Question 6

Georgios Pediaditis

3/5/2021

1. Summarising price column

Copenhagen

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    10.0    77.0   105.0   118.4   144.0  5620.0
```

London

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     7.0    54.0    95.0   128.1   160.2  9534.0
```

From the data above we see that both London prices and Copenhagen prices are highly skewed.

Location The median in Copenhagen(105) prices is higher compared to London(95)

Spread The IQR in Copenhagen is smaller (67) compared to London (106.2).
That means there is higher spread on London prices

Shape On both cities Median is closer to 1st Quartile, meaning the data are highly skewed to the right. Another evidence our data are not symmetric is the difference between Median and Mean

2. Confidence Interval

Copenhagen

Our data above show us that the price distribution is skewed to the right. So in order to proceed we would apply a log transformation in order to normalize the distribution.

```
##
##  One Sample t-test
##
## data:  copenhagen$price2
## t = 1378.9, df = 21300, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  2.015004 2.020741
## sample estimates:
## mean of x
##  2.017872
```

Our one-sample confidence interval for the log transformed data is (2.015004,2.020741) or (103.5152,104.8917) for the original data.

London

Our summary data above show us that the price distribution is skewed to the right. So in order to proceed we would apply a log transformation in order to normalize the distribution.

```
##
##  One Sample t-test
##
## data:  london$price2
## t = 1609.6, df = 64143, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.983397 1.988233
## sample estimates:
## mean of x
##  1.985815
```

Our one-sample confidence interval for the log transformed data is (1.983397,1.988233) or (96.24917,97.32692) for the original data.

3.Two-sample t-test

HO: The population mean prices are the same (m1=m2)

H1: The population mean prices are different (m1!=m2)

```
##
##  Two Sample t-test
##
## data:  prices by towns
## t = 13.932, df = 85443, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.02754753 0.03656728
## sample estimates:
## mean in group Copenhagen     mean in group London
##                 2.017872                 1.985815
```

From the two sample t-test we can conclude that since the confidence interval doesnt contain 0 and the p-value is really small it means that there is a statistical significant difference between the mean price in Copenhagen and London. Since confidence interval is entirely on the positive range it means that the mean price in Copenhagen is higher than in London

The mean difference is highly likely to lie between 0.02754753 and 0.03656728 (log transformed) or between 1.065485 and 1.087846 (price difference)