica dividido em muitas partes dificultando a interpretação e a comunicação ao leitor.

- Histogramas

O histograma é um gráfico usado para representar dados agrupados em intervalos de classe. Este gráfico é formado por um conjunto de retângulos justapostos, de forma que a área de cada retângulo seja diretamente proporcional à freqüência da classe que ele representa. Assim, a soma dos valores correspondentes às áreas dos retângulos será igual a freqüência total.

O histograma é construído tomando-se como referência dois eixos coordenados. No eixo das abscissas são anotados os valores da variável em estudo (limites das classes) e no eixo das ordenadas os valores relativos ao número de observações das classes (freqüências).

- Polígonos de freqüências

O polígono de freqüência é obtido unindo por linhas retas os pontos médios das bases superiores dos retângulos do histograma. Para realmente obtermos um polígono (linha fechada), devemos completar a figura, ligando os extremos da linha obtida aos pontos médios da classe anterior à primeira e da posterior à última, da distribuição.

- Polígonos de freqüências acumuladas: Ogiva de Galton

A Ogiva de Galton ou polígono de freqüências acumuladas tem por finalidade a representação gráfica das tabelas de freqüências acumuladas. Esse gráfico é obtido unindo-se os limites inferiores e superiores das classes das freqüências acumuladas (F_j).

<u>Exemplo</u>

1) Verifique quais gráficos são recomendados nos exemplos de distribuições de freqüências apresentados anteriormente.

<u>Exercício</u>

1) Verifique quais gráficos são recomendados nos exercícios de distribuições de freqüências apresentados anteriormente. Construa esses gráficos no Excel.

<u>MEDIDAS DE POSIÇÃO</u>

Por meio da distribuição de freqüência se estabelece um sistema de classificação que descreve o padrão de variação de um determinado fenômeno estatístico, porém na maioria das vezes se torna difícil trabalhar com a distribuição de freqüência completa, razão pela qual costuma-se lançar mão de determinadas medidas que resumem características importantes da distribuição. Há diversas medidas que condensam as informações, vamos concentrar nosso estudo nos dois tipos mais importantes: as medidas de posição, especialmente as de tendência central e as medidas de dispersão ou de variabilidade.

A moda, a média aritmética e a mediana são as três medidas de posição mais usadas para resumir o conjunto de valores representativos do fenômeno que se deseja estudar em função da tendência de os dados observados se agruparem em torno delas.

- Média aritmética

A medida de tendência central mais usada para descrever uma distribuição de freqüência é a média aritmética que é representada pelo símbolo: $\bar{x}$. A média aritmética pode ser simples ou ponderada.

A média aritmética simples é o quociente entre a soma de todos os valores observados e o número total de observações. Assim, se um conjunto X é formado pelos elementos x_1, x_2, x_3, ..., x_n , a média aritmética dos elementos desse conjunto será:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}.$$

A média aritmética é considerada ponderada quando os valores do conjunto tiverem pesos diferentes. Obtém-se a média aritmética ponderada através do quociente entre o produto dos valores da variável pelos respectivos pesos e a soma dos pesos. Genericamente, se os valores do conjunto x_1, x_2, x_3, ..., x_k ocorrem f_1, f_2, f_3, ..., f_k vezes, respectivamente, a média aritmética do conjunto será:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = \frac{x_1 . f_1 + x_2 . f_2 + x_3 . f_3 + \ldots + x_k . f_k}{n}$$

onde n é o numero total de observações, k é o número de classes ou de valores individuais diferentes da variável.

- Propriedades da média aritmética

Com o recurso de algumas propriedades da média aritmética é possível desenvolver um processo breve de cálculo da média, menos direto, mas que proporciona o mesmo resultado. As principais propriedades da média são:

P1) A soma algébrica dos desvios (diferença entre cada valor do conjunto e a média) de um conjunto tomados em relação a média aritmética é zero.

P2) Somando-se (ou subtraindo-se) um valor constante e arbitrário a cada um dos elementos de um conjunto de números, a média aritmética fica somada (ou subtraída) por essa constante.

P3) Multiplicando-se (ou dividindo-se) cada elemento de um conjunto de números por um valor constante e arbitrário, a média fica multiplicada (ou dividida) por essa constante.

Exemplos

1) Em um escritório de contabilidade trabalham 5 contínuos que recebem salários mensais de R$ 350,00, R$ 370,00, R$ 380,00, R$ 340,00 e R$ 400,00. Qual a média aritmética mensal dos salários desses contínuos?

2) A distribuição dos salários de uma empresa é dada na seguinte tabela:

SALÁRIO EM R$	NÚMERO DE FUNCIONÁRIOS
500	10
1000	5
1500	1
2000	10
5000	4
10500	1
TOTAL	31

Qual é a média dos salários dessa empresa?

3) A nota final dos alunos de uma escola é composta pelas provas p_1, p_2 e p_3 com pesos, respectivamente, 2, 4 e 4. Se um aluno tirou 8 na p_1, 4 na p_2 e 4,5 na p_3, qual a sua média final?

4) A tabela abaixo apresenta os salários pagos a 100 funcionários de uma empresa.

N° de salários mínimos		N° de operários (f_j)
0 ├ 2		40
2 ├ 4		30
4 ├ 6		10
6 ├ 8		15
8 ├ 10		5
Total		100

Determine o salário médio.

5) Um comerciante mistura 4 kg de café do tipo A, que custa R$ 6,00 o kg; 10 kg do café B, que custa R$ 5,60 o kg; e 6 kg do café C, que custa R$ 5,00 o kg. Qual o preço por kg da mistura?

6) Define-se média aritmética de n números dados como o resultado da divisão por n da soma dos n números dados. Sabe-se que 3,6 é a média aritmética de 2,7; 1,4; 5,2 e x. Qual o número x?

7) A média aritmética de 100 números é igual a 40,19. Retirando-se um desses números, a média aritmética dos 99 restantes passará a ser 40,05. Qual foi o número retirado?

8) A média aritmética das idades de um grupo de 120 pessoas é de 40 anos. Se a média aritmética das idades das mulheres é de 35 anos e a dos homens é de 50 anos, qual o número de pessoas de cada sexo, no grupo?

9) É dado um conjunto com 20 números cuja média aritmética é 32. Cada número desse conjunto é multiplicado por 3 e, em seguida, acrescido de 6 unidades. Qual é a média aritmética dos 20 números assim obtidos?

Exercícios

1) Os gastos diários (em reais) de 10 turistas em Ubatuba estão indicados a seguir: 70 – 80 – 90 – 80 – 50 – 65 – 40 – 70 – 100 – 55. Qual o gasto médio dos turistas?

2) Nas eleições em 1º turno em todo o país, no dia 3 de outubro de 1996, inaugurou-se o voto eletrônico. Numa determinada seção eleitoral, cinco eleitores demoraram para votar, respectivamente: 1 min 04s, 1 min 32 s, 1 min 12s, 1 min 52s e 1 min 40s. Qual a média aritmética do tempo de votação desses eleitores?

3) Em um dia de vendas, um vendedor anotou a quantidade de produtos espécie e o preço pelo qual eram vendidos.

Produtos	Quantidade em kg	Preço em reais por kg
A	30	5,00
B	20	3,00
C	10	2,50
D	6	2,00

Qual o preço médio dos produtos vendidos nesse dia?

4) Uma prova foi aplicada em duas turmas distintas. Na primeira turma, com 30 alunos, a média aritmética das notas foi 6,40. Na Segunda, com 50 alunos, foi 5,20. Qual a média aritmética das notas dos 80 alunos?

5) Em uma classe de 40 alunos as notas obtidas em um teste formaram a seguinte distribuição:

Notas	1	2	3	4	5	6	7	8	9	10
Número de alunos	4	4	8	1	2	7	7	5	1	1

Nesse caso, qual a nota média?

6) A tabela abaixo representa a distribuição das exportações de 50 empresas em 2004.

Volume exportado(US$)	Nº de empresas
50000 ├ 60000	5
60000 ├ 70000	10
70000 ├ 80000	20
80000 ├ 90000	10
90000 ├ 100000	5
Total	50

Calcule a média de exportação, em reais, dessas empresas, sabendo que o dólar na época estava cotado a R$ 2,31.

7) Um comerciante atacadista vende determinado produto em sacas que deveriam conter 16,50 kg. A pesagem de 40 sacas revelou os resultados representados na tabela abaixo:

Massas (em kg)	Nº de sacas
14,55 ├ 15,05	1
15,05 ├ 15,55	3
15,55 ├ 16,05	8
16,05 ├ 16,55	9
16,55 ├ 17,05	10
17,05 ├ 17,55	6
17,55 ├ 18,05	3
Total	40

Qual a média da distribuição?

8) O salário médio dos quatro funcionários de uma pequena empresa é R$ 1080,00. Se esta empresa contratar um quinto funcionário com salário de R$ 880,00, qual será a nova média salarial dos funcionários dessa empresa?

9) Um professor de Estatística, após verificar que toda a classe obteve nota baixa, eliminou as questões que não foram respondidas por nenhum de seus alunos. Com isso, as notas de todos os alunos foram aumentadas de três pontos. Então, o que aconteceu com a média das notas dos alunos?

10) Numa pesquisa em determinada cidade foram obtidos os seguintes dados relativos ao número de crianças por família:

Nº de crianças por família	Porcentagem de famílias na cidade
0	5
1	25
2	30
3	20
4	10
5 ou mais	10

O número médio de crianças nas famílias com 5 ou mais filhos é 5,8. Qual o número médio de crianças por família nesta cidade?

- Moda (Mo)

A Moda de um conjunto de valores é a ocorrência mais freqüente entre os valores observados. A moda também é chamada de valor dominante, norma ou valor típico. Quando dizemos, por exemplo, que o salário modal de uma empresa é R$ 1500,00, queremos dizer que esse é o salário recebido pelo maior número de funcionários dessa empresa.

Exemplo

1) Calcular a moda dos seguintes conjuntos de valores:
a) X = {3, 4, 4, 6, 6, 6, 7, 7, 8, 8}
b) Y = {2, 2, 6, 6, 9, 9}
c) Z = {1, 2, 2, 2, 6, 6, 6, 8, 8}
d) W = {6, 7, 8, 9}

Para se determinar a moda quando os valores da variável estão dispostos em uma tabela de distribuição de freqüência com dados não agrupados, basta localizar, na tabela, o valor que possui a maior freqüência.

Exemplo

1) A distribuição dos salários de uma empresa é dada na seguinte tabela:

SALÁRIO EM R$	NÚMERO DE FUNCIONÁRIOS
500	11
1000	5
1500	1
2000	9
5000	4
10500	1
TOTAL	31

Qual é a moda dos salários dessa empresa?

Quando os valores estão agrupados em classes, o procedimento não é imediato, sendo disponíveis alguns métodos de cálculos distintos. Destacamos a seguir alguns desses métodos:

- Moda bruta: Localizamos a classe que apresenta a maior freqüência (classe modal) e calculamos o ponto médio dessa classe. Esse ponto médio é a moda bruta da distribuição.

- Método de King: Esse método baseia-se na influencia das freqüências das classes adjacentes na classe modal.

f_{ant} = freqüência da classe anterior

f_{post} = freqüência da classe posterior

l = limite inferior da classe modal

c = amplitude do intervalo de classe

$$Mo = l + c.\frac{f_{post}}{(f_{ant} + f_{post})}$$

- Método de Czuber: Esse método leva em consideração a freqüência da classe modal e também as freqüências das classes adjacentes e para o seu cálculo consideramos o seguinte:

f_{mo} = freqüência da classe modal

f_{ant} = freqüência da classe anterior

f_{post} = freqüência da classe posterior

l = limite inferior da classe modal

c = amplitude do intervalo de classe

$$Mo = l + c.\frac{f_{mo} - f_{ant}}{2.f_{mo} - (f_{ant} + f_{post})}$$

Exemplo

1) A tabela abaixo representa a distribuição das exportações de 50 empresas em 2004.

Volume exportado(US$)	Nº de empresas
50000 ├ 60000	6
60000 ├ 70000	7
70000 ├ 80000	20
80000 ├ 90000	12
90000 ├ 100000	5
Total	50

Calcule a moda bruta, a moda pelo método de King e a moda pelo método de Czuber da exportação, em dólares, dessas empresas.

Exercícios

1) Os gastos diários (em reais) de 10 turistas em Ubatuba estão indicados a seguir: 70 – 80 – 90 – 80 – 50 – 65 – 40 – 70 – 100 – 55. Qual a moda do gasto dos turistas?

2) Em uma classe de 40 alunos as notas obtidas em um teste formaram a seguinte distribuição:

Notas	1	2	3	4	5	6	7	8	9	10
Número de alunos	4	4	8	1	2	7	7	5	1	1

Nesse caso, qual a nota modal?

3) A tabela abaixo apresenta os salários pagos a 100 funcionários de uma empresa.

Nº de salários mínimos	Nº de operários (f_i)
0 ├ 2	20
2 ├ 4	40
4 ├ 6	15
6 ├ 8	20
8 ├ 10	5
Total	100

Determine a moda bruta, a moda pelo método de King e a moda pelo método de Czuber dos salários dos funcionários dessa empresa.

4) Um comerciante atacadista vende determinado produto em sacas que deveriam conter 16,50 kg. A pesagem de 40 sacas revelou os resultados representados na tabela abaixo:

Massas (em kg)	Nº de sacas
14,55 ├ 15,05	1
15,05 ├ 15,55	3
15,55 ├ 16,05	8
16,05 ├ 16,55	9
16,55 ├ 17,05	10
17,05 ├ 17,55	6
17,55 ├ 18,05	3
Total	40

Determine a moda bruta, a moda pelo método de King e a moda pelo método de Czuber dos salários dos funcionários dessa empresa.

- **Mediana (Md)**

Mediana é o valor que divide uma série ordenada em dois subconjuntos de mesmo número de elementos, sendo que todos os valores do primeiro conjunto são todos menores que ela e os valores do segundo conjunto são todos maiores que ela.

- Mediana para valores não tabulados: A determinação da mediana para valores não tabulados faz-se a partir de um rol.

Exemplo

1) Calcular a mediana dos seguintes conjuntos de números:
a) X = {1, 3, 6, 12, 16, 17, 18, 20, 26}
b) Y = {1, 3, 6, 12, 16, 17, 18, 20, 5, 4}

- Mediana para valores tabulados: Quando os valores já estiverem tabulados sem intervalos de classe, o procedimento a ser adotado será idêntico ao anterior.

Exemplo

1) A distribuição dos salários de uma empresa é dada na seguinte tabela:

SALÁRIO EM R$	NÚMERO DE FUNCIONÁRIOS
500	5
1000	5
1500	2
2000	10
5000	4
10500	1
TOTAL	27

Qual é o salário mediano dessa empresa?

- Mediana para valores agrupados em classes: Quando os valores estão agrupados em intervalos de classe, podemos obter a mediana por interpolação ou utilizando a seguinte fórmula:

$$Md = l + c.\frac{\frac{n}{2} - F_{ant}}{f_{md}}$$, onde:

l é o limite inferior da classe mediana, c é a amplitude do intervalo, n é o número total de observações, F_{ant} é a freqüência acumulada anterior à classe mediana e f_{md} é a freqüência da classe mediana.

Exemplo

1) A tabela abaixo representa a distribuição das exportações de 50 empresas em 2004.

Volume exportado(US$)	N^o de empresas
50000 ├ 60000	5
60000 ├ 70000	10
70000 ├ 80000	19
80000 ├ 90000	11
90000 ├ 100000	5
Total	50

Calcule o valor mediano da exportação dessas empresas.

Exercícios

1) Calcule a mediana dos seguintes conjuntos:
a) X = {70, 80, 90, 80, 50, 65, 40, 68, 100, 55}
b) Y = {12, 14, 18, 3, 15, 15, 40}

2) Em uma classe de 40 alunos as notas obtidas em um teste formaram a seguinte distribuição:

Notas	1	2	3	4	5	6	7	8	9	10
Número de alunos	4	4	8	1	2	7	7	5	1	1

Nesse caso, qual a nota mediana?

3) A tabela abaixo apresenta os salários pagos a 100 funcionários de uma empresa.

Nº de salários mínimos	Nº de operários (f_j)
0 ⊢ 2	20
2 ⊢ 4	40
4 ⊢ 6	15
6 ⊢ 8	20
8 ⊢ 10	5
Total	100

Determine o salário mediano dos funcionários dessa empresa.

4) Os valores da hora de trabalho de cinco funcionários de uma empresa são: R$ 7,50, R$ 9,00, R$ 8,30, R$ 14,20 e R$ 8,80. Determine:
a) a média dos salários por hora de trabalho.
b) a mediana dos salários por hora de trabalho.
c) a moda dos salários por hora de trabalho.

5) Numa pesquisa de opinião, 80 pessoas são favoráveis ao divórcio, 50 são desfavoráveis, 30 são indiferentes e 20 ainda não têm opinião formada a respeito do assunto. Então, a média aritmética será:
a) 180, porque todos opinaram somente uma vez.
b) 40, porque é a média entre os valores centrais 50 e 30.
c) 45
d) 1, porque todos opinaram somente uma vez.
e) não há média aritmética.

Bibliografia:

CRESPO, A. A. **Estatística fácil**. São Paulo: Saraiva, 1993.

OVALLE, I. I., TOLEDO, G. L. **Estatística básica**. São Paulo: Atlas, 1985.

SPIEGEL, M.R. **Estatística**. São Paulo: McGraw-Hill, 1985.

Autor:
JOSE JOÃO DE MELO: *Licenciado em Matemática, Especialista em Educação Matemática pela PUCCAMP, Mestre em Educação Matemática pela PUC-SP, Professor universitário nas áreas de Probabilidade, Estatística, Álgebra Linear, Cálculo Numérico, Custos e Matemática Comercial e Financeira.*