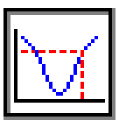
CHAPTER 11



Visualizing Power and Type I / Type II Error

CONCEPTS

- Power, Type I Error, Type II Error, β Level, Power Curve, α Level, Null Hypothesis, Alternative Hypothesis, True Parameter Value

OBJECTIVES

- Understand relationships between Type I error, Type II error, and power
- Recognize the shape of a power curve and how to compare points along the parameter axis
- Understand how a point on the power curve is obtained
- Understand how power is affected by alpha level, sample size, and true parameter values

Overview of Concepts

In hypothesis testing, we compare the **null hypothesis** and the **alternative hypothesis**. These statements often concern a mean (μ) or a proportion (π). For example:

Application	*Hypotheses*
Manufacturing quality assurance	H_0 : wire tensile strength meets or exceeds specification : H_1 : wire tensile strength is below specification
Medical clinic client relations	H_0 : patient complaints are at or below the historical level : H_1 : patient complaint rate has risen above historical level

Testable hypotheses must be mutually exclusive and collectively exhaustive. They may be right-tail, left-tail, or two tail tests. When we make a decision, we seek to avoid both **Type I error** (rejecting a true null hypothesis) and **Type II error** (accepting a false alternative hypothesis). The probability of Type I error is the **α level**. The probability of Type II error is the **β level**. If neither error is committed, we have reached a correct decision.

Even if the null hypothesis is accepted, there is only a probability of $1 - \alpha$ that a correct decision was made. If the null hypothesis is true, the sum of the probabilities of its acceptance or rejection ($1 - \alpha$ plus α) must be 1. Similarly, if the null hypothesis is false, the sum of the probabilities of its acceptance or rejection (β plus $1 - \beta$) must be 1. Just as there are two types of *incorrect* decisions, there are two ways a *correct* decision can be made: a true null hypothesis can be correctly accepted, or a false null hypothesis can be rejected. The probability of correctly rejecting the null hypothesis is called **power**.

		True State	
		Null Hypothesis is True	*Null Hypothesis is False*
Decision	*Accept H_0*	Correct Decision $1 - \alpha$	Type II Error β
	Reject H_0	Type I Error α	Correct Decision (Power) $1 - \beta$

A statistical test that has a high probability of correctly rejecting a false H_0 (high power) is a desirable test. When the power is calculated for all possible true values of the true parameter and plotted on a graph, a **power curve** is generated. Since the **true parameter value** is unknown (otherwise why would we be testing a hypothesis?) there is no way of knowing which error, if any, we have committed when we make our decision. But we can define the relationships among α, β, and power. Power is lowest (near α) when the true parameter differs only slightly from the value hypothesized in H_0, and rises toward unity as H_0 becomes more obviously false. If we increase α, the power curve shifts up at all parameter values. Increasing the sample size increases power at every level of the true parameter (though only slightly near H_0). In testing a mean, a smaller standard deviation σ will improve power.

Illustration of Concepts

The diameter of a piston is designed to be 100.12 mm. The assembly line should be stopped if the piston is either too large or too small. In this case the **null hypothesis** and **alternative hypothesis** would be

$$H_0: \mu = 100.12 \text{ mm}$$
$$H_1: \mu \neq 100.12 \text{ mm}$$

Unless there is enough evidence to refute the null hypothesis, it is assumed to be true. If the diameter is not 100.12 mm and the hypothesis is accepted (i.e., not rejected) then an error has been made. This is **Type II error** and its probability is the **β level**. (the probability of incorrectly accepting the null hypothesis). In the example Type II error would mean that the piston is either too small (insufficient compression) or too large (excessive friction) leading to poor engine performance or premature piston failure. If detected during manufacture, rework may be possible, but often the result is scrappage. If the problem is not detected in manufacture, the company will experience extra repair claims (if failure occurs during the warranty period) or loss of consumer goodwill (if failure occurs after the warranty period or goes unreported). Type II error is costly to any company. Because the **true parameter value** is *unknown,* Type II error is generally unknown.

Costs arise if we continue operating an assembly line that is making inferior parts, but **Type I error** is also costly, because it means shutting down the line when adjustment is unnecessary (downtime, lost wages, disrupted schedules). Because the parameter's value under H_0 is *known, the* **α level** can be specified. It is generally set to 0.10, 0.05, or 0.01. In this example, if $\alpha = 0.01$, it would mean that there is 1 chance in 100 that the assembly line would be stopped to realign the equipment when the line is actually producing proper-sized pistons.

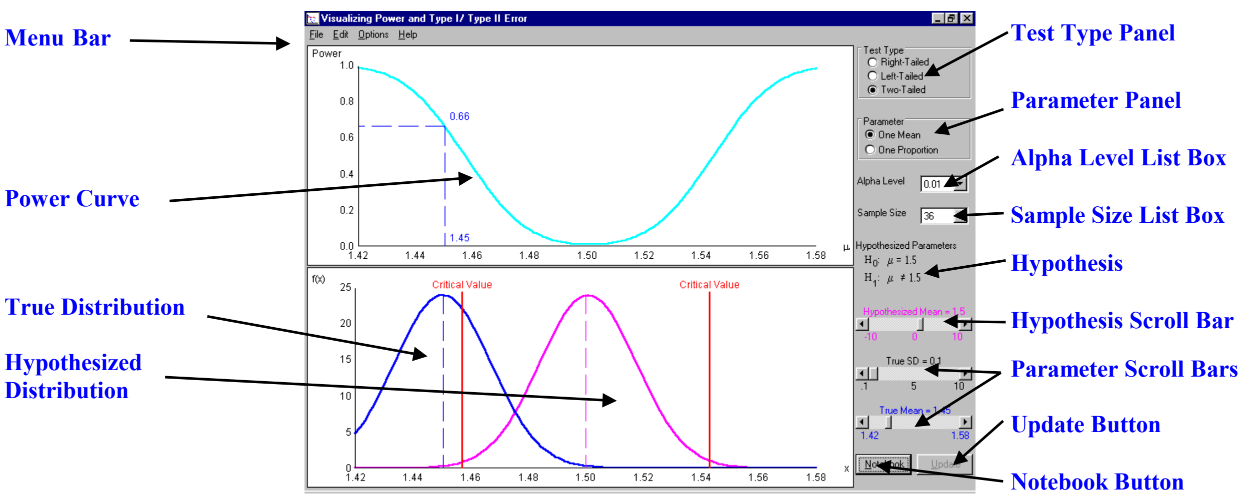
Since the statistician can specify any α that is desired it is natural to suppose that α should always be kept low. But choosing lower α will, *ceteris paribus*, increase β and hence reduce **power** (the converse of β). Saying that low β is preferable is the same as saying that high power is preferable. In testing a mean, there are two ways to reduce both α and β simultaneously: (1) to increase the sample size (but in quality control, the sample size is often kept small), or (2) to decrease the variance of the piston-making process. Manufacturers are well-aware of this latter strategy, and variance reduction is one of their major goals. Low σ will make the product perform more consistently in its intended use, and will lend power to quality tests. Thus, reducing σ is a good way to raise the **power curve** (near the mean specified in H_0 the difference may be small, but slight departures from the specification are intrinsically less damaging).

In many applications, the tradeoff between α and β is elusive. For example, has excessive F.D.A. concern over adverse side effects caused the U.S. to approve potentially useful new drugs too slowly, or has their policy saved consumers from disasters such as Thalidomide, which caused horrifying birth defects in some Western European nations (but not the U.S., which withheld approval)? Responding to criticism of its caution, the FDA has changed some of its testing requirements to allow drug companies to offer apparently effective drugs to the public with warnings that the long-term side effects are still being tested. This policy change has probably been a good one that allows some individuals access to potentially life-saving drugs, even if in the long run it exposes them to unknown risks. At least the individuals may live long enough to see what the long run has to offer.

Orientation to Basic Features

This module enables the user to visualize how Power and Type II error are affected by the alpha level (Type I error), sample size, type of test (two-tailed, left-tailed, or right-tailed), standard deviation, and the value of the true parameter. It also illustrates what a power curve is and how it is generated. These concepts can be shown using either a test for the mean or a test for the proportion.

1. **Opening Screen**
   Start the module by clicking on the module's icon, title, or chapter number in the *Visual Statistics* menu and pressing the Run Module button. When the module is loaded, you will be on the introduction page of the Notebook. Read the questions and then click the Concepts tab to see the concepts that you will learn. Click on the Scenarios tab. Select Mean from the list of choices. Select a scenario, read it, and press OK. A screen similar to the one below will appear. The top left of the screen shows a power curve. The bottom left of the screen shows two distributions. The true distribution is shown in blue and the hypothesized distribution in gray. The module's control panel is on the right. Other module features are controlled on the menu bar at the top of the screen. The amount of power illustrated in the bottom display is shown on the top graph.



2. **Hint**
   A hint message appears on the upper diagram. It says, "Hold down the Alt key and click the left mouse button on the top graph to show power values." After reading the message, press the OK button. Try it. There are several ways you can remove these lines: (1) click on Options on the menu bar at the top of the screen and select Clear Power Values, or (2) press Esc, or (3) press Ctrl-P, or (4) change an aspect of the problem and press the flashing Update button. If you forget what the hint said while you are in the module, select Help and select Show Hint.

3. **True Mean**
   Try changing the scroll bar for the True Mean. Note how the true distribution moves right or left. Each power illustration on the lower diagram represents a single point on the upper diagram. If you hold down the mouse button on the scroll bar (the arrow for small steps, or the bar for large steps), the true distribution will shift continuously until you release it.

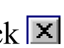
4. **Change the Experiment**
   a. Try changing the scroll bars for the Hypothesized Mean or True SD (standard deviation). Changing either of these controls represents a new experiment, and the power curve is redrawn after pressing the flashing Update button. Similarly, if you change the Sample Size list box or the Alpha Level list box, a new power curve will be redrawn once you press the Update button.
   b. You can also press the Notebook button to return to the Notebook to select a new scenario, read a section of the Notebook, or select the Do-It-Yourself tab. Pressing OK on the Do-It-Yourself section returns you to the module without descriptive labels on the horizontal axis.

5. **Test Types**
   The Test Type panel contains three option buttons: Right-Tailed, Left-Tailed, or Two-Tailed. These options change the null and alternative hypotheses. The power curve and the critical values in the lower diagram are redrawn after you press the Update button.

6. **Options: Graph Labels**
   In addition to clearing power values, several basic graph labeling options are available from the Options menu on the menu bar at the top of the screen.
   a. Click on Show Numeric Critical Value(s) to provide the numeric value for the critical value(s).
   b. Click on Options and select Label Reject/Accept Region to label these regions on the lower graph.
   c. Click on Options and select either Show Type I and Type II Error, or Show Power and Type II Error to reveal these regions as shaded areas on the lower graph.

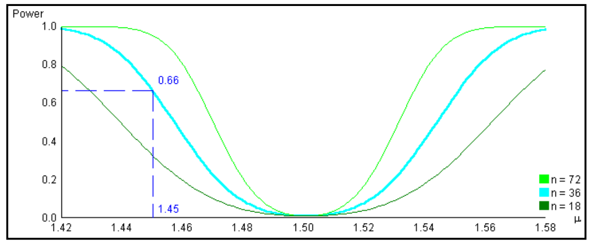
7. **Copying a Graph**
   When you click on a graph, black handles will appear indicating it has been selected (they can be removed by pressing "Esc"). Select Copy under Edit on the menu bar at the top of the screen (or press Ctrl-C) to copy the display. It can then be pasted into other applications, such as Word or WordPerfect, so it can be printed. Graphs are copied as bitmaps

8. **Help**
   Click on Help on the menu bar at the top of the screen. Search for Help lets you search an index for this module, Contents shows a table of contents for this module, Using Help gives instructions on how to use Help, and About gives licensing and copyright information about *Visual Statistics*.

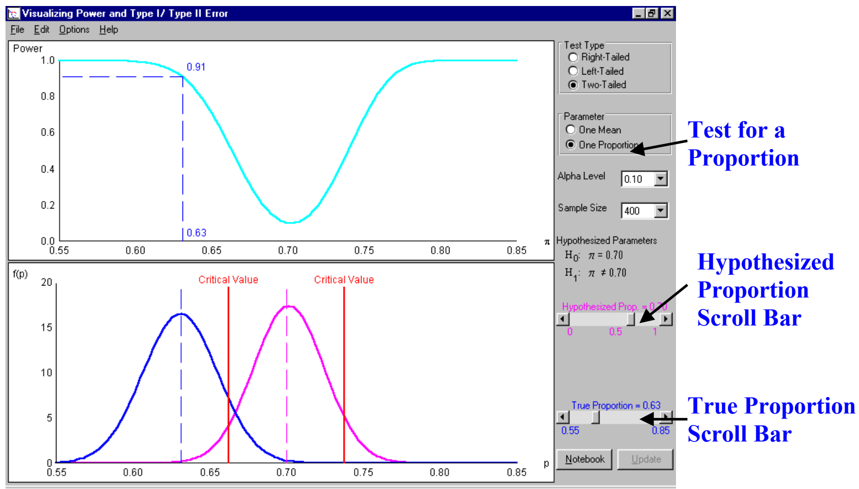
9. **Exit**
   Close the module by selecting Exit in the File menu (or click  in the upper right-hand corner of the window). You will be returned to the *Visual Statistics* main menu.

Orientation to Additional Features

1. **Options: Families of Curves**
   Click on Options and select Families of Power Curves to compare the positions of power curves for different sample sizes, alpha levels, or standard deviations. For example, here are power curves representing 2n (light green), n (cyan), and n/2 (dark green).



2. **Proportions**
   Click on the Notebook button, select the Scenario tab, and click on Proportion. Select a scenario. A screen similar to the one on the right appears. The display illustrates the power of a test for the proportion (π). The Control Panel has changed to reflect a test for a proportion. There no longer is a True SD scroll bar, and the True Mean and Hypothesized Mean scroll bars have been replaced with the True Proportion and Hypothesized Proportion scroll bars. As an alternative to using the Notebook to switch between a test for a proportion and a test for a mean, you can click on the One Mean or One Proportion option buttons in the Parameter panel.



Basic Learning Exercises Name ________________

Reject/Accept Regions

1. Press the Notebook button, select the Scenarios tab, and click on Mean. Select the Quiz Scores scenario, read the scenario, and click OK. What are the null and alternative hypotheses in this scenario if you want a left-tailed test?

2. Select Show Numeric Critical Values from Options on the menu bar. What is the value of the critical value? Explain how it is calculated. **Hint:** Select Contents from Help on the menu bar and click on Power for a Sample Mean under Equations. Read Step 1.

3. On which side of the critical value is the rejection region? **Hint:** Select Options on the menu bar and select Label Rejection Reject/Accept Region.

4. Change the test type from left-tailed to right-tailed by clicking on the Right-Tailed option button and clicking on Update. State the null and alternative hypotheses for this test. On which side of the critical value is the rejection region now?

5. What is the relationship between the rejection region and the direction of the alternate hypothesis?

6. Change the test type from left-tailed to two-tailed by clicking on the Two-Tailed option button and clicking on Update. Why are there now two critical values? Where are the rejection regions in the two-tailed test?

Power and the Power Curve

7. Power is the probability of *correctly* rejecting a null hypothesis that is *false*. Press the Notebook button, select Quiz Scores, read the scenario. Answer the first question after you have pressed the OK button (this resets all of the controls to their original settings). Interpret this answer in the context of this scenario. How is power illustrated on the bottom diagram? Is power in the rejection or acceptance region? **Hint:** Select Options on the menu bar and select Show Power and Type II Error.

8. Answer the second question in the scenario. What is the power if the true mean is 6.6? 6.4? 6.2? 7.0? **Hint:** Click the True Mean scroll bar to change the value of the true mean (or Alt-Click the desired μ values along the X axis of the upper diagram).

9. If you connect these points you have created a power curve. The number on the horizontal axis is the value of the true mean μ. The value on the vertical axis is the power. This graph tells us that if the true value of μ is 6.6, the test will reject the null hypothesis 57% of the time. In this scenario, explain why power *decreases* as the value of the true mean increases from 6.0 to 7.0. Why is the power 0.05 at $\mu = 7.0$? Why is Type II error 0.95 at $\mu = 7.0$?

10. What do you think the power would be when the true mean equals the critical value? Why?

11. Change the test type to Two-Tailed and the True Mean (μ) to 6.6. What is the power? Why is it different from the answer using a one-tailed test (exercise 8)? **Hint:** Select Show Numeric Critical Values from Option on the menu bar and look at your answer to exercise 2.

Intermediate Learning Exercises Name ________________

Alpha Level

12. What is the alpha level of the current test? Alpha level is shown on the lower graph by the area under the null hypothesis (gray distribution) in the rejection region (left of the left critical value and right of the right critical value). Change Alpha Level to 0.01. Notice how this area becomes much smaller. Change True Mean to 6.6. What is the power? **Hint:** Click on Options on the menu bar and select Label Accept/Reject Region.

13. Change Alpha Level to 0.10. What is the power when the true mean μ equals 6.6? Why has the power increased from exercises 11 and 12?

14. *At any value of μ the power increases with an increase in alpha.* This relationship is illustrated by clicking Options on the menu bar, selecting Families of Power Curves, selecting Alpha and clicking OK. This will display three power curves (for $\alpha = 0.01, 0.05$, and 0.10). Use the scroll bar to change the true mean μ to 6.5. Reading from the graph, approximately what is the power if alpha equals 0.01, 0.05, and 0.10?

    $\alpha = 0.01$ ___________ $\alpha = 0.05$ ___________ $\alpha = 0.10$ ___________

15. What rule would you suggest regarding the relationship between power and alpha?

16. Click Families of Power Curves and select None to remove the three power curves. What happens to Type II error when Type I error increases from 0.01 to 0.05 to 0.10? **Hint:** Select Show Type I and Type II Error from Options on the menu bar.

17. What rule would you suggest regarding the relationship between Type I and Type II error?

Sample Size

18. Press the Notebook button, select Battery Voltage, and read the scenario. Answer the first question after you press the OK button. Interpret this answer in the context of the scenario. What does 1.45 represent on the bottom graph? How is power illustrated on the bottom diagram? **Hint:** From the Options menu select Show Power and Type II Error.

19. Click the Notebook button, reread the scenario and press OK. Answer the last question in the scenario by setting Sample Size to 100 (press Update). Use the scroll bar to set the true mean to 1.45 and determine the power.

20. *At any value of μ, power increases with an increase in sample size.* Click Options on the menu bar, select Families of Power Curves, select Sample Size and click OK. This will display three power curves (for a sample size n/2 = 50, n = 100, and 2n = 200). Use the scroll bar to set the true mean to 1.48. Reading from the graph, approximately what is the power when the sample size is 50, 100, and 200?

    n/2 = 50 ____________ n = 100 ____________ 2n = 200 ____________

21. What rule would you suggest regarding the relationship between power and sample size?

22. Click Families of Power Curves and select None to remove the three power curves. What happens to Type II error when sample size is increased from 100 to 200 to 400? **Hint:** Select Show Type I and Type II Error from Options on the menu bar.

23. What rule would you suggest regarding Type II error and sample size?

24. Total error is Type I error plus Type II error. What rule would you suggest regarding the relationship between total error and sample size?

Standard Deviation

25. Press the Notebook button, select Beer Mug Diameter and read the scenario. Answer the first question after you press the OK button. Interpret this answer in the context of the scenario. What does 3.60 represent on the bottom graph? How is power illustrated on the bottom diagram? **Hint:** From the Options menu select Show Power and Type II Error.

26. Click the Notebook button, reread the scenario and press OK. Answer the last question in the scenario by changing the True SD (Standard Deviation) scroll bar to 0.4 and pressing Update. What is the power if $\mu = 3.60$?

27. *At any value of μ, power decreases with an increase in the standard deviation.* This relationship is illustrated by selecting Families of Power Curves and Standard Deviation. This will display three power curves (for a standard deviation equal to $\sigma/2$, σ, or 2σ). Use the scroll bar to set the true mean μ to 3.65. Reading from the graph, approximately what is the power when the standard deviation is 0.2, 0.4, and 0.8?

    $\sigma/2 = 0.2$ ____________ $\sigma = 0.4$ ____________ $2\sigma = 0.8$ ____________

28. What rule would you suggest regarding the relationship between power and standard deviation?

29. Click Families of Power Curves and select None to remove the three power curves. What happens to Type II error when standard deviation is increased from 0.2 to 0.4 to 0.8? **Hint:** Select Show Type I and Type II Error from Options on the menu bar.

    Type II error increases.

30. What rule would you suggest regarding the relationship between Type II error and power? Do you believe that this rule would work with a two-tailed hypothesis about μ? Try it.

Hypothesized Mean

31. Press the Notebook button, select the Do-It-Yourself tab and click OK. The null hypothesis is that $\mu = 2$ against the alternative that $\mu > 2$. Give an example of a situation where a statistician would want to test this null and alternative hypothesis. What is the standard error in this situation? State the critical value and the true mean of the distribution. **Hint:** Click on Options and select Show Numerical Critical Values.

    Example:
    Standard Error = $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ =___________ Critical Value = _______ True Mean = ______

32. Power in this situation is calculated by finding $\Pr(Z > (CV - \mu_1)/\ \sigma_{\bar{x}})$, where CV is the critical value, μ_1 is the true mean, and $\sigma_{\bar{x}}$ is the standard error. Use the numbers from exercise 40 to calculate the Z value. What is the power of the test? **Hint:** A complete derivation of power is given in Help.

33. Since in this case $CV = \mu_0 + Z_{0.05} \times \sigma_{\bar{x}}$, substituting into the probability statement in exercise 41 gives $\Pr(Z > \mu_0 + Z_{0.05} \times \sigma_{\bar{x}} - \mu_1)/\ \sigma_{\bar{x}} = \Pr(Z > (\mu_0 - \mu_1))/\ \sigma_{\bar{x}} + Z_{0.05})$. Use this formula to find the Z value and report the power if the true mean is 2.2 and the hypothesized mean is 1.8? 1.6? 1.4? 2.2? **Hint:** $Z_{0.05} = 1.645$.

Hypothesized Mean = 1.8	CV = _______	Z value = _________	Power = _________
Hypothesized Mean = 1.6	CV = _______	Z value = _________	Power = _________
Hypothesized Mean = 1.4	CV = _______	Z value = _________	Power = _________
Hypothesized Mean = 2.2	CV = _______	Z value = _________	Power = _________

34. a) What rule would you suggest regarding the relationship between power and the difference of the hypothesized mean and the true mean? b) These calculations have been done with a right-tailed test. Does this relation change if it is a left-tailed test? A two-tailed test?

35. Why in exercise 42 is power equal to 0.05 (the alpha level) when the Hypothesized Mean scroll bar is set to 2.2? Instead of power, can you suggest a better label? Explain.

Advanced Learning Exercises Name ________________

Test for Proportion

36. Press the Notebook button, select the Scenarios tab, and click on Proportion. Select the Certification Exam scenario, read the scenario, and click OK. a) State the null and alternative hypotheses in this scenario. b) Answer the three questions in the scenario. c) Set $\alpha = 0.05$ and note the power. What happens to power if you decrease the Sample Size to 200? **Hint:** Press the Notebook button to read the scenario, and then press Cancel to return to the diagram.

Exploring the Dispersion in a Test for a Proportion

37. The rules regarding sample size, alpha level, and the difference between the true and hypothesized proportion are the same when examining the power of a proportion instead of a mean. However, there is no standard deviation scroll bar because the standard deviation of a binomial is always $\sqrt{\frac{\pi\,(1-\pi)}{n}}$. Set the experiment values to the following: Left-Tailed, Alpha Level = 0.05, Sample Size = 100, and Hypothesized Proportion = 0.13. Press Update. Select Show Power and Type II Error from Options on the menu bar. What is the power when the true proportion is 0.12? Briefly describe the true and hypothesized distribution.

38. Click on the left arrow of True Proportion and hold it down. Notice what happens to the true distribution relative to the hypothesized distribution. Stop at 0.04; what is the power? Briefly describe the true and hypothesized distribution.

39. The hypothesized distribution has not changed in exercises 37 and 38, yet the true distribution has changed dramatically. The mean of the true distribution is 0.12 in exercise 32 and 0.04 in exercise 33, while the mean of the hypothesized distribution is 0.13 in both cases (the values specified on the respective scroll bars). Calculate the standard error for both the true and hypothesized distribution for both cases. **Hint**: To find the formula for standard error, click Help, select Contents, click on Equation: Power of a Sample Proportion, and read Step 1.

True Distribution	$\pi = 0.12$ ____________	$\pi = 0.04$ ____________
Hypothesized Distribution	$\pi = 0.13$ ____________	$\pi = 0.13$ ____________

Exploring the Normality Assumption in a test for a Proportion

40. Press the Notebook button, select the Scenarios tab, and click on Proportion. Select the Certification Exam scenario, read the scenario, and click OK. Examine the power curve. Then change Hypothesized Proportion to 0.04 by clicking on its scroll bar. Why did the message come up that the proportion estimate may not be normally distributed?

41. Look at the area under the hypothesized distribution to the left of the critical value. Why does the distribution stop at zero and not meet the horizontal axis?

42. Change Hypothesized Proportion to 0.03. Has the hypothesized distribution become less normally distributed? Why do you believe this? Change the hypothesized mean to 0.02; note the placement of the critical value under the normality assumption. Where should the critical value be? What does this mean about the Type I error and power (look at the power curve)?

43. If you changed to a right-tailed test and did the same experiment using hypothesized proportions of 0.96, 0.97, and 0.98, would the same problem occur? Why?

44. Why is it important that $n\,\pi_0 \geq 5$ and $n\,(1-\pi_0) \geq 5$?

Individual Learning Projects

Write a report on one of the three topics listed below. Use the cut-and-paste facilities of the module to place the appropriate graphs in your report.

1. Investigate the relationship between a one-tailed and a two-tailed power curve. Using either a test for a mean or a proportion, explain and illustrate how the power curve for a two-tailed hypothesis test conducted at a 0.10 alpha level is related to the power curves for a right-tailed and left-tailed hypothesis test conducted at a 0.05 alpha level. The comparison should be made at seven points spread along the horizontal axis and common to the two-tailed power curve and one of the one-tailed power curves (be sure that you examine a point slightly greater and slightly less than the hypothesized value). For each comparison, the probability distributions that generated the power curve should be used in your explanation.
2. Investigate how a power curve for a proportion is created. Using a test for a proportion and at least five different values of the true proportion, explain and illustrate how power is calculated. Also calculate the range of the true proportion over which the power curve has a non-zero slope (power less than one). How is this range affected by the sample size and hypothesized value of the proportion? How is it affected by a two-tailed test, left-tailed test, or right-tailed test?
3. Investigate how power is affected by the standard error in a test for μ. Using a test for the mean, explain and illustrate how the standard deviation and the sample size interact in determining the power curve. Use at least six illustrations in your project. The project should culminate in a precise rule involving the relationship between power, sample size, and standard deviation.

Team Learning Projects

Select one of the three projects listed below. In each case, produce a team project that is suitable for an oral presentation. Use presentation software or large poster boards to display your results. Graphs should be large enough for your audience to see. Each team member should be responsible for producing some of the graphs. Ask your instructor if a written report is also expected.

1. A team of two should investigate how a power curve is created. Create your own two-tailed scenario that can be illustrated with the module. On a large poster board, draw the power curve. For each value of the true parameter along this power curve (using the True Mean or True Proportion scroll bar) make a copy of the probability distributions. Use these illustrations to show how power is calculated at each point along your power curve. Explain what a power curve illustrates.
2. A team of two should investigate the peculiarities of a power curve for a proportion. Select a sample size of 100 for a two-tailed test when $\pi = 0.40$. Try all the different true values of π. Notice what happens to the hypothesized and true distributions. Draw the power curve on a poster board using at least seven points (each point should have an illustration of the probability distributions associated with it). Redo the experiment using the same sample size but $\pi = 0.05$. Try all the different true values of π. Notice what happens to the hypothesized and true distributions. Draw the power curve on another poster board using at least 11 points (each point should have an illustration of the probability distributions associated with it). Explain why the hypothesized and true distributions behave so differently in the two examples. Why would this not be true in a test for μ?
3. A team of three should investigate the range of values over which a power curve exists. Using a one-tailed test for the mean, illustrate four different power curves (use two different hypothesized means and, for each hypothesized mean, two different true standard deviations). Each power curve should consist of at least six points (each point should have an illustration of the probability distributions associated with it). Derive a rule for the range of values over which the power curve has a non-zero slope (power is less than one).

Self-Evaluation Quiz

1. Type II error is
   a. the probability of correctly rejecting the null hypothesis.
   b. the probability of correctly accepting the null hypothesis.
   c. the probability of incorrectly rejecting the null hypothesis.
   d. the probability of incorrectly accepting the null hypothesis.
   e. none of the above.

2. Power is
   a. the probability of correctly rejecting the null hypothesis.
   b. the probability of correctly accepting the null hypothesis.
   c. the probability of incorrectly rejecting the null hypothesis.
   d. the probability of incorrectly accepting the null hypothesis.
   e. none of the above.

3. Which of these relationships is correct?
   a. Alpha + Power = 1
   b. Power + Beta = 1
   c. Alpha + Beta = 1
   d. All of the above are true.
   e. None of the above is true.

4. In a hypothesis test for the mean μ, which represents movement along the power curve as opposed to a shift of the power curve?
   a. Using a different α level.
   b. Using a different value for true μ to calculate the power.
   c. Using a different value for the true variance.
   d. Using a different sample size.
   e. Using a different hypothesized value for μ.

5. The power of a statistical test for the mean (with known variance) is affected by
   a. alpha level.
   b. variance of the variable.
   c. sample size.
   d. true value of the mean.
   e. all of the above.

6. Which statement is most nearly correct concerning the level of significance α?
   a. If α is increased from 0.05 to 0.10, the power curve will shift up.
   b. If α is increased from 0.05 to 0.10, the power curve will shift down.
   c. If α is increased from 0.05 to 0.10, the power curve will not shift.
   d. Changing α is a movement along the power curve, not a shift.
   e. None of the above is correct.

7. Which statement is most nearly correct concerning the sample size?
   a. If the sample size is increased, the power curve will shift up.
   b. If the sample size is increased, the power curve will shift down.
   c. If the sample size is increased, the power curve will not shift.
   d. Changing the sample size is a movement along the power curve, not a shift.
   e. None of the above is correct.

8. In conducting a hypothesis test for the mean, doubling the standard deviation
   a. doubles the power.
   b. increases the power.
   c. leaves the power unchanged.
   d. reduces the power.
   e. halves the power.

9. In conducting a hypothesis test for the mean μ, which statement is correct?
   a. If the variance increases, power decreases.
   b. If the true value of μ increases, the test's power changes.
   c. Increasing the sample size increases a test's power.
   d. Increasing the α level will increase the test's power, all other things held constant.
   e. All of the statements are true.

10. A firm wants to know whether at least half of its 617 hourly employees are satisfied with the company's health care plan. A sample of 25 employees shows that 16 are satisfied with the company's health care plan. If the true proportion of satisfied employees is 0.51, the right-tailed test for $\pi > 0.5$ would have
    a. relatively low power.
    b. relatively high power.
    c. relatively low β level.
    d. relatively low α level.
    e. both a and c are correct.

11. In the preceding problem, if the sample size were increased to 100 employees and the alpha level stayed the same,
    a. the β risk would decrease.
    b. the Type I would decrease.
    c. both power and β risk would decrease.
    d. neither Type I error nor β risk would decrease.
    e. we cannot say without knowing the true proportion.

12. Other things equal, in the preceding problem, if the α level is decreased from 0.05 to 0.01,
    a. the β error is decreased.
    b. the power is increased.
    c. the power is decreased.
    d. the β error does not change.
    e. a and b only.

Glossary of Terms

Acceptance region Portion of the hypothesized distribution that is bounded by the critical value(s). A given sample can disprove (but cannot prove) the null hypothesis, so it is more accurate (but awkward) to call it the "non-rejection region." See **Type I error**.

Alpha level Desired probability of Type I error. It is set by the researcher and is denoted α. Typical values are 0.10, 0.05, and 0.01. Other things equal, power increases as α increases.

Alternative hypothesis Denoted H_1, it is the converse of the null hypothesis. For example, H_0: $\mu = 5$ might suggest H_1: $\mu \neq 5$. If the sample evidence contradicts H_0, we would reject H_0 in favor of the alternative hypothesis, H_1. Often, a hypothesis test is motivated by the suspicion that the alternative hypothesis may be correct.

Beta level Denoted β, it is the probability of accepting a false null hypothesis. See **Type II error**.

Critical value Value on the X-axis that defines the rejection region for a decision rule. In a right-tailed test, the critical value defines a right-tail area, and conversely for a left-tailed test. A two-tailed test requires two critical values to define a rejection region in each tail. The critical value is determined by the α level. If H_0 is true, it is unlikely that the test statistic will lie beyond the critical value(s). When this does occur, it is unlikely that the null hypothesis is true.

Hypothesized mean Denoted μ_0, it is the assumed value of the population mean that is specified in the null hypothesis (e.g., H_0: $\mu_0 = 5$).

Hypothesized proportion Denoted π_0 (or p_0), it is the assumed value of the population proportion that is specified in the null hypothesis (e.g., H_0: $\pi_0 = 5$). See **Proportion**.

Mean Population parameter characterizing central tendency. It is the expected value of X and is denoted μ.

Null hypothesis Denoted H_0, it is a maintained statement that we try to reject (for example, H_0: $\mu_0 = 5$). The null hypothesis is not necessarily chosen because we believe it to be true, but rather to serve as a reference point (e.g., do at least half the voters support campaign reform?). If the sample evidence contradicts H_0, the null hypothesis is rejected. Otherwise, it awaits further testing that could disprove it at a later time.

Power Probability of correctly rejecting a null hypothesis that is false. Ideal power would be near 1. Power is $1 - \beta$ (the complement of Type II error). Power is the area under the true distribution that is also in the rejection region.

Power curve Result of calculating power for all possible values of the true parameter and plotting the results on a graph. Its minimum value is α. For a two-tailed test, it resembles a valley between two plateaus. For a left-tailed test it increases from 0 toward 1, going from right to left (and conversely for a right-tailed test).

Proportion Population parameter representing the probability that a randomly chosen population item will have a particular characteristic (e.g., proportion of computer chips that are defective). Some textbooks call it π (following the convention that Greek letters denote population parameters) while others call it p to avoid confusion with the trigonometric constant π (the ratio of a circle's circumference to its diameter).

Rejection region Area under the hypothesized distribution that lies beyond the critical value(s). It is determined by α, the probability of Type I error. The rejection region usually is small, since we usually choose a low value for α. If the test statistic falls within this region, we will reject the null hypothesis. See **Alpha level**.

Sample size Number of observations taken at random from the population. Other things equal, power increases as the sample size increases.

Standard deviation Population parameter characterizing dispersion. It is denoted σ. In a test involving a sample mean, power increases as the standard deviation decreases, other things being equal.

True mean Power is low when the true mean μ is close to the hypothesized mean μ_0, but increases the farther μ is from μ_0 as we move along the power curve. See **Mean**.

True parameter value Specific value of the true mean in a test for μ or the true proportion in a test for π.

True proportion Power is low when the true proportion π is close to the hypothesized proportion π_0, but increases the farther π is from π_0 as we move along the power curve. See **Proportion**.

Type I error Error of rejecting a null hypothesis that is true. The probability of Type I error is denoted α. It is the area under the hypothesized distribution that is beyond the critical value(s). When it is set by the researcher it is called the alpha level.

Type II error Error of accepting a null hypothesis that is false. The probability of Type II error is denoted β (beta). The probability of Type II error is the area under the true distribution that is not in the rejection region. It is also called β risk.

Solutions to Self-Evaluation Quiz

1. d Do Exercises 7–9. Read the Overview of Concepts.
2. a Do Exercise 7. Read the Overview of Concepts.
3. b Do Exercise 9. Read the Overview of Concepts.
4. b Do Exercises 7–9.
5. e Do Exercises 8, 12, 18, 19, 26, and 33.
6. a Do Exercises 12–17.
7. a Do Exercises 17–24.
8. d Do Exercises 25–30.
9. e Do Exercises 12–30.
10. a Do Exercises 9 and 36–37.
11. a Do Exercises 18–24 and 36–37.
12. c Do Exercises 12–16 and 36–37.