Topic 13 a

Confidence Interval for Population Mean, μ, based on the sample mean when σ is known

Definitely review Table 1 on the web page.

As a consequence, for a population, not necessarily normal, with unknown mean, μ, but with known standard deviation, $\sigma = 17.6$, if we take samples of size 32 then the distribution of the sample means, $\bar{x}$'s, will be $N(\mu, 17.6/\sqrt{32})$. But, for a N(0,1) population we can find a value y such that $P(-y < X < y) = 0.95$ by using qnorm((1-0.95)/2, lower.tail=FALSE).

```
      # find y such that P( -y < X < y ) = 0.95
qnorm( (1-0.95)/2, lower.tail=FALSE)
```

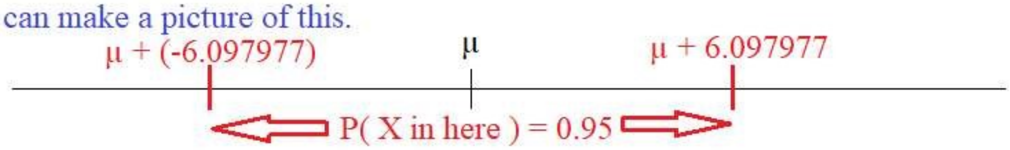
```
>      # find y such that P( -y < X < y ) = 0.95
> qnorm( (1-0.95)/2, lower.tail=FALSE)
[1] 1.959964
```

This means that in a N(0, 1) the probability of getting a random value between -1.959964 and 1.959964 is 0.95, i.e., 95%. If we translate that back to a $N(\mu, 17.6/\sqrt{32})$ then we are looking at the range from $-1.959964 * (17.6/\sqrt{32})$ to $1.959964 * (17.6/\sqrt{32})$. We can compute this as shown:

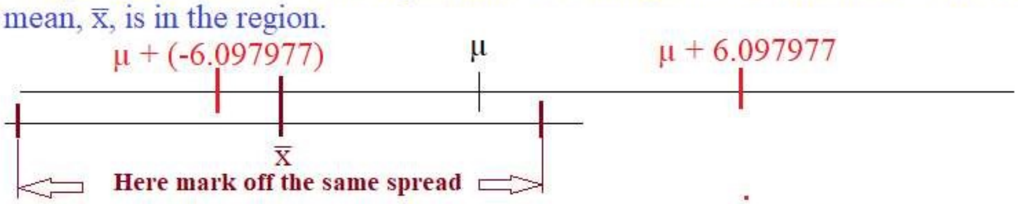
```
      # bring this back to a N( mu, 17.6/sqrt(32))
      # first recompute but save the value
y <- qnorm( (1-0.95)/2, lower.tail=FALSE)
y
-y*17.6/sqrt( 32 )
y*17.6/sqrt( 32 )
```

```
>      # bring this back to a N( mu, 17.6/sqrt(32))
>      # first recompute but save the value
> y <- qnorm( (1-0.95)/2, lower.tail=FALSE)
> y
[1] 1.959964
> -y*17.6/sqrt( 32 )
[1] -6.097977
> y*17.6/sqrt( 32 )
[1] 6.097977
```

We can make a picture of this.

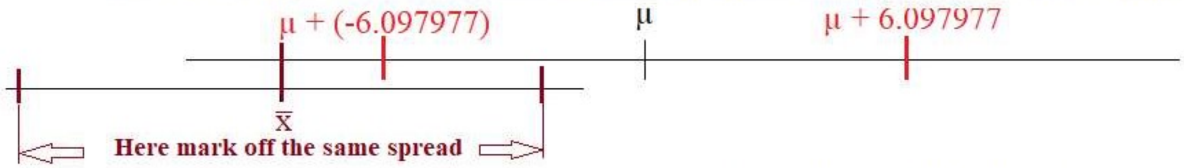


Now, we actually take a sample of size 32 and we find the sample mean, $\bar{x}$. 95% of the time that sample mean will be in the region that we identified. We will look at two cases. First if the mean, $\bar{x}$, is in the region.



Note that if the sample mean is in the region surrounding the true mean, μ, then μ is in the region, of the same width, surrounding $\bar{x}$.

Now we will go back and look at the case where the sample mean, $\bar{x}$, is not in the region.



Note that if the sample mean is outside the region surrounding the true mean, μ, then μ is outside the region, of the same width, surrounding $\bar{x}$.

Here is an important distinction. 95% of the samples of size 32 that we take will have a sample mean, $\bar{x}$, that is beween μ - 6.097977 and μ + 6.097977. On the other hand, if we take a specific sample, its sample mean, $\bar{x}$, either is or is not in that region. There is no probability here. What we can say is that 95% of the samples we take will have the sample mean in that region. Therefore, 95% of the samples that we take will have the true mean, μ, in a region of the same width around the sample mean, $\bar{x}$.

Such a region around the sample mean, $\bar{x}$, is the 95% confidence interval for the true mean of the population, μ.

If we have a population with a known standard deviation, σ, then we can find the α confidence interval by taking a sample of size n, finding the sample mean, $\bar{x}$, then finding the standard N(0,1) value, z, such that $P(X > z) = (1 - \alpha)/2$, then computing $\bar{x} \pm z * \sigma / \sqrt{n}$.

```
     # we can do a few problems
     # find the 95% confidence interval for the mean of a
     # population with known standard deviation 3.67 based
     # on a sample of size 28 with sample mean = 76.9
     #
sigma <- 3.67  #population standard deviation
n <- 28        # sample size
x_bar <- 76.9  # sample mean
alpha <- 0.95  # desired confidence level
z <- qnorm( (1-alpha)/2, lower.tail = FALSE ) # z value for N(0,1)
     # get low and high sides of the confidence interval
x_bar - z * sigma / sqrt( n )  # low side
x_bar + z * sigma / sqrt( n )  # high side
```

```
>      # we can do a few problems
>      # find the 95% confidence interval for the mean of a
>      # population with known standard deviation 3.67 based
>      # on a sample of size 28 with sample mean = 76.9
>      #
> sigma <- 3.67  #population standard deviation
> n <- 28        # sample size
> x_bar <- 76.9  # sample mean
> alpha <- 0.95  # desired confidence level
> z <- qnorm( (1-alpha)/2, lower.tail = FALSE ) # z value for N(0,1)
>      # get low and high sides of the confidence interval
> x_bar - z * sigma / sqrt( n )  # low side
[1] 75.54064
> x_bar + z * sigma / sqrt( n )  # high side
[1] 78.25936
```

So the 95% confidence interval for the population mean is (75.54064, 78.25936). Does this contain the true mean, μ, of the population? We do not know. But we do know that 95% of the confidence intervals that we generate using this methodology will contain the true population mean.

```
23       # find the 93% confidence interval for the mean of a
24       # population with known standard deviation 8.06 based
25       # on a sample of size 35 with sample mean = 1.27
26       #
27 sigma <- 8.06  #population standard deviation
28 n <- 35        # sample size
29 x_bar <- 1.27  # sample mean
30 alpha <- 0.93  # desired confidence level
31 z <- qnorm( (1-alpha)/2, lower.tail = FALSE ) # z value for N(0,1)
32 # get low and high sides of the confidence interval
33 x_bar - z * sigma / sqrt( n )  # low side
34 x_bar + z * sigma / sqrt( n )  # high side
```

```
>       # find the 93% confidence interval for the mean of a
>       # population with known standard deviation 8.06 based
>       # on a sample of size 35 with sample mean = 1.27
>       #
> sigma <- 8.06  #population standard deviation
> n <- 35        # sample size
> x_bar <- 1.27  # sample mean
> alpha <- 0.93  # desired confidence level
> z <- qnorm( (1-alpha)/2, lower.tail = FALSE ) # z value for N(0,1)
> # get low and high sides of the confidence interval
> x_bar - z * sigma / sqrt( n )  # low side
[1] -1.198527
> x_bar + z * sigma / sqrt( n )  # high side
[1] 3.738527
```

So the 93% confidence interval for the population mean is (-1.198527, 3.738527).

```
35 # find the 98.5% confidence interval for the mean of a
36 # population with known standard deviation 15.4 based
37 # on a sample of size 11 with sample mean = 26.7
38 #
39 sigma <- 15.4  #population standard deviation
40 n <- 11        # sample size
41 x_bar <- 26.7  # sample mean
42 alpha <- 0.985 # desired confidence level
43 z <- qnorm( (1-alpha)/2, lower.tail = FALSE ) # z value for N(0,1)
44 # get low and high sides of the confidence interval
45 x_bar - z * sigma / sqrt( n )  # low side
46 x_bar + z * sigma / sqrt( n )  # high side
```

```
> # find the 98.5% confidence interval for the mean of a
> # population with known standard deviation 15.4 based
> # on a sample of size 11 with sample mean = 26.7
> #
> sigma <- 15.4  #population standard deviation
> n <- 11        # sample size
> x_bar <- 26.7  # sample mean
> alpha <- 0.985 # desired confidence level
> z <- qnorm( (1-alpha)/2, lower.tail = FALSE ) # z value for N(0,1)
> # get low and high sides of the confidence interval
> x_bar - z * sigma / sqrt( n )  # low side
[1] 15.4058
> x_bar + z * sigma / sqrt( n )  # high side
[1] 37.9942
```

So the 98.5% confidence interval for the population mean is (15.4058, 37.9942).

```
48       # It is clear that we could do sample after sample in
49       # this way.  There are a few points to observe here, however,
50       # Each time we do the pair of lines, 45 and 46, we
51       # have R compute  z * sigma / sqrt( n ) two times.
52       # We will call this value the Margin of Error, or MOE.
53       # Then we can compute it once.
54 #
55       # Find the 85% confidence interval for the mean of a
56       # population with known standard deviation 6.97 based on
57       # a sample of size 26 with sample mean = 437.2.
```

```
58 z <- qnorm( 0.15/2, lower.tail=FALSE)
59 MOE <- z*6.97/sqrt( 26 )
60 437.2 - MOE  # the low value
61 437.2 + MOE  # the high value
```

```
>      # It is clear that we could do sample after sample in
>      # this way.  There are a few points to observe here, however.
>      # Each time we do the pair of lines, 45 and 46, we
>      # have R compute  z * sigma / sqrt( n ) two times.
>      # We will call this value the Margin of Error, or MOE.
>      # Then we can compute it once.
> #
>      # Find the 85% confidence interval for the mean of a
>      # population with known standard deviation 6.97 based on
>      # a sample of size 26 with sample mean = 437.2.
> z <- qnorm( 0.15/2, lower.tail=FALSE)
> MOE <- z*6.97/sqrt( 26 )
> 437.2 - MOE  # the low value
[1] 435.2323
> 437.2 + MOE  # the high value
[1] 439.1677
```

In this example we cut our number of statements to 4 (lines 58-61) from the 7 (liens 39-46) we had before. Still, we are doing the same steps every time we solve this problem. That is a perfect situation for capturing the steps in a function.

```
62      # do this again with our function
63 source("../ci_known.R")
64 ci_known( 6.97, 26, 437.2, 0.85 )
```

```
>      # do this again with our function
> source("../ci_known.R")
> ci_known( 6.97, 26, 437.2, 0.85 )
    CI Low    CI High        MOE  Std Error
435.232262 439.167738   1.967738   1.366929
```

Be sure to go to the web page to look at examples of repeated samples from a population with known sigma to see the confidence intervals generated by those samples and to see that sometimes we do and sometimes we don't have the true population within our interval.

Also, look at the following examples to see what happens as we increase the confidence interval percent. Also, what happens when we increase the sample size.

```
>      # for a population with known sigma=13.4, taking a sample
>      # of size 15, getting a sample of sample mean=8.74
>      # look at changes for different levels of confidence
> ci_known( 13.4, 15, 8.74, 0.80 )    # 80%
   CI Low   CI High       MOE Std Error
 4.306004 13.173996  4.433996  3.459865
> ci_known( 13.4, 15, 8.74, 0.90 )    # 90%
   CI Low   CI High       MOE Std Error
 3.049028 14.430972  5.690972  3.459865
> ci_known( 13.4, 15, 8.74, 0.98 )    # 98%
    CI Low    CI High       MOE  Std Error
 0.6911501 16.7888499  8.0488499  3.4598651
> ci_known( 13.4, 15, 8.74, 0.999 )  # 99.9%
   CI Low   CI High       MOE Std Error
-2.644779 20.124779 11.384779  3.459865
```

```
>      # do the same thing but keep level at 0.90 and change
>      # the sample size
> ci_known( 13.4, 25, 8.74, 0.90 )    # n=25
   CI Low   CI High       MOE Std Error
 4.331792 13.148208  4.408208  2.680000
> ci_known( 13.4, 45, 8.74, 0.90 )    # n=45
   CI Low   CI High       MOE Std Error
 5.454316 12.025684  3.285684  1.997554
> ci_known( 13.4, 85, 8.74, 0.90 )    # n=85
   CI Low   CI High       MOE Std Error
 6.349314 11.130686  2.390686  1.453434
```