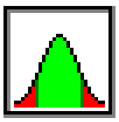
CHAPTER 9



Visualizing One-Sample Hypothesis Tests

CONCEPTS

- Null Hypothesis, Alternative Hypothesis, One-Tail Test, Two-Tail Test, Confidence Interval, Decision Rule, Level of Significance, Sampling Distribution, Critical Value, Test Statistic, P-Value, Type I Error, Type II Error, Power

OBJECTIVES

- Learn to distinguish between one-tail and two-tail null and alternative hypotheses

- Understand the roles of the critical value and test statistic in testing a hypothesis

- Be able to explain the meaning of level of significance and power

- Know the normality assumption and recognize the effects of violating it

Overview of Concepts

Hypothesis testing weighs theory against experimental evidence. Statistical techniques for hypothesis testing were pioneered by Karl Pearson (1857–1936) and R. A. Fisher (1890–1935). Its role in science was explained by Karl R. Popper in his *Logic of Scientific Discovery* (1934). The first step is to create two mutually exclusive hypotheses that exhaust all possibilities (i.e., if one is true, the other must be false). For example, the manufacturer of a CD ROM drive for a certain PC specifies the mean access time for a 2K block of data as 124 milliseconds (ms). If drives are repeatedly tested using randomly chosen 2K blocks of data, the hypotheses are

H_0: $\mu = 124$ ms (access time conforms to the performance standard)
H_1: $\mu \neq 124$ ms (access time fails to conform to the performance standard)

The **null hypothesis** (H_0) is a statement or belief that we try to reject. The **alternative hypothesis** (H_1) is the exact converse of the null hypothesis. Evidence to test the hypothesis is obtained from a sample. If the evidence contradicts H_0, then H_0 will be rejected in favor of H_1. This example depicts a **two-tail test** because a significant deviation in either direction indicates nonconformance to design. If someone (for example, a PC retailer) cares only whether access time is greater than claimed, a **one-tail test** (H_0: $\mu \leq 124$ and H_1: $\mu > 124$) might be useful.

One way to test a hypothesis is to construct a **confidence interval** around a sample estimate to see whether or not it includes the hypothesized parameter value ($\mu = 124$). Another way to test a hypothesis is to compare the sample statistic with the hypothesized parameter value, using our knowledge of the statistic's **sampling distribution**. For a hypothesis about a mean, the sampling distribution of the **test statistic** is normal (for known variance) or Student's t with $n - 1$ degrees of freedom (for unknown variance). The test statistic is compared with a **critical value** based upon the sampling distribution. A **decision rule** is established in advance, stating the criterion for rejecting H_0. If H_0 is not rejected, it is tentatively accepted. Some statisticians prefer to say simply that H_0 is "not rejected" because, although H_0 can be disproved, it *cannot* be proved once and for all.

If H_0 is true and we mistakenly reject it, we have committed **Type I error**, while if a false H_0 is mistakenly accepted we have committed **Type II error**. Otherwise, no error has been committed. The probability of Type I error is denoted α. The probability of Type II error is denoted β. The probability of correctly rejecting a false null hypothesis is called **power** and is equal to $1 - \beta$. The statistician prefers low values for α and β (and hence a high value for power). The value of α (called the **level of significance**) is chosen in advance (typically 0.10, 0.05, or 0.01) and is embodied in the decision rule. The value of β cannot be chosen because it depends on the true parameter value, which is usually unknown. However, power may be found for any possible true parameter value. This module lets you estimate Type I error or power empirically by sampling a known population many times.

Another way of thinking about a hypothesis test is to calculate its **p-value**, which is the probability that a given sample result (or one more extreme) would arise, assuming H_0 is true. The smaller the p-value, the stronger the evidence against H_0. Until recently, p-values were rarely used because they require calculating an exact area under the sampling distribution. Since modern computers can easily calculate such areas, the p-value approach now is widely used. An advantage of the p-value approach is that it avoids the need to specify α in advance (a choice the statistician must make with little basis except tradition).

Illustration of Concepts

To protect baby scallops, the U.S. Fisheries and Wildlife Service requires that in each "harvest" the average meat per scallop shall weigh at least 1/36 pound (12.6 grams). The population is normally distributed with a historic variance of 1.0 grams. A vessel arrives at a Massachusetts port with 11,000 bags of scallops. It is not feasible to weigh each bag, so the harbormaster chooses 18 bags at random. The **null hypothesis** (H_0) is that the average scallop weighs at least 12.6 grams, while the **alternative hypothesis** (H_1) is that this requirement is not met. This is a **one-tail test**, because only underweight scallops are of concern (whereas a **two-tail test** might be used by a biologist who is merely tracking mean scallop weight in general).

H_0: $\mu \geq 12.6$ (average scallop meets the minimum weight requirement)
H_1: $\mu < 12.6$ (average scallop falls short of the minimum weight requirement)

From each bag, a large scoop of scallops is taken and the average meat per scallop is weighed. For the 18 bags that are examined, the sample mean is 12.10 grams. The **sampling distribution** of the **test statistic** for the mean is normal because the population variance is known. Should the null hypothesis be rejected? In the dot plot (Figure 1) the sample mean is a bit below 12.6. The right end of the 95% **confidence interval** (Figure 2) does not include 12.6, which suggests that H_0 should be rejected. However, the decision is fairly close.

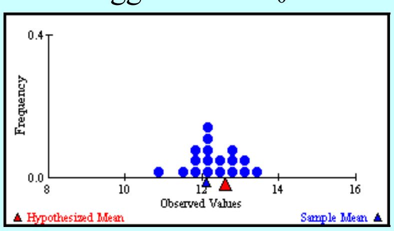


Figure 1: Dot Plot

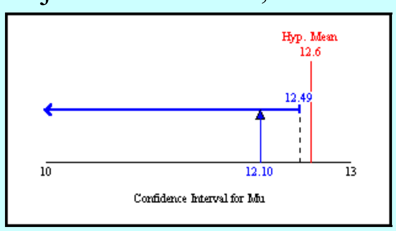


Figure 2: Confidence Interval

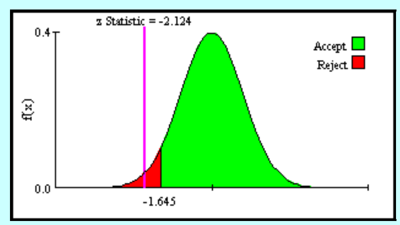


Figure 3: Decision Rule

Another approach is to create a **decision rule**, which will lead to the same conclusion as the confidence interval method. At the 5% **level of significance** we will reject H_0 if the test statistic is to the left of the **critical value** $z_{.05} = -1.645$ (for a normal distribution). Since the test statistic is $z = -2.124$, we reject H_0 (Figure 3). Thus, the captain may face fines for an underweight catch.

Could different samples lead to different results? Suppose this sampling experiment is repeated four more times, using 18 bags each time, with the results summarized below. Samples 1, 3, and 4 lead to rejection (and possible **Type I error**), while samples 2 and 5 lead to acceptance (and possible **Type II error**). The low **p-values** for samples 1, 3, and 4 suggest that these sample means are unlikely if H_0 is true, while samples 2 and 5 are more consistent with H_0. Which decision is correct? We don't know. The **power** of the test is unknown since the true mean is unknown.

Sample	Mean	Upper Limit	Test Statistic	P-Value	Decision
1	12.10	12.49	-2.124	0.016	Reject
2	12.48	12.87	-0.518	0.302	Accept
3	11.55	11.94	-4.450	0.000	Reject
4	11.93	12.32	-2.852	0.020	Reject
5	12.53	12.92	-0.299	0.383	Accept

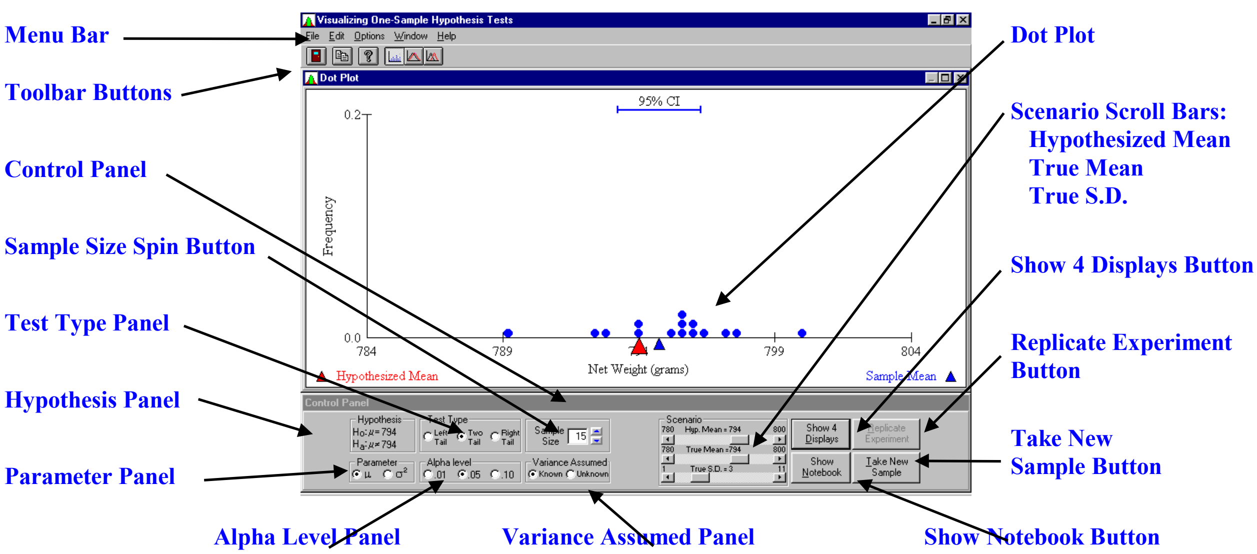
[*] Based on A. Barnett, "Misapplications Reviews: Jail Terms," *Interfaces* 25, 2 (March–April 1995), pp. 18–24.

Orientation to Basic Features

This module helps you learn how to test hypotheses about a single mean (with known or unknown population variance) and about a single variance. It permits one-tail and two-tail tests and shows the relationship between confidence intervals and hypothesis tests using appropriate scenarios. You may also choose a population and specify its parameters. Using replication, you can study the meaning of power, Type I error, and Type II error.

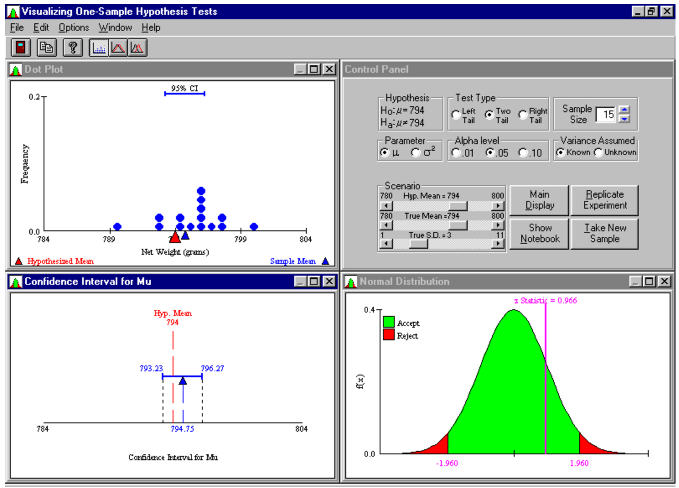
1. **Select a Scenario**
   Start the module by clicking on the module's icon, title, or chapter number in the *Visual Statistics* menu and pressing the Run Module button. When the module is loaded, you will be on the introduction page of the Notebook. Read the questions and then click the Concepts tab to see the concepts that you will learn. Click the Scenarios tab. Click on Tests for Mu with Known Variance. Select a scenario, read it, and press OK.

2. **Main Display**
   The main display opens with a large Dot Plot and a Control Panel. Click Take New Sample and observe the changes in the sampled items (blue dots) and confidence interval (blue line segment labeled 95% CI). The hypothesized mean (red fulcrum) does not change, but the sample mean (blue fulcrum) will change to reflect each new sample. The Replicate button will remain inactive until you press Show 4 Displays (but do not do it yet).



3. **Toolbar Buttons**
   Place the cursor over the fifth toolbar button located above the dot plot. A label will appear (Show True and Hypothesized Distribution). Click the button and see an overlay of the population to be sampled. This distribution does not change when you take samples, but will reflect any changes you make in the true or hypothesized parameters. Click the sixth toolbar button (Show True and Hypothesized Sampling Distribution) and observe the narrower sampling distribution. The sampling distribution will not change when you take samples, but will reflect any changes you make in the sample size or in the true and hypothesized parameters. Click the fourth toolbar button (Show Dot Plot Only) to restore the simple dot plot.

4. **Show 4 Displays**
Click Show 4 Displays. Read the Hint and press OK. The Control Panel will move to the upper right and its controls will be rearranged. The Dot Plot will become smaller and will be in the upper left. Two new quadrant displays will be on the bottom (Confidence Intervals for Mu, Normal Distribution). The Confidence Interval for Mu display shows more detail than the 95% CI on the Dot Plot. The Normal Distribution shows the decision rule. Right-click on the Confidence Intervals for Means display and select Summary Statistics to see a list of sample statistics and parameter values. Right-click on the Summary Statistics display and select Analysis of Experiment to see a verbal description of the results of your experiment. Right-click the Analysis of Experiment display and select Confidence Interval to return to the original quadrant display. Any graph can be enlarged by maximizing its window.



5. **Using the Control Panel**
In the Test Type panel, click each option (Right-Tail, Left-Tail, Two-Tail). The normal distribution display will match the test type you have selected. Click each option (.10, .05, .01) in the Alpha Level panel and verify that the normal distribution tail area and the confidence intervals change to reflect your choices. Click Unknown in the Variance Assumed panel. The normal distribution is replaced by Student's t, and the confidence intervals will change. Click σ^2 in the Parameter panel. The Student's t distribution is replaced by a chi-square distribution and the confidence intervals now show variances instead of means. Click μ in the Parameter panel, and the displays again show a mean. Click Known in the Variance Assumed panel to return to the original scenario. Click the Sample Size spin button to alter the sample size (2 to 99). Click the fifth toolbar button to show an overlay of the population to be sampled. Change the position of any scroll bar in the Scenario panel (Hypothesized Mean, True Mean, True S.D.) and press Take New Sample to see its effect on the distributions.

6. **Copying a Display**
Click on the display you wish to copy. Its window title will be highlighted. Select Copy from the Edit menu (on the menu bar) or Ctrl-C. It can then be pasted into other applications.

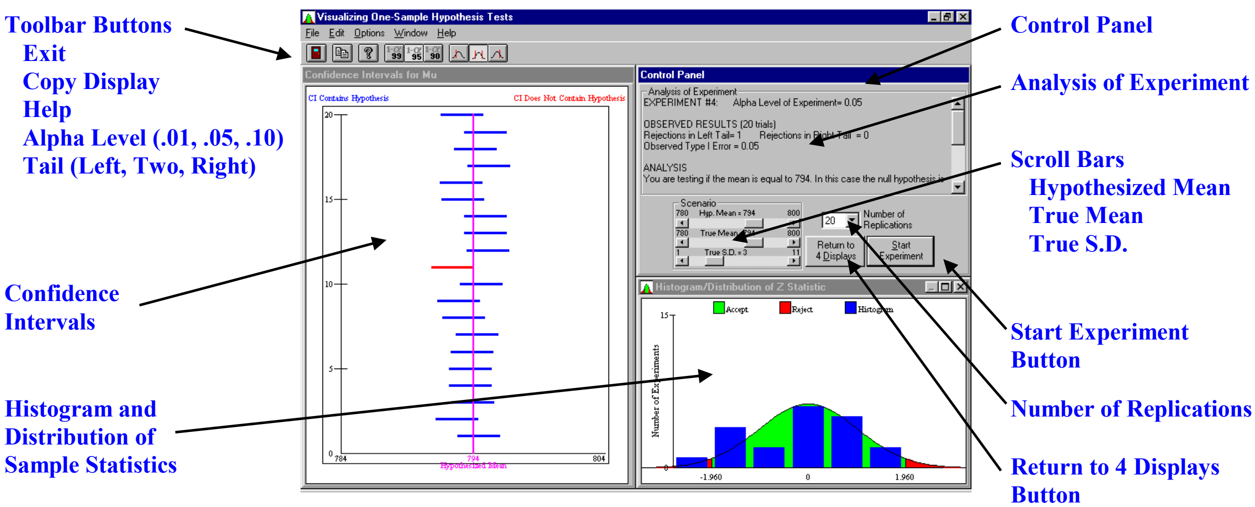
7. **Help**
Click on Help on the menu bar at the top of the screen. Search for Help lets you search an index for this module, Contents shows a table of contents for this module, Using Help gives instructions on how to use Help, and About gives licensing and copyright information.

8. **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

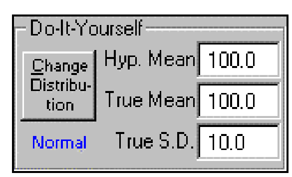
1. **Replication**

   Click the Replicate Experiment button on the Control Panel. At the top are nine toolbar buttons. The new ones control the confidence level (99%, 95%, 90%) and the test type (Left-tail, Two-Tail, Right-Tail). On the left is the Confidence Intervals for Mu display. On the right is a Control Panel containing the Analysis of Experiment window and scroll bar controls (Hypothesized Mean, True Mean, True S.D.) and a Histogram/Distribution display. Click Number of Replications to vary the number of replications (20, 50, 100, 200, 500). If you click the Start Experiment button, it becomes Finish Experiment and Return to 4 Displays becomes Pause Experiment. If you click Pause Experiment, the displays are frozen in their current state and Pause Experiment becomes Continue Experiment. If you click Finish Experiment, the confidence interval and histogram displays go to their final state without showing any intermediate steps. Click Return to 4 Displays to exit replication.



2. **Do-It-Yourself Controls**

   Click Show Notebook, choose the Do-It-Yourself tab, and click OK. Default values for the means and standard deviation are given on the Do-It-Yourself panel. You may change them using the edit boxes. Press Change Distribution and select any of six populations to sample (Normal, Uniform, Skewed Left, Skewed Right, Very Skewed Left, Very Skewed Right). The replication control panel will also show the Do-It-Yourself panel instead of scroll bars.



3. **Options**

   Click Options on the menu bar. You can change terminology from Accept Null Hypothesis to Do Not Reject Null Hypothesis. The latter option may be preferred by some, since we do not actually prove a null hypothesis.

Basic Learning Exercises Name ________________

Press the Show Notebook button, select the Scenarios tab, and click on Tests for Mu with Known Variance. Select the Filling Jars scenario, read it, and press OK.

Tests for a Mean with Known Variance

1. Record the hypothesized mean and true mean that are shown on the control panel. Based on this information, is the null hypothesis true or false?

   Hypothesized mean _____________ True Mean ___________

2. Click Take New Sample 10 times. For each sample, estimate visually the lower and upper end of the range for individual sample items (blue dots) and the sample mean (blue fulcrum shown below the axis). Which shows less variation? Why?

   Smallest sample observation _______ Largest sample observation _______
   Smallest sample mean _______ Largest sample mean _______

3. Click the toolbar button for Show True and Hypothesized Distribution, and record its approximate end points (where it nearly reaches the X-axis). Then click the toolbar button for Show True and Hypothesized Sampling Distribution, and record its approximate end points. Do they agree with your answers in the previous exercise? Explain.

   Minimum of X distribution ______ Maximum of X distribution _____
   Minimum of sampling dist. ______ Maximum of sampling dist. _____

4. The decision rule is in the lower right. a) Why does the decision rule use a normal distribution? b) What are the critical values? c) Do the critical values change when you press Take New Sample? d) Does the test statistic change when you press Take New Sample?

5. Click Show 4 Displays. Click Take New Sample 20 times, and observe the confidence interval display in the lower left each time to see whether it includes the true mean. What percentage of the time did the confidence interval include the true mean? Does this agree with what you would expect, given the level of significance of this two-tailed hypothesis test?

6. Click Take New Sample. Compare the decision about H_0 using the confidence interval (does the interval exclude the hypothesized mean?) and the normal decision rule (does the test statistic fall in the rejection region?). Repeat. Do the decisions always agree?

Tests for a Mean with Unknown Variance

7. Click Return to 4 Displays and Press the Show Notebook button. Select the Scenarios tab and click on Tests for Mu with Unknown Variance. Select the GRE Test Scores scenario, read it, and click OK. Record the hypothesized mean, true mean, true standard deviation, and sample size. Based on these control panel settings, is the null hypothesis true or false?

Hypothesized mean	____________	True Std. Dev.	___________
True mean	____________	Sample size	___________

8. a) Change Variance Assumed from Unknown to Known and watch what happens to the displays. Does changing this assumption of unknown population variances have a substantial impact on the decision rule and its critical values? b) Would your answer be the same if the sample size were larger? Explain.

9. Be sure the Confidence Interval display is showing. Set Variance Assumed to Known. a) Press Take New Sample several times, each time assessing the width of the confidence interval for μ. Does the width vary from sample to sample? b) Set the variance assumption to Unknown and press Take New Sample several times. Does the width vary from sample to sample? Explain. c) Do statisticians have a choice about variances in actual sampling?

Intermediate Learning Exercises Name ________________

Press the Show Notebook button, select the Scenarios tab, and click on Tests for Mu with Known Variance. Select the Filling Jars scenario, read it, and press OK. Make sure all four displays are showing.

Type I/Type II Error and Power with Known Variance

10. Press Replicate. Set the number of replications to 100 and press Start Experiment. If the confidence interval fails to include the true value (shown in red), this indicates rejection of the null hypothesis (μ = 794 grams). How many times did this occur? Repeat this experiment nine more times, entering the results below. What is the mean number of rejections? The range? Why does the number of rejections vary? Does it vary more than you expected?

Experiment:	1	2	3	4	5	6	7	8	9	10
Rejections:	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____

11. Explain what these findings say about Type I error in the context of the jar-filling scenario. If the level of significance were reduced, would that be helpful to the company? Explain.

12. Based purely on the samples, should the assembly line be shut down? **Hint:** Click Show Notebook and review the question posed in the scenario if necessary.

13. Press Replicate. Set the number of replications to 100 and press Start Experiment. Examine the Histogram/Distribution for your 100 replications. Describe its appearance. What does the Analysis of Experiment window tell you about empirical Type I error?

14. Move the True Mean scroll bar four clicks to the left (True Mean = 790 grams). Replicate the sampling experiment 100 times. Look at all three exhibits (Confidence Intervals, Analysis of Experiment, Histogram/Distribution) and describe how these results differ from the previous replication experiments. Would you say it is easy or difficult to detect a difference of 4 grams in this particular scenario? Is it safe to generalize your conclusion about the difficulty of detecting a difference of 0.1?

15. Click Return to 4 Displays and click the Show True and Hypothesized Sampling Distributions toolbar button. How do they help explain the results of exercise 11?

16. Set the True S.D. scroll bar to its largest value (σ = 11 grams) but leave Hypothesized Mean at 794 and True Mean at 790. Click the Take New Sample button. Describe the hypothesized and true sampling distribution. What is the implication for detecting a difference of 4 between the true and hypothesized means in this scenario?

17. Press Replicate. Set the number of replications to 100 and press Start Experiment. Look at all three displays (Confidence Intervals, Analysis of Experiment, Histogram/Distribution) and describe the results of your replication experiment. a) In this scenario (with σ = 11 grams) how often was the difference of 4 grams detected? b) What is the approximate power of the test? c) Why is Type I error not relevant in this case?

Type I/Type II Error and Power with Unknown Variance

18. Set Variance Assumed to Unknown (as in the original scenario) and press Replicate. Choose 500 replications. What is the empirical power? What is the empirical Type II error? What does the histogram of test statistics tell you? Interpret these results.

19. Click Return to 4 Displays, change the sample size to 25, and repeat exercise 18. Interpret your results. Does a larger sample size help?

20. Based purely on the sample evidence, would you conclude that the students' average GRE scores differ from the national average?

Tests for a Variance

21. Press the Show Notebook button, select the Scenarios tab, and click on Tests for Variance. Select the Vending Machine scenario, read it, and press OK. Record the hypothesized variance, true variance, sample size, and test type (left-tail, right-tail, two-tail) that are shown in the control panel. Based on the position of the scroll bars, is the null hypothesis true or false?

    Hypothesized variance _____________ True variance _____________
    Sample size _____________ Test type _____________

22. a) Why is the chi-square distribution the relevant distribution? How many degrees of freedom (DF) does it have, and why? What is the critical value? Why is this a right-tail test? Why is the mean of secondary importance in this scenario? **Hint:** Right-click to see the Statistics and Parameters or the Analysis of Experiment window if you need further information.

    Degrees of freedom _____________ Critical value _____________

23. If it is not already displayed, bring up the Confidence Interval window. Why does this confidence interval extend all the way to the right? Is this true for a one-tailed test of a mean?

24. Press Take New Sample. Should H_0 be rejected? That is, did the confidence interval exclude the hypothesized value of the variance or did the test statistic fall in the rejection region of the chi-square distribution? Do these two approaches to hypothesis testing agree? Press Take New Sample several times. Did the two approaches always agree?

25. Press Replicate. Set Number of Replications to 200 and press Start Experiment. What is the empirical power for your experiment? Why is Type I error not relevant?

26. Press Replicate. Set Number of Replications to 500 and press Start Experiment. What is the empirical power for your experiment?

27. a) Does the number of replications affect the empirical power? b) Which experiment (200 replications or 500 replications) would give the best estimate of empirical power? c) Why do you suppose the maximum number of replications was limited to 500? **Hint:** Look at the display of confidence intervals.

Advanced Learning Exercises

Name ________________

Press the Show Notebook button, select the Do-It-Yourself tab, and click OK. Accept the default parameters ($\mu_0 = 100$, $\mu = 100$, $\sigma = 10$) with a normal distribution. Use a right-tail test for one mean at $\alpha = 0.05$ for a normal distribution with known variance. Make sure all four displays are showing.

Tests of a Mean: Non-Normality and Type I Error

28. Set Sample Size to n = 4, press Replicate, choose 100 replications, press Start Experiment, and then Finish Experiment. Record the empirical Type I error (from the Analysis of Experiment window). Repeat twice and take the average. Compare the *shape* of the histogram of test statistics with its hypothesized normal sampling distribution. Use the Change Distribution button to do similar experiments for Uniform, Skewed Right, and Very Skewed Right populations. Briefly summarize your findings about the effects of population type on the shape of the histogram and on Type I error. How much do your results vary?

Distribution	Experiment 1	Experiment 2	Experiment 3	Avg Type I Error
Normal	________	________	________	________
Uniform	________	________	________	________
Skewed Right	________	________	________	________
Very Skewed R	________	________	________	________

29. Press Return to 4 Displays, set Sample Size to n = 25, and repeat exercise 28. What is the effect of a larger sample size on the *shape* of the histogram of test statistics and on the risk of Type I error when the population is non-normal? What do your results reveal about robustness of the sample mean as an estimator of the true mean? Do results vary less than in the previous experiment? Discuss. **Hint:** See the Glossary definition of robustness.

Distribution	Experiment 1	Experiment 2	Experiment 3	Avg Type I Error
Normal	________	________	________	________
Uniform	________	________	________	________
Skewed Right	________	________	________	________
Very Skewed R	________	________	________	________

Tests of a Mean: Non-Normality and Power

30. Press Return to 4 Displays. To create a false hypothesis, reduce the true mean to 96 and leave everything else the same ($\mu_0 = 100$, $\mu = 96$, $\sigma = 10$, two-tail test for one mean, $\alpha = 0.05$, known variance). Set Sample Size to n = 4 and click the Show True and Hypothesized Sampling Distributions toolbar button. From the appearance of these two distributions, do you expect the test of means to have high power? Explain.

31. Press Replicate and do three experiments of 100 replications for normal, uniform, skewed right, and very skewed right populations. Record the empirical power (from the Analysis of Experiment window) and take the average. Note the *position* of each histogram of test statistics in comparison with the hypothesized sampling distribution. Briefly summarize your findings about the effects of population shape on the histogram and empirical power.

Distribution	Experiment 1	Experiment 2	Experiment 3	Avg Power
Normal	________	________	________	________
Uniform	________	________	________	________
Skewed Right	________	________	________	________
Very Skewed R	________	________	________	________

32. Press Return to 4 Displays, set Sample Size to n = 25, and repeat exercise 31. What is the effect of increased sample size on power? Why? How much did the experiments vary?

Distribution	Experiment 1	Experiment 2	Experiment 3	Avg Power
Normal	________	________	________	________
Uniform	________	________	________	________
Skewed Right	________	________	________	________
Very Skewed R	________	________	________	________

Tests of a Variance: Skewness and Type I Error

33. Change Parameter to σ^2. Use the Do-It-Yourself controls to create a two-tailed test of a true null hypothesis with $\alpha = 0.05$, hypothesized variance of 100, true variance of 100, and sample size of 99 (the mean is irrelevant). Press Replicate Experiment and choose 100 replications. Start with a normal population. Do three replication experiments, record the Type I error, and take the average. Repeat the experiment for a skewed right population and a very skewed right population. Briefly summarize your findings about the effects of population skewness on Type I error in a test for a variance. Did the large sample size offer protection against the ill effects of population skewness? Would the direction of skewness matter?

Distribution	Experiment 1	Experiment 2	Experiment 3	Avg Type I Error
Normal	____________	____________	____________	____________
Skewed Right	____________	____________	____________	____________
Very Skewed R	____________	____________	____________	____________

34. Investigate the effect of population skewness on the distribution of the test statistic in a test of one variance with a true null hypothesis. Use the same setup as in exercise 33. Start with 100 replications using a normal population. Compare the *shape* of the histogram of test statistics with the hypothesized chi-square sampling distribution. Should they be the same? Are they? Use the Change Distribution button to do a replication experiment for a skewed right population and a very skewed right population. Briefly summarize your findings about the effects of population skewness on the shape of the histogram. How does this result relate to the previous exercise?

35. How does this conclusion differ from exercise 29 (a test for a mean)? Why do you suppose a test for a variance is so sensitive to the effects of skewness in the population? What implication, if any, is there for real-world tests of variances? **Hint:** Think about the dot plot as you take samples from various distributions.

Tests of a Variance: Sample Size and Power

36. Investigate the effects of sample size on power in a test of one variance. Create a false hypothesis by using the Do-It-Yourself controls to set up a right-tailed test at $\alpha = 0.05$ with hypothesized variance of 100, true variance of 150, and sample size of 6 (the mean is irrelevant). Be *sure* the population is normal. Use 200 replications, record the power of the test, and repeat the experiment. Average the power. Increase the sample size to 25 and repeat the process. Increase the sample size to 96 and repeat the process. What happens to power each time you quadruple the sample size? Should you generalize from these results concerning power? Would you say that H_0 is "very false" in this case?

Sample Size	Experiment 1	Experiment 2	Average
n = 6	___________	___________	___________
n = 24	___________	___________	___________
n = 96	___________	___________	___________

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. Include in your report a copy of graphs and/or tables you feel are relevant for each different experimental setup.

1. Investigate the effects of sample size and the variance assumption on power in a test for the mean. Use a two-tailed test. Create a moderately false hypothesis using the Do-It-Yourself controls. Be sure the population is normal and Variance Assumed is set to Known. Start with a small sample size ($n < 6$). Choose the hypothesized mean, true mean, true standard deviation, and sample size in such a way that the difference in means is about one standard error so that the sampling distributions overlap substantially (use the Show True and Hypothesized Sampling Distribution toolbar button to display the sampling distributions). Use 500 replications to assess the power of the test (from the Analysis of Experiment window). Progressively double the sample size (n, 2n, 4n, 8n, 16n). Repeat, keeping every control the same, except set Variance Assumed to Unknown. Describe the effect sample size has on power, and discuss the effect of the assumption about variances (known or unknown).
2. Use the Do-It-Yourself controls to set up a two-tailed test of one mean at $\alpha = 0.05$ with a hypothesized mean of 100, a true mean of 97, and a *known* standard deviation of 10. Start with a small sample size ($n < 5$). Use 500 replications to assess the power of the test (from the Analysis of Experiment window). Through trial and error, find out approximately what sample size is required to achieve power of 0.10, 0.25, 0.50, 0.75, and 0.90. If you cannot achieve the desired power, speculate on the reason(s). Repeat the process with an *unknown* standard deviation. What generalizations can you draw?
3. Investigate the effects of level of significance (α) on power in a two-tailed test of one mean by creating a moderately false hypothesis using the Do-It-Yourself controls. Set Variance Assumed to Known. Choose a small sample size ($n < 5$). Choose the true mean, hypothesized mean, standard deviation, and sample size in such a way that the difference in means is about two standard errors. Use 500 replications to assess the power of the test (from the Analysis of Experiment window). Repeat two or three times and take the average. Try all available α values (0.01, 0.05, 0.10). Keeping everything else the same, change Variance Assumed to Unknown and repeat. Write a summary of your findings about the effect of level of significance on power and the effect of the unknown variance on power. What do you think would happen if sample size were increased (say to $n = 30$)?

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. Include in your report a copy of all graphs and statistics that you evaluated. Ask your instructor if a written report is also expected.

1. This is a project for a team of five. Investigate the effects on power of population shape, sample size, and variance assumption in a two-tailed test of one mean at $\alpha = 0.05$. Start with *known* variance. Use the Do-It-Yourself controls to create a moderately false hypothesis about a mean by choosing the hypothesized mean, true mean, standard deviation, and sample size in such a way that the difference in means is about one standard error (use the Show True and Hypothesized Sampling Distribution toolbar button to display the sampling distributions). Each team member should choose a different population (normal, skewed right, very skewed right, skewed left, very skewed left) and should replicate the sampling experiment 500 times for sample sizes of 4, 10, and 30. Record the empirical power (from the Analysis of Experiment window). Compare each histogram of test statistics with the hypothesized sampling distribution. How does the population affect empirical power? Repeat for *unknown* variance. How does the assumption about the variance affect empirical power?
2. This is a project for a team of three. Investigate the effects of sample size and population skewness on power and Type I error in a test of the variance. First, create a true hypothesis by using the Do-It-Yourself controls to set up a two-tailed test at $\alpha = 0.05$ with hypothesized variance of 200 and true variance of 200 (the mean is irrelevant). Each team member should choose a different population shape (normal, skewed, very skewed) and should use sample sizes of 10, 25, and 50. Use 500 replications and record the Type I error of the test for each sample size. Repeat the experiment with the hypothesized variance set to 100 (so the null hypothesis is false) and this time record power. Discuss your findings.
3. This is a project for a team of three or four. Construct power curves for a right-tailed test of one mean with *unknown* variance with $\alpha = 0.05$. Using the Do-It-Yourself controls, choose a normal population. Each team member should choose a different sample size so that the range from 5 to 100 is covered. Create a true mean that exceeds the hypothesized mean by three standard errors (use the Show True and Hypothesized Sampling Distribution toolbar button to display the sampling distributions). Do 500 replications three times and find the average power. Record the true mean and average power. With other factors constant, repeat the replication for four other values of the true mean, moving the true mean progressively closer to the hypothesized mean until the sampling distributions are identical. Display the team results on a single graph. For each sample size, plot the power (on the Y-axis) against the true mean (on the X-axis) and connect the points. Discuss what the power curves tell you.

Self-Evaluation Quiz

1. In a right-tail test, the rejection region refers to
   a. the area to the left of the right-tail critical value.
   b. the area to the right of the left-tail critical value.
   c. the area between the left-tail and right-tail critical values.
   d. the area outside the left-tail and right-tail critical values.
   e. none of the above.

2. Which statement is *not* correct regarding the level of significance?
   a. It denotes the probability of Type I error.
   b. It is usually indicated by the symbol α.
   c. If it is reduced, it is harder to reject the null hypothesis.
   d. It is usually set at a high level (such as 90%, 95%, 99%).
   e. It indicates the percent of the time a true null hypothesis will be rejected.

3. Other things equal, in a left-tail test, if α is decreased from 0.05 to 0.01 the critical value
   a. will shift left.
   b. will shift right.
   c. could shift either left or right.
   d. will not shift.
   e. moves to the right tail.

4. In a normal population with known variance the distribution of the sample mean is
   a. normal.
   b. Student's t.
   c. chi-square.
   d. F.
   e. impossible to determine.

5. In a normal population with unknown variance the distribution of the sample mean is
   a. normal.
   b. Student's t.
   c. chi-square.
   d. F.
   e. impossible to determine.

6. Degrees of freedom for a sample mean with $n = 30$ from a normal population with a known variance will be
   a. 30
   b. 29
   c. 28
   d. 31
   e. irrelevant.

7. Type I 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.

8. In a test of one mean, if the null hypothesis is true, the histogram of sample means in a replicated sampling experiment
   a. will be shifted to the right of the hypothesized sampling distribution.
   b. will be shifted to the left of the hypothesized sampling distribution.
   c. could be shifted either left or right of the hypothesized sampling distribution.
   d. will not be shifted relative to the hypothesized sampling distribution.
   e. has none of the above behaviors.

9. The distribution of the sample variance is
   a. always symmetric.
   b. always skewed right.
   c. always skewed left.
   d. more than one of the above.
   e. none of the above.

10. In a normal population the theoretical distribution of the sample variance is
    a. normal.
    b. Student's t.
    c. chi-square.
    d. F.
    e. impossible to determine.

11. Degrees of freedom for a sample variance with $n = 10$ from a normal population will be
    a. 10
    b. 9
    c. 8
    d. irrelevant.
    e. none of the above.

12. In terms of Type I error (if H_0 is true) or power (if H_0 is false), a test of a hypothesis about a mean is fairly insensitive (robust) to non-normality of the population.
    a. True.
    b. False.

Glossary of Terms

Acceptance region Portion of the hypothesized distribution that is bounded by the critical value(s). Its area is $1 - \alpha$ where α is the chosen level of significance. Since a given sample can disprove (but cannot prove) the null hypothesis, it is more accurate to call it the *non-rejection region*. See **Type I error**.

Alternative hypothesis Denoted H_1, it is the converse of the null hypothesis (e.g., H_0: $\mu = 5$ and 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 test is motivated by the suspicion that the null hypothesis may be false.

Chi-square distribution When testing one sample variance against a hypothesized value of the population variance, the sampling distribution is chi-square with $n - 1$ degrees of freedom. More generally, the chi-square distribution describes the sum of squared independent identically distributed normal random variables (e.g., in the numerator of the sample variance).

Confidence interval Range of values that would enclose a true (generally unknown) population parameter (such as a population mean or variance) a given percentage of the time.

Confidence level Desired probability of enclosing an unknown population parameter when creating a confidence interval from sample data. The confidence level, denoted $1 - \alpha$, is chosen by the researcher and is usually expressed as a percent (typically 90%, 95%, or 99%). The higher the confidence level, the wider the confidence interval. See **Level of significance**.

Critical value Value on the X-axis that defines the rejection region for a hypothesis. In a one-tail test, the critical value defines a right-tail or left-tail area. In a two-tail test, there are critical values defining rejection regions in each tail. The critical value is determined by the level of significance. A test statistic beyond the critical value(s) is unlikely if the null hypothesis is true. Critical values may be found in a table or may be generated by a computer algorithm.

Decision rule Diagram that illustrates the criterion for rejection of the null hypothesis. It shows the hypothesized sampling distribution of the test statistic with its critical value(s) labeled and a shaded rejection region for the specified level of significance.

Degrees of freedom For a Student's t-test for one mean with unknown population variance, degrees of freedom will be $n - 1$. For a chi-square test of one variance, degrees of freedom will also be $n - 1$.

Level of significance The desired probability of Type I error. It is set by the researcher (typical values are 0.10, 0.05, and 0.01) and is denoted α. Other things equal, the power of a hypothesis test increases as the level of significance increases. See **Confidence level**.

Null hypothesis Denoted H_0, it is a statement that we try to reject (for example, H_0: $\sigma^2 = 5$). The null hypothesis is not necessarily chosen because we believe it to be true, but rather as an important reference point. If the sample evidence contradicts H_0, the null hypothesis is rejected. Otherwise, it awaits further testing and could be rejected at a later time.

One-tail test In a one-tail test, the alternative hypothesis always contains > (for a right-tail test) or < (for a left-tail test). For example, the hypotheses H_0: $\mu \geq 5$ and H_1: $\mu < 5$ imply a left-tail test.

P-value Probability that a result as extreme as (or more extreme than) the observed sample statistic would arise by chance if the null hypothesis were true.

Power If the null hypothesis is false, *theoretical* power is the probability of rejecting the null hypothesis H_0 when it is false (or $1 - \beta$ where β is the probability of Type II error). For example, in a test of $\mu = 5$, power will be low if the true mean is near 5, but will be higher if the true mean differs substantially from 5. The ideal power is near 1. *Empirical* power is the ratio of the number of rejections to the number of times the test is performed.

Rejection region Area under the hypothesized sampling distribution that lies beyond the critical value(s). It is determined by α, the probability of Type I error. If the test statistic falls within this region, we will reject the null hypothesis. See **Level of significance**.

Robustness Quality of being unaffected by a violation of an assumption. For example, the Student's t test is robust to non-normality in the population if the sample is moderately large.

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

Sampling distribution Theoretical distribution of the estimator or the test statistic assuming the null hypothesis is true.

Test statistic Calculated value using the sample statistic and a hypothesized population parameter. Its distribution is known if the null hypothesis is true. The test statistic is compared with a critical value to see whether the null hypothesis should be rejected.

Two-tail test In a two-tail test, the alternative hypothesis always contains $\neq$. For example, the hypotheses H_0: $\mu = 5$ and H_1: $\mu \neq 5$ imply a two-tail test.

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 sampling distribution that is beyond the critical value(s). See **Level of significance**.

Type II error Error of accepting a null hypothesis that is false. The probability of Type II error is denoted β. The probability of Type II error is the area under the true sampling distribution that is not in the rejection region. See **Power**.

Solutions to Self-Evaluation Quiz

1. e Read the Overview of Concepts and Illustration of Concepts. Consult the Glossary.
2. d Read the Overview of Concepts. Consult the Glossary.
3. a Read the Overview of Concepts. Do Individual Learning Project 3.
4. a Do Exercises 1–6. Read the Overview of Concepts.
5. b Do Exercises 7–9. Read the Overview of Concepts.
6. e Do Exercises 5 and 6. Consult the Glossary.
7. c Do Exercises 10–13. Read the Overview of Concepts.
8. d Do Exercises 13–16.
9. b Do Exercises 21–25. Consult the Glossary. Read the Overview of Concepts.
10. c Do Exercises 21–25. Consult the Glossary.
11. b Do Exercises 21 and 22. Consult the Glossary.
12. a Do Exercises 28–32. Read the Overview of Concepts.