NDH802 Solutions to rec. exercises Chap 6 and 7

Huong

6.16

Assume $\sigma = 40, n = 100$.

a. SE of the sample mean

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{100}} = 4$$

```
se = 40/sqrt(100)
```

b. $P(\bar{X} - \mu > 5) = P(\bar{X} > \mu + 5)$. Note that this probability does not depend on μ. Imagine the bell curve of the same variance sliding on the x-axis, the area under the curve does not change.

Now recall the Central Limit Theorem? "When we draw samples from the population that is normally distribution with mean μ and SD σ, the sample mean is normally distributed with mean μ and SD $\sigma/\sqrt{n}$" (which is the SE we found in a.)

To find this probability, we can plug in any μ. For example, these two give the same result:

```
pnorm(q = 5, mean = 0, sd = se, lower.tail = FALSE)
```

```
## [1] 0.1056498
```

```
pnorm(q = 105, mean = 100, sd = se, lower.tail = FALSE)
```

```
## [1] 0.1056498
```

c. $P(\mu - \bar{X} > 4) = P(\bar{X} < \mu - 4)$. Similar to b.,

```
pnorm(q = -4, mean = 0, sd = se)
```

```
## [1] 0.1586553
```

d. $P(\mu - \bar{X} > 3)$ or $P(\bar{X} - \mu > 3)$. Equivalently, we can write $P(\bar{X} < \mu - 3)$ or $P(\bar{X} > \mu + 3)$. Because $\bar{X}$ follows normal distribution, these probabilities should be equal. For fun, let's give it a try.

```
pnorm(q = -3, mean = 0, sd = se)
```

```
## [1] 0.2266274
```

```
pnorm(q = 3, mean = 0, sd = se, lower.tail = FALSE)
```

```
## [1] 0.2266274
```

6.75

Let X be the number of students preferring brand A. We can write that $X \sim Binomial(n = 200, p = 0.6)$.

```r
n = 200
p = 0.6
```

a. We want to compute $P(X > 2/3 * 200)$. We can do it in 2 ways:

The direct way:

```r
a1 = pbinom(q = 2/3 * n, size = n, p = p, lower.tail = FALSE)
```

The approximation way:

As we learn, when we have substantially large number of trials in a binomial distribution, we can approximate it with a normal distribution $N(\mu = np, \sigma^2 = np(1-p))$.

```r
mu = n*p
sigma = sqrt(n*p*(1-p))
a2 = pnorm(q = 2/3 * n, mean = mu, sd = sigma, lower.tail = F)
```

a1 and a2 are closed but not identical.

b. We want to compute $P(0.55 * 200 < X < 0.65 * 200)$. We can do it in 2 ways:

The direct way:

Similar to Assignment 2 Q2, we find $P(X < 0.65 * 200)$ then subtract $P(0.55 * 200 < X)$ from it.

```r
pbinom(q = 0.65 * n, size = n, p = p) - pbinom(q = 0.55 * n, size = n, p = p)
```

```
## [1] 0.8503738
```

The approximation way: very similar to Assignment 2 Q2

```r
pnorm(q = 0.65 * n, mean = mu, sd = sigma) - pnorm(q = 0.55 * n, mean = mu, sd = sigma)
```

```
## [1] 0.8510853
```

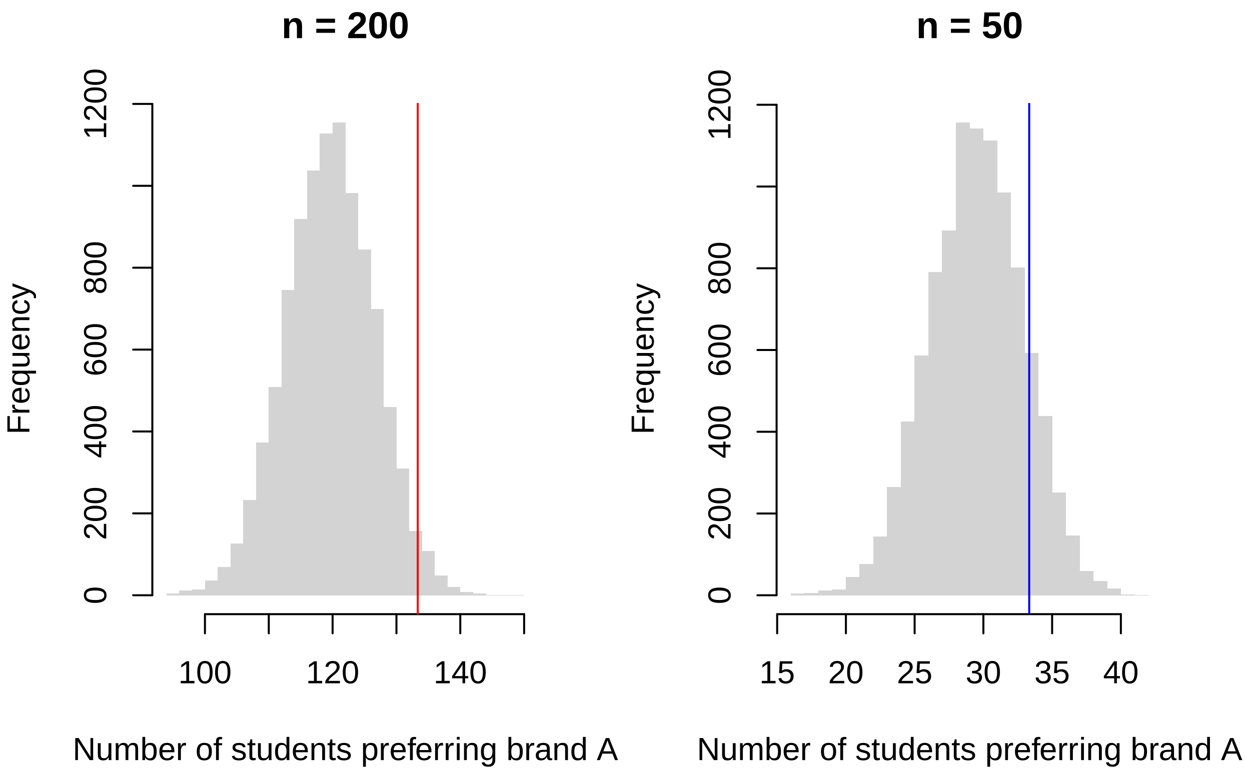
c. We do the same thing we did in question a, with n = 50.

```r
n50 = 50
c1 = pbinom(q = 2/3 * n50, size = n50, p = p, lower.tail = FALSE)
c2 = pnorm(q = 2/3 * n50, mean = n50*p, sd = sqrt(n50*p*(1-p)), lower.tail = F)
c(a1, c1)
```

```
## [1] 0.02471612 0.15609060
```

As we can see from the numbers, when the sample size decreases, the probability of observing 2/3 of the sample members preferring brand A increases. You can refer to the plots below for the visualization. The vertical lines represent the cut off points (2/3 of the students in the sample). Note that the area on the right hand side of the blue line (question c) is visually larger than the area on the right hand side of the red line (question a).

Note: You're not expected to be able to plot this in this course including the exam. Simply run the whole code chunk for the plots to show. But if you'd like to learn, we're here to help.



6.76

Let X be the scores. $X \sim N(\mu, \sigma^2 = 2500)$. A random sample of 25 scores was taken. Let $\bar{x}$ denote the sample mean. Find $P(\bar{x} - 10 < \mu < \bar{x} + 10)$.

Small notes: We use small $\bar{x}$ here because the random variable is defined in such way in the question. Typically in the course we use capital $\bar{X}$ to denote the random variable the sample mean.

Back to the the question:

$$P(\bar{x} - 10 < \mu < \bar{x} + 10) = P(\mu - 10 < \bar{x} < \mu + 10)$$

```
pnorm(10, mean = 0, sd = 10) - pnorm(-10, mean = 0, sd = 10)
```

```
## [1] 0.6826895
```

7.14

It is known that the standard deviation in the volumes of 20-ounce (591-millliliter) bottles of natural spring water bottled by a particular company is 5 millliliters. One hundred bottles are randomly sampled and measured.

```
#Fist thing first, summarize the assumptions
sigma = 5
n = 100
```

(a) *Calculate the standard error of the mean*

The formula to find the standard error SE:

$$SE = \frac{\sigma}{\sqrt{n}}$$

```
standard_error = sigma/sqrt(n)
#by hand, it is exactly the same. the square root of 100 is 10, therefore
standard_error_byhand = 5/10
```

(b) *Find the margin of error of a 90% confidence interval estimate for the population mean volume. From the question, we know the population standard deviation. We therefore use the z-score to calculate the confidence interval.*

The formula to find the margin of error ME

$$ME = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = z_{\alpha/2} * \text{SE}$$

We first define α, which is 1-CI=1-0.9=0.1. To find the z-score in R, we use `qnorm()` with $p = \alpha/2 = 0.1/2 = 0.05$. To get the positive z-score, we add `lower.tail = F`.

```
z_score_90 = qnorm(0.05, lower.tail = F)
```

Otherwise, you can look up the z table if that suits you better. Now we know that the z score is 1.6448536, we just need to plug it in the formula.

```
ME_90 = z_score_90*sigma/sqrt(n)
#alternatively
ME_90_2 = z_score_90*standard_error
```

The margin of error of a 90% confidence interval estimate for the population mean volume is 0.8224268

(c) *Calculate the width for a 98% confidence interval for the population mean volume*

Similar to b, we first find the z score corresponding to CI = 98%. In this case, $p = \alpha/2 = 0.02/2 = 0.01$. Then we find the ME:

```
z_score_98 = qnorm(0.01, lower.tail = F)
ME_98 = z_score_98*sigma/sqrt(n)
```

The width is twice the ME.

```
width_98 = 2*ME_98
```

The z score is 2.3263479, ME is 1.1631739, and the width is 2.3263479

7.21

A random sample of 16 tires was tested to estimate the average life of this type of tire under normal driving conditions. The sample mean and sample standard deviation were found to be 47,500 miles and 4,200 miles, respectively.

```
#Fist thing first, summarize the assumptions
n = 16
x_bar = 47500
s = 4200
```

(a) *Calculate the margin of error for a 95% confidence interval estimate of the mean lifetime of this type of tire if driven under normal driving conditions.*

The formula to find the ME when population variance is unknown.

$$ME = t_{\alpha/2}\frac{s}{\sqrt{n}}$$

We first ask R (or the t-table) the t-score.

```
t_score_95 = qt(0.025, df = n-1, lower.tail = F)
pt(t_score_95, df = n-1, lower.tail = F)
```

```
## [1] 0.025
```

Now that we have t_score = 2.1314495, we plug in the formula

```
ME_95 = t_score_95*s/sqrt(n)
```

The ME is 2238.

(b) *Find the UCL and the LCL of a 90% confidence interval estimate of the mean lifetime of this type of tire if driven under normal driving conditions.*

We first find the ME at 90% confidence level. As you already know the drill, we'll go fast.

```
t_score_90 = qt(0.05, df = n-1, lower.tail = F)
ME_90 = t_score_90*s/sqrt(n)
```

Theoretically, $CI = \bar{x} \pm ME$. We already know $\bar{x}$ and ME, we just need to plug them in. The LCL and UCL are then 45659.3 and 49340.7, respectively.

7.85

A random sample of 174 college students was asked to indicate the number of hours per week that they surf the Internet for either personal information or material for a class assignment. The sample mean response was 6.06 hours and the sample standard deviation was 1.43 hours. Based on these results, a confidence interval extending from 5.96 to 6.16 was calculated for the population mean. Find the confidence level of this interval

```
#Fist thing first, summarize the assumptions
n = 174
x_bar = 6.06
s = 1.43
lcl = 5.96
ucl = 6.16
```

Recall the formula

$$LCL = \bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \text{ and } UCL = \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

Equivalently,

$$LCL = \bar{x} - t_{\alpha/2} * SE \text{ and } UCL = \bar{x} + t_{\alpha/2} * SE$$

In this case, we already know everything and we want to find $t_{\alpha/2}$. With a little transformation,

$$t_{\alpha/2} = \frac{\bar{x} - LCL}{SE} = \frac{UCL - \bar{x}}{SE}$$

The rest are just plugging in numbers

```
SE = s/sqrt(n)
t_score = (x_bar-lcl)/SE
#t_score = (ucl-x_bar)/SE
```

Now that we know the t-score, we can ask R the corresponding probability using `pt()`, which gives us $\frac{\alpha}{2}$.

```
half_alpha = pt(t_score, df = n-1, lower.tail = F)
CL = 1-2*half_alpha
```

The confidence level is then 64.24%.

8.30

Summarize what are given:

```
n_A = 40; s_A = 20; x_bar_A = 340;
n_B = 50; s_B = 30; x_bar_B = 285;
alpha = 1 - 0.9
df = n_A+n_B-2
```

We assume the independent random samples are taken from the normal populations with unknown variances.

Case 1. If the population variances are assumed to be equal, we use formula 8.7 (Newbold 2019, p. 340) to first find the polled sample variance:

```
s_p_squared = ((n_A-1)*s_A^2 + (n_B-1)*s_B^2)/df
```

The ME is:

```
t_score = qt(p = alpha/2, df = df, lower.tail = F)
ME = t_score*sqrt(s_p_squared/n_A + s_p_squared/n_B)
```

The confidence interval is the mean difference ± ME:

```
d_bar = x_bar_A - x_bar_B
UCL_8.30 = d_bar + ME
LCL_8.30 = d_bar - ME
```

Case 2. If the population variances are assumed to be **not** equal, we use formula 8.10 (Newbold 2019, p. 342)

```
nu = ((s_A^2)+(s_B^2))^2 / ((s_A^2/n_A)^2/(n_A-1)+(s_B^2/n_B)^2/(n_B-1))
```

Similarly we find the ME, then the CI

```
t_score_ne = qt(p = alpha/2, df = nu, lower.tail = FALSE)
ME_ne = t_score*sqrt(s_p_squared/n_A + s_p_squared/n_B)
UCL_8.30_ne = d_bar + ME_ne
LCL_8.30_ne = d_bar - ME_ne
```

Key takeaways

Theoretically,

- When population variance/SD is known, we use z-distribution. To find z-score in R, `qnorm(p=alpha/2, lower.tail = F)`
- When population variance/SD is unknown, we use t-distribution. To find t-score in R, `qt(p=alpha/2, df=n-1, lower.tail = F)`
- Pay attention to the difference between matched and independent samples.