SDGB 7844 HW 4: One-Sample t-Test

Jiayin Hu

2018/11/15

Q1

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

Through the experiment, the company evaluates whether the list makes an effect of increasing sales. Then the company can make a decision whether to buy the list in order to make profit.

Because buying the list needs money, it is better to determine whether the list can bring more profit first. If the company buys the list, but it does not make more profit, the company will suffer a loss.

Q2

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

Because though the potential purchasers could run an experiment, they could not get the list actually. Instead, the list holder provides the promotional mailing service. Therfore, the experiment would not disclose the list information. Under this condition, if the potential purchasers get desirable result, they are highly possible to buy the list. In the contrast, if they do not have opportunity to make an experiment, it is possible that they do not believe the information and give up the purchase.

Q3

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

The null hypothesis is $H_0: \mu = 3$ and the alternative hypothesis is $H_1: \mu > 3$, where μ is the average profit (not including the \$3 cost of the name) for each individual.

Q4

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

Population is all people on the mailing list. The parameter is the mean of profit for every person on the mailing list. Sample is the 225 names selected in the experiment. Statistic is the the mean of profit for every person which is selected in the experiment.

Q5

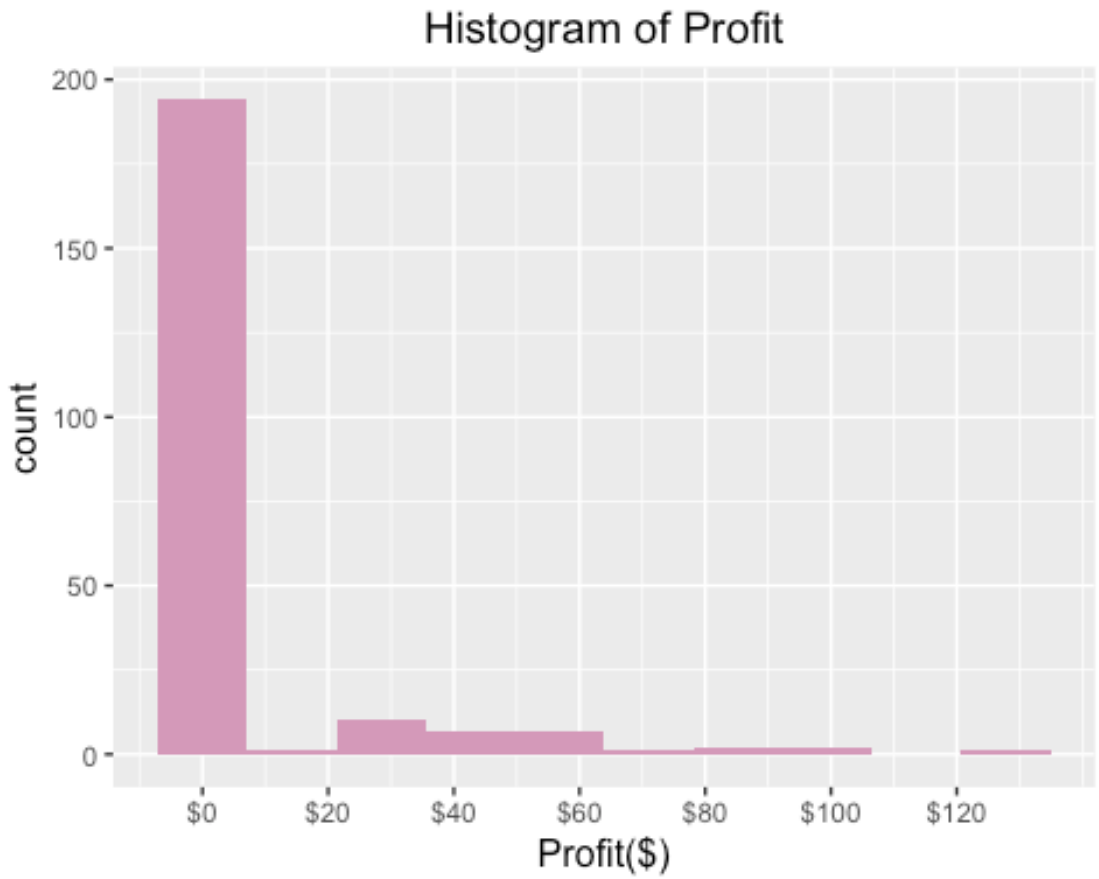
In your hypotheses in question 3, what would it mean to make a Type I error in this context? What is the probability of making such an error?

Type I error in this case means that mean of profit on each person is not greater than $3, but the company thinks the list can help them make profit and buy it. The probability of making Type I error is equal to the significant level 0.05.

Q6

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.

```
require(tidyverse)
require(readxl)
require(knitr)

data <- read_excel("direct_mail.xlsx")

profit <- data * 0.2
profit <- rename(profit, profit = order_cost)

histogram <- profit %>% ggplot(aes(profit)) +
  geom_histogram(bins = 10, fill = "#D499B9") +
  ggtitle("Histogram of Profit") +
  scale_x_continuous(name = "Profit($)", breaks = seq(0, 130, by = 20),
                     labels = paste("$", c(0, 20, 40, 60, 80, 100, 120),
                                    sep = "")) +
  theme(plot.title = element_text(size = 14, hjust = 0.5),
        axis.title = element_text(size = 12))
histogram
```




```
options(tibble.width = Inf)
summaryTable <- profit %>% summarize(min(profit), median(profit),
                                     mean(profit), max(profit), sd(profit))
kable(summaryTable, caption = "Summary Statistics", align = "c")
```

Summary Statistics

min(profit)	median(profit)	mean(profit)	max(profit)	sd(profit)
0	0	6.732871	127.784	19.40391

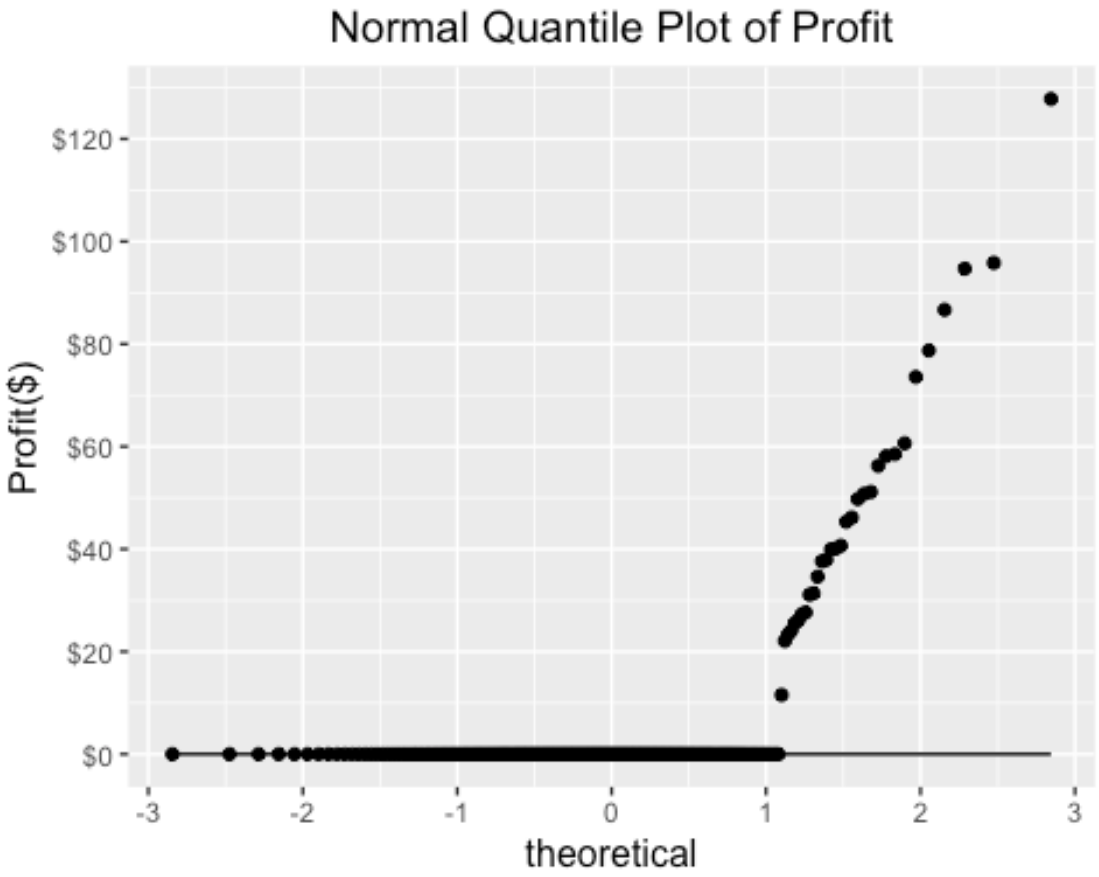
```
num_nothing <- nrow(filter(profit, profit == 0))
num_total <- nrow(profit)
prop_nothing <- num_nothing/num_total
prop_nothing

## [1] 0.8622222
```

The histogram shows right-skewness. The proportion of people who bought nothing from the company is 86.22%

Q7

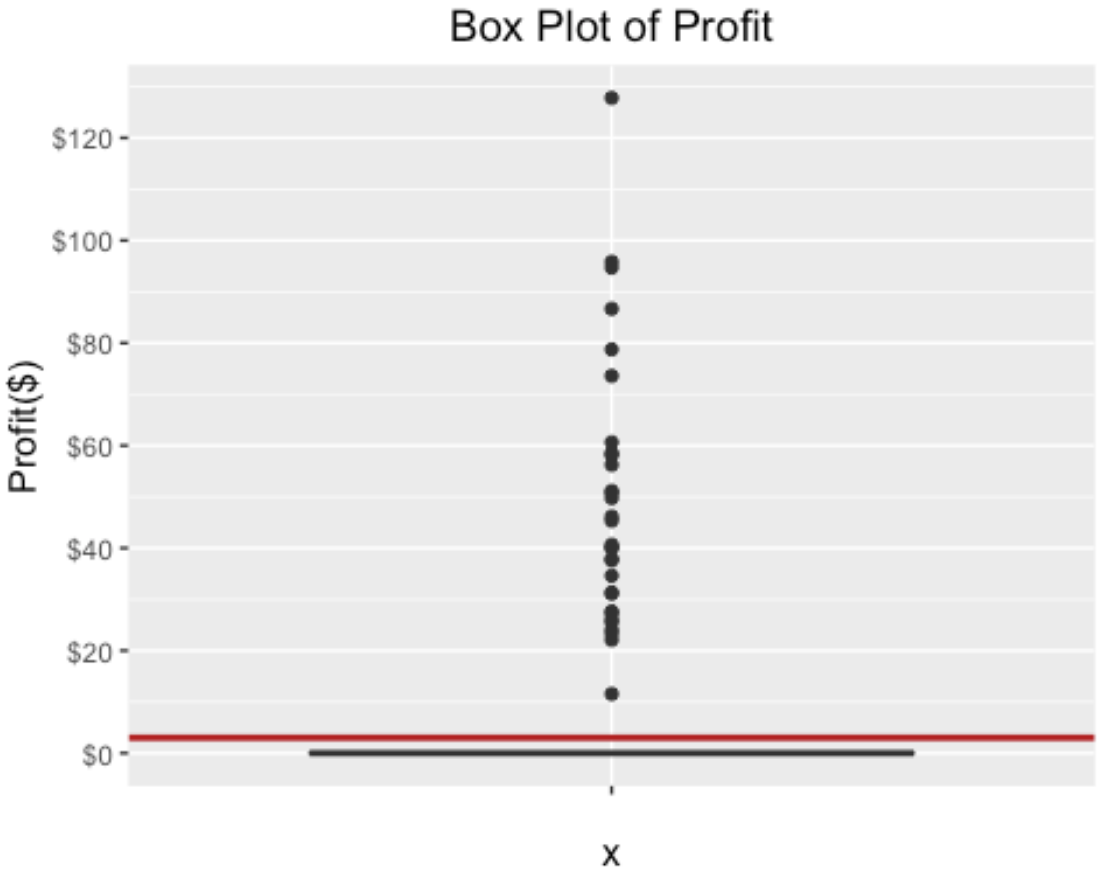
Check the assumptions for a one-sample t-test. Are they satisfied for this data? Explain your answer.

```
profit %>% ggplot(aes(sample = profit)) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Normal Quantile Plot of Profit") +
  scale_y_continuous(name = "Profit($)", breaks = seq(0, 130, by = 20),
                     labels = paste("$", c(0, 20, 40, 60, 80, 100, 120), sep
= "")) +
  theme(plot.title = element_text(size = 14, hjust = 0.5),
        axis.title = element_text(size = 12))
```




```
profit %>% ggplot(aes(x = "", y = profit)) +
  geom_boxplot() +
  geom_hline(yintercept = 3, color = "firebrick", size = 1) +
  ggtitle("Box Plot of Profit") +
  scale_y_continuous(name = "Profit($)", breaks = seq(0, 130, by = 20),
                     labels = paste("$", c(0, 20, 40, 60, 80, 100, 120), sep
= "")) +
```

```
  theme(plot.title = element_text(size = 14, hjust = 0.5),
        axis.title = element_text(size = 12))
```



First, the 225 names are random sampled from the list, which matches the assumption.

Second, according to the normal quantile plot, we can see that the data does not comply normal distribution. Although the profit is not nornally distributed, but n=225, which is great enough to ignore the normally distribution assumption.

Then, the box plot of profit shows that all points except the no purchase are outliers, which can influence the mean and standard deviation. These points, however, are reasonable, because of individuals' different preferences.

Q8

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(x = profit$profit, alternative = "greater",
       mu = 3, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  profit$profit
## 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
```

According the output of R, t statistic is 2.8857. Degrees of freedom is 224. P-value is 0.002144, which is less than 0.05. The conclusion is the true mean is greater than 3.

Q9

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

```
require(pwr)
```

```
## Loading required package: pwr
```

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

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

```
type2error <- 1 - power$power
type2error
```

```
## [1] 0.9921446
```

If the average profit is $2, the probability of making a Type II error is 0.9921446.