

## Where did Table 15.1 in Fundamentals text come from.

I generated Table 15.1 (and Appendix D.5) in *Fundamental Statistics for the Behavioral Sciences* many years ago, and when I was recently asked about how I generated it I had to think for a while. But actually that is pretty easy after a bit of thought.

First, keep in mind that the approach in that chapter is based on an approximation to power. That approximation is based on the standard normal distribution ( $z$ ). (I will give a more exact solution below, but the normal approximation is really good enough for just about any purpose we want.) The advantage of the normal distribution is that the shape of the distribution does not depend on sample size, so I don't have to worry about my degrees of freedom. When I come to using the (noncentral)  $t$  distribution to give a more exact solution I will have to pay attention to the degrees of freedom. The distribution will still be symmetric, but it will not be normal.

I will start with an example that I expect to have a small effect size ( $d = 0.20$ ). Imagine that I want to study to examine a product advertised to help you gain weight. If the product did not work I would expect a mean weight gain of 0 pounds (my null hypothesis), but from previous research I expect that the mean weight gain from 25 participants will be 3 pounds with a standard deviation of 15. (I have such a large standard deviation because many people will probably lose weight or even gain a large amount of weight even though the average gain is 3.) I want to compute the power of this

experiment to find a significant mean weight gain if, as I expect, the true mean gain is 3 pounds with a standard deviation of 15.

$$d = \frac{\mu_1 - \mu_0}{\sigma} = \frac{3 - 0}{15} = .20$$

$$\delta = d\sqrt{n} = .20\sqrt{25} = 1.00$$

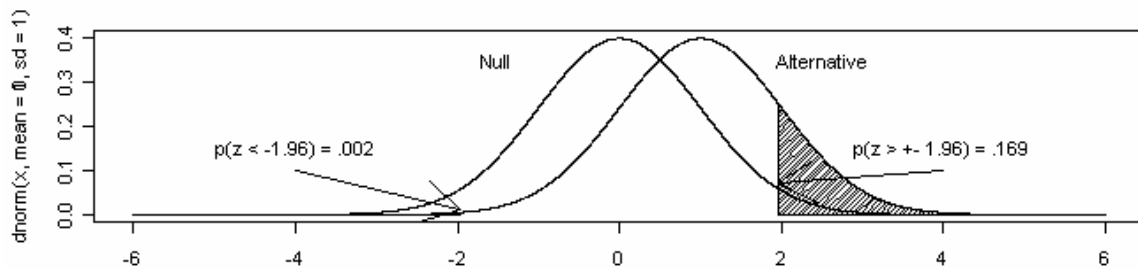
Notice that I have an effect size of .20, which Cohen would think of as a small effect.

When we combine the effect size and the sample size to produce  $\delta$ , which is required for use of Table 15.1, we have  $\delta = 1.00$ . (When we come to working with  $t$  instead of  $z$ ,  $\delta$  will be known as the “noncentrality parameter.”)

From Table 15.1 we see that for a two tailed test as  $\alpha = .05$  the predicted power for our experiment is only .17, meaning that if we are correct in our expectations, only 17% of the time that we run this experiment will we get a significant result. But where did that .17 come from?

First of all remember that with a two-tailed  $z$  test (see Chapter 12, Section 12.2) the critical value at  $\alpha = .05$  is  $\pm 1.96$ . In other words we will reject the null whenever we calculate a  $z$  on our sample data in excess of  $\pm 1.96$ . Also, if the  $\mu_1$  really equals 3 with a standard deviation of 5, as we expect, the distribution of  $z$  values for a huge number of replications of this experiment would be have a standard normal distribution around  $\delta (= 1)$  with a standard deviation of 5. So all we really need to ask is “How often would we get

an obtained value of  $z$  greater than  $\pm 1.96$  from a normal distribution with mean = 1 and standard deviation = 5?" The diagram that follows illustrates that probability. Notice that on the left I have drawn the distribution of expected  $z$  values under the null, and on the right I have drawn expected  $z$  values under the alternative hypothesis. I have also given the areas under the right hand distribution for values outside the range of  $\pm 1.96$ . If you add these two areas you get  $.002 + .169 = .171$ , which is the value given in the table for  $\delta = 1$ ,  $\alpha = .05$ , two-tailed.

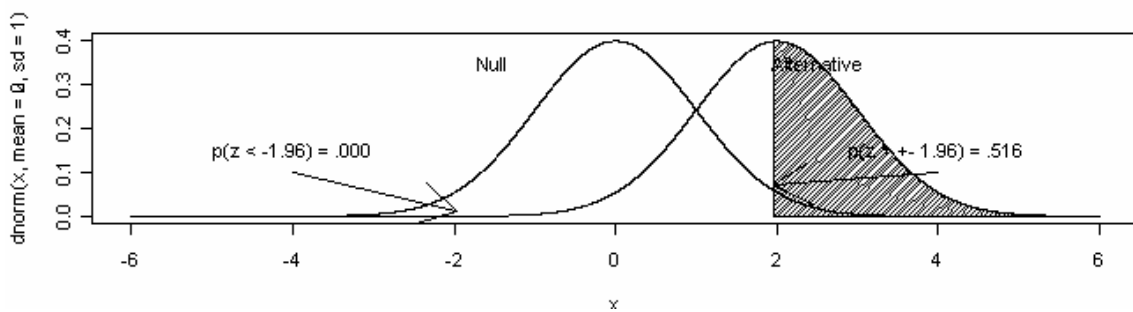
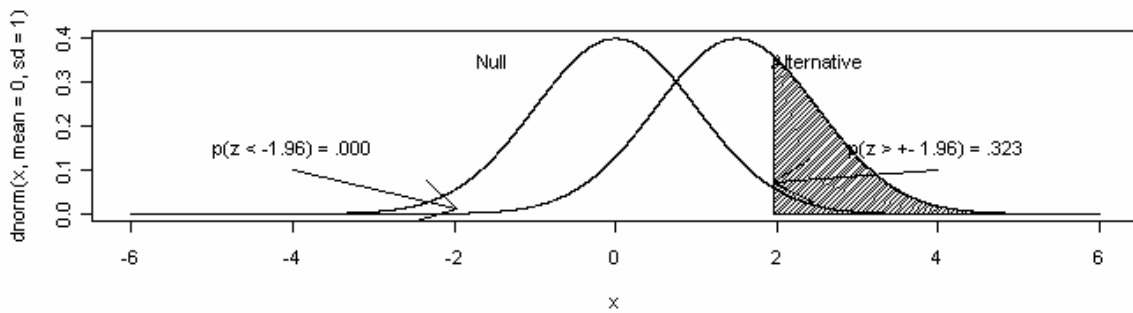
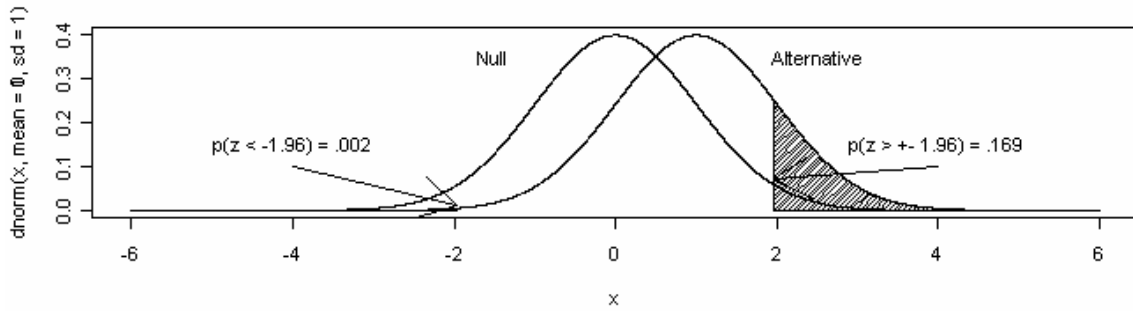


### What about other values of $\delta$ ?

I now want to repeat the above but consider other values of  $\delta$ , which is the effect size adjusted for  $n$ . I could alter  $\delta$  in several ways. I could change the expected mean, the expected standard deviation, or the sample size ( $n$ ). I am going to do the latter, and I will aim for  $\delta = 1.50$  and  $2.50$ , which will require sample sizes of 56 and 100, respectively.

The logic is exactly the same as it was for the first example except that I substitute the new values of  $\delta$ . This will cause the alternative hypothesis to be displaced to the right, with more and more of the area under the alternative exceeding  $\pm 1.96$ , resulting in greater power.

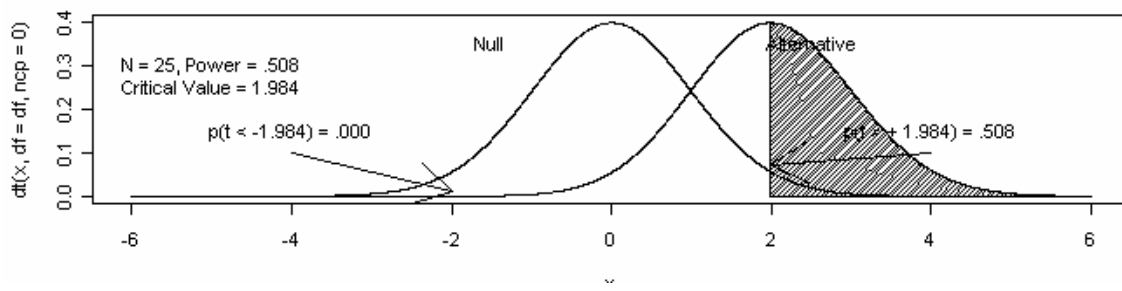
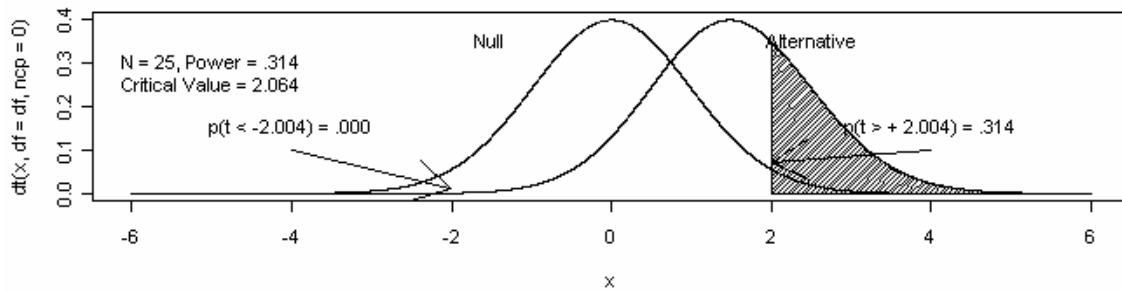
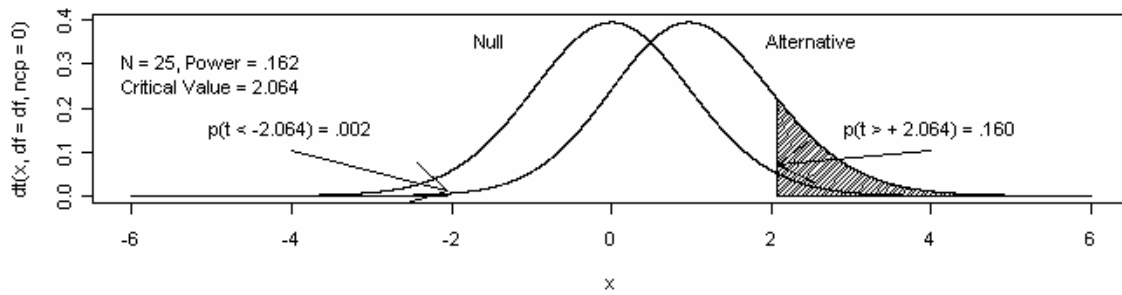
The results of these runs are shown in the figure below, where I have included the case with  $\delta = 1$  simply to make it easier to compare. If you add the areas in both tails of the distribution you will get .32 and .52 for  $\delta = 1.5$  and 2.0, respectively. These agree with the values shown in Table 15.1.



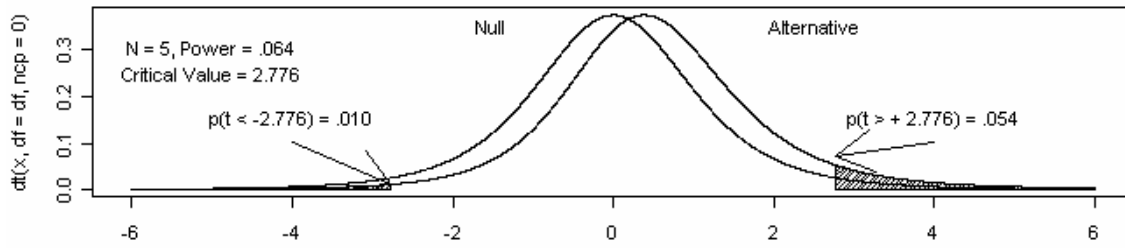
### **But I Want to be More Exact—I Don't Want an Approximation.**

The calculations given above assume that you are going to use the normal distribution to approximate the power of our test. But that means that the critical values are always  $\pm 1.96$  regardless of the sample size. The correct way to deal with this problem is to use the (noncentral)  $t$  distribution. It is called the noncentral  $t$  because it will be centered over the noncentrality parameter ( $\delta$ ) instead of over 0, as it is in the normal  $t$  tables.

Because I will get a different distribution for each value of sample size and for each value of  $\delta$ , I could generate a huge number of figures similar to the ones above. Instead I am going to stick with  $d = 0.20$  and vary  $n$  from 25 to 56 to 100. This will give me  $\delta$  values of 1.0, 1.5, and 2.0. The only real change that I have to make to my program, other than to change the critical values in line with the changing degrees of freedom, is to replace commands to draw normal distributions with commands to draw  $t$  distributions.



From this figure you can see how power increases as sample size increases. You can also see that in the top plot power is .162, which is close to what we obtained with our approximation. However, the approximation is not good for very small sample sizes. Below I have included the same plot as above but with  $n = 5$ . Here you can see that the true power is .064 rather than the .17 that our approximation gives. However when you are working with such low power values you probably should not run that experiment to begin with.



The R program that generated these plots is available if you would like it, but it was assembled quickly and is far from elegant.

dch: 10/30/2008