

The Sampling Distribution of the Mean

January 9, 2021

Contents

- The Central Limit Theorem
- The sampling distribution of the mean of IQ scores
- Example 1
- Example 2
- Example 3
- Questions

Happy birthday to Jasmine Nichole Morales!

This tutorial should be easy to understand if you understand the z-table tutorial and the normal distribution tutorial.

In science, we often want to estimate the mean of a population. But all we can typically do is sample members of the population and calculate sample means. How well does a sample mean represent the population mean?

The mean is an **unbiased** statistic, which means that on average a sample mean will be equal to the population mean. Of course, any given sample mean will typically be different from the population mean, but since it's unbiased we can be sure that it won't on average be higher or lower than the population mean.

How close is a typical sample mean to the population mean? You probably have the intuition that this answer depends on the size of the sample. The larger the sample, the more confident you can be that your sample mean is close to the population mean.

The Central Limit Theorem

The Central Limit Theorem is a formal description of this intuition. It's a theorem that tells you about the distribution of sample means.

Let's take a moment to think about the term "distribution of sample means". Every time you draw a sample from a population, the mean of that sample will be different. Some means will be more likely than other means. So it makes sense to think about means having their own distribution, which we call the **sampling distribution of the mean**. The Central Limit Theorem tells us how the shape of the sampling distribution of the mean relates to the distribution of the population that these means are drawn from.

To define some terms, if samples from a population are labeled with the variable X , we define the parameters of mean as μ_x and the standard deviation as σ_x . Remember, the greek letter is the parameter, and the subscript is the name of the thing that we're talking about.

Now consider the sampling distribution of the mean. You know that sample means are written as \bar{x} . Using the same notation, the sampling distribution of the mean has its own mean, called $\mu_{\bar{x}}$, and its own standard deviation, called $\sigma_{\bar{x}}$.

There are three parts to the Central Limit Theorem:

1) The sampling distribution of the mean will have the same mean as the population mean. Formally, we state: $\mu_{\bar{x}} = \mu_x$.

This just means what I said earlier, that the mean is unbiased, so that sample means will be, on average, equal to the population mean.

2) For a sample size n , the standard deviation of the sampling distribution of the mean will be $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$.

The name for $\sigma_{\bar{x}}$ is sometimes shortened to the **standard error of the mean**, and sometimes shortened even more to 's.e.m.'.

This is a formalization of the intuition above. Since \sqrt{n} is in the denominator, it means that as your sample size gets bigger, the standard deviation of the distribution of means, $\sigma_{\bar{x}}$, gets smaller. So as you increase sample size, any given sample mean will be on average closer to the population mean.

3) The sampling distribution of the mean will tend to be close to normally distributed. Moreover, the sampling distribution of the mean will tend towards normality as (a) the population tends toward normality, and/or (b) the sample size increases.

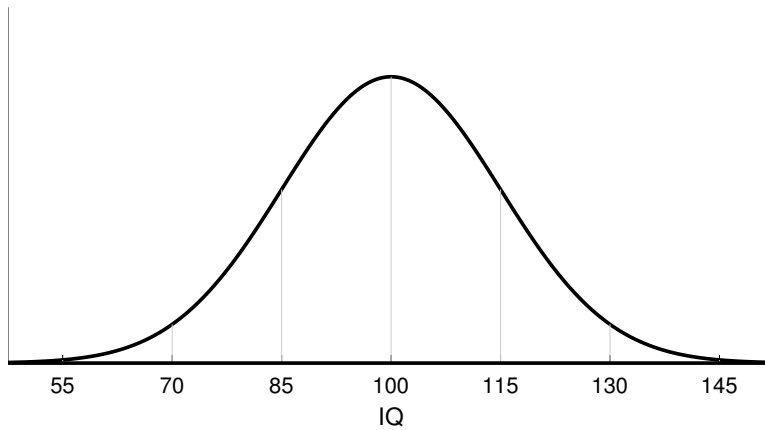
This last part is the most remarkable. It means that even if the population is not normally distributed, the sampling distribution of the mean will be roughly normal if your sample size is large enough.

The Central Limit Theorem is powerful because, as we've learned from previous chapters, if you know that a distribution is normal, and you know its mean and standard deviation, then you know everything about this distribution.

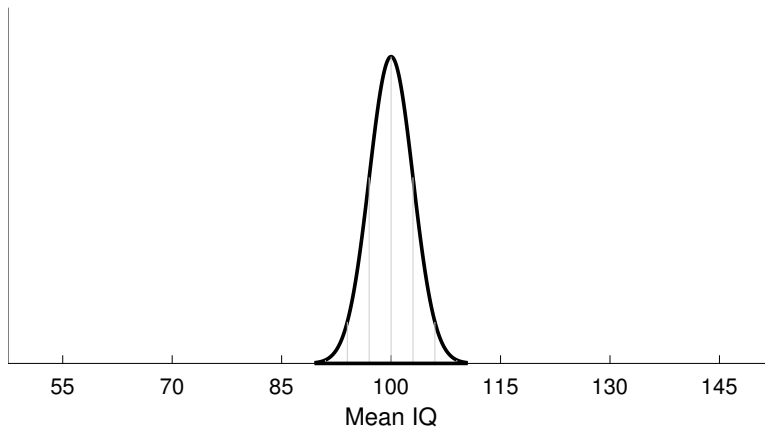
This is why we've been doing all this work converting scores to z-scores and looking things up in Table A. In science, we typically grab samples and calculate means to get an estimate of the population mean. With the Central Limit Theorem, we can now say something about how close a given mean should be to the population mean.

The sampling distribution of the mean of IQ scores

Here is the distribution of IQ scores which are normally distributed with a mean μ_x of 100 and a standard deviation σ_x of 15:



Suppose you repeatedly drew 25 samples from this distribution and calculated the mean each time. According to the Central Limit Theorem, the means will be distributed normally too, and will have a mean $\mu_{\bar{x}}$ equal to the population mean $\mu_x = 100$ and a standard deviation of $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{15}{\sqrt{25}} = 3$.



Notice that the sampling distribution of the mean is normal, and notice also how tight it is. It's much less likely to get a mean IQ of, say 115, than it is for an individual to have this IQ.

Think about what the sampling distribution of the mean will look like if we had a larger or smaller sample size. What about a sample size of 1?

Now we can cash in on all the work we've put into calculating areas under normal distributions. Because of the Central Limit Theorem, we can estimate how close our sample mean should be to the population mean, in terms of probabilities.

Example 1

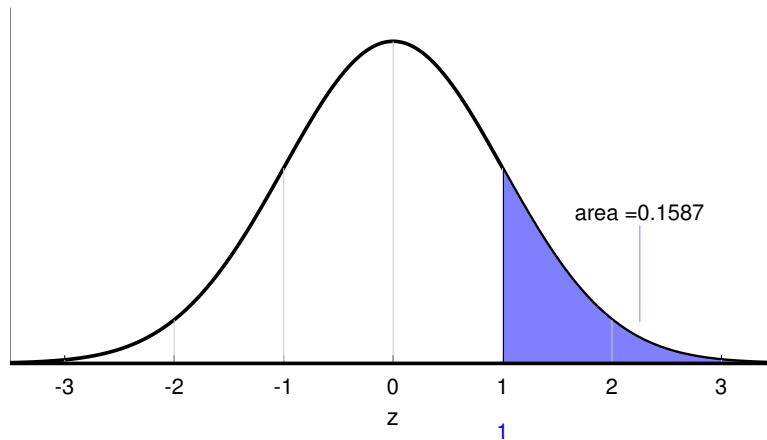
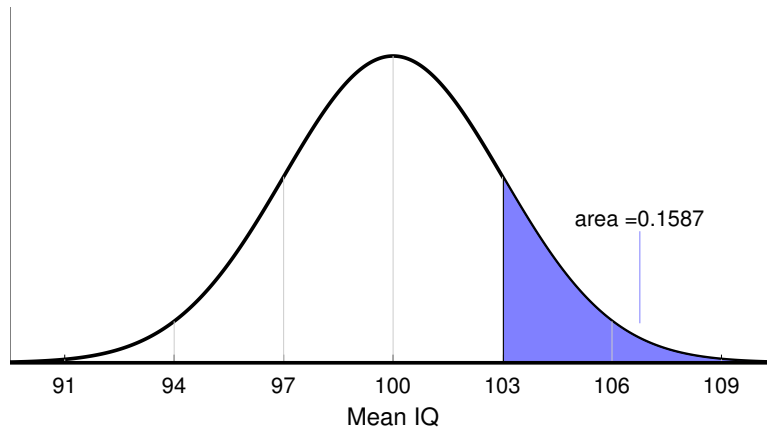
What is the probability that a mean drawn from a sample of 25 IQ scores will exceed 103

points?

All we need to do is convert our score of 103 to a z score and use Table A. This is just like the problems that we did earlier, but we're now talking about \bar{x} instead of X .

Remember to convert to a z score, we first subtract the mean and then divide by the standard deviation. But now, the mean is $\mu_{\bar{x}}$ and the standard deviation is $\sigma_{\bar{x}}$.

$$z = \frac{X - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{X - \mu_x}{\frac{\sigma_x}{\sqrt{n}}} = \frac{103 - 100}{3} = 1$$

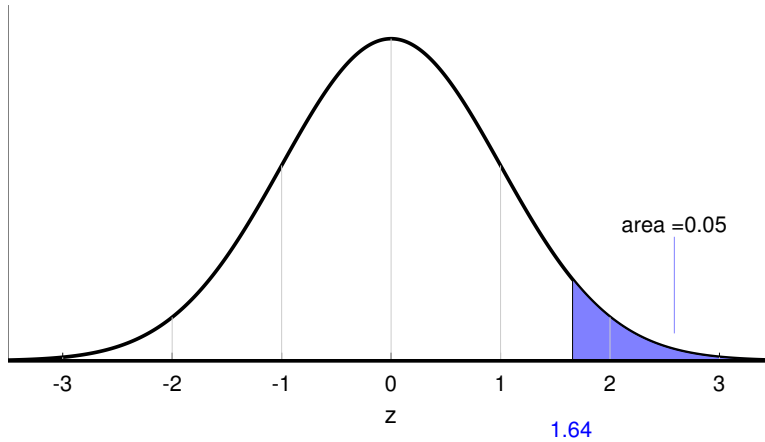


From table A, the area under the standard normal distribution above $z = 1$ is 0.1587. So the probability of obtaining a mean of 103 or more is 0.1587.

Example 2

Suppose you measure the heights of 100 randomly sampled women from a population that has a standard deviation of 64 inches and a standard deviation of 2.5 inches. For what mean height will 5% of the heights fall above?

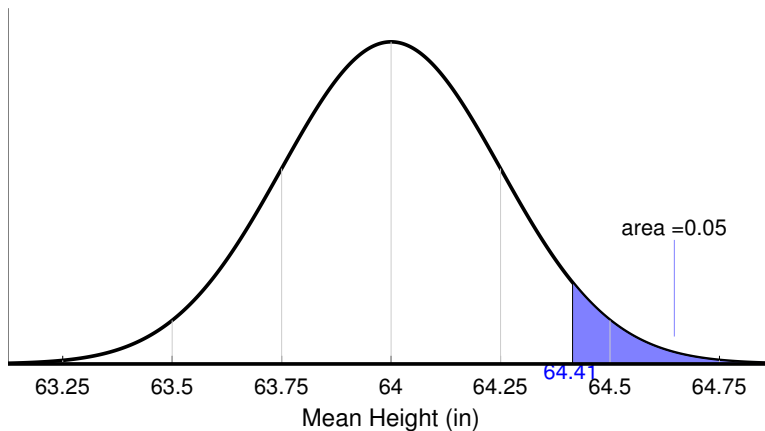
We need to first find the z-score for which 5% of the area under the standard normal distribution lies above. Using table A, $z = 1.64$.



To convert this to the distribution of mean heights, we use the Central Limit Theorem. With a sample size of 100, we can assume that mean heights will be normally distributed with a mean of 64 and a standard deviation of $\sigma_{\bar{x}} = \frac{2.5}{\sqrt{100}} = 0.25$.

So our desired mean height will be:

$$\bar{x} = \mu_{\bar{x}} + z\sigma_{\bar{x}} = \mu_x + z\frac{\sigma_x}{\sqrt{n}} = 64 + (1.64)(0.25) = 64.41$$

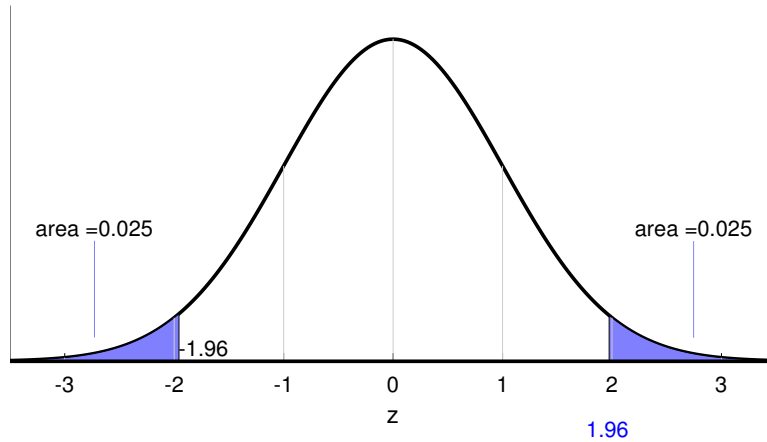


So we can conclude that 5% of the mean heights will fall above a mean of 64.41 inches.

Example 3

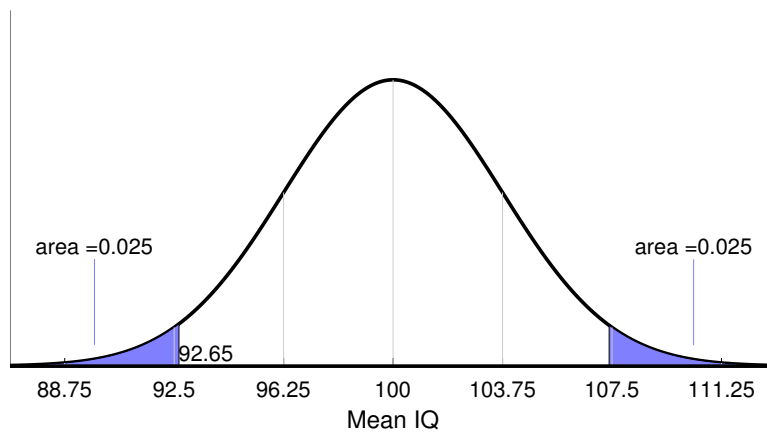
When drawing 16 samples from a distribution of IQ's, what is the range of means that will cover the middle 95% of sample means?

This time we work the other way around. We first find the z-scores for which the tails of the standard normal contain $\frac{100-95}{2} = 2.5\%$ of the population.



From table A, this corresponds to z-scores of plus or minus 1.96.

We then convert these z-scores to mean IQ scores, knowing that the standard error of the mean is $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{15}{\sqrt{16}} = 3.75$.



The range of mean IQ scores will therefore be between:

$$\mu_x - z\sigma_{\bar{x}} = 100 - (1.96)(3.75) = 92.65$$

and

$$\mu_x + z\sigma_{\bar{x}} = 100 + (1.96)(3.75) = 107.35$$

In other words, for a sample size of 16, we can expect to obtain a mean IQ between 92.65 and 107.35 points 95% of the time.

Questions

Your turn again. Here are 10 questions and answers, including R commands. See how similar these problems are to those in the normal distribution tutorial. The main difference is that since talking about distributions of means, we're using standard errors of the mean instead of standard deviations.

1) Suppose you draw 23 samples a population with mean of 32 and a standard deviation of 2.5.

What is the probability of obtaining a mean of 33 or less?

Given $\mu_x = 32$ and $\sigma_x = 2.5$, Find $Pr(\bar{x} < 33)$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{2.5}{\sqrt{23}} = 0.5213$$

$$z = \frac{\bar{x} - \mu_x}{\sigma_{\bar{x}}} = \frac{33 - 32}{0.5213} = 1.9183$$

$$Pr(z < 1.9183) = 0.9725$$

Answer: $p = 0.9725$

```
sem <- 2.5/sqrt(23)
print(sem)
[1] 0.521286
p <- pnorm(33,32,sem)
print(p)
[1] 0.9724656
```

2) Suppose you draw 21 samples a population with mean of 78 and a standard deviation of 8.8.

What is the probability of obtaining a mean of 78.2 or more?

Given $\mu_x = 78$ and $\sigma_x = 8.8$, Find $Pr(\bar{x} > 78.2)$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{8.8}{\sqrt{21}} = 1.9203$$

$$z = \frac{\bar{x} - \mu_x}{\sigma_{\bar{x}}} = \frac{78.2 - 78}{1.9203} = 0.1042$$

$$Pr(z > 0.1042) = 0.4585$$

Answer: $p = 0.4585$

```
sem <- 8.8/sqrt(21)
print(sem)
[1] 1.920317
p <- 1-pnorm(78.2,78,sem)
print(p)
```


[1] 0.4585254

3) Suppose you draw 43 samples a population with mean of 103 and a standard deviation of 5.8.

For what mean do 7.08 percent of the sample means fall below?

Given $\mu_x = 103$ and $\sigma_x = 5.8$, find \bar{x} so that $Pr(\bar{x} < \mu_x) = 0.0708$

$Pr(z < -1.4698) = 0.0708$, so $z = -1.4698$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{5.8}{\sqrt{43}} = 0.8845$$

$$\bar{x} = \mu_x + (z)(\sigma_{\bar{x}}) = 103 + (-1.4698)(0.8845) = 101.7$$

Answer: $\bar{x} = 101.7$

```
sem <- 5.8/sqrt(43)
print(sem)
[1] 0.8844917
x <- qnorm(0.0708, 103, sem)
print(x)
[1] 101.6999
```

4) Suppose you draw 46 samples a population with mean of 62 and a standard deviation of 10.3.

For what mean do 44.76 percent of the sample means fall above?

Given $\mu_x = 62$ and $\sigma_x = 10.3$, find \bar{x} so that $Pr(\bar{x} > \mu_x) = 0.5524$

$Pr(z > 0.1317) = 0.4476$, so $z = 0.1317$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{10.3}{\sqrt{46}} = 1.5187$$

$$\bar{x} = \mu_x + (z)(\sigma_{\bar{x}}) = 62 + (0.1317)(1.5187) = 62.2$$

Answer: $\bar{x} = 62.2$

```
sem <- 10.3/sqrt(46)
print(sem)
[1] 1.518652
x <- qnorm(0.5524, 62, sem)
print(x)
[1] 62.20005
```

5) Suppose you draw 26 samples a population with mean of 13 and a standard deviation of 2.5.

What range of means covers the middle 65.83 percent of means?

Given $\mu_x = 13$, $\sigma_x = 2.5$, and $p = 0.6583$

The area in each tail is: $\frac{1-0.6583}{2} = 0.17085$

$Pr(z > 0.9508) = 0.17085$, so $z = -0.9508$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{2.5}{\sqrt{26}} = 0.4903$$

$$\bar{x}_1 = \mu_x - (z)(\sigma_{\bar{x}}) = 13 - (0.9508)(0.4903) = 12.53$$

$$\bar{x}_2 = \mu_x + (z)(\sigma_{\bar{x}}) = 13 + (0.9508)(0.4903) = 13.47$$

Answer: 65.83 percent of the means fall between 12.53 and 13.47.

```
sem <- 2.5/sqrt(26)
print(sem)
[1] 0.4902903
p <- (1-0.6583)/2
print(p)
[1] 0.17085
x <- c(qnorm(p,13,sem),qnorm(1-p,13,sem))
print(x)
[1] 12.53383 13.46617
```

6) Suppose you draw 24 samples a population with mean of 63 and a standard deviation of 8.7.

What is the probability of obtaining a mean of 58.8 or less?

Given $\mu_x = 63$ and $\sigma_x = 8.7$, Find $Pr(\bar{x} < 58.8)$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{8.7}{\sqrt{24}} = 1.7759$$

$$z = \frac{\bar{x} - \mu_x}{\sigma_{\bar{x}}} = \frac{58.8 - 63}{1.7759} = -2.365$$

$$Pr(z < -2.365) = 0.009$$

Answer: $p = 0.009$

```
sem <- 8.7/sqrt(24)
print(sem)
[1] 1.77588
p <- pnorm(58.8,63,sem)
print(p)
```

[1] 0.009014437

7) Suppose you draw 23 samples a population with mean of 11 and a standard deviation of 5.8.

What is the probability of obtaining a mean of 13.4 or more?

Given $\mu_x = 11$ and $\sigma_x = 5.8$, Find $Pr(\bar{x} > 13.4)$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{5.8}{\sqrt{23}} = 1.2094$$

$$z = \frac{\bar{x} - \mu_x}{\sigma_{\bar{x}}} = \frac{13.4 - 11}{1.2094} = 1.9845$$

$$Pr(z > 1.9845) = 0.0236$$

Answer: $p = 0.0236$

```
sem <- 5.8/sqrt(23)
print(sem)
```

```
[1] 1.209384
```

```
p <- 1-pnorm(13.4,11,sem)
```

```
print(p)
```

```
[1] 0.02360107
```

8) Suppose you draw 21 samples a population with mean of 17 and a standard deviation of 6.2.

For what mean do 81.24 percent of the sample means fall below?

Given $\mu_x = 17$ and $\sigma_x = 6.2$, find \bar{x} so that $Pr(\bar{x} < \mu_x) = 0.8124$

$$Pr(z < 0.8869) = 0.8124, \text{ so } z = 0.8869$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{6.2}{\sqrt{21}} = 1.353$$

$$\bar{x} = \mu_x + (z)(\sigma_{\bar{x}}) = 17 + (0.8869)(1.353) = 18.2$$

Answer: $\bar{x} = 18.2$

```
sem <- 6.2/sqrt(21)
print(sem)
```

```
[1] 1.352951
```

```
x <- qnorm(0.8124,17,sem)
```

```
print(x)
```

```
[1] 18.19976
```

9) Suppose you draw 35 samples a population with mean of 79 and a standard deviation of 7.9.

For what mean do 61.78 percent of the sample means fall above?

Given $\mu_x = 79$ and $\sigma_x = 7.9$, find \bar{x} so that $Pr(\bar{x} > \mu_x) = 0.3822$

$Pr(z > -0.2996) = 0.6178$, so $z = -0.2996$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{7.9}{\sqrt{35}} = 1.3353$$

$$\bar{x} = \mu_x + (z)(\sigma_{\bar{x}}) = 79 + (-0.2996)(1.3353) = 78.6$$

Answer: $\bar{x} = 78.6$

```
sem <- 7.9/sqrt(35)
print(sem)
[1] 1.335344
x <- qnorm(0.3822, 79, sem)
print(x)
[1] 78.59979
```

10) Suppose you draw 28 samples a population with mean of 103 and a standard deviation of 10.5.

What range of means covers the middle 1.33 percent of means?

Given $\mu_x = 103$, $\sigma_x = 10.5$, and $p = 0.0133$

The area in each tail is: $\frac{1-0.0133}{2} = 0.49335$

$Pr(z > 0.0167) = 0.49335$, so $z = -0.0167$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{10.5}{\sqrt{28}} = 1.9843$$

$$\bar{x}_1 = \mu_x - (z)(\sigma_{\bar{x}}) = 103 - (0.0167)(1.9843) = 102.97$$

$$\bar{x}_2 = \mu_x + (z)(\sigma_{\bar{x}}) = 103 + (0.0167)(1.9843) = 103.03$$

Answer: 1.33 percent of the means fall between 102.97 and 103.03.

```
sem <- 10.5/sqrt(28)
print(sem)
[1] 1.984313
p <- (1-0.0133)/2
print(p)
[1] 0.49335
x <- c(qnorm(p, 103, sem), qnorm(1-p, 103, sem))
print(x)
```

[1] 102.9669 103.0331

