

Report on Coke and Snickers price differences

Introduction

This is a report on *Coke and Snickers price differences*.

The motivation of our analysis is to summarize and organize main characteristics of our Coke and Snickers dataset. Our descriptive statistics include measures of central tendency, that are means and medians, conditional on city part (inner or outer district). It also includes measures of variability of prices, that are standard deviations, range, quartiles, minimum and maximum values of prices of selected products conditional on districts, in order to describe how the prices are distributed within the dataset. For the price variables, we used tables and density plots for better understanding the meaning of our analyzed data.

Data

The data was collected by using random sampling. We randomly dropped a few pins on Google Maps in two selected districts (where one is in the “inner city” and another is in the “outer city”). We chose a random direction and walked until we bumped into a store to collect prices of Snickers and Coke.

Short description of each variable:

1. *prices* for two products, that are Coke and Snickers (integer);
2. *name* of randomly selected shop (character);
3. unique *ID* for the shop, as there are duplicated shop names on different addresses (integer);
4. *city* part where the shop is located, that are outer or inner cities (character);
5. *district*, where one was selected from inner city and another is from outer city (character);
6. *number of cashiers*, implicitly measuring the store size (integer);
7. *open on Sundays* or not (binary variable);
8. *membership in a store chain*, that brings in economics of scale (binary variable);
9. *opening hours*, that were calculated from closing hours minus opening hours (integer).

During the data preparation, we cleaned and re-calculated some of our variables for further analysis. In this short report we only concentrate on price differences among city parts, so most of the variables aren't used.

Descriptive statistics

We investigated the data on the collected prices and we have the following descriptive statistics on Snickers and Coke prices in inner and outer districts.

Table 1: Descriptive statistics of Coke and Snickers prices

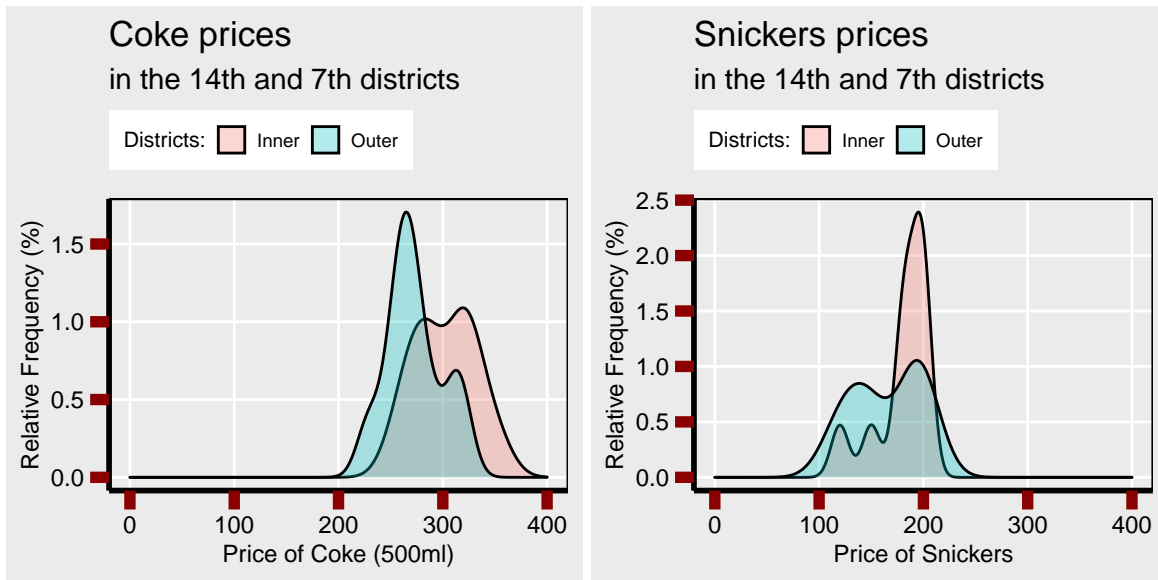
		Mean	Median	SD	Min	Max	Range	P05	P95	N
Coke price	Inner	302.40	304.50	29.15	259	350	91.00	263.95	340.55	10
	Outer	272.80	269.00	27.15	229	319	90.00	238.45	314.95	10
Snickers price	Inner	180.00	187.00	26.10	120	200	80.00	133.50	200.00	10
	Outer	165.30	169.50	33.04	120	199	79.00	120.00	199.00	10

Coke: Price for Coke in both inner and outer cities is distributed in the range of 90 HUF, while the averages, maximum and minimum values are 30 HUF more in outer city than in inner city. The standard deviation for both inner and outer cities is approximately the same, around 27-29 HUF.

Snickers: Price for Snickers in both inner and outer city is distributed in the range of 80 HUF, spreading from 120 to 200 HUF in both. The standard deviation for inner city is around 26 and for outer city it is around 33. The averages for Snickers price in inner city are 180-190 HUF, while for outer city it's around 170 HUF.

The next Figure shows the density plots for prices of Coke and Snickers in inner city and outer city.

We choose kernel density plots to display price differences, because it allows us to compare conditional distributions on the same chart. The X axis was defined in a way that the 2 products can be compared next to each other.



Conclusion

From this exercise we learned how to do perform descriptive analysis in proper order, that is we organized the dataset, by cleaning and filtering it, decided which descriptive statistics to use and which graphs to build in order to better describe main characteristics of our set of data.