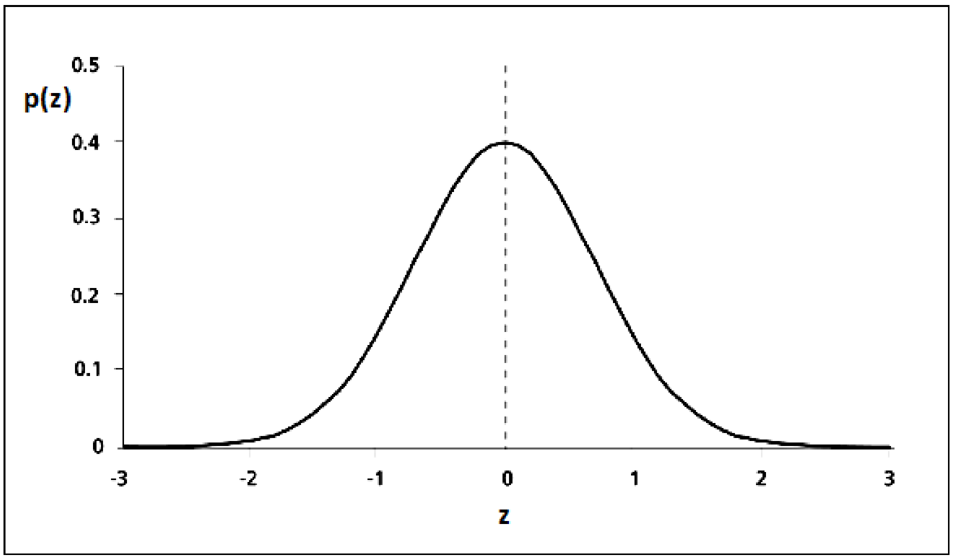
Why the Normal Distribution is Actually a Distribution

Since a great deal of statistics is calculus – based, every once in a while, we need to (in class) take something on faith. One of those "somethings" is the fact that the area between the normal curve and its horizontal axis is 1. It's something I dropped in your lap when we first studied this wondrous curve, and you all seemed to not mind the statement a bit. In this paper, I'll formally demonstrate, in two ways, why I wasn't full of it.

1. **Using the idea of additive areas (Riemann Sums)**

Without calculus, this is really the only way to show you why this sum has to work. It might seem a bit brute force, but this is precisely the method that calculus uses to find areas of non – traditional regions.

Imagine the graph of the standard normal distribution:



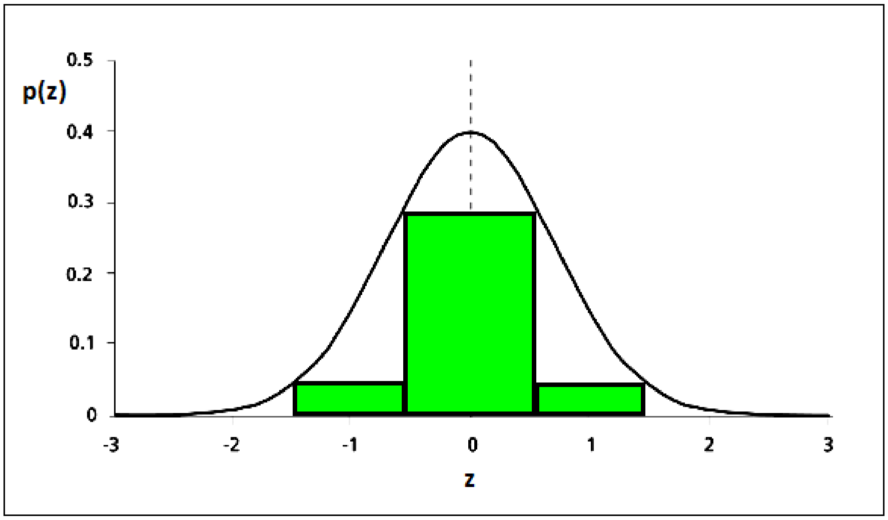
Ah, she's beautiful, no? To find the area we desire, we would need to shade between the curve and that z - axis. There is one immediate issue: the curve, being based in an exponential function, never ***touches*** the z – axis[1]. That's OK; since it's asymptotic at its tails, we can pretty much ignore the area left of -3 and right of 3, since there isn't a whole lot of area there, anyway (we'll just have to realize that our area estimates will be off a little because of this).

Now...the idea of Riemann sums is pretty simple: since we can't find the area of this curve directly using geometry (since its shape isn't the kind we studied way back when), we'll fill it up with shapes we do remember...rectangles[2]!

---

[1] Its equation, in case you've forgotten, is $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

[2] If you've taken MTH 241 or MTH 251, you've done this before.

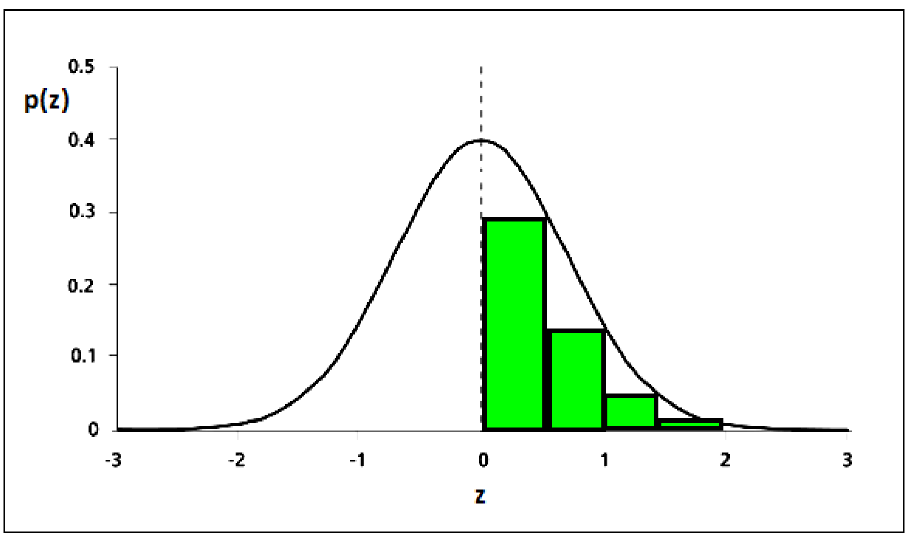


The idea is simple: we add up the areas of those rectangles. Their area will be close to the area we need. These three areas, using the function rule from footnote 1 above, are as follows:

$$Area_1 + Area_2 + Area_3 = \frac{1}{\sqrt{2\pi}}\left(e^{-\frac{(-1.5)^2}{2}} + e^{-\frac{(-0.5)^2}{2}} + e^{-\frac{(1.5)^2}{2}} \right) \approx 0.611$$

Wow...that's not very close to 1 at ***all***. What the hell?

Oh, wait...I only used 3 rectangles, and they were really, really **w i d e** . I should use more rectangles, and make them skinnier, so they "fit" the area better (I'll also only do half of the curve, and then double it, since it's symmetric):



$$Area_1 + Area_2 + Area_3 + Area_4 = \frac{1}{2\sqrt{2\pi}}\left(e^{-\frac{(0.5)^2}{2}} + e^{-\frac{(1)^2}{2}} + e^{-\frac{(1.5)^2}{2}} + e^{-\frac{(2)^2}{2}}\right) \approx 0.389$$

Doubling yields an area of about 0.778.

Better, but still not close enough to 1. How to proceed? You guessed it: skinnier rectangles, so you can fit more in. I'll let Excel take it from here:

With this many rectangles...	...here's the area approximation:
24	**0.899**
60	**0.958**
120	**0.978**
600	**0.993**
1200	**0.995**

See how it's getting closer and closer to 1? It has to, as the rectangles get closer[3]. So, Riemann sums, especially with technology, are a great way for us to get this area.

2. **Using the idea of integration**

If you've had calculus (integral calculus), then you've seen the idea of Riemann Sums applied to the "infinite rectangles" case. This technique is called integration (if you haven't had calculus, you might want to skip this section).

To find the area we need, we'll just integrate the function from 0 to infinity, then double (much like we did above):

$$\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{x^2}{2}} dx$$

One teeny problem...it's been shown that the integral form we're using has no closed form. Fortunately, its numerical approximation is known, and we can use that instead:

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \frac{1}{2} \sqrt{\frac{\pi}{\frac{1}{2}}} = \frac{1}{2\sqrt{2\pi}} \sqrt{2\pi} = \frac{1}{2}$$

Which, upon doubling, equals one. Much easier than rectangles!

[3] Actually, since I'm only going from -3 to 3, it's getting closer to 0.9974...but this argument could be applied with infinite endpoints as well.