

Estatística para Ciência de Dados  
Resolução do trabalho 06

Eduardo Façanha Dutra  
Giovanni Brígido

# Contents

<b>1</b>	<b>Enunciado</b>	<b>3</b>
<b>2</b>	<b>Resolução</b>	<b>3</b>
2.1	Tarefa 1 . . . . .	3

# 1 Enunciado

Você é um analista de dados no setor de compras da Casa Pio, uma grande rede de lojas de calçados.

Ela é uma empresa com sede em Fortaleza estabelecido há 70 anos. Atualmente, também opera em várias cidades do interior do Ceará. A empresa vende sapatos de preços médios, variando de 120 a 200 reais. Embora os sapatos sejam de alta qualidade, existem lotes e lotes no estoque que nunca são vendidos. Em outras palavras, os sapatos estragando nas prateleiras das lojas.

Por isto, o gerenciamento de estoque é um problema muito comum e importante.

Muitas, senão a maioria das lojas, não conseguem determinar número correto de itens que precisam manter em estoque. O problema oposto também surge: lojas sem estoque necessário de mercadorias não irão atender à demanda de mercado. Por exemplo, você certamente já entrou em uma loja de sapatos, mas não conseguiu comprar porque eles não tinham o número em estoque.

Neste estudo, vamos examinar o problema oposto - ter muito estoque.

Este é um problema mais significativo para a empresa, pois indica que investiu na produção ou compra de um produto, mas não foi capaz de vendê-lo.

Temos um banco de dados com as informações de vendas das lojas durante os anos 2014 a 2016, com várias informações úteis.

Nosso problema é estimar o volume de compra de sapatos para não ficar com estoque muito alto, nem perder muitas vendas.

**Tarefa 1** Responda as seguintes perguntas:

- 1- Os dados fornecidos são amostrais ou populacionais?
- 2- Você separaria os dados em quantos grupos?
- 3- Esses grupos são dependentes ou independentes?
- 4- Como você agrupará os dados para ter uma melhor visão do problema a ser resolvido?
- 5- Os dados fornecidos representam uma distribuição Normal?
- 6- Qual o intervalo de confiança que será usado?
- 7- Qual a estatística será usada? Z ou t?
- 8- Com base na sua resposta a pergunta 7, qual a sua justificativa?
- 9- Quantos pares de cada sapato devemos ter em estoque?

## Tarefa 2

Digamos que queremos usar um intervalo de confiança para ver se duas lojas estão vendendo o mesmo número de sapatos. Além disso, queremos saber com 90% confiança quanto uma loja supera a outra em termos de vendas. Você tem duas tabelas representando as vendas de calçados femininos em duas lojas. Seus códigos são ARA 1 e ARA 2. Temos dados para 2016.

Justifique suas escolhas e aplique uma estatística para determinar qual das lojas vende mais produtos, e quais.

Verifique com 90% de confiança, se ambas lojas podem ter um estoque único.

# 2 Resolução

## 2.1 Tarefa 1

- 1- Os dados fornecidos são amostrais ou populacionais?

Os dados fornecidos são amostrais, uma vez que se referem a apenas 3 anos, de 2014 a 2016 e Casa Pio já está estabelecida em Fortaleza há 70 anos.

**2- Você separaria os dados em quantos grupos?**

Separaria os dados por gênero, por tamanho de sapato, por mês e ano e por cidade.

**3- Esses grupos são dependentes ou independentes?**

Os grupos são independentes.

**4- Como você agrupará os dados para ter uma melhor visão do problema a ser resolvido?**

Os dados serão agrupados por meses de 1 a 12, por ano, e por tamanho de sapato, em ordem crescente, sendo criadas duas tabelas, uma para cada gênero. Dessa maneira, obtém-se uma visão melhor do problema a ser resolvido.

**5- Os dados fornecidos representam uma distribuição Normal?**

De acordo com o Teorema do Limite Central, as médias de amostras grandes e aleatórias são aproximadamente normais. Logo, os dados fornecidos representam uma distribuição normal.

**6- Qual o intervalo de confiança que será usado?**

O intervalo de confiança será de 95%

**7- Qual a estatística será usada? Z ou t?**

A estatística que será usada será a t.

**8- Com base na sua resposta a pergunta 7, qual a sua justificativa?**

Como são três anos de dados e os dados estão divididos em meses, totalizando 36 meses de dados, optou-se pela distribuição t, uma vez que a amostra tem tamanho próximo de 30. Ademais, como a variância populacional é desconhecida, sugere-se o uso da estatística t.

**9- Quantos pares de cada sapato devemos ter em estoque?**

Aplicando uma estratégia agressiva de mercado, adotou-se o arredondamento para cima do limite superior do intervalo de confiança para definir quantos pares de sapato deve-se ter em estoque.

Digamos que queremos usar um intervalo de confiança para ver se duas lojas estão vendendo o mesmo número de sapatos. Além disso, queremos saber com 90% de confiança quanto uma loja supera a outra em termos de vendas. Você tem duas tabelas representando as vendas de calçados femininos em duas lojas. Seus códigos são ARA 1 e ARA 2. Temos dados para 2016.

Justifique suas escolhas e aplique uma estatística para determinar qual das lojas vende mais produtos, e quais.

Não se pode concluir que uma loja vende mais que a outra para qualquer tamanho de sapato, uma vez que, para todos os tamanhos de sapatos, no intervalo de confiança, o limite inferior é um valor negativo e o limite superior é um valor positivo.

- Temos variância populacional desconhecida;
- Temos uma amostra pequena;
- Então devemos usar a estatística: T

$$s_p^2 = \frac{(n_x - 1) s_x^2 + (n_y - 1) s_y^2}{n_x + n_y - 2}$$

$$(\bar{x} - \bar{y}) \pm t_{n_x + n_y - 2, \alpha/2} \sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}$$

Table 1: Vendas de sapatos masculinos de 2014 a 2016 em todo o Ceará

Tamanho	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez	Total
6	8	6	3	2	11	8	4	14	10	26	9	4	30
6.5	10	7	5	9	1	3	8	19	7	9	6	6	40
7	4	3	10	4	10	16	12	8	16	12	9	10	49
7.5	13	11	8	3	23	11	20	9	23	23	14	13	13
8	30	30	35	25	30	24	45	19	35	31	22	28	87
8.5	50	49	54	56	49	63	62	82	57	72	43	41	455
9	97	71	88	111	111	148	141	121	151	131	69	108	1230
9.5	133	127	134	152	147	182	154	183	154	183	155	186	8613
10	96	101	96	104	89	134	130	135	117	128	112	105	1005
10.5	86	94	100	117	117	113	130	114	134	115	99	89	350
11	32	49	55	49	51	46	56	44	56	43	33	38	245
11.5	30	26	30	20	27	51	36	34	26	54	35	30	30
12	15	14	9	16	24	29	17	35	30	24	18	9	84
13	4	2	6	4	9	13	7	4	6	10	14	8	30
14	8	13	14	8	14	10	15	12	4	12	9	7	15
15	3	10	5	10	8	11	14	11	11	4	14	10	52
Total	619	613	652	690	721	862	851	844	837	877	661	692	48507

Table 2: Vendas de sapatos femininos de 2014 a 2016 em todo o Ceará

Tamanho	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez	Total
4.5	1	8	3	0	2	17	5	3	2	5	1	4	10
5	5	0	1	0	3	2	6	7	10	1	0	1	15
5.5	5	3	0	6	1	3	6	4	5	5	8	17	143
6	4	18	7	22	12	3	0	1	5	4	1	4	4
6.5	23	17	20	21	23	30	34	16	20	13	17	18	123
7	22	23	26	41	34	41	28	34	47	47	37	43	715
7.5	74	61	59	60	53	80	63	59	84	84	57	46	1740
8	105	113	92	103	106	94	117	129	103	138	106	96	1005
8.5	57	55	70	66	120	93	116	96	88	92	99	86	2145
9	73	60	67	86	92	88	90	91	98	70	70	69	355
9.5	38	31	42	50	37	22	40	39	34	42	47	34	180
10	15	20	19	22	13	18	22	23	19	26	15	19	85
10.5	12	20	14	19	13	17	16	27	12	23	16	9	42
11	8	3	2	5	4	2	3	5	0	2	2	9	17
11.5	2	12	9	6	7	4	5	5	7	1	2	0	3
12	1	4	0	0	8	5	11	10	6	17	15	1	1
Total	445	448	431	507	528	519	562	549	540	570	493	456	5406