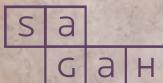


# SIMULAÇÃO DE SISTEMAS PRODUTIVOS

Gabriela Fonseca Parreira Gregorio



SOLUÇÕES  
EDUCACIONAIS  
INTEGRADAS

# Tratamento dos dados

## Objetivos de aprendizagem

Ao final deste texto, você deve apresentar os seguintes aprendizados:

- Definir o processo de tratamento de dados.
- Descrever os principais métodos estatísticos para tratamento de dados.
- Enumerar os testes estatísticos de aderência.

## Introdução

A coleta e o tratamento de dados são muito importantes. Afinal, a confiabilidade das informações geradas para conduzir a tomada de decisão depende desses processos. Existem várias formas de tratamento de dados: algumas são quantitativas e outras são qualitativas. Assim, surge a seguinte questão: qual é a forma mais adequada de tratar os dados relacionados a determinado projeto de modelagem e simulação?

O tratamento de dados busca avaliar a representatividade de uma amostra e determinar atributos e propriedades que facilitem o processo de análise. Neste capítulo, você vai conhecer as etapas que conduzem o tratamento de dados. Você também vai ver quais são as principais técnicas estatísticas aplicadas a esse processo e identificar testes estatísticos de aderência.

## Processo de tratamento dos dados

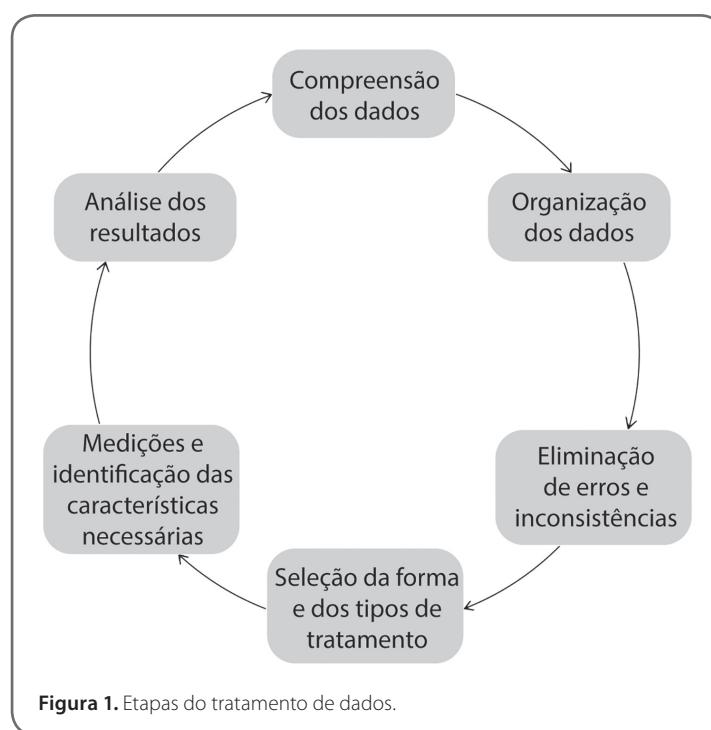
Tratar dados consiste em realizar medições e testes a fim de detectar erros e inconsistências, determinar características e familiarizar-se com elas. Em síntese, o tratamento de dados tem como objetivo facilitar o manuseio, a visualização e a compreensão dos elementos (FARIAS; SOARES; CÉSAR, 1996).

O tratamento estatístico de dados busca gerar informações que contribuam com o processo de tomada de decisões nas organizações. Por meio do tratamento, é possível identificar comportamentos e tendências. No entanto, como afirmam por Clark e Whitfield (1993 *apud* CARVALHO, 2003), para que

boas informações sejam geradas, é necessário ter em mente que o tratamento não melhora dados ruins e que tais dados podem conduzir a interpretações inadequadas.

Assim, o tratamento se inicia com a verificação da qualidade do conjunto de dados obtidos. A ideia é identificar erros e inconsistências e familiarizar-se com eles. Sem essas ações prévias, interpretações inadequadas ou inconclusivas podem ser realizadas (CARVALHO, 2003). Além disso, o tratamento de dados deve ser adequado aos propósitos do projeto. Para o alcance dos objetivos, é necessário coletar, tratar e interpretar os dados. Assim, é importante correlacionar os objetivos do projeto às formas de alcançá-los (VERGARA, 1998).

Os dados podem ser tratados de forma qualitativa ou quantitativa. Como exemplo da forma qualitativa, considere a codificação ou a apresentação de modo estruturado. Já a forma quantitativa faz uso de procedimentos estatísticos. Tratar os dados consiste em compreendê-los, organizá-los, detectar erros e inconsistências, eliminar esses erros e inconsistências e realizar as medições e determinações de características necessárias. Na Figura 1, você pode ver as etapas do processo de tratamento de dados.



A seguir, veja como são descritas as etapas mostradas na Figura 1.

1. **Compreensão dos dados:** é o contato inicial com os dados, em busca de familiarização e de melhor compreensão.
2. **Organização dos dados:** é a busca por estruturação, projeção, classificação e agrupamento.
3. **Eliminação de erros e inconsistências:** é a identificação de discrepâncias para que os dados sejam representativos e possam conduzir a informações assertivas.
4. **Seleção da forma e dos tipos de tratamento:** implica definir se os dados serão tratados de forma qualitativa e/ou quantitativa. Também implica estabelecer o tipo de tratamento que será dado, ou seja, as técnicas de tratamento utilizadas.
5. **Medições e identificação das características necessárias:** é a aplicação das técnicas selecionadas no passo anterior a fim de identificar, por exemplo, associações e comparações.
6. **Análise dos dados:** é o exame e a investigação dos resultados gerados ao se aplicarem as técnicas selecionadas.



### Saiba mais

As ferramentas de Big Data fazem o tratamento de dados por meio de algoritmos inteligentes que conduzem a tomada de decisões. Ele possui uma grande capacidade de tratamento e análise de dados. Existem também alguns softwares estatísticos que auxiliam no tratamento de dados. É o caso do Minitab, que ajuda na identificação de tendências e na melhoria de processos.

## Métodos estatísticos para tratamento de dados

Existem inúmeros métodos estatísticos para tratamento de dados. Assim, a questão central é: “[...] dado um conjunto de dados, como tratar os valores, numéricos ou não, a fim de extrair informações a respeito de uma ou mais características de interesse?” (MAGALHÃES; LIMA, 2011, p. 5).

A seleção dos métodos mais adequados depende dos objetivos do projeto de modelagem e simulação, dos dados coletados e das características a serem avaliadas. Segundo Evans e Lindsay (2005 *apud* TOLEDO, 2018), as técnicas

e ferramentas podem ser classificadas de acordo com a sua complexidade. Veja a seguir.

- Básicas: diagrama de dispersão, média, mediana, análise de Pareto, estratificação, entre outras.
- Intermediárias: cartas de controle, regressão, análise de capacidade do processo.
- Avançadas: análise de variância, regressão múltipla e testes de hipóteses.

A seguir, você vai conhecer melhor os seguintes métodos de tratamento de dados: média aritmética simples, média aritmética ponderada, mediana, moda, desvio padrão, variância e gráficos.

## Medidas estatísticas

As medidas estatísticas desportam um valor representativo em volta do qual os dados tendem a se agrupar, com maior ou menor frequência. Elas são utilizadas para resumir em um único número um conjunto de dados analisados (FARIAS; SOARES; CÉSAR, 1996). As medidas mais utilizadas são: média aritmética simples, média aritmética ponderada, mediana e moda.

### Média aritmética simples

A média aritmética simples de um conjunto  $n$  de observações  $x_1, x_2, \dots, x_n$  é o quociente da divisão da soma dos valores dessas observações pelo número de observações. É representada por (FARIAS; SOARES; CÉSAR, 1996):

$$\text{Média aritmética} = \frac{\sum_{i=1}^n (x_i)}{n}$$

### Média aritmética ponderada

A média aritmética ponderada dos números  $x_1, x_2, \dots, x_n$ , com pesos  $p_1, p_2, \dots, p_n$ , é representada por (FARIAS; SOARES; CÉSAR, 1996):

$$\text{Média aritmética ponderada} = \frac{x_1 p_1 + x_2 p_2 + \dots + x_n p_n}{p_1 + p_2 + \dots + p_n}$$

## Mediana

A mediana é o valor médio de uma distribuição em ordem. Esse valor apresenta o mesmo número de valores acima e abaixo de si (FEIJOO, 2010). Se  $n$  é um valor ímpar, esse valor é único. Se  $n$  é um valor par, a mediana é a média simples dos valores centrais. Veja o exemplo a seguir.

Considere os seguintes dados apresentados em ordem crescente: 1,2; 1,7; 2,0; 2,0; 2,1; 2,4. Como a amostra apresenta um número par de valores, ou seja, seis, a mediana é a média aritmética dos valores centrais:

$$\text{Mediana} = \frac{2,0+2,0}{2} = 2,0$$

## Moda

A moda é o valor ou os valores da variável que têm a maior probabilidade de ocorrência (MAGALHÃES; LIMA, 2011).

## Desvio padrão e variância

O desvio padrão e a variância são medidas de dispersão. A principal utilização dessas medidas é para avaliar as variações de determinada amostra ou grupo de dados. Eles podem ser medidos, respectivamente, pelas fórmulas a seguir:

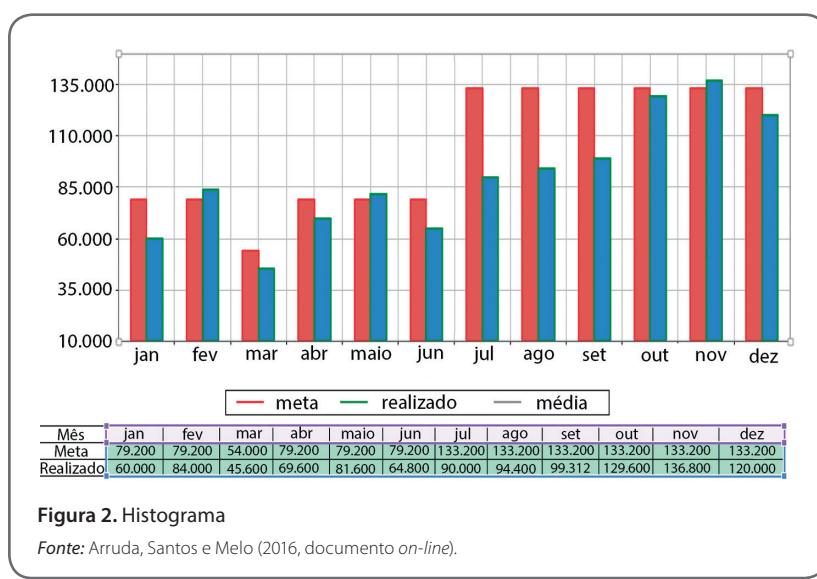
$$\begin{aligned}\text{Desvio padrão} &= \sqrt{\left( \frac{\sum(x_i - \bar{x}_{\text{médio}})^2}{n-1} \right)} \\ \text{Variância} &= \frac{\sum(x_i - \bar{x}_{\text{médio}})^2}{n}\end{aligned}$$

## Gráficos

Os gráficos têm o objetivo de representar os dados, facilitando a visualização dos resultados. Um dos modelos de gráfico é o histograma.

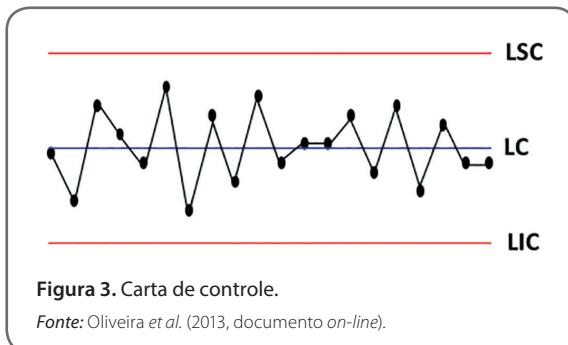
## Histograma

Segundo Farias, Soares e César (1996), trata-se de um conjunto de retângulos com bases sobre um eixo, dividido segundo os tamanhos das classes, com centros nos pontos médios das classes e áreas proporcionais às frequências. Na Figura 2, você pode ver um modelo de histograma que apresenta a média programada e a realizada de produção de café.



## Cartas de controle

O gráfico de controle é elaborado por meio de três linhas paralelas: uma linha média, que representa o nível de operação do processo, e outras duas linhas externas identificadas como limite superior de controle (LSC) e limite inferior de controle (LIC). A Figura 3 representa uma carta de controle.



**Figura 3.** Carta de controle.

Fonte: Oliveira et al. (2013, documento on-line).

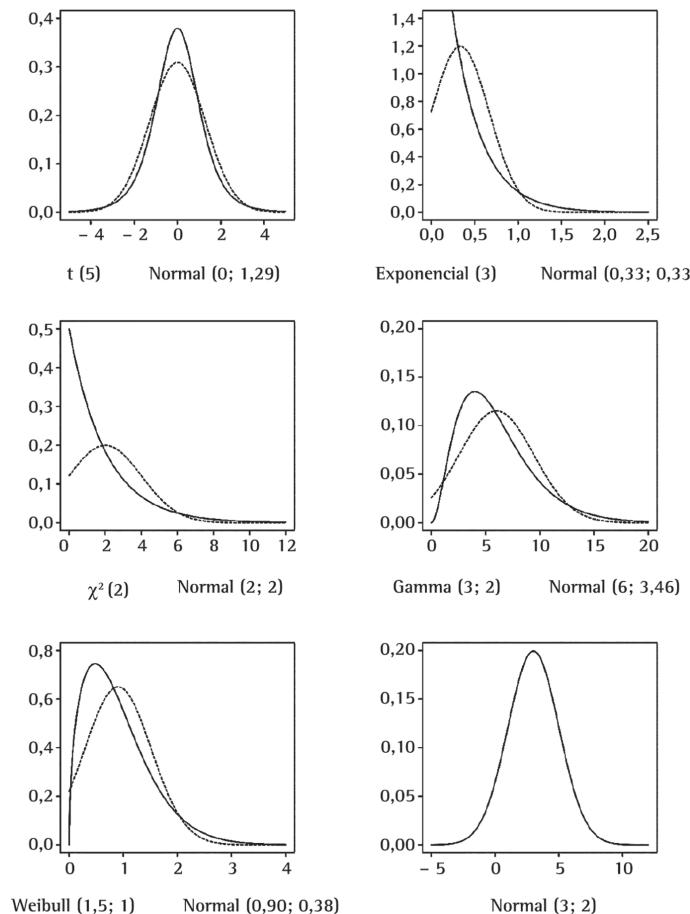
Para a construção da carta de controle, são necessárias as seguintes ações:

- a) definir a amostra aleatória;
- b) calcular a média e o desvio padrão;
- c) representar na carta a média, o limite superior de controle e o limite inferior de controle (os limites são a média  $+/-$  três vezes o desvio padrão — carta de Shewhart);
- d) marcar na carta o valor de cada item da amostra;
- e) avaliar se os valores estão em torno do valor central (média).

Caso os pontos estejam alternando em torno do valor central, considera-se que os itens estão controlados. No entanto, caso existam pontos que excedam os limites superiores e inferiores ou tendências de crescimento ou decréscimo, é necessário identificar as causas de possíveis problemas.

## Testes estatísticos de aderência

Uma variável aleatória assume uma distribuição de frequência específica. Essa distribuição pode apresentar inúmeras formas. Assim, é necessário encontrar uma forma específica de função de distribuição de probabilidade que reflita o comportamento da variável (FARIAS; SOARES; CÉSAR, 1996). Na Figura 4, você pode ver alguns modelos de distribuição de probabilidade.



**Figura 4.** Distribuições de probabilidade.

Fonte: Korzenowski e Werner (2012, documento on-line).

A distribuição de probabilidade pauta, por meio de modelos matemáticos, determinado valor de uma variável com sua respectiva probabilidade de ocorrência. E quando não se conhece a distribuição de probabilidades da variável? Nesse caso, é necessário testar a hipótese de que a variável pertence a uma distribuição de probabilidade específica.

Os testes de aderência são utilizados para testar modelos e avaliar se os dados de uma amostra conformam-se ou não a uma distribuição de probabilidade específica. Assim, eles avaliam duas hipóteses ( $H$ ), como você pode ver a seguir.

- $H_0$ : os dados da amostra conformam-se a determinada distribuição de probabilidade.
- $H_1$ : os dados da amostra não se conformam a determinada distribuição de probabilidade.

## Teste qui-quadrado

Uma situação frequente é ter observações de uma variável aleatória cuja distribuição é desconhecida. Em casos como esse, uma opção é tentar, por meio de algum modelo teórico, identificar o comportamento da variável. Em outras situações, é possível incorporar informações de outros fenômenos aleatórios similares que tenham distribuição conhecida. Existem ainda casos em que não se tem ideia sobre o comportamento da variável e recorre-se, inicialmente, ao histograma para representar as frequências de ocorrências (MAGALHÃES; LIMA, 2011).

Em todas as situações, o problema consiste em estabelecer um procedimento para avaliar se o modelo proposto deve ou não ser aceito. O teste qui-quadrado (teste  $X^2$ ) pode ser utilizado como teste de aderência (MAGALHÃES; LIMA, 2011).

Imagine que, em uma amostra preestabelecida, observou-se que um conjunto de eventos possíveis  $E_1, E_2, \dots, E_n$  apresentou as frequências/ocorrências  $O_1, O_2, \dots, O_n$  (frequências observadas/medidas) e que, de acordo com as regras de probabilidade, acreditava-se que elas ocorressem com as frequências  $e_1, e_2, \dots, e_n$ , denominadas frequências teóricas ou esperadas. Questiona-se, então: as frequências observadas diferem, consideravelmente, das esperadas? (SPIEGEL, 1993).

Sendo  $X^2$  as discrepâncias entre as frequências observadas e teóricas (esperadas), ele pode ser dado por meio de (SPIEGEL, 1993):

$$X^2 = \sum_{j=1}^k \frac{(o_j - e_j)^2}{e_j}$$

Se  $X^2 = 0$ , as frequências teóricas e observadas concordam exatamente. Já se  $X^2 > 0$ , as frequências teóricas e observadas não concordam exatamente. Quanto maior o valor de  $X^2$ , maiores as discrepâncias entre as ocorrências notadas e as esperadas.

Os testes de aderência são feitos considerando determinado nível de significância, que é a probabilidade de se rejeitar, de forma inadequada, determinada hipótese. Para a realização de comparações entre o qui-quadrado tabelado e

o qui-quadrado calculado, é importante conhecer também o número de grau de liberdade.

O número de grau de liberdade ( $v$ ) é dado por (SPIEGEL, 1993):

- $v = k - 1$  se as frequências esperadas puderem ser calculadas, sem que se façam estimativas dos parâmetros populacionais, por meio de estatísticas amostrais.
- $v = k - 1 - m$  se as frequências esperadas somente podem ser calculadas mediante a estimativa de  $m$  parâmetros populacionais, por meio de estatísticas amostrais.

De posse do qui-quadrado calculado e do qui-quadrado tabelado, se estabelece que:

- se  $X^2$  calculado  $> X^2$  tabelado, rejeita-se  $H_0$ , ou seja, existem discrepâncias consideráveis entre os valores observados e os esperados;
- se  $X^2$  calculado  $< X^2$  tabelado, não se rejeita  $H_0$ , ou seja, não existem discrepâncias consideráveis entre os valores observados e os esperados.

Considere o número de pessoas observadas e esperadas em determinado centro de vacinação, conforme o Quadro 1.

**Quadro 1.** Dados de centro de vacinação.

| Dia | Frequência observada | Frequência esperada |
|-----|----------------------|---------------------|
| 1   | 97                   | 100                 |
| 2   | 104                  | 100                 |
| 3   | 99                   | 100                 |
| 4   | 107                  | 100                 |
| 5   | 100                  | 100                 |
| 6   | 101                  | 100                 |
| 7   | 98                   | 100                 |

Existe discrepância entre os dados observados e esperados, considerando um nível de significância de 0,05? Veja as hipóteses a seguir.

- $H_0$ : não existe discrepância entre o número de pessoas observadas e esperadas no centro de vacinação.
- $H_1$ : existe discrepancia entre o número de pessoas observadas e esperadas no centro de vacinação.

### O teste

Veja o cálculo do  $X^2$ :

$$X^2 = \frac{(97-100)^2}{100} + \frac{(104-100)^2}{100} + \frac{(99-100)^2}{100} + \frac{(107-100)^2}{100} + \frac{(100-100)^2}{100} + \frac{(101-100)^2}{100} + \frac{(98-100)^2}{100} = 0,09 + 0,36 + 0,01 + 0,49 + 0 + 0,01 + 0,04 = 1$$

O valor do  $X^2$  tabelado, considerando o nível de significância de 0,05 e o número de grau de liberdade igual a 6 ( $V = 7 - 1 = 6$ ), é 12,6.

Assim, como o  $X^2$  calculado é menor do que o  $X^2$  tabelado, conclui-se que não existe discrepancia considerável entre as frequências observadas e as frequências esperadas.



#### Saiba mais

Para ler mais sobre a correção de dados agrometeorológicos, busque na internet o artigo "Correção de dados agrometeorológicos utilizando métodos estatísticos", de Ricardo Baba, Maria Vaz e Jéssica da Costa.

## Outros testes de aderência

O qui-quadrado é um teste de aderência importante e amplamente utilizado. No entanto, existem outros testes que podem ser úteis para o tratamento e a análise de dados. O **teste de Ryan-Joiner**, por exemplo, avalia se os dados de uma amostra são oriundos de uma distribuição de probabilidade normal ou não. A construção de um modelo de regressão que verifica quanto bom é o ajuste é a base do método (BESSEGATO, 2010). Quando o coeficiente de correlação é próximo de 1, existe normalidade na população.

O **teste Shapiro-Wilks** também verifica se os dados apresentam uma distribuição normal. Ele é mais indicado quando o número de elementos de uma amostra é pequeno, ou seja, inferior a 50. Outro teste de aderência é denominado **teste de Anderson-Darling**. Esse teste compara a função de distribuição acumulada empírica dos elementos de uma amostra com a distribuição esperada se os dados fossem normais. Quando a diferença é grande, conclui-se que não há normalidade da população (MINITAB, c2019).

Já o **teste Levene** é útil para fazer a comparação de variâncias de grupos de observações originadas de distribuições contínuas e não obrigatoriamente normais (ALMEIDA; ELIAN; NOBRE, 2008). Por sua vez, o **teste M de Box** permite avaliar variações em diversas amostras e testar a homogeneidade de matrizes de variância-covariância. Assim, cabe ao analista identificar os testes mais adequados que conduzirão a geração de informações úteis ao processo de modelagem e simulação.



### Saiba mais

Para ler mais sobre métodos estatísticos e big data, busque na internet o artigo “Métodos estatísticos no contexto de Big Data”, de Carolina Schleger, Benjamim Zucolotto e Rafael Frantz.



## Referências

- ALMEIDA, A.; ELIAN, S.; NOBRE, J. Modificações e alternativas aos testes de Levene e de Brown e Forsythe para igualdade de variâncias e médias. *Revista Colombiana de Estadística*, [s. l.], v. 31, n. 2, p. 241-260, dic. 2008.
- ARRUDA, A. I. B.; SANTOS, E. C. A.; MELO, L. S. S. Análise da gestão da qualidade em uma indústria de alimentos em Caruaru - PE: estudo sobre a utilização das ferramentas da qualidade. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 36., 2016, João Pessoa. *Anais* [...]. João Pessoa: [s. n.], 2016. Disponível em: [http://www.abepro.org.br/biblioteca/TN\\_STO\\_227\\_328\\_29552.pdf](http://www.abepro.org.br/biblioteca/TN_STO_227_328_29552.pdf). Acesso em: 09 abr. 2019.
- BESSEGATO, L. *Testes de aderência: testes de bondade de ajuste*. Belo Horizonte: Universidade Federal de Minas Gerais, [2010]. Disponível em: <http://www.bessegato.com.br/UFMG/bondade.htm>. Acesso em: 09 abr. 2019.
- CARVALHO, M. A. G. *Métodos estatísticos para análise de dados de monitoração ambiental*. 2003. Tese (Doutorado em Ciências) – IPEN, Universidade de São Paulo, São Paulo, 2003.
- FARIAS, A.; SOARES, J.; CÉSAR, C. *Introdução à estatística*. 2. ed. Rio de Janeiro: LTC, 1996.
- FEIJOO, A. Medidas de Tendência Central. In: FEIJOO, A. *A pesquisa e a estatística na psicologia e na educação*. Rio de Janeiro: Centro Edelstein de Pesquisas Sociais, 2010.
- KORZENOWSKI, A. L.; WERNER, L. Probabilidade do erro tipo I nas cartas X e S de Shewhart sob não normalidade. *Produção*, São Paulo, v. 22, n. 4, p. 807-816, set./dez. 2012. Disponível em: [http://www.scielo.br/pdf/prod/v22n4/aop\\_t6\\_0002\\_0127.pdf](http://www.scielo.br/pdf/prod/v22n4/aop_t6_0002_0127.pdf). Acesso em: 09 abr. 2019.
- MAGALHÃES, M.; LIMA, A. *Noções de probabilidade e estatística*. São Paulo: Universidade de São Paulo, 2011.
- MINITAB. [Site]. c2019. Disponível em: [www.minitab.com](http://www.minitab.com). Acesso em: 09 abr. 2019.
- OLIVEIRA, C. C. et al. *Manual para elaboração de cartas de controle para monitoramento de processos de medição quantitativos em laboratórios de ensaio*. São Paulo: SES/SP, 2013. Disponível em: [http://www.ial.sp.gov.br/resources/editorinplace/ial/2016\\_3\\_19/manual-carta-controle\\_ial\\_2013.pdf](http://www.ial.sp.gov.br/resources/editorinplace/ial/2016_3_19/manual-carta-controle_ial_2013.pdf). Acesso em: 09 abr. 2019.
- SPIEGEL, M. *Estatística*. 3. ed. São Paulo: Pearson, 1993.
- TOLEDO, J.; LIZARELLI, F.; BORRÁS, M.; MARTINS, M. *O uso da estatística em grandes e médias empresas de autoparcelas no estado de São Paulo*, Brasil. Innovar. 2018.
- VERGARA, S. *Projetos e relatórios de pesquisa em administração*. São Paulo: Atlas, 1998.

## Leituras recomendadas

BABA, R. K.; VAZ, M. S. M. G.; COSTA, J. Correção de dados agrometeorológicos utilizando métodos estatísticos. *Revista Brasileira de Meteorologia*, v. 29, n. 4, p. 515-526, 2014. Disponível em: <https://www.scielo.br/j/rbmet/a/TJPzfbvqdFbXpvHVkYRTxHk/?format=pdf&lang=pt>. Acesso em: 10 mar. 2022.

LOPES, M. M.; BRANCO, V. T. F. C.; SOARES, J. B. Utilização dos testes estatísticos de Kolmogorov-Smirnov e Shapiro-Wilk para verificação da normalidade para materiais de pavimentação. *Transportes*, São Paulo, v. 21, n. 1, p. 59-66, 2013. Disponível em: <https://www.revistatransportes.org.br/anpet/article/view/566/467>. Acesso em: 09 abr. 2019.

SCHLEGER, C. H.; ZUCOLOTTO, B.; FRANTZ, R. Z. Métodos estatísticos no contexto de Big Data. *Salão do Conhecimento*, v. 7, n. 7, 2021. Disponível em: <https://publicacoeseventos.unijui.edu.br/index.php/salaconhecimento/article/view/20747/19459>. Acesso em: 10 mar. 2022.

UNIVERSIDADE DE SÃO PAULO. Instituto de Física de São Carlos. Laboratório Avançado de Física. *Tabela-qui-quadrado*. c2017. Disponível em: <http://www.ifsc.usp.br/~lavfis/lavfiswp/flutuacao-estatistica/tabela-qui-quadrado/>. Acesso em: 09 abr. 2019.

WERKEMA, M. C. *Ferramentas estatísticas básicas para o gerenciamento de processos*. Belo Horizonte: Fundação Christiano Ottoni, 2006.

Encerra aqui o trecho do livro disponibilizado para esta Unidade de Aprendizagem. Na Biblioteca Virtual da Instituição, você encontra a obra na íntegra.

Conteúdo:



SOLUÇÕES  
EDUCACIONAIS  
INTEGRADAS