

6.3: Confidence Intervals for Means (Sigma Known)

Learning Objectives

- Motivate the use of the z -score transformation to determine margin of error
- Define and compute critical values
- Determine the margin of error
- Construct confidence intervals, interpret their meaning, and apply them to contextual questions
- Utilize sample size as a means to balance confidence and margin of error

Confidence Intervals: A Quick Review

When we select a random sample and study it, we do not expect that the computed sample statistic is equal to the population parameter. The distance between the sample statistic and the population parameter is called the error. We want an idea of how far off our sample statistic might be from the population parameter and provide an interval of possible parameter values using the information from our sample. Through our knowledge of sampling distributions, we can provide a level of confidence that we have caught the population parameter in our interval. If the confidence level is 80%, the construction method successfully catches the population parameter for 80% of all the samples of that given size. In other words, if we repeatedly sampled the population randomly with the same sample size, we would expect 80% of the samples to produce confidence intervals with the population parameter in them. To maintain our level of confidence, we determine the distance (the margin of error), such that the percentage of samples have sample statistics that fall within that distance from the population parameter. If we then center our confidence interval at our computed sample statistic and extend our interval out by our margin of error in both directions, we produce a confidence interval that catches the population parameter with a success rate that is equal to our confidence level. We now dive into the details.

Confidence Intervals for Means

Let us frame our task within the particular context of this section: constructing confidence intervals for the population mean. We are constructing a confidence interval using information collected from a random sample of size n from our population. The form of our confidence interval will be $(\bar{x} - \text{ME}, \bar{x} + \text{ME})$, where \bar{x} is the computed sample mean from the random sample of size n and ME is the margin of error. We must select a level of confidence for the confidence interval. This is the percentage of samples of size n that we want to be within the margin of error of the population mean. We need to determine ME, the margin of error, using the sampling distribution of sample means, which is normal, or at least approximately normal, under certain conditions. Those conditions must be met. If you cannot remember the conditions, you can review the section on the [sampling distribution of sample means](#) and commit the conditions to memory. Examine the figure below for a visual representation. Consider which of the symbols below will have a fixed, known value in an actual research situation.

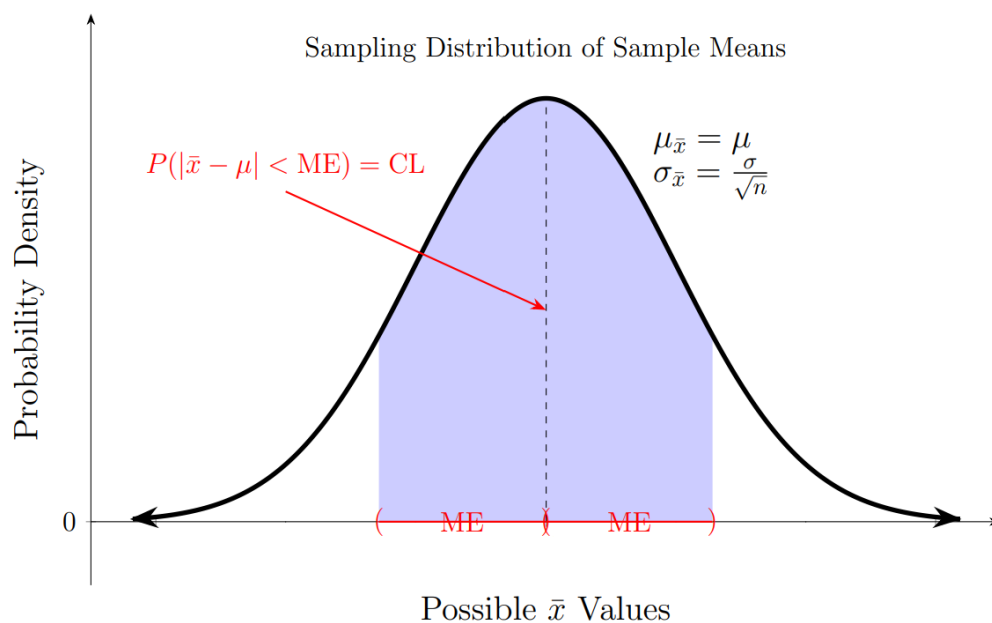


Figure 6.3.1: Sampling distribution of sample means

When considering the sampling distribution of sample means, \bar{x} is a variable with no fixed value for us to know while determining the margin of error. Once we collect a sample and compute its sample mean, we will have a value of \bar{x} . It is important to remember that the logic of computing the margin of error requires us to treat \bar{x} as a variable. We determine CL and n , so they are known to us. In general, we do not know anything about the population; that is why we are studying it. So, μ and σ are generally unknown as well. As such, it seems like our unknown symbols outnumber our known symbols. That is okay; we will be able to manage.

At this stage, we make one assumption for the sake of pedagogy. Let us assume that we know the value of σ , the population standard deviation. This is a rather large assumption because, as we all know, the population mean is an integral part of the computation of the population standard deviation. How could we know the population standard deviation without knowing the population mean? Perhaps in some situations, a past known population standard deviation may make a sufficient approximation for a current population standard deviation, but making such a claim is highly context dependent and beyond the scope of this book. For now, know we are making a simplifying assumption so that we can better understand the notion and construction of confidence intervals.

Determining the Margin of Error (σ known)

Recall that every normal distribution can be transformed into the standard normal distribution using the z -score transformation which preserves the area bounded beneath the probability density curves. We will use this transformation to determine what our margin of error needs to be. Consider the following sequence of figures that illustrates the transformation process.

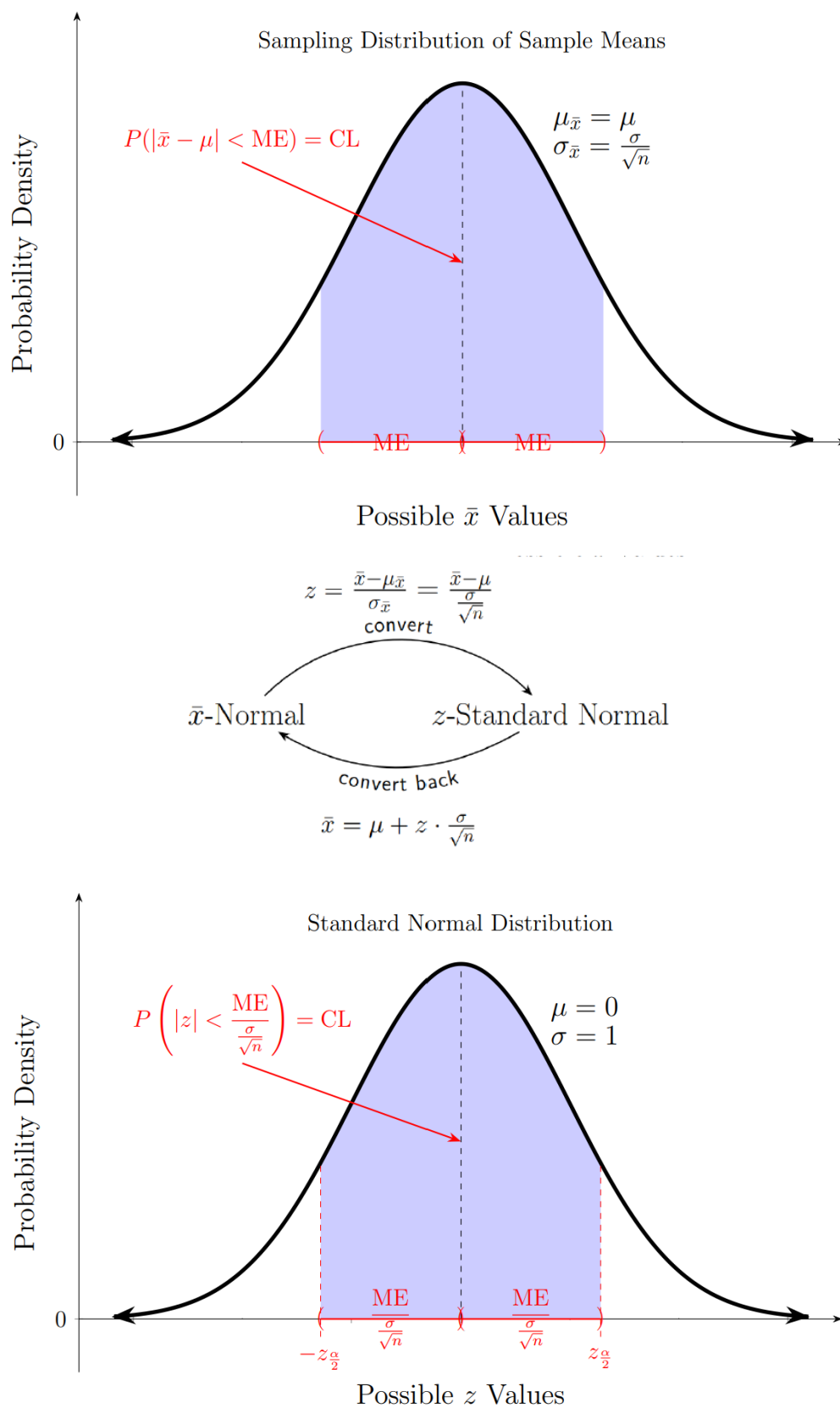


Figure 6.3.2: Sampling distribution of sample means transformed into the standard normal distribution

We started with the same figure as before, underwent the z -score transformation, and now want to determine the margin of error necessary to get the desired level of confidence. Notice that under the z -score transformation, the margin of error is scaled by a factor of one over the standard deviation of the sampling distribution. Since we are now considering the standard normal distribution, we know the mean and standard deviation of the distribution. Therefore, using technology, we can find the boundary points that will produce the desired confidence. These are denoted as $\pm z_{\frac{\alpha}{2}}$ in the figure above. We call these points **critical values**. Note that $z_{\frac{\alpha}{2}}$ represents the z -value where the area under the standard normal distribution to the right is $\frac{\alpha}{2}$ and $(-z_{\frac{\alpha}{2}})$ represents the z -value where the area under the standard normal distribution to the left is $\frac{\alpha}{2}$.

? Text Exercise 6.3.1

Remaining in the context of constructing confidence intervals for population means when σ is known, determine the critical values for the indicated level of confidence by first sketching the problem in a standard normal distribution and then using technology to compute the critical values.

1. Confidence level: 90%

Answer

We first sketch a standard normal curve and then form an interval that is centered at the mean 0 and label the boundary points. The area under the curve between these two points is our confidence level. Notice that the critical values are equal in magnitude but opposite in sign so we can find one value and then take the positive and negative values as our critical values. Since we are using technology, we need to find the area to the left of one of the points. There are several ways of achieving this goal. We illustrate a different way for each of the first three problems of this text exercise; though, all three methods work for each of the problems.

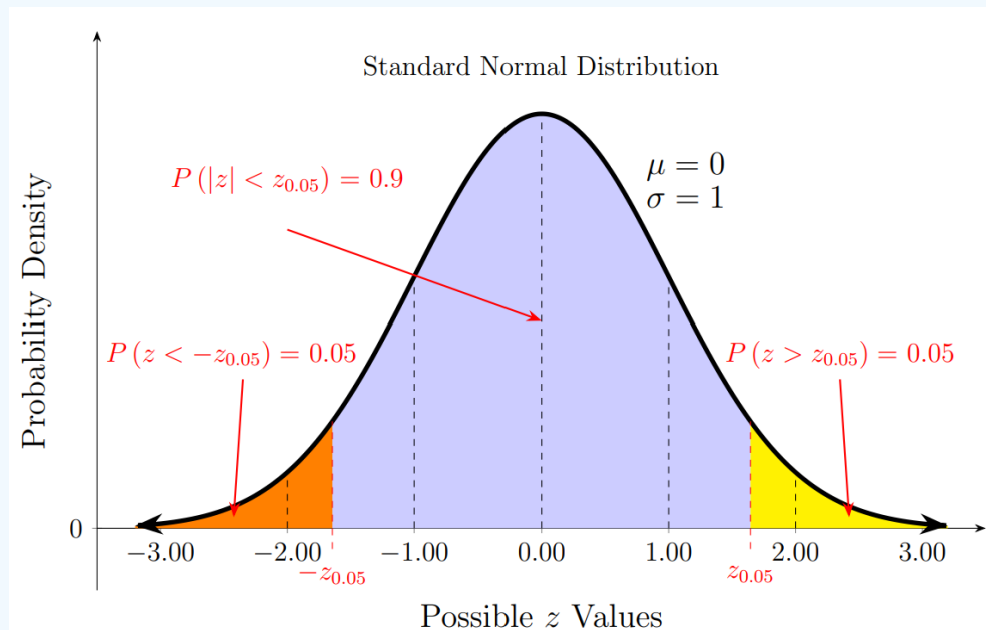


Figure 6.3.3 Standard normal distribution with 90% confidence interval

For our first example, we find the negative critical value first. We can determine the area outside of our critical values because the total area underneath the curve is 1, and the area between our critical values is 0.9. The area outside of our critical values is $1 - 0.9 = 0.1$. Note that this is what we have been calling the α value. Since normal distributions are symmetric about the mean and the critical values are equally far from the mean, the two tails of the distribution (the values less than the negative critical value and then the values greater than the positive critical value) have the same area. To find the area to the left of the negative critical value, we split the area of the two tails in half. $\frac{0.1}{2} = 0.05$. Notice the labels for the critical values; we replaced the $\frac{\alpha}{2}$ in the subscript with the value of $\frac{\alpha}{2}$ in the context of the problem. We can use technology to determine the left critical value.

$$-z_{0.05} = \text{NORM.S.INV}(0.05) \approx -1.6449$$

We thus have our critical values: $\pm z_{0.05} \approx \pm 1.6449$

2. Confidence level: 95%

Answer

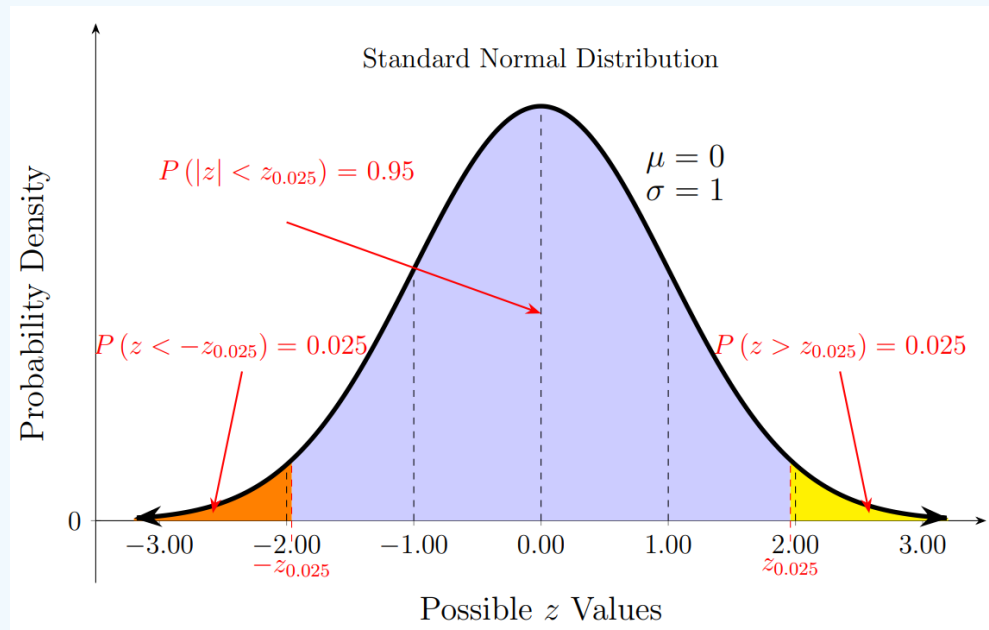


Figure 6.3.4 Standard normal distribution with 95% confidence interval

We find the positive critical value for our second and third examples. Note that the area to the left of the positive critical value is the confidence level 0.95 and the area in the left tail, which we know from the last exercise is half of the α value $\frac{0.05}{2} = 0.025$. The area to the left of the positive critical value is $0.95 + 0.025 = 0.975$. We can use technology to determine the right critical value.

$$z_{0.025} = \text{NORM.S.INV}(0.975) \approx 1.96$$

Now we have our critical values: $\pm z_{0.025} \approx \pm 1.96$

3. Confidence level: 99%

Answer

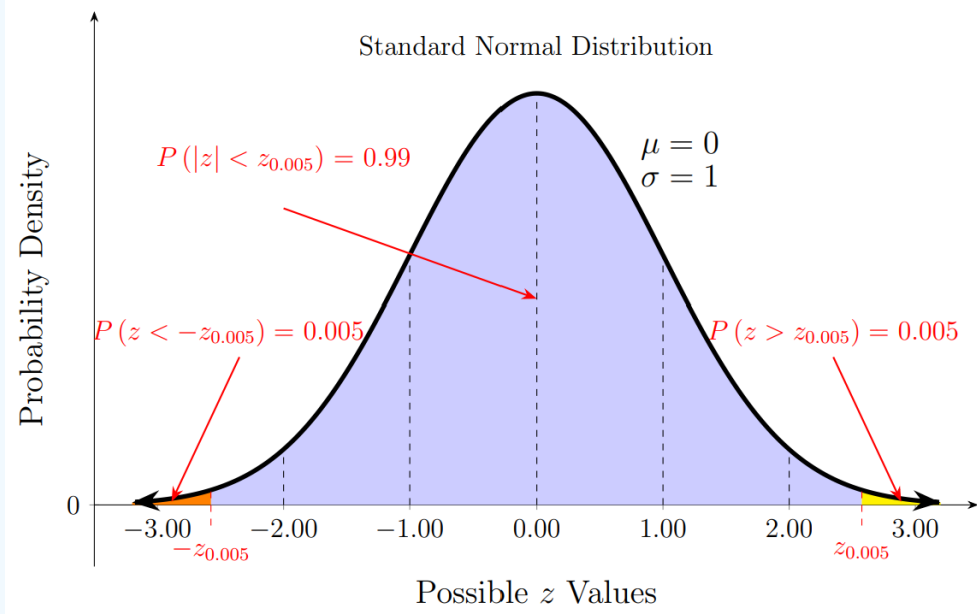


Figure 6.3.5 Standard normal distribution with 99% confidence interval

Another way to find the area to the left of the positive critical value is to use the complementary relationship between the area to the left and the area to the right of a point. The total area is 1. The area to the right of the positive critical value is half of the α value $\frac{0.01}{2} = 0.005$. So the area to the left of the positive critical value is $1 - 0.005 = 0.995$. We can use technology to determine the right critical value.

$$z_{0.005} = \text{NORM.S.INV}(0.995) \approx 2.5758$$

Now have our critical values: $\pm z_{0.005} \approx \pm 2.5758$

4. For a general confidence level CL

Answer

We generally care about the positive critical value as it represents the number of standard deviations that must be traversed in both directions from the mean to gain the desired confidence level. In this solution, we will use each of the three methods above to find the right critical value for the general confidence level CL.

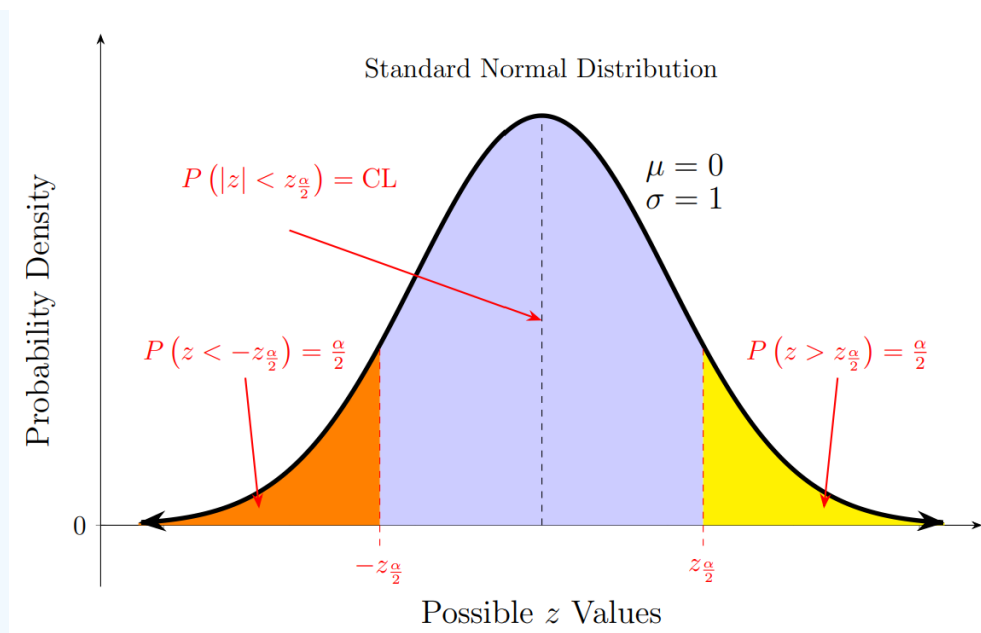


Figure 6.3.6 Standard normal distribution with general confidence interval

$$\begin{aligned} z_{\frac{\alpha}{2}} &= -1 \cdot \text{NORM.INV}\left(\frac{\alpha}{2}\right) \\ &= \text{NORM.INV}\left(\text{CL} + \frac{\alpha}{2}\right) \\ &= \text{NORM.INV}\left(1 - \frac{\alpha}{2}\right) \end{aligned}$$

With the critical values in hand, we make the final step by noticing that the positive critical value is equal to the length of the scaled margin of error.

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

Given our simplifying assumption (that we know the population standard deviation), we know the factor by which the margin of error was scaled and what the scaled length is. From this, we determine the margin of error.

$$\text{ME} = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Constructing Confidence Intervals for Means (σ known)

We now have all the pieces to construct a confidence interval for the population mean when the population standard deviation is known.

$$(\bar{x} - \text{ME}, \bar{x} + \text{ME}) = \left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

We often write these confidence intervals as $\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$.

? Text Exercise 6.3.2

The 2024 Toyota Camry Hybrid LE gets 52¹ miles per gallon when considering both highway and city driving with a standard deviation of 5.1 miles per gallon. In designing the 2025 Toyota Camry, the engineers would like to assert that the fuel efficiency in the newest model exceeds that of the previous model. The engineers randomly test-drove 50 2025 models and

recorded an average of 54 miles per gallon from the sample. Assuming the standard deviation remained the same, construct a 98% confidence interval to predict the population mean fuel economy. Does this bode well for the engineer's desires? Explain.

¹This is the only statistic based on actual, substantial data. The remainder of the numbers in this problem were contrived loosely based on available data.

Answer

First, we check that the conditions for constructing a confidence interval for means are satisfied. We want our sample to be randomly selected and the sampling distribution to be approximately normal. Since the engineers randomly test drove 50 cars, we have both conditions met ($n > 30$).

We next connect the values in the problem statement with the variables at play: $CL = 98\%$, $n = 50$, $\bar{x} = 54$, and $\sigma = 5.1$. Sometimes, we do not use all the numbers in a problem statement. 52 comes into play at the end, not while constructing the confidence interval, because the engineers want the population mean of the 2025 Camry to be greater than the previous model, which was 52 miles per gallon.

We need to find the positive critical value $z_{\frac{\alpha}{2}}$. Since $CL = 98\% = 0.98$, $\alpha = 1 - 0.98 = 0.02$, and $\frac{\alpha}{2} = \frac{0.02}{2} = 0.01$. The distribution that we find the critical value from is the standard normal distribution because we are considering means with the population standard deviation known. We encourage sketching pictures.

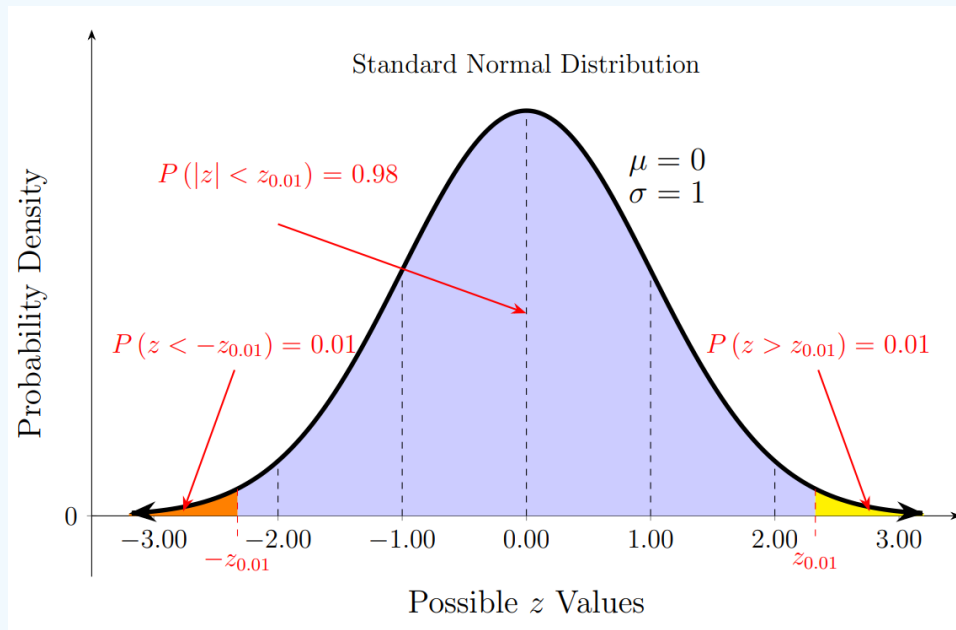


Figure 6.3.7 Standard normal distribution with 98% confidence interval

$$z_{0.01} = -1 \cdot \text{NORM.S.INV}(0.01) \approx 2.3264$$

$$\left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right) \approx \left(54 - 2.3264 \cdot \frac{5.1}{\sqrt{50}}, 54 + 2.3264 \cdot \frac{5.1}{\sqrt{50}} \right) \approx (52.3221, 55.6779)$$

So, we have constructed the confidence interval based on the results of the random sample of 50 cars. Our conclusion is that, at a 98% confidence level, the population mean, μ , of the 2025 Camry Hybrid LE fuel economy is somewhere between 52.3221 miles per gallon and 55.6779 miles per gallon. We then notice that 52.3221 miles per gallon is greater than 52 miles per gallon. The engineers can feel confident that the fuel economy of the newest model exceeds the fuel economy of the previous model.

The Margin of Error and Sample Size

At the beginning of this section, we mentioned a balancing act at play in constructing confidence intervals. As we saw, the higher the confidence level, the larger the positive critical value. The larger the critical value, the larger the margin of error. At the same

time, we want our confidence interval to give us a pretty good idea of the population mean. The larger the margin of error, the wider the range of values we conclude our population mean falls. For example, we can be 100% confident that the population parameter falls in the interval $(-\infty, \infty)$, but that interval does not yield any useful information; similarly, we could very precisely estimate that $\mu = \bar{x}$, but we would have 0% confidence in this estimate. These desires conflict, but there is another variable at play in determining the margin of error: the sample size, n . We do not have control over σ , the population standard deviation, but we do have control over the size of the sample we select.

$$ME = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

? Text Exercise 6.3.3

1. Explain what happens to the margin of error as the sample size n increases.

Answer

Since $ME = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$, we have that n is in the denominator. As n increases, \sqrt{n} also increases. When dividing by larger and larger numbers, the resulting number is smaller and smaller. Consider $\frac{1}{1} = 1$, $\frac{1}{10} = 0.1$, $\frac{1}{100} = 0.01$, and so on. As we increase the denominator by a factor of 10 each time, the value decreases getting closer and closer to 0. Thus, the margin of error goes to 0 as the sample size gets larger. This should match our intuition that larger samples are more likely to produce statistics close to the population parameter.

2. If the margin of error for a 95% confidence interval for means with σ known was 4 with a sample size of 35, how large of a sample must be taken to have a margin of error of 1 while maintaining the same level of confidence?

Answer

Since both confidence intervals are being constructed at the same level of confidence and from the same population, $z_{\frac{\alpha}{2}}$ and σ will be the same. We will have two margins of error and two sample sizes. $ME_1 = 4$, $n_1 = 35$, $ME_2 = 1$, and n_2 . The last one is unknown. This yields the following system of equations.

$$4 = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{35}}$$

$$1 = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n_2}}$$

We are only interested in finding n_2 . So, we want to eliminate the critical value and standard deviation. We note that if we multiply the second equation by 4 on both sides, we can set the two equations equal to each other.

$$z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{35}} = 4 \cdot z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n_2}}$$

$$\cancel{z_{\frac{\alpha}{2}}} \cdot \frac{\cancel{\sigma}}{\sqrt{35}} = 4 \cdot \cancel{z_{\frac{\alpha}{2}}} \cdot \frac{\cancel{\sigma}}{\sqrt{n_2}}$$

$$\frac{1}{\sqrt{35}} = \frac{4}{\sqrt{n_2}}$$

$$\sqrt{n_2} = 4\sqrt{35}$$

$$(\sqrt{n_2})^2 = (4\sqrt{35})^2$$

$$n_2 = 16 \cdot 35 = 560$$

? Text Exercise 6.3.4

1. Suppose the engineers at Toyota decided that they wanted a confidence interval with a margin of error of 0.25 miles per gallon while maintaining the confidence level of 98%, how large of a sample of 2025 Toyota Camry LE cars would need to be taken?

Answer

Recall that the population standard deviation was given to be 5.1 miles per gallon and that the positive critical value was $z_{\frac{\alpha}{2}} = -1 \cdot \text{NORM.S.INV}(0.01) \approx 2.3264$. The engineers have set the desired margin of error to $\text{ME} = 0.25$. Given the fact that $\text{ME} = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$, we can solve for the unknown sample size.

$$\begin{aligned} 0.25 &\approx 2.3264 \cdot \frac{5.1}{\sqrt{n}} \\ \sqrt{n} &\approx \frac{2.3264 \cdot 5.1}{0.25} \\ (\sqrt{n})^2 &\approx \left(\frac{2.3264 \cdot 5.1}{0.25} \right)^2 \\ n &\approx 2252.214 \end{aligned}$$

We now must remember the context behind the situation. We are trying to determine the minimum sample size necessary to result in a 98 confidence interval with a margin of error of 0.25 miles per gallon, and we have deduced that n must be at least 2252.214. We, therefore, decide that a sample of size 2253 cars would be necessary.

- Let us now solve the problem in general. If we set the margin of error, confidence level, and know the population standard deviation, how large of a sample is necessary to construct a confidence interval at that level of confidence and margin of error?

Answer

$$\begin{aligned} \text{ME} &= z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \\ \sqrt{n} &= \frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\text{ME}} \\ (\sqrt{n})^2 &= \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\text{ME}} \right)^2 \\ n &= \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\text{ME}} \right)^2 \end{aligned}$$

Now, just as before, we need our sample size to be a whole number. When it is not a whole number, we always round up so that we are within the threshold of our margin of error tolerance. It is better to more precise than less precise.

Formula for the mathematically inclined

$$n = \left\lceil \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\text{ME}} \right)^2 \right\rceil$$

This formula introduces the ceiling function $\lceil x \rceil$ which returns the smallest integer value that is greater than or equal to x .

6.3: Confidence Intervals for Means (Sigma Known) is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by The Math Department at Fort Hays State University.

- 10.7: Confidence Interval for Mean by David Lane is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.