

Homework 11, due April 22 2010.

9.32, 9.36 (page 440)

10.6, 10.12 (pages 480/481)

Note: In 10.6, don't just take the "95% CI for difference" and the "Test for difference" as given: show how they were derived.

Question 9.79, page 462.

For a quantitative variable you want to test $H_0 : \mu = 0$ against $H_a : \mu \neq 0$. The 10 observations are 3,7,3,3,0,8,1,12,5,8.

1. Show: (a) $\bar{x} = 5.0$, (b) $s = 3.71$, (c) standard error is 1.17, (d) test statistic is 4.26, (e) $df=9$.
2. The P-value is 0.002. Interpret, and make a decision using a significance level of 0.05.
3. If you had instead used $H_a : \mu > 0$, what would the P-value be? Interpret.
4. If you had instead used $H_a : \mu < 0$, what would the P-value be? Interpret.

Solution

1. (a) and (b) are direct calculation (you could also use Excel).
(c) $\frac{3.71}{\sqrt{10}} = 1.17$. (d) $t = \frac{5-0}{1.17} = 4.27$ (slight rounding error).
(e) $df = n - 1 = 10 - 1 = 9$.
2. Since $0.002 < 0.05$, we reject the null hypothesis. Based on the data, there seems to be significant evidence that the mean is not 0.
3. For a one-sided test, the P-value is half that for a two-sided test, i.e. 0.001. This also leads to the conclusion that the mean is not 0.
4. The P-value for $H_a : \mu < 0$ would be the probability that $t < 4.27$ based on $df = 9$. This is 1 minus the probability that $t > 4.27$, i.e. about 0.999. In this case, since the only alternative of interest is $\mu < 0$ and the data support $\mu > 0$, we have no choice but to accept H_0 .

Limitations of Significance Tests

1. Statistical significance does not necessarily mean practical significance.
2. Significance tests are generally less useful than confidence intervals.

Example. Suppose we take a group of 200 patients whose mean cholesterol level is 250 or higher. We given them a cholesterol-reducing drug and measure the reduction in cholesterol after some period of time. Over all the patients, the mean reduction in cholesterol is 10 points and the standard deviation is 50.

Is this a significant result?

The standard error is $\frac{50}{\sqrt{200}} = 3.53$ so we calculate $z = \frac{10}{3.53} = 2.83$. The associated P-value is approximately 0.002 (one-sided). So it is (statistically) significant.

But is it practically significant? Most doctors recommend a cholesterol level lower than 200 — a reduction of 10 points is not necessarily of practical significance for someone over 250.

In other words, the difference is big enough that it could not have occurred by chance (that's the definition of statistical significance), but that doesn't mean the drug is effective in practice.

The alternative approach is to calculate a confidence interval for the mean reduction in cholesterol.

Since $\bar{x} = 10$ and the standard error is 3.53, a 95% confidence interval is $10 \pm 1.96 \times 3.53 = (3.1, 16.9)$. In the opinion of most statisticians, that gives a much clearer picture of the uncertainty in the data than simply saying we rejected the null hypothesis $\mu = 0$.

Common Misinterpretations of Significance Tests

1. If a significance test results in the conclusion “do not reject H_0 ”, that does not mean H_0 is true.
2. The P-value is *not* the probability that H_0 is true.
3. Many medical researchers only report their results if they are statistically significant. This leads to *publication bias*.
4. Some results are statistically significant by chance. Therefore, you cannot conclude that every “significant” result published in the literature is true.

Type I and Type II Errors

Type I Error: H_0 true but rejected

Type II Error: H_0 false but not rejected

The significance level is the probability of a type I error.

Example

A new drug therapy is proposed to reduce the risk of heart attack.

Among the category of patients for whom the drug is intended, the chance of a heart attack within 5 years is considered to be 25%.

We test the drug on 120 patients so that the type I error is 0.01 in a one-sided test ($H_0 : p = 0.25$ against $H_a : p < 0.25$). Suppose the drug is effective, so that for a patient who takes the drug, the chance of a heart attack within 5 years is only 10%. What is the probability of a type II error?

First define the test for a significance level of 0.01.

The standard error is $\sqrt{\frac{0.25 \times 0.75}{120}} = 0.0395$.

Reject H_0 if $\hat{p} < 0.25 - 2.33 \times 0.0395 = 0.158$.

(The z value 2.33 comes from the 0.01 significance level, via Table A.)

If $p = 0.1$, the standard error is now $\sqrt{\frac{0.1 \times 0.9}{120}} = 0.0274$.

The probability that $\hat{p} < 0.158$ is derived from $z = \frac{0.158 - 0.1}{0.0274} = 2.12$.

The probability associated with that value of z is 0.983.

The *probability of type II error* is $1 - 0.983 = 0.017$.

Example: Question 9.63, page 458.

For the example of an astrologer trying to identify someone's personality profile, the probability that the astrologer is correct if she is randomly guessing is $\frac{1}{3}$. Suppose we test $H_0 : p = \frac{1}{3}$ against $H_a : p \neq \frac{1}{3}$ when the sample size is $n = 116$ and the significance level is 0.05. Suppose also the true value of p is $\frac{1}{2}$.

1. Show that a type II error occurs if $0.248 < \hat{p} < 0.419$.
2. Show that when $p = \frac{1}{2}$, the probability that $\hat{p} < 0.248$ is 0 and the probability that $\hat{p} > 0.419$ is 0.96.
3. Show that the probability of a type II error is 0.04.

Solution

1. For null hypothesis $p_0 = \frac{1}{3}$, the standard error is $\sqrt{\frac{p_0(1-p_0)}{116}} = .0438$ so we reject H_0 if $\hat{p} < 0.333 - 1.96 \times .0438 = 0.248$ or if $\hat{p} > 0.333 + 1.96 \times .0438 = 0.419$ (you may get slightly different answers because of rounding error)
2. If $p = 0.5$, the standard deviation of $\hat{p}$ is $\sqrt{\frac{0.5 \times 0.5}{116}} = .0464$. For the probability that $\hat{p} < x$ when $x = .248$, we calculate $z = \frac{.248 - .5}{.0464} = -5.43$, for which the associated left-tail probability is 0.00 to 2 decimal places. For $x = .419$, compute $z = \frac{.419 - .5}{.0464} = -1.75$, for which the left-tail probability is 0.04. Hence the right-tail probability is 0.96, as required.
3. If the true p is 0.5, then the probability that $\hat{p}$ is between 0.248 and 0.419 is $0.04 - 0.00 = 0.04$. Hence this is the probability of a type II error.

Comparing Two Groups

In this chapter (Chapter 10), we consider problems of estimation or hypothesis testing when we have to compare two groups of observations, instead of comparing one group with a predetermined standard.

Comparing Two Proportions

Dermatology example: among “light runners”, 5 out of 78 subjects had to be referred for a skin lesion conditions.

Among “heavy runners”, the number was 6 out of 31.

Is this significant evidence that the two groups are different?

Suppose the true proportions of individuals who have the critical skin condition are p_1 for the light runners group, and p_2 for the heavy runners group.

I: Form a confidence interval for $p_2 - p_1$.

II: Test the hypothesis $H_0 : p_1 = p_2$ against $H_a : p_1 \neq p_2$.

Confidence interval calculation:

$$\hat{p}_1 = \frac{5}{78} = 0.0641,$$

$$\text{standard error } SE_1 = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}} = \sqrt{\frac{0.0641 \times 0.9359}{78}} = .0277.$$

$$\hat{p}_2 = \frac{6}{31} = 0.1935,$$

$$\text{standard error } SE_2 = \sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{0.1935 \times 0.8065}{31}} = .0710.$$

The combined standard error is defined to be

$$SE = \sqrt{SE_1^2 + SE_2^2} = .0762.$$

Note that $\hat{p}_2 - \hat{p}_1 = 0.129$.

95% confidence interval is $0.129 \pm 1.96 \times .0762 = (0.043, 0.341)$.