Mini-Lecture 10.1

The Language of Hypothesis Testing

Objectives

1. Determine the null and alternative hypotheses
2. Understand Type I and Type II errors
3. State conclusions to hypothesis tests

Examples

1. The Graduate Record Examination (GRE) is a test required for admission to many U.S. graduate schools. The Educational Testing Service (ETS), which produces the test, claims that students' scores on the verbal reasoning portion of the GRE are normally distributed with a mean of 465 and a standard deviation of 117. (Source: http://www.ets.org/Media/Tests/GRE/pdf/994994.pdf) An admissions officer at a large graduate school believes the mean is less than 465.
   a. Determine the null and alternative hypotheses (H_0: $\mu = 465$; H_1: $\mu < 465$)
   b. Explain what it would mean to make a Type I error. (The mean is 465, but the null hypothesis is rejected)
   c. Explain what it would mean to make a Type II error. (The mean is less than 465, but the null hypothesis is not rejected)

2. Many health care workers believe the mean body temperature of healthy adults is 98.6 degrees Fahrenheit. Philip A. Mackowiak, a doctor at the University of Maryland Center for Vaccine Development, suspects this number might be incorrect. (Source: Mackowiak, PA, Wasserman, SS, and Levine, MM (1992) A Critical Appraisal of 98.6 Degrees F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich. Journal of the American Medical Association (268) pp. 1578-1580.)
   a. Determine the null and alternative hypotheses. (H_0: $\mu = 98.6$; H_1: $\mu \neq 98.6$)
   b. Sample data indicated that the null hypothesis should be rejected. State the conclusion of the researcher. (There is sufficient evidence to suggest that the mean body temperature of healthy adults is not equal to 98.6 degrees Fahrenheit.)
   c. Suppose the mean body temperature of healthy adults is actually 98.6 degrees Fahrenheit. Was a Type I or Type II error committed? (Type I)

3. Pork producers and their neighbors are concerned about potential health safety and environmental issues associated with swine barns. The Veterinary Infectious Disease Organization (VIDO) conducted a 5-month air-quality study and found that there was no statistically significant increase in airborne pollutants 600 meters (m) downwind from a swine barn. Let *X* be the increase in airborne pollutants measured 600 m downwind from the swine barn. (Source: Swine barns are not a downwind health threat (2002) Canadian Veterinary Journal (43) p. 344.)
   a. Determine the null and alternative hypotheses. (H_0: $\mu = 0$; H_1: $\mu > 0$)
   b. Sample data indicated that the null hypothesis should not be rejected. State the conclusion of the researcher. (There is insufficient evidence to suggest that there is an increase in airborne pollutants 600m downwind from a swine barn.)
   c. Suppose, in fact, there is an increase in airborne pollutants 600 meters downwind from a swine barn. Was a Type I or Type II error committed? (Type II)

4. A group of researchers studied the effect of global environmental change on grazing lands. Their article reported the results of 171 different hypothesis tests. Of these tests, 4 were determined to be statistically significant at the 0.05 level. (Source: Asner GP, Elmore AJ, Olander LP, Martin RE, and Harris AT (2004) Grazing systems, ecosystem responses, and global change. Annual Review of Environment and Resources (29) .)
   a. For each of the tests, what was the probability of committing a Type I error? (0.05)
   b. Based on this information, what conclusions might be drawn? (We expect that these significant results might be due to chance alone. Assuming all the null hypotheses are true, and the 171 hypothesis tests are all independent, we would expect about 5%, or 8.55, would result in Type I errors at the 0.05 level of significance.)

Mini-Lecture 10.2

Testing Claims for a Population Mean – Population Standard Deviation Known

Objectives

1. Understand the logic of hypothesis testing
2. Test hypotheses about a population mean with σ known using the classical approach
3. Test hypotheses about a population mean with σ known using *P*-values
4. Test hypotheses about a population mean with σ known using confidence intervals
5. Understand the difference between statistical significance and practical significance

Examples

1. The Graduate Management Admission Test (GMAT) is a test required for admission into many masters of business administration (MBA) programs. Total scores on the GMAT are normally distributed and historically have a standard deviation of 113. The Graduate Management Admission Council, who administers the test, claims that the mean total score is 529. (Source: http://www.mba.com/NR/rdonlyres/55DF55BA-4F4E-4DB9-A5BE-39DC98C46551/0/ExamineeScoreGuide5.pdf.) Suppose a random sample of 8 students took the test, and their scores are given below.

   699, 560, 414, 570, 521, 663, 727, 413

   Test the claim that the mean total score is different from 529 using either the classical approach or the *P*-value approach at the 0.05 level of significance. (Fail to reject the null hypothesis)

2. The Graduate Record Examination (GRE) is a test required for admission to many U.S. graduate schools. Students' scores on the verbal reasoning portion of the GRE follow a normal distribution with a standard deviation of 117. The Educational Testing Service (ETS), the exams administrators, claim that the mean verbal reasoning score is 465. (Source: http://www.ets.org/Media/Tests/GRE/pdf/994994.pdf) Suppose a random sample of 10 students took the test, and their scores are given below.

   489, 564, 624, 284, 388, 424, 515, 361, 398, 546

   Test the claim that the mean verbal reasoning score is different from 465 using either the classical approach or the *P*-value approach at the 0.10 level of significance. (Fail to reject the null hypothesis)

Mini-Lecture 10.3

Hypothesis Tests for a Population Mean – Population Standard Deviation Unknown

Objective

1. Test hypotheses about a population mean with σ unknown

Examples

1. Assume the resting metabolic rate (RMR) of healthy males in complete silence is 5710 kJ/day. Researchers measured the RMR of 11 healthy males who were listening to calm classical music. The data are given in kJ/day:

   4900, 6900, 6600, 5800, 6600, 5500, 5600, 6700, 3200, 5500, 5500.

   (Based on data from: Carlsson E, Helgegren H and Slinde F. (2005) Resting energy expenditure is not influenced by classical music. Journal of Negative Results in BioMedicine, 4(6).)
   a. Assuming the RMR is normally distributed in the population, is there significant evidence to support the claim that the mean RMR of males listening to calm classical music is different than 5710 kJ/day? Use the 0.05 level of significance. (Fail to reject the null hypothesis)
   b. Test the claim by constructing a 95% confidence interval. ((5001.6, 6416.5); since this confidence interval contains 5710, fail to reject the null hypothesis)

2. The heights of adult men are normally distributed. The heights (in inches) of 9 adult males are given below.

   72, 71, 71, 68, 68.75, 70.25, 72, 70.25, 68.25

   A physician believes that the mean height of adult men is larger than 69.9 inches. Test the physician's claim at the $\alpha = 0.05$ level of significance. (Fail to reject the null hypothesis)

Mini-Lecture 10.4
Hypothesis Test for a Population Proportion

Objectives
1. Test hypotheses about a population proportion
2. Test hypotheses about a population proportion using the binomial probability distribution

Examples
1. It has been reported that the probability that an individual will develop schizophrenia over their lifetime is 0.004. In a random sample of 3,000 individuals, it was determined that 17 developed schizophrenia. Is there evidence to support the claim that the true proportion of people who will develop schizophrenia is different from 0.004 at the $\alpha = 0.05$ level of significance? (Source: Saha S, Chant D, Welham J, McGrath J (2005) A systematic review of the prevalence of schizophrenia. PLoS Med 2(5): e141.) (Fail to reject the null hypothesis)

2. In 2007, the New York Mets won 41 of their 81 home baseball games. (Source: www.baseball-almanac.com) Is there sufficient evidence that the proportion of wins at home is greater than 0.5 at the $\alpha = 0.10$ significance level? (Fail to reject the null hypothesis)

3. A coin was tossed 10 times and "heads" appeared exactly 2 times. Is there sufficient evidence that the coin is not fair at the $\alpha = 0.05$ significance level? (Note: the sample size is small.) (Fail to reject the null hypothesis)

Mini-Lecture 10.5

Hypothesis Test for a Population Standard Deviation

Objective

1. Test hypotheses about a population standard deviation

Examples

1. The Airsoft Elite Precision 0.2g BBs are projectiles used in BB guns. The manufacturer advertises that these BB's are very accurate in size. Their tolerances state that the diameter of the BBs must be between 5.97 and 5.99 mm. (www.airsplat.com/Categories/B1-B4.htm.) If the manufacturer uses the 6-sigma methodology, the process standard deviation must not exceed 0.0017 mm. Suppose that during one 8-hour shift, the quality control manager at Airsoft randomly selected seven BBs from the production line for thorough testing, and the diameters of these BBs are given below.

   5.9808, 5.9801, 5.9819, 5.9785, 5.9769, 5.9796, 5.9778

   Assuming the diameters of the BBs follow a normal distribution, is there sufficient evidence for the quality control manager to conclude that the standard deviation exceeds 0.0017 at the $\alpha = 0.05$ significance level? (Fail to reject the null hypothesis)

2. The Graduate Record Examination (GRE) is a test required for admission to many U.S. graduate schools. Students' scores on the verbal reasoning portion of the GRE follow a normal distribution with a standard deviation of 117. (Source: http://www.ets.org/Media/Tests/GRE/pdf/994994.pdf.) Suppose a random sample of 10 students took the test, and their scores are given below.

   489, 564, 624, 284, 388, 424, 515, 361, 398, 546

   Is there sufficient evidence to conclude that the standard deviation is different from 117 at the $\alpha = 0.05$ significance level? (Fail to reject the null hypothesis)

3. The test for a population standard deviation can be used to assess the reproducibility of a result. If measurements from a piece of equipment have a large standard deviation, then the data are not very reproducible, and repeated measurements could vary widely. A group of researchers at the University of California-San Francisco investigated the use of a new technology, diffusion tensor tractography (DTT) as a means of assessing the development of premature newborns. Assume that the data were drawn from normally distributed populations. It was determined that the standard deviation obtained using this new technology is lower than the previous standard. (Source: Partridge SC, et al. (2005) Tractography-based quantitation of diffusion tensor imaging parameters in white matter tracts of preterm newborns. Journal of Magnetic Resonance Imaging (22)4.)
   a. What were the null and alternative hypotheses? (H_0: $\sigma = \sigma_0$, H_a: $\sigma < \sigma_0$)
   b. Was the P-value greater than or less than the level of significance? (The *P*-value was less than the level of significance, α)
   c. What was the conclusion of the test? (Reject the null hypothesis)

Mini-Lecture 10.6

Putting It Together: Which Method Do I Use?

Objective

1. Determine the appropriate hypothesis test to perform

Examples

1. A small municipality in Tennessee is interested in gauging public interest for developing an indoor recreational complex in the city. They surveyed 350 residents and of those, 261 stated that they "agree" or "strongly agree" that the city should build an indoor recreational facility. The remaining 109 people stated that they "disagree" or "strongly disagree" that the city should construct such a facility. The mayor knows this sample only represents a fraction of the residents. Which hypothesis test should be performed? Use this procedure to test the claim that over 67% of the residents in the city would support such an action at the $\alpha = 0.05$ significance level. (Test for a population proportion; reject the null hypothesis)

2. The quality control manager for a filling operation in a bottling plant is concerned with the variability in the volume of milk dispensed into gallon jugs. The filling process results in jugs whose volumes are normally distributed with a mean of 1.02 gallons. The process standard deviation should be less than 0.004 gallons. A sample of 25 jugs was selected and the sample standard deviation was determined to be 0.0036 gallons. Which hypothesis test should be performed? Use this procedure to test the claim that the standard deviation of the amount of milk in the jugs is less than 0.004 at the $\alpha = 0.05$ level of significance. (Test for a population standard deviation; fail to reject the null hypothesis.)

3. A group of researchers questioned the common belief that the body temperature of healthy adults is centered at 98.6 degrees Fahrenheit. What type of test should be performed? (Source: Mackowiak, PA, Wasserman, SS, and Levine, MM (1992) A Critical Appraisal of 98.6 Degrees F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich. Journal of the American Medical Association (268).) (A test for a population mean)

Mini-Lecture 10.7

The Probability of a Type II Error and the Power of the Test

Objectives

1. Determine the probability of making a Type II error
2. Compute the power of the test

Examples

1. A group of researchers investigated the common belief that the mean body temperature of healthy adults is 98.6 degrees Fahrenheit. They measured the body temperature of 148 healthy men and women. (Source: Mackowiak, PA, Wasserman, SS, and Levine, MM (1992) A Critical Appraisal of 98.6 Degrees F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich. Journal of the American Medical Association (268) pp. 1578-1580.) Consider the test of H_0: $\mu = 98.6$ versus H_1: $\mu < 98.6$.

   Assume that the true mean and standard deviation of body temperatures are 98.25 and 0.73, respectively.

   a. What would it mean to make a Type II error for this test? (A Type II error is committed if the true body temperature is less than 98.6 degrees Fahrenheit and we fail to reject the null hypothesis.)
   b. If the researcher tests this null hypothesis at the $\alpha = 0.05$ level of significance, compute the probability of making a Type II error if the true population mean is 98.25. What is the power of the test? (less than 0.0001; over 0.9999. Answers may vary due to rounding.)
   c. Redo part (b.) if the true population mean is 98.50 degrees Fahrenheit. (0.4914; 0.5086. Answers may vary due to rounding.)

2. A deputy sheriff in Washington County believes that the mean speed of cars on a particular county road exceeds the posted speed limit of 55 mph. Assume that the population standard deviation for speeds on this road is 12 mph. To test the deputy's theory, a random sample of the speeds of 40 cars on this road was collected.

   a. What would it mean to make a Type II error for this test? (A Type II error is committed if the true mean speed exceeds 55 mph and we fail to reject the null hypothesis.)
   b. If the sheriff decides to test this hypothesis at the $\alpha = 0.01$ level of significance, compute the probability of making a Type II error if the true population mean speed is 57 mph. What is the power of the test? (0.8984; 0.1016. Answers may vary due to rounding.)
   c. Redo part (b.) if the true mean speed on this road is 60 mph. (0.3787; 0.6213. Answers may vary due to rounding.)