Web Activity: Power of a One-Sample *t*-Test

In this activity we will develop an intuition about the *power* of a hypothesis test, as well as the relationship between power and significance level, effect size, and sample size, in the context of a one-sample, one-sided *t*-test. The *power* of a test is defined as the probability of rejecting the null hypothesis when the null hypothesis is indeed false. Thus, high values of power are desirable. Power depends on multiple factors, such as the true population mean, the sample size, and the variance of the population.

Recall that a *t*-test is a hypothesis test in which the test statistic follows a *t*-distribution. In this activity you will be testing whether the mean of a Normally-distributed population is equal to a specific value μ_0 or larger than it (that is, the test is one-sided). The value μ_0 is determined by the experimenter before collecting data. For this test, the test statistic is defined as

$$t=\frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}},$$

where $\bar{x}$ is the sample mean, s^2 is the sample variance, and n is the sample size. Since the test is one-sided, we reject the null hypothesis if the *t*-statistic is too large. In particular, we reject H_0 if t is larger than a *critical value* that depends on the significance level and the sample size. You will use an online interactive resource to visualize how the power of a one-sample *t*-test changes with respect to various conditions.

The resource can be accessed at

`https://shiny-apps.stat.ubc.ca/FlexibleLearning/Power.`

The resource allows you to specify the value of μ_0 and the true population mean μ_1 (and thus the difference of these two means). You can also specify the population variance and the significance level. If you specify the sample size, the resource will calculate the power. If you specify the desired power, the resource will calculate the smallest sample size needed to achieve that power.

Part I – difference in means

For this part of the activity, you will use the *Density plots* panel of the resource, accessed via the top of the resource. The curves in the panel show the distribution of the test statistic t under both the null and alternative hypotheses—the latter assuming that μ_1 is the true population mean.

1) Using the controls on the tab on the left side of the resource, set the mean under the null hypothesis μ_0 to be 0 and the true population mean μ_1 to be 1. Set the variance to 1, the sample size to 10, and the significance level to 0.05.
   a) For these settings, what is the null hypothesis? What is the alternative hypothesis?
   b) We reject the null hypothesis if the *t*-statistic is larger than a critical value. What is this critical value, according to the resource?
   c) How is the significance level $\alpha = 0.05$ represented in the top plot?
   d) What is the power of the *t*-test when the population mean is equal to 1? How is this value represented in the bottom plot?
      (Hint: the power is displayed on top of the blue area in the lower curve of the panel.)
2) What do you think will happen to the power if the sample comes from a population with mean $\mu_1 = 0.5$? Don't use the resource yet—just think about this.
3) Now use the resource. Suppose the true population mean is $\mu_1 = 0.5$ and modify the μ_1 setting on the left tab to reflect this. Leave the other values fixed. Did the power of the test decrease, stay the same, or increase? Does this match what you expected to see in question 2)?
4) Suppose that you now want to test the null hypothesis that the population mean is -0.5. Modify the μ_0 setting on the left tab to reflect this and leave the other values fixed. Compare the power of the test to the power of the test in question 3). Is it smaller, equal, or larger? Why do you think this is so?
5) Suppose that you now want to test again the null hypothesis that the population mean is 0. Repeat the exercise in question 3) for the following values of the true population mean μ_1: 0.2, 0.8, 1.2, and 1.5. Write down the power of the test for each μ_1 value. Based on this exercise, you can conclude that for larger differences between μ_0 and μ_1 the power of the test will be _________.
   (Fill in the blank with one of smaller, unchanged, or larger.)

Part II – variance

For this part of the activity, you will use the *Density plots* panel of the resource again.

1) Set the mean under the null hypothesis μ_0 to be 0 and the true population mean μ_1 to be 1. Set the variance to 1, the sample size to 10, and the significance level to 0.05. Write down the power of the test.
2) What do you think will happen to the power of the test if the sample came from a population with variance 1.5? Don't use the resource yet—just think about this.
3) Now use the resource. Suppose the sample does come from a population with variance 1.5. Modify the σ^2 setting on the left tab to reflect this and leave the other values fixed. Did the power of the test decrease, stay the same, or increase? Does this match what you expected to see in question 2)?

4) Repeat the exercise in question 3) for the following values of the variance: 0.5, 2, 5, and 10. Write down the power of the test for each value of the variance. Based on this exercise, you can conclude that the larger the variance of the population the power of the test will be _________. *(Fill in the blank with one of smaller, unchanged, or larger.)*

Part III – effect size

For this part of the activity, you will use the *Density plots* panel of the resource again.

You sell 0.4 gallon containers of milk and you suspect that your employees are overfilling containers. You randomly choose 10 milk containers and measure the volume of each. You want to test the null hypothesis that the average volume is indeed 0.4 gallons versus the alternative hypothesis that the average volume is larger than 0.4 gallons. You would like to test with a significance of 0.2. Furthermore, you know that the variability of the filling machines means that the variance of the milk in your factory's containers is 0.1 gallons^2.

1) Modify the settings on the left tab of the resource to reflect this test.
2) Assuming that the true average volume of the containers is actually 0.5 gallons, what is the difference between the true mean μ_1 and the mean under the null μ_0? What is the power of the test?
3) In the lower-most part of the left tab a value called *effect* size is printed. Write down this value.
4) Now assume you instead measure the volume in litres instead of gallons. 0.4 gallons are approximately 1.5L, and 0.5 gallons are approximately 1.9L. Modify the μ_0 and μ_1 settings on the left tab to reflect this. Furthermore, if the variance in gallons is 0.1 gallons^2 then, by variance rules, the variance in litres is approximately 1.6L^2. Modify the σ^2 in the left tab to reflect this.
5) Do you expect the power of this test to be smaller than, equal to, or larger than the one in question 2)?
6) With the settings in question 4), what is the difference between the true mean μ_1 and the mean under the null μ_0? What is the power of the test?
7) Does the answer in question 6) match what you expected to see in question 5)?
8) Observe the *effect size* value printed in the left tab. Is this value smaller than, equal to, or larger than the one in question 3)?

As you saw, both tests have the same power despite having different variances and differences between μ_0 and μ_1. The reason is that, when dealing with different units, the variance accounts for the change in units. Due to this, sometimes it is useful to work with the *effect size*, a "units-free" measure that combines both difference in means and variance.

In this activity, the *effect* size is obtained by first calculating the difference between the mean under the null and the true population mean studied in Part I and then dividing that difference by the square root of the variance, that is, by the standard deviation, of the population.

1) What do you think is the relationship between each of effect size and variance, effect size and difference in means, and effect size and power?

2) Set the mean under the null hypothesis μ_0 to be 0 and the true population mean μ_1 to be 1. Set the variance to 1, the sample size to 10, and the significance level to 0.05. Write down the effect size and the power of the test.
3) Suppose the population variance is now 2 and modify the left tab to reflect this. Leave the other values fixed. What is the new value of the effect size? Did the power decrease, stay the same, or increase?
4) Now suppose the population variance is 0.5 and modify the left tab to reflect this. What is the new value of the effect size? Did the power decrease, stay the same, or increase?
5) Based on questions 3) and 4), what is the relationship between variance and effect size? Does this match what you expected to see in question 1)?
6) Now return the variance setting to 1 and suppose the true value of the population mean is 0.5. What is the new value of the effect size? Did the power decrease, stay the same, or increase compared to the test in question 2)?
7) If the true population mean is instead 2, what is the value of the effect size? Did the power decrease, stay the same, or increase?
8) Based on questions 6) and 7), what is the relationship between difference in means and effect size? Does this match what you expected to see in question 1)?
9) Based on questions 2) through 8), what is the relationship between effect size and power? Does this match what you expected to see in question 1)?

Part IV – sample size

For this part of the activity, you will use the *Density plots* panel of the resource again.

1) Set the mean under the null hypothesis μ_0 to be 0 and the true population mean μ_1 to be 1. Set the variance to 1, the sample size to 10, and the significance level to 0.05. Write down the power of the test.
2) What do you think will happen to the power of the test if the sample size were 5 instead?
3) Now suppose the sample size is 5 and modify the appropriate setting on the left tab to reflect this. Leave the other values fixed. Did the power of the test decrease, stay the same, or increase? Does this match what you expected to see in question 2)?
4) Repeat the exercise in question 3) for the following values of the sample size: 4, 8, 15, and 20. Write down the power of the test for each value of sample size. Based on this exercise, you can conclude that the larger the sample size the power of the test will be __________.
   (Fill in the blank with one of <u>smaller</u>, <u>unchanged</u>, or <u>larger</u>.)

Part V – significance level

For this part of the activity, you will use the *Density plots* panel of the resource again.

1) Set the mean under the null hypothesis μ_0 to be 0 and the true population mean μ_1 to be 1. Set the variance to 1, the sample size to 10, and the significance level to 0.05. Write down the power of the test.
2) What do you think will happen to the power of the test if the significance level of the test were 0.03?

3) Now suppose you want to test at a significance level of 0.03 and modify the appropriate setting on the left tab to reflect this. Leave the other values fixed. Did the power of the test decrease, stay the same, or increase? Does this match what you expected to see in question 2)?
4) Repeat the exercise in question 3) for the following values of significance level: 0.01, 0.1, 0.2, and 0.5. Write down the power of the test for each significance level. Based on this exercise, you can conclude that the larger the significance level of the test the power of the test will be _________.
   (Fill in the blank with one of smaller*,* unchanged*, or* larger*.)*

Part VI – sample size calculation in practice

For this part of the activity, you will first use the *Density plots* panel of the resource again.

A researcher will carry out an experiment to collect data, and afterwards she will perform a one-sided *t*-test to test the null hypothesis that $\mu=0$ against $\mu>0$. Standards in her research area dictate that a significance level of 0.05 should be used. She believes that the population variance is 1.1 and that the true population mean is $\mu_1=0.5$. Furthermore, she would like to reject the null hypothesis with probability of at least 0.8 if her guess of $\mu_1=0.5$ is correct. You are going to help the researcher determine the size of the sample that she needs to collect.

1) What power does the researcher want her test to have if the true population mean is 0.5?
2) Modify the controls on the left tab to reflect the test that the researcher wants to perform. For now, select a sample size of 10. What is the power of this test? Does this satisfy the requirement of question 1)?
3) What will the power of the test be if the sample is of size 20 instead? Does this power satisfy the power requirement of question 1)?
4) What will the power of the test be if the sample is of size 30? Does this power satisfy the power requirement of question 1)?
5) What is the *smallest* sample size such that the resulting test satisfies the power requirement of question 1)? This is the sample size you would recommend to the researcher.

The procedure of determining the smallest sample size that guarantees a desired power given a significance level, μ_0, μ_1, and σ^2 is already implemented in the ***Sample size calculator*** panel. Go to that panel now. Notice that the sample size control is no longer available, and that a new control titled *Desired power* is included at the top of the panel. The curve in the panel shows the relationship between power and sample size.

1) Using the controls on the left side of the resource, set the mean under the null hypothesis μ_0 to be 0 and the true population mean μ_1 to be 0.5. Set the variance to 1.1, the significance level to 0.05, and the desired power to 0.8. What is the smallest sample size that guarantees the desired power? Does this match the previous exercise?
   (*Hint: the required sample size is displayed on top of the plot.*)
2) Now suppose that the true population mean is instead 1. What would be the smallest sample size be in order to achieve the same power of 0.8? Is this value the same as the one in question 1)? Why do you think this is so?
3) With the same configuration as in question 1), suppose the researcher wants a test with a power of 0.9.

a) What do you think the required sample size would be compared to the one in question 1)?
b) Increase the desired power to 0.9 and write down the required sample size. Does it match what you expected to see in question 3)?

4) With the same configuration as in question 1), suppose the test is instead carried out at a significance level of 0.01.
   a) What do you think the required sample size would be compared to the one in question 1)?
   b) Decrease the significance level to 0.01 and write down the required sample size. Does it match what you expected to see in question 5)?
5) Based on questions 1) through 6), what is the impact of each of difference in means, desired power, and significance on the required sample size?