Comparing Two Means and Two Proportions

Comparing Two Population Means by Using Independent Samples

- Suppose a random sample has been taken from each of two different populations
- Suppose that the populations are independent of each other
  - Then the random samples are independent of each other
- Then the sampling distribution of the difference in sample means is normally distributed

Sampling Distribution of the Difference of Two Sample Means #1

- Suppose population 1 has mean μ_1 and variance σ_1^2
  - From population 1, a random sample of size n_1 is selected which has mean x_1 and variance s_1^2
- Suppose population 2 has mean μ_2 and variance σ_2^2
  - From population 2, a random sample of size n_2 is selected which has mean x_2 and variance s_2^2
- Then the sample distribution of the difference of two sample means…

Sampling Distribution of the Difference of Two Sample Means #2

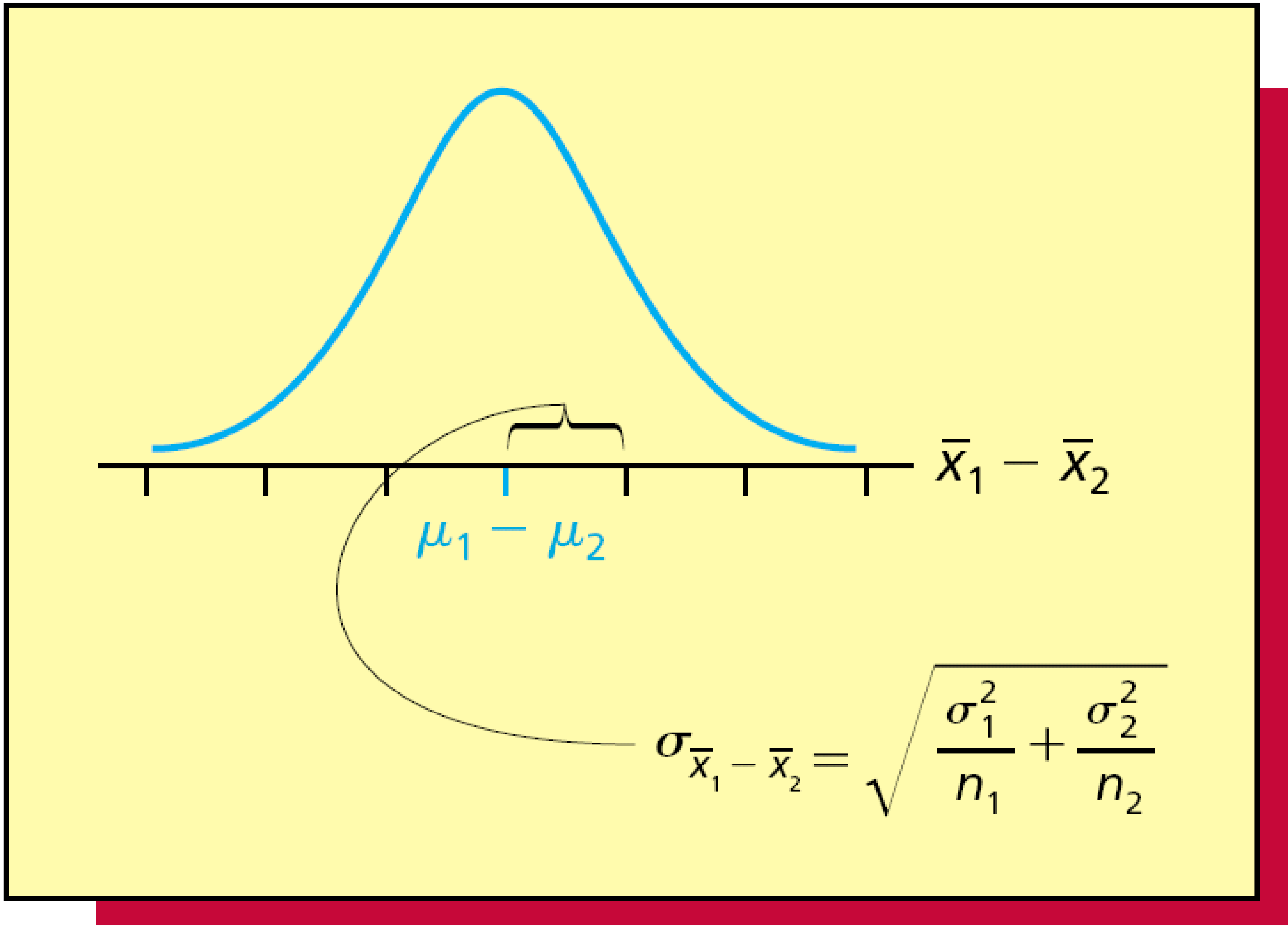
- Is normal, if each of the sampled populations is normal
  - Approximately normal if the sample sizes n_1 and n_2 are large

- Has mean $\mu_{x_1 - x_2} = \mu_1 - \mu_2$

- Has standard deviation

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

Sampling Distribution of the Difference of Two Sample Means #3



t-Based Confidence Interval for the Difference in Means: Equal Variances

A 100(1 − α) percent confidence interval for $\mu_1 - \mu_2$ is

$$\left[(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right] \quad \text{where} \quad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

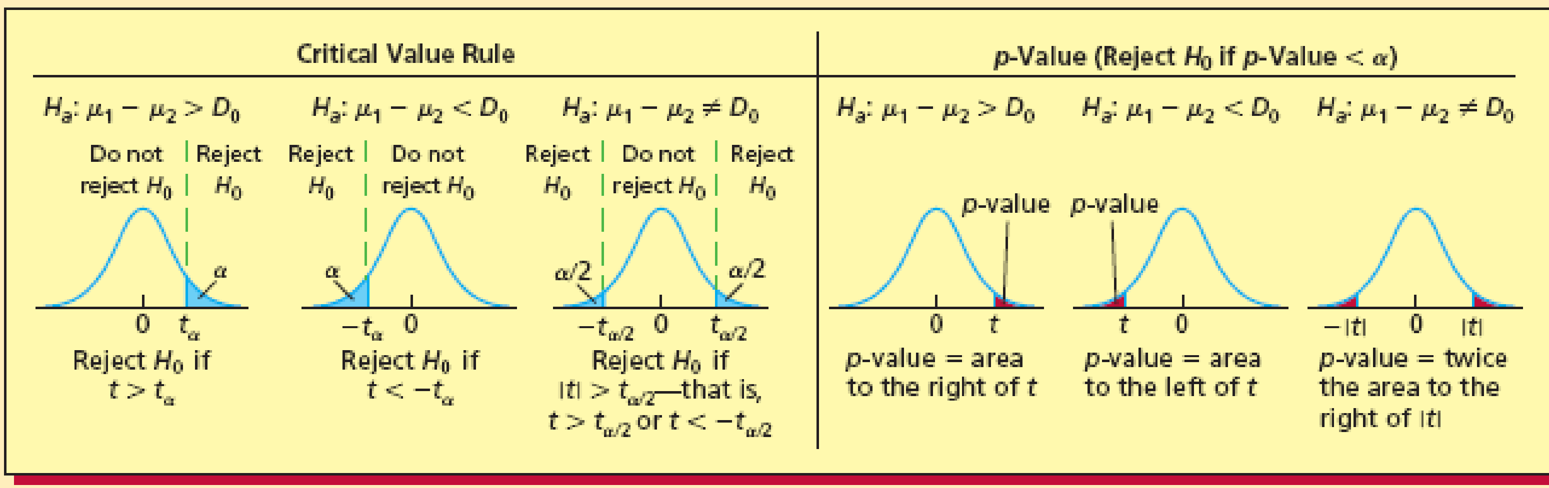
and $t_{\alpha/2}$ is based on $(n_1 + n_2 - 2)$ degrees of freedom.

t-Based Test About the Difference in Means: Equal Variance

Null Hypothesis $H_0: \mu_1 - \mu_2 = D_0$

Test Statistic $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

Assumptions
Independent samples
and
Equal variances
and either
Normal populations
or
Large sample sizes



Here t_α, $t_{\alpha/2}$, and the p-values are based on $n_1 + n_2 - 2$ degrees of freedom.

t-Based Confidence Intervals and t Tests: Unequal Variances

Paired Difference Experiments

- Before, drew random samples from two different populations
- Now, have two different processes (or methods)
- Draw one random sample of units and use those units to obtain the results of each process

Paired Difference Experiments Continued

- For instance, use the same individuals for the results from one process vs. the results from the other process
  - E.g., use the same individuals to compare “before” and “after” treatments
- Using the same individuals, eliminates any differences in the individuals themselves and just comparing the results from the two processes

Paired Difference Experiments #3

- Let μ_d be the mean of population of paired differences
  - $\mu_d = \mu_1 - \mu_2$, where μ_1 is the mean of population 1 and μ_2 is the mean of population 2
- Let $\bar{d}$ and s_d be the mean and standard deviation of a sample of paired differences that has been randomly selected from the population
  - $\bar{d}$ is the mean of the differences between pairs of values from both samples

t-Based Confidence Interval for Paired Differences in Means

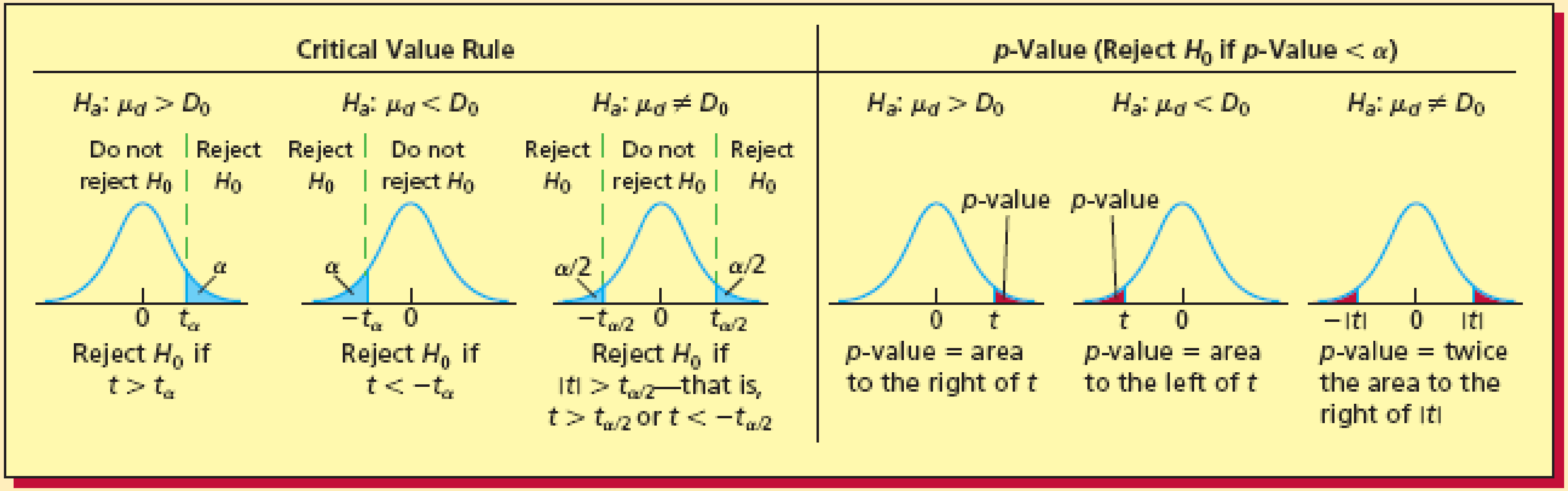
- If the sampled population of differences is normally distributed with mean μ_d
- A (1- α)100% confidence interval for $\mu_d = \mu_1 - \mu_2$ is…

$$\left[\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}\right]$$

- where for a sample of size n, $t_{\alpha/2}$ is based on n – 1 degrees of freedom

Test Statistic for Paired Differences

Null Hypothesis	$H_0: \mu_d = D_0$	Test Statistic	$t = \frac{\bar{d} - D_0}{s_d / \sqrt{n}}$ $\quad df = n - 1$	Assumptions	Normal population of paired differences or Large sample size



Comparing Two Population Proportions by Using Large, Independent Samples

- Select a random sample of size n_1 from a population, and let $\hat{p}_1$ denote the proportion of units in this sample that fall into the category of interest
- Select a random sample of size n_2 from another population, and let $\hat{p}_2$ denote the proportion of units in this sample that fall into the same category of interest
- Suppose that n_1 and n_2 are large enough
  - $n_1 \cdot p_1 \geq 5$, $n_1 \cdot (1 - p_1) \geq 5$, $n_2 \cdot p_2 \geq 5$, and $n_2 \cdot (1 - p_2) \geq 5$

Comparing Two Population Proportions Continued

- Then the population of all possible values of $\hat{p}_1 - \hat{p}_2$
- Has approximately a normal distribution if each of the sample sizes n_1 and n_2 is large
  - Here, n_1 and n_2 are large enough so $n_1 \cdot p_1 \geq 5$, $n_1 \cdot (1 - p_1) \geq 5$, $n_2 \cdot p_2 \geq 5$, and $n_2 \cdot (1 - p_2) \geq 5$
- Has mean $\mu_{\hat{p}1 - \hat{p}2} = p_1 - p_2$
- Has standard deviation

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}$$

Confidence Interval for the Difference of Two Population Proportions

If the random samples are independent of each other, then the following is a $100(1-\alpha)$ percent confidence interval for $\hat{p}_1$ - $\hat{p}_2$

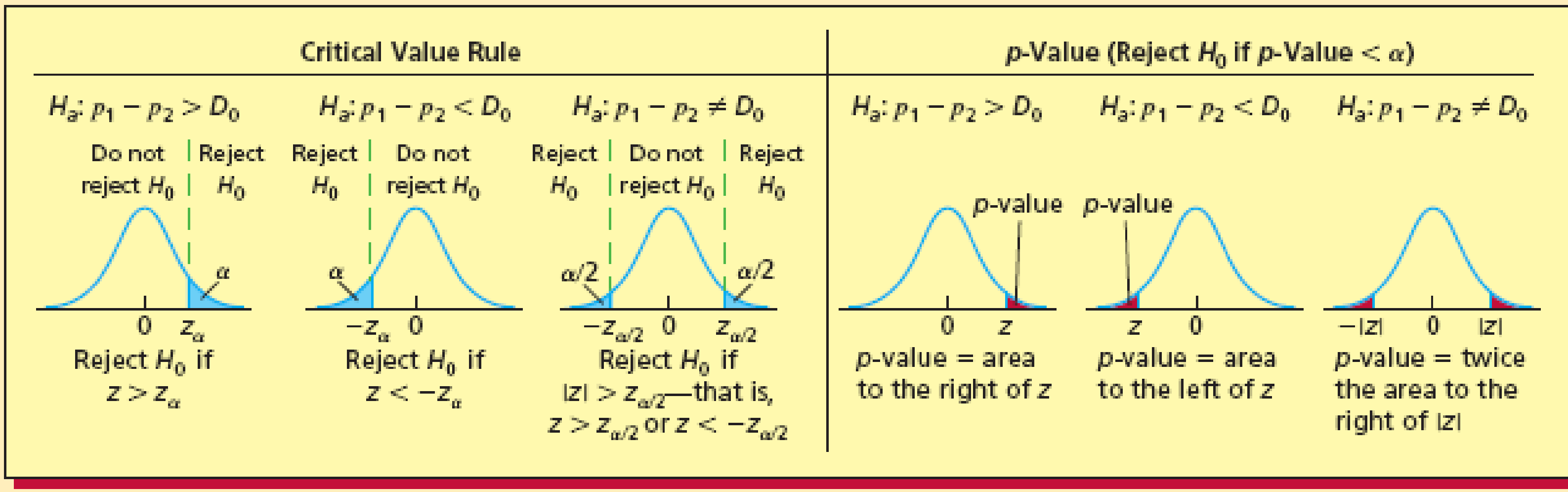
$$\left[\left(\hat{p}_1 - \hat{p}_2\right) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{n_1} + \frac{\hat{p}_2\left(1-\hat{p}_2\right)}{n_2}}\right]$$

Test Statistic for the Difference of Two Population Proportions

Null Hypothesis $H_0: p_1 - p_2 = D_0$

Test Statistic $z = \dfrac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sigma_{\hat{p}_1 - \hat{p}_2}}$

Assumptions Independent samples and Large sample sizes



$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$