TESTING HYPOTHESES WITH ONE SAMPLE

(DA_2022)

Andrea Giansanti
Dipartimento di Fisica, Sapienza Università di Roma

Andrea.Giansanti@roma1.infn.it

Lecture n. 14, Rome 20th April 2022

DIPARTIMENTO DI FISICA

SAPIENZA
UNIVERSITÀ DI ROMA

Outline DA_2022 L 14

- One sample/two sample tests
- One sided test/two sided test
- Parametric/non parametric tests
- Type I and Type II errors, Power of a test
- Test flowchart (R p.268)
- One sample test for the mean of a normal variable (one sided/two sided test R 7.3/R.7.4))
- Acceptance/rejection regions
- P-values

Rosner's chapter 7 and Whitlock's chap. 6

As an home work for next mondy please complete by yourself the study of the first paragraphs (7.1,7.2, 7.3 and 7.4 by Rosner)

In particular consider the review questions 7A p.222 of R IN THE LOGBOOK

Summary

From W&S chap. 6

- The four steps of hypothesis testing are (1) state the hypotheses; (2) compute the test statistic; (3) determine the *P*-value; and (4) draw the appropriate conclusions.
- Hypothesis testing uses data to decide whether a parameter equals the value stated in a null hypothesis. If the data are too unusual, assuming the null hypothesis is true, then we reject the null hypothesis.
- The null hypothesis (H_0) is a specific claim about a parameter. The null hypothesis is the default hypothesis, the one assumed to be true unless the data lead us to reject it. A good null hypothesis would be interesting if rejected.
- The alternative hypothesis (H_A) usually includes all values for the parameter other than that stated in the null hypothesis.
- The test statistic is a quantity calculated from data, used to evaluate how compatible the data are with the null hypothesis.
- The null distribution is the sampling distribution of the test statistic under the assumption that the null hypothesis is true.
- The *P*-value is the probability of obtaining a difference from the null expectation as great as or greater than that observed in the data if the null hypothesis were true. If *P* is less than or equal to α, then H_0 is rejected.
- The threshold α is called the significance level of a test. Typically, α is set to 0.05.
- The *P*-value is not the probability that the null hypothesis is true or false.
- The *P*-value reflects the weight of evidence against the null hypothesis, but *P* does not measure the size of the effect. Use confidence intervals to put bounds on the magnitude of effect.

- A Type I error is rejecting a true null hypothesis. A Type II error is failing to reject a false null hypothesis:

	Reality	
Decision	**H_0 true**	**H_0 false**
Reject H_0	Type I error	(no error)
Do not reject H_0	(no error)	Type II error

- The probability of making a Type I error is set by the significance level, α. If $\alpha = 0.05$, then the probability of making a Type I error is 0.05.
- The power of a test is the probability that a random sample, when analyzed, leads to rejection of a false null hypothesis.
- Increasing sample size increases the power of a test.
- In a two-sided test, the alternative hypothesis includes parameter values on both sides of the parameter value stated by the null hypothesis. In a one-sided test, the alternative hypothesis includes parameter values on only one side of the parameter value stated by the null hypothesis.
- Most hypothesis tests are two-sided. One-sided tests should be restricted to rare instances in which a parameter value on one side of the null value is inconceivable.

THE BASIC STEPS

Hypothesis testing: an example

To show you the basic concepts and terminology of hypothesis testing, we'll take you through all the steps by using an example. Our goal is to illuminate the basic process without distraction from the details of the probability calculations. We'll get to plenty of the details in later chapters.

Four basic steps are involved in hypothesis testing:

1. State the hypotheses.
2. Compute the test statistic.
3. Determine the *P*-value.
4. Draw the appropriate conclusions.

We'll define the new terms we just used in this section.

Example 6.2 tests a hypothesis about a proportion, but hypothesis testing can address a wide variety of quantities, such as means, variances, differences in means, correlations, and so on. We'll try to emphasize the general over the specific here. Further details of how to test hypotheses about proportions are discussed in Chapter 7.

From: Michael C. Whitlock and Dolph Schluter - The Analysis of Biological Data-W. H. Freeman and Company (2015)

THE ROADMAP I

7.13 SUMMARY

In this chapter some of the fundamental ideas of hypothesis testing were introduced: (1) specification of the null (H_0) and alternative (H_1) hypotheses; (2) type I error (α), type II error (β), and power ($1 - \beta$) of a hypothesis test; (3) the *p*-value of a hypothesis test; and (4) the distinction between one-sided and two-sided tests. Methods for estimating the appropriate sample size for a proposed study as determined by the prespecified null and alternative hypotheses and type I and type II errors were also discussed.

These general concepts were applied to several one-sample hypothesis-testing situations:

(1) The mean of a normal distribution with unknown variance (one-sample *t* test)

(2) The mean of a normal distribution with known variance (one-sample *z* test)

(3) The variance of a normal distribution (one-sample χ^2 test)

(4) The parameter *p* of a binomial distribution (one-sample binomial test)

(5) The expected value μ of a Poisson distribution (one-sample Poisson test)

Each of the hypothesis tests can be conducted in one of two ways:

(1) Specify critical values to determine the acceptance and rejection regions (critical-value method) based on a specified type I error α.

(2) Compute *p*-values (*p*-value method).

These methods were shown to be equivalent in the sense that they yield the same inferences regarding acceptance and rejection of the null hypothesis.

Furthermore, the relationship between the hypothesis-testing methods in this chapter and the CI methods in Chapter 6 was explored. We showed that the inferences that can be drawn from using these methods are usually the same.

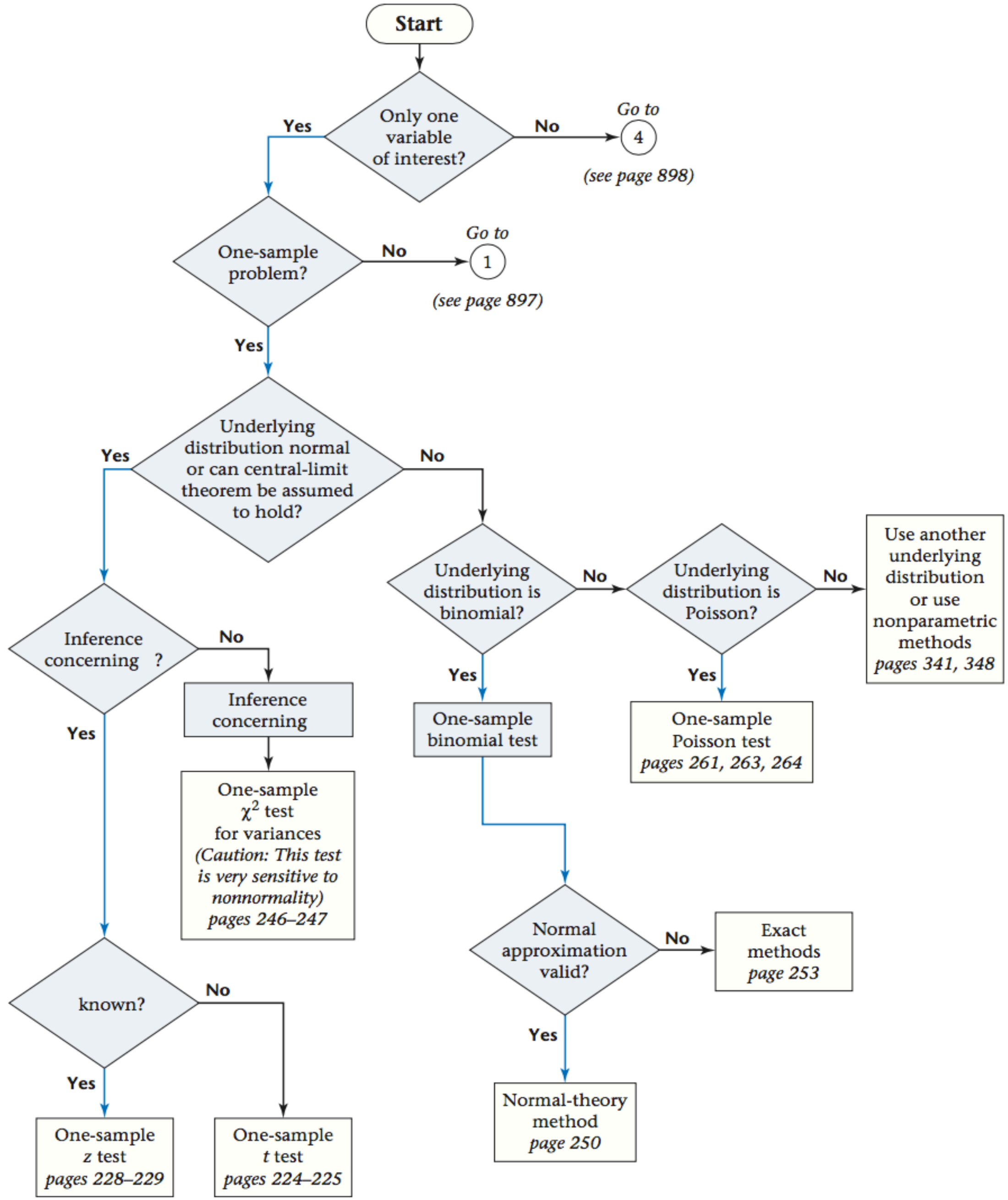
THE ROADMAP II

Many hypothesis tests are covered in this book. A master flowchart (pp. 895–902) is provided at the back of the book to help clarify the decision process in selecting the appropriate test. The flowchart can be used to choose the proper test by answering a series of yes/no questions. The specific hypothesis tests covered in this chapter have been presented in an excerpt from the flowchart shown in Figure 7.18 and have been referred to in several places in this chapter. For example, if we are interested in performing hypothesis tests concerning the mean of a normal distribution with known variance, then, beginning at the "Start" box of the flowchart, we would answer *yes* to each of the following questions: (1) only one variable of interest? (2) one-sample problem? (3) underlying distribution normal or can central-limit theorem be assumed to hold? (4) inference concerning μ? (5) σ known? The flowchart leads us to the box on the lower left of the figure, indicating that the one-sample z test should be used. In addition, the page number(s) where a specific hypothesis test is discussed is also provided in the appropriate box of the flowchart. The boxes marked "Go to 1" and "Go to 4" refer to other parts of the master flowchart in the back of the book.

The study of hypothesis testing is extended in Chapter 8 to situations in which two different samples are compared. This topic corresponds to the answer *yes* to the question (1) only one variable of interest? and no to (2) one-sample problem?

The FLOW CHART

FIGURE 7.18 Flowchart for appropriate methods of statistical inference



Type I and Type II errors

There are two kinds of errors in hypothesis testing, prosaically named Type I and Type II.

Rejecting a true null hypothesis is a **Type I error**. Failing to reject a false null hypothesis is a **Type II error**. Both types of error are summarized in Table 6.3-1.

Type I error is rejecting a true null hypothesis. The significance level α sets the probability of committing a Type I error.

Type II error is failing to reject a false null hypothesis.

TABLE 6.3-1 Types of error in hypothesis testing.

	Reality	
Conclusion	H_0 true	H_0 false
Reject H_0	Type I error	Correct
Do not reject H_0	Correct	Type II error

The significance level, α, gives us the probability of committing a Type I error. If we go along with convention and use a significance level of $\alpha = 0.05$, then we reject H_0 whenever P is less than or equal to 0.05. This means that, if the null hypothesis were true, we would reject it mistakenly one time in 20. Biologists typically regard this as an acceptable error rate.

EQUATION 7.4

Guidelines for Judging the Significance of a p-Value

If $.01 \le p < .05$, then the results are *significant*.
If $.001 \le p < .01$, then the results are *highly significant*.
If $p < .001$, then the results are *very highly significant*.
If $p > .05$, then the results are considered *not statistically significant* (sometimes denoted by NS).
However, if $.05 < p < .10$, then a trend toward statistical significance is sometimes noted.

EQUATION 7.5

Determination of Statistical Significance for Results from Hypothesis Tests

Either of the following methods can be used to establish whether results from hypothesis tests are statistically significant:

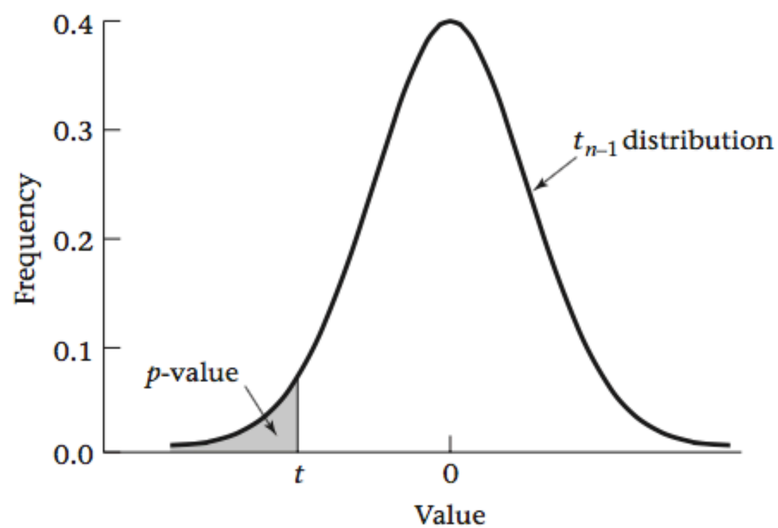
(1) The test statistic t can be computed and compared with the critical value $t_{n-1,\,\alpha}$ at an α level of .05. Specifically, if $H_0: \mu = \mu_0$ vs. $H_1: \mu < \mu_0$ is being tested and $t < t_{n-1,.05}$, then H_0 is rejected and the results are declared *statistically significant* ($p < .05$). Otherwise, H_0 is accepted and the results are declared *not statistically significant* ($p \ge .05$). We have called this approach the **critical-value method** (see Definition 7.12).

(2) The exact p-value can be computed and, if $p < .05$, then H_0 is rejected and the results are declared *statistically significant*. Otherwise, if $p \ge .05$, then H_0 is accepted and the results are declared *not statistically significant*. We will refer to this approach as the **p-value method.**

DEFINITION 7.13 The ***p*-value** for any hypothesis test is the α level at which we would be indifferent between accepting or rejecting H_0 given the sample data at hand. That is, the p-value is the α level at which the given value of the test statistic (such as t) is on the borderline between the acceptance and rejection regions.

DEFINITION 7.14 The ***p*-value** can also be thought of as the probability of obtaining a test statistic as extreme as or more extreme than the actual test statistic obtained, given that the null hypothesis is true.

Graphic display of a *p*-value



We know that under the null hypothesis, the t statistic follows a t_{n-1} distribution. Hence, the probability of obtaining a t statistic that is no larger than t under the null hypothesis is $Pr(t_{n-1} \le t) = p$-value, as shown in Figure 7.1.

7.4 ONE-SAMPLE TEST FOR THE MEAN OF A NORMAL DISTRIBUTION: TWO-SIDED ALTERNATIVES

See for this VI case also W&S chap 111

Cardiovascular Disease Suppose we want to compare fasting serum-cholesterol levels among recent Asian immigrants to the United States with typical levels found in the general U.S. population. Suppose we assume cholesterol levels in women ages 21–40 in the United States are approximately normally distributed with mean 190 mg/dL. It is unknown whether cholesterol levels among recent Asian immigrants are higher or lower than those in the general U.S. population. Let's assume that levels among recent female Asian immigrants are normally distributed with unknown mean μ. Hence we wish to test the null hypothesis $H_0: \mu = \mu_0 = 190$ vs. the alternative hypothesis $H_1: \mu \neq \mu_0$. Blood tests are performed on 100 female Asian immigrants ages 21–40, and the mean level ($\bar{x}$) is 181.52 mg/dL with standard deviation = 40 mg/dL. What can we conclude on the basis of this evidence?

The type of alternative given in Example 7.20 is known as a *two-sided* alternative because the alternative mean can be either less than or greater than the null mean.

A two-tailed test is a test in which the values of the parameter being studied (in this case μ) under the alternative hypothesis are allowed to be either *greater than or less than* the values of the parameter under the null hypothesis (μ_0).

A reasonable decision rule to test for alternatives on *either* side of the null mean is to *reject H_0 if t is either too small or too large*. Another way of stating the rule is that H_0 will be rejected if t is either $< c_1$ or $> c_2$ for some constants c_1, c_2 and H_0 will be accepted if $c_1 \le t \le c_2$.

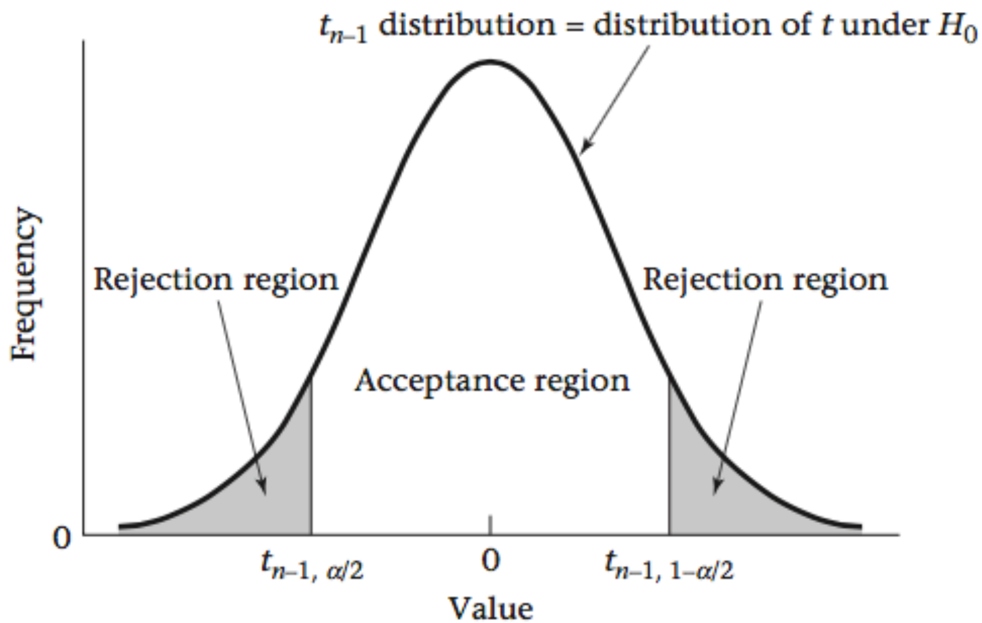
The question remains: What are appropriate values for c_1 and c_2? These values are again determined by the type I error (α). The constants c_1, c_2 should be chosen such that

$$Pr(\text{reject } H_0 | H_0 \text{ true}) = Pr(t < c_1 \text{ or } t > c_2 | H_0 \text{ true})$$
$$= Pr(t < c_1 | H_0 \text{ true}) + Pr(t > c_2 | H_0 \text{ true}) = \alpha$$

Half of the type I error is assigned arbitrarily to each of the probabilities on the left side of the second line of Equation 7.8. Thus, we wish to find c_1, c_2 so that

$$Pr(t < c_1 | H_0 \text{ true}) = Pr(t > c_2 | H_0 \text{ true}) = \alpha/2$$

FIGURE 7.3 One-sample t test for the mean of a normal distribution (two-sided alternative)



We know t follows a t_{n-1} distribution under H_0. Because $t_{n-1,\alpha/2}$ and $t_{n-1,1-\alpha/2}$ are the lower and upper $100\% \times \alpha/2$ percentiles of a t_{n-1} distribution, it follows that

$$Pr\left(t < t_{n-1,\alpha/2}\right) = Pr\left(t > t_{n-1,1-\alpha/2}\right) = \alpha/2$$

Therefore,

$$c_1 = t_{n-1,\alpha/2} = -t_{n-1,1-\alpha/2} \qquad \text{and} \qquad c_2 = t_{n-1,1-\alpha/2}$$

Summary

One-Sample t Test for the Mean of a Normal Distribution with Unknown Variance (Two-Sided Alternative)

To test the hypothesis H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$, with a significance level of α, the best test is based on $t = (\bar{x} - \mu_0)/(s/\sqrt{n})$.

If $|t| > t_{n-1,1-\alpha/2}$
then H_0 is rejected.

If $|t| \leq t_{n-1,1-\alpha/2}$
then H_0 is accepted.

The acceptance and rejection regions for this test are shown in Figure 7.3.

Cardiovascular Disease Test the hypothesis that the mean cholesterol level of recent female Asian immigrants is different from the mean in the general U.S. population, using the data in Example 7.20.

Solution: We compute the test statistic

$$
\begin{aligned}
t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \\
&= \frac{181.52 - 190}{40/\sqrt{100}} \\
&= \frac{-8.48}{4} = -2.12
\end{aligned}
$$

For a two-sided test with $\alpha = .05$, the critical values are $c_1 = t_{99,.025}$, $c_2 = t_{99,.975}$.

From Table 5 in the Appendix, because $t_{99,.975} < t_{60,.975} = 2.000$, it follows that $c_2 < 2.000$. Also, because $c_1 = -c_2$ it follows that $c_1 > -2.000$. Because $t = -2.12 < -2.000 < c_1$, it follows that we can reject H_0 at the 5% level of significance. We conclude that the mean cholesterol level of recent Asian immigrants is significantly different from that of the general U.S. population.

Alternatively, we might want to compute a p-value as we did in the one-sided case. The p-value is computed in two different ways, depending on whether t is less than or greater than 0.

p-Value for the One-Sample *t* Test for the Mean of a Normal Distribution (Two-Sided Alternative)

Let $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

$$p = \begin{cases} 2 \times Pr(t_{n-1} \le t), \text{ if } t \le 0 \\ 2 \times [1 - Pr(t_{n-1} \le t)], \text{ if } t > 0 \end{cases}$$

Thus, in words, if $t \le 0$, then $p = 2$ times the area under a t_{n-1} distribution to the left of t; if $t > 0$, then $p = 2$ times the area under a t_{n-1} distribution to the right of t. One way to interpret the p-value is as follows.

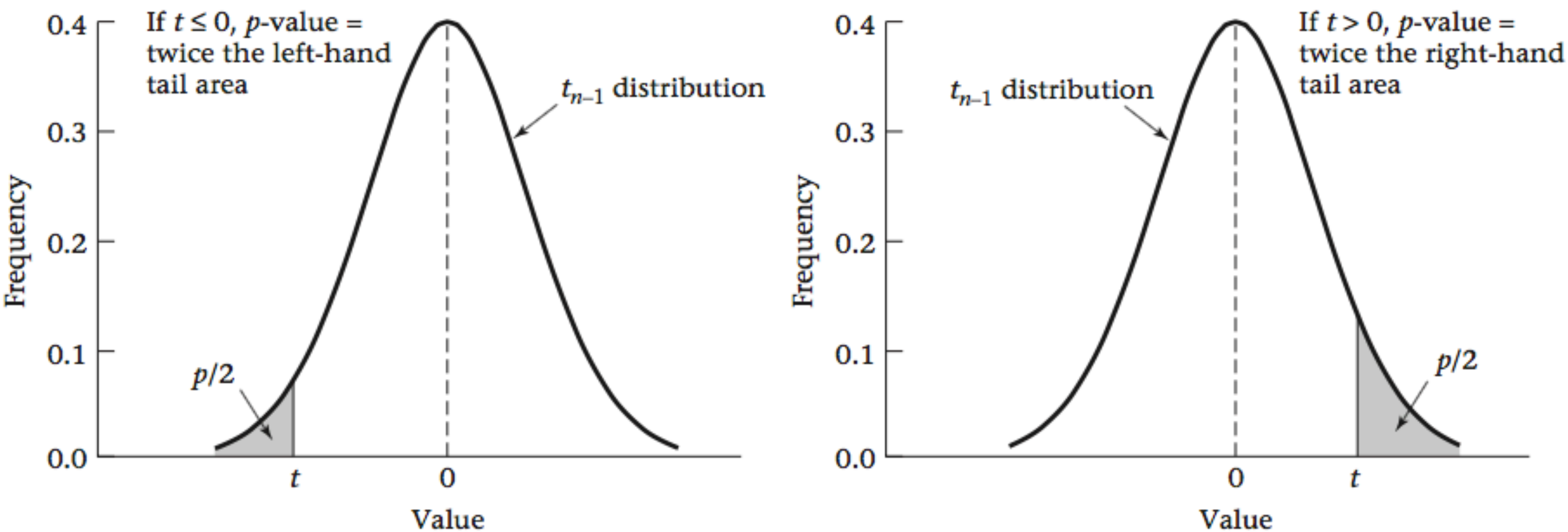
The ***p*-value** is the probability under the null hypothesis of obtaining a test statistic as extreme as or more extreme than the observed test statistic, where, because a two-sided alternative hypothesis is being used, extremeness is measured by the **absolute value** of the test statistic.

Hence, if $t > 0$, the p-value is the area to the right of t plus the area to the left of $-t$ under a t_{n-1} distribution.

However, this area simply amounts to twice the right-hand tail area because the t distribution is symmetric around 0. A similar interpretation holds if $t < 0$.

These areas are illustrated in Figure 7.4.

FIGURE 7.4 Illustration of the p-value for a one-sample t test for the mean of a normal distribution (two-sided alternative)



When is a one-sided test more appropriate than a two-sided test? Generally, the sample mean falls in the expected direction from μ_0 and it is *easier* to reject H_0 using a one-sided test than using a two-sided test. However, this is not necessarily always the case. Suppose we guess from a previous review of the literature that the cholesterol level of Asian immigrants is likely to be lower than that of the general U.S. population because of better dietary habits. In this case, we would use a one-sided test of the form $H_0: \mu = 190$ vs. $H_1: \mu < 190$. From Equation 7.3, the one-sided p-value $= Pr(t_{99} < -2.12) = pt(-2.12, 99) = .018 = \frac{1}{2}$ (two-sided p-value). Alternatively, suppose we guess from a previous literature review that the cholesterol level of Asian immigrants is likely to be higher than that of the general U.S. population because of more stressful living conditions. In this case, we would use a one-sided test of the form $H_0: \mu = 190$ vs. $H_1: \mu > 190$. From Equation 7.6, the p-value $= Pr(t_{99} > -2.12)$ $= .982$. Thus, we would accept H_0 if we use a one-sided test and the sample mean is on the opposite side of the null mean from the alternative hypothesis. Generally, a two-sided test is always appropriate because there can be no question about the conclusions. Also, as just illustrated, a two-sided test can be more conservative because you need not guess the appropriate side of the null hypothesis for the alternative hypothesis. However, in certain situations only alternatives on one side of the null mean are of interest or are possible, and in this case a one-sided test is better because it has more power (that is, it is easier to reject H_0 based on a finite sample if H_1 is actually true) than its two-sided counterpart. In all instances, it is important to decide whether to use a one-sided or a two-sided test *before* data analysis (or preferably before data collection) begins so as not to bias conclusions based on results of hypothesis testing. In particular, do not change from a two-sided to a one-sided test *after* looking at the data.

For today that’s all folks

As an home work for next mondy please complete by yourself the study of the first paragraphs (7.1,7.2, 7.3 and 7.4 by Rosner) In particular consider the review questions 7A p.222 of R IN THE LOGBOOK