Unidade III

5 MEDIDAS DE DISPERSÃO

Nas seções anteriores, apresentamos grandezas que nos trazem um valor significativo dos dados de um conjunto. Nesta seção, vamos discutir como aprofundar nossa análise, enfocando agora a dispersão dos dados, ou seja, analisaremos se se eles estão concentrados em torno do valor médio ou se estão espalhados, dispersos em torno dele.

É bastante provável que no exemplo a respeito dos gastos de um conjunto de amigos na lanchonete você tenha pensado que não seria justo dividir as despesas igualmente, já que os gastos são bem diferentes. É também provável que você não tenha considerado injusta a divisão das despesas para fazer o jantar.

A origem do provável desconforto no primeiro caso e da sensação de justiça no segundo se explica estatisticamente pelas medidas de dispersão.

No jantar, assumimos que todos têm acesso às mesmas comidas e bebidas e, portanto, o gasto efetivo de cada um seria parecido, visto que se supõe que os que gastaram mais serão ressarcidos e a despesa será igual para todos. No caso da lanchonete, há diferenças significativas no valor do consumo, e, portanto, o valor efetivamente gasto pode diferir significativamente do valor pago.

O objetivo desta seção é começar a transformar essa sensação intuitiva em conhecimento quantitativo para a grandeza correspondente. Vamos abordar aqui somente as medidas de dispersão mais frequentemente utilizadas: o intervalo e o desvio-padrão.

5.1 Intervalo

O conceito de intervalo é bastante intuitivo e coincide com o significado da linguagem cotidiana. O intervalo de valores é aquele que vai do menor ao maior valor.

Retomemos os exemplos das tabelas 12, 13 e 14 para determinar os intervalos de dados.

5.1.1 Dados da tabela 12

Nos dados da tabela 12 temos idades de pessoas em uma família. A mais nova, Ana, tem 3 anos e a mais velha, Gabriel, tem 56. Assim sendo, o intervalo vai de 3 a 56.

Quando não é de interesse saber exatamente qual o maior e qual o menor valor, mas apenas saber qual o tamanho do intervalo, este será encontrado, subtraindo-se o menor valor do maior valor do conjunto. No exemplo, o intervalo tem tamanho 53, que é a diferença de idade da pessoa mais velha para a mais nova.

5.1.2 Dados da tabela 13

Tomando agora os dados da tabela 13, vemos que o maior gasto foi de R$ 47,00 e o menor foi de R$19,00. Desse modo, o intervalo seria de 19 a 47, o que nos dá um intervalo de tamanho 28.

No entanto, como comentamos anteriormente, nesse exemplo as diferenças de gasto eram apenas circunstanciais e, após feito o acerto de contas, todos teriam gastado exatamente o mesmo valor: R$ 30,00.

Se utilizarmos em nossa análise o gasto efetivo, teremos então R$ 30,00 como a maior e a menor despesa, simultaneamente, e o intervalo indo de 30 a 30, com tamanho zero.

Ou seja, se todos os dados têm o mesmo valor, dizemos que são dados sem qualquer dispersão, ou dizemos, formalmente, que a dispersão é nula.

5.1.3 Dados da tabela 14

Na tabela 14 temos os dados para os gastos na lanchonete. Nesse caso, a menor despesa individual foi de R$ 12,00 e a maior foi de R$ 18,00. O intervalo vai, portanto, de 12 a 18 e tem tamanho 6.

5.2 Variância e desvio-padrão

O intervalo é uma medida de dispersão fácil de calcular, porém de utilidade limitada, pois traz apenas informações sobre os valores extremos.

A exemplo do que discutimos para as medidas de posição, quando abordamos a média, vamos agora definir uma medida de dispersão que considere cada um dos valores individuais.

5.2.1 Estabelecendo critérios

Nesta seção vamos discutir alguns aspectos importantes referentes à definição de variância. O intuito não é trazer uma discussão formal, mas apenas discutir alguns aspectos de modo a facilitar a compreensão.

Tomemos um exemplo similar àquele dos gastos na lanchonete, em que analisaremos os gastos de 5 amigos em um restaurante.

Para analisar o quanto a divisão igualitária das despesas foi justa (ou injusta), o primeiro passo é comparar cada valor consumido com o valor efetivamente pago. Formalmente, isso significa subtrair o valor médio do valor específico.

Como essa análise é um refinamento do estudo da média, continuaremos seguindo um procedimento similar para a realização dos cálculos. Assim, retomamos a tabela já com a linha e a coluna adicionais para o cálculo da média e acrescentamos uma nova coluna com o valor dessas diferenças, conforme colocamos na tabela 24.

O cálculo da média se dá simplesmente pela soma das despesas individuais dividida pelo número de pessoas, com resultado 30. Na terceira coluna, temos as diferenças entre os gastos individuais e a média. Essas diferenças são obtidas pela subtração $x_i - \bar{x}$, logo, um resultado positivo significa que a pessoa consumiu mais do que pagou; ao contrário, quando o resultado é menor do que zero, isso significa que a pessoa consumiu menos do que pagou.

Por exemplo, Alberto gastou R$ 25,00 e pagou R$ 30,00, a diferença citada é de R$ -5,00. Já Beatriz consumiu R$ 31,00 e pagou R$ 30,00, portanto, sua diferença é de R$ 1,00.

Para analisar a dispersão dos dados, uma primeira proposta seria a de somar as diferenças. No entanto, como podemos ver na última linha da terceira coluna, a soma das diferenças dá zero. Esse resultado era esperado, visto que houve equilíbrio entre os consumos maiores ou menores que a média.

Uma segunda proposta seria considerar as diferenças sem o sinal, ou seja, o valor absoluto delas e dividir pelo número de pessoas para ter uma média desses desvios. Existe de fato uma medida de dispersão assim calculada, mas sua utilização não é frequente, então passaremos a discutir outra proposta matematicamente mais significativa: a variância.

Pessoa	xi	$xi - \bar{x}$
Alberto	25	-5
Beatriz	31	1
Carlos	47	17
Diana	19	-11
Edgar	28	-2
Soma	**150**	**0**
Média	**30**	

Tabela 24 – Cálculo da comparação de valores individuais e média para o exemplo de gastos na lanchonete.

A variância utiliza as distâncias entre os valores individuais e as médias, mas o faz elevando esses valores ao quadrado. Não cabe aqui discutir os motivos de tal definição, dada a complexidade da matemática envolvida, mas vale uma discussão qualitativa dessa escolha.

Ao utilizarmos os quadrados das diferenças, garantimos que os parâmetros de desvio serão sempre positivos, já que o produto de números com sinais iguais é sempre positivo.

Além disso, esse critério faz com que tenhamos ainda mais rigor na medida da "injustiça", pois quanto maior a diferença, maior o peso com que ela será contada. Por exemplo: a diferença referente a Beatriz, que é de 1, entrará na "conta da injustiça" com o seu quadrado, que é também 1, e a diferença devida a Alberto, que é de -5, entrará na "conta da injustiça" com o seu quadrado, que é 25.

Nas próximas seções, vamos utilizar os critérios anteriores para definir os parâmetros de dispersão que são efetivamente utilizados na maioria dos estudos estatísticos. Em seguida, passaremos a construir passo a passo o procedimento para efetuar seus cálculos.

5.3 Definições formais

Há duas grandezas que descrevem a dispersão dos dados utilizando o critério acima: a variância e o desvio-padrão. O desvio-padrão é a raiz quadrada da variância e será denotado pela letra sigma minúscula (σ). A variância é, portanto, o quadrado do desvio-padrão e se denota por σ^2.

5.3.1 Variância e desvio-padrão para populações

Quando nossos dados trazem a totalidade da população estudada, como é o caso do exemplo anterior, a variância é definida como a média dos quadrados das diferenças entre o valor individual e o valor médio, conforme formalizado a seguir:

$$\sigma^2 = \frac{\sum\left(x - \bar{x}\right)^2}{N}$$

O desvio-padrão será simplesmente a raiz da variância:

$$\sigma = \sqrt{\sigma^2}$$

Geralmente, a grandeza mais utilizada para analisar dispersões é o desvio-padrão. Não detalharemos ainda o seu significado, isso ficará para quando estudarmos a distribuição normal de probabilidades. Por hora, basta saber que uma maior dispersão dos dados implica ter um valor maior para o desvio-padrão.

Nos casos em que temos os dados em tabelas de frequência, precisamos que lembrar que cada valor x_i aparece f_i vezes e, a exemplo do que fizemos para o cálculo da média, é preciso multiplicar a quantidade de vezes que cada valor aparece para que todas as diferenças sejam computadas.

Assim sendo, teremos:

$$\sigma^2 = \frac{\sum f_i.\left(x - \bar{x}\right)^2}{\sum f_i}$$

E, novamente, o desvio-padrão será simplesmente a raiz da variância.

Lembrete

Lembre que, embora as fórmulas sejam escritas de maneira diferente quando colocamos as frequências explicitamente, a grandeza calculada é a mesma nos dois casos.

5.3.2 Variância e desvio-padrão para amostras

Nos casos em que temos dados para uma amostra e não para toda a população, em lugar de dividirmos por N, a divisão será feita por N-1.

Não cabe aqui discutir os motivos dessa definição, mas apenas comentar dois aspectos dessa diferença.

Quando o número de dados é grande, o valor final não será muito afetado pela substituição de N por N-1. Quando o número de dados é pequeno, a dispersão encontrada dividindo-se por N-1 é maior, refletindo o fato de que se espera que haja maior diversidade de valores na população em geral que na observada na amostra.

Por fim, note que como na maioria das vezes temos amostras em lugar de populações, é mais comum utilizarmos as definições na forma que estão colocadas a seguir:

$$\sigma^2 = \frac{\sum \left(x - \bar{x}\right)^2}{N - 1}$$

ou

$$\sigma^2 = \frac{\sum f_i.\left(x - \bar{x}\right)^2}{\left(\sum f_i\right) - 1}$$

e, como nas outras vezes:

$$\sigma = \sqrt{\sigma^2}$$

5.4 Realização dos cálculos

Vamos agora completar o cálculo da variância e do desvio-padrão que havíamos iniciado na tabela 24, e para que o procedimento de cálculo e o significado fiquem mais claros, vamos resolver ainda outros exemplos adicionais.

5.4.1 Dados não agrupados

Para efetuar os cálculos, é preciso lembrar que nesse caso temos uma população, já que temos os valores referentes a todos os ocupantes da mesa estudada.

Utilizaremos aqui um procedimento similar ao usado no cálculo das médias, colocando os valores intermediários que precisamos para os cálculos em colunas adicionais e realizando as somas pertinentes nas colunas respectivas. Vejamos como fazer isso na tabela 25.

Pessoa	xi	$xi-\overline{x}$	$(xi-\overline{x})$^2
Alberto	25	-5	25
Beatriz	31	1	1
Carlos	47	17	289
Diana	19	-11	121
Edgar	28	-2	4
Total	**150**	**–**	**440**
Média	**30**	**variância**	**88**
		desvio	**9,38**

Tabela 25 – Cálculo da variância e do desvio-padrão para o primeiro exemplo de gastos na lanchonete.

Como havíamos feito antes, na terceira coluna estão as diferenças entre os valores individuais e a média. Não efetuamos a soma para esses valores porque a variância utilizará somente o somatório dos quadrados desses números, os quais estão colocados na coluna seguinte.

Para maior clareza, vejamos passo a passo como foi feito o cálculo do desvio-padrão depois de já termos calculado a média.

- Como já visto anteriormente, a diferença entre o valor individual e a média, dado por $x_i - \overline{x}$, tem valor -5.

- Elevamos o valor encontrado ao quadrado para determinar $(x_i - \overline{x})^2$, obtendo o valor 25, que é colocado na tabela.

- Repetimos esse procedimento para todos os dados x_i.

- Somamos todos os valores da coluna $(x_i - \overline{x})^2$, obtendo o valor 440, que é colocado na linha adicional da coluna.

- Como se trata de uma população, dividimos o valor obtido por N, que, nesse caso, é 5. O resultado é a variância de valor 88.

- Para a obtenção do desvio-padrão, calcula-se a raiz quadrada de 88 e temos finalmente o valor aproximado σ =9,38.

Utilizaremos o mesmo procedimento para calcular o desvio-padrão de outros dois novos conjuntos de dados. Para tornar mais imediata a comparação entre os valores das dispersões nos três exemplos, mantivemos o tipo de situação descrita e o valor da média inalterados.

Observando os dados da tabela 26, temos imediatamente a sensação de que, nesse caso, a divisão da conta em partes iguais seria mais injusta que no caso anterior.

De modo formal, temos inicialmente um intervalo que vai de 11 a 48 e é maior neste caso que no primeiro exemplo. Essa medida de dispersão apoia, portanto, essa sensação.

Mas, como já dissemos, o intervalo só nos informa a respeito de extremos e é mais interessante analisar o desvio-padrão. O cálculo da variância e do desvio-padrão foi efetuado seguindo o mesmo procedimento já descrito. O valor obtido é de 15,19, maior que o encontrado anteriormente.

Pessoa	xi	xi-$\overline{x}$	(xi-$\overline{x}$)^2
1	36	6	36
2	42	12	144
3	48	18	324
4	11	-19	361
5	13	-17	289
Total	**150**	**–**	**1.154**
Média	**30**	**variância**	**230,8**
		desvio	**15,19**

Tabela 26 – Cálculo da variância e do desvio-padrão para o segundo exemplo de gastos na lanchonete.

Analisando agora os dados da tabela 27, vemos que, assim como no segundo conjunto, o intervalo vai de 11 a 48. No entanto, os dados nos mostram que três pessoas pagaram praticamente o mesmo que consumiram, sendo que as outras duas tiveram consumo mais distante da média.

O cálculo do desvio-padrão, efetuado novamente seguindo o procedimento descrito, tem por resultado o valor 11,71. Esse valor é menor que o do segundo exemplo e maior que o do primeiro, mostrando que a dispersão no último caso é intermediária, ficando entre as dispersões dos dois casos anteriores.

Pessoa	xi	xi-$\overline{x}$	(xi-$\overline{x}$)^2
1	30	0	0
2	31	1	1
3	48	18	324
4	11	-19	361
5	30	0	0
Total	**150**	**0**	**686**
Média	**30**	**variância**	**137,2**
		desvio	**11,71**

Tabela 27 – Cálculo da variância e do desvio-padrão para o segundo exemplo de gastos na lanchonete.

5.4.2 Dados agrupados

No intuito de construir o procedimento para o caso de dados agrupados, vamos calcular a variância e o desvio-padrão para um conjunto de dados em que haja valores repetidos.

A partir de agora, passaremos a utilizar a fórmula da variância para amostras, visto que sua utilização é mais frequente, pois é mais comum termos estudos que utilizam amostras que estudos que trazem informações sobre toda a população. Esse cálculo está colocado na tabela 28.

Dado	xi	xi-$\overline{x}$	(xi-$\overline{x}$)^2
1	10	-10	100
2	30	10	100
3	10	-10	100
4	30	10	100
5	30	10	100
6	30	10	100
7	20	0	0
8	20	0	0
9	0	-20	400
10	10	-10	100
11	30	10	100
Somas	**220**	**–**	**1.200**
Média	**20**	**variância**	**120**
		desvio	**10,95**

Tabela 28 – Cálculo do desvio-padrão para um conjunto de dados brutos de uma amostra.

Assim como no cálculo da média quando há valores repetidos, podemos agrupar os dados e fazer com que os termos referentes a valores iguais sejam multiplicados pelas suas frequências. Desse modo, todos os dados serão incluídos na conta de uma maneira mais prática e concisa.

Vamos ver agora um procedimento para o cálculo do desvio-padrão quando os dados estão agrupados em uma tabela de frequências. Para isso tomaremos os mesmos dados da tabela 28, mas já agrupados numa tabela de frequências. Os dados agrupados e a tabela com os valores utilizados no cálculo do desvio-padrão estão na tabela 29.

xi	fi	fi*xi	xi-$\overline{x}$	(xi-$\overline{x}$)^2	fi*(xi-$\overline{x}$)^2
0	1	0	-20	400	400
10	3	30	-10	100	300
20	2	40	0	0	0
30	5	150	10	100	500
Soma	**11**	**220**			**1.200**
Média	**20**			**variância**	**120**
				desvio	**10,95**

Tabela 29 – Cálculo do desvio-padrão para um conjunto de dados amostrais agrupados.

Comparando o procedimento utilizado na tabela 28, em que tínhamos os dados brutos, vemos que para a tabela de frequências acrescentamos duas colunas necessárias para colocar os resultados dos produtos pelas frequências. Descrevemos, a seguir, o uso dessa tabela para realizar os cálculos:

- A primeira e a segunda colunas trazem os dados para valor e respectiva frequência.
- O somatório de f_i, apresentado na primeira linha adicional da segunda coluna, dá-nos o total de dados no conjunto, ou seja, $N = 11$.
- Como cada valor x_i aparece no conjunto f_i vezes, o somatório de fi*xi, apresentado na primeira linha adicional da terceira coluna, dividido por N, tem como resultado a média: $\bar{x} = \frac{220}{11}$, logo, $\bar{x} = 20$.
- A quarta coluna traz a diferença entre cada valor e a média.
- A quinta coluna traz o quadrado dessa diferença.
- A última coluna traz o produto de cada valor da coluna anterior pela frequência respectiva.
- Como cada valor x_i aparece no conjunto f_i vezes, também a diferença entre o valor x_i e a média aparecerá f_i vezes. Para o cálculo da variância utilizaremos o somatório de $f_i.(xi\text{-}x)^2$, cujo resultado está colocado na primeira linha adicional. Assim, lembrando que ao utilizar a fórmula para amostras devemos subtrair 1 no denominador, teremos a variância dada pela expressão: $\sigma^2 = \frac{1.200}{11-1}$, logo, $\sigma^2 = 120$.
- Para a obtenção do desvio-padrão, calcula-se a raiz quadrada de 120 e temos finalmente o valor aproximado $\sigma = 10{,}95$.

Vale enfatizar que os resultados obtidos dessa maneira teriam necessariamente que ser iguais aos obtidos quando utilizamos os dados brutos da tabela 27, afinal em ambos os casos temos o mesmo conjunto de dados, apenas com apresentação diferente.

Vejamos agora um exemplo adicional em que os dados amostrais são apresentados diretamente numa tabela de frequências. Além disso, nesse exemplo utilizaremos dados cuja média não é inteira, o que é mais frequente na realidade. Os dados e os resultados estão colocados na tabela 30. Procure identificar ali cada um dos passos do procedimento descritos anteriormente.

xi	fi	fi*xi	xi-$\bar{x}$	(xi-$\bar{x}$)^2	fi*(xi-$\bar{x}$)^2
5	5	25	-4,68	21,9	109,51
8	8	64	-1,68	2,82	22,58
12	9	108	2,32	5,38	48,44
15	3	45	5,32	28,3	84,91
Soma	**25**	**242**			**265,44**
Média	**9,68**			**variância**	**10,62**
				desvio	**3,26**

Tabela 30 – Cálculo do desvio-padrão para um conjunto de dados amostrais agrupados numa tabela de frequências.

5.4.3 Médias ponderadas

Vamos ver agora um procedimento para o cálculo quando os dados têm pesos diferentes.

Lembramos que, para efeito dos cálculos, o peso e a frequência têm papéis similares. Assim sendo, montaremos a tabela e efetuaremos os cálculos seguindo os mesmos passos descritos para a tabela 29, mas tendo em lugar das frequências f_i os pesos p_i.

Os dados e os cálculos estão esquematicamente apresentados na tabela 31 colocada a seguir.

xi	pi	pi*xi	xi-$\overline{x}$	(xi-$\overline{x}$)^2	pi*(xi-$\overline{x}$)^2
5	5	25	-4,68	21,9	109,51
8	8	64	-1,68	2,82	22,58
12	9	108	2,32	5,38	48,44
15	3	45	5,32	28,3	84,91
Soma	**25**	**242**			**265,44**
Média	**9,68**			**variância**	**10,62**
				desvio	**3,26**

Tabela 31 – Cálculo do desvio-padrão para um conjunto de dados amostrais com ponderação.

6 NOÇÕES GERAIS DE PROBABILIDADE

Utilizamos o conceito intuitivo de probabilidade em diversas situações de nossas vidas, diariamente. Antes de sair de casa, analisamos a probabilidade de chover para levarmos ou não o guarda-chuva. Não pensamos, porém, se deveríamos levar uma roupa elegante para o caso de uma festa formal inesperada.

Em ambos os casos, lançamos mão do conceito intuitivo de probabilidade para tomar nossa decisão.

O objetivo desta seção é utilizar esse conceito intuitivo como ponto de partida para a construção do conceito formal de probabilidade enquanto frequência relativa de determinados eventos de interesse.

6.1 Variáveis aleatórias

As probabilidades dizem respeito a situações em que existe aleatoriedade. Ou seja, em que o resultado a ser obtido depende de fatores imponderáveis do acaso.

Em estatística, quando falamos em um resultado, ele se expressa no valor de uma variável. Se o valor depende do acaso, a variável que expressa esse valor é chamada de variável aleatória.

Podemos chamar de variável aleatória, por exemplo, o resultado de um jogo de par ou ímpar, sendo que a variável "resultado" poderia assumir os valores "par" ou "ímpar".

Cada resultado de uma variável aleatória terá uma chance, maior ou menor, de ser observado. Estabelecer a magnitude dessas chances é o que se busca no cálculo de probabilidades.

6.2 Noções intuitivas de probabilidade

Todos temos alguma noção intuitiva a respeito de probabilidades, pois sabemos avaliar que coisas serão mais ou menos prováveis em diversas situações, ainda que de maneira subjetiva.

Um exemplo de utilização dessa capacidade é a compra de presente de um amigo-secreto ao qual não conhecemos bem. Para essa compra, buscamos avaliar que tipo de coisas a pessoa em questão deve gostar, baseando-nos no conhecimento que temos de outras pessoas ou de determinados grupos sociais. Uma decisão possível, dado que a maioria das pessoas gosta de música, seria comprar um CD que estivesse fazendo sucesso no momento. Ou seja, pensando assim estamos assumindo que o evento "gostar do presente" será mais provável se escolhermos algo que é frequentemente apreciado por outras pessoas.

Observação

Uma boa análise das probabilidades envolvidas em uma situação nos ajuda a tomar boas decisões.

A situação descrita anteriormente reflete a noção intuitiva da probabilidade de um evento. A partir de agora, refinaremos essa noção, transformando-a em algo que se possa quantificar.

6.3 Probabilidade enquanto frequência relativa

Diz-nos a intuição que o que é mais comum é mais provável. Se escrevermos essa ideia em termos do formalismo matemático, teremos a definição de probabilidade enquanto frequência relativa.

6.3.1 Formalização

Para determinar a probabilidade de que ocorra um determinado evento **E** como resultado de uma variável aleatória, precisamos analisar quantos são os resultados possíveis em geral e quantos são aqueles favoráveis ao evento **E**.

A probabilidade de o evento **E** ocorrer, que será denotada por P(E), será a razão entre o número específico de eventos que são favoráveis a **E**, ao qual chamaremos n_E, pelo número total de eventos possíveis, ao qual chamaremos n_{tot}. O conjunto de todos os eventos possíveis também é chamado de "espaço amostral".

Formalmente, escrevemos:

$$P_E = \frac{n_E}{n_{tot}}$$

Decorre daí que uma probabilidade P será necessariamente um número entre 0 e 1, pelos motivos que seguem:

- A menor probabilidade possível está relacionada ao menor número possível de eventos favoráveis a **E**. O número de eventos favoráveis a **E** será, no mínimo, zero, visto que uma contagem de eventos não pode ser negativa. Assim sendo, a menor probabilidade possível é zero.

- A maior probabilidade possível está relacionada ao maior número possível de eventos favoráveis a **E**. O número de eventos favoráveis a **E** será, no máximo, igual ao número total de eventos possíveis. Dessa forma, n_E será igual a n_{tot} e a divisão de um pelo outro será igual a 1.

É bastante comum falar de porcentagens utilizando a notação percentual. Assim, uma probabilidade 0,6 seria descrita como uma probabilidade de 60%. Note, porém, que temos aqui somente duas maneiras de escrever o mesmo valor. No decorrer da apostila utilizaremos ambas as notações para descrever probabilidades a depender da conveniência da discussão.

Lembrete

A porcentagem é somente uma maneira de escrever uma fração cujo denominador é 100, logo, não há diferença entre uma probabilidade de 0,5 e de 50%.

Exemplo

Numa festa de escola são realizados alguns sorteios de brindes entre os alunos, cujas idades estão apresentadas na tabela 32.

Idade	Quantidade
6	12
7	20
8	17
9	21
10	15
Total	**85**

Tabela 32 – Frequências para as idades das crianças participantes de uma festa escolar.

Para calcular a probabilidade de um aluno de 8 anos de idade ganhar o brinde num determinado sorteio seguimos os passos descritos abaixo:

- Temos 85 alunos no total, ou seja, o número de eventos possíveis nesse caso é 85, ou, seguindo a notação proposta: $n_{tot} = 85$.

- Como queremos analisar o evento "aluno de 8 anos sorteado", chamamos de "número de eventos favoráveis" ao número de crianças dessa idade. Dos 85 alunos, 17 têm 8 anos de idade, assim, há 17 eventos favoráveis no conjunto. Novamente, de acordo com a notação proposta: $n_8 = 17$.

- O cálculo da probabilidade se dá dividindo-se o número de eventos favoráveis pelo número total de eventos possíveis. Chamaremos P(8) a probabilidade em questão. Assim, temos:

$$P(8) = \frac{n_8}{n_{tot}} = \frac{17}{85}$$

E, portanto:

P(8) = 0,2.

Se houver preferência pela notação percentual, podemos dizer que essa probabilidade é de 20%.

Suponhamos agora que haja um brinde que seja de interesse apenas dos alunos maiores, com mais de 7 anos. Assim, o sorteio desse brinde seria feito somente com parte das crianças. Como calcularíamos a probabilidade de um aluno de 8 anos ser sorteado?

Para esse sorteio, embora haja um mesmo número de alunos de 8 anos, temos um menor número de eventos possíveis, o que aumenta a probabilidade de um deles ser sorteado. Formalmente, realizamos os cálculos seguindo um procedimento similar ao citado anteriormente:

- Temos agora somente 53 alunos com mais de 7 anos, ou seja, o número de eventos possíveis nesse caso é 53, ou $n_{tot} = 53$.

- O número de crianças com 8 anos permanece o mesmo, portanto, $n_8 = 17$.

- O cálculo da probabilidade se dá dividindo-se o número de eventos favoráveis pelo número total de eventos possíveis. Agora temos:

$$P(8) = \frac{n_8}{n_{tot}} = \frac{17}{53}$$

E, portanto:

P(8) = 0,32.

Por simplicidade, arredondamos o resultado para 2 casas decimais. Novamente, se houver preferência pela notação percentual, podemos dizer que essa probabilidade é de aproximadamente 32%.

6.4 Alguns cuidados na interpretação de uma probabilidade

Elencamos alguns cuidados para a interpretação de resultados de probabilidade que podem levar a conclusões equivocadas.

6.4.1 Caso geral

Um primeiro cuidado, e talvez o mais importante, é lembrar que a portabilidade dos resultados para as probabilidades calculadas a partir de certo conjunto de dados só vale se a situação descrita for similar àquela em questão.

Retomando o exemplo inicial, o do amigo-secreto, realizado entre os membros de uma orquestra, é mais provável que o presenteado, dadas as características específicas desse grupo, goste de ganhar um CD de música clássica e não se anime com o lançamento dos últimos sucessos populares, ao contrário do que se observaria na maioria dos casos.

Retomando o exemplo das crianças da festa, podemos pensar que as mesmas pessoas, 10 anos depois, não gostariam de ganhar aqueles mesmos brindes pelos quais tanto se interessaram na década anterior.

É claro que nos casos citados a distinção é óbvia e dificilmente alguém utilizaria dados de uma situação na outra. No entanto, é comum que se utilizem estudos gerados em um país para analisar a economia de outro, ou produtos com diferentes especificações etc. Há vezes em que a utilização é válida, mas em outras não. Assim, busque ter um olhar crítico ao trabalhar com informações que envolvam probabilidades.

6.4.2 Quando a probabilidade é 0

Quando a probabilidade de um evento é zero, isso não quer dizer que ele não ocorrerá. Quer dizer somente que entre os dados disponíveis não havia nenhum que correspondesse ao evento em questão.

Temos como exemplos de casos assim todos os eventos historicamente novos ou aqueles que são extremamente raros.

No entanto, tudo aquilo que é impossível terá, necessariamente, probabilidade nula.

6.4.3 Quando a probabilidade é 1

Do mesmo modo que a probabilidade nula (zero) não quer dizer que algo seja totalmente impossível, também a probabilidade de valor 1 (ou 100%) não significa certeza absoluta de que algo acontecerá.

Entram nessa categoria os eventos cuja não ocorrência é extremamente rara ou são aqueles que acabam não ocorrendo por causa de um evento imponderável e imprevisível.

Do mesmo modo, algo que seja certeza terá probabilidade igual a um.

6.5 Origem dos dados

Quando estudamos probabilidades, podemos analisar situações em que os valores conhecidos das variáveis são empíricos ou analíticos. Na sequência definiremos cada um deles.

Os dados analíticos e os empíricos são tratados de maneira diferente. Passamos agora a discutir essa distinção, mostrando como utilizar os dados de ambos os tipos.

6.5.1 Dados empíricos

Dados empíricos são aqueles cujos valores são observados na prática. Fazem parte dessa classificação todos os dados oriundos de pesquisas de campo, como a idade das pessoas de certo grupo, os valores de preços de mercado etc.

Para efeitos didáticos, os dados do tipo empírico utilizados não foram retirados da realidade, mas simulam valores que poderiam ter sido encontrados dessa maneira.

6.5.2 Dados analíticos

Os dados analíticos têm um caráter diferente, eles não precisam ser medidos diretamente, visto que a análise das características do sistema estudado já nos dá os valores possíveis da variável aleatória, bem como a proporção em que eles se encontram.

Como exemplo dessa classe de dados temos os jogos de azar, como o jogo de uma moeda, o jogo de dados ou o sorteio de cartas, por exemplo.

Nesta seção vamos realizar a análise das probabilidades referentes a alguns desses sistemas.

1. Jogo de uma moeda

Quando jogamos uma moeda, sabemos que haverá dois resultados possíveis: face cara ou face coroa. Em princípio, podemos assumir que a moeda é equilibrada e que a ocorrência de uma ou outra face dependerá somente do acaso e com igual proporção.

Para analisar as probabilidades referentes a esse caso, chamemos o resultado "face cara" de K e o resultado "face coroa" de C, para facilitar a identificação.

Chamando o número de eventos possíveis cara, coroa e total de n_K, n_C e n_{tot}, respectivamente, teremos: $n_K = 1$, $n_C = 1$ e $n_{tot} = 2$.

A partir desses dados, podemos calcular as probabilidades de cada evento possível:

- Probabilidade de face cara: $P(K) = n_K / n_{tot}$, logo $P(K) = 1/2$.

- Probabilidade de face coroa: $P(C) = n_C / n_{tot}$, logo $P(C)=1/2$.

2. Jogo de um dado

Podemos fazer uma análise similar para o jogo de um dado. Novamente supomos o dado equilibrado, em que todas as faces são igualmente prováveis.

Chamaremos de P(n) a probabilidade de termos o resultado "face n". Assim, P(1) simboliza a probabilidade de termos o resultado 1 no jogo do dado.

Como há 6 faces e todas são igualmente prováveis, teremos $n_{tot} = 6$ e, para qualquer face:

- Probabilidade de face 1: $P(1) = n_1 / n_{tot}$, logo $P(1) = 1/6$.

- Probabilidade de face 4: $P(4) = n_4 / n_{tot}$, logo $P(4) = 1/6$.

Podemos ainda calcular as probabilidades de algumas situações adicionais, como a de sortearmos um número par. As faces cujo resultado atende a esse critério são as faces 2, 4 e 6. Logo, temos 3 resultados favoráveis ao critério escolhido. O número total de resultados possíveis é o mesmo ($n_{tot} = 6$), logo:

- Probabilidade de face par: $P(par) = n_{par} / n_{tot}$, logo, $P(par) = 3/6$.

- Probabilidade de face par: $P(par) = 1/2$.

Exemplos de aplicação

1) Um vendedor de automóveis deseja impressionar possíveis compradores de certo modelo. Para isso apregoa a grande possibilidade de personalização do automóvel, fornecendo opções de escolha entre três tipos de motor, dois tipos de transmissão, cinco cores externas e duas internas. Quantas opções ele pode oferecer ao cliente?

2) Se um torneio de basquete consiste na participação de 36 times, supondo que todos os times tenham exatamente o mesmo nível, pergunta-se:

   a) De quantas maneiras podem ser conquistados os três primeiros lugares?

   b) De quantas maneiras podem ser conquistadas as três vagas para a participação num campeonato internacional?

3) Um fabricante estima haver defeito em cerca de 2% de seus produtos. Se tal suspeita é correta, estime a probabilidade de que, numa amostra de 9 unidades, haja:

   a) Uma defeituosa.

b) Nenhuma defeituosa.

4) Doze por cento dos que reservam lugares em um voo, sistematicamente faltam ao embarque. O avião comporta 15 passageiros.

a) Determine a probabilidade de todos os que reservaram voo comparecerem ao embarque.

b) Se houver 16 pedidos de reserva, determine a probabilidade de:

i) Uma pessoa ficar de fora.

ii) Nenhuma pessoa ficar de fora.

5) Um exame é constituído de 100 testes com 5 alternativas cada. Qual a nota esperada de um estudante que não sabe nada da matéria? Qual a variância correspondente a essa distribuição?

6) Um teste de escolaridade tem distribuição normal com média 100 e desvio-padrão 10. Determine a probabilidade de um indivíduo que se submeta ao teste ter nota:

a) Maior que 120.

b) Maior que 80.

c) Entre 85 e 115.

d) Maior que 100.

Resolução

1) Utilizando o princípio da multiplicação: $N=3 * 2 * 5 * 2$, logo $N = 60$.

2)

a) $N = 36 * 35 * 34 = 42840$

b) $N = 42840/3! = 42840/6 = 7140$

3) Distribuição binomial com $n = 9$ e $p = 0{,}02$.

a) $k = 1$, $P(1) = 0{,}1531$

b) $k = 0$, $P(0) = 0{,}8337$

4) Binomial com $p = 0{,}12$ (probabilidade de faltar).

a) $n = 15$, $k = 0$: $P(0) = 0{,}1470$

b) $n = 16$

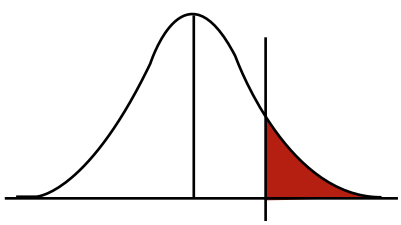
i) Só fica alguém de fora se ninguém faltar, logo, $k = 0$. $P(0) = 0{,}1293$.

ii) Resultado complementar daquele discutido no item anterior, logo:
$P = 1 - 0{,}1293 = 0{,}8707$

5) Média: $\bar{x} = 100.0{,}2 = 20$; desvio-padrão: $\sigma = 100.0{,}2.0{,}8 = 16$.

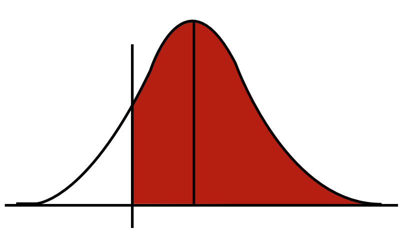
6)

a) Maior que 120.



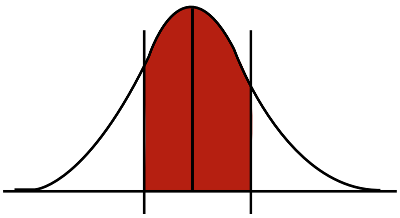
$z = (120-100)/10 = 2$
$A(2) = 0{,}4772$
$P(x>120) = P(z>2) = 0{,}5 - A(2) = 0{,}0228$

b) Maior que 80.



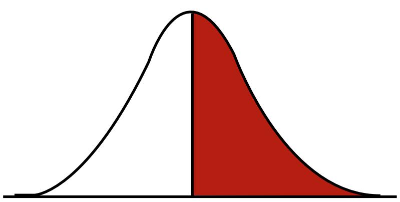
$z = (120-100)/10 = 2$
$A(2) = 0{,}4772$
$P(x>120) = P(z>2) = 1-A(2) = 0{,}5228$

c) Entre 85 e 115.



z1 = (85-100)/10 = -1,5, z2 = (115-100)/10 = 1,5
A(1,5) = 0,4332
P(x>120) = P(-1,5<z>1,5) = A (1,5)+A(1,5) = 0,8664

d) Maior que 100.



z = (100-100)/10 = 0
P(x>100) = P(z>0) = 0,5

Resumo

Medidas de posição e dispersão

1. As medidas de posição e dispersão têm o papel de resumir as características principais de um conjunto.

2. São medidas de posição, entre outras, a moda, a mediana e a média.

3. As medidas de dispersão nos informam a respeito da concentração dos valores individuais em torno dos valores médios.

4. São medidas de dispersão o intervalo, a variância e o desvio-padrão.

Exercícios

Questão 1. Muitos países enfrentam sérios problemas com seu elevado crescimento populacional. Em alguns foi proposta (e por vezes colocada em efeito) a proibição de as famílias terem mais de um filho. Algumas vezes, no entanto, essa política teve consequências trágicas (por exemplo, em alguns países houve registros de famílias de camponeses que abandonaram suas filhas recém-nascidas para terem mais uma chance de ter um filho do sexo masculino). Por essa razão, outras leis menos restritivas foram consideradas. Uma delas foi a de que as famílias teriam direito a um segundo (e último) filho, caso o primeiro fosse do sexo feminino. Suponha que essa última regra fosse seguida por todas as famílias de certo país (isto é, sempre que o primeiro filho fosse do sexo feminino, fariam uma segunda e última tentativa para ter um menino). Suponha ainda que, em cada nascimento sejam iguais as chances de nascer menino ou menina. Examinando os registros de nascimento, após alguns anos de a política ter sido colocada em prática, seria esperado que:

A) O número de nascimentos de meninos fosse aproximadamente o dobro do de meninas.

B) Em média, cada família tivesse 1,25 filho.

C) Aproximadamente 25% das famílias não tivessem filhos do sexo masculino.

D) Aproximadamente 50% dos meninos fossem filhos únicos.

E) Aproximadamente 50% das famílias tivessem um filho de cada sexo.

Resposta correta: alternativa C.

Análise das alternativas

Nas condições do exposto no exercício, seria possível a existência de quatro tipos de famílias:

- Famílias com um único filho homem (o primeiro e único a nascer). A probabilidade de essa família existir é de 50%, pois as probabilidades de nascer um homem ou uma mulher são iguais, logo, no primeiro nascimento existe uma probabilidade de 50% de que o filho seja homem.
- Famílias com uma primeira filha mulher e um segundo filho homem. A probabilidade de essa família existir é de 25%, porque 50% das famílias têm a primeira filha mulher, e dessas, metade (50%) terá um segundo filho homem, ou seja, 50% de 50% é igual a 25%.
- Famílias com duas filhas mulheres, ou seja, nenhum filho homem. É o que sobra de probabilidade, ou seja, 25%. Assim sendo, a resposta correta corresponde à alternativa C.

A) Alternativa incorreta.

Justificativa: o próprio enunciado diz que a probabilidade de nascerem meninas e meninos é igual.

B) Alternativa incorreta.

Justificativa: na situação descrita teríamos 50% de famílias com dois filhos e 50% de famílias com um filho, ou seja, em média 1,5 filho e não 1,25 filho.

C) Alternativa correta.

Justificativa: são os casos das famílias com duas filhas mulheres, como visto acima, existem 25% de probabilidades de isso ocorrer.

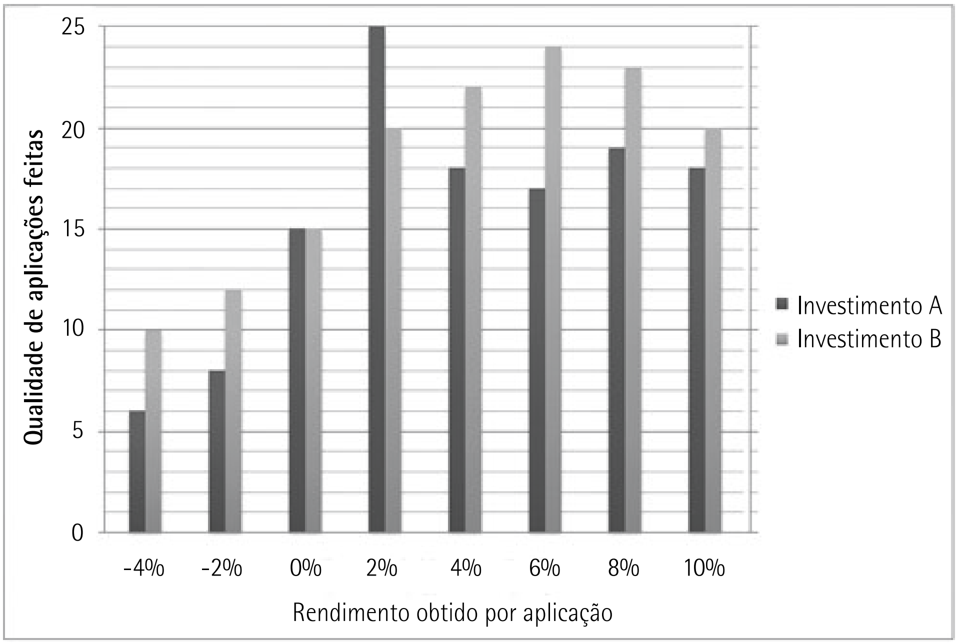
D) Alternativa incorreta.

Justificativa: nas condições dadas, 50% das famílias terão filhos únicos homens e não 50% dos meninos serão filhos únicos.

E) Alternativa incorreta.

Justificativa: apenas 25% das famílias terão um filho de cada sexo, como visto acima.

Questão 2. Determinado investidor resolve mandar fazer um estudo estatístico antes de aplicar parte de suas reservas. Para tanto, foram levantados os dados relacionados no gráfico abaixo referente a dois investimentos diferentes.



Ele deseja escolher o investimento baseado não só na rentabilidade média, mas também na estabilidade do investimento (variações pequenas). O especialista encarregado de fazer o estudo analisa os dados e faz as afirmações:

I. O investimento B é mais recomendado que o investimento A porque tem maior rendimento médio.

II. O investimento A é mais rentável e tem menor variabilidade.

III. Apesar de ambos os investimentos terem rentabilidades próximas, o investimento A tem maior variabilidade, o destaca preferencialmente.

IV. As rentabilidades dos dois investimentos são muito próximas, mas ambas têm uma dispersão muito elevada, o que torna as análises de baixa qualidade.

V. O investimento A é mais recomendado que o investimento B porque tem um desvio-padrão menor.

Selecione a alternativa correta:

A) I, II e III.

B) II, III e IV.

C) III, IV e V.
D) I, IV e V.

E) II, IV e V.

Resolução desta questão na Plataforma.