Report on Billion Price Project

Name of author

Introduction

This is a report on *The Billion Price Project*.

HERE COMES THE MOTIVATION WHY THIS IS A MEANINGFUL PROJECT AND WHAT IS THE MAIN GOAL!

For more details on the project see: http://www.thebillionpricesproject.com/ or this embedded link.

Data

HERE COMES A DETAILLED EXPLANATION ABOUT WHERE THE DATA COMES FROM AND IF IT IS REPRESENTATIVE OR NOT.

Our main interest is whether online prices are lower or higher than simple retail store prices. We investigated the data on the collected prices and we have the following descriptive statistics on online, in-store prices and in their differences.

Table 1: Descriptive statistics of prices

	Mean	Median	SD	Min	Max	P05	P95
Retail	55.22	14.49	135.49	0.25	970.00	1.99	219.00
Online	54.74	13.99	133.86	0.25	970.00	1.99	225.00
Price difference	−0.48	0.00	16.27	−380.13	450.01	−2.30	1.34

Data are available from: https://osf.io/yhbr5/

The number of observations is 7891 for all of our key variables.

DESCRIPTION OF THE SUMMARY STATS: WHAT CAN WE LEARN FROM THEM?

As the focus is the price difference, the next Figure shows the histogram for this variable.

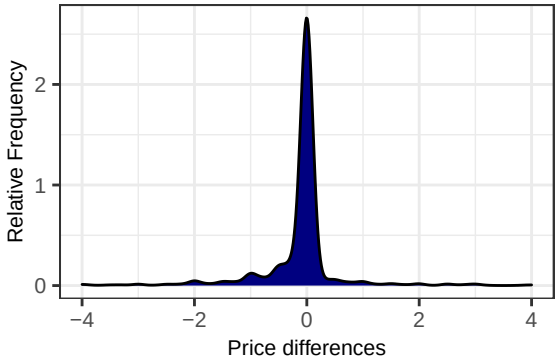

Figure 1: Distribution of price differences

DESCRIPTION OF THE FIGURE. WHAT DOES IT TELS US?

(May change the order of descriptive stats and graph.)

Testing Price Differences

We test the hypothesis, whether the price difference is zero, therefore there is no difference between retail and online prices:

$$H_0 := \text{price online} - \text{price retail} = 0$$
$$H_A := \text{price online} - \text{price retail} \neq 0$$

Running a two-sided t-test, we have the t-statistic as -2.62 and the p-value as 0.01. The 95% confidence intervals are: -0.84 and -0.12. Based on these results with 95% confidence we can reject the hypothesis that the two price would be the same in this particular sample.

Robustness check / 'Heterogeneity analysis'

Task: calculate and report t-tests for each countries.

You should report: Country, mean of p_diff, se of the mean for p_diff, number of observations in each country, t-statistic, p-value.

Hint: use 'kable()' and to hold the table position you can define the following argument: 'position = "H"' . Take care of caption, number of digits you use and the name of variables you report! You may check how the output changes if you use 'booktabs = TRUE' input for kable! In case of html output use something like:

```
kable(...,
      "html", booktabs = F,  position = "H") %>%
      kable_classic(full_width = F, html_font = "Cambria")
```

Table 2: Online and retail price differences by countries and t-tests

Country	Mean	SE	Num.Obs	t-stat	p-val
Brazil	-0.9053	0.7847	122	-1.1537	0.1254
China	-0.5105	0.8411	19	-0.6070	0.2757
Germany	3.6797	1.8658	420	1.9722	0.0246
Japan	-11.9829	2.1467	350	-5.5820	0.0000
South Africa	-2.5297	0.8319	541	-3.0408	0.0012
USA	0.0545	0.1246	6439	0.4372	0.3310

Extra: In words, select those countries, where you can not reject the alternative that the prices are different. With the command 'this is red you can highlight these countries!

Countries, where we can not reject the alternative with 95% confidence (or with 5% significance level), that the prices are different, hence retail and online prices might differ: Germany, Japan, South Africa

Conclusion

HERE COMES WHAT WE HAVE LEARNED AND WHAT WOULD STRENGHTEN AND WEAKEN OUR ANALYSIS.