GESTÃO
E NEGÓCIOS

Estatística EMPRESARIAL

Juliane Silveira Freire da Silva

UNIVERSIDADE LaSalle EAD

ESTATÍSTICA EMPRESARIAL

Juliane Silveira Freire da Silva

Universidade LaSalle EAD

2ª Edição
Julho | 2019

S586e Silva, Juliane Silveira Freire da.
Estatística empresarial / Juliane Silveira Freire da Silva. – Canoas, RS : Universidade La Salle EAD, 2018.
138 p. : il. ; 30 cm. – (Gestão e negócios)

Bibliografia.

1. Matemática. 2. Estatística. 3. Estatística empresarial. 4. Administração de empresas – Métodos estatísticos. I. Título. II. Série.

CDU: 519.22

Bibliotecário responsável: Samarone Guedes Silveira - CRB 10/1418

Universidade La Salle Canoas | Av. Victor Barreto, 2288 | Canoas - RS
CEP: 92010-000 | 0800 541 8500 | ead@unilasalle.edu.br

APRESENTAÇÃO

Prezado estudante,

A equipe de gestão da EaD Unilasalle sente-se honrada em entregar a você este material didático. Ele foi produzido com muito cuidado para que cada Unidade de estudos possa contribuir com seu aprendizado da maneira mais adequada possível à modalidade que você escolheu para estudar: a modalidade a distância. Temos certeza de que o conteúdo apresentado será uma excelente base para o seu conhecimento e para sua formação. Por isso, indicamos que, conforme as orientações de seus professores e tutores, você reserve tempo semanalmente para realizar a leitura detalhada dos textos deste livro, buscando sempre realizar as atividades com esmero a fim de alcançar o melhor resultado possível em seus estudos. Destacamos também a importância de questionar, de participar de todas as atividades propostas no ambiente virtual e de buscar, para além de todo o conteúdo aqui disponibilizado, o conhecimento relacionado a esta disciplina que está disponível por meio de outras bibliografias e por meio da navegação online.
Desejamos a você um excelente módulo e um produtivo ano letivo. Bons estudos!

Gestão de EaD Unilasalle

APRESENTANDO A AUTORA

Juliane Silveira Freire da Silva

Bacharel em Estatística, pela Universidade Federal do Rio Grande do Sul (2001) e mestre em Engenharia de Produção, pela Universidade Federal do Rio Grande do Sul (2006). Atualmente, atua como professora da Universidade La Salle – Canoas e como professora do Instituto Brasileiro de Gestão de Negócios. Possui experiência em análises estatísticas, banco de dados, engenharia de produção, área da qualidade, pesquisa em marketing, análises financeiras.

APRESENTANDO A DISCIPLINA

Caros alunos, sejam bem-vindos à disciplina de Estatística Empresarial. Trataremos, ao longo do percurso de estudos, sobre conceitos básicos de Estatística. Inicialmente, vamos situá-los quanto ao pensamento estatístico, apresentando alguns exemplos simples para podermos aprofundar esse conhecimento.

A Estatística traz um enorme ganho aos profissionais de todas as áreas, e suas técnicas ajudam muito o profissional para a tomada de decisão.

Muitas vezes, as empresas necessitam de uma tomada de decisão, pois, se os responsáveis não tiverem os dados anotados e organizados, as decisões serão tomadas com base no "achismo", o que, no mercado competitivo atual, é impensável. Assim, podemos, por exemplo, querer analisar a sazonalidade de algum produto. Isso é possível acompanhando a evolução da média de vendas mensais por meio de estatísticas descritivas e gráficos. Talvez seja necessário calcular, também, uma projeção de gastos para o próximo ano, através da análise de correlação. Como são muitas as ferramentas estatísticas que podem fazer parte do dia a dia da empresa, veremos algumas ao longo do curso.

Nesse sentido, vamos construir um raciocínio lógico de conceitos estatísticos que, de maneira conjunta ou separadamente, trará o diferencial profissional de vocês.

Sumário

UNIDADE 3

UNIDADE 4

GESTÃO E NEGÓCIOS

unidade 1

Representações Estatísticas, Banco de Dados e Estatística Descritiva

Prezado(a) estudante.

Estamos começando uma unidade desta disciplina. Os textos que a compõem foram organizados com cuidado e atenção, para que você tenha contato com um conteúdo completo e atualizado tanto quanto possível. Leia com dedicação, realize as atividades e tire suas dúvidas com os tutores. Dessa forma você, com certeza, alcançará os objetivos propostos para essa disciplina.

OBJETIVO GERAL

Aprender a organizar um banco de dados, extrair tabelas e gráficos e calcular as medidas de posição central (média, moda e mediana) e as medidas de variabilidade (variância, desvio-padrão e coeficiente de variação).

OBJETIVOS ESPECÍFICOS

- Exercitar o raciocínio lógico na interpretação e solução de situações-problema envolvendo estatística.
- Compreender conceitos e nomenclaturas básicas de estatística.
- Aprender a organizar um banco de dados, bem como analisar tabelas e gráficos.
- Propiciar o entendimento de cálculos de medidas de posição e de variabilidade para dados dispostos em rol e em tabelas de distribuição de frequências.
- Analisar e interpretar as medidas descritivas.

UNIVERSIDADE LaSalle EAD

1.1 Conceitos Básicos

Você já se deparou com uma planilha de Excel cheia de informações, cheia de dados, e precisou tomar alguma decisão com base nessa planilha? Os dados brutos não nos possibilitam tirar conclusões antes de organizarmos as informações, para que, então, possamos elaborar tabelas, gráficos e calcular as medidas descritivas das variáveis. Para isso, aprenderemos todos esses conceitos básicos nesta unidade.

Vamos, nesse primeiro momento, familiarizar-nos com alguns conceitos e nomenclaturas da estatística.

1.1.1 Divisões da Estatística

A Estatística costuma ser dividida em três áreas: estatística descritiva, estatística inferencial e probabilidade.

A estatística descritiva é utilizada quando queremos resumir os nossos dados, quando, por exemplo, possuímos bancos de dados e precisamos apresentá-los de forma que seja fácil interpretar os fenômenos que estão acontecendo a partir deles. É a parte da estatística que cuida da descrição dos dados, quer sejam amostrais ou populacionais. A estatística descritiva os apresenta de forma inteligível e resumida.

A inferência estatística, ou indução estatística, faz a generalização do que é estudado na estatística descritiva, a partir das amostras, para as populações. A inferência estatística utiliza-se de dados amostrais para fazer generalizações para toda a população em estudo, desde que os pressupostos de amostragem probabilística sejam cumpridos. Um exemplo típico e mais próximo de nossa realidade são as pesquisas eleitorais: ocorre quando apenas uma amostra (pequena parte dos eleitores) é entrevistada e o resultado é anunciado como sendo a preferência de toda a população de uma cidade.

Já a probabilidade estuda a incerteza dos fenômenos, quando lidamos com o acaso. Por exemplo, a probabilidade de chuva para amanhã ou a probabilidade de um equipamento eletrônico estragar depois de passar o tempo de garantia.

1.1.2 Definições Básicas

No exemplo dos eleitores, citado anteriormente, vimos que a estatística se utiliza de alguns conceitos como população e amostra. A seguir, então, vamos entender melhor o que eles significam e veremos alguns outros conceitos relacionados.

1.1.2.1 População

O termo **população** é utilizado quando temos um grupo de pessoas, animais ou objetos que tenham pelo menos um atributo comum a todos os elementos do conjunto que a compõe. Dessa forma, quando nos referimos aos eleitores, no caso de uma pesquisa eleitoral, estamos identificando a população como o conjunto de pessoas que votam – sendo assim, a característica comum a todos os elementos da população é ser eleitor. Uma população pode ter mais de uma característica comum quando queremos restringir ainda mais o universo pesquisado, por exemplo: uma população pode ser definida como o conjunto de empresas da área do comércio, do estado do Rio Grande do Sul, que tenham mais de vinte empregados e que estejam no mercado há pelo menos oito anos.

1.1.2.2 Amostra

O termo **amostra** é definido como um subconjunto da população. Amostra é o subconjunto de uma população cujas propriedades se estuda com o fim de generalizá-las para a população.

Uma consideração importante sobre a amostra é a de que ela não seja tendenciosa e seja representativa do universo. Não ser tendenciosa depende do processo de amostragem que vai ser utilizado.

1.1.2.3 Censo e Amostragem

Sempre que realizamos um **censo**, estamos coletando os dados referentes a toda a população. Quando fazemos uma **amostragem**, estamos coletando os dados referentes à nossa amostra.

1.1.2.4 Parâmetro e Estatística

Sempre que tivermos uma medida numérica que descreve a característica de uma população, temos um **parâmetro** dessa população. Por exemplo: representamos a média de uma população pela letra grega μ (mi), o desvio-padrão pela letra grega σ (sigma) e o tamanho dessa população por N. Os parâmetros populacionais serão representados por letras do alfabeto grego ou letras maiúsculas do alfabeto latino.

Quando representarmos uma medida numérica que descreve uma característica de uma amostra, teremos uma **estatística** ou um **estimador**. Por exemplo: representamos a média de uma amostra com o $\overline{x}$ (a letra *x*, com uma barra em cima), o desvio-padrão pela letra minúscula s e o tamanho da amostra n. Os estimadores serão representados por letras minúsculas do alfabeto latino.

1.1.3 Tipos de Variáveis

Uma variável é uma característica de uma população que difere de um indivíduo para outro e a qual temos interesse em estudar. Cada unidade amostral da população que é coletada como parte de uma amostra proporciona uma medida de uma ou mais variáveis, a que chamamos de observações.

Variáveis são as observações ou as características investigadas (ex.: sexo, número de acidentes, peso). As variáveis dividem-se em dois grandes grupos, as qualitativas e as quantitativas, conforme pode ser observado a seguir:

$$\textit{Váriáveis}\begin{cases} \textit{Qualitativas}\begin{cases}\text{Nominais}\\ \text{Ordinais}\end{cases} \\ \textit{Quantitativas}\begin{cases}\text{Discretas}\\ \text{Contínuas}\end{cases}\end{cases}$$

1.1.3.1 Variáveis Qualitativas

As **variáveis qualitativas** são as observações que apresentam qualidades ou atributos e se dividem em variáveis qualitativas nominais e qualitativas ordinais.

As **variáveis qualitativas nominais** são as observações para as quais não existe ordenação na sua organização; existe apenas uma qualidade atribuída à resposta, como:

- Sexo dos empregados de uma empresa (opções de resposta: masculino, feminino).
- Modelos de automóveis fabricados pela empresa X (opções de resposta: modelo Beta, modelo Alfa, modelo Sigma etc.).
- Tipos de defeito em um circuito eletrônico (opções de resposta: falha de solda, capacitor queimado etc.).
- Raça de cachorros (opções de resposta: boxer, yorkshire, maltês, pug etc.).

As **variáveis qualitativas ordinais** são as observações para as quais existe certa ordenação possível na sua organização, como:

- Escala de satisfação com o serviço prestado (opções de resposta: muito satisfeito, satisfeito, indiferente, insatisfeito, muito insatisfeito).
- Classificação de porte de uma empresa (opções de resposta: pequeno porte, médio porte, grande porte).

As escalas de **Likert** são sempre qualitativas ordinais. Essas escalas sempre terão um número ímpar de opções de resposta, pois elas sempre propõem o mesmo número de pontos positivos e negativos e ainda dão uma opção neutra.

GLOSSÁRIO

Escala de Likert: é um tipo de escala, utilizada em questionários, que apresenta alternativas em níveis de concordância com a questão ou afirmação proposta.

1.1.3.2 Variáveis Quantitativas

As **variáveis quantitativas** são as observações que apresentam números, resultantes de contagem ou mensuração e se dividem em variáveis quantitativas discretas e quantitativas nominais.

As **variáveis quantitativas discretas** são valores que formam um conjunto finito ou enumerável de números e resultam, frequentemente, de uma contagem, assumindo valores inteiros. Por exemplo:

- número de acidentes por mês em uma empresa;
- número de carros vendidos por ano no Brasil etc.

As **variáveis quantitativas contínuas** são valores que formam um intervalo de números reais e resultam, normalmente, de uma mensuração, ou seja, os resultados podem assumir qualquer valor na reta dos reais. Por exemplo:

- tempo de atuação de uma empresa;
- taxas de juros aplicados no mercado financeiro no ano passado etc.

1.1.4 Dados Estatísticos

Podemos definir dados estatísticos como sendo representações numéricas das variáveis estudadas. A princípio, qualquer dado pode ser quantificado.

Cada pergunta de um questionário ou levantamento vai gerar uma variável e, por consequência, uma quantidade de dados estatísticos. Nós podemos trabalhar esses dados de duas maneiras: podemos ter dados brutos ou dados elaborados. Os dados brutos resultam de observação direta, levantamento ou mediante observação pura e simples, não sofrendo qualquer alteração. Os dados dispostos dessa maneira não podem ajudar muito no entendimento do fenômeno, mas eles são o ponto de partida para os dados elaborados, que resultam da análise dos dados brutos.

Também podemos classificar os dados pela sua maneira de apresentação: podemos ter dados em rol, como veremos adiante, e em tabelas de distribuição de frequência (por ponto ou por intervalo).

Vejamos um exemplo de um banco de dados e consideremos uma pesquisa de satisfação com o atendimento de uma empresa de telefonia celular. Em seguida, serão abordadas as maneiras de se representar os dados estatísticos.

Tabela 1.1: Exemplo de banco de dados de uma empresa de telefonia quanto à satisfação do cliente.

Entrevista	Sexo	Tempo (em anos) como cliente	Possui internet?	Tipo de plano	Nota de Satisfação com a Operadora (0 a 10)
1	F	4	Não	Pré-pago	7
2	M	5	Sim	Pós-pago	6
3	M	1	Não	Pré-pago	6
4	F	9	Sim	Pré-pago	8
5	F	8	Não	Pré-pago	7
6	M	6	Sim	Pós-pago	9
7	M	2	Sim	Pós-pago	8
8	F	5	Sim	Pós-pago	8
9	F	6	Não	Pré-pago	5
10	M	2	Sim	Pós-pago	6
11	F	8	Não	Pré-pago	7
12	M	5	Sim	Pós-pago	9
13	M	3	Sim	Pós-pago	8
14	F	6	Sim	Pós-pago	8
15	F	1	Não	Pré-pago	5
16	M	3	Sim	Pós-pago	6

Fonte: Elaborada pela autora (2016).

1.1.4.1 Dados em Rol

Quando tivermos os **dados em rol**, estaremos diante da forma mais simples de apresentarmos os dados pesquisados. Nesse tipo de apresentação, os dados são dispostos de forma que a informação coletada se apresenta individualmente. É mais indicado para quando a quantidade de dados não for muito grande, ou seja, a quantidade de dados precisa deixar fácil o entendimento: não existe quantidade específica e o bom senso é o que indica isso.
Nesse tipo de disposição dos dados, sempre existe a possibilidade de os dados serem agrupados por suas frequências, desde que o nível de mensuração da variável permita. Quando trabalhamos com dados em rol, temos condições de saber exatamente o valor correspondente entre cada unidade pesquisada e a fonte da informação.

Porém, em muitas situações fica complicado estudarmos os dados em rol, pois o tamanho de amostra normalmente é muito grande. Quando temos amostras grandes fica trabalhoso e pouco interessante apresentarmos os dados de forma isolada.

Segundo Stevenson (2001), em sua forma não processada os dados podem quase não ter sentido. Grandes quantidades de números tendem a confundir, ao invés de esclarecer, simplesmente porque nossa mente não é capaz de abranger a variedade e os detalhes inerentes a grandes conjuntos de números.

1.1.4.2 Tabela de Distribuição de Frequências por Ponto

Um dado em uma **tabela de distribuição de frequência** é organizado de forma a não se precisar trabalhar com uma quantidade muito grande de informações, ou seja, os dados que estavam em rol passam a ser agrupados em uma tabela de distribuição de frequências, diminuindo a informação visual e facilitando muito o entendimento do fenômeno com os dados organizados.

Na Tabela 1.2, temos o exemplo de uma distribuição de frequência por ponto para a variável “quantidade de alunos por turma” em uma escola de informática. No exemplo, temos cinquenta turmas.

Tabela 1.2: Tabela de distribuição de frequências do número de alunos por turma em uma escola de informática.

Número de alunos por turma	Frequência observada (f_i)	Frequência relativa (f_{ri})
9	14	28,0 = (14/50*100)
12	12	24,0 = (12/50*100)
13	8	16,0 = (8/50*100)
15	9	18,0 = (9/50*100)
18	7	14,0 = (7/50*100)
Total	50	100

Fonte: Elaborada pela autora (2016).

Quando f_i é a frequência simples absoluta (que resulta da contagem das frequências dos dados), pode ser chamada também de frequência observada. E f_{ri} é a frequência simples relativa, que resulta do percentual obtido da frequência observada pelo total de respostas.

1.1.4.3 Tabela de Distribuição de Frequências por Classes (ou intervalos)

Existem situações em que a tabela de distribuição de frequências por pontos não é suficiente para representar os dados. Isso pode ocorrer quando temos variáveis quantitativas contínuas ou quando temos uma variabilidade muito alta entre os dados. Então, recomenda-se que, quando a tabela de distribuição de frequências por ponto ultrapasse 10 linhas, a variável passe a ser disposta em uma tabela de distribuição de frequências por classes (também conhecida como tabela de distribuição de frequências por intervalos). Pode ocorrer, também, de haver casos em que não há muita repetição de dados, ou mesmo haja frequências muito pequenas, o que também não reduziria a quantidade de dados pesquisados.

Tomemos como exemplo as faixas etárias dos funcionários de uma empresa.

Tabela 1.3: Tabela de distribuição de frequências para a faixa etária dos funcionários de uma empresa.

Faixa etária	Frequência observada (f_i)	Frequência relativa (f_{ri})	Frequência acumulada (F_i)	Frequência acumulada relativa (F_{ri})
18 ⊢ 28	12	24,0	12	24,0
28 ⊢ 38	18	36,0	30	60,0
38 ⊢ 48	13	26,0	43	86,0
48 ⊢ 58	7	14,0	50	100,0
Total	50	100,0	-	-

Fonte: Elaborada pela autora (2016).

Em que:

A ⊢ B (Inclui A e não inclui B).

A ⊣ B (Inclui B e não inclui A).

A ⊢⊣ B (Inclui A e B).

A — B (Não inclui A e B).

Observemos que a Tabela 1.3 inclui mais duas colunas. Na apresentação de resultados, elas não são obrigatórias, pois se tem como padrão apenas as três primeiras; porém, as duas últimas podem auxiliar o pesquisador na interpretação dos dados e não precisam ser apresentadas para efeitos de relatórios de pesquisa.

Essas duas últimas colunas representam as frequências acumuladas: uma referente às frequências observadas (f_i), em que repetimos a primeira frequência na primeira linha, e depois vamos acumulando essas observações. Essa seria a frequência acumulada absoluta (F_i). A outra é referente à frequência relativa (f_{ri}), em que fazemos o mesmo procedimento com os percentuais e obtemos a frequência acumulada relativa (F_{ri}).

1.1.5 Tipos de Gráficos

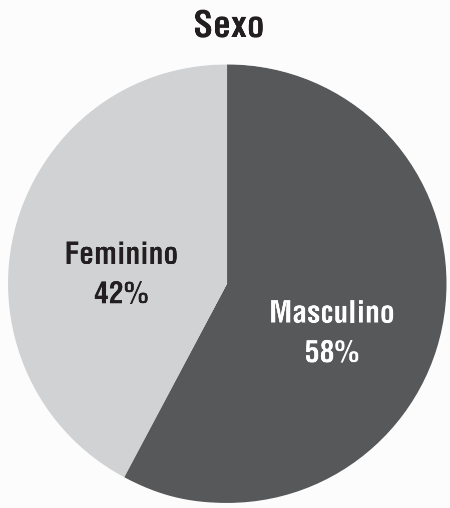
Muitas vezes precisamos de recursos que permitem uma leitura mais rápida dos dados. Nesses casos, fazemos uso de representações gráficas. Pensem comigo: ficaria fácil de fazer uma apresentação para a diretoria de uma empresa apenas com tabelas? Seria possível, mas também seria provável que os ouvintes ficassem mais tempo desvendando as tabelas do que prestando atenção nos resultados que importam.

Devemos levar em consideração a simplicidade nos recursos gráficos, a clareza e a veracidade dos dados. Vejamos alguns exemplos a seguir.

1.1.5.1 Gráfico de Setores

O **gráfico de setores**, também conhecido como gráfico pizza, é um dos mais simples e um dos mais utilizados. Sua construção é baseada no fato de que um círculo possui 360º, e é dividido em fatias de acordo com o percentual em cada categoria. É um gráfico utilizado para representar variáveis nominais ou apresentadas em categorias de respostas, ou seja, é utilizado para variáveis qualitativas.

Figura 1.1: Exemplo de gráfico de setores.



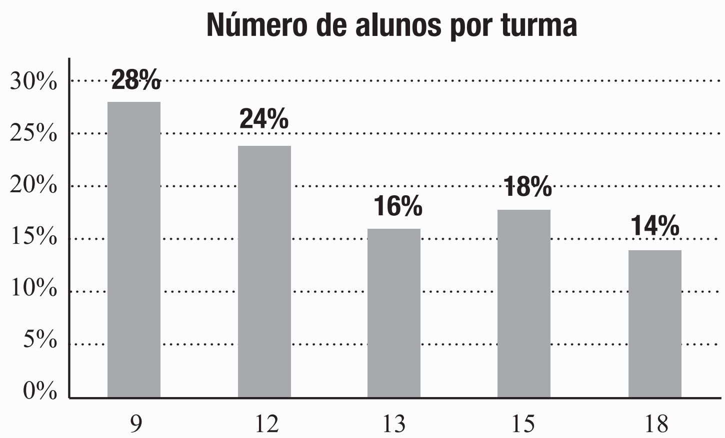
Fonte: Elaborada pela autora (2016).

Note que a leitura dos dados fica mais rápida do que se tivéssemos uma tabela. Vejamos outro tipo de gráfico.

1.1.5.2 Gráfico de Colunas

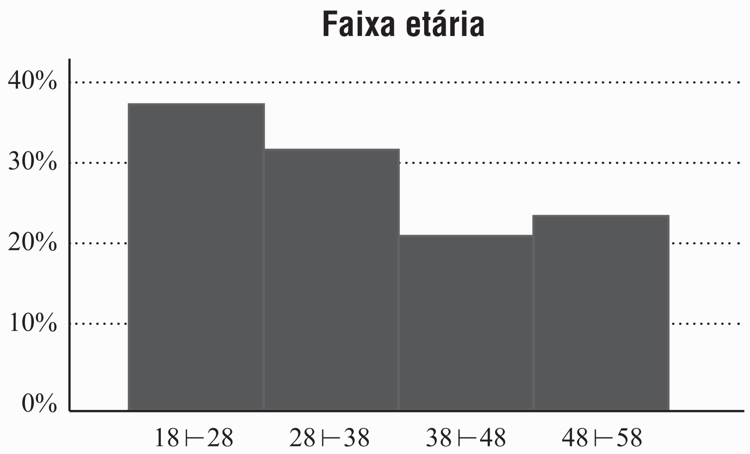
O **gráfico de colunas** tem a sua representação através de retângulos dispostos verticalmente. A altura desses retângulos é proporcional às suas respectivas frequências. Esse gráfico pode ser utilizado para representar qualquer tipo de variável. No caso de uma variável quantitativa disposta em uma distribuição de frequências por classes, essas colunas devem estar "grudadas" umas às outras. Com isso, o gráfico passa a se chamar histograma.

Figura 1.2: Exemplo de gráfico de colunas.



Fonte: Elaborada pela autora (2016).

Figura 1.3: Exemplo de histograma.

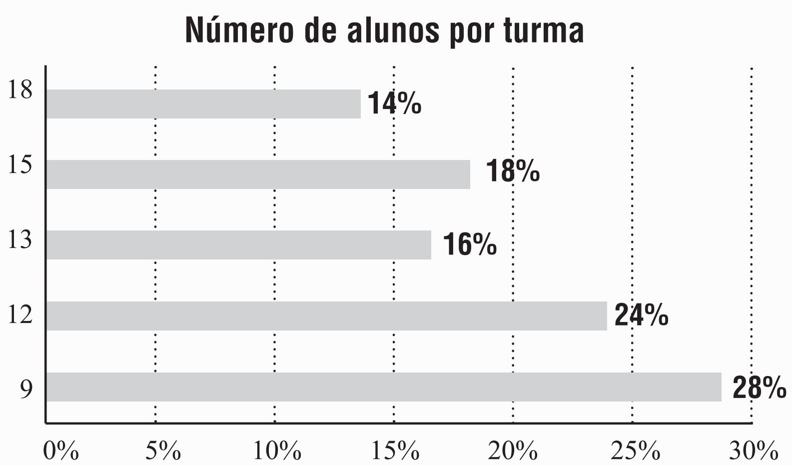


Fonte: Elaborada pela autora (2016).

1.1.5.3 *Gráfico de Barras*

O **gráfico de barras** segue a mesma lógica do gráfico de colunas, com a diferença de as barras estarem dispostas horizontalmente.

Figura 1.4: Exemplo de gráfico de barras.

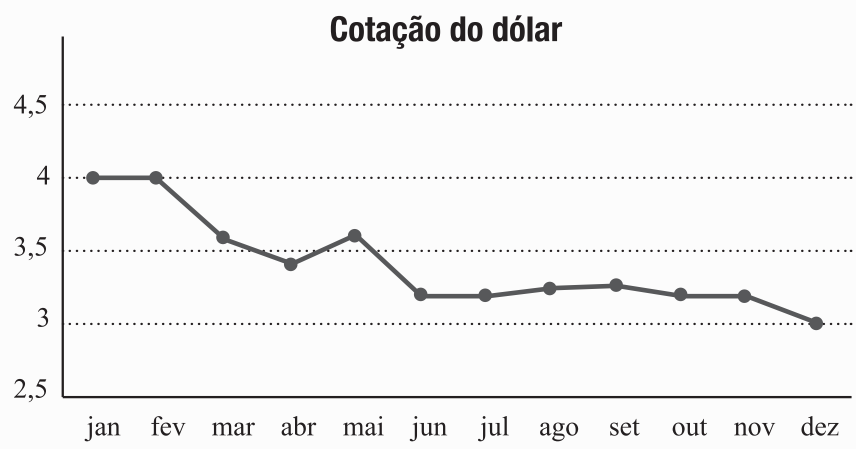


Fonte: Elaborada pela autora (2016).

1.1.5.4 *Gráfico de Linhas*

O **gráfico de linhas** é utilizado para representar séries estatísticas, ou seja, ele serve para representar uma variável quantitativa ao longo do tempo, para acompanhar a sua evolução. Seu traçado deve ser realizado considerando o eixo "*x*" (horizontal), a escala de tempo e o eixo "*y*" (vertical), de acordo com a frequência observada dos valores.

Figura 1.5: Exemplo de gráfico de linhas.



Fonte: Elaborada pela autora (2016).

1.1.6 Medidas de Tendência Central (ou posição)

Outra forma de representar os dados é fazendo alguns cálculos para resumir dados com um ou mais valores que representam uma série. Iniciamos com os cálculos cujas medidas representam a posição aproximada na qual os dados estão "circulando".

1.1.6.1 Média Aritmética ($\overline{x}$ ou μ)

A média aritmética é a medida de posição mais utilizada e a mais conhecida. Pensem: quem nunca fez a média das notas no colégio para saber com quanto ficaria?

Essa medida busca encontrar um ponto de equilíbrio entre os valores da variável. O cálculo da média é simples e de alguma maneira todos nós já a calculamos. Basta somarmos todos os valores da nossa série e dividirmos pela quantidade de valores considerados.

Média aritmética populacional: μ_x. Fórmula: $\mu_x = \dfrac{\sum X_i}{N}$

Em que:
X_i = cada um dos valores da minha população;
N = tamanho da minha população.

Média aritmética amostral: $\overline{x}$ Fórmula: $\overline{x} = \dfrac{\sum x_i}{n}$

Em que:
x_i = cada um dos valores da minha amostra;
n = tamanho da minha amostra.

Observe que, se a média utiliza em seu cálculo todos os valores da série, logo, quando existem pontos discrepantes (números que destoam da maioria dos valores), eles podem distorcer o resultado da média.

Lembre-se que podemos ter a representação dos dados em rol e em tabelas de distribuição de frequências. Então, para efeitos de cálculo, será apresentada a maneira de calcular as medidas de posição para os dados em rol e para os dados em tabelas de distribuição de frequências por ponto.

Exemplo 1.1: Para dados em rol, consideremos a pesquisa de satisfação com o atendimento de uma empresa de telefonia celular e a variável da nota de satisfação com a operadora (de 0 a 10).

Considere os dados de nota para a empresa de telefonia celular: 7, 6, 6, 8, 7, 9, 8, 8, 5, 6.

$$\text{Resolução: } \overline{x} = \frac{\sum x_i}{n} = \frac{7+6+6+8+7+9+8+8+5+6}{10} = \frac{70}{10} = 7$$

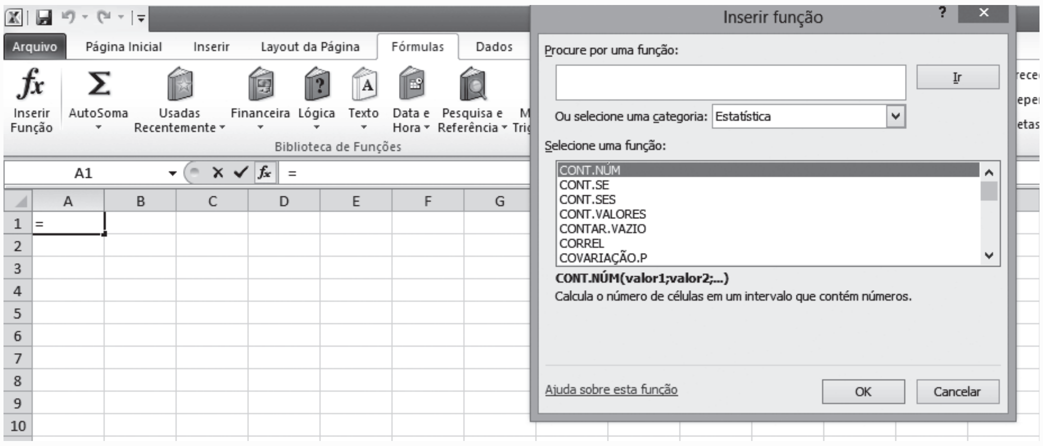
A interpretação desse resultado é que a nota média que os clientes da empresa de telefonia deram para a satisfação com a operadora foi 7. Perceba que o valor da média tem que estar entre o maior e o menor valor da série.

AMPLIANDO CONHECIMENTO

No *link* a seguir você acessa um vídeo sobre os tipos de médias.
Não deixe de conferir! https://goo.gl/41PGLS. Acesso em: 9 jul. 2017.

Veremos como calcular a média para dados em rol no Excel. Para todas as medidas de posição, existe uma fórmula já criada no Excel. Assim, é necessário que se procure a aba “fórmulas” e depois se insira a função no canto esquerdo. Com isso, surgirá a janela da figura a seguir, bastando selecionar a categoria estatística. Desse modo, serão apresentadas várias funções estatísticas, basta procurar pela média.

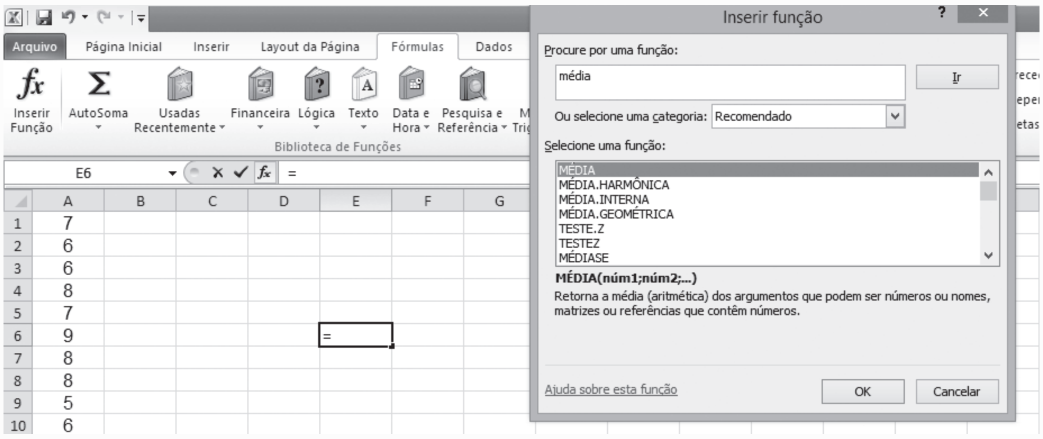
Figura 1.6: Inserir função no Excel.



Fonte: Elaborada pela autora (2016).

Selecionando a função Média, basta selecionar os dados da amostra e obter a média.

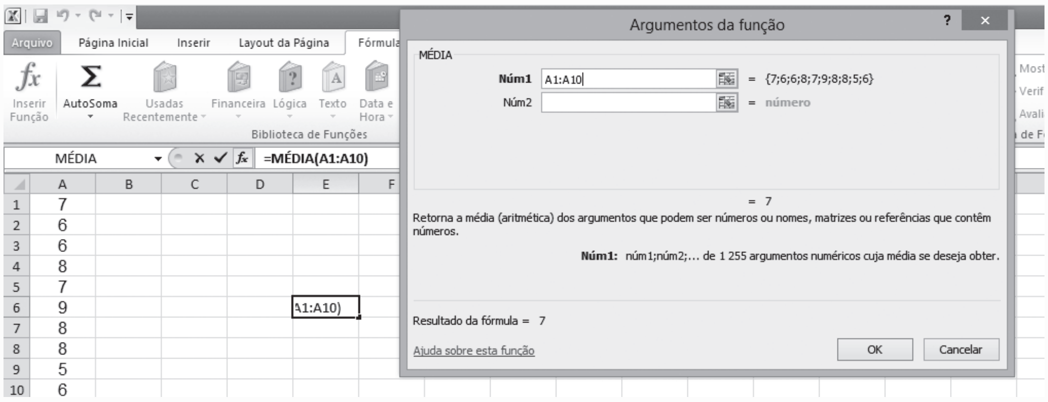
Figura 1.7: Função média.



Fonte: Elaborada pela autora (2016).

Nessa janela, clique em "OK" e selecione os dados.

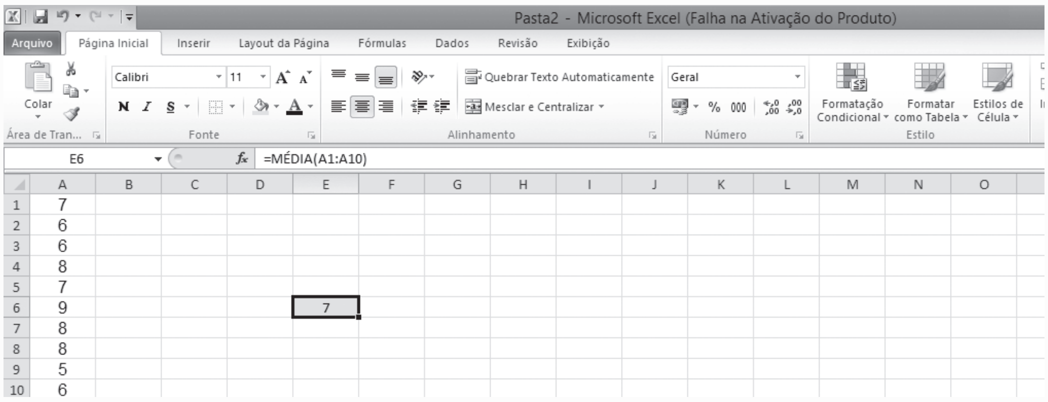
Figura 1.8: Selecionar dados.



Fonte: Elaborada pela autora (2016).

Ao clicar em "OK", a média aparecerá na célula selecionada.

Figura 1.9: Resultado da média.



Fonte: Elaborada pela autora (2016).

Vamos, agora, ver um exemplo de cálculo de média para dados em tabela de distribuição de frequências por ponto, considerando - como nosso exemplo - a quantidade de alunos por turma em uma escola de informática.

Tabela 1.4: Tabela de distribuição de frequências para o número de alunos por turma.

Número de alunos por turma	Frequência observada (f_i)	$(x_i \cdot f_i)$
9	14	(9 x 14) = 126
12	12	(12 x 12) = 144
13	8	(13 x 8) = 104
15	9	(15 x 9) = 135
18	7	(18 x 7) = 126
Total	50	635

Fonte: Elaborada pela autora (2016).

Resolução:

$$\overline{x} = \frac{\sum x_i f_i}{n} = \frac{(9\,.\,14) + (12\,.\,12) + (13\,.\,8) + (15\,.\,9) + (18\,.\,7)}{50} = \frac{635}{50} = 12,7$$

Ao interpretarmos esse resultado, vemos que, em média, as turmas da escola possuem 12,7 alunos. Ou, melhor, a média de alunos por turma é igual a 13.

Observe que, como a variável é quantitativa discreta, podemos arredondar o resultado para um número inteiro.

Então, na fórmula da média para os dados em tabela de distribuição de frequências, precisamos multiplicar cada um dos resultados que obtivemos pela quantidade de vezes em que ele se repete na nossa série.

Média aritmética populacional: μ_x Fórmula: $\mu_x = \frac{\sum X_i f_i}{N}$

Média aritmética amostral: $\bar{x}$ Fórmula: $\bar{x} = \frac{\sum x_i f_i}{n}$

1.1.6.2 Mediana (m_d)

É o valor que divide o conjunto de dados ao meio, desde que o conjunto de dados esteja ordenado de forma crescente ou decrescente. Então, a mediana é a medida que leva em consideração o valor central da série estatística ordenada.

Prefere-se empregar a mediana quando se deseja obter o ponto que divide a distribuição dos valores em duas partes iguais ou quando temos uma série de dados com valores extremos (muito destoantes do geral da amostra) que afetam de uma maneira acentuada a média.

Para acharmos a mediana após ordenar a nossa série, precisamos calcular a posição que ela ocupa e depois verificar qual valor está ocupando a posição. Como podemos ter amostras de tamanho par ou de tamanho ímpar, então temos dois cálculos para encontrar a posição no nossa mediana.

Posição da mediana para n é ímpar: $m_d = x_{n+1/2}$.

No caso de n ser ímpar, a mediana será igual ao valor da variável que se encontra na posição mediana.

Posição da mediana para n par: $m_d = \frac{x_{n/2} + x_{(n/2)+1}}{2}$.

No caso de n ser par, o valor da mediana será a média aritmética dos valores que ficam nas posições centrais.

Para obter a mediana para dados em rol, vamos considerar a pesquisa de satisfação com o atendimento de uma empresa de telefonia celular, sendo que a variável da nota de satisfação com a operadora (de 0 a 10) é igual a: 7, 6, 6, 8, 7, 9, 8, 8, 5, 6.

O primeiro passo para obter a mediana é ordenar os nossos dados: 5, 6, 6, 6, 7, 7, 8, 8, 8, 9.

Nesse caso, nosso n é par ($n = 10$), então utilizamos a seguinte fórmula:

Resolução:

$$m_d = \frac{x_{n/2} + x_{(n/2)+1}}{2} = \frac{x_{10/2} + x_{(10/2)+1}}{2} = \frac{x_5 + x_{5+1}}{2} = \frac{x_5 + x_6}{2} = \frac{7+7}{2} = 7$$

Ao interpretarmos esse resultado, temos que a metade das pessoas avaliou a satisfação com a operadora com nota 7 ou menos, e a outra metade com nota 7 ou mais.

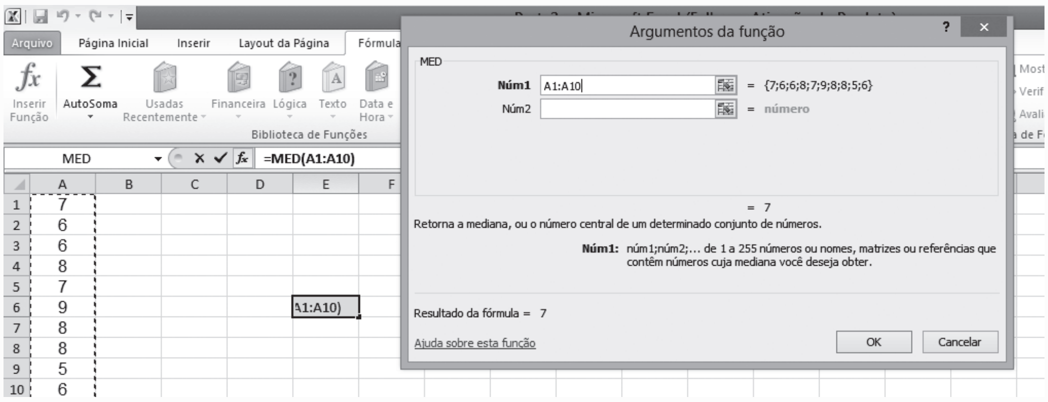
> Preste atenção, pois na mediana você pode utilizar os seguintes termos: inferior, superior, no mínimo, no máximo, pelo menos, mais de, menos de etc. Assim sendo, a maneira ERRADA de interpretar é dizer que metade das notas é igual a 7.

Exemplo 1.2: Para dados em tabela de distribuição de frequências por ponto, considerando nosso exemplo da quantidade de alunos por turma em uma escola de informática.

Para calcular a mediana no Excel, é necessário que se procure a aba fórmulas depois que for inserida a função, no canto esquerdo, bastando selecionar a categoria estatística. Assim, serão apresentadas várias funções estatísticas. Para isso, basta procurar pela Mediana que aparece a função Med.

Ao selecionarmos a função Med, basta selecionar os dados da amostra para obter a mediana.

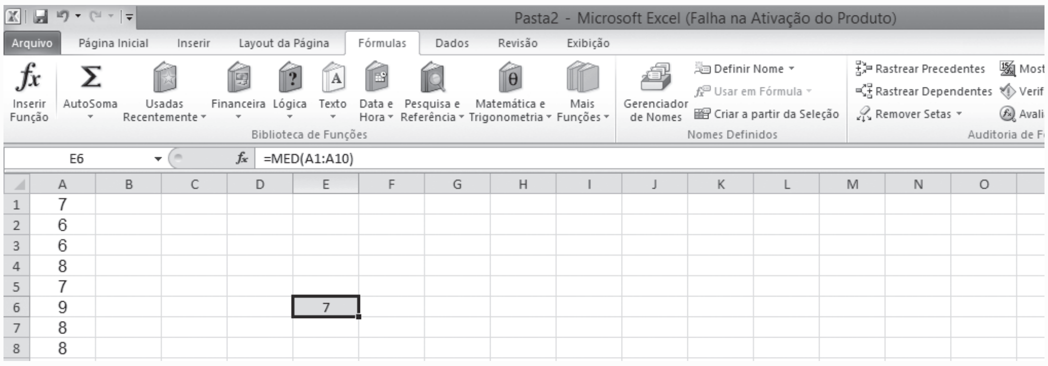
Figura 1.10: Selecionar dados.



Fonte: Elaborada pela autora (2016).

Clique em “OK” e obtenha a mediana para os dados.

Figura 1.11: Resultado da mediana.



Fonte: Elaborada pela autora (2016).

Tabela 1.5: Tabela de distribuição de frequências para o número de alunos por turma.

Número de alunos por turma	Frequência observada (f_i)	
9	14	14 (repito o primeiro f_i)
12	12	26 (14+12 = 26)
13	8	34 (26+8 = 34)
15	9	43 (34+9 = 43)
18	7	50 (43+7 = 50)
Total	50	-

Fonte: Elaborada pela autora (2016).

Quando temos uma tabela de distribuição de frequências por ponto, nossos dados já estão, consequentemente, ordenados. Então, precisamos calcular a nossa frequência acumulada para podermos localizar a posição da mediana.

Aqui, novamente, temos um número de elementos par, então calculamos a posição da mediana. Assim:

$$\textbf{Resolução:}\ m_d = \frac{x_{50/2} + x_{(50/2)+1}}{2} = \frac{x_{25} + x_{25+1}}{2} = \frac{x_{25} + x_{26}}{2} = \frac{12+12}{2} = 12$$

Ao interpretarmos esse resultado, temos que metade das turmas possuem doze alunos ou menos e a outra metade possui doze alunos ou mais.

Precisamos encontrar, então, qual valor ocupa as posições 25 e 26. Observemos a tabela na coluna F_i: na primeira linha temos do primeiro ao décimo quarto elemento; na linha 2 temos do décimo quinto ao vigésimo sexto elemento, e assim sucessivamente, mas para o que precisamos podemos parar na segunda linha, pois já encontramos as posições 25 e 26. Basta verificar quem é o valor correspondente a essa linha, em que encontramos o valor doze.

> Sempre que tivermos os dados em rol, primeiro precisamos ordenar os dados de forma crescente ou decrescente. Sempre que tivermos os dados em tabelas de distribuição de frequências por ponto, precisamos calcular a frequência acumulada para podermos achar quem ocupa a posição da mediana.

1.1.6.3 Moda (m_o)

É o valor da variável que possui maior frequência. Corresponde, simplesmente, ao valor ou aos valores que mais se repetem. Nem sempre é possível obter a moda, já que pode acontecer de nenhum valor se repetir mais de uma vez. Por outro lado, podem existir várias modas, cujos valores predominam sobre os demais.

Para obter a moda para dados em rol, vamos considerar a pesquisa de satisfação com o atendimento de uma empresa de telefonia celular e a variável da nota de satisfação com a operadora (de 0 a 10): 7, 6, 6, 8, 7, 9, 8, 8, 5, 6.

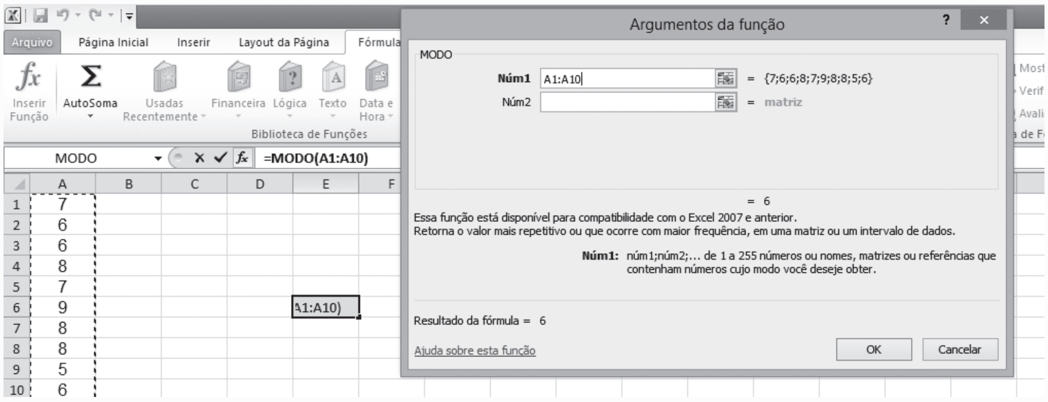
$m_o =$ 6 e 8 (valores com maior frequência)

Ao interpretarmos esse resultado, temos que a maior parte dos clientes atribuiu nota 6 ou 8 para a sua satisfação com a operadora.

Para calcular a moda no Excel, é necessário que se procure a aba ”Fórmulas”; depois, que se insira a função no canto esquerdo, bastando selecionar a categoria estatística. Assim, serão apresentadas várias funções estatísticas. Para isso, basta procurar por Modo, a fim de que surja a função Modo.

Ao selecionarmos a função Modo, basta selecionar os dados da amostra e obter a moda.

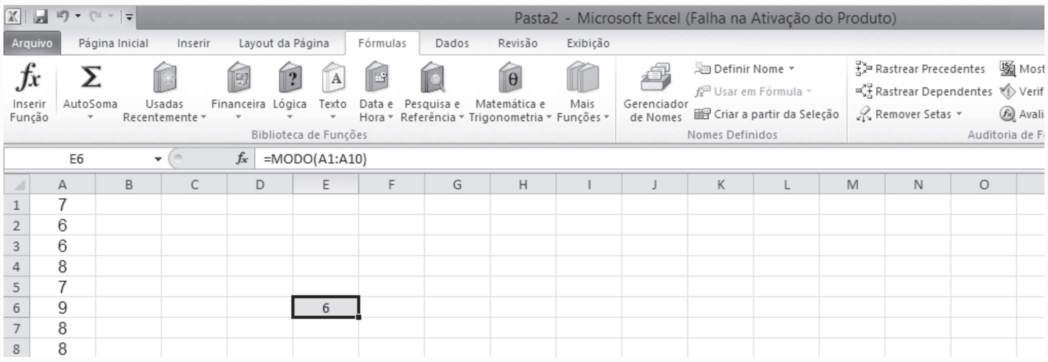
Figura 1.12: Selecionar dados.



Fonte: Elaborada pela autora (2016).

Clicar em "OK" e obter a moda para os dados.

Figura 1.13: Resultado da moda.



Fonte: Elaborada pela autora (2016).

É importante ressaltar que o Excel apresenta apenas a primeira moda da série dos dados, assim se a série tiver mais de uma moda, o Excel apresentará somente uma delas.

Para dados em tabela de distribuição de frequências por ponto, temos como nosso exemplo a quantidade de alunos por turma em uma escola de informática.

Tabela 1.6: Tabela de distribuição de frequências para o número de alunos por turma.

Número de alunos por turma	Frequência observada (f_i)
9	14
12	12
13	8
15	9
18	7
Total	50

Fonte: Elaborada pela autora (2016).

Podemos perceber que a maior frequência observada é a de 14, então, a moda será nove, ou $m_o = 9$.

A moda é simplesmente o valor, ou os valores, da variável que mais se repete.

1.1.7 Medidas de Variabilidade

Sempre que analisamos as medidas de posição, é interessante analisarmos em conjunto as medidas de variabilidade, até para saber qual a medida de posição melhor representa os nossos dados. Pensem comigo: vamos comparar dois jogadores de futebol, imagine que ambos tenham uma média de dois gols por partida, então são jogadores bons não é mesmo? Sim, mas eles têm o mesmo desempenho? Não necessariamente, pois pode ser que um dos jogadores tenha essa média porque faça muitos gols em uma partida e fique outras tantas sem marcar; já o outro jogador marca constantemente pelo menos um gol por partida. Então, o nosso segundo jogador é mais interessante e mais eficiente, ou seja, ele é mais constante, sendo a variabilidade de gols por partida menor. Então, esse jogador é melhor, apesar de ter a mesma média que o outro jogador.

Sempre analisamos as medidas de posição em conjunto com as de variabilidade. As medidas de variabilidade estão associadas às distâncias entre os valores dentro de um conjunto de dados.

1.1.7.1 Amplitude

A amplitude é a medida de variabilidade com menor grau de percepção da variabilidade de um conjunto, pois ela é calculada com distância entre o maior e o menor valor do conjunto. Dessa forma, não consideram os valores intermediários do conjunto, porém, muitas vezes, dá uma ideia da variabilidade geral da série de dados.

$$h = \text{máximo}(x) - \text{mínimo}(x)$$

Para obter a amplitude para dados em rol, vamos considerar a pesquisa de satisfação com o atendimento de uma empresa de telefonia celular e a variável da nota de satisfação com a operadora (de 0 a 10): 7, 6, 6, 8, 7, 9, 8, 8, 5, 6.

Resolução: $h = máximo(x) - mínimo(x) = 9 - 5 = 4$.

Como exemplo para dados em tabela de distribuição de frequências por ponto, temos a quantidade de alunos por turma em uma escola de informática.

Tabela 1.7: Tabela de distribuição de frequências para o número de alunos por turma.

Número de alunos por turma	Frequência observada (f_i)
9	14
12	12
13	8
15	9
18	7
Total	50

Fonte: Elaborada pela autora (2016).

Resolução: $h = máximo(x) - mínimo(x) = 18 - 9 = 9$.

1.1.7.2 Variância σ_x^2 ou s_x^2

A variância é a medida de variabilidade obtida através da média dos quadrados das distâncias de cada ponto com a média do conjunto. Com essa frase, parece complicado, mas, na verdade, precisamos pegar cada um dos valores e subtraí-los pela média e elevar o resultado ao quadrado. Depois, soma-se tudo e se divide pelo número de elementos. Como estamos elevando as diferenças de cada um dos valores em relação à média ao quadrado, não temos variância negativa. Como essa medida avalia as distâncias entre todos os pontos e a média, ela dá uma noção maior da variabilidade dos dados do que a amplitude. Porém, se elevamos essas diferenças ao quadrado, a nossa unidade também ficaria elevada ao quadrado, então a variância não nos permite uma comparação direta com a média.

Vocês observarão que, para dados populacionais, a variância é dividida por N, e para dados amostrais, dividida por n-1. Isso se deve a uma correção por estarmos lidando com dados amostrais.

Variância populacional: σ_x^2

$$\sigma_x^2 = \frac{\sum (x_i - \mu_x)^2}{N}$$

Variância amostral: s_x^2

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Outra forma de se obter a variância é expressa através das fórmulas a seguir:

Variância populacional: σ_x^2

$$\sigma_x^2 = \frac{\sum x_i^2 - N\mu_x^2}{N}$$

Variância amostral: s_x^2

$$s_x^2 = \frac{\sum x_i^2 - n\overline{x}^2}{n\text{-}1}$$

Ambas as fórmulas acima levam ao mesmo resultado.

Para obter a variância para dados em rol, vamos considerar a pesquisa de satisfação com o atendimento de uma empresa de telefonia celular e a variável da nota de satisfação com a operadora (de 0 a 10): 6, 6, 8, 7, 9, 8, 8, 5, 6.

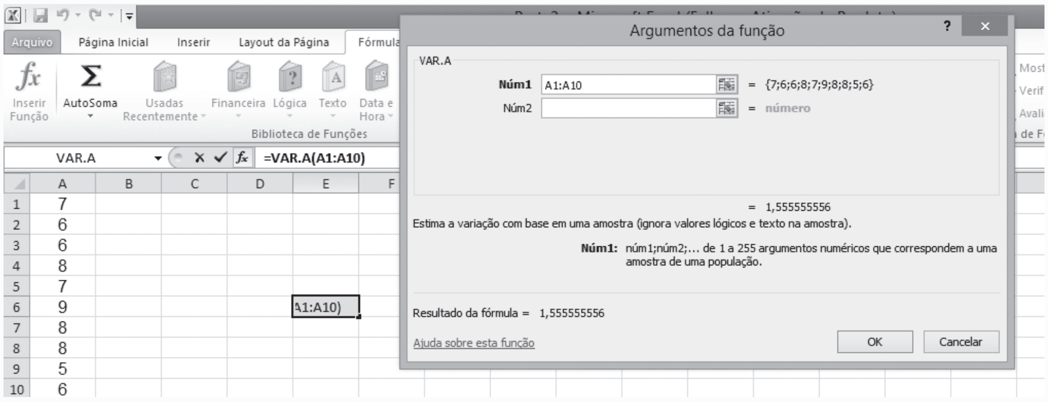
Resolução:

$$s_x^2 = \frac{\sum\left(x_i - \overline{x}\right)^2}{n\text{-}1} = \frac{(7-7)^2 + (6-7)^2 + ...+(5-7)^2 + (6-7)^2}{10-1} = \frac{14}{9} = 1{,}56$$

Para calcular a variância no Excel, é necessário que se procure a aba “Fórmulas”; depois, que se insira a função no canto esquerdo, bastando selecionar a categoria estatística. Assim, serão apresentadas várias funções estatísticas. Para isso, basta procurar pela Variância, a fim de que surja a função VAR.A. Para a variância, temos duas possibilidades: a variância da amostra e a variância da população. No nosso caso, temos a da amostra.

Ao selecionarmos a função VAR.A., basta selecionar os dados da amostra e obter a variância.

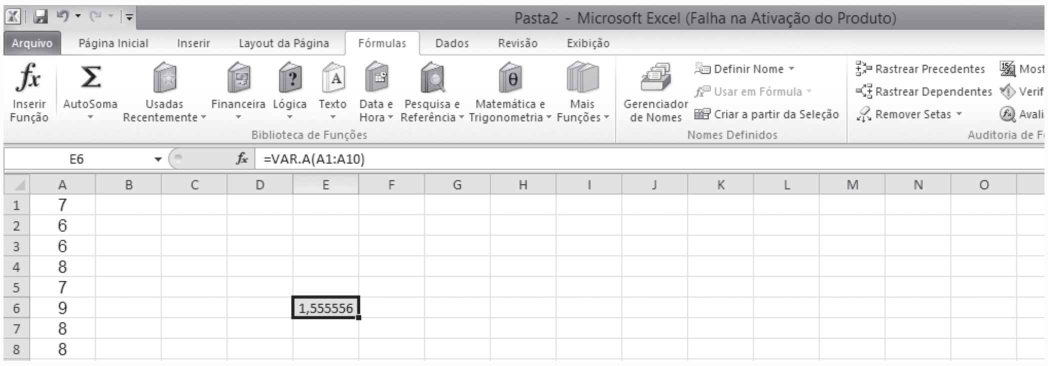
Figura 1.14: Selecionar dados.



Fonte: Elaborada pela autora (2016).

Clique em “OK” e obtenha a variância para os dados.

Figura 1.15: Resultado da variância.



Fonte: Elaborada pela autora (2016).

Como exemplo para dados em tabela de distribuição de frequências por ponto, temos a quantidade de alunos por turma em uma escola de informática.

Tabela 1.8: Tabela de distribuição de frequências para o número de alunos por turma.

Número de alunos por turma	Frequência observada (f_i)	$(x_i - \bar{x})^2 f_i$
9	14	$(9 - 12{,}72)^2.14 = 191{,}66$
12	12	$(12 - 12{,}72)^2.12 = 5{,}88$
13	8	$(13 - 12{,}72)^2.8 = 0{,}72$
15	9	$(15 - 12{,}72)^2.9 = 47{,}61$
18	7	$(18 - 12{,}72)^2.7 = 196{,}63$
Total	50	442,5

Fonte: Elaborada pela autora (2016).

Resolução:

$$s_x^2 = \frac{\sum x_i - \bar{x}\;^{2}.f_i}{n\text{-}1} = \frac{(9-12,7)^2.14 + (12-12,7)^2.12 + ... + (18-12,7)^2.18}{50-1} =$$

$$\frac{191,66+5,88+0,72+47,61+196,63}{49} = \frac{442,5}{49} = 9,03$$

Assim como fizemos na média, aqui também multiplicamos cada valor pela quantidade de vezes em que ele aparece (f_i).

Variância populacional: σ_x^2

$$\sigma_x^2 = \frac{\sum (x_i - \mu_x)^2 f_i}{N}$$

Variância amostral: s_x^2

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{n\text{-}1}$$

Outra forma de se obter a variância é expressa através das fórmulas a seguir:

Variância populacional: σ_x^2

$$\sigma_x^2 = \frac{\sum x_i^2 f_i - N\mu_x^2}{N}$$

Variância amostral: s_x^2

$$s_x^2 = \frac{\sum x_i^2 f_i - n\bar{x}^2}{n\text{-}1}$$

1.1.7.3 Desvio-Padrão σ_x ou s_x

O desvio-padrão é a raiz quadrada da variância. Lembrem-se quando falamos que a unidade de medida da variância estaria diferente da unidade da média? Pois, então, calcularemos a raiz quadrada da variância e ela passará a se chamar "desvio-padrão". Assim, poderemos comparar a variabilidade em torno da média.

Desvio-padrão populacional: σ_x

$$\sigma_x = \sqrt{\frac{\sum (x_i - \mu_x)^2}{N}}$$

Desvio-padrão amostral: s_x^2

$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n\text{-}1}}$$

Outra forma de se obter o desvio-padrão é expressa através das fórmulas a seguir:

Desvio-padrão populacional: σ_x^2

$$\sigma_x = \sqrt{\frac{\sum x_i^2 - N\mu_x^2}{N}}$$

Desvio-padrão amostral: s_x^2

$$s_x = \sqrt{\frac{\sum x_i^2 - n\bar{x}^2}{n\text{-}1}}$$

Para obter o desvio-padrão para dados em rol, vamos considerar a pesquisa de satisfação com o atendimento de uma empresa de telefonia celular e a variável da nota de satisfação com a operadora (de 0 a 10): 7, 6, 6, 8, 7, 9, 8, 8, 5, 6.

Resolução:

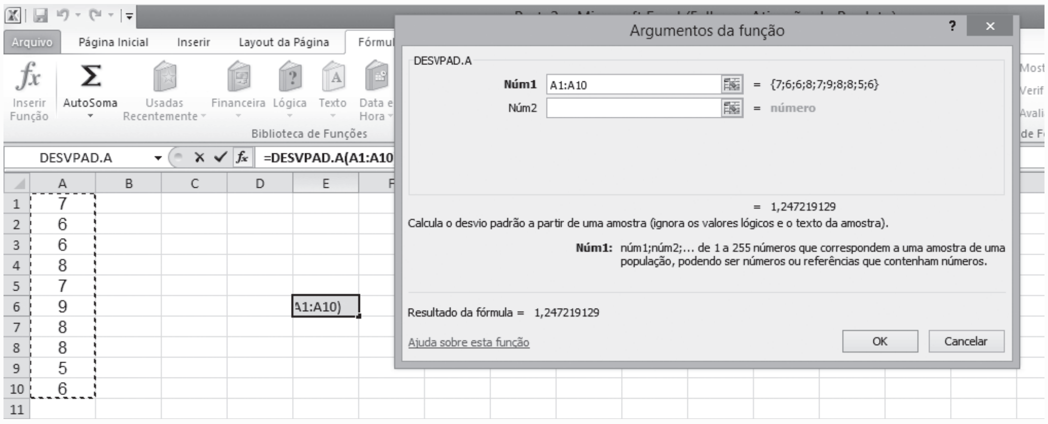
$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n\text{-}1}} = \sqrt{\frac{(7-7)^2 + (6-7)^2 + ... + (5-7)^2 + (6-7)^2}{10-1}} =$$
$$\sqrt{\frac{24}{9}} = \sqrt{1,56} = 1,25$$

Ao interpretarmos esse resultado, verificamos que existe uma variação de 1,25 pontos em torno da média de satisfação.

Para calcular o desvio-padrão no Excel, é necessário que se procure a aba "Fórmulas"; depois, que se insira a função no canto esquerdo, bastando selecionar a categoria estatística. Assim, serão apresentadas várias funções estatísticas. Para isso, basta procurar por Desvio-padrão e escolher a função DESVPAD.A. Para o desvio-padrão também temos mais de uma possibilidade, no entanto, no nosso caso, precisamos do desvio-padrão para a amostra.

Ao selecionarmos a função DESVPAD.A, basta selecionar os dados da amostra e obter o desvio-padrão.

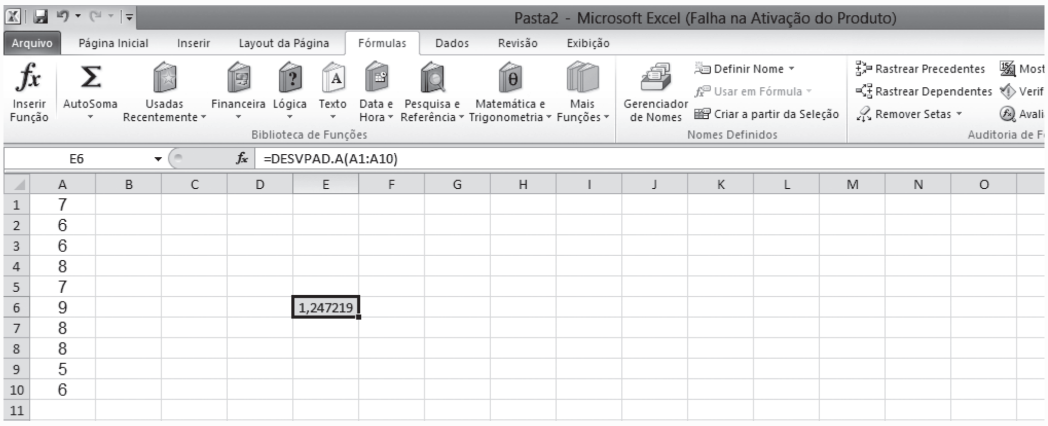
Figura 1.16: Selecionar dados.



Fonte: Elaborada pela autora (2016).

Clique em em “OK” e obtenha o desvio-padrão para os dados.

Figura 1.17: Resultado do desvio-padrão.



Fonte: Elaborada pela autora (2016).

Como exemplo para dados em tabela de distribuição de frequências por ponto, temos a quantidade de alunos por turma em uma escola de informática.

Tabela 1.9: Tabela de distribuição de frequências para o número de alunos por turma.

Número de alunos por turma	Frequência observada (f_i)	$(x_i - \bar{x})^2 . f_i$
9	14	$(9 - 12{,}72)^2.14 = 191{,}66$
12	12	$(12 - 12{,}72)^2.12 = 5{,}88$
13	8	$(13 - 12{,}72)^2.8 = 0{,}72$
15	9	$(15 - 12{,}72)^2.9 = 47{,}61$
18	7	$(18 - 12{,}72)^2.7 = 196{,}63$
Total	50	442,5

Fonte: Elaborada pela autora (2016).

Resolução:

$$s_x = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{(9-12,7)^2.14 + (12-12,7)^2.12 + ...+(18-12,7)^2.12}{50-1}} =$$

$$= \sqrt{\frac{191,55+5,88+0,72+47,61+196,63}{49}} = \sqrt{\frac{442,5}{49}} = \sqrt{9,03} = 3,0$$

Ao interpretarmos esse resultado, verificamos que existe uma variação de três alunos em torno da média de alunos por turma.

> Perceba que se a variância eleva as diferenças ao quadrado – não possibilitando, assim, valores negativos –, então não existe variância negativa e nem desvio-padrão negativo.

1.1.7.4 Coeficiente de Variação CV

O coeficiente de variação é uma variabilidade relativa (percentual) e, por esse motivo, independe das unidades de medida das variáveis, sendo muito útil quando precisamos determinar qual conjunto de dados é mais homogêneo ou heterogêneo e as variáveis que têm unidades de medidas diferentes ou médias muito diferentes.

Dessa forma, quanto menor for o coeficiente de variação, mais homogêneos serão os dados estudados. Na comparação de duas amostras, a amostra mais homogênea será a que tiver o menor CV como resultado.

Fórmula: $CV_x = \frac{s_x}{\overline{x}}.100$

Para obter o coeficiente de variação para dados em rol, vamos considerar a pesquisa de satisfação com o atendimento de uma empresa de telefonia celular, a variável nota de satisfação com a operadora (de 0 a 10): 7, 6, 6, 8, 7, 9, 8, 8, 5, 6.

Resolução:

$$CV_x = \frac{s_x}{\overline{x}}.100 = \frac{1,25}{7}.100 = 17,86\%$$

Ao interpretarmos esse resultado, verificamos que existe uma variação de 17,86% em torno da nota média de satisfação com a operadora.

Como exemplo para dados em tabela de distribuição de frequências por ponto, temos a quantidade de alunos por turma em uma escola de informática.

Tabela 1.10: Tabela de distribuição de frequências para o número de alunos por turma.

Número de alunos por turma	Frequência observada (f_i)
9	14
12	12
13	8
15	9
18	7
Total	50

Fonte: Elaborada pela autora (2016).

Resolução:

$$CV_x = \frac{s_x}{\bar{x}}.100 = \frac{3}{12,7}.100 = 23,62\%$$

Ao interpretarmos esse resultado, verificamos que existe uma variação de 23,62% em torno da nota média de alunos por turma na escola de informática.

Quanto menor for o **Coeficiente de Variação (CV)**, mais homogêneos serão os dados da série.

Vimos, ao longo desta unidade, as maneiras de apresentar os dados. Aprendemos, também, as estatísticas descritivas para os dados quantitativos.

Na próxima, aprenderemos sobre probabilidade. Até lá!

SÍNTESE DA UNIDADE

Nesta unidade, vimos os conceitos básicos de estatística que estarão sempre presentes nas análises de dados, como:

- a população que representa o total de informações;
- a amostra de uma parte da minha população;
- os tipos de variáveis que temos: as variáveis qualitativas, que podem ser nominais ou ordinais, e as variáveis quantitativas, que podem ser discretas ou contínuas;
- algumas representações gráficas;
- medidas de posição que são a média, a mediana e a moda;
- medidas de variabilidade que são a amplitude, a variância, o desvio-padrão e o coeficiente de variação.

REFERÊNCIAS

STEVENSON, W. J. **Estatística Aplicada à Administração**. São Paulo: Harbra, 2001.

ANOTAÇÕES

Probabilidade

Prezado(a) estudante.

Estamos começando uma unidade desta disciplina. Os textos que a compõem foram organizados com cuidado e atenção, para que você tenha contato com um conteúdo completo e atualizado tanto quanto possível. Leia com dedicação, realize as atividades e tire suas dúvidas com os tutores. Dessa forma você, com certeza, alcançará os objetivos propostos para essa disciplina.

OBJETIVO GERAL

Introduzir os conceitos de probabilidade.

OBJETIVOS ESPECÍFICOS

- Mostrar e exemplificar os conceitos iniciais da teoria de probabilidade.
- Aplicar conceitos básicos de probabilidade.
- Apresentar algumas aplicações para as distribuições de probabilidade discretas.

2.1 Introdução à Probabilidade

Quem nunca utilizou a palavra “probabilidade” quando queria informar qual a chance de alguma coisa acontecer? Acho que todos nós já passamos por isso. Vejamos algumas frases do tipo:

- “A chance de eu ir ao cinema hoje é de 99%”.
- “A probabilidade de chover hoje, segundo a previsão do tempo, é de 87%”.
- “A probabilidade de que eu saia na véspera do feriado é muito pequena”.
- “A probabilidade de ganhar na Mega Sena é quase zero, mas mesmo assim eu jogo”.

Todas essas frases têm algo em comum, pois todas demonstram uma incerteza, ou seja, não podemos dizer exatamente o que vai acontecer.

Aqui entra, então, a teoria das probabilidades. Ela nos permite calcular essa incerteza.

Para o estudo das probabilidades, vamos introduzir alguns conceitos básicos.

2.1.1 Experimento

Experimento é um fenômeno que poder ser repetido sob as mesmas condições várias vezes.

Esse experimento pode ser aleatório, ou seja, não podemos, de antemão, descobrir qual resultado efetivamente acontecerá em uma repetição em particular. Nesse sentido, serão os experimentos aleatórios que iremos tratar nesta unidade.

O experimento também pode ser determinístico, ou seja, no experimento determinístico independe o número de vezes que ele seja repetido – ele sempre levará ao mesmo resultado.

Vejamos, a seguir, outro conceito importante.

2.1.2 Espaço Amostral (s ou Ω)

O espaço amostral é o conjunto de todos os resultados possíveis de um experimento aleatório.

Exemplo 2.1: Cara ou Coroa?

Experimento 1: Lançar uma moeda e observar a face para cima:

$$\Omega_1 = \{\text{cara, coroa}\}$$
$$\text{Cara} = K$$
$$\text{Coroa} = C$$

Experimento 2: Lançar uma moeda duas vezes e observar a sequência de faces:

$$\Omega_2 = \{KK, CK, KC, CC\}$$

Experimento 3: Lançar um dado e observar a face voltada para cima

$$\Omega_3 = \{1, 2, 3, 4, 5, 6\}$$

Experimento 4: Jogar uma moeda quatro vezes e observar o número de caras obtido.

$$\Omega_4 = \{0, 1, 2, 3, 4\}$$

Experimento 5: Jogar uma moeda quatro vezes e observar a sequência de caras e coroas obtidas.

$$\Omega_5 = \left\{\begin{array}{l} CCCC, CCCK, CCKC, CKCC, KCCC, CCKK, KKCC, CKCK, KCCK, \\ CKKC, CKKK, KCKK, KCKC, KKCK, KKKC, KKKK \end{array}\right\}$$

Experimento 6: Uma lâmpada nova é ligada e observa-se o tempo gasto até queimar.

$$\Omega_6 = \{t \in \Re \,/\, t \geq 0\}$$

Experimento 7: Lança-se uma moeda até que ocorra uma cara e conta-se, então, o número de lançamentos necessários.

$$\Omega_7 = \{1, 2, 3, 4, 5, ...\}$$

Experimento 8: Lançam-se dois dados e anota-se o total de pontos obtidos.

$$\Omega_8 = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

Experimento 9: Lançam-se dois dados e anota-se o par obtido.

$$\Omega_9 = \left\{\begin{array}{l} (1,1), (1,2), (1,3), (1,4), (1,5), (1,6) \\ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6) \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6) \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6) \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6) \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \end{array}\right\}$$

Como podemos observar nos exemplos, ao descrever um espaço amostral de um experimento deve-se ficar atento para o que se está mensurando. Devemos verificar “um” espaço amostral associado a um experimento e não “o” espaço amostral. Outro ponto importante a observarmos é que nem sempre o espaço amostral é composto por números.

Ao detalharmos um pouco mais os espaços amostrais dos exemplos, temos os espaços amostrais $\Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5, \Omega_8$ e Ω_9, que são considerados como finitos, pois sabemos exatamente quantas são as possibilidades do nosso experimento, ou seja, sabemos o tamanho do espaço amostral.

Os espaços amostrais Ω_6 e Ω_7 são infinitos; sabemos o que pode acontecer, mas não sabemos qual o tamanho exato do espaço amostral. Sendo que não conseguimos contar Ω_6, já que o tempo pode assumir qualquer valor na reta dos reais e não sabemos até quando podemos contar Ω_7 – mas poderíamos calcular para saber qual o tamanho do espaço da amostra, já que a quantidade de lançamentos necessários só pode assumir números inteiros.

2.1.3 Métodos de Enumeração

Muitas vezes é difícil pensar em todas as combinações possíveis de um experimento. Para isso, pedimos ajuda a algumas teorias da Matemática. Nesse momento, apelamos para conceitos básicos estudados em Análise Combinatória.

I. **Regra da multiplicação**

Suponha que se possa fazer n escolhas independentes com:

m_1 maneiras de fazer a escolha **1**
m_2 maneiras de fazer a escolha **2**
....,
m_n maneiras de fazer a escolha **n**.

Então, existem $m_1.m_2.(...).m_n$ maneiras diferentes de fazer a sequência inteira de escolhas.

Exemplo 2.2: Se em uma eleição há cinco candidatos a prefeito e quinze a vereador, os dois cargos podem ser preenchidos de quantos modos?

Resolução:

$5.\,15 = 75$ modos

II. **Permutação (fatorial)**

Em uma livraria, temos quatro novos títulos e queremos colocá-los lado a lado em uma vitrine. Quantas maneiras diferentes existem para colocar os cinco livros?

Para o primeiro espaço, existem quatro escolhas possíveis, uma para cada livro. Uma vez colocado o primeiro livro, restam três escolhas para o segundo espaço, e assim por diante. Então o número de escolhas diferentes é: $4 \cdot 3 \cdot 2 \cdot 1 = 24$. Esse tipo especial de multiplicação tem um símbolo próprio: 4!. De um modo geral, se dispomos de um número *n*, então o produto anterior é representado por $n!$ e é lido "ene fatorial", isto é:

$$P_n = n! = n \cdot (n-1) \cdot (n-2)...$$

Temos, também, que:

0! = 1

III. **Permutações com itens duplicados**

Permutações também podem ser realizadas com itens duplicados. Por exemplo, de quantas maneiras diferentes se pode arranjar a palavra voo sem considerarmos o acento circunflexo? Agora, estamos lidando com itens duplicados e não com repetições. Uma vez que dois "os" podem ser arranjados em 2! de diferentes maneiras, o número de permutações diferentes (ou distinguíveis) é:

$$\frac{3!}{2!} = 3 \text{ (voo, ovo, oov)}$$

Ou seja, $P_n = \dfrac{n!}{(n_1!) \cdot (n_2!) ... (n_k!)}$

Exemplo 2.3: Quantos arranjos distintos podem ser feitos com as letras da palavra "estatística"?

Resolução:

Nesse caso, temos um total de 11 letras, das quais $n_1 = 2$ (o "*s*" ocorre duas vezes), $n_2 = 3$ (o "*t*" ocorre três vezes), $n_3 = 2$ (o "*a*" ocorre duas vezes) e $n_4 = 2$ (a letra "*i*" ocorre duas vezes).

Então, existem: $\dfrac{11!}{2! \cdot 3! \cdot 2! \cdot 2!} = 831.600$ arranjos distintos de letras da palavra "estatística".

IV. **Arranjos**

Um arranjo consiste no número de possíveis maneiras de agrupar ou arranjar certos conjuntos de objetos.

O número de arranjos de n objetos distintos – tomados em grupos de p, em que p é menor que n – é representado por:

$$A_{n,p} = A_p^n = \frac{n!}{(n-p)!}$$

Exemplo 2.4: Quantos números de dois algarismos podemos formar com os algarismos 2, 3, 7 e 8?

Resolução:

$$A_{4,2} = \frac{4!}{(4-2)!} = 12$$

V. **Permutações (arranjos) com repetição**

Vamos considerar n elementos tomados como p a p, que podem se repetir, isto é, o mesmo elemento pode ocorrer mais de uma vez. Então o número de arranjos, não necessariamente distintos, é dado por: n^p, isto é:

$$A_{(n,p)} = n^p$$

Exemplo 2.5: Uma urna contém bolas vermelhas, brancas e pretas. Uma bola é extraída e, em seguida, anotada a sua cor, volta para a urna. Então uma segunda bola é extraída e é anotada igualmente a cor. Quantas são as possíveis sequências de cores observadas?

Resolução:

Observe que cada extração fornece uma cor entre $\{V, B, P\}$, sendo que o número de sequências possíveis é, pelo princípio fundamental da contagem:

$$3 \cdot 3 = 3^2 = 9$$

VI. **Combinações**

Existem certos arranjos em que a ordem entre os elementos não é importante. Por exemplo, para calcular a probabilidade de acertar os números sorteados na Mega-Sena ou na Quina, não é necessário saber a ordem em que foram sorteados, uma vez que o que importa é a combinação deles.

Então, arranjos em que a ordem não interessa são denominadas de combinações.

$$\binom{n}{p} = C_{n,p} = C_n^p = \frac{n!}{(n-p)!\cdot p!}$$

Exemplo 2.6: Quantos números de dois algarismos podemos formar com os algarismos de zero a nove (0, 1, 2, 3, 4, 5, 6, 7, 8, 9), não importando a ordem?

Resolução:

$$\binom{9}{2} = \frac{9!}{(9-2)!\cdot 2!} = 36$$

Agora que sabemos como calcular o tamanho do espaço amostral, vamos aos próximos conceitos.

2.1.4 Eventos

Evento é qualquer subconjunto de um espaço amostral Ω. Assim, temos:

- O espaço amostral - Ω é o evento certo;
- $\varnothing$ é o evento impossível.

Exemplo 2.7: Seja o experimento, jogar um dado e observar a face voltada para cima.
Temos o espaço amostral: $\Omega = \{1,2,3,4,5,6\}$

Sejam os seguintes eventos:

Evento 1: cair a face com número par $\rightarrow$ $A = \{2,4,6\}$.

Evento 2: o valor da face ser maior que 4 $\rightarrow B = \{5,6\}$.

Evento 3: cair a face com número ímpar $\rightarrow$ $C = \{1,3,5\}$.

Evento 4: cair a face com número menor que 6 $\rightarrow$ $D = \{1,2,3,4,5\}$.

Evento 5: cair uma face com número maior que 10 $\rightarrow E = \{\varnothing\}$ $\Rightarrow$ evento impossível.

Evento 6: cair uma face com número maior ou igual a 1 $\rightarrow$ $F = \{1,2,3,4,5,6\}$ $\Rightarrow$ evento certo.

Sabendo dessas definições, podemos começar a exercitar algumas operações com os eventos.

2.1.5 Operações com Eventos

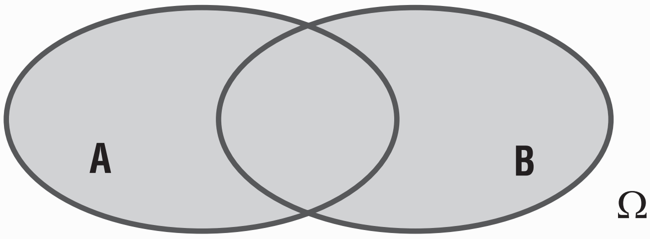
Podemos realizar operações com eventos da mesma forma com que realizamos operações entre conjuntos.

Considere E um experimento com um espaço amostral associado Ω, e estabeleceremos A e B como sendo eventos quaisquer do nosso espaço amostral Ω.

Então, sejam A e B dois eventos de um mesmo espaço amostral Ω. Nesse sentido, podemos ter as seguintes operações com eventos:

- **União**
  Evento A em união com o evento B; ou também podemos dizer: evento A soma evento B, anotado por $A \cup B$, ou seja, A ocorre ou B ocorre.

Figura 2.1: União.

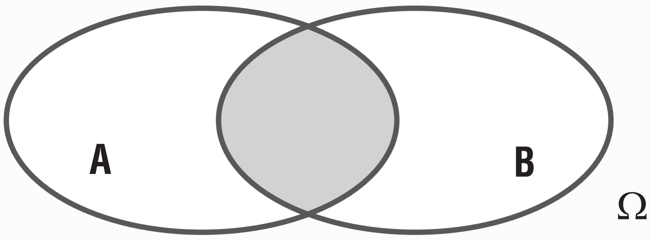


Fonte: Elaborada pela autora (2016).

- **Interseção**

  Evento A em interseção com B, representado por $A \cap B$ ou AB, ou seja, A ocorre e B ocorre, simultaneamente.

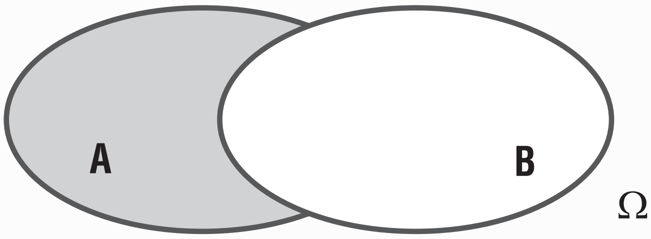
Figura 2.2: Interseção.



Fonte: Elaborada pela autora (2016).

- **Um evento menos outro evento**
  Evento A menos evento B ou, ainda, A diferença B; anota-se $A - B$, ou seja, se A ocorre, B não ocorre.

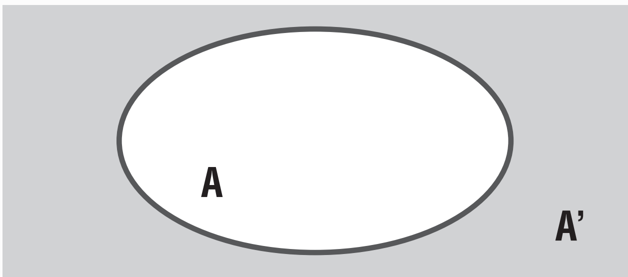
Figura 2.3: Evento A menos evento B .



Fonte: Elaborada pela autora (2016).

- **Complementação**
  A complementação do evento A, anotado por $\overline{A}, A^C$ ou, ainda, A', indica que A não ocorre.

Figura 2.4 :Complementação.



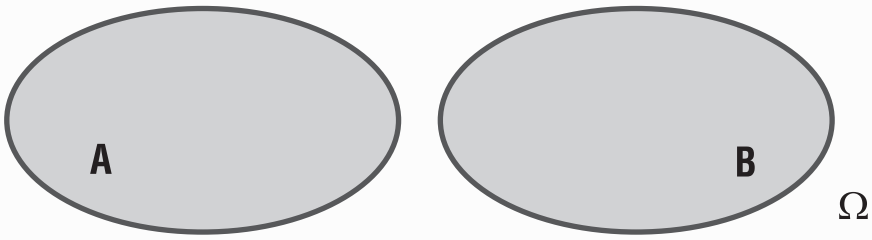
Fonte: Elaborada pela autora (2016).

Podemos ter mais algumas definições sobre eventos.

2.1.6 Eventos Mutuamente Excludentes

Para que dois eventos A e B sejam denominados mutuamente exclusivos ou excludentes, eles não podem ocorrer simultaneamente, isto é, se $A B \cap = \varnothing$, ou seja, não existe interseção entre os eventos A e B .

Figura 2.5: Eventos mutuamente exclusivos.



Fonte: Elaborada pela autora (2016).

Exemplo 2.8: Seja o experimento, jogar um dado e observar a face voltada para cima.

Temos o espaço amostral: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Sejam os seguintes eventos:

Evento 1: cair a face com número par $\rightarrow A = \{2, 4, 6\}$.

Evento 2: o valor da face ser maior que 4 $\rightarrow B = \{5, 6\}$.

Evento 3: cair a face com número ímpar $\rightarrow C = \{1, 3, 5\}$.

Evento 4: cair a face com número menor que 6 $\rightarrow D = \{1, 2, 3, 4, 5\}$.

Evento 5: cair uma face com número maior que 10 $\rightarrow E = \{\varnothing\}$.

Evento 6: cair uma face com número maior ou igual a 1 $\rightarrow \Omega = \{1, 2, 3, 4, 5, 6\}$.

Determine:

$A \cup B = \{2, 4, 5, 6\}$ $\quad$ $A \cup C = \{1, 2, 3, 4, 5, 6\}$

$A \cap B = \{6\}$

$D \cap F = \{1, 2, 3, 4, 5\}$

$\overline{A} = \{1, 3, 5\}$

$\overline{B} = \{1, 2, 3, 4\}$

$\overline{A \cup B} = \{1, 3\}$

$\overline{A \cap B} = \{1,\ 2,\ 3,\ 4,\ 5\}$

$\overline{A} \cup \overline{D} = \{1, 3, 5, 6\}$

$A - B = \{2, 4\}$

$A \cap \overline{A} = \{\varnothing\}$

Agora que temos as premissas anteriores, podemos começar a calcular as probabilidades.

2.1.7 Conceitos de Probabilidade

Definição **clássica de probabilidade**, seja E um experimento aleatório e Ω um espaço amostral associado formado por n resultados igualmente prováveis. A probabilidade de A, anotada por $P(A)$, (lê-se pe de A), é definida como sendo:

$$P(A) = \frac{a}{n}$$

Em que:

a é o número de casos favoráveis ao evento A.

n é o número de casos possíveis do experimento.

Exemplo 2.9: Seja o experimento, jogar um dado, e observar a face voltada para cima.

Temos o espaço amostral: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Sejam os seguintes eventos:

Evento 1: cair a face com número par $\rightarrow$ $A = \{2, 4, 6\}$.

Evento 2: o valor da face ser maior que 4 $\rightarrow$ $B = \{5, 6\}$.

Evento 3: cair a face com número ímpar $\rightarrow$ $C = \{1, 3, 5\}$.

Evento 4: cair a face com número menor que 6 $\rightarrow$ $D = \{1, 2, 3, 4, 5\}$.

Evento 5: cair uma face com número maior que 10 $\rightarrow$ $E = \{\varnothing\}$.

Evento 6: cair uma face com número maior ou igual a 1 $\rightarrow$ $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Para os eventos do nosso exemplo, temos:

$$P(A) = \frac{a}{n} = \frac{\textit{casos favoráveis ao evento face par}}{\textit{casos possíveis do experimento}} = \frac{3}{6} = 0,5 = 50\%$$

$$P(B) = \frac{a}{n} = \frac{2}{6} = 0,3333 = 33,33\%$$

$$P(C) = \frac{a}{n} = \frac{3}{6} = 0,5 = 50\%$$

$$P(D) = \frac{a}{n} = \frac{5}{6} = 0,8333 = 83,33\%$$

$$P(E) = \frac{a}{n} = \frac{0}{6} = 0,0 = 0,0\%$$

$$P(F) = \frac{a}{n} = \frac{6}{6} = 1,000 = 100\%$$

Observe que o número de casos possíveis é o tamanho do espaço amostral.

A definição não pode ser aplicada quando o espaço amostral é infinito.

Na definição frequencial de probabilidade, temos que, na prática, nem sempre é possível determinar a probabilidade de um evento. Assim, precisamos de um método de aproximação dessa probabilidade. Um dos métodos utilizados é a experimentação. A probabilidade determinada através desse processo é chamada de probabilidade empírica.

Seja E um experimento e A um evento de um espaço amostral Ω associado ao experimento E. Suponha que esse experimento E seja repetido n vezes e seja m o número de vezes em que A ocorre nas n repetições do experimento E. Então, a frequência relativa do evento A, anotada por f_{rA} é o quociente entre o número de vezes que o evento A ocorreu pelo número de vezes em que o evento E é repetido:

$$f_{rA} = \frac{m}{n}$$

Em que:

m é o número de vezes em que A ocorre.

n é o número de vezes em que o experimento E é repetido.

Exemplo 2.10: Seja o experimento lançar uma moeda e observar a face voltada para cima, a moeda foi lançada 500 vezes, e, desses lançamentos, 272 vezes apareceu a face cara. Então, a frequência relativa de "caras" é:

Resolução:

$$f_{rA} = \frac{m}{n} = \frac{\textit{número de vezes em que o evento A ocorreu}}{\textit{número de vezes em que o experimento foi repetido}} = \frac{272}{500}$$

$$f_{rA} = 0,5440 = 54,40\%$$

Seja o experimento lançar um dado e observar a face voltada para cima, o dado foi lançado 200 vezes e a face 5 apareceu 35 vezes. Então, a frequência relativa do evento A = { face 5 } é:

Resolução:

$$f_{rA} = \frac{m}{n} = \frac{35}{200} = 0,1750 = 17,50\%$$

Agora, com esses conceitos iniciais de probabilidade, podemos verificar as propriedades das probabilidades.

AMPLIANDO CONHECIMENTO

Você pode ver exemplos ilustrados de espaços amostrais no seguinte endereço: https://goo.gl/u1F2pQ. Acesso em: 9 jul. 2017.

2.1.8 Propriedades das Probabilidades

Segundo Stevenson (2001), as probabilidades são utilizadas para exprimir a chance de ocorrência de determinado evento.

Vejamos, a seguir, algumas propriedades das probabilidades:

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

II. $P(\Omega) = 1$.

III. $P(A \cup B) = P(A) + P(B)$ desde que A e B sejam eventos mutuamente exclusivos.

IV. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ para eventos quaisquer.

V. $P(\overline{A}) = 1 - P(A)$ para eventos complementares.

VI. $P(A - B) = P(A) - P(A \cap B)$.

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

VIII. $P(A \cap B) = P(A) \cdot P(B)$ desde que A e B sejam eventos independentes.

2.1.8.1 Eventos Equiprováveis

São chamados de eventos equiprováveis quando todas as possibilidades dentro do espaço amostral têm a mesma probabilidade de ocorrer.

Temos como exemplo o dado, no qual cada face tem a mesma probabilidade de cair voltada para cima em um lançamento; ou a moeda, na qual cara e coroa têm a mesma chance de ocorrer; ou, ainda, as cartas de um baralho, que, em um sorteio, têm a mesma probabilidade de ocorrência.

Exemplo 2.11: Retira-se ao acaso uma carta de um baralho. Sabendo que o baralho possui 52 cartas, tendo quatro símbolos que representam os naipes (♣,♦,♥,♠) e cada naipe possui 13 cartas (do 2 ao 10, mais 4 letras – A, J, Q, K).

Ω = {A♣, 2♣, 3♣,4♣, 5♣, 6♣, 7♣, 8♣, 9♣, 10♣, J♣, Q♣, K♣,
A♦, 2♦, 3♦,4♦, 5♦, 6♦, 7♦, 8♦, 9♦, 10♦, J♦, Q♦, K♦,
A♠, 2♠, 3♠,4♠, 5♠, 6♠, 7♠, 8♠, 9♠, 10♠, J♠, Q♠, K♠,
A♥, 2♥, 3♥,4♥, 5♥, 6♥, 7♥, 8♥, 9♥, 10♥, J♥, Q♥, K♥}

ou seja, Ω = { 52 cartas do baralho }

Qual a probabilidade de sair um ás ou uma carta de espadas?

Resolução:
Espaço amostral - Ω = { 52 cartas }, então n = 52

Evento A = tirar um ás = { A♣,A♦,A♥,A♠ }

$$P(A) = \frac{a}{n} = \frac{4}{52}$$

Evento B = tirar uma carta de espadas = { A♠,2♠, ..., K♠}

$$P(B) = \frac{a}{n} = \frac{13}{52}$$

Percebam que o evento A e o evento B não são mutuamente exclusivos, ou seja, existe uma interseção no A♠.

$$P(A \cap B) = \frac{a}{n} = \frac{1}{52}$$

Então, utilizamos a propriedade (iv):

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

$$P(A \cup B) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = 0,3077 = 30,77\%$$

Então, a probabilidade de se retirar um ás ou uma carta de espadas é de 30,77%.

2.1.8.2 Probabilidade Condicional e Independência

Sejam A e B dois eventos de um espaço amostral Ω, associado a um experimento E, em que $P(A) > 0$. A probabilidade de B ocorrer condicionada a

A ter ocorrido será representada por $P(B/A)$, que lemos como: "probabilidade de B dado A", e calculada por:

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

Exemplo 2.12: Suponha a fabricação de calçados, em que se deseja extrair calçados ao acaso de um lote com 100 peças, dos quais 90 calçados estão em perfeito estado e 10 são defeituosos (possuem algum tipo de defeito), de acordo com os critérios (a) com reposição e (b) sem reposição. Definem-se os seguintes eventos:

A = { O primeiro calçado é defeituoso}.
B = { O segundo calçado é defeituoso}.

Então, se a extração for **com** reposição, temos:

$$P(A) = P(B) = \frac{10}{100} = \frac{1}{5} = 0{,}1 = 10\%$$

porque existem 10 calçados defeituosos em um total de 100.

Agora, se a extração for **sem** reposição, temos que:

$$P(A) = \frac{10}{100} = 0{,}1 = 10\%$$

Mas o mesmo não se aplica para $P(B)$. Nesse caso, é necessário verificar o lote no momento da extração da segunda peça, isto é, é preciso saber se a primeira peça retirada era realmente defeituosa. Nesse caso, é necessário saber se A ocorreu ou não, e isso mostra a necessidade do conceito de probabilidade condicionada.

Então, $P(B/A) = 9/99$, pois se A ocorreu (isto é, se saiu peça defeituosa na primeira retirada) existirão no lote apenas 99 peças, das quais são nove defeituosas.

Observe que quando calculamos $P(B/A)$ estamos calculando a probabilidade do evento B em relação ao espaço amostral reduzido do evento A, e não em relação ao espaço amostral total Ω, já que B ocorre somente se A ocorrer .

Verificamos, então, as seguintes propriedades de $P(B/A)$ para A fixado:

I. $0 \le P(B/A) \le 1$,
II. $P(\Omega/A) = 1$

Exemplo 2.13: Consideremos uma empresa com 300 funcionários, distribuídos conforme a tabela a seguir:

Tabela 2.1: Distribuição de funcionários por área e sexo.

	Diretoria	Gerência	Técnico	Total
Homens	7	15	158	180
Mulheres	4	12	104	120
Total	**11**	**27**	**262**	**300**

Fonte: Elaborada pela autora (2016).

Um funcionário é sorteado ao acaso. Pergunta-se:

a. Qual a probabilidade de ser homem?

$$P(H) = \frac{180}{300} = 0,6 = 60\%$$

b. Qual a probabilidade de ser mulher?

$$P(M) = \frac{120}{300} = 0,4 = 40\%$$

c. Qual a probabilidade de ser da Diretoria?

$$P(D) = \frac{11}{300} = 0,0267 = 3,67\%$$

d. Dado que é da Gerência, qual a probabilidade de ser homem?

$$P(H / G) = \frac{15}{27} = 0,5556 = 55,56\%$$

e. Dado que é da Gerência, qual a probabilidade de ser mulher?

$$P(M / G) = \frac{12}{27} = 0,4444 = 44,44\%$$

f. Dado que é homem, qual a probabilidade de ser Técnico?

$$P(T / H) = \frac{158}{180} = 0,8778 = 87,78\%$$

g. Dado que é mulher, qual a probabilidade de ser Técnico?

$$P(T / M) = \frac{104}{120} = 0,8667 = 86,67\%$$

Muitas vezes o espaço amostral torna-se muito complicado de ser calculado, então, podemos associar ao experimento uma função matemática. Assim, temos as funções de distribuição de probabilidade.

2.1.9 Funções de Distribuição de Probabilidades

Vejamos alguns conceitos relacionados à distribuição de probabilidades.

2.1.9.1 Variável Aleatória

Temos uma variável aleatória, quando podemos associar a cada um dos resultados do espaço amostral a um valor numérico.

Como observamos, nem sempre nosso espaço amostral é composto por números. Por exemplo, o experimento lançar uma moeda três vezes e observar a face voltada para cima, na qual temos o espaço amostral Ω={CCC, CCK, CKC, KCC, KKC, KCK, CKK, KKK}. Nesse experimento precisamos associar um valor numérico para podermos calcular a probabilidade de ocorrência de uma cara, por exemplo.

2.9.1.2 Variável Aleatória Discreta

Sempre que uma variável aleatória X puder assumir apenas valores isolados ao longo de uma escala ela será classificada como discreta. E, para cada possível valor da variável, podemos associar a sua probabilidade definida por $f(x) = P(X = x)$. Estaremos, assim, construindo a **distribuição de probabilidade**. A função $f(x)$ é chamada **função (massa) de probabilidade** e satisfaz às condições:

I. $f(x) \geq 0$

II. $\sum f(x_i) = 1$

Exemplo 2.14 - Seja o exemplo anterior: lançar uma moeda 3 vezes e observar a face voltada para cima, temos o espaço amostral:

Ω={CCC, CCK, CKC, KCC, KKC, KCK, CKK, KKK}

Em que K é cara e C é coroa.

Seja X o número de caras obtido, cada um dos resultados tem probabilidade de $1/8$. Assim, a distribuição de probabilidade de X é:

$$X = 0 \rightarrow f(0) = P(CCC) = \frac{1}{8}$$

$$X = 1 \rightarrow f(1) = P(CCK, CKC, KCC) = \frac{3}{8}$$

$$X = 2 \rightarrow f(2) = P(KKC, KCK, CKK) = \frac{3}{8}$$

$$X = 3 \rightarrow f(3) = P(KKK) = \frac{1}{8}$$

Tabela 2.2: Exemplo 2.14.

X	0	1	2	3
$f(x)$	1/8	3/8	3/8	1/8

Fonte: Elaborada pela autora (2016).

Exemplo 2.15: Dois dados são lançados e observa-se o par obtido. O espaço amostral é formado por 36 resultados equiprováveis. Seja a variável X a soma do par obtido, temos então:

$$X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\}$$

$$f(2) = P(X = 2) = P(\{1,1\}) = {}^{1}\!/_{36}$$
$$f(3) = P(X = 3) = P(\{(2,1),(1,2)\}) = {}^{2}\!/_{36}$$
$$f(4) = P(X = 4) = P(\{(1,3),(2,2),(3,1)\}) = {}^{3}\!/_{36}$$
$$f(5) = P(X = 5) = P(\{(1,4),(2,3),(3,2),(4,1)\}) = {}^{4}\!/_{36}$$
$$f(6) = P(X = 6) = P(\{(1,5),(2,4),(3,3),(4,2),(5,1)\}) = {}^{5}\!/_{36}$$
$$f(7) = P(X = 7) = P(\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}) = {}^{6}\!/_{36}$$
$$f(8) = P(X = 8) = P(\{(2,6),(3,5),(4,4),(5,3),(6,2)\}) = {}^{5}\!/_{36}$$
$$f(9) = P(X = 9) = P(\{(3,6),(4,5),(5,4),(6,3)\}) = {}^{4}\!/_{36}$$
$$f(10) = P(X = 10) = P(\{(4,6),(5,5),(6,4)\}) = {}^{3}\!/_{36}$$
$$f(11) = P(X = 11) = P(\{(5,6),(6,5)\}) = {}^{2}\!/_{36}$$
$$f(12) = P(X = 12) = P(\{(6,6)\}) = {}^{1}\!/_{36}$$

Tabela 2.3: Exemplo 2.15.

X	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	1/36	2/36	3/36	4/36	5/36	6/37	5/36	4/36	3/36	2/36	1/36

Fonte: Elaborada pela autora (2016).

Existem algumas funções de probabilidade que possuem uma função matemática conhecida e que se adequam a vários experimentos.

2.1.10 Distribuição de Probabilidades Discreta

Como podemos assumir um valor numérico para cada função de X, podemos então construir uma função matemática que representa cada um dos valores de x da nossa distribuição.

Aqui, abordaremos duas funções de distribuição discretas para variáveis aleatórias discretas (VAD). São elas a distribuição Binomial e a distribuição de Poisson.

2.1.10.1 Distribuição Binomial

Temos um experimento E, com um espaço amostral associado a Ω, um evento desse espaço amostral A.

Para a distribuição binomial, temos n, que é o número de vezes em que o experimento E é repetido, e seja p a probabilidade de A ocorrer **em cada uma das n repetições** desse experimento, de modo que p permaneça constante durante as n repetições do experimento E.

Para a distribuição Binomial, precisamos de uma variável que possa assumir apenas duas situações: A ocorre ou A não ocorre.

Então definimos a probabilidade de A ocorrer como sendo p; e a probabilidade de A não ocorrer $q = 1 - p$ (ou seja, A complementar).

Normalmente chamamos de p a probabilidade de "sucesso" e q a probabilidade de fracasso.

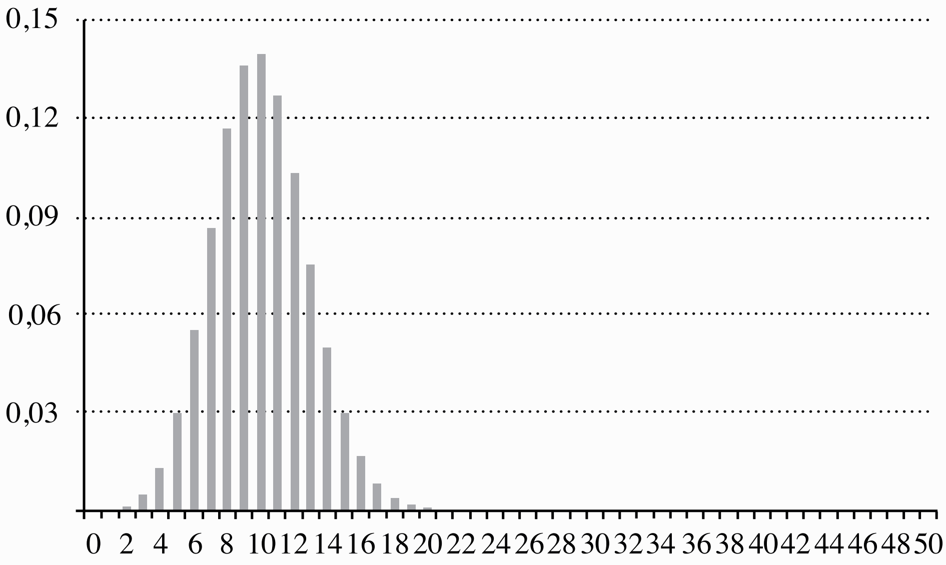
Então, seja X uma VAD definida por X = número de vezes em que A ocorreu nas n repetições de E.

Essa variável aleatória X é denominada de variável aleatória Binomial. O conjunto de valores de X, isto é, $X(\Omega)$, é:

$$X(\Omega) = \{0, 1, 2, 3, ..., n\}$$

Vamos exemplificar a forma gráfica de uma função binomial, utilizando 50 repetições (n), para uma probabilidade de sucesso de 30% ($p = 0,3$). Representamos essa distribuição por $B(n, p)$ para o exemplo $B(50; 0,3)$.

Figura 2.6: Distribuição binomial.



Fonte: Elaborada pela autora (2016).

Se X é uma variável aleatória com um comportamento binomial, então a probabilidade de X assumir um dos valores do conjunto $X(\Omega)$ é calculada por:

$$f(x) = P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x} \text{ para } x = 0, 1, 2, ..., n.$$

Exemplo 2.16: Considerando X como sendo a variável aleatória discreta (VAD) igual ao "número de vezes em que ocorre face cara em 10 lançamentos de uma moeda equilibrada", determine a probabilidade de ocorrerem:

a. duas caras.
b. no máximo duas caras.
c. mais de uma cara.

Resolução:

Primeiro, é melhor lermos as informações do enunciado e as anotarmos.

Então:

O número de repetições do experimento → $n = 10$

Com relação à probabilidade de sucesso (obter cara em cada um dos lançamentos), sabemos que a moeda tem duas faces. Logo, a probabilidade de cair uma face cara é de 1/2 → $p = 0,5$.

O número de caras nos 10 lançamentos (aqui pensamos em todos os resultados que X pode assumir) → $X(\Omega) = \{0,1,2,3,4,5,6,7,8,9,10\}$.

Agora, aplicando a função binomial, temos:

$$f(x) = P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

a. Aqui, basta calcular o $x = 2$

$$P(X = 2) = \binom{10}{2} \cdot 0,5^2 \cdot 0,5^{10-2} = 0,0439 = 4,39\%$$

b. Aqui, precisamos de, no máximo, duas caras, então pode acontecer $X = \{0,1,2\}$. Assim, temos:

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X \leq 2) = \binom{10}{0} \cdot 0,5^0 \cdot 0,5^{10-0} + \binom{10}{1} \cdot 0,5^1 \cdot 0,5^{10-1} + \binom{10}{2} \cdot 0,5^2 \cdot 0,5^{10-2}$$
$$= 0,000977 + 0,009766 + 0,043945 = 0,0547 = 5,47\%$$

c. Agora, precisamos de uma cara ou mais, então pode acontecer $X = \{1,2,3,4,5,6,7,8,9,10\}$. Aqui, ao invés de fazermos todos esses cálculos, podemos utilizar a mesma lógica do complementar, então podemos calcular da seguinte maneira:

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0)$$

$$P(X \geq 1) = 1 - \binom{10}{0} \cdot 0,5^0 \cdot 0,5^{10-0} = 0,000977 = 0,097\%$$

Na função, aonde aparece $\binom{n}{x}$, temos a combinação de n com x. Isso significa que o cálculo que precisamos é o da combinação, que é dado por:

$$\binom{n}{x} = \frac{n!}{x! \cdot (n - x)}$$

Relembrando: o ponto de exclamação significa fatorial. Na maioria das calculadoras científicas resolvemos essa combinação com a tecla nCr.

A vantagem de termos um modelo matemático conhecido é que podemos determinar suas características de um modo geral. Vejamos as propriedades da distribuição binomial a seguir.

- **Propriedades da distribuição binomial**
  Média, expectância ou valor esperado:

  $\mu = E(X) = n \cdot p$, isto é, a média de uma variável aleatória com distribuição binomial é igual ao produto dos parâmetros n e p.

  Variância:

  $\sigma^2 = E(X^2) = VAR(X) = n \cdot p \cdot q$, isto é, a variância de uma variável aleatória com distribuição binomial é igual ao produto dos parâmetros n e p multiplicados, ainda, por q.

  O desvio-padrão:

  $\sigma = \sqrt{n \cdot p \cdot q}$

Exemplo 2.17: A probabilidade de encontrar uma peça defeituosa em um lote é de 10%. Considerando X a variável "número de peças defeituosas em um lote ocasional de 10 unidades, utilizando as propriedades da binomial", determinaremos:

a. o número médio de peças defeituosas na amostra;
b. o desvio-padrão do número de peças defeituosas na amostra.

Resolução:

a. $E(X) = n \cdot p = 10 \cdot 0{,}10 = 1$ item defeituoso.

Espera-se encontrar, em média, uma peça defeituosa por lote de 10 unidades.

b. $\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{20 \cdot 0{,}10 \cdot 0{,}90} = 0{,}99$ itens defeituosos.

Espera-se uma variação em torno da média de 0,99 peças defeituosas.

Exemplo 2.18: A proporção de rebatidas corretas de um jogador de beisebol é de 30%. Se ele bate quatro vezes, qual é a probabilidade de ele obter:

a. dois acertos.
b. pelo menos um acerto.

Solução:

a. $P(X=2)=\binom{4}{2}\cdot 0,3^2\cdot 0,7^{4-2}=0,2646=26,46\%$

b. $P(pelo\ menos\ um\ acerto)=P(X=1)+P(X=2)+P(X=3)+P(X=4)$

Ao invés de calcular dessa forma, é mais conveniente utilizar o complementar. Assim:

$P(X\geq 1)=1-P(X=0)=1-0,2401=75,99\%$

Outra distribuição a ser vista no curso é a distribuição de Poisson.

2.1.10.2 A Distribuição de Poisson

Enquanto que na distribuição binomial temos o número de ocorrências do evento *A*, o que representa nosso número de sucessos, em um número de repetições (que resultam de uma contagem, um intervalo discreto) na distribuição de Poisson também teremos um número de sucessos, porém, os contabilizaremos não mais em um número de repetições, mas sim em um intervalo contínuo, seja de tempo ou de espaço.

Para se caracterizar uma distribuição que leve em conta o número de sucessos em um intervalo contínuo, será suposto que:

I. Em intervalos de mesmo comprimento, as probabilidades de sucessos são equiprováveis.
II. Em intervalos muito pequenos, a probabilidade de mais de um sucesso é muito baixa.
III. Em intervalos muito pequenos, a probabilidade de um sucesso é proporcional ao comprimento do intervalo.

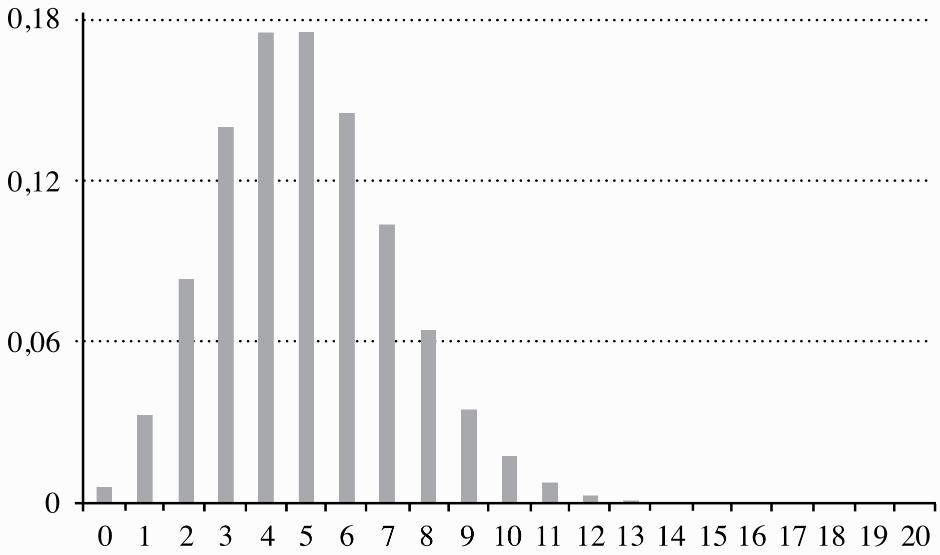
Então, seja X uma variável aleatória discreta (VAD) definida por um processo de Poisson, assumindo os valores: $0,1,2,....$, com taxa λ.

Temos:

$f(x)=P(X=x)=\dfrac{e^{-\lambda}\cdot\lambda^x}{x!}$ para $x=0,1,2,....$, em que x é o número de eventos que ocorrem em um intervalo contínuo, que se espera uma taxa λ de ocorrências.

A distribuição de Poisson será representada por $P(\lambda)$, a seguir a representação gráfica para $P(5)$.

Figura 2.7: Distribuição de Poisson.



Fonte: Elaborada pela autora (2016).

Exemplo 2.19: Em uma central telefônica chegam ligações que deixam de ser atendidas a uma taxa de quatro ligações por hora. Seja X o número de chamadas não atendidas por hora, calcule as probabilidades:

a. de não haver chamadas não atendidas em 30 minutos;
b. de termos no máximo duas chamadas não atendidas por hora;
c. de termos mais de quatro chamadas não atendidas em duas horas.

Resolução:
Novamente, é melhor primeiro lermos as informações do enunciado e a anotarmos. Então:

λ = Taxa de ligações perdidas por hora (60 minutos).
X = número de ligações perdidas.
x = 0, 1, 2, 3, [...].

Então:

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

a. Aqui, queremos a probabilidade de não termos chamadas não atendidas $X = \{0\}$. Mas temos de observar que no enunciado a taxa é dada por hora, e agora queremos calcular a probabilidade em um intervalo de meia hora. Podemos, então, calcular uma regra de três e obter o valor da taxa por meia hora.

Assim, a taxa de ligações perdidas em meia hora é $\lambda = 2$.

$$P(X=0)=\frac{e^{-2}\cdot 2^{0}}{0!}=0{,}1353=13{,}53\%$$

b. Agora, temos a probabilidade de, no máximo, duas chamadas. Então, $X=\{0,1,2\}$. A uma taxa de quatro ligações perdidas por hora, temos:

$$P(X\leq 2)=P(X=0)+P(X=1)+P(X=2)$$

$$P(X\leq 2)=\frac{e^{-4}\cdot 4^{0}}{0!}+\frac{e^{-4}\cdot 4^{1}}{1!}+\frac{e^{-4}\cdot 4^{2}}{2!}=0{,}01831+0{,}07326+0{,}146525=$$
$$0{,}2381=23{,}81\%$$

c. Queremos agora, pelo menos, quatro chamadas não atendidas: $X=\{4,5,6,7,8,9,........\}$. Observe que temos uma taxa por duas horas; usando regra de três novamente, temos uma taxa de oito chamadas não atendidas em duas horas.

$$P(X\geq 4)=P(X=4)+P(X=5)+P(X=6)+P(X=7)+P(X=8)+P(X=9)+.......$$

Outra observação importante é que utilizaremos a teoria do complementar para resolver essa função.

$$P(X\geq 4)=1-P(X<4)=1-\left(P(X=0)+P(X=1)+P(X=2)+P(X=3)\right)$$

$$P(X\geq 4)=1-\frac{e^{-8}\cdot 8^{0}}{0!}+\frac{e^{-8}\cdot 8^{1}}{1!}+\frac{e^{-8}\cdot 8^{2}}{2!}+\frac{e^{-8}\cdot 8^{3}}{3!}=$$
$$0{,}000335+0{,}002684+0{,}010735+0{,}028626=0{,}0424=4{,}24\%$$

Lembrem-se de que $x!$ é a representação para o fatorial de x.

- **Propriedades da Distribuição de Poisson**

  Se X for uma variável aleatória discreta (VAD) com distribuição de Poisson, então:

  Média, expectância ou valor esperado:

  $$\mu=E(X)=\lambda$$

Variância:

$$\sigma^2 = VAR(X) = \lambda$$

O desvio-padrão:

$$\sigma = \sqrt{\lambda}$$

Podemos também fazer uma relação entre as duas distribuições quando seguirmos alguns pressupostos.

- **Relação entre as distribuições binomial e Poisson**
  Seja X uma VAD com distribuição binomial de parâmetros n e p. Isto é:

$$f(x) = P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

Admita-se que quando *n* tende *a* $\propto$ e *p* tende a 0, de modo que *n.p* permanece constante; nessas condições, tem-se, então:

$$\lim_{n \to a} P(X = x) = \lim_{n \to a} \binom{n}{x} \cdot p^x \cdot q^{n-x} = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

Traduzindo, o teorema diz que é possível obter uma aproximação das probabilidades binomiais com as probabilidades da distribuição de Poisson, toda vez que n seja grande e p seja pequeno.

Exemplo 2.20: Uma amostra de 400 peças é retirada da produção de uma máquina que trabalha com um índice de defeitos de 1%. Determine a probabilidade de se encontrarem duas peças defeituosas na amostra.

Resolução:
Pela distribuição binomial, tem-se:

$$P(X = 2) = \binom{400}{2} \cdot 0,01^2 \cdot 0,99^{400-2} = 0,1462 = 14,62\%$$

Para usarmos a aproximação pela distribuição de Poisson precisamos estimar o λ, então podemos utilizar a média da binomial e estimarmos, assim, a média da Poisson. Isso podemos relembrar ao retomarmos as propriedades da binomial e as propriedades da Poisson, basta olhar nas fórmulas do valor esperado para cada uma delas: $\mu = n \cdot p = 400 \cdot 0,01 = 4$ é o valor esperado da binomial, já que, pelo teorema, podemos estimar a binomial pela Poisson e sabemos que o valor esperado da distribuição de Poisson é dado por $\mu = \lambda$ então $\mu = \lambda = 4$

Pela distribuição de Poisson, temos:

$$P(X = 2) = \frac{e^{-4} \cdot 4^2}{2!} = 0,1465 = 14,65\%$$

Como podemos observar, os valores das probabilidades calculados pela binomial e pela Poisson são muito próximos, comprovando assim o teorema.

Vimos aqui duas distribuições de probabilidade discretas. Claro que existem muitas outras, que podem ser funções conhecidas ou funções elaboradas especialmente para um experimento determinado. Continuaremos os estudos aprendendo a mais importante distribuição contínua de probabilidade, a Distribuição Normal.

SÍNTESE DA UNIDADE

Nesta unidade, vimos as primeiras noções de probabilidade e alguns conceitos iniciais. Estudamos:

- os conceitos iniciais de experimento, espaço amostral, eventos;
- métodos de enumeração de espaço amostral;
- operações com eventos;
- propriedades das probabilidades;
- a função distribuição de probabilidades binomial, suas aplicações e propriedades;
- a função da distribuição de probabilidades de Poisson, suas aplicações e propriedades.

REFERÊNCIAS

STEVENSON, W. J. **Estatística Aplicada à Administração**. São Paulo: Harbra, 2001.

ANOTAÇÕES

Distribuição Normal

Prezado(a) estudante.

Estamos começando uma unidade desta disciplina. Os textos que a compõem foram organizados com cuidado e atenção, para que você tenha contato com um conteúdo completo e atualizado tanto quanto possível. Leia com dedicação, realize as atividades e tire suas dúvidas com os tutores. Dessa forma você, com certeza, alcançará os objetivos propostos para essa disciplina.

OBJETIVO GERAL

Apresentar a distribuição de probabilidade Normal.

OBJETIVOS ESPECÍFICOS

- Elucidar como calcular probabilidades de acordo com a distribuição Normal, a mais importante distribuição de probabilidades contínua.
- Exemplificar aplicações e proporcionar o uso da distribuição nas mais diversas variáveis.

3.1 Distribuição Normal

Já imaginou como calcular o tempo de garantia de um produto produzido em uma empresa para que esta troque, no máximo, 10% das peças fabricadas? Ou, então, saber qual a probabilidade de uma financeira ficar sem dinheiro para emprestar, em um ano que ela reservou determinado valor para empréstimo?

Essas são algumas das questões que você irá aprender a responder com os conceitos e ferramentas que vamos estudar nesta Unidade.

Vimos, na Unidade anterior, algumas distribuições de probabilidade discretas, ou seja, para variáveis com nível de mensuração discretas. Não existem apenas essas, mas abordamos as mais importantes.

Na prática, temos mais variáveis com nível de mensuração contínua e temos, também, algumas distribuições de probabilidade predeterminadas para alguns tipos de variáveis.

Abordaremos, aqui, a mais importante função de distribuição de probabilidades, a distribuição normal, também conhecida como a Distribuição de Gauss ou Gaussiana.

A distribuição normal tem grande importância na inferência estatística, pois dela surge o **teorema do limite central**. Muitas teorias estatísticas surgem da suposição de normalidade dos dados em estudo.

GLOSSÁRIO

Teorema do limite central: basicamente determina que valores individuais podem não seguir uma distribuição Normal, mas se somarmos e calcularmos a média de grupos de valores, estes tenderão a seguir o modelo Normal, independentemente da distribuição que os dados seguiam de modo individual.

Com a distribuição normal, podemos realizar testes de hipóteses para médias. Além disso, podemos estabelecer limites de confiança, calcular tamanhos de amostras probabilísticas, além de tantas outras aplicações.

3.1.1 Função Densidade de Probabilidade (FDP) Normal

A função matemática que descreve a distribuição Normal está descrita a seguir:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ para } -\infty < x < \infty \text{ , } -\infty < \mu < \infty \text{ e } \sigma^2 > 0$$

Os parâmetros da distribuição Normal são a média (μ) e o desvio-padrão (σ).

A média é o parâmetro de localização da distribuição, e o desvio-padrão, o parâmetro de variabilidade, ou seja, o desvio-padrão define a forma da distribuição (mais achatada ou mais fina).

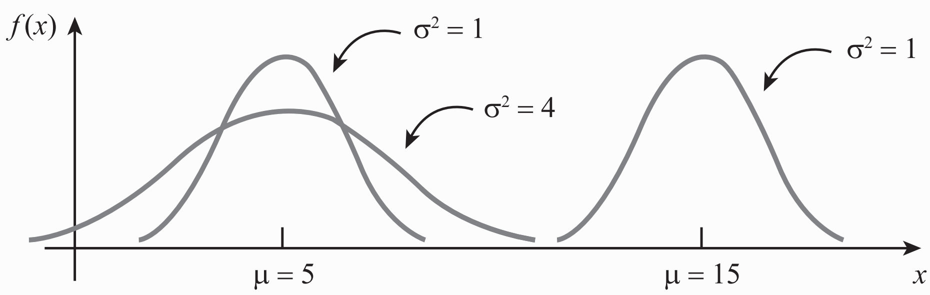
Podemos representar uma variável seguindo uma distribuição Normal dessa forma:

$$X \sim N(\mu, \sigma)$$

Ou seja, a variável X segue uma distribuição Normal, com média μ e desvio-padrão σ .

A forma da distribuição Normal é semelhante a um sino: ela tem o pico mais alto na média, o centro da distribuição e o formato conforme o desvio-padrão, sendo que quanto maior o desvio-padrão, mais achatada ela será, e quanto menor o desvio-padrão, mais estreita, conforme pode ser visto na Figura 3.1

Figura 3.1: Formatos da distribuição Normal.



Fonte: Adaptada de Montgomery e Runger (2012).

Outra informação importante sobre a distribuição normal é que ela é simétrica em torno da média.

Então, as propriedades da distribuição Normal são:

i. A distribuição Normal é simétrica em torno da média (μ).

ii. Quanto maior for o desvio-padrão, mais achatado será o gráfico da distribuição Normal.

iii. A área total abaixo da curva soma 1 (1 corresponde a 100%).

iv. Os parâmetros são a média (μ) e o desvio-padrão (σ).

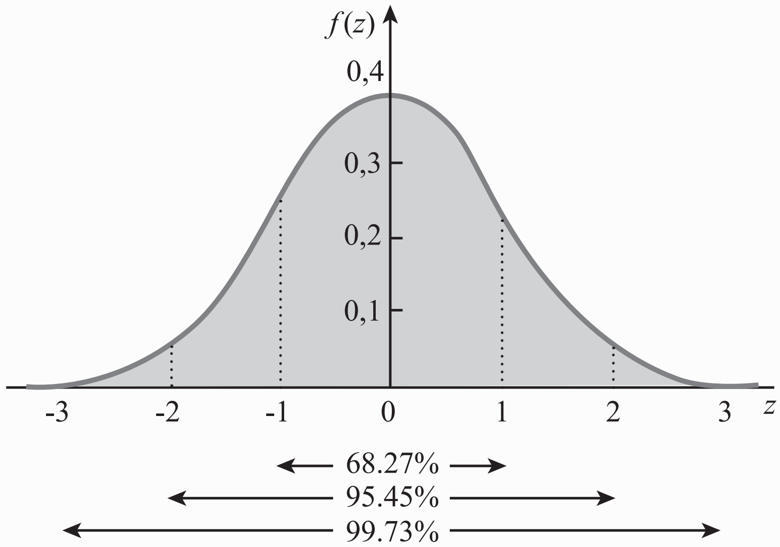
v. Não existe probabilidade menor do que zero, nem maior do que 1.

Assim, para calcular a probabilidade de uma função contínua precisamos matematicamente calcular a sua integral. Mas a integral da função densidade da Normal é de difícil resolução analítica.

Como essa resolução é complicada, passamos a utilizar uma distribuição chamada de Normal Padrão, em que assumimos média igual a 0 ($\mu = 0$) e desvio-padrão igual a 1 ($\sigma = 1$).

Na Figura 3.2 temos o gráfico da distribuição Normal Padrão. Observe que o centro está no zero, já que temos $\mu = 0$. Já os valores no eixo $f(z)$ e z no gráfico resultam da função distribuição de probabilidade Normal vista anteriormente.

Figura 3.2: Distribuição Normal Padrão.



Fonte: Adaptada de Spiegel, Schiller e Srinivasan (2015).

Então, como não queremos calcular a integral complexa para cada variável que estudarmos, precisamos utilizar uma padronização para os dados, assumindo que a variável tenha média zero e desvio-padrão igual a um. Efetuando essa padronização aos dados, passamos de uma variável X para uma variável padronizada Z. Essa padronização é feita com o intuito de termos uma tabela que fornecerá, no seu interior, a probabilidade associada ao valor de Z, cuja área começa no zero e vai até o ponto especificado.

Precisamos utilizar essa padronização para que se torne possível utilizar uma tabela para qualquer variável que seja normalmente distribuída, independentemente da sua média e do seu desvio-padrão.

A distribuição Normal com média zero ($\mu = 0$) e desvio-padrão igual a um ($\sigma = 1$) é chamada de Normal Padrão.

Na figura a seguir temos a tabela de probabilidades da distribuição Normal (uma versão ampliada da tabela a seguir você encontra no Anexo I, ao final do texto).

Figura 3.3: Probabilidades da curva Normal Padrão.

Áreas sob a curva normal
(Cauda da esquerda)

z	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	0,01	0,00
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,6	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,4	0,0002	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003
-3,3	0,0003	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0005	0,0005	0,0005
-3,2	0,0005	0,0005	0,0005	0,0006	0,0006	0,0006	0,0006	0,0006	0,0007	0,0007
-3,1	0,0007	0,0007	0,0008	0,0008	0,0008	0,0008	0,0009	0,0009	0,0009	0,0010
-3,0	0,0010	0,0010	0,0011	0,0011	0,0011	0,0012	0,0012	0,0013	0,0013	0,0013
-2,9	0,0014	0,0014	0,0015	0,0015	0,0016	0,0016	0,0017	0,0018	0,0018	0,0019
-2,8	0,0019	0,0020	0,0021	0,0021	0,0022	0,0023	0,0023	0,0024	0,0025	0,0026
-2,7	0,0026	0,0027	0,0028	0,0029	0,0030	0,0031	0,0032	0,0033	0,0034	0,0035
-2,6	0,0036	0,0037	0,0038	0,0039	0,0040	0,0041	0,0043	0,0044	0,0045	0,0047
-2,5	0,0048	0,0049	0,0051	0,0052	0,0054	0,0055	0,0057	0,0059	0,0060	0,0062
-2,4	0,0064	0,0066	0,0068	0,0069	0,0071	0,0073	0,0075	0,0078	0,0080	0,0082
-2,3	0,0084	0,0087	0,0089	0,0091	0,0094	0,0096	0,0099	0,0102	0,0104	0,0107
-2,2	0,0110	0,0113	0,0116	0,0119	0,0122	0,0125	0,0129	0,0132	0,0136	0,0139
-2,1	0,0143	0,0146	0,0150	0,0154	0,0158	0,0162	0,0166	0,0170	0,0174	0,0179
-2,0	0,0183	0,0188	0,0192	0,0197	0,0202	0,0207	0,0212	0,0217	0,0222	0,0228
-1,9	0,0233	0,0239	0,0244	0,0250	0,0256	0,0262	0,0268	0,0274	0,0281	0,0287
-1,8	0,0294	0,0301	0,0307	0,0314	0,0322	0,0329	0,0336	0,0344	0,0351	0,0359
-1,7	0,0367	0,0375	0,0384	0,0392	0,0401	0,0409	0,0418	0,0427	0,0436	0,0446
-1,6	0,0455	0,0465	0,0475	0,0485	0,0495	0,0505	0,0516	0,0526	0,0537	0,0548
-1,5	0,0559	0,0571	0,0582	0,0594	0,0606	0,0618	0,0630	0,0643	0,0655	0,0668
-1,4	0,0681	0,0694	0,0708	0,0721	0,0735	0,0749	0,0764	0,0778	0,0793	0,0808
-1,3	0,0823	0,0838	0,0853	0,0869	0,0885	0,0901	0,0918	0,0934	0,0951	0,0968
-1,2	0,0985	0,1003	0,1020	0,1038	0,1056	0,1075	0,1093	0,1112	0,1131	0,1151
-1,1	0,1170	0,1190	0,1210	0,1230	0,1251	0,1271	0,1292	0,1314	0,1335	0,1357
-1,0	0,1379	0,1401	0,1423	0,1446	0,1469	0,1492	0,1515	0,1539	0,1562	0,1587
-0,9	0,1611	0,1635	0,1660	0,1685	0,1711	0,1736	0,1762	0,1788	0,1814	0,1841
-0,8	0,1867	0,1894	0,1922	0,1949	0,1977	0,2005	0,2033	0,2061	0,2090	0,2119
-0,7	0,2148	0,2177	0,2206	0,2236	0,2266	0,2296	0,2327	0,2358	0,2389	0,2420
-0,6	0,2451	0,2483	0,2514	0,2546	0,2578	0,2611	0,2643	0,2676	0,2709	0,2743
-0,5	0,2776	0,2810	0,2843	0,2877	0,2912	0,2946	0,2981	0,3015	0,3050	0,3085
-0,4	0,3121	0,3156	0,3192	0,3228	0,3264	0,3300	0,3336	0,3372	0,3409	0,3446
-0,3	0,3483	0,3520	0,3557	0,3594	0,3632	0,3669	0,3707	0,3745	0,3783	0,3821
-0,2	0,3859	0,3897	0,3936	0,3974	0,4013	0,4052	0,4090	0,4129	0,4168	0,4207
-0,1	0,4247	0,4286	0,4325	0,4364	0,4404	0,4443	0,4483	0,4522	0,4562	0,4602
0,0	0,4641	0,4681	0,4721	0,4761	0,4801	0,4840	0,4880	0,4920	0,4960	0,5000

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Fonte: Elaborada pela autora (2016).

Observe que temos a área negativa e positiva, que são simétricas em torno da média.

Essa tabela fornece as probabilidades da área compreendida do $-\infty$ até o ponto X ,ou seja, a integral foi calculada da seguinte forma:

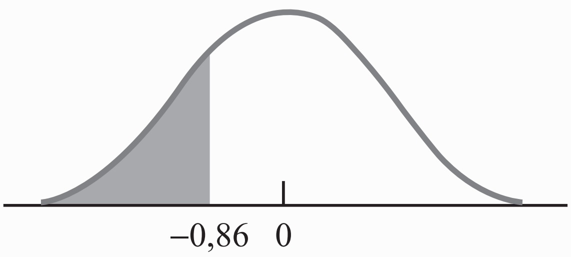
$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Não utilizaremos essa integral por causa da sua complexidade, mas, sim, pela padronização para a variável *Z*.

Antes de aprendermos a padronizar a variável X, vamos ver como olhar a probabilidade na tabela. Para aprendermos a utilização, precisamos considerar a variável já padronizada z com $\mu = 0$ e $\sigma = 1$.

i. Probabilidade da variável Z ser menor ou igual a -0,86 $P(Z \leq -0,86)$

Figura 3.4: Curva Normal Padrão.



Fonte: Adaptada de Montgomery e Runger (2012).

Observe que queremos a probabilidade inferior ao ponto -0,86, então, essa é justamente a probabilidade que a tabela nos fornece. Podemos observar diretamente na tabela, da linha -0,8 até a coluna 0,06. Cabe destacar que essa tabela nos fornece os valores das áreas correspondentes de $-\infty$ até o ponto Z.

Figura 3.5: Probabilidades da curva normal padrão.

z	0,09	0,08	0,07	0,06	0,05
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001
-3,6	0,0001	0,0001	0,0001	0,0001	0,0001
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002
-3,4	0,0002	0,0003	0,0003	0,0003	0,0003
-3,3	0,0003	0,0004	0,0004	0,0004	0,0004
-3,2	0,0005	0,0005	0,0005	0,0006	0,0006
-3,1	0,0007	0,0007	0,0008	0,0008	0,0008
-3,0	0,0010	0,0010	0,0011	0,0011	0,0011
-2,9	0,0014	0,0014	0,0015	0,0015	0,0016
-2,8	0,0019	0,0020	0,0021	0,0021	0,0022
-2,7	0,0026	0,0027	0,0028	0,0029	0,0030
-2,6	0,0036	0,0037	0,0038	0,0039	0,0040
-2,5	0,0048	0,0049	0,0051	0,0052	0,0054
-2,4	0,0064	0,0066	0,0068	0,0069	0,0071
-2,3	0,0084	0,0087	0,0089	0,0091	0,0094
-2,2	0,0110	0,0113	0,0116	0,0119	0,0122
-2,1	0,0143	0,0146	0,0150	0,0154	0,0158
-2,0	0,0183	0,0188	0,0192	0,0197	0,0202
-1,9	0,0233	0,0239	0,0244	0,0250	0,0256
-1,8	0,0294	0,0301	0,0307	0,0314	0,0322
-1,7	0,0367	0,0375	0,0384	0,0392	0,0401
-1,6	0,0455	0,0465	0,0475	0,0485	0,0495
-1,5	0,0559	0,0571	0,0582	0,0594	0,0606
-1,4	0,0681	0,0694	0,0708	0,0721	0,0735
-1,3	0,0823	0,0838	0,0853	0,0869	0,0885
-1,2	0,0985	0,1003	0,1020	0,1038	0,1056
-1,1	0,1170	0,1190	0,1210	0,1230	0,1251
-1,0	0,1379	0,1401	0,1423	0,1446	0,1469
-0,9	0,1611	0,1635	0,1660	0,1685	0,1711
-0,8	0,1867	0,1894	0,1922	0,1949	0,1977
-0,7	0,2148	0,2177	0,2206	0,2236	0,2266
-0,6	0,2451	0,2483	0,2514	0,2546	0,2578
-0,5	0,2776	0,2810	0,2843	0,2877	0,2912

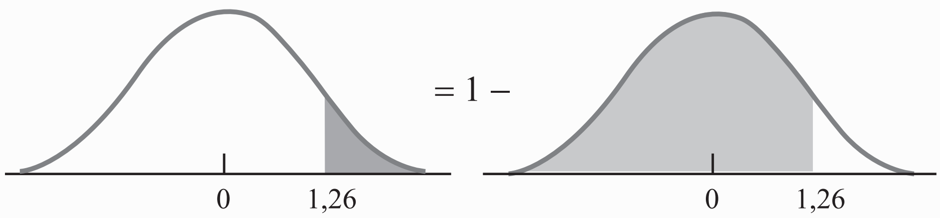
Fonte: Elaborada pela autora (2016).

Assim sendo, a probabilidade da variável Z ser menor ou igual a -0,86 é $P(Z \leq -0{,}86) = 0{,}1949$.

ii. Probabilidade da variável Z ser maior do que 1,26

$P(Z > 1{,}26)$.

Figura 3.6: Curva normal padrão.



Fonte: Adaptada de Montgomery e Runger (2012).

Observe que agora queremos a probabilidade acima do ponto 1,26, então precisamos usar da teoria das probabilidades para podermos obter o valor dessa probabilidade. Assim, utilizaremos a propriedade do complementar. Sabemos que a área total abaixo da curva soma 1, logo, se temos o valor na tabela para o ponto menor do que 1,26, podemos então fazer 1 – essa probabilidade tabelada – e obteremos o nosso resultado. Isso se faz necessário, pois essa tabela foi construída do $-\infty$ até o ponto Z, e o que queremos é a probabilidade do ponto z até o $+\infty$.

Novamente, precisamos conferir a tabela, dessa vez da linha do 1,2 até a coluna do 0,06.

Figura 3.7: Probabilidades da curva normal padrão.

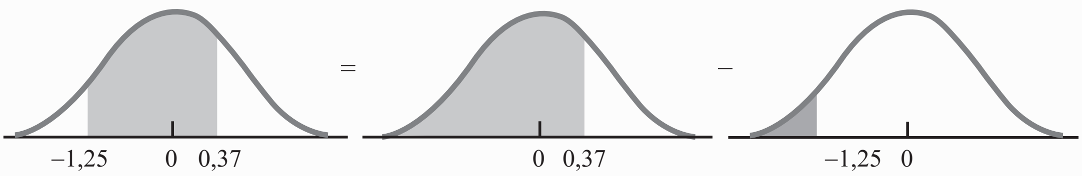
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319

Fonte: Elaborada pela autora (2016).

Assim sendo, a Probabilidade da variável Z ser maior do que 1,26 é $P(Z > 1,26) = 1 - P(Z < 1,26) = 1 - 0,8962 = 0,1038$.

iii. Probabilidade da variável Z estar entre -1,25 e 0,37, ou seja, $P(-1,25 < Z < 0,37)$.

Figura 3.8: Curva normal padrão.



Fonte: Adaptada de Montgomery e Runger (2012).

Agora queremos a probabilidade de um intervalo, então precisamos observar dois valores na tabela – o referente ao ponto 0,37 e o ponto -1,25. Para isso, subtraímos uma área da outra para podermos obter o intervalo. Cabe destacar que fazemos dessa maneira porque a tabela foi construída para fornecer probabilidade do $-\infty$ até o ponto Z .

Novamente, precisamos observar a tabela, pois dessa vez haverá dois valores.

Figura 3.9: Probabilidades da curva normal padrão.

z	0,09	0,08	0,07	0,06	0,05	0,04
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,6	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,4	0,0002	0,0003	0,0003	0,0003	0,0003	0,0003
-3,3	0,0003	0,0004	0,0004	0,0004	0,0004	0,0004
-3,2	0,0005	0,0005	0,0005	0,0006	0,0006	0,0006
-3,1	0,0007	0,0007	0,0008	0,0008	0,0008	0,0008
-3,0	0,0010	0,0010	0,0011	0,0011	0,0011	0,0012
-2,9	0,0014	0,0014	0,0015	0,0015	0,0016	0,0016
-2,8	0,0019	0,0020	0,0021	0,0021	0,0022	0,0023
-2,7	0,0026	0,0027	0,0028	0,0029	0,0030	0,0031
-2,6	0,0036	0,0037	0,0038	0,0039	0,0040	0,0041
-2,5	0,0048	0,0049	0,0051	0,0052	0,0054	0,0055
-2,4	0,0064	0,0066	0,0068	0,0069	0,0071	0,0073
-2,3	0,0084	0,0087	0,0089	0,0091	0,0094	0,0096
-2,2	0,0110	0,0113	0,0116	0,0119	0,0122	0,0125
-2,1	0,0143	0,0146	0,0150	0,0154	0,0158	0,0162
-2,0	0,0183	0,0188	0,0192	0,0197	0,0202	0,0207
-1,9	0,0233	0,0239	0,0244	0,0250	0,0256	0,0262
-1,8	0,0294	0,0301	0,0307	0,0314	0,0322	0,0329
-1,7	0,0367	0,0375	0,0384	0,0392	0,0401	0,0409
-1,6	0,0455	0,0465	0,0475	0,0485	0,0495	0,0505
-1,5	0,0559	0,0571	0,0582	0,0594	0,0606	0,0618
-1,4	0,0681	0,0694	0,0708	0,0721	0,0735	0,0749
-1,3	0,0823	0,0838	0,0853	0,0869	0,0885	0,0901
-1,2	0,0985	0,1003	0,1020	0,1038	0,1056	0,1075
-1,1	0,1170	0,1190	0,1210	0,1230	0,1251	0,1271
-1,0	0,1379	0,1401	0,1423	0,1446	0,1469	0,1492

Fonte: Elaborada pela autora (2016).

Figura 3.10: Probabilidades da curva normal padrão.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844

Fonte: Elaborada pela autora (2016).

Assim sendo, há probabilidade da variável Z estar entre -1,25 e 0,37. Por isso, precisamos calcular o valor da probabilidade do limite superior do intervalo menos o valor da probabilidade do limite inferior do intervalo. Então, temos:

$$P(-1,25 < Z < 0,37) = P(Z < 0,37) - P(Z < -1,25) = 0,6443 - 0,1056 = 0,5387$$

Aprendemos, até aqui, como utilizar essa tabela da Normal Padrão, e agora que já sabemos como olhar as probabilidades na tabela, podemos avançar e aprender a padronizar os dados, transformando-os em uma média igual a zero, e um desvio-padrão igual a 1.

Os dados disponíveis na natureza dificilmente seguirão essa média e esse desvio-padrão. Então, a padronização de dados, quando a média de uma variável x não é zero é feita da seguinte forma:

$$z = \frac{X - \mu_x}{\sigma_x}$$

AMPLIANDO CONHECIMENTO

No vídeo a seguir você pode aprofundar o seu entendimento a respeito da Distribuição Normal. Não deixe de assistir. Para acessá-lo, clique no *link* a seguir: https://goo.gl/uYmkBY. Acesso em: 9 jul. 2017.

Exemplo 3.1: Suponha que uma pequena financeira faça empréstimos pessoais com um valor médio de R$ 3.500,00, com um desvio-padrão de R$ 1.200,00. Calcule as probabilidades:

a. Emprestar menos de R$ 3.100,00.

b. De a financeira emprestar mais de R$ 4.000,00.

c. Emprestar entre R$ 2.900,00 e R$ 3.800,00.

Resolução:

Temos que a média da distribuição é igual a 3.500 ($\mu_x = 3.500$) e que o desvio-padrão é igual a 1.200 ($\sigma_x = 1.200$).

Como temos a média diferente de zero e o desvio-padrão é diferente de 1, precisamos primeiramente padronizar a variável para que possamos olhar na tabela Normal Padrão e encontremos as probabilidades correspondentes.

Cabe destacar que a tabela nos fornece os valores de $-\infty$ até o ponto Z e, nesse caso, queremos a probabilidade de X ser menor do que 3.100. Assim sendo, já estamos no formato necessário para observar na tabela da distribuição Normal Padrão após a padronização. Precisamos, então, partir para a padronização.

$$z = \frac{X - \mu_x}{\sigma_x}$$

a. Temos que:

$$X = 3.100$$
$$\mu_x = 3.500$$
$$\sigma_x = 1.200$$

$$P(X < 3.100) = P\left(Z < \frac{3.100 - 3.500}{1.200}\right) = P(Z < -0,33)$$

Nesse ponto, precisamos olhar o valor -0,33 na tabela. Esse é o nosso valor padronizado (Z), lembrando que ele serve para encontrarmos a probabilidade associada a esse ponto na tabela de distribuição Normal Padrão.

Figura 3.11: Probabilidades da curva normal padrão.

z	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,6	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,4	0,0002	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003
-3,3	0,0003	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0005
-3,2	0,0005	0,0005	0,0005	0,0006	0,0006	0,0006	0,0006	0,0006
-3,1	0,0007	0,0007	0,0008	0,0008	0,0008	0,0008	0,0009	0,0009
-3,0	0,0010	0,0010	0,0011	0,0011	0,0011	0,0012	0,0012	0,0013
-2,9	0,0014	0,0014	0,0015	0,0015	0,0016	0,0016	0,0017	0,0018
-2,8	0,0019	0,0020	0,0021	0,0021	0,0022	0,0023	0,0023	0,0024
-2,7	0,0026	0,0027	0,0028	0,0029	0,0030	0,0031	0,0032	0,0033
-2,6	0,0036	0,0037	0,0038	0,0039	0,0040	0,0041	0,0043	0,0044
-2,5	0,0048	0,0049	0,0051	0,0052	0,0054	0,0055	0,0057	0,0059
-2,4	0,0064	0,0066	0,0068	0,0069	0,0071	0,0073	0,0075	0,0078
-2,3	0,0084	0,0087	0,0089	0,0091	0,0094	0,0096	0,0099	0,0102
-2,2	0,0110	0,0113	0,0116	0,0119	0,0122	0,0125	0,0129	0,0132
-2,1	0,0143	0,0146	0,0150	0,0154	0,0158	0,0162	0,0166	0,0170
-2,0	0,0183	0,0188	0,0192	0,0197	0,0202	0,0207	0,0212	0,0217
-1,9	0,0233	0,0239	0,0244	0,0250	0,0256	0,0262	0,0268	0,0274
-1,8	0,0294	0,0301	0,0307	0,0314	0,0322	0,0329	0,0336	0,0344
-1,7	0,0367	0,0375	0,0384	0,0392	0,0401	0,0409	0,0418	0,0427
-1,6	0,0455	0,0465	0,0475	0,0485	0,0495	0,0505	0,0516	0,0526
-1,5	0,0559	0,0571	0,0582	0,0594	0,0606	0,0618	0,0630	0,0643
-1,4	0,0681	0,0694	0,0708	0,0721	0,0735	0,0749	0,0764	0,0778
-1,3	0,0823	0,0838	0,0853	0,0869	0,0885	0,0901	0,0918	0,0934
-1,2	0,0985	0,1003	0,1020	0,1038	0,1056	0,1075	0,1093	0,1112
-1,1	0,1170	0,1190	0,1210	0,1230	0,1251	0,1271	0,1292	0,1314
-1,0	0,1379	0,1401	0,1423	0,1446	0,1469	0,1492	0,1515	0,1539
-0,9	0,1611	0,1635	0,1660	0,1685	0,1711	0,1736	0,1762	0,1788
-0,8	0,1867	0,1894	0,1922	0,1949	0,1977	0,2005	0,2033	0,2061
-0,7	0,2148	0,2177	0,2206	0,2236	0,2266	0,2296	0,2327	0,2358
-0,6	0,2451	0,2483	0,2514	0,2546	0,2578	0,2611	0,2643	0,2676
-0,5	0,2776	0,2810	0,2843	0,2877	0,2912	0,2946	0,2981	0,3015
-0,4	0,3121	0,3156	0,3192	0,3228	0,3264	0,3300	0,3336	0,3372
-0,3	0,3483	0,3520	0,3557	0,3594	0,3632	0,3669	0,3707	0,3745
-0,2	0,3859	0,3897	0,3936	0,3974	0,4013	0,4052	0,4090	0,4129
-0,1	0,4247	0,4286	0,4325	0,4364	0,4404	0,4443	0,4483	0,4522

Fonte: Elaborada pela autora (2016).

Encontramos, então, a probabilidade de 0,3707. Terminando os cálculos, temos:

$$P(X < 3.100) = P\ (Z < -0,33) = 0,3707$$

Podemos deixar o valor dessa maneira, ou então multiplicar por 100, para transformarmos em um valor percentual.

Então, a probabilidade de a financeira emprestar menos de R$ 3.100,00 é de 37,07%.

b. $P(X > 4.000) = 1 - P\ (X < 4.000)$

Observe que, primeiramente, deixamos em um formato que podemos conferir na tabela, conforme vimos anteriormente, já que o formato que a tabela foi construída é de $-\infty$ até o ponto Z. Assim, utilizamos a propriedade complementar. Sabemos que a área total abaixo da curva soma 1, logo, se temos o valor na tabela para o ponto menor do que Z, podemos, então, considerar 1 – essa probabilidade tabelada e obter o nosso resultado.

Agora, precisamos padronizar e observar a tabela.

Temos:

$$X = 4.000$$
$$\mu_x = 3.500$$
$$\sigma_x = 1.200$$

$$P(X > 4.000) = 1 - P\ (X < 4.000) = 1 - P\left(z < \frac{4.000 - 3.500}{1.200}\right) = 1 - P\ (Z < 0{,}42)$$

Nesse ponto, precisamos olhar o valor 0,42 na tabela antes de fazer qualquer outra operação. Olhamos na tabela para poder encontrar a probabilidade associada ao ponto 0,42.

Figura 3.12: Probabilidades da curva normal padrão.

z	0,00	0,01	0,02	0,03	0,04
0,0	0,5000	0,5040	0,5080	0,5120	0,5160
0,1	0,5398	0,5438	0,5478	0,5517	0,5557
0,2	0,5793	0,5832	0,5871	0,5910	0,5948
0,3	0,6179	0,6217	0,6255	0,6293	0,6331
0,4	0,6554	0,6591	0,6628	0,6664	0,6700
0,5	0,6915	0,6950	0,6985	0,7019	0,7054

Fonte: Elaborada pela autora (2016).

Somente após encontrarmos a probabilidade na tabela é que podemos continuar o cálculo.

Assim, terminando os cálculos, temos:

$$P(X > 4.000) = 1 - P\ (Z < 0{,}42) = 1 - 0{,}6628 = 0{,}3372$$

Então, a probabilidade de a financeira emprestar mais de R$ 4.000,00 é de 0,3372, ou, em termos percentuais, de 33,72%.

c. $P(2.900 < X < 3.800) = P(X < 3.800) - P(X < 2.900)$

Aqui, queremos calcular a probabilidade de um intervalo. Observe que novamente deixamos em um formato que podemos observar na tabela, conforme estudamos anteriormente. Precisamos, assim, padronizar dois valores referentes ao limite inferior e ao superior e, em seguida, olhar a tabela com as probabilidades referentes a esses valores, subtraindo um do outro.

Agora, precisamos padronizar e observar a tabela.

Assim, temos:

$$X_1 = 2.900$$
$$X_2 = 3.800$$
$$\mu_x = 3.500$$
$$\sigma_x = 1.200$$

$$P(2.900 < X < 3.800) = P(X < 3.800) - P(X < 2.900) =$$
$$= P\left(Z < \frac{3.800 - 3.500}{1.200}\right) - P\left(Z < \frac{2.900 - 3.500}{1.200}\right) = P(Z < 0{,}25) - P(Z < -0{,}50)$$

Precisamos, agora, antes de qualquer operação, observar os valores 0,25 e -0,50 na tabela.

Figura 3.13: Probabilidades da curva normal padrão.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123

Fonte: Elaborada pela autora (2016).

Figura 3.14: Probabilidades da curva normal padrão.

z	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	0,01	0,00
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,6	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,4	0,0002	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003
-3,3	0,0003	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0005	0,0005	0,0005
-3,2	0,0005	0,0005	0,0005	0,0006	0,0006	0,0006	0,0006	0,0006	0,0007	0,0007
-3,1	0,0007	0,0007	0,0008	0,0008	0,0008	0,0008	0,0009	0,0009	0,0009	0,0010
-3,0	0,0010	0,0010	0,0011	0,0011	0,0011	0,0012	0,0012	0,0013	0,0013	0,0013
-2,9	0,0014	0,0014	0,0015	0,0015	0,0016	0,0016	0,0017	0,0018	0,0018	0,0019
-2,8	0,0019	0,0020	0,0021	0,0021	0,0022	0,0023	0,0023	0,0024	0,0025	0,0026
-2,7	0,0026	0,0027	0,0028	0,0029	0,0030	0,0031	0,0032	0,0033	0,0034	0,0035
-2,6	0,0036	0,0037	0,0038	0,0039	0,0040	0,0041	0,0043	0,0044	0,0045	0,0047
-2,5	0,0048	0,0049	0,0051	0,0052	0,0054	0,0055	0,0057	0,0059	0,0060	0,0062
-2,4	0,0064	0,0066	0,0068	0,0069	0,0071	0,0073	0,0075	0,0078	0,0080	0,0082
-2,3	0,0084	0,0087	0,0089	0,0091	0,0094	0,0096	0,0099	0,0102	0,0104	0,0107
-2,2	0,0110	0,0113	0,0116	0,0119	0,0122	0,0125	0,0129	0,0132	0,0136	0,0139
-2,1	0,0143	0,0146	0,0150	0,0154	0,0158	0,0162	0,0166	0,0170	0,0174	0,0179
-2,0	0,0183	0,0188	0,0192	0,0197	0,0202	0,0207	0,0212	0,0217	0,0222	0,0228
-1,9	0,0233	0,0239	0,0244	0,0250	0,0256	0,0262	0,0268	0,0274	0,0281	0,0287
-1,8	0,0294	0,0301	0,0307	0,0314	0,0322	0,0329	0,0336	0,0344	0,0351	0,0359
-1,7	0,0367	0,0375	0,0384	0,0392	0,0401	0,0409	0,0418	0,0427	0,0436	0,0446
-1,6	0,0455	0,0465	0,0475	0,0485	0,0495	0,0505	0,0516	0,0526	0,0537	0,0548
-1,5	0,0559	0,0571	0,0582	0,0594	0,0606	0,0618	0,0630	0,0643	0,0655	0,0668
-1,4	0,0681	0,0694	0,0708	0,0721	0,0735	0,0749	0,0764	0,0778	0,0793	0,0808
-1,3	0,0823	0,0838	0,0853	0,0869	0,0885	0,0901	0,0918	0,0934	0,0951	0,0968
-1,2	0,0985	0,1003	0,1020	0,1038	0,1056	0,1075	0,1093	0,1112	0,1131	0,1151
-1,1	0,1170	0,1190	0,1210	0,1230	0,1251	0,1271	0,1292	0,1314	0,1335	0,1357
-1,0	0,1379	0,1401	0,1423	0,1446	0,1469	0,1492	0,1515	0,1539	0,1562	0,1587
-0,9	0,1611	0,1635	0,1660	0,1685	0,1711	0,1736	0,1762	0,1788	0,1814	0,1841
-0,8	0,1867	0,1894	0,1922	0,1949	0,1977	0,2005	0,2033	0,2061	0,2090	0,2119
-0,7	0,2148	0,2177	0,2206	0,2236	0,2266	0,2296	0,2327	0,2358	0,2389	0,2420
-0,6	0,2451	0,2483	0,2514	0,2546	0,2578	0,2611	0,2643	0,2676	0,2709	0,2743
-0,5	0,2776	0,2810	0,2843	0,2877	0,2912	0,2946	0,2981	0,3015	0,3050	0,3085
-0,4	0,3121	0,3156	0,3192	0,3228	0,3264	0,3300	0,3336	0,3372	0,3409	0,3446

Fonte: Elaborada pela autora (2016).

Somente depois de encontrar as probabilidades referentes aos dois pontos na tabela é que podemos subtrair um do outro. Assim, utilizamos a probabilidade associada ao limite superior e a calculamos, subtraída ao valor da probabilidade associada ao limite inferior.

Assim, terminando os cálculos, temos:

$$P(2.900 < X < 3.800) = P(Z < 0,25) - P(Z < -0,50) = 0,5987 - 0,3085 = 0,2902$$

Então, a probabilidade de a financeira emprestar entre R$ 2.900,00 e R$ 3.800,00 é de 29,02%.

Exemplo 3.2: Um fabricante de sapatos tem uma produção diária normalmente distribuída, com média de 1.000 pares e desvio-padrão de 100 pares. O gerente pretende implantar uma bonificação aos funcionários toda vez que a produção ultrapassar 1.150 pares. Qual a probabilidade de a empresa não pagar essa bonificação?

Resolução:

Precisamos considerar que aqui está sendo pedida a probabilidade de a empresa NÃO pagar a bonificação. Então, isso ocorrerá quando a produção não ultrapassar 1.150 pares, ou seja, com uma produção inferior a 1.150.

Assim, temos:

$$X = 1.150$$
$$\mu_x = 1.000$$
$$\sigma_x = 100$$

Já estamos com o formato necessário para poder padronizar e conferir na tabela. Dessa forma, utilizamos a padronização, procuramos o valor de Z na tabela e encontramos a probabilidade associada a esse valor.

$$z = \frac{X - \mu_x}{\sigma_x}$$

Assim:

$$P(X < 1.150) = P\left(Z < \frac{1.150 - 1.000}{100}\right) = P\ (Z < 1,50)$$

Então, observamos o valor 1,50 na tabela.

Figura 3.15: Probabilidades da curva normal padrão.

z	0,00	0,01	0,02
0,0	0,5000	0,5040	0,5080
0,1	0,5398	0,5438	0,5478
0,2	0,5793	0,5832	0,5871
0,3	0,6179	0,6217	0,6255
0,4	0,6554	0,6591	0,6628
0,5	0,6915	0,6950	0,6985
0,6	0,7257	0,7291	0,7324
0,7	0,7580	0,7611	0,7642
0,8	0,7881	0,7910	0,7939
0,9	0,8159	0,8186	0,8212
1,0	0,8413	0,8438	0,8461
1,1	0,8643	0,8665	0,8686
1,2	0,8849	0,8869	0,8888
1,3	0,9032	0,9049	0,9066
1,4	0,9192	0,9207	0,9222
1,5	0,9332	0,9345	0,9357
1,6	0,9452	0,9463	0,9474
1,7	0,9554	0,9564	0,9573
1,8	0,9641	0,9649	0,9656

Fonte: Elaborada pela autora (2016).

Ao fazermos os cálculos, temos que:

$$P(X < 1.150) = P(Z < 1,50) = 0,9332$$

Então, a probabilidade de a empresa não pagar a bonificação é de 93,32%.

Exemplo 3.3: Uma fábrica de televisores sabe que os equipamentos de sua fábrica tem duração normal com média de 150.000 horas e desvio-padrão de 5.000 horas. Qual a probabilidade de que um televisor, escolhido ao acaso, dure:

a. menos de 170.000 horas?
b. entre 140.000 e 160.000 horas?
c. se o fabricante deseja oferecer uma garantia, tal que ele tenha que substituir no máximo 1% dos televisores, qual deve ser o valor desta garantia?

Resolução:

a. Temos:

$$X = 170.000$$
$$\mu_x = 150.000$$
$$\sigma_x = 5.000$$

Utilizamos a seguinte padronização:

$$z = \frac{X - \mu_x}{\sigma_x}$$

$$P(X < 170.000) = P\left(Z < \frac{170.000 - 150.000}{5.000}\right) = P\,(Z < 4,00)$$

Aqui, já podemos padronizar e conferir na tabela, pois está na forma que a tabela fornece as probabilidades.

Figura 3.16: Probabilidades da curva normal padrão.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Fonte: Elaborada pela autora (2016).

Observe que o valor 4,00 não está na tabela, mas o valor 3,99 já chegou ao 1 (ou seja, os 100% da área abaixo da curva), então todos os valores que ultrapassarem esse número serão iguais a 1,000. Então:

$$P(X < 170.000) = P\left(Z < \frac{170.000 - 150.000}{5.000}\right) = P(Z < 4,00) = 1,0000$$

Dessa forma, a probabilidade do televisor durar menos de 170.000 horas é de 100%.

b. Temos, assim:

$$X_1 = 140.000$$
$$X_2 = 160.000$$
$$\mu_x = 150.000$$
$$\sigma_x = 5.000$$

Aqui, precisamos calcular a probabilidade de ser menor que o limite superior menos a probabilidade do limite inferior, como já fizemos anteriormente.

$$P(140.000 < X < 160.000) = P(X < 160.000) - P(X < 140.000) =$$

$$= P\left(Z < \frac{160.000 - 15.000}{5.000}\right) - P\left(Z < \frac{140.000 - 150.000}{5.000}\right) =$$
$$= P(Z < 2,00) - P(Z < -2,00)$$

Precisamos observar os dois valores na tabela antes de fazer qualquer operação matemática.

Figura 3.17: Probabilidades da curva normal padrão.

z	0,00	0,01	0,02	0,03
0,0	0,5000	0,5040	0,5080	0,5120
0,1	0,5398	0,5438	0,5478	0,5517
0,2	0,5793	0,5832	0,5871	0,5910
0,3	0,6179	0,6217	0,6255	0,6293
0,4	0,6554	0,6591	0,6628	0,6664
0,5	0,6915	0,6950	0,6985	0,7019
0,6	0,7257	0,7291	0,7324	0,7357
0,7	0,7580	0,7611	0,7642	0,7673
0,8	0,7881	0,7910	0,7939	0,7967
0,9	0,8159	0,8186	0,8212	0,8238
1,0	0,8413	0,8438	0,8461	0,8485
1,1	0,8643	0,8665	0,8686	0,8708
1,2	0,8849	0,8869	0,8888	0,8907
1,3	0,9032	0,9049	0,9066	0,9082
1,4	0,9192	0,9207	0,9222	0,9236
1,5	0,9332	0,9345	0,9357	0,9370
1,6	0,9452	0,9463	0,9474	0,9484
1,7	0,9554	0,9564	0,9573	0,9582
1,8	0,9641	0,9649	0,9656	0,9664
1,9	0,9713	0,9719	0,9726	0,9732
2,0	0,9772	0,9778	0,9783	0,9788
2,1	0,9821	0,9826	0,9830	0,9834
2,2	0,9861	0,9864	0,9868	0,9871

Fonte: Elaborada pela autora (2016).

Figura 3.18: Probabilidades da curva normal padrão.

z	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	0,01	0,00
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,6	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,4	0,0002	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003
-3,3	0,0003	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0005	0,0005	0,0005
-3,2	0,0005	0,0005	0,0005	0,0006	0,0006	0,0006	0,0006	0,0006	0,0007	0,0007
-3,1	0,0007	0,0007	0,0008	0,0008	0,0008	0,0008	0,0009	0,0009	0,0009	0,0010
-3,0	0,0010	0,0010	0,0011	0,0011	0,0011	0,0012	0,0012	0,0013	0,0013	0,0013
-2,9	0,0014	0,0014	0,0015	0,0015	0,0016	0,0016	0,0017	0,0018	0,0018	0,0019
-2,8	0,0019	0,0020	0,0021	0,0021	0,0022	0,0023	0,0023	0,0024	0,0025	0,0026
-2,7	0,0026	0,0027	0,0028	0,0029	0,0030	0,0031	0,0032	0,0033	0,0034	0,0035
-2,6	0,0036	0,0037	0,0038	0,0039	0,0040	0,0041	0,0043	0,0044	0,0045	0,0047
-2,5	0,0048	0,0049	0,0051	0,0052	0,0054	0,0055	0,0057	0,0059	0,0060	0,0062
-2,4	0,0064	0,0066	0,0068	0,0069	0,0071	0,0073	0,0075	0,0078	0,0080	0,0082
-2,3	0,0084	0,0087	0,0089	0,0091	0,0094	0,0096	0,0099	0,0102	0,0104	0,0107
-2,2	0,0110	0,0113	0,0116	0,0119	0,0122	0,0125	0,0129	0,0132	0,0136	0,0139
-2,1	0,0143	0,0146	0,0150	0,0154	0,0158	0,0162	0,0166	0,0170	0,0174	0,0179
-2,0	0,0183	0,0188	0,0192	0,0197	0,0202	0,0207	0,0212	0,0217	0,0222	0,0228
-1,9	0,0233	0,0239	0,0244	0,0250	0,0256	0,0262	0,0268	0,0274	0,0281	0,0287
-1,8	0,0294	0,0301	0,0307	0,0314	0,0322	0,0329	0,0336	0,0344	0,0351	0,0359

Fonte: Elaborada pela autora (2016).

Note que a soma dos dois valores é igual a 1, e isso se deve à propriedade de simetria em torno da média na Curva Normal.

Ao continuarmos os cálculos, temos:

$$P(140.000 < X < 160.000) = P(Z < 2,00) - P(Z < -2,00) = 0,9772 - 0,0228 = 0,9544$$

Então, a probabilidade do televisor durar entre 140.000 e 160.000 horas é de 95,44%.

c. $P(X < G) = 0.01$

Aqui, já temos como resultado a probabilidade, então precisamos fazer o caminho inverso: conferir na tabela e resolver a padronização com a incógnita no lugar do x para encontrar o tempo para garantia.

Figura 3.19: Probabilidades da curva normal padrão.

z	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	0,01	0,00
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,6	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,4	0,0002	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003
-3,3	0,0003	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0005	0,0005	0,0005
-3,2	0,0005	0,0005	0,0005	0,0006	0,0006	0,0006	0,0006	0,0006	0,0007	0,0007
-3,1	0,0007	0,0007	0,0008	0,0008	0,0008	0,0008	0,0009	0,0009	0,0009	0,0010
-3,0	0,0010	0,0010	0,0011	0,0011	0,0011	0,0012	0,0012	0,0013	0,0013	0,0013
-2,9	0,0014	0,0014	0,0015	0,0015	0,0016	0,0016	0,0017	0,0018	0,0018	0,0019
-2,8	0,0019	0,0020	0,0021	0,0021	0,0022	0,0023	0,0023	0,0024	0,0025	0,0026
-2,7	0,0026	0,0027	0,0028	0,0029	0,0030	0,0031	0,0032	0,0033	0,0034	0,0035
-2,6	0,0036	0,0037	0,0038	0,0039	0,0040	0,0041	0,0043	0,0044	0,0045	0,0047
-2,5	0,0048	0,0049	0,0051	0,0052	0,0054	0,0055	0,0057	0,0059	0,0060	0,0062
-2,4	0,0064	0,0066	0,0068	0,0069	0,0071	0,0073	0,0075	0,0078	0,0080	0,0082
-2,3	0,0084	0,0087	0,0089	0,0091	0,0094	0,0096	0,0099	0,0102	0,0104	0,0107
-2,2	0,0110	0,0113	0,0116	0,0119	0,0122	0,0125	0,0129	0,0132	0,0136	0,0139
-2,1	0,0143	0,0146	0,0150	0,0154	0,0158	0,0162	0,0166	0,0170	0,0174	0,0179
-2,0	0,0183	0,0188	0,0192	0,0197	0,0202	0,0207	0,0212	0,0217	0,0222	0,0228
-1,9	0,0233	0,0239	0,0244	0,0250	0,0256	0,0262	0,0268	0,0274	0,0281	0,0287
-1,8	0,0294	0,0301	0,0307	0,0314	0,0322	0,0329	0,0336	0,0344	0,0351	0,0359
-1,7	0,0367	0,0375	0,0384	0,0392	0,0401	0,0409	0,0418	0,0427	0,0436	0,0446
-1,6	0,0455	0,0465	0,0475	0,0485	0,0495	0,0505	0,0516	0,0526	0,0537	0,0548
-1,5	0,0559	0,0571	0,0582	0,0594	0,0606	0,0618	0,0630	0,0643	0,0655	0,0668
-1,4	0,0681	0,0694	0,0708	0,0721	0,0735	0,0749	0,0764	0,0778	0,0793	0,0808

Fonte: Elaborada pela autora (2016).

$$P(X < G) = P(Z < \left(\frac{G - 150.000}{5.000}\right) = 0,01 = P(Z < -2,33)$$

Então:

$$\left(\frac{G - 150.000}{5.000}\right) = -2,33 \quad \text{isolando o G (tempo de garantia):}$$

$$G = (-2,33 \cdot 5.000) + 150.000 = 138.350 \text{ horas}$$

A garantia deve ser de 138.350 horas para substituir, no máximo, 1% dos televisores.

Assim, finalizamos mais uma unidade em que conhecemos uma importante forma de análise estatística: a distribuição Normal. Na unidade seguinte, daremos continuidade ao assunto com a Correlação e Regressão. Até lá!

Áreas sob a curva normal
(Cauda da esquerda)

z	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	0,01	0,00
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,6	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,4	0,0002	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003
-3,3	0,0003	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0005	0,0005	0,0005
-3,2	0,0005	0,0005	0,0005	0,0006	0,0006	0,0006	0,0006	0,0006	0,0007	0,0007
-3,1	0,0007	0,0007	0,0008	0,0008	0,0008	0,0008	0,0009	0,0009	0,0009	0,0010
-3,0	0,0010	0,0010	0,0011	0,0011	0,0011	0,0012	0,0012	0,0013	0,0013	0,0013
-2,9	0,0014	0,0014	0,0015	0,0015	0,0016	0,0016	0,0017	0,0018	0,0018	0,0019
-2,8	0,0019	0,0020	0,0021	0,0021	0,0022	0,0023	0,0023	0,0024	0,0025	0,0026
-2,7	0,0026	0,0027	0,0028	0,0029	0,0030	0,0031	0,0032	0,0033	0,0034	0,0035
-2,6	0,0036	0,0037	0,0038	0,0039	0,0040	0,0041	0,0043	0,0044	0,0045	0,0047
-2,5	0,0048	0,0049	0,0051	0,0052	0,0054	0,0055	0,0057	0,0059	0,0060	0,0062
-2,4	0,0064	0,0066	0,0068	0,0069	0,0071	0,0073	0,0075	0,0078	0,0080	0,0082
-2,3	0,0084	0,0087	0,0089	0,0091	0,0094	0,0096	0,0099	0,0102	0,0104	0,0107
-2,2	0,0110	0,0113	0,0116	0,0119	0,0122	0,0125	0,0129	0,0132	0,0136	0,0139
-2,1	0,0143	0,0146	0,0150	0,0154	0,0158	0,0162	0,0166	0,0170	0,0174	0,0179
-2,0	0,0183	0,0188	0,0192	0,0197	0,0202	0,0207	0,0212	0,0217	0,0222	0,0228
-1,9	0,0233	0,0239	0,0244	0,0250	0,0256	0,0262	0,0268	0,0274	0,0281	0,0287
-1,8	0,0294	0,0301	0,0307	0,0314	0,0322	0,0329	0,0336	0,0344	0,0351	0,0359
-1,7	0,0367	0,0375	0,0384	0,0392	0,0401	0,0409	0,0418	0,0427	0,0436	0,0446
-1,6	0,0455	0,0465	0,0475	0,0485	0,0495	0,0505	0,0516	0,0526	0,0537	0,0548
-1,5	0,0559	0,0571	0,0582	0,0594	0,0606	0,0618	0,0630	0,0643	0,0655	0,0668
-1,4	0,0681	0,0694	0,0708	0,0721	0,0735	0,0749	0,0764	0,0778	0,0793	0,0808
-1,3	0,0823	0,0838	0,0853	0,0869	0,0885	0,0901	0,0918	0,0934	0,0951	0,0968
-1,2	0,0985	0,1003	0,1020	0,1038	0,1056	0,1075	0,1093	0,1112	0,1131	0,1151
-1,1	0,1170	0,1190	0,1210	0,1230	0,1251	0,1271	0,1292	0,1314	0,1335	0,1357
-1,0	0,1379	0,1401	0,1423	0,1446	0,1469	0,1492	0,1515	0,1539	0,1562	0,1587
-0,9	0,1611	0,1635	0,1660	0,1685	0,1711	0,1736	0,1762	0,1788	0,1814	0,1841
-0,8	0,1867	0,1894	0,1922	0,1949	0,1977	0,2005	0,2033	0,2061	0,2090	0,2119
-0,7	0,2148	0,2177	0,2206	0,2236	0,2266	0,2296	0,2327	0,2358	0,2389	0,2420
-0,6	0,2451	0,2483	0,2514	0,2546	0,2578	0,2611	0,2643	0,2676	0,2709	0,2743
-0,5	0,2776	0,2810	0,2843	0,2877	0,2912	0,2946	0,2981	0,3015	0,3050	0,3085
-0,4	0,3121	0,3156	0,3192	0,3228	0,3264	0,3300	0,3336	0,3372	0,3409	0,3446
-0,3	0,3483	0,3520	0,3557	0,3594	0,3632	0,3669	0,3707	0,3745	0,3783	0,3821
-0,2	0,3859	0,3897	0,3936	0,3974	0,4013	0,4052	0,4090	0,4129	0,4168	0,4207
-0,1	0,4247	0,4286	0,4325	0,4364	0,4404	0,4443	0,4483	0,4522	0,4562	0,4602
0,0	0,4641	0,4681	0,4721	0,4761	0,4801	0,4840	0,4880	0,4920	0,4960	0,5000

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

SÍNTESE DA UNIDADE

Nessa unidade você aprendeu:

- a calcular probabilidades utilizando a mais importante das distribuições, a distribuição Normal;
- a utilizar a tabela Normal, que é uma ferramenta importante para encontrar as probabilidades da distribuição Normal, uma vez que o cálculo da integral da probabilidade não é nada trivial.

REFERÊNCIAS

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística Aplicada e Probabilidade para Engenheiros**. 5. ed. São Paulo: LTC, 2012.

SPIEGEL, M. R.; SCHILLER, J. J.; SRINIVASAN, R. A. **Probabilidade e Estatística.** 3. ed. Porto Alegre: Bookman, 2015. (Coleção Schaum).

STEVENSON, W. J. **Estatística Aplicada à Administração**. São Paulo: Harbra, 2001.

ANOTAÇÕES

Análise de Correlação e Regressão

Prezado(a) estudante.

Estamos começando uma unidade desta disciplina. Os textos que a compõem foram organizados com cuidado e atenção, para que você tenha contato com um conteúdo completo e atualizado tanto quanto possível. Leia com dedicação, realize as atividades e tire suas dúvidas com os tutores. Dessa forma você, com certeza, alcançará os objetivos propostos para essa disciplina.

OBJETIVO GERAL

Aprender a calcular o coeficiente de correlação, o coeficiente de determinação e estimar a reta de regressão.

OBJETIVOS ESPECÍFICOS

- Propiciar o raciocínio para previsão de valores através da análise de regressão.
- Tornar inteligível a análise de correlação simples.
- Clarear a análise de diagramas de dispersão.

4.1 Correlação e Regressão

Muitas vezes estamos interessados em saber o quanto uma variável influencia no resultado final. Imagine a seguinte situação: em um banco, quero oferecer planos de poupança para pessoas físicas e preciso acertar em um valor que seja interessante e factível para o cliente. Como podemos fazer essa análise?

Outra situação, dentre tantas possíveis, é avaliar o desenvolvimento e crescimento de crianças. Por exemplo, como podemos avaliar se o peso de um bebê de dois meses está dentro de um valor esperado e considerado saudável?

São questões como essas que os assuntos que estudaremos a seguir nos permitirão responder.

4.1.1 Correlação Linear Simples

A correlação mede a relação conjunta entre duas ou mais variáveis. No caso do nosso curso, aprenderemos a correlação entre duas variáveis, que é designada como sendo correlação simples. No caso de termos mais de duas variáveis, designamos como sendo correlação múltipla.

Utilizaremos no nosso curso apenas as equações lineares, ou seja, quando a correlação simples pode ser explicada por uma reta.

Teremos, então, duas variáveis em estudo: a variável independente (ou explicativa) “ x ” e a variável dependente (ou variável explicada) “ y ”.

4.1.1.1 Coeficiente de Correlação de Pearson

A correlação é medida a partir de um coeficiente r_{xy} denominado coeficiente de correlação, que assume valores no intervalo entre -1 e 1.

O coeficiente de correlação mede a intensidade da correlação, pois quanto mais próximo de 1 ou de -1, maior a correlação entre as variáveis. Quando – ou quando tende a esse valor – o coeficiente demonstra que, provavelmente, não existe relação entre as variáveis pesquisadas.

O coeficiente de correlação, além de medir a intensidade, mede também a direção dessa correlação. Assim, sendo a correlação negativa, teremos uma correlação inversa, ou seja, conforme a variável X cresce, a variável Y decresce. Já se a correlação for positiva, ela será direta, ou seja, conforme a variável X cresce, a variável Y também cresce.

Tabela 4.1: Intensidade e direção do coeficiente de correlação.

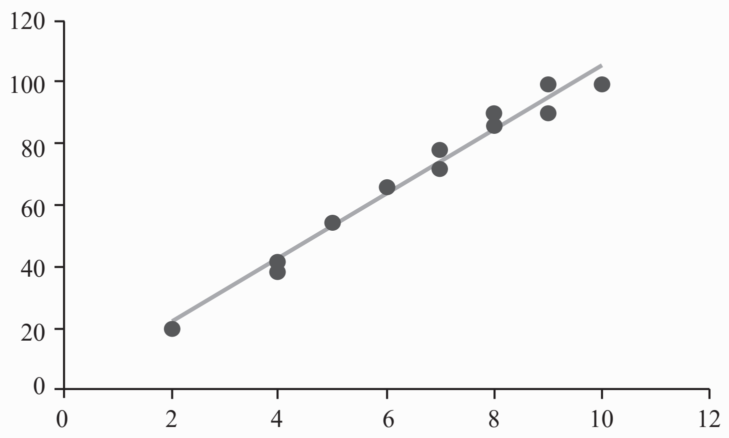
r_{xy}	Intensidade e direção
-1	Correlação perfeita negativa
-1--| -0,9	Correlação muito forte negativa
-0,9 --|-0,8	Correlação forte negativa
-0,8 --|0,6	Correlação razoável negativa
-0,6 --| -0,3	Correlação fraca negativa
-0,3 – 0,0	Correlação muito fraca negativa
0	Correlação nula
0 – 0,3	Correlação muito fraca positiva
0,3 |-- 0,6	Correlação fraca positiva
0,6 |-- 0,8	Correlação razoável positiva
0,8 |-- 0,9	Correlação forte positiva
0,9 |-- 1,0	Correlação muito forte positiva
1,00	Correlação perfeita positiva

Fonte: Elaborada pela autora (2016).

Segundo Martins (2012), o coeficiente de correlação é utilizado, principalmente, devido à vantagem que representa, que é decorrente da facilidade de interpretação e de seu intervalo compreender valores em um intervalo com uma escala reduzida.

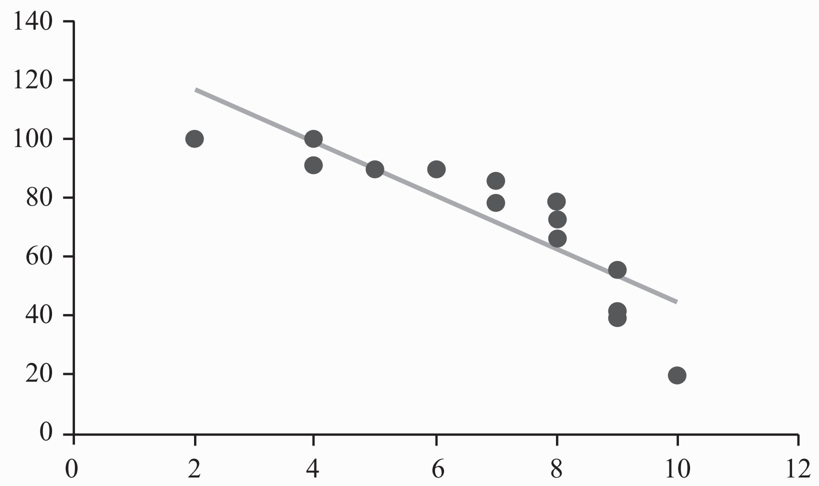
Nos gráficos a seguir, podemos ver modelos de correlação direta (ou positiva) ou correlação inversa (ou negativa). Lembrando que quando duas variáveis têm correlação positiva, isso significa dizer que, à medida que uma variável aumenta, a outra também aumenta, mesmo que em proporções diferentes. Por outro lado, quando a correlação é negativa, à medida que uma variável aumenta, a outra tende a diminuir.

Figura 4.1: Diagrama de dispersão.



Fonte: Elaborada pela autora (2016).

Figura 4.2: Diagrama de dispersão.



Fonte: Elaborada pela autora (2016).

Para calcularmos a correlação linear simples de Pearson, a fórmula é a seguinte:

$$r_{xy} = \frac{n \cdot \sum xy - \sum x \cdot \sum y}{\sqrt{\left[n \cdot \sum x^2 - \left(\sum x\right)^2\right] \cdot \left[n \cdot \sum y^2 - \left(\sum y\right)^2\right]}}$$

Os somatórios dessa fórmula resultam das duas variáveis observadas.

Exemplo 4.1: Suponha uma amostra de oito pessoas, na qual foi identificado o valor do salário e o valor economizado na poupança. Calcule o coeficiente de correlação para esses dados.

Tabela 4.2: Tabela de dados.

Salário	Poupança
1.250,00	990,00
880,00	600,00
5.000,00	2.250,00
2.700,00	1.500,00
1.000,00	900,00
4.200,00	1.880,00
1.600,00	1.100,00
3.500,00	2.000,00

Fonte: Elaborada pelo autor (2016).

Resolução:

Primeiramente, precisamos identificar a variável X e a variável Y. Nesse caso, esperamos que quanto maior o salário, maior seja o valor da poupança. Assim, a poupança depende de quanto a pessoa ganha de salário.

Identificado isso, precisamos fazer os somatórios para aplicar na fórmula.

Tabela 4.3: Tabela de somatórios.

Salário (x)	Poupança (y)	$x \cdot y$	x^2	y^2
1.250,00	990,00	$1.250 \cdot 990 = 1.237.500$	$1.250^2 = 1.562.500$	$990^2 = 980.100$
880,00	600,00	$880 \cdot 600 = 528.000$	$880^2 = 774.400$	$600^2 = 360.000$
5.000,00	2.250,00	$5.000 \cdot 2.250 = 11.250.000$	$5.000^2 = 25.000.000$	$2.250^2 = 5.062.500$
2.700,00	1.500,00	$2.700 \cdot 1.500 = 4.050.000$	$2.700^2 = 7.290.000$	$1.500^2 = 2.250.000$
1.000,00	900,00	$1.000 \cdot 900 = 900.000$	$1.000^2 = 1.000.000$	$900^2 = 810.000$
4.200,00	1.880,00	$4.200 \cdot 1.880 = 7.896.000$	$4.200^2 = 17.640.000$	$1.880^2 = 3.534.400$
1.600,00	1.100,00	$1.600 \cdot 1.100 = 1.760.000$	$1.600^2 = 2.560.000$	$1.100^2 = 1.210.000$
3.500,00	2.000,00	$3.500 \cdot 2.000 = 7.000.000$	$3.500^2 = 12.250.000$	$2.000^2 = 4.000.000$
20.130	11.220	34.621.500	68.076.900	18.207.000

Fonte: Elaborada pela autora (2016).

Note que os somatórios estão apresentados na última linha da tabela anterior. Agora, com os somatórios prontos, podemos calcular a fórmula do coeficiente de correlação.

$$r_{xy} = \frac{n \cdot \sum xy - \sum x \cdot \sum y}{\sqrt{\left[n \cdot \sum x^2 - \left(\sum x\right)^2\right] \cdot \left[n \cdot \sum y^2 - \left(\sum y\right)^2\right]}} =$$

$$= \frac{8 \cdot 34.621.500 - 201.30 \cdot 11.220}{\sqrt{\left[8 \cdot 68.076.900 - (20.130)^2\right] \cdot \left[8 \cdot 18.207.000 - (11.220)^2\right]}} =$$

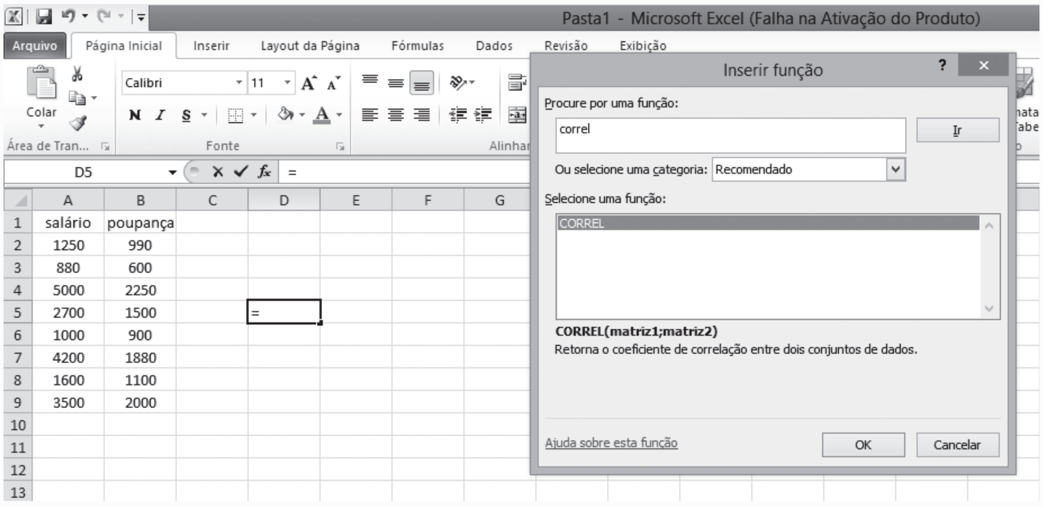
$$= \frac{51.113.400}{\sqrt{(139.398.300).(19.767.600)}} = 0{,}9737$$

Com esse resultado, podemos concluir que a correlação é muito forte e positiva (ver Tabela 4.1), ou seja, quanto maior o valor do salário, maior o valor investido na poupança.

Podemos calcular o coeficiente de correlação com o auxílio do Excel. Uma das maneiras de calcular, é clicar no ícone “inserir função” (f_x) e digitar a palavra “correl”.

Ao final desta unidade, apresentarei a outra maneira de utilizar o Excel para análise de correlação e regressão.

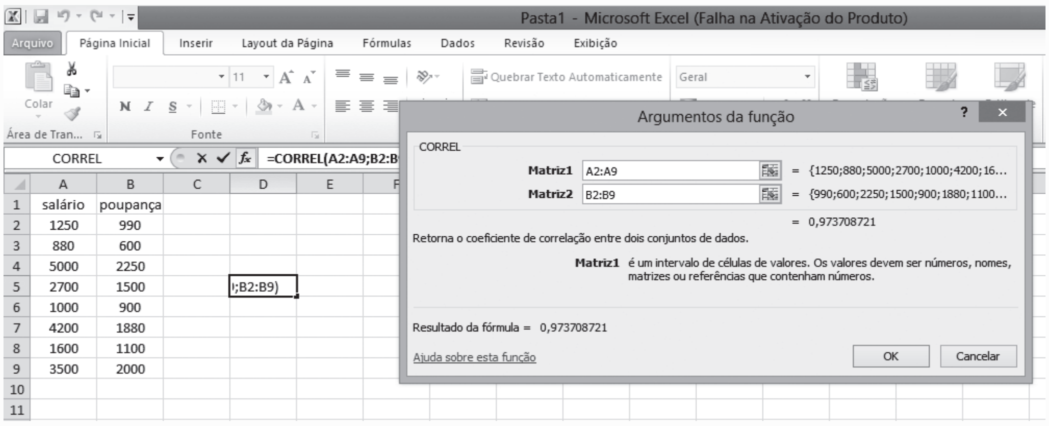
Figura 4.3: Função “correl” do Excel.



Fonte: Elaborada pela autora (2016).

Clicando em “OK”, podemos selecionar os dados. A matriz 1 é a variável x e a matriz 2 é a variável y .

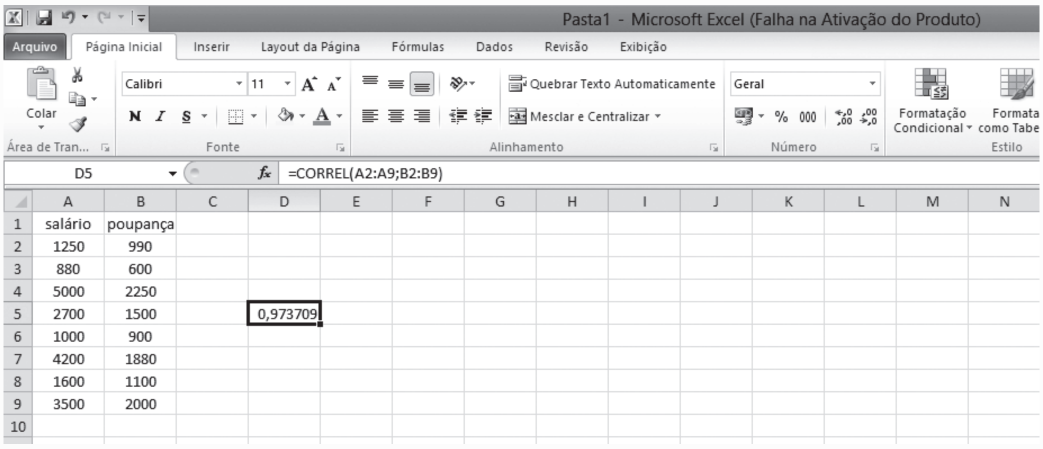
Figura 4.4: Função “correl” do Excel.



Fonte: Elaborada pela autora (2016).

Depois de selecionados os dados, clicamos em “OK” e obtemos o mesmo resultado que calculamos com a fórmula anteriormente.

Figura 4.5: Função "correl" do Excel.



Fonte: Elaborada pela autora (2016).

Em algumas situações, podemos ficar em dúvida se a correlação realmente é válida, se podemos realmente afirmar que as duas variáveis estão correlacionadas. Para isso, podemos fazer um teste de hipóteses para validar essa correlação e continuar com as análises, como veremos a seguir.

4.1.1.2 Teste para o Coeficiente de Correlação de Pearson

Vamos inserir aqui o conceito de teste de hipóteses, exemplificando esse teste específico. A teoria de teste de hipóteses está ligada à probabilidade.

Assim, utilizaremos a tabela t, que tem formato semelhante à tabela Normal Padrão, já estudada na unidade anterior. Quando n (número de elementos) tende ao infinito, a tabela t tem os mesmos valores encontrados na tabela Normal Padrão.

Vamos, agora, ao teste. Nesse caso, precisamos testar a correlação entre as variáveis. As hipóteses formuladas são as seguintes:

$H_0 : \rho = 0$ não existe correlação entre as variáveis;

$H_0 : \rho \neq 0$ existe correlação entre as variáveis.

Formuladas as hipóteses, calculamos a estatística de teste, um valor a ser verificado como significativo ou não.

A estatística de teste para essa situação específica é:

$$t_{calculado} = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$$

Quando o valor dessa conta, que é a nossa estatística de teste, for maior que o valor tabelado na tabela t, então a decisão é rejeitar H_0. Se a estatística de teste for menor que o valor tabelado, não rejeitamos H_0 com a significância (α) estabelecida, conforme dados a seguir:

Se $t_{calculado} > t_{tabelado} \rightarrow$ rejeitamos H_0

Se $t_{calculado} < t_{tabelado} \rightarrow$ não rejeitamos H_0

Para olharmos na tabela t precisamos do nível de significância que está nas colunas da tabela e dos graus de liberdade (GL). No caso desse teste de hipóteses, os graus de liberdade são obtidos por $GL = n-2$. A tabela encontra-se ao final dessa unidade.

Observe, então, que nesse teste desejamos sempre rejeitar H_0, pois desejamos que exista a correlação entre as variáveis.

Exemplo 4.2: (utilizando os dados do Exemplo 4.1): Suponha uma amostra de 8 pessoas, na qual foi identificado o valor do salário e o valor economizado na poupança. Teste a hipótese de existência da correlação, com 5% de significância.

Resolução:
Como primeiro passo, devemos formular as hipóteses.

- H_0: não existe correlação entre as variáveis;
- H_1: existe correlação entre as variáveis.

Em seguida, devemos calcular a estatística de teste:

$$t_{calculado} = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,9737 \cdot \sqrt{8-2}}{\sqrt{1-0,9737^2}} = \frac{2,3850}{0,2278} = 10,47$$

Agora, precisamos encontrar o valor tabelado. Temos $\alpha = 5\% = 0,05$. Nesse caso, temos um teste bilateral, então dividimos $\alpha / 2$, assim, $\alpha / 2 = 0,025$ e os graus de liberdade $GL = n-2 = 8-2 = 6$.

Figura 4.6: Tabela *t – Student*.

	Nível de significância - alfa					
GL	**0,250**	**0,100**	**0,050**	**0,025**	**0,010**	**0,005**
1	1,000	3,078	6,314	12,706	31,821	63,657
2	0,816	1,886	2,920	4,303	6,965	9,925
3	0,765	1,638	2,353	3,182	4,541	5,841
4	0,741	1,533	2,132	2,776	3,747	4,604
5	0,727	1,476	2,015	2,571	3,365	4,032
6	0,718	1,440	1,943	2,447	3,143	3,707
7	0,711	1,415	1,895	2,365	2,998	3,499
8	0,706	1,397	1,860	2,306	2,896	3,355
9	0,703	1,383	1,833	2,262	2,821	3,250
10	0,700	1,372	1,812	2,228	2,764	3,169
11	0,697	1,363	1,796	2,201	2,718	3,106
12	0,695	1,356	1,782	2,179	2,681	3,055
13	0,694	1,350	1,771	2,160	2,650	3,012
14	0,692	1,345	1,761	2,145	2,624	2,977
15	0,691	1,341	1,753	2,131	2,602	2,947
16	0,690	1,337	1,746	2,120	2,583	2,921
17	0,689	1,333	1,740	2,110	2,567	2,898
18	0,688	1,330	1,734	2,101	2,552	2,878
19	0,688	1,328	1,729	2,093	2,539	2,861
20	0,687	1,325	1,725	2,086	2,528	2,845
21	0,686	1,323	1,721	2,080	2,518	2,831
22	0,686	1,321	1,717	2,074	2,508	2,819
23	0,685	1,319	1,714	2,069	2,500	2,807
24	0,685	1,318	1,711	2,064	2,492	2,797
25	0,684	1,316	1,708	2,060	2,485	2,787
26	0,684	1,315	1,706	2,056	2,479	2,779
27	0,684	1,314	1,703	2,052	2,473	2,771
28	0,683	1,313	1,701	2,048	2,467	2,763
29	0,683	1,311	1,699	2,045	2,462	2,756
inf	0,674	1,282	1,645	1,960	2,326	2,576

Fonte: Elaborada pela autora (2016).

Agora, comparamos o valor calculado $t_{calculado} = 10,47$ com o valor tabelado $t_{tabelado} = 2,447$.

Então, se $t_{calculado} > t_{tabelado} \rightarrow$ devemos rejeitar H_0, ou seja, há evidências de que a correlação entre as variáveis exista, ao nível de significância de 5%.

Agora, sim, temos maior certeza de que exista a correlação entre as variáveis salário e o valor economizado na poupança, ou seja, faz sentido dizer que existe uma correlação positiva, pois quanto maior for o valor do salário, maior será o valor economizado na poupança.

AMPLIANDO CONHECIMENTO

Para ver mais exemplos de testes de hipóteses, comece acessando o vídeo a seguir: https://goo.gl/tvEmTz. Acesso em: 9 jul. 2017.

4.1.1.3 Coeficiente de Determinação

Além do coeficiente de correlação, podemos também calcular o coeficiente de determinação, também conhecido como coeficiente de explicação.

Esse coeficiente é obtido elevando-se ao quadrado o coeficiente de correlação calculado anteriormente. Então, se o coeficiente de correlação pode assumir valores entre -1 e 1, se esse valor for elevado ao quadrado, consequentemente, o coeficiente de determinação poderá assumir valores entre 0 e 1.

Então, $0 < r^2 < 1$, e esse valor pode ser expresso em percentual, em que 0 indica 0% de explicação e 1 representa 100% de explicação.

Nesse caso, se a variável x é a variável explicativa e y é a variável explicada, podemos dizer que o coeficiente de explicação informa quanto, percentualmente, as variações de x explicam as variações de y.

Um exemplo genérico para essa interpretação seria: ?% (qual o percentual) do aumento da variável x explica o aumento da variável y.

Observe que na interpretação genérica consideramos uma correlação positiva, em que a variável x e a variável y crescem.

Caso tenhamos uma correlação negativa, em que a variável x cresce e a variável y decresce, a interpretação ficaria assim:

?% (qual o percentual) do aumento da variável x explica a diminuição da variável y.

Exemplo 4.3 (utilizando os dados do Exemplo 4.1): Calcular o coeficiente de determinação.

Nos dados do nosso primeiro exemplo encontramos um coeficiente de correlação $r = 0,9737$. Para calcularmos o coeficiente de determinação, precisamos apenas elevar esse valor ao quadrado.

Resolução:

$$r^2 = 0,9737^2 = 0,9481$$

Interpretamos o resultado da seguinte maneira:

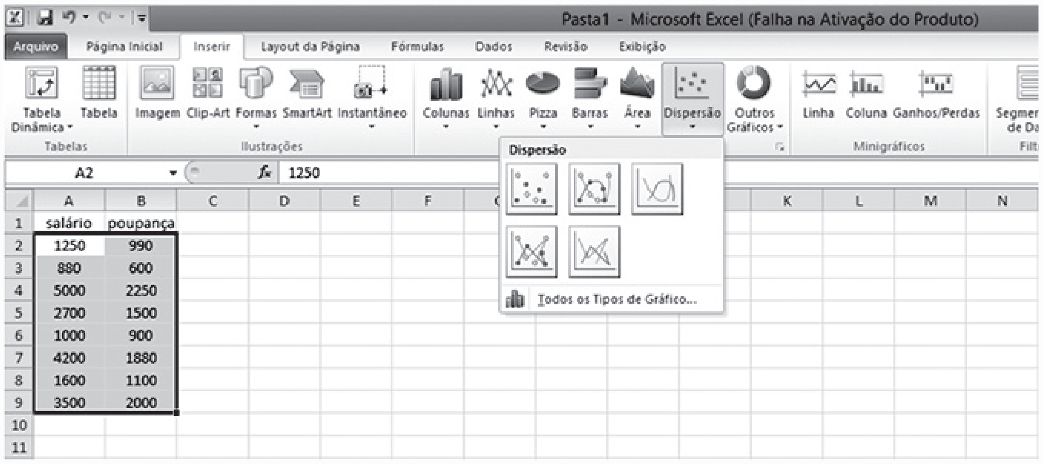
94,81% do aumento do salário explica o aumento do valor economizado com a poupança.

Ou

94,81% do aumento do valor economizado com a poupança é explicado pelo aumento do salário.

Podemos fazer esse cálculo com o auxílio do Excel. Primeiramente, pegamos os dados do Exemplo 4.1. Nesse caso, precisamos obrigatoriamente deixar na primeira coluna a variável x e na segunda coluna a variável y.

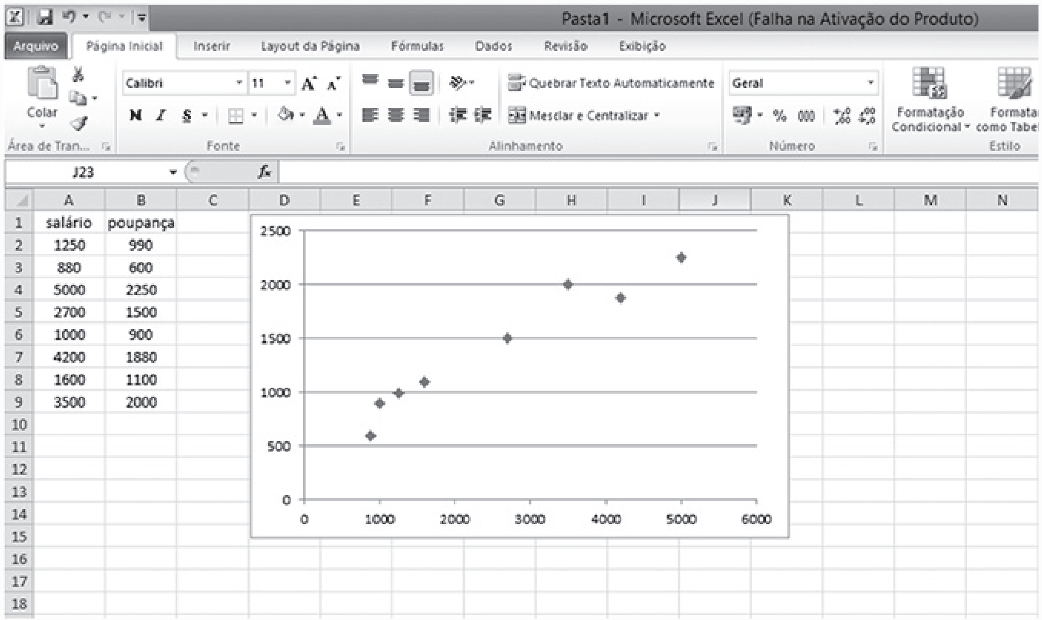
Figura 4.7: Inserir gráfico no Excel.



Fonte: Elaborada pela autora (2016).

Ao clicarmos no primeiro gráfico do diagrama de dispersão, vemos o seguinte:

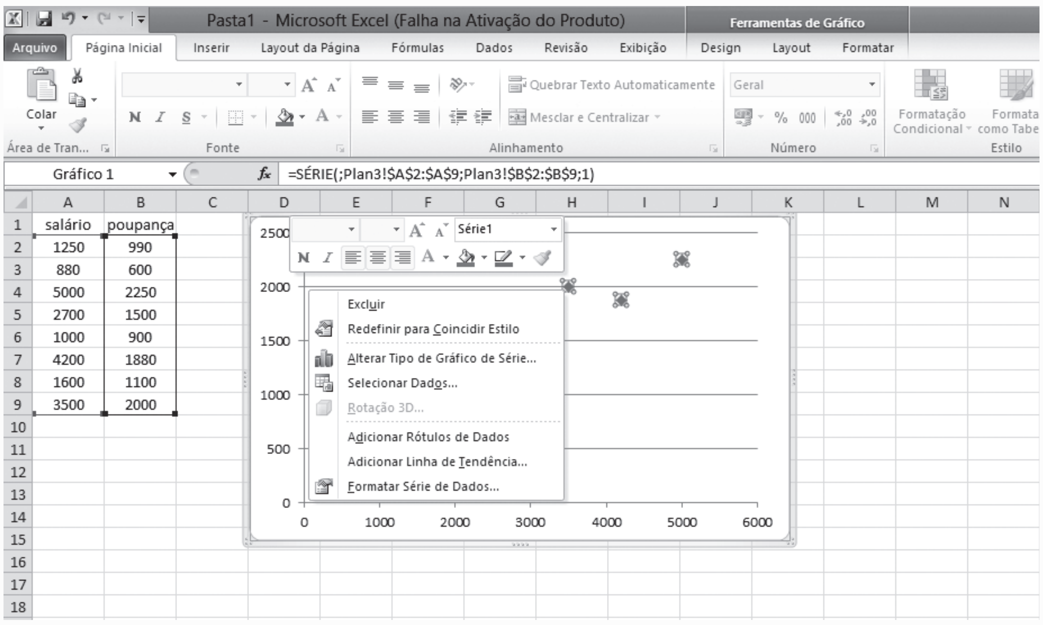
Figura 4.8: Diagrama de dispersão.



Fonte: Elaborada pela autora (2016).

Agora, com o botão direito do *mouse* sobre os pontos do diagrama de dispersão, temos.

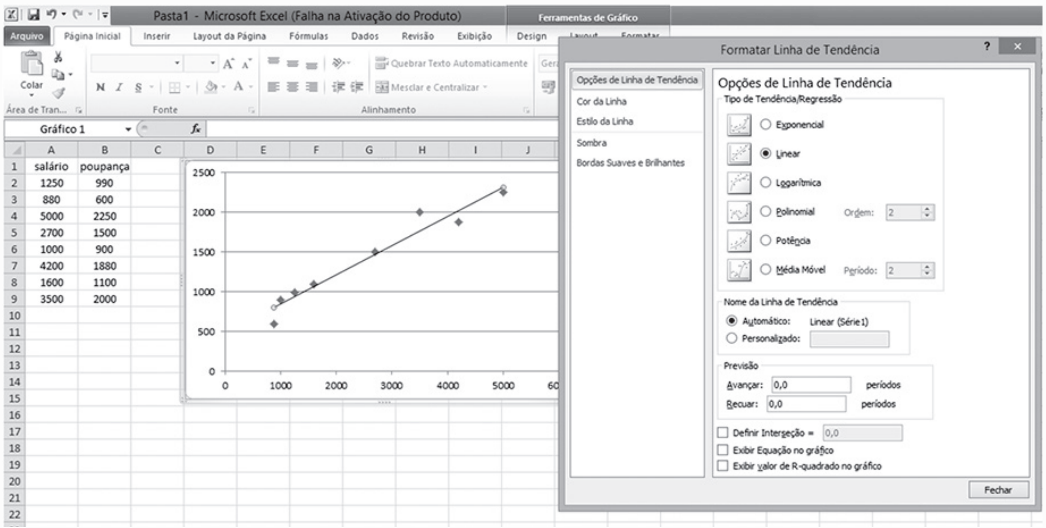
Figura 4.9: Diagrama de dispersão.



Fonte: Elaborada pela autora (2016).

Ao clicarmos em "adicionar linha de tendência", vemos o seguinte:

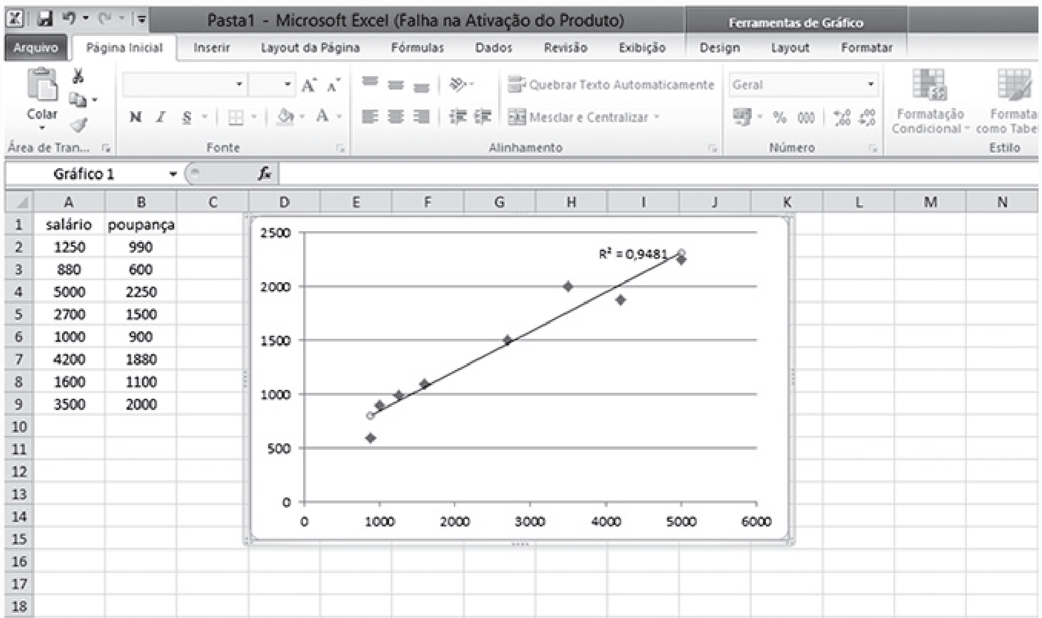
Figura 4.10: Janela de inserção de linha de tendência.



Fonte: Elaborada pela autora (2016).

Temos a regressão linear e, por hora, clicamos em exibir o valor de R-quadrado no gráfico. Assim, temos:

Figura 4.11: Diagrama de dispersão com o valor do r^2.



Fonte: Elaborada pela autora (2016).

4.1.1.4 Regressão Linear Simples

Em nosso curso abordaremos apenas a regressão linear simples, em que temos uma reta como aproximação e duas variáveis. Caso existam mais de duas variáveis, teremos a regressão múltipla aonde temos uma variável explicada y e várias explicativas x. Temos, também, outros ajustes além do ajuste linear abordado aqui.

O modelo de regressão linear simples fornece uma reta para resumir nossos dados e podermos fazer previsões para valores desconhecidos ou futuros.

Assim, a reta de regressão linear simples é dada por:

$$y = a + b \cdot x$$

Temos, assim, duas constantes a que representam o coeficiente da intercessão no eixo y. A outra constante b representa o coeficiente angular, sendo que esse coeficiente determina o ângulo da reta.

O sinal do valor de b na reta de regressão deve ser igual ao sinal do coeficiente de correlação. Dessa forma, quando a correlação é negativa, a função deve ser decrescente; e quando for positiva, a função deve ser crescente.

Assim, calculamos os coeficientes de seguinte maneira:

$$b = \frac{n \cdot \sum xy - \sum x \cdot \sum y}{n \cdot \sum x^2 - \left(\sum x\right)^2}$$

E

$$a = \frac{\sum y}{n} - b \cdot \frac{\sum x}{n}$$

Exemplo 4.4 (utilizando os dados do Exemplo 4.1): Vamos calcular a reta de regressão para as variáveis salário e valor economizado na poupança.

Tabela 4.4: Exemplo 4.1.

Salário (x)	Poupança (y)	$x \cdot y$	x^2
1.250,00	990,00	1.237.500	1.562.500
880,00	600,00	528.000	774.400
5.000,00	2.250,00	11.250.000	25.000.000
2.700,00	1.500,00	4.050.000	7.290.000
1.000,00	900,00	900.000	1.000.000
4.200,00	1.880,00	7.896.000	17.640.000
1.600,00	1.100,00	1.760.000	2.560.000
3.500,00	2.000,00	7.000.000	12.250.000
20.130	11.220	34.621.500	68.076.900

Fonte: Elaborada pela autora (2016).

Resolução: Aplicando as expressões apresentadas anteriormente, a partir dos dados obtidos da tabela, temos:

$$b = \frac{n \cdot \sum xy - \sum x \cdot \sum y}{n \cdot \sum x^2 - \left(\sum x\right)^2} = \frac{8 \cdot 34.621.500 - 20.130 \cdot 11.220}{8 \cdot 68.076.900 - (20.130)^2} = 0,3667$$

$$a = \frac{\sum y}{n} - b \cdot \frac{\sum x}{n} = \frac{11.220}{8} - 0,3667 \cdot \frac{20.130}{8} = 479,86$$

Note que esse último valor foi calculado considerando todas as casas decimais do coeficiente b .

Então, a reta de regressão será a seguinte:

$$y = 479,86 + 0,3667 \cdot x$$

Quando se pede para estimar a reta de regressão, esse é o resultado final, ficando o x e o y assim inalterados.

Essa reta de regressão serve para que possamos fazer uma projeção para valores distintos dos que possuímos nos dados. Com base em um valor qualquer da variável independente x, podemos estimar a variável dependente y. Veja o exemplo a seguir.

Exemplo 4.5: Com base na reta obtida no exemplo anterior, calcule qual seria o valor esperado na poupança para uma pessoa que tenha um salário de R$ 6.000,00.

Resolução:

Temos a reta $y = 479,86 + 0,3667 \cdot x$

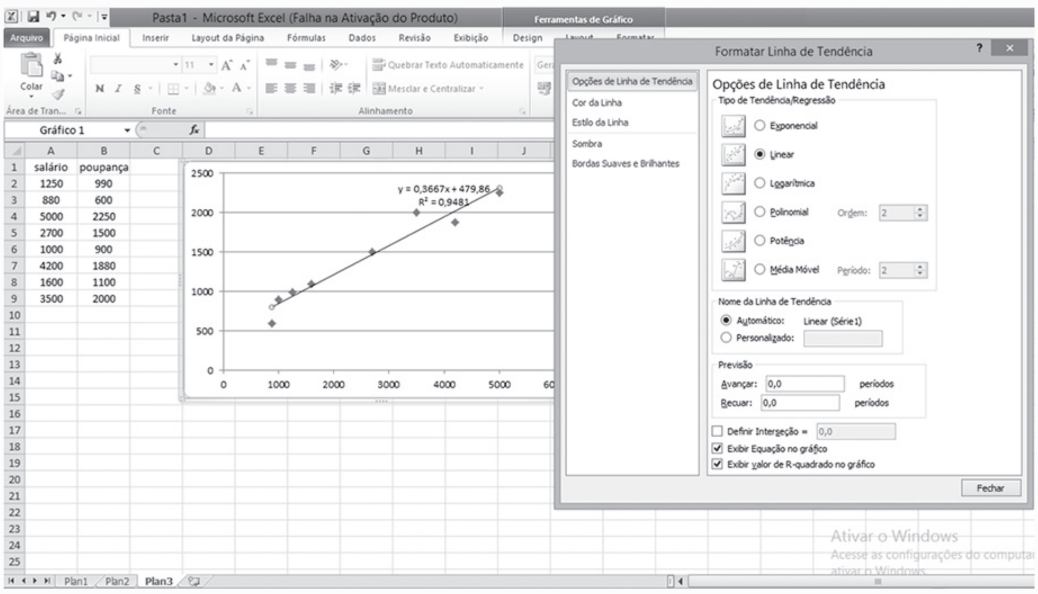
Agora, no lugar do x,colocamos o valor de 6.000.

$y = 479,86 + 0,3667 \cdot 6.000 = 2.680,06$

Então, o valor esperado em uma poupança de uma pessoa que tenha um salário de R$ 6.000,00 é de R$ 2.680,00.

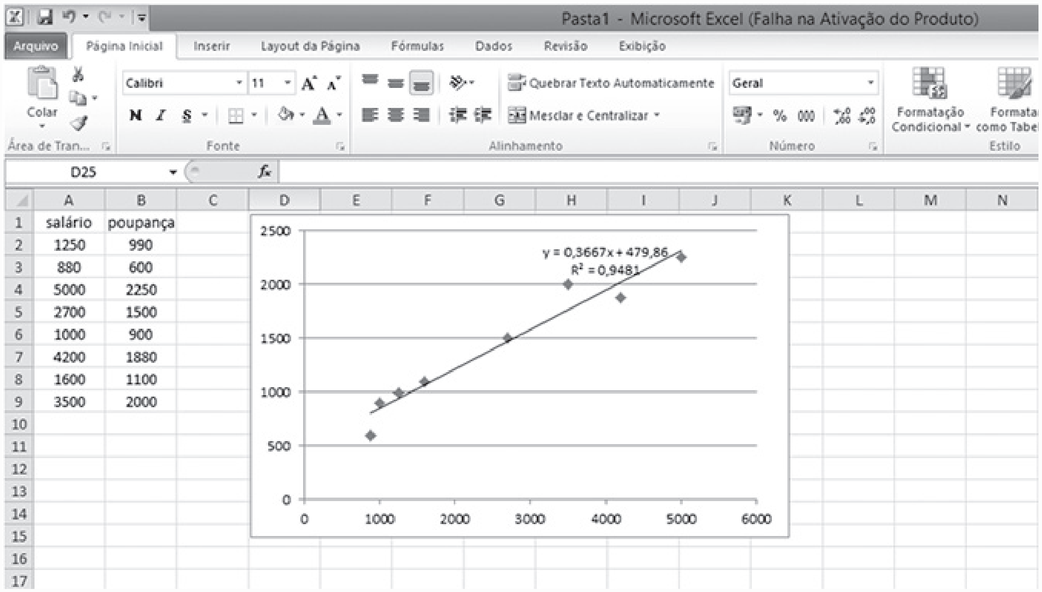
Podemos obter a reta de regressão no Excel. Precisamos, então, seguir os passos das Figuras 4.7, 4.8, 4.9 e 4.10, como mostrado na Figura 4.12, na qual se seleciona “exibir equação no gráfico”.

Figura 4.12: Janela de inserção de linha de tendência.



Fonte: Elaborada pela autora (2016).

Figura 4.13: Diagrama de dispersão.



Fonte: Elaborada pela autora (2016).

Observe que o Excel exibe a equação como $y = b \cdot x + a$, mas são os mesmos coeficientes a de intercessão e b angular.

Existe mais de uma maneira de obtermos a análise de correlação e regressão no Excel. A maneira apresentada aqui é a de mais fácil acesso. Apresentaremos, então, a partir de agora, outra maneira de fazermos essa análise: mais rápida e mais completa do que vimos até o momento.

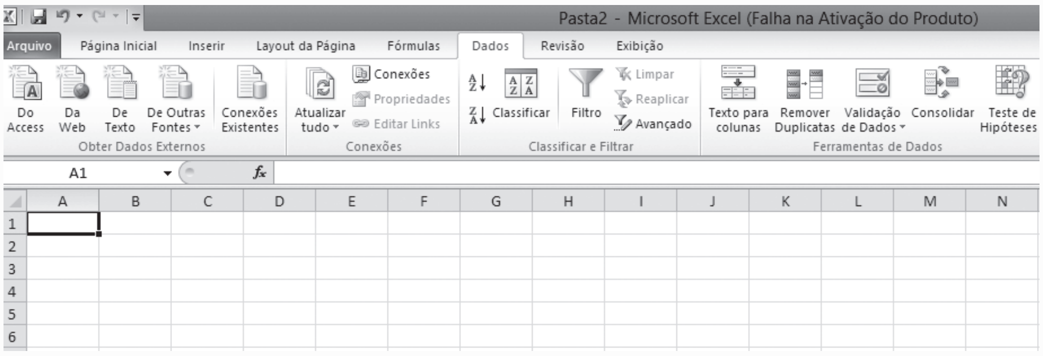
Porém, é preciso habilitar um suplemento do Excel, e isso precisa ser feito manualmente.

AMPLIANDO CONHECIMENTO

Para habilitar o suplemento de análise de dados no Excel, basta seguir as etapas descritas neste *link*: https://goo.gl/wXuUZx. Acesso em: 9 jul. 2017.

Após a habilitar a análise de dados, ela deve aparecer na aba “Dados do Excel”. Na figura a seguir, ela aparece no canto superior direito.

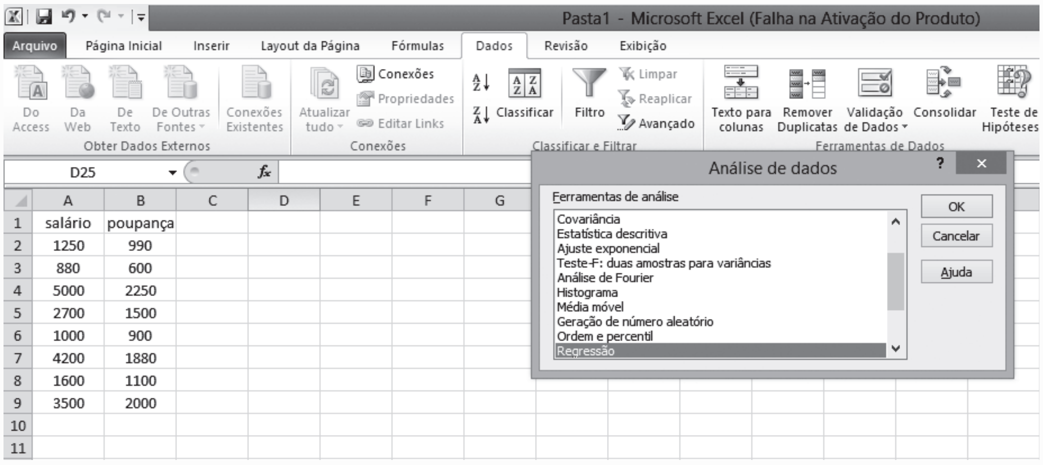
Figura 4.14: Análise de dados.



Fonte: Elaborada pela autora (2016).

Ao instalar o suplemento, passamos à análise propriamente dita. Assim, de posse dos dados, selecionamos a análise de dados e, nessa janela, clicamos em "Regressão".

Figura 4.15: Análise de dados.



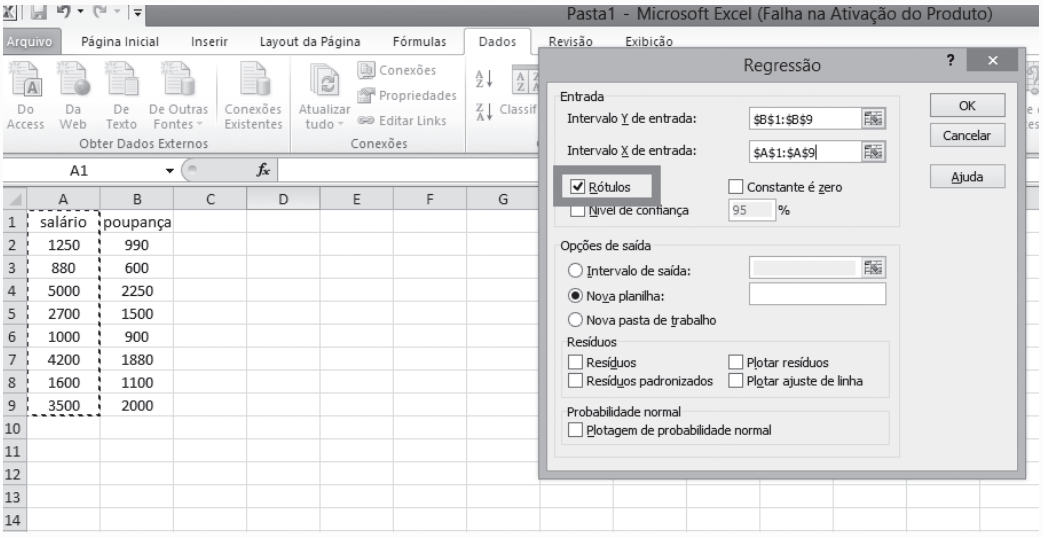
Fonte: Elaborada pela autora (2016).

Observem que nesse suplemento temos muitos outros tipos de análise estatística disponíveis.

Feito isso e uma vez clicado em "Regressão", selecionamos os dados: primeiramente, a coluna da variável y e, posteriormente, a coluna da variável x .

Nessa análise, podemos selecionar, inclusive, o rótulo dos dados, a célula com o nome da variável. Para que o Excel entenda que a primeira linha é um rótulo, precisamos marcar a opção "Rótulos", conforme a figura a seguir.

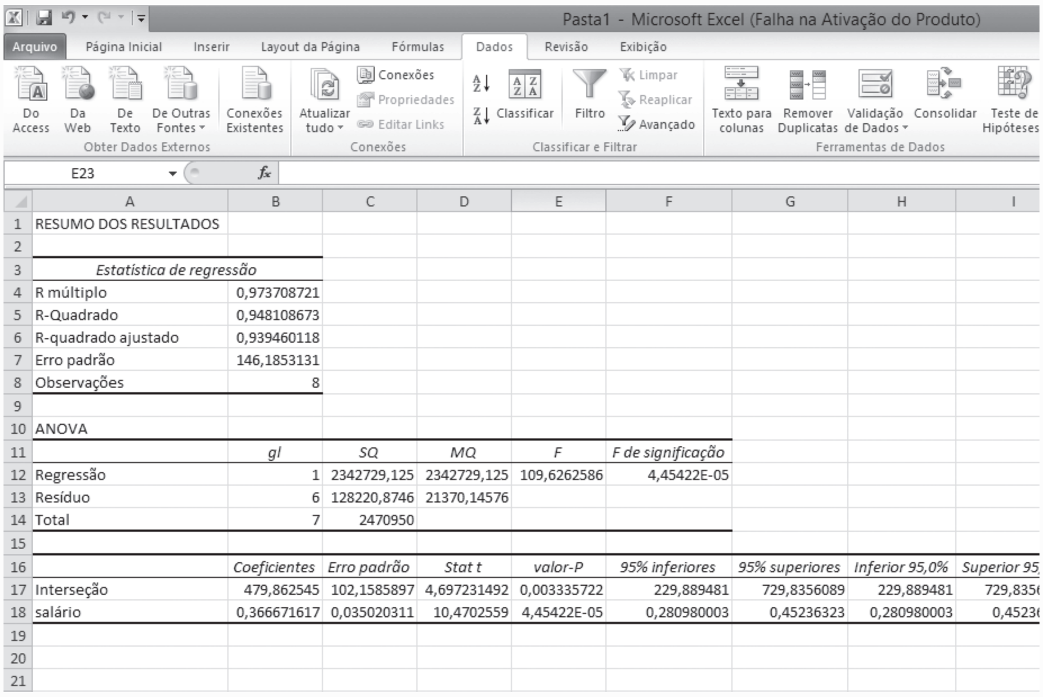
Figura 4.16: Análise de dados.



Fonte: Elaborada pela autora (2016).

Após selecionarmos os dados, clicamos em "OK" e uma nova planilha surgirá com o resultado da análise.

Figura 4.17: Resultado da análise de Regressão.



RESUMO DOS RESULTADOS

Estatística de regressão	
R múltiplo	0,973708721
R-Quadrado	0,948108673
R-quadrado ajustado	0,939460118
Erro padrão	146,1853131
Observações	8

ANOVA

	gl	*SQ*	*MQ*	*F*	*F de significação*
Regressão	1	2342729,125	2342729,125	109,6262586	4,45422E-05
Resíduo	6	128220,8746	21370,14576		
Total	7	2470950			

	Coeficientes	*Erro padrão*	*Stat t*	*valor-P*	*95% inferiores*	*95% superiores*	*Inferior 95,0%*	*Superior 95,*
Interseção	479,862545	102,1585897	4,697231492	0,003335722	229,889481	729,8356089	229,889481	729,835
salário	0,366671617	0,035020311	10,4702559	4,45422E-05	0,280980003	0,45236323	0,280980003	0,4523

Fonte: Elaborada pela autora (2016).

Vamos entender como esses dados estão apresentados. Nesse resumo, temos todas as informações necessárias.

Na primeira tabela apresentada, temos:

- **R múltiplo**: é o valor do coeficiente de correlação, conforme havíamos calculado anteriormente $r = 0,9737$, ou seja, temos uma correlação forte. É importante ressaltar que aqui não temos o sinal do coeficiente de correlação, pois, para obtermos essa informação, podemos fazer a análise de correlação no mesmo suplemento, ou então observar o sinal do coeficiente b.

- **R quadrado:** é o coeficiente de determinação (ou explicação), como calculamos anteriormente: $r^2 = 0,9737^2 = 0,9481$.

- **R quadrado ajustado**: útil quando temos mais de uma variável independente. No nosso caso, não precisamos analisar esse valor.

- **Erro padrão**: é uma medida de variabilidade, semelhante ao conceito do desvio-padrão. Enquanto o desvio-padrão mede as distâncias com relação à média, o erro padrão mede as distâncias entre cada valor da variável dependente e o valor ajustado do modelo gerado. Em nosso exemplo, cada valor previsto se afasta, em média, R$ 146,18 do valor economizado na poupança que foi observado.

- **Observações:** número de observações coletadas. No nosso caso, $n = 8$.

Na segunda tabela temos a análise de variância – ANOVA – que fornece um teste de hipóteses para validar o modelo. A ANOVA é um teste de hipóteses que leva em consideração as variações existentes no modelo e compara suas médias. No caso da ANOVA na análise de correlação e regressão, esse teste avalia se a variável independente x exerce influencia sobre a variável dependente y.

A ANOVA testa, nesse caso, a significância do modelo ajustado.

Temos nas colunas dessa tabela os graus de liberdade (gl), as somas de quadrados (SQ), os quadrados médios (QM) que estudam a variabilidade desse modelo, e temos, ainda, a estatística de teste (F) e o valor de p,que é a probabilidade associada ao valor da estatística de teste (F *de significação*).

Nessa tabela, precisamos analisar justamente o valor do F *de significação* comparando esse valor com o nível de significância de 5% (ou seja, $\alpha = 0,05$).

Se o F *de significação* for menor do que 0,05, o modelo é significativo. Já se o F *de significação* for maior do que 0,05, o modelo não é significativo.

No nosso caso, o valor do F *de significação* é $F\ de\ significação = 0,00004454 < 0,05$, então o modelo é significativo. Assim, podemos dizer que o salário explica significativamente o valor economizado na poupança.

Na terceira tabela precisamos observar a coluna dos coeficientes, na qual teremos os coeficientes da reta. O primeiro é o coeficiente da interseção, ou seja, o coeficiente a da reta; no nosso caso, como $a = 479,86$, o segundo coeficiente é o angular b, o coeficiente que acompanha a variável x. Assim, $b = 0,3667$.

Dessa forma, podemos escrever a equação $y = 479,86 + 0,3667 \cdot x$.

Tal equação nos diz que o valor da poupança é composto de uma parcela fixa de R$ 479,86 e a partir desse valor a poupança aumenta, em média, 0,3667 por real acrescido no salário.

Observe que o coeficiente angular b tem um valor positivo, o que indica que a correlação é positiva.

Vale atentar, também, para os valores de p nessa última tabela. Eles também precisam ser menores do que 0,05 para serem significativos. Somente no caso de serem inferiores a 0,05 é que os coeficientes poderão ser utilizados. Nesse caso, a reta será uma boa aproximação para os dados.

No caso de uma correlação com mais de uma variável independente, pode ocorrer de nem todas as variáveis sejam significativas para predizer o valor da variável dependente.

Assim, terminamos mais uma unidade, na qual vimos a análise de correlação linear e regressão simples.

Também finalizamos a disciplina na qual abordamos diversas ferramentas, como saber avaliar os tipos de variáveis, apresentar os dados. Aprendemos a resumir os dados com as medidas de posição e de tendência central, e aprendemos também sobre os cálculos de probabilidade básica e a utilização da distribuição Normal.

Esperamos que você tenha compreendido os importantes conceitos e as teorias estatísticas que aqui foram tratadas, fundamentais na prática de quaisquer administradores e gestores. Desejamos muito sucesso nas próximas etapas de sua carreira profissional e pessoal!

Distribuição de Student - cauda da direita

Pr (t > t_{alfa}) = alfa

	Nível de significância - alfa					
GL	**0,250**	**0,100**	**0,050**	**0,025**	**0,010**	**0,005**
1	1,000	3,078	6,314	12,706	31,821	63,657
2	0,816	1,886	2,920	4,303	6,965	9,925
3	0,765	1,638	2,353	3,182	4,541	5,841
4	0,741	1,533	2,132	2,776	3,747	4,604
5	0,727	1,476	2,015	2,571	3,365	4,032
6	0,718	1,440	1,943	2,447	3,143	3,707
7	0,711	1,415	1,895	2,365	2,998	3,499
8	0,706	1,397	1,860	2,306	2,896	3,355
9	0,703	1,383	1,833	2,262	2,821	3,250
10	0,700	1,372	1,812	2,228	2,764	3,169
11	0,697	1,363	1,796	2,201	2,718	3,106
12	0,695	1,356	1,782	2,179	2,681	3,055
13	0,694	1,350	1,771	2,160	2,650	3,012
14	0,692	1,345	1,761	2,145	2,624	2,977
15	0,691	1,341	1,753	2,131	2,602	2,947
16	0,690	1,337	1,746	2,120	2,583	2,921
17	0,689	1,333	1,740	2,110	2,567	2,898
18	0,688	1,330	1,734	2,101	2,552	2,878
19	0,688	1,328	1,729	2,093	2,539	2,861
20	0,687	1,325	1,725	2,086	2,528	2,845
21	0,686	1,323	1,721	2,080	2,518	2,831
22	0,686	1,321	1,717	2,074	2,508	2,819
23	0,685	1,319	1,714	2,069	2,500	2,807
24	0,685	1,318	1,711	2,064	2,492	2,797
25	0,684	1,316	1,708	2,060	2,485	2,787
26	0,684	1,315	1,706	2,056	2,479	2,779
27	0,684	1,314	1,703	2,052	2,473	2,771
28	0,683	1,313	1,701	2,048	2,467	2,763
29	0,683	1,311	1,699	2,045	2,462	2,756
inf	0,674	1,282	1,645	1,960	2,326	2,576

SÍNTESE DA UNIDADE

Nesta Unidade, você aprendeu que:

- O coeficiente de correlação nos indica a direção e a intensidade da correlação.
- O ajuste de uma equação de regressão, para os dados.
- Para estimar valores da variável dependente, para valores desconhecidos da variável independente, basta usar a expressão e o respectivo valor na expressão da reta ($y = a+bx$).
- Há dois métodos mais usuais para realizar a análise de correlação e regressão no Excel.

REFERÊNCIAS

MARTINS, G. A.; TOLEDO, G. L.; FONSECA, J. da. **Estatística Aplicada**. 2. ed. São Paulo: Atlas, 2012.

ANOTAÇÕES

UNIVERSIDADE LaSalle EAD

Av. Victor Barreto, 2288
Canoas - RS
CEP: 92010-000 | 0800 541 8500
ead@unilasalle.edu.br