Coding 1 Team Project

Team Hong Kong (Maryam, Fatima, Hassan, Oszkar)

11/7/2021

Introduction

A data collection survey was conducted to collect data about products from multiple shops located across two districts of Budapest. The next segment displays the mean, median and standard deviation (among other descriptive statistic measures) of the price of the two products.

Data

Products chosen for the survey were Coca Cola (0.5 L) and Orbit gum. The main variable examined in this report is the price of the products per district.

The data was collected in person by the team.

Table 1: Descriptive statistics prices

District		Mean	Median	SD	Max	Min	N	range_df	P5	P25	P75	P95
9	Orbit	168.00	165.00	38.82	260.00	120.00	10	140.00	124.50	150.00	177.50	228.50
10	Orbit	179.60	174.00	50.04	299.00	129.00	10	170.00	129.45	146.25	198.75	254.45
9	Coke	270.50	260.00	33.04	340.00	220.00	10	120.00	233.50	252.50	286.25	322.00
10	Coke	291.90	284.50	64.13	439.00	219.00	10	220.00	219.00	259.00	316.25	389.95

Data are available from: Please refer to link in the introduction

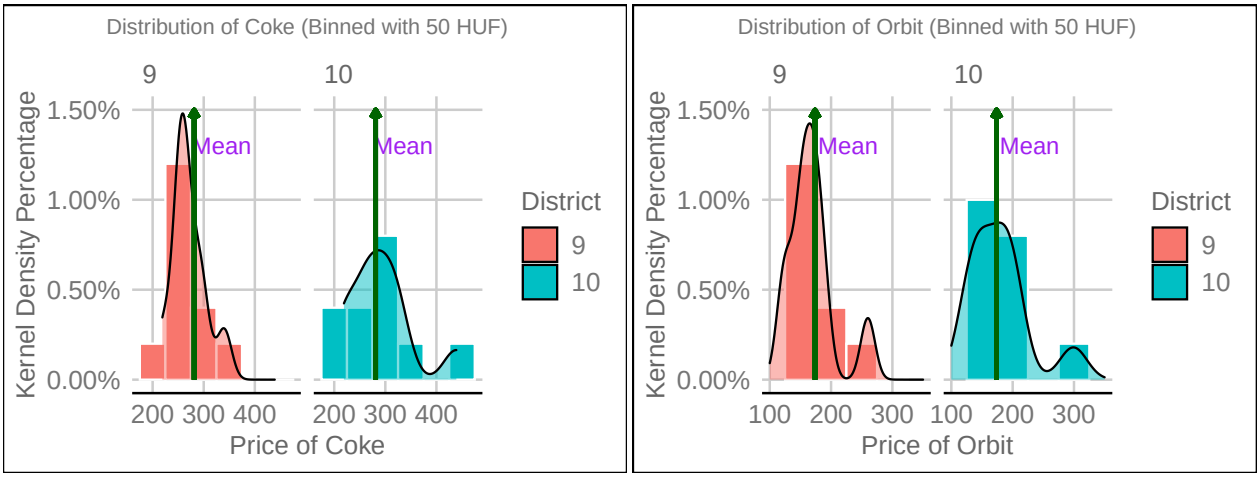
The number of observations is 20 for all of our key variables.

Data cleaning

Part of the data cleaning steps were changing the column names, removing the ID column and converting the district to a factor variable.

Price distribution

The following histograms show the price distribution for each product per district.



We chose 50 HUF binwidth as it is a good aggregation of the individual measurements at this pricerange.

The plots show us the distribution of the measurements and the mean per product and per district.

Conclusion

Based on our measurements we can conclude, that the prices were higher for both products in the 10th district, that means higher mean, standard deviation, minimum and maximum price. Having more measurements could strengthen our analysis.