GUIA MANGÁ
ANÁLISE DE REGRESSÃO

SHIN TAKAHASHI,
IROHA INOUE E
TREND-PRO CO., LTD.

Novatec

The Manga Guide to Regression Analysis is a translation of the japanese original, *Manga de wakaru tōkei-gaku kaiki bunseki-hen*, published by Ohmsha, Ltd. of Tokyo, Japan, © 2005 by Shin Takahashi and TREND-PRO Co., Ltd. The english edition is co-published by No Starch Press, Inc. and Ohmsha, Ltd. Portuguese-language rights arranged with Ohmsha, Ltd. and No Starch Press, Inc. for *Guia Mangá Análise de Regressão* ISBN 978-85-7522-698-8, published by Novatec Editora Ltda.

Edição original em japonês *Manga de wakaru tōkei-gaku kaiki bunseki-hen*, publicado pela Ohmsha, Ltd. de Tóquio, Japão © 2005 por Shin Takahashi e TREND-PRO Co., Ltd. Edição em inglês *The Manga Guide to Regression Analysis*, co-publicação da No Starch Press, Inc. e Ohmsha, Ltd. Direitos para a edição em português acordados com a Ohmsha, Ltd. e No Starch Press, Inc. para *Guia Mangá Análise de Regressão* ISBN 978-85-7522-698-8, publicada pela Novatec Editora.

Editor: Rubens Prates
Tradução: BrodTec
Revisão gramatical: Tássia de Carvalho
Editoração eletrônica: Carolina Kuwabata

ISBN: 978-85-7522-698-8

Histórico de impressões:
Setembro/2018 Primeira edição

NOVATEC EDITORA LTDA.

Rua Luís Antônio dos Santos 110

02460-000 – São Paulo, SP – Brasil
Tel.: +55 11 2959-6529
E-mail: novatec@novatec.com.br
Site: www.novatec.com.br
Twitter: twitter.com/novateceditora
Facebook: facebook.com/novatec
LinkedIn: linkedin.com/in/novatec

SUMÁRIO

PREFÁCIO

Este livro é uma introdução à análise de regressão, cobrindo a análise de regressão simples, múltipla e logística.

As análises de regressão simples e múltiplas são métodos estatísticos para prever valores: por exemplo, você pode utilizar a regressão simples para prever quantos chás gelados serão pedidos, baseando-se na temperatura máxima do dia ou utilizar a análise de regressão múltipla para prever os salários mensais de uma loja, baseando-se no seu tamanho e na distância até a estação de trem mais próxima.

A análise de regressão logística é um método para prever a probabilidade de se vender um bolo específico baseando-se no dia da semana, por exemplo. O público-alvo deste livro são estudantes de matemática e estatística ou qualquer pessoa que esteja começando a aprender sobre previsões estatísticas e probabilidades. Você precisará de alguns conceitos básicos sobre estatística antes de começar. O *Guia Mangá de Estatística* (Novatec Editora, 2010) é um excelente guia para prepará-lo para este livro. Esta obra consiste em quatro capítulos:

- Capítulo 1: Um Copo Refrescante de Matemática
- Capítulo 2: Análise de Regressão Simples
- Capítulo 3: Análise de Regressão Múltipla
- Capítulo 4: Análise de Regressão Logística

Cada capítulo tem uma seção em mangá e uma parte mais técnica em texto. Você pode obter uma visão geral a partir do mangá e alguns detalhes mais úteis bem como definições específicas a partir das seções de texto.

Gostaria de mencionar alguns pontos sobre o Capítulo 1. Embora muitos leitores tenham conhecimento prévio sobre os tópicos deste capítulo, como diferenciação e operações com matrizes, o capítulo revisa estes tópicos dentro do conceito de análise de regressão, o que será útil nas seções seguintes. Se o Capítulo 1 for uma simples revisão para você, ótimo. Se nunca estudou esses tópicos ou faz muito tempo que estudou sobre eles, vale a pena se empenhar um pouco e ter a certeza de que entendeu o Capítulo 1 antes de continuar.

Neste livro, a matemática para os cálculos é tratada em detalhes. Se você é bom em matemática, será capaz de acompanhar e entender os cálculos. Se não for tão bom em matemática, pode obter uma visão geral dos procedimentos e utilizar as instruções passo a passo para encontrar as respostas. Você não precisa obrigar-se a entender a parte matemática agora. Mantenha-se relaxado, mas dê uma olhada no desenvolvimento dos cálculos.

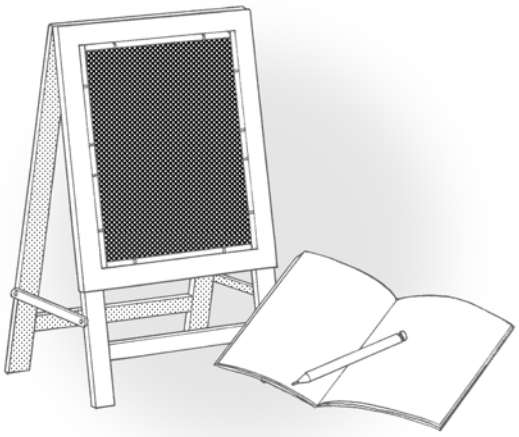
Nós arredondamos alguns números neste livro para facilitar a leitura, o que significa que alguns valores podem ser inconsistentes com os valores que você encontrará fazendo os cálculos por conta, mas eles ficarão próximos. Pedimos sua compreensão.

Gostaria de agradecer ao meu editor, Ohmsha, por me dar a oportunidade de escrever este livro. Também gostaria de agradecer a TREND-PRO, Co., Ltd. por transformar o meu manuscrito neste mangá, ao escritor de cenário re _akino e ao ilustrador Iroha Inoue. Por último, mas não menos importante, gostaria de agradecer ao Dr. Sakaori Fumitake da Faculdade de Relações Sociais Rikkyo University. Ele me deu conselhos imensuráveis, mais ainda do que já havia me dado quando eu estava preparando o meu livro anterior. Gostaria de expressar a minha profunda admiração.

Shin Takahashi
Setembro, 2005

1

UM COPO REFRESCANTE DE MATEMÁTICA




CONSTRUINDO UMA BASE
UFA!
O CHEFE FINALMENTE FOI EMBORA, E NÓS TEMOS QUE IR TAMBÉM!
UHM, RISA, PODEMOS COMEÇAR AS AULAS HOJE À NOITE?
SÉRIO?
VOCÊ QUER COMEÇAR AGORA?
こく
ACENO
EU NUNCA VI VOCÊ TÃO EMPOLGADA EM APRENDER! GERALMENTE VOCÊ DORME NAS AULAS!
É, BEM.. É QUE...
DESCULPE...
EU NÃO QUERIA ENVERGONHÁ-LA...

REGRAS PARA EXPONENCIAIS E LOGARITMOS

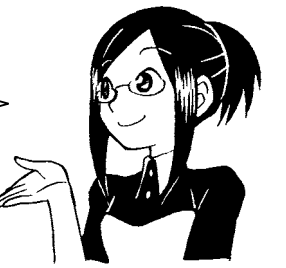
1. REGRA DA POTÊNCIA $(e^a)^b$ E $e^{a\times b}$ SÃO IGUAIS.

Vamos fazer uma tentativa. Nós vamos confirmar que $(e^a)^b$ e $e^{a\times b}$ são iguais quando $a = 2$ e $b = 3$.

$$\left(e^2\right)^3 = \underbrace{e^2 \times e^2 \times e^2}_{3} = \underbrace{\left(e\times e\right)\times\left(e\times e\right)\times\left(e\times e\right)}_{3} = \underbrace{e\times e\times e\times e\times e\times e}_{6} = e^{2\times 3}$$

Isso também significa que $(e^a)^b = e^{a\times b} = (e^b)^a$.

2. REGRA DO QUOCIENTE $\dfrac{e^a}{e^b}$ E e^{a-b} SÃO IGUAIS.



Agora, vamos tentar o seguinte: Nós vamos confirmar que $\dfrac{e^a}{e^b}$ e e^{a-b} são iguais quando $a = 3$ e $b = 5$.

$$\frac{e^3}{e^5} = \frac{e\times e\times e}{e\times e\times e\times e\times e} = \frac{\cancel{e}\times\cancel{e}\times\cancel{e}}{e\times e\times\cancel{e}\times\cancel{e}\times\cancel{e}} = \frac{1}{e^2} = e^{-2} = e^{3-5}$$


PRIMEIROS PASSOS
ISSO SIGNIFICA QUE....
EXISTEM UMA LIGAÇÃO ENTRE AMBOS, CERTO?
ISSO MESMO!
ONDE VOCÊ APRENDEU TANTO SOBRE ANÁLISE DE REGRESSÃO, MIU?
MIU!
はっ
PISCA PISCA
TERRA CHAMANDO MIU! TEM ALGUÉM AÍ?

Passo 3

Derive S_e em relação a a e b, e iguale a derivada a zero. A derivada de $y = (ax + b)^{n-1}$ em relação a x é $\frac{dy}{dx} = n(ax + b)^{n-1} \times a$.

- Derive em relação a a.

$$\frac{dS_e}{da} = 2\left[77 - (29a + b)\right] \times (-29) + \cdots + 2\left[84 - (30a + b)\right] \times (-30) = 0 \quad ❶$$

- Derive em relação a b.

$$\frac{dS_e}{db} = 2\left[77 - (29a + b)\right] \times (-1) + \cdots + 2\left[84 - (30a + b)\right] \times (-1) = 0 \quad ❷$$

Passo 4

Rearranje as equações ❶ e ❷.

Rearranje ❶.

$$2\left[77 - (29a + b)\right] \times (-29) + \cdots + 2\left[84 - (30a + b)\right] \times (-30) = 0$$

$$\left[77 - (29a + b)\right] \times (-29) + \cdots + \left[84 - (30a + b)\right] \times (-30) = 0$$

DIVIDA AMBOS OS LADOS POR 2

$$29\left[(29a + b) - 77\right] + \cdots + 30\left[(30a + b) - 84\right] = 0$$

$$(29 \times 29a + 29 \times b - 29 \times 77) + \cdots + (30 \times 30a + 30 \times b - 30 \times 84) = 0$$

MULTIPLIQUE POR -1

$$\left(29^2 + \cdots + 30^2\right)a + (29 + \cdots + 30)\,b - (29 \times 77 + \cdots + 30 \times 84) = 0$$

MULTIPLIQUE

❸

ISOLE a E b

Rearranje ❷.

$$2\left[77 - (29a + b)\right] \times (-1) + \cdots + 2\left[84 - (30a + b)\right] \times (-1) = 0$$

$$\left[77 - (29a + b)\right] \times (-1) + \cdots + \left[84 - (30a + b)\right] \times (-1) = 0$$

DIVIDA AMBOS OS LADOS POR 2

$$\left[(29a + b) - 77\right] + \cdots + \left[(30a + b) - 84\right] = 0$$

MULTIPLIQUE POR -1

$$(29 + \cdots + 30)a + \underbrace{b + \cdots + b}_{14} - (77 + \cdots + 84) = 0$$

ISOLE a E b

$$(29 + \cdots + 30)a + 14b - (77 + \cdots + 84) = 0$$

$$14b = (77 + \cdots + 84) - (29 + \cdots + 30)a$$

SUBTRAIA $14b$ DE AMBOS OS LADOS E MULTIPLIQUE POR -1

$$b = \frac{77 + \cdots + 84}{14} - \frac{29 + \cdots + 30}{14}a$$

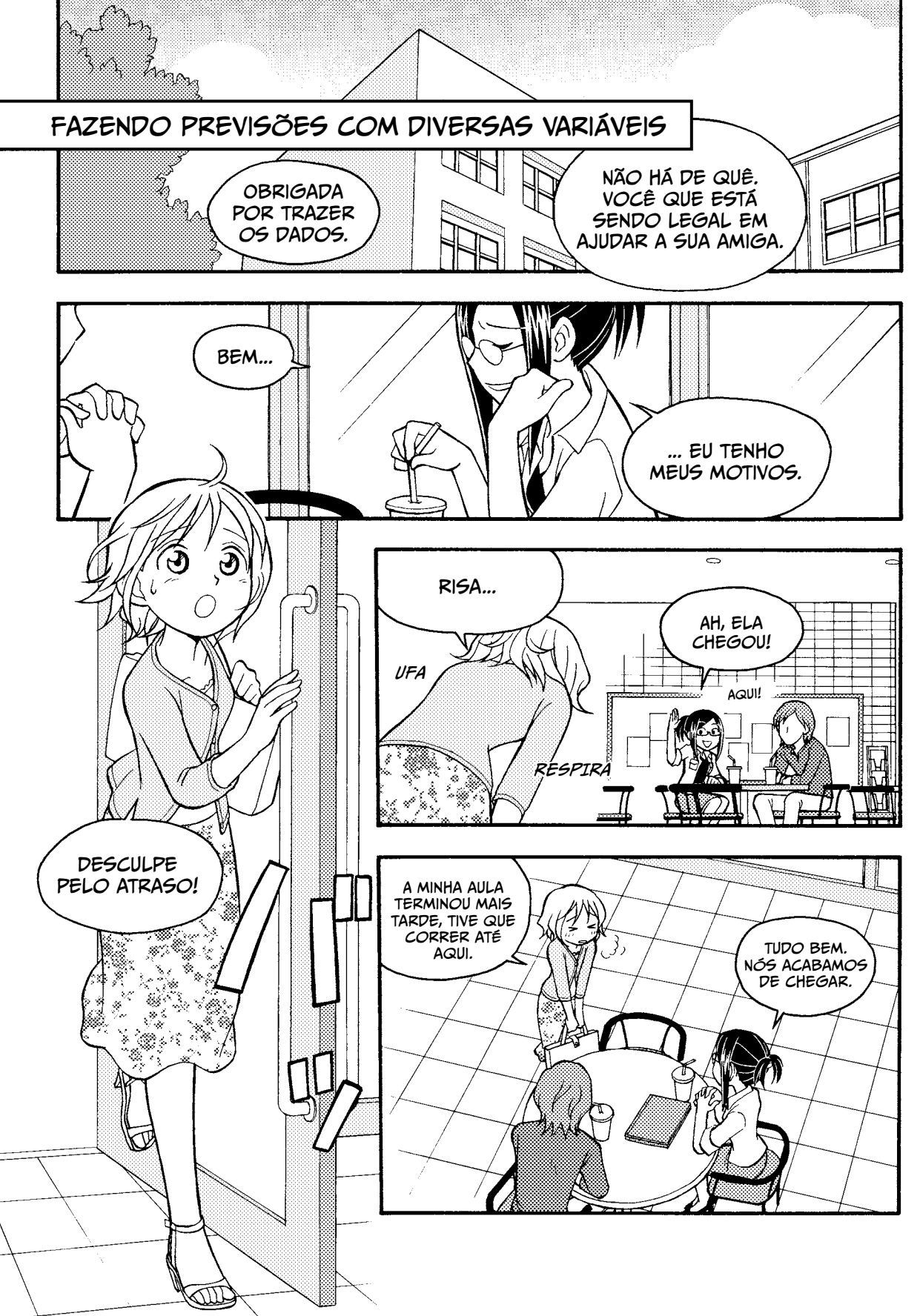
$$b = \bar{y} - \bar{x}a$$

ISOLE b NO LADO ESQUERDO DA EQUAÇÃO

❹ $b = \bar{y} - \bar{x}a$

❺

OS COMPONENTES EM ❹ SÃO AS MÉDIAS DE y E x


FAZENDO PREVISÕES COM DIVERSAS VARIÁVEIS
OBRIGADA POR TRAZER OS DADOS.
NÃO HÁ DE QUÊ. VOCÊ QUE ESTÁ SENDO LEGAL EM AJUDAR A SUA AMIGA.
BEM...
... EU TENHO MEUS MOTIVOS.
RISA...
UFA
RESPIRA
AH, ELA CHEGOU!
AQUI!
DESCULPE PELO ATRASO!
A MINHA AULA TERMINOU MAIS TARDE, TIVE QUE CORRER ATÉ AQUI.
TUDO BEM. NÓS ACABAMOS DE CHEGAR.

ANALISANDO A POPULAÇÃO COM A ANÁLISE DE REGRESSÃO MÚLTIPLA

Vamos revisar o procedimento utilizado na análise de regressão múltipla, mostrado na página 112.

1. Desenhe um gráfico de dispersão de cada variável preditora e da variável resultado para analisar se elas estão relacionadas.
2. Calcule a equação de regressão múltipla.
3. Examine a acurácia da equação de regressão múltipla.
4. Faça a análise do teste de variância (ANOVA).
5. Calcule o intervalo de confiança da população.
6. Faça uma previsão!

Assim como no Capítulo 2, falamos sobre todos os passos como se eles fossem obrigatórios. Porém, na realidade os passos 4 e 5 podem ser ignorados na análise de determinados conjuntos de dados.

As Padarias Kazami são, atualmente, compostas de 10 filiais, e entre essas filiais apenas uma (a padaria Yumenooka) possui área de 10 tsubo[1] e está a 80 metros da estação mais próxima. Porém, Risa calculou o intervalo de confiança para uma população de padarias que possuem área de 10 tsubo e que estão a 80 metros da estação mais próxima. Por que ela fez isso?

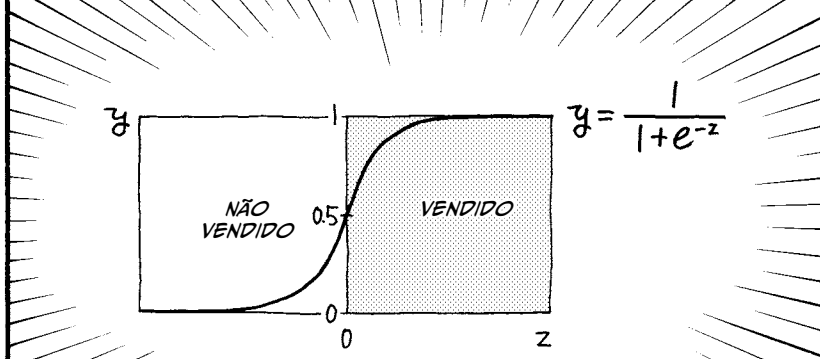
Bem, é possível que as Padarias Kazami inaugurem outra padaria com 10 tsubo de área e que também estejam a 80 metros da estação de trem mais próxima. Se as Padarias Kazami continuarem crescendo, poderá haver dezenas de padarias que se encaixarão nessa descrição. Quando Risa fez essa análise, ela presumiu que mais padarias de 10 tsubo a 80 metros de uma estação de trem seriam inauguradas no futuro.

A utilidade dessa hipótese é discutida. A padaria Yumenooka vende mais do que qualquer outra padaria, o que pode fazer com que a família Kazami abra mais padarias iguais a ela. Porém, a próxima padaria que será inaugurada, a Padaria Isebashi, terá 10 tsubo de área e estará a 110 metros da estação mais próxima. Na verdade, provavelmente não seria necessário analisar uma população tão específica de padarias, pois Risa poderia ter ignorado o cálculo do R^2 ajustado para realizar a previsão. Porém, ela o calculou para demonstrar a Miu todos os passos.

1. Lembre-se de que 1 tsubo é equivalente a aproximadamente 3,34 metros quadrados.

A NORNS NÃO VENDE O BOLO ESPECIAL TODOS OS DIAS. AS PESSOAS SÓ O COMPRAM PARA OCASIÕES ESPECIAIS, COMO UM ANIVERSÁRIO.

OU UM CASAMENTO...



ENTÃO HOJE VAMOS ENCONTRAR UMA EQUAÇÃO DE REGRESSÃO LOGÍSTICA PARA PREVER SE O PRATO ESPECIAL DA NORNS VAI VENDER EM UM DIA ESPECÍFICO.

AH! ISSO PARECE TÃO DIVERTIDO!

ACHO QUE VAMOS PRECISAR DE DADOS AMOSTRAIS. MAS O QUE VAMOS MEDIR? O QUE AFETA A VENDA?

ÓTIMA PERGUNTA!

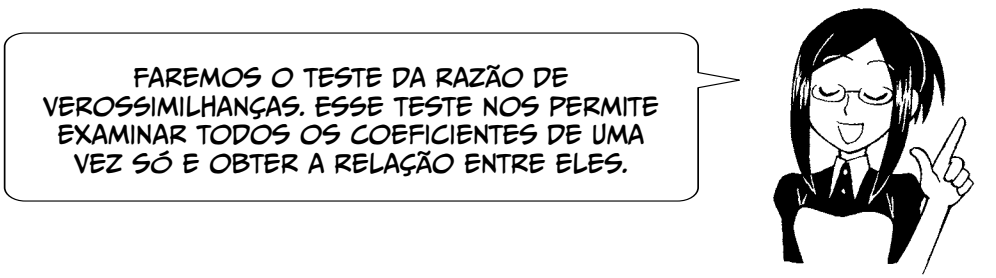
TENHO TENTADO DESCOBRIR ISSO FAZ UM TEMPO. PERCEBI QUE MAIS PESSOAS COMPRAM O PRATO ESPECIAL EM DIAS QUENTES. ALÉM DISSO, TENHO A IMPRESSÃO DE QUE VENDEMOS MAIS BOLOS NAS QUARTAS, SÁBADOS E DOMINGOS.

É MESMO?

SIM. TEMOS MAIS CLIENTES NO FINAL DE SEMANA...

HORA DA FESTA!

E NAS QUARTAS, UM GRANDE GRUPO DE FÃS DE MANGÁ FAZ UMA REUNIÃO AQUI. ELES SEMPRE FAZEM MUITOS PEDIDOS!



OS PASSOS DO TESTE DA RAZÃO DE VEROSSIMILHANÇAS

Passo 1	Defina a população.	Todos os dias em que o prato especial da Norns é vendido, comparando quartas, sábados e domingos com os dias restantes, para cada temperatura.
Passo 2	Defina uma hipótese nula e uma hipótese alternativa.	Hipótese nula é $A_1 = 0$ e $A_2 = 0$. Hipótese alternativa $A_1 \neq 0$ ou $A_2 \neq 0$.
Passo 3	Escolha qual teste de hipóteses conduzir.	Vamos calcular o teste da razão de verossimilhanças.
Passo 4	Escolha o nível de significância	Vamos usar um nível de significância de 0,05.
Passo 5	Calcule a estatística de teste a partir dos dados amostrais.	A estatística de teste é: $2[L - n_1\log_e(n_1) - n_0\log_e(n_0) + (n_1 + n_0)\log_e(n_1 + n_0)]$ Ao colocarmos nossos dados, obtemos: $2[-8.9010 - 8\log_e 8 - 13\log_e 13 + (8 + 13)\log_e(8 + 13)] = 10.1$ A estatística de teste segue uma distribuição qui-quadrado com 2 graus de liberdade (o número de variáveis preditoras), caso a hipótese nula seja verdadeira.
Passo 6	Determine se o *p*-valor para a estatística de teste obtida no Passo 5 é menor do que o nível de significância.	O nível de significância é 0,05. O valor da estatística de teste é 10,1, então o *p*-valor é 0,006. Finalmente, 0,006<0,05.*
Passo 7	Decida se você pode rejeitar a hipótese nula.	Uma vez que o *p*-valor é menor do que o nível de significância, rejeitamos a hipótese nula.

* Os passos para obter o *p*-valor em uma distribuição qui-quadrado estão explicados na página 205.