

4.6: Accumulation Functions And Area Measures in Normal Distributions

Learning Objectives

- Define an accumulation function for continuous probability distributions
- Use an accumulation function for the standard normal (z -)distribution to find area measures of regions in the standard normal distribution
- Use the inverse of an accumulation function for the standard normal (z -)distribution to find the location for the specified region's area measures
- Standardize non-standard normal distributions to find area measures and scale locations
- Use an accumulation function for general normal distributions to find area measures of regions in those distributions
- Use the inverse of an accumulation function for general normal distributions to find scale location for specified region's area measures
- Use spreadsheet functions of NORM.S.DIST, NORM.S.INV, NORM.DIST, and NORM.INV appropriately for finding needed values in normal distributions

Review and Preview

We have discussed the relationship between the area of regions within a continuous random variable's probability distribution and the probability of occurrence in relation to that variable. We also examined several families of distributions. Lastly, we noted how any region of interest in these distributions could always be related to left-regions. We now focus on how to produce these left-region area measures on normal distributions using technology. Once we reasonably master these concepts in relation to normal distributions, similar ideas are used in t -distributions and χ^2 -distributions, as well as many other specialized distributions.

Accumulation Functions of Area

We have discussed the importance of the determining the area of regions within probability distributions since the probability of selecting an outcome in the region formed by an interval is equal to its area. It became more difficult to determine the area if the regions of interest were not basic geometric shapes. Specifically, if our regions were rectangular, we could easily compute the area of such regions.

In general, the area of a region in our distributions can be sliced up into thin slices. We can approximate the area of each of these thin slices by summing the areas of rectangles that are close in height to the thin slices. This is illustrated in Figure 4.6.1 below.

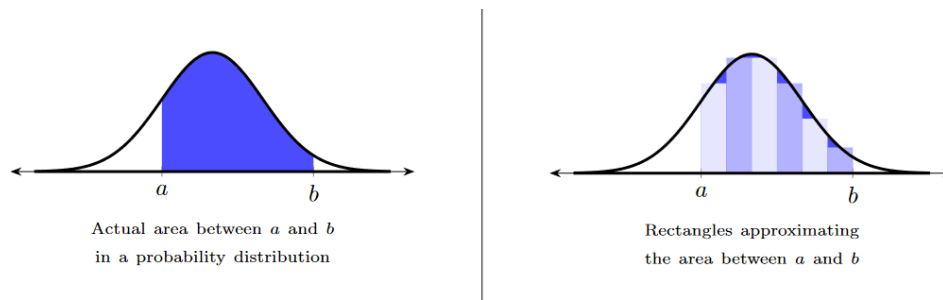


Figure 4.6.1: Probability distribution region being approximated by thin rectangles

Although it is messy work for humans to complete this process with only ten or twenty slices, computing technology efficiently calculates with the use of hundreds or even thousands of thinner and thinner slices on the region. As the number of slices for a fixed region gets larger, the approximating rectangles get thinner, and the approximation from the sum of areas on the thin rectangles gets closer to the actual area of the original region. At one time, large tables of area values were produced to list the approximation sums. Now this process has been programmed for several probability distributions, producing specialized **accumulation functions (also called cumulative distribution functions)** that provide highly accurate approximations to the area of regions in most common probability distributions. Different statistical software might name and program their accumulation functions differently; we will focus on the accumulation functions within spreadsheets in our following work. We start with an accumulation function for finding accurate left-tail regions in the standard normal probability distribution.

Area Measures for the Standard Normal Distribution

We begin with an **accumulation function for the standard normal distribution**. The name and syntax of this function can vary depending on the technology being used. The name of the accumulation function in Excel is NORM.S.DIST. We note that the spreadsheet function name corresponds to the distribution we are discussing. This function requires we provide it with a specific z -score value; the function will then return the area of the region to the left of that z -value when choosing the TRUE option to accumulate left area.

Due to the symmetry of the standard normal curve, we know that $P(z < 0) = 0.5000 = 50\%$ as shown in Figure 4.6.2. If we enter `=NORM.S.DIST(0, TRUE)` as a function in a spreadsheet cell, the spreadsheet returns the left-area measure of 0.5000 with appropriate cell formatting as shown in the spreadsheet image.

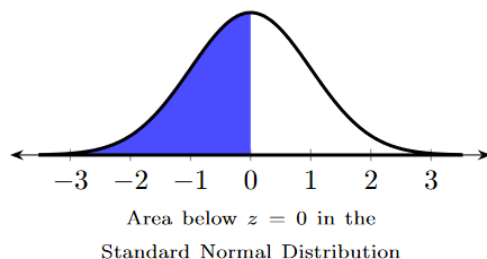


Figure 4.6.2: Standard normal distribution with shaded area

As a shortcut, we can enter the digit 1 instead of typing out the word TRUE when using this accumulation function. If we use FALSE, the function returns only the height measure of the density function but not an area measure. We will almost always want to use this function for area accumulation, but we must remember the function only returns left-tail area measures. If we find any other region, we must adjust our computation work as discussed in Section 4.5. In general, the syntax of this accumulation function is `=NORM.S.DIST(z -score, TRUE)` or the slightly shorter version of `=NORM.S.DIST(z -score, 1)`.

Suppose we wish to find $P(z \leq -2)$, another left-tail region as shown by our graph of the standard normal distribution below. We found an area measure approximated at $0.0250 = 2.5\%$ in Section 2.7 through the use of the Empirical (68 – 95 – 99.5) Rule.

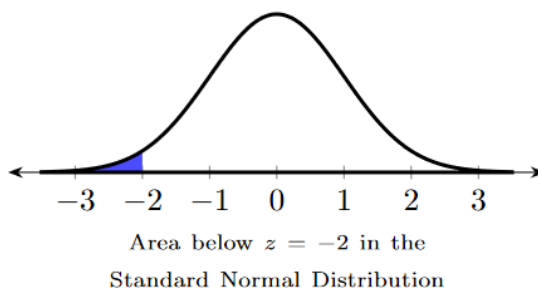


Figure 4.6.3: Standard normal distribution with shaded area

We can also use our spreadsheet accumulation function to find this area:

$$\begin{aligned} P(z \leq -2) &= \text{NORM.S.DIST}(-2, \text{TRUE}) \\ &= \text{NORM.S.DIST}(-2, 1) \\ &\approx 0.02275 = 2.275\%. \end{aligned}$$

For a random variable that possesses the standard normal distribution, we consider it unusual (since $2.275\% \leq 5.000\%$) to have a z -score that is at most -2 . We note that this value of 2.275% from our technology's accumulation function is more accurate than the estimate from the Empirical Rule. Generally, we will use our technology to generate more accurate measures instead of the less accurate values computed by the Empirical Rule.

Suppose we want to find $P(z > 1.25)$, a right-tail region in the standard normal distribution as shown below. We must turn this into a left-tail region calculation to use our accumulation function.

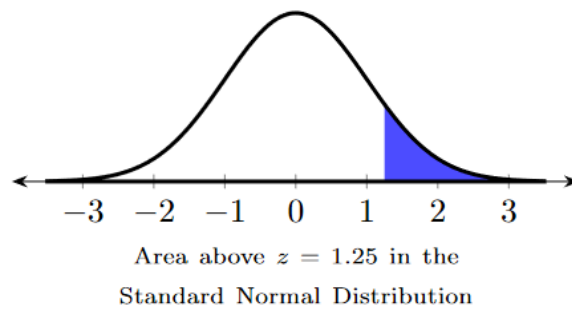


Figure 4.6.4 Standard normal distribution with shaded area

We should recognize the need to use the complement to relate to left-tail regions, producing:

$$\begin{aligned} P(z > 1.25) &= 1 - P(z \leq 1.25) \\ &= 1 - \text{NORM.S.DIST}(1.25, 1) \\ &\approx 1 - 0.89435 \\ &= 0.10565 = 10.565\%. \end{aligned}$$

Since it can be easy to make entry errors when using our spreadsheet functions, we compare our value to the region shaded in the graph. The shaded region does seem to be a small portion of the entire distribution, and our resulting value of 10.565% appears aligned to our graph. Until we have a strong mastery of the ideas, we will sketch the graph of the region of interest to aid us in proper accumulation function use and to verify the reasonableness of our computed area. Once mastery of these ideas is achieved, we encourage mental visualization of the graph of the distribution and showing work with the accumulation function.

? Text Exercise 4.6.1

Sketch graphs and determine the solutions of the following probability problems.

1. Find $P(z \geq -2)$.

Answer

After producing our graph representing $P(z \geq -2)$ (shown below), we notice we are working in a right-tailed region, but recall that the NORM.S.DIST function is for producing left-tailed area measures only.

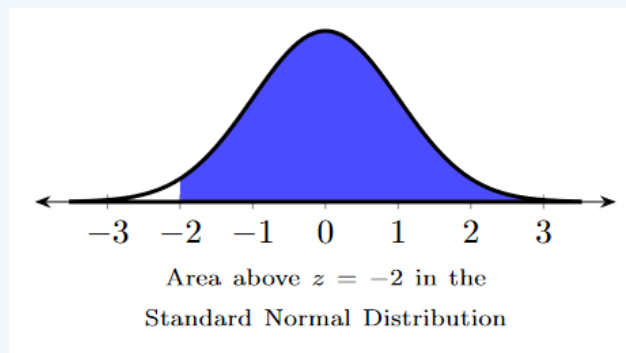


Figure 4.6.5 Standard normal distribution with shaded area

We must adjust the computation with the complement to find the right-tailed area in this situation.

$$\begin{aligned}
 P(z \geq -2) &= 1 - P(z < -2) \\
 &= 1 - \text{NORM.S.DIST}(-2, 1) \\
 &\approx 1 - 0.0228 \\
 &= 0.9772 = 97.72\%
 \end{aligned}$$

Remembering our quick check, we notice that the size of the shaded region in the graph seems to align with this proportional measure of 97.72%. Thus, 97.72% of the standard normal distribution's area is to the right of -2 . Or equivalently, there is a 97.72% probability of randomly selecting a z -score outcome that is at least -2 in value.

2. Find $P(-2 \leq z \leq 1)$.

Answer

After producing our graph (shown below), we notice we are working in a "between" region. In this situation, we again must make a computational adjustment with a difference calculation.

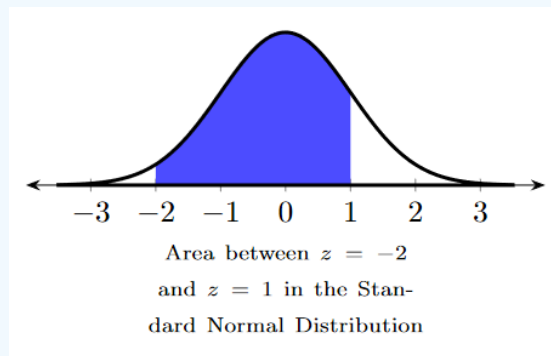


Figure 4.6.6 Standard normal distribution with shaded area

We subtract two left-tail areas to find the desired region's area measure.

$$\begin{aligned}
 P(-2 \leq z \leq 1) &= P(z \leq 1) - P(z < -2) \\
 &= \text{NORM.S.DIST}(1, 1) - \text{NORM.S.DIST}(-2, 1) \\
 &\approx 0.8413 - 0.0228 \\
 &= 0.8185 = 81.85\%
 \end{aligned}$$

There is an 81.85% probability of randomly selecting a z -score that is between -2 and 1 in value.

3. Find $P(z \leq -1.25)$.

Answer

After producing our graph (shown below), we notice we are working in a left-tail region. No computational adjustment is needed to use our accumulation function in this situation.

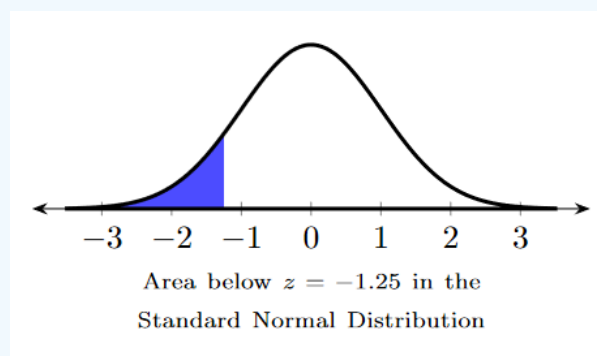


Figure 4.6.7 Standard normal distribution with shaded area

To find the area, we compute it with our spreadsheet function.

$$P(z \leq -1.25) = \text{NORM.S.DIST}(-1.25, 1) \\ \approx 0.1056 = 10.56\%$$

After a quick check, we can confidently say that 10.56% of the standard normal distribution's area is to the left of -1.25 . We also note that such an outcome in the random variable is not unusual since the probability measure is more than 5%.

4. Find the proportion of the standard normal curve between -1.5 and 0.5 .

Answer

Based on our graph (shown below) of the given information, we notice we are again working in a "between" region. We must make the computational adjustment using another difference calculation.

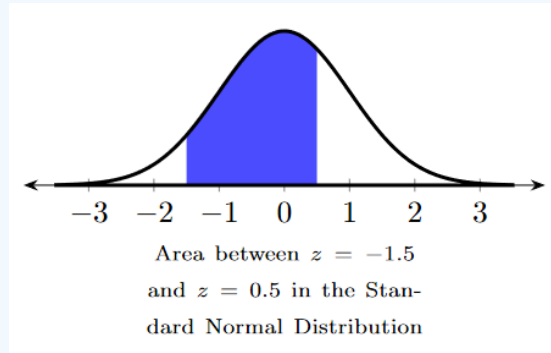


Figure 4.6.8 Standard normal distribution with shaded area

To find the area, we subtract the two left-tail areas.

$$P(-1.5 < z < 0.5) = P(z < 0.5) - P(z \leq -1.5) \\ = \text{NORM.S.DIST}(0.5, 1) - \text{NORM.S.DIST}(-1.5, 1) \\ \approx 0.6915 - 0.0668 \\ = 0.6247 = 62.47\%$$

There is a 62.47% probability of randomly selecting a z -score outcome that is between -1.5 and 0.5 in value. Notice that this is the proportion of outcomes in this interval.

Working with decimal or fractional valued z -scores requires no adjustment in our thinking or work. The "messiness" of the numbers involved or produced should not impact our established reasoning or computational work.

Now, there will also be occasions in which we need to reverse the process above; that is, given the description of a region and its area measure, what is/are the z -score(s) that produce that region? For example, suppose we wish to know the one z -score that separates the lower 5% region of the standard normal distribution from the upper 95% region. Stated another way: what is the 5th percentile of the standard normal distribution? This is illustrated in Figure 4.6.9 below:

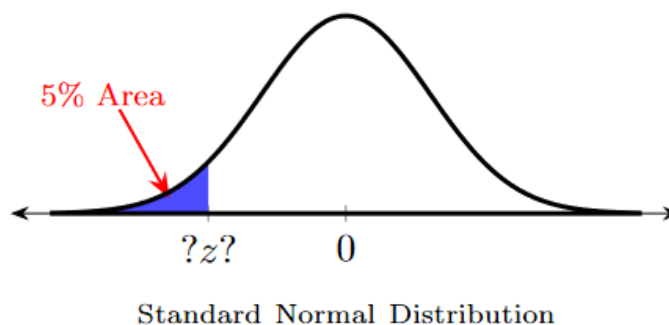


Figure 4.6.9: A region of the standard normal distribution with unknown z -score

With experiences from above, one might just guess-and-test reasonable z -score values with our NORM.S.DIST function to find an approximate value, a , for which $P(z < a) = 5\%$. This approach may take some time. Instead, we are blessed with the mathematics of inverse functions. We have a spreadsheet function called NORM.S.INV. Given any left-tail region's area, this function will compute the associated right boundary z -score that forms that region, an inverse process to the accumulation function. This function has the syntax = NORM.S.INV(left-tail area measure between 0 and 1). Since our left-tail area is $5\% = 0.05$, we compute in our spreadsheet to produce these results.

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

So a z -score of approximately -1.6449 separates the lower 5% area in the standard normal distribution from the upper 95% area. These z -scores are often called critical z -scores as they are critical boundary values for specific area measures in the standard normal distribution. We now attempt similar text exercises.

? Text Exercise 4.6.2

After sketching the regions described, find z -score(s) that produce the area described in the standard normal distribution.

1. Find the z -score associated with the 65^{th} percentile of the standard normal distribution.

Answer

After producing a sketch indicating the concept of 65^{th} percentile in a standard normal distribution graph (shown below), we notice we are working with a left-tailed region.

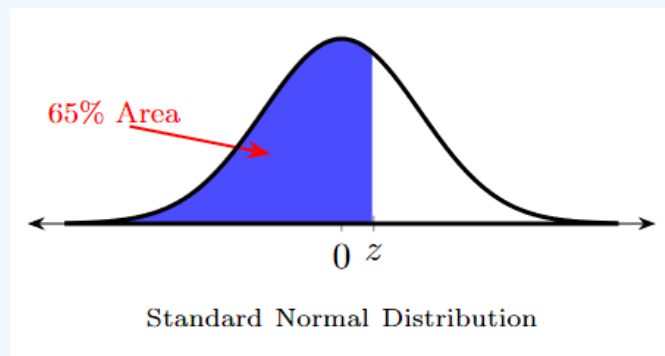


Figure 4.6.10 Standard normal distribution with shaded area

To find the boundary z -score value associated with this left-tailed area measure, we can go directly to our inverse accumulation function with no adjustment.

$$z = \text{NORM.S.INV}(0.65) \\ \approx 0.3853$$

Thus, 65% of the standard normal distribution's area is to the left of a z -score of 0.3853 . Equivalently, there is a 65% probability of randomly selecting a z -score outcome that is less than 0.3853 in value. We also note that such decimal values for z -scores will occur more frequently in our computation results. We must note the type of value we are computing/measuring and not depend on what the value looks like to control our interpretation. Remember that probabilities are numbers between 0 and 1; whereas, z -scores can be any real number, including the numbers between 0 and 1.

2. Find the z -score so that 10% of the standard normal distribution is above that z value.

Answer

After producing a sketch indicating the described region in the exercise (shown below), we notice we are working with a right-tailed region.

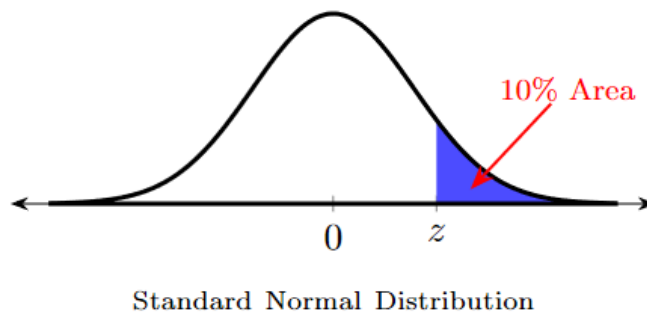


Figure 4.6.11: Standard normal distribution with shaded area

We must make complement adjustments in our work using our inverse accumulation function to find the boundary z -score value associated with this right-tailed area measure. We notice that the left-tailed (white) region under our curve must be $1 - 0.10 = 0.90 = 90\%$ of the total area based on our complement rule. We find our boundary z -score value by:

$$\begin{aligned} z &= \text{NORM.S.INV}(0.90) \\ &\approx 1.2816 \end{aligned}$$

Thus, 10% of the standard normal distribution's area is to the right of a z -score of 1.2816. Or equivalently, there is a 10% probability of randomly selecting a z -outcome that is at least 1.2816 in value. We should be careful with the syntax here. Notice that entering $1 - \text{NORM.S.INV}(0.1)$ does not produce the correct answer.

3. Find the value of a so that $P(z \leq a) = 3\%$.

Answer

After producing a sketch indicating the 3% left-tailed region in a standard normal distribution graph (shown below), we go directly to our inverse accumulation function to compute the related z -score labeled with a .

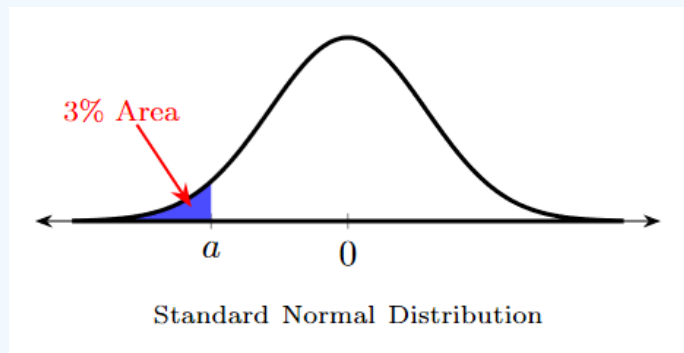


Figure 4.6.12: Standard normal distribution with shaded area

$$\begin{aligned} a &= \text{NORM.S.INV}(0.03) \\ &\approx -1.8808 \end{aligned}$$

Thus, 3% of the standard normal distribution's area is to the left of a z -score of -1.8808 . Equivalently, there is a 3% probability of randomly selecting a z -score that is most -1.8808 in value; stated symbolically, $P(z \leq -1.8808) = 3\%$. Note that we would label this as an unusual outcome.

4. Find a for which $P(-a < z < a) = 80\%$; that is find the z -scores that capture the central 80% of the standard normal distribution.

Answer

We must adjust our computation work for left-tailed regions. After producing a sketch indicating the central 80% region in a standard normal distribution graph (shown below).

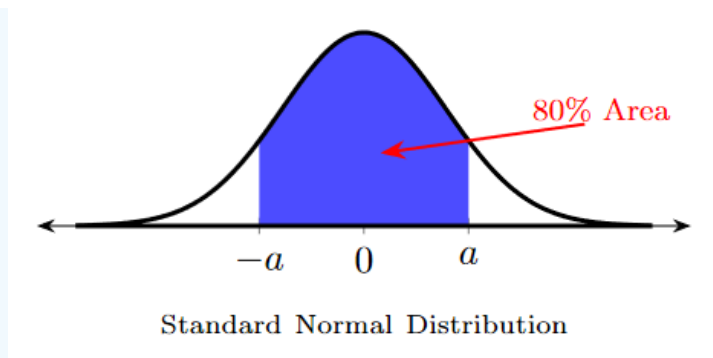


Figure 4.6.13 Standard normal distribution with shaded area

With 80% in the central region, that leaves $100\% - 80\% = 20\%$ area measure for the two tails. This means that there is an area of $\frac{20\%}{2} = 10\%$ in each tail since the standard normal distribution is symmetric. To find the left-boundary z -score value (labeled as $-a$ in the diagram), we use our inverse accumulation function with the 10% left-region area measure.

$$\begin{aligned} z &= \text{NORM.S.INV}(0.10) \\ &\approx -1.2816 \end{aligned}$$

To find the right-boundary z -score value (labeled as a in the diagram), we can use the symmetry of the standard normal curve about the central value 0 to reason that $a = 1.2816$. For extra practice, we can also compute with our inverse accumulation function with a $80\% + 10\% = 90\%$ left-region area measure.

$$\begin{aligned} z &= \text{NORM.S.INV}(0.90) \\ &\approx 1.2816 \end{aligned}$$

Thus, 80% of the standard normal distribution's area is between z -scores of -1.2816 and 1.2816 .

We have now found ways to use a technology accumulation function and its inverse to produce various area and scale measures of the standard normal distribution. However, many normal distributions are not the standard normal distribution. We now examine the same ideas for any normal distribution.

Area Measures in Non-standard Normal Distributions

We discuss two methods for finding probabilities as well as the inverse action when working with normal distributions that are not standard normal (the mean is not zero and/or the standard deviation is not one). We will be using both methods on this concept in the remainder of our text, so it is important for us to learn the methods well now before adding more concepts.

Conversion to Standard Normal

As reviewed in [Section 4.5](#), we can convert any normally distributed random variable, x , into the scale of the standard normal variable, z , using our standardization calculation: $z = \frac{x - \mu}{\sigma}$. This implies that we can compute any needed areas and z -values for any normal distribution by using this conversion process first and then applying the concepts from the standard normal distribution.

For example, suppose that the time for various college students to complete a specific task is normally distributed with $\mu = 25$ minutes and $\sigma = 5$ minutes, and we want to know what proportion of the students spent less than 15 minutes to complete the task. We recall that a normal probability distribution is determined by its mean and standard deviation values, allowing us to quickly sketch the distribution, including reasonably accurate scaling of our horizontal axis. Based on the provided information, we graph this non-standard normal distribution in Figure 4.6.14, along with its standardization into the z -distribution.

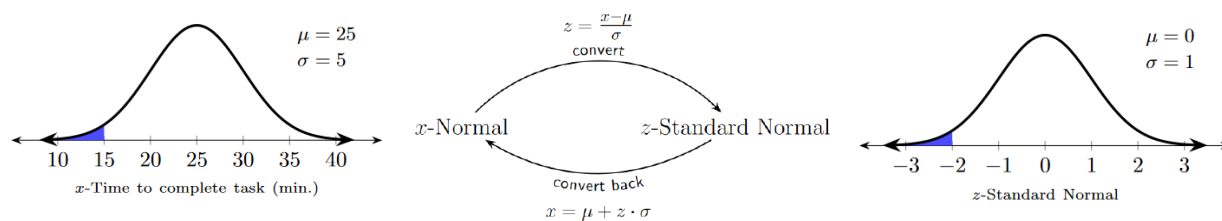


Figure 4.6.14 Standardization of a normal distribution

Often, to save space and sketching time, we do not write the vertical scale on our normal distributions as seen above (we do note that the vertical scaling is different between the two distributions since the horizontal scaling is different, but for our current purposes knowing the difference is not essential). Often, we will place the natural/raw and standardized scaling in the same distribution sketch, as shown in Figure 4.6.15 below.

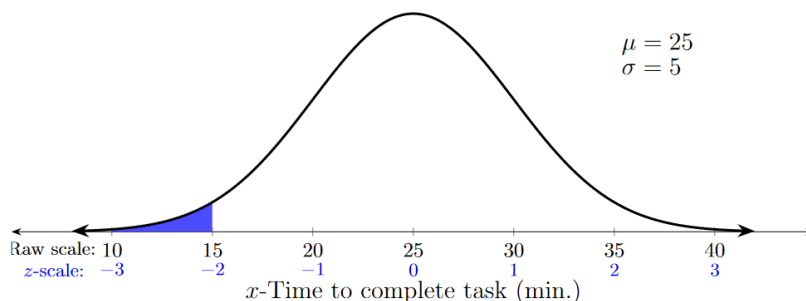


Figure 4.6.15 Standardized scaling on a non-standard normal distribution

Since we are seeking $P(x < 15 \text{ min.})$, we standardize 15 min. to $z = \frac{15-25}{5} = -2$. Intuitively, this means that 15 minutes is 2 standard deviations below the mean. As can be seen in either Figure 4.6.14 or 4.6.15, the area of the region to the left of 15 in the specified normal distribution is the same as the area to the left of -2 in the standard normal distribution; that is, $P(x < 15 \text{ min.}) = P(z < -2)$. Since we can compute $P(z < -2)$ by $= \text{NORM.S.DIST}(-2, 1)$ in our spreadsheet, producing the value 0.0228 we know that $P(x < 15 \text{ min.}) = 0.0228 = 2.28\%$. That is, about 2.28% of those college students spent less than 15 minutes to complete the task.

The key idea here is that we can convert back and forth between any given normal distribution scale and the standard normal distribution scale to handle probability questions related to the non-standard normal distribution. We illustrate one more example below with an "inverse" function problem where we "convert back" to find our needed measures.

Suppose, in the same random variable context, we want to know the time interval the central 80% of those students took to complete the task. As shown in Figure 4.6.16 below, we need to find the scale values, labeled as a and b , that captures the central 80% of the distribution's entire region. Notice that, even though we technically have the same horizontal scale values known to us, we have very little of the x - or z -axes scaled in our sketch as compared to our Figures 4.6.14 and 4.6.15. This is because, initially, we are unsure where these boundary values for the 80% region are exactly located until computed in our later work. Again, a reasonable sense of the figure is essential for answering this "inverse" question.

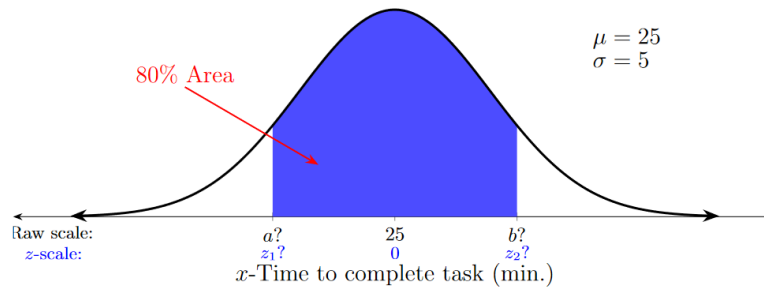


Figure 4.6.16 Inverse conversion to find scale value

We did such "inverse" or "convert back" work above in the standard normal distribution with related left-area measures producing the two results of

$$z_1 = \text{NORM.S.INV}(0.10) \approx -1.2816$$

$$z_2 = \text{NORM.S.INV}(0.90) \approx 1.2816.$$

If we reverse the conversion process, taking these z -scores back to the related x -scores using inverse formula $x = \mu + z \cdot \sigma$, we can produce the related raw scale values of a and b :

$$a = \mu + z_1 \cdot \sigma \approx 25 + (-1.2816) \cdot (5) = 18.592$$

$$b = \mu + z_2 \cdot \sigma \approx 25 + (1.2816) \cdot (5) = 31.408.$$

Finally, thinking about the contextual interpretation of these results, we know that the central 80% of those college students (that is, a large majority of them) took between 18.6 minutes and 31.4 minutes to complete the task. Knowing such information might be useful in planning the time one should allot so that most can complete the task on time.

? Text Exercise 4.6.3

Sketch the distributions described and find the desired value(s).

1. If a random variable x has a normal distribution with $\mu = 18.2$ and $\sigma = 3.4$, find $P(x > 22)$, that is, find the proportion of this distribution that is above 22.

Answer

First, we sketch a diagram of the described normal probability distribution with the standardized scale on the distribution.

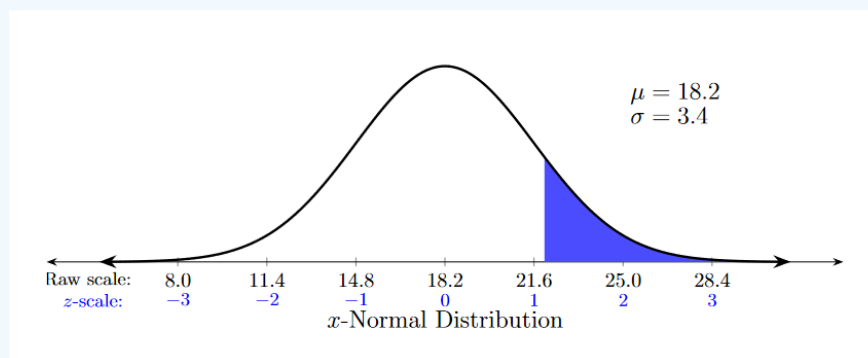


Figure 4.6.17 Normal distribution with shaded area

We note that this region is right-tailed, with a complement left-tailed region. To determine $P(x > 22)$, we must move from $x = 22$ to the related standardized value of $z = \frac{22-18.2}{3.4} \approx 1.1176$. Hence,

$$\begin{aligned}
 P(x > 22) &= P(z > 1.1176) \\
 &= 1 - P(z \leq 1.1176) \\
 &= 1 - \text{NORM.S.DIST}(1.1176, 1) \\
 &\approx 1 - 0.8681 \\
 &= 0.1319 = 13.19\%.
 \end{aligned}$$

That is, the proportion of this normal distribution with values above 22 is about 13.19%.

2. If a random variable x has a normal distribution with $\mu = 18.2$ and $\sigma = 3.4$, find the 25th percentile value for the distribution.

Answer

Sketching a new diagram of the given normal distribution and given conditions on that distribution, remembering that the 25th percentile is equivalent in meaning to the boundary value that separates the lower 25% of the distribution from the upper 75% :

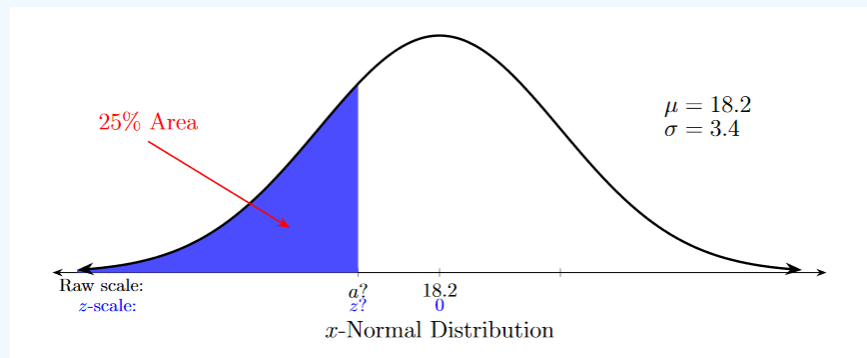


Figure 4.6.18 Normal distribution with shaded area

We seek our shaded region's boundary value a . We must determine the related z -score first through the use of our NORM.S.INV function and then convert that z -score back to our raw-scale score. We first compute our critical z -score by

$$\begin{aligned}
 z &= \text{NORM.S.INV}(0.25) \\
 &\approx -0.6745,
 \end{aligned}$$

then convert to raw value by

$$\begin{aligned}
 a &= \mu + z \cdot \sigma \\
 &\approx 18.2 + (-0.6745) \cdot 3.4 \\
 &= 15.9067.
 \end{aligned}$$

Hence, the 25th percentile value in this given normal distribution is approximately at the value $x = 15.91$.

3. A soft drink bottler has data that suggest that the amount of drink placed in their 12-ounce cans by a specific bottling machine is normally distributed with $\mu = 12.1$ ounces and $\sigma = 0.5$ ounces (the machine is slightly over-filling on average from designed specifications).
- What proportion of cans are under-filled from the labeled amount by more than 1 ounce?
 - What amount of soft drink in the cans accounts for the central 90% of all cans filled by this specific machine?

Answer

- Below is our sketch of the situation, noting we are involved with a left-tailed region:

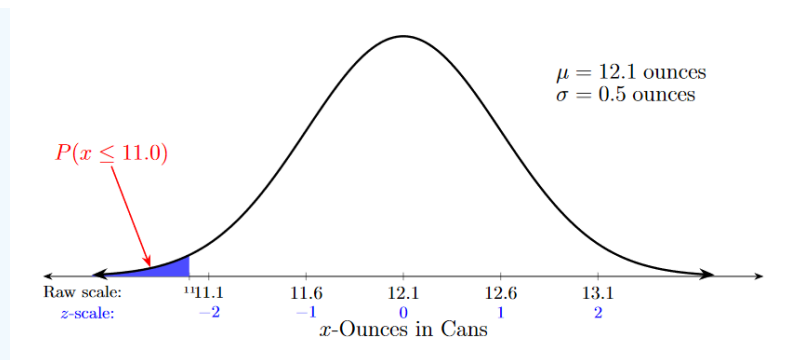


Figure 4.6.19 Normal distribution with shaded area

To find the proportion of under-filled cans by this machine from the labeled amount of 12 ounces by more than 1 ounce, we need to find $P(x < 11)$. First we convert to z -scale:

$$\begin{aligned} z &= \frac{x - \mu}{\sigma} \\ &= \frac{12.1 - 11}{0.5} \\ &= -2.20, \end{aligned}$$

then find our area measure in the standard normal distribution:

$$\begin{aligned} P(x < 11) &= P(z < -2.20) \\ &= \text{NORM.S.DIST}(-2.20, 1) \\ &\approx 0.0139 = 1.39\%. \end{aligned}$$

So about 1.39% of the cans are being under-filled by more than one ounce from the desired specifications. That value shows that under-filling by more than one ounce is unusual for this machine.

b. We sketch a diagram of the given information:

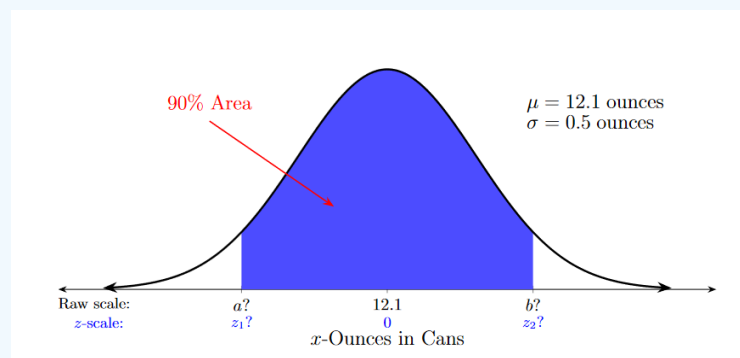


Figure 4.6.20 Normal distribution with shaded area

We seek the boundary values a and b in our raw scaled axis to capture the central 90% of the normal probability distribution. However, we again must first get the related z_1 - and z_2 -scores through the use of our NORM.S.INV function and then convert them back to our raw-scale score. So, after noticing we have 5% of the area in the white regions of the two tails in our distribution, we compute our two symmetrical critical z -scores by

$$\begin{aligned} \pm z &= \pm \text{NORM.S.INV}(0.05) \\ &\approx \pm 1.6449, \end{aligned}$$

then convert to raw scale by

$$\begin{aligned}
 a &= \mu + z_1 \cdot \sigma & \text{and} & & b &= \mu + z_2 \cdot \sigma \\
 &\approx 12.1 + (-1.6449) \cdot 0.5 & & & &\approx 12.1 + (1.6449) \cdot 0.5 \\
 &= 11.2776 & & & &= 12.9224.
 \end{aligned}$$

Hence, 90% of the cans being filled by this machine have between 11.28 and 12.92 ounces in them.

By working in the standard normal distribution with left-tail regions, we can determine the areas' related z -scale values. These z -scale values can then be "converted back" into the scaled values of the non-standard normal distribution. This back-and-forth conversion work between the x -scale and the z -scale can get tedious, but it is a beneficial strategy for working with normal distributions. We all likely need more practice by doing several homework problems to get reasonable mastery of these ideas. For some of our later work in inferential statistics, this type of conversion work and the meaning of this conversion action will be extremely important.

Nonetheless, in the following subsection, we explore our second method and two new spreadsheet functions that hide/automate this conversion process, allowing us to keep within the natural/raw scale of the given normal probability distribution.

Hidden/Automated Conversion to Standard Normal

We introduce an **accumulation function for any normal distribution**. The name and syntax of this function can vary depending on the technology one uses, but the name of the accumulation function in Excel is NORM.DIST. We note that the spreadsheet function name here only misses the ".S" required for the standard normal distribution function. This function requires we provide it with a specific x -scale value in the distribution as well as the mean μ and the standard deviation σ of the normal distribution. In general, the syntax of this accumulation function is **= NORM.DIST(x -score, μ , σ , TRUE)** or the slightly shorter version of **= NORM.DIST(x -score, μ , σ , 1)**. The function will return the area of the region to the left of that x -value if we choose the TRUE option. Similar to the standard normal distribution's function, we can enter the digit 1 instead of typing out the word TRUE when using this accumulation function. Also, similarly, if we use FALSE with the function, it returns only the height of the density function at that specific x -value, not an area.

Let us re-examine the example problems from the last section in which we converted back and forth between the normal distribution of interest and the standard normal distribution. Recall our given context in which the time for various college students to complete a specific task is normally distributed with $\mu = 25$ minutes and $\sigma = 5$ minutes. We again ask, what proportion of the students spent less than 15 minutes to complete the task? We graph this in Figure 4.6.21 based on this information. In our second method, we do not include scaling with the related standardized z -scores:

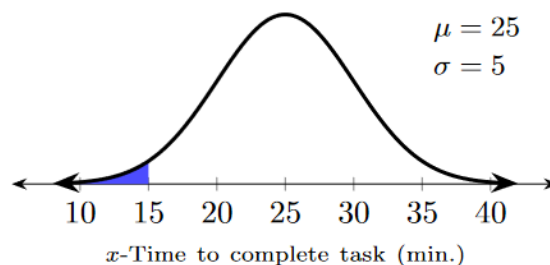


Figure 4.6.21: Standardization of a specific normal distribution

Since we are seeking $P(x < 15 \text{ min.})$, we need to compute the area of the region to the left of 15. In this approach, we use our general normal distribution accumulation function instead of standardizing values. We can compute $P(x < 15)$ through **= NORM.DIST(15, 25, 5, 1)** in our spreadsheet producing the value 0.0228. This is the same value we computed using the conversion process. We have found that about 2.28% of those college students spent less than 15 minutes to complete the task.

As with the standard normal distribution function, we must remember that this function always produces only a left-tail area measure. If our regions of interest are central or right-tail regions, adjustments must be made similarly to our previous work.

For one more example, suppose in the context of the college student's time to complete a task, we wish to know the probability of randomly selecting a student who took over 37.5 minutes on the task. We produce a quick sketch again for this question:

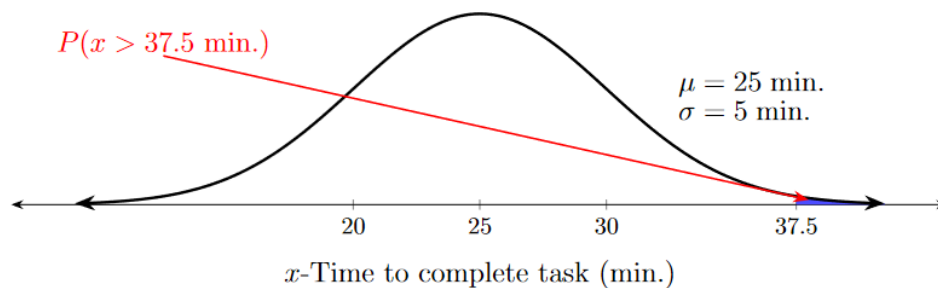


Figure 4.6.22 Probability in a non-standard normal distribution without standardizing

Since we are seeking $P(x > 37.5 \text{ min.})$, our figure shows we need to compute a right-tailed region and use of our complement rule:

$$\begin{aligned} P(x > 37.5) &= 1 - P(x \leq 37.5) \\ &= 1 - \text{NORM.DIST}(37.5, 25, 5, 1) \\ &\approx 1 - 0.9938 = 0.0062 = 0.62\%. \end{aligned}$$

The probability of randomly selecting a student from this group who took over 37.5 minutes on the task is less than 1% and considered an unusual event.

? Text Exercise 4.6.4

Sketch graphs of and determine the designated measures in the following:

- Find $P(x \geq 122)$ and $P(75 < x < 110)$ in a normal distribution with $\mu = 100$ and $\sigma = 15$.

Answer

First we find $P(x \geq 122)$. Our sketched graph is shown below...we are working in a right-tailed region. We are using an approach that does not require standardization.

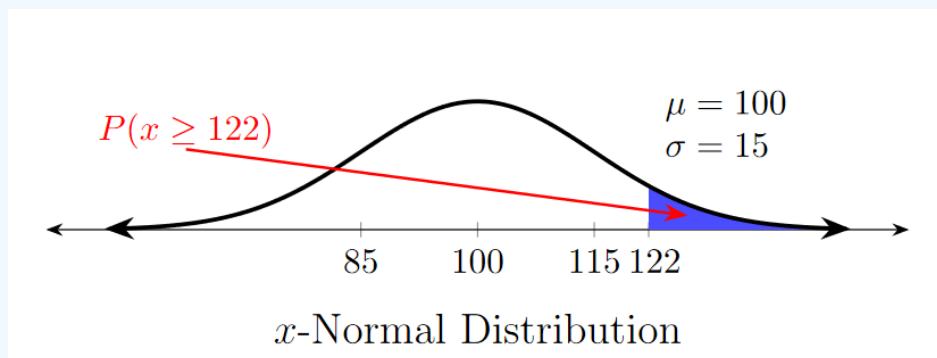


Figure 4.6.23 Normal distribution with shaded region

To find the right-tailed area measure, we make our complement adjustment.

$$\begin{aligned} P(x \geq 122) &= 1 - P(x < 122) \\ &= 1 - \text{NORM.DIST}(122, 100, 15, 1) \\ &\approx 1 - 0.9288 \\ &= 0.0712 = 7.12\% \end{aligned}$$

Remembering our quick check, we notice that the size of the shaded region in the graph seems to align with this proportional measure of 7.12%. Thus, 7.12% of the normal distribution's area is to the right of 122. Or equivalently, there is a 7.12% probability of randomly selecting a x -outcome from this distribution that is at least 122 in value.

Next, we find $P(75 < x < 110)$. Our sketched graph is shown below; noticing that we are in a central region, we must subtract two left-tailed area measures.

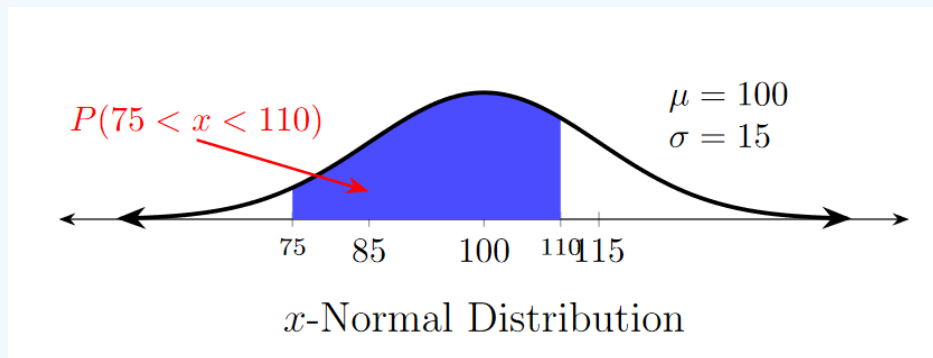


Figure 4.6.24 Normal distribution with shaded region

We subtract two left-tail areas to find the desired region's area measure.

$$\begin{aligned} P(75 < x < 110) &= P(x < 110) - P(x \leq 75) \\ &= \text{NORM.DIST}(110, 100, 15, 1) - \text{NORM.DIST}(75, 100, 15, 1) \\ &\approx 0.7475 - 0.0478 \\ &= 0.6997 = 69.97\% \end{aligned}$$

There is a 69.97% probability of randomly selecting a x -score that is between 75 and 110 in value.

2. A soft drink bottler has data that suggest that the amount of drink placed in their 12-ounce cans by a specific bottling machine is normally distributed with $\mu = 12.1$ ounces and $\sigma = 0.5$ ounces. What proportion of cans are under-filled from the labeled amount by more than 1 ounce?

Answer

After carefully reading the context about filling the cans with soft drinks, we produce the volume of soft drink probability distribution graph below.

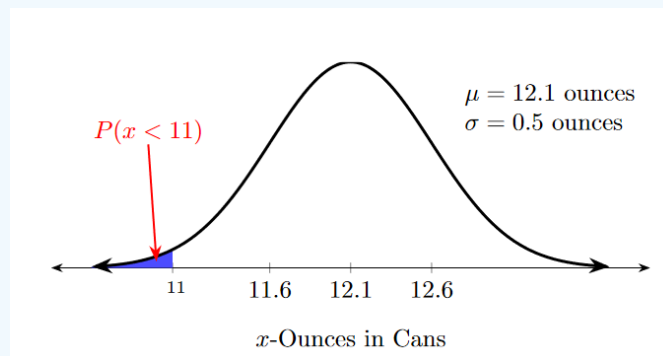


Figure 4.6.25 Normal distribution with shaded region

To determine the proportion of cans that are under-filled from the labeled amount by more than 1 ounce, we find $P(x < 11)$. We compute the area of the left-tail region.

$$\begin{aligned} P(x < 11) &= 1 - \text{NORM.DIST}(11, 12.1, 0.5, 1) \\ &\approx 0.0139 = 1.39\% \end{aligned}$$

Since the probability of randomly selecting a can filled by this machine with less than 11 ounces is less than 5%, such an outcome would be considered unusual. We note that this is the same value we produce using the standardization conversion method.

3. The average consumption of electricity by electric four-door passenger vehicles is believed to be normally distributed with $\mu = 0.346$ kWh per mile and $\sigma = 0.022$ kWh per mile, where kWh stands for kilowatt hour. Is our vehicle considered unusual if we own such an electric vehicle that achieves 0.400 kWh per mile?

Answer

After carefully reading the electricity car context, we produced the electricity consumption probability distribution graph below.

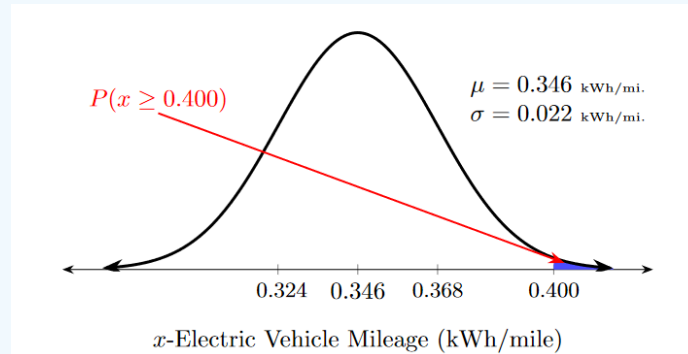


Figure 4.6.26 Normal distribution with shaded region

To determine if our vehicle is getting "unusually" high mileage, we need to determine the probability measure of having a mileage of 0.400 kWh per mile or higher; that is, we need to compute $P(x \geq 0.400)$. Per the graphic, even without computation, it appears that the probability is small. However, scales can be deceiving, so we compute the value to have measurement evidence to base our conclusion. As this is a right-tail region, we use our complement adjustment.

$$\begin{aligned} P(x \geq 0.400) &= 1 - P(x < 0.400) \\ &= 1 - \text{NORM.DIST}(0.400, 0.346, 0.022, 1) \\ &\approx 1 - 0.9929 \\ &= 0.0071 = 0.71\% \end{aligned}$$

Since the probability of getting 0.400 kWh per mile or higher is less than 1%, our electric vehicle would be considered unusual. We are getting unusually high mileage compared to similar electric vehicles.

4. Based on data taken from a large group of healthy humans in the United States, human body temperatures seem to be normally distributed with $\mu = 98.3^\circ\text{F}$ and standard deviation $\sigma = 0.92^\circ\text{F}$. If a local hospital uses 100.5°F as the lowest temperature indicating a likely fever and illness, what percentage of healthy humans will be classified as ill by this hospital?

Answer

We produce the following graph of the normal distribution of healthy body temperatures. Noting a person would be considered feverish by this hospital if they have body temperatures over 100.5°F , shown in the shaded region of the probability distribution. To compute this region, we must use the complement.

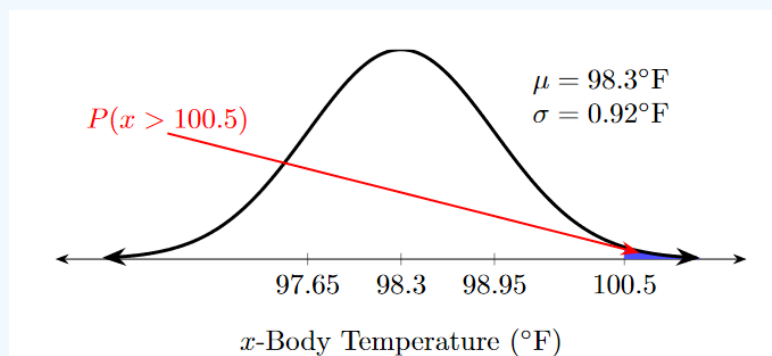


Figure 4.6.27 Normal distribution with shaded region

$$\begin{aligned}
 P(x \geq 100.5) &= 1 - P(x < 100.5) \\
 &= 1 - \text{NORM.DIST}(100.5, 98.3, 0.92, 1) \\
 &\approx 1 - 0.9916 \\
 &= 0.0084 = 0.84\%
 \end{aligned}$$

Less than 1% of healthy individuals will be classified by this hospital as feverish when they are not ill. The hospital will not likely run into this situation very often.

There will also be occasions in which we need to reverse the area/probability process above. Given the description of a region and its area measure, can we find the horizontal scale measure(s) that serve as boundary(ies) of the described region? We can use our standardization process, but there are functions for this inverse process that take care of the calculations for us and leave us within the raw/natural scale of the situation.

We return to one of our earlier questions: what are the x -scores in the normal distribution of student times for completing the specific task that produces the central 80% region of that distribution? This time, as we did in an earlier analysis, we want to avoid converting to standard normal distribution measures on our axis. As shown in Figure 4.6.28 we need to find the scale values labeled as a and b in this normal distribution that captures the central 80% of the distribution's region.

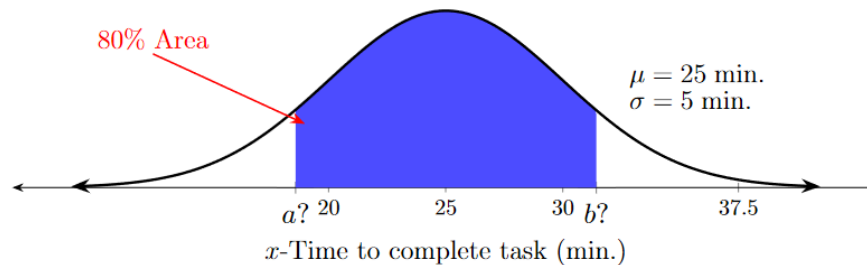


Figure 4.6.28 Inverse conversion to find scale value(s) in a non-standard normal distribution

We are saved from manually doing all the conversion work by a new inverse function, NORM.INV, that behaves similarly to our already familiar NORM.S.INV function from the standard normal distribution. If given any left-tail region's area measure, this function will compute the associated right boundary x -scale value forming that region, provided the mean μ and standard deviation σ are both known. This function has the syntax = NORM.INV(left-tail area measure between 0 and 1, μ, σ). Again, we emphasize that the function provides values only for left-tailed regions.

For the boundary value a in our diagram, we note a left-tail area measure of $10\% = 0.10$. We compute in our spreadsheet:

$$\begin{aligned}
 a &= \text{NORM.INV}(\text{left area measure}, \mu, \sigma) \\
 &= \text{NORM.INV}(0.10, 25, 5) \\
 &\approx 18.5922 \text{ min.}
 \end{aligned}$$

An x -score of approximately 18.5922 separates the lower 10% area in the given normal distribution from the upper 90% area. We also need to determine our right boundary value b in Figure 4.6.28, which has a left-tail area measure of 90%.

$$\begin{aligned}
 b &= \text{NORM.INV}(\text{left area measure}, \mu, \sigma) \\
 &= \text{NORM.INV}(0.90, 25, 5) \\
 &\approx 31.4078 \text{ min.}
 \end{aligned}$$

As a completed result, the central 80% of the students had completion times for the task between approximately 18.6 and 31.4 minutes.

We can compare these results with our earlier work (which included inverse conversion work from the standard normal distribution) to see that they are the same. We now attempt similar text exercises involving inverse distribution methods.

? Text Exercise 4.6.5

After sketching the regions described, find x -score(s) that produce the area measures described in the normal distribution.

1. When designing a building, a common requirement is to design for 95% of the population that will be using that building. To be safe as well as cost effective, and since men on average are taller than women, a building's doorways are to be designed so that all but the tallest 5% of men can walk through the doorway without having to stoop. If the heights of men are normally distributed with a mean of 161.29 cm with standard deviation of 8.3 cm, determine the design height needed for doorways.

Answer

Our sketch of the height distribution for men is shown below, with shading indicating the separation between the lower 95% and upper 5% in the distribution.

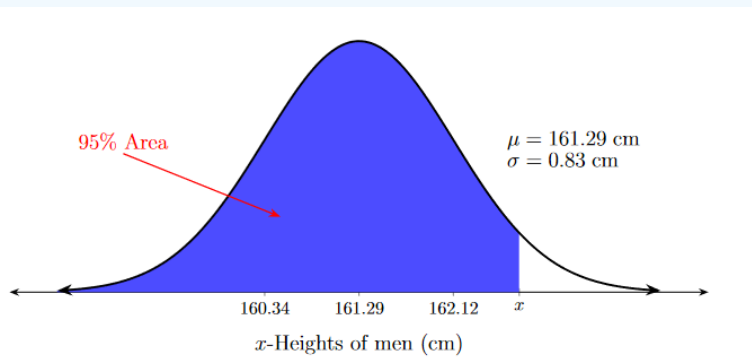


Figure 4.6.29 Normal distribution with shaded region

To find the boundary x -height value associated with the left-tailed 0.95 area value, we can go directly to our inverse accumulation function.

$$x = \text{NORM.INV}(0.95, 161.29, 0.83) \\ \approx 162.6552$$

Thus, 95% of men have heights below 162.66 cm and 5% have heights above. The building should be designed with doorways that have heights of at least 162.66 cm to fit the design specifications.

2. The average consumption of electricity by electric four-door passenger vehicles is believed to be normally distributed with $\mu = 0.346$ kWh per mile with $\sigma = 0.022$ kWh per mile, where kWh stands for kilowatt hour. What is the central 90% expected average consumption for these types of electric vehicles?

Answer

After sketching our normal distribution (shown below), we are seeking the two boundary values of a and b that separate the central 90% of our distribution.

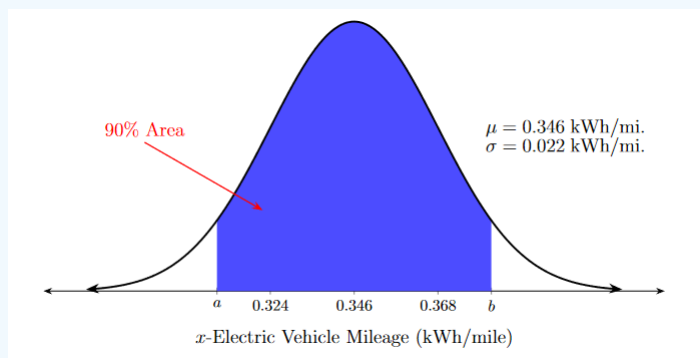


Figure 4.6.30 Normal distribution with shaded region

To find the left boundary value a , we use our inverse accumulation function with the 5% left-region area measure.

$$a = \text{NORM.INV}(0.05, 0.346, 0.022) \\ \approx 0.3098$$

To find the right-boundary b , we again use the inverse accumulation function with a $90\% + 5\% = 95\%$ total left-region area measure.

$$b = \text{NORM.INV}(0.95, 0.346, 0.022) \\ \approx 0.3822$$

Thus, the central 90% average consumption of electricity by these electric vehicles is expected to be between 0.310 and 0.382kWh per mile.

3. A tire company is about to begin large-scale manufacturing of a new tire made of newly developed materials. The tire's tread life has been tested, the research team found the tread life in miles produced a normal distribution with $\mu = 72,000$ miles and $\sigma = 7,000$ miles. The company must develop a consumer warranty policy and only wants to replace tires that do not last sufficiently to tested expectations. The company decides to only set the mileage warranty to cover the lowest 5% of their tires. What is the mileage number they will need to place on the warranty policy?

Answer

We sketch the described tire-life distribution (shown below). Next, we go directly to our inverse accumulation function to compute the related x -mileage value for the boundary value establishing the lowest 5% of the distribution's area.

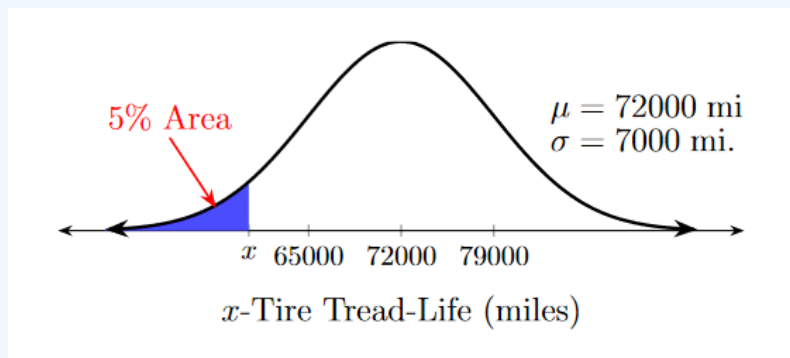


Figure 4.6.31: Normal distribution with shaded region

$$x = \text{NORM.INV}(0.05, 72000, 7000) \\ \approx 60,486.02$$

Thus, 5% of the tires can be expected to last less than 60,486 miles. The tire company should set their replacement warranty value near this value, likely at 60,000 miles just to round to an easier value for customers.

4. Established in 1946, Mensa, currently a global community of around 150,000 people, requires individuals first score in the upper 2% (in relation to the general population) on an IQ test before being considered for membership. If the general population produces normally distributed scores with $\mu = 100$ points and $\sigma = 15$ points on the IQ test, what must we score on the exam to be considered for membership in Mensa?

Answer

Our sketch of the distribution for IQ scores is shown below, with shading indicating the upper 2% region of the scores.

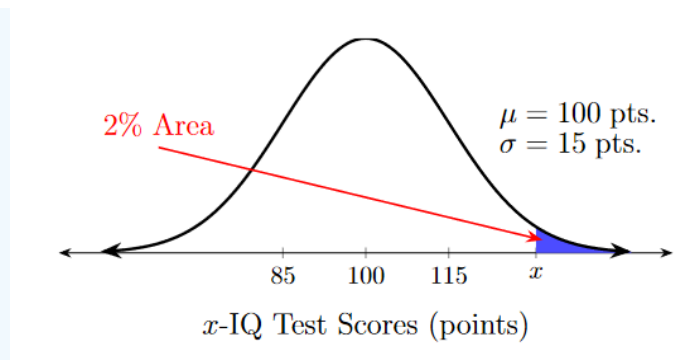


Figure 4.6.32 Normal distribution with shaded region

To find the boundary x -height value associated with the right-tailed 0.02 area value, we can apply complement action ($1 - 0.02 = 0.98$) within the use of the inverse accumulation function.

$$\begin{aligned} x &= \text{NORM.INV}(0.98, 100, 15) \\ &\approx 130.8062 \end{aligned}$$

We must score at least 130.8 points on the IQ test to be within the top 2%.

We have now found ways, using another technology accumulation function and its inverse, to be able to produce various area or x -scale measures of any x -normal distribution. These new methods eliminated in our work the converting back and forth between a general normal distribution and the standard normal distribution. Yes, the above technology does make our work less intense by hiding the conversion work (in the programming of the functions the conversion work is actually still happening). This is a blessed simplification for us humans as we prefer to eliminate calculation work when possible. However, we again warn that applying the conversion process is necessary in some of our future work, so we should practice both approaches.

Summary

We now have the tools to answer practically any probability related question tied to normal distributions. Our technology's accumulation functions will produce accurate measures of left-region areas of all types of normal distributions. The inverse functions will allow us to find scale measures tied to given regions of a normal probability distribution. In the future, we will also examine similar accumulation functions for other common probability distributions, such as the t - and χ^2 -distributions.

4.6: Accumulation Functions And Area Measures in Normal Distributions is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by The Math Department at Fort Hays State University.

- **3.3: Measures of Central Tendency** by [David Lane](#) is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.