GUIA MANGÁ DE
ESTATÍSTICA

SHIN TAKAHASHI
TREND-PRO, CO., LTD.

Ohmsha

no starch press

novatec

Original Japanese-language edition Manga de Wakaru Toukeigaku ISBN 4-274-06570-7 © 2004 by Shin Takahashi and TREND-PRO Co., Ltd., published by Ohmsha, Ltd.

English-language edition The Manga Guide to Statistics ISBN 978-1-59327-189-3 © 2009 by Shin Takahashi and TREND-PRO Co., Ltd., 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 de Estatística ISBN 978-85-7522-168-6 © 2009 by Shin Takahashi and TREND-PRO Co., Ltd., published by Novatec Editora Ltda.

Edição original em Japonês Manga de Wakaru Toukeigaku ISBN 4-274-06570-7 © 2004 por Shin Takahashi e TREND-PRO Co., Ltd., publicado pela Ohmsha, Ltd.

Edição em Inglês The Manga Guide to Statistics ISBN 978-1-59327-189-3 © 2009 por Shin Takahashi e TREND-PRO Co., Ltd., 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 de Estatística ISBN 978-85-7522-168-6 © 2009 por Shin Takahashi e TREND-PRO Co., Ltd., publicado pela Novatec Editora Ltda.

Copyright © 2010 da Novatec Editora Ltda.

Todos os direitos reservados e protegidos pela Lei 9.610 de 19/02/1998.
É proibida a reprodução desta obra, mesmo parcial, por qualquer processo, sem prévia autorização, por escrito, do autor e da Editora.

Editor: Rubens Prates
Ilustração: Iroha Inoue
Tradução e revisão gramatical: Lia Gabriele Regius
Revisão técnica: Dennis Cintra Leite
Editoração eletrônica: Camila Kuwabata e Carolina Kuwabata

ISBN: 978-85-7522-168-6

Histórico de impressões:

Junho/2011 Primeira reimpressão
Janeiro/2010 Primeira edição

NOVATEC EDITORA LTDA.
Rua Luís Antônio dos Santos 110
02460-000 – São Paulo, SP – Brasil
Tel.: +55 11 2959-6529
Fax: +55 11 2950-8869
E-mail: novatec@novatec.com.br
Site: www.novatec.com.br
Twitter: twitter.com/novateceditora
Facebook: facebook.com/novatec
LinkedIn: linkedin.com/in/novatec

Dados Internacionais de Catalogação na Publicação (CIP) (Câmara Brasileira do Livro, SP, Brasil)

Takahashi, Shin
Guia mangá de estatística / Shin Takahashi, Trend-pro Co. ; [tradução e revisão Lia Gabriele Regius]. -- São Paulo : Novatec Editora ; Tokyo : Ohmsha ; São Francisco : No Starch Press, 2010.

Título original: The manga guide to statistics
ISBN 978-85-7522-168-6

1. Estatística matemática - História em quadrinhos 2. Estatística matemática - História em quadrinhos - Obras de divulgação I. Trend-pro Co.. II. Título.

09-11945 CDD-519.5

Índices para catálogo sistemático:

1. Estatística : Matemática em quadrinhos 519.5

OGF24052011

SUMÁRIO

Prefácio vii
Nosso prólogo: ♥ apaixone-se pela estatística ♥ 1

1 13
Determinação de tipos de dados 13

1. Dados categóricos e dados numéricos 14
2. Um exemplo de dados categóricos de difícil classificação 20
3. Como as respostas de múltipla escolha são administradas na prática 28
Exercícios e respostas 29
Resumo 29

2 31
Compreendendo o quadro geral: a essência dos dados numéricos 31

1. Tabelas de distribuição de frequências e histogramas 32
2. Média (valor médio) 40
3. Mediana 44
4. Desvio-padrão 48
5. intervalo de classe de uma tabela de frequência 54
6. Inferência estatística e estatística descritiva 57
Exercícios e respostas 57
Resumo 58

3 59
Compreendendo o quadro geral: a essência dos dados categóricos 59

1. Tabulações cruzadas 60
Exercícios e respostas 64
Resumo 64

4 65
Valor-padrão e valor do desvio 65

1. Normalização e valor-padrão 66
2. Características do valor-padrão 73
3. Valor do desvio 74
4. Interpretação do valor do desvio 76
Exercícios e respostas 78
Resumo 80

5 81
Vamos calcular a probabilidade 81

1. Função de densidade de probabilidade 82
2. Distribuição normal 86

3. Distribuição normal padrão....89
Exemplo I....95
Exemplo II....97
4. Distribuição qui-quadrado....99
5. Distribuição T....106
6. Distribuição f....106
7. Distribuições e Excel....107
Exercícios e respostas....108
Resumo....109

6....111
Estudo da relação entre duas variáveis....111

1. Coeficiente de correlação....116
2. Taxa de correlação....121
3. Coeficiente de Cramer....127
Exercícios e respostas....138
Resumo....142

7....143
Explorando os testes de hipótese....143

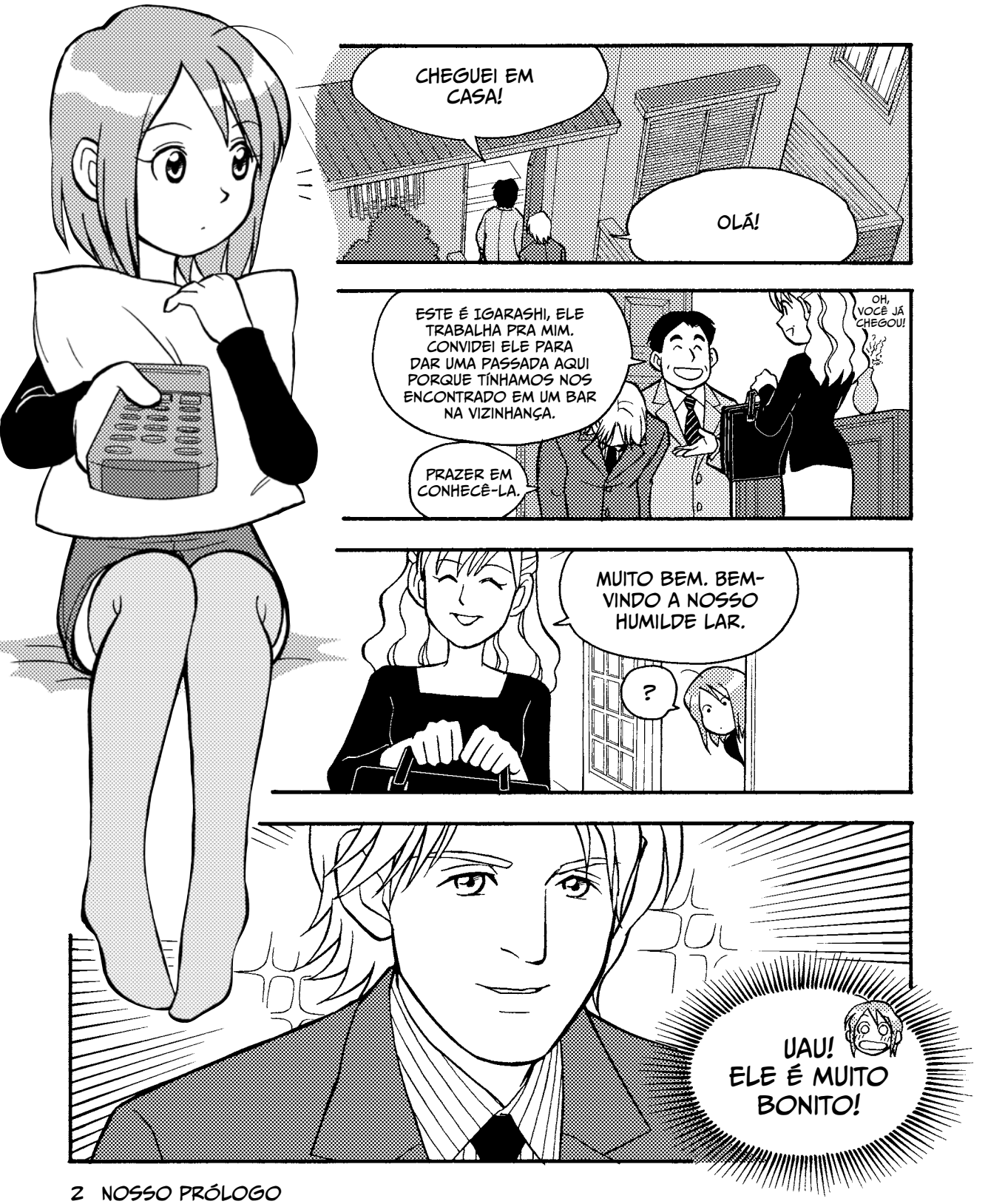
1. Testes de hipótese....144
2. Teste qui-quadrado de independência....151
Explicação....152
Exercício....157
Reflexão....158
Resposta....160
3. Hipóteses nulas e hipóteses alternativas....170
4. Valor P e procedimento para testes de hipótese....175
5. Testes de independência e testes de homogeneidade....184
Exemplo....184
Procedimento....185
6. Conclusões de testes de hipótese....187
Exercícios e respostas....188
Resumo....189

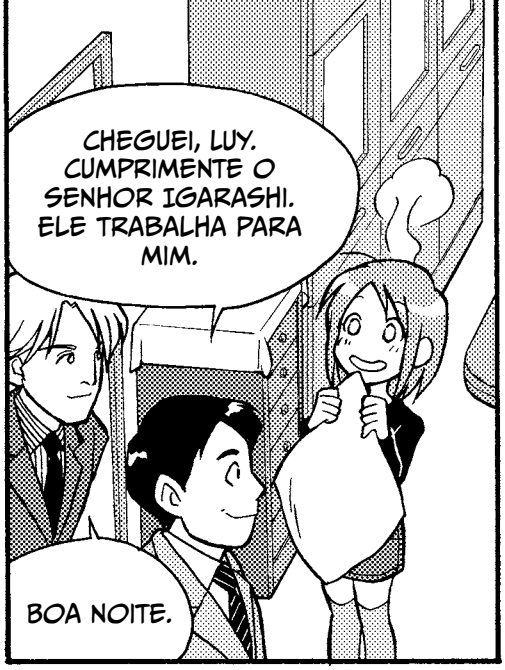
apêndice
Cálculos com o uso do Excel....191

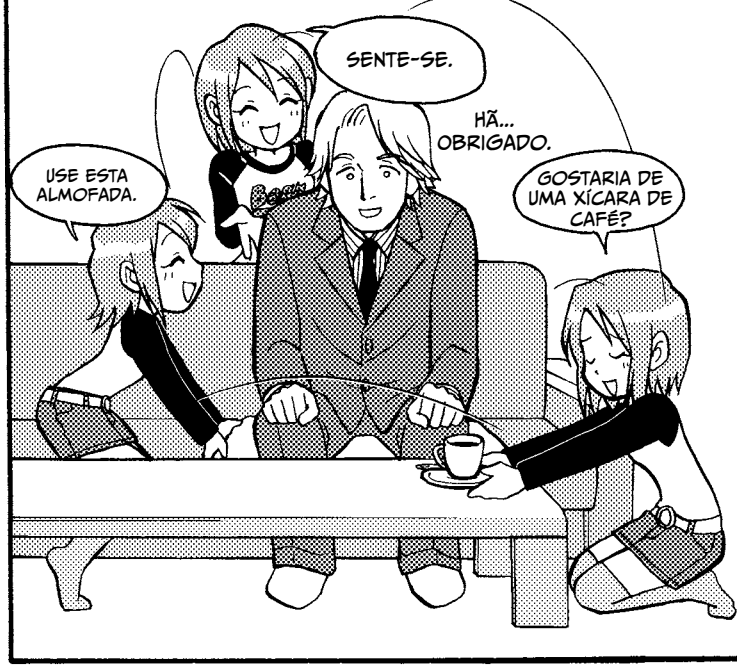
1. Criação de uma tabela de frequência....192
2. Cálculo da média aritmética, da mediana e do desvio-padrão....195
3. Criação de uma tabulação cruzada....197
4. Cálculo do valor-padrão e do padrão do desvio....199
5. Cálculo da probabilidade da distribuição normal padrão....204
6. Cálculo do ponto no eixo horizontal da distribuição qui-quadrado....205
7. Cálculo do coeficiente de correlação....206
8. Realização de testes de independência....208

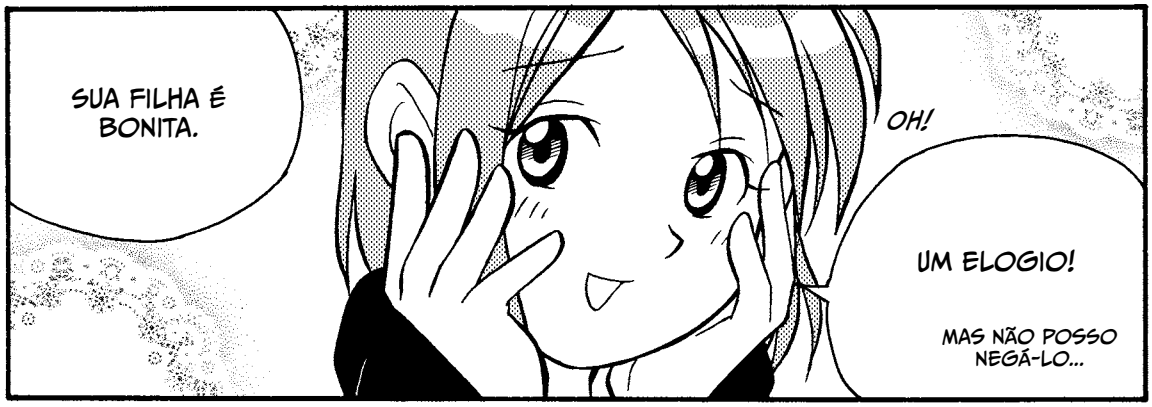
Índice....213

NOSSO PRÓLOGO:
♥ APAIXONE-SE PELA ESTATÍSTICA ♥

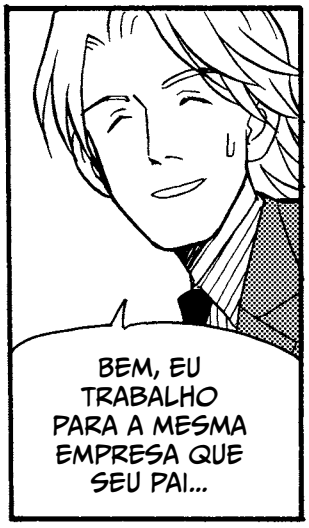

CHEGUEI EM CASA!
OLÁ!
ESTE É IGARASHI, ELE TRABALHA PRA MIM. CONVIDEI ELE PARA DAR UMA PASSADA AQUI PORQUE TÍNHAMOS NOS ENCONTRADO EM UM BAR NA VIZINHANÇA.
OH, VOCÊ JÁ CHEGOU!
PRAZER EM CONHECÊ-LA.
MUITO BEM. BEM-VINDO A NOSSO HUMILDE LAR.
?
UAU! ELE É MUITO BONITO!


CHEGUEI, LUY. CUMPRIMENTE O SENHOR IGARASHI. ELE TRABALHA PARA MIM.
BOA NOITE.


SENTE-SE.
HÃ... OBRIGADO.
USE ESTA ALMOFADA.
GOSTARIA DE UMA XÍCARA DE CAFÉ?


SUA FILHA É BONITA.
OH!
UM ELOGIO!
MAS NÃO POSSO NEGÁ-LO...

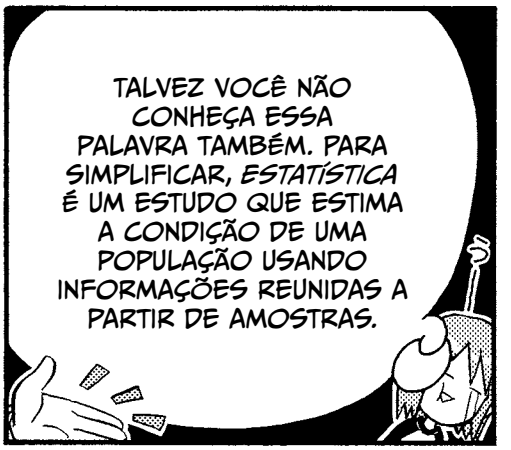
SENHOR IGARASHI, COMO É SEU TRABALHO?

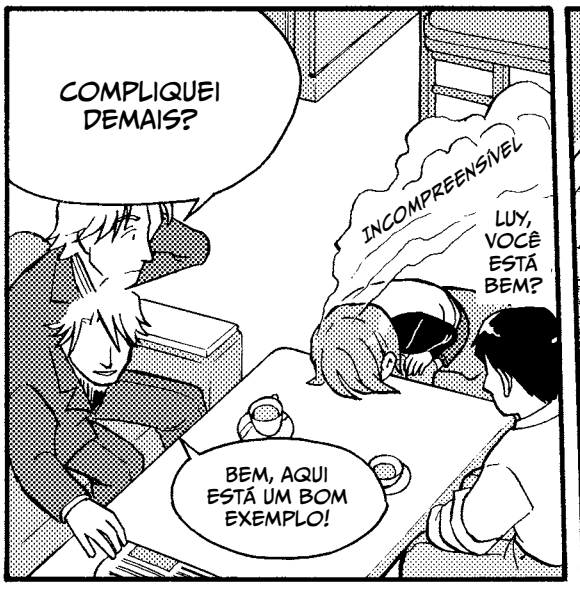

BEM, EU TRABALHO PARA A MESMA EMPRESA QUE SEU PAI...

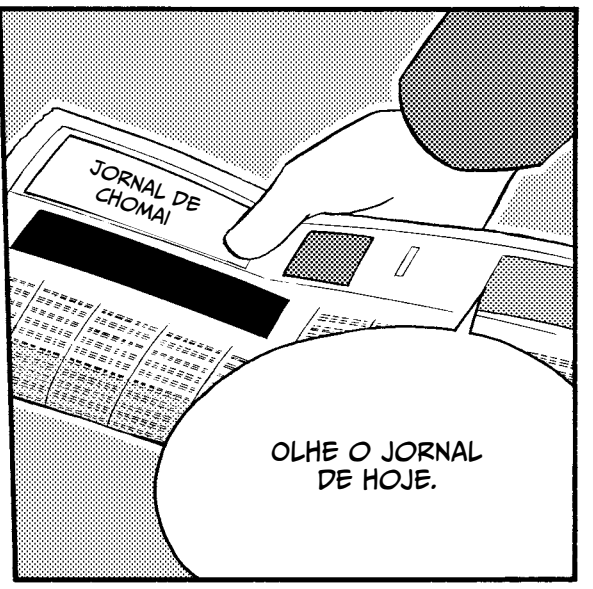
RESUMINDO, EU TRABALHO COM MARKETING.
MARKETING?

PARA SER MAIS EXATO, FAÇO PESQUISAS DE MERCADO USANDO ESTATÍSTICA... MAS ACHO QUE A PALAVRA MARKETING NÃO É MUITO CLARA PARA UMA JOVEM GAROTA COMO VOCÊ.
DESCULPE, NUNCA OUVI FALAR DELA.

VOCÊ É SINCERA. SABE O QUE SIGNIFICA ESTATÍSTICA, ENTÃO?
HMMMM...

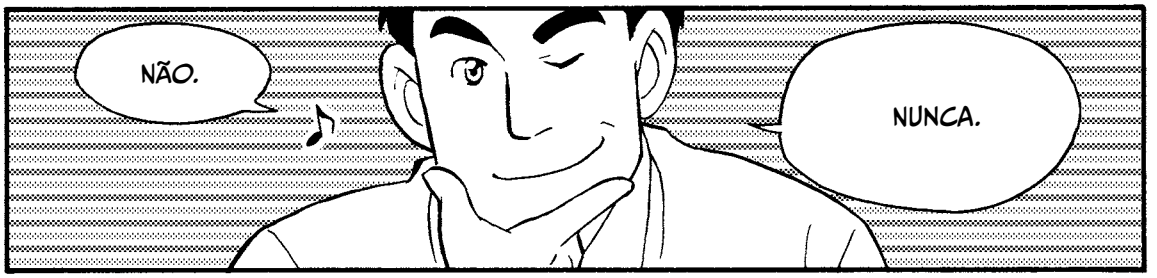

TALVEZ VOCÊ NÃO CONHEÇA ESSA PALAVRA TAMBÉM. PARA SIMPLIFICAR, *ESTATÍSTICA* É UM ESTUDO QUE ESTIMA A CONDIÇÃO DE UMA POPULAÇÃO USANDO INFORMAÇÕES REUNIDAS A PARTIR DE AMOSTRAS.


COMPLIQUEI DEMAIS?
INCOMPREENSÍVEL
LUY, VOCÊ ESTÁ BEM?
BEM, AQUI ESTÁ UM BOM EXEMPLO!

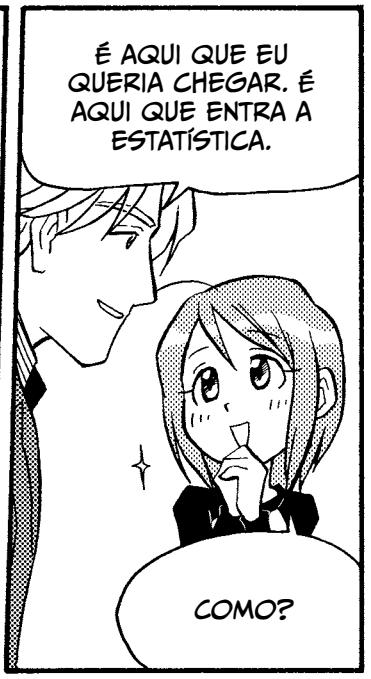

JORNAL DE CHOMAI
OLHE O JORNAL DE HOJE.

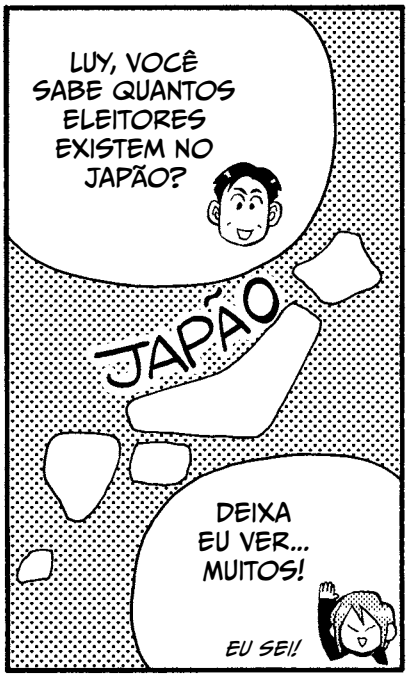
AQUI DIZ QUE "DE ACORDO COM UMA PESQUISA DO JORNAL DE CHOMAI, A TAXA DE APROVAÇÃO DO GOVERNO É DE 39%."
E O QUE ISSO QUER DIZER?

NUNCA FUI ENTREVISTADO PELO JORNAL DE CHOMAI SOBRE O GOVERNO.
E VOCÊ, TAKATSU?


NÃO.
NUNCA.

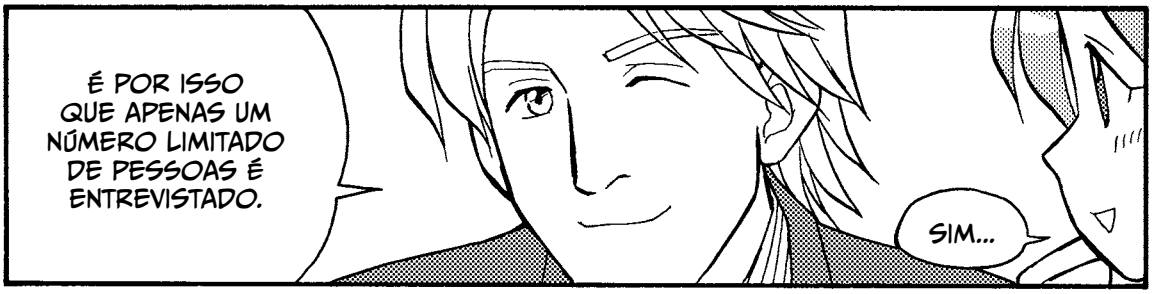
HUM... NENHUM DE VOCÊS FOI ENTREVISTADO, MAS A TAXA DE APROVAÇÃO DO GOVERNO ESTÁ NO JORNAL.
ESTRANHO. VOCÊS TÊM O DIREITO DE VOTAR, NÃO?


É AQUI QUE EU QUERIA CHEGAR. É AQUI QUE ENTRA A ESTATÍSTICA.
COMO?


LUY, VOCÊ SABE QUANTOS ELEITORES EXISTEM NO JAPÃO?
JAPÃO
DEIXA EU VER... MUITOS!
EU SEI!

CERTO.
É POSSÍVEL OBTER A TAXA DE APROVAÇÃO PRECISA DO GOVERNO SE VOCÊ PUDESSE ENTREVISTAR CADA UM DOS ELEITORES.
SIM!

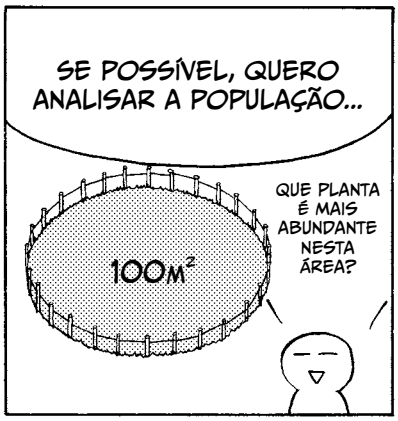
NO ENTANTO, É IMPOSSÍVEL ENTREVISTAR TODOS OS ELEITORES. É MUITA GENTE!
DEFINITIVAMENTE.
DESISTO!
ISSO NÃO FAZ SENTIDO!
CONCORDO!

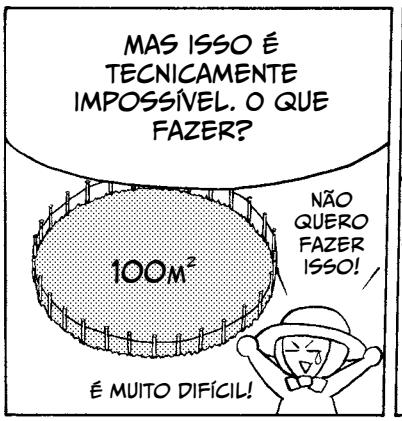

É POR ISSO QUE APENAS UM NÚMERO LIMITADO DE PESSOAS É ENTREVISTADO.
SIM...

ENTENDEU, LUY? O GRUPO QUE REALMENTE DEVERIA SER PESQUISADO CHAMA-SE POPULAÇÃO. UM GRUPO FORMADO POR AMOSTRAS SELECIONADAS DE UMA POPULAÇÃO CHAMA-SE AMOSTRA. SÃO TERMOS ESTATÍSTICOS.
POPULAÇÃO?
AMOSTRA?

MEU PAI ESTÁ ME TORTURANDO FALANDO DE ASSUNTOS TÃO DIFÍCEIS!
AI!
GLUP! NÃO ESTOU TORTURANDO VOCÊ!

O QUE ELE ESTÁ DIZENDO É... NO CASO DA TAXA DE APROVAÇÃO DO GOVERNO, A POPULAÇÃO É COMPOSTA POR TODOS OS ELEITORES.
AQUI DIZ QUE A PESQUISA FOI REALIZADA COM 2.000 PESSOAS. ASSIM, A AMOSTRA É FORMADA POR ESSAS 2.000 PESSOAS.
TODOS OS ELEITORES
AMOSTRAGEM
2,000 ELEITORES
POPULAÇÃO
AMOSTRA
ENTENDO.

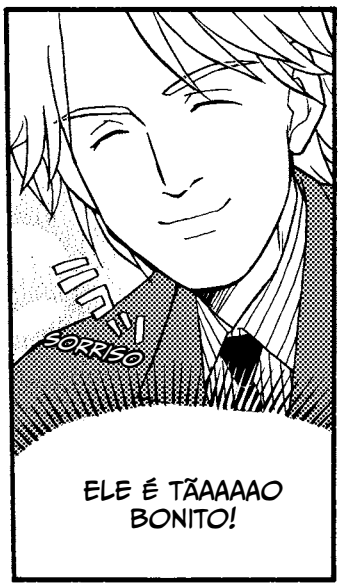

SE POSSÍVEL, QUERO ANALISAR A POPULAÇÃO...
100M²
QUE PLANTA É MAIS ABUNDANTE NESTA ÁREA?

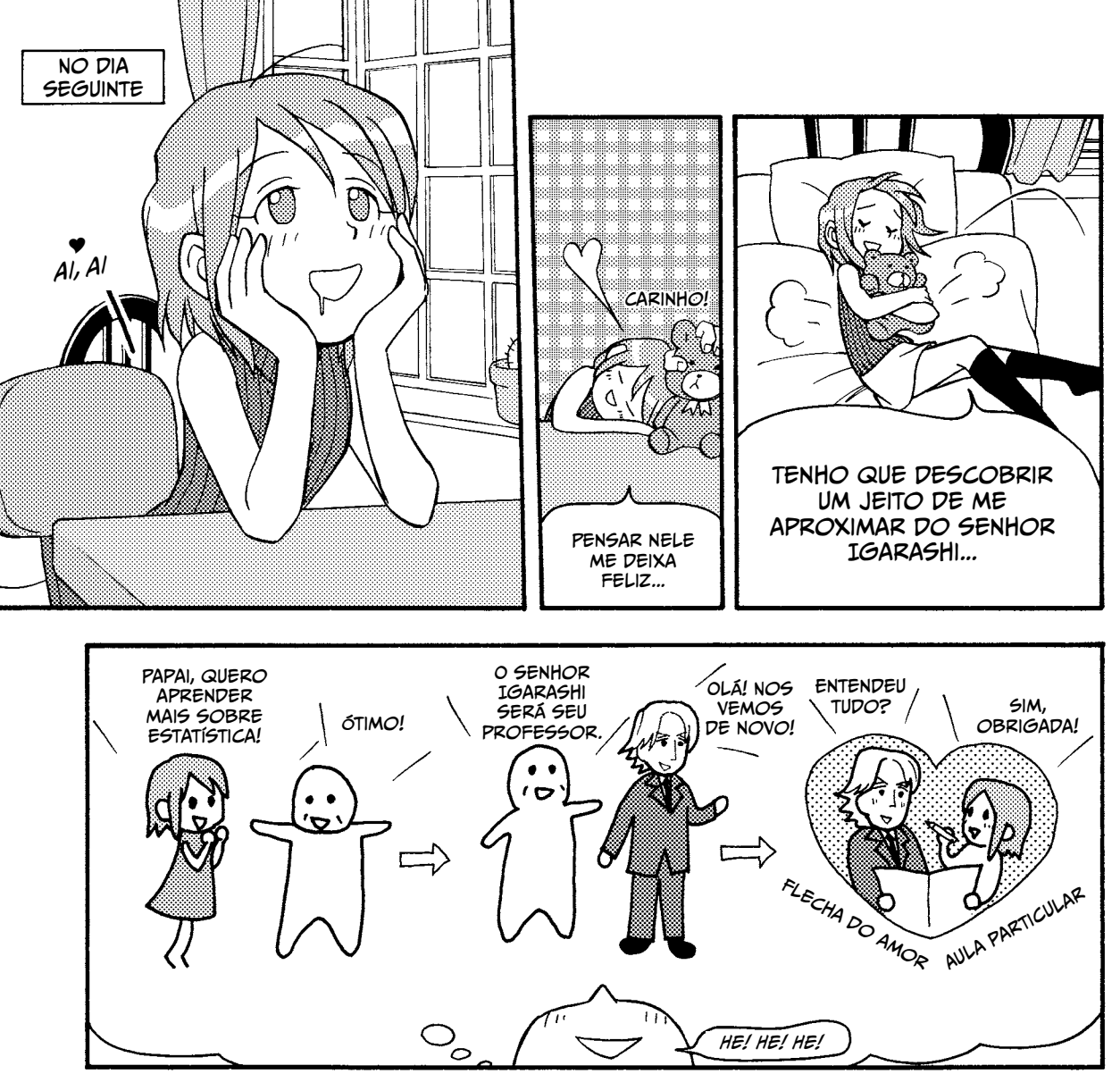

MAS ISSO É TECNICAMENTE IMPOSSÍVEL. O QUE FAZER?
100M²
NÃO QUERO FAZER ISSO!
É MUITO DIFÍCIL!

COMO POSSO TER UMA IDEIA DO STATUS DA POPULAÇÃO? NÃO PRECISA SER RIGOROSAMENTE PRECISO, MAS TINHA QUE SER O MAIS ACURADO POSSÍVEL.
100M²
HMMMM...

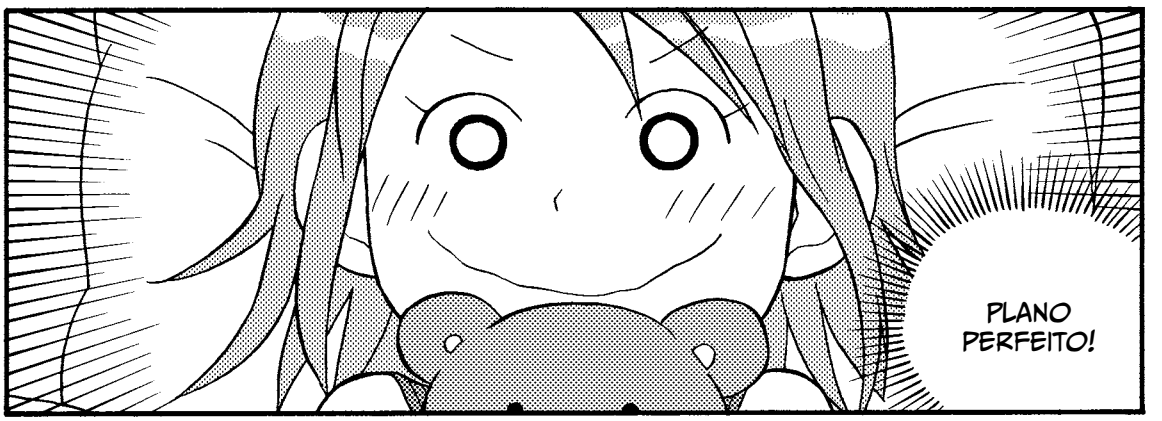
VOU ANALISAR APENAS 1 METRO QUADRADO
1M²
E É AQUI QUE A ESTATÍSTICA PODE AJUDAR MUITO.
OH! POR FAVOR, ME CONTE MAIS!

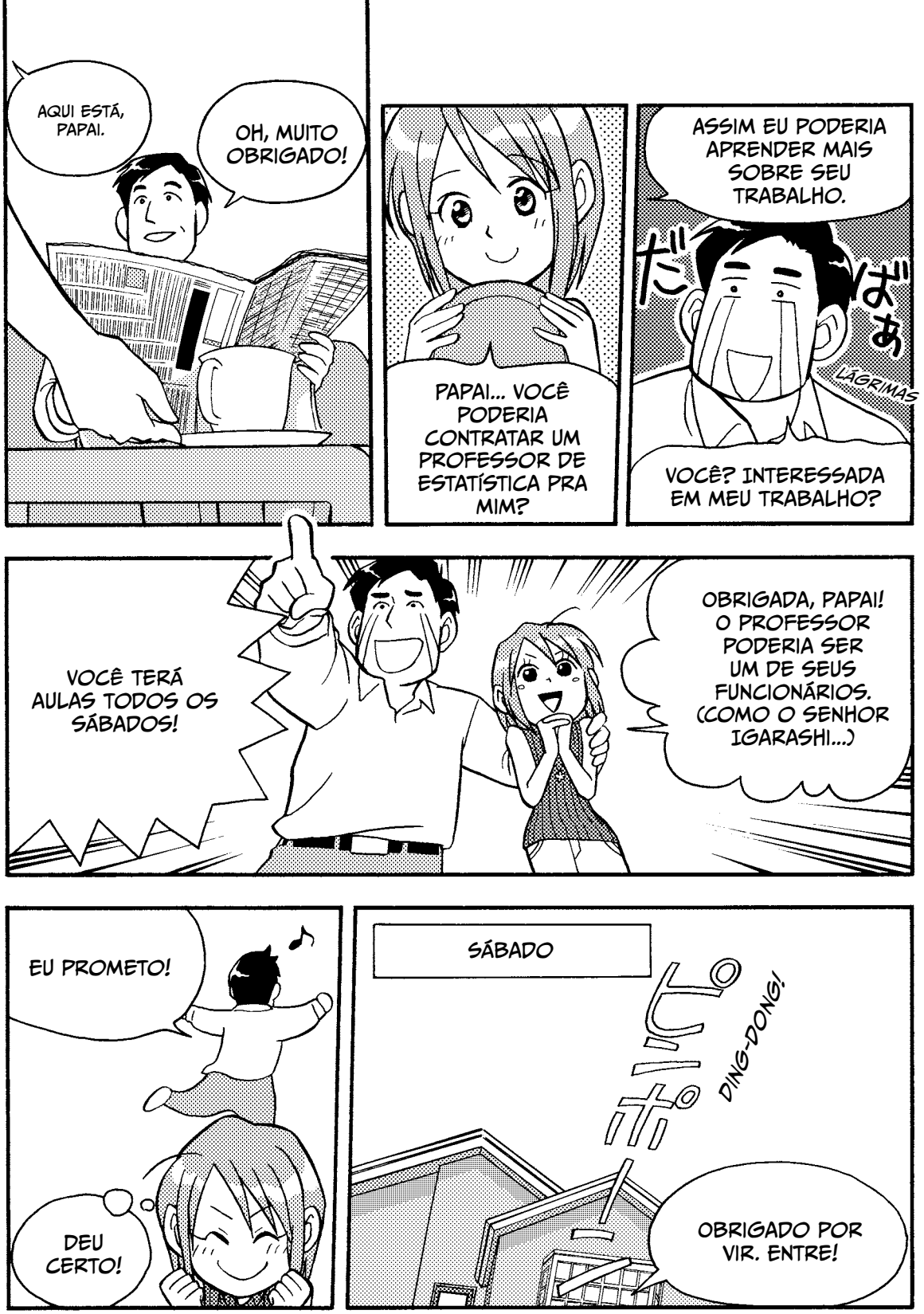
BEM, TALVEZ NA PRÓXIMA VEZ.


SORRISO
ELE É TÃAAAAO BONITO!


NO DIA SEGUINTE
AI, AI
CARINHO!
PENSAR NELE ME DEIXA FELIZ...
TENHO QUE DESCOBRIR UM JEITO DE ME APROXIMAR DO SENHOR IGARASHI...

PAPAI, QUERO APRENDER MAIS SOBRE ESTATÍSTICA!
ÓTIMO!
O SENHOR IGARASHI SERÁ SEU PROFESSOR.
OLÁ! NOS VEMOS DE NOVO!
ENTENDEU TUDO?
SIM, OBRIGADA!
FLECHA DO AMOR
AULA PARTICULAR
HE! HE! HE!

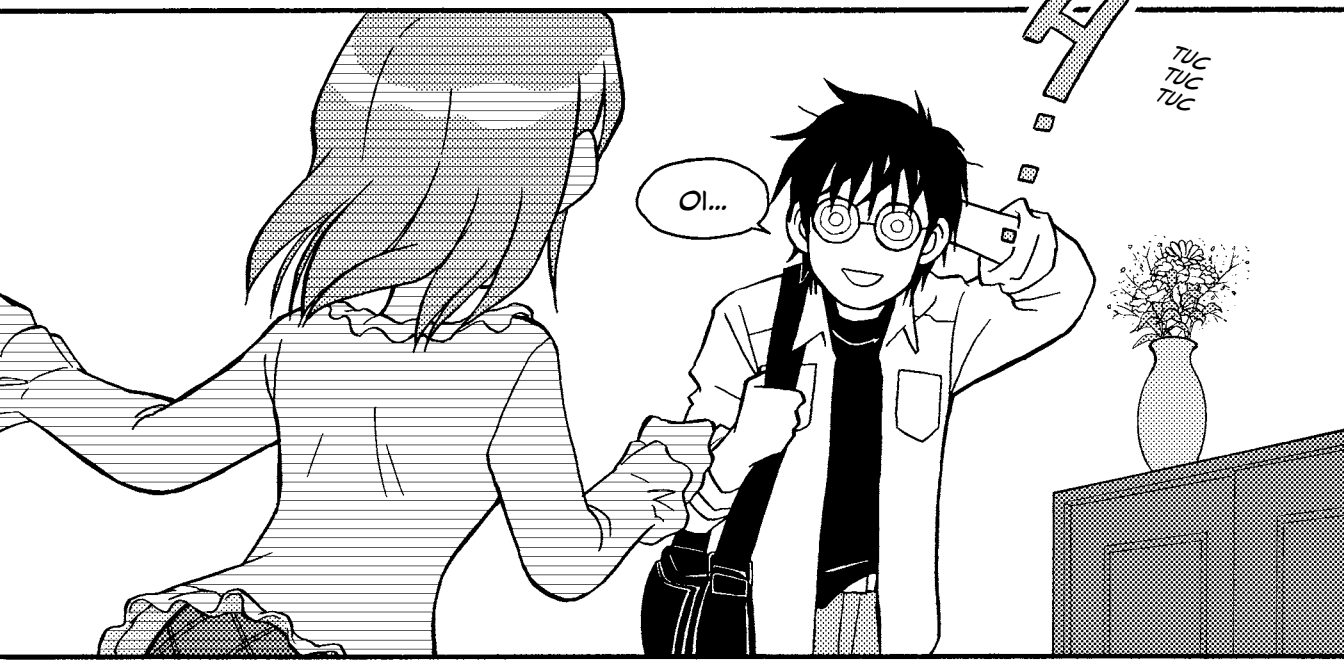

PLANO PERFEITO!


AQUI ESTÁ, PAPAI.
OH, MUITO OBRIGADO!
PAPAI... VOCÊ PODERIA CONTRATAR UM PROFESSOR DE ESTATÍSTICA PRA MIM?
ASSIM EU PODERIA APRENDER MAIS SOBRE SEU TRABALHO.
だばあ
LÁGRIMAS
VOCÊ? INTERESSADA EM MEU TRABALHO?
VOCÊ TERÁ AULAS TODOS OS SÁBADOS!
OBRIGADA, PAPAI! O PROFESSOR PODERIA SER UM DE SEUS FUNCIONÁRIOS. (COMO O SENHOR IGARASHI...)
EU PROMETO!
DEU CERTO!
SÁBADO
ピンポーン
DING-DONG!
OBRIGADO POR VIR. ENTRE!

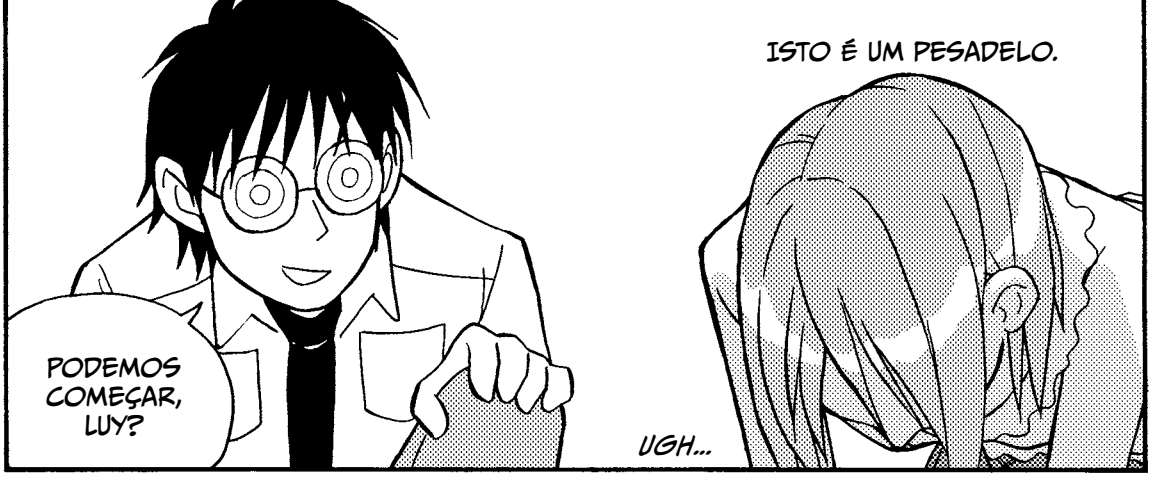
ELE CHEGOU!

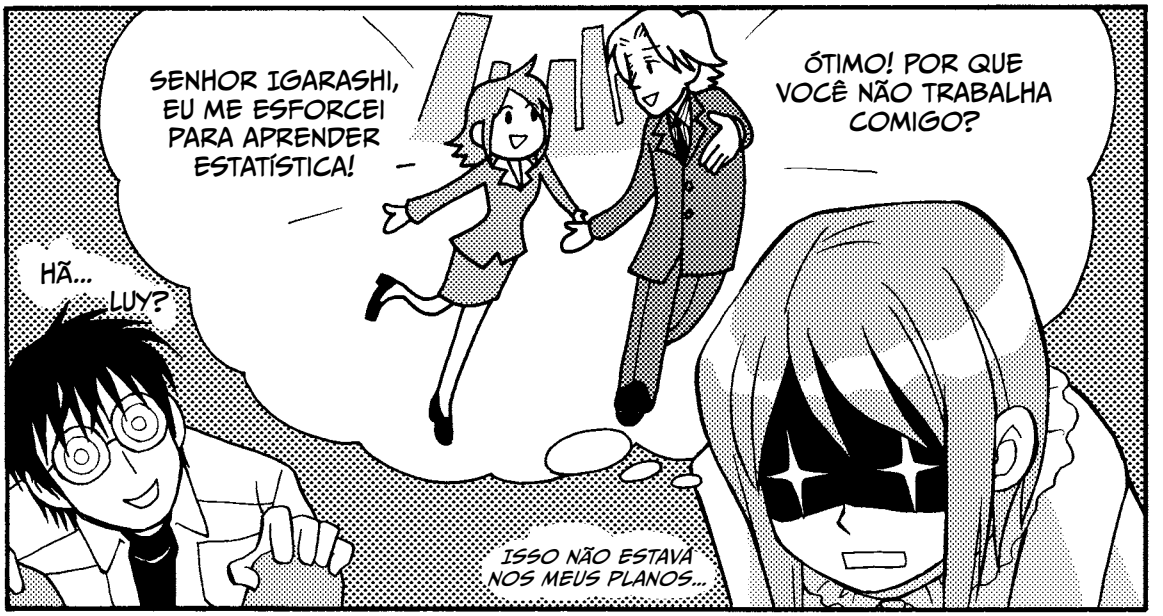
LUY, SEU PROFESSOR CHEGOU!

ESTOU INDO!
タッタッタ


TUC TUC TUC
OI...

QUEM É ESSE CARA?!
カッ!
LUY, ESTE É MEU FUNCIONÁRIO, MAMORU YAMAMOTO.
COMO VAI?
PAI... O SENHOR IGARASHI NÃO VEM?
IGARASHI? MAMORU MORA MAIS PERTO DAQUI. E ELE ENSINA MUITO BEM TAMBÉM.
ESTUDE BASTANTE!
HO
HO
HO!


PODEMOS COMEÇAR, LUY?
ISTO É UM PESADELO.
UGH...


SENHOR IGARASHI, EU ME ESFORCEI PARA APRENDER ESTATÍSTICA!
ÓTIMO! POR QUE VOCÊ NÃO TRABALHA COMIGO?
HÃ... LUY?
ISSO NÃO ESTAVA NOS MEUS PLANOS...

TENHO UMA IDEIA!
?

APROVEITAREI ESSA CHANCE PARA APRENDER TUDO SOBRE ESTATÍSTICA, E DEPOIS PROCURAREI O SENHOR IGARASHI!
NUNCA DESISTIREI!

VAMOS COMEÇAR AGORA!
ER... OK.
ASSIM, A LIÇÃO COMEÇA.

4. INTERPRETAÇÃO DO VALOR DO DESVIO

Preste atenção ao interpretar valores do desvio. Como explicado na página 74, a definição do valor do desvio é:

$$\text{valor do desvio} = \text{valor-padrão} \times 10 + 50 = \frac{(\text{cada valor} - \text{média})}{\text{desvio-padrão}} \times 10 + 50$$

Como informado na página 62, a turma de Luy tem um total de 40 alunos, e como indicado na página 40, há 18 meninas na classe. O exemplo do valor de desvio na página 69 não é para a classe toda, somente para as meninas. Se a turma inteira fosse investigada, a média e o desvio-padrão teriam sido diferentes daqueles para as meninas isoladamente. Naturalmente, os valores de desvio de Luy e Yumi teriam sido diferentes também. Na verdade, quando todos os alunos da classe são considerados, Luy tem o maior valor do desvio. A Tabela 4-1 mostra os resultados da prova para a classe toda. Tente calcular o valor do desvio.

Antecipando a resposta, saiba que o valor do desvio para o teste de história do Luy é 59,1, e da prova de biologia da Yumi é 56,7.

Suponha que a mesma prova seja aplicada aos alunos das classes 1 e 2. A média e o desvio-padrão da classe 1 são calculados individualmente, e os valores do desvio são obtidos de acordo com esses resultados. Da mesma forma, a média, o desvio-padrão e os valores do desvio da turma 2 são obtidos. O aluno A da classe 1 tem um valor do desvio de 57. O aluno B na turma 2 tem o mesmo valor do desvio de 57. Aparentemente, os alunos A e B parecem ter o mesmo desempenho. No entanto, a média e o desvio-padrão usados para calcular esses dois valores do desvio diferem, porque são de duas classes diferentes. A menos que a média e o desvio-padrão das duas classes sejam iguais, você não pode comparar os valores do desvio dos dois alunos.

Aqui está outro exemplo. Suponha que o aluno A faça uma prova de pré-vestibular em um curso de preparação em abril e obtenha um valor do desvio de 54. Depois de dar duro em um curso adicional de estudos, o aluno A faz um teste de admissão numa escola preparatória diferente em setembro. O valor do desvio é 62. Pode parecer que o desempenho do aluno melhorou. No entanto, a prova e os alunos inscritos em abril são diferentes do teste e dos alunos inscritos em setembro. Assim, você não pode comparar os valores do desvio desses dois testes, porque os dados utilizados para calcular a média e o desvio-padrão das provas de abril e setembro são diferentes. Em situações de testes, você pode comparar apenas valores do desvio para um grupo de alunos que recebe a mesma prova. Leve isso em consideração ao interpretar os valores do desvio.

TABELA 4-1: RESULTADOS DOS TESTES DE HISTÓRIA E BIOLOGIA (TODA A TURMA DE LUY)

Meninas	História	Biologia	Meninos	História	Biologia
Luy	73	59	a	54	2
Yumi	61	73	b	93	7
A	14	47	c	91	98
B	41	38	d	37	85
C	49	63	e	44	100
D	87	56	f	16	29
E	69	15	g	12	57
F	65	53	h	44	37
G	36	80	i	4	95
H	7	50	j	17	39
I	53	41	k	66	70
J	100	62	l	53	14
K	57	44	m	14	97
L	45	26	n	73	39
M	56	91	o	6	75
N	34	35	p	22	80
O	37	53	q	69	77
P	70	68	r	95	14
			s	16	24
			t	37	91
			u	14	36
			v	88	76
Média da turma toda				48,0	54,9
Desvio-padrão da turma toda				27,5	26,9

EXERCÍCIO

Confira a seguir os resultados da corrida de 100 m das meninas da escola.

Atleta	Corrida de 100 m (segundos)
A	16,3
B	22,4
C	18,5
D	18,7
E	20,1
Média	19,2
Desvio-padrão	2,01

1. Demonstre que a média dos valores-padrão da corrida de 100 m é 0.
2. Demonstre que o desvio-padrão do valor-padrão da corrida de 100 m é 1.

RESPOSTA

1. Média do valor-padrão da corrida de 100 m

$$= \frac{\left(\frac{16,3 - 19,2}{2,01}\right) + \left(\frac{22,4 - 19,2}{2,01}\right) + \left(\frac{18,5 - 19,2}{2,01}\right) + \left(\frac{18,7 - 19,2}{2,01}\right) + \left(\frac{20,1 - 19,2}{2,01}\right)}{5}$$

$$= \frac{\left\{\frac{(16,3 - 19,2) + (22,4 - 19,2) + (18,5 - 19,2) + (18,7 - 19,2) + (20,1 - 19,2)}{2,01}\right\}}{5}$$

O numerador foi simplificado.

$$= \frac{\left\{\frac{16,3 + 22,4 + 18,5 + 18,7 + 20,1 - 19,2 - 19,2 - 19,2 - 19,2 - 19,2}{2,01}\right\}}{5}$$

O numerador foi reorganizado para que cada valor e (-19,2) fossem separados.

$$= \frac{\left\{\frac{96 - 19,2 \times 5}{2,01}\right\}}{5}$$

$$= \frac{\left\{\frac{96 - 96}{2,01}\right\}}{5}$$

$$= \frac{0}{5}$$

$$= 0$$

2. Desvio-padrão do valor-padrão da corrida de 100 m

$$= \sqrt{\frac{\left(\frac{16,3 - 19,2}{2,01} - 0\right)^2 + \left(\frac{22,4 - 19,2}{2,01} - 0\right)^2 + \left(\frac{18,5 - 19,2}{2,01} - 0\right)^2 + \left(\frac{18,7 - 19,2}{2,01} - 0\right)^2 + \left(\frac{20,1 - 19,2}{2,01} - 0\right)^2}{5}}$$

$$= \sqrt{\frac{\left(\frac{16,3 - 19,2}{2,01}\right)^2 + \left(\frac{22,4 - 19,2}{2,01}\right)^2 + \left(\frac{18,5 - 19,2}{2,01}\right)^2 + \left(\frac{18,7 - 19,2}{2,01}\right)^2 + \left(\frac{20,1 - 19,2}{2,01}\right)^2}{5}}$$

$$= \sqrt{\frac{\left\{\frac{(16,3 - 19,2)^2 + (22,4 - 19,2)^2 + (18,5 - 19,2)^2 + (18,7 - 19,2)^2 + (20,1 - 19,2)^2}{2,01^2}\right\}}{5}}$$

O numerador foisimplificado.

$$= \sqrt{\frac{1}{2,01^2} \times \frac{(16,3 - 19,2)^2 + (22,4 - 19,2)^2 + (18,5 - 19,2)^2 + (18,7 - 19,2)^2 + (20,1 - 19,2)^2}{5}}$$

O numerador foisimplificado.

$$= \frac{1}{2,01} \times \sqrt{\frac{(16,3 - 19,2)^2 + (22,4 - 19,2)^2 + (18,5 - 19,2)^2 + (18,7 - 19,2)^2 + (20,1 - 19,2)^2}{5}}$$

$$= \frac{1}{\text{desvio-padrão da corrida de 100 m}} \times \text{desvio-padrão da corrida de 100 m}$$

Veja atentamente a tabela da página 78.

$$= 1$$

RESUMO

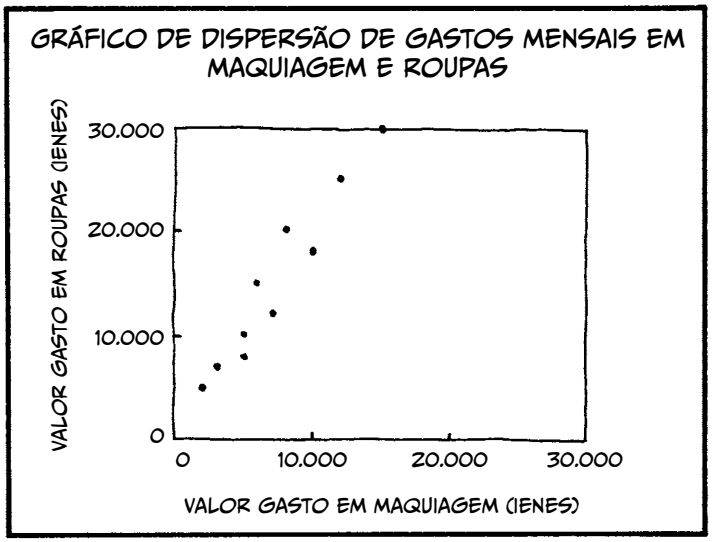
- A normalização ajuda a examinar o valor de um ponto de dados em relação ao resto dos dados usando sua distância entre a média e "o grau de dispersão" dos dados.
- Use a normalização para:
  - Comparar variáveis com diferentes intervalos
  - Comparar variáveis que usam unidades diferentes de medidas
- Um ponto de dados que foi padronizado denomina-se valor-padrão para esse contexto. O valor do desvio é uma aplicação do valor-padrão.

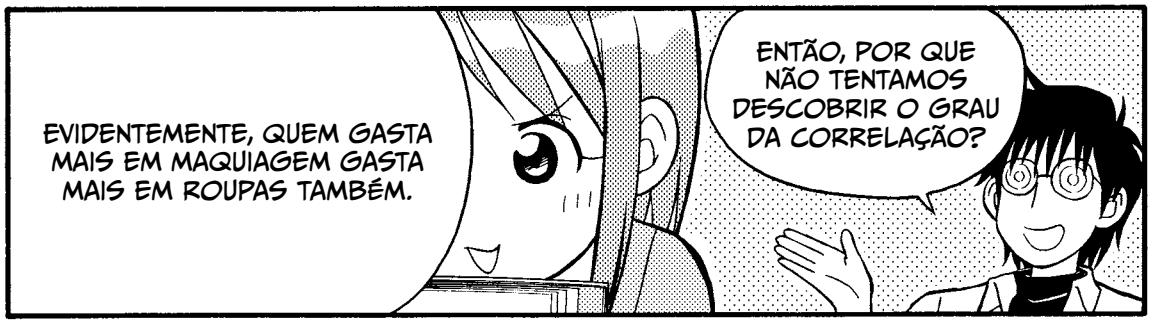
1. COEFICIENTE DE CORRELAÇÃO

Pesquisa de rua

Dez mulheres na faixa dos 20 anos respondem
Gastos mensais com maquiagem e roupas

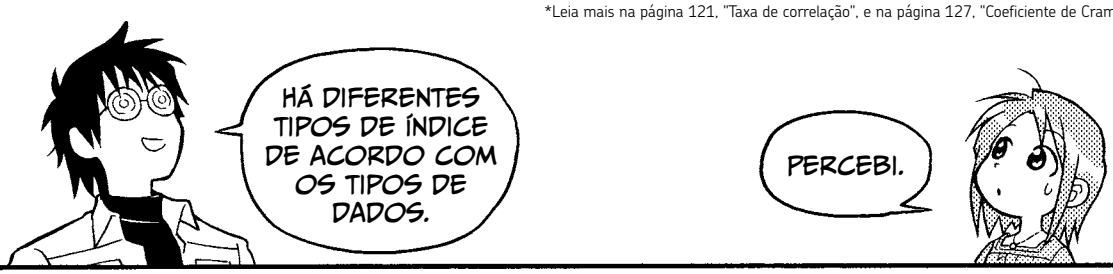
Entrevistada	Valor gasto com maquiagem (ienes)	Valor gasto em roupas (ienes)
Sra. A	3,000	7,000
Sra. B	5,000	8,000
Sra. C	12,000	25,000
Sra. D	2,000	5,000
Sra. E	7,000	12,000
Sra. F	15,000	30,000
Sra. G	5,000	10,000
Sra. H	6,000	15,000
Sra. I	8,000	20,000
Sra. J	10,000	18,000

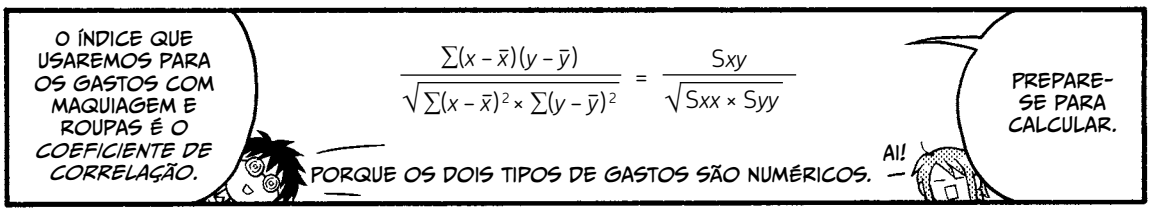


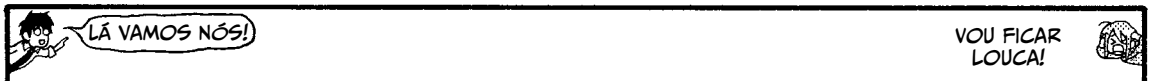


Tipos de dados	Índice	Intervalo de valor	Fórmula
Numéricos e numéricos	Coeficiente de correlação	–1 – 1	$\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \times \sum(y-\bar{y})^2}} = \frac{Sxy}{\sqrt{Sxx \times Syy}}$
Numéricos e categóricos	Taxa de correlação*	0 – 1	$\frac{\text{variação interclasse}}{\text{variância intraclasse + variância interclasse}}$
Categóricos e categóricos	Coeficiente de Cramer*	0 – 1	$\sqrt{\frac{\chi_0^2}{\text{número total de valores} \times (\text{mín. } \{\text{nº de linhas na tabulação cruzada, nº de colunas na tabulação cruzada}\} - 1)}}$

*Leia mais na página 121, "Taxa de correlação", e na página 127, "Coeficiente de Cramer".



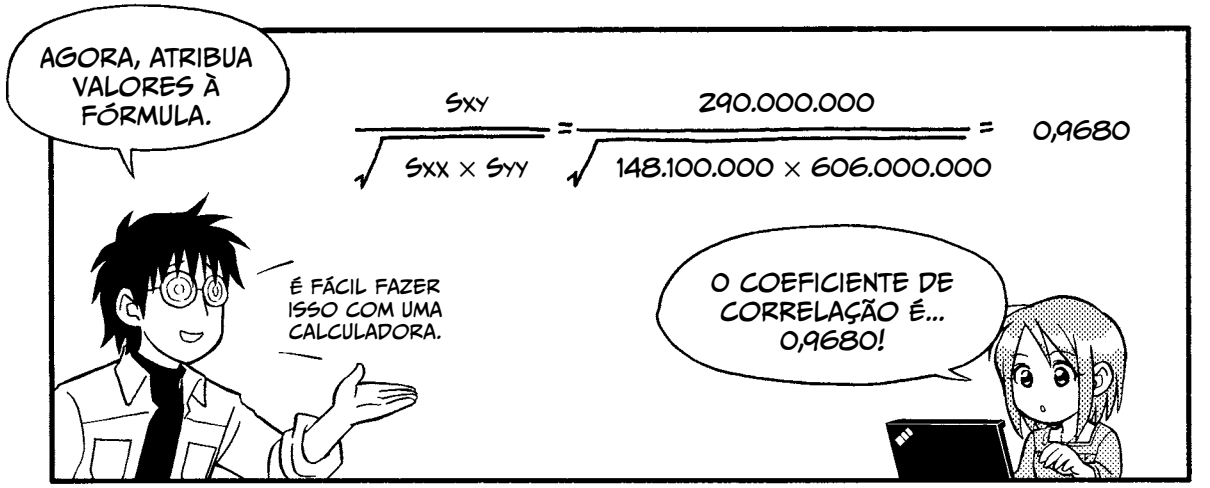




O PROCESSO PARA CALCULAR O COEFICIENTE DE CORRELAÇÃO PARA GASTOS MENSAIS EM MAQUIAGEM E ROUPAS

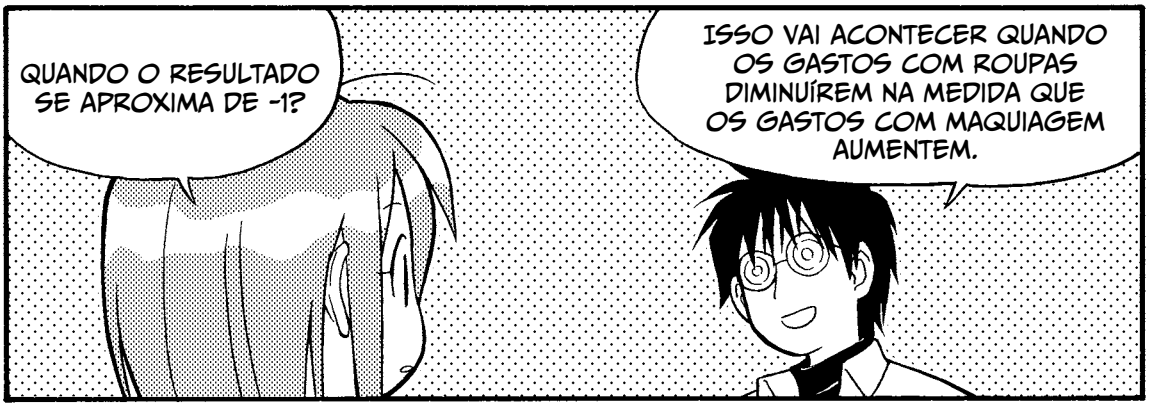
	Valor gasto em maquiagem (ienes)	Valor gasto em roupas (ienes)					
	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
A	3.000	7.000	-4.300	-8.000	18.490.000	64.000.000	34.400.000
B	5.000	8.000	-2.300	-7.000	5.290.000	49.000.000	16.100.000
C	12.000	25.000	4.700	10.000	22.090.000	100.000.000	47.000.000
D	2.000	5.000	-5.300	-10.000	28.090.000	100.000.000	53.000.000
E	7.000	12.000	-300	-3.000	90.000	9.000.000	900.000
F	15.000	30.000	7.700	15.000	59.290.000	225.000.000	115.500.000
G	5.000	10.000	-2.300	-5.000	5.290.000	25.000.000	11.500.000
H	6.000	15.000	-1.300	0	1.690.000	0	0
I	8.000	20.000	700	5.000	490.000	25.000.000	3.500.000
J	10.000	18.000	2.700	3.000	7.290.000	9.000.000	8.100.000
Soma	73.000	150.000	0	0	148.100.000	606.000.000	290.000.000
Média	7.300	15.000					

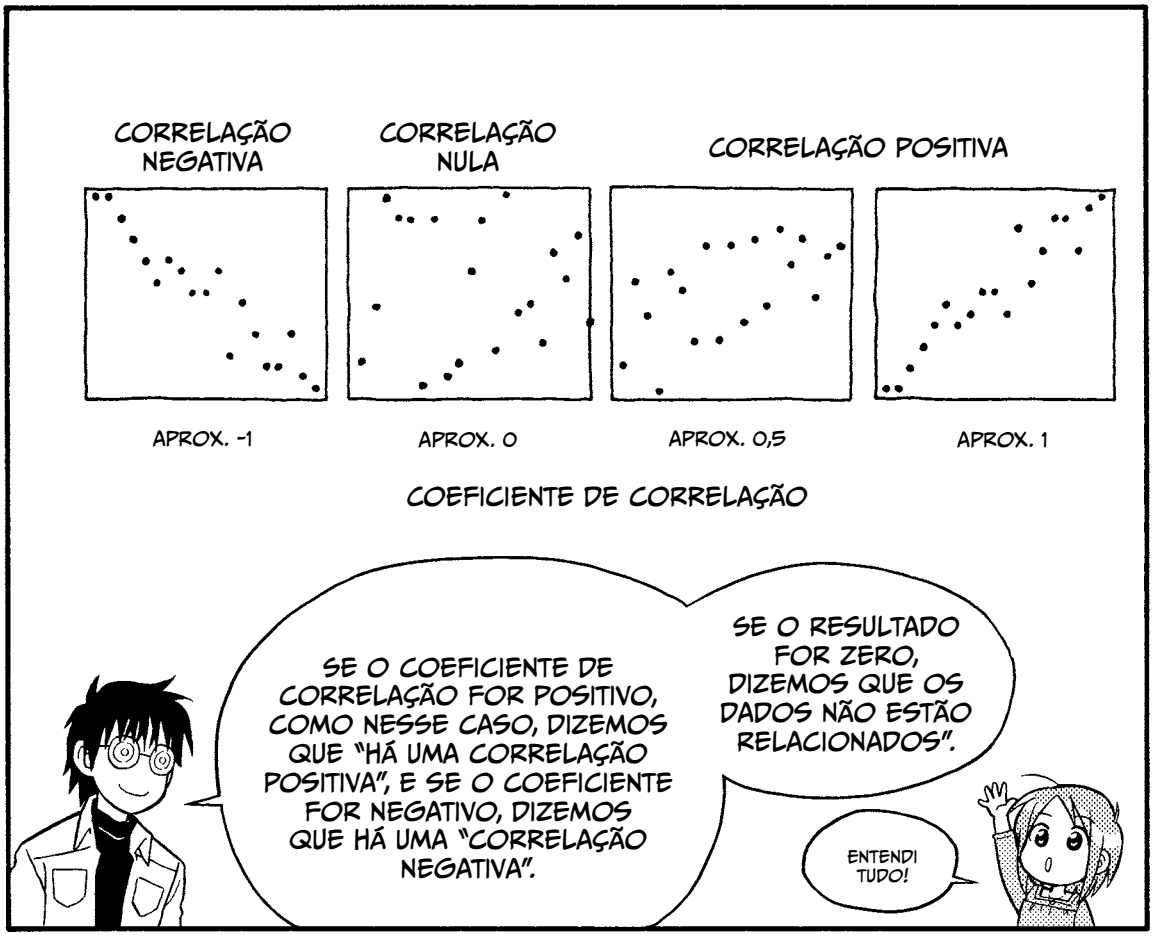
$\bar{x}$ $\bar{y}$ Sxx Syy Sxy


AGORA, ATRIBUA VALORES À FÓRMULA.
$\frac{Sxy}{\sqrt{Sxx \times Syy}} = \frac{290.000.000}{\sqrt{148.100.000 \times 606.000.000}} = 0{,}9680$
É FÁCIL FAZER ISSO COM UMA CALCULADORA.
O COEFICIENTE DE CORRELAÇÃO É... 0,9680!

O COEFICIENTE DE CORRELAÇÃO SE APROXIMA DE ±1 SE A RELAÇÃO LINEAR ENTRE AS DUAS VARIÁVEIS FOR MAIS FORTE.
QUANDO A RELAÇÃO SE TORNA MAIS FRACA, ELE SE APROXIMA DE 0.
-1 ← 0 → +1
INTERESSANTE...

O RESULTADO QUE CALCULEI É BEM PRÓXIMO DE 1, O QUE SIGNIFICA QUE A RELAÇÃO ENTRE OS GASTOS COM MAQUIAGEM E COM ROUPAS É BEM GRANDE!
P-girls
VOCÊ ESTÁ CERTA.

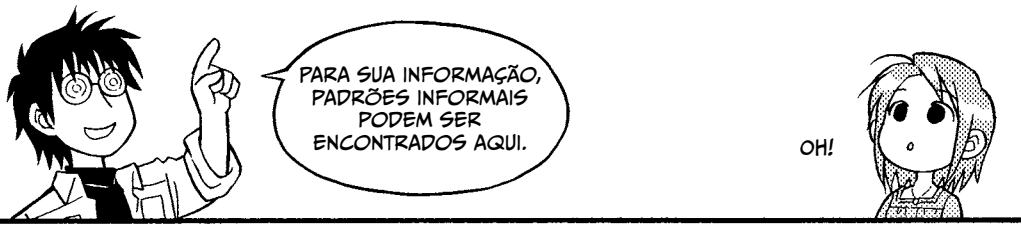

QUANDO O RESULTADO SE APROXIMA DE -1?
ISSO VAI ACONTECER QUANDO OS GASTOS COM ROUPAS DIMINUÍREM NA MEDIDA QUE OS GASTOS COM MAQUIAGEM AUMENTEM.


CORRELAÇÃO NEGATIVA
CORRELAÇÃO NULA
CORRELAÇÃO POSITIVA
APROX. -1
APROX. 0
APROX. 0,5
APROX. 1
COEFICIENTE DE CORRELAÇÃO
SE O COEFICIENTE DE CORRELAÇÃO FOR POSITIVO, COMO NESSE CASO, DIZEMOS QUE "HÁ UMA CORRELAÇÃO POSITIVA", E SE O COEFICIENTE FOR NEGATIVO, DIZEMOS QUE HÁ UMA "CORRELAÇÃO NEGATIVA".
SE O RESULTADO FOR ZERO, DIZEMOS QUE OS DADOS NÃO ESTÃO RELACIONADOS".
ENTENDI TUDO!

AGORA, SOBRE O COEFICIENTE DE CORRELAÇÃO...
INFELIZMENTE, NÃO HÁ PADRÕES ESTATÍSTICOS QUE GARANTAM QUE AS DUAS VARIÁVEIS APRESENTAM UMA RELAÇÃO FORTE.
QUE ÍNDICE INÚTIL...

PADRÕES INFORMAIS DO COEFICIENTE DE CORRELAÇÃO

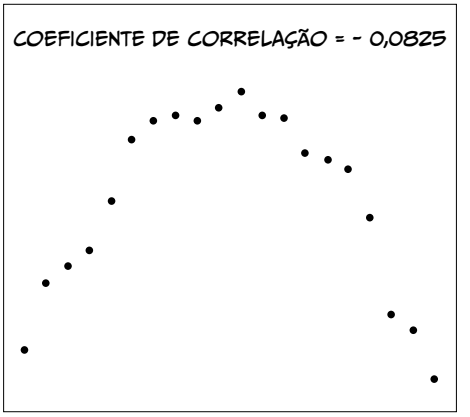
Valor absoluto do coeficiente de correlação		Descrição detalhada	Descrição resumida
1,0–0,9	⇨	Relação muito forte	
0,9–0,7	⇨	Relação razoavelmente forte	Relacionados
0,7–0,5	⇨	Relação um pouco fraca	
Abaixo de 0,5	⇨	Relação muito fraca	Não relacionados



ATENÇÃO

Mencionei anteriormente que o coeficiente de correlação é um índice que mostra o grau de relação linear entre duas variáveis numéricas.

EXEMPLO DE DADOS INADEQUADOS PARA O COEFICIENTE DE CORRELAÇÃO



Por exemplo, as duas variáveis estão evidentemente relacionadas neste gráfico. No entanto, o coeficiente de correlação é quase 0 porque a relação é não-linear.

2. TAXA DE CORRELAÇÃO

Pesquisa de público em Everyhills

Idade e grife favorita

Entrevistada	Idade	Marca
A	27	Theremes
B	33	Channelior
C	16	Bureperry
D	29	Bureperry
E	32	Channelior
F	23	Theremes
G	25	Channelior
H	28	Theremes
I	22	Bureperry
J	18	Bureperry
K	26	Channelior
L	26	Theremes
M	15	Bureperry
N	29	Channelior
O	26	Bureperry

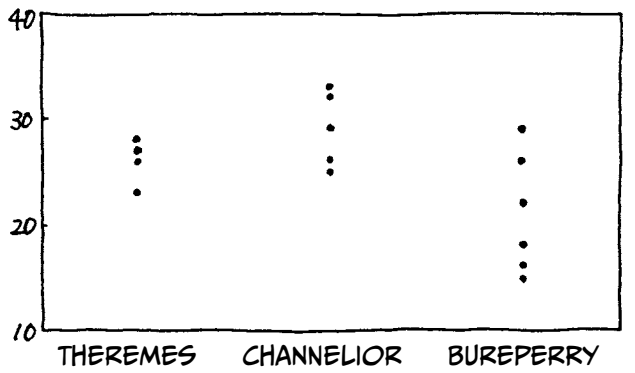
VOU REORGANIZAR A TABELA.

HUM...

MARCA DE ROUPAS PREFERIDA E IDADE

THEREMES	CHANNELIOR	BUREPERRY	
23	25	15	
26	26	16	
27	29	18	
28	32	22	
	33	26	
		29	
SOMA	104	145	126	375
MÉDIA	26	29	21	25

GRÁFICO DE DISPERSÃO DE MARCA FAVORITA E IDADE



AGORA, VAMOS FAZER UM GRÁFICO.

UAU! TENHO A IMPRESSÃO DE QUE EXISTE ALGUMA CORRELAÇÃO!

NESSE PONTO, VAMOS CALCULAR O VALOR DA TAXA DE CORRELAÇÃO.

ISSO, VAMOS ADIANTE!

O valor da taxa de correlação pode ser calculado pelas etapas 1 a 4, a seguir.

Etapa 1

Faça os cálculos abaixo.

		Soma
(Theremes – média para Theremes)^2	$(23 - 26)^2 = (-3)^2 = 9$ $(26 - 26)^2 = 0^2 = 0$ $(27 - 26)^2 = 1^2 = 1$ $(28 - 26)^2 = 2^2 = 4$	14 S_{TT}
(Channelior – média para Channelior)^2	$(25 - 29)^2 = (-4)^2 = 16$ $(26 - 29)^2 = (-3)^2 = 9$ $(29 - 29)^2 = 0^2 = 0$ $(32 - 29)^2 = 3^2 = 9$ $(33 - 29)^2 = 4^2 = 16$	50 S_{CC}
(Bureperry – média para Bureperry)^2	$(15 - 21)^2 = (-6)^2 = 36$ $(16 - 21)^2 = (-5)^2 = 25$ $(18 - 21)^2 = (-3)^2 = 9$ $(22 - 21)^2 = 1^2 = 1$ $(26 - 21)^2 = 5^2 = 25$ $(29 - 21)^2 = 8^2 = 64$	160 S_{BB}

Etapa 2

Calcule a variação intraclasse (STT + SCC + SBB = o quanto os dados dentro de cada categoria variam).

$$S_{TT} + S_{CC} + S_{BB} = 14 + 50 + 160 = 224$$

Etapa 3

Calcule a variação interclasse, ou o quanto as categorias diferem umas das outras.

(número de votos para Theremes) × (média da Theremes − média para todos os dados)^2
+ (número de votos para Channelior) × (média da Channelior − média para todos os dados)^2
+ (número de votos para Bureperry) × (média da Bureperry − média para todos os dados)^2

$4 \times (26 - 25)^2 + 5 \times (29 - 25)^2 + 6 \times (21 - 25)^2$

$= 4 \times 1 + 5 \times 16 + 6 \times 16$

$= 4 + 80 + 96$

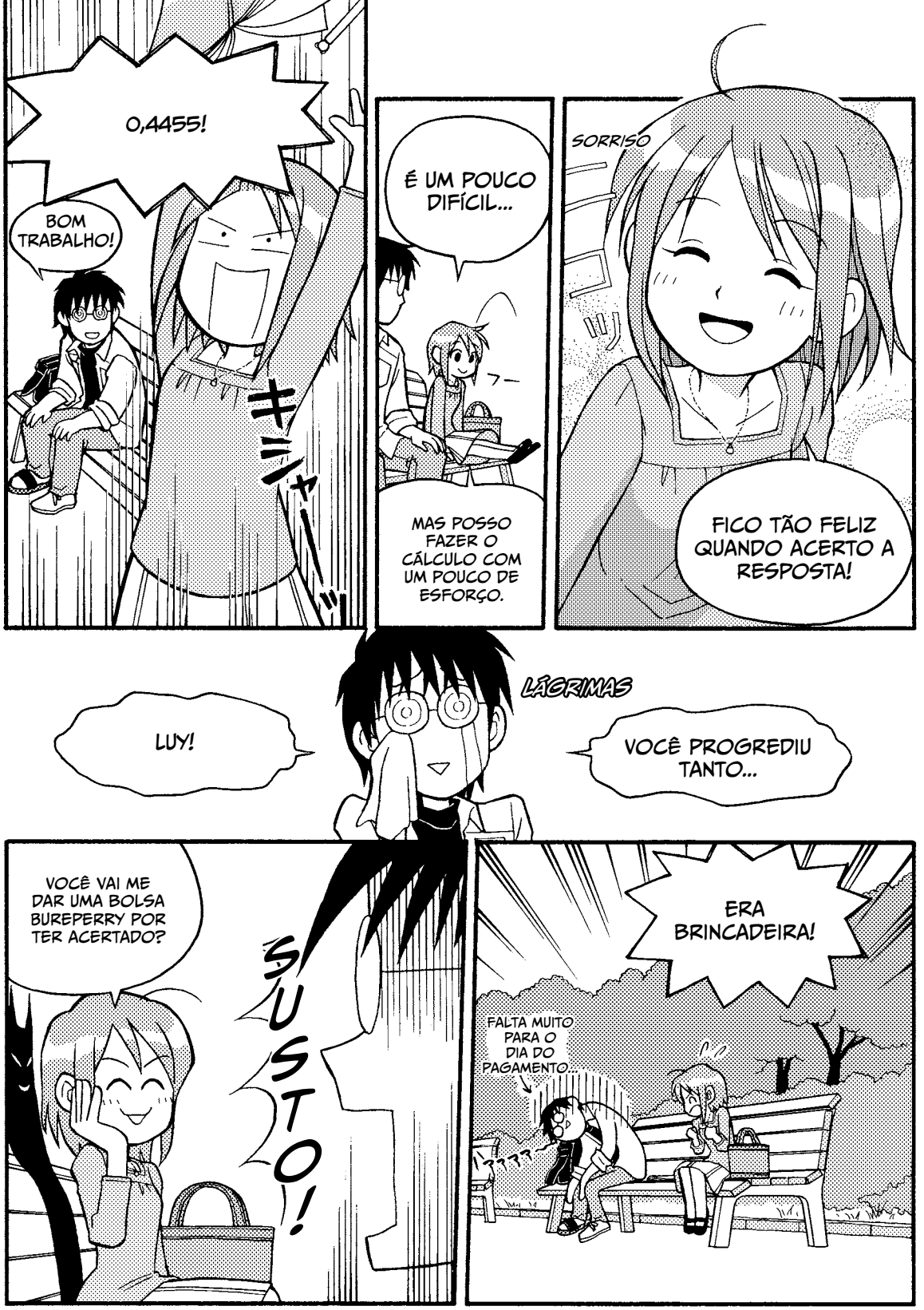
$= 180$

Etapa 4

Calcule o valor da taxa de correlação.

$$\frac{\text{variação interclasse}}{\text{variação intraclasse + variação interclasse}}$$

$$\frac{180}{224 + 180} = \frac{180}{404} = 0{,}4455$$


0,4455!
BOM TRABALHO!
キニャーーッ
É UM POUCO DIFÍCIL...
MAS POSSO FAZER O CÁLCULO COM UM POUCO DE ESFORÇO.
SORRISO
FICO TÃO FELIZ QUANDO ACERTO A RESPOSTA!
LÁGRIMAS
LUY!
VOCÊ PROGREDIU TANTO...
VOCÊ VAI ME DAR UMA BOLSA BUREPERRY POR TER ACERTADO?
SUSTO!
ERA BRINCADEIRA!
FALTA MUITO PARA O DIA DO PAGAMENTO...
アアアア