Assignment2

Fan Chen

2022-09-20

```
time_data = read.table("DeliveryTimes.csv",header=TRUE, sep=",")
time_data
```

Problem 1

a. The results are shown as follows:

   The sample average is **45.10607**

   The sample standard deviation is **2.469763**

   The sample variance is **6.099727**

```
sample_avg = mean(time_data$DeliveryTime)
sample_sd = sd(time_data$DeliveryTime)
sample_var = var(time_data$DeliveryTime)
sample_avg ## 45.10607
sample_sd ## 2.469763
sample_var ## 6.099727
```

b. As we all know that the $Var(x) = \frac{1}{12}(b-a)^2$ and therefore the $Std(x) = \frac{1}{2\sqrt{3}}(b-a)$ Supposing that it conforms to the uniform distribution, the variance is **75**

```
var_sp = 1/12 * (60-30) **2
var_sp ## 75
```

   And also, the std is **8.660254**

```
std_sp = 1/(2*sqrt(3))*(60-30)
std_sp #8.660254
```

   In this case, we can find that the difference is so large between the theoretical value and our calculative values. And it means that it's highly possible that it doesn't conform to the uniform distribution.

c. Supposing that it conforms to the normal distribution, it approximately conforms to the $N(45.1, 2.47^2)$. For the prob of 95%, the time locates in **[40.25889, 49.94111]**

```
qnorm(0.975, 45.1, 2.47) ## 49.94111
qnorm(0.025, 45.1, 2.47) ## 40.25889
```

   And for the prob of 80%, the time locates between **[41.93457, 48.26543]**

```r
qnorm(0.9, 45.1, 2.47) ## 48.26543
qnorm(0.1, 45.1, 2.47) ## 41.93457
```

On the other side, Supposing that it conforms to the uniform distribution, it conforms to the $U(30, 60)$. For the prob of 95%, the time locates between **[30.75, 59.25]**

```r
qunif(0.975, 30, 60) ## 59.25
qunif(0.025, 30, 60) ## 30.75
```

And for the prob of 80%, the time locates between **[33, 57]**

```r
qunif(0.9, 30, 60) ## 57
qunif(0.1, 30, 60) ## 33
```

Therefore, when comparing the 95% and 80% PI, the 80% PI in both distribution are narrower. And when comparing these two distributions, the intervals are much narrower for the normal distribution than the uniform one.

d. For the normal distribution,

```r
P(X <= 45) = pnorm(45, 45.10607, 2.469763) ## 0.4828717
P(X <= 40) = pnorm(40, 45.10607, 2.469763) ## 0.01934668
P(X >= 50) = pnorm(50, 45.10607, 2.469763, lower.tail = F) ## 0.02376547
```

For the uniform distribution,

```r
P(X <= 45) = punif(45, 30, 60) ## 0.5
P(X <= 40) = punif(40, 30, 60) ## 0.3333333
P(X >= 50) = punif(50, 30, 60, lower.tail = F) ## 0.3333333
```

For the prob of less than 45min, the two answers are close. However, when comparing the rest two (less than 40min or more than 50min), the answer varies significantly and the prob of normal distribution is much less than that of uniform one.

e. It really matters. Because if it conforms to the normal distribution, it's a reasonable strategy. Because only about 2% of the pizzas are free and it is acceptable for the pizzeria. However, if it conforms to the uniform distribution between 30 and 60min. It means that one third of the pizzas are free. It's not reasonable for the pizzeria to keep running.

f. Here, the 95% confidence interval can be calculated as:

```r
left = 45.10607 - qnorm(0.975)*2.469763/sqrt(length(time_data$DeliveryTime)) # 44.9767
right = 45.10607 + qnorm(0.975)*2.469763/sqrt(length(time_data$DeliveryTime)) ## 45.23544
```

We find that the confidence interval **[44.9767, 45.23544]** is much narrower than the normal distribution. Because the confidence interval actually is an estimate to the population mean, which is influenced by the number of sample. In this case, the sample size is 1400, which is a large value. As a result, the sample std error will be much smaller than choosing only one value every time. Therefore, the interval is much narrower than the prediction interval.

g. No, it won't change. Because when the population is large enough, the $\frac{\overline{X}-\mu_X}{\sigma_{\overline{X}}} \to N(0,1)$ makes sense for any distribution of X. Therefore, when our μ_X and $\sigma_{\overline{X}}$ remains unchanged. The confidence interval won't change.

h. Let's consider the two-sided null hypothesis that the average time is 50min:

$H_0 : \mu_m = 50$ vs. $H_A : \mu_m \neq 50$

To test it, we can calculate as follows:

```
mean_m = 50 - 45.10607 #4.89393
SE_m = sqrt(2.469763 **2/length(time_data$DeliveryTime)) ## 0.06600719
t_act = mean_m/SE_m # 74.14238
p_value = 2*pnorm(-t_act) #0
```

Actually, the p value is so small that r studio automatically equals it to 0. Or we can compare the t_act, which is 74.14238 to 1.96(95%) and 2.58(99%), we can find that it is larger than any of them. Therefore, we can reject the hypothesis both at 1% and 5% levels. To sum up, the average delivery time is not 50minutes.