Unidade II

3 TABELAS DE FREQUÊNCIA E SEUS GRÁFICOS

Nesta seção começamos a análise dos dados por meio da construção de tabelas de frequência.

As tabelas de dados brutos, embora tragam todas as características do conjunto, não permitem que tenhamos um entendimento dele, pois os dados assim dispersos somente nos permitem formar uma imagem fragmentada. Tomando a tabela do exemplo da Unidade I, com apenas 12 clientes, verifica-se um amontoado de fatos sem significado, imagine em pesquisas reais, em que podemos ter até mesmo milhões de pessoas entrevistadas, como é o caso do censo. Assim sendo, para que os dados possam se converter em informação é preciso analisá-los.

Como saber que espécie de informação uma tabela de dados brutos esconde? A resposta é até mesmo intuitiva: contando quantas vezes cada valor aparece. De maneira formal, dizemos que a primeira análise a ser feita é a construção das tabelas de frequência para cada uma das variáveis, as quais agrupam os dados com valores iguais ou similares em uma mesma classe e nos dão as frequências com que cada valor ou intervalo de valores aparece no conjunto estudado.

O procedimento para a construção de uma tabela de frequências é simples: tomam-se os valores possíveis da variável e contam-se quantas vezes cada um aparece. Também é comum calcular a frequência relativa, ou seja, qual o percentual relativo a cada valor.

3.1 Construção das tabelas

Para melhor compreensão, vamos construir tabelas de frequência e de frequência relativa para os dados do exemplo a respeito de vendas de suco.

3.1.1 Variável "sabor"

Os valores possíveis para essa variável são: caju, maracujá e uva. É usual chamar os valores das variáveis de x. Utilizando essa notação, teremos: x_1 = caju, x_2 = maracujá, x_3 = uva. Pelo fato de a variável ser nominal, não há ordem específica a ser seguida pelos valores, mas é usual nesses casos, para facilitar sua localização, ordenar os valores alfabeticamente.

Dos diversos clientes, dois compraram suco de caju, seis compraram de maracujá e quatro o de uva. Chamando as frequências de f, teremos: $f_1 = 2$, $f_2 = 6$, $f_3 = 4$. As frequências relativas serão os valores percentuais de cada sabor. Ou seja: $f_1 = 100 * 2/12$, $f_2 = 100 * 6/12$, $f_3 = 100 * 4/12$. Fazendo as contas e apresentando os resultados com aproximação a partir da primeira casa decimal: $f_1 = 16{,}7\%$, $f_2 = 50{,}0\%$, $f_3 = 33{,}3\%$.

Além dos dados anteriores, é comum que tenhamos uma linha adicional com a quantidade total de dados.

Colocamos, então, essas informações na tabela 2, como segue:

xi	fi	fi(%)
caju	2	16,7
maracujá	6	50
uva	4	33,3
Total	**12**	**100**

Tabela 2 – Frequências e frequências relativas para a variável "sabor".

Comparando a tabela acima com os valores dados na tabela de dados brutos, notamos que essa nova organização nos mostra de maneira muito mais rápida e direta o grau de preferência dos clientes com relação ao sabor.

3.1.2 Variável "tamanho"

Fazendo a mesma coisa para a segunda variável, teremos a tabela 3. Assim como para o caso da variável "sabor", também aqui a distribuição de preferências dos clientes fica muito mais visível que na tabela de dados brutos. Atente agora para o seguinte detalhe: como a variável é ordinal, a tabela deve seguir a ordem natural dos valores, ou seja, ordeno os valores do menor para o maior, assim: x_1 = pequeno, x_2 = médio, x_3 = grande.

xi	fi	fi(%)
Pequeno	3	25
Médio	5	41,67
Grande	4	33,33
Total	**12**	**100**

Tabela 3 – Frequências e frequências relativas para a variável "tamanho".

3.1.3 Variável "preço"

Empregamos o mesmo procedimento para a variável "preço". Utilizamos os valores existentes para os preços, pois, como a variável é contínua, não é possível determinar antes os valores possíveis, pois eles são inúmeros. Assim procedendo, teremos construída a tabela 4. Note que não incluímos as frequências relativas por uma questão de concisão.

xi	fi
1,20	1
1,35	1
1,50	2
1,75	3
2,00	1
4,80	1
4,95	1
5,00	1
5,30	1
Total	**12**

Tabela 4 – Frequências sem perda de informação para a variável "preço".

A tabela feita dessa maneira é dita "sem perda de informações", pois cada valor individual está ali representado. No entanto, se olharmos para os dados que há nela, nesse caso constataremos que não há aí o mesmo ganho de clareza que nos casos anteriores, pois há muitos valores possíveis para a variável. Isso faz com que haja uma grande dispersão dos dados entre todos esses valores. Essa característica nos faz buscar uma nova maneira de construir essa tabela.

O procedimento apropriado em casos como esse é utilizar intervalos de valores em lugar de valores individuais. Como os valores das variáveis oscilam variam entre 1,20 e 5,30, por exemplo, escolho utilizar intervalos que agrupem de 1 em 1 real. Perceba a necessidade de definir o intervalo de maneira que não haja dúvidas quanto à colocação de certo dado, razão pela qual o final de cada intervalo refere-se a 99 centavos. Com isso, temos uma nova distribuição de frequências, conforme dados da tabela 5.

Preço	Frequência
1,00 – 1,99	7
2,00 – 2,99	1
3,00 – 3,99	0
4,00 – 4,99	2
5,00 – 5,99	2
Total	**12**

Tabela 5 – Frequências para a variável "preço", com agrupamento por intervalos.

Olhando para a nova tabela, percebe-se mais claramente haver uma concentração maior de compras dos artigos mais baratos, o que não ficava claro na construção anterior.

3.1.4 Variável "vendas"

Utilizamos finalmente um procedimento similar para a variável "vendas", sem perdas de informação. Como a variável é discreta, podemos admitir que os valores xi serão os números inteiros entre 1 e 12, que englobam todos os dados coletados. Desse modo construímos a tabela 6.

xi	fi
1	4
2	0
3	1
4	2
5	0
6	1
7	0
8	2
9	1
10	0
11	0
12	1
Total	**12**

Tabela 6 – Frequências sem perda de informação para a variável "vendas".

Novamente, a tabela traz dados muito dispersos pelos valores, não permitindo uma compreensão muito melhor do que a tabela de dados brutos. De novo, buscamos a construção de intervalos de valores em lugar de valores individuais. Note-se que a construção de intervalos requer uma escolha, e diferentes escolhas podem gerar diferentes leituras. Para exemplificar, foram construídas duas tabelas com escolhas diferentes de intervalos: tabelas 7 e 8.

xi	fi
1 – 5	7
6 – 10	4
11 – 15	1
Total	**12**

Tabela 7 – Frequências para a variável "vendas", com agrupamento por intervalos de tamanho 5.

xi	fi
1 – 4	7
5 – 8	3
9 – 12	2
Total	**12**

Tabela 8 – Frequências para a variável "vendas", com agrupamento por intervalos de tamanho 3.

Para evitar que as diferenças geradas pela escolha dos intervalos possa levar a conclusões errôneas, há critérios de escolha, mas a discussão desses critérios vai além do escopo e dos objetivos desta disciplina. Neste momento, basta saber que em estudos reais tomam-se os cuidados necessários para evitar esse problema, muitas vezes, inclusive, utilizando-se de intervalos de tamanhos diferentes, quando preciso.

3.2 Gráficos de tabelas de frequência

A apresentação de dados estatísticos utiliza-se não somente de tabelas, mas também de gráficos. O objetivo de uma apresentação gráfica é tornar as características importantes visíveis em um tempo bastante curto. Essa característica dos gráficos é que torna sua utilização tão frequente na apresentação de conjuntos de valores, principalmente quando se discutem aspectos técnicos, por exemplo, nos cadernos de economia dos jornais.

Neste momento, vamos nos ater à construção de gráficos de colunas e diagramas circulares, pois eles são os mais comumente utilizados para representar frequências. Um mesmo conjunto de dados pode ser apresentado em diferentes tipos de gráficos. Novamente, utilizaremos o exemplo de dados a respeito da venda de sucos e discutiremos a construção de gráficos a partir de sua elaboração para as tabelas de frequência construídas na seção anterior.

Após a construção dos gráficos, discutiremos brevemente a conveniência da utilização de cada um para cada situação.

Observação

A escolha apropriada do tipo de gráfico dependerá das características do conjunto que se queira enfatizar.

3.2.1 Gráficos de colunas

Os gráficos de colunas são construídos tendo como eixo horizontal os valores da variável e na vertical a frequência. Assim sendo, as colunas serão tanto mais altas quanto maior a frequência daquele valor.

Para a grandeza "sabor", apresentada no gráfico 1, como não há ordenamento natural dos valores, podemos escolher a ordem que nos pareça mais adequada. Por coerência, utiliza-se a mesma ordem já definida na construção da tabela 2.

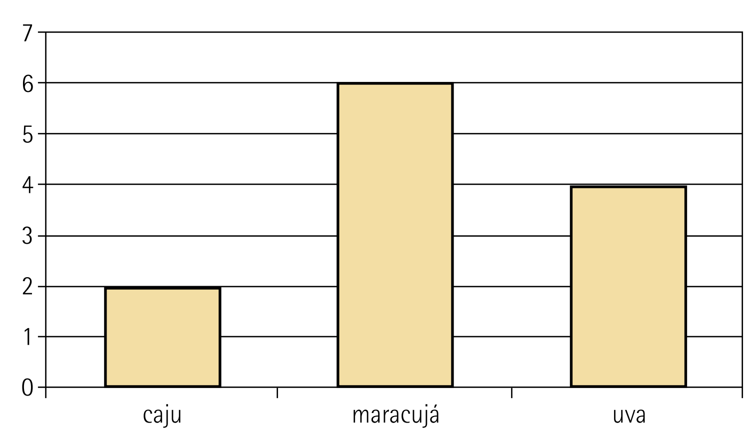


Gráfico 1 – Colunas para as frequências referentes à variável "sabor".

Seguindo as definições de significado dadas anteriormente, vemos que para os valores "caju", "maracujá" e "uva" temos colunas de altura 2, 6 e 4, respectivamente, o que corresponde ao número de clientes entrevistados que compraram sucos de cada um desses sabores.

No caso da variável "tamanho", temos como eixo horizontal os valores "pequeno", "médio" e "grande". Como a grandeza é ordinal, o eixo deve seguir o valor crescente de tamanho, definindo a ordem de aparecimento das colunas referentes a cada valor, conforme apresentamos no gráfico 2.

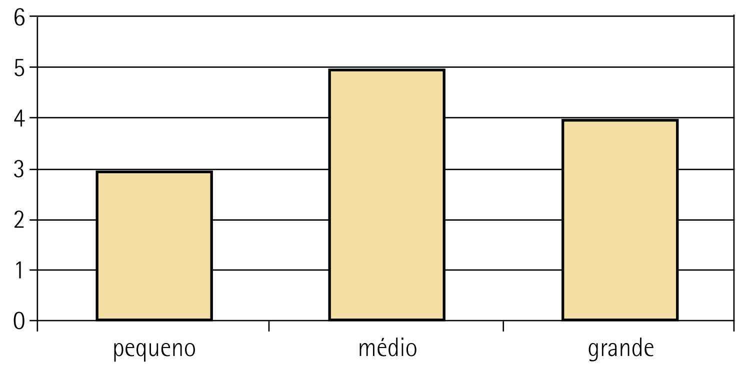


Gráfico 2 - Colunas para as frequências referentes à variável "tamanho".

Para a variável "preço", utilizamos como referência de cada categoria o centro do intervalo correspondente. Como os valores são contínuos, as colunas foram construídas sem espaçamento, de modo a deixar claro o fato de que um elemento que pertença a uma coluna terá valor no intervalo correspondente. Por exemplo, o elemento contado como fazendo parte da coluna 2,5 terá valor entre 1 e 2. Observe como essas informações são apresentadas no gráfico 3 a seguir.

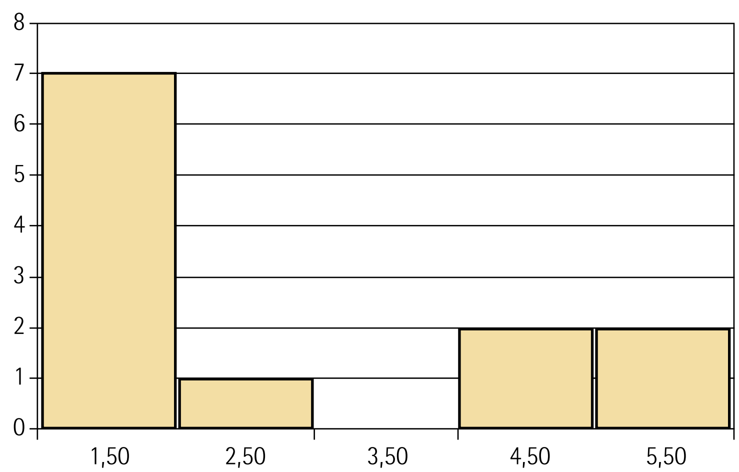


Gráfico 3 - Colunas para as frequências referentes à variável "preço".

Por fim, trazemos o gráfico da variável "vendas", para o qual escolhemos a distribuição de intervalos de tamanho 3. Desta vez, optamos por utilizar diretamente o intervalo como rótulo dos dados. Esses valores de frequência estão apresentados no gráfico 4.

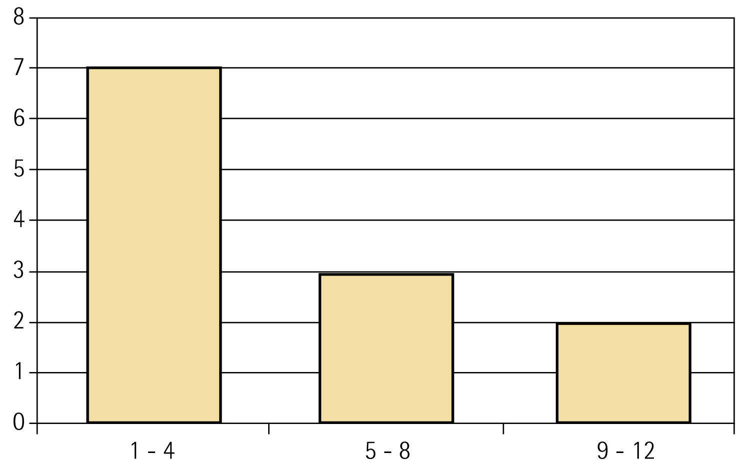


Gráfico 4 - Colunas para as frequências referentes à variável "vendas".

3.2.2 Diagramas circulares

Os diagramas circulares são construídos de tal modo que, quanto maior a proporção de uma categoria no conjunto, maior será a área do círculo que se refere a ela, sendo que a categoria pode referir-se a um valor ou a um intervalo de valores. Devido aos cortes radiais que definem fatias da circunferência, esses gráficos também são conhecidos por "gráficos de pizza".

Por sua natureza, são bastante utilizados nos casos em que mostrar a quantidade específica de cada categoria não é importante, mas sim mostrar clara e rapidamente as proporções de cada uma.

Como exemplo, utilizaremos as mesmas tabelas de frequência para as quais fizemos os gráficos de coluna.

Para a variável "sabor", temos que o valor "maracujá" que responde por metade das vendas, logo, ele será representado por uma "fatia" correspondente, ou seja, pela metade do círculo; em outras palavras, por uma seção circular de 180°, visto que o círculo todo tem 360°. O restante será dividido na proporção de 4 para 2 entre os sabores uva e caju; desse modo, o primeiro terá uma fatia de 120° e o segundo uma de 60°, conforme representado no gráfico 5.

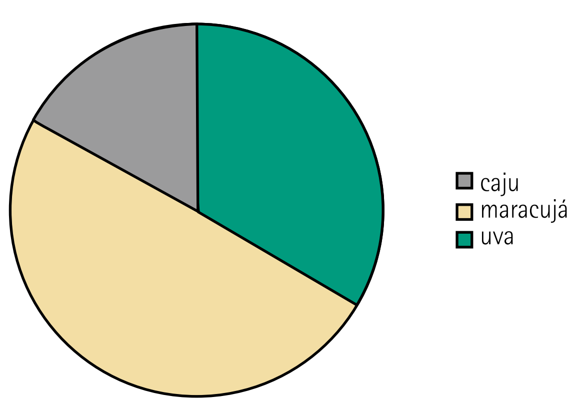


Gráfico 5 – Diagrama circular para a variável "sabor", de acordo com os dados da tabela 2.

De maneira equivalente foram construídos os gráficos para as variáveis "tamanho", "preço" e "vendas", os quais estão apresentados nos gráficos 6, 7 e 8, respectivamente.

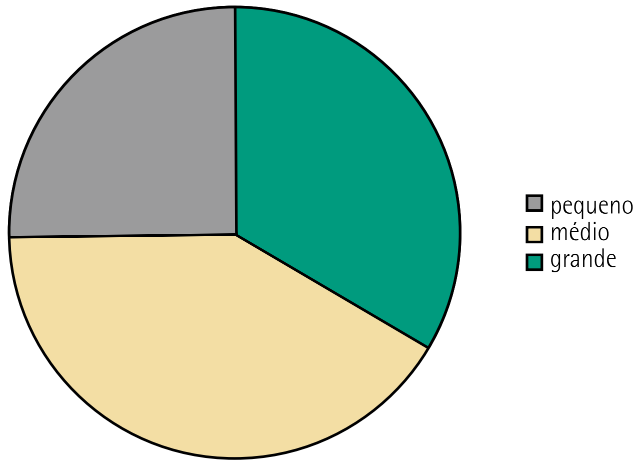


Gráfico 6 – Diagrama circular para a variável "tamanho", de acordo com os dados da tabela 3.

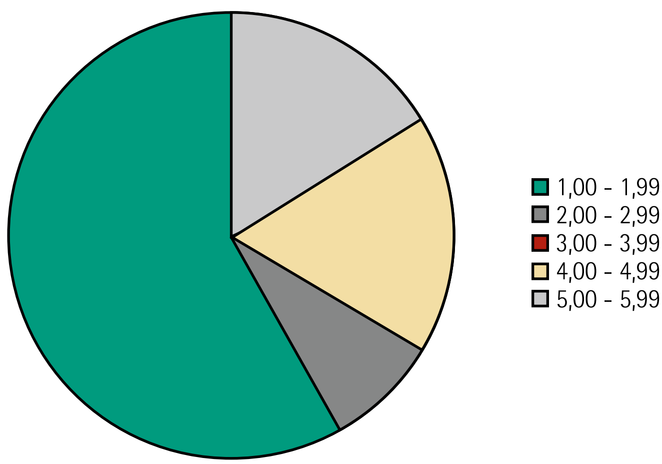


Gráfico 7 – Diagrama circular para a variável "preço", de acordo com os dados da tabela 5.

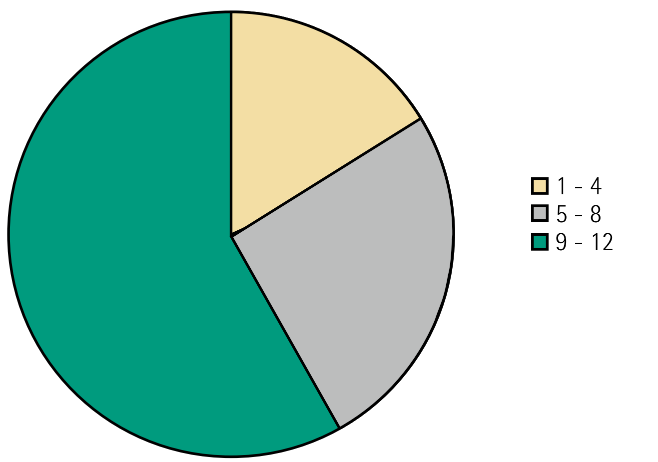


Gráfico 8 – Diagrama circular para a variável "vendas", de acordo com os dados da tabela 6.

3.2.3 Escolha do tipo de gráfico

Como dissemos no início desta seção, o papel de um gráfico é trazer alguma informação de maneira clara e direta, logo, o critério para se escolher um gráfico ou outro quando vamos apresentar dados depende de qual é o objetivo dessa figura em nossa argumentação. Ou seja, se quisermos enfatizar os valores específicos de cada frequência para cada variável, o gráfico mais apropriado será o de colunas; no entanto, se for interessante enfatizar a proporção de cada um, devemos utilizar o diagrama circular.

Tomemos como exemplo o caso da variável "sabor". No gráfico 1, a característica mais marcante é o fato de maracujá ser o valor mais frequente, mas não fica claro que isso diz respeito à metade dos clientes. No gráfico 5, no entanto, a informação de que um dos sabores responde por metade das vendas salta aos olhos em poucos instantes. Logo, se o objetivo for ressaltar a preferência por maracujá, é melhor usar o diagrama circular, mas se a intenção é fazer uma discussão que englobe os três sabores, pode ser mais interessante utilizar o gráfico de colunas.

Do mesmo modo, se tomamos como exemplo a variável "preço", vemos que o diagrama circular nos dá, de maneira imediata, a característica de o artigo mais barato ter um número muito maior de vendas. Entretanto, como o diagrama circular não enfatiza a ordem, não se percebe que há uma segunda concentração das vendas na faixa de maior preço, o que se vê facilmente no gráfico de colunas. Logo, o diagrama circular será adequado numa discussão em que a informação de qual a faixa de preços mais procurada for mais interessante, mas, se for importante conhecer o comportamento para todas as faixas, ele não apresenta bem os dados.

3.2.4 Comentários adicionais

Apresentamos aqui somente dois tipos de gráficos, por considerá-los os mais significativos para as discussões de interesse da disciplina. No entanto, existem diversos outros tipos, mas a lógica de construção que os rege é geralmente derivada da mesma lógica de construção dos tipos vistos aqui. Portanto, não deve haver dificuldade na compreensão ou na construção de algum outro formato.

Note ainda que atualmente é raro que tenhamos que construir gráficos manualmente, visto que todas as planilhas de dados já possuem ferramentas de elaboração dos gráficos em diversos formatos, e ainda há programas específicos de excelente qualidade desenvolvidos para esse fim.

4 MEDIDAS DE POSIÇÃO

No primeiro passo de transformar dados em informação, construímos as tabelas de frequência. Mas, para irmos adiante em nossas análises, é preciso encontrar maneiras de encontrar características que se apresentem de forma ainda mais resumida.

Nesta seção apresentaremos três grandezas que resumem características de um conjunto em apenas um valor: a moda, a mediana e a média.

4.1 Moda

A palavra moda é utilizada por nós cotidianamente com um significado parecido com a definição em estatística. No dia a dia, dizemos que algo está na moda se muita gente o está usando ou o está fazendo. Em estatística, moda é o valor que mais aparece no conjunto, aquele que é a característica da maioria. Quando há um valor que se sobressai em frequência, com relação aos demais valores, dizemos que o conjunto tem uma moda, é **modal**.

Tomemos como exemplo os dados da tabela 9. Eles dizem a respeito às vendas mensais de impressoras e multifuncionais em uma loja.

Tipo	Quantidade
Impressora jato de tinta	112
Impressora *laser* monocromática	263
Impressora *laser* colorida	185
Multifuncional jato de tinta	488
Multifuncional *laser* monocromática	124
Multifuncional *laser* colorida	93
Total	**1.265**

Tabela 9 – Quantidade de impressoras e multifuncionais vendida na loja A.

Vemos que o artigo que mais aparece é a multifuncional jato de tinta, com 488 unidades. Um erro comum é atribuir à moda a frequência do valor, mas ressalto que a moda será o valor da variável e não a quantidade de vezes que o mesmo aparece. Ou seja, a moda desse conjunto será, portanto, “multifuncional jato de tinta”.

Há situações em que o conjunto tem dois valores mais frequentes em lugar de um. Nesses casos, dizemos que o conjunto tem duas modas. Um conjunto com essa característica também é chamado de bimodal.

Tomando o mesmo tipo de dados que no exemplo anterior, temos na tabela 10 um caso em que o conjunto é **bimodal**.

Tipo	Quantidade
Impressora jato de tinta	102
Impressora *laser* monocromática	485
Impressora *laser* colorida	183
Multifuncional jato de tinta	488
Multifuncional *laser* monocromática	85
Multifuncional *laser* colorida	92
Total	**1.435**

Tabela 10 – Quantidade de impressoras e multifuncionais vendida na loja A.

Nessa ocorrência, muito embora os valores de vendas não sejam rigorosamente iguais, nota-se haver dois produtos que são muito mais procurados que os demais, portanto, dizemos que o conjunto é bimodal, tendo como modas "impressora laser monocromática" e "multifuncional jato de tinta".

Existe ainda a possibilidade de não haver nenhum valor que realmente se sobressaia com relação aos demais. Quando isso acontece, dizemos que o sistema não tem moda ou, mais formalmente, que o conjunto é **amodal**.

Utilizando novamente o tema dos exemplos anteriores, mostramos na tabela 11 um caso em que o conjunto é amodal, não havendo nenhum valor que realmente se destaque dos demais em número de vezes que aparece.

Tipo	Quantidade
Impressora jato de tinta	98
Impressora *laser* monocromática	115
Impressora *laser* colorida	121
Multifuncional jato de tinta	118
Multifuncional *laser* monocromática	95
Multifuncional *laser* colorida	72
Total	**619**

Tabela 11 – Quantidade de impressoras e multifuncionais vendida na loja A.

A moda é uma grandeza de fácil determinação e de grande utilidade quando é interessante verificar se o conjunto tem alguma característica marcante, um valor que seja muito comum. No entanto, essa é a única informação que a grandeza moda pode nos dar. Ela não traz qualquer informação a respeito dos demais valores do conjunto, logo, é importante que definamos outras grandezas para podermos aprofundar nossas análises.

4.2 Mediana

A segunda grandeza que utilizaremos como um resumo do conjunto é a mediana, que definimos como sendo o valor que corresponde ao ponto central do conjunto. Ou seja, a mediana é o valor que divide o conjunto em duas metades, isto é, haverá um mesmo número de valores inferiores e superiores ao valor da mediana.

Tomemos como exemplo um conjunto de pessoas cujas idades estão colocadas na tabela 12.

Nome	Idade
Ana	3
Bruno	32
Denise	19

Gabriel	56
José	12
Maria	16
Pedro	11

Tabela 12 – Idades dos integrantes de uma família.

A mediana será a idade que divide o conjunto em dois, ou seja, a idade acima e abaixo da qual haverá o mesmo número de pessoas. Metaforicamente, podemos imaginar que colocamos essas pessoas em fila, da mais nova à mais velha. Como temos um número ímpar de elementos no conjunto, a mediana será a idade da pessoa que ficou no meio da fila. Como há 7 elementos no conjunto, a pessoa do meio da fila será a quarta mais velha. Colocando as idades em ordem, temos: 3, 11, 12, 16, 19, 32 e 56. Nesse caso, a mediana é 16.

Suponhamos, agora, que uma pessoa de 5 anos seja adicionada ao conjunto. Com isso, a sequência crescente de idades passa a ser 3, 5, 11, 12, 16, 19, 32 e 56. A divisão em dois subconjuntos faria com que o primeiro tivesse as quatro pessoas mais novas e o segundo, as quatro mais velhas. Define-se como mediana, nesse caso, o valor intermediário entre o maior do primeiro conjunto e o menor do segundo. Para o exemplo teríamos as idades de 12 e 16. A mediana seria o valor equidistante de 12 e 16, que é 14.

Note que a mediana não é influenciada pelos valores não centrais do conjunto. Por exemplo, se Gabriel tivesse 40 anos em lugar de 56, não haveria qualquer mudança no valor da mediana.

Vale ainda comentar que essa grandeza só pode ser definida para grandezas quantitativas ou para qualitativas ordinais, visto que não há possibilidade de haver ordenação das variáveis qualitativas nominais.

Observação

É comum o aluno confundir os conceitos de média e mediana. Assim, antes de passar ao próximo tópico, tenha certeza de ter assimilado bem o conceito de mediana.

4.3 Média

As medidas de posição dadas anteriormente podem traduzir algumas características interessantes dos conjuntos estudados, mas não permitem um estudo analítico mais profundo. Para que isso seja feito, é preciso definir uma terceira medida de posição do conjunto que é a média.

Assim como outros conceitos, também o de média faz parte de nossa linguagem cotidiana, pois costumamos realizar com bastante frequência cálculos de médias, ainda que de maneira informal e intuitiva. Por exemplo, quando saímos para trabalhar, sabemos aproximadamente quanto tempo demoraremos no trajeto, e assim também sabemos quais são, em média, os nossos gastos com determinados itens comumente comprados etc.

Imagine a seguinte situação: um grupo de amigos organiza um jantar e cada um fica responsável por realizar parte das compras. Ao final, cada um deles comunica o valor gasto e eles dividem igualmente as despesas. O valor da contribuição de cada um para a festa é justamente a média de gastos do grupo.

Visto o conceito, passaremos agora a formalizar a maneira de realizar os cálculos de médias e aprofundaremos o conceito discutindo as médias ponderadas. Em seguida, comentaremos algumas utilizações de interesse na área econômica.

Observação

As médias são as grandezas mais comumente utilizadas nas análises estatísticas.

4.3.1 Média aritmética

A chamada média aritmética, ou média ou média simples, é a mais utilizada no cotidiano e estabelece que todos os elementos têm a mesma importância.

Se detalharmos o exemplo citado anteriormente, na tabela 13 teremos listados os gastos individuais. Para saber o valor do gasto efetivo de cada um, temos que somar todos os gastos individuais e dividi-los pelo número de pessoas.

Pessoa	Gasto
Alberto	25
Beatriz	31
Carlos	47
Diana	19
Edgar	28

Tabela 13 – Gastos efetuados para a realização do jantar por participante do grupo de amigos.

Nesse caso, teremos:

$$\text{Gasto médio} = \frac{25+31+47+19+28}{5} = 30$$

Dessa maneira, determinamos que a média de gastos foi de R$ 30,00, ou seja, esse é o valor correspondente aos gastos de cada um. É interessante observar que o valor da média não precisa ser um valor encontrado no conjunto.

Utilizando o exemplo como guia, vamos agora formalizar o cálculo da média aritmética.

Em primeiro lugar, utilizaremos a letra N para denotar a quantidade de pessoas, x para indicar a variável "gasto" e o índice i para indicar cada gasto individual. Ou seja, teremos que N = 5, ou seja, o gasto efetuado por Beatriz poderia ser escrito como $x_2 = 31$. Para indicar que estamos falando da média, utilizaremos uma barra superior. Com isso, a expressão anterior pode ser reescrita de maneira mais geral como:

$$\overline{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{N}$$

Mas, essa notação não será prática se tivermos um grande número de elementos, assim, utilizamos um indicador de que temos que somar todos os valores de x_i. A isso chamamos "somatório de x", ou seja, soma de todos os valores de x. Para indicar tal operação, convencionou-se utilizar a letra grega sigma maiúscula, como segue:

$$\overline{x} = \frac{\sum x_i}{N}$$

Note que utilizamos aqui a notação mais concisa. Em muitos textos, utiliza-se na expressão a indicação explícita da soma de todos os valores de x entre 1 e N. A notação explícita é absolutamente necessária em casos em que pode haver valores não considerados, mas, como no contexto desta disciplina sempre utilizaremos somatórios com todos os elementos, tornou-se mais conveniente usar a notação resumida em lugar da mais explicativa.

Passamos agora a discutir o cálculo da média quando os dados estão apresentados em uma tabela de frequência. Para isso, começaremos calculando a média da mesma maneira utilizada anteriormente, mas agora para uma situação ligeiramente diferente. Suponhamos agora que haja uma mesa com 10 pessoas em uma lanchonete, as quais combinaram dividir igualmente a despesa final. Os dados a respeito do consumo de cada um dos ocupantes estão apresentados na tabela 14.

Pessoa	Gasto
1	12
2	15
3	15
4	18
5	13
6	12
7	13
8	15
9	12
10	15

Tabela 14 – Gastos efetuados por ocupante da mesa da lanchonete.

Não é difícil perceber que montar a conta da maneira utilizada anteriormente é pouco prático. Assim, vamos inicialmente montar a tabela de frequências, o que se apresenta na tabela

15. Temos aqui como valores xi os gastos individuais e como fi o número de pessoas que gastou esse valor.

xi	fi
12	3
13	2
15	4
18	1

Tabela 15 – Frequência para os gastos efetuados por ocupante da mesa da lanchonete.

Para calcular a média precisamos, em primeiro lugar, do número de pessoas, que é o somatório das frequências, como segue:

$$N = \sum f_i$$

O segundo valor a ser determinado é o valor total da conta, o qual é calculado considerando o seguinte:

- 3 pessoas gastaram 12, logo, somam-se 36 à conta.
- 2 pessoas gastaram 13, logo, somam-se 26 à conta.
- 4 pessoas gastaram 15, logo, somam-se 60 à conta.
- 1 pessoa gastou 18, logo, somam-se 18 à conta.

Formalmente, podemos escrever de maneira geral que se efetua o somatório do produto do valor da variável pela sua frequência. O resultado será então dividido pelo número de dados N, que é o somatório das frequências. O cálculo da média é dado, portanto, pela expressão:

$$\overline{x} = \frac{\sum f_i . x_i}{\sum f_i}$$

No exemplo, como a soma das frequências nos dá o total de pessoas na mesa, que são 10, temos uma conta de R$ 140,00 a ser dividida por 10 pessoas. A média, portanto, será de R$ 14,00. Na tabela 16 é apresentada uma forma prática de se efetuar a conta. Acrescentamos uma coluna à tabela, na qual, para cada valor de x se coloca o resultado do produto $x_i * f_i$. Coloca-se também uma linha adicional ao final da tabela, na qual se apresentam os resultados dos somatórios que serão utilizados no cálculo da média.

Observação

Note que estamos calculando exatamente a mesma coisa e que utilizar as frequências explicitamente apenas muda a maneira de fazer as contas.

xi	fi	xi*fi
12	3	36
13	2	26
15	4	60
18	1	18
Soma	**10**	**140**

Tabela 16 – Quadro esquemático para cálculo de médias a partir de tabela de frequências.

Dessa maneira, o cálculo da média é feito dividindo-se o valor da soma da terceira coluna, R$ 140,00, pela soma da segunda coluna, 10 pessoas. Como vimos anteriormente, a média será de R$ 14,00.

Daremos agora um exemplo genérico, em que os dados de frequências serão apresentados diretamente na tabela. Veja a seguir:

xi	fi
5	5
6	8
7	9
8	3

Tabela 17 – Frequências para os valores de uma variável x genérica.

Seguindo o procedimento descrito anteriormente, acrescentaremos a coluna com os produtos $x_i * f_i$, de modo a obter a soma de todos os valores de x do conjunto (no exemplo, era o total da conta). Uma linha adicional será utilizada para colocar o resultado dos somatórios de cada coluna, deixando claros os valores da soma anteriormente descrita e o número total de dados N (no exemplo, o número de pessoas na mesa). Por fim, acrescentamos ainda uma última linha, que trará o resultado final, consequência da divisão do somatório da terceira coluna pelo somatório da segunda coluna. O procedimento aqui descrito está colocado na tabela 18. Vemos aí que a situação equivale a termos um gasto de R$ 160,00 dividido igualmente por 25 pessoas, obtendo a média de 6,4.

xi	fi	xi*fi
5	5	25
6	8	48
7	9	63
8	3	24
Soma	**25**	**160**
Média	**6,4**	

Tabela 18 – Cálculo do valor médio de um conjunto de dados para uma variável x genérica a partir de sua tabela de frequências.

4.3.2 Cálculo da média para dados apresentados em intervalos

Quando a tabela de frequências traz classes de dados em intervalos em lugar de valores individuais, assumimos que o valor que melhor representa a classe é o valor referente ao meio do intervalo.

Tomando como exemplo os intervalos da tabela 7, os intervalos teriam por valores representativos:

- O intervalo 1-5 tem como valor representativo $x_1 = 3$.
- O intervalo 6-10 tem como valor representativo $x_2 = 8$.
- O intervalo 11-15 tem como valor representativo $x_3 = 13$.

Para efeito de cálculo, remontamos a tabela tendo em lugar dos intervalos os valores centrais e calculamos a média da mesma maneira já descrita anteriormente, conforme apresentamos na tabela 19.

Intervalo	xi	fi	xi*fi
1-5	3	7	21
6-10	8	4	32
11-15	13	1	13
	Soma	**12**	**66**
	Média	**5,5**	

Tabela 19 – Cálculo do valor médio de um conjunto de dados a partir da tabela de frequências dada por intervalos de valores.

Como os valores tratados dessa maneira não consideram exatamente o valor de cada um dos dados, a média será um valor aproximado. Portanto, quando houver acesso aos dados brutos e também necessidade de um cálculo preciso, é preferível utilizar os dados brutos para efetuar o cálculo da média.

4.3.3 Média ponderada

Até aqui tratamos todos os dados como tendo a mesma importância para o cálculo da média. No entanto, sabemos que na realidade há gradações de importância que fazem com que esse cálculo de média simples não seja uma descrição apropriada da realidade.

Pensemos no seguinte exemplo cotidiano: para avaliar o desempenho de um aluno, é usual que se utilize mais de um instrumento de apreciação. Digamos que um professor aplique duas avaliações no bimestre: a primeira seria um trabalho em grupo a respeito de algum tópico desenvolvido em aula e a segunda uma prova a respeito de todo o conteúdo. Nesse caso, o segundo instrumento de avaliação traz mais informações a respeito do desempenho de cada aluno que o trabalho, e é necessário que essa diferença de importância esteja refletida na nota final, o que se faz estabelecendo pesos para cada nota.

O cálculo de uma média ponderada se faz de maneira similar ao cálculo em que são utilizadas tabelas de frequência.

Tomando o exemplo das notas, digamos que o professor assuma que a prova é quatro vezes mais importante que o trabalho; ele então atribuirá peso 1 ao trabalho e peso 4 à prova. Para o cálculo da nota final, chamaremos de n_1 e p_1 a nota e o peso do trabalho, respectivamente. Do mesmo modo, chamaremos de n_2 e p_2 a nota e o peso da prova. Em analogia ao cálculo de médias pela tabela de frequências, poderíamos imaginar que a situação equivalente seria termos 1 trabalho e 4 provas e calcularíamos a nota final como se fossem 5 avaliações em lugar de duas: um trabalho e quatro provas sendo, as quatro com a mesma nota. O cálculo da média final do aluno seria dado, então, por:

$$M = \frac{p_1.n_1 + p_2.n_2}{p_1 + p_2}$$

Com os pesos citados anteriormente, teríamos, portanto:

$$M = \frac{1.n_1 + 4.n_2}{1 + 4}$$

Ou, simplesmente:

$$M = \frac{n_1 + 4n_2}{5}$$

Na tabela 20 apresentamos alguns exemplos de cálculo de médias para diversos conjuntos de notas considerando os pesos citados acima.

Aluno	n_1	n_2	n_1*p_1	n_2*p_2	N
1	0	10	0	40	8
2	10	0	10	0	2
3	8	2	8	8	3,2
4	2	8	2	32	6,8
5	7	7	7	28	7
6	9	6	9	24	6,6

Tabela 20 – Exemplo de cálculo de médias com ponderação: notas de alunos.

A partir do exemplo anterior, podemos agora apresentar a fórmula geral para o cálculo de médias ponderadas:

$$\bar{x} = \frac{\sum p_i.x_i}{\sum p_i}$$

Vejamos agora outro exemplo de interesse, em que trataremos da dinâmica de cálculo de inflação. Os números e os artigos aqui abordados são claramente irreais, uma vez que nosso propósito é exclusivamente didático.

Suponhamos que o cálculo de inflação levasse em conta somente os aumentos de cinco artigos num dado período, conforme apresentado na tabela 21.

	Artigo	Aumento (xi)
1	arroz	1
2	açúcar	1
3	café	1
4	feijão	1
5	uva	20

Tabela 21 – Exemplo de cálculo de média aritmética para um conjunto de aumento de preços (em valores percentuais).

Como a uva teve aumento de 20% e os demais artigos somente de 1%, teríamos, num cálculo sem ponderação:

$$\overline{x} = \frac{1+1+1+1+20}{5} = 4{,}8$$

O que resulta é de uma inflação de 4,8% no período.

É imediato pensar, porém, que esse valor não é representativo, uma vez que o consumo de uvas é geralmente esporádico e o valor de 4,8% não reflete o impacto que tal aumento traria à economia da região.

Em outras palavras, considerar que o aumento do preço da uva tenha o mesmo peso ou importância do aumento do preço do arroz não é razoável e é preciso que o valor médio do aumento de preços leve em conta essas diferenças.

Estabelecemos, então, pesos fictícios para os artigos citados, buscando refletir a importância atribuída a cada um deles. Cálculos reais de inflação têm critérios para a escolha dos pesos, os valores aqui escolhidos apenas buscam ilustrar o efeito dos pesos.

Segundo nossas estimativas, o arroz seria o item mais relevante, com grau de importância 50; o açúcar e o feijão viriam empatados em segundo lugar, ambos com importância 35; o café teria importância 25 e a uva, por ter pequeno impacto, teria apenas importância 1. Esses valores estão apresentados na tabela 22.

Temos, assim, um novo cálculo de inflação, em que o aumento da uva se torna muito menos importante que os demais aumentos, fazendo a descrição do aumento médio dos preços ser mais condizente com a realidade.

A maneira de realizar o cálculo é similar à utilizada quando tínhamos os dados na tabela de frequência:

- Acrescentamos uma coluna onde colocaremos os resultados dos produtos $p_i.x_i$, na qual p_i é o peso do artigo i e xi é o aumento de preço do mesmo artigo.
- Acrescentamos uma linha adicional, na qual serão colocados os somatórios dos valores das colunas p_i e $p_i.x_i$.
- Efetuamos os somatórios dos pesos e colocamos o resultado na linha adicional da coluna p_i.
- Efetuamos os somatórios dos produtos $p_i.x_i$ e colocamos o resultado na linha adicional da coluna $p_i.x_i$.
- Efetuamos a divisão dos dois somatórios para obter a média.

O procedimento descrito está esquematizado na tabela 22.

	Artigo	Aumento (xi)	Peso (pi)	pi*xi
1	arroz	1	50	50
2	açúcar	1	35	35
3	café	1	25	25
4	feijão	1	35	35
5	uva	20	1	20
		Soma	**146**	**165**
			Média	**1,13**

Tabela 22 – Exemplo de cálculo de médias com ponderação para aumento de preços (em valores percentuais).

Calculando a média dessa maneira, a inflação recalculada nesses moldes é de 1,13%, o que é muito mais condizente do que o valor anterior de 4,8%, pois temos um número próximo àquele do aumento dos itens importantes.

4.3.4 Propriedades da média

Para terminar nossa discussão a respeito das médias, vale comentar duas de suas características:

1. Se somamos um mesmo número a todos os elementos do conjunto, a média será acrescida de mesmo valor.

- É fácil entender essa propriedade se pensarmos em uma média de idade. Um conjunto que tenha, por exemplo, uma média de idade de 6 anos, passados 3 anos, todos terão envelhecido o mesmo tempo e a idade média do conjunto será acrescida de 3 anos.

2. Se multiplicarmos todos os elementos do conjunto por um mesmo número, a média será acrescida de mesmo valor.

- Podemos compreender essa propriedade facilmente se pensarmos no exemplo da lanchonete. Se todos se reunirem uma vez por semana na mesma lanchonete e consumirem a mesma coisa todas as vezes, o gasto mensal de cada um será multiplicado por quatro e o gasto médio será a soma dos gastos médios diários, resultando num gasto mensal 4 vezes maior.

4.3.5 Médias ponderadas da área econômica: índices de inflação

Os indicadores da área econômica são calculados levando-se em conta a relevância dos diversos fatores abordados no universo estudado. Em outras palavras, os índices econômicos são geralmente uma ponderação de diversos dados com pesos diferentes.

Como discutimos anteriormente, o cálculo da inflação precisa ponderar a importância dos diversos artigos. Além disso, também há outros fatores a considerar. Por exemplo, é mais importante o valor referente a uma região cuja população seja maior, dado que haverá um número maior de pessoas que sentirá aquelas variações de preço.

Para que possa haver confiabilidade nos resultados, os institutos de pesquisa econômica disponibilizam os detalhes de suas atribuições de peso para que estudiosos que utilizem esses números possam saber como ele foi obtido.

É interessante notar que existem diversos índices, cada um com características próprias, de acordo com o interesse em se conhecer algum aspecto da economia.

No quadro abaixo, tabela 23, citamos alguns exemplos de índices de preços reais.

Símbolo	Instituição	Descrição sumária
ICV	**Dieese**	**Índice do Custo de Vida** • Variação do custo de vida das famílias com renda de 1 a 30 salários-mínimos. • Município de São Paulo.
IGP-M	**FGV**	**Índice Geral de Preços do Mercado** • Índice de Preços por Atacado (IPA), 60%. • Índice de Preços ao Consumidor (IPC), 30%. • Índice Nacional de Custo da Construção (INCC), 10%.
INCC-M	**FGV**	**Índice Nacional do Custo da Construção – Mercado** • Evolução dos custos de construções habitacionais. • 18 municípios das maiores capitais de estados do país.
INPC	**IBGE**	**Índice Nacional de Preços ao Consumidor** • Pesquisa de preços nas 11 regiões de maior produção econômica. • Pesquisa de Orçamento Familiares (POF), que abrange famílias com renda de 1 a 8 salários-mínimos.
INCTL	**Fipe**	**Índice Nacional da Variação de Custos do Transporte Rodoviário de Carga Lotação** • Custos do transporte rodoviário de carga lotação e de carga fracionada. • A carga é denominada lotação quando a mercadoria é suficiente para lotar um caminhão; nesse caso, a transportadora encaminha um veículo até o depósito, realiza a operação de carregamento e transfere a carga diretamente ao seu destino.

Tabela 23 – Exemplos de índices de preços.

Note haver índices gerais, tais como o IGP-M, e índices específicos, como é o caso do INCTL. A variedade dos índices reflete a própria variedade da economia, pois setores diferentes precisam ter informações detalhadas a respeito de sua área de atuação.

Vale ainda comentar que o Brasil não tem um índice oficial de inflação. Alguns dos comumente utilizados são calculados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), pela Fundação Getúlio Vargas (FGV), pela Fundação Instituto de Pesquisas Econômicas (Fipe) e pelo Departamento Intersindical de Estatística e Estudos Socioeconômicos (Dieese).

Indicadores econômicos

- São geralmente médias ponderadas que refletem alguma característica ou tendência da economia de uma região.
- Os indicadores econômicos servem de parâmetro de compreensão da economia e de aspectos importantes das sociedades e auxiliam na análise de conjuntura para a tomada de decisões.
- Há uma grande variedade de indicadores econômicos. Conhecer os índices principais de sua área de atuação pode auxiliá-lo a tomar decisões mais bem embasadas na realidade.

Saiba mais

Os portais oficiais do Banco Central do Brasil (http://www.bcb.gov.br/) e do Instituto de Pesquisa Econômica Aplicada (http://www.ipea.gov.br/) trazem informações a respeito de diversos índices, bem como seus valores em séries históricas. Você encontrará também informações detalhadas sobre cada índice.

Observação

Utilizar um índice econômico mais indicado para o caso, ao analisar uma determinada situação importante, demonstra competência na área de atuação.

Exemplos de aplicação

1) Extrai-se só uma carta de um baralho de 52 cartas. Determine a probabilidade de obter:

a) Um valete.

b) Uma carta vermelha.

c) Um dez de paus.

d) Uma figura.

e) Uma carta de ouros.

f) Um nove vermelho ou um oito preto.

2) Uma urna tem 20 bolas azuis, 15 bolas vermelhas, 10 bolas laranja e 5 bolas verdes. Misturam-se as bolas e escolhe-se uma aleatoriamente. Determine a probabilidade de a bola escolhida ser:

a) Verde.

b) Azul.

c) Azul ou verde.

d) Não vermelha.

e) Vermelha ou verde.

f) Amarela.

g) Não amarela.

3) Dez fichas são numeradas de 0 a 9 e colocadas em uma urna. Determine a probabilidade de a ficha escolhida ser:

a) O número 3.

b) Um número ímpar.

c) Um número menor do que 4.

4) Uma população consome três marcas de papel: A, B e C. Feita uma pesquisa de mercado, colheram-se os resultados mostrados na tabela abaixo.

A	B	C	A e B	B e C	A e C	A, B e C	Nenhum
129	146	146	75	68	59	23	132

Supondo que esses números sejam significativos para descrever o consumo dessa população, determine a probabilidade de um consumidor:

a) Comprar papel da marca A.

b) Comprar papel da marca B.

c) Comprar papel das três marcas.

d) Comprar somente papel da marca C.

e) Não comprar papel da marca B.

f) Não comprar nenhuma das marcas.

g) Comprar de qualquer das marcas.

h) Comprar de apenas uma marca.

i) Um consumidor da marca A usar também papel da marca B.

j) Um consumidor da marca B usar também papel da marca A.

5) Utilizando o conceito geral de probabilidades e as propriedades apresentadas, calcule as probabilidades solicitadas.

a) A meteorologia prevê que probabilidade de chuva é de 60%. A probabilidade de Marta receber uma ligação de Joana é de 80%. Qual a probabilidade de chover e Joana telefonar?

b) Uma prova de múltipla escolha tem 5 alternativas por questão. Qual a probabilidade de um aluno acertar uma questão que não saiba?

c) Qual a probabilidade de adivinhar em que dia da semana nasceu Adam Smith?

d) Um filme foi bem avaliado por 70% do público. Qual a probabilidade de encontrar na plateia alguém que o avalia mal?

e) Uma revista sorteia brindes para os compradores que enviarem um cupom pelo correio. Apenas 10% dos leitores enviam o cupom. Dos leitores dessa revista, 90% são maiores de 30 anos. Qual a probabilidade de um comprador com menos de 30 anos ser a pessoa sorteada?

Resolução

1) Em todos os casos, temos como espaço amostral as 52 cartas do baralho, ou seja: $n_{tot} = 52$.

a) Há 4 valetes no baralho, logo, P = 4/52, ou P = 0,08 (valor arredondado).

b) Metade das cartas é vermelha, metade é preta, logo, P = 0,5.

c) 1 carta específica, logo, P = 1/52, P = 0,02 (valor arredondado).

d) Cada naipe tem 3 figuras, e num baralho há 4 naipes, perfazendo 12 figuras, logo, P = 12/52, P = 0,23 (valor arredondado).

e) Há 13 cartas de ouros, logo, P = 13/52, P = 0,25.

f) Duas cartas específicas, logo, P = 2/52, P = 0,038.

2) Total de bolas: 50.

a) Há 5 bolas verdes, logo, P = 5/50, P = 0,1.

b) Há 20 bolas azuis, logo, P = 20/50, P = 0,4.

c) Há 5 bolas verdes e 20 bolas azuis, logo, P = 25/50, P = 0,5.

d) Há 15 bolas vermelhas, portanto as demais são não vermelhas e: P = (50-15)/50 = 35/50, P = 0,7.

e) Há 15 bolas vermelhas e 5 verdes, logo, P = 20/50, P = 0,4.

f) Não há bolas amarelas, logo, P = 0/50, P = 0.

g) Não há bolas amarelas, logo, todas são não amarelas e: P = 50/50, P = 1.

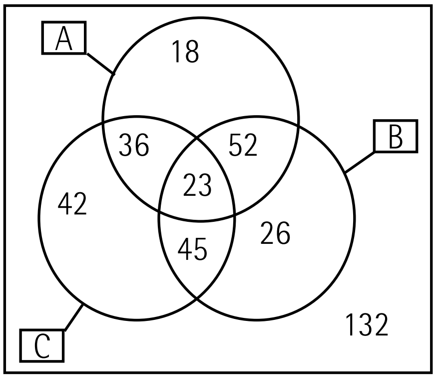
3)

a) 1 ficha em 10, logo, P = 1/10, P = 0,1.

b) Metade dos números é ímpar, logo, P = 0,5.

c) Há 4 números menores que 4, logo, P = 0,4.

4) Para resolver as questões, é preciso montar o diagrama:



Com isso temos o total de compradores, que é de 400. Esse é o espaço amostral referente aos itens a-h. As contas são indicadas exatamente, mas os resultados finais podem estar arredondados.

a) Comprar papel da marca A: P = 129/400, P = 0,32.

b) Comprar papel da marca B: P = 146/400, P = 0,37.

c) Comprar papel das 3 marcas: P = 23/400, P = 0,06.

d) Comprar somente papel da marca C: P = 42/400, P = 0,11.

e) Não comprar papel da marca B: P = 1 - 0,37, P = 63.

f) Não comprar nenhuma das marcas P = 132/400, P = 0,33.

g) Comprar qualquer das marcas: P = 1 - 0,33, P = 0,67.

h) Comprar de apenas uma marca: P = (18 + 26 + 42)/400, P = 0,22.

i) Um consumidor da marca A usar também papel da marca B: P = 75/129, P = 0,58.

j) Um consumidor da marca B usar também papel da marca A: P = 75/146, P = 0,51.

5) Utilizando o conceito geral de probabilidades e as propriedades apresentadas, as probabilidades solicitadas são:

a) Eventos independentes: P = 0,6 * 0,8, P = 0,48.

b) 5 alternativas, apenas 1 certa: P = 0,2.

c) 7 dias na semana, nasceu em um deles: P = 0,14.

d) Eventos mutuamente excludentes: p = 0,7, q = 1 - p, q = 0,3.

e) Envio de cupom e idade são eventos independentes; ter menos de 30 anos é evento complementar de ter mais de 30: P = 0,1 * 0,1, P = 0,01.

Resumo

Tabelas de frequência e seus gráficos

1. Para a organização e análise de dados estatísticos, é interessante verificar com que frequência aparece cada valor, ou intervalo de valores, de uma variável e constroem-se tabelas de frequência.

2. As tabelas de frequência podem trazer valores absolutos (número de vezes que o valor aparece) ou relativos (qual a porcentagem de cada categoria de dados).

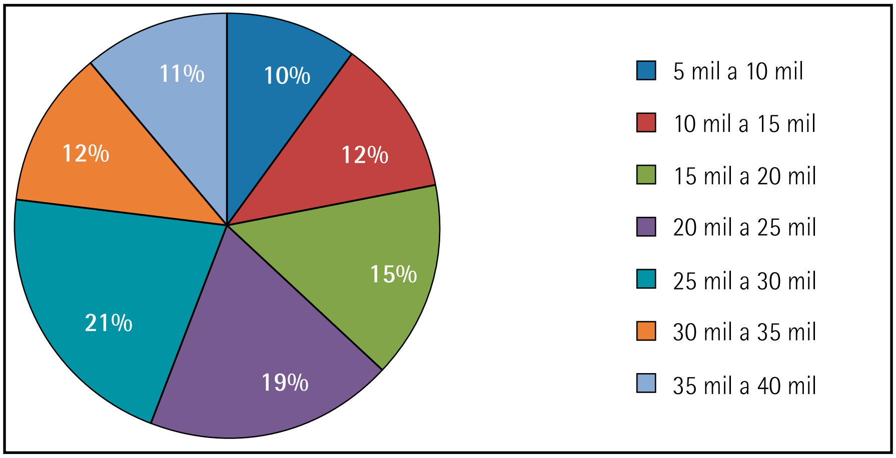
3. Quando queremos apresentar dados estatísticos, de modo a trazer alguma informação de maneira imediata, fazemos isso por meio da utilização de gráficos.

4. Existem diversos tipos de gráfico, e cada um poderá deixar mais claras algumas das características diferentes do conjunto, logo, a escolha do tipo de gráfico a ser utilizado em uma situação depende de qual ou quais características se queira mostrar com maior ênfase.

Exercícios

Questão 1. Uma agência de publicidade relacionou o valor das últimas 119 campanhas publicitárias que realizou até determinada data e representou-as no gráfico a seguir, agrupando os valores em categorias.

Custos das últimas 119 campanhas realizadas



Baseando-se nesses dados, podemos afirmar que o custo médio de uma campanha publicitária dessa agência é de:

A) R$ 22.290,00

B) R$ 22.920,00

C) R$ 23.290,00

D) R$ 24.190,00

E) R$ 21.690,00

Resposta correta: alternativa B.

Análise das alternativas

O gráfico dado apresenta a porcentagem de campanhas para cada categoria de custos. Assim, por exemplo, informa que 10% das campanhas têm custo entre 5 mil e 10 mil reais. Portanto, podemos chegar à conclusão de que 12 campanhas (10% de 119) apresentaram custos entre 5 mil e 10 mil reais (informações retiradas da primeira categoria).

Raciocínio semelhante podemos fazer com as outras categorias, chegando-se ao quadro abaixo, no qual também são feitos cálculos acessórios para a determinação da média:

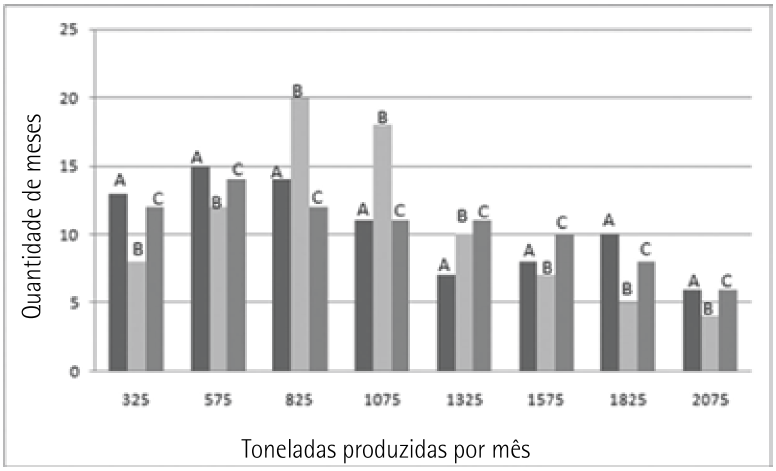
Custo das campanhas publicitárias realizadas (em reais)	Quantidade de campanhas	Ponto médio de classe	Ponto médio vezes frequência simples
5 mil a 10 mil	12	R$ 7.500	R$ 90.000
10 mil a 15 mil	14	R$ 12.500	R$ 175.000
15 mil a 20 mil	18	R$ 17.500	R$ 315.000
20 mil a 25 mil	23	R$ 22.500	R$ 517.500
25 mil a 30 mil	25	R$ 27.500	R$ 687.500
30 mil a 35 mil	14	R$ 32.500	R$ 455.000
35 mil a 40 mil	13	R$ 37.500	R$ 487.500
Somatórios	119		R$ 2.727.500

Cálculo do custo médio**:** $\overline{x} = \dfrac{\sum p_{ix}\, x_i}{\sum p_i} = \dfrac{2.727.500}{119} = 22.920$

A alternativa que corresponde a esse resultado é a B.

Questão 2. O gráfico a seguir representa a produção, em toneladas, atingida ao longo dos meses por uma empresa em suas três linhas de produção (A, B, C).

Produção mensal



Com base nesse gráfico **não** se pode afirmar que:

A) Em cerca de 20 meses, o volume produzido na linha B é de 825 toneladas.

B) A produção de 1.825 toneladas foi atingida no mesmo número de meses pelas linhas A e B.

C) A produção de 1.575 toneladas foi atingida ao longo de um período de 10 meses na linha de produção C.

D) A produção mais frequente do produto A é de 575 toneladas.

E) A produção mais rara para o produto B é de 2.075 toneladas.

Resolução desta questão na Plataforma.