Unit 6: Comparing Two Populations or Groups

Section 10.2

Comparing Two Means

Section 10.2
Comparing Two Means

Learning Objectives

After this section, you should be able to…

- ✓ DESCRIBE the characteristics of the sampling distribution of the difference between two sample means
- ✓ CALCULATE probabilities using the sampling distribution of the difference between two sample means
- ✓ DETERMINE whether the conditions for performing inference are met
- ✓ USE two-sample *t* procedures to compare two means based on summary statistics or raw data
- ✓ INTERPRET computer output for two-sample *t* procedures
- ✓ INTERPRET the results of inference procedures
- ✓ PERFORM a significance test to compare two means
- ✓ INTERPRET the results of inference procedures

- **What does the CLT say about the Sampling Distribution of a Difference Between Two Means**

The Sampling Distribution of the Difference Between Sample Means

Choose an SRS of size n_1 from Population 1 with mean μ_1 and standard deviation σ_1 and an independent SRS of size n_2 from Population 2 with mean μ_2 and standard deviation σ_2.

Shape When the population distributions are Normal, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is approximately Normal. In other cases, the sampling distribution will be approximately Normal if the sample sizes are large enough ($n_1 \geq 30, n_2 \geq 30$).

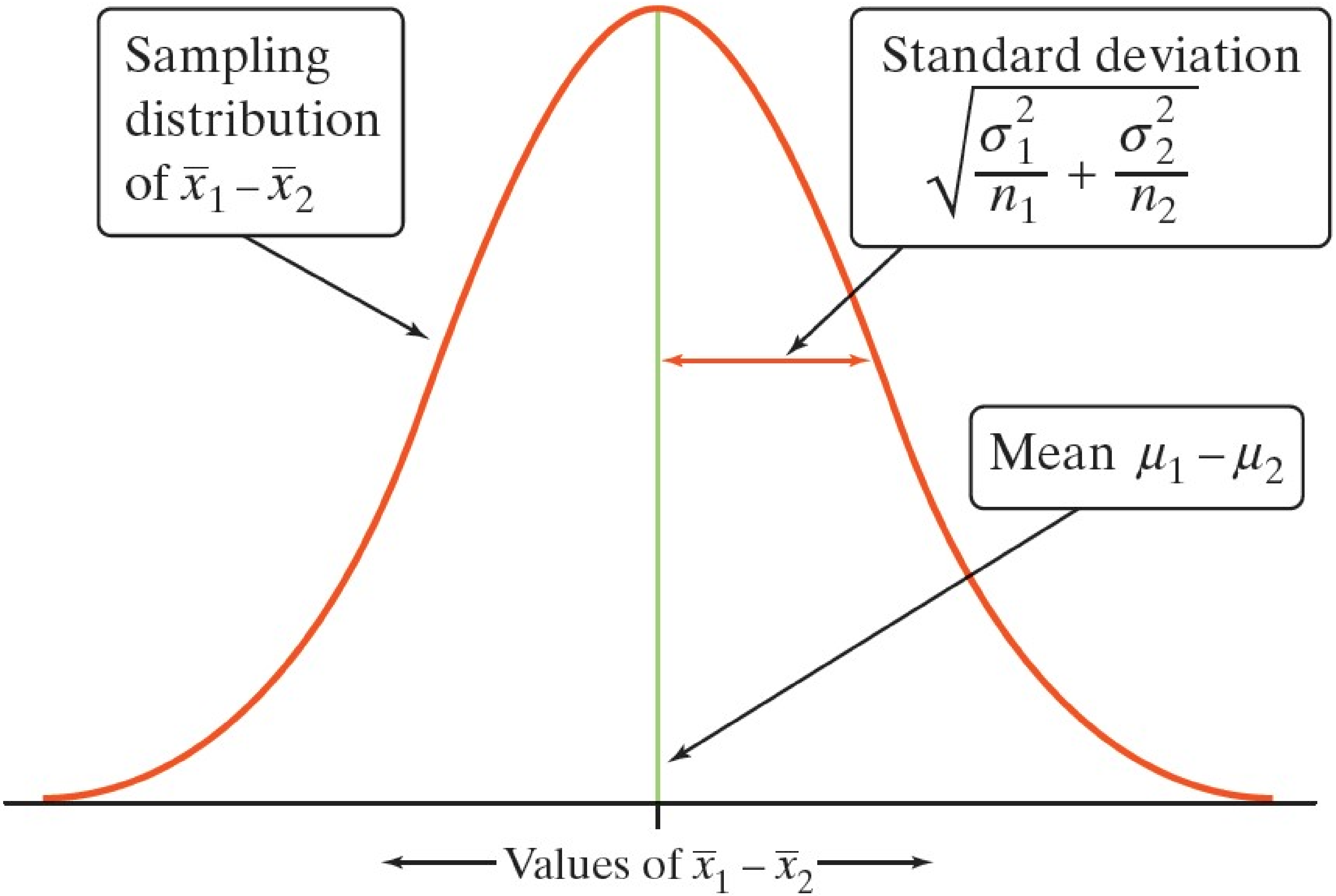
Center The mean of the sampling distribution is $\mu_1 - \mu_2$. That is, the difference in sample means is an unbiased estimator of the difference in population means.

Spread The standard deviation of the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

as long as each sample is no more than 10% of its population (10% condition).

- **The Sampling Distribution of a Difference Between Two Means**



Example: Who's Taller at Ten, Boys or Girls?

- Based on information from the U.S. National Health and Nutrition Examination Survey (NHANES), the heights (in inches) of ten-year-old girls follow a Normal distribution *N*(56.4, 2.7). The heights (in inches) of ten-year-old boys follow a Normal distribution *N*(55.7, 3.8). A researcher takes independent SRSs of 12 girls and 8 boys of this age and measures their heights. After analyzing the data, the researcher reports that the sample mean height of the boys is larger than the sample mean height of the girls.

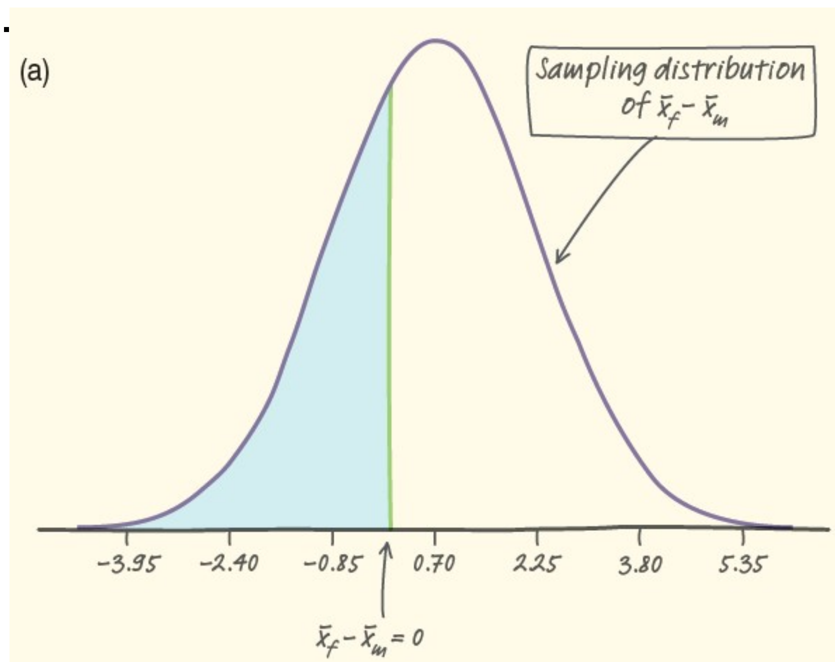
a) Describe the shape, center, and spread of the sampling distribution of $\bar{x}_f - \bar{x}_m$.

Because both population distributions are Normal, the sampling distribution of $\bar{x}_f - \bar{x}_m$ is Normal.

Its mean is $\mu_f - \mu_m = 56.4 - 55.7 = 0.7$ inches.

Its standard deviation is

$$\sqrt{\frac{2.7^2}{12} + \frac{3.8^2}{8}} = 1.55 \text{ inches.}$$



The Two-Sample *t* Statistic

When data come from two random samples or two groups in a randomized experiment, the statistic $\bar{x}_1 - \bar{x}_2$ is our best guess for the value of $\mu_1 - \mu_2$.

When the Independent condition is met, the standard deviation of the statistic $\bar{x}_1 - \bar{x}_2$ is:

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}$$

Since we don't know the values of the parameters σ_1 and σ_2, we replace them in the standard deviation formula with the sample standard deviations. The result is the **standard error** of the statistic $\bar{x}_1 - \bar{x}_2$:

$$\sqrt{\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}}$$

The two-sample *t* statistic has approximately a *t* distribution. We can use technology to determine degrees of freedom OR we can use a conservative approach, using the smaller of $n_1 - 1$ *and* $n_2 - 1$ for the degrees of freedom.

Confidence Intervals for $\mu_1 - \mu_2$

Two-Sample *t* Interval for a Difference Between Means

When the Random, Normal, and Independent conditions are met, an approximate level C confidence interval for $(\bar{x}_1 - \bar{x}_2)$ is

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where t^* is the critical value for confidence level C for the t distribution with degrees of freedom from either technology or the smaller of $n_1 - 1$ and $n_2 - 1$.

Random The data are produced by a random sample of size n_1 from Population 1 and a random sample of size n_2 from Population 2 or by two groups of size n_1 and n_2 in a randomized experiment.

Normal Both population distributions are Normal OR both sample group sizes are large ($n_1 \geq 30$ and $n_2 \geq 30$).

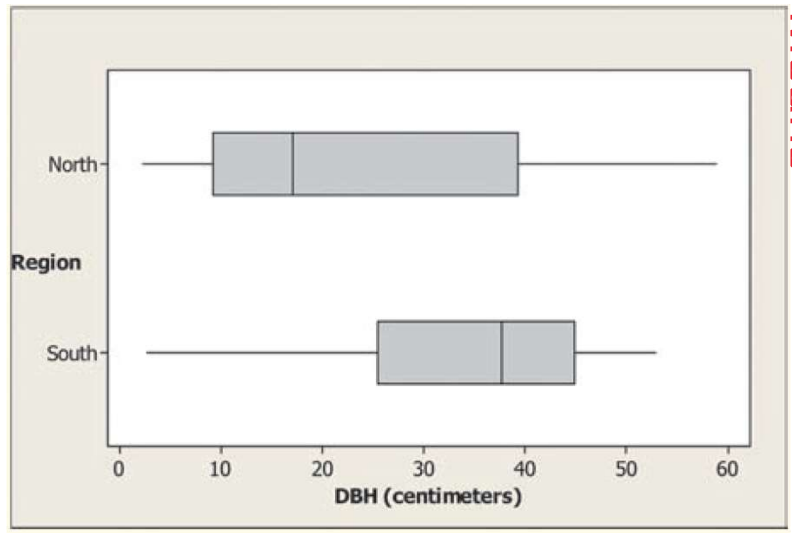
Independent Both the samples or groups themselves and the individual observations in each sample or group are independent. When sampling without replacement, check that the two populations are at least 10 times as large as the corresponding samples (the 10% condition).

Big Trees, Small Trees, Short Trees, Tall Trees

The Wade Tract Preserve in Georgia is an old-growth forest of longleaf pines that has survived in a relatively undisturbed state for hundreds of years. One question of interest to foresters who study the area is "How do the sizes of longleaf pine trees in the northern and southern halves of the forest compare?" To find out, researchers took random samples of 30 trees from each half and measured the diameter at breast height (DBH) in centimeters. Comparative boxplots of the data and summary statistics from Minitab are shown below. Construct and interpret a 90% confidence interval for the difference in the mean DBH for longleaf pines in the northern and southern halves of the Wade Tract Preserve.

Descriptive Statistics: North, South

Variable	N	Mean	StDev
North	30	23.70	17.50
South	30	34.53	14.26



State: We want to estimate the difference $\mu_1 - \mu_2$ at a 90% confidence level. Our parameters of interest are μ_1 = the true mean DBH of all trees in the southern half of the forest and μ_2 = the true mean DBH of all trees in the northern half of the forest.

Big Trees, Small Trees, Short Trees, Tall Trees

Plan: We should use a two-sample *t* interval for $\mu_1 - \mu_2$ if the conditions are satisfied.

✓ **Random** The data come from a random samples of 30 trees each from the northern and southern halves of the forest.

✓ **Normal** The boxplots give us reason to believe that the population distributions of DBH measurements may not be Normal. However, since both sample sizes are at least 30, we are safe using *t* procedures.

✓ **Independent** Researchers took independent samples from the northern and southern halves of the forest. Because sampling without replacement was used, there have to be at least 10(30) = 300 trees in each half of the forest. This is pretty safe to assume.

Do: Since the conditions are satisfied, we can construct a two-sample *t* interval for the difference $\mu_1 - \mu_2$. We'll use the conservative df = 30-1 = 29.

Upper-tail probability *p*

df	.10	.05	.025
28	1.313	1.70	2.048
29	1.311	1.699	2.045
30	1.310	1.697	2.042
	80%	90%	95%

Confidence level *C*

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}} = (34.5 - 23.70) \pm 1.699\sqrt{\frac{14.26^2}{30} + \frac{17.50^2}{30}}$$

$$= 10.83 \pm 7.00 = (3.83,\ 17.83)$$

STEP 4

Big Trees, Small Trees, Short Trees, Tall Trees

Conclude: We are 90% confident that the interval from 3.83 to 17.83 centimeters captures the difference in the actual mean DBH of the southern trees and the actual mean DBH of the northern trees. This interval suggests that the mean diameter of the southern trees is between 3.83 and 17.83 cm larger than the mean diameter of the northern trees.

Significance Tests for $\mu_1 - \mu_2$

An observed difference between two sample means can reflect an actual difference in the parameters, or it may just be due to chance variation in random sampling or random assignment. Significance tests help us decide which explanation makes more sense. The null hypothesis has the general form

$$H_0: \mu_1 - \mu_2 = \text{hypothesized value}$$

We're often interested in situations in which the hypothesized difference is 0. Then the null hypothesis says that there is no difference between the two parameters:

$$\boldsymbol{H_0: \mu_1 - \mu_2 = 0 \text{ or, alternatively, } H_0: \mu_1 = \mu_2}$$

The alternative hypothesis says what kind of difference we expect.

$$\boldsymbol{H_a: \mu_1 - \mu_2 > 0,\ H_a: \mu_1 - \mu_2 < 0, \text{ or } H_a: \mu_1 - \mu_2 \neq 0}$$

If the Random, Normal, and Independent conditions are met, we can proceed with calculations.

Significance Tests for $\mu_1 - \mu_2$

To do a test, standardize $\bar{x}_1 - \bar{x}_2$ to get a two-sample t statistic:

$$\text{test statistic} = \frac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

To find the *P*-value, use the *t* distribution with degrees of freedom given by technology or by the conservative approach (df = smaller of n_1 - 1 *and* n_2 - 1).

- **Two-Sample *t* Test for The Difference Between Two Means**

- If the following conditions are met, we can proceed with a two-sample *t* test for the difference between two means:

Random The data are produced by a random sample of size n_1 from Population 1 and a random sample of size n_2 from Population 2 or by two groups of size n_1 and n_2 in a randomized experiment.

Normal Both population distributions (or the true distributions of responses to the two treatments) are Normal OR both sample group sizes are large ($n_1 \geq 30$ and $n_2 \geq 30$).

Independent Both the samples or groups themselves and the individual observations in each sample or group are independent. When sampling without replacement, check that the two populations are at least 10 times as large as the corresponding samples (the 10% condition).

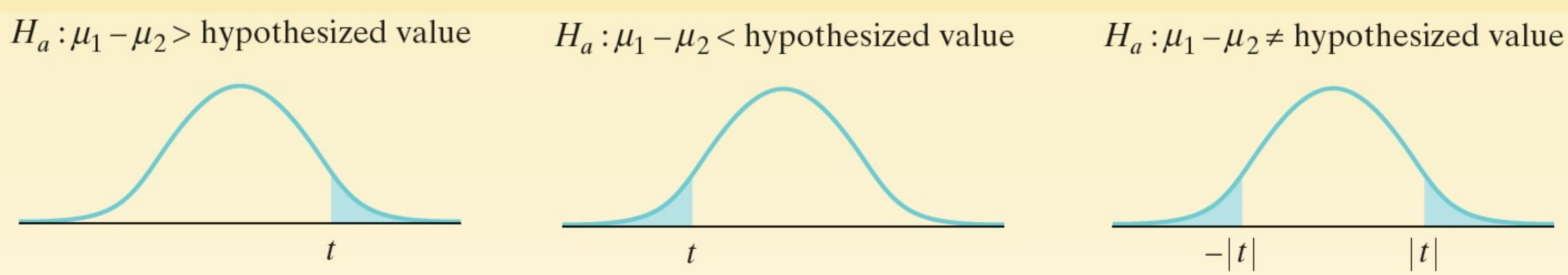
Two-Sample *t* Test for The Difference Between Two Means

Two-Sample *t* Test for the Difference Between Two Means

Suppose the Random, Normal, and Independent conditions are met. To test the hypothesis $H_0: \mu_1 - \mu_2 =$ hypothesized value, compute the t statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^{\ 2}}{n_1} + \dfrac{s_2^{\ 2}}{n_2}}}$$

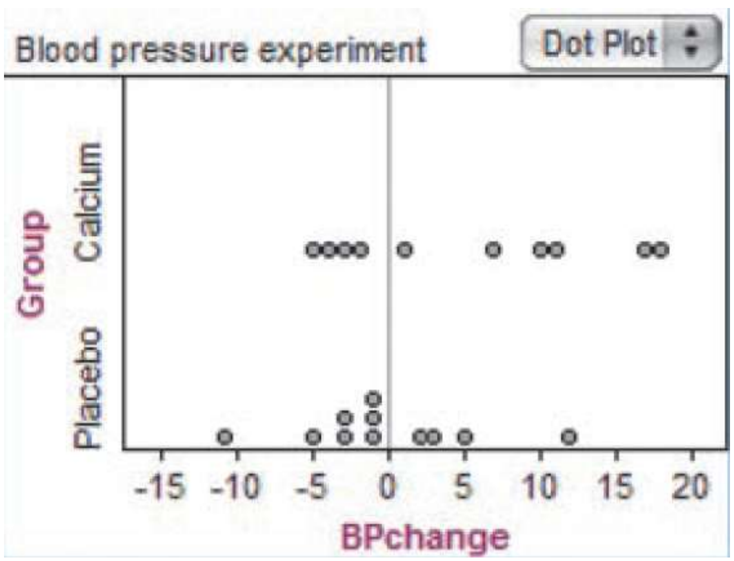
Find the P - value by calculating the probabilty of getting a t statistic this large or larger in the direction specified by the alternative hypothesis H_a. Use the t distribution with degrees of freedom approximated by technology or the smaller of $n_1 - 1$ and $n_2 - 1$.



Example: Calcium and Blood Pressure

Does increasing the amount of calcium in our diet reduce blood pressure? Examination of a large sample of people revealed a relationship between calcium intake and blood pressure. The relationship was strongest for black men. Such observational studies do not establish causation. Researchers therefore designed a randomized comparative experiment. The subjects were 21 healthy black men who volunteered to take part in the experiment. They were randomly assigned to two groups: 10 of the men received a calcium supplement for 12 weeks, while the control group of 11 men received a placebo pill that looked identical. The experiment was double-blind. The response variable is the decrease in systolic (top number) blood pressure for a subject after 12 weeks, in millimeters of mercury. An increase appears as a negative response Here are the data:

Group 1 (calcium):	7	−4	18	17	−3	−5	1	10	11	−2	
Group 2 (placebo):	−1	12	−1	−3	3	−5	5	2	−11	−1	−3



State: We want to perform a test of

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_a: \mu_1 - \mu_2 > 0$$

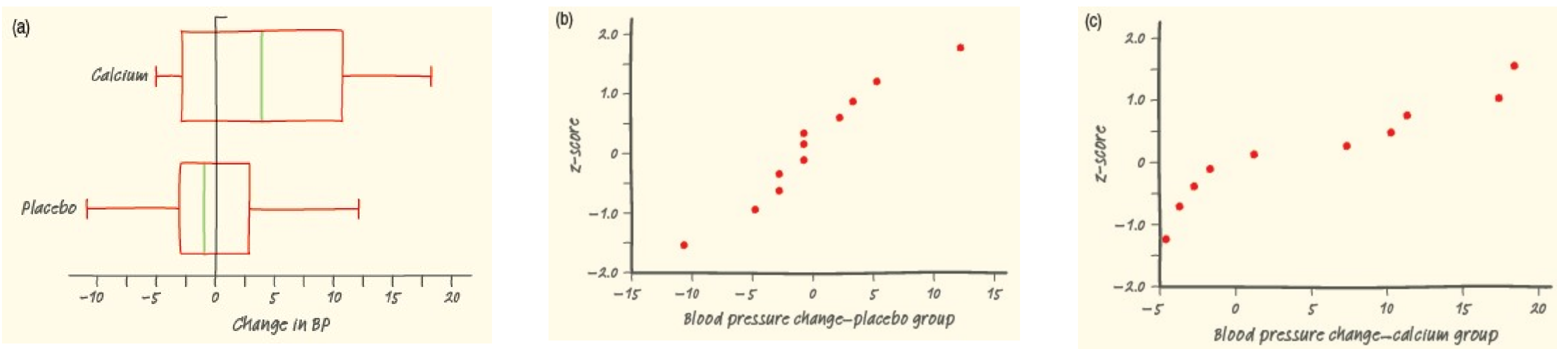
where μ_1 = the true mean decrease in systolic blood pressure for healthy black men like the ones in this study who take a calcium supplement, and μ_2 = the true mean decrease in systolic blood pressure for healthy black men like the ones in this study who take a placebo.
We will use $\alpha = 0.05$.

Example: Calcium and Blood Pressure

Plan: If conditions are met, we will carry out a two-sample *t* test for $\mu_1 - \mu_2$.

• **Random** The 21 subjects were randomly assigned to the two treatments.

• **Normal** With such small sample sizes, we need to examine the data to see if it's reasonable to believe that the actual distributions of differences in blood pressure when taking calcium or placebo are Normal. Hand sketches of calculator boxplots and Normal probability plots for these data are below:



The boxplots show no clear evidence of skewness and no outliers. The Normal probability plot of the placebo group's responses looks very linear, while the Normal probability plot of the calcium group's responses shows some slight curvature. With no outliers or clear skewness, the *t* procedures should be pretty accurate.

• **Independent** Due to the random assignment, these two groups of men can be viewed as independent. Individual observations in each group should also be independent: knowing one subject's change in blood pressure gives no information about another subject's response.

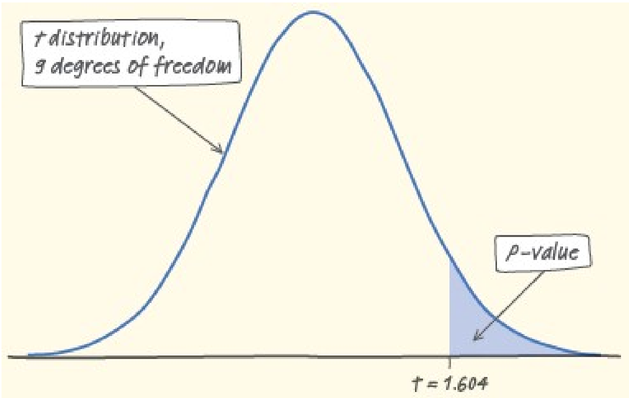
Example: Calcium and Blood Pressure

Do: Since the conditions are satisfied, we can perform a two-sample *t* test for the difference $\mu_1 - \mu_2$.

Test statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{[5.000 - (-0.273)] - 0}{\sqrt{\frac{8.743^2}{10} + \frac{5.901^2}{11}}} = 1.604$$

	Upper tail probability *p*		
df	**.10**	**.05**	**.025**
8	1.397	1.860	2.306
9	1.383	1.833	2.262
10	1.372	1.812	2.228



P-value Using the conservative df = 10 – 1 = 9, we can use Table B to show that the *P*-value is between 0.05 and 0.10.

Conclude: Because the *P*-value is greater than $\alpha = 0.05$, we fail to reject H_0. The experiment provides some evidence that calcium reduces blood pressure, but the evidence is not convincing enough to conclude that calcium reduces blood pressure more than a placebo.

Example: Calcium and Blood Pressure

We can estimate the difference in the true mean decrease in blood pressure for the calcium and placebo treatments using a two-sample *t* interval for $\mu_1 - \mu_2$. To get results that are consistent with the one-tailed test at $\alpha = 0.05$ from the example, we'll use a 90% confidence level. The conditions for constructing a confidence interval are the same as the ones that we checked in the example before performing the two-sample *t* test.

With df = *9*, the critical value for a 90% confidence interval is $t^* = 1.833$.

Upper-tail probability *p*			
df	**.10**	**.05**	**.025**
8	1.397	1.860	2.306
9	1.383	1.833	2.262
10	1.372	1.812	2.228

The interval is:

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = [5.000 - (-0.273)] \pm 1.833\sqrt{\frac{8.743^2}{10} + \frac{5.901^2}{11}}$$

$$= 5.273 \pm 6.027$$

$$= (-0.754, 11.300)$$

We are 90% confident that the interval from -0.754 to 11.300 captures the difference in true mean blood pressure reduction on calcium over a placebo. Because the 90% confidence interval includes 0 as a plausible value for the difference, we cannot reject H_0: $\mu_1 - \mu_2 = 0$ against the two-sided alternative at the $\alpha = 0.10$ significance level or against the one-sided alternative at the $\alpha = 0.05$ significance level.

Using Two-Sample *t* Procedures Wisely

The two-sample *t* procedures are more robust against non-Normality than the one-sample *t* methods. When the sizes of the two samples are equal and the two populations being compared have distributions with similar shapes, probability values from the *t* table are quite accurate for a broad range of distributions when the sample sizes are as small as $n_1 = n_2 = 5$.

Using the Two-Sample *t* Procedures: The Normal Condition

•*Sample size less than 15:* Use two-sample *t* procedures if the data in both samples/groups appear close to Normal (roughly symmetric, single peak, no outliers). If the data are clearly skewed or if outliers are present, do not use *t*.

• *Sample size at least 15:* Two-sample *t* procedures can be used except in the presence of outliers or strong skewness.

• *Large samples:* The two-sample *t* procedures can be used even for clearly skewed distributions when both samples/groups are large, roughly $n \geq 30$.

Using Two-Sample *t* Procedures Wisely

Here are several cautions and considerations to make when using two-sample *t* procedures.

✓ **In planning a two-sample study, choose equal sample sizes if you can.**

✓ **Do not use "pooled" two-sample *t* procedures!**

✓ **We are safe using two-sample *t* procedures for comparing two means in a randomized experiment.**

✓ **Do not use two-sample *t* procedures on paired data!**

✓ **Beware of making inferences in the absence of randomization. The results may not be generalized to the larger population of interest.**

Homework

Chapter 10, #'s 33, 39b&c, 40b&c, 43, 49, 53
Topic 25 additional notes

Section 10.2
Comparing Two Means

Summary

In this section, we learned that…

- ✓ Choose an SRS of size n_1 from Population 1 and an independent SRS of size n_2 from Population 2. The sampling distribution of the difference of sample means has:

  Shape Normal if both population distributions are Normal; approximately Normal otherwise if both samples are large enough ($n \geq 30$).

  Center The mean $\mu_1 - \mu_2$.

  Spread As long as each sample is no more than 10% of its population (10% condition), its standard deviation is $\sqrt{\frac{s_1^{\ 2}}{n_n} + \frac{s_2^{\ 2}}{n_2}}$.

- ✓ Confidence intervals and tests for the difference between the means of two populations or the mean responses to two treatments $\mu_1 - \mu_2$ are based on the difference between the sample means.

- ✓ If we somehow know the population standard deviations σ_1 and σ_2, we can use a z statistic and the standard Normal distribution to perform probability calculations.

Section 10.2
Comparing Two Means

Summary

✓ The conditions for two-sample *t* procedures are:

Random The data are produced by a random sample of size n_1 from Population 1 and a random sample of size n_2 from Population 2 or by two groups of size n_1 and n_2 in a randomized experiment.

Normal Both population distributions (or the true distributions of responses to the two treatments) are Normal OR both sample/group sizes are large ($n_1 \geq 30$ and $n_2 \geq 30$).

Independent Both the samples or groups themselves and the individual observations in each sample or group are independent. When sampling without replacement, check that the two populations are at least 10 times as large as the corresponding samples (the 10% condition).

Section 10.2
Comparing Two Means

Summary

✓ The level C **two-sample *t* interval for $\mu_1 - \mu_2$** is

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where *t** is the critical value for confidence level *C* for the *t* distribution with degrees of freedom from either technology or the conservative approach.

+

Section 10.2
Comparing Two Means

Summary

- Since we almost never know the population standard deviations in practice, we use the **two-sample *t* statistic**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- where *t* has approximately a *t* distribution with degrees of freedom found by technology or by the conservative approach of using the smaller of $n_1 - 1$ and $n_2 - 1$.

- The conditions for two-sample *t* procedures are:

  Random The data are produced by a random sample of size n_1 from Population 1 and a random sample of size n_2 from Population 2 or by two groups of size n_1 and n_2 in a randomized experiment.

  Normal Both population distributions (or the true distributions of responses to the two treatments) are Normal OR both sample/group sizes are large ($n_1 \geq 30$ and $n_2 \geq 30$).

  Independent Both the samples or groups themselves and the individual observations in each sample or group are independent. When sampling without replacement, check that the two populations are at least 10 times as large as the corresponding samples (the 10% condition).

Section 10.2
Comparing Two Means

Summary

- To test H_0: $\mu_1 - \mu_2$ = hypothesized value, use a **two-sample *t* test for $\mu_1 - \mu_2$**. The test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

P-values are calculated using the *t* distribution with degrees of freedom from either technology or the conservative approach.

Section 10.2
Comparing Two Means

Summary

- ✓ The two-sample *t* procedures are quite robust against departures from Normality, especially when both sample/group sizes are large.

- ✓ Inference about the difference $\mu_1 - \mu_2$ in the effectiveness of two treatments in a completely randomized experiment is based on the **randomization distribution** of the difference between sample means. When the Random, Normal, and Independent conditions are met, our usual inference procedures based on the sampling distribution of the difference between sample means will be approximately correct.

- ✓ Don't use two-sample *t* procedures to compare means for paired data.