

# Lecture 9: Hypothesis testing: Two-sample case

Ernesto F. L. Amaral

March 03–19, 2025

Advanced Methods of Social Research (SOCI 420)

[www.ernestoamaral.com](http://www.ernestoamaral.com)

Source: Healey, Joseph F. 2015. "Statistics: A Tool for Social Research." Stamford: Cengage Learning. 10th edition. Chapter 9 (pp. 216–246).



# Outline

- Identify and cite examples of situations in which the two-sample test of hypothesis is appropriate
- Explain the logic of hypothesis testing, as applied to the two-sample case
- Explain what an independent random sample is
- Perform a test of hypothesis for two sample means or two sample proportions, following the five-step model and correctly interpret the results
- List and explain each of the factors (especially sample size) that affect the probability of rejecting the null hypothesis
- Explain the differences between statistical significance and importance



# Basic logic

- We analyze a difference between two sample statistics
  - We compare means or proportions of two samples from specific sub-groups of the population
- This is the question under consideration
  - “Is the difference between the samples large enough to allow us to conclude (with a known probability of error) that the populations represented by the samples are different?”



# Null hypothesis

- The  $H_0$  indicates that the populations are the same
  - Assuming that the  $H_0$  is true, there is no difference between the parameters of the two populations
- On the other hand, we reject the  $H_0$  and say there is a difference between the populations
  - If the difference between the sample statistics is large enough
  - Or if the size of the estimated difference is unlikely



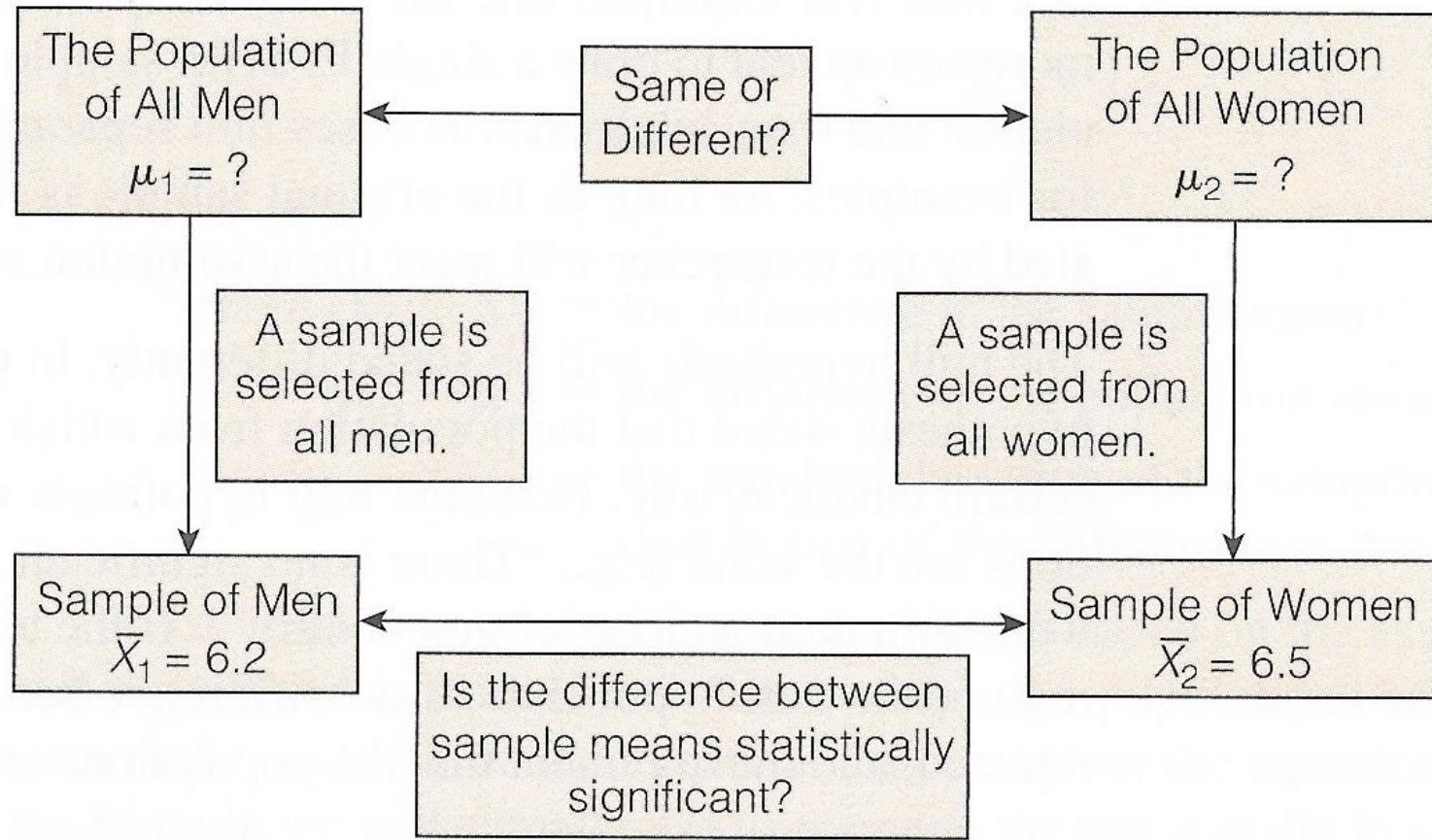
# $H_0$ , $\alpha$ , $Z$ score, $p$ -value

- The  $H_0$  is a statement of “no difference”
- The 0.05 level ( $\alpha$ ) will continue to be our indicator of a significant difference
- We change the sample statistics to a  $Z$  score
  - Place the  $Z(\textit{obtained})$  on the sampling distribution
- Estimate probability ( $p$ -value) above  $Z(\textit{obtained})$ 
  - $p$ -value is the probability of not rejecting the null hypothesis
  - Compare the  $p$ -value to the  $\alpha$
  - If  $p < \alpha$ , we reject  $H_0$
  - If  $p > \alpha$ , we do not reject  $H_0$





# Test of hypothesis for two sample means



# The five-step model

1. Make assumptions and meet test requirements
2. Define the null hypothesis ( $H_0$ )
3. Select the sampling distribution and establish the critical region
4. Compute the test statistic
5. Make a decision and interpret the test results

# Changes from one-sample case

- Step 1
  - In addition to samples selected according to EPSEM principles
  - Samples must be selected independently of each other: independent random sampling
- Step 2
  - Null hypothesis statement will state that the two populations are not different
- Step 3
  - Sampling distribution refers to difference between the sample statistics





# Two-sample test of means (large samples)

- Do men and women significantly differ on their support of gun control?
- For men (sample 1)
  - Mean = 6.2
  - Standard deviation = 1.3
  - Sample size = 324
- For women (sample 2)
  - Mean = 6.5
  - Standard deviation = 1.4
  - Sample size = 317

# Step 1: Assumptions, requirements

- Independent random sampling
  - The samples must be independent of each other
- Level of measurement is interval-ratio
  - Support of gun control is assessed with an interval-ratio level scale, so the mean is an appropriate statistic
- Sampling distribution is normal in shape
  - Total  $n \geq 100$  ( $n_1 + n_2 = 324 + 317 = 641$ )
  - Thus, the Central Limit Theorem applies and we can assume a standard normal distribution ( $Z$ )



# Step 2: Null hypothesis

- Null hypothesis,  $H_0: \mu_1 = \mu_2$ 
  - The null hypothesis asserts there is no difference between the populations
- Alternative hypothesis,  $H_1: \mu_1 \neq \mu_2$ 
  - The research hypothesis contradicts the  $H_0$  and asserts there is a difference between the populations

# Step 3: Distribution, critical region

- Sampling distribution
  - Standard normal distribution ( $Z$ )
- Significance level
  - Alpha ( $\alpha$ ) = 0.05 (two-tailed)
  - The decision to reject the null hypothesis has only a 0.05 probability of being incorrect
- $Z(\text{critical}) = \pm 1.96$ 
  - If the probability ( $p$ -value) is less than 0.05
  - $Z(\text{obtained})$  will be beyond  $Z(\text{critical})$



# Step 4: Test statistic

- Sample outcomes for support of gun control

Sample 1 (men)	Sample 2 (women)
$\bar{X}_1 = 6.2$	$\bar{X}_2 = 6.5$
$s_1 = 1.3$	$s_2 = 1.4$
$n_1 = 324$	$n_2 = 317$

- Pooled estimate of the standard error

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}} = \sqrt{\frac{(1.3)^2}{324 - 1} + \frac{(1.4)^2}{317 - 1}} = 0.107$$

- Obtained Z score

$$Z(\text{obtained}) = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{6.2 - 6.5}{0.107} = -2.80$$



# Step 5: Decision, interpret

- $Z(\text{obtained}) = -2.80$ 
  - This is beyond  $Z(\text{critical}) = \pm 1.96$
  - The obtained  $Z$  score falls in the critical region, so we **reject** the  $H_0$
  - Therefore, the  $H_0$  is false and must be rejected
- The difference between men's and women's support of gun control is statistically significant
  - The difference between the sample means is so large that we can conclude (at  $\alpha = 0.05$ ) that a difference exists between the populations represented by the samples





# Two-sample test of means (small samples)

- Do families that reside in the center-city have more children than families that reside in the suburbs?
- For suburbs (sample 1)
  - Mean = 2.37
  - Standard deviation = 0.63
  - Sample size = 42
- For center-city (sample 2)
  - Mean = 2.78
  - Standard deviation = 0.95
  - Sample size = 37



# Step 1: Assumptions, requirements

- Independent random sampling
  - The samples must be independent of each other
- Level of measurement is interval-ratio
  - Number of children can be treated as interval-ratio
- Population variances are equal
  - As long as the two samples are approximately the same size, we can make this assumption
- Sampling distribution is normal in shape
  - Because we have two small samples ( $n < 100$ ), we have to add the previous assumption in order to meet this assumption



# Step 2: Null hypothesis

- Null hypothesis,  $H_0: \mu_1 = \mu_2$ 
  - The null hypothesis asserts there is no difference between the populations
- Alternative hypothesis,  $H_1: \mu_1 < \mu_2$ 
  - The research hypothesis contradicts the  $H_0$  and asserts there is a difference between the populations



# Step 3: Distribution, critical region

- Sampling distribution
  - Student's  $t$  distribution
- Significance level
  - Alpha ( $\alpha$ ) = 0.05 (one-tailed)
- Degrees of freedom
  - $n_1 + n_2 - 2 = 42 + 37 - 2 = 77$
- Critical  $t$ 
  - $t(\text{critical}) = -1.671$



# Step 4: Test statistic

- Sample outcomes for number of children

Sample 1 (suburban)	Sample 2 (center-city)
$\bar{X}_1 = 2.37$	$\bar{X}_2 = 2.78$
$s_1 = 0.63$	$s_2 = 0.95$
$n_1 = 42$	$n_2 = 37$

- Pooled estimate of the standard error

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}} = \sqrt{\frac{(42)(0.63)^2 + (37)(0.95)^2}{42 + 37 - 2}} \sqrt{\frac{42 + 37}{(42)(37)}} = 0.18$$

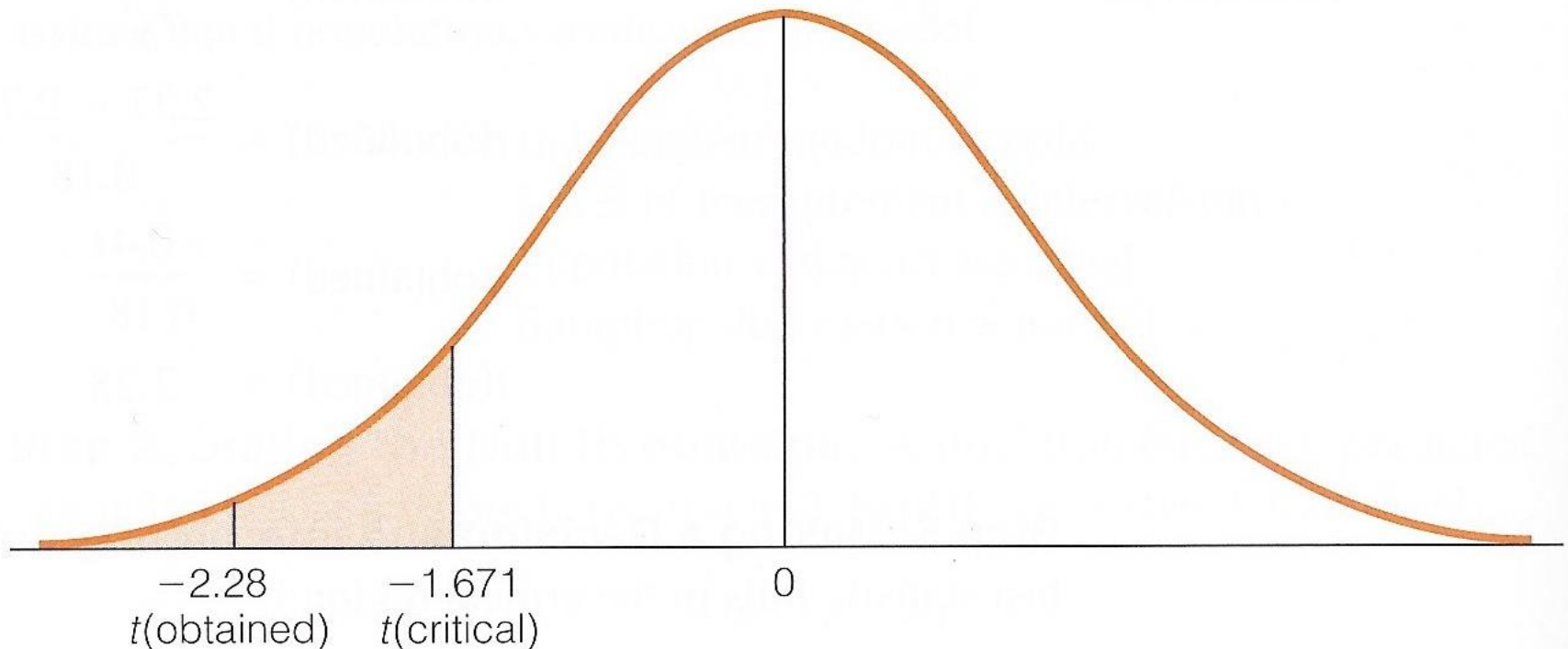
- Obtained  $t$

$$t(\text{obtained}) = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{2.37 - 2.78}{0.18} = -2.28$$



# $t(\text{obtained})$ & $t(\text{critical})$

- Sampling distribution with critical region and test statistic displayed





# Step 5: Decision, interpret

- $t(\text{obtained}) = -2.28$ 
  - This is beyond  $t(\text{critical}) = -1.671$
  - The obtained test statistic falls in the critical region, so we **reject** the  $H_0$
- The difference between the number of children in center-city families and the suburban families is statistically significant
  - The difference between the sample means is so large that we can conclude (at  $\alpha = 0.05$ ) that a difference exists between the populations represented by the samples



# Example from GSS: *t*-test

- We know the average income by sex from the 2016 GSS

```
. table sex, c(mean conrinc)
```

respondents sex	mean(conrinc)
male	<b>41583.52814</b>
female	<b>28353.34628</b>

- What causes the difference between male income of \$41,583.53 and female income of \$28,353.35?
- Real difference? Or difference due to random chance?



# Example from GSS: Result

- Men have an average income that is significantly higher than the female average income
  - The difference between male income (\$41,583.53) and female income (\$28,353.35) was large and unlikely to have occurred by random chance ( $p < 0.05$ ) in 2016

```
. ttest conrinc, by(sex)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
male	798	41583.53	1433.963	40507.87	38768.74	44398.32
female	834	28353.35	1049.496	30308.45	26293.38	30413.31
combined	1,632	34822.52	897.5571	36259.53	33062.03	36583
diff		13230.18	1765.955		9766.402	16693.96

```
diff = mean(male) - mean(female)                                t = 7.4918
Ho: diff = 0                                                    degrees of freedom = 1630
```

```
Ha: diff < 0
Pr(T < t) = 1.0000
```

```
Ha: diff != 0
Pr(|T| > |t|) = 0.0000
```

```
Ha: diff > 0
Pr(T > t) = 0.0000
```



# Edited table

**Table 1. Two-sample *t*-test of individual average income of the U.S. adult population by sex, 2004, 2010, and 2016**

<b>Sex</b>	<b>2004</b>	<b>2010</b>	<b>2016</b>
Male	45,741.48 (1,343.92)	37,864.34 (1,359.39)	41,583.53 (1,433.96)
Female	29,264.54 (972.15)	26,141.60 (972.97)	28,353.35 (1,049.50)
Difference	16,476.94*** (1,665.71)	11,722.74*** (1,643.94)	13,230.18*** (1,765.96)
Sample size	1,688	1,202	1,632

Note: Standard errors are reported in parentheses. \*Significant at  $p < 0.10$ ; \*\*Significant at  $p < 0.05$ ; \*\*\*Significant at  $p < 0.01$ .

Source: 2004, 2010, 2016 General Social Surveys.



# Two-sample test of proportions (large samples)

- Do Black and White senior citizens differ in their number of memberships in clubs and organizations?
  - Using the proportion of each group classified as having a “high” level of membership
- For Black senior citizens (sample 1)
  - Proportion = 0.34
  - Sample size = 83
- For White senior citizens (sample 2)
  - Proportion = 0.25
  - Sample size = 103



# Step 1: Assumptions, requirements

- Independent random sampling
  - The samples must be independent of each other
- Level of measurement is nominal
  - We have measured the proportion of each group classified as having a “high” level of membership
- Population variances are equal
  - As long as the two samples are approximately the same size, we can make this assumption
- Sampling distribution is normal in shape
  - Total  $n \geq 100$  ( $n_1 + n_2 = 83 + 103 = 186$ )
  - Thus, the Central Limit Theorem applies and we can assume a standard normal distribution





# Step 2: Null hypothesis

- Null hypothesis,  $H_0: P_{u1} = P_{u2}$ 
  - The null hypothesis asserts there is no difference between the populations
- Alternative hypothesis,  $H_1: P_{u1} \neq P_{u2}$ 
  - The research hypothesis contradicts the  $H_0$  and asserts there is a difference between the populations



# Step 3: Distribution, critical region

- Sampling distribution
  - Standard normal distribution ( $Z$ )
- Significance level
  - Alpha ( $\alpha$ ) = 0.05 (two-tailed)
  - The decision to reject the null hypothesis has only a 0.05 probability of being incorrect
- $Z(\text{critical}) = \pm 1.96$ 
  - If the probability ( $p$ -value) is less than 0.05
  - $Z(\text{obtained})$  will be beyond  $Z(\text{critical})$



# Step 4: Test statistic

- Sample outcomes for club memberships

Sample 1 (Black senior citizens)	Sample 2 (White senior citizens)
$P_{s1} = 0.34$	$P_{s2} = 0.25$
$n_1 = 83$	$n_2 = 103$

- Population proportion

$$P_u = \frac{n_1 P_{s1} + n_2 P_{s2}}{n_1 + n_2} = \frac{(83)(0.34) + (103)(0.25)}{83 + 103} = 0.29$$

- Pooled estimate of the standard error

$$\sigma_{p_1 - p_2} = \sqrt{P_u(1 - P_u)} \sqrt{\frac{n_1 + n_2}{n_1 n_2}} = \sqrt{(0.29)(0.71)} \sqrt{\frac{83 + 103}{(83)(103)}} = 0.07$$

- Obtained Z score

$$Z(\text{obtained}) = \frac{P_{s1} - P_{s2}}{\sigma_{p_1 - p_2}} = \frac{0.34 - 0.25}{0.07} = 1.29$$



# Step 5: Decision, interpret

- $Z(\text{obtained}) = 1.29$ 
  - This is below the  $Z(\text{critical}) = 1.96$
  - The obtained test statistic does not fall in the critical region, so we **do not reject** the  $H_0$
- The difference between the memberships of Black and White senior citizens is not significant
  - The difference between the sample means is small enough that we can conclude (at  $\alpha = 0.05$ ) that no difference exists between the populations represented by the samples

# Example from GSS: proportion

- We know the proportion of pro-immigrants by political party from the 2016 GSS

```
. table democrat, c(mean proimmig)
```

Political party	mean(proimmig)
Republicans	<b>.117096</b>
Democrats	<b>.4559471</b>

- What causes the difference between the percentage of Republicans who are pro-immigration (11.7%) and the percentage of Democrats who are pro-immigration (45.6%)?
  - Real difference? Or difference due to random chance?



# Example from GSS: Result

- Republicans are less pro-immigration than Democrats
  - The difference between the percentage of Republicans who are pro-immigration (11.7%) and the percentage of Democrats who are pro-immigration (45.6%) was large and unlikely to have occurred by random chance ( $p < 0.05$ ) in 2016

```
. prtest proimmig, by(democrat)
```

```
Two-sample test of proportions                Republicans: Number of obs =    427
                                             Democrats: Number of obs =    454
```

Variable	Mean	Std. Err.	z	P> z	[95% Conf. Interval]
Republicans	.117096	.0155602			.0865987 .1475934
Democrats	.4559471	.0233749			.4101332 .5017611
diff	-.3388511	.0280803			-.3938875 -.2838147
	under Ho:	.0306428	-11.06	0.000	

```
diff = prop(Republicans) - prop(Democrats)                z = -11.0581
Ho: diff = 0
```

```
Ha: diff < 0
Pr(Z < z) = 0.0000
```

```
Ha: diff != 0
Pr(|Z| > |z|) = 0.0000
```

```
Ha: diff > 0
Pr(Z > z) = 1.0000
```



# Edited table

**Table 2. Test of proportions of pro-immigrants among the U.S. adult population by political party, 2004, 2010, and 2016**

<b>Political Party</b>	<b>2004</b>	<b>2010</b>	<b>2016</b>
Republican	0.0911 (0.0124)	0.1429 (0.0193)	0.1171 (0.0156)
Democratic	0.2164 (0.0178)	0.2761 (0.0223)	0.4559 (0.0234)
Difference	-0.1253*** (0.0217)	-0.1333*** (0.0295)	-0.3389*** (0.0281)
Sample size	1,074	731	881

Note: Standard errors are reported in parentheses. \*Significant at  $p < 0.10$ ; \*\*Significant at  $p < 0.05$ ; \*\*\*Significant at  $p < 0.01$ .

Source: 2004, 2010, 2016 General Social Surveys.





# Statistical significance vs. importance (magnitude)

- As long as we work with random samples, we must conduct a test of significance
- Statistical significance is not the same thing as importance
  - Importance is also known as magnitude of the effect
- Differences that are otherwise trivial or uninteresting may be significant



# Influence of sample size

- When working with large samples, even small differences may be statistically significant
- The larger the sample size ( $n$ )
  - The greater the value of the test statistic
  - The more likely it will fall in the critical region and be declared statistically significant
- In general, when working with random samples, statistical significance is a necessary but not a sufficient condition for importance



# Sample size & test statistic

Test Statistics for Single-Sample Means Computed from Samples of Various Sizes ( $\bar{X} = 80$ ,  $\mu = 79$ ,  $s = 5$  throughout)

Sample Size ( $N$ )	Computing the Test Statistic	Test Statistic, $Z(\text{Obtained})$
50	$Z(\text{obtained}) = \frac{\bar{X} - \mu}{\sigma/\sqrt{N-1}} = \frac{80 - 79}{5/\sqrt{49}} = \frac{1}{0.71} =$	1.41
100	$Z(\text{obtained}) = \frac{\bar{X} - \mu}{\sigma/\sqrt{N-1}} = \frac{80 - 79}{5/\sqrt{99}} = \frac{1}{0.50} =$	2.00
500	$Z(\text{obtained}) = \frac{\bar{X} - \mu}{\sigma/\sqrt{N-1}} = \frac{80 - 79}{5/\sqrt{499}} = \frac{1}{0.22} =$	4.55
1000	$Z(\text{obtained}) = \frac{\bar{X} - \mu}{\sigma/\sqrt{N-1}} = \frac{80 - 79}{5/\sqrt{999}} = \frac{1}{0.16} =$	6.25
10,000	$Z(\text{obtained}) = \frac{\bar{X} - \mu}{\sigma/\sqrt{N-1}} = \frac{80 - 79}{5/\sqrt{9999}} = \frac{1}{0.05} =$	20.00



# Outcomes of hypothesis testing

- Result of a specific analysis could be
  - Statistically significant and
    - Important (large magnitude)
  - Statistically significant, but
    - Unimportant (small magnitude)
  - Not statistically significant, but
    - Important (large magnitude)
  - Not statistically significant and
    - Unimportant (small magnitude)



# Factors influencing the decision

1. The size of the observed difference
  - For larger differences, we are more likely to reject  $H_0$
2. The value of alpha
  - Usually the decision to reject the null hypothesis has only a 0.05 probability of being incorrect
  - The higher the alpha
    - The more likely we are to reject the  $H_0$
    - But we would have a higher chance of being incorrect
3. The use of one- vs. two-tailed tests
  - We are more likely to reject  $H_0$  with a one-tailed test
4. The size of the sample ( $n$ )
  - For larger samples, we are more likely to reject  $H_0$





TEXAS A&M  
UNIVERSITY.