Lecture 08: t-tests 1: Paired Sample and Independent Sample

Outline

- Introduction to t-statistics
- Two-sample hypothesis testing:
  - Paired t-tests
  - Independent sample t-tests (equal variance only)

Introduction to t-statistics

The t Statistic

- The t statistic allows researchers to use sample data to test hypotheses about an unknown population mean.
- The particular advantage of the t statistic, is that the t statistic does not require any knowledge of the population standard deviation.
- Thus, the t statistic can be used to test hypotheses about a *completely unknown* population; that is, both μ and σ are unknown, and *the only available information about the population comes from the sample*.

The t Statistic (cont.)

- All that is required for a hypothesis test with t is a sample and a reasonable hypothesis about the population mean.
- There are two general situations where this type of hypothesis test is used:

1. The t statistic is used when a researcher wants to determine whether or not a treatment causes a change in a population mean. In this case you must know the value of μ for the original, untreated population. A sample is obtained from the population, and the treatment is administered to the sample. If the resulting sample mean is significantly different from the original population mean, you can conclude that the treatment has a significant effect.

The t Statistic (cont.)

2. Occasionally a theory or other prediction will provide a hypothesized value for an unknown population mean. A sample is then obtained from the population and the t statistic is used to compare the actual sample mean with the hypothesized population mean. A significant difference indicates that the hypothesized value for μ should be rejected.

The Estimated Standard Error and the t Statistic

- Whenever a sample is obtained from a population you expect to find some discrepancy or "error" between the sample mean and the population mean.
- This general phenomenon is known as **sampling error**.
- The goal for a hypothesis test is to evaluate the *significance* of the observed discrepancy between a sample mean and the population mean.

The Estimated Standard Error and the t Statistic (cont.)

The hypothesis test attempts to decide between the following two alternatives:

1. Is it reasonable that the discrepancy between M and μ is simply due to sampling error and not the result of a treatment effect?
2. Is the discrepancy between M and μ more than would be expected by sampling error alone? That is, is the sample mean significantly different from the population mean?

The Estimated Standard Error and the t Statistic (cont.)

- The critical first step for the t statistic hypothesis test is to calculate exactly how much difference between M and μ is reasonable to expect.
- However, because the population standard deviation is unknown, it is impossible to compute the standard error of M as we did with z-scores in Lecture 6.
- Therefore, **the t statistic** requires that you use the sample data to compute an **estimated standard error of M**.

The Estimated Standard Error and the t Statistic (cont.)

- This calculation defines standard error exactly as it was defined in previous lectures, but now we must use the sample variance, s^2, in place of the unknown population variance, σ^2 (or use sample standard deviation, s, in place of the unknown population standard deviation, σ).

The Estimated Standard Error and the t Statistic (cont.)

- The t statistic (like the z-score) forms a ratio.
- The top of the ratio contains the obtained difference between the sample mean and the hypothesized population mean.
- The bottom of the ratio is the standard error which measures how much difference is expected by chance.

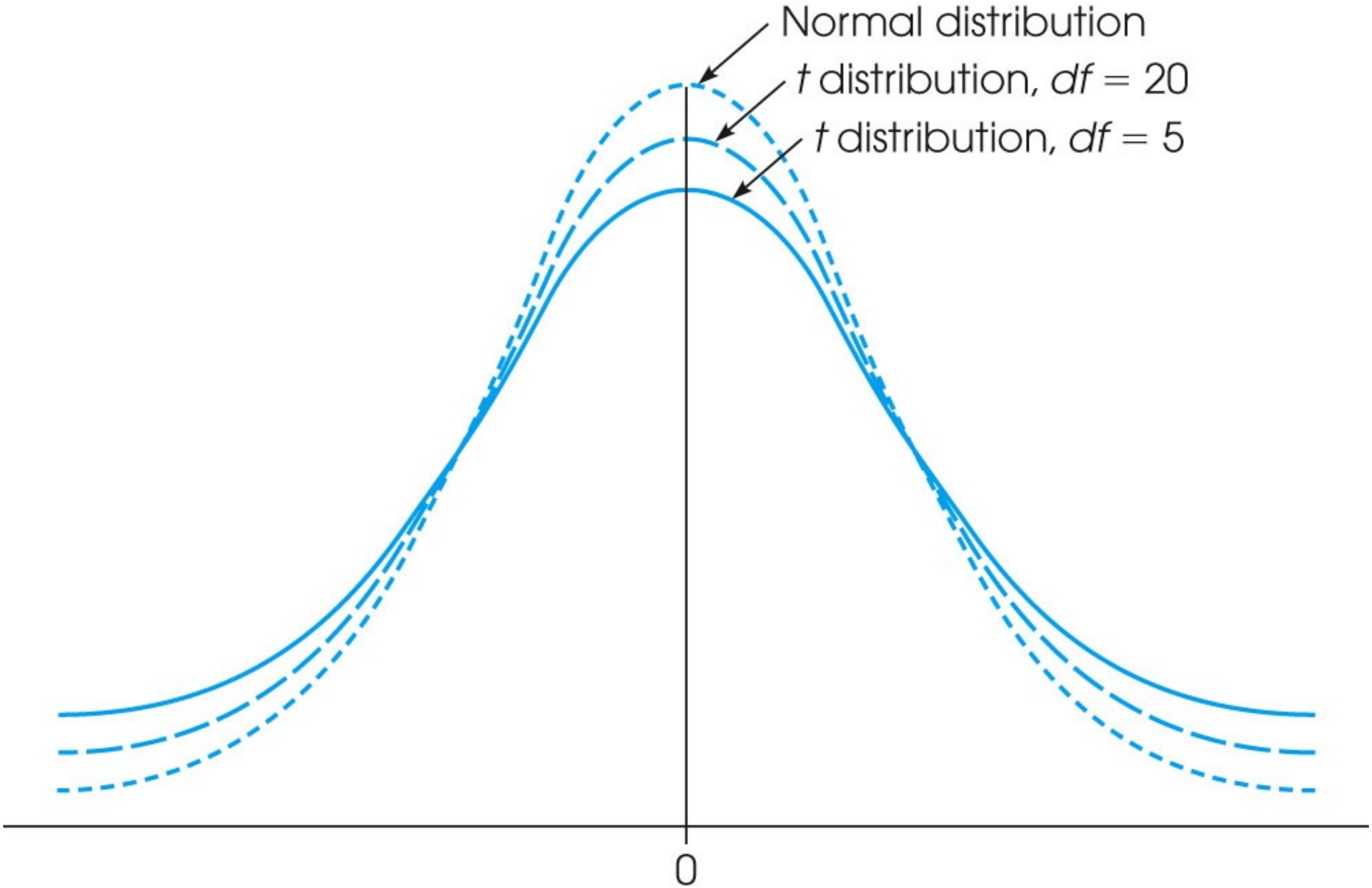
$$t = \frac{M - u}{S_M}$$

$$t = \frac{\underset{\text{(from the data)}}{\text{sample mean}} - \underset{\text{(hypothesized from } H_0\text{)}}{\text{population mean}}}{\underset{\text{(computed from the sample data)}}{\text{estimated standard error}}}$$

- A large value for t indicates that the obtained difference between the data and the hypothesis is greater than that would be expected if the treatment has no effect.

The t Distributions and Degrees of Freedom

- You can think of the t statistic as an "estimated z-score."
- The estimation comes from the fact that we are using the sample variance to estimate the unknown population variance.
- With a large sample, the estimation is very good and the t statistic will be very similar to a z-score.
- With small samples, however, the t statistic will provide a relatively poor estimate of z.



Normal distribution
t distribution, $df = 20$
t distribution, $df = 5$
0

The t Distributions and Degrees of Freedom (cont.)

- The value of **degrees of freedom**, df = n - 1, is used to describe how well the t statistic represents a z-score.
- Also, the value of df will determine how well the distribution of t approximates a normal distribution.
- For large values of df, the **t distribution** will be nearly normal, but with small values for df, the t distribution will be flatter and more spread out than a normal distribution.

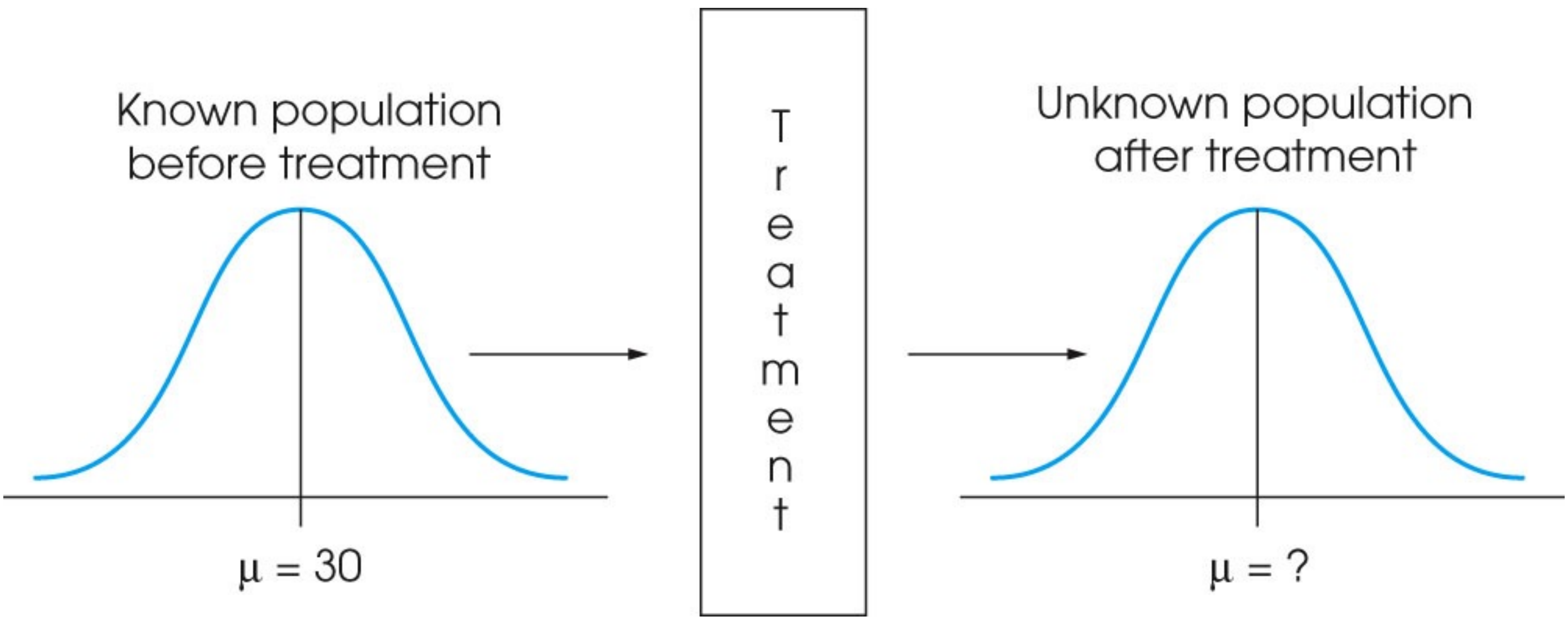
The t Distributions and Degrees of Freedom (cont.)

- To evaluate the t statistic from a hypothesis test, you must select an α level, find the value of df for the t statistic, and consult the t distribution table.
- If the obtained t statistic is larger than the critical value from the table, you can reject the null hypothesis.
- In this case, you have demonstrated that the obtained difference between the data and the hypothesis (numerator of the ratio) is significantly larger than the difference that would be expected if there was no treatment effect (the standard error in the denominator).

Hypothesis Tests with the t Statistic

The hypothesis test with a t statistic follows the same four-step procedure that was used with z-score tests:

1. State the hypotheses and select a value for α. (Note: The null hypothesis always states a specific value for μ.)
2. Locate the critical region. (Note: You must find the value for df and use the t distribution table.)
3. Calculate the test statistic.
4. Make a decision (Either "reject" or "fail to reject" the null hypothesis).


Known population
before treatment
μ = 30
Treatment
Unknown population
after treatment
μ = ?

Summary for one-sample t-test

- $H_1 : \mu > \mu_o$ reject H_0 if $\bar{x} > \mu_0 + t_{n-1,1-\alpha} \cdot s / \sqrt{n}$

  Or equivalently: $t_{obs} = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}} > t_{n-1,1-\alpha}$

- $H_1 : \mu < \mu_o$ reject H_0 if $\bar{x} < \mu_0 + t_{n-1,\alpha} \cdot s / \sqrt{n}$

  Or equivalently: $t_{obs} = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}} < t_{n-1,\alpha}$

- $H_1 : \mu \neq \mu_o$ reject H_0 if

  $\bar{x} > \mu_0 + t_{n-1,1-\alpha/2} \cdot s / \sqrt{n} \quad or \quad \bar{x} < \mu_0 + t_{n-1,\alpha/2} \cdot s / \sqrt{n}$

Or equivalently: $t_{obs} = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}} < t_{n-1,\alpha/2} \; or \; t_{obs} = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}} > t_{n-1,1-\alpha/2}$

Measuring Effect Size with the t Statistic

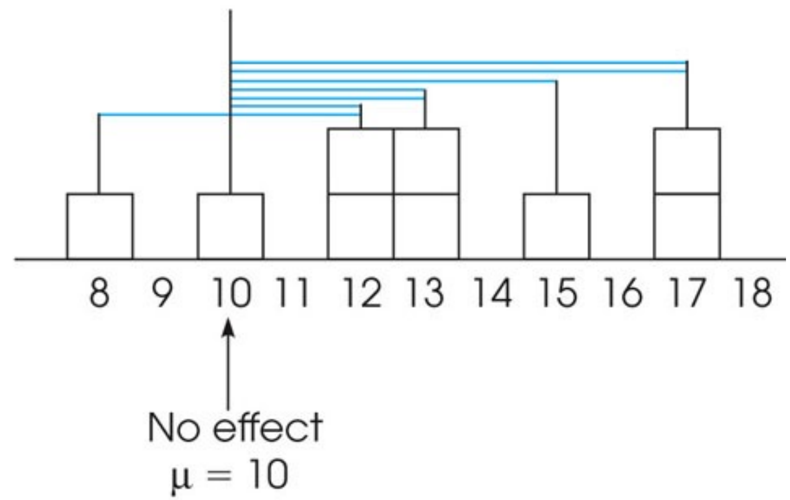
- Because the significance of a treatment effect is determined partially by the size of the effect and partially by the size of the sample, you cannot assume that a significant effect is also a large effect.
- Therefore, it is recommended that a measure of effect size be computed along with the hypothesis test.

Measuring Effect Size with the t Statistic (cont.)

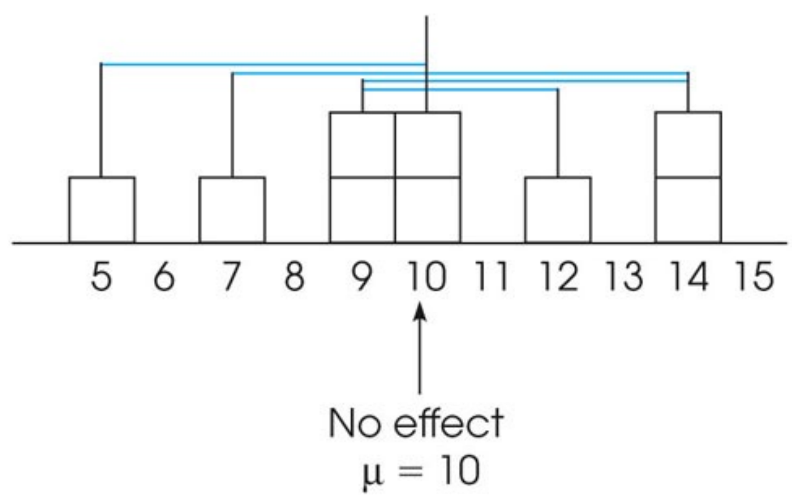
- For the t test it is possible to compute an estimate of Cohen's d just as we did for the z-score test in previous lecture. The only change is that we now use the sample standard deviation instead of the population value (which is unknown).

$$\text{estimated Cohen's d} = \frac{\text{mean difference}}{\text{standard deviation}} = \frac{M - \mu}{s}$$

(a) Original scores, including the treatment effect



(b) Adjusted scores with the treatment effect removed



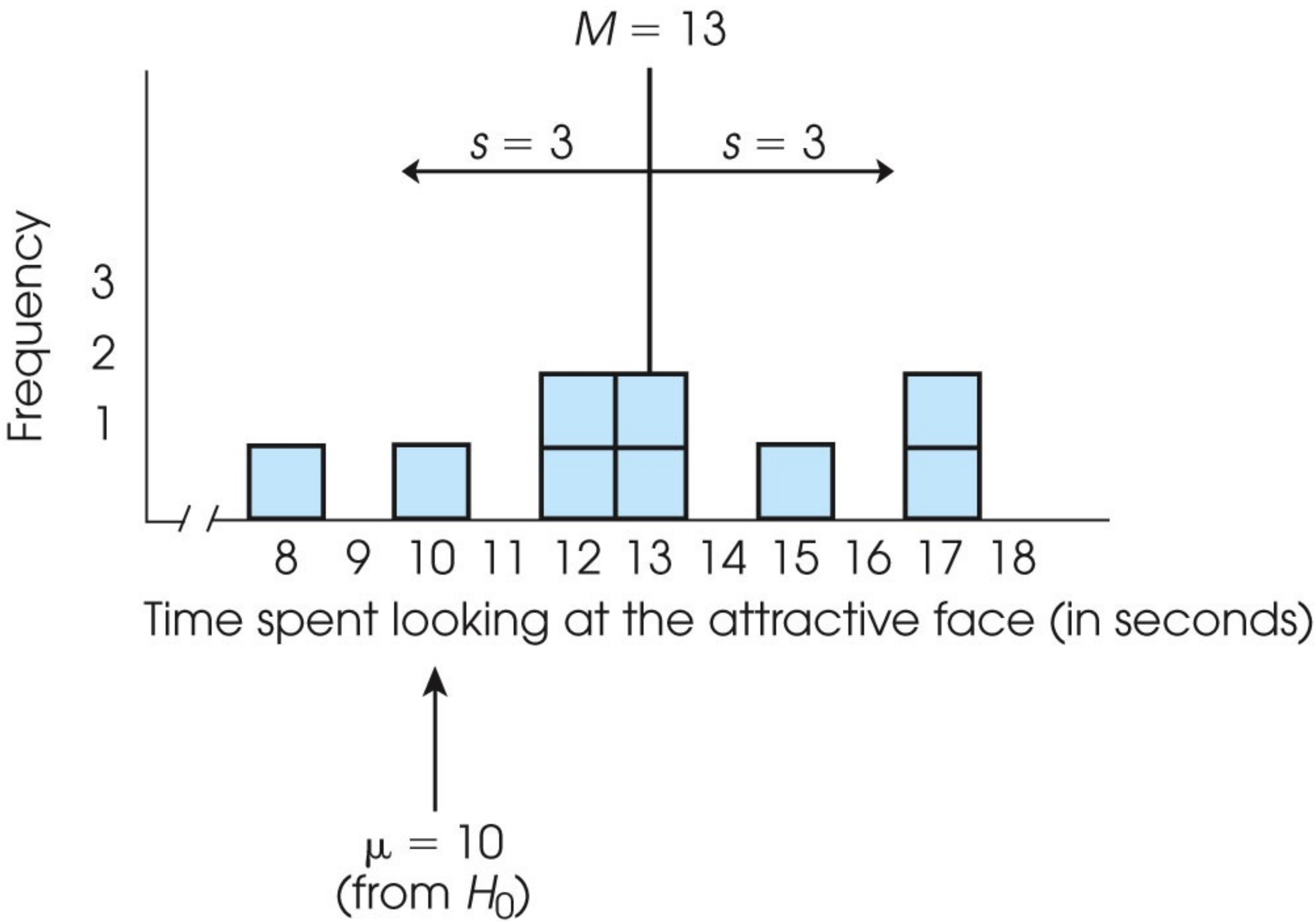
Measuring Effect Size with the t Statistic (cont.)

- As before, Cohen's d measures the size of the treatment effect in terms of the standard deviation.
- With a t test it is also possible to measure effect size by computing the **percentage of variance accounted for** by the treatment.
- This measure is based on the idea that the treatment causes the scores to change, which contributes to the observed variability in the data.

Measuring Effect Size with the t Statistic (cont.)

- By measuring the amount of variability that can be attributed to the treatment, we obtain a measure of the size of the treatment effect. For the t statistic hypothesis test,

$$\text{percentage of variance accounted for} = r^2 = \frac{t^2}{t^2 + df}$$



M = 13
s = 3
s = 3
Frequency
3
2
1
8 9 10 11 12 13 14 15 16 17 18
Time spent looking at the attractive face (in seconds)
μ = 10
(from H_0)

One-sample t-test
Vs
Two-sample test

Scenarios of Two-Sample Hypothesis Testing

- Recap: The parameter of a population from which the sample was drawn is compared with a specific value, which is supposed to be the known parameter value of certain larger population.

- In many studies, we want to compare the underlying parameters of two different populations, neither of whose values are known.

- Typical study designs:

  Between-group (independent-measures) design

  Within-group (dependent-measures) design

Within-group Study

(or called paired study, repeated measures study, within-subject study, follow-up study, longitudinal study, crossover study; 组内设计, 重复测量,被试内设计，相关样本等等)

The same group of subjects are followed over time.

- Repeated observations of the same subjects (or matched subjects) over time
- Paired-sample design: the differences observed in these subjects are less likely to be the result of confounding factors such as cultural differences across generations

Examples: Wealth and Satisfaction of life

1. Identify a group of age-matched, gender-matched (a lot of other things matched) poor college graduates. Measure their satisfaction of life.
2. Re-screen 4 years later to obtain a subgroup who becomes really rich. Measure their satisfaction again.
3. Compare them.

Between-group Study

(or called independent-sample, cross-section study, between-subject study, 独立指标设计，组间设计, 被试间设计等)

Multiple groups of subjects are seen at <u>one time point</u>.

- One observation of different subjects at one time point
- Independent-sample: groups are independent to each other

Examples: Wealth and Satisfaction of life

1. Identify a group of poor people and a group of rich people and measure their satisfaction of life.
2. Compare the two groups.

Paired （匹配）v.s. Independent （独立）t tests

- Two samples are said to be <u>paired</u> when each data point of first sample is matched to a unique point of the second sample. (e.g., 2 measurements on same person, or measuring the same property of a twin).
- Two samples are <u>independent</u> when data points in one sample are unrelated to data points in the second sample.

The t Test for Two Related Samples: paired sample t-test

Repeated-Measures Designs

- The related-samples hypothesis test allows researchers to evaluate the mean difference between two treatment conditions using the data from a single sample.
- In a **repeated-measures design**, a single group of individuals is obtained and each individual is measured in both of the treatment conditions being compared.
- Thus, the data consist of two scores for each individual.

Hypothesis Tests with the Repeated-Measures t

- The repeated-measures t statistic allows researchers to test a hypothesis about the population mean difference between two treatment conditions using sample data from a repeated-measures research study.
- In this situation it is possible to compute a **difference score** for each individual:

$$\text{difference score} = D = x_2 - x_1$$

Where x_1 is the person's score in the first treatment and x_2 is the score in the second treatment.

Hypothesis Tests with the Repeated-Measures t (cont.)

- The related-samples t test can also be used for a similar design, called a **matched-subjects design**, in which each individual in one treatment is matched one-to-one with a corresponding individual in the second treatment.
- The matching is accomplished by selecting pairs of subjects so that the two subjects in each pair have identical (or nearly identical) scores on the variable that is being used for matching.
- For a matched-subjects design, a difference score is computed for each matched pair of individuals.

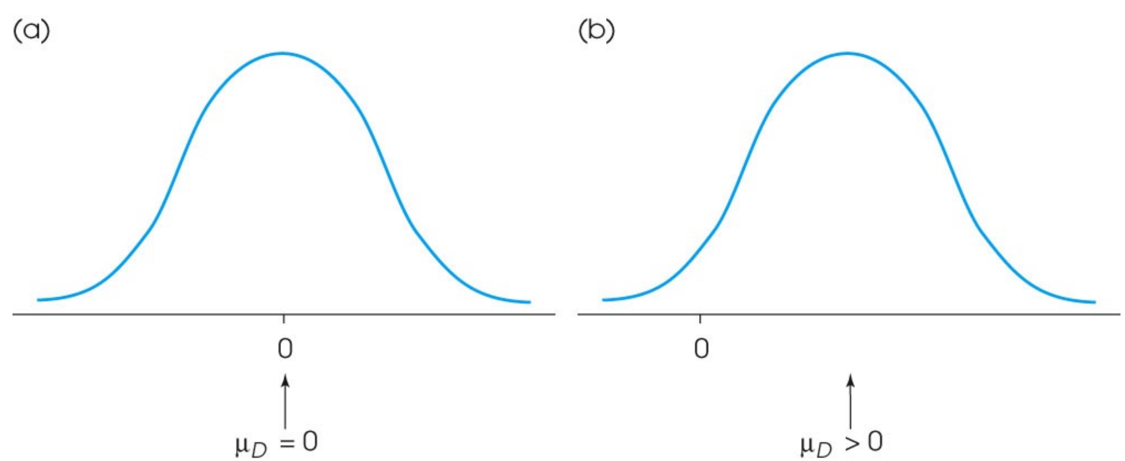
Hypothesis Tests with the Repeated-Measures t (cont.)

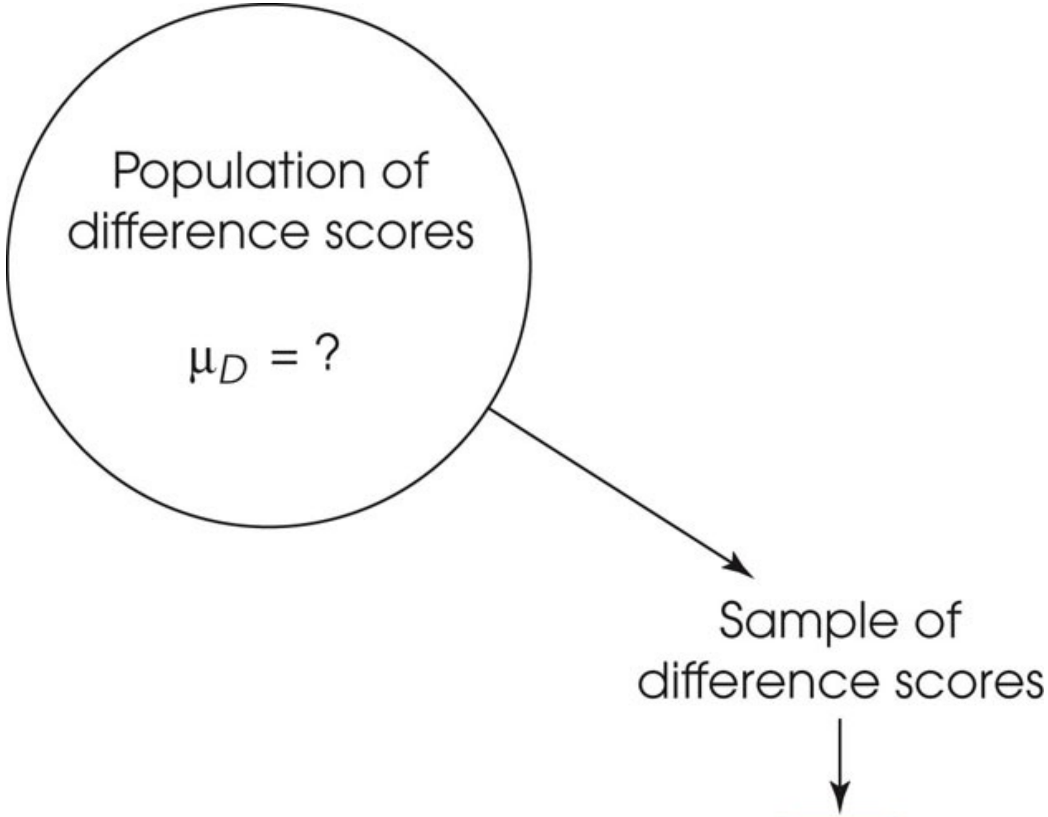
- However, because the matching process can never be perfect, matched-subjects designs are relatively rare.
- As a result, **repeated-measures designs (using the same individuals in both treatments)** make up the vast majority of related-samples studies.

Hypothesis Tests with the Repeated-Measures t (cont.)

- The sample of difference scores is used to test hypotheses about the population of difference scores. The null hypothesis states that the population of difference scores has a mean of zero,

$$H_0: \ \mu_D = 0$$





Subject	I	II	*D*
A	10	14	4
B	15	13	–2
C	12	15	3
D	11	12	1

The logic behind paired t-test

Using mean rule and variance rule

$$\Delta = \mu_D = \mu_{X-Y} = \mu_X - \mu_Y$$

$$\sigma_D^2 = Var(X-Y) = \sigma_X^2 + \sigma_Y^2 - 2r\sigma_X\sigma_Y$$

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

相关系数,以后再教

- $D_i = X_{2i} - X_{1i} \sim N(\Delta, \sigma_D^2)$
- So, the testing problem has been reduced to a one-sample t-test for D_i

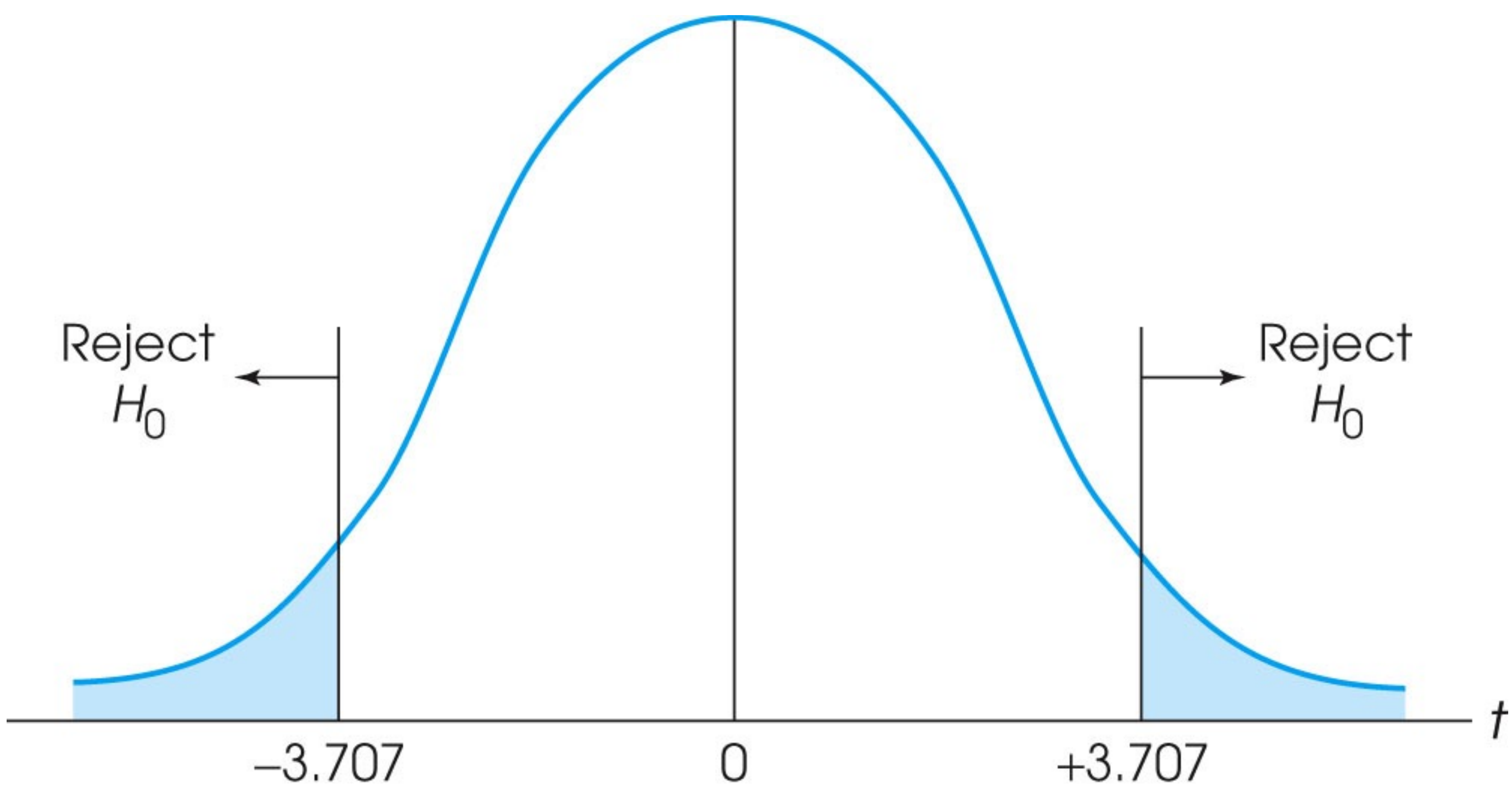
→ Is $N(\Delta, \sigma_D^2)$ distinguishable from 0?

Hypothesis Tests with the Repeated-Measures t (cont.)

- The null hypothesis says that there is no consistent or systematic difference between the two treatment conditions.
- Note that the null hypothesis does not say that each individual will have a difference score equal to zero.
- Some individuals will show a positive change from one treatment to the other, and some will show a negative change.

Hypothesis Tests with the Repeated-Measures t (cont.)

- On average, the entire population will show a **mean difference** of zero.
- Thus, according to the null hypothesis, the sample mean difference should be near zero.
- Remember, the concept of sampling error states that samples are not perfect, and we should always expect small differences between a sample mean and the population mean.



Reject
H_0
Reject
H_0
–3.707
0
+3.707
t

Hypothesis Tests with the Repeated-Measures t (cont.)

- The alternative hypothesis states that there is a systematic difference between treatments that causes the difference scores to be consistently positive (or negative) and produces a non-zero mean difference between the treatments:

$$H_1: \ \mu_D \neq 0$$

- According to the alternative hypothesis, the sample mean difference obtained in the research study is a reflection of the true mean difference that exists in the population.

Hypothesis Tests with the Repeated-Measures t (cont.)

- The repeated-measures t statistic forms a ratio with exactly the same structure as the single-sample t statistic presented before.
- The numerator of the t statistic measures the difference between the sample mean and the hypothesized population mean.

Hypothesis Tests with the Repeated-Measures t (cont.)

- The bottom of the ratio is the standard error, which measures how much difference is reasonable to expect between a sample mean and the population mean if there is no treatment effect; that is, how much difference is expected simply by sampling error.

$$t = \frac{M_D - u_D}{S_{MD}}$$

df = n-1

Hypothesis Tests with the Repeated-Measures t (cont.)

- For the repeated-measures t statistic, all calculations are done with the sample of difference scores.
- The mean for the sample appears in the numerator of the t statistic and the variance of the difference scores is used to compute the standard error in the denominator.
- As usual, the standard error is computed by

$$s_{MD}^2 = \frac{s^2}{n} \qquad s_{MD} = \frac{s}{\sqrt{n}}$$

$s^2 = \frac{\sum_{i=1}^{n}(D_i - \bar{D})^2}{n-1}$ Is the sample variance of D.

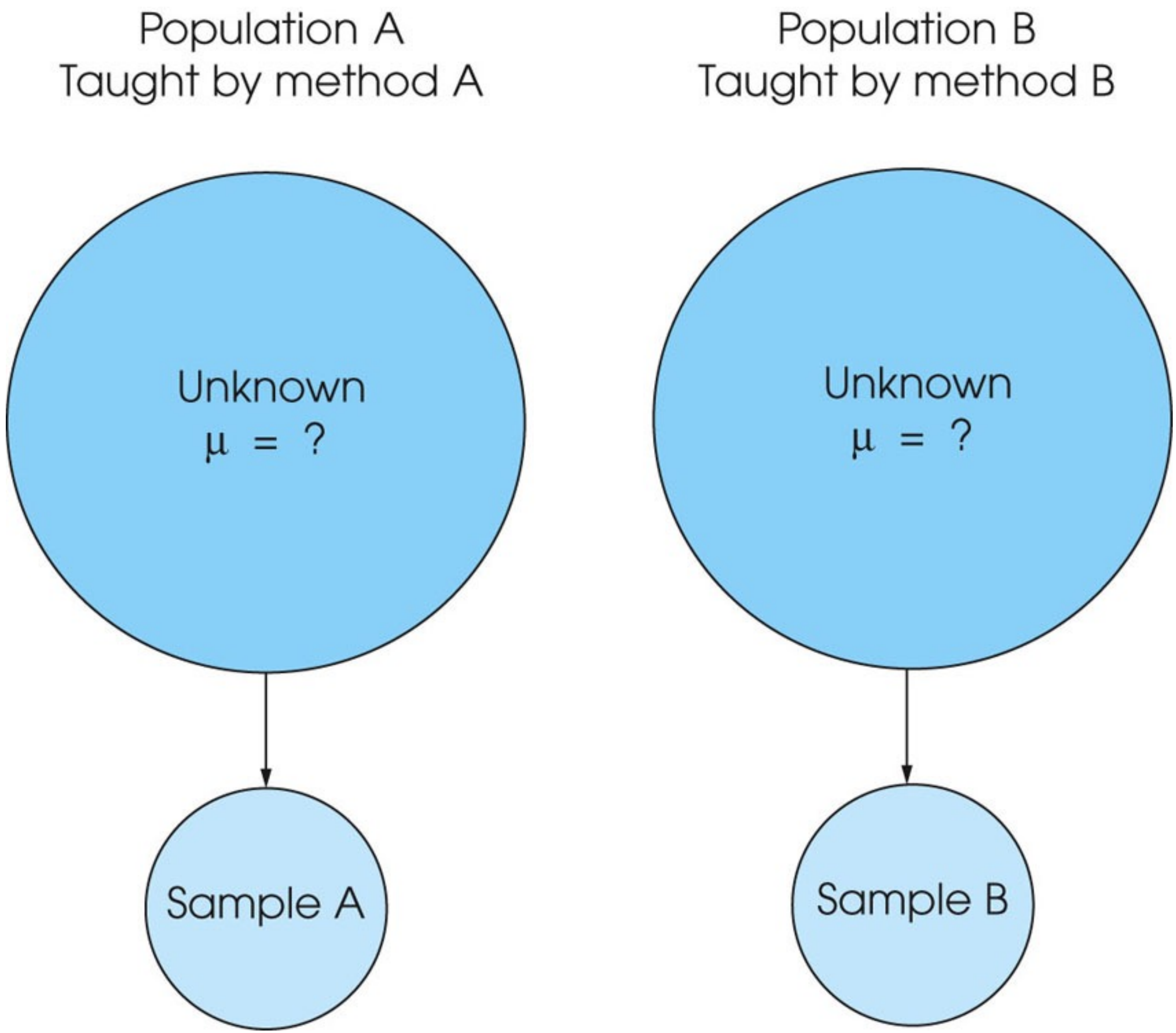
Measuring Effect Size for the Repeated-Measures t

- Effect size for the repeated-measures t is measured in the same way that we measured effect size for the single-sample t and the repeated-measures t.
- Specifically, you can compute an estimate of **Cohen's d** to obtain a standardized measure of the mean difference, or you can compute $\mathbf{r^2}$ to obtain a measure of the percentage of variance accounted for by the treatment effect.

The t-test for Two Independent Samples: independent sample t-test

Independent-Measures Designs

- The independent-measures hypothesis test allows researchers to evaluate the mean difference between two populations using the data from two separate samples.
- The identifying characteristic of the **independent-measures** or **between-subjects** design is the existence of two separate or independent samples.
- Thus, an independent-measures design can be used to test for mean differences between two distinct populations (such as men versus women) or between two different treatment conditions (such as drug versus no-drug).


Population A
Taught by method A
Unknown
μ = ?
Sample A
Population B
Taught by method B
Unknown
μ = ?
Sample B

Independent-Measures Designs (cont.)

- The independent-measures design is used in situations where a researcher has no prior knowledge about either of the two populations (or treatments) being compared.
- In particular, the population means and standard deviations are all unknown.
- Because the population variances are not known, these values must be estimated from the sample data.

Hypothesis Testing with the Independent-Measures t Statistic

- As with all hypothesis tests, the general purpose of the independent-measures t test is to determine whether the sample mean difference obtained in a research study indicates a real mean difference between the two populations (or treatments) or whether the obtained difference is simply the result of sampling error.
- Remember, if two samples are taken from the same population and are given exactly the same treatment, there still will be some difference between the sample means.

Two-sample test for <u>independent samples</u>

In other words

- Suppose the measurements from group one are $N(\mu_l,\sigma_1^2)$, and group two as $N(\mu_2,\sigma_2^2)$

- H_0: $\mu_l=\mu_2$; H_1: $\mu_l \neq \mu_2$ (or H_1: $\mu_l>\mu_2$ or H_1: $\mu_l<\mu_2$)

=>*logic:*

equivalent to say: H_0: $\mu_l-\mu_2=0$; H_1: $\mu_l-\mu_2 \neq 0$

Hypothesis Testing with the Independent-Measures t Statistic (cont.)

- The hypothesis test provides a standardized, formal procedure for determining whether the mean difference obtained in a research study is significantly greater than can be explained by sampling error
- To prepare the data for analysis, the first step is to compute the sample mean and SS (or s, or s^2) for each of the two samples.

Hypothesis Testing with the Independent-Measures t Statistic (cont.)

1. State the hypotheses and select an α level. For the independent-measures test, H_0 states that there is no difference between the two population means.
2. Locate the critical region. The critical values for the t statistic are obtained using degrees of freedom that are determined by adding together the df value for the first sample and the df value for the second sample.

Hypothesis Testing with the Independent-Measures t Statistic (cont.)

3. Compute the test statistic. The t statistic for the independent-measures design has the same structure as the single sample t introduced in Chapter 9. However, in the independent-measures situation, all components of the t formula are doubled: there are two sample means, two population means, and two sources of error contributing to the standard error in the denominator.
4. Make a decision. If the t statistic ratio indicates that the obtained difference between sample means (numerator) is substantially greater than the difference expected by chance (denominator), we reject H_0 and conclude that there is a real mean difference between the two populations or treatments.

Two-sample t-test of Normal mean

for independent samples with equal & unknown variances

- *Only when S_1 and S_2 are close to each other (so-called equal);*
- Homogeneity of variance方差同质性; This is the usual case!

Also, we then have two estimates of the variance.

Use pooled estimate of variance:

$$S_p^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$$

$$\text{pooled variance} = s_p^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2}$$

Eq. 10.3

Which is a weighted average of the two sample variances.

Why weighted?

Two-sample t-test of Normal mean

for independent samples with equal & unknown variances

Want to test $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$

at level α for two Normally distributed populations

σ^2 assumed equal for both groups $\left(\sigma_1^2 = \sigma_2^2 = \sigma^2\right)$**.BUT unknown**

$$t = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - 0}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$$

Reject H_0 if $t > t_{n_1+n_2-2,1-\alpha/2}$ or $t < -t_{n_1+n_2-2,1-\alpha/2}$

$(equivalently\ ,reject\ \ if\ \ |t| > t_{n_1+n_2-2,1-\alpha/2})$

df = n-1

TABLE 10.1

The basic elements of a t statistic for the single-sample t and the independent-measures t

	Sample Data	Hypothesized Population Parameter	Estimated Standard Error	Sample Variance
Single-sample t statistic	M	μ	$\sqrt{\frac{s^2}{n}}$	$s^2 = \frac{SS}{df}$
Independent-measures t statistic	$(M_1 - M_2)$	$(\mu_1 - \mu_2)$	$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$	$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$

$df_1 = n_1 - 1$
$df_2 = n_2 - 1$

$$\text{pooled variance} = s_p^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2}$$

Summary

- **Paired test:**

  simple, reduced to one-sample test.

- **independent test:**

  equal variance?