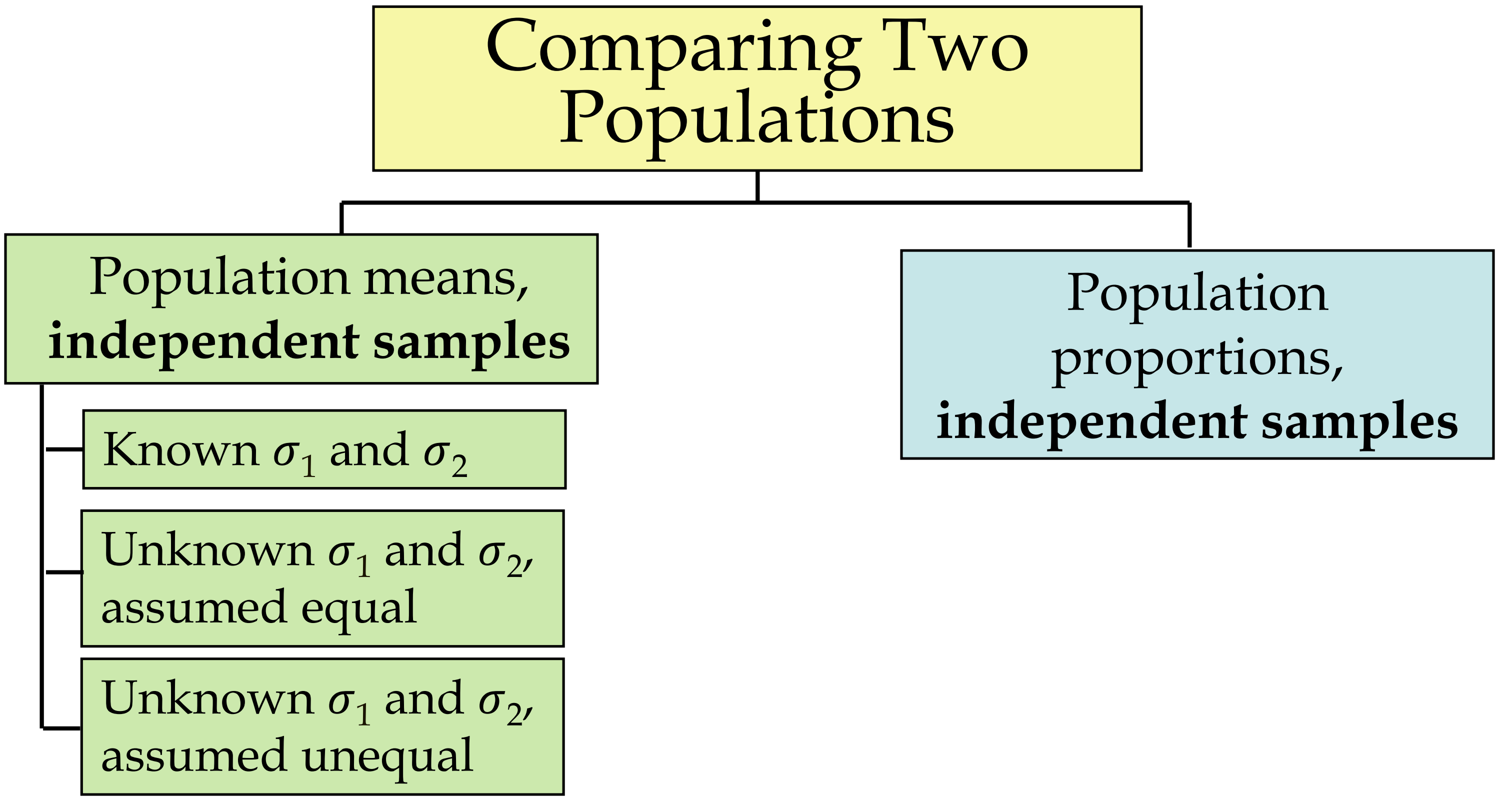
Inference About Two Populations

- Comparing Two Population Means:
  - σ_1 and σ_2 are Known
  - σ_1 and σ_2 are Unknown and Assumed Equal
  - σ_1 and σ_2 are Unknown and Assumed Unequal
- Comparing Two Population Proportions (*if time permits*)
- Reading: Chapter 10 (skip Section 10.3; Section 10.4 only if we talk about last bullet point in class)

Hypothesis Tests to Compare Two Populations



Inferences About the Difference Between Two Population Means

- Are the checkout times different at *Kroger* and *Walmart*?
- Did the recession of 2008-2009 cause the sale prices of homes to fall? (prices *before* and *after* the recession)
- Do television commercials increase purchase potential of the consumers? (those who *watch* vs those who *do not watch*)
- Are the annual cost of attending *public colleges* lower as compared with *private colleges*?

Our Goal

⇒ Our goal is to make an inference about difference between **population means** $\mu_1 - \mu_2$ having information about the difference between **sample means** $\bar{x}_1 - \bar{x}_2$ from random samples taken from Population 1 and Population 2

Hypothesis Tests to Compare Two Means

Three possible hypothesis statements:

Two-tail test:	One-tail test: (Upper tail)	One-tail test: (Lower tail)
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$H_0: \mu_1 \leq \mu_2$ $H_1: \mu_1 > \mu_2$	$H_0: \mu_1 \geq \mu_2$ $H_1: \mu_1 < \mu_2$

These tests are equivalent to hypothesis tests about the **difference** between two population means

$H_0: \mu_1 - \mu_2 = 0$ $H_1: \mu_1 - \mu_2 \neq 0$	$H_0: \mu_1 - \mu_2 \leq 0$ $H_1: \mu_1 - \mu_2 > 0$	$H_0: \mu_1 - \mu_2 \geq 0$ $H_1: \mu_1 - \mu_2 < 0$

Where Are We Going?

- To test the hypotheses, we need an appropriate test statistic
- Previously, we looked at the test statistic:

$$Test\ Statistic = \frac{Point\ Estimate\ - Hypothesized\ Value}{Standard\ Error}$$

Here, we test **the difference** between population means

$\Rightarrow$ Need to discuss the point estimate of the **difference** between two population means and the standard error for the **difference** between two means

$\Rightarrow$ Generally, need to discuss the sampling distribution of $\bar{x}_1 - \bar{x}_2$

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

The **sampling distribution** of $\bar{x}_1 - \bar{x}_2$, the difference in means, is the result of subtracting the sampling distribution for the mean of one population from the sampling distribution for the mean of a second population

The **point estimate** of $\mu_1 - \mu_2$, the difference between the means of Populations 1 and 2:

$$\bar{x}_1 - \bar{x}_2$$

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

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

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

Standard Error of the Difference between Two Means

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

where σ_1 and σ_2 = standard deviations of Populations 1 and 2

n_1 and n_2 = sample sizes from Populations 1 and 2

Comparing Two Population Means: Known Standard Deviations

Assumptions:

- Independent Samples
  - When samples are independent, the results we observe when sampling from one population have no impact on the results we observe when sampling from another population
  - In other words, the process that generates one sample is completely separate from the process that generates the other sample
- If the sample sizes are small ($n < 30$), populations should follow the normal distribution
- If the populations are not normally distributed, the sample sizes should be large ($n \geq 30$)

Hypothesis Test to Compare Two Means: σ_1 and σ_2 are Known

Test Statistic:

- z-test statistic for a hypothesis test for the difference between two means

$$z_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

where $(\mu_1 - \mu_2)_{H_0}$ = The hypothesized difference

$\bar{x}_1 - \bar{x}_2$ = The difference in sample means

$\sigma_{\bar{x}_1 - \bar{x}_2}$ = The standard error of $\bar{x}_1 - \bar{x}_2$

Critical Value and the *p*-value:

- Use the standard normal distribution

Excel Exercise >>

Hypothesis Test to Compare Two Means

Example: Is there a difference between the amount a fan spends on food at a baseball game in Chicago compared to a fan in New York?

- A random sample of 40 Chicago fans had mean spending of $\bar{x}_1$ = \$13.60; a sample of 36 New York fans had mean $\bar{x}_2$ = \$14.75
- Assume that the *population* standard deviations are known, σ_1 = \$1.80 for Chicago and σ_2 = \$2.50 for New York

Hypothesis Test to Compare Two Means

Step 1: Identify the null and alternative hypotheses

- Let Population 1 = Chicago fan spending on food
  Population 2 = New York fan spending on food

$H_0: \mu_1 - \mu_2 = 0$ (no difference in spending between cities)

$H_1: \mu_1 - \mu_2 \neq 0$ (spending amounts are different)

Step 2: Set a value of the significance level, α

- Suppose that $\alpha = 0.05$ is chosen

Hypothesis Test to Compare Two Means

Step 3: Calculate the appropriate test statistic

- z-test statistic for a hypothesis test for the difference between two means when σ_1 and σ_2 are known

$$z_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

For the example,

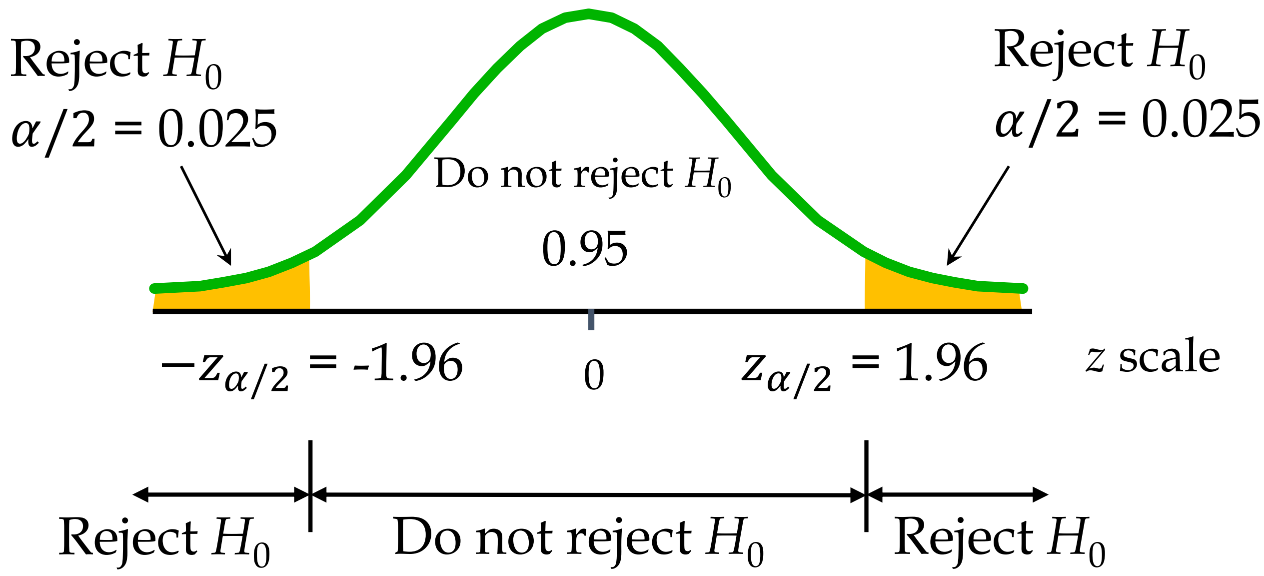
$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{\$1.80^2}{40} + \frac{\$2.50^2}{36}} = \$0.505$$

$$z_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(\$13.60 - \$14.75) - 0}{\$0.505} = -2.28$$

Hypothesis Test to Compare Two Means

Step 4: Determine the appropriate critical value

- σ_1 and σ_2 are known $\Rightarrow$ use the critical z-score
- Critical z-score identifies borders of the rejection region
- Since this is a two-tail test, α = 0.05 is split evenly between two tails:



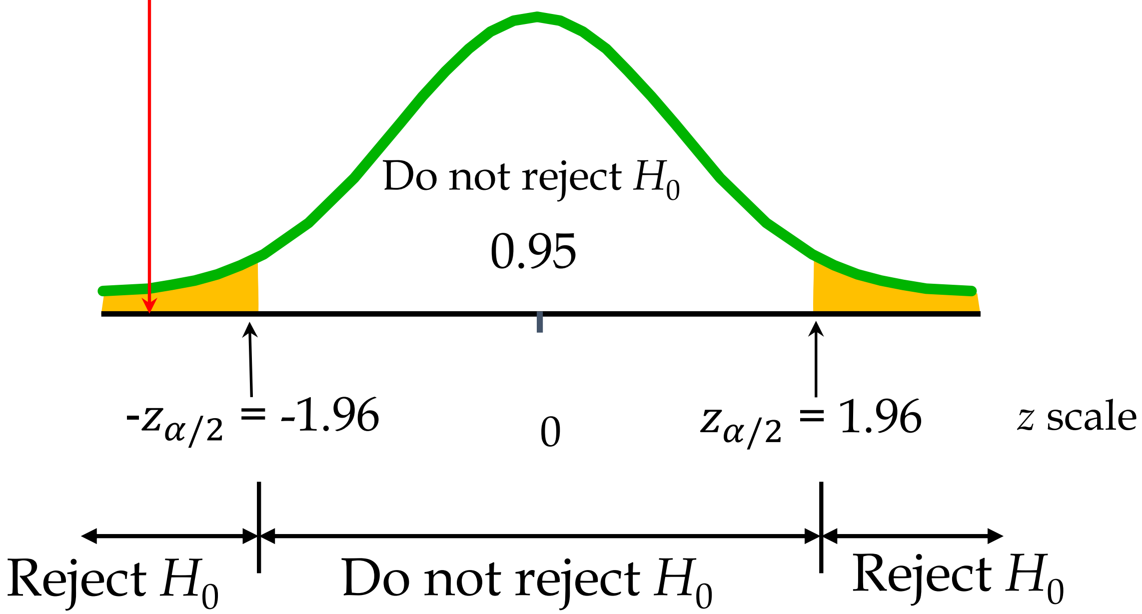
Hypothesis Test to Compare Two Means

Step 5: Compare the test statistic $z_{\bar{x}}$ and the critical value

- For a two-tail test, reject the null hypothesis if

$$|z_{\bar{x}}| > |z_{\alpha/2}|$$

- Here, $|z_{\bar{x}}| = 2.28$ is greater than $|z_{\alpha/2}| = 1.96 \Rightarrow$ **reject H_0**



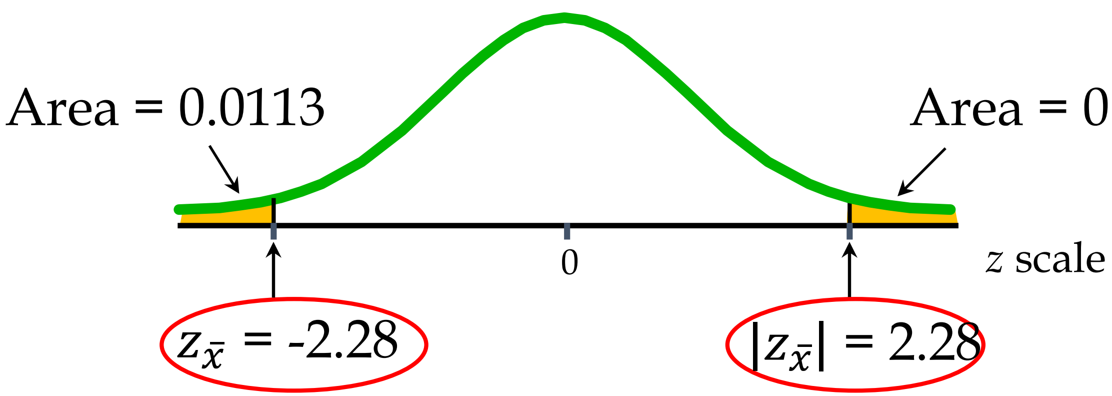
Hypothesis Test to Compare Two Means

Step 4: Calculate the *p*-value

Since this is a two-tail test and $z_{\bar{x}} = -2.28 < 0$:

$$p - \text{value} = 2 \times P(z < z_{\bar{x}}) = 2 \times P(z < -2.28)$$

$$= 2 \times (1 - 0.9887) = 0.0226$$



Step 5: Compare *p*-value and the significance level

Because *p*-value = 0.0226 < α = 0.05 $\Rightarrow$ **we reject H_0**

Hypothesis Test to Compare Two Means

Conclusion

By rejecting the null hypothesis, we have support for the alternative hypothesis.

According to these two samples, we have evidence to conclude that the average spending on food by fans in Chicago is **not equal** to the average in New York

Hypothesis Test to Compare Two Means: σ_1 and σ_2 are Unknown

When population standard deviations (σ_1 and σ_2) are **unknown**, then:

- We use the sample standard deviations (s_1 and s_2) to *estimate* the standard error of the sampling distribution
- The Student's t-distribution is used to calculate the critical value and the p-value
- Both populations need to be normally distributed or both sample sizes should be 30 or larger

Two cases to consider:

- Case 1: The population variances are equal ($\sigma_1^2 = \sigma_2^2$)
- Case 2: The population variances are not equal ($\sigma_1^2 \neq \sigma_2^2$)

Case 1: Population Variances Are Equal

- Independent Samples
- Unknown Population Standard Deviations (σ_1 and σ_2)
- The population variances are **equal** ($\sigma_1^2 = \sigma_2^2$)

Hypothesis testing procedures follow the same steps as when σ_1 and σ_2 are known, except for these differences:

1. The test statistic is a *t* test statistic
2. The critical value and the *p*-value are calculated from the *t* distribution with $df = n_1 + n_2 - 2$
3. A **pooled variance** is calculated to estimate the standard error of the sampling distribution of $\bar{x}_1 - \bar{x}_2$

Case 1: Population Variances Are Equal

- Independent Samples
- Unknown Population Standard Deviations (σ_1 and σ_2)
- The population variances are **equal** ($\sigma_1^2 = \sigma_2^2$)

Pooled Variance

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

where s_1^2 = The variance of the sample from Population 1

s_2^2 = The variance of the sample from Population 2

n_1 and n_2 = The sample sizes from Populations 1 and 2

Case 1: Population Variances Are Equal

- Independent Samples
- Unknown Population Standard Deviations (σ_1 and σ_2)
- The population variances are **equal** ($\sigma_1^2 = \sigma_2^2$)

Test Statistic:

- t test statistic for a hypothesis test for the difference between two means (σ_1 and σ_2 are unknown but assumed equal)

$$t_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where s_p^2 = The pooled variance

Excel Exercise >>

Case 2: Population Variances Are Unequal

- Independent Samples
- Unknown Population Standard Deviations (σ_1 and σ_2)
- The population variances are **unequal** ($\sigma_1^2 \neq \sigma_2^2$)

When σ_1^2 and σ_2^2 are **unequal**, a pooled variance is not appropriate, since the sample variances are not estimates of a common population variance

1. The sample values s_1 and s_2 are used as separate estimates of the unknown (but different) values of σ_1 and σ_2
2. The test statistic is a *t*-test statistic
3. The critical value and the *p*-value are calculated from the *t* distribution
4. The degrees of freedom are determined differently

Case 2: Population Variances Are Unequal

- Independent Samples
- Unknown Population Standard Deviations (σ_1 and σ_2)
- The population variances are **unequal** ($\sigma_1^2 \neq \sigma_2^2$)

Degrees of freedom for the *t* distribution:

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

Note: Always **round down** the degrees of freedom. This makes it more challenging to reject the null hypothesis, which is a more conservative approach.

Case 2: Population Variances Are Unequal

- Independent Samples
- Unknown Population Standard Deviations (σ_1 and σ_2)
- The population variances are **unequal** ($\sigma_1{}^2 \neq \sigma_2{}^2$)

Test Statistic:

- t test statistics for a hypothesis test for the difference between two means (σ_1 and σ_2 are unknown and assumed unequal)

$$t_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

Excel Exercise >>

Hypothesis Test to Compare Two Means

Example: Specific Motors of Detroit has developed a new automobile known as the M car. 24 M cars and 28 J cars (from Japan) were road tested to compare miles-per-gallon (mpg) performance. The sample statistics:

M Cars	J Cars	
24 cars	28 cars	Sample Size
29.8 mpg	27.3 mpg	Sample Mean
2.56 mpg	1.81 mpg	Sample Std. Dev.

Can you conclude, using 5% level of significance, that the miles-per-gallon (*mpg*) performance of M cars is greater than the miles-per-gallon performance of J cars assuming that population standard deviations are different for M and J cars?

Note: Assume that Population 1 and 2 are normally distributed

Hypothesis Test to Compare Two Means

Step 1: Identify the null and alternative hypotheses

Let μ_1= mean mpg for the Population of M cars
μ_2= mean mpg for the Population of J cars

H_0: $\mu_1 \leq \mu_2$ (mpg of M cars is no greater than J cars)
H_1: $\mu_1 > \mu_2$ (mpg of M cars is greater than J cars)

In other words,
H_0: $\mu_1 - \mu_2 \leq 0$
H_1: $\mu_1 - \mu_2 > 0$

Step 2: Significance level α = 0.05 is given

Hypothesis Test to Compare Two Means

Step 3: Calculate the appropriate test statistic

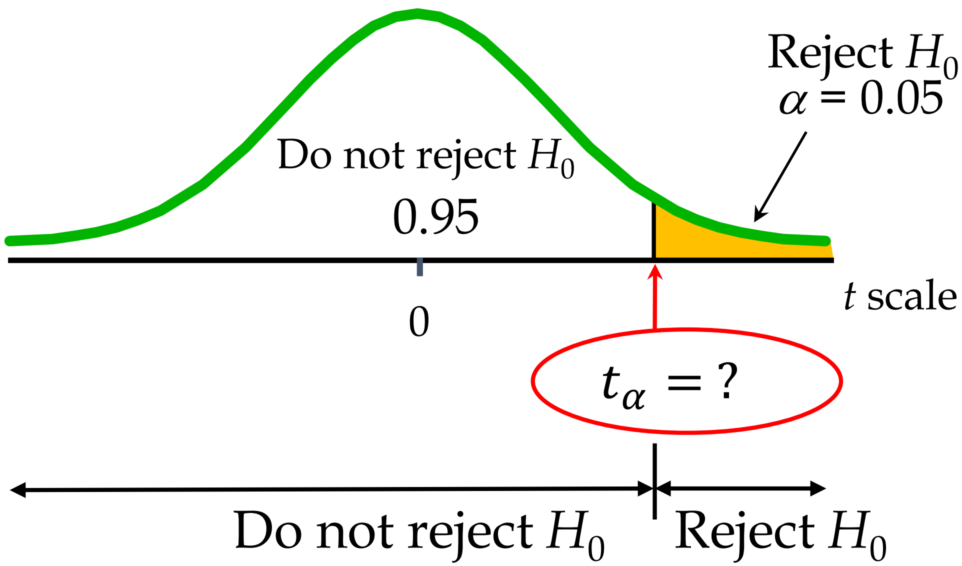
- *t* test statistic for a hypothesis test for the difference between two means when the population standard deviations are not known and assumed unequal

$$t_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(29.8 - 27.3) - 0}{\sqrt{\frac{(2.56)^2}{24} + \frac{(1.81)^2}{28}}} = 4.003$$

Hypothesis Test to Compare Two Means

Step 4: Determine the appropriate critical value

- σ_1 and σ_2 are unknown and assumed unequal $\Rightarrow$ use the t score
- Since this is a one-tail test (upper) the area for $\alpha = 0.05$ is put in the upper or the right tail of the distribution



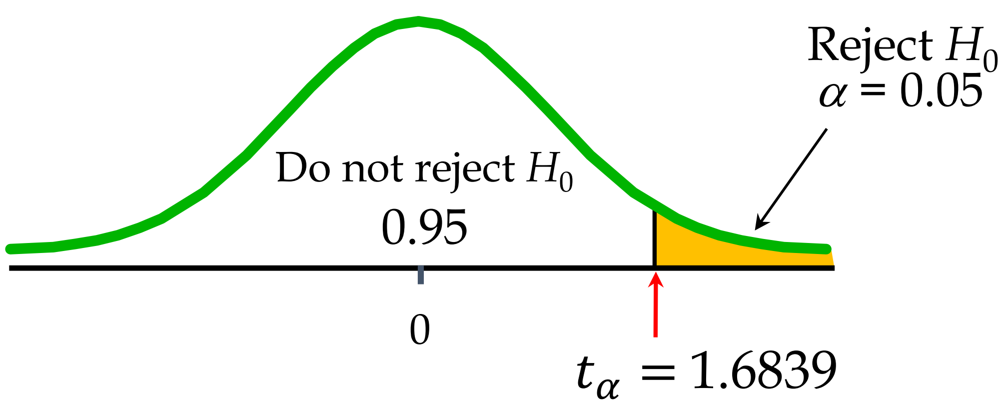
Hypothesis Test to Compare Two Means

Step 4: Determine the appropriate critical value

- Calculate the degrees of freedom for the *t* distribution

$$df = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}} = \frac{\left(\frac{(2.56)^2}{24}+\frac{(1.81)^2}{28}\right)^2}{\frac{\left(\frac{(2.56)^2}{24}\right)^2}{24-1}+\frac{\left(\frac{(1.81)^2}{28}\right)^2}{28-1}} = 40.566 = 40$$

$\Rightarrow$ Use $df = 40$ to find the critical value and the *p*-value



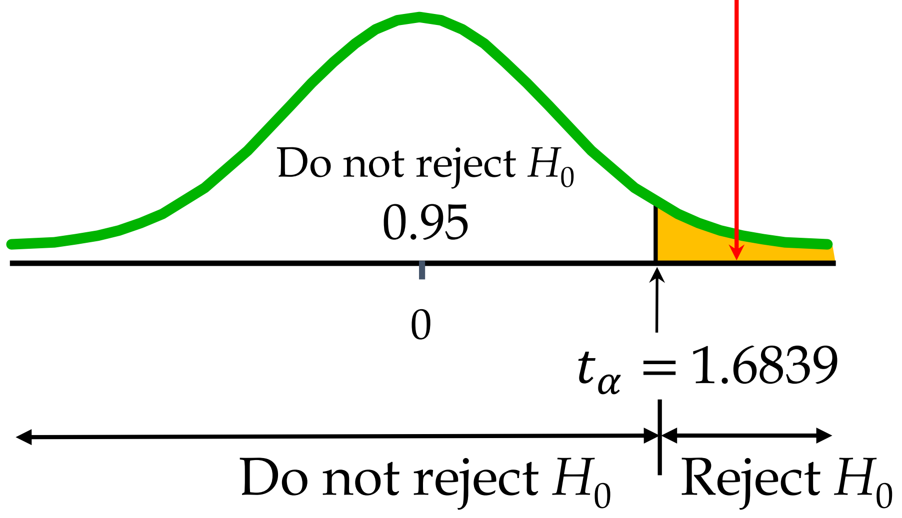
Hypothesis Test to Compare Two Means

Step 5: Compare the test statistic $t_{\bar{x}}$ and the critical value

- For a one-tail test (upper), reject the null hypothesis if

$$t_{\bar{x}} > t_{\alpha}$$

- Here, $t_{\bar{x}} = 4.003$ is greater than $t_{\alpha}=1.6839 \Rightarrow$ **reject H_0**

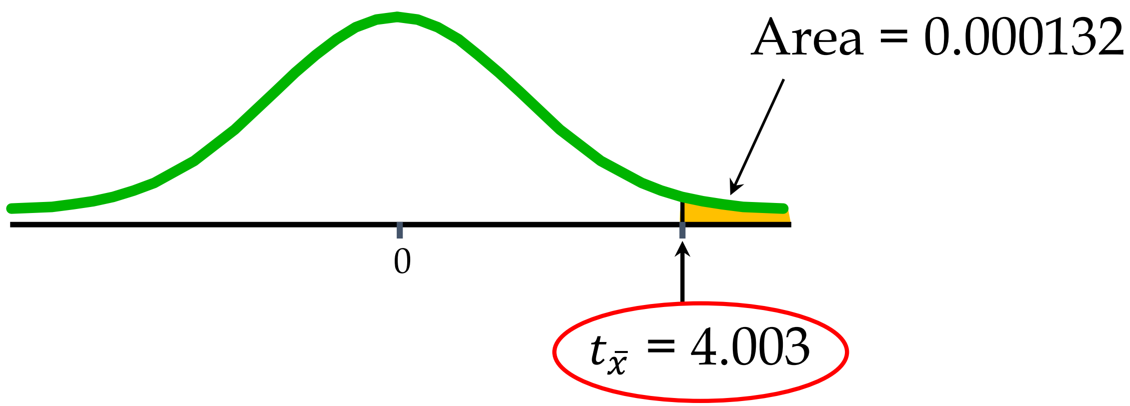


Hypothesis Test to Compare Two Means

Step 4: Calculate the *p*-value

Since this is a one-tail test (upper tail):

$$p\text{-value} = P(t > t_{\bar{x}}) = P(t > 4.003) = 0.000132$$



Step 5: Compare the *p*-value and the significance level

Because p-value = 0.000132 < α = 0.05 $\Rightarrow$ **we reject H_0**

Hypothesis Test to Compare Two Means

Conclusion

By rejecting the null hypothesis, we have support for the alternative hypothesis. Therefore, there is enough evidence to conclude that the average mpg performance of M cars is **greater** than J cars.

Some Notes: Excel

1. We can use the Data Analysis tool for the hypothesis tests in Excel *only if we have raw data*
   - I.e. if we have values of all observations which we include in the samples

2. It the question does not provide raw data, i.e. if the sample means and/or sample standard deviations are already precomputed and given to you in the question, then the Data Analysis tool cannot be used
   - In this case, one needs to perform the test “by-hand” using formulas on the lecture slides

Inferences About the Difference Between Two Population Proportions

(We'll talk about it if class time permits)

My Favorite Example: Smoking Bans

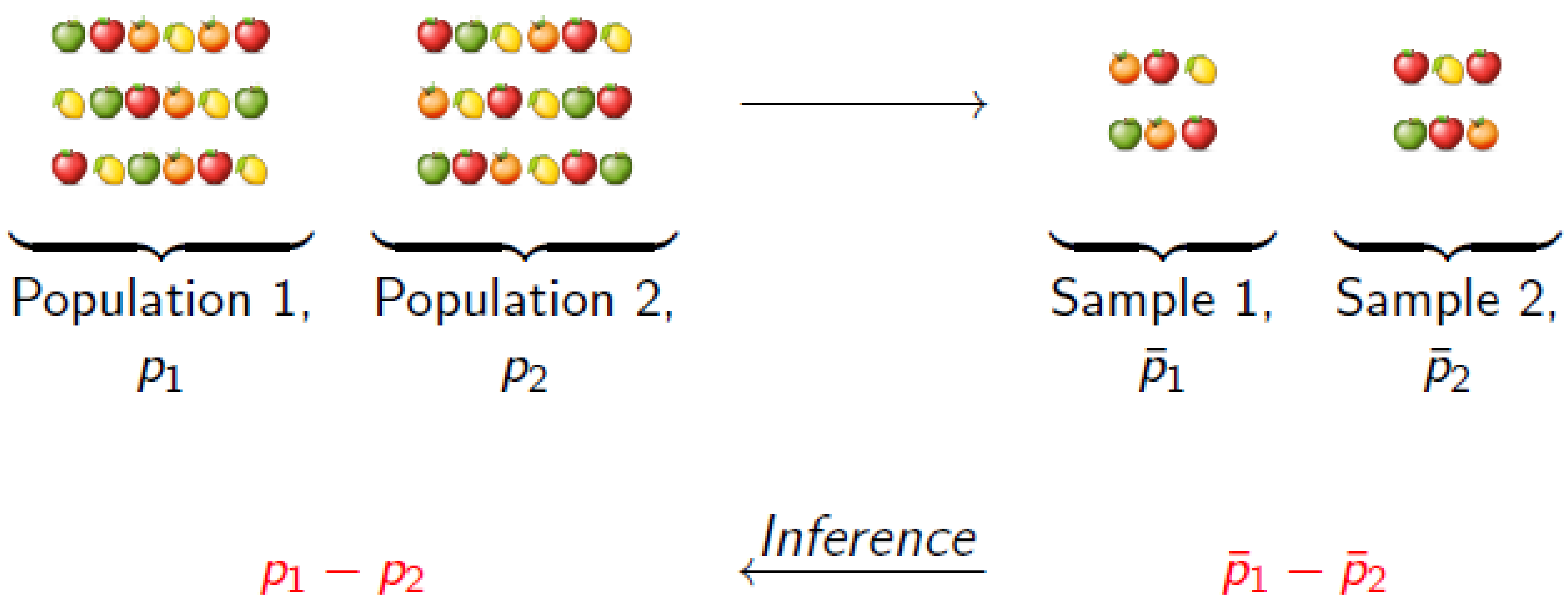
- Imagine you are a governor or a state official
- You decide to introduce a ban on smoking in the public places like restaurants, bus stops, etc.
  - $\downarrow$ indirect effect of second-hand smoking on non-smokers
  - You also hope that it will force some people to quit smoking!
- How do you know if the ban was effective to reduce smoking?

$\Rightarrow$ Compare proportions of smoking population before and after the ban was introduced!!!

More Examples...

- Does the Federal college subsidy ↑ college graduation rate?
- Does subsidizing childcare ↑ female labor force participation?
- Do the proportions of people experiencing difficulties in the relationships differ among unemployed and employed?
- Are vegetarians less likely to experience heart attack by the age of 70?
- ☺ Is there a difference between proportions of red M&M's and Skittles?

Our Goal



$\Rightarrow$ Our goal is to make an inference about the difference between population proportions $p_1 - p_2$ having information about sample proportions $\bar{p}_1 - \bar{p}_2$ for random samples from Population 1 and Population 2

Hypothesis Test to Compare Two Proportions

Three possible hypothesis statements:

Two-tail test:	One-tail test: (Upper tail)	One-tail test: (Lower tail)
$H_0: p_1 = p_2$ $H_1: p_1 \neq p_2$	$H_0: p_1 \leq p_2$ $H_1: p_1 > p_2$	$H_0: p_1 \geq p_2$ $H_1: p_1 < p_2$

These tests are equivalent to hypothesis tests about the **difference** between two population proportions

$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 \neq 0$	$H_0: p_1 - p_2 \leq 0$ $H_1: p_1 - p_2 > 0$	$H_0: p_1 - p_2 \geq 0$ $H_1: p_1 - p_2 < 0$

Sampling Distribution of $\bar{p}_1 - \bar{p}_2$

The **sampling distribution** of the difference between proportions is the result of subtracting the sampling distribution for the proportion of one population from the sampling distribution for the proportion of another population

- Sampling distribution for the difference in proportions is the distribution of $\bar{p}_1 - \bar{p}_2$

The **point estimate** of the difference between proportions of the Populations 1 and 2 is:

$$\bar{p}_1 - \bar{p}_2$$

Sampling Distribution of $\bar{p}_1 - \bar{p}_2$

To conduct a test about the difference between two population proportions, we need a sampling distribution of $\bar{p}_1 - \bar{p}_2$

- If $\bar{p}_1$ and $\bar{p}_2$ have the normal distribution

$\Rightarrow$ $\bar{p}_1 - \bar{p}_2$ has the normal distribution:

1. $n_1 p_1 \geq 5$ $\qquad$ $n_2 p_2 \geq 5$
2. $n_1(1 - p_1) \geq 5$ $\qquad$ $n_2(1 - p_2) \geq 5$

- Because p_1 and p_2 are not known $\Rightarrow$ use $\bar{p}_1$ and $\bar{p}_2$ to check these conditions

Sampling Distribution of $\bar{p}_1 - \bar{p}_2$

The Mean of the Sampling Distribution

$$\mu_{\bar{p}_1-\bar{p}_2} = \mu_{\bar{p}_1} - \mu_{\bar{p}_2} = p_1 - p_2$$

Standard Error of the Sampling Distribution

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

Sampling Distribution of $\bar{p}_1 - \bar{p}_2$ under H_0

Under the assumption that H_0 is true as equality:

$$p_1 = p_2 = p$$

$\Rightarrow$ Standard Error of the Sampling Distribution

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

Estimate of $\sigma_{\bar{p}_1 - \bar{p}_2}$

1. *Pooled* estimate for the overall proportion

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

   where x_1 and x_2 are the number of elements that have a characteristic of interest in samples from Populations 1 and 2

2. Estimated standard error of the sampling distribution, $\hat{\sigma}_{\bar{p}_1 - \bar{p}_2}$

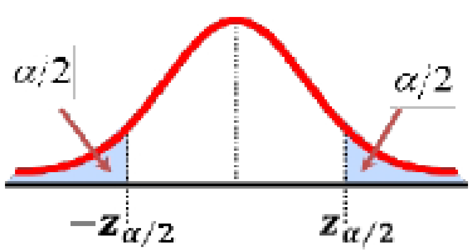
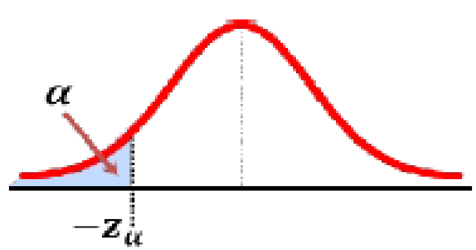
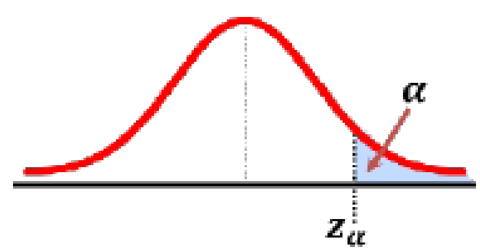
$$\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = \sqrt{\hat{p}\,(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Test Statistic: Difference Between Two Proportions

- *z*-test statistic

$$z_p = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)_{H_0}}{\hat{\sigma}_{\bar{p}_1 - \bar{p}_2}} = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)_{H_0}}{\sqrt{\hat{p}\,(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

- Use the standard normal distribution to find the critical value and the *p*-value. *For example*, **reject $\mathbf{H_0}$** if

Two tail	One (lower) tail	One (upper) tail
		
$\lvert z_p \rvert > \lvert z_{\alpha/2} \rvert$	$z_p < -z_\alpha$	$z_p > z_\alpha$

Excel Time: Exercise 10.7

Suppose the Bureau of Labor Statistics would like to investigate if the average retirement age for a worker in Japan is higher than the average retirement age for a worker in the U.S. A random sample of 30 retired U.S. workers had an average retirement age of 64.6 years. A random sample of 30 Japanese workers had an average retirement age of 67.5 years. Assume that the *population* standard deviation for the retirement age in the U.S. is 4 years and for Japan is 4.5 years.

Using α = 0.05, test if the average retirement age in Japan is *higher* than it is in the U.S. Use either:

- Use the critical value approach
- Use the *p*-value approach

Back <<

Excel Time: Exercise 10.48

McDonald's would like to compare the wait times its drive-through customers experience vs. the wait times its customers using the restaurants' inside counter experience. The data representing the wait time experience (in minutes) for randomly selected customers in the two types of groups can be found in the Excel file **wait_times.xlsx** (*Excel Files* → *Ch 10*). Assume the population variances of the wait times for both locations are **equal**.

Perform a hypothesis test using $\alpha = 0.05$ to determine if the average wait time differs for customers in these two locations.

Hint: you can use the Data Analysis tool for this problem

Back <<

Excel Time: Exercise 10.64 (Modified)

Citibank, which has a major credit card division, has instituted a promotion designed to encourage its customers to increase their credit card usage. To test the effectiveness of this promotion, Citibank monitored the monthly credit card balances of the customers who were informed and were unaware about the promotion. The results can be found in the Excel file **Citibank_ 2.xlsx** (*Excel Files → Ch 10*). Assume the population variances of the credit card balances are **unequal**.

Perform a hypothesis test using $\alpha = 0.05$ to determine if the average credit card balance is higher for the customers aware of the promotion.

Hint: be careful assigning Populations 1 and 2

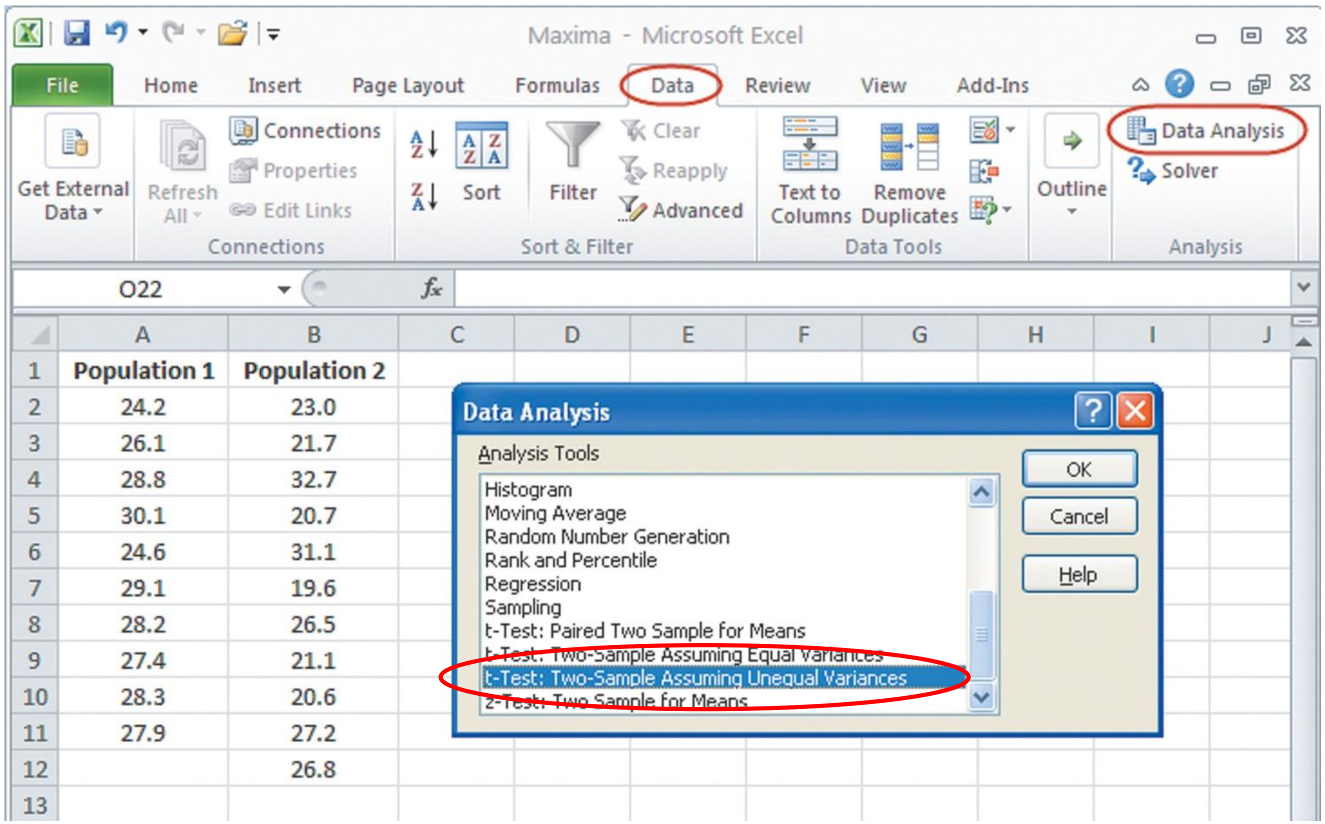
Back <<

Excel Time: Hypothesis Test to Compare Two Means

Screenshots are for the example with independent samples, σ_1 and σ_2 are unknown and assumed **unequal** (two-tail test)

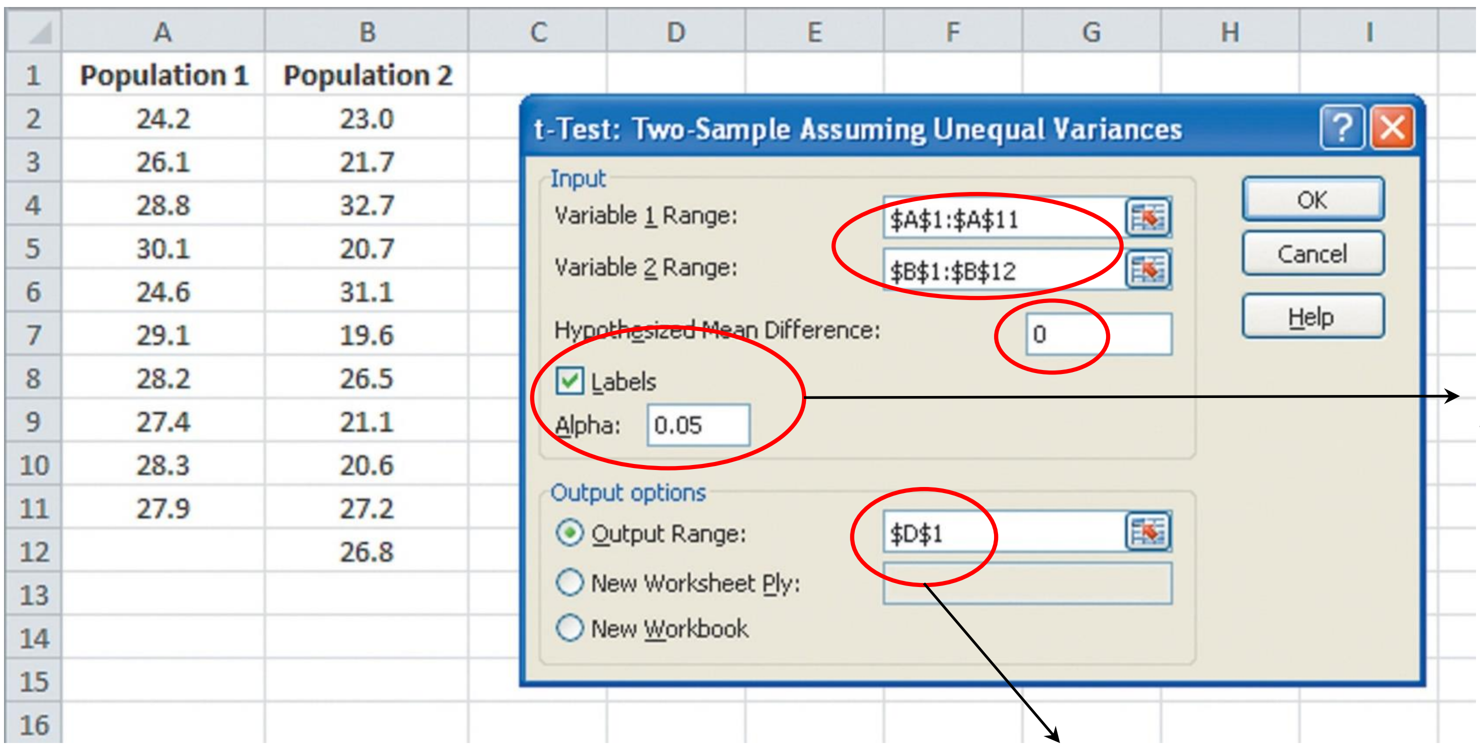
1. Enter the sample data from two populations in separate columns

2. Go to **Data > Data Analysis > t-test: Two-Sample Assuming Unequal Variances >** Click **OK**



Excel Time: Hypothesis Test to Compare Two Means

3. Fill in the ***t*-test: Two-Sample Assuming Unequal Variances** dialog box and click **OK**



Check Labels if data is selected with column titles

Upper left corner of the output

Excel Time: Hypothesis Test to Compare Two Means

Excel output:

p-value and the critical value for one tail test

	A	B	C	D	E	F
1	**Population 1**	**Population 2**		t-Test: Two-Sample Assuming Unequal Variances		
2	24.2	23.0				
3	26.1	21.7			*Population 1*	*Population 2*
4	28.8	32.7		Mean	27.47	24.63636364
5	30.1	20.7		Variance	3.729	20.24854545
6	24.6	31.1		Observations	10	11
7	29.1	19.6		Hypothesized Mean Difference	0	
8	28.2	26.5		df	14	
9	27.4	21.1		t Stat	1.904526432	
10	28.3	20.6		P(T<=t) one-tail	0.038797843	
11	27.9	27.2		t Critical one-tail	1.761310136	
12		26.8		P(T<=t) two-tail	0.077595685	
13				t Critical two-tail	2.144786688	
14						

p-value and the critical value for two tail test

Excel Time: Exercise 10.52 (Extra Practice)

The airline industry measures fuel efficiency by calculating how many miles one seat can travel, whether occupied or not, on one gallon of jet fuel. The data showing the fuel economy, in miles per seat, for 15 randomly selected flights on Delta and US Airways can be found in the Excel file **fuel_efficiency.xlsx** (*Excel Files → Ch 10*). Assume the population variances for the fuel efficiency for these two airlines are **unequal**.

Perform a hypothesis test using $\alpha = 0.05$ to determine if the fuel efficiency differs between these two airlines.

What assumptions need to be made to perform the hypothesis test?

Exercises that follow require performing *t*-tests manually

☹

Excel Time: Exercise 10.6 (Extra Practice)

The following data shows the average monthly utility bills for a random sample of households in Baltimore and for a random sample of households in Houston (the bills include phone, television, Internet, electricity, and natural gas).

	Baltimore	**Houston**
Sample Mean	\$390.44	\$359.52
Sample Size	33	36
Population Stand. Dev.	\$64	\$58

Using $\alpha = 0.01$, test the hypothesis to determine if there is a difference between the mean utility bills in these two cities.

Excel Time: Exercise 10.18 (Extra Practice)

The following data shows the average hourly wage rates for day-care centers from the Northeast and Midwest based on two random samples:

	Northeast	**Midwest**
Sample Mean	\$9.60	\$8.60
Sample Size	26	31
Sample Stand. Dev.	\$1.25	\$1.20

Perform a hypothesis test using α = 0.03 to determine if the average hourly wage for the day-care in the Northeast is **at least \$0.5 per hour higher** than the day-care in the Midwest. Assume that the population variances for wage rates in each region are equal.

Excel Time: Exercise (Extra Practice)

Are nursing salaries in Tampa, Florida, lower than those in Dallas, Texas? As reported by the *Tampa Tribune*, salary data show that staff nurses in Tampa earn less than staff nurses in Dallas. Suppose that in the follow up study of 40 staff nurses in Tampa and 50 staff nurses in Dallas you obtained the following results.

	Tampa	**Dallas**
Sample Mean	$56,100	$59,400
Sample Size	40	50
Sample Stand. Dev.	$6,000	$7,000

Using $\alpha = 0.02$, test the hypothesis to determine if the staff nurses in Tampa earn less than staff nurses in Dallas. Assume the population variances for salaries in these two locations are not equal.

Excel Time: Exercise 10.17 (Extra Practice)

The Transportation Security Administration would like to compare average amount of time it takes passengers to pass though airport security at Philadelphia vs Orlando during peak times. A random sample of 25 travelers in Philadelphia spent in line an average of 14.6 minutes with sample standard deviation 5.8 minutes. A random sample of 27 travelers in Orlando spent an average of 11.5 minutes with standard deviation 4.9 minutes.

Perform a hypothesis test using $\alpha = 0.05$ to determine if the average time to pass through security is Philadelphia is more than in Orlando. Assume that the population variances at these locations are equal.

Excel Time: Exercise 10.20 (Extra Practice)

Suppose a student organization at the University of Illinois collected data for a study involving class sizes from different departments. A random sample of 11 classes in the business department had an average size of 38.1 students with a sample standard deviation of 10.6 students. A random sample of 12 classes in the engineering department had an average size of 32.6 students with a sample standard deviation of 13.2 students.

Perform a hypothesis test using $\alpha = 0.02$ to determine if the class size differs between these departments. Assume that the population variances for the number of students per class are not equal.

Inferences About the Difference Between Two Population Proportions (*if time permits*)

Excel Time: Exercise

Winter visitors are extremely important to the economy of Southwest Florida. Hotel occupancy is an often-reported measure of visitor volume and visitor activity. Hotel occupancy data for February in two consecutive years are as follows:

	Current Year	**Previous Year**
Occupied Rooms	1470	1458
Total Rooms	1750	1800

Perform a hypothesis test using $\alpha = 0.05$ to determine if there has been an increase in the proportion of rooms occupied over the one-year period.

Excel Time: Exercise 10.38 (Extra Practice)

The major challenge for retailers like Best Buy is "show-rooming" which describes customer who browse through their stores and then make their purchase from an on-line competitor. In an effort to measure the volume of show-rooming, Best Buy observes the percentage of customers who leave their store with a purchase. A random sample of 170 female shoppers was selected and it is found that 73 left the store with a purchase, while a random sample of 150 male showed that 56 left with a purchase.

Using α = 0.1, test the hypothesis to determine if the percentage of women who leave the store with a purchase is higher than the percentage of men.

Excel Time: Exercise 10.49 (Extra Practice)

The on-time performances of flights in the airline industry are an important measurement that can attract customers. The following data show the number of late flights from random samples taken from United Airlines and Southwest.

	United	**Southwest**
Arrived late	130	119
Total flights	175	140

Perform a hypothesis test using $\alpha = 0.05$ to determine if the performances of these two airline are different.