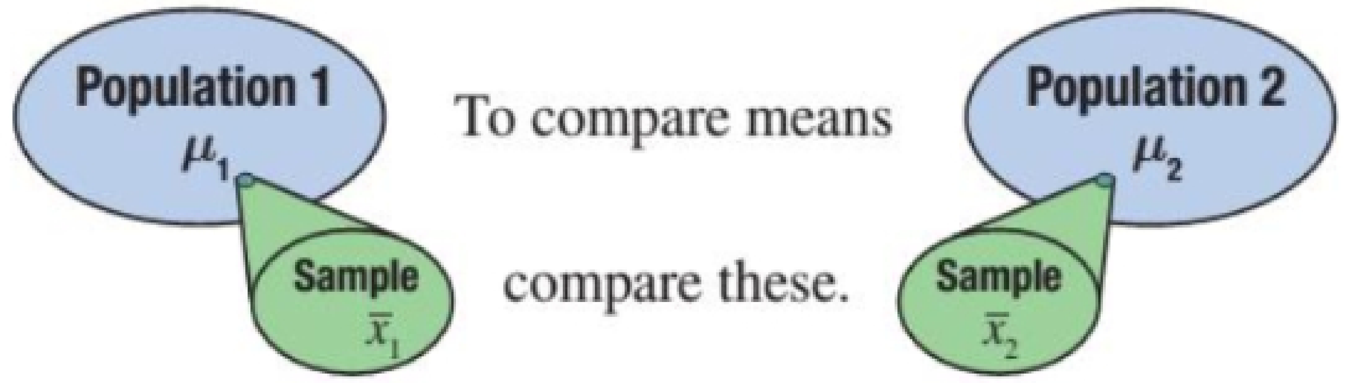
“An honorable man will not be bullied by a hypothesis.”

— Bergen Evans

12 CHAPTER Inferences about Two Samples

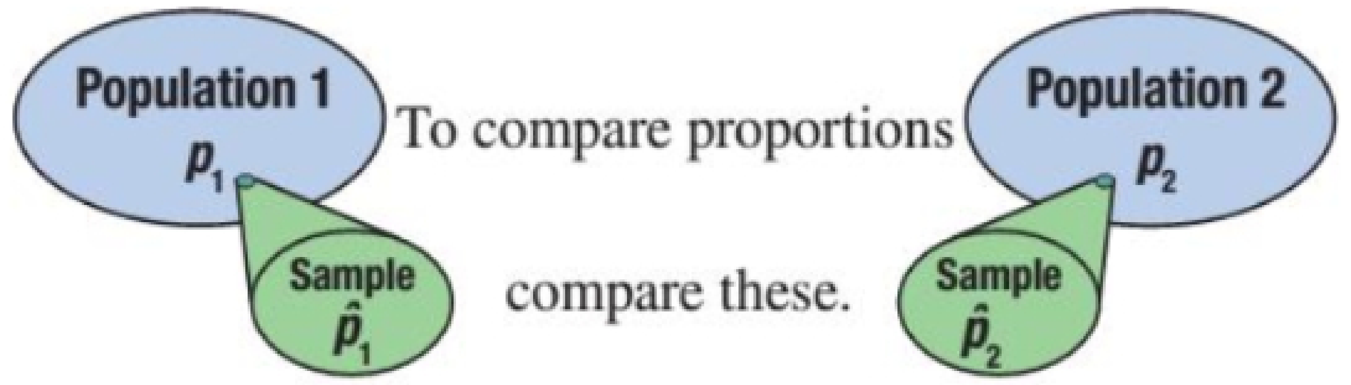
With respect to comparing two population means, we will develop methods for comparing two population means, and try to answer questions like the following.

- Is $\mu_1 > \mu_2$?
- Is $\mu_1 < \mu_2$?
- Is $\mu_1 = \mu_2$?



Similarly, we will develop methods for comparing two population proportions, and try to answer questions like the following.

- Is $p_1 > p_2$?
- Is $p_1 < p_2$?
- Is $p_1 = p_2$?



Properties of the Sampling Distribution of $\bar{x}_1 - \bar{x}_2$

If $n_1 > 30$ and $n_2 > 30$, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ has an approximately normal distribution. As well, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is normally distributed if both samples come from populations that are normally distributed.

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$ if the two samples are independent.

σ_1 and σ_2 are known.

PROPERTIES

100(1 − α)% Confidence Interval for $\mu_1 - \mu_2$

The 100(1 − α)% confidence interval estimate for the difference in the population means of two independent populations is given by

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

where $z_{\alpha/2}$ is the critical value of the standard normal distribution with an area of $\alpha/2$ in the upper tail. σ_1 and σ_2 are known, and the independent samples are drawn from normally distributed populations or the samples sizes are large; i.e., $n_1 > 30$ and $n_2 > 30$.

FORMULA

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Service Outage Time (in minutes)			
	n	$\bar{x}$	σ
Cable	50	8.8	3
Satellite	50	9.5	2

95%

$$n_1 = 50,\ n_2 = 50,\ \sigma_1 = 3,\ \sigma_2 = 2,\ \bar{x}_1 = 8.8,\ \text{and}\ \bar{x}_2 = 9.5$$

$-1.7 \longleftrightarrow 0.3$

Table 12.1.1 - Hypotheses Concerning a Test About Two Means			
	Are the population means different?	**Is the population mean in group 1 greater than the population mean in group 2?**	**Is the population mean in group 1 less than the population mean in group 2?**
Null hypothesis, H_0	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 = 0$
Alternative Hypothesis, H_a	$\mu_1 - \mu_2 \neq 0$	$\mu_1 - \mu_2 > 0$	$\mu_1 - \mu_2 < 0$
Type of Hypothesis Test	Two-tailed	Right-tailed	Left-tailed

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Service Outage Time (in minutes)			
	n	$\bar{x}$	σ
Cable	50	8.8	3
Satellite	50	9.5	2

$\alpha = 0.05$

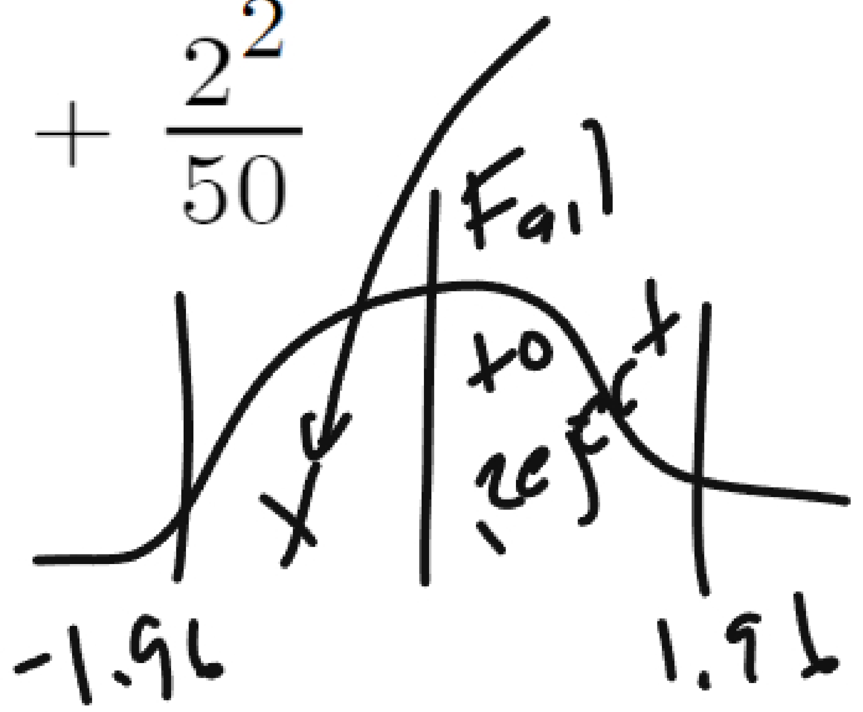
95%

$$n_1 = 50,\ n_2 = 50,\ \sigma_1 = 3,\ \sigma_2 = 2,\ \bar{x}_1 = 8.8, \text{ and } \bar{x}_2 = 9.5$$

$H_0 \rightarrow \mu_1 - \mu_2 = 0$

$H_a \rightarrow \mu_1 - \mu_2 \neq 0$

$$z = \frac{8.8 - 9.5}{\sqrt{\frac{3^2}{50} + \frac{2^2}{50}}} = -1.37$$



An elementary school teacher is interested in knowing if fifth-grade girls take longer to read a passage from a book than fifth-grade boys. She randomly selects 40 fifth-grade boys and 40 fifth-grade girls for the study. She gives each student several pages of the sameσ book to read. The time it takes each group to complete the reading is recorded in minutes. The results of the study are shown in the following table.

$$\frac{11-10}{\sqrt{\frac{2^2}{40}+\frac{3^2}{40}}}$$

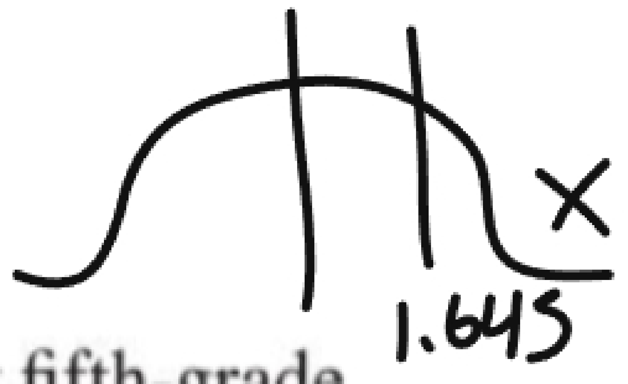
$= 1.75$

Reading Times (in minutes)

	n	$\bar{x}$	σ
Girls	40	11	2
Boys	40	10	3

$H_0 \rightarrow \mu_1 - \mu_2 = 0$

$H_a \rightarrow \mu_1 - \mu_2 > 0$



Reject

Is there persuasive evidence for the teacher to conclude at $\alpha = 0.05$ that fifth-grade girls have longer reading times than fifth-grade boys?

Hypothesis Testing

Testing a Hypothesis about Two Independent Population Means, σ_1 and σ_2 Unknown but Equal

Assumptions:

1. The samples are independent of one another.
2. Both samples are simple random samples.
3. The populations from which the two samples are drawn are normally distributed or the sample sizes are both large ($n_1 > 30$ and $n_2 > 30$).
4. Both of the populations have approximately equal, but unknown standard deviations, i.e., $\sigma_1 = \sigma_2 = \sigma$.

Continued…

PROCEDURE

Testing a Hypothesis about Two Independent Population Means, σ_1 and σ_2 Unknown but Equal (Continued)

Test Statistic:

If the assumptions outlined above are satisfied, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ has a t-distribution with $n_1 + n_2 - 2$ degrees of freedom. The t-test statistic is given by

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ is the **pooled sample variance**.

Military Training Programs			
	n	$\bar{x}$	**s**
Program A	50	85	10
Program B	55	87	9

$\alpha = 0.05$

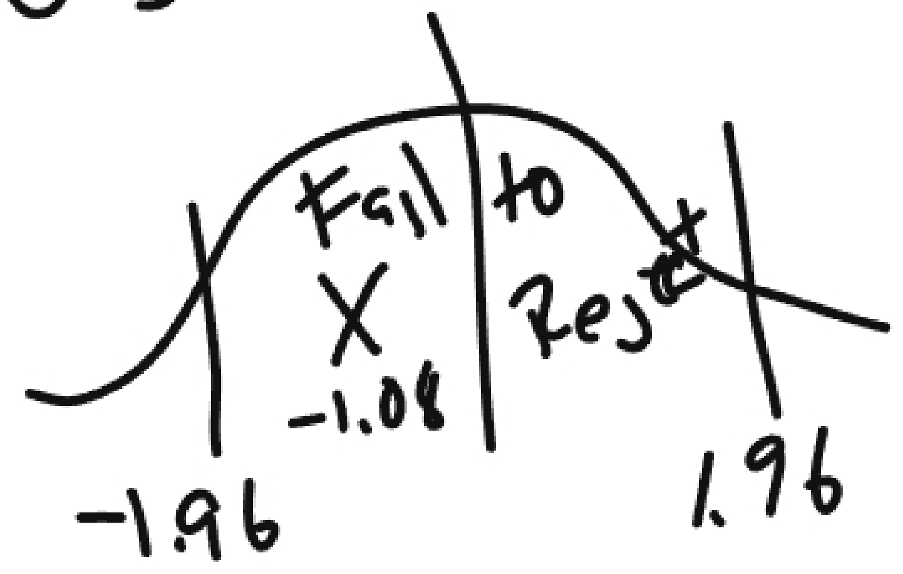
$H_0 \rightarrow \mu_1 - \mu_2 = 0$

$H_a \rightarrow \mu_1 - \mu_2 \neq 0$

$$\text{d.f.} \Rightarrow 50 + 55 - 2 = 103$$

$$s_p^2 = \frac{(50-1)(10^2) + (55-1)(9)^2}{103} = 90.04$$

$$t = \frac{85 - 87}{\sqrt{90.04\left(\frac{1}{50} + \frac{1}{55}\right)}} = -1.08$$



Hypothesis Testing of $\mu_1 - \mu_2$ with σ_1 and σ_2 Unknown and Unequal

Assumptions:

1. The samples are independent of one another.
2. Both samples are simple random samples.
3. The populations from which the two samples are drawn are approximately normally distributed or the sample sizes are both large ($n_1 > 30$ and $n_2 > 30$).
4. The values of σ_1 and σ_2 are unknown and we do not assume they are equal.

Test Statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \text{[redacted]}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

The test statistic follows a *t*-distribution and there are two options for calculating the degrees of freedom.

Option 1: Use the smaller of $n_1 - 1$ and $n_2 - 1$ for the degrees of freedom.

A biologist is interested in comparing the size of blue crabs in two river basins: (1) Cooper and (2) Stono River. Based on the health of the rivers, she believes the crabs in the Stono will have a higher average weight. She samples 32 crabs from the Stono River, and this sample has a mean weight of 800 g with a standard deviation of 225 g. She takes a random sample of 35 crabs from the Cooper River and finds the crabs have a mean weight of 700 g with a standard deviation of 175 g.

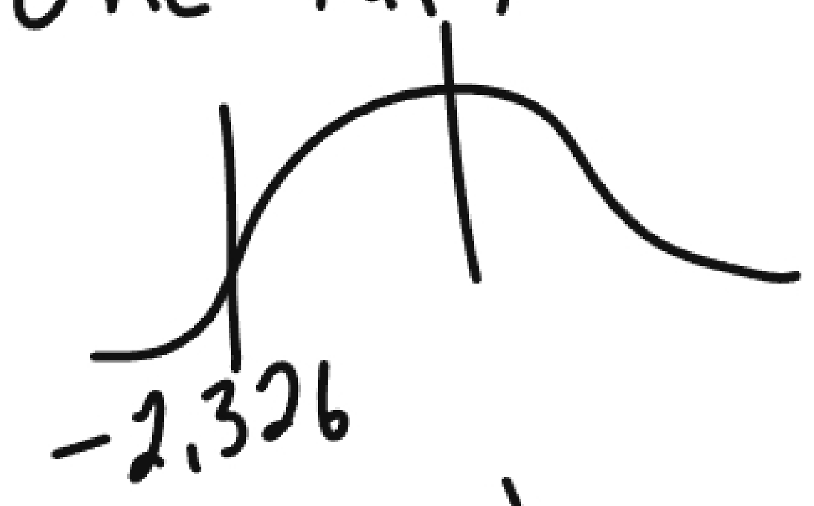
$H_0 \to \mu_1 - \mu_2 = 0$

$H_a \to \mu_1 - \mu_2 < 0$

$df = 32 - 1 = 31$

$\alpha = 0.01$

one-tail



$$t = \frac{700 - 800}{\sqrt{\frac{175^2}{35} + \frac{225^2}{32}}} = -2.017$$

Fail to Reject

A cereal manufacturer has advertised that its product, Fiber Oat Flakes, has a lower fat content than its competitor, Bran Flakes Plus. Because of complaints from the manufacturers of Bran Flakes Plus, the FDA has decided to test the claim that Fiber Oat Flakes has a lower average fat content than Bran Flakes Plus. Several boxes of each cereal are selected and the fat content per serving is measured. The results of the study are as follows. Assume that the

Fat Content (Grams)			
	n	$\bar{x}$	s
Fiber Oat Flakes	16	5	1
Bran Flakes Plus	15	6	2