

Research Report on Cotti Coffee

Christie, Elijah, Eric, Kyle and Warren*

December 13, 2024

Contents

1	Introduction	1
2	One-variable analysis	1
3	Two variable analysis	2
4	Contributions	3

*listed in alphabetical order.

1 Introduction

Coffee is one of the world's most cherished beverages, known for its rich flavor and stimulating effects, and it is becoming more and more popular in China. On this occasion, our group did research on Cotti coffee. With the core question, "How do changes in commodity prices affect the quantity of products?" We tried to explore and analyze the issue from an mathematical point of view, using our knowledge of statistics.

2 One-variable analysis

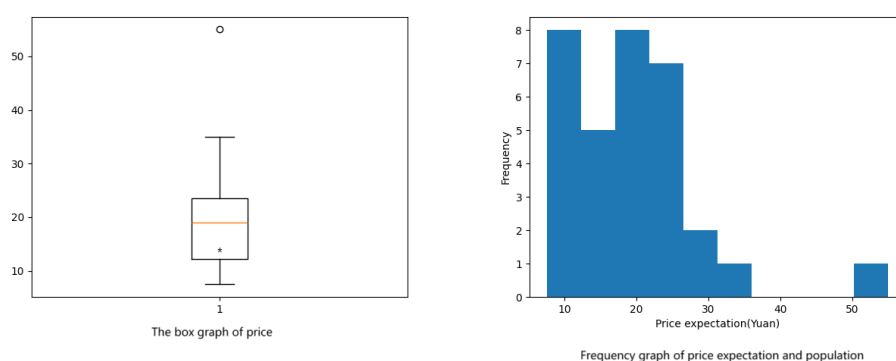


Figure 1: Price expectation

In our survey, we first studied the price expectation of customers. Out of our 274 questionnaires, only 32 people volunteered to provide a reasonable price expectation of a 500ml cup of coffee. We tried to conduct one variable analysis on the collected data. The processing code is provided in appendix. As is shown in Figure 1, there's one outlier among all data, which is a reasonable amount. If the outlier is excluded, we can find data in Table 1.

Table 1: Mean, std. deviation and five number summary

Name	Sample size	Mean*	Mean	σ^{\wedge}	σ^2 (variance)	Range
Value	32	19.28593	18.13387	6.93969	48.1596	27.5
Name	Min.	Max.	IQR	Q1	Q2(median)	Q3
Value	7.5	35	18	13	23	31

* Not excluding the outlier.

\wedge Standard deviation.

In this chart, the mean (μ), variation (σ^2), standard deviation (σ), interquartile range (IQR), range are given by the formulae:

$$\mu = \frac{\sum_{i=1}^n a_i}{n}$$

$$\sigma^2 = \frac{\sum_{i=1}^n (a_i - \mu)^2}{n}$$

$$\sigma = \sqrt{\sigma^2}$$

$$\text{IQR} = Q_3 - Q_1$$

$$\text{Range} = \max\{a_i\} - \min\{a_i\}$$

where a_i refers to the i -th number of the data, n is the size of the sample space, and Q_1 and Q_3 are the quartiles respectively.

The average price of a cup of Cotti coffee is 14 yuan (denoted by a small asterisk mark in Figure 1), located in the lower 25%-50% quartile. This means that Cotti coffee is a relatively affordable brand.

3 Two variable analysis

As is shown in Figure 2, we plotted the price against the quantity by economics convention. This graph shows a significant negative correlation between P , the price, and Q , the quantity sold. This can be confirmed by a correlation coefficient of -0.93828826 , which is close enough to -1 to show a strong negative correlation. This coefficient is reasonable because customers are less likely to purchase a product when the price go up. However, there is still a point that does not follow this rule. It might because people do not think carefully when it comes to “50% off” and “-10 yuan”, where latter is a greater discount but the former sounds more attractive. This reminds managers of the importance of a

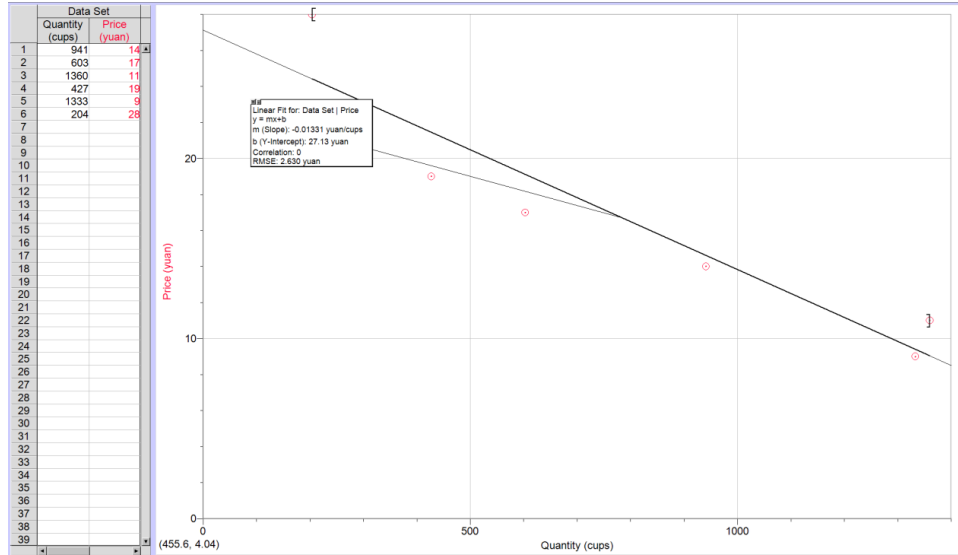


Figure 2: Demand curve according to the data we collected

well-organised advertising strategy.

Also, we found a linear fit of $y = -0.1331x + 27.13$, where x is the quantity demand and y is the price. This equation fits well when y does not deviate much, and the gradient means if the price goes up by 0.1331 yuan, Cotti is likely to lose one customer. However, the y-intercept of 27.13 yuan, which indicates if the price is higher than or equals to 27.13 yuan, no one will be willing to purchase this product, is not reasonable. This means the linear fit model does not apply anymore when y is significantly bigger than the current value.

4 Contributions

Our work distribution is listed here. Members are listed in alphabetical order.

- Christie Chen Haonan: PPT making, evaluation of age distribution, 1/8 of math part.
- Elijah Dong Yicheng (group leader): PPT design and spread of questionnaires. Time management.
- Eric Zhou Changhui: Formatting, coding(calculation of PED), 3/4 of the math part.
- Kyle Dong Zeguang: Summary, spread of questionnaires, PPT design, photo taking, economics evaluation of portion of income.
- Warren Wang Yuanhao: Design of question paper, spread of question paper, introduction, 1/8 of math part.

Appendix

Listing 1: One variable analysis code

```
1 import matplotlib.pyplot as plt
2 import numpy as np
3 data = [55, 17.5, 27.5, 21.05, 25, 25, 10, 13, 30, 20, 9.9, 15, 15, 9.9, 18, 20, 8,
         7.5, 25, 23, 23, 25, 21, 10, 35, 20, 18, 15, 22, 9.9, 9.9, 13]
4 plt.boxplot(data)
5 plt.show()
6 print(f"""mean:{sum(data)/len(data)},
7         mean excluding outlier:{sum(data[1:])/len(data[1:])}
8         min excl.{min(data[1:])}
9         max excl.{max(data[1:])}
10        std excl.{np.std(data[1:])}
11        median {np.median(data[1:])}""")
12 data = sorted(data[1:])
13 print(data[int(len(data)/4)+1],data[int(3*len(data)/4)], len(data))
```

Listing 2: Two variable analysis code

```
1 import numpy as np
2 data = """
3 941 14
4 603 17
5 1360 11
6 427 19
7 1333 9
8 204 28
9 """.split()
10 x, y = [], []
11 for dx in data:
12     if len(x) == len(y): y += [int(dx)]
13     else: x += [int(dx)]
14 xy = [(x[i], y[i]) for i in range(len(x))]
15 xy.sort(key = lambda t: t[0])
16 x,y = ([xyi[z] for xyi in xy] for z in [0,1])
17 ped = []
18 print("x and y are respectively",x,y)
19 print("Their correlation coefficient is", np.corrcoef(x,y=y)[0][1])
20 for i in range(len(x)-1):
21     dx = (x[i+1] - x[i]) / x[i]
22     dy = (y[i+1] - y[i]) / y[i]
23     ped += [dy/dx]
24 print("PED of the data is", ped)
```