TESTING HYPOTHESES WITH TWO SAMPLES (DA_2022)

Andrea Giansanti
Dipartimento di Fisica, Sapienza Università di Roma

Andrea.Giansanti@roma1.infn.it

Lecture n. 16&17, Rome Mon. 2nd and 4th of May 2022

DIPARTIMENTO DI FISICA

SAPIENZA
UNIVERSITÀ DI ROMA

OutlineDa_2022_L 16

One sample vs two sample tests MOTIVATION

General reference for this topic: Rosner's chapter 8 and Whitlock's chap. 12

Assumption of normality (see Whitlock chap. 13)

Two samples (cross-sectional, synchronic studies) vs longitudinal paired tests (longitudinal, diachronic studies)

The paired t-test (R 8.2) [paired t-test statistic, acceptance region, p-value, interval estimation (R.8.3)]

Two sample test for independent samples with equal variances: acceptance region, p-value (R. 8.4), interval estimation (R.8.5)

& L17

- Testing for the equality of variances (R section 8.6): the F distribution, The F-test
- Two sample t test for independent samples with different variances (R. 8.7) (self study)

MOTIVATION

Biological data are often gathered to compare different group or treatment means. Do female hyenas differ from male hyenas in body size? Do patients treated with a new drug live significantly longer than those treated with the old drug? Do students perform better on tests if they stay up late studying or get a good night's rest? In Chapter 8, we presented methods to compare proportions of a categorical variable between different groups. In this chapter, we develop procedures for comparing means of a numerical variable between two treatments[1] or groups. We also include methods to compare two variances. All of the methods in the current chapter assume that the measurements are normally distributed in the populations.

We show analyses for two different study designs. In the paired design, both treatments have been applied to every sampled unit, such as a subject or a plot, at different times or on different sides of the body or of the plot. In the two-sample design, each group constitutes an independent random sample of individuals. In both cases, we make use of the *t*-distribution to

from W&S chap.12

Paired sample versus two independent samples

There are two study designs to measure and test differences between the means of two treatments. To describe them, let's use an example: does clear-cutting a forest affect the number of salamanders present? Here we have two treatments (clear-cutting/no clear-cutting), and we want to know if the mean of a numerical variable (the *number of salamanders*) differs between them. "Clear-cut" is the treatment of interest, and "no clear-cut" is the control. This is the same as asking whether these two variables, *treatment* (a categorical variable) and *salamander number* (a numerical variable), are associated.

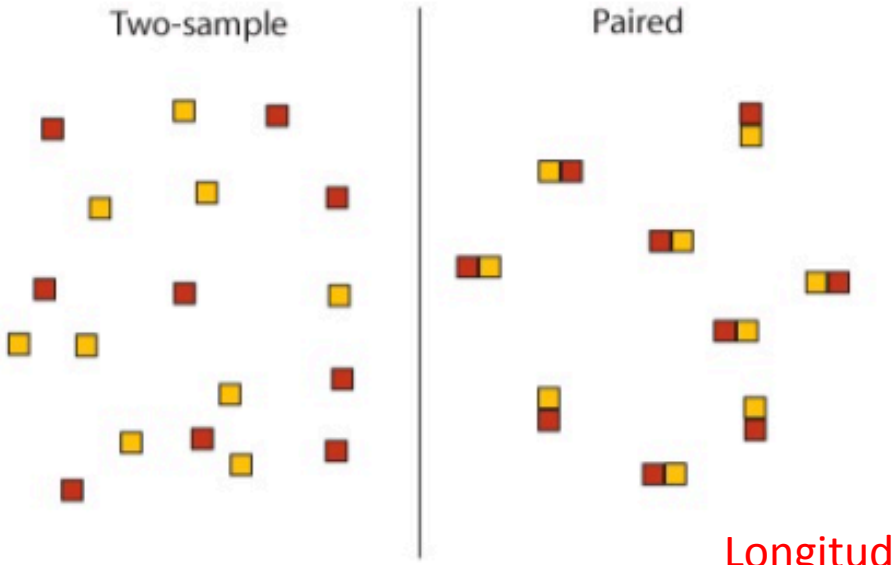
We can design this kind of a study in either of two ways: a two-sample design (see the left panel in Figure 12.1-1) or a paired design (see the right panel in Figure 12.1-1). In the *two-sample* design, we take a random sample of forest plots from the population and then randomly assign either the clear-cut treatment or the no-clear-cut treatment to each plot. In this case, we end up with two independent samples, one from each treatment. The difference in the mean number of salamanders between the clear-cut and no-clear-cut areas estimates the effect of clear-cutting on salamander number.[2]

In the *paired* design, we take a random sample of forest plots and clear-cut a randomly chosen half of each plot, leaving the other half untouched. Afterward, we count the number of salamanders in each half. The mean difference between the two sides estimates the effect of clear-cutting.

from W&S chap.12

Paired sample versus two independent samples

Cross sectional study



Longitudinal follow-up study

Figure 12.1-1
Whitlock et al., *The Analysis of Biological Data*, 2e, © 2015
W. H. Freeman and Company

FIGURE 12.1-1 Alternative designs to compare two treatments. A two-sample design is on the left; the paired design is on the right. Freestanding blocks represent sampling units, such as plots. The red and gold areas represent different treatments (e.g., clear-cut and not clear-cut). In the paired design (*right*), both treatments are applied to every unit. In the two-sample design (*left*), different treatments are applied to separate units.

from W&S chap.12

Paired comparison of means

- Comparing patient weight before and after hospitalization
- Comparing fish species diversity in lakes before and after heavy metal contamination?
- Testing effects of sunscreen applied to one arm of each subject compared with a placebo applied to the other arm
- Testing effects of smoking in a sample of smokers, each of which is compared with a nonsmoker closely matched by age, weight, and ethnic background

Paired measurements are converted to a single measurement by taking the difference between them.

PAIRED TESTS AND CONFOUNDING VARIABLES…

Paired designs are usually more powerful than unpaired designs, because they control for a lot of the extraneous variation between plots or sampling units that sometimes obscures the effects we are looking for. It is easier to see a real difference between two treatments if nearly everything else is similar between sides of the same plot or sampling unit. Very often, though, a paired design is just not possible.

from W&S chap.12

DEFINITION 8.1 In a **two-sample hypothesis-testing problem,** the underlying parameters of two different populations, *neither of whose values is assumed known,* are compared.

Comparison means looking for differences which should be statistically tested
We can use either a LONGITUDINAL (diachronic) design or a CROSS SECTIONAL (synchronic) study

PAIRED SAMPLES vs INDEPENDENT SAMPLES

As previously discussed in the MOTIVATION

CASE/CONTROL APPROACH

Hypertension Let's say we are interested in the relationship between oral contraceptive (OC) use and blood pressure in women.

Two different experimental designs can be used to assess this relationship. One method involves the following design:

Longitudinal Study

(1) Identify a group of nonpregnant, premenopausal women of childbearing age (16–49 years) who are not currently OC users, and measure their blood pressure, which will be called the *baseline blood pressure*.

(2) Rescreen these women 1 year later to ascertain a subgroup who have remained nonpregnant throughout the year and have become OC users. This subgroup is the study population.

(3) Measure the blood pressure of the study population at the follow-up visit.

(4) Compare the baseline and follow-up blood pressure of the women in the study population to determine the difference between blood pressure levels of women when they *were* using the pill at follow-up and when they *were not* using the pill at baseline.

From Rossner's chap 8

Cross-Sectional Study

(1) Identify both a group of OC users and a group of non-OC users among non-pregnant, premenopausal women of childbearing age (16–49 years), and measure their blood pressure.

(2) Compare the blood pressure level between the OC users and nonusers.

In a **longitudinal** or **follow-up study** the same group of people is followed *over time*.

In a **cross-sectional study**, the participants are seen at only one point in time.

There is another important difference between these two designs. The longitudinal study represents a *paired-sample* design because each woman is used as her own control. The cross-sectional study represents an *independent-sample* design because two completely different groups of women are being compared.

Two samples are said to be **paired** when each data point in the first sample is matched and is related to a unique data point in the second sample.

The paired samples may represent two sets of measurements on the same people. In this case each person is serving as his or her own control, as in Equation 8.1. The paired samples may also represent measurements on different people who are chosen on an individual basis using matching criteria, such as age and sex, to be very similar to each other.

Two samples are said to be **independent** when the data points in one sample are unrelated to the data points in the second sample.

The samples in Equation 8.2 are completely independent because the data are obtained from unrelated groups of women.

Controlling confounding factors

Which type of study is better in this case? The first type of study is probably more definitive because most other factors that influence a woman's blood pressure at the first screening (called confounders) will also be present at the second screening and will not influence the comparison of blood-pressure levels at the first and second screenings. However, the study would benefit from having a control group of women who remained non-OC users throughout the year. The control group would allow us the chance of ruling out other possible causes of blood pressure change besides changes in OC status. The second type of study, by itself, can only be considered suggestive because other confounding factors may influence blood pressure in the two samples and cause an apparent difference to be found where none is actually present.

THE MORAL: EVERYBODY IS SEARCHING FOR SUGGESTIONS (CROSS SECTION STUDIES, EXPLORATORY, NOT SO EXPENSIVE, THEN ASK A GRANT TO MAKE A MORE EXTENSIVE DIACHRONIC LONGITUDINAL STUDY IN WHICH CONFOUNDING FACTORS ARE "CANCELLED"

8.4 TWO-SAMPLE *t* TEST FOR INDEPENDENT SAMPLES WITH EQUAL VARIANCES

Hypertension Suppose a sample of eight 35- to 39-year-old nonprenant, premenopausal OC users who work in a company and have a mean systolic blood pressure (SBP) of 132.86 mm Hg and sample standard deviation of 15.34 mm Hg are identified. A sample of 21 nonpregnant, premenopausal, non-OC users in the same age group are similarly identified who have mean SBP of 127.44 mm Hg and sample standard deviation of 18.23 mm Hg. What can be said about the underlying mean difference in blood pressure between the two groups?

Assume SBP is normally distributed in the first group with mean μ_1 and variance σ_1^2 and in the second group with mean μ_2 and variance σ_2^2. We want to test the hypothesis H_0: $\mu_1 = \mu_2$ vs. H_1: $\mu_1 \neq \mu_2$. Assume in this section that the underlying variances in the two groups are the same (that is, $\sigma_1^2 = \sigma_2^2 = \sigma^2$). The means and variances in the two samples are denoted by $\bar{x}_1$ $\bar{x}_2$, s_1^2, s_2^2, respectively.

It seems reasonable to base the significance test on the difference between the two sample means, $\bar{x}_1 - \bar{x}_2$. If this difference is far from 0, then H_0 will be rejected;

otherwise, it will be accepted. Thus, we wish to study the behavior of $\bar{x}_1 - \bar{x}_2$ under H_0. We know $\bar{X}_1$ is normally distributed with mean μ_1 and variance σ^2/n_1 and $\bar{X}_2$ is normally distributed with mean μ_2 and variance σ^2/n_2. Hence, from Equation 5.10, because the two samples are independent, $\bar{X}_1 - \bar{X}_2$ is normally distributed with mean $\mu_1 - \mu_2$ and variance $\sigma^2(1/n_1 + 1/n_2)$. In symbols,

$$\bar{X}_1 - \bar{X}_2 \sim N\left[\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]$$

Under H_0, we know that $\mu_1 = \mu_2$. Thus, Equation 8.7 reduces to

$$\bar{X}_1 - \bar{X}_2 \sim N\left[0, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]$$

If σ^2 were known, then $\bar{X}_1 - \bar{X}_2$ could be divided by $\sigma\sqrt{1/n_1 + 1/n_2}$. From Equation 8.8,

$$\frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$$

and the test statistic in Equation 8.9 could be used as a basis for the hypothesis test. Unfortunately, σ^2 in general is unknown and must be estimated from the data. How can σ^2 be best estimated in this situation?

From the first and second samples, the sample variances are s_1^2 and s_2^2, respectively, each of which could be used to estimate σ^2. The average of s_1^2 and s_2^2 could simply be used as the estimate of σ^2. However, this average will weight the sample variances equally even if the sample sizes are very different from each other. The sample variances should not be weighted equally because the variance from the larger sample is probably more precise and should be weighted more heavily. The best estimate of the population variance σ^2, which is denoted by s^2, is given by a weighted average of the two sample variances, where the weights are the number of *df* in each sample.

The **pooled estimate of the variance** from two independent samples is given by

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

In particular, s^2 will then have $n_1 - 1$ *df* from the first sample and $n_2 - 1$ *df* from the second sample, or

$$(n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2 \text{ } df$$

overall. Then s can be substituted for σ in Equation 8.9, and the resulting test statistic can then be shown to follow a t distribution with $n_1 + n_2 - 2$ *df* rather than an $N(0,1)$ distribution because σ^2 is unknown. Thus, the following test procedure is used.

Two-Sample *t* Test for Independent Samples with Equal Variances

Suppose we wish to test the hypothesis $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$ with a significance level of α for two normally distributed populations, where σ^2 is assumed to be the same for each population.

Compute the test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s = \sqrt{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right] / (n_1 + n_2 - 2)}$

If $t > t_{n_1 + n_2 - 2, 1-\alpha/2}$ or $t < -t_{n_1 + n_2 - 2, 1-\alpha/2}$

then H_0 is rejected.

If $-t_{n_1 + n_2 - 2, 1-\alpha/2}$ or $t \leq t_{n_1 + n_2 - 2, 1-\alpha/2}$

then H_0 is accepted.

The acceptance and rejection regions for this test are shown in Figure 8.3.

Similarly, a p-value can be computed for the test. Computation of the p-value depends on whether $\bar{x}_1 \le \bar{x}_2 (t \le 0)$ or $\bar{x}_1 > \bar{x}_2 (t > 0)$. In each case, the p-value corresponds to the probability of obtaining a test statistic at least as extreme as the observed value t. This is given in Equation 8.12.

Computation of the p-Value for the Two-Sample *t* Test for Independent Samples with Equal Variances

Compute the test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

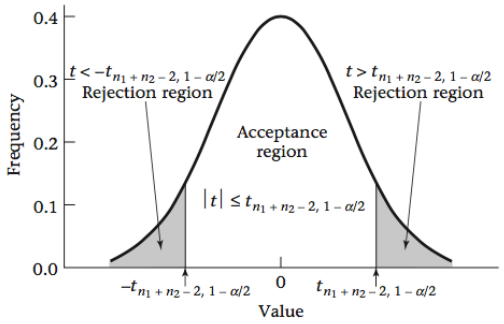
where $s = \sqrt{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right] / (n_1 + n_2 - 2)}$

If $t \le 0$, $p = 2 \times$ (area to the left of t under a $t_{n_1+n_2-2}$ distribution).

If $t > 0$, $p = 2 \times$ (area to the right of t under a $t_{n_1+n_2-2}$ distribution).

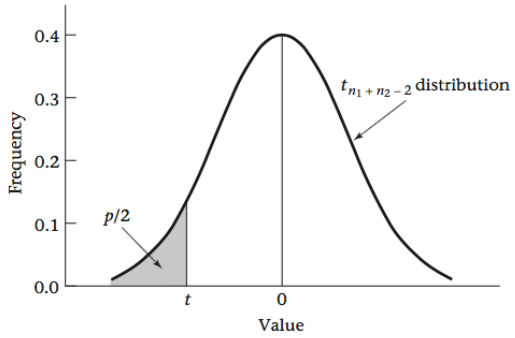
The computation of the p-value is illustrated in Figure 8.4.

Acceptance and rejection regions for the two-sample *t* test for independent samples with equal variances

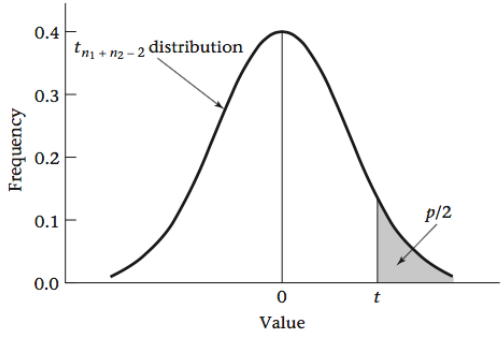


Distribution of t in Equation 8.11 under $H_0 = t_{n_1+n_2-2}$ distribution

Computation of the *p*-value for the two-sample *t* test for independent samples with equal variances



If $t = (\bar{x}_1 - \bar{x}_2) \Big/ \left(s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) \le 0$, then $p = 2 \times$ (area to the left of t under a $t_{n_1+n_2-2}$ distribution).



If $t = (\bar{x}_1 - \bar{x}_2) \Big/ \left(s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) > 0$, then $p = 2 \times$ (area to the right of t under a $t_{n_1+n_2-2}$ distribution).

Hypertension Assess the statistical significance of the data in Example 8.9.

Solution: The common variance is first estimated:

$$s^2 = \frac{7(15.34)^2 + 20(18.23)^2}{27} = \frac{8293.9}{27} = 307.18$$

or $s = 17.527$. The following test statistic is then computed:

$$t = \frac{132.86 - 127.44}{17.527\sqrt{1/8 + 1/21}} = \frac{5.42}{17.527 \times 0.415} = \frac{5.42}{7.282} = 0.74$$

Conclusions accept H0, at variance with the longitudinal study!
In this case intuition and good will should be used in order not to accept a negative conclusion with possible heavy consequencxes for public health

If the critical-value method is used, then note that under H_0, t comes from a t_{27} distribution. Referring to Table 5 in the Appendix, we see that $t_{27,.975} = 2.052$. Because $-2.052 \le 0.74 \le 2.052$, it follows that H_0 is accepted using a two-sided test at the 5% level, and we conclude that the mean blood pressures of the OC users and non-OC users do not significantly differ from each other. In a sense, this result shows the superiority of the longitudinal design in Example 8.5. Despite the similarity in the magnitudes of the mean blood-pressure differences between users and nonusers in the two studies, significant differences could be detected in Example 8.5, in contrast to the nonsignificant results that were obtained using the preceding cross-sectional design. The longitudinal design is usually more efficient because it uses people as their own controls.

To compute an approximate p-value, note from Table 5 that $t_{27,.75} = 0.684$, $t_{27,.80} = 0.855$. Because $0.684 < 0.74 < 0.855$, it follows that $.2 < p/2 < .25$ or $.4 < p < .5$. The exact p-value obtained from MINITAB is $p = 2 \times P(t_{27} > 0.74) = .46$.

HOMEWORK DUE WEDNESDAY 11 May 2022 in GROUPS

REVIEW QUESTIONS 8A

1 How do a paired-sample design and an independent-sample design differ?

2 A man measures his heart rate before using a treadmill and then after walking on a treadmill for 10 minutes on 7 separate days. His mean heart rate at baseline and 10 minutes after treadmill walking is 85 and 93 beats per minute (bpm), respectively. The mean change from baseline to 10 minutes is 8 bpm with a standard deviation of 6 bpm.

(a) What test can we use to compare pre- and post-treadmill heart rate?

(b) Implement the test in Review Question 8A.2a, and report a two-tailed *p*-value.

(c) Provide a 90% confidence interval (CI) for the mean change in heart rate after using the treadmill for 10 minutes.

(d) What is your overall conclusion concerning the data?

8.6 TESTING FOR THE EQUALITY OF TWO VARIANCES

In Section 8.4, when we conducted a two-sample t test for independent samples, we assumed the underlying variances of the two samples were the same. We then estimated the common variance using a weighted average of the individual sample variances. In this section we develop a significance test to validate this assumption. In particular, we wish to test the hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ vs. $H_1: \sigma_1^2 \neq \sigma_2^2$, where the two samples are assumed to be independent random samples from an $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ distribution, respectively.

sample variance of the case group is about four times as large as that of the control group:

$$35.6^2/17.3^2 = 4.23$$

See the case of the rabbits…

What should we do?

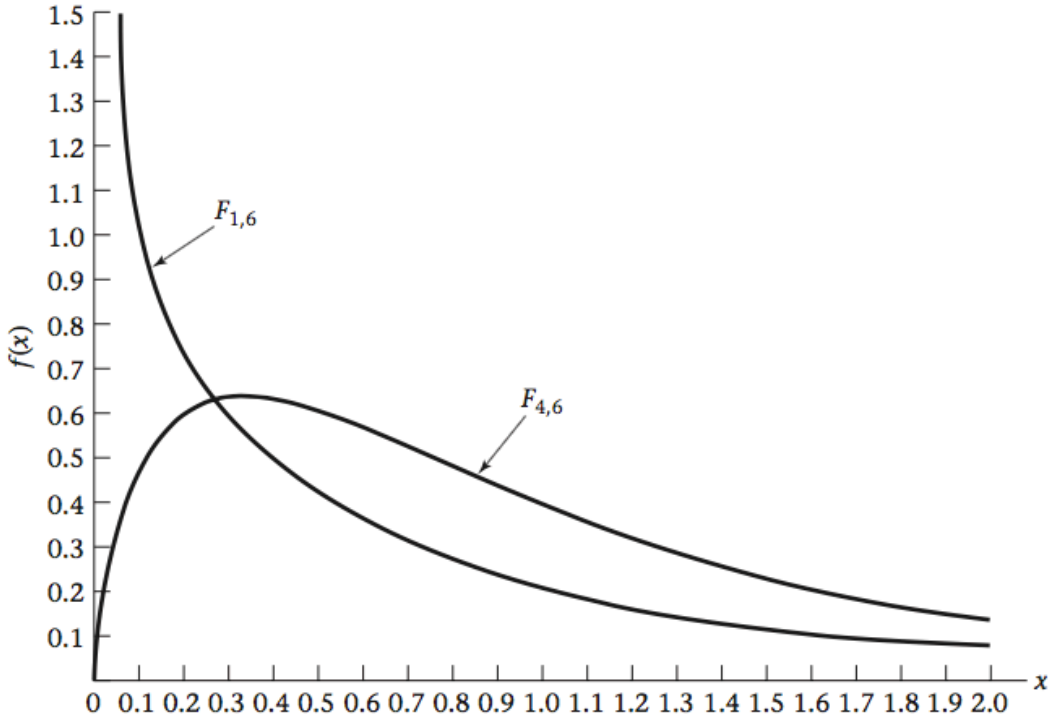
What we need is a significance test to determine whether the underlying variances are in fact equal; that is, we want to test the hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ vs. $H_1: \sigma_1^2 \neq \sigma_2^2$. It seems reasonable to base the significance test on the relative magnitudes of the sample variances $\left(s_1^2, s_2^2\right)$. The best test in this case is based on the ratio of the sample variances $\left(s_1^2/s_2^2\right)$ rather than on the difference between the sample variances $\left(s_1^2 - s_2^2\right)$. Thus, H_0 would be rejected if the variance ratio is either too large or too small and accepted otherwise. To implement this test, the sampling distribution of s_1^2/s_2^2 under the null hypothesis $\sigma_1^2 = \sigma_2^2$ must be determined.

The *F* Distribution

The distribution of the variance ratio $\left(S_1^2/S_2^2\right)$ was studied by statisticians R. A. Fisher and G. Snedecor. It can be shown that the variance ratio follows an ***F* distribution** under the null hypothesis that $\sigma_1^2 = \sigma_2^2$. There is no unique F distribution but instead a family of F distributions. This family is indexed by two parameters termed the *numerator* and *denominator degrees of freedom*, respectively. If the sizes of the first and second samples are n_1 and n_2 respectively, then the variance ratio follows an F distribution with $n_1 - 1$ (numerator *df*) and $n_2 - 1$ (denominator *df*), which is called an F_{n_1-1,n_2-1} distribution.

The F distribution is generally positively skewed, with the skewness dependent on the relative magnitudes of the two degrees of freedom. If the numerator *df* is 1 or 2, then the distribution has a mode at 0; otherwise, it has a mode greater than 0. The distribution is illustrated in Figure 8.5. Table 8 in the Appendix gives the percentiles of the F distribution for selected values of the numerator and denominator *df*.

Probability density for the *F* distribution



SYMMETRY OF THE F PERCENTILES : F TABLES

Generally, F distribution tables give only upper percentage points because the symmetry properties of the F distribution make it possible to derive the lower percentage points of any F distribution from the corresponding upper percentage points of an F distribution with the degrees of freedom reversed. Specifically, note that under H_0, S_2^2/S_1^2 follows an F_{d_2,d_1} distribution. Therefore,

$$Pr\left(S_2^2/S_1^2 \geq F_{d_2,d_1,1-p}\right) = p$$

By taking the inverse of each side and reversing the direction of the inequality, we get

$$Pr\left(\frac{S_1^2}{S_2^2} \leq \frac{1}{F_{d_2,d_1,1-p}}\right) = p$$

Under H_0, however, S_1^2/S_2^2 follows an F_{d_1,d_2} distribution. Therefore,

$$Pr\left(\frac{S_1^2}{S_2^2} \leq F_{d_1,d_2,p}\right) = p$$

It follows from the last two inequalities that

$$F_{d_1,d_2,p} = \frac{1}{F_{d_2,d_1,1-p}}$$

This principle is summarized as follows.

Computation of the Lower Percentiles of an *F* Distribution

The **lower *p*th percentile** of an F distribution with d_1 and d_2 *df is* the reciprocal of the **upper *p*th percentile** of an F distribution with d_2 and d_1 *df*. In symbols,

$$F_{d_1,d_2,p} = 1/F_{d_2,d_1,1-p}$$

The **$100 \times p$th percentile of an F distribution** with d_1 and d_2 degrees of freedom is denoted by $F_{d_1,d_2,p}$. Thus,

$$Pr\left(F_{d_1,d_2} \leq F_{d_1,d_2,p}\right) = p$$

The F table is organized such that the numerator df (d_1) is shown in the first row, the denominator df (d_2) is shown in the first column, and the various percentiles (p) are shown in the second column.

The *F* Test

We now return to the significance test for the equality of two variances. We want to test the hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ vs. $H_1: \sigma_1^2 \neq \sigma_2^2$. We stated that the test would be based on the variance ratio S_1^2/S_2^2, which under H_0 follows an F distribution with $n_1 - 1$ and $n_2 - 1$ *df*. This is a two-sided test, so we want to reject H_0 for both small and large values of S_1^2/S_2^2. This procedure can be made more specific, as follows.

F Test for the Equality of Two Variances

Suppose we want to conduct a test of the hypothesis H_0: $\sigma_1^2 = \sigma_2^2$ vs. $H_1: \sigma_1^2 \neq \sigma_2^2$ with significance level α.

Compute the test statistic $F = s_1^2/s_2^2$.

If $F > F_{n_1-1,n_2-1,1-\alpha/2}$ or $F < F_{n_1-1,n_2-1,\alpha/2}$

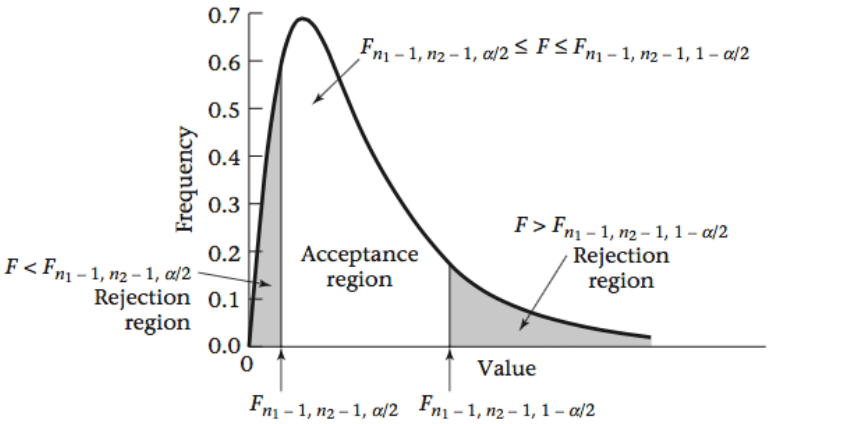
then H_0 is rejected.

If $F_{n_1-1,n_2-1,\alpha/2} \leq F \leq F_{n_1-1,n_2-1,1-\alpha/2}$

then H_0 is accepted. The acceptance and rejection regions for this test are shown in Figure 8.6.

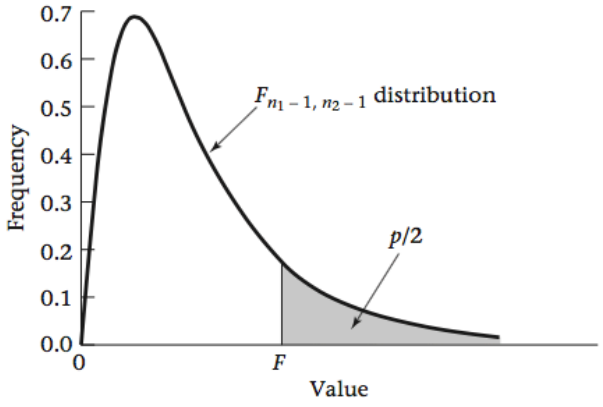
Alternatively, the exact *p*-value is given by Equation 8.16.

FIGURE 8.6 Acceptance and rejection regions for the *F* test for the equality of two variances

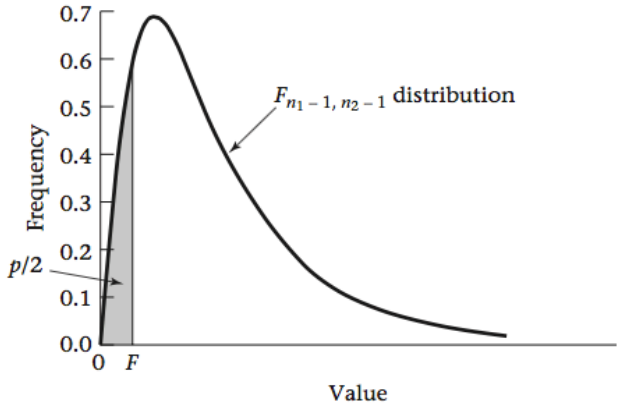


$F_{n_1-1,\,n_2-1}$ distribution = distribution of $F = S_1^2/S_2^2$ under H_0

FIGURE 8.7 Computation of the *p*-value for the *F* test for the equality of two variances



If $F = s_1^2/s_2^2 \ge 1$, then $p = 2 \times$ (area to the right of F under an $F_{n_1-1,\,n_2-1}$ distribution)



If $F = s_1^2/s_2^2 < 1$, then $p = 2 \times$ (area to the left of F under an $F_{n_1-1,\,n_2-1}$ distribution)

EQUATION 8.16 **Computation of the *p*-Value for the *F* Test for the Equality of Two Variances**

Compute the test statistic $F = s_1^2/s_2^2$.

If $F \ge 1$, then $p = 2 \times Pr\left(F_{n_1-1,n_2-1} > F\right)$

If $F < 1$, then $p = 2 \times Pr\left(F_{n_1-1,n_2-1} < F\right)$

This computation is illustrated in Figure 8.7.

Let us consider the story in the example 8.13

Cardiovascular Disease, Pediatrics Consider a problem discussed earlier, namely the familial aggregation of cholesterol levels. In particular, suppose cholesterol levels are assessed in 100 children, 2 to 14 years of age, of men who have died from heart disease and it is found that the mean cholesterol level in the group $(\bar{x}_1)$ is 207.3 mg/dL. Suppose the sample standard deviation in this group (s_1) is 35.6 mg/dL. Previously, the cholesterol levels in this group of children were compared with 175 mg/dL, which was assumed to be the underlying mean level in children in this age group based on previous large studies.

A better experimental design would be to select a group of control children whose fathers are alive and do not have heart disease and who are from the same census tract as the case children, and then to compare their cholesterol levels with those of the case children. If the case fathers are identified by a search of death records from the census tract, then researchers can select control children who live in the same census tract as the case families but whose fathers have no history of heart disease. The case and control children come from the same census tract but are *not* individually matched. Thus, they are considered as two independent samples rather than as two paired samples. The cholesterol levels in these children can then be measured. Suppose the researchers found that among 74 control children, the mean cholesterol level $(\bar{x}_2)$ is 193.4 mg/dL with a sample standard deviation (s_2) of 17.3 mg/dL. We would like to compare the means of these two groups using the two-sample t test for independent samples given in Equation 8.11, but we are hesitant to assume equal variances because the

sample variance of the case group is about four times as large as that of the control group:

$35.6^2/17.3^2 = 4.23$

What should we do?

What we need is a significance test to determine whether the underlying variances are in fact equal; that is, we want to test the hypothesis H_0: $\sigma_1^2 = \sigma_2^2$ vs. H_1: $\sigma_1^2 \neq \sigma_2^2$. It seems reasonable to base the significance test on the relative magnitudes of the sample variances $\left(s_1^2, s_2^2\right)$. The best test in this case is based on the ratio of the sample variances $\left(s_1^2/s_2^2\right)$ rather than on the difference between the sample variances $\left(s_1^2 - s_2^2\right)$. Thus, H_0 would be rejected if the variance ratio is either too large or too small and accepted otherwise. To implement this test, the sampling distribution of s_1^2/s_2^2 under the null hypothesis $\sigma_1^2 = \sigma_2^2$ must be determined.

Cardiovascular Disease, Pediatrics Test for the equality of the two variances given in Example 8.13.

Solution: $F = s_1^2 / s_2^2 = 35.6^2 / 17.3^2 = 4.23$

Because the two samples have 100 and 74 people, respectively, we know from Equation 8.15 that under H_0, $F \sim F_{99,73}$. Thus, H_0 is rejected if

$$F > F_{99,73,.975} \quad \text{or} \quad F < F_{99,73,.025}$$

Note that neither 99 *df* nor 73 *df* appears in Table 8 in the Appendix. One approach is to obtain the percentiles using a computer program. In this example, we want to find the value $c_1 = F_{99,73,.025}$ and $c_2 = F_{99,73,.975}$. such that

$$Pr\left(F_{99,73} \le c_1\right) = .025 \quad \text{and} \quad Pr\left(F_{99,73} \ge c_2\right) = .975$$

We can use the qf function of R for this purpose. We have:

c_1 = qf(0.025, 99, 73),

c_2 = qf(0.975, 99, 73).

The result is shown as follows:

```
> qf(0.025, 99, 73)
[1] 0.65476
> qf(0.975, 99, 73}
[1] 1.549079.
```

Thus, $c_1 = 0.655$, $c_2 = 1.549$. Because $F = 4.23 > c_2$ it follows that $p < 0.05$. Alternatively, we could compute an exact *p*-value. This is given by:

$p = 2 \times Pr(F_{99,73} > 4.23) = 2 \times [1 - pf(4.23, 99, 73)]$. The result is shown as follows:

```
> p.value < -2 * (1 - pf(4.23, 99, 73))
> p.value
[1] 8.839514e-10
```

CONCLUSION & REMARK

Thus, the p-value = 8.8×10^{-10} indicates that the variances are significantly different. Therefore, the two-sample t test with equal variances given in Section 8.4 cannot be used, because this test depends on the assumption that the variances are equal.

A question often asked about the F test is whether it makes a difference which sample is selected as the numerator sample and which is selected as the denominator sample. The answer is that, for a two-sided test, it does *not* make a difference because of the rules for calculating lower percentiles given in Equation 8.14. A variance ratio > 1 is usually more convenient, so there is no need to use Equation 8.14. Thus, the larger variance is usually put in the numerator and the smaller variance in the denominator.

You can Find a discussion of the themes presented today also in the notebook Associated to lecture n. 11 by R. Di Leonardo.

FIGURE 8.13 Flowchart summarizing two-sample statistical inference—normal-theory methods