1. **Problem definition and description**

The main target of the project is to i) find the population proportion of call types of taxis and ii) to provide its error range with a 95% level of confidence. The characteristics of the population could be figured out through these processes using the given sample.

1. **Core code**

# define a calculator

class prop\_calculator():

def \_\_init\_\_(self, prop\_num, sample\_num):

# a number of the counted call types

self.\_prop = prop\_num

# a total number of trajected samples

self.\_samples = sample\_num

# calculate sample ratio

def sample\_prop(self):

return round(self.\_prop / self.\_samples, 5)

# calculate sample standard deviation

def sigma\_calc(self, p\_hat = None):

if p\_hat == None:

p\_hat = self.sample\_prop()

return round(np.sqrt((p\_hat \* (1 - p\_hat)) / self.\_samples), 5)

# calculate confidence interval

def confidence\_interval(self, p\_hat = None, sigma\_hat = None):

if p\_hat == None:

p\_hat = self.sample\_prop()

sigma\_hat = self.sigma\_calc(p\_hat)

return round(p\_hat - 1.96 \* sigma\_hat, 5), round(p\_hat + 1.96 \* sigma\_hat, 5)

# draw pie charts

import matplotlib.pyplot as plt

ratio = [a\_ratio, b\_ratio, c\_ratio]

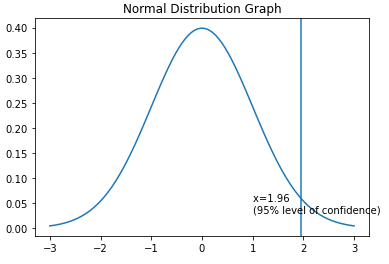
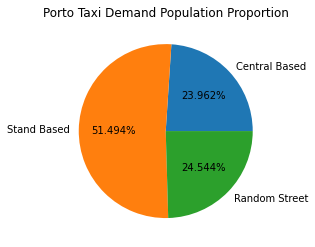
labels = ["Central Based", "Stand Based", "Random Street"]

plt.pie(ratio, labels = labels, autopct='%.3f%%')

plt.title("Porto Taxi Demand Population Proportion")

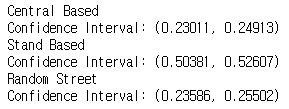
plt.show()

1. **Results and plots**



**Figure 1, 2 (left) Pie chart of Porto Taxi Demand**

**(right) Normal Distribution Graph with x=1.96**

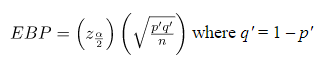


**Figure 2 Confidence Interval of the Proportion (with a 95% level of confidence)**

1. **Discussion**

The taxi demand population proportion was shown in the pie chart above (Figure 1). Since the chart was drawn only by dividing the number of each demand by a total number of demands, it perfectly explains the given sample but not the whole population. Therefore, to figure out the characteristics of the original population, we need a type of estimate that gives the range of values for the proportion of each demand of the original population.

The intervals were estimated based on the assumption that the original population’s distribution resembles the normal distribution (Figure 2). If the error bound for a proportion is provided, we can figure out the minimum sample size that can represent the original population effectively. (See the formula below.)



Not the error bound for a proportion but only the level of confidence was provided, but the sample size was considered large enough to conduct population estimation since it was above 7000.

Usually, the confidence intervals and levels are frequently misunderstood like: the given realized interval there is a 95% possibility that the population parameter lies within the interval. However, the confidence intervals only relate to the reliability of the estimation procedure.

1. **Refernces**

Barbara Illowsky and Susan Dean, 2019. *Introductory Statistics*. Available online: https://opentextbc.ca/introsta  
topenstax/chapter/a-population-proportion/