# Discussion 11

## Practice Problem

### Problem 1

A scientist wishes to investigate whehter glucose level in the diabetes patients has changed or not after the patients have taken a kind of new drug. Previous records shows that the average glucose level in the bloodstream of patients is 4.5 ($mg/dLi$) and the true standard deviation among loads is 2 ($mg/dLi$). Now we want to detect a change of the percent of glucose level to 2.40 ($mg/dLi$). What's the sample size if we hope to achieve a power of 0.9 and the significance level $\alpha$ is 5%?

### Problem 2

Mcdonalds wants to know whether there is a significant (or only random) difference in the average cycle time to deliver an order between itself and its oppent Pizza Hut. Here is the data collected from a sample of deliveries of two companies.

| Mcdonalds | Pizza Hut |
|---|---|
| 11.13 | 6.23 |
| 10.98 | 5.98 |
| 10.72 | 6.33 |
| 10.81 | 6.92 |
| 10.24 | 7.41 |
| 12.93 | 5.84 |
| 11.22 | 6.71 |
| 12.14 | 8.32 |
| 11.77 | |

1. First draw a dotplot that shows the delivery time for two companies at the same graph. What do you find?
2. Take a *logarithm* transformation about the data, and then draw the dotplot of the data, what do you find?
3. Use the *logarithm* transformed data to construct 95% confidence interval for the difference in mean delivery time between two companies by first log-transforming the data and then back-transforming the interval.

## Solution

1. (a) Suppose the sample size is $n$. First we state the hypothesis of this Hypothesis Testing Problem, and then calculate the rejection region under the $H_0$.

$$H_0 : \mu = 4.5$$

$$H_A : \mu \neq 4.5$$

   (b) The test statistics we use is $\bar{X}$, The sampling distribution of $\bar{X}$ under $H_0$ is a normal distribution with mean 4.5 and standard deviation is $\frac{2}{\sqrt{n}}$. Thus the 0.975 and 0.025 quantile are respectively $4.5 - 1.96 * \frac{2}{\sqrt{n}}$and $4.5 + 1.96 * \frac{2}{\sqrt{n}}$ The rejection region at a significance level $\alpha = 0.05$ is $(-\infty, 4.5 - 1.96 * \frac{2}{\sqrt{n}})$ and $(4.5 + 1.96 * \frac{2}{\sqrt{n}}, \infty)$.

(c)

$$Power = P(\bar{X} \in RejectionRegion|H_A)$$

Under the condition of $H_A$, $\bar{X} \sim N(2.5, \frac{2}{\sqrt{n}})$. Thus $Power = P(\bar{X} < 4.5 - 1.96 * \frac{2}{\sqrt{n}}|H_A) + P(\bar{X} > 4.5 + 1.96 * \frac{2}{\sqrt{n}}|H_A)$

One can verify that the second term on the right-handside is negelible even when $n$ is quite small compared to the first term, thus we can focus on the first term only. If we want the power at $\mu = 2.5$ is larger than 0.9, we should set $P(\bar{X} < 4.5 - 1.96 * \frac{2}{\sqrt{n}}|H_A) \geq 0.9$.

(d) Under the condition of $H_A$, the sampling distribution of $\bar{X}$ is a normal distribution with mean 2.5 and standard deviation of $\frac{2}{\sqrt{n}}$. Thus $4.5 - 1.96 * \frac{2}{\sqrt{n}}$ is the 0.9 quantile a random variable distributed as $N(2.5, \frac{2}{\sqrt{n}})$. 0.9 quantile for a standard normal distribution is 1.28, thus correspondently, the 0.9 quantile for random variable distributed as $N(2.5, \frac{2}{\sqrt{n}})$ is $2.5+1.28*\frac{2}{\sqrt{n}}$.

(e) Set two expressions above so that to determine how large the sample size should be to reach a power of 0.9.

$$2.5 + 1.28 * \frac{2}{\sqrt{n}} = 4.5 - 1.96 * \frac{2}{\sqrt{n}}$$

we have $n \geq 10.497$. Thus the sample size should be at least 11 to achieve a power of 0.9 at $\mu = 2.5$.

2. (a) R codes for drawing dotplot of data is as follows:

```
Mc=c(11.13,10.98,10.72,10.81,10.24,12.93,11.22,12.14,11.77)
Ph=c(6.23,5.98,6.33,6.92,7.41,5.84,6.71,8.32)
data=c(Mc, Ph) #import data of Mcdonalds and Pizza Hut
type=c(rep(1,length(Mc)),rep(2,length(Ph))) #indicator of type of company
type=factor(type)
library(lattice)
dotplot(data~type,main="Before log transformation")
```

From the **Figure 1** we can see that the data of delivery time is highly skewed, and also the variance of the two population is not equal, which violates the assumption of constructing confidence interval for the difference of two population mean.

(b)
```
data=log(c(Mc,Ph))
dotplot(data~type,main="After log transformation")
```

After we take *log* transformation, from **Figure 2** data is not skewed, and approximately symmetric distributed. So the assumption for constructing confidence interval for the difference of two population mean is satisfied.

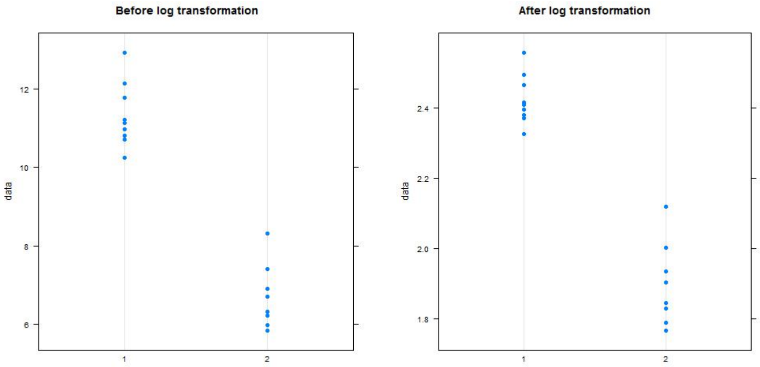
(c)
```
t.test(log(Mc),log(Ph),var.equal=T)

        Two Sample t-test

data:  log(Mc) and log(Ph)
t = 11.2631, df = 15, p-value = 1.023e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.4268337 0.6260906
sample estimates:
mean of x mean of y
```

```
 2.424879  1.898417
```

Thus the 95% confidence interval of the difference is $(exp(0.4268337), exp(0.6260906)) = (1.53, 1.87)$.



(a) Figure 1

(b) Figure 2