Disciplina: Matemática

TÓPICO 13: Probabilidade e Estatística

Ninguém imaginou que o mundo viveria uma pandemia em 2020. Nesse ano, o novo vírus "SARS-CoV-2" fez algo que pensamos que não seria mais possível: parar o mundo e fazê-lo refletir sobre seu modo de vida. Em um cenário como esse, não sabemos ao certo como a situação estará daqui a um mês, uma semana, quiçá, um dia. No entanto, o que podemos fazer é prever possíveis realidades, resultados. É nesse momento que a Probabilidade e a Estatística entram em ação, que nos permitem entender e lidar com a variabilidade e a incerteza. (Prepara Enem, 2020).

1 Probabilidade

A partir de agora, vamos construir formas de analisar processos que envolvam aleatoriedade e incerteza pela via da matemática.

1.1 Definições

A probabilidade é uma medida da chance de algo acontecer. Toda vez que não temos certeza sobre o resultado de alguma coisa (a qual denominaremos, em breve, de evento), estamos tratando da probabilidade de certos resultados acontecerem—ou quais as chances de eles acontecerem, o que é muito útil na tomada de decisões.

Vamos analisá-la por meio de um exemplo.

Exemplo: Albert Einstein, um estudante do ensino médio, decidiu fazer um experimento que seu professor de matemática lhe pediu. Consiste em pegar uma moeda não viciada e lançá-la para cima uma única vez. Após isso, Albert deve prever a probabilidade de a face de cima ser cara.

Certo, antes mesmo de ajudar Einstein, é intuitivo pensar em dizer que tal "chance" de sair cara é de 50%. E a nossa intuição está certa, mas como traduzir esse raciocínio para a matemática?

Vamos lá.

Como Einstein lançou a moeda uma vez só, o cálculo da probabilidade de sair cara é bastante direto:

$$P\left(cara\right) = \frac{\text{número de formas de obtermos cara}}{\text{número total de resultados possíveis}}$$

Nesse caso,

$$P\left(cara\right) = \frac{1}{2} = 0,5 = 50\%$$

Note que há, num só lançamento de moeda, apenas uma forma de cara ser obtida e há dois resultados possíveis no total, cara ou coroa.

Com base nesse exemplo, podemeos estabelecer algumas definições acerca da Probabilidade:

Espaço Amostral (S): No caso daquele experimento que o jovem Einstein fez, tínhamos dois possíveis resultados: cara e coroa. Esses resultados formam o **Espaço Amostral**, o conjunto de todos os possíveis resultados do experimento (Barbetta, Reis e Bornia, 2010). Esse conjunto será denominado pela letra "S".

FIQUE LIGADO

Quando falamos que S é um conjunto, queremos dizer um conjunto matemático, o mesmo visto na Aula 1 de nosso curso. Veremos a seguir que os eventos também o são. Ou seja, uma vez que são conjuntos matemáticos, aquelas operações, como união, intersecção, complementar e diferença, serão de grande importância e úteis aqui.

Exemplo: São realizados dois experimentos: 1) Lançar de uma moeda e observar a face voltada para cima. 2) Lançar um dado de seis faces e observar a face voltada para cima. Quais os espaços amostrais dos experimentos?

Lembre-se de que o espaço amostral S consiste na enumeração de todos os resultados possíveis de um experimento analisado.

O experimento lançar uma moeda, como visto anteriormente, possui dois resultados possíveis em sua totalidade, cara ou coroa. Então, chamando seu espaço amostral de $S1$, temos:

$$S1 = \{cara, coroa\}$$

Para o experimento, lançar um dado, seis resultados são possíveis no total: 1, 2, 3, 4, 5 e 6. Assim, denominando seu espaço amostral de $S2$, temos

$$S2 = \{1, 2, 3, 4, 5, 6\}$$

Evento (A, B, C...): Muitas vezes estamos interessados na probabilidade de ocorrência de um ou mais resultados do espaço amostral. Esse conjunto de resultados de interesse chamamos de Evento. Por definição, um evento é qualquer subconjunto do espaço amostral. Essa ideia pode ser representada por:

$$A \subseteq S \qquad (1)$$

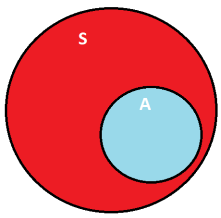


Figura 1: *O evento é um subconjunto do espaço amostral S.*

A relação da Equação 1 nos diz que o nosso subconjunto A (evento) está contido em S (espaço amostral) e pode ter um "tamanho", no máximo, equivalente ao próprio espaço amostral. Depois do exemplo a seguir, veremos que conseguimos retirar algumas importantes propriedades a partir disso.

Exemplo: Seja o experimento do lançamento de um dado. Temos que seu espaço amostral é:

$$S = \{1, 2, 3, 4, 5, 6\}$$

Vamos agora supor que queremos analisar o evento em que a face virada para cima do dado seja um número par. Com isso, nosso subconjunto, ou seja, nosso evento, ficaria:

$$A1 = \{2,\ 4,\ 6\}$$

Se o interesse for nos números ímpares:

$$A2 = \{1,\ 3,\ 5\}$$

Ou números primos, etc. A ideia é pensar que o evento é aquilo que temos interesse em calcular uma probabilidade, o que está a nosso favor. Um evento ocorre quando um dos resultados que o compõe ocorre. E não se esqueça de que os elementos pertencentes ao evento também devem pertencer ao espaço amostral.

Depois de toda essa conversa, podemos fazer a definição clássica de probabilidade.

A probabilidade de ocorrer um evento A de um espaço amostral $S \neq \emptyset$ pode ser representado por:

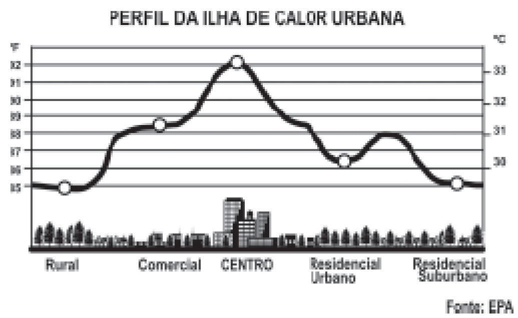
$$P(A) = \frac{n(A)}{n(S)} \qquad (2)$$

Em que $n(A)$ representa o número de resultados do nosso evento A e $n(S)$, o número de resultados possíveis de nosso espaço amostral S.

Também, podemos pensar que a probabilidade é o quociente entre o número de casos favoráveis, que é $n(A)$, e o número de casos possíveis, que é $n(S)$.

Problema 1

(ENEM 2011) Rafael mora no Centro de uma cidade e decidiu se mudar, por recomendações médicas, para uma das regiões: Rural, Comercial, Residencial Urbano ou Residencial Suburbano. A principal recomendação médica foi com as temperaturas das "ilhas de calor" da região, que devem ser inferiores a 31°C. Tais temperaturas são apresentadas no gráfico:



Escolhendo, aleatoriamente, uma das outras regiões para morar, a probabilidade de ele escolher uma região que seja adequada às recomendações médicas é:

a) 1/5
b) 1/4
c) 2/5
d) 3/5
e) 3/4

RESOLUÇÃO: .
Nesse caso, o evento de interesse é "uma região que seja adequada às recomendações médicas", que, de acordo com o enunciado, deve apresentar temperatura abaixo de 31°C.
Assim, tempos que o nosso evento A é dado pelas cidades Rural, Residencial Urbano e Residencial Suburbano, ou seja, $n(A) = 3$. E, de uma maneira geral, Rafael tem, num total, 4 possíveis escolhas de regiões para morar; isto é, $n(S) = 4$.

Assim, aplicando a definição de probabilidade (Equação 2):

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{4} = 75\%$$

Logo, a probabilidade de Rafael escolher uma região que seja adequada às recomendações médicas é de 3/4 (e).

Problema 2

(ENEM 2015) Em uma central de atendimento, cem pessoas receberam senhas numeradas de 1 até 100. Uma das senhas é sorteada ao acaso. Qual a probabilidade de a senha sorteada ser um número de 1 a 20.

a) 1/100
b) 19/100
c) 20/100
d) 21/100
e) 80/100

RESOLUÇÃO: .
Novamente, nesse exemplo, basta aplicar a definição de probabilidade (Equação 2), em que $n(S) = 100$ e $n(A) = 20$ (pois são 20 números favoráveis ao nosso objetivo). Assim, :

$$P(A) = \frac{20}{100}$$

Portanto, (c).

1.2 Propriedades básicas da probabilidade

Agora analisaremos propriedades básicas, que são de extrema importância.

Consideremos $S \neq \emptyset$, um espaço amostral não vazio, e A e B dois eventos quaisquer contidos em S. Então, são válidas as seguintes propriedades:

a) A probabilidade de qualquer evento está entre 0 e 1 (0% a 100%). Ou seja, se no momento do cálculo de probabilidade, encontrarmos valores negativos ou maiores do que 1, cometemos algum erro.

$$0 \leq P(A) \leq 1$$

b) A probabilidade de um evento qualquer A ocorrer será zero se, e somente se, tal evento for o conjunto vazio.

$$P(A) = 0 \leftrightarrow A = \emptyset$$

c) A probabilidade de um evento qualquer A ocorrer será um se, e somente se, tal subconjunto for o próprio espaço amostral S.

$$P(A) = 1 \leftrightarrow A = S$$

Com base na teoria dos conjuntos, conseguimos extrair importantes propriedades também:

d) Evento Complementar de A ($\bar{A}$ ou A')): Como $\bar{A}$ significa "não-A", ou seja, tudo aquilo que não pertença ao conjunto A, podemos inferir a seguinte ideia a partir da Figura 2:

$$P(A) + P(\bar{A}) = 1 \therefore$$
$$P(\bar{A}) = 1 - P(A) \qquad (3)$$

A Equação 3 é muito utilizada, pois, muitas vezes, é mais fácil calcular a probabilidade o complementar de um evento ($\bar{A}$) do que determinar a probabilidade o evento em si (A).

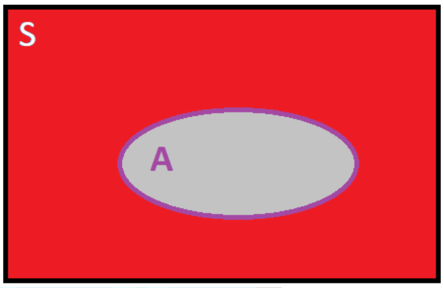


Figura 2: *A região destacada em vermelho representa o complementar de A, ou seja, $\bar{A}$.*

FIQUE LIGADO

Diagramas como o da Figura 2 são conhecidos por "Diagramas de Venn". Nesses diagramas, o espaço amostral é representado pelo retângulo. Os eventos são representados por círculos, elipses ou outras formas inseridas no retângulo. Esses diagramas são muito úteis na resolução de problemas envolvendo probabilidade.

e) União de dois eventos ($A \cup B$): Com essa propriedade, podemos estabelecer a regra da soma das probabilidades. A probabilidade de ocorrer A ou B (lembre-se de que a ideia de união pode ser representada pelo conectivo "ou") pode ser dada por:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (4)$$

Seria bastante intuitivo se a probabilidade de A ou B ocorrerem fosse apenas $P(A) + P(B)$. Para visualizar essa ideia nem tanto intuitiva, observe a Figura 3.

Note que, quando queremos determinar a probabilidade de A ou B ocorrer, claramente, devemos somar $P(A)$ com $P(B)$. Entretanto, ao fazer isso, estaríamos somando uma região duas vezes: a intersecção entre

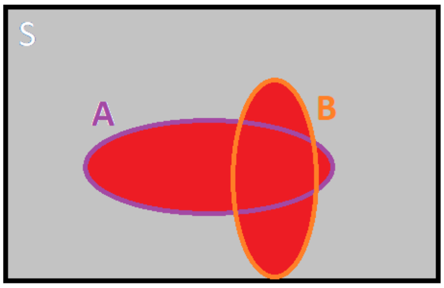


Figura 3: *União de dois eventos A e B contidos num espaço amostral S. Representada pela região destacada em vermelho.*

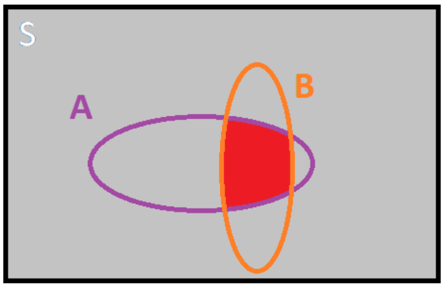


Figura 4: *Região de intersecção $A \cap B$ representada pela região em vermelho.*

os dois eventos. Por isso, é necessário a subtração do termo $P(A \cap B)$.

FIQUE LIGADO

Uma pergunta pode vir à tona: será que sempre haverá essa região de intersecção entre eventos? A resposta é não. Há casos que dois eventos não possuem intersecção alguma. Nesses, podemos dizes que A e B são **eventos mutuamente exclusivos**, como mostrado na Figura 5. Matematicamente, essa ideia pode ser representada por $A \cap B = \emptyset$. Nesses casos, Já que $P(A \cap B) = 0$ a regra da soma das probabilidades pode ser resumida em:

$$P(A \cup B) = P(A) + P(B) \tag{5}$$

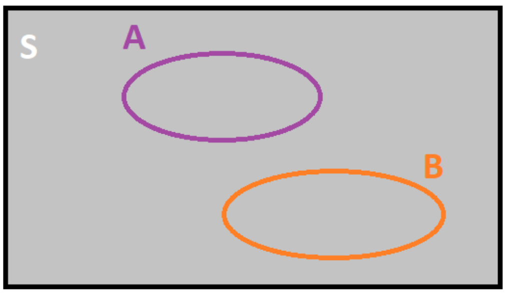


Figura 5: *Note que os eventos A e B não possuem interseção entre si. Por isso, são mutuamente exclusivos.*

Nessa perspectiva acima, um outro detalhe importante pode ser notado. Se os eventos A e B forem mutuamente exclusivos (Equação 5) e a união deles resultar no próprio espaço amostral ($A \cup B = S$), como ilustrado na Figura 6), então A e B, além de mutuamente exclusivos, são **eventos exaustivos**. Consequentemente,

$$P(A \cup B) = P(A) + P(B) = 1 \tag{6}$$

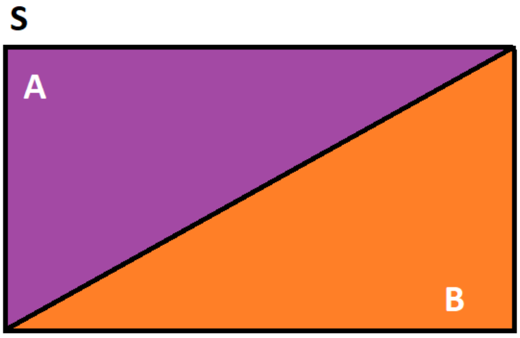


Figura 6: *Eventos A e B, nesse caso, são mutuamente exclusivos (não têm intersecção entre si) e também exaustivos (união de ambos resulta no espaço amostral).*

FIQUE LIGADO

A definição de eventos exaustivos é apenas que a união deles resulta no espaço amostral. Mas eventos exaustivos não são necessariamente mutuamente exclusivos, pois eles podem ter interseção. Considere o lançamento de um dado. Os eventos $A = \{4, 5, 6\}$ e $B = \{1, 2, 3, 4\}$ são exaustivos, mas não mutuamente exclusivos.

Estendendo para n eventos ($A1$, $A2$, $A3$, ..., An) contidos em um espaço amostral S e considerando que $A1 \cup A2 \cup A3 \cup ... \cup An = S$, logo,

$$P(A1 \cup A2 \cup A3 \ldots \cup An) = P(S) = 1 \tag{7}$$

Ademais, se analisando os eventos dois a dois, não identificarmos interseções, os eventos, além de exaustivos, serão mutuamente exclusivos. Para visualizarmos de maneira mais clara esse espaço amostral genérico de n eventos exaustivos e mutuamente exclusivos, vejamos a Figura 7.

FIQUE LIGADO

Cuidado com a sutil diferença entre as Equações 5 e 6. Na equação 5, não sabemos o seu resultado, apenas que os eventos são mutuamente exclusivos. Já, pela Equações 6, sabemos seu valor, pois, além de ser mutuamente exclusivos, os eventos são exaustivos também.

Agora, veremos alguns exemplos sobre o que foi visto até o momento para fixarmos as ideias.

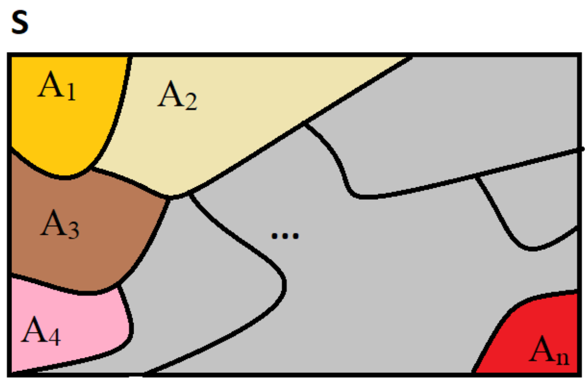


Figura 7: *Espaço amostral S formado por n eventos ($A1$, $A2$, $A3$, ..., An). Note que não a união de todos os eventos resulta em S (eventos exaustivos). Também não há intersecção a cada dois eventos analisados (eventos mutuamente exclusivos).*

Problema 3

A análise da série histórica de pluviometria dos últimos 20 anos, em uma praia do litoral cearense, mostrou que a probabilidade de chover 5 ou mais dias no mês de outubro é de 33% e a probabilidade de chover 5 ou menos dias nesse mesmo mês é de 81%. Qual é a probabilidade de chover exatamente 5 dias, nessa praia, no próximo mês de outubro? [3]

RESOLUÇÃO: .
Esse é um caso em que apenas a aplicação da definição de união de dois eventos (Equação 4) nos ajuda. O que pode ser trabalhoso ao se trabalhar com tal equação é saber se o termo $P(A \cap B)$ é ou não igual a zero. Vamos analisar, então, os eventos!
Temos dois eventos que definem todo o nosso espaço amostral: chover 5 ou mais dias no mês de outubro (A) e 5 ou menos dias nesse mesmo mês (B); e a intersecção ($A \cap B$) corresponde a chover em exatamente 5 dias. Matematicamente, organizando as informações, temos:

$$A = \ \{5,\ 6,\ 7,\ ...,\ 31\}\ ;\ P\,(A) = 0,33$$
$$B = \{0,\ 1,\ 2,\ 3,\ 4,\ 5\}\ ;\ P\,(B) = 0,81$$
$$A\ \cap\ B = \ \{5\}\ ;\ P\,(A\ \cap\ B) = ?$$
$$A \cup B = S = \{0,\ 1,\ 2,\ ...,\ 31\}\ ; P\,(A\ \cup\ B) = 1$$

Assim, aplicando a Equação 4, tem-se:

$$P\,(A \cup B) = P(A) + P\,(B) - P(A \cap B)$$
$$P(A \cap B) = P(A) + P\,(B) - P\,(A \cup B)$$
$$P\,(A \cap B) = 0,33 + 0,81 - 1 = 0,14$$

Logo, a probabilidade de chover exatamente 5 dias é de 14%.

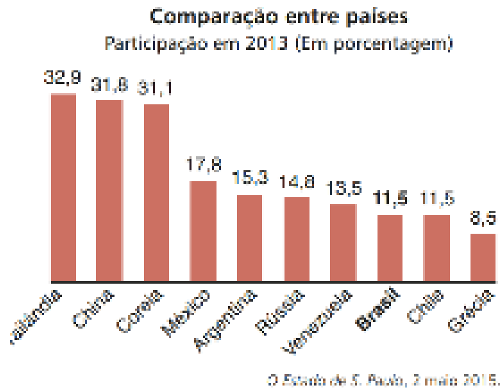
FIQUE LIGADO

Você já deve ter percebido que podemos expressar uma probabilidade em termos de valores entre 0 e 1 ou em termos percentuais (entre 0 e 100%). Por exemplo: Uma probabilidade de 0,14 é equivalente a uma probabilidade de 14%. No entanto, lembre-se que todos os cálculos devem sempre ser feitos usando valores entre 0 e 1.

Problema 4

O gráfico abaixo compara a participação da indústria de transformação no PIB de alguns países (Iezzi et al., 2016). Sorteando-se ao acaso dois países dessa relação, qual é a probabilidade de que:

a) ambos tenham percentual de participação no PIB menor que 15%?
b) ao menos um dos países selecionados tenha participação percentual maior que 20%?



RESOLUÇÃO: Esse exemplo é de muita importância, pois nos mostra que conceitos vistos em análise combinatória, tais como princípio fundamental da contagem, permutação, combinação e arranjo também podem ser utilizados como ferramentas para o cálculo de probabilidades. Vamos lá!
Para resolvermos o item (a), necessitaremos de auxílio de nosso amigo p.f.c. (princípio fundamental da contagem). Vamos sortear dois países ao caso, obtendo

$$\underbrace{10}_{1^{\text{a}}\text{ escolha}} \times \underbrace{9}_{2^{\text{a}}\text{ escolha}} = 90 \text{ possibilidades}$$

Assim, a quantidade de resultados no espaço amostral é $n(S) = 90$. O nosso evento de interesse A são países que tenham um percentual de participação menor que 15%. Cuidado ao dizer que $n(A) = 5$, pois, primeiramente, devemos realizar o sorteio aleatório. Ou seja, de forma análoga,

$$\underbrace{5}_{1^{\text{a}}\text{ escolha}} \times \underbrace{4}_{2^{\text{a}}\text{ escolha}} = 20 \text{ possibilidades}$$

Logo, a quantidade de resultados no evento A é $n(A) = 20$. Aplicando a Equação 2, obtemos

$$P(A) = \frac{20}{90} = \frac{2}{9}$$

Portanto, a probabilidade pedida é de 2/9 (22,2%).
Note que poderíamos ter chegado aos resultados de $n(A)$ e de $n(S)$ por meio da combinação simples, visto que a ordem dos elementos no sorteio não importa. Como sugestão de estudo, tente fazer por esse caminho.
Para o item (b), a nossa amiga análise combinatória será necessária novamente. Primeiramente, devemos nos atentar ao que o item quer: que ao menos um dos dois países tenha uma participação superior a 20%. Temos 3 países que tem participação superior a 20% e 7 que não.
Analisando o nosso caso favorável, podemos ter 1 ou 2 países com participação acima de 20%. A partir da análise desses dois casos, conseguimos obter $n(A)$, enquanto $n(S)$ permanece o mesmo do item anterior, uma vez que o espaço amostral não foi alterado. Logo,

- Com 1 maior e 1 menor, temos 21 casos favoráveis (3×7=21), por meio do p.f.c.;
- Com 2 maiores, temos que

$$C_{3,\,2} = \frac{3!}{2!\,(3-2)!} = 3 \text{ casos favoráveis}$$

Assim, no total, temos 24 casos favoráveis, ou seja, $n(A) = 24$. Pela Equação 2, obtemos

$$P(A) = \frac{24}{45} = \frac{8}{15}$$

Portanto, a probabilidade é de 8/15 (53,3%).
Há uma segunda forma de se resolver esse item. Poderíamos calcular primeiro a probabilidade de os dois países sorteados terem uma participação menor que 20%:

$$P(A) = \frac{C_{7,\,2}}{45} = \frac{21}{45} = \frac{7}{15}$$

Logo, o que queremos é o complementar desse valor, isto é, $1 - P(A)$, resultando novamente em 8/15.

FIQUE LIGADO

Esse exemplo nos mostrou, com bastante ênfase, a importância da análise combinatória na determinação da probabilidade. Acostume-se a isso, esse "crossover" entre tais temas é bem comum.

1.3 Resultados igualmente prováveis

Antes de entrarmos num dos principais ramos da probabilidade, é importante nos atentar a um tipo de espaço amostral, o espaço amostral equiprovável.

Consideremos um espaço amostral S finito composto por a_n eventos unitários (ou resultados):

$$S = a_1,\ a_2,\ a_3,\ \ldots,\ a_n$$

Um espaço amostral S é equiprovável se todos os eventos unitários possuem a mesma chance de ocorrer.

Vamos pensar em um exemplo bem simples para ilustrar essa ideia.

Num lançamento de um dado não viciado, qual é a chance de virar o número 1?

Intuitivamente, você deve pensar 1/6, e está certo. E se eu lhe pedisse a chance do 2, 3, 4, etc.? A resposta continuaria a mesma: 1/6. Ou seja, cada face tem a mesma chance de ocorrer, $P(1) = P(2) = \ldots = P(6)$.

Com isso, retornando ao nosso espaço equiprovável S, podemos matematicamente enunciar que:

$$P(a_1) = P(a_2) = \ldots = P(a_n) = \frac{1}{n}$$

Como são unitários,

$$P(a_1) + P(a_2) + \ldots + P(a_n) = 1$$

Note que, no caso do lançamento do dado, teríamos:

$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$

Agora, vamos considerar um evento A formado por r resultados igualmente prováveis, em que $r \leq n$, ou seja, $A = a_1, a_2, a_3, \ldots, a_r$.

Assim,

$$P(A) = \frac{n(A)}{n(S)} = \frac{r}{n} \tag{8}$$

Problema 5

De um baralho comum, com 52 cartas, extraímos, ao acaso, uma carta. Qual é a probabilidade de não ser sair um ás? (Iezzi et al., 2016)

RESOLUÇÃO:
Inicialmente, vemos que o nosso espaço amostral (conjunto com 52 cartas) é equiprovável,

visto que a probabilidade de se pegar qualquer uma carta é a mesma: 1/52.
Sabe-se, também, que o evento de interesse é "não retirar uma carta que é um ás". Nesse caso, é mais fácil calcular a probabilidade de interesse por meio da propriedade complementar. A probabilidade de retirar uma carta de ás é dada por

$$P\left(A\right)=\frac{n(A)}{n(S)}=\frac{4}{52}=\frac{1}{13}$$

Assim, a probabilidade de interesse pode ser obtida obtida pela Equação 3

$$P\left(\bar{A}\right)=1-P\left(A\right)=1-\frac{1}{13}=\frac{12}{13}$$

Logo, a probabilidade é de 12/13 ou, aproximadamente, 92,3%.

1.4 Probabilidade Condicional

Agora vamos estudar um dos principais temas envolvendo a probabilidade. Em muitas questões de vestibulares a ideia de probabilidade condicional é utilizada. Então, vamos lá!

Muitas vezes, temos interesse em calcular a probabilidade de ocorrência de um evento A dada a ocorrência de um evento B, como na reflexão a seguir:

Qual a probabilidade de chover amanhã em Joinville, sabendo que choveu hoje?

Perceba que, para o cálculo da probabilidade acima, estamos "limitados", "condicionados" a uma situação. Como um joinvilense, você deve pensar "100% de certeza". Brincadeira à parte, como podemos traduzir essa ideia para a matemática?

Dados dois eventos A e B de um espaço amostral S não vazio, chamamos a probabilidade de ocorrer A dado que B ocorreu ou ocorrerá de $P(A|B)$. Com essa ideia, podemos ilustrar o exemplo a cima como:

P(chover amanhã em Jvlle | choveu hoje em Jlle)

O símbolo " | " é lido como "dado que".
$P(A|B)$ pode ser definido como:

$$P\left(A|B\right)=\frac{n\left(A\cap B\right)}{n\left(B\right)}$$

Dividindo tanto o numerador quanto o denominador pela quantidade de elementos do espaço amostral S, isto é,

$$P\left(A|B\right)=\frac{\frac{n(A\cap B)}{n(S)}}{\frac{n(B)}{n(S)}}$$

Obtemos:

$$P\left(A|B\right)=\frac{P\left(A\cap B\right)}{P\left(B\right)} \tag{9}$$

A Equação 9 é uma ferramenta extremamente poderosa, pois, escrita de uma outra maneira, ela nos fornece a Regra do Produto. Se isolarmos o termo $P(A\cap B)$, temos

$$P\left(A\cap B\right)=P\left(A|B\right)\cdot P\left(B\right) \tag{10}$$

A Equação 10 é conhecida como **Regra do Produto**. Ela permite calcular a probabilidade de ocorrência de eventos simultâneos (intersecção).

FIQUE LIGADO

E se a condição fosse B dado A? Analogamente à Equação 9, teríamos

$$P\left(B|A\right)=\frac{P\left(B\cap A\right)}{P\left(A\right)}$$

Mas $P\left(B\cap A\right)=P(A\cap B)$ Assim,

$$P\left(A\cap B\right)=P(B|A).P(A) \tag{11}$$

Observe que fato interessante. Apesar da comutatividade em relação à intersecção, dependendo da condição imposta ao cálculo da probabilidade (A dado B ou B dado A), há diferença na determinação de $P=(A\cap B)$, fato observado ao compararmos as Equações 10 e 11, isto é, $P(A|B)\neq P(B|A)$.

1.5 Eventos independentes

Dois eventos são independentes quando a ocorrência de um não influencia a probabilidade de ocorrência do outro, ou seja, matematicamente,

$$P\left(A|B\right)=P\left(A\right)$$

$$P\left(B|A\right)=P(B)$$

Assim, a regra do produto pode ser simplificada para

$$P\left(A\cap B\right)=P(A)\cdot P(B) \tag{12}$$

A Equação 12 é válida se, e somente se, A e B forem eventos independentes.

Depois dessa conversa, vamos fazer alguns exemplos sobre o que acabamos de conhecer. Vamos lá!

Problema 6

(ENEM – 2019) Em um determinado ano, os computadores da receita federal de um país identificaram como inconsistentes 20% das declarações de imposto de renda que lhe foram encaminhadas. Uma declaração é classificada

como inconsistente quando apresenta algum tipo de erro ou conflito nas informações prestadas. Essas declarações consideradas inconsistentes foram analisadas pelos auditores, que constataram que 25% delas eram fraudulentas. Constatou-se ainda que, dentre as declarações que não apresentaram inconsistências, 6,25% eram fraudulentas.
Qual é a probabilidade de, nesse ano, a declaração de um contribuinte ser considerada inconsistente, dado que ela era fraudulenta?

a) 0,0500
b) 0,1000
c) 0,1125
d) 0,3125
e) 0,5000

RESOLUÇÃO: Perceba que se trata de uma probabilidade condicional. A primeira coisa recomendada a se fazer é transcrever as informações do texto para uma linguagem matemática. Assim, temos:

$$P\,(\text{inconsistente}) = P(I) = 20\% = 0,2$$

$$P\,(\text{não inconsistente}) = P(\bar{I}) = 80\% = 0,8$$

Ademais, dentre as declarações "inconsistentes", 25% das declarações são fraudulosas. Dentre as "não inconsistentes", 6,25% são fraudulentas.

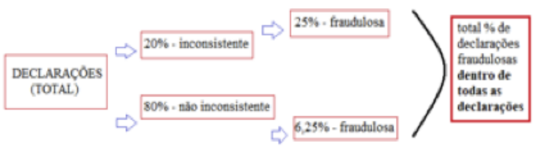
$$P\,(F|\,I) = 25\% = 0,25$$

$$P\,\left(F|\bar{I}\right) = 6,25\% = 0,0625$$

Logo, a probabilidade de qualquer declaração ser fraudulosa, $P(F)$, será dada pela soma das proporções de declarações fraudulentas encontradas naqueles dois conjuntos, ou seja,

$$P(F) = \underbrace{0,20 \cdot 0,25}_{\text{parcela em } I} + \underbrace{0,80 \cdot 0,0625}_{\text{parcela em} \bar{I}} = 0,10$$

Para esclarecer tal raciocínio, veja a ilustração abaixo. Perceba que a quantidade total de declarações fraudulosas não é somente somar 25%. com 6,25%.



O queremos é a probabilidade de uma declaração ser inconsistente sabendo que é fraudulenta, ou seja, $P(I|F)$. Para tanto, devemos fazer algumas manipulações algébricas com as equações de $P(I|F)$ e $P(F|I)$. Pela definição (Equação 9):

$$P\,(I|F) = \frac{P\,(I \cap F)}{P\,(F)}$$

$$P\,(F|I) = \frac{P\,(I \cap F)}{P\,(I)}$$

Isolando $P(I \cap F)$ nas duas equações, igualando-as e isolando $P(I|F)$:

$$P\,(I|F) = \frac{P\,(F|I) \cdot P\,(I)}{P\,(F)}$$

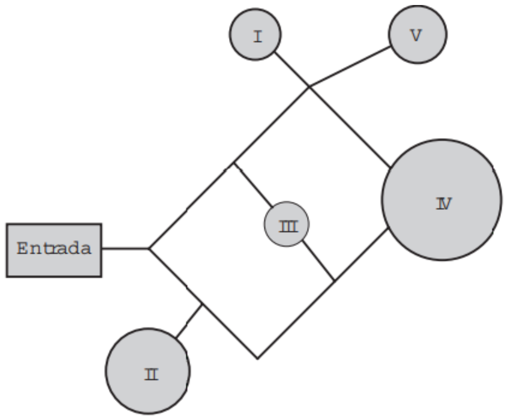
Substituindo os valores:

$$P\,(I|F) = \frac{0,25 \cdot 0,20}{0,1} = 0,50 = 50\%$$

Enfim, chegamos ao resultado (e).

Problema 7

(ENEM 2016) Um adolescente vai a um parque de diversões tendo, prioritariamente, o desejo de ir a um brinquedo que se encontra na área IV, dentre as áreas I, II, III, IV e V existentes. O esquema ilustra o mapa do parque, com a localização da entrada, das cinco áreas com os brinquedos disponíveis e dos possíveis caminhos para se chegar a cada área. O adolescente não tem conhecimento do mapa do parque e decide ir caminhando da entrada até chegar à área IV.

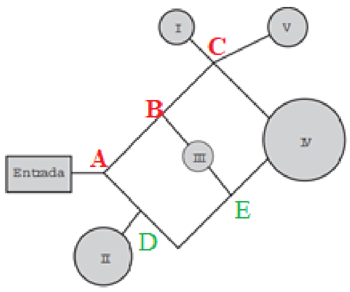


Suponha que relativamente a cada ramificação, as opções existentes de percurso pelos caminhos apresentem iguais probabilidades de escolha, que a caminhada foi feita escolhendo ao acaso os caminhos existentes e que, ao tomar um caminho que chegue a uma área distinta

da IV, o adolescente necessariamente passa por ela ou retorna.
Nessas condições, a probabilidade de ele chegar à área IV sem passar por outras áreas e sem retornar é igual a:

a) 1/96
b) 1/64
c) 5/24
d) 1/4
e) 5/12

RESOLUÇÃO:
Nesse problema, devemos calcular a probabilidade de tal jovem chegar à área IV. Sem passar por outras regiões e sem retornar, temos dois caminhos possíveis, onde as letras representam as bifurcações:



Passando por A-B-C ou A-D-E. Sabemos que as opções existentes de percurso pelos caminhos apresentem iguais probabilidades de escolha e que ele comece pela entrada, vamos calcular essas duas probabilidades.
Caminho 1: A-B-C

- Bifurcação A (2 opções): $P(A) = \frac{1}{2}$
- Bifurcação B (2 opções): $P(B) = \frac{1}{2}$
- Bifurcação C (3 opções): $P(C) = \frac{1}{3}$

Note que, para chegar à IV por esse jeito, o jovem tem que passar por A e por B e por C. Como comentado anteriormente nessa aula, os conectivos "ou" e "e" significam, respectivamente, união e intersecção em teoria dos conjuntos. Assim, vamos trabalhar com a ideia de intersecção. Como em uma dada bifurcação, a probabilidade de escolher um caminho é sempre a mesma, independente da escolha que fez no cruzamento anterior, trata-se de eventos independentes. Logo, com base na Equação 12, podemos calcular a probabilidade desse caminho como:

$$P\,(\text{Cam. } 1) = P\,(A \cap B \cap C) = P\,(A) {\cdot} P\,(B) {\cdot} P\,(C)$$

$$P(\text{Cam. } 1) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{12}$$

Caminho 2: A-D-E

- Bifurcação A (2 opções): $P(A) = \frac{1}{2}$
- Bifurcação D (2 opções): $P(D) = \frac{1}{2}$
- Bifurcação E (2 opções): $P(E) = \frac{1}{2}$

Assim, de maneira semelhante ao primeiro caso,

$$P\,(\text{Cam. } 2) = P\,(A \cap D \cap E) = P\,(A) {\cdot} P\,(D) {\cdot} P\,(E)$$

$$P(\text{Cam. } 2) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

O próximo passo é calcular a probabilidade de ele ir pelo caminho 1 ou pelo caminho 2. Note que agora devemos trabalhar com a ideia de união. Perceba que esses dois caminhos são mutuamente exclusivos, ou seja, não possuem intersecção. O que faz sentido, visto que seria fisicamente impossível desse jovem fazer os dois caminhos ao mesmo tempo. Com essa ideia, temos:

$$P\,(\text{Cam. } 1 \cup \text{Cam. } 2) =$$

$$P\,(\text{Cam. } 1) + P\,(\text{Cam. } 2) - P(\text{Cam. } 1 \cap \text{Cam. } 2)$$

Mas $P\,(\text{Cam. } 1 \cap \text{Cam. } 2) = 0$ (mutuamente exclusivos). Assim,

$$P\,(\text{Cam. } 1 \cup \text{Cam. } 2) = P\,(\text{Cam. } 1) + P\,(\text{Cam. } 2)$$

$$P\,(\text{Cam. } 1 \cup \text{Cam. } 2) = \frac{1}{2} + \frac{1}{8} = \frac{5}{24}$$

Portanto, a probabilidade de o jovem chegar à área IV é de 5/24 (c).

FIQUE LIGADO

É muitíssimo recorrente trabalharmos com as equações de união e intersecção de eventos em probabilidade duma maneira geral. Saber identificar se um problema pede intersecção ou união de eventos é bem importante. Um meio de auxílio é notar se o problema quer a probabilidade de A **ou** B ocorrer (união) ou de A **e** B ocorrer (interseção). Acostume-se também às definições de eventos independentes e mutuamente exclusivos, pois eles facilitam muitos cálculos, como pôde ser visto nesse último exemplo.
Faça bastantes exercícios sobre probabilidade para que essas ideias estejam bem fixadas.

2 Estatística

Nesse momento, vamos entrar na segunda e última parte de nossa aula. Estudaremos noções de **Estatística**. Essa área da matemática trata da **coleta**, **apresentação**, **análise** e **uso** de **dados** para que possamos tomar decisões, resolver problemas, planejar processos e produtos, etc. (Barbetta, Reis e Bornia, 2010)

Aqui vamos nos concentrar em técnicas para organização e apresentação e interpretação de dados em tabelas e gráficos e no cálculo de alumas medidas que resumem um conjunto de dados.

2.1 Distribuição de frequências

Para extrairmos o máximo de informação dos dados coletados é muito útil resumí-los na forma de tabelas e gráficos de distribuição de frequências.

Para melhor explicar, vamos ver um exemplo.

Exemplo: Marie Curie, dona de uma nova pizzaria em Joinville, decidiu fazer um levantamento de pizzas vendidas por dia ao longo de 1 mês (30 dias). Ou seja, a **variável** de interesse é a quantidade de pizzas vendidas por dia e os **dados** são os valores dessa varável que Marie Curie coletou ao longo de 1 mês. Vamos chamar essa variável (quantidade de pizzas vendidas por dia) de PD Abaixo, consta o conjunto de dados obtidos pela empresária:

FIQUE LIGADO

Uma **variável** é uma uma característica de interesse. Uma variável pode ser **quantitativa**, com dados representados por números, como a quantidade de pizzas vendidas ou a altura de estudantes de uma escola ou **qualitativas**, com dados não-numéricos, como o sabor de pizza preferido de cada cliente da pizzaria ou a cor dos cabelos dos estudantes de uma escola.

$$PD = \{25, 46, 29, 36, 37, 32, 46, 32, 43, 40, 50, 29, 27, 32, 35,$$
$$39, 48, 32, 48, 43, 53, 41, 29, 39, 44, 52, 44, 54, 40, 55\}$$

Para dados quantitativos, um primeiro passo que sempre é muito útil, é **ordenar** dos dados (de forma crescente ou decrescente). O conjunto ordenado de forma crescente fica:

$$PD = \{25, 27, 29, 29, 29, 32, 32, 32, 32, 35, 36, 37, 39, 39$$
$$40, 40, 41, 43, 43, 44, 44, 46, 46, 48, 48, 50, 52, 53, 54, 55\}$$

Agora, com os dados organizados, fica mais fácil de se construir uma poderosa ferramenta de análise, a **tabela de frequências**. Vamos lá!

2.1.1 Tabela de frequências

Trata-se de uma tabela que organiza e resume o conjunto de dados coletados de uma pesquisa (Iezzi et al., 2016).

Numa tabela de frequências podemos incluir as seguintes informações.

1. **Valor** x_i: corresponde ao i-ésimo valor que queremos incluir em nossa tabela de frequências, sendo que esses valores devem estar ordenados do menor para o maior se a variável for quantitativa. Em nosso exemplo, vamos incluir todos os valores diferentes do nosso conjunto de dados. Veja que, temos 19 valores diferentes. Então, o valor x_1 é 25, o valor x_2 é 27, o valor x_3 é 29, o valor x_4 é 32 e assim por diante.
2. **Frequência absoluta** f_i: corresponde ao número de vezes que um determinado valor x_i da variável aparece no conjunto de dados.
3. **Frequência relativa** f_r: a razão entre f_i e o número total de dados (n).

$$f_r = \frac{f_i}{n} \tag{13}$$

Também, podemos multiplicar f_r por 100%, obtendo, assim, a **Frequência relativa percentual** ($f_{r\%}$):

$$f_{r\%} = f_r \cdot 100\% \tag{14}$$

4. **Frequência absoluta acumulada** (F_i): soma da frequência absoluta do valor x_i da variável com todas as frequências absolutas anteriores.
5. **Frequência relativa acumulada** (F_r): soma da frequência relativa de um valor x_i da variável com todas as frequências relativas anteriores (o mesmo processo feito na absoluta acumulada). Também, podemos fazer a mesma coisa com a relativa percentual, resultando na Frequência Relativa Percentual Acumulada ($F_{r\%}$).

Agora estamos prontos para montarmos nossa tabela de frequências. Veja a Tabela 1.

Tabela 1: *Tabela de frequências da venda diária de pizzas.*

x_i	f_i	f_r	$f_{r\%}$	F_i	$F_{r\%}$
25	1	1/30	3,34%	$0+1=1$	3,34%
27	1	1/30	3,34%	$1+1=2$	6,68%
29	3	3/30	10%	$2+3=5$	16,68%
32	4	4/30	13,34%	$5+4=9$	30,02%
35	1	1/30	3,34%	$9+1=10$	33,36%
36	1	1/30	3,34%	$10+1=11$	36,7%
37	1	1/30	3,34%	$11+1=12$	40,04%
39	2	2/30	6,67%	$12+2=14$	40,71%
40	2	2/30	6,67%	$14+2=16$	53,38%
41	1	1/30	3,34%	$16+1=17$	56,72%
43	2	2/30	6,67%	$17+2=19$	63,39%
44	2	2/30	6,67%	$19+2=21$	70,06%
46	2	2/30	6,67%	$21+2=23$	76,73%
48	2	2/30	6,67%	$23+2=25$	83,4%
50	1	1/30	3,34%	$25+1=26$	86,74%
52	1	1/30	3,34%	$26+1=27$	90,06%
53	1	1/30	3,34%	$27+1=28$	93,42%
54	1	1/30	3,34%	$28+1=29$	96,76%
55	1	1/30	3,34%	$29+1=30$	100%
$\sum$	30	1	100%		

Note que, ao final, a frequência relativa percentual acumulada deve resultar em 100%, e a frequência absoluta acumulada em n (número total de dados). Olhando para a tabela, vemos rapidamente 29 e 32 foram as quantidades de pizza mais frequentes.

E se Marie Curie quisesse analisar suas vendas ao longo de 6 meses? Imagine a quantidade de valores diferentes que teríamos. Seria bem trabalhoso montar uma tabela de frequências. Além disso, talvez não conseguíssemos mais distinguir tão vem valores mais frequentes. Nesses casos, é uma boa prática organizar os dados em intervalos, os quais denominaremos de **classes**.

A construção de uma tabela de frequências usando classes é idêntica à que é feita com valores individuais, como fizemos anteriormente. Para mostrar isso, veremos um breve exemplo.

Exemplo: Considere uma amostra de 100 alunos da UFSC Joinville. O nosso objetivo é analisar as alturas dos estudantes. Após a coleta dos dados, vamos condensá-los em classes para que o nosso trabalho fique mais "leve". Vejamos a tabela abaixo.

Tabela 2: *Altura alunos UFSC - JOI*

Classes [altura (m)]	f	f_r	$f_r\%$
$1,50 \vdash 1,58$	5	0,05	5%
$1,58 \vdash 1,66$	18	0,18	18%
$1,66 \vdash 1,74$	42	0,42	42%
$1,74 \vdash 1,82$	27	0,27	27%
$1,82 \vdash 1,90$	8	0,08	8%

Note que os "tamanhos" de todas as classes são iguais. Esse tamanho corresponde à **amplitude** de cada classe.

Amplitude (A) é a diferença entre o maior e o menor elemento de um conjunto. Nesse exemplo, para uma classe, seria a diferença entre os limites do intervalo. Assim, para a primeira, temos:

$$A = 1,58 - 1,50 = 0,08$$

FIQUE LIGADO

A simbologia da classe "$\vdash$" é bastante importante de ser entendida. Quando dizemos $1,51 \vdash 1,58$, significa que nessa classe estão dados que estão todos os dados que são iguais ou maiores que 1,51 e menores que 1,58. Se, no exemplo, houvesse um aluno cuja altura é 1,58m, ele seria posto dento da classe $1,58 \vdash 1,66$.

Uma vez tendo tabelas de distribuições prontas, podemos fazer **gráficos**!

2.1.2 Representações gráficas de distribuição de frequências

O objetivo dos gráficos é apresentar as informações acerca dos dados coletados de uma maneira visual e compacta. Há várias tipos de gráficos. No entanto, nessa aula, focaremos em três tipos: histograma, polígonos de frequência e gráfico de setores.

1. **Histograma:** uma representação gráfica de uma distribuição de frequências por meio de retângulos justapostos.

Vamos criar um histograma com os dados da Tabela 2. Veja-o na Figura 8.

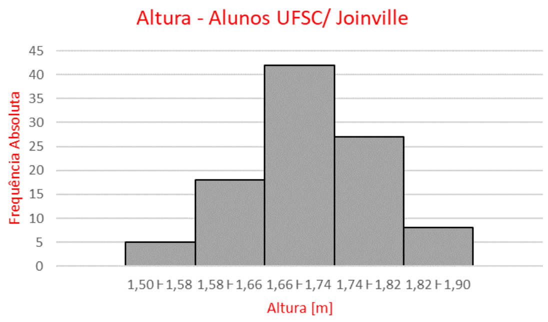


Figura 8: *Histograma representando a distribuição de frequências das alturas dos 100 alunos da UFSC/Joinville.*

A largura dos retângulos corresponde às amplitudes das classes e as áreas dos retângulos são proporcionais às frequências de cada classes. Ou seja, quanto maior for a frequência de uma classe, maior será a área do retângulo correspondente.

Normalmente se utiliza histogramas quando há uma grande quantidade de dados. Recomenda-se que sejam utilizadas entre 5 e 20 classes.

2. **Polígono de frequência:** um gráfico de linhas que é obtido unindo os pontos médios de cada classe do histograma. Ponto médio (x_i) é a média aritmética entre o limite inferior e o limite superior de cada classe.

Veja na Figura 9 um polígono de frequência para o mesmo exemplo anterior.

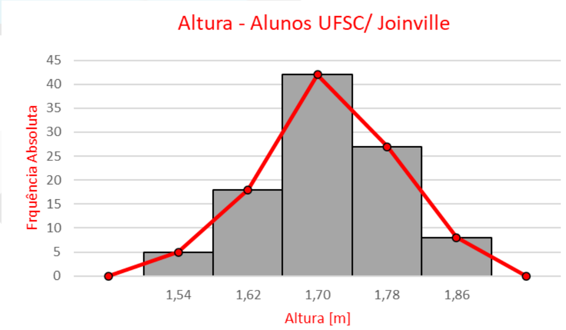


Figura 9: *Um polígono de frequência passando pelos pontos médios de cada classe (retângulos) do histograma.*

Observe que, para cada classe, temos um ponto médio marcado ao topo dos retângulos, cujos valores são mostrados no eixo x. Uma linha passa por cada um desses pontos.

3. **Gráfico de setores (de pizza):** esse gráfico consiste em representar as frequências na forma de setores circulares. Para tanto, devemos fazer a seguinte relação: 100% dos dados equivalem a todo círculo, ou seja, equivale a uma setor circular de 360°.

Para exemplificar tal ideia, consideremos novamente o último exemplo sobre alturas dos estudantes. Se quisermos representar a frequência da primeira classe ($1,50 \vdash 1,58$) em ângulo de um setor circular, basta fazermos regra de três.

$$100\% - 360^\circ$$
$$5\% - x$$

Assim, multiplicando em cruz e isolando x, temos:

$$x = \frac{360 \cdot 5}{100} = 18^\circ$$

Ou seja, num gráfico de setores, uma frequência de 5% corresponde a uma fatia de 18°. Fazendo a mesma coisa para as classes restantes, chegamos aos seguintes resultados

Classes	f_i	$f_{r\%}$	ângulo
$1,50 \vdash 1,58$	5	5%	18°
$1,58 \vdash 1,66$	18	18%	64,8°
$1,66 \vdash 1,74$	42	42%	151,2°
$1,74 \vdash 1,82$	27	27%	97,2°
$1,82 \vdash 1,90$	8	8%	28,8°

Agora podemos montar o nosso gráfico de pizza com base nesses ângulos. Veja abaixo.

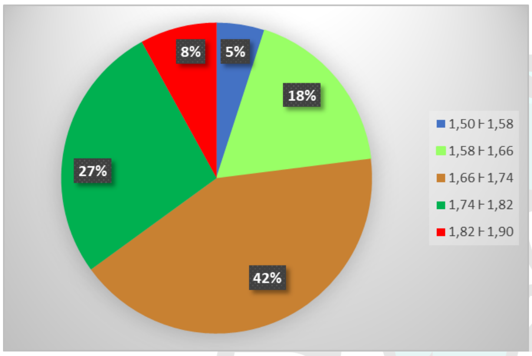


Figura 10: *Gráfico de setores (de pizza) da distribuição de frequências das alturas dos alunos.*

Até o momento, entendemos o que é Estatística e aprendemos a construir tabelas de frequências e a construir alguns gráficos.

Agora, vamos aprender a calcular algumas **medidas descritivas**, ou seja, números que descrevem ou representam os dados. Vamos lá!

2.2 Medidas descritivas

Medidas descritivas são muito úteis, pois resumem um conjunto de dados de uma determinada variável. Medidas descritivas só podem ser calculadas para variáveis quantitativas. Por isso, devemos sempre que possível dar preferência a esse tipo de variável. Diferenciaremos dois tipos de medidas descritivas: as medidas de **tendência central** e as **medidas de dispersão**.

2.2.1 Medidas de tendência central

Essas medidas revelam a posição de um ponto central de um conjunto de dados. Também podem ser entendidas como "valores típicos" de um conjunto de dados de uma variável quantitativa. Estudaremos três tipos de medidas de tendência central: **média**, **mediana** e **moda**.

1. **Média Aritmética ($\bar{x}$)**: a média aritmética de um conjunto de n números $x_1, x_2, \ldots, x_n$ é dada por:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{15}$$

Ou seja, a média aritmética é a soma de todos os dados dividido pela quantidade total de dados da variável de interesse.

Problema 8

(UFSC 2017.1 – questão 25, item 16). Suponha que na tabela abaixo estejam as estaturas da Mafalda e da sua turma (personagens da Mafalda).

Personagem	Altura (cm)
Miguelito	117,5
Susanita	125,4
Libertad	107,3
Mafalda	120
Manolito	116,4
Guille	108,7
Filipe	117,5
Mamã (mãe)	169
Papá (pai)	179,2

Com base nos dados acima, é correto afirmar que a estatura média dos personagens da Mafalda é de 129 cm.

RESOLUÇÃO:

Essa questão exige basicamente o uso da Equação 15.

$$\bar{x} = \frac{1}{9} \cdot (117,5 + 125,4 + 107,3 + 120 + 116,4 +$$
$$+108,7 + 117,5 + 169 + 179,2)$$

$$\bar{x} = \frac{1161}{9} = 129$$

Assim, essa proposição está correta.

Temos também a **média aritmética ponderada**. Ela é usada quando queremos dar pesos diferentes a cada um dos dados. É definida como:

$$\bar{x}_p = \frac{x_1 \cdot p_1 + x_2 \cdot p_2 + \ldots + x_n \cdot p_n}{p_1 + p_2 + \ldots + p_n} \tag{16}$$

Em que $p_1, p_2, \ldots, p_n$ são chamados de fatores de ponderação ou pesos, cujos significado depende do problema em questão. Para ilustrar melhor, vejamos alguns exemplos abaixo.

Problema 9

(ENEM 2017) A avaliação de rendimento de alunos de um curso universitário baseia-se na média ponderada das notas obtidas nas disciplinas pelos respectivos números de créditos, como mostra o quadro:

Avaliação	Média de notas (M)
Excelente	$9 < M \leq 10$
Bom	$7 \leq M \leq 9$
Regular	$5 \leq M < 7$
Ruim	$3 \leq M < 5$
Péssimo	$M < 3$

Quanto melhor a avaliação de um aluno em determinado período letivo, maior sua prioridade na escolha de disciplinas para o período seguinte.
Determinado aluno sabe que se obtiver avaliação "Bom" ou "Excelente" conseguirá matrícula nas disciplinas que deseja. Ele já realizou as provas de 4 das 5 disciplinas em que está matriculado, mas ainda não realizou a prova da disciplina I, conforme o quadro.

Disciplinas	Notas	Número de créditos
I		12
II	8,00	4
III	6,00	8
IV	5,00	8
V	7,50	10

Para que atinja seu objetivo, a nota mínima que ele deve conseguir na disciplina I é:

a) 7,00.
b) 7,38.
c) 7,50.
d) 8,25.
e) 9,00.

RESOLUÇÃO:
Como queremos a *nota mínima*, a média do aluno deve ser **maior ou igual** a 7. Porém, que média devemos usar? Usaremos a ponderada, pois, embora o exercício já tenha nos informado, cada disciplina possui um **peso** diferente, uma **importância** diferente, que é dados pelo respectivo número de créditos. Enquanto a disciplina I tem um peso 12, a V possui 10, por exemplo. Assim, caso fizéssemos uma média aritmética apenas, estaríamos considerando que todas as matérias tenham um mesmo peso, o que nos levaria a um erro.
Fazendo a média ponderada (Equação 16) do aluno, sendo que deve ser maior ou igual que 7, temos que

$$\bar{x}_p = \frac{x \cdot 12 + 8 \cdot 4 + 6 \cdot 8 + 5 \cdot 8 + 7,5 \cdot 10}{12 + 4 + 8 + 8 + 10} \geq 7$$

Sendo que x representa a nota na disciplina I.

$$\bar{x}_p = \frac{x \cdot 12 + 195}{42} \geq 7$$

Isolando x,

$$12x \geq 294 - 195$$

$$x \geq \frac{99}{12} = 8,25$$

Portanto, a nota mínima para que o estudante atinja seu objetivo é 8,25 (d).

Problema 10

(ENEM 2018) A Comissão Interna de Prevenção de Acidentes (CIPA) de uma empresa, observando os altos custos com os frequentes acidentes de trabalho ocorridos, fez, a pedido da diretoria, uma pesquisa no número de acidentes sofridos por funcionários. Essa pesquisa, realizada com uma amostra de 100 funcionários, norteará as ações da empresa na política de segurança do trabalho.
Os resultados obtidos estão no quadro abaixo:

Número de acidentes sofridos	Número de trabalhadores
0	50
1	17
2	15
3	10
4	6
5	2

A média do número de acidentes por funcionário na amostra que a CIPA apresentará à diretoria da empresa é

a) 0,15
b) 0,30
c) 0,50
d) 1,11
e) 2,22

RESOLUÇÃO: Esse exemplo é bem interessante, pois ele nos estabelece o jeito que vamos calcular a média: "A média do número de acidentes por funcionário[...]". Então, basta fazer saber o número total de acidentes e dividir pelo total de funcionários. Assim,

$$\bar{x}_p = \frac{0 \cdot 50 + 1 \cdot 17 + 2 \cdot 15 + 3 \cdot 10 + 4 \cdot 6 + 5 \cdot 2}{100}$$

Logo,

$$\bar{x}_p = \frac{111}{100} = 1,11$$

Portanto, a média de acidentes por funcionário é de 1,11 (d).

2. **Mediana (Me)**: diante de dados ordenados de maneira crescente, a **mediana** corresponde ao elemento que ocupa a *posição central*. Dessa forma, podemos entender a mediana como sendo o valor que *divide o conjunto de dados ao meio*.

A sua obtenção depende da quantidade de dados que temos. Consideremos um conjunto com n valores.

- Se n for ímpar, a mediana corresponderá ao elemento que está na posição $\frac{n+1}{2}$;
- Se n for par, a mediana será dada pela média aritmética dos elementos de posição $\frac{n}{2}$ e $\frac{n}{2} + 1$.

FIQUE LIGADO

Antes de calcular a mediana, não se esqueça de ordenar os dados de maneira crescente.

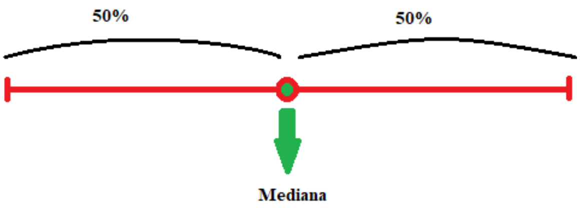


Figura 11: *Em um conjunto de dados, note que a mediana se encontra bem no meio; é a fronteira que separa os 50% menores dos 50% maiores valores.*

Vamos ver alguns exemplos!

Problema 11

(ENEM 2014) Os candidatos K, L, M, N e P estão disputando uma única vaga de emprego em uma empresa e fizeram provas de português, matemática, direito e informática. A tabela apresenta as notas obtidas pelos cinco candidatos.

Candidatos	Português	Matemática	Direito	Informática
K	33	33	33	34
L	32	39	33	34
M	35	35	36	34
N	24	37	40	35
P	36	16	26	41

Segundo o edital de seleção, o candidato aprovado será aquele para o qual a mediana das notas obtidas por ele nas quatro disciplinas for a maior. O candidato aprovado será

a) K
b) L
c) M
d) N
e) P

RESOLUÇÃO:
Primeiramente, para analisar a mediana de cada candidato, devemos **ordenar os dados de maneira crescente**. Assim,

K	33	33	33	34
L	32	33	34	39
M	34	35	35	36
N	24	35	37	40
P	16	26	36	41

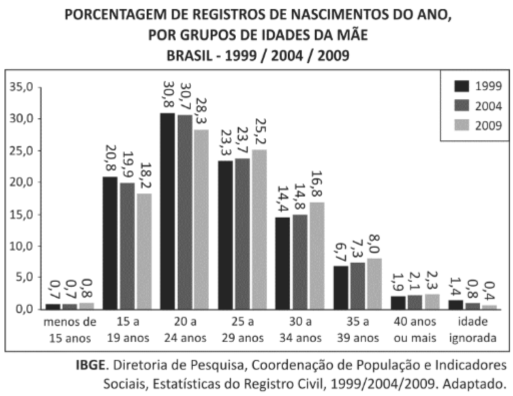
Após isso, percebemos que temos um número par de valores (n=4). Logo, a mediana será dada pela média aritmética dos termos de posições $\frac{n}{2}$ e $\frac{n}{2}+1$. Nesse caso, posições 2 e 3 (colunas destacadas). Basta calcular a mediana de cada um dos candidatos. Consequentemente, temos que

$$K : Me = 33$$
$$L : Me = 33,5$$
$$M : Me = 35$$
$$N : Me = 36$$
$$P : Me = 31$$

Portanto, N será o candidato selecionado (d).

Problema 12

(USP 2015 – 1ª Fase) Examine o gráfico.



Com base nos dados do gráfico, pode-se afirmar corretamente que a idade

a) mediana das mães das crianças nascidas em 2009 foi maior que 27 anos.
b) mediana das mães das crianças nascidas em 2009 foi menor que 23 anos.
c) mediana das mães das crianças nascidas em 1999 foi maior que 25 anos.
d) média das mães das crianças nascidas em 2004 foi maior que 22 anos.
e) média das mães das crianças nascidas em 1999 foi menor que 21 anos.

RESOLUÇÃO:

- **Item a)**

Devemos calcular a mediana dos dados em questão. Como? Diferentemente do exemplo anterior, cuja mediana foi calculada apenas pela análise dos termos, aqui, trabalhamos com frequências dos dados coletados. Por definição, a mediana é localizada bem ao meio da distribuição, como visto anteriormente. Assim, precisamos encontrar os primeiros 50% dos dados para ter uma ideia de sua localização.
Logo, para 2009, temos que

$$0,8\% + 18,2\% + 28,3\% = 47,3\%$$

Ou seja, a mediana não se encontra nas três primeiras classes, visto a soma de suas frequências relativas não ultrapassou 50%. Se somarmos a esse valor a próxima classe,

$$47,3\% + 25,2\% = 72,5\%$$

O valor de 50% é ultrapassado, significando que a mediana se encontra dentro dessa classe.
Assim, a mediana, para 2009, se encontra no grupo "25 a 29 anos". Analisando o item (a), nada podemos afirmar sobre a posição da mediana **dentro** de tal classe. Ela pode estar perto da idade 25 anos, 26 anos... não temos certeza se está acima de 27 anos como propôs o item.
Logo, não vamos marcar o item (a).

- **Item b)**

Pela análise do item (a), o (b) está errado, visto que a mediana, para 2009, se encontra no grupo de mulheres com idade mínima de 25 anos e máxima de 29 anos.
Logo, não vamos marcar o item (b).

- **Item c)**

A análise do item (c) é igual à do item (a), mas com os dados de 1999. Assim,

$$0,7\% + 20,8\% = 21,5\%$$

$$21,5\% + 30,8\% = 52,3\%$$

Ao somarmos a frequência da classe “20 a 24 anos”, a frequência relativa absoluta ultrapassou os 50%, indicando que a presença da mediana está inserida no grupo de mulheres com idade mínima de 20 anos e **máxima** de 24 anos.
Logo, não vamos marcar o item (c).

- **Item d)**

Para esse item, temos que trabalhar com **média aritmética ponderada** com os dados de 2004, cujos pesos serão as frequências de cada classe e as variáveis, a idade mínima de cada classe. Ademais, não vamos considerar a primeira e a última classe, visto que suas frequências relativas são bem pequenas comparadas às demais. Assim, calculando a média, por meio da Equação 16, temos que

$$\bar{x}_p = \frac{19,9 \cdot 15 + 30,7 \cdot 20 + 23,7 \cdot 25 + 14,8 \cdot 30 + 7,3 \cdot 35 + 2,1 \cdot 40}{19,9 + 30,7 + 23,7 + 14,8 + 7,3 + 2,1}$$

$$\bar{x}_p = \frac{2288,5}{98,5} \approx 23,23$$

Assim, a idade média das mães das crianças nascidas em 2004 foi maior que 22 anos
Logo, a resposta correta é a (d).

- **Item e)**

Mesma análise feita ao item (d) utilizando dados de 1999. Recomendo como um exercício de casa. A média resultante é, aproximadamente, 23,01 anos.
Logo, o item (e) está errado.
Apesar desse problema ter sido trabalhoso, ele é bem completo. Note a importância de termos estudado as ideias de distribuição de frequência para interpretarmos tal gráfico. Além disso, tivemos que trabalhar tanto com o conceito de média aritmética ponderada quanto o de mediana.

3. **Moda (Mo)**: valor mais frequente de um conjunto de dados, ou seja, o valor mais comum.

A moda pode não existir e, caso exista, pode haver mais de uma.
Por exemplo, o conjunto abaixo

$$3 - 4 - 5 - 5 - 5 - 6 - 8$$

Possui moda o valor 5, visto que é o que possui maior frequência de ocorrência.
Para este,

$$2 - 3 - 4 - 5 - 6 - 7 - 8$$

Não há moda, todos os elementos têm uma mesma frequência de aparições.
Já, neste,

$$1 - 2 - 4 - 5 - 5 - 6 - 6$$

Há duas modas: 5 e 6, denominando, então, um conjunto **bimodal**.
Vamos ver alguns exemplos.

Problema 13

(ENEM 2014) Uma loja que vende sapatos recebeu diversas reclamações de seus clientes relacionadas à venda de sapatos de cor branca ou preta. Os donos da loja anotaram as numerações dos sapatos com defeito e fizeram um estudo estatístico com intuito de reclamar com o fabricante.
A tabela contém a média, a mediana e a moda desses dados anotados pelos donos.

Estatísticas sobre as numerações dos sapatos com defeito			
	Média	Mediana	Moda
Numerações dos sapatos com defeito	36	37	38

Para quantificar os sapatos pela cor, os donos representaram a cor branca pelo número 0 e a cor preta pelo número 1. Sabe-se que a média da distribuição desses zeros e uns é igual a 0,45.
Os donos da loja decidiram que a numeração dos sapatos com maior número de reclamações e a cor com maior número de reclamações não serão mais vendidas.
A loja encaminhou um ofício ao fornecedor dos sapatos, explicando que não serão mais encomendados os sapatos de cor

a) branca e os de número 38.
b) branca e os de número 37.
c) branca e os de número 36.
d) preta e os de número 38.
e) preta e os de número 37.

RESOLUÇÃO: Pela tabela, vemos que a moda das numerações dos sapatos com defeitos é o tamanho 38, ou seja, meus caros, esse tamanho é o mais frequente a apresentar defeitos. Assim, já podemos eliminar os itens (b), (c) e (e).
Pelo enunciado, a média da distribuição de zeros e uns é igual a 0,45. Uma vez que 0,45 está mais próximo de 0 do que de 1, temos mais valores 0 do que 1, indicando que os sapatos brancos têm mais defeitos que os sapatos de cor preta.
Logo, a resposta correta é (a).

Problema 14

(ENEM 2011) Uma equipe de especialistas do centro meteorológico de uma cidade mediu a temperatura do ambiente, sempre no mesmo horário, durante 15 dias intercalados, a partir do primeiro dia de um mês. Esse tipo de procedimento é frequente, uma vez que os dados coletados servem de referência para estudos e verificação de tendências climáticas ao longo dos meses e anos.
As medições ocorridas nesse período estão indicadas no quadro:

Dia do mês	Temperatura (em °C)
1	15,5
3	14
5	13,5
7	18
9	19,5
11	20
13	13,5
15	13,5
17	18
19	20
21	18,5
23	13,5
25	21,5
27	20
29	16

Em relação à temperatura, os valores da média, mediana e moda são, respectivamente, iguais a

a) 17°C, 17°C e 13,5°C.
b) 17°C, 18°C e 13,5°C.
c) 17°C, 13,5°C e 18°C.
d) 17°C, 18°C e 21,5°C.
e) 17°C, 13,5°C e 21,5°C.

RESOLUÇÃO: Nesse exemplo, as três medidas de tendência central são pedidas. Porém, note que não é necessário calcular a média, visto que, para os itens, a média possui o mesmo valor: 17°C.
Para obter a mediana, a primeira coisa a ser feita é ordenas os dados de maneira crescente. Assim,
13,5; 13,5; 13,5; 13,5; 14; 15,5; 16; 18; 18; 18,5; 19,5; 20; 20; 20; 21,5
Como $n = 15$, a mediana será dada pelo termo de posição 8: 18°C.
Para a moda, vemos que só há um elemento que aparece quatro vezes (a maior frequência para esse conjunto de dados): 13,5°C.
Logo, com $\bar{x} = 17$°C, $Me = 18$°C e $Mo = 13,5$°C, o item correto é o (b).

2.2.2 Medidas de dispersão (ou variabilidade)

Para entendermos a necessidade de se ter um outro conjunto de medidas de dispersão para caracterizar um conjunto de dados, considere a seguinte situação abaixo.

Seja um animador de festas infantis que seleciona as atividades de acordo com a média das idades das crianças convidadas para uma festa. Consideremos as idades de dois grupos de crianças que irão participar de duas festas diferentes:

$$Festa\ A: 1; 2; 2; 12; 12; 13.\ \bar{x}_A = 7\ anos$$

$$Festa\ B: 5; 6; 7; 7; 8; 9.\ \bar{x}_B = 7\ anos$$

Nos dois casos, a média de idade foi a mesma. Porém, ao analisar os grupos, seria correto afirmar que ambos terão a mesma atividade ofertada? É uma resposta que a média não consegue responder sozinha. O animador, para escolher a atividade mais adequada para cada um, terá de usar um novo conjunto de ferramentas para realizar essa análise: as **medidas de dispersão**.

A situação descrita acima nos mostra que precisamos de novas ferramentas para analisar dados, além das medidas de tendência central. Agora, vamos conhecer as **medidas de dispersão**, cujo objetivo é quantificar a variabilidade de um conjunto de dados. Em nossa aula, estudaremos três tipos: **amplitude**, **variância** e **desvio padrão**.

1. **Amplitude (A)**: diferença entre o maior e o menor valor de um conjunto de dados.

Para os as crianças nas festas A e B, suas amplitudes são

$$A_A = 13 - 1 = 12$$

$$A_B = 9 - 5 = 4$$

FIQUE LIGADO

A amplitude é uma medida de dispersão que nos diz pouco em relação à dispersão dos dados, pois note que ela depende de apenas dois valores, os quais são os extremos do conjunto.
Por conta dos valores extremos considerados para o seu cálculo, a amplitude é bem sensível a valores discrepantes também. Por exemplo, se em uma festa tivéssemos 5 crianças com 13 anos e apenas 1 criança com 1 ano, a amplitude seria 12 anos. Mas se essa criança pequena não fosse à festa, a amplitude seria 0. Uma enorme diferença...

2. **Variância (σ^2)**: uma medida de dispersão que mostra o quão distante, em média, cada valor de um conjunto de dados está da média aritmética Brasil Escola, 2020.

Para calculá-la, devemos seguir alguns passos. Vamos lá!
Sejam $x_1,\ x_2,\ \dots,\ x_n$ valores de um conjunto de dados cuja média é $\bar{x}$.
1°: Façamos, para cada dado, $x_1,\ x_2,\ \dots,\ x_n$, a diferença entre ele e a média. E, depois, elevamos essa diferença ao quadrado. Essa diferença é denominada de *desvio quadrático*. Matematicamente, fazemos isso

$$(x_1 - \bar{x})^2$$

$$(x_2 - \bar{x})^2$$

$$\dots$$

$$(x_n - \bar{x})^2$$

2°: Somamos essas diferenças quadráticas e dividimos pela quantidade total de dados do conjunto n. Com isso, temos a definição matemática de variância, representada por σ^2 (sigma ao quadrado). Assim,

$$\sigma^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+\ (x_n-\bar{x})^2}{n} \quad (17)$$

Ou seja, a variância é a média aritmética dos desvios quadráticos (Iezzi et al., 2016).

No entanto, temos um "prejuízo" ao trabalhar com a variância. A desvantagem é a perda de sensibilidade em relação à sua interpretação, visto que sua unidade de medida é a **unidade dos dados ao quadrado**. Por exemplo, se tivéssemos medindo o desvio de altura de alunos de uma sala de aula em metros (m), a variância nos daria um resultado cuja unidade seria m^2, o que não faria muito sentido, uma vez que trabalhamos com altura no exemplo.

Para resolver esse empecilho, devemos tirar a raiz quadrada da variância, o que dá origem à nossa última medida de dispersão da nossa aula, o **desvio padrão**.

3. **Desvio padrão (σ)**: possui a mesma interpretação que a variância. Seu valor é, a raiz quadrada da variância. Assim:

$$\sigma = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+\ (x_n-\bar{x})^2}{n}} \quad (18)$$

Uma vez tirando a raiz quadrada da variância, não teremos mais aquele problema com unidades ao quadrado, o que facilita a interpretação dessa medida de dispersão.

Para o cálculo do desvio padrão deve-se seguir os mesmos passos do cálculo da variância e, ao final, calcular a raiz quadrada do resultado.

Com isso, estamos preparados para exemplos.

Vamos lá!

Problema 15

(UDESC 2016.2) Sejam a e b $\in \mathbb{R}$ O valor do desvio padrão, de modo que o conjunto de dados ordenados {4, 17, 22, a, b, 37} tenha média e mediana iguais a 24, é:

a) $\sqrt{59}$
b) $\sqrt{62}$
c) $\sqrt{58}$
d) $\sqrt{57}$
e) $\sqrt{\frac{19}{3}}$

RESOLUÇÃO:
Primeiro, precisamos encontrar os valores de a e b. Para isso, usamos o fato de a média e a mediana serem iguais a 24. Como é um conjunto com 6 elementos, a mediana será dada pela média aritmética do terceiro com o quarto termo. Assim,

$$\frac{22+a}{2} = 24$$

Isolando a,

$$a = 26$$

Analisando a média aritmética do conjunto, temos

$$\frac{14+17+22+26+b+37}{6} = 24$$

Isolando b,

$$b = 28$$

Agora, vamos ao cálculo do desvio padrão. Como a média do conjunto já foi dada, vamos ao segundo passo: calcular a diferença de cada elemento com a média. Logo,

$$14-24 = 10$$

$$17-24 = -7$$

$$22-24 = -2$$

$$26-24 = 2$$

$$28-24 = 4$$

$$37-24 = 13$$

Agora, devemos elevar cada resultado ao quadrado,

$$10^2 = 100$$

$$(-7)^2 = 49$$

$$(-2)^2 = 4$$

$$2^2 = 4$$

$$4^2 = 16$$

$$13^2 = 169$$

Com isso, vamos somar os valores resultantes, dividir pelo número total de elementos e tirar sua raiz, conforme a Equação 18. Assim,

$$\sigma = \sqrt{\frac{100+49+4+4+16+169}{6}}$$

$$\sigma = \sqrt{\frac{342}{6}} = \sqrt{57}$$

Portanto, o desvio padrão para aquele conjunto é $\sqrt{57}$ (d).

Problema 16

(ENEM – 2016) O Procedimento de perda rápida de "peso" é comum entre os atletas dos esportes de combate. Para participar de um torneio, quatro atletas da categoria até 66 kg, Peso-Pena, foram submetidos a dietas balanceadas e atividades físicas. Realizaram três "pesagens" antes do início do torneio. Pelo regulamento do torneio, a primeira luta deverá ocorrer entre o atleta mais regular e o menos regular quanto aos "pesos". As informações com base nas pesagens dos atletas estão no quadro. Quais atletas participarão da luta?

Atleta	1ª pesagem (kg)	2ª pesagem (kg)	3ª pesagem (kg)	Média	Mediana	Desvio padrão
I	78	72	66	72	72	4,90
II	83	65	65	71	65	8,49
III	75	70	65	70	70	4,08
IV	80	77	62	73	77	7,87

a) I e III.
b) I e IV.
c) II e III.
d) II e IV.
e) III e IV.

RESOLUÇÃO:
Apesar de a tabela ter nos dados três medidas (duas de tendência central e uma de dispersão), o que nos interessa é apenas o desvio padrão. Lembre-se de que o desvio padrão é uma medida de dispersão que nos mostra o quão disperso os dados estão do valor central de uma distribuição. Ou seja, caros, quanto maior for seu valor, *maior será a dispersão* dos dados analisados; quanto menor for, *menor será a dispersão*. Assim, analisando a coluna de desvio padrão, percebemos que o atleta mais regular é aquele cujo desvio é o menor, logo, atleta III. E o menos regular, será aquele de maior desvio, logo, atleta II. Portanto, a luta será entre II e III (c).

Certo, caros *padawans*, a nossa caminhada pelo universo da Estatística e Probabilidade foi longa, exigiu trabalho, mas cá estamos na linha de chegada. Esperamos que vocês tenham conseguido assimilar e compreender as ideias presentes nessa aula. E, claro, como tudo na vida, para ganharmos mais experiência e sabedoria, devemos praticar. Esse é o motivo pelo qual colocamos a seguir exercícios de aplicação sobre os conteúdos abordados aqui, cujo gabarito se encontra ao final. No entanto, recomendamos que façam mais exercícios por fora para que fixem melhor as ideias. É isso aí, pessoal!

COLABORADORES DESTA AULA

- **Texto**:
  Murilo Brunazzo Medeiros
- **Diagramação**:
  Maurício de Campos Porath
- **Revisão**:
  Maurício de Campos Porath

Referências Bibliográficas

Barbetta, P.A., M.M. Reis e A.C. Bornia (2010). *Estatística para Cursos de Engenharia e Informática: 3a Ed.* Vol. 1. São Paulo: Atlas.

Brasil Escola (2020). *Medidas de dispersão: variância e desvio padrão*. URL: https://brasilescola.uol.com.br/matematica/medidas-dispersao-variancia-desvio-padrao.htm (acesso em 01/08/2020).

Iezzi, G. et al. (2016). *Matemática: ciência e aplicações – 2º ano (Ensino Médio): 9. Ed*. Vol. 1. São Paulo: Saraiva Educação.

Prepara Enem (2020). *Estatística*. URL: https://www.preparaenem.com/matematica/estatistica.htm (acesso em 11/06/2020).

3 Lista de Problemas

Os problemas da lista a seguir foram retirados diretamente dos cadernos de prova dos referidos processos seletivos.

1. (ENEM 2019) O dono de um restaurante situado às margens de uma rodovia percebeu que, ao colocar uma placa de propaganda de seu restaurante ao longo da rodovia, as vendas aumentaram. Pesquisou junto aos seus clientes e concluiu que a probabilidade de um motorista perceber uma placa de anúncio é 1/2.
Com isso, após autorização do órgão competente, decidiu instalar novas placas com anúncios de seu restaurante ao longo dessa rodovia, de maneira que a probabilidade de um motorista perceber pelo menos uma das placas instaladas fosse superior a 99/100.
A quantidade mínima de novas placas de propaganda a serem instaladas é

   a) 99.
   b) 51.
   c) 50.
   d) 6.
   e) 1.

2. (ENEM 2015) Uma competição esportiva envolveu 20 equipes com 10 atletas cada. Uma denúncia à organização dizia que um dos atletas havia utilizado substância proibida. Os organizadores, então, decidiram fazer um exame antidoping. Foram propostos três modos diferentes para escolher os atletas que irão realizá-lo:
Modo I: sortear três atletas dentre todos os participantes;
Modo II: sortear primeiro uma das equipes e, desta, sortear três atletas;
Modo III: sortear, primeiro, três equipes e, então, sortear um atleta de cada uma dessas três equipes.
Considere que todos os atletas têm igual probabilidade de serem sorteados e que P(I), P(II) e P(III) sejam as probabilidades de o atleta que utilizou a substância proibida seja um dos escolhidos para o exame no caso do sorteio ser feito pelo modo I, II ou III. Comparando-se essas probabilidades, obtém-se

   a) $P(I) < P(III) < P(II)$
   b) $P(II) < P(I) < P(III)$
   c) $P(I) < P(II) = P(III)$
   d) $P(I) = P(II) < P(III)$
   e) $P(I) = P(II) = P(III)$

3. (ENEM 2015) Em uma escola, a probabilidade de um aluno compreender e falar inglês é de 30%. Três alunos dessa escola, que estão em fase final de seleção de

intercâmbio, aguardam, em uma sala, serem chamados para uma entrevista. Mas, ao invés de chamá-los um a um, o entrevistador entra na sala e faz, oralmente, uma pergunta em inglês que pode ser respondida por qualquer um dos alunos. A probabilidade de o entrevistador ser entendido e ter sua pergunta oralmente respondida em inglês é

a) 23,7%
b) 30,0%
c) 44,1%
d) 65,7%
e) 90,0%

4. (ENEM 2014) O psicólogo de uma empresa aplica um teste para analisar a aptidão de um candidato a determinado cargo. O teste consiste em uma série de perguntas cujas respostas devem ser "verdadeiro" ou "falso" e termina quando o psicólogo fizer a décima pergunta ou quando o candidato der a segunda resposta errada.
Com base em testes anteriores, o psicólogo sabe que a probabilidade de o candidato errar uma resposta é 0,20.
A probabilidade de o teste terminar na quinta pergunta é

a) 0,02048.
b) 0,08192.
c) 0,24000.
d) 0,40960.
e) 0,49152.

5. (UFSC 2017.1 – Questão 28, proposição 08) Uma urna contém 3 bolas brancas, numeradas de 1 a 3, e 6 bolas pretas, numeradas de 1 a 6. Uma bola é extraída ao acaso. Se for sorteado um número ímpar, então a probabilidade de ter saído uma bola branca é de 2/9.
6. (UFSC 2017.1 – Questão 28, proposição 16) A probabilidade de um casal com quatro filhos ter dois meninos e duas meninas é menor do que a probabilidade de dois casais com dois filhos terem, cada casal, um menino e uma menina.
7. (UFSC 2019.2 – Questão 29, proposição 04) Em certa universidade foi realizado um levantamento acerca do número de reprovações dos estudantes em duas disciplinas. Constatou-se que entre os alunos de engenharia 25% reprovaram na disciplina de Cálculo, 15% reprovaram na disciplina de Álgebra e 10% reprovaram em ambas as disciplinas. Ao selecionar, ao acaso, um dos alunos de engenharia, a probabilidade de ele não ter reprovado em Álgebra sabendo que reprovou em Cálculo será de 60%.
8. (UFSC 2019.2 – Questão 29, proposição 16) A urna A tem três bolas vermelhas e quatro brancas e a urna B tem seis bolas vermelhas e duas brancas. Uma urna é escolhida ao acaso e dela, também ao acaso, é sorteada uma bola. Se a bola escolhida for vermelha, então a probabilidade de que ela seja da urna A é igual a 4/11.
9. (UFSC 2018.2 – Questão 26, proposição 01) Na final do concurso de Miss Brasil, estão classificadas as Misses do Ceará, de Goiás, da Paraíba, do Rio Grande do Sul e de Santa Catarina. Em determinada prova devem ser sorteadas, ao acaso, duas candidatas para um teste fotográfico. O sorteio consiste em retirar de uma urna, sem reposição, dois papéis grafados com a sigla do estado que cada candidata representa. Se na primeira retirada foi sorteada a representante do estado da Paraíba, a probabilidade de a outra candidata ser da Região Sul do Brasil é igual a 2/5.
10. (UFSC 2018.1 – Questão 27, proposição 16) A maioria dos sistemas de regras de RPG usa dados para testar as habilidades dos personagens. As formas mais comuns de dados utilizados são os sólidos de Platão, isto é, dados de 4, 6, 8, 12 e 20 faces, conhecidos como d4, d6, d8, d12 e d20, respectivamente, conforme a figura abaixo. Se forem lançados aleatoriamente dois dados "d12", a probabilidade de não serem obtidos números iguais nas duas faces é de 11/12.

11. (UERJ 2019.2) Um menino vai retirar ao acaso um único cartão de um conjunto de sete cartões. Em cada um deles está escrito apenas um dia da semana, sem repetições: segunda, terça, quarta, quinta, sexta, sábado, domingo. O menino gostaria de retirar sábado ou domingo. A probabilidade de ocorrência de uma das preferências do menino é:

a) 1/49
b) 2/49
c) 1/7
d) 2/7

12. (ENEM 2019 PPL) Uma locadora possui disponíveis 120 veículos da categoria que um cliente pretende locar. Desses, 20% são da cor branca, 40% são da cor cinza, 16 veículos são da cor vermelha e o restante, de outras cores. O cliente não gosta da cor vermelha e ficaria contente com qualquer outra cor, mas o sistema de controle disponibiliza os veículos sem levar em conta a escolha da cor pelo cliente.
Disponibilizando aleatoriamente, qual é a probabilidade de o cliente ficar contente com a cor do veículo?

a) 16/120
b) 32/120
c) 72/120
d) 101/120
e) 104/120

13. (UFPR 2020) Uma adaptação do Teorema do Macaco afirma que um macaco digitando aleatoriamente num teclado de computador, mais cedo ou mais tarde, escreverá a obra "Os Sertões", de Euclides da Cunha. Imagine que um macaco digite sequências aleatórias de 3 letras em um teclado que tem apenas as seguintes letras: S, E, R, T, O.
Qual é a probabilidade de esse macaco escrever a palavra "SER" na primeira tentativa?

a) 1/5.
b) 1/15.
c) 1/75.
d) 1/125.
e) 1/225.

14. (UDESC 2019.2). Dois dados de cores diferentes com 6 faces numeradas de 1 a 6, e não viciados, são lançados. Analise as sentenças.

I) O resultado mais provável para a soma de suas faces superiores é 7, que tem probabilidade 1/6.

II) A soma de suas faces superiores com resultado 5 tem probabilidade 1/9, assim como o resultado 9.
III) O produto de suas faces superiores com resultado 4 tem probabilidade 1/12.
IV) O 9 é um dos resultados menos prováveis para o produto de suas faces superiores e tem probabilidade 1/36.

Assinale a alternativa correta.

a) Somente a sentença I é verdadeira.
b) Somente as sentenças I e IV são verdadeiras.
c) Somente as sentenças I, II e III são verdadeiras.
d) Somente as sentenças II e III são verdadeiras.
e) Todas as sentenças são verdadeiras.

15. (ENEM 2019 PPL) O quadro apresenta a relação dos jogadores que fizeram parte da seleção brasileira de voleibol masculino nas Olimpíadas de 2012, em Londres, e suas respectivas alturas, em metro.

Nome	Altura (m)
Bruninho	1,90
Dante	2,01
Giba	1,92
Leandro Vissoto	2,11
Lucas	2,09
Murilo	1,90
Ricardinho	1,91
Rodrigão	2,05
Serginho	1,84
Sidão	2,03
Thiago Alves	1,94
Wallace	1,98

Disponível em: www.cbv.com.br. Acesso em: 31 jul. 2012 (adaptado).

A mediana das alturas, em metro, desses jogadores é

a) 1,90.
b) 1,91.
c) 1,96.
d) 1,97.
e) 1,98.

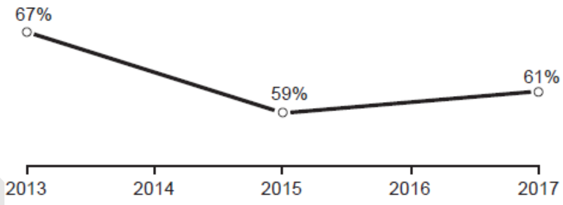
16. (ENEM 2010) Marco e Paulo foram classificados em um concurso. Para classificação no concurso o candidato deveria obter média aritmética na pontuação igual ou superior a 14. Em caso de empate na média, o desempate seria em favor da pontuação mais regular. No quadro a seguir são apresentados os pontos obtidos nas provas de Matemática, Português e Conhecimentos Gerais, a média, a mediana e o desvio padrão dos dois candidatos. Dados dos candidatos no concurso

	Matemática	Português	Conhecimentos Gerais	Média	Mediana	Desvio Padrão
Marco	14	15	16	15	15	0,32
Paulo	8	19	18	15	18	4,97

O candidato com pontuação mais regular, portanto mais bem classificado no concurso, é

a) Marco, pois a média e a mediana são iguais.
b) Marco, pois obteve menor desvio padrão.
c) Paulo, pois obteve a maior pontuação da tabela, 19 em Português.
d) Paulo, pois obteve maior mediana.
e) Paulo, pois obteve maior desvio padrão.

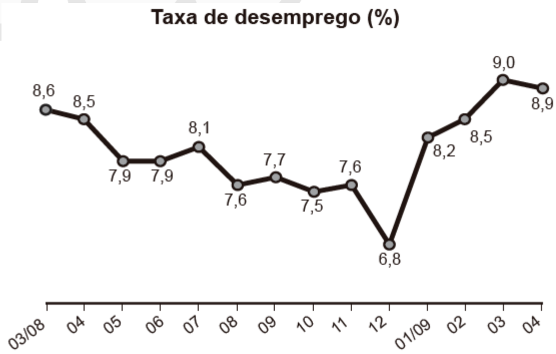
17. (ENEM 2018) A raiva é uma doença viral e infecciosa, transmitida por mamíferos. A campanha nacional de vacinação antirrábica tem o objetivo de controlar a circulação do vírus da raiva canina e felina, prevenindo a raiva humana. O gráfico mostra a cobertura (porcentagem de vacinados) da campanha, em cães, nos anos de 2013,2015 e 2017, no município de Belo Horizonte, em Minas Gerais. Os valores das coberturas dos anos de 2014 e 2016 não estão informados no gráfico e deseja-se estimá-los. Para tal, levou-se em consideração que a variação na cobertura de vacinação da campanha antirrábica, nos períodos de 2013 a 2015 e de 2015 a 2017, deu-se de forma linear.
Qual teria sido a cobertura dessa campanha no ano de 2014?



Disponível em: http://pni.datasus.gov.br. Acesso em: 5 nov. 2017.

a) 62,3%
b) 63,0%
c) 63,5%
d) 64,0%
e) 65,5%

18. (ENEM 2017) O gráfico apresenta a taxa de desemprego (em %) para o período de março de 2008 a abril de 2009, obtida com base nos dados observados nas regiões metropolitanas de Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre.



A mediana dessa taxa de desemprego, no período de março de 2008 a abril de 2009, foi de

a) 8,1%
b) 8,0%
c) 7,9%
d) 7,7%
e) 7,6%

19. (ENEM 2017) Cinco regiões de um país estão buscando recursos no Governo Federal para diminuir a taxa de desemprego de sua população. Para decidir qual região receberia o recurso, foram colhidas as taxas de

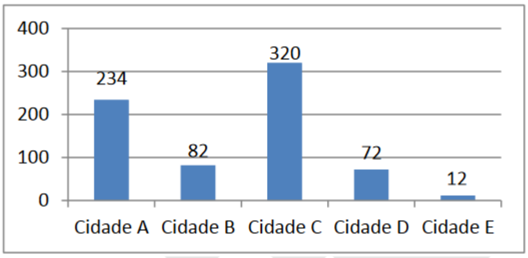
desemprego, em porcentagem, dos últimos três anos. Os dados estão apresentados na tabela.

Taxa de desemprego (%)					
	Região A	Região B	Região C	Região D	Região E
Ano I	12,1	12,5	11,9	11,6	8,2
Ano II	11,7	10,5	12,7	9,5	12,6
Ano III	12,0	11,6	10,9	12,8	12,7

Ficou decidido que a região contemplada com a maior parte do recurso seria aquela com a maior mediana das taxas de desemprego dos últimos três anos.
A região que deve receber a maior parte do recurso é a

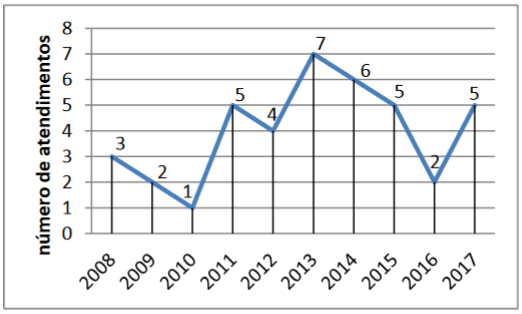
a) A.
b) B.
c) C.
d) D.
e) E.

20. (UFSC 2018.2 – Questão 30, proposição 04) Em uma campanha de prevenção contra a proliferação do mosquito Aedes aegypti, a Secretaria de Saúde do Estado resolveu fazer panfletos informativos que indicassem as cidades com maior número de indivíduos infectados com o vírus da dengue. Para tanto, foi construído o seguinte gráfico, que apresenta o número de infectados em cada cidade da região:



Foi decidido que o gráfico de barras acima deveria ser substituído por um gráfico de setores. Então, o ângulo do maior setor desse novo gráfico deve ser de 160°?

21. (UFSC 2018.2 – Questão 30, proposição 08) Em uma unidade de saúde de determinada região do estado, foi registrado, a cada ano, o número de atendimentos a pacientes que relataram terem sido picados por cobras, conforme indica o gráfico abaixo:



Considerando os dados indicados no gráfico acima e que m_1, m_2, m_3 representam, respectivamente, moda, mediana e média, é correto afirmar que $m_3 < m_2 < m_1$.

4 Gabarito

1. D
2. E
3. D
4. B
5. Falsa
6. Falsa
7. Certa
8. Certa
9. Falsa
10. Certa
11. D
12. E
13. D
14. E
15. C
16. B
17. B
18. B
19. E
20. Certa
21. Certa

"Só se pode alcançar um grande êxito quando nos mantemos fiéis a nós mesmos"