

Significance Tests

The basic setup of statistical hypothesis testing is important to understand. The pieces are

1. A pair of hypotheses H_0 and H_A .
2. A test statistic with known sampling distribution to judge the validity of an assumption that H_0 is true.
3. A data set that yields an observed value of the test statistic.
4. An analysis computing a p -value, or at least comparing it to a significance level α .

We learned the following tests:

1. Test for proportion
2. Test for mean
3. Test of difference of proportions
4. Test for difference of means

The steps you will need to know to perform any of these significance tests are

1. Specify the null and alternative hypotheses: H_0 , H_A .
2. Collect the data.
3. Compute the observed value of the appropriate test statistic
4. Either use the table to find the critical value(s) for the problem or the p -value if using the computer. The critical value(s) make use of α and the alternative hypothesis.
5. Finally, compare the observed value with the critical value(s) or the p -value with α to determine if the difference is statistically significant.

Each test has a different test statistic and a different set of assumptions. All the test statistics have the form:

$$\frac{\text{observed} - \text{expected}}{\text{SE}}$$

where the Standard Error, SE, takes a number of different forms depending on the specifics of the problem. Possibilities include:

$$s/\sqrt{n}, \sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}, \sqrt{s_1^2/n_1 + s_2^2/n_2}, \sqrt{\hat{p}(1-\hat{p})/n}, \dots$$

Some sample problems follow:

1. Are women more likely to smoke than men? A random sample of 210 men found that 65 smoked, while a similar sample of women found that 87 of 240 smoked. Are the percentages of smokers statistically different for men and women? Formulate Null and Alternative hypothesis. Test the hypothesis - determine p -values - report your results.
2. Lucky Steel Works produces high precision engine castings for General Motors. A quality control team finds that of 250 castings produced on machine A, 19 were out of tolerance, therefore defective. The same team found that 19 of 300 castings produced on machine C were defective. Are the machines different? Should one be taken out of service?
 - (a) Calculate individual sample proportions for the two machines.
 - (b) Calculate the 99% confidence interval for the difference in proportions of defective parts produced by the two machines.
 - (c) Test the hypothesis that the machines produce different proportions of defective parts. Is there a statistical difference at the $\alpha = 0.05$ level?
3. Have gas prices gone up from last month? Suppose it was well known that average gas prices last month were \$2.35. This month a random sample of 10 gas stations found a mean of \$2.60 with a standard deviation of \$0.12. Is the difference statistically significant at the $\alpha = 0.05$ level?
4. According to a new report from Israel, taking aspirin up to the day of a coronary bypass grafting, as opposed to stopping one week prior to surgery, seems to speed lung function recovery after the surgery without increasing the risk of bleeding significantly.

Suppose the data summarizing the study are

Group	n	xbar	s
took aspirin	14	5	3
stopped aspirin	18	7	2.5

The variable x records the time spent on a ventilator after surgery.

Is there statistically significant evidence, at the $\alpha = 0.05$ level that the mean time spent on the ventilator has decreased?

5. Is Weight Watchers a better diet than the Atkins diet for long-term weight loss? According to a May 2005 *Consumer Reports* this is the case. Suppose the data used to support the claim came from a study of two groups of dieters for one year. The amount of weight lost over that one-year period is measured, and summarized below.

Group	n	xbar	s
Weight Watchers	25	4.5	3.5
Atkins Diet	20	4.2	5.0

Is the difference in weight loss statistically significant at the $\alpha = 0.05$ level?

6. A news article states

BOSTON - Two research reports suggest a possible link between two bone-building drugs and irregular heart rhythms in a small number of women who take the medicine.

This study of 7,736 postmenopausal women with bone-thinning osteoporosis focuses on Reclast. Novartis recently won approval to sell Reclast, known generically as zoledronic acid, for Paget's disease, another bone condition. The company hopes to get an OK later this year to sell it for osteoporosis use.

The new study, funded by Novartis, shows that Reclast works at least as well as existing drugs in the same class, researchers say. However, the risk of a serious case of irregular heart rhythm was more than double that in the other patients, 50 cases in the drug-taking half compared to 20 cases in the control study.

- (a) In the control case 20 out of half of 7736 developed issues. Is this in agreement with the known value of $p = 0.005$? To see if it is, find 90% and 95% confidence intervals for the population proportion based on the sample data: 20 out of 7736/2.
- (b) Does taking Reclast increase the chances of irregular heart rhythm? Do a test of

$$H_0 : \pi = 0.005, \quad H_a : \pi > 0.005$$

for the population proportion of those taking Reclast. Their data is

Reclast	n	x
	3868	50