Task 2

Team A

October 25, 2020

Introduction

In this exercise, we want to compare the margherita pizza price and beverage price from pizza restaurants in inner and outer Budapest. Data in this set were collected from different online sources, such as restaurant website and www.netpincer.hu. There are 44 observations from inner and outer Budapest pizza restaurants on their 32cm Margherita Pizza and 0.5l beverage price.

All three members of the team recorded the observations in one Google Sheet. However, some duplicates were found. We decided to leave one such observation and drop the other duplicate observations as they had the same value for all variables. Deciding which method to use for measuring the distance was problematic. We decided to use "Measure distance" functionality of Google Map as it works as a "ruler on the map" manner and shows the same distance between two points no matter when the distance was checked. Deciding which size of pizza to choose was problematic, too. As the pizza of size 30cm. 31cm, and 32cm were observed most frequently, we decided to record these pizzas in one variable, in column marpizza_price.
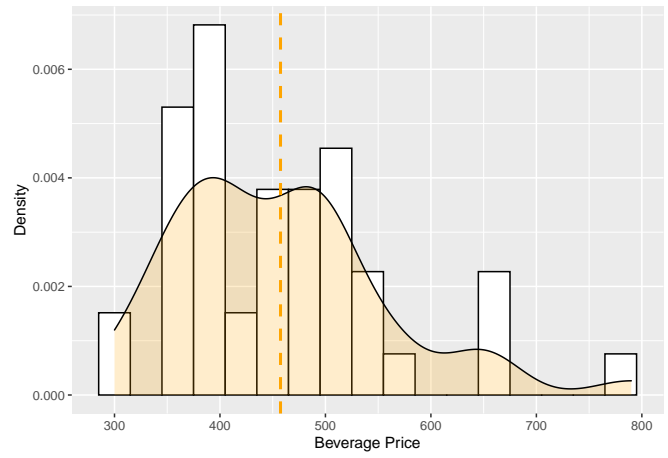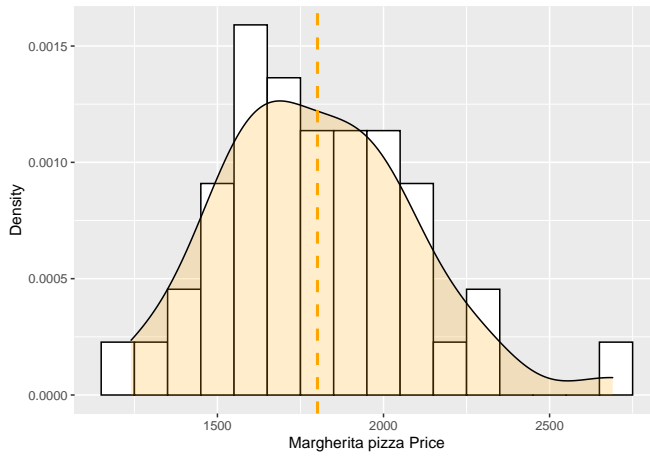
Descriptive Statistics

The following table shows some important summary in the dataset on margherita pizza price and beverage price. Average price for margherita pizza is HUF1801, and HUF457 for beverage. The mean is a little bit higher than the median in both pizza and beverage prices, which suggests a somewhat skewed distribution with a samll right tail.

Obs	n	mean	median	min	max	sd	skew
Pizza	44	1801.18	1790	1240	2690	291.30	0.55
Beverage	44	457.41	450	300	790	101.78	1.00

Descriptive Graphs

The following two histograms side by side shows the price distributions of margherita pizza and beverage price in the whole data. A kernel density estimate is also added to the histogram for completeness. The bin size for margherita pizza is HUF100, and HUF30 for beverage. The histograms reveal many important features: Both price distributions have a single mode. The price distribution for beverage is more spread out, and has a large difference between the minimum and maximum price.

Conditional Descriptive Graphs

Hypothesis Test

Our question here is whether the average Margherita Pizza price in Inner and Outer Budapest the same? The following is our defined statistic, null hypothesis, and its alternative: $\mu = \bar{p}_{inner} - \bar{p}_{outer}$

$$H_0 : \mu = 0$$
$$H_A : \mu \neq 0$$

Our pre-set significance level is at 5%. The following table statistics are obtained from the hypothesis test:

t	df	p-value	95 % confidence interval	mean of x
1.31	21	0.20	-68.96 to 303.14	117.09

The t-statistic is 1.31, well within $\pm 2SE$ (Do not Reject region), thus we cannot reject the null hypothesis of zero difference. In addition, from P-value of 0.2047, which means that the smallest significance at which we can reject the null is 20.47%, far larger than our pre-set initial target of 5%, we cannot reject the null.

The 95% confidence interval [-68.95722, 303.13904] contains H_0 value of zero, which means we are 95% confident that the true average price will not fall outside this range. We can expect that on average, the null hypothesis is rejected 5% of the time given a 5% level of confidence and that on average, a type I

error will occur 5% of the time.

In one word, the average Margherita Pizza price in Inner and Outer Budapest is the same, based on our study.