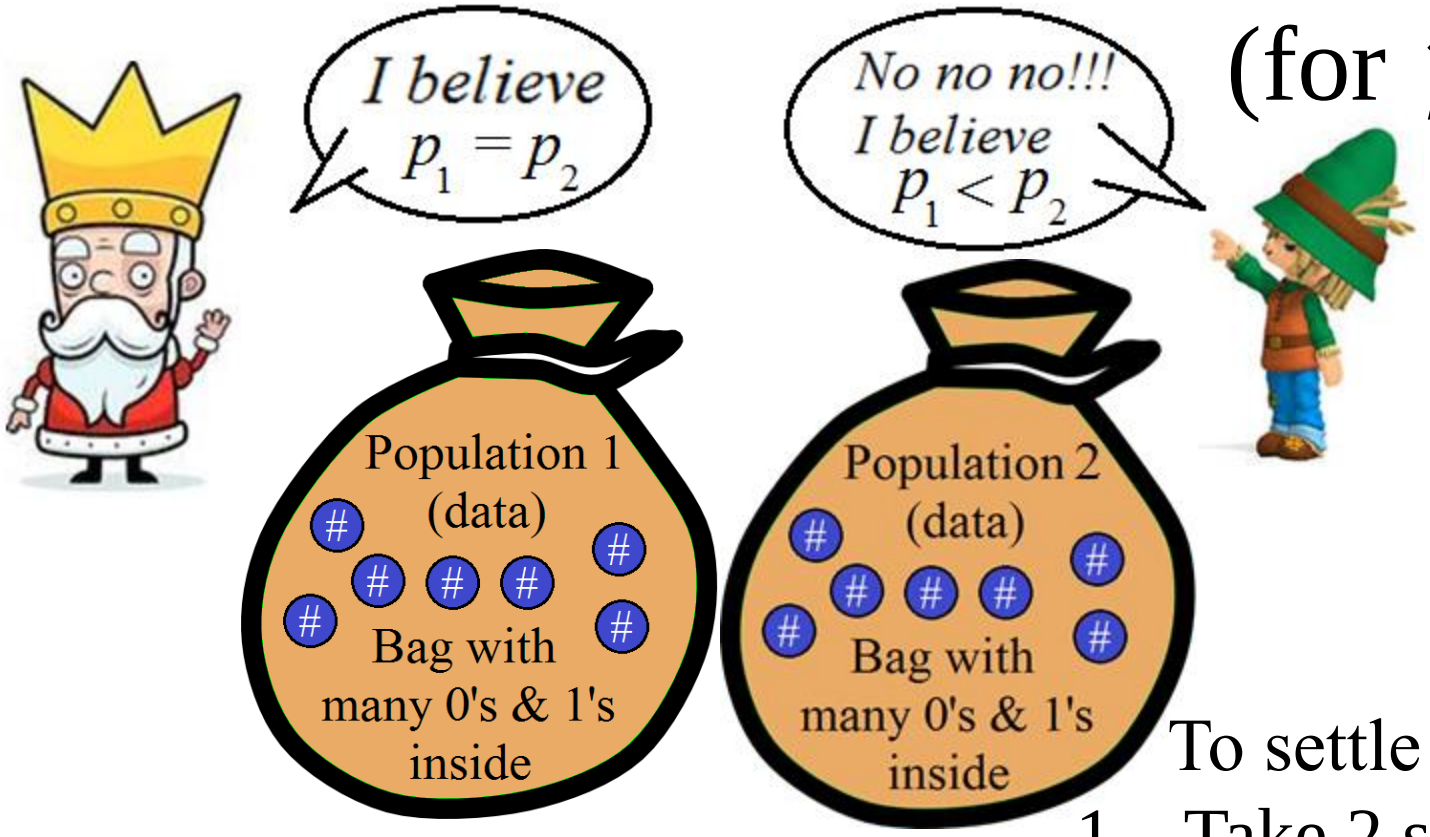
Sec. 11.1: Hypothesis Tests for 2 Population Proportions p_1 & p_2

Idea Behind 2 Population Hypothesis Tests (for p_1 & p_2)



To settle the dispute…

1. Take 2 samples (one from each population)
2. Calculate $\hat{p}_1$ & $\hat{p}_2$

If $\hat{p}_1 = \hat{p}_2$, believe the king

If $\hat{p}_1 > \hat{p}_2$, believe the king

If $\hat{p}_1 < \hat{p}_2$, it depends!

If $\hat{p}_1 < \hat{p}_2$ but $\hat{p}_1$ is close to $\hat{p}_2$, still believe the king

If $\hat{p}_1 < \hat{p}_2$ but $\hat{p}_1$ is far from $\hat{p}_2$, then believe the peasant

Independent Samples vs. Dependent Samples

Determine if the individuals in one sample were used to determine the individuals in the other sample

If no → Independent Samples

If yes → Dependent Sample

Independent Samples vs. Dependent Samples

Independent Samples

- If the individuals selected for one sample do not dictate which individuals are selected for the second samples
- The 2 samples are in no way related to each other
- In all problems in this class, if the 2 sample sizes are different, they must be independent
- Sometimes if the sample sizes are the same, the samples can still be independent

Independent Samples vs. Dependent Samples

Dependent (Matched-Pairs) Samples

- If the individuals selected for one sample dictate which individuals are selected for the second samples
- The 2 samples are very closely related to each other
- If the samples are dependent, they will have the same sample size (but not the other way around)
- In this section, the 2 samples will almost always be THE SAME SAMPLE
- In this section, the data will be given in a contingency table

		Answer to Question 1 (from population 1)	
		Yes	No
Answer to Question 2 (from population 2)	Yes	f_{11}	f_{12}
	No	f_{21}	f_{22}

Independent Samples vs. Dependent Samples Examples

Example: Are these samples independent or dependent?

Sample 1 = 35 randomly selected PCC students

Sample 2 = 31 randomly selected Rio Hondo students

Answer: Independent

Independent Samples vs. Dependent Samples Examples

Example: Are these samples independent or dependent?

Sample 1 = 18 randomly selected straight married American Men

Sample 2 = The 18 women that are the wives of the men from sample 1

Answer: Dependent

Independent Samples vs. Dependent Samples Examples

Example: Are these samples independent or dependent?

Sample 1 = 50 randomly selected Americans

Sample 2 = The same 50 randomly selected Americans

Answer: Dependent

Independent Samples vs. Dependent Samples Examples

Example: Are these samples independent or dependent?

Sample 1 = 20 randomly selected Americans

Sample 2 = 20 randomly selected Europeans

Answer: Independent

Independent Samples vs. Dependent Samples Examples

Example: Are these samples independent or dependent?

Sample 1 = The 900 Rio Hondo students who enrolled in math 30 this semester

Sample 2 = The 300 Rio Hondo students who enrolled in math 33 this semester

Answer: Independent

Independent Samples vs. Dependent Samples Examples

Example: Are these samples independent or dependent?

Sample 1 = 75 randomly selected kids between the ages of 10 and 18 that have exactly one sibling

Sample 2 = The siblings of the kids in sample 1

Answer: Dependent

Hypothesis Tests for 2 Population Proportions
p_1 & p_2 Formulas & Info (independent samples)

<u>Population 1</u>

Population parameter: p_1

<u>Sample 1</u>

Sample size: n_1

# of yes's in sample: x_1

Sample statistic: $\hat{p}_1 = \frac{x_1}{n_1}$

<u>Population 2</u>

Population parameter: p_2

<u>Sample 2</u>

Sample size: n_2

# of yes's in sample: x_2

Sample statistic: $\hat{p}_2 = \frac{x_2}{n_2}$

Pooled sample proportion: $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

$\hat{q}$: $\hat{q} = 1 - \hat{p}$

Hypothesis Tests for 2 Population Proportions p_1 & p_2 Formulas & Info (independent samples)

Quantities you are performing a hypothesis test for: p_1 & p_2

Probability distribution: z - distribution

Test statistic formula: $$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}\hat{q}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

(helps you determine how many steps apart $\hat{p}_1 - \hat{p}_2$ and $p_1 - p_2$ are)

Condition: $n_1\hat{p}_1\hat{q}_1 \geq 10$ and $n_2\hat{p}_2\hat{q}_2 \geq 10$

Ex 1 (Sec. 11.1 Hw #19, pg. 540): **Prevnar** The drug Prevnar is a vaccine meant to prevent certain types of bacterial meningitis. It is typically administered to infants starting around 2 months of age. In randomized, double-blind clinical trials of Prevnar, infants were randomly divided into two groups. Subjects in group 1 received Prevnar, while subjects in group 2 received a control vaccine. After the first dose, 107 of 710 subjects in the experimental group (group 1) experienced fever as a side effect. After the first dose, 67 of 611 of the subjects in the control group (group 2) experienced fever as a side effect. Does the evidence suggest that a higher proportion of subjects in group 1 experienced fever as a side effect than subjects in group 2 at the $\alpha = 0.05$ level of significance?

a) Use the P-value method

b) Use the rejection region method

Ex 2: Peter the gambler feels like his local casino is cheating at the game of blackjack. To test his claim he goes to the casino and plays 500 hands of blackjack. Of the 500 hands he played at the casino, he won 211 of them. Then he goes home (where the deck of cards is fair) and plays another 500 hands of blackjack. Of the 500 hands he played at home, he won 239 of them. Perform the appropriate hypothesis test at the $\alpha = 0.01$ level to determine if the casino is cheating at blackjack (i.e. if the probability of winning a hand of blackjack at the casino is less than the probability of winning a hand of blackjack at home with a fair deck).

a) Use the P-value method

b) Use the rejection region method

Ex 3: In order to figure out if PCC or Rio Hondo has better students, Greg the stats teacher performed an experiment in Fall 2011. He taught 2 stats classes at each school and gave identical exams, quizzes, and homework in all 4 classes. Of the 87 students Greg started with at PCC, 51 passed the class. Of the 75 students Greg started with at Rio Hondo, 33 passed the class. Perform the appropriate test at the $\alpha = 0.04$ significance level to test the claim that the percentage of students who can pass Greg's stats class at PCC is the same as the percentage of students who can pass Greg's stats class at Rio Hondo.

a) Use the P-value method

b) Use the rejection region method

Hypothesis Tests for 2 Population Proportions
p_1 & p_2 Formulas & Info (dependent samples)

Population 1	Population 2
Population parameter: p_1	Population parameter: p_2
Sample 1	**Sample 2**
Sample size: n_1	Sample size: n_2
# of yes's in sample: x_1	# of yes's in sample: x_2
Sample statistic: $\hat{p}_1 = \frac{x_1}{n_1}$	Sample statistic: $\hat{p}_2 = \frac{x_2}{n_2}$

What's the difference?

H_1 will always be $H_1: p_1 \neq p_2$ $\qquad$ $n_1 = n_2$

Usually: Population 1 = Population 2 & Sample 1 = Sample 2

Hypothesis Tests for 2 Population Proportions
p_1 & p_2 Formulas & Info (dependent samples)

Quantities you are performing a hypothesis test for: p_1 & p_2

Probability distribution: z - distribution

Contingency table:

Answer to Question 2 from population 2	Answer to Question 1 (from population 1): Yes	Answer to Question 1 (from population 1): No
Yes	f_{11}	f_{12}
No	f_{21}	f_{22}

Test statistic formula (McNemar's Test):

$$z = \frac{|f_{12} - f_{21}| - 1}{\sqrt{f_{12} + f_{21}}}$$

Condition: $f_{12} + f_{21} \geq 10$

Ex 4 (Sec. 11.4 Book Example 4 pg. 537): A recent General Social Survey asked the following two questions of a random sample of 1492 adult Americans under the hypothetical scenario that the government suspected that a terrorist act was about to happen:

- Do you believe the authorities should have the right to tap people's telephone conversations?
- Do you believe the authorities should have the right to stop and search people on the street at random?

Ex 4 (Sec. 11.4 Book Example 4 pg. 537):

The results of the survey are in the following table:

		Random Stop	
		Agree	Disagree
Tap Phone	Agree	494	335
	Disagree	126	537

Do the proportions who agree with each scenario differ significantly? Use the $\alpha = 0.05$ level of significance.

a) Use the P-value method

b) Use the rejection region method