Métodos Quantitativos Estatísticos

Denise Maria Martins

Denise Maria Martins

Métodos Quantitativos Estatísticos

Edição revisada

IESDE Brasil S.A.
Curitiba
2012

CIP-BRASIL. CATALOGAÇÃO-NA-FONTE
SINDICATO NACIONAL DOS EDITORES DE LIVROS, RJ

M341m

Martins, Denise Maria
Métodos quantitativos estatísticos / Denise Maria Martins. - 1.ed., rev. - Curitiba, PR : IESDE Brasil, 2012.
138p. : 24 cm

Inclui bibliografia
ISBN 978-85-387-2986-0

1. Estatística matemática. 2. Probabilidades. I. Título.

12-5023. CDD: 519.5
CDU: 519.2

16.07.12 31.07.12 037526

Capa: IESDE Brasil S.A.
Imagem da capa: Shutterstock

Inteligência Educacional

IESDE Brasil S.A.
Al. Dr. Carlos de Carvalho, 1.482. CEP: 80730-200
Batel – Curitiba – PR
0800 708 88 88 – www.iesde.com.br

Denise Maria Martins

Mestre em Administração Estratégica pela Universidade Cidade de São Paulo (Unicid). Especialista em Engenharia da Qualidade pela Universidade Católica de Minas Gerais (PUC-Minas). Graduada em Estatística pelo Centro Universitário Capital (Unicapital). Atua como gestora de processos.

sumário

Estatística com aplicações e análise exploratória 11

11 | Estatística: definição e aplicações
13 | Conceitos e regras
20 | Análise exploratória de dados: o problema

Medidas de tendência central e posição 35

35 | Definição
47 | Quartis, Decis e Percentis

Medidas de variabilidade 55

55 | Definição
60 | O problema

Introdução à probabilidade e distribuição de probabilidade discretas 75

Distribuição de probabilidade contínua 113

113 | Definição

132 | Apêndice n.º 1 – Tabela 1: Áreas sob a curva normal

Referências 137

Apresentação

O assunto da estatística pode ser apresentado em diversos níveis de dificuldades matemáticas e orientado para aplicações em vários campos de pesquisa. A consequência é apresentação de textos sobre estatística média, estatística para administração, estatística educacional, estatística psicológica e até mesmo estatística para historiadores. Embora os problemas que surgem nessas diversas áreas por vezes exijam técnicas estatísticas especiais, nenhum dos métodos básicos apresentados neste livro está restrito a qualquer campo particular de aplicação.

Os objetivos gerais da Estatística aplicados à Administração são os seguintes:

- Desenvolver a confiança dos alunos ao lidar com dados numéricos;
- Expor o leitor a uma ampla variedade de técnicas estatísticas introdutórias para uso na interpretação e análise de dados empresariais.

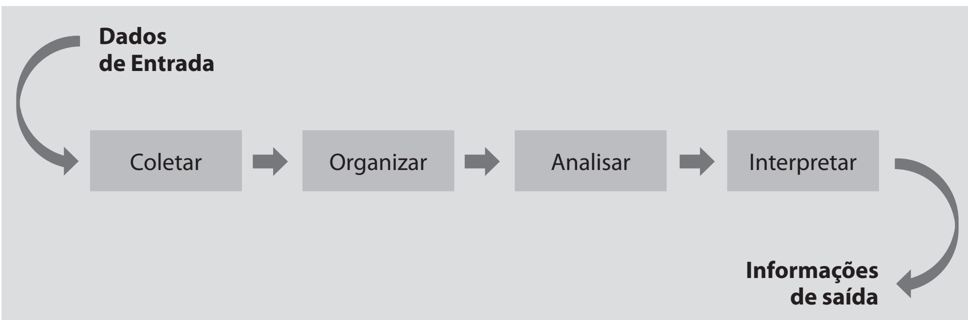
No livro, temos a descrição das principais ferramentas e métodos, apresentando a estatística e a análise exploratória, em que são indicados conceitos básicos e ferramentas para análise gráfica. A aplicação da estatística envolve medidas de tendência central e de posição, evidenciando métodos e ferramentas para indicar valores que representam a maioria dos dados de forma resumida. O desenvolvimento e o entendimento de medidas de variabilidade que permitem ao leitor acrescentar a sua interpretação e análise de ferramentas para medir as dispersões de amostras analisadas. Temos a "Introdução à probabilidade e distribuições de probabilidade discretas" demonstrando métodos para tratamento de dados discretos e indicando cálculos de probabilidade. Como encerramento é apresentada a distribuição de probabilidade contínua, representada pela distribuição normal como meio para a realização de previsões em determinados contextos onde temos variável contínua como dados.

Estatística com aplicações e análise exploratória

Estatística: definição e aplicações

No moderno ambiente administrativo e econômico global, dispõe-se de uma vasta quantidade de informações estatísticas. Os gerentes e tomadores de decisão de maior sucesso são aqueles capazes de entender a informação e usá-la de forma eficaz. As empresas precisam de informações para tomar as decisões: parte dessas informações será transformada em dados e em análise estatística.

Uma definição de dicionário afirma que estatística é a apresentação de fatos numéricos coletados sistematicamente, ordenados e estudados. Para aqueles que tomam decisões, o principal papel da estatística é fornecer-lhes os métodos para obter e converter dados (valores, fatos, observações, medições) em dados úteis.



(SMAILES, 2007)

Figura 1 – Definição de Estatística.

O objetivo da estatística é proporcionar conhecimento a partir de dados. Os dados são números que representam um contexto. Exemplo: o número 4,8, não dá, por si só, qualquer informação. Mas se o filho recém-nascido de um amigo pesa 4,8 quilos, congratula-se pelo tamanho do filho. O contexto motiva o conhecimento fundamental e permite fazerem-se julgamentos.

Do ponto de vista prático, pode-se dividir a estatística em três partes.

- **Análise de dados** (estatística descritiva): consiste em métodos e ideias para organizar e descrever dados mediante a utilização de gráficos, resumos numéricos e descrições matemáticas mais elaboradas. A revolução do computador colocou a análise de dados no centro da prática estatística.

- A **produção de dados** fornece métodos para produzir dados que podem dar respostas mais claras a questões específicas. Os conceitos básicos de como selecionar amostras (técnicas de amostragem) e planejar experimentos são talvez as mais importantes de todas as ideias estatísticas.

- A **inferência estatística** vai além dos dados disponíveis e procura tirar conclusões sobre um universo mais amplo. A inferência estatística não só formula conclusões, como também as acompanha, indicando seu grau de confiabilidade.

Os dados vêm em diferentes formas, cada uma das quais é tratada de maneira um pouco diferente na conversão em informações.

Quadro 1 – Tipo de dados

Dados			
Qualitativos (Categóricos)		Quantitativos	
Nominais	**Ordinais**	**Contínuos**	**Discretos**
↓	↓	↓	↓
Categorias nomeadas	Números agem como categorias/ ordenações	Qualquer valor dentro de um intervalo possível	Somente valores fixos são possíveis (valores absolutos)

(SMAILES, 2007)

Supondo as situações a seguir pode-se entender a importância da estatística e sua aplicação nas atividades em que as decisões devem ser tomadas em um curto espaço de tempo e a um baixo custo.

Contexto A

Uma empresa que está se preparando para lançar um novo produto precisa conhecer as preferências dos consumidores no mercado de interesse. Para isso, pode fazer uma pesquisa de mercado entrevistando um número de residências escolhidas aleatoriamente. Poderá então usar os resultados para estimar as preferências de toda a população.

Contexto B

Um auditor deve verificar os livros de uma empresa, para se certificar de que os lançamentos refletem efetivamente a situação financeira da companhia. Ele deve examinar pilhas de documentos originais, como notas de venda, ordens de compra e requisições. Seria um trabalho incalculável consultar todos os documentos originais; em vez disso, o auditor pode verificar uma amostra de documentos escolhidos aleatoriamente, com base nessa amostra, fazer inferências sobre toda a população.

Contexto C

Um técnico de controle da qualidade deve realizar um ensaio para garantir que o produto fabricado está funcionando conforme as especificações do cliente. Por se tratar de palitos de fósforos, o ensaio destrói o produto, o técnico utiliza seleção de amostras para realizar e concluir sua análise.

Os contextos permitem refletir como analisar uma situação sem a necessidade de levantar todos os dados, isso é possível através da estatística e a aplicação de métodos na coleta de dados que estimam as características da população com base na amostra. Uma das vantagens em utilizar-se de amostras, encontra-se no fato de chegar a uma conclusão e/ou decisão em um tempo menor a um custo reduzido, levando-se também em consideração que em algumas situações o processo de pesquisa destrói o elemento pesquisado.

- Como é dispendioso, difícil e por vezes impraticável ter acesso a toda uma população, costuma-se escolher uma amostra e estudá-la.
- Para evitar predições imprecisas, é essencial que a amostra represente efetivamente a população da qual foi extraída.

Conceitos e regras

Estatística

É uma coleção de métodos para planejar experimentos, obter dados e organizá-los, resumi-los, analisá-los, interpretá-los e deles extrair conclusões.

População (N)

É uma coleção de todos os elementos (valores, pessoas, medidas etc.) a serem estudados.

Amostra (n)

É uma subcoleção de elementos extraídos de uma população.

População **Amostra**

Figura 2 – Definição de população e amostra.

Parâmetro e Estatística

Um Parâmetro é uma medida numérica que descreve uma característica de uma população, enquanto Estatística é uma medida numérica que descreve uma característica de uma amostra.

Dados quantitativos

Consistem em números que representam contagens ou medidas. Ex.: duração de uma música.

Dados discretos

Resultam de um conjunto finito de valores possíveis, ou de um conjunto enumerável desses valores (o número de valores possíveis é 0, ou 1, ou 2 etc.), ou seja, quando os dados representam contagens são discretos.

Ex.: contagens, números de mensagens em uma secretária eletrônica, número de visitas a Londres.

Dados contínuos (numéricos)

Resultam de um número infinito de valores possíveis que podem ser associados a pontos em uma escala contínua de tal maneira que não haja lacunas ou interrupções, ou seja, quando os dados representam mensurações, são contínuos. Ex.: alturas, salários, horários, pesos.

Dados qualitativos (ou dados categóricos, ou atributos)

Podem ser separados em diferentes categorias que se distinguem por alguma característica não numérica. Ex.: estilo de música.

Níveis de mensuração de dados

Outra forma de classificar dados também muito comum.

Quadro 2 – Níveis de mensuração de dados

Nível	Sumário	Exemplo
Nominal	Categorias somente. Os dados não podem ser dispostos em um esquema ordenado.	Carros de alunos 10 Mercedez 20 Ferraris 40 Porsches Categorias ou nomes somente
Ordinal	As categorias são ordenadas, mas não podem estabelecer diferenças, ou estas não têm sentido.	Carros de alunos 10 Compactos 20 Médios 40 Grandes Está determinada uma ordem: "compacto"; "médio" e "grande".
Intervalo	Podemos determinar diferenças entre valores, mas não há ponto de partida inerente. As razões não têm sentido.	Temperatura no campus 45º F 80º F 90º F 90º F não é duas vezes mais quente do que 45ºF
Razão	Como intervalo, mas com um ponto de partida inerente. As razões têm sentido.	Peso de jogadores em uma faculdade 115 lb 195 lb 300 lb 300 lb é duas vezes 150 lb

(TRIOLA, 1999)

Tabelas de frequências

Relaciona categorias (ou classes) de valores, juntamente com contagens (ou frequências) do número de valores que se enquadram em cada categoria.

Exemplo: tabela de frequência da idade dos alunos de Administração.

idade	frequência
De 18 a menos de 19	16
De 19 a menos de 20	7
De 20 a menos de 21	7
De 21 a menos de 22	3
22 ou mais	3
36	

Rol

Um rol é um arranjo de dados numéricos brutos em ordem crescente ou decrescente de grandeza.

Exemplo: 2, 3, 4, 5, 6, 7, 8, 9, 10.

Limites inferiores de classes

São os menores números que podem efetivamente pertencer às diferentes classes.

De 18 a menos de 19

Amplitude de classe + limite inferior da primeira classe

1 + 18 = 19 (limite inferior da segunda classe)

idade	frequência
De 18 a menos de 19	16

limite inferior de classes

Limites superiores de classes

São os maiores números que podem efetivamente pertencer às diferentes classes.

Exemplo:

classes	idade	frequência
1	De 18 a menos de 19	16

limite superior de classes

Amplitude de classe

É a diferença entre dois limites de classe inferiores consecutivos ou entre duas fronteiras inferiores de classe consecutivas.

Exemplo:

	idade	frequência
Diferença entre dois limites de classes inferiores	De 18 a menos de 19	16
	De 19 a menos de 20	7

Fronteiras de classes

São os números usados para separar classes, mas sem as lacunas criadas pelos limites de classe, podendo ser representados pelos intervalos de classes ajustados.

Exemplo:

classes	intervalo de classes
1	18 ├ 19

fronteira de classes

Intervalo de classe

É a amplitude de uma classe, ou intervalo de valores que ela pode conter e é dado pela diferença entre seus limites ou fronteiras.

Exemplo:

classes	intervalo de classes	ponto médio
1	18 ├ 19	18,5
2	19 ├ 20	19,5

Distribuição de frequência

É o agrupamento dos dados em um certo número de classes, intervalos ou categorias.

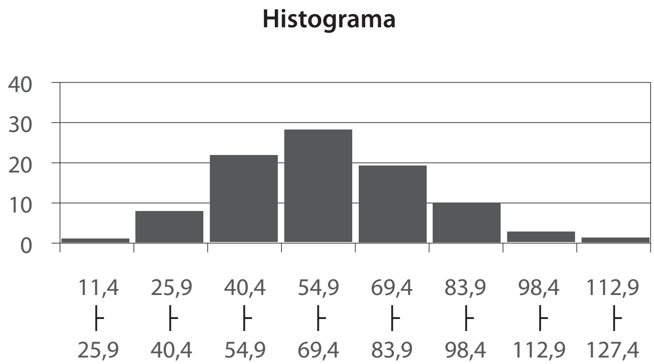
Exemplo:

Idade	frequência
De 18 a menos de 19	16
De 19 a menos de 20	7

Histograma

Consiste em uma escala vertical para as frequências e barras para representar os valores das frequências das diversas classes.

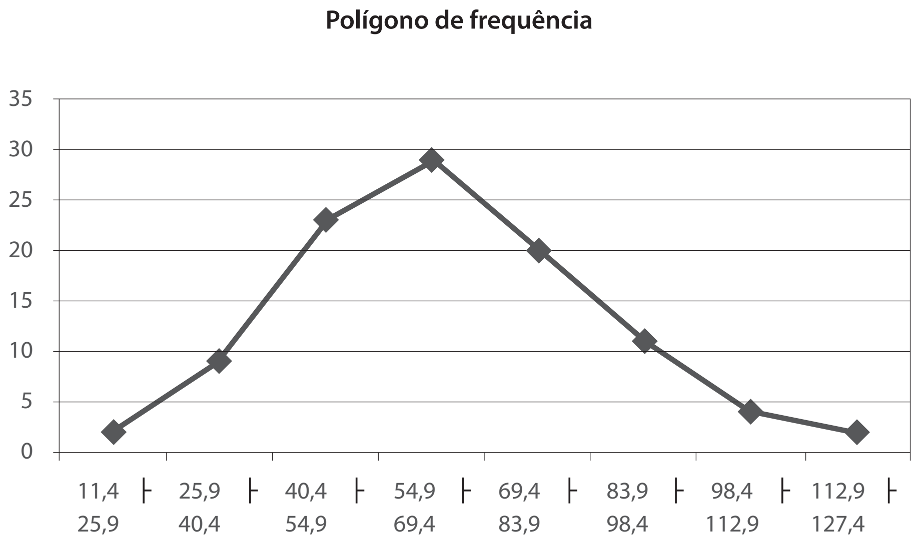
Exemplo:



Polígono de frequência

É uma variante do histograma, sendo que as frequências são marcadas nos pontos médios, e os valores são unidos por segmentos retilíneos.

Exemplo:



Ponto médio

Obtém-se adicionando os limites inferior e superior de uma classe e dividindo-se o resultado por dois.

Exemplo:

classes	intervalo de classes	ponto médio
1	18 ├ 19	18,5
2	19 ├ 20	19,5

Ramo e folhas

É uma tabela em que cada linha representa a posição de um ramo e cada algarismo à direita da reta pode ser considerado uma folha.

Exemplo:

ramo	folhas																					
1	1	5																				
2	6	7	8																			
3	7																					
4	0	0	0	0	0	1	1	1	2	2	2	2	3	3	4	9						
5	0	2	2	3	3	3	3	4	5	5	5	5	5	5	6	6	6					
6	0	0	0	0	0	0	1	3	3	4	6	6	7	7	7	7	8	8	8	8	9	9
7	4	5	5	5	6	6	7	7														
8	0	1	2	2	2	2	3	3	3	4	5	5	6	6	9							
9	0	3	4	5	5	6	7	7	7													
10	0	3	8																			

(TRIOLA, 1999)

Análise exploratória de dados: o problema

Contexto A

A organização e apresentação de informações numéricas é a primeira etapa para entender um problema. Como situação típica, considerando os valores a seguir, que representam o tempo de percurso até o trabalho de empregados de um grande escritório localizado no centro de uma cidade.

Os tempos são em minutos e cada valor representa o tempo médio gasto por um empregado em cinco dias úteis. A simples coleta de dados por si só já não é tarefa simples, mas é claro que é preciso muito mais para tornar os números compreensíveis. O que se pode fazer para tornar esta massa de informações mais utilizável?

Tabela 1 – Coleta de dados de tempo de percurso ao trabalho em minutos

49,6	42,6	43,2	44	50,6	52,6	53,6	42,2	52,4	43
55,4	53,6	54,6	55	55,4	55,8	55,8	53,6	55,6	53,8
60,2	56,4	60	60,2	60,4	60,8	61	56,2	60,4	56,4
67	63,8	66,8	66,8	67,2	67,4	68	63,4	67,4	64
74	68,2	69,8	69,8	75,4	75,8	76	68	75,8	68,8
82	77,4	80	81,2	82,4	82,6	82,8	76	82,6	77,8
85,4	83,2	84	85	86,4	89,6	95,2	83	86,4	83,4
29,4	15,8	37	28,8	30,2	36,8	38,4	11,4	35,8	26
41	40,4	40,6	40,8	41	42	42	40,2	41,4	40,4
103,4	97	97,8	100	108,2	123,8	125,4	96,8	110,2	97

Explorando o problema

Ao analisar um conjunto de dados, deve-se determinar se é uma **amostra** ou uma **população**. Essa determinação afetará não somente os métodos utilizados, mas também suas conclusões. Na sequência utilizar métodos de estatística descritiva (análise dos dados) para resumir ou descrever as características importantes de um conjunto conhecido de dados, considerando a tabela de frequência e o histograma como um método para organizar e resumir os dados coletados.

O objetivo do uso da tabela de frequência e histograma é identificar a natureza ou forma da distribuição dos dados, como forma de sino, uniforme ou assimétrica.

Equacionando o problema

O processo de construção de uma tabela de frequência e gráficos que envolvem os seguintes passos:

Passo 1:

Organizar os dados em ordem crescente ou decrescente de grandeza.

Arranjar os dados em forma de Rol.

Aplicando no exemplo:

Coleta de dados em Rol – tempo de percurso ao trabalho em minutos

11,4	15,8	26	28,8	29,4	30,2	35,8	36,8	37	38,4
40,2	40,4	40,4	40,6	40,8	41	41	41,4	42	42
42,2	42,6	43	43,2	44	49,6	50,6	52,4	52,6	53,6
53,6	53,6	53,8	54,6	55	55,4	55,4	55,6	55,8	55,8
56,2	56,4	56,4	60	60,2	60,2	60,4	60,4	60,8	61
63,4	63,8	64	66,8	66,8	67	67,2	67,4	67,4	68
68	68,2	68,8	69,8	69,8	74	75,4	75,8	75,8	76
76	77,4	77,8	80	81,2	82	82,4	82,6	82,6	82,8
83	83,2	83,4	84	85	85,4	86,4	86,4	89,6	95,2
96,8	97	97	97,8	100	103,4	108,2	110,2	123,8	125,4

Passo 2:

Decidir o número de classes de sua tabela de frequência.

A título de orientação, o número de classes deve ficar entre 5 e 20.

O número efetivo de classes pode depender da conveniência de utilizar números arredondados ou outros fatores subjetivos.

Importante: a relação entre o número de classes (**k**) e o tamanho da amostra (**n**), foi estudada por Sturges, o qual estabeleceu a relação:

Cálculo do número de classes (**k**)

$$\text{k (número de classes)} = 1 + \frac{\log_n}{\log_2}$$

Aplicando no exemplo:

$$k = 1 + \frac{\log_{100}}{\log_2} = 764 \quad \text{ou seja, } \mathbf{k = 7{,}64}$$

Passo 3:

Determinar o maior e o menor valor dos dados organizados.

Calcular a amplitude total do Rol (diferença entre o maior e o menor dos valores coletados).

Aplicando no exemplo:

Amplitude total do Rol = 125,4 – 11,4 ou seja, amplitude total do Rol **= 114**

Passo 4:

Dividir a amplitude total do rol pelo número de classes.

Arredondar o resultado para mais, até um número conveniente. Esse arredondamento para mais, não somente é conveniente como também garante que todos os valores sejam incluídos na tabela de frequências.

Fórmula 1 – Cálculo da amplitude de classe (h)

$$\text{Amplitude de classe (h)} = \frac{\text{R (Amplitude)}}{\text{k (Número de classes)}} \quad \text{(arredondar para mais)}$$

Aplicando no exemplo:

$$\text{Amplitude de classe (h)} = \frac{114}{8} = 14{,}25, \quad \text{ou seja, } \mathbf{14{,}5}$$

(os dados originais trabalham com uma casa decimal)

Passo 5:

Escolher como limite inferior da primeira classe o menor valor observado ou um valor ligeiramente inferior a ele.

Esse valor serve como ponto de partida.

Aplicando no exemplo:

Limite inferior da primeira classe com intervalo fechado para incluí-lo = **11,4**

Passo 6:

Somar a amplitude de classe ao ponto de partida, obtendo o segundo limite inferior de classe.

Adicionar a amplitude de classe ao segundo limite inferior para obter o terceiro; e assim por diante.

Aplicando no exemplo:

Amplitude de classe + limite inferior da primeira classe

11,4 + 14,5 = **25,9** (limite inferior da segunda classe)

Obs.: fazer o mesmo até completar as oito classes.

Passo 7:

Relacionar os limites inferiores de classe em uma coluna e introduzir os limites superiores, que podem ser facilmente determinados a esta altura.

Aplicando no exemplo:

Tabela 2 – Frequência do tempo de percurso ao trabalho em minutos

classes	intervalo de classes
1	11,4 ├ 25,9
2	25,9 ├ 40,4
3	40,4 ├ 54,9
4	54,9 ├ 69,4
5	69,4 ├ 83,9
6	83,9 ├ 98,4
7	98,4 ├ 112,9
8	112,9 ├ 127,4

Passo 8:

Representar cada observação por um pequeno traço na classe apropriada e, com auxílio desses traços, determinar a frequência total de cada classe.

Aplicando no exemplo:

Tabela 3 – Frequência com classes ajustadas: tempo de percurso até o trabalho em minutos

classes	intervalo de classes	intervalo de classe ajustado
1	11,4 ├ 25,9	11,4 a 25,8
2	25,9 ├ 40,4	25,9 a 40,3
3	40,4 ├ 54,9	40,4 a 54,8
4	54,9 ├ 69,4	54,9 a 69,3
5	69,4 ├ 83,9	69,4 a 83,8
6	83,9 ├ 98,4	83,9 a 98,3
7	98,4 ├ 112,9	98,4 a 112,8
8	112,9 ├ 127,4	112,9 a 127,3

Tabela 4 – Frequência com frequência total: tempo de percurso ao trabalho em minutos

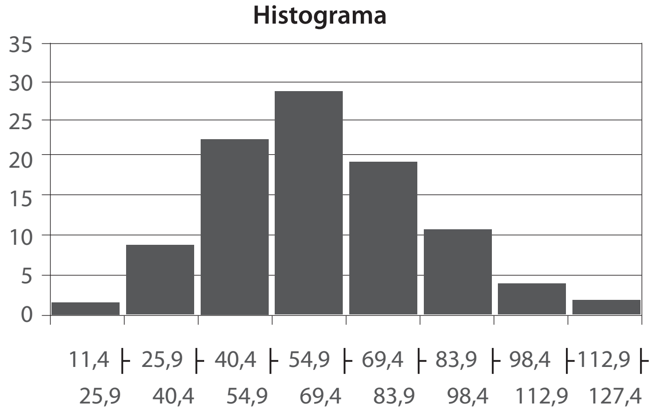
classes	intervalo de classes	frequência
1	11,4 ├ 25,9	02
2	25,9 ├ 40,4	09
3	40,4 ├ 54,9	23
4	54,9 ├ 69,4	29
5	69,4 ├ 83,9	20
6	83,9 ├ 98,4	11
7	98,4 ├ 112,9	04
8	112,9 ├ 127,4	02

Passo 9:

Representar as distribuições de frequência (dados da tabela de frequência) em forma gráfica, conhecida como **histograma**. Um histograma é construído representando-se as medidas ou observações que são agrupadas em uma escala horizontal, e as frequências de classes em uma escala vertical; traçam-se os retângulos, onde as bases são iguais aos intervalos de classes e cujas alturas são as frequências de classe correspondentes.

Aplicando no exemplo:

Gráfico 1 – Histograma dos tempos de percurso ao trabalho



Passo 10:

A finalidade do histograma é a de ajudar a entender os dados. Após ter sido construído deve-se perguntar: "Que é que estou vendo?" – Procurar não só um padrão global, como também desvios acentuados em relação ao mesmo. No caso de um histograma, o padrão global é a forma geral da distribuição. Os valores discrepantes são um tipo importante de desvio em relação ao padrão global. Uma vez localizados os valores discrepantes, deve-se procurar uma explicação. Muitos valores discrepantes são provenientes de erros, outros revelam a natureza especial de algumas observações. A explicação dos valores discrepantes, em geral, requer alguma informação básica do contexto.

Passo 11:

Outra forma, não tanto utilizada de apresentação gráfica é o **polígono de frequência**, onde as frequências são marcadas nos pontos médios de cada intervalo de classe, e os valores sucessivos são unidos por segmentos retilíneos.

Fórmula 2 – Cálculo do ponto médio

$$\text{Ponto médio} = \frac{(X_{\text{maior}} + X_{\text{menor}})}{2}$$

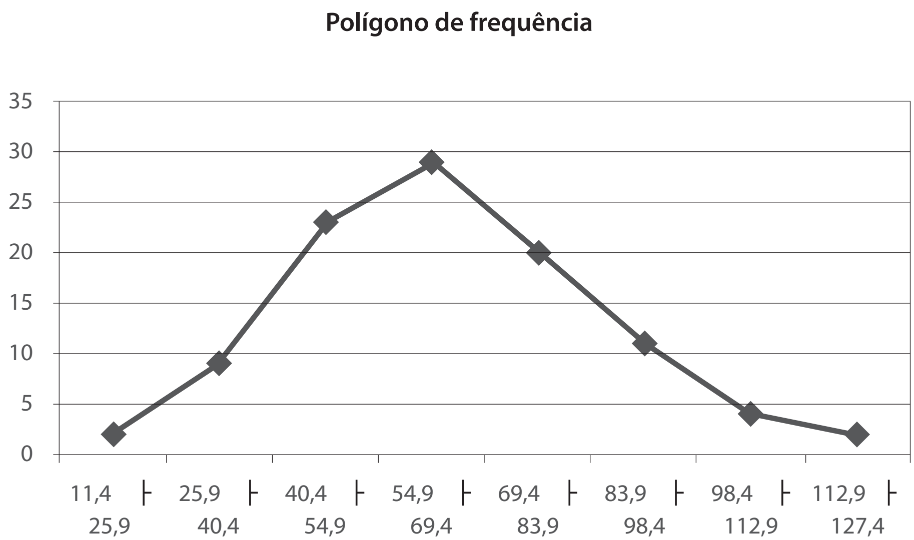
(considerar os valores dos intervalos)

Aplicando no exemplo:

Tabela 5 – Frequência com ponto médio – Tempo de percurso ao trabalho

classes	intervalo de classes	ponto médio	frequência
1	11,4 ├ 25,9	18,65	02
2	25,9 ├ 40,4	33,15	09
3	40,4 ├ 54,9	47,65	23
4	54,9 ├ 69,4	62,15	29
5	69,4 ├ 83,9	76,65	20
6	83,9 ├ 98,4	91,15	11
7	98,4 ├ 112,9	105,65	04
8	112,9 ├ 127,4	120,15	02

Gráfico 2 – Polígono de frequência dos tempos de percurso ao trabalho



Passo 11:

Uma técnica elaborada recentemente, a apresentação **ramo e folhas**, oferece uma boa visualização global dos dados. Primeiramente decompõem-se cada número em seus algarismos das dezenas e das unidades, marcando juntos os valores que têm o mesmo algarismo das dezenas. Considerando os 100 tempos de percurso registrados, esses números foram dados em décimos de minutos. Ao fazer o gráfico **ramo e folhas**, recomenda-se ignorar os décimos, ao invés de arredondá-los para o próximo minuto.

Aplicando no exemplo:

Gráfico 3 – Ramo e folhas – Tempos de percurso ao trabalho

Ramo	Folhas																					
1	1	5																				
2	6	8	9																			
3	0	5	6	7	8																	
4	0	0	0	0	0	1	1	1	2	2	2	2	3	3	4	9						
5	0	2	2	3	3	3	3	4	5	5	5	5	5	5	6	6	6					
6	0	0	0	0	0	0	1	3	3	4	6	6	7	7	7	7	8	8	8	8	9	9
7	4	5	5	5	6	6	7	7														
8	0	1	2	2	2	2	2	3	3	3	4	5	5	6	6	9						
9	5	6	7	7	7																	
10	0	3	8																			
11	0																					
12	3	5																				

Nota: 7 | 4 significa 74 minutos
12 | 3 significa 123 minutos

Ampliando seus conhecimentos

Uma forma conveniente de indicar relações entre os dados qualitativos (categorias) é a construção de um diagrama de Pareto. Um diagrama de Pareto é um gráfico em barras para dados qualitativos, com as barras ordenadas de acordo com a frequência. As escalas verticais em um diagrama de Pareto podem representar frequências absolutas ou relativas. A barra mais alta fica à esquerda, e as barras menores na extrema direita.

Dispondo as barras por ordem de frequência, o diagrama de Pareto focaliza a atenção sobre as categorias mais importantes.

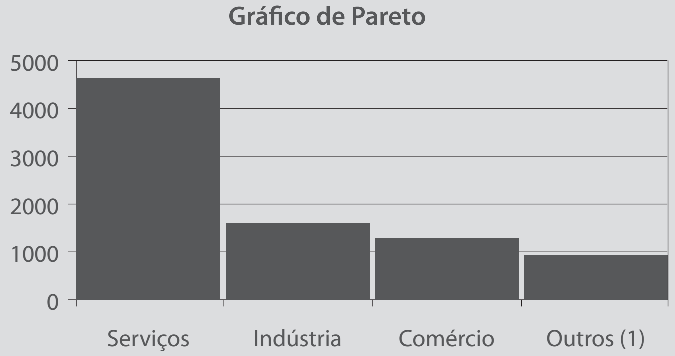
Construção do gráfico de Pareto

Número de ocupados por setor de atividades - Região Metropolitana SP - Jul/2007

setores de atividade	frequência
Serviços	4663
Indústria	1626
Comércio	1384
Outros (1)	978
Total	861

(1) incluem construção civil, serviços domésticos etc.
Fonte: SEP. Convênio Seade – Dieese e TEM/FAT
<www.dieese.org.br>.

Número de ocupados por setor de atividades – Região Metropolitana SP – Jul/2007



(1) incluem construção civil, serviços domésticos etc.

Fonte: SEP. Convênio Seade – Dieese e TEM/FAT <www.dieese.org.br>.

Análise do gráfico de Pareto

Conforme mostra o gráfico de Pareto, evidenciou-se uma concentração de pessoas em atividades no setor de serviços na região metropolitana em São Paulo no mês de julho/2007, comparado com outros setores.

Curiosidades

John W. Tukey começou como químico e tornou-se matemático, especializando-se na estatística, em virtude das "experiências com problemas reais e experiências com dados reais no trabalho da Segunda Guerra Mundial". [...] Tukey dedicou muito do seu tempo ao estudo estatístico de problemas confusos com dados complexos: a segurança dos anestésicos utilizados por muitos médicos e hospitais e em muitos pacientes, o monitoramento da concordância com uma proibição de testes nucleares, e a qualidade do ar e a poluição ambiental. Inventou alguns instrumentos simples como o diagrama em caixa e os diagramas ramo e folhas. Mais importante, modificou a maneira de tratar os dados, enfatizando a necessidade de uma abordagem flexível, exploratória, que procure não só responder a questões específicas, mas também formular questões como "o que os dados nos dizem?".

Atividades de aplicação

1. Escolha a alternativa correta quanto às características/variáveis: cor de olhos, número de filhos, peso líquido e idade podem ser classificadas respectivamente, como:

   a) variáveis qualitativas, quantitativa discreta, quantitativa contínua e quantitativa contínua.

   b) variáveis quantitativa discreta, quantitativa discreta, quantitativa contínua e qualitativa.

   c) variáveis qualitativas, quantitativa contínua, quantitativa contínua e qualitativa.

   d) variáveis qualitativas, quantitativa discreta, quantitativa contínua e quantitativa contínua.

2. Identifique a resposta correta.

   a) Uma pesquisa efetuada com 1 015 pessoas indica que 40 delas são assinantes de um serviço de computador on-line. Trata-se de uma variável quantitativa contínua.

   b) Rendas anuais de enfermeiras tratam-se uma variável qualitativa.

   c) As cores de uma amostra de confeitos M&M são variáveis quantitativas.

   d) O número de inscrições do INSS trata-se de uma variável quantitativa discreta.

3. Escolha a alternativa correta quanto ao conjunto de características: intenção de votos dos eleitores de uma cidade, opiniões dos telespectadores sobre um filme de ficção exibido num canal de televisão e atitudes de clientes de supermercados face a um novo detergente. Trata-se de:

   a) um conjunto de amostras.

   b) um conjunto de amostragem.

c) um conjunto de população

d) um conjunto de amostragem e população.

4. Calcule os pontos médios conforme tabela de frequência abaixo:

peso (kg)	frequência
0 ├ 2,0	20
2,0 ├ 4,0	32
4,0 ├ 6,0	49
6,0 ├ 8,0	31
8,0 ├ 10,0	18

5. Determine o número ideal de classes (**k**) para um conjunto de dados, conforme orientação de *Sturges* (arredondando para cima com número inteiro).

a) 50 dados (elementos) e

b) 150 dados (elementos).

Gabarito

1. A

2. D

3. C

4.

$$\text{Ponto médio} = \frac{(\text{X maior} + \text{X menor})}{2}$$

$$\text{Ponto médio 1.ª classe} = \frac{(2 + 0)}{2} = 1$$

$$\text{Ponto médio 2.ª classe} = \frac{(4 + 2)}{2} = 3$$

$$\text{Ponto médio 3.ª classe} = \frac{(6 + 4)}{2} = 5$$

$$\text{Ponto médio 4.ª classe} = \frac{(8 + 6)}{2} = 7$$

$$\text{Ponto médio 5.ª classe} = \frac{(10 + 8)}{2} = 9$$

5. k (número de classes) = $1 + \dfrac{\log n}{\log 2}$

para 50 dados: k (número de classes) = $1 + \dfrac{\log 50}{\log 2} = 1 + 5{,}6438 = 6{,}6438 \sim$ **7 classes**

para 150 dados: k (número de classes) = $1 + \dfrac{\log 150}{\log 2} = 1 + 7{,}\ 2288 = 8{,}2288 \sim$ **9 classes**

Medidas de tendência central e posição

Definição

As medidas de tendência central têm como finalidade principal a de informar sobre onde se localiza o centro da distribuição.

É um dado importante para o estabelecimento de um esquema de trabalho, para a efetivação de uma compra, para a avaliação de um projeto ou de um produto qualquer, etc. Por exemplo, suponha-se uma variável que seja o número de lâmpadas vendidas por dia em uma casa comercial. Esse número é uma variável X que assume valores possivelmente diferentes ao longo do tempo, mas que se distribuirão em torno de um valor central, o qual fixa e caracteriza as vendas em um determinado número de unidades por dia. Na verdade, esse centro seria o valor que representaria o número de unidades vendidas, caso ele fosse uma constante ao longo do tempo, ou seja: vender seis lâmpadas hoje e dez amanhã seria equivalente a vender oito em cada um dos dois.

Determinar o valor exato do centro de uma distribuição é muitas vezes impraticável, ou mesmo impossível, seja pela evolução natural da população em função do tempo, seja por deficiência dos aparelhos, dos métodos, dos observadores. Por isso, em muitos casos, poder-se-á contar apenas com uma estimativa do total, obtida por meio de uma amostra.

Há diferentes maneiras de definir o centro e, assim, há diferentes definições de medidas de tendência central. As medidas de tendência central frequentemente utilizadas são: média aritmética, mediana, moda e ponto médio. Qual dessas medidas é a melhor? Infelizmente, não há uma resposta única, porque não há critérios objetivos para determinar a medida mais representativa para todos os conjuntos de dados. As diversas medidas de tendência central têm diferentes vantagens e desvantagens. Uma vantagem importante da média aritmética é que se levam em conta todos os valores, mas uma grande desvantagem é que, às vezes, pode ser seriamente afetada por alguns valores extremos.

Tabela1 – Comparação entre média, mediana e moda

Medida	Definição	Existência	Vantagens	Desvantagens
Média	$\overline{X} = \frac{\sum_i^n =_1 x}{n}$	Existe sempre	Reflete cada valor. Possui propriedades matemáticas atraentes.	É influenciada por valores extremos.
Mediana	Valor do meio	Existe sempre	Menos sensível a valores extremos do que a média.	Difícil de determinar para grande quantidade de dados.
Moda	Valor de maior frequência	Pode não existir; Pode haver mais de uma moda	Valor "típico": maior quantidade de valores concentrados neste ponto.	Não se presta a análise matemática. Pode não ser moda para certos conjuntos de dados.
Ponto Médio	$\frac{\text{alto} + \text{baixo}}{2}$	Existe sempre	_________x_________	Muito sensível a valores externos Raramente usada.

Comentários gerais

- Para um conjunto de dados aproximadamente simétrico com uma moda, a média, a mediana, a moda e o ponto médio tendem a coincidir.
- Para um conjunto de dados obviamente assimétricos, convém levar em conta a média e a mediana.
- A média é relativamente confiável: ou seja, quando as amostras são extraídas da mesma população, as médias tendem a ser mais constantes do que outras medidas (constantes no sentido de que as médias amostrais extraídas da mesma população não variam tanto quanto as outras medidas).

Conceitos e regras

Média aritmética da amostra ($\overline{X}$):

A média fornece uma medida central de um conjunto de valores. Se os dados são de uma amostra a média é denotada por $\mathbf{\bar{x}}$ (Lê-se: "x barra").

Talvez a medida de tendência central mais importante seja a média de uma variável. A média fornece uma medida da posição central. Se os dados são de uma amostra, a média é denotada por $\bar{\mathbf{x}}$; se os dados são de uma população, a média é denotada pela letra grega $\boldsymbol{\mu}$. Nas fórmulas estatísticas, é costume denotar o valor da primeira observação por $\mathbf{x}_1$, o valor da segunda observação por $\mathbf{x}_2$ e assim por diante. Em geral, o valor da *i-ésima* observação é denotado por $\mathbf{x}_i$. Para uma amostra com **n** observações.

Fórmula 1 – Média aritmética da amostra ($\bar{x}$)

$$\bar{X} = \frac{\sum_{i=1}^{n} x}{n}$$

Notação:

Σ: denota somatório de um conjunto de valores;

x: é a variável usada para representar valores individuais dos dados;

n: representa o número de valores em uma amostra.

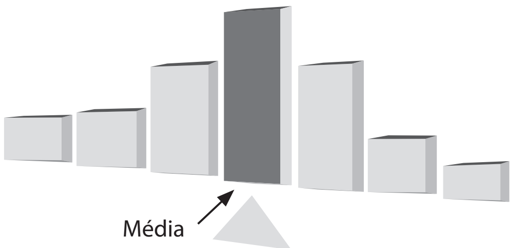


Figura 1 – A média como ponto de equilíbrio.

Média aritmética da população (μ):

Também fornece uma medida central de um conjunto de valores, mas se os dados são de uma população, a média é denotada pela letra grega **μ** (Lê-se: "mi").

Fórmula 2 – Média aritmética da população (μ)

$$\mu = \frac{\sum_{i=1}^{N} x}{N}$$

Notação:

Σ: denota somatório de um conjunto de valores;

x: é a variável usada para representar valores individuais dos dados;

N: representa o número de valores de uma população.

Média ponderada ($\overline{X}$):

Às vezes associam-se os números a certos fatores de ponderação ou pesos, que dependem do significado ou importância atribuída aos números. Nesse caso tem a denominação de média ponderada.

Fórmula 3 – Média ponderada ($\overline{x}$)

$$\overline{X} = \frac{\sum_{i=1}^{n} w_i \cdot x_i}{\sum_{i=1}^{n} w_i}$$

Notação:

Σ: denota somatório de um conjunto de valores;

$\mathbf{X_i}$: é a variável usada para representar valores individuais dos dados;

$\mathbf{W_i}$: é o peso da observação de ordem i.

Mediana ($\tilde{X}$):

Em um conjunto de valores é o valor do meio desse conjunto, quando os valores estão dispostos em ordem crescente ou decrescente (lê-se "x til").

A mediana é outra medida de centralização de uma variável. A mediana é o valor que fica no meio da sequência quando os dados são arranjados na ordem ascendente (classificação do menor para o maior). Com um número ímpar de observações, a mediana é o valor do meio. Um número par de observações não tem um valor único no meio. Neste caso, segue-se a convenção de definir a mediana como sendo a média dos valores das duas observações do meio.

Processo para determinar a mediana:

- ordenar os valores de forma crescente ou decrescente;

- se o número de valores é impar, a mediana é o número localizado exatamente no meio da lista;
- se o número de valores é par, a mediana é a média dos dois valores do meio.

Moda (M)

Em um conjunto de dados é o dos que ocorre com maior frequência.

Podem surgir situações em que a maior frequência ocorra em dois ou mais valores diferentes. Nesses casos existe mais de uma moda. Se os dados têm exatamente duas modas, diz-se que são bimodais. Se os dados têm mais de duas modas, diz-se que são multimodais. Nos casos multimodais, a moda quase nunca é considerada, porque listar três ou mais modas não seria particularmente útil para descrever a posição dos dados. A moda é uma importante medida de posição para os dados qualitativos, em que se observa a característica que apresentou maior frequência na análise de um conjunto de dados.

Quando nenhum valor é repetido, o conjunto não tem moda.

Ponto Médio (P_m)

É o valor que está a meio caminho entre o maior e o menor valor.

Embora o Ponto Médio não seja muito usado, é importante enfatizar que existem maneiras diferentes de definir o centro de um conjunto de dados.

Fórmula 4 – Ponto médio (P_m)

$$P_m = \frac{x\text{ maior} + x\text{ menor}}{2}$$

Processo para determinar o ponto médio:

- identificar o maior valor de todos os valores;
- identificar o menor valor de todos os valores;
- somar os dois valores;
- dividir por dois.

Média geométrica (G)

De um conjunto de **n** números $x_1, x_2, x_3, ..., x_n$ é a raiz de ordem **n** do produto desses números:

Fórmula 5 – Média geométrica

$$G = \sqrt[n]{x_1\ x_2\ x_3 \ldots x_n}$$

A média geométrica é usada em administração e economia para achar taxas médias de variação, de crescimento, ou razões médias. Dados **n** valores (todos positivos), a média geométrica de 2, 4, 10 multiplicando-se os três valores – o que dá 80, e tomando-se a raiz cúbica do resultado, cúbica porque há três valores.

Exemplo:

Admita-se que, nos últimos quatro anos, o produto interno bruto (**PIB**) de um determinado país cresceu 2,5%, 1,7%, 2,2% e 3,5%. Denota-se por $\mathbf{PIB_0}$ o valor do **PIB** no ano relativo a este período, o seu valor do último ano será dado por:

$$PIB_4 = (R_1 . R_2 . R_3 . R_4) . PIB_0$$

Onde as razões de crescimento, R_n, são:

$R_1 = 1 + 0{,}025$

$R_2 = 1 + 0{,}017$

$R_3 = 1 + 0{,}022$

$R_4 = 1 + 0{,}035$

Se o crescimento fosse constante nos quatro anos e, globalmente fosse idêntico ao verificado, a razão **R** entre valores sucessivos do **PIB** deveria satisfazer a seguinte condição:

$$R^4 = R_1 . R_2 . R_3 . R_4$$

Fórmula 6 – Média geométrica

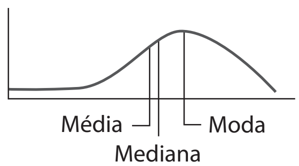
$$G = \sqrt[n]{x_1 \, x_2 \, x_3 \ldots x_n}$$

$$R = \sqrt[4]{R_1 \; R_2 \; R_3 \; R_4} = \sqrt[4]{1{,}025 \;.\; 1{,}017 \;.\; 1{,}022 \;.\; 1{,}035} = \mathbf{1{,}0247}$$

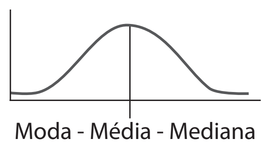
O valor de **R** é a média geométrica das razões de crescimento $R_1 . R_2 . R_3 . R_4$ a partir de **R= 1,0247** obtém-se a taxa média anual de crescimento do PIB, que é de **2,47%**.

Assimetria

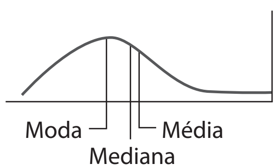
Uma distribuição de dados é assimétrica quando não é simétrica, estendendo-se mais para um lado do que para o outro (uma distribuição de dados é simétrica quando a metade esquerda do seu histograma é aproximadamente a imagem-espelho da metade direita).



Assimétrica para a esquerda (negativamente assimétrica).
A média e a mediana estão à esquerda da moda.



Simétrica (assimetria zero).
A média, a mediana e a moda coincidem.



Assimétrica para a direita (positivamente assimétrica).
A média e a mediana estão à direita da moda.

Figura 2 – Assimetria.

Na prática, muitas distribuições de dados são simétricas. As distribuições assimétricas para a direita são mais comuns do que as assimétricas para a esquerda, porque em geral é mais fácil obter valores excepcionalmente grandes do que valores excepcionalmente pequenos. Com as rendas anuais, por exemplo, é impossível termos valores abaixo do limite inferior zero, mas há algumas pessoas que ganham milhões de reais em um ano. Assim, as rendas anuais tendem a ser assimétricas para a direita.

O problema

Contexto A

Há uma grande variedade de bebidas alcoólicas espalhadas pelo mundo, fazendo do álcool a substância psicoativa mais popular do planeta. O Brasil detém o primeiro lugar do mundo no consumo de destilados de cachaça e é o quinto maior produtor de cerveja. O álcool é a droga preferida dos brasileiros (68,7% do total). Motoristas alcoolizados são responsáveis por 65% dos acidentes fatais em São Paulo. A maioria das fatalidades relacionadas ao consumo de álcool ocorre entre 18 e 25 anos. Com esta preocupação foi coletada a idade de 15 motoristas envolvidos em acidentes fatais.

Tabela 1 – Idades de motoristas envolvidos em acidentes fatais

Idades	16	18	19
	17	18	20
	45	20	22
	22	15	18
	21	19	19

Explorando o problema

De posse de uma grande lista de números, pouco proveito se pode tirar dela, a menos que possamos reduzi-la a uma ou algumas medidas numéricas que resumem todo o conjunto. Tais medidas são de mais fácil manejo e compreensão do que os dados originais. Uma característica importante dos dados é o valor central ou mais típico do conjunto. O objetivo deste problema é apresentar os métodos mais úteis para resumir dados.

Equacionando o problema

As medidas de posição referem-se a valores de uma variável que são típicos ou representativos de um conjunto de dados, isto é, eles são um valor em torno do qual uma grande proporção de outros valores está centralizada. Existem alguns métodos de se obter a medida de posição. Utilizando os valores do problema (Contexto A) serão explicadas as diferentes abordagens.

Calculando a média aritmética da amostra ($\overline{X}$)

Para calcular-se a média aritmética da amostra das idades dos motoristas, conforme o contexto A, tem-se a seguinte condição:

$$\overline{X} = \frac{\sum_{i=1}^{n} x}{n}$$

Passo 1:

Somam-se todas as idades.

16 + 17 + 45 + 22 + 21 + 18 + 18 + 20 + 15 + 19 + 19 + 20 + 22 +18 + 19 = **309**

Passo 2:

O total das idades divide-se pelo total de motoristas (15).

$$\overline{X} = \frac{309}{15}$$

$$\overline{X} = 38{,}625 = \mathbf{39}$$

Passo 3:

Concluir e interpretar o resultado (média).

A média aritmética pode ser calculada para qualquer conjunto de dados e, assim, sempre existe. Leva-se em conta todos os elementos de um conjunto de dados, mas pode ser influenciada por um valor extremo.

Calculando a média ponderada

O cálculo da média ponderada deve levar em conta a quantidade em que cada idade dos motoristas aparece:

$$\overline{X} = \frac{\sum_{i=1}^{n} w_i \,.\, x_i}{\sum_{i=1}^{n} w_i}$$

Passo 1:

Multiplica-se a idade pela sua respectiva frequência.

15	x	1 vez	=	15
16	x	1 vez	=	16
17	x	1 vez	=	17
18	x	3 vezes	=	54
19	x	3 vezes	=	57
20	x	2 vezes	=	40
21	x	1 vez	=	21
22	x	2 vezes	=	44
45	x	1 vez	=	45

Passo 2:

Somam-se os resultados.

$$15 + 16 + 17 + 54 + 57 + 40 + 21 + 44 + 45 = \mathbf{309}$$

Passo 3:

O total divide-se pelo total de idades (**15**).

$$\overline{X} = \frac{309}{15}$$

$$\overline{X} = 38{,}625 = \mathbf{39}$$

Passo 4:

Concluir e interpretar o resultado (média ponderada).

Calculando a mediana ($\tilde{X}$)

A mediana é o valor que divide um conjunto de dados em duas partes iguais. Então, para calcular-se a Mediana das idades dos motoristas, tem-se a seguinte condição:

Passo 1:

Ordenar os valores de forma crescente ou decrescente.

15	16	17	18	18	18	19	19	19	20	20	21	22	22	45

Passo 2:

O total de valores é impar (15 motoristas). Pega-se o elemento que separa o grupo ao meio (oitavo elemento).

15	16	17	18	18	18	19	**19**	19	20	20	21	22	22	45

$$\bar{X} = 19$$

Passo 3:

Concluir e interpretar o resultado (Mediana).

Mediana ($\tilde{X}$) é o valor central, quando os valores encontram-se ordenados. O valor do elemento do meio se **n** é ímpar, ou a média dos dois valores do meio se **n** é par.

Calculando a moda (M)

A moda não é genericamente considerada a medida mais eficiente de tendência central, porque existem muitas situações em que múltiplos valores ou nenhum valor distinto ocorre.

Passo 1:

Ordenar os valores de forma crescente ou decrescente.

15	16	17	18	18	18	19	19	19	20	20	21	22	22	45

Passo 2:

Selecionar o valor que ocorre com maior frequência.

15	16	17	18	18	18	**19**	**19**	**19**	20	20	21	22	22	45

M = 18 e 19

Passo 3:

Concluir e interpretar os resultados (Moda).

O conjunto de valores é bimodal, possui dois valores que ocorrem com a mesma frequência máxima.

Calculando o ponto médio (P_m)

Para calcular-se o Ponto Médio das idades dos motoristas, conforme o contexto A, tem-se a seguinte condição:

$$P_m = \frac{X_{maior} + X_{menor}}{2}$$

Passo 1:

Identificar a maior idade e a menor idade.

Idades	16	18	19
	17	18	20
	45	20	22
	22	**15**	18
	21	19	19

Passo 2:

Somam-se os dois valores.

$$45 + 15 = 60$$

Passo 3:

Dividir a soma por dois.

$$P_m = \frac{60}{2} = 30$$

Passo 4:

Concluir e interpretar o resultado (Ponto Médio).

Extremamente simples, o Ponto Médio (P_m), porém, deverá ser usado com cuidado como medida de tendência central. É uma medida de localização de centro quando as distribuições forem simétricas.

Quartis, Decis e Percentis

Assim como a mediana divide os dados em duas partes iguais, os três *quartis*, denotados por Q_1, Q_2 e Q_3, dividem as observações ordenadas (dispostas em ordem crescente) em quatro partes iguais. Grosso modo, Q_1 separa os 25% inferiores dos 75% superiores dos valores ordenados; Q_2 é a mediana e Q_2 separa os 75% inferiores dos 25% superiores dos dados. Mais precisamente, ao menos 25% dos dados serão no máximo iguais a Q_1 e ao menos 75% dos dados serão no mínimo iguais a Q_1. Ao menos 75% dos dados serão no máximo iguais a Q_3, enquanto ao menos 25% serão, no mínimo, iguais a Q_3. Analogamente, há nove *decis*, denotados por D_1, D_2, D_3,..., D_9, que dividem os dados em dez grupos com cerca de 10% deles em cada grupo. Há, finalmente, 99 *percentis*, que dividem os dados em 100 grupos com cerca de 1% em cada grupo. (Os *quartis, decis e percentis* são exemplos de *fractis*, que dividem os dados em partes aproximadamente iguais). Um estudante que se submeteu ao vestibular para ingresso em uma faculdade é informado de que está no 92.° percentil. Isso não significa, entretanto, que ele tenha obtido 92% no exame; indica, apenas, que qualquer que tenha sido a nota obtida, ela foi superior a 92% (e inferior a 8%) das notas de toda a turma. O 92.° percentil é, pois, uma excelente classificação em relação aos outros que fizeram o exame.

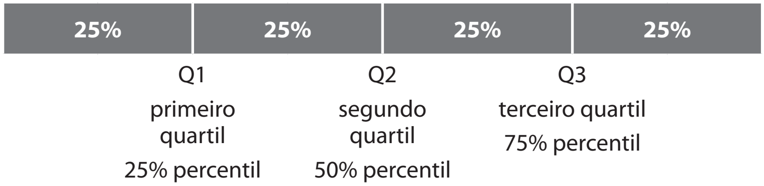


Figura 2 – Posição dos quartis.

Para calcular os quartis:

- dispomos as observações em ordem crescente e localizamos a Mediana ($\bar{X}$) na lista ordenada de observações;
- o primeiro quartil Q_1 é a mediana das observações que estão à esquerda da mediana global na lista ordenada de observações;
- o terceiro quartil Q_3 é a mediana das observações que estão à direita da mediana global na lista ordenada de observações.

Fórmula 6 – Percentil

$$\text{Cálculo do índice (i)} = \left(\frac{p}{100}\right) \cdot n$$

Onde **p** é o percentil de interesse e o **n** é o número de observações.

Se não for um número inteiro, arredonda-se para cima. O próximo inteiro maior que **i** denota a posição do **p**-ésimo percentil.

Se **i** é um inteiro, o p-ésimo percentil é a média dos valores de dados nas posições **i** e **i+1**.

Exemplo de aplicação

Frequentemente é desejável dividir os dados em quatro partes, cada parte contendo aproximadamente um quarto, ou 25% das observações.

Os dados representam o salário inicial e estão arranjados em ordem ascendente. Q_2, o segundo quartil (mediana), já foi identificado como 2405.

2210	2255	2350	2380	2380	2390	2420	2440	2450	2550	2630	2825

Cálculo do quartil

Os cálculos dos quartis Q_1 e Q_3 exigem o uso da regra para encontrar o 25.º e o 75.º percentis. Estes cálculos são:

Para Q_1

Fórmula 7 – Percentil

$$\text{Cálculo do índice (i)} = \left(\frac{p}{100}\right) . \ n$$

$$i = \left(\frac{p}{100}\right) n = \left(\frac{25}{100}\right) 12 = 3$$

Como **i** é um inteiro, indica que o primeiro quartil, ou 25% percentil, é a média do terceiro e do quarto valor dos dados; assim, Q_1= (2350+2380)/2 = 2365.

Para Q_3

Fórmula 8 – Percentil

$$\text{Cálculo do índice (i)} = \left(\frac{p}{100}\right) . \ n$$

$$i = \left(\frac{p}{100}\right) n = \left(\frac{75}{100}\right) 12 = 9$$

Como **i** é um inteiro, indica que o terceiro quartil, ou 75% percentil, é a média do nono e do décimo valores de dados: assim, Q_3= (2450+2550)/2 = 2500. Os quartis dividem os dados dos salários iniciais em quatro partes, com cada parte contendo 25% das observações.

2210	2255	2350	2380	2380	2390	2420	2440	2450	2550	2630	2825

q_1 = 2365 **q_2 = 2405** **q_3 = 2500**

Assim, calculam-se os quartis do mesmo modo que os percentis. No entanto, outras convenções podem ser usadas para calcular os quartis. Os valores reais atribuídos aos quartis podem variar levemente, dependendo da convenção usada. Contudo, o objetivo de todos os procedimentos para o cálculo dos quartis é dividir os dados em quatro partes iguais.

Ampliando seus conhecimentos

Fundada em 1997, a Small Fry Design é uma empresa de brinquedos e de acessórios que projeta e importa produtos para crianças. A linha de produtos da empresa inclui ursinhos, móbiles, brinquedos musicais, chocalhos e cobertores de segurança, caracterizando-se por projetos de brinquedos delicados e de alta qualidade, com ênfase na cor, textura e som. Os produtos são projetados nos Estados Unidos e fabricados na China.

A Small Fry Design utiliza representantes independentes para vender os produtos para as crianças, fornecendo para varejistas, lojas de roupas e acessórios infantis, lojas de presentes, lojas de departamento de grande porte e principais empresas de catálogo. Atualmente, os produtos da Small Fry Design são distribuídos em mais de mil canais de varejo por todo o território dos Estados Unidos.

O gerenciamento do fluxo de caixa é uma das mais críticas atividades na operação do dia a dia dessa jovem empresa. Assegurar a suficiente entrada de caixa para satisfazer tanto as obrigações de débito correntes como as vindouras pode significar a diferença entre o sucesso e o fracasso no negócio. Um fator crítico no gerenciamento do fluxo de caixa é a análise e o controle das contas a receber. Avaliando-se o período médio e o valor em dólares das faturas pendentes, os gerentes podem prever a disponibilidade de caixa e monitorar as mudanças na posição das contas a receber. A empresa estabeleceu os seguintes objetivos: o tempo médio de atraso no pagamento das faturas não deve exceder a 45 dias e o valor das faturas com mais de 60 dias não deve exceder a 5% do de todas as contas a receber.

Em um recente sumário da posição das contas a receber, as seguintes estatísticas descritivas foram fornecidas para o período das faturas pendentes:

Média: 40 dias	Mediana: 35 dias	Moda: 31 dias

A interpretação dessas estatísticas mostra que o período médio de uma fatura é de 40 dias. A mediana mostra que metade das faturas tem ficado pendente 35 dias ou mais. A moda de 31 dias é o período mais frequente de fatura, indicando que a extensão de tempo mais comum que uma fatura tem ficado pendente é 31 dias. O sumário estatístico também mostrou que somente 3% do valor monetário de todas as contas a receber ficaram acima de 60 dias. Baseada na informação estatística, a administração ficou satisfeita de que as contas a receber e a entrada de caixa estejam sob controle.

Atividades de aplicação

1. Valores de vendas diárias de pizzas do tipo calabresa durante um período de nove dias:

15	7	7	11	9	13	14	12	2

Relacione as colunas de acordo com as respostas.

a) Média ( ) 7

b) Mediana ( ) 11

c) Moda ( ) 8,5

d) Ponto Médio ( ) 10

2. Escolha a alternativa correta quanto aos resultados obtidos do conjunto a seguir:

22	19	22	19	18	20	21	22

a) média aritmética é igual a 17,6

mediana é igual a 20

moda é igual a 22

ponto médio é igual a 3

b) média aritmética é igual a 17,6

mediana é igual a 20,5

moda é igual a 19

ponto médio é igual a 3

c) média aritmética é igual a 20,4

mediana é igual a 20,5

moda é igual a 22

ponto médio é igual a 20,0

d) média aritmética é igual a 16,6

mediana é igual a 20

moda é igual a 19 e 22

ponto médio é igual a 3

3. Para os 20 valores de cargas axiais de latas de alumínio, relacionados abaixo, determine (a) a média aritmética, (b) a mediana, (c) a moda e (d) o ponto médio.

225	200	201	223	209	230	209	217	234	209

217	218	220	217	200	219	201	225	200	236

4. Escolha a alternativa correta quanto aos resultados obtidos do conjunto a seguir:

35	25	28	32	31	31	29	30

a) A distribuição dos dados é assimétrica para a esquerda.

b) A distribuição dos dados é simétrica.

c) A distribuição dos dados é assimétrica para a direita.

5. Para as 30 idades de motoristas envolvidos em acidentes fatais, ordenadas da mais nova até a mais velha. Determine o percentil correspondente a 21.

17	17	17	18	18	18	18	19	19	19
19	19	21	22	22	22	23	23	24	25
27	27	28	29	31	31	32	35	39	40

Gabarito

1.

a) Média (C) 7

b) Mediana (B) 11

c) Moda (D) 8,5

d) Ponto Médio (A) 10

2. C

3.

a) média $(\overline{X}) = \frac{\sum x}{n} = \frac{4310}{20} = 215{,}5$

b) mediana (X)

200	200	200	201	201	209	209	209	217	217
217	218	219	220	223	225	225	230	234	236

$$\frac{217 + 217}{2} = 217$$

c) moda (M) = 200, 209 e 217 (multimodal)

d) Ponto médio (P_m)

225	200	201	223	209	230	209	217	234	209
217	218	220	217	200	219	201	225	200	236

$$P_m = \frac{X\ \text{maior} + X\ \text{menor}}{2} = \frac{236 + 200}{2} = 218$$

4. A

5. Percentil de 21 $= \frac{12}{30} \times 100 = 40$

A idade de 21 é o 40.º percentil

Definição

Um conjunto de valores pode ser convenientemente resumido, por meio de procedimentos estatísticos, em poucos valores representativos – média aritmética, mediana e moda. Tais valores podem servir de comparação para dar a posição de qualquer elemento do conjunto. Porém, não é o bastante dar uma das medidas de posição para caracterizar perfeitamente um conjunto de valores, pois, mesmo sabendo, por exemplo, que a temperatura média de duas cidades é a mesma, e igual a 28°C, ainda assim, se é levado a pensar a respeito do clima dessas cidades. Em uma delas a temperatura poderá variar entre limites de muito calor e de muito frio e possuir, ainda, uma temperatura média de 28°C. A outra poderá ter uma variação pequena de temperatura e possuir, portanto, no que se refere à temperatura, um clima mais favorável.

Por essa razão a média – ainda que considerada como um número que tem a faculdade de representar uma série de valores – não pode, por si mesma, destacar o grau de homogeneidade ou heterogeneidade que existe entre os valores que compõem o conjunto.

Considerando os seguintes valores das variáveis x, y e z:

X: 60, 60, 60, 60, 60

Y: 58, 59, 60, 61, 62

Z: 5,5, 30, 110, 150

Calculando a média aritmética de cada um desses conjuntos, obtém-se:

Fórmula 1 – Média aritmética da amostra

$$\text{Média Aritmética da amostra } (\bar{x}) = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\text{variável X } (\bar{x}) = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{300}{5} = 60$$

$$\text{variável Y } (\bar{y}) = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{300}{5} = 60$$

$$\text{variável Z } (\bar{z}) = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{300}{5} = 60$$

Observa-se, então, que os três conjuntos apresentam a mesma média aritmética: 60.

É, porém, fácil de notar que o conjunto de **X** é mais homogêneo que os conjuntos **Y** e **Z**, já que todos os valores são iguais à média.

O conjunto **Y**, por sua vez, é mais homogêneo que o conjunto **Z**, pois há menor diversificação entre cada um de seus valores e a média representativa. Chamando de dispersão ou variação a maior ou menor diversificação dos valores de uma variável, em torno de um valor de tendência central tomado como ponto de comparação, pode-se dizer, então, que o conjunto **X** apresenta dispersão ou variação nula e que o conjunto **Y** apresenta uma dispersão ou variação menor que o conjunto **Z**. Portanto, para qualificar os valores de uma dada variável, ressaltando a maior ou menor dispersão ou variação entre esses valores e a sua medida de posição, a Estatística recorre a medidas de dispersão ou de variação.

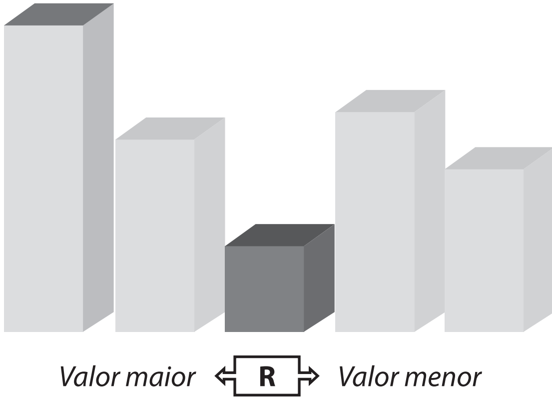
Conceitos e regras

Dispersão ou variação

O grau aos quais os dados numéricos tendem a dispersar-se em torno de um valor médio chama-se variação ou dispersão dos dados.

Amplitude (R)

De um conjunto de dados é a diferença entre o maior e o menor valor. É claro que o valor de R está relacionado com a dispersão dos dados. Quanto maior a Amplitude maior a dispersão dos dados. Entretanto, por depender de apenas dois valores do conjunto de dados, a amplitude contém relativamente pouca informação quanto à dispersão.



Amplitude (R)

$$R = X_{máximo} - X_{mínimo}$$

Figura 1 – Representação da amplitude em um conjunto de dados.

Processo para determinar a amplitude:

- ordenar os valores de forma crescente ou decrescente;
- identificar o maior valor de todos os dados da amostra;
- identificar o menor valor de todos os dados da amostra;
- tomar a diferença entre os dois valores.

Variância amostral (s^2)

De um conjunto de dados é, por definição, a média dos quadrados das diferenças dos valores em relação à sua média.

Quando se tratar de uma amostra, a simbologia é $\mathbf{s^2}$ e quando se tratar de população é $\boldsymbol{\sigma^2}$, que é a letra minúscula sigma. Não confundir o $\boldsymbol{\sigma}$ (sigma) minúsculo com o $\boldsymbol{\Sigma}$ (sigma) maiúsculo, este usado para representar um somatório.

Propriedades da variância

- A Variância de uma constante é zero.
- Multiplicando-se uma variável aleatória por uma constante, sua Variância fica multiplicada pelo quadrado da constante.

- Somando-se ou subtraindo-se uma constante à uma variável aleatória, sua Variância não se altera.
- A Variância da soma ou diferença de duas variáveis aleatórias independentes é a soma das respectivas variâncias.

Fórmula 2 – Variância da amostra (s_x^2)

$$(s_x^2) = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$

Processo para determinar a variância da amostra:

- Calcular a média ($\overline{x}$);
- Subtrair a média a cada valor do conjunto ($x_i - \overline{x}$);
- Elevar ao quadrado cada desvio $(x_i - \overline{x})^2$;
- Somar os quadrados dos desvios $\sum_{i=1}^{n} (x_i - \overline{x})^2$; e
- Dividir a soma por (n – 1) quando forem dados amostrais, ou simplesmente por (N) para somar o conjunto ou todos os valores da população, conhecido como Variância da população e média da população (μ).

Fórmula 3 – Variância da população (σ^2)

$$\text{Variância da população } (\sigma_x^2) = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}, \text{ onde}$$

(σ)lê-se sigma e (μ) lê-se Mi

Desvio-padrão amostral (s):

De um conjunto de valores amostrais é uma medida da variação dos valores em relação à média. Define-se desvio-padrão como a raiz quadrada positiva da variância. Seu cálculo é feito por meio da variância. Ao contrário da amplitude, o desvio-padrão leva em conta todos os valores, mas essa vantagem torna o cálculo mais difícil.

Fórmula 4 – Desvio-padrão da amostra (s_x)

$$(s_x) = \sqrt[2]{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

Fórmula 5 – Desvio-padrão da população (σ_x)

$$(\sigma_x) = \sqrt[2]{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$$

Regras para o desvio-padrão

- o desvio-padrão (s): mede a dispersão em torno da média e só deve ser usado quando a média é tomada como medida de centro.
- o desvio-padrão (s) igual a zero: somente quando não há dispersão. E isto ocorre quando todas as observações têm o mesmo valor. Em caso contrário, **s > 0**. Na medida em que as observações se tornam mais dispersas em torno da média, **s** aumenta.
- Assim como a média ($\overline{x}$), o Desvio-padrão *(s)* é fortemente influenciado por observações extremas. Uns poucos valores discrepantes podem acarretar um grande valor de **s**.

Arredondamento de dados

O resultado do arredondamento de um número como 72,8 para o inteiro mais próximo é 73, posto que 72,8 é mais próximos de 73 do que de 72. Semelhantemente em 72,8146 o arredondamento para o centésimo mais próximo, ou com duas decimais, é 72,81, porque 72,8146 é mais próximo de 72,81 do que 72,82.

Ao arredondar 72,465 para o centésimo mais próximo, entretanto, depara-se com um dilema, pois 72,465 dista igualmente de 72,46 e de 72,47. Usa-se, na prática, em tais casos, aproximar para o número par que precede o cinco. Assim, 72,465 é arredondado para 72,46, onde 183,575 é arredondado para 183,58. Essa prática é especialmente valiosa para reduzir ao mínimo os erros acumulados por arredondamento, quando se tratar de grande número de operações.

O problema

Contexto A

Muitos bancos comerciais costumavam exigir que os clientes formassem filas separadas para os diversos guichês, mas recentemente passaram a adotar fila única. Qual o motivo dessa modificação? O tempo médio de espera não se modifica, porque a fila de espera não afeta a eficiência dos caixas. A adoção de fila única se deveu ao fato de os clientes preferirem tempos de espera mais consistentes com menor variação. Assim é que milhares de bancos efetuaram uma modificação que resultou em uma variação menor (e clientes mais satisfeitos), mesmo que a média não tenha sido afetada. Considera-se agora uma amostra de dados bancários, onde os valores relacionados são tempos de espera (em minutos) de clientes.

Tabela 1 – Tempos de espera em minutos em filas de banco

Banco Jefferson Valley (fila única)	7,1	7,7	6,7	6,8	6,5	7,3	7,4	7,7	7,7	6,6
Banco da Providência (fila múltipla)	7,7	5,4	9,3	6,7	6,2	7,7	4,2	8,5	5,8	10,0

Os clientes do Banco Jefferson Valley entram em uma fila única que é atendida por três caixas. Os clientes do Banco da Providência podem entrar em qualquer uma das três filas que conduzem a três guichês.

Explorando o problema

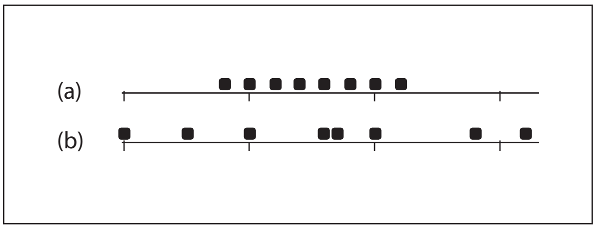
São necessários dois tipos de medidas para descrever adequadamente um conjunto de dados. Além da informação quanto ao "meio" de um conjunto de números, é conveniente dispor também de um método que permita exprimir a dispersão, pois a sumarização de um conjunto de dados, por meio de uma única medida de representação central, esconde toda a informação sobre a variação do conjunto de valores. Então se nota a conveniência de criar-se uma medida que sumariza a variação de uma série de valores e que permite comparar conjuntos diferentes de valores. A Estatística recorre às medidas de dispersão ou variação para atender esse objetivo.

Alguns conceitos-chave são fundamentais para dominar a análise de um conjunto de dados:

- a variação se refere a quanto os valores podem diferir entre si e pode ser medida por números específicos;
- os números relativamente próximos uns dos outros têm baixas medidas de variação, enquanto os valores mais dispersos têm maior medida de variação.

Equacionando o problema

As medidas de dispersão indicam se os valores estão relativamente próximos uns dos outros, ou separados. Esta situação é ilustrada esquematicamente na figura abaixo. As observações (a) apresentam valores relativamente próximos dos outros, em comparação com os da (b). As formas de encontrar essas medidas de dispersão podem ser tratadas através da amplitude, variância amostral e desvio-padrão amostral.



(a) pequena dispersão e (b) grande dispersão

Figura 2 – Representação da dispersão de dados.

Calculando a amplitude (R)

Para medir-se a amplitude dos tempos de espera em filas de bancos, conforme o contexto A, tem-se a seguinte condição:

Fórmula 6 – Amplitude (R)

$$\text{Amplitude (R)} = X_{\text{máximo}} - X_{\text{mínimo}}$$

Passo 1:

Ordenar os valores de forma crescente ou decrescente.

Tabela 2 – Tempos de espera em minutos em filas de banco – ordenados

Banco Jefferson Valley (fila única)	6,5	6,6	6,7	6,8	7,1	7,3	7,4	7,7	7,7	7,7
Banco da Providência (fila múltipla)	4,2	5,4	5,8	6,2	6,7	7,7	7,7	8,5	9,3	10,0

Passo 2:

Identificar o maior valor ($X_{máximo}$) e o menor valor ($X_{mínimo}$) para cada banco.

Tabela 3 – Tempos de espera em minutos em filas de banco – identificação dos valores extremos

Banco Jefferson Valley (fila única)	$X_{máximo} = 7{,}7$	$X_{mínimo} = 6{,}5$
Banco da Providência (fila múltipla)	$X_{máximo} = 10{,}0$	$X_{mínimo} = 4{,}2$

Passo 3:

Aplicar a fórmula para cada banco.

$$\text{Amplitude (R)} = X_{máximo} - X_{mínimo}$$

Tabela 4 – Tempos de espera em minutos em filas de banco – valor da amplitude

Banco Jefferson Valley (fila única)	R = 1,2
Banco da Providência (fila múltipla)	R = 5,8

Passo 4:

Concluir e interpretar os resultados (Amplitude).

A Amplitude (R) fornece uma ideia do afastamento entre o maior valor e o menor valor, mas não é, na realidade, uma boa medida de dispersão de toda a distribuição. A amplitude não fornece qualquer informação de toda a distribuição, ou seja, não dá qualquer informação sobre qualquer elemento na relação, exceto seus valores extremos. No contexto apresentado dos bancos, nota-se uma distância maior dos valores (máximo e mínimo) do Banco da Providência (fila múltipla) necessita-se ainda de uma medida de dispersão que leve em conta todos os números da relação.

Calculando a variância amostral (s_x^2)

A variância da amostra (s_x^2) é uma medida de dispersão extremamente importante na teoria estatística. Do ponto de vista prático, ela tem o inconveniente de se expressar numa unidade quadrática em relação à da variável em questão.

Fórmula 5 – Variância da amostra (s_x^2)

$$\text{Variância da amostra } (s_x^2) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$

Passo 1:

Calcular a média aritmética para cada banco.

Fórmula 6 – Média aritmética da amostra ($\bar{x}$)

$$\text{Média Aritmética da amostra } (\bar{x}) = \frac{\sum_{i=1}^{n} x_i}{n}$$

Tabela 5 – Média dos tempos de espera em minutos em filas de banco

Banco Jefferson Valley (fila única)	$\bar{x}_{BJV} = \frac{6,5 + 6,6 + 6,7 + 6,8 + 7,1 + 7,3 + 7,4 + 7,7 + 7,7 + 7,7}{10} = \mathbf{7,15}$
Banco da Providência (fila múltipla)	$\bar{x}_{BPro} = \frac{4,2 + 5,4 + 5,8 + 6,2 + 6,7 + 7,7 + 7,7 + 8,5 + 9,3 + 10,0}{10} = \mathbf{7,15}$

Passo 2:

Subtrair a média de cada valor do conjunto (amostra).

Tabela 6 – Tempos de espera em minutos no Banco Jefferson Valley

x_i	$\bar{x}$	$(x_i - \bar{x})$
6,5	7,15	-0,65
6,6	7,15	-0,55
6,7	7,15	-0,45
6,8	7,15	-0,35
7,1	7,15	-0,05
7,3	7,15	0,15
7,4	7,15	0,25
7,7	7,15	0,55
7,7	7,15	0,55
7,7	7,15	0,55

Tabela 7 – Tempos de espera em minutos no Banco Providência

x_i	$\bar{x}$	$(x_i - \bar{x})$
4,2	7,15	-2,95
5,4	7,15	-1,75
5,8	7,15	-1,35
6,2	7,15	-0,95
6,7	7,15	-0,45
7,7	7,15	0,55
7,7	7,15	0,55
8,5	7,15	1,35
9,3	7,15	2,15
10,0	7,15	2,85

Passo 3:

Elevar ao quadrado cada desvio.

Tabela 8 – Tempos de espera em minutos no Banco Jefferson Valley para cálculo da Variância da Amostra (s_x^2)

x_i	$\bar{x}$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
6,5	7,15	-0,65	0,4225
6,6	7,15	-0,55	0,3025
6,7	7,15	-0,45	0,2025
6,8	7,15	-0,35	0,1225
7,1	7,15	-0,05	0,0025
7,3	7,15	0,15	0,0225
7,4	7,15	0,25	0,0625
7,7	7,15	0,55	0,3025
7,7	7,15	0,55	0,3025
7,7	7,15	0,55	0,3025

Tabela 9 – Tempos de espera em minutos no Banco Providência para cálculo da Variância da Amostra (s_x^2)

x_i	$\bar{x}$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
4,2	7,15	-2,95	8,7025
5,4	7,15	-1,75	3,0625
5,8	7,15	-1,35	1,8225
6,2	7,15	-0,95	0,9025
6,7	7,15	-0,45	0,2025
7,7	7,15	0,55	0,3025
7,7	7,15	0,55	0,3025
8,5	7,15	1,35	1,8225
9,3	7,15	2,15	4,6225
10,0	7,15	2,85	8,1225

Passo 4:

Somar os quadrados dos desvios.

Tabela 10 – Somatória dos tempos de espera em minutos no Banco Jefferson Valley

x_i	$\bar{x}$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
6,5	7,15	-0,65	0,4225
6,6	7,15	-0,55	0,3025
6,7	7,15	-0,45	0,2025
6,8	7,15	-0,35	0,1225
7,1	7,15	-0,05	0,0025
7,3	7,15	0,15	0,0225
7,4	7,15	0,25	0,0625
7,7	7,15	0,55	0,3025
7,7	7,15	0,55	0,3025
7,7	7,15	0,55	0,3025
		(somatória) **Σ =**	**2,0450**

Tabela 11 – Somatória dos tempos de espera em minutos no Banco Providência

x_i	$\bar{x}$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
4,2	7,15	-2,95	8,7025
5,4	7,15	-1,75	3,0625
5,8	7,15	-1,35	1,8225
6,2	7,15	-0,95	0,9025
6,7	7,15	-0,45	0,2025
7,7	7,15	0,55	0,3025
7,7	7,15	0,55	0,3025
8,5	7,15	1,35	1,8225
9,3	7,15	2,15	4,6225
10,0	7,15	2,85	8,1225
		(somatória) **Σ =**	**29,8650**

Passo 5:

Dividir a soma por (n - 1) se tratar de dados amostrais ou simplesmente por **N** se os dados representam todos os valores de uma população. No caso dos bancos trata-se de dados amostrais dividindo-se por (n - 1), o valor de n=10 elementos/valores. Encontrar a variância da amostra (s^2).

Variância da amostra *(s_x^2)*

$$\text{Variância da amostra } (s_x^2) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$

Tabela 12 – Valor da variância da amostra para cada banco (s_x^2)

Banco Jefferson Valley (fila única)	$s_{BJV}^2 = \dfrac{2{,}0450}{10 - 1} = \mathbf{0{,}227222}$
Banco da Providência (fila múltipla)	$s_{BP_{RO}}^2 = \dfrac{29{,}8650}{10 - 1} = \mathbf{3{,}318333}$

Passo 6:

Concluir e interpretar os resultados (variância amostral).

A variância de uma amostra é a média dos quadrados dos desvios dos valores a contar da média. No contexto dos bancos percebe-se uma maior variância quanto ao tempo de espera na fila do Banco Providência. Porém, a variância representa um valor (resultado) elevado ao quadrado, sendo difícil interpretar o valor numérico. Esse inconveniente é sanado com a definição do desvio-padrão amostral (s).

Calculando o desvio-padrão da amostra (s_x)

O Desvio-Padrão da amostra é simplesmente a raiz quadrada positiva da variância da amostra. Assim, para determinar o desvio-padrão, calcula-se a variância e toma-se a raiz quadrada positiva do resultado. O desvio-padrão cresce quando a dispersão dos dados aumenta.

Fórmula 7 – Desvio-padrão da amostra (s_x)

$$\text{Desvio-padrão da amostra } (s_x) = \sqrt[2]{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

Passo 1:

Calcular a raiz quadrada positiva da variância da amostra, isto é, calcular o desvio-padrão da amostra (s). Tomar uma casa a mais dos dados originais.

Tabela 13 – Valor do desvio-padrão da amostra para cada banco (s_x)

Banco Jefferson Valley (fila única)	$s^2_{BJV} = \frac{2{,}0450}{10 - 1} = \mathbf{0{,}227222}$	$s = \sqrt[2]{0{,}227222} = \mathbf{0{,}48}$
Banco da Providência (fila múltipla)	$s^2_{BPRO} = \frac{29{,}8650}{10 - 1} = \mathbf{3{,}318333}$	$s = \sqrt[2]{3{,}318333} = \mathbf{1{,}82}$

Passo 6:

Concluir e interpretar os resultados (variância amostral).

O desvio-padrão da amostra (**s**) é uma das medidas mais comumente utilizadas para distribuições e desempenha papel relevante em toda a estatística. Cabe observar que a unidade do desvio-padrão é a mesma unidade da média aritmética, permitindo confirmar que entre os bancos o que apresenta uma variação (dispersão) maior quanto ao tempo de espera nas filas é o banco Providência, com um desvio-padrão de 1,821629, reforçando a conclusão de que o sistema de fila única utilizado na amostra do Banco Jefferson Valley tem variação muito menor.

Ampliando seus conhecimentos

Uma desvantagem do desvio-padrão como medida de variação é que ele depende das unidades de medida. Por exemplo, os pesos de determinado objeto podem ter um desvio-padrão de 0,1 miligramas, o valor não informa ou traduz se representa uma grande variação ou pequena variação, somente se existe uma comparação entre duas amostras ou mais. O que interessa nessa situação é uma medida de variação relativa, como o coeficiente de variação.

Coeficiente de variação % (CV%)

É definido como o quociente entre o desvio-padrão e a média. É frequentemente expresso em porcentagem. Sua vantagem é a de caracterizar a dispersão dos dados em termos relativos a seu valor médio. Assim, uma pequena dispersão absoluta pode ser, na verdade, considerável quando comparada com a ordem de grandeza dos valores da variável e vice-versa. Além disso, por ser adimensional, o coeficiente de variação fornece uma maneira de se compararem as dispersões de variáveis cujas unidades são irredutíveis.

Fórmula 9 – Coeficiente de variação % (CV%)

$$CV\% = \frac{s_x}{\overline{x}} \cdot 100$$

No exemplo dos bancos tem-se a análise da variação de cada um confirmada pelo coeficiente de variação.

Tabela 14 – Valor do coeficiente de variação para cada banco

Banco Jefferson Valley (fila única)	$CV\% = \frac{0,48}{7,15} \cdot 100 =$ **6,71%**

Banco da Providência (fila múltipla)	$CV\% = \frac{1,82}{7,15} \cdot 100 =$ **25,87%**

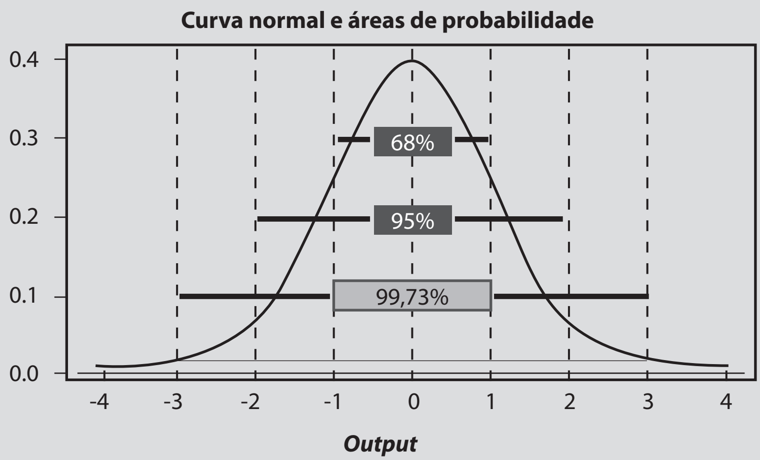
Regra empírica para os dados

Uma regra que auxilia a interpretação do valor de um desvio-padrão é a regra empírica, aplicável somente a conjuntos de dados com distribuição aproximadamente em forma de sino.

A regra 68-95-99 para os dados com distribuição em forma de sino.

- **Cerca de 68%** dos valores estão a menos de um desvio-padrão a contar da média.
- **Cerca de 95%** dos valores estão a menos de dois desvios-padrão a contar da média.
- **Cerca de 99,7%** dos valores estão a menos de três desvios-padrão a contar da média.

Gráfico 1 – A regra empírica para dados com distribuição em forma de sino



O entendimento do gráfico acima evidencia valores presumivelmente em percentagens (*output* = saída – são os registros dos dados de uma determinada variável), sendo razoável entender que são aproximadamente **68%**, aproximadamente **95%**, ou aproximadamente **99,73%** dos valores em torno da média (no gráfico a média é representada pelo ponto zero).

Curiosidade

No Brasil o Instituto Brasileiro de Geografia e Estatística (IBGE) se constitui no principal provedor de dados e informações do país, que atendem às necessidades dos mais diversos segmentos da sociedade civil, bem como dos órgãos das esferas governamentais federal, estadual e municipal.

O IBGE oferece uma visão completa e atual do país, através do desempenho de suas principais funções:

- produção e análise de informações estatísticas;
- coordenação e consolidação das informações estatísticas;
- produção e análise de informações geográficas;
- coordenação e consolidação das informações geográficas;
- estruturação e implantação de um sistema de informação ambiental;
- documentação e disseminação de informações;
- coordenação do sistema estatístico e cartográfico nacionais.

Histórico do IBGE

Durante o período imperial, o único órgão com atividades exclusivamente estatísticas era a Diretoria Geral de Estatística, criada em 1871. Com o advento da República, o governo sentiu necessidade de ampliar essas atividades, principalmente depois da implantação do registro civil de nascimentos, casamentos e óbitos.

Com o passar do tempo, o órgão responsável pelas estatísticas no Brasil mudou de nome e de funções algumas vezes até 1934, quando foi extinto o Departamento Nacional de Estatística, cujas atribuições passaram aos ministérios competentes.

A carência de um órgão capacitado a articular e coordenar as pesquisas estatísticas, unificando a ação dos serviços especializados em funcionamento no País, favoreceu a criação, em 1934, do Instituto Nacional de Estatística (INE), que iniciou suas atividades em 29 de maio de 1936. No ano seguinte, foi instituído o Conselho Brasileiro de Geografia, incorporado ao INE, que passou a se chamar, então, Instituto Brasileiro de Geografia e Estatística.

Há 69 anos, o IBGE cumpre a sua missão: identifica e analisa o território, conta a população, mostra como a economia evolui através do trabalho e da produção das pessoas, revelando ainda como elas vivem.

Atividades de aplicação

1. Observam-se, a seguir, os tempos (em segundos) de reação a um alarme de incêndio, após a liberação de fumaça de uma fonte fixa:

12	9	11	7	9	14	6	10

Escolha a alternativa correta quanto ao valor da amplitude da amostra.

a) Amplitude é igual a 10.

b) Amplitude é igual a 4.

c) Amplitude é igual a 8.

d) Amplitude é igual a 6.

2. Identifique das alternativas abaixo a resposta correta para a situação descrita:

Situação: suponha que três grupos de alunos submetem-se a um teste, obtendo as seguintes notas:

Grupo A: notas = 3 – 4 – 5 – 6 – 9

Grupo B: notas = 1 – 3 – 5 – 7 – 9

Grupo C: notas = 5 – 5 – 5 – 5 – 5

a) O grupo A é o que apresenta maior dispersão/variação, com amplitude igual a 2.

b) O grupo B é o que apresenta menor dispersão/variação, com amplitude igual a 4.

c) O grupo A e B apresentam maior dispersão/variação, do que o grupo C .

d) O grupo C apresenta maior dispersão/variação, com amplitude igual a 0 (zero).

3. Escolha a alternativa correta quanto aos resultados obtidos da amostra abaixo:

Amostra n.º 1:

4	2	9	8	7

a) A amplitude é igual a 3 e a variância é igual a 5,5.

b) A amplitude é igual a 5 e a variância é igual a 8,5.

c) A amplitude é igual a 7 e a variância é igual a 5,5.

d) A amplitude é igual a 7 e a variância é igual a 8,5.

4. Calcule a variância da amostra dos valores abaixo:

Situação: para facilitar um projeto de ampliação da rede de esgotos de uma região em uma pequena cidade, as autoridades tomaram uma amostra de oito quarteirões que compõem a região, e identificaram os seguintes números de casas por quarteirões:

12	6	7	3	15	10	18	5

5. Calcule o desvio-padrão da amostra dos valores abaixo:

Situação: para facilitar um projeto de ampliação da rede de esgotos de uma região em uma pequena cidade, as autoridades tomaram uma amostra de oito quarteirões que compõem a região, e identificaram os seguintes números de casas por quarteirões:

12	6	7	3	15	10	18	5

Gabarito

1. C
2. C
3. D
4. Média Aritmética da amostra $(\bar{x}) = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{76}{8} = 9{,}5$

   Variância da amostra $(s_x^2) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$

x_i	$\bar{x}$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
3	9,5	-6,5	42,25
5	9,5	-4,5	20,25
6	9,5	-3,5	12,25
7	9,5	-2,5	6,25
10	9,5	0,5	0,25
12	9,5	2,5	6,25
15	9,5	5,5	30,25
18	9,5	8,5	72,25
	Σ		**190**

   Variância da amostra $(s_x^2) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{190}{7} = 27{,}14286$

5. Média Aritmética da amostra $(\bar{x}) = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{76}{8} = 9{,}5$

   Variância da amostra $(s_x^2) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$

x_i	$\bar{x}$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
3	9,5	-6,5	42,25
5	9,5	-4,5	20,25
6	9,5	-3,5	12,25
7	9,5	-2,5	6,25
10	9,5	0,5	0,25
12	9,5	2,5	6,25
15	9,5	5,5	30,25
18	9,5	8,5	72,25
	Σ		**190**

   Variância da amostra $(s_x^2) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{190}{7} = 27{,}14286$

   Desvio-padrão da amostra $(s_x) = \sqrt[2]{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = \sqrt[2]{27{,}14286} = 5{,}20988$

Introdução à probabilidade e distribuição de probabilidade discretas

Muitas pessoas agem com base na chance da ocorrência de eventos. Alguns viajam em aviões, reconhecendo que, embora haja a chance de uma colisão com outro aparelho, essa chance é realmente mínima. Alguns aceleram seus automóveis durante uma tempestade, sabendo que podem ser atingidos por um raio, mas, novamente, a chance de tal evento é também mínima. Muitos compram bilhetes de loteria com a esperança de ganhar, no entanto, a chance de ganhar na loteria com uma aposta é menor do que a chance de ser atingido por um raio no período de um ano.

As probabilidades são úteis porque auxiliam a desenvolver estratégias, pois exprimem a chance de ocorrência de determinado evento.

As decisões nos negócios são frequentemente baseadas na análise de incertezas, tais como.

- Quais são as chances de as vendas decrescerem se aumentarmos os preços?
- Qual é a probabilidade do projeto terminar no prazo?
- Quais são as chances de um novo investimento ser lucrativo?

A probabilidade é uma medida numérica da chance de um evento ocorrer. Assim, as probabilidades podem ser usadas como medidas do grau de incerteza associadas aos eventos, por exemplo, aos citados acima. Se as probabilidades estiverem disponíveis, pode-se determinar a chance de cada evento ocorrer.

Na natureza encontram-se dois tipos de fenômenos: determinísticos e aleatórios.

Os fenômenos determinísticos são aqueles em que os resultados são sempre os mesmos, qualquer que seja o número de ocorrências dos mesmos.

Quando se toma um determinado sólido, sabemos que a certa temperatura haverá a passagem para o estado líquido. Esse exemplo caracteriza um fenômeno determinístico.

É importante notar que a definição de um evento determinístico exige que os resultados tenham a mesma chance. Se os resultados não têm a mesma chance, evento aleatório deve-se apelar para a estimativa através da frequência relativa.

Ao calcular probabilidades pelo método da frequência relativa, obtemos uma aproximação em lugar de um valor exato. À medida que o número de observações aumenta, as aproximações tendem a ficar cada vez mais próximas da probabilidade efetiva. Essa propriedade é enunciada como um teorema comumente conhecido como a **lei dos grandes números**.

Mas não é absolutamente essencial realizar um experimento para obter dados amostrais. Em muitos casos dispõem-se de informações históricas, que podem ser utilizadas precisamente da mesma maneira. Essas informações históricas podem apresentar-se sob a forma de dados publicados, ou resultados de testes prévios, ou simplesmente informações acumuladas no arquivo de uma companhia.

Por exemplo, os arquivos de uma companhia imobiliária revelam que, num período de 16 dias, a frequência de casas vendidas por dia foi:

Tabela 1 – Exemplo de frequência relativa

número vendido	número de dias
0	3
1	2
2	5
3	6
Total	**16**

Admitindo-se que o passado é representativo do futuro (o que nem sempre é o caso), pode-se determinar as seguintes probabilidades:

$$P(0) = \frac{3}{16},\ P(1) = \frac{2}{16},\ P(2) = \frac{5}{16},\ P(3) = \frac{6}{16}$$

Assim, de acordo com a conceituação de frequência de probabilidade, imagina-se uma recorrência desse mesmo conjunto de condições, e procura-se responder à pergunta: "Que porcentagem das vezes ocorreu o evento em questão?" Por exemplo, duas casas vendidas em cinco dos 16 dias, de modo que a estimativa da probabilidade de tal ocorrência seria $\frac{5}{16}$. Analogamente, pode-se estimar em $\frac{6}{16}$ a probabilidade de vender três casas. Verifica-se então que, do ponto de vista empírico, a probabilidade pode ser encarada como uma proporção, ou uma frequência relativa, com que ocorre um evento.

Ao adotar-se o método através de frequência relativa, é importante reconhecer os seguintes pontos:

- A probabilidade assim determinada é apenas uma estimativa do verdadeiro valor. O simples fato de obter-se cara quatro vezes em 10 lances de moeda não autoriza a afirmação de que isso ocorrerá sempre. A evidência empírica não nos dá uma probabilidade exata.

- Quanto maior a amostra, melhor a estimativa da probabilidade (lei dos grandes números). O número de observações é importante, de modo geral, quanto maior for esse número (isto é, o tamanho da amostra), melhor será a estimativa da frequência relativa.

- A probabilidade só é válida para um conjunto de condições idênticas àquelas sob as quais se originaram os dados. A validade do método da frequência relativa depende da coincidência dos dois conjuntos de condições. Naturalmente, a não ser no campo das ciências físicas, frequentemente é difícil, ou mesmo impossível, coincidirem as condições exatamente. Infelizmente, na maioria das situações de administração, não se pode controlar todos os fatores relevantes. A implicação é que as proporções resultantes devem ser encaradas como aproximações menos precisas do que as que poderiam obter-se através de experimentos mais controlados. Consequentemente, o grau de confiança nessas probabilidades deve levar em conta o grau de discrepância entre as condições originais e as condições em que tais probabilidades vão ser aplicadas.

Do ponto de vista prático, em geral não é necessário calcular as probabilidades individuais para obter uma distribuição de probabilidades. Existem tabelas e fórmulas para isso. Consequentemente, o problema não é "como se deduzem os valores?", mas sim "como se usam as distribuições para resolver problemas?"

Fora o fato de que as distribuições de probabilidades proporcionam um método simples para a determinação de certas probabilidades, os tipos de distribuições podem ser considerados como modelos para descrever situações que envolvem resultados gerados pela chance.

As distribuições discretas de probabilidade envolvem variáveis aleatórias relativas a dados que podem ser contados, como o número de ocorrências por amostra, ou o número de ocorrências por unidade num intervalo de tempo, de área, ou de distância.

As distribuições discretas podem ser definidas como Binomial e de Poisson, em que o termo Binomial é utilizado para designar situações em que os resultados de uma variável aleatória podem ser agrupados em duas classes ou categorias. Há muitos exemplos de variáveis aleatórias que podem ser classificadas como varáveis Binomiais: respostas a um teste do tipo F ou V, respostas do tipo sim ou não a um questionário, produtos manufaturados classificados como perfeitos ou defeituosos, alunos de uma escola – vacinados ou não vacinados, exames do tipo passa ou não passa. É comum referir-se às duas categorias de uma distribuição Binomial como "sucesso" ou "fracasso/falha", muito embora não importe, para fins de cálculo, qual categoria seja considerada sucesso e qual seja considerada fracasso/falha, pois as duas são complementares. Por exemplo, no caso de um jogo de chance, o sucesso para um parceiro é falha para o outro. Note-se, porque "sucesso" e "fracasso/falha" são mutuamente excludentes e coletivamente exaustivos, P(sucesso) + P(fracasso/falha) = 1,00. A distribuição Binomial é útil para determinar a probabilidade de certo número de sucessos num conjunto de observações.

Por exemplo, suponha-se que se saiba que 80% dos eleitores registrados numa seção eleitoral têm mais de 30 anos. Pode-se querer saber a probabilidade de, numa amostra de 4 eleitores registrados, encontrarem-se 3 ou mais eleitores com mais de 30 anos. Em tal caso, sucesso = eleitor com mais de 30 anos e P(sucesso) = 0,8.

- Há **n** observações ou provas idênticas.

- Cada prova tem dois resultados possíveis, um chamado "sucesso" e o outro "falha/fracasso".
- As probabilidades **p** de sucesso e **1-p** de falha/fracasso permanecem constantes em todas as provas.
- Os resultados das provas são independentes uns dos outros.

Para calcular uma probabilidade Binomial é preciso especificar **n**, o número de provas, **x**, o número de sucessos, e **p**, a probabilidade de sucesso em cada prova. Suponha-se que p = 0,80 (como no exemplo acima, representam eleitores com mais de 30 anos), e consequentemente, **P**(falha/fracasso) = 0,20, e que queira se calcular a probabilidade de três sucessos (isto é, três eleitores com mais de 30 anos) em uma amostra de quatro observações.

Essas observações acham-se tabeladas a seguir, com as respectivas probabilidades.

Tabela 2 – Exemplo binomial

disposição	probabilidade	
SSSF	(0,8).(0,8).(0,8).(0,2) =	0,1024
SSFS	(0,8).(0,8).(0,2).(0,8) =	0,1024
SFSS	(0,8).(0,2).(0,8).(0,8) =	0,1024
FSSS	(0,2).(0,8).(0,8).(0,8) =	0,1024
	Soma	**0,4096**

A probabilidade de três sucessos e uma falha é a soma das probabilidades de todas as maneiras de se obter três sucessos em quatro observações.

No caso acima a soma é 0,4096. Nota-se que cada situação tem a mesma probabilidade de ocorrência, porque os fatores são os mesmos, apenas sua ordem é diferente. Isso é **sempre** verdadeiro. Essa observação leva às seguintes diretrizes: as probabilidades de resultados Binomiais podem ser determinadas levando-se duas coisas em consideração: o número de maneiras como a situação pode ocorrer e a probabilidade de uma dessas maneiras.

No caso da distribuição discreta de probabilidade Poisson, é útil para descrever as probabilidades do número de ocorrências num campo ou intervalo contínuo (em geral tempo ou espaço). Eis alguns exemplos de variáveis que podem ter como modelo a distribuição de Poisson: defeitos por centímetro quadrado, acidentes por dia, clientes por hora, chamada telefônica por minuto etc. Nota-se que a unidade de medida (tempo, área) é contínua, mas a variável aleatória (número de ocorrências) é discreta. Além disso, as falhas não são contáveis. Não é possível contar os acidentes que não ocorreram, nem tampouco o número de chamadas que não foram feitas, nem o número de defeitos por centímetro quadrado que não ocorreram.

A utilização da distribuição de Poisson baseia-se nas seguintes hipóteses:

- A probabilidade de uma ocorrência é a mesma em todo o campo de observação.
- A probabilidade de mais de uma ocorrência num único ponto é aproximadamente zero.
- O número de ocorrências em qualquer intervalo é independente do número de ocorrências em outros intervalos.

O limite inferior do número de ocorrências, em todas essas situações, é zero, enquanto o limite superior é – ao menos teoricamente – infinito, muito embora, na maioria dos exemplos acima, seja difícil imaginar um número ilimitado de ocorrências. A distribuição de Poisson fica completamente caracterizada por um único parâmetro – a média do processo. Assim é que, sabendo que a variável aleatória tem resultados distribuídos segundo Poisson, e conhecendo o número médio de ocorrências por unidade, pode-se determinar a probabilidade de qualquer dos resultados possíveis.

Se uma variável aleatória é descrita por uma distribuição de Poisson, então a probabilidade de realizar (observar) qualquer número dado de ocorrências por unidade de medida (minuto, hora, centímetro, jarda quadrada etc.) é dada pela fórmula:

Distribuição de Poisson

$$P\ (X \cong k) = \frac{e^{-\lambda} . \lambda^k}{k!}$$

Exemplo

Um processo mecânico produz tecido para tapetes com uma média de dois defeitos por metro. Determine a probabilidade de um metro ter exatamente um defeito, admitindo-se que o processo possa ser bem aproximado por uma distribuição de Poisson.

É dado μ= 2, e pela aplicação da fórmula calcula-se:

$$P\ (X \cong 1) = \frac{e^{-2} \,.\, 2^{1}}{1\,!} = \frac{0{,}135\ .\ (2)}{1} = \mathbf{0{,}270}$$

A probabilidade de encontrar-se em um metro um defeito é de 0,27, ou seja, de 27%, tendo como referência a média de ocorrências (μ).

Conceitos e regras

Experimento

É qualquer processo que permite ao pesquisador fazer observações.

Evento

É uma coleção de resultados de um experimento. Exemplo de evento: pode ser chuva, lucro, cara, rendimento de pelo menos 6%, terminar o curso, notas etc.

Evento simples

É um resultado ou evento que não comporta mais qualquer decomposição. Exemplo: o arremesso de um dado é um experimento, e o resultado três é um evento. O resultado três é um evento simples porque não pode ser decomposto: e o espaço amostral consiste nesses eventos simples: 1, 2, 3, 4, 5, 6.

Evento composto

É qualquer evento que combina dois ou mais eventos simples. Exemplo: o arremesso de um par de dados é um experimento, o resultado sete é um evento, mas sete não é um evento simples porque pode ser decomposto em eventos mais simples, como 3-4 e 6-1. Na jogada de um par de dados, o espaço amostral consiste em 36 eventos simples, 1-1, 1-2, ..., 6-6.

Eventos mutuamente excludentes

Os eventos A e B dizem-se mutuamente excludentes quando não podem ocorrer simultaneamente.

Espaço amostral

De um experimento consiste em todos os eventos simples possíveis; ou seja, o espaço amostral consiste em todos os resultados que não comportam mais qualquer decomposição. Representa-se o espaço amostral por Ω (letra grega: ômega). Exemplo:

a) lançamento de uma moeda honesta: $\Omega = \{cara, coroa\}$

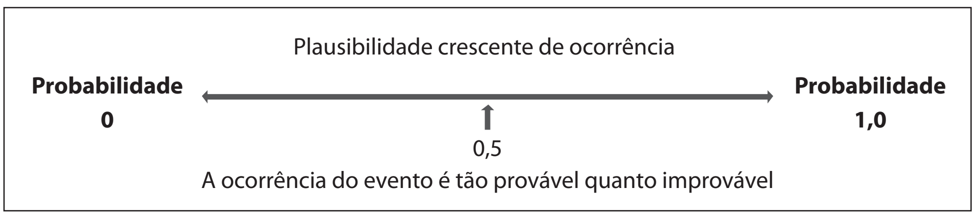
b) lançamento de um dado: $\Omega = \{1,2,3,4,5,6\}$

c) lançamento de duas moedas:

$\Omega = \{(cara, coroa), (cara, coroa), (cara, coroa), (cara, coroa)\}$

Probabilidade

É uma medida numérica da plausibilidade de que um evento ocorrerá. Os valores da probabilidade são sempre atribuídos numa escala de 0 a 1. A probabilidade próxima de zero indica um evento improvável de ocorrer e a probabilidade próxima de 1 indica um evento quase certo. Por exemplo: ao considerarmos o evento "chover amanhã", entendemos que quando a previsão do tempo indica "uma probabilidade próxima de zero de chover" significa "quase sem chance de chover". No entanto, se uma probabilidade 9,90 de chuva é anunciada, sabe-se que é provável que chova.



Probabilidade pela frequência relativa

Realizar um experimento um grande número de vezes e contar quantas vezes o evento A ocorre efetivamente. Então P(A) é estimada como:

Fórmula 1 – Probabilidade relativa

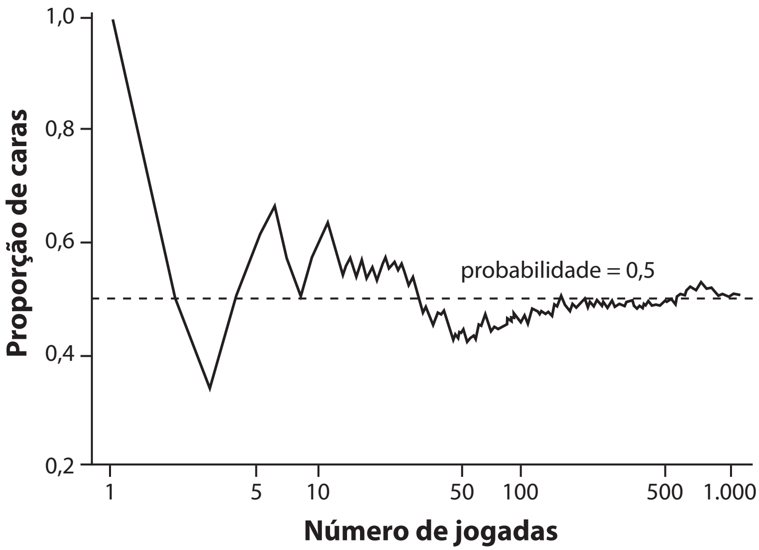
$$P\ (A) = \frac{\text{número de ocorrências de } \mathbf{A}}{\text{número de repetições do experimento}}$$

Lei dos grandes números

Ao se repetir um experimento um grande número de vezes, a probabilidade pela frequência relativa de um evento tende para a probabilidade teórica. A lei dos grandes números afirma que a aproximação pela frequência relativa tende a melhorar quando o número de observações aumenta. Essa lei reflete uma noção bastante simples apoiada pelo senso comum: uma estimativa probabilística baseada apenas em poucas observações pode apresentar grande divergência, mas com um número crescente de provas a estimativa tende a ser mais precisa. Por exemplo, se fizermos uma pesquisa entrevistando apenas algumas pessoas, os resultados podem acusar grande erro; mas se entrevistamos milhares de pessoas *selecionadas aleatoriamente*, os resultados amostrais estarão muito mais próximos dos verdadeiros valores populacionais. Em uma amostra aleatória de um elemento de uma população, todos os elementos da mesma têm igual chance de serem escolhidos; uma amostra de **n** elementos é uma amostra aleatória (ou uma amostra aleatória simples) se é escolhida de tal maneira que toda amostra possível de **n** elementos da população tem a mesma chance de ser escolhida. O conceito geral de aleatoriedade é extremamente importante em estatística. Ao fazerem-se inferências (induções) baseadas em amostras, deve-se ter um processo de amostragem que seja representativo, imparcial e não tendencioso. Se uma amostra não é selecionada cuidadosamente, pode ser totalmente inútil.

Por exemplo: considerando os resultados do lançamento de uma moeda, trata-se de uma variável aleatória, portanto, o comportamento da proporção de lançamentos de uma moeda que dão cara, de 1 a 1000 lançamentos, em longo prazo, a proporção de caras tende para 0,5, sendo a probabilidade de dar cara.

Gráfico 1 – Proporção no lançamento de moedas



Probabilidade clássica

Supondo que um experimento tenha **n** eventos simples diferentes, cada um dos quais com a mesma chance de ocorrer. Se o evento A pode ocorrer em **s** dentre as **n** maneiras, então:

Fórmula 2 – Probabilidade clássica

$$P\,(A) = \frac{\text{número de maneiras como } \mathbf{A} \text{ poder ocorrer}}{\text{número de eventos simples diferentes}} = \frac{s}{n}$$

Comparação entre frequência relativa e abordagem clássica

- **Abordagem pela frequência relativa**: ao procurar determinar **P** (tachinha cair com a ponta para cima) deve-se repetir o experimento (jogar a tachinha) muitas vezes e determinar a razão do número de vezes que a ponta fica pra cima para o número total de jogadas. Essa razão é a estimativa da probabilidade.

- **Abordagem clássica**: ao procurar determinar **P**(2) com um dado equilibrado, cada uma das faces tem a mesma chance de aparecer.

Fórmula 3 – Probabilidade clássica aplicada no lançamento de um dado

$$P\,(2) = \frac{\text{número de possibilidades de ocorrência de 2}}{\text{número total de eventos simples}} = \frac{1}{6}$$

Notação para probabilidade

P denota uma probabilidade.

A, B, C denotam eventos específicos.

P(A) denota a probabilidade de ocorrência do evento **A**.

Arredondamento de probabilidade

O valor de probabilidade deve dar a fração ordinária ou a expressão decimal exata, ou arredondar o resultado final para três algarismos significativos. Sugestão: quando uma probabilidade não é uma fração simples como 2/3 ou 5/9, deve-se expressá-la na forma decimal, como por exemplo: 0,667 ou 0,556 respectivamente.

Notação de espaço amostral:

Representa-se o espaço amostral por Ω.

Classe dos eventos aleatórios

É o conjunto formado de todos os eventos (subconjunto) do espaço amostral.

Exemplo:

Considerando um espaço amostral finito:

$\Omega = \{e_1, e_2, e_3, e_4\}$

A classe dos eventos aleatórios é:

$$F(\Omega) = \left\{ \begin{array}{c} \emptyset \\ \{e_1\} - \{e_2\} - \{e_3\} - \{e_4\} - \{e_1,\ e_2,\ e_3,\ e_4\} \\ \{e_1, e_2\} - \{e_1, e_3\} - \{e_1,\ e_4\} - \{e_2,\ e_3\} - \{e_2,\ e_4\} - \{e_3, e_4\} \\ \{e_1,\ e_2,\ e_3\} - \{e_1,\ e_2,\ e_4\} - \{e_1,\ e_3,\ e_4\} - \{e_2, e_3, e_4\} \\ \{e_1,\ e_{2,},\ e_3,\ e_4\} \end{array} \right\}$$

Para determinar o número de elementos (eventos) de F(Ω) observa-se que:

$\emptyset$ corresponde a $\binom{4}{0}$

$\{e_1,\ e_{2,},\ e_3,\ e_4\}$ corresponde a $\binom{4}{1}$

$\{e_1, e_2\} - \{e_1, e_3\} - \{e_1,\ e_4\} - \{e_2,\ e_3\} - \{e_2,\ e_4\} - \{e_3,\ e_4\}$ corresponde a $\binom{4}{2}$

$\{e_1,\ e_2,\ e_3\} - \{e_1,\ e_2,\ e_4\} - \{e_1,\ e_3,\ e_4\} - \{e_2,\ e_3,\ e_4\}$ $\binom{4}{3}$

$\{e_1,\ e_{2,},\ e_3,\ e_4\}$ corresponde a $\binom{4}{4}$

Portanto, $n(F) = \binom{4}{0} + \binom{4}{1} + \binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 16$

Genericamente, se o número de pontos amostrais de um espaço amostral finito é **n**, então o número de eventos de **F** é **2ⁿ**, pois

$$n(F) = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n$$

Operações com eventos aleatórios

Considerando um espaço amostral finito $\Omega = \{e_1,\ e_{2,},\ e_3,\ \ldots,\ e_n\}$. Sejam A e B dois eventos de F(Ω). As seguintes operações são definidas:

Reunião

$(A \cup B) = \{e_i \in \Omega \mid e_i \in A$ ou $e_i \in B\},\ i\ +\ 1,\ 2,\ 3,\ \ldots,\ n.$

O evento reunião é formado pelos pontos amostrais que pertencem a **pelo menos um** dos eventos.

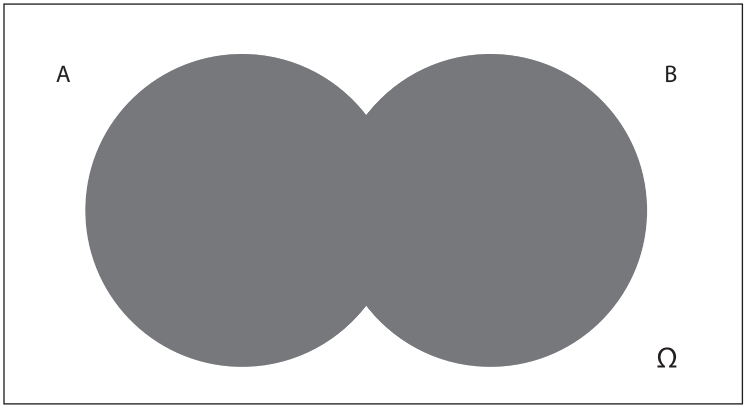


Figura 1 – Reunião.

Intersecção

$(\mathbf{A} \cap \mathbf{B}) = \{e_i \in \Omega \mid e_i \in \mathbf{A} \text{ e } e_i \in \mathbf{B}\},\ \mathbf{i} + \mathbf{1, 2, 3, \ldots, n}.$

O evento intersecção é formado pelos pontos amostrais que pertencem **simultaneamente** aos eventos **A** e **B**.

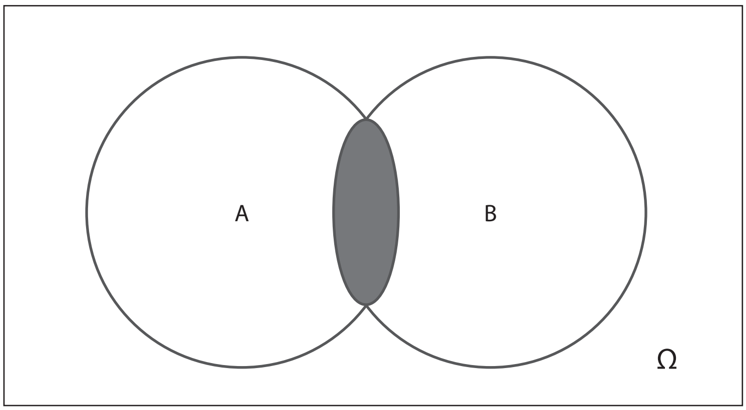


Figura 2– Intersecção.

Complementação

$\Omega - A = \bar{A} = \{e_i \in \Omega \mid e_i \notin A\}.$

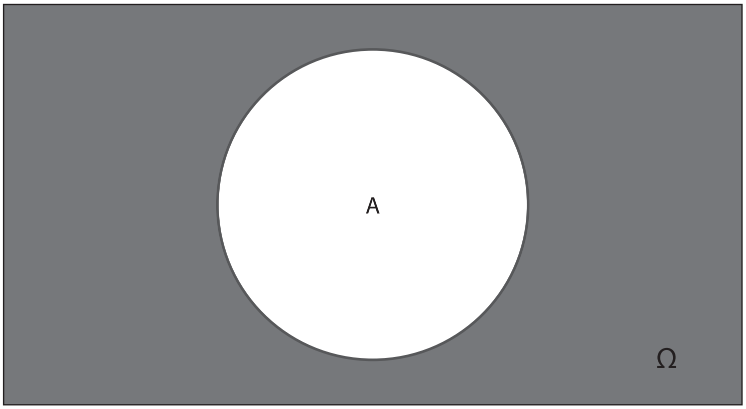


Figura 3 – Complementação.

Regra da adição

Cálculo da probabilidade de ocorrência do evento **A** ou do evento **B** ocorrer, encontrando o total de maneiras como **A** pode ocorrer e o total de maneiras que **B** pode ocorrer, mas de modo que nenhum resultado seja contado mais de uma vez.

Regra formal da adição

Fórmula 4 – Regra da adição

$$P\ (A\ \text{ou}\ B) = P\ (A) + P\ (B) - P\ (A\ \text{e}\ B)$$

Ou

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

Aplicação da regra de adição

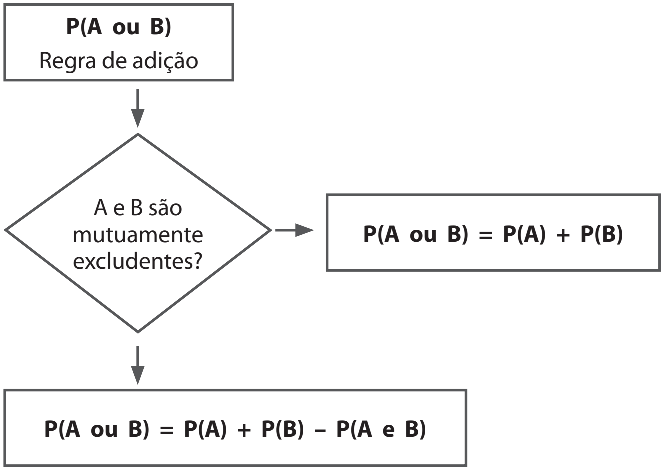


Figura 4 – Fluxograma para aplicação da regra de adição.

Exemplo

Em um teste com o antialérgico Seldane, 49 dos 781 usuários de Seldane experimentaram dores de cabeça, 49 dos 665 que usaram placebo experimentaram dores de cabeça e 24 dos 626 indivíduos do grupo de controle experimentaram dores de cabeça (fonte: *Merrell Doww Pharmaceutical*).

Nessa afirmação, os dados são um tanto difíceis de serem compreendidos, mas se tornam muito mais claros se forem reorganizados em forma tabular.

Tabela 3 – Teste de Seldane

	Seldane	Placebo	Grupo de Controle	Total
Dor de cabeça	49	49	24	122
Não dor de cabeça	732	616	602	1950
Total	781	665	626	2072

Se um dos indivíduos é escolhido aleatoriamente, determine a probabilidade de se obter alguém que fez uso do placebo ou estava no grupo de controle.

Fórmula 5 – Regra da adição (eventos excludentes)

$$P(A \text{ ou } B) = P(A) + P(B)$$

$$P(\text{placebo ou controle}) = \left(\frac{665}{2072}\right) + \left(\frac{626}{2072}\right) = \frac{1291}{2072} = \mathbf{0{,}623}$$

Regra da multiplicação

Para determinar a probabilidade de ocorrência do evento **A** em uma prova, e de ocorrência do evento **B** na próxima prova, deve-se multiplicar a probabilidade de **A** pela probabilidade de **B**, não esquecendo de que a probabilidade do evento **B** deve levar em conta a ocorrência prévia do evento **A**.

Regra formal da multiplicação

Regra da multiplicação para eventos independentes

$P(A \text{ e } B) = P(A) \;.\; P(B)$ se A e B são independentes

Ou

$P(A \cap B) = P(A) \;.\; P(B)$ se A e B são independentes

Regra da multiplicação para eventos dependentes

$P(A \text{ e } B) = P(A) \;.\; P(B \mid A)$ se A e B são dependentes

Ou

$P(A \cap B) = P(A) \;.\; P(B \mid A)$ se A e B são dependentes

Eventos independentes

Se a ocorrência de um deles não afeta a probabilidade de ocorrência do outro.

Eventos dependentes

Se **A** e **B** não são independentes, isto é, a ocorrência de um deles afeta a probabilidade de ocorrência do outro.

Exemplo:

Na extração de duas cartas de um baralho bem misturado, determinar a probabilidade de que a primeira carta seja um ás e a segunda seja um rei (admite-se que a primeira carta extraída não seja reposta antes da extração da segunda carta). Utilizando a regra de multiplicação tem-se a seguinte condição: existem 4 ases nas 52 cartas distintas, tem-se **P**(ás) = $\frac{4}{52}$. Para a segunda extração, supõem-se que tenha obtido um às na primeira extração, de modo que têm-se agora 4 reis entre apenas 51 cartas, em que **P**(rei) = $\frac{4}{51}$. A probabilidade de obter um ás na primeira extração e um rei na segunda é:

Regra da multiplicação para eventos dependentes

$$P\ (\text{ás e rei}) = \left(\frac{4}{52}\right) \cdot \left(\frac{4}{51}\right) = \mathbf{0{,}00603}$$

Regra dos eventos complementares

A definição de eventos complementares implica que eles devem ser mutuamente excludentes, pois é impossível que um evento ocorra e não ocorra simultaneamente.

Regra dos eventos complementares

Fórmula 6 – Regras dos eventos complementares

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

Exemplo:

Se P(chuva) = 0,4 determine P(não chuva). Pela regra dos eventos complementares, tem-se: P(não chuva) = 1 – 0,4 = **0,6**.

Regra fundamental da contagem

Dados dois eventos, o primeiro dos quais pode ocorrer de **m** maneiras distintas e o segundo pode ocorrer de **n** maneiras distintas, então os dois eventos conjuntamente podem ocorrer de **m.n** maneiras distintas.

Notação fatorial

Denota o produto dos inteiros positivos em ordem distinta. Por exemplo, $4! = 4 \,.\, 3 \,.\, 2 \,.\, 1 = 24$. Por definição $0! = 1$.

Regra do fatorial

Uma coleção de **n** objetos diferentes pode ser ordenado de **n!** maneiras distintas. (Esta regra do fatorial traduz o fato de que o primeiro objeto pode ser escolhido de **n** maneiras diferentes, o segundo objeto pode ser escolhido de **n-1** maneiras distintas, e assim por diante).

Exemplo: os problemas de roteamento costumam envolver aplicações da regra do fatorial. Suponha que um vendedor de computadores deva visitar três cidades distintas denotadas por A, B e C, quantos caminhos são possíveis? Pela regra do fatorial, vê-se que as três diferentes cidades (A, B e C) podem ser dispostas de $3! = 6$ maneiras distintas, como mostra a figura abaixo.

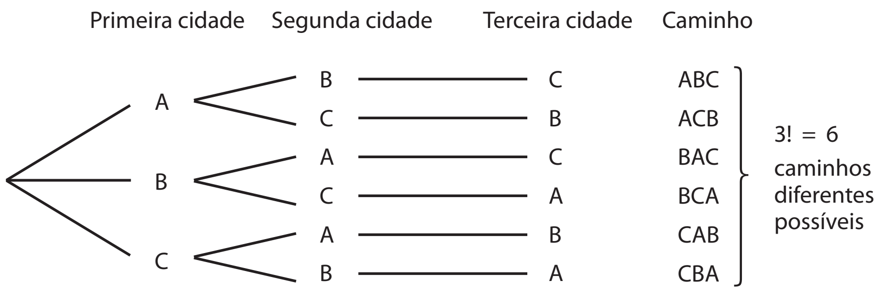


Figura 5 – Diagrama em árvore para roteamentos.

Nota-se que existem três escolhas para a primeira cidade e duas escolhas para a segunda, com isto, resta apenas uma escolha para a terceira cidade. O número de arranjos possíveis para as três cidades é, pois $3 \,.\, 2 \,.\, 1 = 6$.

Regra dos arranjos

O número de arranjos (ou sequência) de **r** elementos, escolhidos **n** elementos (sem repetição).

Regra do arranjo para elementos distintos

$$ {}_{n}P_{r} = \frac{n!}{(n-r)!} $$

Exemplo de aplicação

No planejamento de um programa noturno da rede de televisão NBC, devem ser escolhidos seis shows dentre 30 disponíveis. Quantas programações diferentes são possíveis?

Selecionar **r** = 6 dentre **n** = 30 programas disponíveis. A ordem tem importância, porque os espectadores são outros mais tarde. Como a ordem influi, deve-se calcular o número de arranjos, como segue:

Fórmula 7 – Regra dos arranjos para elementos distintos

$$_{30}P_{6} = \frac{30\,!}{(30-6)\,!} = 427.518.000 \text{ programações}$$

Regra de contagem para combinações

O número de combinações de **N** objetos que são tomados **n** de cada vez é:

Fórmula 8 – Regra de contagem para combinações

$$C_n^N = \binom{N}{n} = \frac{N\,!}{n\,!\,.\,(N-n)\,!}$$

Exemplo

Considerando um procedimento de controle da qualidade em que um inspetor seleciona, aleatoriamente, duas de cinco peças para testar, com relação a defeitos. Em um grupo de cinco peças, quantas combinações de duas peças podem ser selecionadas?

A regra da contagem mostra que com N=5 e n=2, tem-se:

Fórmula 9 – Regra de contagem para combinações

$$C_2^5 = \binom{5}{2} = \frac{5\,!}{2\,!\,.(5-2)\,!} = \frac{(5)\,.\,(4)\,.\,(3)\,.\,(2)\,.\,(1)}{(2)\,.\,(1)\,.\,(3)\,.\,(2)\,.\,(1)} = \frac{120}{12} = \mathbf{10}$$

Variável aleatória discreta

É uma variável (geralmente representada por um **x**) que tem um valor numérico único (determinado aleatoriamente) para cada resultado de um experimento que admite um número finito de valores ou tem quantidade enumerável de valores.

Exemplo: o número de espectadores que assistem a um filme é um número inteiro, sendo, portanto, uma variável aleatória discreta.

Variável aleatória contínua

É uma variável (geralmente representada por um **x**) que tem um valor numérico único (determinado aleatoriamente) para cada resultado de um experimento que admite tomar número infinito de valores, e esses valores podem ser associados a mensurações em uma escala contínua, de tal forma que não haja lacunas ou interrupções.

Exemplo: a voltagem na pilha de um detector de fumaça pode ser qualquer valor entre 0 volts e 9 volts, sendo, por conseguinte, uma variável aleatória contínua.

Distribuição de probabilidade

É a probabilidade de cada valor de uma variável aleatória.

Exemplo: suponha que uma companhia aérea detenha 20% de todas as linhas aéreas domésticas e que todos os voos apresentam a mesma chance de um acidente.

Se a variável aleatória **x** representa o número de acidentes com a companhia aérea dentre sete acidentes, escolhidos aleatoriamente, então a distribuição de probabilidade é dada pela tabela abaixo:

Tabela 4 – Distribuição de probabilidade do número de acidentes, dentre sete acidentes

x	P(x)
0	0,21
1	0,367
2	0,275
3	0,115
4	0,029
5	0,004
6	0
7	0

Obtendo-se as probabilidades da tabela, por exemplo, a probabilidade de 0 acidentes com a companhia aérea é de 0,210, a probabilidade de um acidente é 0,367 etc.

Condições para uma distribuição de probabilidade

$$\sum P(x) = 1$$
$$0 \leqslant P(x) \leqslant 1 \text{ para todo } x$$

Valor esperado

De uma variável aleatória discreta é denotado por E, e representa o valor médio dos resultados. É dado por $\Sigma x \,.\, P(x)$.

Exemplo

A probabilidade de um investidor vender uma propriedade com um lucro de R$2.500,00 de R$1.500,00 de R$500,00 ou com um prejuízo de R$500,00 são 0,22, 0,36, 0,28 e 0,14 respectivamente. Qual é o lucro esperado do investidor?

Considerando:

$a_1 = 2.500{,}00$ $a_2 = 1500{,}00$ $a_3 = 500{,}00$ $a_4 = -500{,}00$.

$P_1 = 0{,}22$, $P_2 = 0{,}36$, $P_3 = 0{,}28$, $P_4 = 0{,}14$

Aplicando a fórmula de E, tem-se

$$E = 2.500{,}00.(0{,}22) + 1.500{,}00.(0{,}36) + 500{,}00.\ (0{,}28) - 500{,}00.(0{,}14) = R\$1.160{,}00$$

Distribuição binomial

A probabilidade de ocorrerem **x** sucessos em **n** provas independentes.

Fórmula 10 – Distribuição binomial

$$P(x = k) = \binom{n}{k} p^k \,.\, (1 - p)^{n-k} \text{ para } k = 0, 1, 2, 3, \ldots, \text{ ou } n.$$

ou seja:

$$P(x = k) = \binom{n}{k} [P\ (\text{sucesso})^k] \,.\, [P\ (\text{fracasso ou falhas})]^{n-k}$$
$$= \text{para } k = 0, 1, 2, 3, \ldots, \text{ ou } n.$$

ou seja:

onde **p** é a probalidade constante de sucesso em cada prova.

Cada tentativa admite apenas dois resultados: fracasso com probabilidade de q e sucesso com probabilidade p, p + q = 1. As probabilidades de sucesso e fracasso são as mesmas para cada tentativa.

Propriedades da distribuição binomial

- O experimento deve comportar um número fixo de provas.
- As provas devem ser independentes (o resultado de qualquer prova não afeta as probabilidades das outras provas).
- Cada prova deve ter todos os resultados classificados em duas categorias.
- As probabilidades devem permanecer constantes para cada prova.

Exemplo:

Uma moeda PE é lançada 20 vezes. Qual é a probabilidade de saírem oito caras?

X: número de sucessos (caras)

$$X = 0, 1, 2, 3, 4, 5, ..., 20 \rightarrow p = P(c) = \frac{1}{2} \rightarrow X: B\left(20; \frac{1}{2}\right)$$

$$P(x = 8) = \binom{20}{8} \cdot \left(\frac{1}{2}\right)^{8} \cdot \left(\frac{1}{2}\right)^{12} = \mathbf{0{,}12013}$$

Esperança (E) e Variância (σ2):

A média de uma variável aleatória discreta é o resultado médio teórico de um número infinito de provas. Diz-se que a média é um valor esperado (esperança) no sentido de que é o valor médio que se espera obter se as provas se prolongassem indefinidamente.

$$\text{Se } X: B(n, p) \rightarrow P(x = k) = \binom{n}{k} \cdot p^{k} \cdot (1 - p)^{n - k}$$

Esperança (E) ou média (μ)e variância $(\sigma)^2$ binomial

$$E(X) = n \cdot p \qquad e \qquad \sigma^2 = n.\ p.\ q$$

Exemplo:

Achar a média e a variância aleatória Y = 3X +2, sendo X: B (20 : 0,3).

Esperança (E) ou média (μ) e variância (σ^2) binomial

$$E(X) = n \cdot p = 20 \cdot 0{,}3 = 6 \qquad e \qquad \sigma^2 = n.\ p.\ q = 20 \cdot 0{,}3 \cdot 0{,}7 = 4{,}2$$

Distribuição de Poisson

Muitas vezes no uso da binomial acontece que **n** é muito grande ($n \to \infty$) e **p** é muito pequeno ($p \to 0$). Nesses casos o cálculo torna-se muito difícil, podendo-se fazer uma aproximação da binomial por meio da distribuição da Poisson. Considera-se a probabilidade de ocorrência de sucessos em um determinado intervalo. A probabilidade da ocorrência de um sucesso no intervalo é proporcional ao intervalo. A probabilidade de mais de um sucesso nesse intervalo é bastante pequena com relação à probabilidade de um sucesso.

Propriedades da distribuição de Poisson

a) $n \to \infty$

b) $p \to 0$ ($p < 0{,}1$)

c) $0 < \mu \leqslant 10$

Quando isso ocorre a média μ = n. p. será tomada como n. p = λ.

Nessas condições obtém-se a distribuição de Poisson:

Fórmula 11 – Distribuição de Poisson

$$P(x \cong k) = \frac{c^{-\lambda} \cdot \lambda^k}{k!}$$

Exemplo:

A probabilidade de uma lâmpada se queimar ao ser ligada é $\frac{1}{100}$. Numa instalação com 100 lâmpadas, qual a probabilidade de 2 lâmpadas se queimarem ao serem ligadas?

X: número de lâmpadas queimadas

$$X: B\left(100, \frac{1}{100}\right)$$

$$P(X = 2) = \binom{100}{2} . (0{,}01)^2 . (0{,}99)^{98} = \mathbf{0{,}183940}$$

Usando a aproximação pelo Poisson

$$\lambda = n.\ p = 100 . 0{,}01 = 1$$

$$P(x = 2) \cong \frac{e^{-1} . 1^2}{2\,!} = \mathbf{0{,}183940}$$

Esperança (E) e Variância (σ^2):

Fórmula 12 – Esperança (E) ou Média (μ) e Variância (σ^2) de Poisson

$$E(X) = \lambda \quad e \quad \sigma^2 = \lambda$$

Contexto A

A companhia aérea “Voar Bem” detém 20% dos voos domésticos e está envolvida em quatro de cada sete acidentes aéreos consecutivos. Supondo que os acidentes aéreos sejam eventos independentes e aleatórios, e admitindo ainda que a empresa seja tão segura quanto as outras companhias de aviação.

a) Qual é a probabilidade de que, em sete acidentes aéreos, quatro ocorram com aviões da “Voar Bem”?

b) Qual é a probabilidade da empresa “Voar Bem” ter ao menos quatro dentre sete acidentes?

Contexto B

Existe uma preocupação com as ocorrências de grandes defeitos em uma estrada um mês depois do recapeamento. Considerando que a probabilidade de um defeito é a mesma para qualquer dos intervalos da estrada de igual comprimento, e a ocorrência ou não ocorrência de um defeito em qualquer outro intervalo. A taxa média de grandes defeitos é de dois para cada 100 km.

a) Qual é a probabilidade de que em 250km ocorram pelo menos três acidentes?

b) Qual é a probabilidade de que em 300km ocorram cinco acidentes?

Explorando o problema

Para encontrar-se uma solução no **contexto A** deve-se, primeiramente, observar se as condições abaixo são atendidas:

- o experimento deve comportar um número fixo de provas;
- as provas devem ser independentes (o resultado de qualquer prova não afeta as probabilidades das outras provas);
- cada prova deve ter todos os resultados classificados em duas categorias;
- as probabilidades devem permanecer constantes para cada prova.

O experimento confirma as condições de um experimento Binomial, conforme critérios indicados acima.

No caso do **contexto B**, deve-se analisar as seguintes condições para aplicação de um modelo de distribuição adequado:

- que a variável aleatória **X** seja o número de ocorrências de um evento em um intervalo;
- que as ocorrências sejam aleatórias;
- que as ocorrências sejam independentes umas das outras;
- que as ocorrências sejam distribuídas uniformemente sobre o intervalo considerado.

O experimento confirma as condições de um experimento como sendo uma Distribuição de Poisson. A distribuição de Poisson difere da Binomial em dois aspectos importantes:

- a distribuição Binomial é afetada pelo tamanho amostral **n** e pela probabilidade **p**, enquanto a distribuição de Poisson é afetada apenas pelo média μ.
- em uma distribuição Binomial, os valores possíveis da variável aleatória de Poisson os valores possíveis de **x** são 0, 1, 2, ..., sem limite superior.

Equacionando o problema

No contexto A trata-se de um experimento Binomial, porque:

1. tem-se um número fixo de provas (7);
2. admite-se que as provas sejam independentes;
3. há duas categorias: cada acidente envolve, ou não envolve, um avião da "Voar Bem". A probabilidade de um acidente envolver um avião da "Voar Bem" (considerada "sucesso" neste experimento) é de 0,20 (porque a "Voar Bem" detém 20% dos voos domésticos), e permanece constante em cada prova (admitindo-se que os acidentes sejam independentes e aleatórios).

Passo 1:

Identificar a variável a ser observada (o sucesso):

X: número de acidentes aéreos com a empresa "Voar Bem "

Passo 2:

Determinar a função sucesso de probabilidade da variável **X**, isto é, P(x): p = P(quando ocorre um acidente com a empresa "Voar Bem") =

$\frac{20}{100} = 0{,}20 \rightarrow$ X: B (7; 0,20), onde n= 7.

Passo 3:

Determinar o fracasso (q): $p+q = 1 \rightarrow q= 1- 0{,}20 = 0{,}80$

Passo 4:

Determinar o experimento observado:

a) Qual é a probabilidade de que, em sete acidentes aéreos, quatro ocorram com aviões da "Voar Bem"?

Fórmula 13 – Distribuição Binomial

$$P(x = k) = \left(\frac{n!}{(n - k)! \,.\, k!} \right) p^k \,.\, (1 - p)^{n - k} \text{ para } k = 0, 1, 2, 3, ..., \text{ ou } n.$$

Para se ter a probabilidade de quatro acidentes com aviões da empresa para sete acidentes:

Tem-se: P(4)

$$X = K = 4 \rightarrow P\ (x = k = 4) = \frac{7!}{(7 - 4)! \,.\, 4!} \,.\, 0{,}20^4 \,.\, (1 - 0{,}20)^{7-4} = \mathbf{0{,}029}.$$

b) Qual é a probabilidade da empresa "Voar Bem" ter ao menos quatro dentre sete acidentes?

Para se ter a probabilidade de ao menos quatro em sete acidentes :

Tem-se: P(4) + P(5) + P(6) + P(7)

$$X = K = 4 \rightarrow P(x = k = 4) = \frac{7!}{(7 - 4)! \,.\, 4!} \,.\, 0{,}20^4 \,.\, (1 - 0{,}20)^{7-4} = 0{,}029$$

$$X = K = 5 \rightarrow P(x = k = 5) = \frac{7!}{(7 - 5)! \,.\, 5!} \,.\, 0{,}20^5 \,.\, (1 - 0{,}20)^{7-5} = 0{,}004$$

$$X = K = 6 \rightarrow P(x = k = 6) = \frac{7!}{(7 - 6)! \,.\, 6!} \,.\, 0{,}20^6 \,.\, (1 - 0{,}20)^{7-6} = 0{,}0$$

$$X = K = 7 \rightarrow P(x = k = 7) = \frac{7!}{(7 - 7)! \,.\, 7!} \,.\, 0{,}20^7 \,.\, (1 - 0{,}20)^{7-7} = 0{,}0$$

$P\ (X \geqslant 4) = \mathbf{0{,}033}$

No contexto B trata-se de um experimento de Poisson visto que:

1. a variável aleatória **X** seja o número de ocorrências de um evento em um intervalo;
2. as ocorrências sejam aleatórias;
3. as ocorrências sejam independentes umas das outras;
4. as ocorrências sejam distribuídas uniformemente sobre o intervalo considerado.

Aplica-se uma distribuição de Poisson, pois se tem a ocorrência de grande defeito em uma estrada em um intervalo (espaço) em quilometragem, onde a chance de ocorrência desse defeito no intervalo está distribuída aleatoriamente.

Passo 1:

Identificar a variável a ser observada (o sucesso).

X: ocorrência de defeitos em uma estrada a cada intervalo em km.

Passo 2:

Definir o experimento observado.

c) Qual é a probabilidade de que em 250km ocorram pelo menos 3 grandes defeitos?

A média de grandes defeitos é de 2 grandes defeitos para cada 100km, portanto para cada 250km têm-se a média de cinco grandes defeitos.

2 grandes defeitos para 100km

λ grandes defeitos para 250km

Portanto: $\lambda = \dfrac{2 \,.\, 250}{100} = 5$

Passo 3:

Calcular o experimento observado.

Distribuição de Poisson

$$P(x \cong k) = \frac{e^{-\lambda} \,.\, \lambda^{k}}{k\,!}$$

$$(X \geqslant 3) = 1 - P(X < 3) = 1 - \{(P(X = 0) + P\ (X = 1) + P(X + 2))\} =$$

$$P(x \cong 0) = \frac{e^{-5} \,.\, 5^{0}}{0\,!} = \frac{0{,}006737947 \,.\, 1}{1} = 0{,}006738$$

$$P(x \cong 1) = \frac{e^{-5} \,.\, 5^{1}}{0\,!} = \frac{0{,}006737947 \,.\, 5}{1} = \frac{0{,}033689735}{1} = 0{,}033690$$

$$P(x \cong 2) = \frac{e^{-5} \cdot 5^{2}}{2!} = \frac{0{,}006737947 \cdot 25}{2} = \frac{0{,}168448675}{2} = 0{,}084224$$

$$= 1 - \left\{ \frac{e^{-5} \cdot 5^{0}}{0!} + \frac{e^{-5} \cdot 5^{1}}{1!} + \frac{e^{-5} \cdot 5^{2}}{2!} \right\} = 1 - \{0{,}006738 + 0{,}033690 + 0{,}084224\} =$$

$= 1 - 0{,}124652 =$ **0,875348**.

No caso da questão abaixo a solução encontra-se em:

Passo 1:

Definir o experimento observado.

Qual é a probabilidade de que em 300km ocorram 5 grandes defeitos?

A média de grandes defeitos é de 2 grandes defeitos para cada 100km, portanto para cada 250km têm-se a média de 5 grandes defeitos.

Dois grandes defeitos para 100km

λ grandes defeitos para 300km

$$\text{Portanto: } \lambda = \frac{2 \cdot 300}{100} = 6$$

Passo 2:

Calcular o experimento observado.

Distribuição de Poisson

$$P(x \cong k) = \frac{e^{-\lambda} \cdot \lambda^{k}}{k!}$$

$$P(x = k = 5) \cong \frac{e^{-6} \cdot 6^{5}}{5!} = \frac{0{,}002478752 \cdot 7776}{120} = \frac{19{,}27477692}{120} = \mathbf{0{,}160623}$$

Notação

μ: média;

σ: desvio-padrão, a raiz quadrada da variância. É a distância, a partir da média, a qualquer dos dois pontos de inflexão; é a variável usada para representar valores individuais dos dados;

e: um número especial, aproximadamente igual a **2,71828**;

π: também um número especial, aproximadamente igual a **3,14159**.

Ampliando seus conhecimentos

Quando há interesse na determinação da probabilidade de um único valor numa distribuição binomial ou distribuição de Poisson, tal como a probabilidade de exatamente quatro sucessos em seis observações, então utiliza-se a tabela de probabilidades binomiais individuais ou as tabelas de probabilidades de Poisson.

Tabelas binomiais individuais

Tal como no caso da fórmula, são necessários três dados: **n**, o número de observações, **p**, a probabilidade de sucesso, e x, o número especificado de sucessos.

Para determinar uma probabilidade de cinco sucessos (x=k=5) em 8 observações (n=8), quando a probabilidade de sucesso é 0,30.

Utiliza-se a tabela como segue.

1. Procurar no topo da tabela o valor de **p** indicado.
2. Localizar o **n** na coluna esquerda da tabela e procurar o número **x** de sucessos observado.
3. A probabilidade de **x** sucessos se encontra na intersecção da linha achada conforme a parte dois com a coluna achada conforme a parte um.
4. Assim, a probabilidade de exatamente dois sucessos em quatro observações, quando a probabilidade de sucesso em cada observação é 0,30, é de 0,2646.

Tabelas binomiais individuais – exemplos

n	Probabilidade de sucesso P	x	P(x)
5	0,20	0	0,1681
6	0,60	3	0,2765
3	0,30	1	0,4410

Probabilidade Binomiais

n	x	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8
2	0	0,902	0,810	0,640	0,490	0,360	0,250	0,160	0,090	0,040
	1	0,095	0,180	0,320	0,420	0,480	0,500	0,480	0,420	0,320
	2	0,002	0,010	0,040	0,090	0,160	0,250	0,360	0,490	0,640
3	0	0,857	0,729	0,512	0,343	0,216	0,125	0,064	0,027	0,008
	1	0,135	0,243	0,384	0,441	0,432	0,375	0,288	0,189	0,096
	2	0,007	0,027	0,096	0,189	0,288	0,375	0,432	0,441	0,384
	3		0,001	0,008	0,027	0,064	0,125	0,216	0,343	0,512
4	0	0,815	0,656	0,410	0,240	0,130	0,062	0,026	0,008	0,002
	1	0,171	0,292	0,410	0,412	0,346	0,250	0,154	0,076	0,026
	2	0,014	0,049	0,154	0,265	0,346	0,375	0,346	0,265	0,154
	3		0,004	0,026	0,076	0,154	0,250	0,346	0,412	0,410
	4			0,002	0,008	0,026	0,062	0,130	0,240	0,410
5	0	0,774	0,590	0,328	0,168	0,078	0,031	0,010	0,002	
	1	0,204	0,328	0,410	0,360	0,259	0,156	0,077	0,028	0,006
	2	0,021	0,073	0,205	0,309	0,346	0,312	0,230	0,132	0,051
	3	0,001	0,008	0,051	0,132	0,230	0,312	0,346	0,309	0,205
	4			0,006	0,028	0,077	0,156	0,259	0,360	0,410
	5				0,002	0,010	0,031	0,078	0,168	0,328
6	0	0,735	0,531	0,262	0,118	0,047	0,016	0,004	0,001	
	1	0,232	0,354	0,393	0,303	0,187	0,094	0,037	0,010	0,002
	2	0,031	0,098	0,246	0,324	0,311	0,234	0,138	0,060	0,015
	3	0,002	0,015	0,082	0,185	0,276	0,312	0,276	0,185	0,082
	4		0,001	0,015	0,060	0,138	0,234	0,311	0,324	0,246
	5			0,002	0,010	0,037	0,094	0,187	0,303	0,393
	6				0,001	0,004	0,016	0,047	0,118	0,262
7	0	0,698	0,478	0,210	0,082	0,028	0,008	0,002		
	1	0,257	0,372	0,367	0,247	0,131	0,055	0,017	0,004	
	2	0,041	0,124	0,275	0,318	0,261	0,164	0,077	0,025	0,004
	3	0,004	0,023	0,115	0,227	0,290	0,273	0,194	0,097	0,029
	4		0,003	0,029	0,097	0,194	0,273	0,290	0,227	0,115
	5			0,004	0,025	0,077	0,164	0,261	0,318	0,275
	6				0,004	0,017	0,055	0,131	0,247	0,367
	7					0,002	0,008	0,028	0,082	0,210

(John E. Freund, 2000)

n	*x*	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8
8	0	0,663	0,430	0,168	0,058	0,017	0,004	0,001		
	1	0,279	0,383	0,336	0,198	0,090	0,031	0,008	0,001	
	2	0,051	0,149	0,294	0,296	0,209	0,109	0,041	0,010	0,001
	3	0,005	0,033	0,147	0,254	0,279	0,219	0,124	0,047	0,009
	4		0,005	0,046	0,136	0,232	0,273	0,232	0,136	0,046
	5			0,009	0,047	0,124	0,219	0,279	0,254	0,147
	6			0,001	0,010	0,041	0,109	0,209	0,296	0,294
	7				0,001	0,008	0,031	0,090	0,198	0,336
	8					0,001	0,004	0,017	0,058	0,168

Tabelas de Poisson Individuais

As tabelas de Poisson são bastante semelhantes às tabelas binomiais, embora à primeira vista possam parecer diferentes. Como a distribuição de Poisson só depende da média do processo, as tabelas são construídas de forma a dar as probabilidades com base nessa média. Os valores escolhidos de **μ**, média do processo (número médio de ocorrências por unidade), constituem a linha do topo da tabela, e os resultados possíveis constam da coluna lateral. O corpo da tabela dá as probabilidades de exatamente **x** ocorrências por unidade. Alguns exemplos de probabilidade de Poisson obtidos da tabela de probabilidades individuais.

Tabela 4.4 – Tabelas de Poisson individuais

μ	x	P(x)
2,1	0	0,1225
2,4	1	0,2177
3,0	2	0,2240
2,2	2	0,2681
3,0	5	0,1008
2,8	10	0,0005

(William J. Stevenson, 1981)

Termos Individuais da Distribuição de Poisson

	μ									
x	**0,1**	**0,2**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,8**	**0,9**	**1,0**
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659	0,3679
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647	0,1839
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494	0,0613
4	0,0000	0,0001	0,0002	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111	0,0153
5	0,0000	0,0000	0,0000	0,0001	0,0002	0,0004	0,0007	0,0012	0,0020	0,0031
6	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0003	0,0005
7	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001

	μ									
x	**1,1**	**1,2**	**1,3**	**1,4**	**1,5**	**1,6**	**1,7**	**1,8**	**1,9**	**2,0**
0	0,3329	0,3012	0,2725	0,2466	0,2231	0,2019	0,1827	0,1653	0,1496	0,1353
1	0,3662	0,3614	0,3543	0,3452	0,3347	0,3230	0,3106	0,2975	0,2842	0,2707
2	0,2014	0,2169	0,2303	0,2417	0,2510	0,2584	0,2640	0,2678	0,2700	0,2707
3	0,0738	0,0867	0,0998	0,1128	0,1255	0,1378	0,1496	0,1607	0,1710	0,1804
4	0,0203	0,0260	0,0324	0,0395	0,0471	0,0551	0,0636	0,0723	0,0812	0,0902
5	0,0045	0,0062	0,0084	0,0111	0,0141	0,0176	0,0216	0,0260	0,0309	0,0361
6	0,0008	0,0012	0,0018	0,0026	0,0035	0,0047	0,0061	0,0078	0,0098	0,0120
7	0,0001	0,0002	0,0003	0,0005	0,0008	0,0011	0,0015	0,0020	0,0027	0,0034
8	0,0000	0,0000	0,0001	0,0001	0,0001	0,0002	0,0003	0,0005	0,0006	0,0009
9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0002

	μ									
x	**2,1**	**2,2**	**2,3**	**2,4**	**2,5**	**2,6**	**2,7**	**2,8**	**2,9**	**3,0**
0	0,1225	0,1108	0,1003	0,0907	0,0821	0,0743	0,0672	0,0608	0,0550	0,0498
1	0,2572	0,2438	0,2306	0,2177	0,2052	0,1931	0,1815	0,1703	0,1596	0,1494
2	0,2700	0,2681	0,2652	0,2613	0,2565	0,2510	0,2450	0,2384	0,2314	0,2240
3	0,1890	0,1966	0,2033	0,2090	0,2138	0,2176	0,2205	0,2225	0,2237	0,2240
4	0,0992	0,1082	0,1169	0,1254	0,1336	0,1414	0,1488	0,1557	0,1622	0,1680
5	0,0417	0,0476	0,0538	0,0602	0,0668	0,0735	0,0804	0,0872	0,0940	0,1008
6	0,0146	0,0174	0,0206	0,0241	0,0278	0,0319	0,0362	0,0407	0,0455	0,0504
7	0,0044	0,0055	0,0068	0,0083	0,0099	0,0118	0,0139	0,0163	0,0188	0,0216
8	0,0011	0,0015	0,0019	0,0025	0,0031	0,0038	0,0047	0,0057	0,0068	0,0081
9	0,0003	0,0004	0,0005	0,0007	0,0009	0,0011	0,0014	0,0018	0,0022	0,0027
10	0,0001	0,0001	0,0001	0,0002	0,0002	0,0003	0,0004	0,0005	0,0006	0,0008
11	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0002	0,0002
12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001

Curiosidade

A estatística tem realmente algum valor?

Com uma venda anual superior a $17 bilhões, a Motorola é uma das maiores produtoras de equipamentos eletrônicos, inclusive telefones celulares, telefones sem fio, *pagers*, rádios transceptores e módulos controladores de transmissão em carros. Nos últimos cinco anos, a Motorola economizou aproximadamente $2,5 bilhões, implementando um plano de melhoria de qualidade que usa extensamente métodos estatísticos. Seus *pagers* e telefones celulares estão sendo fabricados com uma taxa projetada de defeitos de 0,00034%. A Motorola visa a um objetivo popularmente conhecido como o nível de qualidade dos "seis sigmas", que corresponde a menos de 3,4 defeitos por milhão de unidades produzidas. A Motorola constatou que a aplicação de métodos estatísticos é imprescindível para a sobrevivência em um mercado cada vez mais competitivo.

Atividades de aplicação

1. No porto chegam navios à razão de $\lambda = 2$ navios por hora. Observa-se o processo durante um período de meia hora $t = \frac{1}{2}$, sendo a variável aleatória o número de navios que chegam no porto por hora. A distribuição adequada para aplicação nesse processo trata-se de. Escolha a alternativa correta:

a) ( ) Distribuição discreta de probabilidade binomial.

b) ( ) Distribuição discreta de probabilidade Poisson.

c) ( ) Distribuição contínua de probabilidade Normal.

d) ( ) Distribuição contínua de probabilidade Poisson.

2. No porto chegam navios à razão de $\lambda = 2$ navios por hora. Observa-se o processo durante um período de meia hora $t = \frac{1}{2}$, sendo a variável aleatória o número de navios que chegam no porto por hora. Determinar a média de ocorrência no tempo de meia hora $\left(t = \frac{1}{2}\right)$.

3. No porto chegam navios à razão de $\lambda = 2$ navios por hora. Observa-se o processo durante um período de meia hora $t = \frac{1}{2}$, sendo a variável aleatória o número de navios que chegam no porto por hora. Determinar a probabilidade de não chegar nenhum navio no porto.

4. Um engenheiro recolhe uma amostra aleatória de 100 interruptores de um lote de 10 000. Suponha (sem que o engenheiro saiba) que 10% dos interruptores do lote apresentem algum defeito. O engenheiro conta o número de X interruptores defeituosos na amostra. Esta é a distribuição binomial com n = 10 e p = 0,1. Os números possíveis de interruptores defeituosos em uma amostra de dez interruptores são os inteiros de 0 a 10. Calcular a média e o desvio-padrão da distribuição binomial.

5. Um engenheiro recolhe uma amostra aleatória de 100 interruptores de um lote de 10 000. Suponha (sem que o engenheiro saiba) que 10% dos interruptores do lote apresentem algum defeito. O engenheiro conta o número de X interruptores defeituosos na amostra. Determinar a probabilidade de no máximo um interruptor não ser aprovado.

Gabarito

1. B

2. Fórmula – Esperança ou Média de Poisson

   $\mu = \lambda \,.\, t$

   Aplicar a fórmula:

   $\mu = 2 \,.\left(\frac{1}{2}\right) =$ 1 navio por meia hora

3. Sendo: $\mu = 2 \,.\left(\frac{1}{2}\right) =$ 1 navio por meia hora

   A probabilidade do número de ocorrência (chegada de navio) = 0

   Fórmula – Distribuição de Poisson

   $P(X = k) \cong \dfrac{e^{-\lambda} \,.\, \lambda^{k}}{k\,!}$

   Aplicar a fórmula:

   $P(X = k = 0) \cong \dfrac{e^{-1} \,.\, 1^{0}}{0\,!} = 0{,}368$

4. Fórmula – Esperança ou Média Binominal

$E(X)$ ou $\mu = n \,.\, p$ e $\sigma = \sqrt[2]{n \,.\, p \,.\, q}$ ou $\sigma = \sqrt[2]{n \,.\, p \,.\, (p-1)}$

Aplicar as fórmulas:

$\mu = n \,.\, p = (10) \,.\, (0{,}1) = 1$

$\sigma = \sqrt[2]{(10) \,.\, (0{,}1) \,.\, (0{,}9)} = \sqrt[2]{0{,}9} = 0{,}9487$

5. Sendo o número X interruptores que não são aprovados na inspeção tem, aproximadamente, distribuição binomial, com $n = 10$ e $p = 0{,}1$. A probabilidade de, no máximo, um interruptor não ser aprovado é:

$P(X \leq 1) = P(X = 1) + P(X = 0)$

Fórmula – Distribuição binominal

$P(x = k) = \binom{n}{k} p^k \,.\, (1 - p)^{n-k}$

Aplicando-se a fórmula:

$= \binom{10}{1} \,.\, (0{,}1)^1 \,.\, (0{,}9)^9 + \binom{10}{0} \,.\, (0{,}1)^0 \,.\, (0{,}9)^{10} =$

$= \dfrac{10!}{1!\,9!} \,.\, (0{,}1) \,.\, (0{,}3874) + \dfrac{10!}{0!\,10!} \,.\, (1) \,.\, (0{,}38740{,}3487) =$

$= (10) \,.\, (0{,}1) \,.\, (0{,}3874) + (1) \,.\, (1) \,.\, (0{,}\ 3487) =$

$= 0{,}3874 + 0{,}3487 = 0{,}7361$

Obs.: nota-se que $0! = 1$ e $a^0 = 1$ para todo e qualquer número **a** diferente de 0.

Distribuição de probabilidade contínua

Definição

Distribuições contínuas

Nos histogramas as frequências, percentagens, proporções ou probabilidades são representadas pelas alturas dos retângulos, ou por suas áreas. No caso contínuo, representam-se as probabilidades por áreas – não por áreas do retângulo – mas por áreas sob curvas contínuas. Um histograma da distribuição de probabilidade de uma variável discreta que assume apenas os valores 0, 1, 2, ... e 10. A probabilidade de a variável assumir o valor 3, por exemplo, é dada pela área do retângulo branco. No diagrama abaixo a variável contínua pode assumir qualquer valor no intervalo de 0 a 10. A probabilidade de a variável assumir um valor no intervalo de 2,5 a 3,5, por exemplo, é dada pela área da região branca sob a curva.

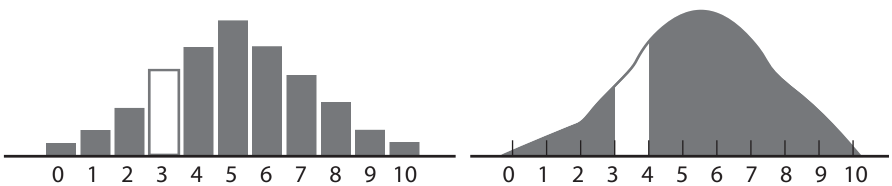


Figura 1 – Histograma de distribuição de probabilidade e gráfico de distribuição contínua.

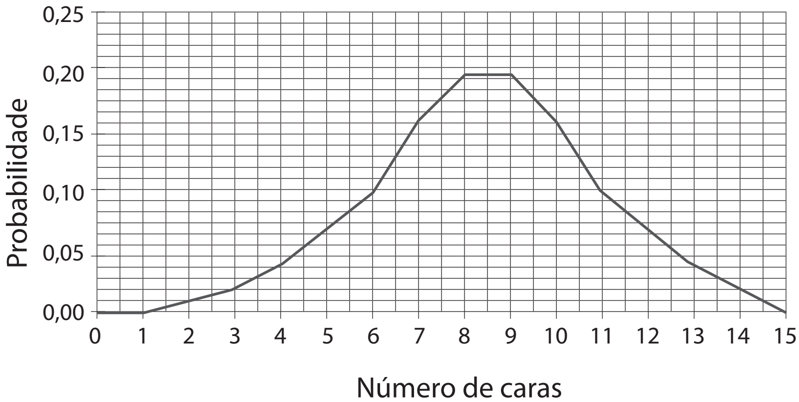
As curvas contínuas são gráficos de funções chamadas de **densidade de probabilidade** ou, informalmente, distribuições contínuas. A expressão densidade de probabilidade tem origem na física, em que os termos **peso e densidade** são empregados precisamente da mesma forma que usamos os termos probabilidade e densidade em estatística. Nas densidades de probabilidade a área sob a curva entre dois valores quaisquer (**a** e **b**) indica a probabilidade de uma variável aleatória com essa distribuição contínua assumir um valor no intervalo de **a** e **b**.

Conceitos e regras

A curva em forma de sino:

Supondo um gráfico de probabilidade dos números de "caras" esperados em 15 jogadas sucessivas de uma moeda. Esta curva em forma de sino, chamada **curva normal,** é a curva mais importante da estatística.

Gráfico 1 – Gráfico de probabilidades dos números de caras – Curva de Sino



Variáveis aleatórias contínuas

Escolha aleatoriamente um nome de catálogo telefônico e meça a altura (em pés) da pessoa assim selecionada. Se **H** é a altura da pessoa, pode-se considerar **H** como uma variável aleatória. Se relacionar todos os valores possíveis de **H** existem valores que, obviamente, não são possíveis. Por exemplo, **H** nunca poderá ser menos de 0,10m nem mais de 3m. Todavia, não poderá listar todos os valores possíveis. A altura pode ser de 1,6m, ou 1,61m, ou 1,600001m. De fato, admitindo-se a possibilidade de se medir a altura com perfeita precisão (isto é, apenas na teoria), há um número infinito de valores possíveis para a altura. Não se pode utilizar uma variável aleatória discreta em um caso como esse, em que o resultado pode ser qualquer número de um determinado intervalo. Deve-se, ao contrário, utilizar uma **variável aleatória contínua**.

Funções de distribuição acumulada contínua

Veja agora como descrever o comportamento de uma variável aleatória contínua. Há muitas semelhanças entre variáveis aleatórias discretas e variáveis aleatórias contínuas; mas há também algumas diferenças importantes. Pode-se estimar a probabilidade de a pessoa escolhida da lista telefônica ter altura inferior a 2m, ou calcular a probabilidade de a pessoa ter altura superior a 10m (que é, logicamente, zero). Portanto, para definir uma função de distribuição acumulada para uma variável aleatória contínua, a letra maiúscula **F** representa uma função de distribuição acumulada.

Fórmula 1 – Função de distribuição acumulada

$$F\ (a) = Pr\ (X \leqslant a)$$

Uma função de distribuição acumulada contínua satisfaz os requisitos:

- **F**(**a**) está entre 0 e 1;
- quando **a** se torna muito grande, **F**(**a**) tende para 1;
- quando **a** se torna muito pequeno (tendendo para -∞), **F**(**a**) tende para 0;
- **F**(**a**) nunca é decrescente.

Seguem duas propriedades práticas importantes:

Para determinar a probabilidade de **X** ser maior do que determinado valor **a**.

Fórmula 2 – Função da probabilidade de X ser maior do que valor

$$F\ (x) > a) = 1 - Pr\ (X < a) = 1 - F\ (a)$$

Para a probabilidade de **X** estar entre dois valores dados **b** e **c**.

Fórmula 3 – Função da probabilidade de X estar entre dois valores

$$F\ (b < x < c) = Pr\ (X < c) - Pr\ (X < b) = F\ (c) - F\ (b)$$

Exemplo de variável aleatória contínua é a variável uniforme, isto é, uma variável que tem a mesma probabilidade de tomar qualquer um dos valores de determinado intervalo. Como exemplo, considerar uma variável aleatória **Y** que tem a mesma chance de tomar qualquer valor entre 0 e 3. Então, a probabilidade de **Y** ser inferior a 1 é 1/3, a probabilidade de **Y** estar entre 1 e 1,5 é 1/6 e assim por diante.

Funções contínuas de densidade de probabilidade

Para determinar a probabilidade de **Y** ser exatamente igual a 2.

Qualquer número de 0 a 3 tem a mesma chance de ser escolhido. Seja, pois, **N** o número de número entre 0 e 3; então, **Pr (Y=2) = 1/N**. Mas existem infinitos números entre 0 e 3 (por exemplo: 0,01; 0,011; 0,01111; etc.).

Esta propriedade vale de modo geral para variáveis aleatórias contínuas: a probabilidade de qualquer variável aleatória contínua tomar valor específico é zero!

Para definir uma função de probabilidade de uma variável aleatória contínua deve-se lembrar que há uma relação entre a função de probabilidade e o diagrama de frequência. Inicialmente, traçar um diagrama de frequência para os pesos das pessoas de determinada amostra.

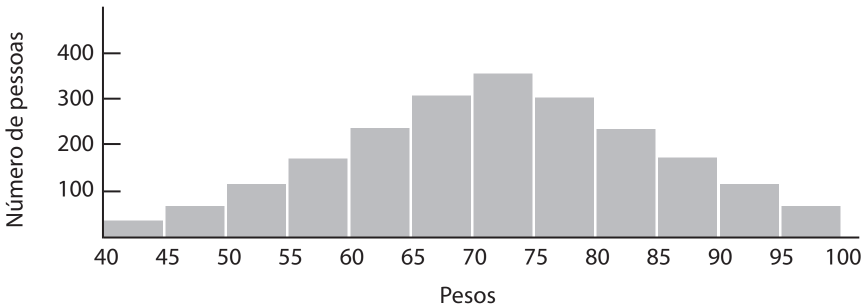


Figura 2 – Diagrama de frequência de peso das pessoas.

Note que a altura de cada barra não representa o número de pessoas cujos pesos estão entre dois valores especificados. Por exemplo, a altura da barra entre 70 e 75 é o número de pessoas na amostra cujos pesos estão entre 70kg e 75kg. A largura de cada barra é designada por Δx (Δ é a letra maiúscula grega **delta**). (Neste caso, Δx = 5).

Por analogia, traçar o gráfico de uma função tal que sua altura em qualquer intervalo seja igual à probabilidade de a variável aleatória tomar um valor naquele intervalo. Entretanto, é mais conveniente fazer a altura de cada barra igual à probabilidade de estar naquele intervalo, dividida por Δx (a largura da barra).

Fórmula 4 – Altura da Barra entre a e a + Δx

$$f\ (a) = \frac{Pr\ (a < X < a = \Delta x)}{\Delta x}$$

Essa função é uma **função de densidade**.

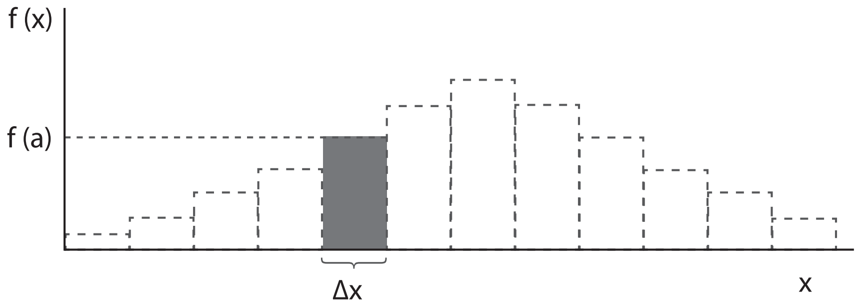


Figura 3 – Representação da função de densidade.

Considerar agora a probabilidade de **X** estar entre dois valores **a** e **b**. Somar as alturas de todas as barras de **a** e **b** e multiplicar o resultado por Δx. Mas como f(x) é a altura de cada barra e Δx é a largura, f(x) Δx é a área da barra. Por conseguinte, a probabilidade de **X** estar entre **a** e **b** é precisamente igual à área de todas as barras entre **a** e **b**. Pr (a < X < b) = área de todos os retângulos entre **a** e **b**.

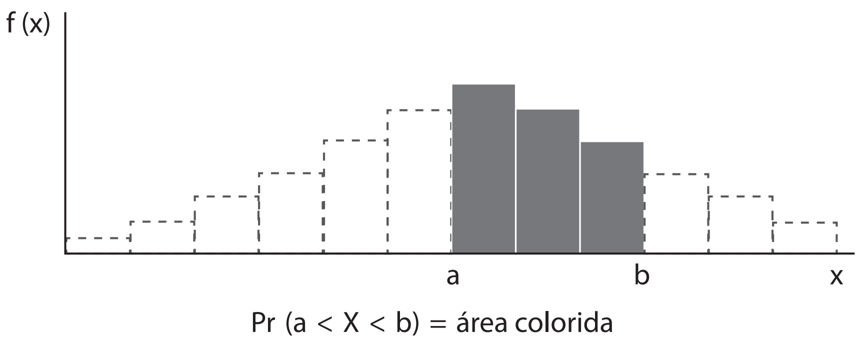


Figura 4 – Função de densidade entre dois valores.

Esta é a característica definidora básica da função de densidade para uma variável aleatória contínua. A área sob a função entre dois valores é a probabilidade de a variável aleatória estar entre esses valores. Todavia, para uma variável aleatória contínua, o diagrama em barras é apenas uma representação aproximada da função de densidade. Para obter melhor aproximação da verdadeira natureza da variável aleatória contínua tornando as barras cada vez mais estreitas. Quando as barras se apresentam extremamente estreitas, a função de densidade se assemelha a uma curva suave.

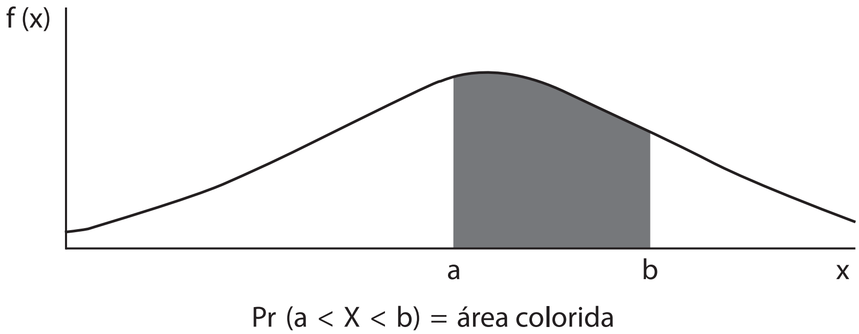


Figura 5 – Curva da função de densidade.

Definição: a função f(x) é uma função de densidade para a variável aleatória **X**, se satisfaz a propriedade de que a área sob a curva **y** = f(x), à esquerda da reta **x = b**, à direita da reta **x = a**, e acima do eixo **x**, é igual a Pr (a < X < b) (recorde que as letras maiúsculas representam variáveis aleatórias, e as letras minúsculas representam variáveis comuns).

Sabe-se que, se F(x) é a função de distribuição acumulada, então

Área sob f(x) de *a* e *b* = Pr (*a* < X < *b*) = F(*b*) – F(*a*)

Considerar o intervalo de – ∞ a + ∞. Sabe-se que F(+∞) – F(–∞ < X +∞), pois o valor de X deve sempre estar em algum ponto entre – ∞ e +∞; não há escolha. Isso significa que a área sob f(x) de – ∞ a +∞ = 1.

Em outras palavras, a área total sob a função f(x) deve ser igual a 1. Se f(x) não tem esta propriedade, então não pode ser uma função legítima de densidade de probabilidade. Pode-se mostrar que esta condição é satisfeita para a função de densidade da variável uniforme Y.

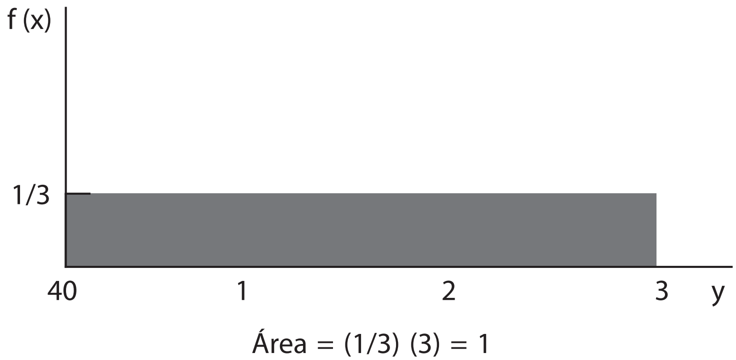


Figura 6 – Função de densidade para variável aleatória contínua uniforme.

Para calcular a probabilidade de uma variável aleatória contínua estar entre dois números, é necessário calcular a área sob a função de densidade entre esses números. O cálculo da área sob uma curva exige uma técnica chamada **integração**. Entretanto, mesmo conhecendo cálculo, não se podem obter fórmulas simples para áreas sob muitas curvas usadas em estatística, tornando-se necessário recorrer a uma tabela para achá-las.

Distribuição normal

É a mais importante distribuição de variável aleatória contínua; sua função de probabilidade tem a forma de um sino que se prolonga indefinidamente em ambas as direções. Muitas populações reais seguem a distribuição normal.

Fórmula 4 – Função da densidade da curva normal

$$f\ (x) = \frac{1}{\sigma\ \sqrt{2\pi}} \times e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Notação

μ: média;

σ: desvio-padrão, a raiz quadrada da variância. É a distância, a partir da média, a qualquer dos dois pontos de inflexão; é a variável usada para representar valores individuais dos dados;

e: um número especial, aproximadamente igual a **2,71828**;

π: também um número especial, aproximadamente igual a **3,14159**.

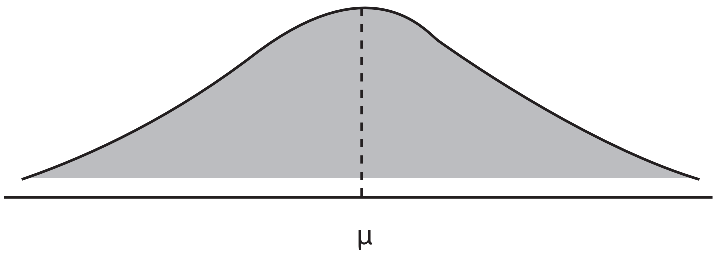
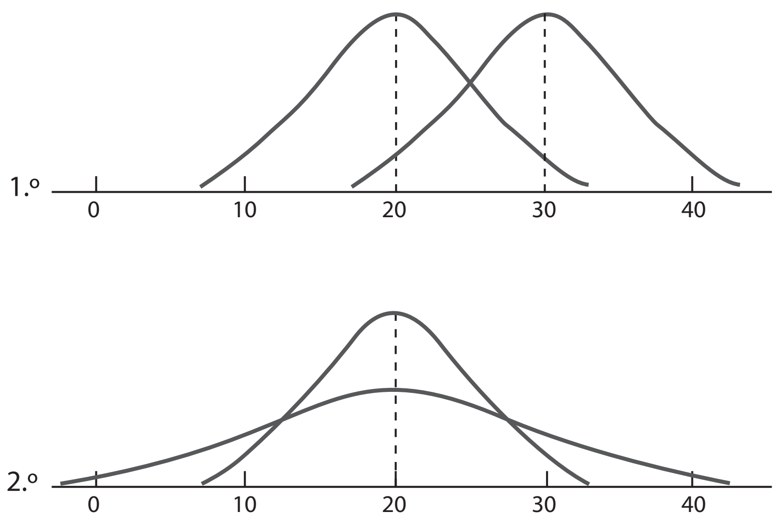


Figura 7 – Curva da distribuição normal.

Característica importante da distribuição normal, visível na equação acima, é que ela depende apenas dos dois valores (ou parâmetros) μ e σ, que são, na realidade, a média e o desvio-padrão. Em outras palavras, há uma e só uma distribuição normal com uma média μ e um dado desvio-padrão σ.

A figura abaixo mostra diferentes curvas, dependendo daqueles dois valores. No primeiro caso, veem-se duas curvas normais com médias diferentes e desvios-padrão iguais; a curva à direita tem maior média. No segundo caso, duas curvas normais com mesma média, mas desvios-padrão diferentes; a mais achatada e mais dispersa tem maior desvio-padrão. Finalmente, no terceiro, duas curvas normais com médias e desvios-padrão diferentes.



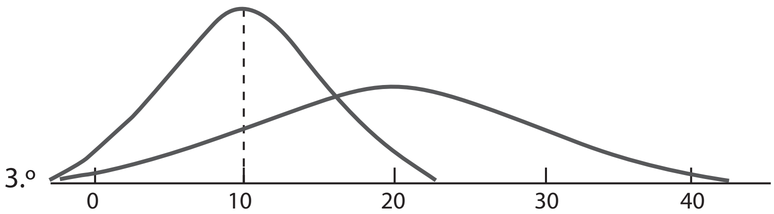


Figura 8 – Três pares de distribuições normais.

Unidades padronizadas

No trabalho com distribuições normais, interessam apenas as áreas sob suas curvas, as chamadas **áreas sob a curva normal**. Como é fisicamente impossível e também desnecessário construir tabelas separadas de áreas sob as curvas normais para todos os pares imagináveis de valores de μ e σ, foram tabeladas (apêndice n.° 2 – tabela 1) apenas as áreas para a distribuição normal com $\boldsymbol{\mu = 0}$ **e** $\boldsymbol{\sigma = 1}$, chamada distribuição normal padronizada. Pode-se obter áreas sob qualquer curva normal fazendo a mudança de escala que transforma as unidades de medida da escala original, ou escala **x**, em unidades padronizadas (escala **z**).

Fórmula 5 – Transformação para unidades padronizadas

$$z = \frac{x - \mu}{\sigma}$$

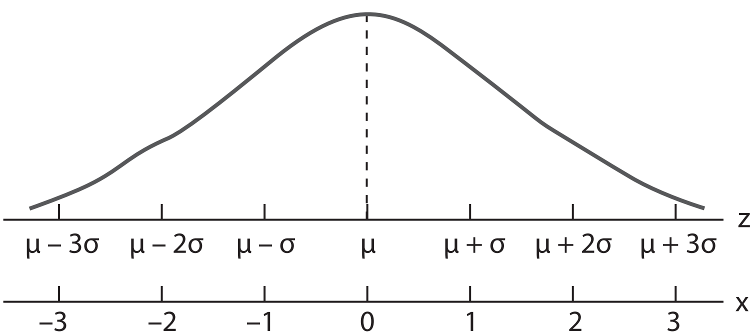


Figura 9 – Transformação da escala para unidades padronizadas.

Nesta nova escala (a escala z), um valor de **z** representa a quantos desvios-padrão o valor correspondente de **x** está da média da distribuição – de um lado ou de outro. O apêndice n.° 2 – Tabela 1 não dá valores correspondentes a valores negativos de **z**, os que são desnecessários em virtude da simetria de qualquer curva normal em relação à sua média.

O problema

Contexto A

Suponha que a fábrica de pneus "Fura menos", que produz pneus para automóveis de passeio, acabou de desenvolver um novo pneu "Roda Mais", que será vendido com desconto por meio de uma cadeia nacional de lojas. Como o pneu é um novo produto, os gerentes da "Fura menos" acreditam que a garantia de quilometragem oferecida com o pneu será um fator importante na aceitação do produto. Antes de concluírem a política de garantia de quilometragem, os gerentes da "Fura menos" desejam informações de probabilidade sobre o número de quilômetros em que os pneus se gastarão.

Explorando o problema

A partir de testes reais com os pneus em situações em estradas não pavimentadas, o grupo de engenharia estimou a quilometragem média do pneu em $\mu = 16.500$ quilômetros e o desvio-padrão em $\sigma = 5.000$. Além disso, os dados coletados indicam que a distribuição normal é uma hipótese razoável. Qual a porcentagem dos pneus que apresenta expectativa de durar mais 20.000 quilômetros?

Equacionando o problema

Uma vez que a distribuição de probabilidade tenha sido estabelecida para uma aplicação em particular, ela pode ser usada rápida e facilmente para obter a informação de probabilidade sobre o problema. A probabilidade não faz diretamente uma recomendação de decisão, mas fornece informações que auxiliam o tomador de decisão a melhor entender os riscos e incertezas associados ao problema e, finalmente, essas informações podem auxiliá-lo a tomar uma boa decisão.

Para transformar o x (20.000) em z, tem-se a seguinte condição:

Fórmula 6

$$z = \frac{x - \mu}{\sigma}$$

Passo 1:

Subtrai-se o valor de **x** (20.000) do valor da média **μ** = 16.500.

$z = 20.000 - 16.500 = 3.500$

Passo 2:

Dividiram-se as sobras dos quilômetros pelo desvio-padrão σ = 2.500.

$$z = \frac{3.500}{5.000}$$

$$z = 0{,}70$$

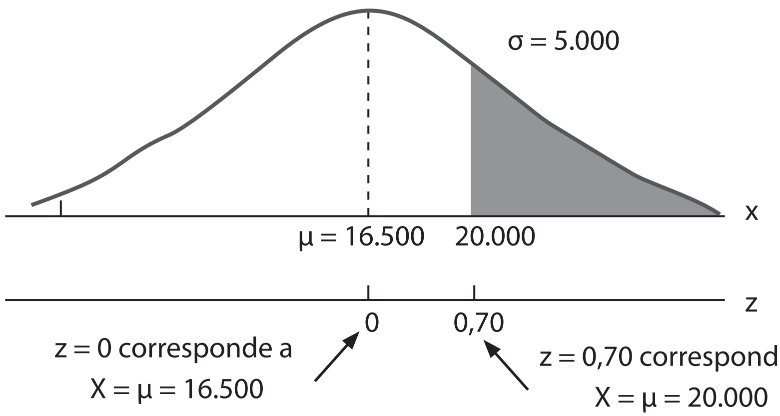


Figura 10 – Distribuição de quilometragem da companhia Ranger.

Passo 3:

Consultar o apêndice n.° 2 – tabela 1.

O valor de **x** = 20.000 na distribuição normal corresponde a um valor de z = 0,70 na distribuição normal-padrão. Usando o apêndice n.° 2 - tabela 1, vemos que a área entre a média e **z** = 0,7 é 0,2580.

Tabela 1 – Modelo do apêndice n.° 2 – tabela 1, z = 0,70

z	0,00	0,01
0,0	0,0000	0,0040
0,1	0,0398	0,0438
0,2	0,0793	0,0832
0,3	0,1179	0,1217
0,4	0,1554	0,1591
0,5	0,1915	0,1950
0,6	0,2257	0,2291
0,7	**0,2580**	0,2611
0,8	0,2881	0,2910
0,9	0,3159	0,3186

Passo 4:

Calcular a probabilidade.

z = 0,5000 – 0,2580 = **0,2420**

Assim: 0,2420 é a probabilidade de **x** exceder 20.000. Podemos concluir que cerca de 24,20% dos pneus excederá 20.000 na quilometragem.

Contexto B

A Ranger Time está considerando agora uma garantia que fornecerá um desconto na substituição dos pneus se os originais não excederem a quilometragem declarada na garantia. Qual deveria ser a quilometragem de garantia se a Ranger Time não quer mais do que 15% dos pneus qualificados elegíveis para a garantia de desconto?

Para o cálculo temos a seguinte condição:

Fórmula 6

$$z = \frac{x - \mu}{\sigma}$$

Passo 1:

Calcular a porcentagem da área.

% = 50% (metade da tabela) – 15% = 35%

Passo 2:

Consultar no apêndice n.° 2 – tabela 1, o valor correspondente.

35% da área precisa estar entre a média e a quilometragem de garantia desconhecida. Na tabela, 0,3500 corresponde aproximadamente a 1,04 desvios-padrões abaixo da média. Isto é, **z** = –1,04 é o valor da variável aleatória normal-padrão que corresponde à desejada garantia de quilometragem na distribuição normal da Ranger Time.

z	0,00	0,01	0,02	0,03	0,04
0,0	0,0000	0,0040	0,0080	0,0120	0,0160
0,1	0,0398	0,0438	0,0478	0,0517	0,0557
0,2	0,0793	0,0832	0,0871	0,0910	0,0948
0,3	0,1179	0,1217	0,1255	0,1293	0,1331
0,4	0,1554	0,1591	0,1628	0,1664	0,1700
0,5	0,1915	0,1950	0,1985	0,2019	0,2054
0,6	0,2257	0,2291	0,2324	0,2357	0,2389
0,7	0,2580	0,2611	0,2642	0,1673	0,2704
0,8	0,2881	0,2910	0,2939	0,2967	0,2995
0,9	0,3159	0,3186	0,3212	0,3228	0,3264
1,0	0,3413	0,3438	0,3461	0,3485	0,3508

Figura 5.13 – Modelo do apêndice n.° 2 – tabela 1, z = 1,04.

Passo 3:

Calcular o valor de **x**.

Para encontrar a quilometragem x correspondendo a z = –1,04

Fórmula 7

$$z = \frac{x - \mu}{\sigma} = -1{,}04$$

$$x - \mu = -1{,}04\ \sigma$$

$$x = \mu - 1{,}04\ \sigma$$

Com $\mu = 16.500$ e $\sigma = 5.000$

$x = 16.500 - 1{,}04 \cdot 5.000 = 11.300$

Assim, uma garantia de 11.300 quilômetros satisfará as exigências de que aproximadamente 15% dos pneus serão qualificados para a garantia? Talvez com essa informação a empresa possa estabelecer sua garantia de quilometragem em 11.000 quilômetros.

Ampliando seus conhecimentos

A seguir, diversas observações sobre as características da distribuição normal de probabilidade.

Há uma família inteira de distribuições normais de probabilidade. Elas são diferenciadas por suas médias μ, e desvios-padrão σ.

O ponto mais alto na curva está na média, que também é a mediana à moda da distribuição.

A média da distribuição pode ser de qualquer valor numérico: negativo, zero ou positivo.

A distribuição normal de probabilidade é simétrica, em que a forma da curva à esquerda da média é uma imagem espelhada da forma da curva à direita da média. Os extremos da curva estendem-se ao infinito em ambas as direções e teoricamente nunca tocam o eixo horizontal.

O desvio-padrão determina a largura da curva, valores maiores do desvio--padrão resultam em curvas mais largas e mais planas, mostrando maior variabilidade nos dados.

A área total sob a curva para a distribuição normal de probabilidade é 1, o que é verdadeiro para todas as distribuições contínuas de probabilidade.

As probabilidades para a variável aleatória normal são dadas por áreas sob a curva. As porcentagens de valores em alguns intervalos comumente usados são:

- **68,26%** dos valores de uma variável aleatória normal estão dentro de um desvio-padrão positivo ou negativo de sua média;
- **95,44%** dos valores de uma variável aleatória normal estão dentro de dois desvios-padrões positivos ou negativos de sua média;
- **99,72%** dos valores de uma variável aleatória normal estão dentro de três desvios-padrões positivos ou negativos de sua média.

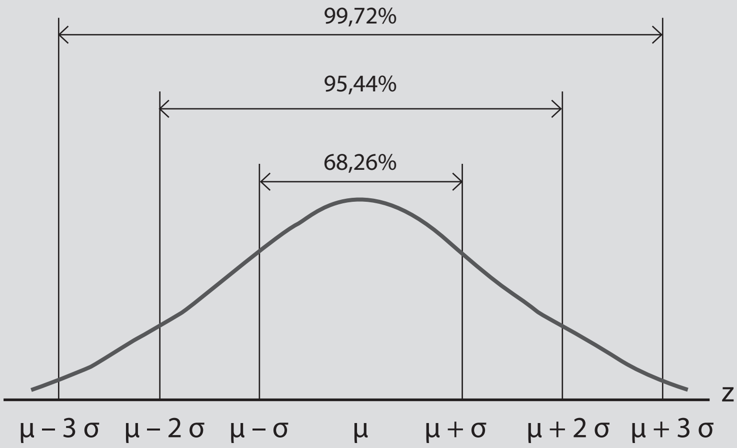


Figura – Áreas sob a curva para qualquer distribuição normal.

Teorema do limite central

Envolve duas distribuições diferentes: a distribuição da população original e a distribuição das médias amostrais. Muitos problemas importantes e de ordem prática podem ser resolvidos com o Teorema do Limite Central.

- **Dados**

A variável aleatória **x** tem distribuição (que pode ser normal, ou não), com média μ e desvio-padrão σ.

A amostra de tamanho **n** é extraída aleatoriamente dessa população.

- **Conclusões**

Na medida em que o tamanho da amostra aumenta, a distribuição das médias amostrais tende para uma distribuição normal.

A média das médias amostrais será a média populacional μ.

O desvio-padrão das médias amostrais será $\frac{\sigma}{\sqrt{n}}$.

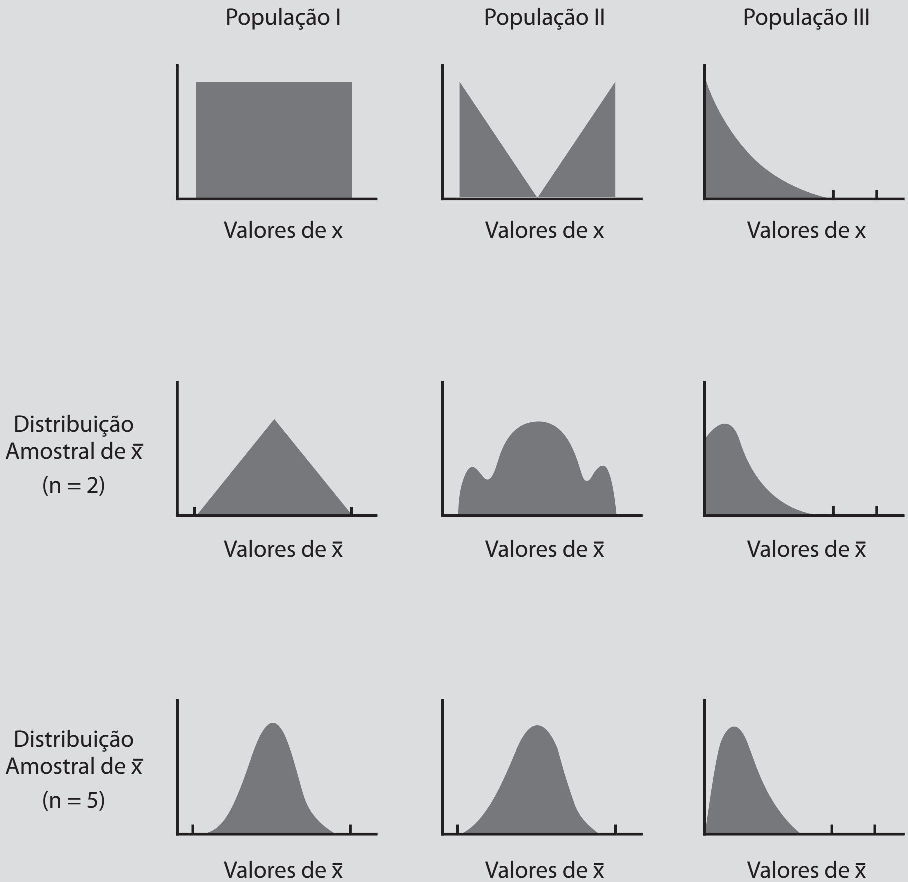
- **Regras práticas de uso comum:**

Para amostras de tamanho n > 30, a distribuição das médias amostrais pode ser aproximada satisfatoriamente por uma distribuição normal. A aproximação melhora à medida que aumenta o tamanho da amostra n.

Se a própria distribuição original tem distribuição normal, então as médias amostrais terão distribuição normal para qualquer tamanho amostral n.

Fórmula – Teorema limite central

$$z = \frac{x - \mu}{\sigma}$$



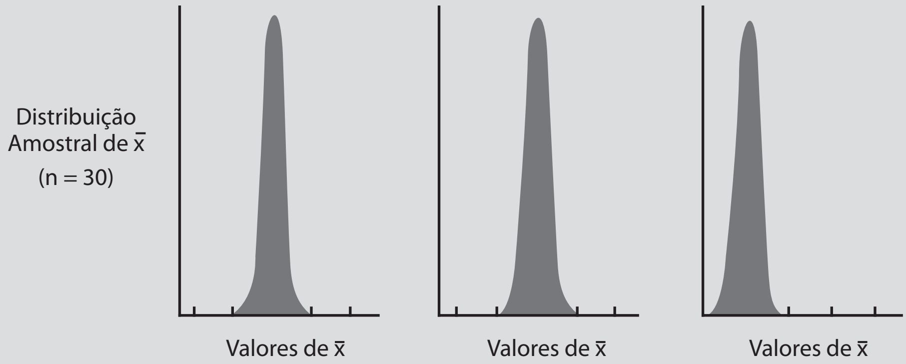


Figura – Ilustração do teorema do limite central para três populações.

Curiosidade

A Procter & Gamble (P&G) está no negócio de produtos de consumo no mundo inteiro. A P&G produz e comercializa produtos tais como detergente, fralda descartável, produtos farmacêuticos, dentifrício, sabão em barra, colutório e toalha de papel. Ela detém a marca líder em mais categorias do que qualquer outra empresa de produtos de consumo.

Líder também na aplicação de métodos estatísticos para a tomada de decisão, a P&G emprega pessoas com formação acadêmica: engenharia, estatística, pesquisa operacional e administrativa. As principais tecnologias quantitativas para as quais esses profissionais fornecem suporte são: decisão probabilística e análise de risco, simulação avançada, melhoria da qualidade e métodos quantitativos (programação linear, análise de regressão, análise de probabilidade).

A Divisão Industrial de Produtos Químicos da P&G é o principal fornecedor de álcoois graxos derivados de substâncias naturais, tais como óleo de coco e derivados do petróleo. A divisão queria avaliar os riscos econômicos e as oportunidades de expandir suas instalações de produção de álcoois graxos, e os peritos da P&G em decisão probabilística e análise de risco foram chamados para auxiliar. Depois de estruturarem e esquematizarem o problema, eles determinaram que a chave da lucratividade e a diferença de custo entre as matérias-primas baseadas no petróleo e no óleo coco. Os custos futuros eram desconhecidos, mas os analistas foram capazes de representá-los com as seguintes variáveis aleatórias contínuas.

x = preço do óleo de coco litro de álcool graxo

y = preço da matéria-prima do petróleo por quilo de álcool graxo

Devido à chave da lucratividade ser a diferença entre essas duas variáveis aleatórias, uma terceira variável aleatória **d = x – y** foi usada na análise. Os peritos foram entrevistados para determinar a distribuição de probabilidade para **x** e para **y**. Essa informação foi usada para desenvolver uma distribuição contínua de probabilidade para a diferença **d**.

A distribuição contínua de probabilidade mostrou 0,90 de que a diferença de preço seria de US$0,0655 ou menos, e uma probabilidade de 0,50 de que a diferença de preço seria de US$0,035 ou menos. Além disso, havia somente uma probabilidade de 0,10 de que a diferença de preço seria de US$0,0045 ou menos. Obs.: As diferenças de preço foram feitas para proteger dados do proprietário.

A Divisão Industrial de Produtos Químicos considerou que, sendo possível quantificar o impacto das diferenças do preço da matéria-prima, teria a chave para atingir um consenso. As probabilidades obtidas foram usadas em uma análise de sensibilidade da diferença no preço da matéria-prima. A análise produziu subsídios suficientes para embasar uma recomendação à administração.

O uso de variáveis aleatórias contínuas e suas distribuições de probabilidade foram úteis à P&G para analisar os riscos econômicos associados a sua produção de álcoois graxos.

Atividades de aplicação

1. Ache a área sob a curva normal padronizada:

   a) à esquerda de $z = 0,90$

   b) à direita de $z = 1,35$

   c) entre $z = 0,62$ e $z = 1,02$

2. Se uma variável aleatória tem distribuição normal com $\mu = 5$ e $\sigma = 5$, qual é a probabilidade de ela assumir um valor no intervalo de 6 a 9?

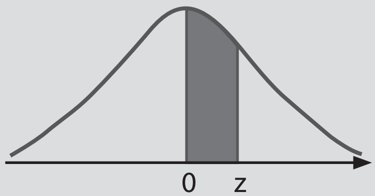
3. Constatou-se que os pesos de uma cidade de 15 000 habitantes têm distribuição normal com média $\mu = 65$ quilos, e desvio-padrão de 5 quilos. Quantas pessoas têm peso entre 65kg e 75kg?

4. Selecionado aleatoriamente um termômetro de uma amostra, determine a probabilidade de acusar uma leitura entre + 1,30° e 2,20°:

5. O adulto médio tem 1,70m de altura. Assuma que o desvio-padrão seja de 0,1m ao responder a seguinte questão. Qual a probabilidade de que um homem adulto tenha entre 1,75m e 1,80m?

Apêndice n.º 1 – Tabela 1: Áreas sob a curva normal

Distribuição normal



Os valores da tabela 1 são as probabilidades de uma variável aleatória com distribuição normal padronizada tomar um valor entre 0 e z; são dadas pela área sombreada entre 0 e z sob a curva na figura ao lado.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,4038	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817

2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Gabarito

1.

a) a área à esquerda de z = 0,90 é igual a 0,5000 mais o valor da Tabela 1 correspondente a z = 0,90, ou **0,5000 + 0,3159 = 0,8159**

b) a área à direita de z = 1,35 é 0,5000 menos o valor da Tabela 1 correspondente a 1,35, ou **0,5000 – 0,4115 = 0,0885**

c) a área entre z = 0,62 e z = 1,02 é a diferença entre os valores da Tabela 1 correspondente a z = 0,62 e z = 1,02, ou **0,3461 – 0,2324 = 0,1137**

2. Transformando x = 6 e x = 9 em unidades padronizadas, vem

$z = \dfrac{6-5}{5} = 0{,}20 \qquad z = \dfrac{9-5}{5} = 0{,}80$

E como os valores correspondentes na Tabela 1 são 0,0793 e 0,2881, a probabilidade procurada é **0,2881 – 0,0793 = 0,2088 ou 20,88%**

3. Transformando x = 65 e x = 75 em unidades padronizadas

$z = \dfrac{65-65}{5} = \mathbf{0{,}00} \qquad z = \dfrac{75-65}{5} = \mathbf{2{,}00}$

E como os valores correspondentes na Tabela 1 são 0,0000 e 0,4772, a probabilidade procurada é 0,4772 – 0,0000 = **0,4772 ou 47,72 %**

Para conhecer a quantidade de pessoas 15.000 × 0,4772 = **7.158**

7 158 pessoas pesam entre 65kg e 75kg.

4. Na Tabela 1 z = 1,30 corresponde a uma área de 0,4032, e que z = 2,30 corresponde a uma área de 0,4893 ou 0,4893 – 0,4032 = **0,0861 ou 8,61%**

5. Transformando x = 1,75 e x = 1,80 em unidades padronizadas

   $z = \dfrac{1,75 - 1,70}{0,1} = 0,50 \qquad z = \dfrac{1,80 - 1,70}{0,1} = 1,00$

   E como os valores correspondentes na Tabela 1 são 0,1915 e 0,3413, a probabilidade procurada é 0,3413 – 0,1915 = **0,1498 ou 14,98%**

Referências

ANDERSON, D. R. *et al*. **Estatística Aplicada à Administração e Economia**. São Paulo: Pioneira, 2000.

COSTA NETO, P. L. de O. **Estatística**. São Paulo: Edgard, Blucher, 1977.

CRESPO, A. A. **Estatística Fácil**. São Paulo: Saraiva, 1987.

DOWNING, D. **Estatística Aplicada**. Tradução de: FARIAS, A. A. São Paulo: Saraiva, 2002.

FEUND, J. E.; SIMON, G. A. **Estatística Aplicada** – Economia, Administração e Contabilidade. Porto Alegre: Bookman, 2000.

GUIMARÃES, R. C.; CABRAL, J. A. S. **Estatística**. Portugal: McGraw-Hill de Portugal, 1998.

MOORE, D. **A Estatística Básica e sua Prática**. Rio de Janeiro: LTC, 2000.

MORETTIN, L. G. **Estatística Básica**. São Paulo: LCTE, 1983.

PAIVA, A. F. **Estatística**. Minas Gerais: UFMG. vol 1.

SMAILES, J. **Estatística Aplicada à Administração com Excel**. São Paulo: Atlas, 2007.

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

_____. **Estatística**. 2. ed. São Paulo: McGraw-Hill, 1984.

STEVENSON, W. J. **Estatística Aplicada à Administração**. São Paulo: Harbra, 1986.

_____. **Estatística Aplicada à Administração**. São Paulo: Harper & Row do Brasil, 1981.

TRIOLA, M. F. **Introdução à Estatística**. Rio de Janeiro: LTC, 1999.

Consulta online

<www.ibge.gov.br/home/disseminacao/eventos/missao/instituicao.shtm.>.

<www.ufrrj.br/instituto/it/de/acidentes/etanol5.html>.

Métodos Quantitativos Estatísticos

IESDE
Inteligência Educacional

www.iesde.com.br

Fundação Biblioteca Nacional
ISBN 978-85-387-2986-0