

## WISE Power Tutorial – All Exercises

### Exercise 1a: Power and Effect Size (Differences between Means)

If you do not feel comfortable with hypothesis testing concepts, you may want to complete the [WISE Hypothesis Testing Tutorial](#) and return to this tutorial later.

How probable is it that a sample of graduates from the ACE training program will provide convincing statistical evidence that ACE graduates perform better than non-graduates on the standardized Verbal Ability and Skills Test (VAST)? What is this probability for a less effective competitor, the DEUCE training program? Power analysis will allow us to answer these questions.

In **Exercise 1** we will use the WISE Power Applet to examine and compare the statistical power of our tests to detect the claims of the ACE and DEUCE training programs. We begin with a test of ACE graduates.

We assume that for the population of non-graduates of a training course, the mean on VAST is 500 with a standard deviation of 100. For the population of ACE graduates the mean is 580 and the standard deviation is 100. Symbolically,  $\mu_0 = 500$ ,  $\mu_1 = 580$ , and  $\sigma = 100$ . Both distributions are assumed to be normal.

The effect size,  $d$ , is defined as the number of standard deviations between the null mean and the alternate mean. In this example the effect size is .80. Symbolically,

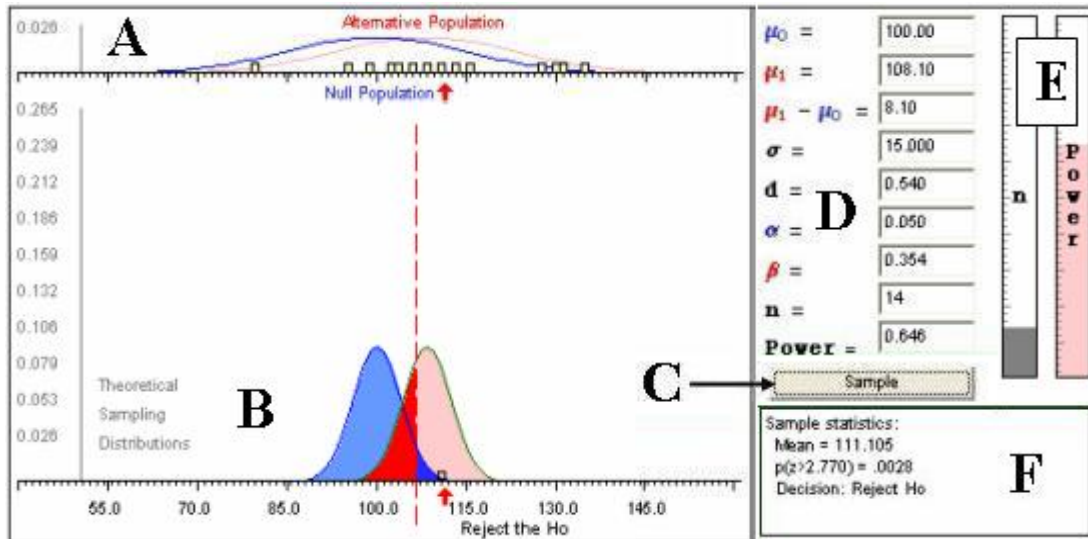
$$d = \frac{\mu_1 - \mu_0}{\sigma} = \frac{580 - 500}{100} = .80$$

Click [here](#) for more information on effect sizes.

### Using the WISE Power Applet

The WISE Power Applet (which is shown below as a static picture) will be used to simulate drawing a sample of graduates from the ACE program. At the top (**Area A**), the blue curve represents the population distribution for non-graduates (**Null Population**) while the red curve represents graduates from the ACE program (**Alternative Population**). For this exercise we assume both populations are normal distributions. In the textboxes to the right (**Area D**), we can set values for the two population means ( $\mu_0$  and  $\mu_1$ ) and the population standard deviation ( $\sigma$ ) by entering values into the textboxes. We can also set  $n$ , the number of cases to be sampled, and our alpha error rate,  $\alpha$ . After changing any of these values, be sure to press **Enter**.

Pressing the **Sample** button (**Area C**) simulates drawing a sample of size  $n$  from the **Alternative Population**. The sample of  $n$  cases is shown as small yellow boxes in **Area A** and the sample mean is shown with a red arrow. The sample mean is also shown below relative to the two theoretical sampling distributions (**Area B**).



We will reject the hypothesis that our sample came from the Null Distribution if our sample mean is far from the center of the blue sampling distribution. In this example, we may reject this hypothesis as unlikely if our sample mean falls in the extreme upper 5% of the blue distribution (one-tailed alpha error = .05, symbolically  $\alpha = .05$ ). The applet (**Area F**) shows the z-value of the sample mean on the null distribution as well as the one-tailed  $p$ -value and the decision: Reject or do not reject the null hypothesis ( $H_0$ ). In the example shown here, the sample mean is 111.105 and z-value on the null sampling distribution of the mean (blue) is 2.770. The probability of finding a z-score greater than 2.770 if we are sampling from the null distribution is  $p = .0028$ . Because this value is less than alpha, our statistical decision is to reject  $H_0$ .

**Area D** shows many statistical values including power and effect size, and **Area E** represents sample size ( $n$ ) and power as ‘thermometers.’ In the actual applet on the next page you will be able to change any of these values.

### Exercise 1b: Sampling 25 ACE Graduates (Mean = 580)

To simulate drawing a random sample of 25 cases from graduates of the ACE program, enter the following information into the applet below:

- $\mu_0 = 500$  (null mean);
- $\mu_1 = 580$  (alternative mean);
- $\sigma = 100$  (standard deviation);
- $\alpha = .05$  (alpha error rate, one tailed);
- $n = 25$  (sample size).

**Press enter/return after placing the new values in the appropriate boxes!**

To simulate drawing one sample of 25 cases, press Sample. The mean and z-score are shown in the applet (bottom right box). Record these values in the first pair of boxes below (you may round the mean to a whole number).

The z-score computed on the null sampling distribution allows us to determine the probability of observing a sample mean this large or larger if the null hypothesis is true. The sample actually came from the alternative population, but is the sample mean large enough to provide convincing evidence that the sample did not come from the null population? The **dashed red line** shows where we have set our alpha

criterion. In this case we set  $\alpha = .05$ , corresponding to the upper 5% of the blue null sampling distribution. If our sample mean is to the right of the dashed line, we can reject the null hypothesis with  $p < .05$ , one-tailed (and correctly conclude that the sample did not come from the null population). If a sample mean falls to the left of the dashed line, we fail to reject the null hypothesis. This would be a Type II error (i.e., failure to reject a false null hypothesis) because we are actually sampling from the alternate distribution.

Now draw nine more samples and record the mean and z for each (mean / z):

	mean	z-score		mean	z-score		mean	z-score		mean	z-score
1	<input type="text"/>	<input type="text"/>	2	<input type="text"/>	<input type="text"/>	3	<input type="text"/>	<input type="text"/>	4	<input type="text"/>	<input type="text"/>
5	<input type="text"/>	<input type="text"/>	6	<input type="text"/>	<input type="text"/>	7	<input type="text"/>	<input type="text"/>	8	<input type="text"/>	<input type="text"/>
9	<input type="text"/>	<input type="text"/>	10	<input type="text"/>	<input type="text"/>						

The power of this statistical test is the probability that the sample mean will be large enough to allow us to correctly reject the null hypothesis. Because we are actually sampling from the **Alternative Population** (red distribution), the probability that we will observe a sample mean large enough to reject  $H_0$  corresponds to the proportion of the red sampling distribution that is to the right of the dashed line. Later we will see how to compute this value. For now, we can use the value provided by the applet, **.991**.

Thus, if we draw a sample of 25 cases from ACE graduates, the probability is 99.1% that our sample mean will be large enough that we can reject the null hypothesis that the population mean is only 500. The probability that we will fail to reject  $H_0$  is only  $1.000 - .991 = .009$ , less than one chance in 100.

1a. How many times could you reject the null hypothesis in your ten samples? \_\_\_\_\_  
 (Use one-tailed alpha  $\alpha = .05$ ,  $z = 1.645$ , so reject  $H_0$  if your z-score is greater than 1.645)

### Exercise 1c: Sampling 25 DEUCE Graduates (Mean = 520)

Now we will test the claims of the DEUCE training program. The mean score for the population of graduates of this program is 520. Again we assume the population distribution is normal with a standard deviation of 100. The population effect size for the DEUCE program is only .20.

$$d = \frac{\mu_1 - \mu_0}{\sigma} = \frac{520 - 500}{100} = .20$$

Recall the effect size for the ACE program was much larger:

$$d = \frac{\mu_1 - \mu_0}{\sigma} = \frac{580 - 500}{100} = .80$$

**1b.** Before drawing samples, consider how the statistical power will differ for a test of DEUCE graduates compared to the power we found for a test of ACE graduates. That is, do you expect you will be more likely or less likely to reject the null hypotheses for a sample of 25 graduates drawn from the DEUCE program compared to a similar test for the ACE program? Explain your response below.

To simulate drawing a sample of 25 from graduates from the DEUCE program, enter the following information into the WISE Power Applet:

- $\mu_0 = 500$  (null mean);
- $\mu_1 = 520$  (alternative mean);
- $\sigma = 100$  (standard deviation);
- $\alpha = .05$  (alpha error rate, one tailed);
- $n = 25$  (sample size).

**Press enter/return after placing the new values in the appropriate boxes!**

Do ten simulations of drawing a sample of 25 cases, and record the results below.

	mean	z-score		mean	z-score		mean	z-score		mean	z-score
1	<input type="text"/>	<input type="text"/>	2	<input type="text"/>	<input type="text"/>	3	<input type="text"/>	<input type="text"/>	4	<input type="text"/>	<input type="text"/>
5	<input type="text"/>	<input type="text"/>	6	<input type="text"/>	<input type="text"/>	7	<input type="text"/>	<input type="text"/>	8	<input type="text"/>	<input type="text"/>
9	<input type="text"/>	<input type="text"/>	10	<input type="text"/>	<input type="text"/>						

**1c.** What is the power for this test as shown in the applet? \_\_\_\_\_

**1d.** How many of your ten simulated samples allowed you to reject the null hypothesis? \_\_\_\_\_  
 (Use one-tailed alpha  $\alpha = .05$ ,  $z = 1.645$ , so reject  $H_0$  if your z-score is greater than 1.645)

**1e.** For the ACE program, power was **.991**. Briefly describe your findings from the two simulations and explain how the difference in population means produced the difference in statistical power.

## Exercise 2: Power and Variability (Standard Deviation)

In this Exercise, we will examine the effect of variability on statistical power. If the standard deviation of the VAST test was only 50 instead of 100, do you think would power be greater or less (assume no other change in a population values)? Think about what will happen before you try the simulation.

To simulate drawing a sample from a DEUCE population with a smaller standard deviation, enter the following values into the WISE Power Applet:

- $\mu_0 = 500$  (null mean);
- $\mu_1 = 520$  (alternative mean);
- $\sigma = 50$  (standard deviation);
- $\alpha = .05$  (alpha error rate, one tailed);
- $n = 25$  (sample size).

Press enter/return after placing the new values in the appropriate boxes.

Do ten simulations of drawing a sample of 25 cases and record the results below.

	mean	z-score		mean	z-score		mean	z-score		mean	z-score
1	<input type="text"/>	<input type="text"/>	2	<input type="text"/>	<input type="text"/>	3	<input type="text"/>	<input type="text"/>	4	<input type="text"/>	<input type="text"/>
5	<input type="text"/>	<input type="text"/>	6	<input type="text"/>	<input type="text"/>	7	<input type="text"/>	<input type="text"/>	8	<input type="text"/>	<input type="text"/>
9	<input type="text"/>	<input type="text"/>	10	<input type="text"/>	<input type="text"/>						

2a. What is the power for this test (from the applet)? \_\_\_\_\_

2b. How many of your ten simulated samples allowed you to reject the null hypothesis? \_\_\_\_\_  
 (Use one-tailed alpha  $\alpha = .05$ ,  $z = 1.645$ , so reject  $H_0$  if your z-score is greater than 1.645) \_\_\_\_\_

2c. Below, compare your results from the DEUCE graduates in **Exercise 1** (where the power was .260, and effect size,  $d = .20$ ). Why does a smaller standard deviation lead to greater power?

## Question A: Effect Size and Power

Which of the following situations would yield the greatest power (assuming alpha is held constant)?

- Null mean = 500, Alternative mean = 510, Standard Deviation = 40
- Null mean = 500, Alternative mean = 540, Standard Deviation = 160
- Null mean = 500, Alternative mean = 520, Standard Deviation = 60

### Exercise 3: Power and Sample Size

3. In this exercise we will examine the effect of sample size on statistical power. If we drew a sample of 100 graduates from the DEUCE program rather than a sample of 25 graduates, do you think would power be greater or less (assume no other change in a population values)? Think about what will happen before you try the simulation.

To simulate drawing a larger sample, enter the following values into the WISE Power Applet:

- $\mu_0 = 500$  (null mean);
- $\mu_1 = 520$  (alternative mean);
- $\sigma = 100$  (standard deviation);
- $\alpha = .05$  (alpha error rate, one tailed);
- $n = 100$  (sample size).

Press enter/return after placing the new values in the appropriate boxes.

Do ten simulations of drawing a sample of 100 cases and record the results below.

	mean	z-score		mean	z-score		mean	z-score		mean	z-score
1	<input type="text"/>	<input type="text"/>	2	<input type="text"/>	<input type="text"/>	3	<input type="text"/>	<input type="text"/>	4	<input type="text"/>	<input type="text"/>
5	<input type="text"/>	<input type="text"/>	6	<input type="text"/>	<input type="text"/>	7	<input type="text"/>	<input type="text"/>	8	<input type="text"/>	<input type="text"/>
9	<input type="text"/>	<input type="text"/>	10	<input type="text"/>	<input type="text"/>						

3a. What is the power for this test? \_\_\_\_\_

Now change  $n$  to 4. Press **enter** on your keyboard. Do ten simulations with samples of size 4.

	mean	z-score		mean	z-score		mean	z-score		mean	z-score
1	<input type="text"/>	<input type="text"/>	2	<input type="text"/>	<input type="text"/>	3	<input type="text"/>	<input type="text"/>	4	<input type="text"/>	<input type="text"/>
5	<input type="text"/>	<input type="text"/>	6	<input type="text"/>	<input type="text"/>	7	<input type="text"/>	<input type="text"/>	8	<input type="text"/>	<input type="text"/>
9	<input type="text"/>	<input type="text"/>	10	<input type="text"/>	<input type="text"/>						

3b. What is the power for this test? \_\_\_\_\_

3c. How many times could you reject the null hypothesis using  $\alpha = .05$  one-tailed ( $z = 1.645$ ) for:

$n = 4$ : \_\_\_\_\_

$n = 100$ : \_\_\_\_\_

3d. What do you conclude about the effect of sample size on power? How is sample size related to effect size? Why?

### Question B: The Impact of Sample Size

Consider the shape of the sampling distributions for samples of size  $n = 4$ ,  $n = 25$ , and  $n = 100$ . What happens to the sampling distribution of the sample mean when  $n$  rises?

- Sampling distribution gets more disperse.
- Sampling distribution gets less disperse.
- Sampling distribution remains the same.

### Exercise 4: Power and Alpha

Now, we will consider the impact of using a different alpha value.

As the researcher, we decide on the value of alpha, typically at .05 or .01. Alpha is the error rate we are willing to accept for the error of rejecting the null hypothesis if it were true. We require stronger evidence to reject the null hypothesis if we set alpha at .01 than if we use alpha of .05.

4. For this example, use one-tailed alpha  $\alpha = .01$  ( $z = 2.326$ ). In this case, we will reject the null hypothesis only if a sample mean is so large that it would occur less than 1% of the time given the null hypothesis is true. You do not need to draw additional samples for this problem; you can use the data recorded for samples drawn in **Exercise 1** ( $\mu_0 = 500$ ,  $\sigma = 100$ ,  $n = 25$ ,  $\alpha = .05$ ,  $z = 1.645$ ).

4a. Using these criteria, how many times could you reject the null hypothesis for your results in **Exercise 1**?

	$\alpha = .05$ (from #1)	$\alpha = .01$
Reject for ACE Program ( $\mu_1 = 580$ )		
Reject for DEUCE Program ( $\mu_1 = 520$ )		

**4b.** Using these criteria, what is the power for each of these tests? You will need to use the applet below to calculate power for the tests using alpha  $\alpha = .01$ .

	$\alpha = .05$ (from #1)	$\alpha = .01$
Power for ACE Program ( $\mu_1 = 580$ )	.991	
Power for DEUCE Program ( $\mu_1 = 520$ )		

You may also examine the effects of changing alpha in the WISE Power Applet.

**4c.** Does power rise or fall using alpha = .01 compared to .05? Why?

### Question C: What Affects Power?

So far you have examined the effect of magnitude of difference between the null mean and the alternative mean, standard deviation, sample size, and alpha level on power. Which of the answers below best summarizes the effect of each on power?

- More power = large magnitude of difference, larger standard deviation, larger sample, larger alpha.
- More power = large magnitude of difference, smaller standard deviation, larger sample, smaller alpha.
- More power = large magnitude of difference, smaller standard deviation, larger sample, larger alpha.
- More power = smaller magnitude of difference, smaller standard deviation, larger sample, smaller alpha.