Chapter 9: Testing a Claim

1. The time that it takes an untrained rat to run a standard maze has a Normal distribution with mean 65 seconds and standard deviation 15 seconds. The researchers want to use a test of hypotheses to determine whether training significantly improves the rats’ completion times. An appropriate alternative hypothesis would have the form

A. *H*a: μ > 65.

AR. Incorrect. The alternative hypothesis corresponds to what the researchers are trying to show (the research hypothesis). If training improves times, then the rats should run the maze in less time, which means that the mean completion time would decrease rather than increase.

B. *H*a: < 65.

BR. Incorrect. Hypotheses are statements about population parameters, not sample statistics.

\*C. *H*a: μ < 65.

CR. Correct. The alternative hypothesis corresponds to what the researchers are trying to show (the research hypothesis). If training improves times, then the rats should run the maze in less time, which means that the mean completion time would decrease.

2. A researcher collects infant mortality data from a random sample of villages in a certain country. It is claimed that the mean death rate in this country is the same as that of a neighboring country, which is known to be 17 deaths per 1000 live births. To test this claim using a test of hypotheses, what should the null and alternative hypotheses be?

\*A. *H*0: μ = 17, *H*a: μ ≠ 17

AR. Correct. When directly testing a claim, the claim should be used as the null hypothesis. No direction was specified in this case, so a two-sided alternative hypothesis should be used.

B. *H*0: μ ≠ 17, *H*a: μ = 17

BR. Incorrect. You have interchanged the two hypotheses. Recall that a null hypothesis is typically the “no difference” hypothesis.

C. *H*0: μ = 17, *H*a: μ > 17

CR. Incorrect. No direction was specified in this case, so a two-sided alternative hypothesis should be used.

3. A social psychologist reports that “in our sample, ethnocentrism was significantly higher

(*P* < 0.05) among church attendees than among non-attendees.” Which of the following statements best describes what this means.

A. Ethnocentrism was at least 5% higher among church attendees than among non-attendees.

AR. Incorrect. The *P*-value tells us nothing about the magnitude of the actual difference in the amount of ethnocentrism in church attendees compared to non-attendees.

B. The probability that the true mean is 17 is less that 0.05.

BR. Incorrect. The *P*-value does not measure the probability that either hypothesis is false or true.

\*C. If there were actually no difference in ethnocentrism between church attendees and non-attendees, then the chance that we would have observed a difference at least as extreme as the one we did is less than 5%.

CR. Correct. Recall that the *P*-value is the probability that the test statistic would take a value at least as extreme as that actually observed if the null hypothesis were true.

4. Which of the following *P*-values obtained from a test of hypotheses constitutes the *least* amount of evidence against the null hypothesis?

A. 0.107

AR. Incorrect. Recall that the *P*-value measures the likelihood of getting a sample statistic like ours if *H0* is true. The smaller the *P*-value, the less plausible *H­0* is.

\*B. 0.207

BR. Correct. Recall that the *P*-value measures the likelihood of getting a sample statistic like ours if *H0* is true. The smaller the *P*-value, the less plausible *H­0* is. A large *P*-value provides less evidence against the null hypothesis.

C. 0.017

CR. Incorrect. Recall that the *P*-value measures the likelihood of getting a sample statistic like ours if *H0* is true. The smaller the *P*-value, the less plausible *H­0* is.

5. Suppose we conduct a test of hypotheses and find that the test results are significant at the α = 0.025 level. Which of the following statements then must be true?

\*A. The results are significant at level α = 0.05.

AR. Correct. For the results to be significant at level α = 0.025, the *P*-value must be at most 0.025. The *P*-value is, therefore, also at most 0.05, and thus the results are significant at level α = 0.05.

B. The results are not significant at level α = 0.01.

BR. Incorrect. It is possible for the results to be significant at level α = 0.025 *and* significant at level α = 0.01, for example, if the *P*-value is 0.008.

C. The test results are important.

CR. Incorrect. Recall that statistical significance is not the same thing as practical significance. An effect of a small magnitude may nonetheless be declared significant at level α = 0.025 if the sample size is sufficiently large.

6. A noted psychic is tested for ESP. The psychic is presented with 400 cards, all face down, and asked to determine if each card is marked with one of four symbols: a star, a cross, a circle, or a square. The psychic is correct in 120 of the 400 cases. Let *p* represent the probability that the psychic correctly identifies the symbol on the card in a random trial. Suppose we wish to see if there is evidence to suggest that the psychic is doing significantly better than he would be if he were just guessing. To do so, we test *H*0: *p* = 0.25 against *H*a: *p* > 0.25. Which of the following is closest to the *P*-value of the test?

\*A. 0.0104.

AR. Correct. This is a 1-sample *z*-test for a proportion, where = 120/400 = 0.30, *n* = 400, and *p*0 = 0.25. So  For a one-sided test, 

B. 0.0146.

BR. Incorrect. This choice suggests that you calculated the standard error incorrectly. Try again!

C. 0.9896.

CR. Incorrect. Make sure that the area you use to calculate the *P*-value is appropriate for your alternative hypothesis.

7. Suppose that we are conducting a *t-*test for a population mean *μ*. We originally do a one-sided test but then decide that a two-sided test might be more appropriate (typically, we use a two-sided test unless there is some reason to believe that an effect in a particular direction exists). Which of the following accurately describes how the results of the test will change?

A. The test statistic will have a larger value, but the *P-*value will not change.

AR. Incorrect. The value of the test statistic is not affected by the form of the alternative hypothesis.

B. The test statistic will not change, but the *P*-value of the test will be smaller.

BR. Incorrect. While the test statistic will not change, a two-sided test requires that your *P-*value measures the probability of getting a result on either side of the null value.

\*C. The test statistic will not change, but the *P*-value of the test will be larger.

CR. Correct. For the same sample statistic, the tests statistic will not change, but the *P*-value of the test will double to take into account the probability of getting a results as extreme as our sample on *either* side of the null value.

8. The Department of Health plans to test the lead level in a public park. The park will be closed if the lead level exceeds the allowed limit. Otherwise, the park will be kept open. The department conducts the test using soil samples gathered at randomly selected locations. Which of the following decisions would constitute making a Type I error in this situation?

A. Keeping the park open when the average lead level exceeds the allowed limit.

AR. Incorrect. A Type I error is committed when a true null hypothesis is incorrectly rejected. In this case, the null hypothesis is that the lead level is acceptable and the alternative is that it exceeds the allowed limit. For the answer you chose, a false null hypothesis has not been rejected, which would constitute a Type II error.

\*B. Closing the park when the average lead level is acceptable.

BR. Correct. A Type I error is committed when a true null hypothesis is incorrectly rejected. In this case, the null hypothesis is that the lead level is acceptable and the alternative is that it exceeds the allowed limit, so the answer you chose would, in fact, correspond to a Type I error.

C. Closing the park when the average lead level exceeds the allowed limit.

CR. Incorrect. In this case, the null hypothesis is that the lead level is acceptable and the alternative is that it exceeds the allowed limit. No error has been committed here: we are rejecting a false null hypothesis!

9. Two species of sunfish, pumpkinseeds and bluegills, are common in Puffer’s Pond. For many years, the proportion of bluegills was 0.42, but a local ecologist suspects that a newly-introduced predator is increasing that proportion. He collects 50 sunfish with a net and finds that 27 of them are bluegills. Assuming that we can treat his net sample as a simple random sample, which of the following is the correct check of the Large Counts condition for a one-sample *z*-test of 

A. 

AR. Incorrect. This is the Normal/Large Sample condition for a test of a *mean*, not a proportion.

B. both of which are ≥ 10.

BR. Incorrect. For a one-sample *confidence interval* for a proportion, this would be the correct calculation. But there is a slight change for the test of significance.

\*C.  both of which are ≥ 10.

CR. Correct. The large counts condition for a test of significance requires the use of the null value for *p*, since the reasoning of the test assumes that *H*0 is true.

10. In a test of hypotheses, the probability that a false null hypothesis is rejected is also known as the

A. probability of committing a Type II error.

AR. Incorrect. A Type II error corresponds to incorrectly *failing* to reject a false null hypothesis.

\*B. power of the test.

BR. Correct. The power of the test (for a particular alternative) is the probability of rejecting *H*0 when H0 is false.

C. significance level of the test.

CR. Incorrect. The significance level, α, is the probability of committing a Type I error, that is, incorrectly rejecting a true null hypothesis.

11. Which of the following will cause the power of a test to increase?

\*A. Increasing the sample size

AR. Correct. Increasing the sample size makes the test more sensitive to the effect and increases the power.

B. Decreasing the significance level α of the test

BR. Incorrect. Decreasing α causes the probability of committing a Type I error to decrease. Type I and Type II error probabilities are inversely related, so this would cause the probability of a Type II error to go up, thereby decreasing the power of the test.

C. Increasing the probability of committing a Type II error.

CR. Incorrect. The power of the test is the complement of the probability of a Type II error. If we increase this probability, power will decrease.

12. Which one of the following sample data sets would satisfy the Normal/Large sample condition for performing a one-sample *t*-procedure?

\*A. A sample of 38 prices for new houses in Sonoma County, California that is moderately skewed to the right but has no outliers.

AR. Correct. This sample is large enough so that the central limit theorem ensures an approximately Normal distribution for the sampling distribution of means in spite of the skew.

B. A sample of diameters for 14 oak trees in Cumberland County, Maine that is somewhat skewed toward larger trees.

BR. Incorrect. Since the sample size is small and the data shows some skew, we cannot safely use *t*-procedures with these data.

C. A sample of 60 hold time lengths for calls to a customer service line that are roughly symmetric except for one high outlier.

CR. Incorrect. The presence of outliers makes the *t*-procedure unreliable, even if the sample size is large.

13. A city school board claims that the mean number of school days missed due to illness by the city’s schoolteachers is 5 per year. The teacher’s union thinks it actual mean is lower than that. A random sample of 28 city school teachers missed an average of 4.5 days last year, with a sample standard deviation of 0.9 days. The distribution of the number of days missed in the sample is roughly symmetric with no outliers. A test of *H*0: μ = 5 and Ha: μ < 5 produces a *P*-value in which of the following intervals?

\*A. Between 0.0025 and 0.005.

AR. Correct. The test statistic is , where = 4.5, *μ*0 = 5, *s* = 0.9, and *n* = 28, so . The test is left-tailed, so the *P*-value is *P*(*t* ≤ −2.94), where *t* has 27 degrees of freedom. By the *t-*table, the *P*-value is between 0.0025 and 0.005.

B. Between 0.001 and 0.0025.

BR. Incorrect. This would be the correct answer if the test procedure were based on the standard Normal distribution, but this should be a *t*-test for a mean.

C. between 0.005 and 0.01.

CR. Incorrect. You incorrectly used a two-sided alternative hypothesis in this case. The actual alternative is one-sided.

14. An advertisement for Food Chain, a regional supermarket chain, claimed that the chain has had consistently lower prices than its regional competitors. As part of a survey conducted by an independent price-checking company, the average weekly grocery bill (based on the prices of approximately 95 commonly purchased items) was recorded for Food Chain and one of its leading competitors during 8 randomly selected weeks. We wish to conduct a test of

*H*0: μ*d* = 0 vs. *H*a: μ*d* < 0, where μ*d* = the mean difference between the weekly Food Chain bill and the weekly bill at the competing chain. Which of the following is the correct Normal/Large Sample condition for conducting this test of significance?

A. The distributions of weekly bills at each of the chains should be approximately Normally distributed.

AR. Incorrect. It is not necessary that the individual bills at each of the chains are Normally distributed.

\*B. The distribution of differences between weekly bills at the two chains should be approximately Normally distributed.

BR. Correct. Since a matched-pairs test of this type is essentially a one-sample *t*-test performed on the individual weekly differences, only those differences need to be approximately Normally distributed.

C. Both the distributions at each of the chains and the distribution of differences between weekly bills should be approximately Normally distributed.

CR. Incorrect. All three distributions do *not* need to be approximately Normal.

15. We would like to test the null hypothesis *H*0: μ = 50 against *H*a: μ ≠ 50. The 95% confidence interval for *μ* is found to be (51.3, 54.7). Assuming all conditions for a one-sample t procedure have been met, which of the following must be *true*?

A. The *P*-value of the test is greater than 0.05.

AR. Incorrect. A two-sided significance test rejects *H*0 at the  = 0.05 level when the null value falls outside a level 95% confidence interval for *μ*. The interval here suggests that we should reject *H*0: μ = 50 in favor of *H*a: μ ≠ 50, since the interval doesn’t include the value 50.

\*B. The *P*-value of the test is less than 0.05.

BR. Correct. A two-sided significance test rejects *H*0 at the  = 0.05 level when the null value falls outside a level 95% confidence interval for *μ*. The interval here suggests that we should reject *H*0: *μ* = 50 in favor of *H*a: *μ* ≠ 50, since the interval doesn’t include the value 50. Because we reject *H*0 at level α = 0.05, the *P*-value of the test must be less than 0.05.

C. The *P*-value of the test could be either greater than or at most 0.05. It can’t be determined without knowing the sample size.

CR. Incorrect A two-sided significance test rejects *H*0 at the  = 0.05 level when the null value falls outside a level 95% confidence interval for *μ*. The interval is given in this case, and we just have to determine whether the null value of 50 is in the interval or not.

16. Does taking garlic tablets twice a day provide significant health benefits? A researcher conducted a study of 50 adult subjects who took garlic tablets twice a day for a period of six months. At the end of the study, 100 variables related to the health of the subjects were measured for each subject, and the means were compared to known means for these variables in the population of all adults. Four of these 100 variables were significantly better (in the sense of statistical significance) at the 5% level for the group taking the garlic tablets compared to the population as a whole. One variable was significantly better at the 1% level for the group taking the garlic tablets compared to the population as a whole. Which of the following is an appropriate conclusion to draw from these results?

A. There is good statistical evidence that taking garlic tablets twice a day provides some health benefits.

AR. Incorrect. Consider this: if you perform 100 tests of significance at the  = 0.05 level, about how often would you expect to make a Type I error?

B. There is good statistical evidence that taking garlic tablets twice a day provides benefits in the case of the variable that was significant at the 1% level. However, we should be somewhat cautious about making claims for the variables that were significant at the 5% level.

BR. Incorrect. Consider this: if you perform 100 tests of significance at the  = 0.01 level, about how often would you expect to make a Type I error?

\*C. Neither (A) nor (B) is true.

CR. Correct. Even if garlic had no effect on any of the 100 variables measured, we would expect to see about five variables appearing to be statistically significant at the 5% level and about one variable appearing to be significant at the 1% level by chance alone. The level of significance is the probability of making a Type I error, so we would expect to reject a true null hypothesis about 5 times in 100 tests when  = 0.05 and once in 100 tests when  = 0.01.

17. A kitchen stove manufacturer is testing a shipment of oven thermostats to see if they keep an oven at 350 Fº. Which of the following tests has the greatest power?

A. 

AR. Incorrect. Consider that both increasing effect size and sample size increase power.

B. 

BR. Incorrect. Consider that both increasing effect size and sample size increase power.

\*C. 

CR. Correct. Since this choice has the greatest effect size (25º) and the largest sample size, this test has the greatest power.