

TutorTube: Z-Scores and the Normal Distribution

Fall 2020

Introduction

Hello, welcome to another edition of TutorTube, where the Learning Center's Lead Tutors help you understand challenging course concepts with easy to understand videos. My name is Kelly Schmidt, Lead Tutor for statistics at the Learning Center. In today's video, we will explore a few of the fundamental topics and skills related to the normal distribution that will help you to be successful in the course. Let's get started!

z-Scores

When we are working with data that follows a normal distribution, one of the most important concepts to understand is the z-score. The formula for z-scores for samples is:

$$z = \frac{x - \bar{x}}{s}$$

Why do we need z-scores? Here's an example: Let's say we have classes of students taking the same class with two different professors, Professor A and Professor B. Professor A gives his class a test and overall the class does pretty well. The average score is a 75%.

However, Professor B also gives a test, but this test is much harder. The class average is only a 40%. Let's say I'm a student in Professor B's class and I know that I got the highest score on the test at a 65%. How do we compare my performance to the performance of students on Professor A's test? My score of 65 was less than the class average for Professor A, but it was still the best in my class. You can't judge my performance on the same scale as the students in Professor A's class. We need some way to represent situation, and that's where z-scores come into play.

Using z-scores allows us to **standardize** scores so that we can compare them across populations that have different means and standard deviations. (As long as they follow the normal (bell-shaped) distribution.)

Practice #1:

Let's do a practice problem to see how this works. Here's a common question that you will see in this section:

- Mark finished 1st in last year's annual race in 16.2 minutes. The average time for last year's race was 18.2 with a standard deviation of 2 minutes.
- Sally finished 1st this year's race in 16.7 minutes. The average for this year's race was 20.3 with a standard deviation of 3 minutes.

Who did relatively better?

Z-Score Formula

$$z = \frac{x - \bar{x}}{s}$$

Observation → x \bar{x} ← Mean
 s ← Standard Deviation

Figure 1: Z-Score Formula

Ok, first off, notice the phrase “relatively” in the question. They aren't asking who had the fastest time, they are asking who had the fastest time *relative* to other runners in their race. Because of this, in order to answer this question we need to find the z-scores for Mark and Sally.

Recall that the z-score formula is: Observation minus Mean divided by Standard Deviation. First, let's find each of these values for each runner.

Mark's Race	Sally's Race
$\bar{x} = 18.2$	$\bar{x} = 20.3$
$s = 2$	$s = 3$
$x = 16.2$	$x = 16.7$

For Mark's race, the mean was 18.2, the standard deviation was 2, and his score was 16.2. For Sally's race, the mean was 20.3, the standard deviation was 3, and her score was 16.7.

Now, let's calculate Mark's z-score, which I'll label z_M :

$$z_M = \frac{16.2 - 18.2}{2} = -1$$

So, we can say that Mark had a standardized score, a z-score, of -1. What does this mean? Well, we know now that Mark's time was 1 standard deviation faster than the average score in his race. Not bad!

Next, for Sally's z-score, which we'll label z_S :

$$z_s = \frac{16.7 - 20.3}{3} = -1.2$$

So, we can say that Sally had a standardized score of -1.2. What do these scores mean? This tells us that Sally's score was 1.2 standard deviations faster than the average time for her race, which is even better than Mark.

So, we can answer the question by saying that even though Sally had a slower time, she actually did **relatively** better than Mark during her race.

Empirical Rule

Next, let's look at the Empirical Rule. This is a concept that will come up a lot when you are working with the normal distribution. The formal definition of the Empirical Rule is this:

If the shape of the distribution is bell-shaped (a.k.a. normal), then

- Approximately 68% of the data lie within 1 standard deviation of the mean
- Approximately 95% of the data lie within 2 standard deviations of the mean
- Approximately 99.7% of the data lie within 3 standard deviations of the mean

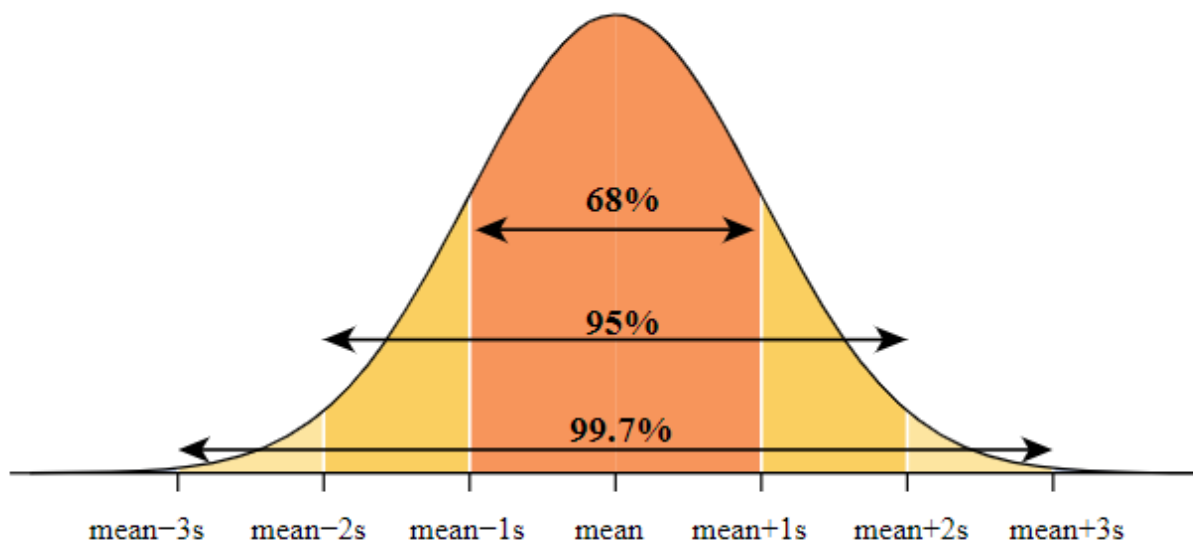


Figure 2: Empirical Rule (Penn State, 2020)

What does this mean? Well, the simplest way to think about it that as we move farther and farther away from the mean of our dataset, the less likely we are to see an observation that extreme. For example, since we know that 99.7% of all our data typically lie within 3 standard deviations of the mean, the odds of seeing an observation more extreme than 3 standard deviations above or below would only be .03% (Which comes from subtracting 99.7% from 100%).

Let's do some practice problems to see how this rule helps us out when we are working with data.

Practice #2:

Scores on an IQ test have a bell-shaped distribution with a mean of 100 and a standard deviation of 14.

a. What percentage of people have an IQ score between 86 and 114?

The first step is to build our distribution using the values they gave us. The mean is 100, so we place this value in the center. Next, since the standard deviation is 14, we find the values of the tick marks on either side of the mean by adding and subtracting 14 from 100.

$$100 - 14 = 86$$

$$100 + 14 = 114$$

So, we know a score of 86 is one standard deviation below the mean, and a score of 114 is one standard deviation above the mean.

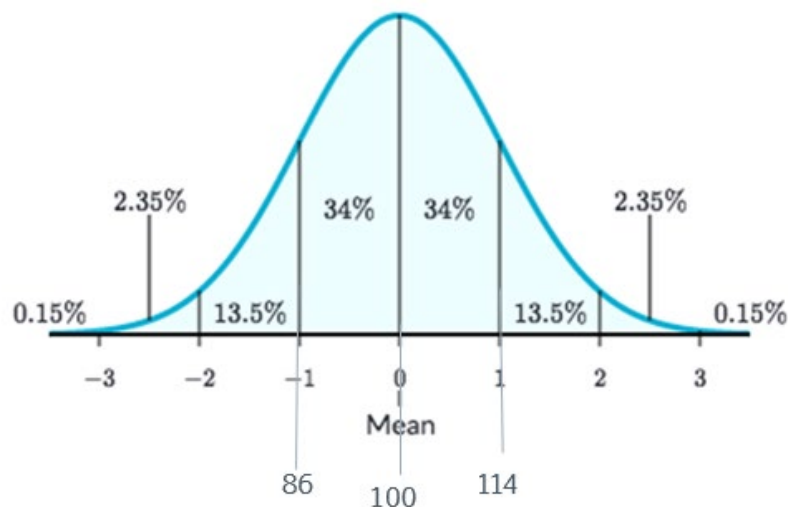


Figure 3: Empirical Rule Segmented (Xie, 2019)

Now back to our question; we want to know the percentage of people between 86 and 114. Using the chart here (which is a simplified version of the Empirical Rule that I prefer to use) all we need to do is add the percentage values between our two values: 34% plus 34% equals a total of **68%**.

Now for Part B:

b. What percentage of people have an IQ score less than 86 or greater than 114?

This question is actually asking us to find two different areas. The first is the percentage of scores less than 86. We can do this by using the same chart we just made. Adding all the percentages to the left of 86 we get:

$$13.5\% + 2.35\% + 0.15\% = 16\%$$

Similarly, if we want to find the percentage greater than 114, we simply add all the percentage values to the right of 114:

$$13.5\% + 2.35\% + 0.15\% = 16\%$$

Combining these two values together, 16 plus 16 gives us a total answer of **32%**.

Next, Part C:

c. What percentage of people have an IQ score greater than 142?

For this question, we need to find the location of a score of 142. We know it is larger than the mean, but we don't know if it's one, two, three, or more standard deviations above. Let's start by finding the values on the next two tick marks on our distribution:

$$114 + 14 = 128$$

So, we know that 128 is two standard deviations above the mean. Now we repeat:

$$128 + 14 = 142$$

We have found the value we were after; 142 is 3 standard deviations above the mean. Now we can simply add the percentage values greater than 142 to find our answer. This is only one value in this case: **0.15%**.

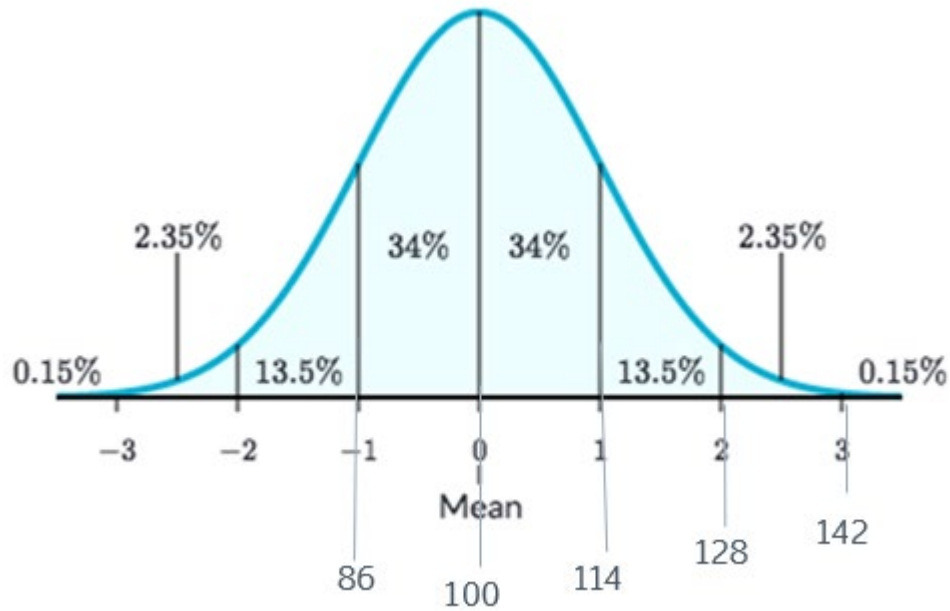


Figure 4: Labeled Distribution

Normal Calculator in StatCrunch

We can also use the normal calculator in StatCrunch to find the answers to the problems we just completed. In StatCrunch, we go to Stat, then Calculators, then Normal.

Figure 5: Stat > Calculators > Normal

This will open up the normal calculator. Let's say we wanted to find the answer to Part C (What percentage of people have an IQ score greater than 142?).

First, we enter the values for the mean and standard deviation we were given. Here that was 100 for the mean and 14 for the standard deviation.

Next, since the problem wants the proportion GREATER than 142, we change the sign to a greater-than sign, and then enter our value of 142 into the box. Now we just hit Compute. We can see that the answer given is 0.0013499, or about **0.135%**. Notice that this is not the same answer that we got when we used the empirical rule which was 0.15%; this is because StatCrunch gives us exact answers while the empirical rule only provides us with estimates.

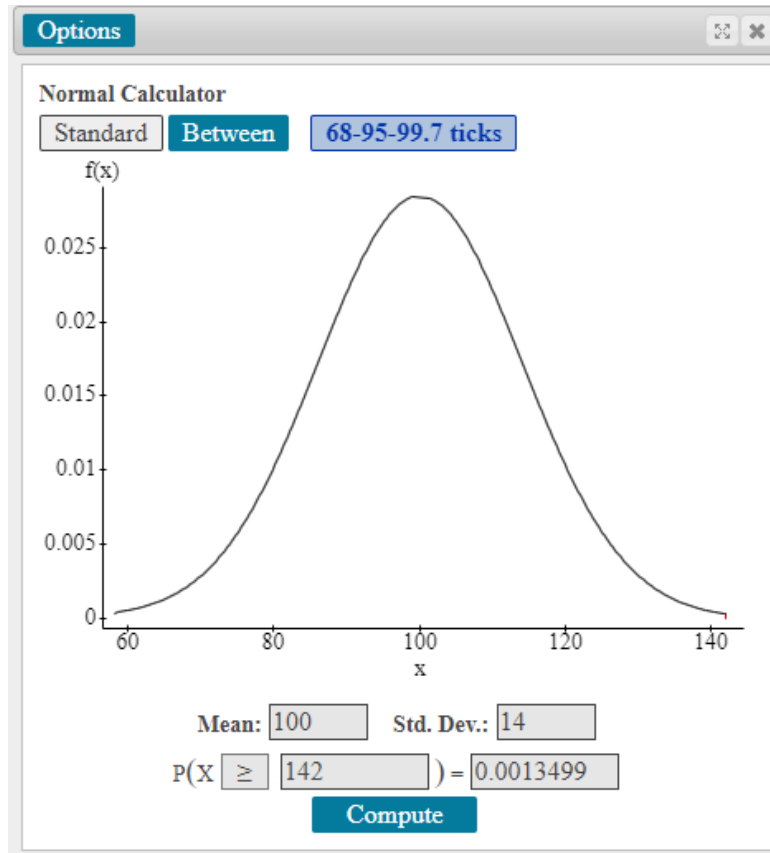


Figure 6: Part C using the Normal Calculator

Now, let's say that we wanted to find the answer to Part A (What percentage of people have an IQ score between 86 and 114?)

The key word in the question here is "between." Since we want the area under the curve between two values, we need to select the "Between" option at the top of the Normal Calculator. From there, we simply enter the two values at the lower and upper end points of our interval, 86 and 114, into our boxes. Then click Compute.

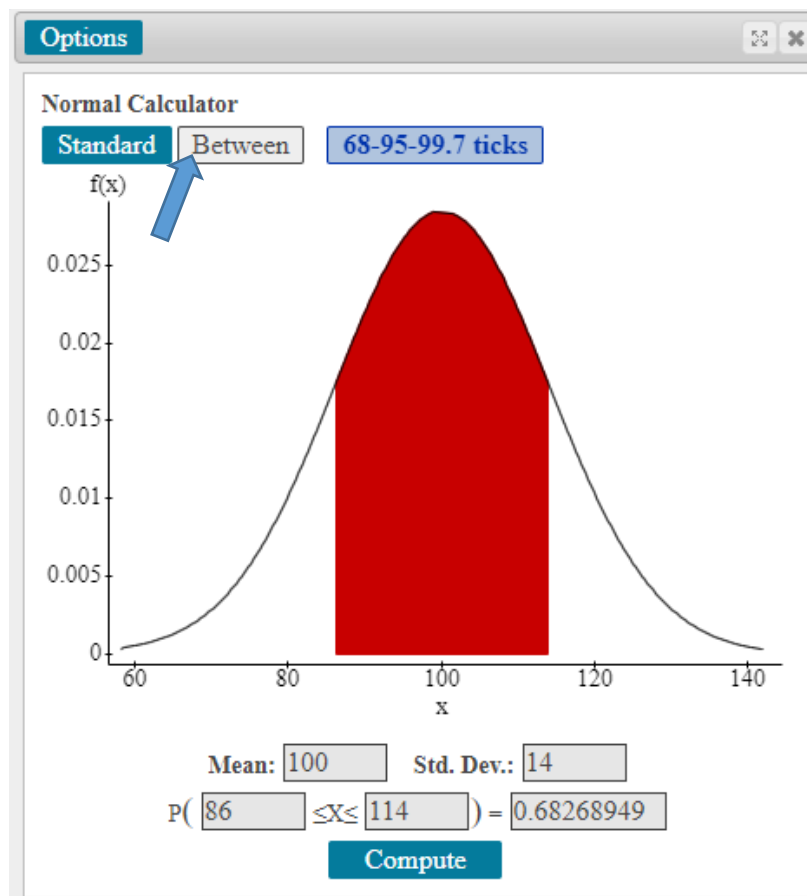


Figure 7: Part A using the Normal Calculator

Again, notice that the answer we come up with here (**0.6826**) is not exactly 68%. However, it's pretty close, which means that the empirical rule is able to give us a pretty good approximation of the exact values without use of technology.

Alright, so with that we've finished all the parts of our problem. I hope you found this to be a good introduction to z-scores and the empirical rule.

Outro

Thank you for watching this TutorTube presentation! Please subscribe to our channel for more exciting videos. Check out the links in the description below for more information about The Learning Center and follow us on social media. See you next time!

References

Pearson. (2020). MyLab: Statistics. *Pearson Higher Education Inc.*

Penn State. (2020). STAT 200 - Elementary Statistics: Standard Normal Distribution. *Pennsylvania State University*. Web.

Xie, S. (2019). Empirical rule (68-95-99.7 Rule). *Medium.com*. Web.

*All calculations in this video were performed with Pearson StatCrunch 2020 software.