Homework 4

SDGB 7844, Prof. Nagaraja

Yuwen Wu (A16770560) 3:30-5:30

11/15/2018

Question 1

Why would a company want to run an experiment? Why not just buy the list and see what happens?

ENTER YOUR ANSWER HERE.

It is because company does not want to invest too much if name list would not bring back enough return. The company want to see if this investment is worthy or not.

Question 2

Why would the holder of the list agree to allow the potential purchaser to run an experiment?

ENTER YOUR ANSWER HERE.

This could be an advertisement of selling name list. Also, the holder of name list has enough confidence to pass the test and to sell their name list.

Question 3

If you wanted to run a hypothesis test on the profitability of the list at the $\alpha = 0.05$ level, what would your hypotheses be? What does μ represent?

ENTER YOUR ANSWER HERE.

H0: mu (the profit of each name) <= 3 (the cost of each name) Ha: mu (the profit of each name) > 3 (the cost of each name)

(mu) represents to the profit per name without cost

Question 4

Identify the population, parameter, sample, and statistic in this scenario.

ENTER YOUR ANSWER HERE.

population: name list sale total sample size: 225 parameter: hypothesis true profit per sale statistic: t-statistic: (sample mean of profit - hypothesis mean of profit)/(sd / sqrt(degree of freedom - 1))

Question 5

In your hypotheses in question 3, what would it mean to make a Type I error in this context? What is the probability of making such an error?

ENTER YOUR ANSWER HERE.

The type I error means that when the null hypothesis is correct we reject the null hypothesis according to the result of test. The probability of making a type I error is alpha which is 0.05. In this senario, type I error means that the actual profit is less than $3 per name in the name list, but we reject H0 and buy this name list with each name cost of 3.

Question 6

With the data you will use to test your hypothesis, (a) construct a histogram, (b) compute summary statistics (minimum, median, mean, maximum, and standard deviation), and (c) compute the fraction of people who bought nothing from Performance Tires. Describe the shape of the data. Remember to include the units of measurement.

Solution:

Purchase from Name List
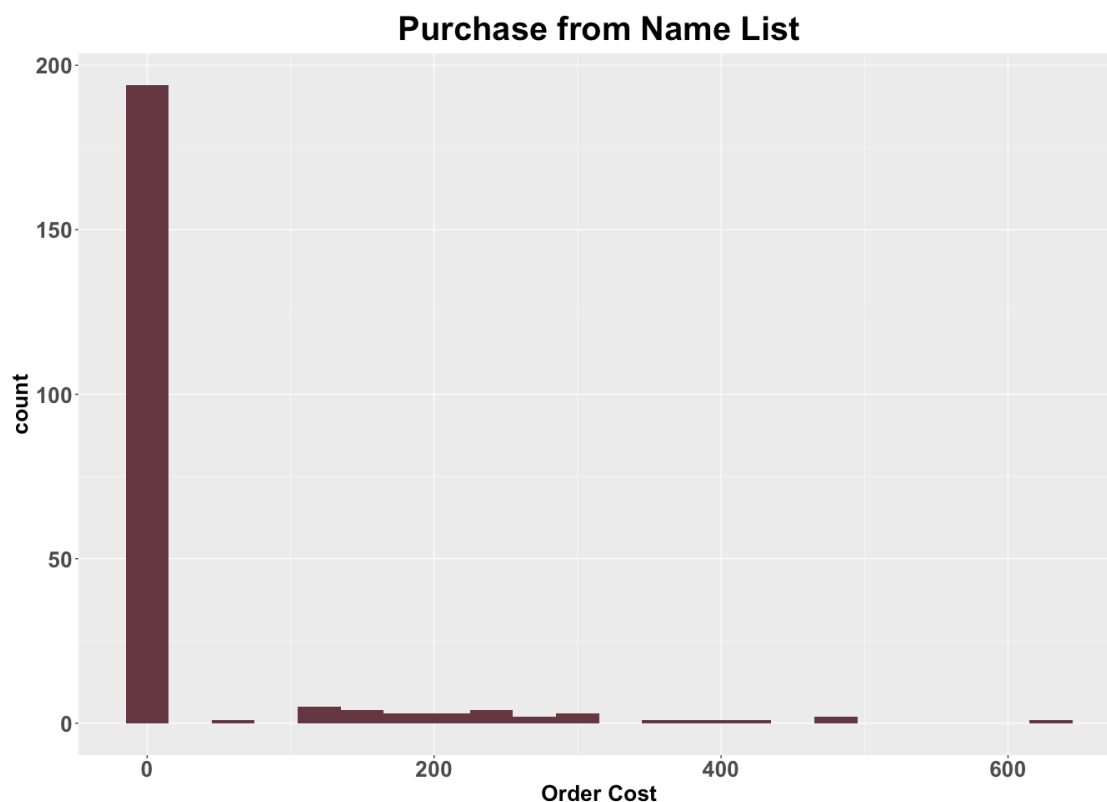
```
##     order_cost
##  Min.    :  0.00
##  1st Qu.:  0.00
##  Median :  0.00
##  Mean   : 33.66
##  3rd Qu.:  0.00
##  Max.   :638.92

## [1] 97.01957

## [1] 0.8622222

##     order_cost
##  Min.    :  0.000
##  1st Qu.:  0.000
##  Median :  0.000
##  Mean   :  6.733
##  3rd Qu.:  0.000
##  Max.   :127.784

## [1] 19.40391

## [1] 0.8622222
```
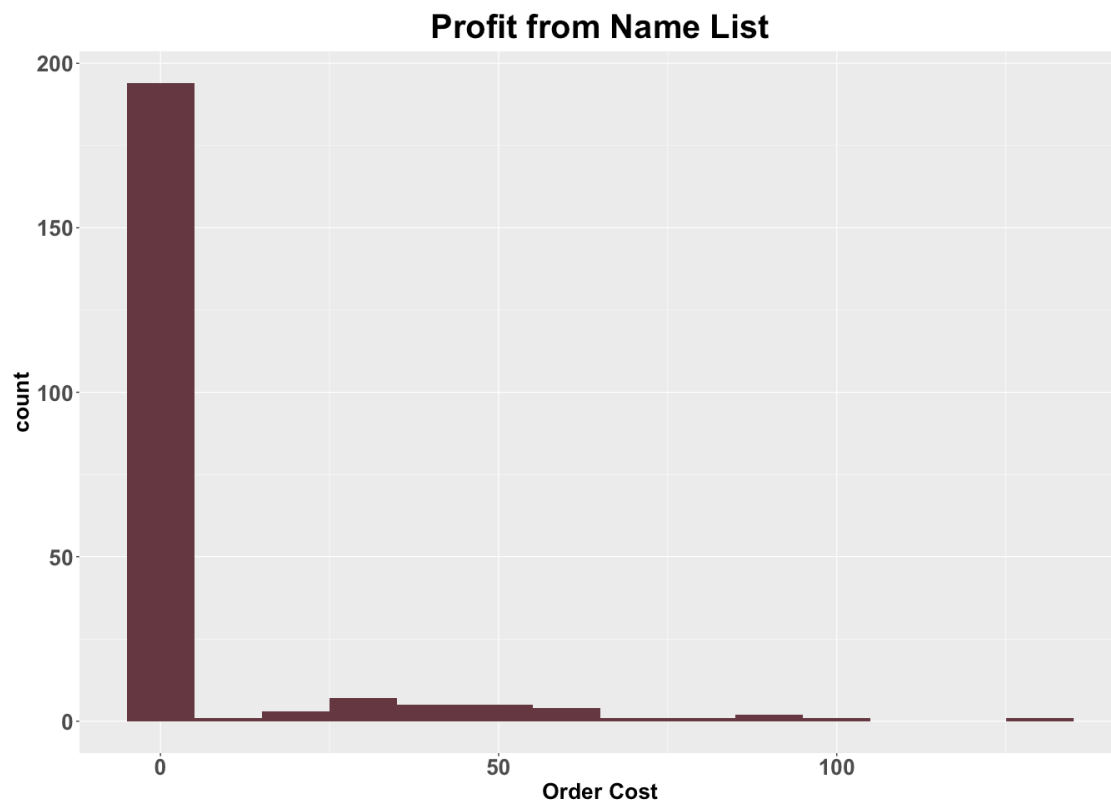
Profit from Name List

For gross order data The Min. of gross order is $0.00
The Median of gross order is $0.00
The Mean of gross order is $33.66

The Max. of gross order is $ 638.92
The standard deviation of gross order is $97.01957 The gross order data is right skewed.

I would use profit data to plot the histogram and to do the test The Min of profit is $0.000
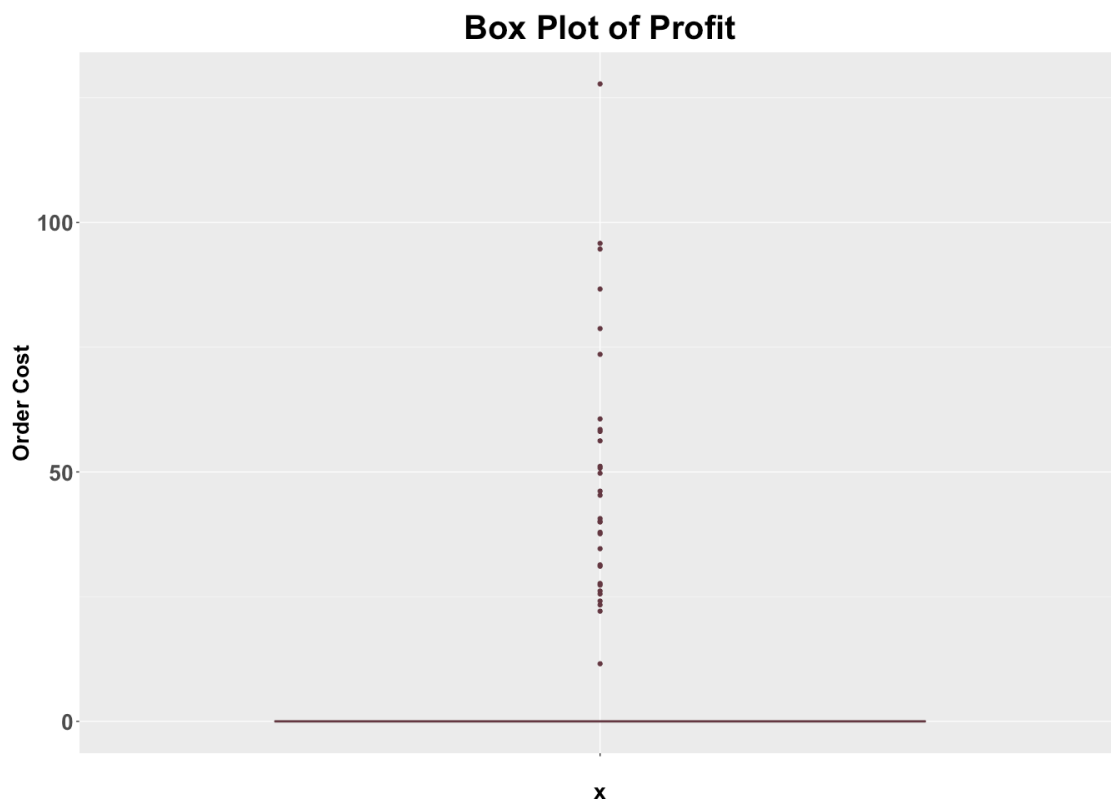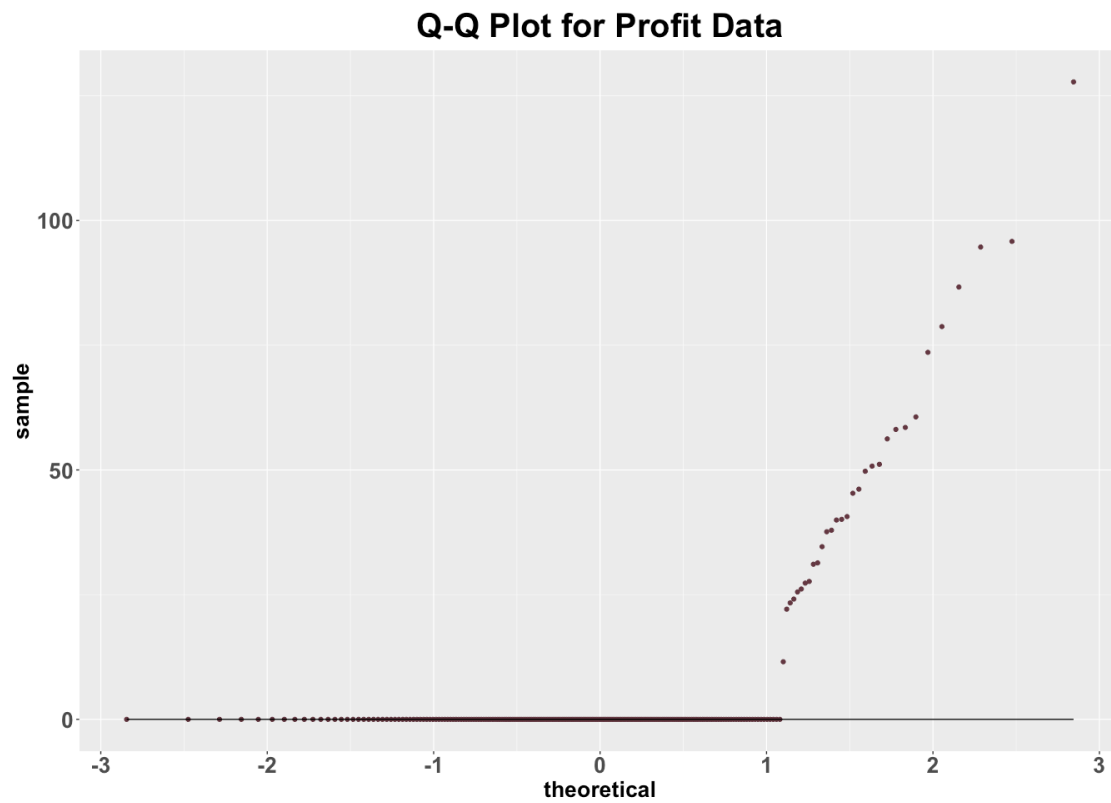The Median of profit is $0.000
The Mean of profit is $6.733
The Max of profit is $127.784
The standard deviation of profit is $19.40391 The shape of the data is right skewed. Almost 86% people bought nothing

Question 7

Check the assumptions for a one-sample t-test. Are they satisfied for this data? Explain your answer.

Solution:

Q-Q Plot for Profit Data

Box Plot of Profit

ENTER YOUR ANSWER HERE.

There are three assumption in the one-sample test 1) data are independent, which means the data should be randam selected 2）data are identically distributed 3）data are normally distributed, or large enought under the Central Limited Theorm From the qq plot, the data is not normally distributed. However, the data is large enough with 225 data and the data are identically distributed. Also, these data all belong to order data. The only thing we cannot make sure is whether this dataset is randomly selected from population. And, from box plot, we can see that outliers would have little affect on the mean and sd of the population because most of data are zero.

Question 8

Test the hypotheses you specified in question 3 and provide a recommendation to the company. Remember to identify the test statistic, degrees of freedom, p-value, and conclusion (don't just show the output of your R code)

```
t.test(y, alternative="greater", mu=3, conf.level = 0.95)

##
##  One Sample t-test
##
## data:  y
## t = 2.8857, df = 224, p-value = 0.002144
## alternative hypothesis: true mean is greater than 3
## 95 percent confidence interval:
##  4.596261      Inf
## sample estimates:
## mean of x
##  6.732871
```

sample mean of profit per one in the name list : 6.732871 degree of freedom: (225-1)=224 t-statistic= (6.732871-3)/(19.40391/sqrt(224)) t-statistics is 2.8857, which means 2.8857 standard errors away from the hypothesized population mean, $3 cost per one p-value 0.002144 (Prob(T > t statistic)) one-sided 95% confidence interval=> 4.596261~Inf

The mean profit of sample, 6.732871, is significantly higher than cost of $3 per name, and in the one-sided 95% confidence interval of (4.566261~inf). The p-value is 0.002144 and is smaller than alpha of 0.05, which is large enough to reject the H0. So, the company should buy the names.

Question 9

What is the probability of making a Type II error with your hypothesis test in question 3 if the average profit was actually $2?

```
pw <- pwr.t.test(d=(2-3)/sd(y$order_cost), n=nrow(x), sig.level=0.05,
type="one.sample",
          alternative="greater")
1-pw$power
```

```
## [1] 0.9921446

pw

##
##       One-sample t test power calculation
##
##              n = 225
##              d = -0.05153599
##      sig.level = 0.05
##          power = 0.007855414
##    alternative = greater
```

The power of test is the probability we will reject the null hypothesis correctly, which is mu<= 3 in my hypothesis. In this case, the probability of rejecting the null hypothesis，which is mu<= 3 in my hypothesis，is 0.78% when the true average profit is 2. Also, the probability of Type II error, which is not rejecting the null hypothesis when the alternative is true, is 99.21% in this hypothesis.