We compare two sample proportions to infer about the populations from which they came. We will test $H_0: p_1 = p_2$.

Statistics
Class Notes
Inference about Two Population Proportions: Independent Sampling (Section 11.1)

In March 2003, The Pew Research Group surveyed 1508 adult Americans asking, "Do you believe the US made the right or the wrong decision to use military force in Iraq?" Seventy-two percent of the respondents said the US made the right decision. Several years later, in August 2010, they asked the same question of (another) 1508 adult Americans but only 41% said the US made the right decision. Is this evidence that the percent of adult Americans who agreed with the military decision really decreased over the years? Or maybe random just happens? Perhaps the true percentage of adult Americans who think the US made the right decision is the same as in 2003 but the sample just happened to get a smaller number in the second poll. How do we know?

Statistics, baby!

Our first task is to identify if the samples are dependent or independent. If they are dependent, we use different statistics we will *not* cover in this course (section 12.3 in the book). If they are independent, then all is well and, in this section, we see how to answer our question.

Determining if Two Samples are Independent:
Recall, independent events in probability theory are ones such that the occurrence of one event does *not* affect the occurrence of the other event. We have a similar definition needed here.

Definition: A sampling method is **independent** when an individual selected for one sample does *not* dictate which individual is to be in a second sample.

A sampling method is **dependent** when an individual selected to be in one sample is used to determine the individual in the second sample. Dependent samples are often referred to as **matched-pairs** samples. It is possible for an individual to be matched against him or herself.

expl 1: Determine whether the sampling method is independent or dependent. Also, determine whether the response variable is qualitative or quantitative.

Again, different statistics are employed to answer this question if the sampling method is dependent.

a.) Are products purchased on Amazon less expensive than those purchased online at Walmart? To answer this question, researchers randomly identified 20 products sold at both stores and determined the selling price at Amazon and the online Walmart store to determine if there was a significant difference in the price of the goods.

This is dependent because an item picked at Walmart will then be picked at Amazon. Variable is quantitative.

b.) Consider the surveys described above for the opinions of adult Americans concerning the Iraq war. These are totally different samples with different people in each – so it's independent. Variable is a yes/no question, so qualitative.

Qualitative versus Quantitative Variables:
If the variable in question is qualitative, we are dealing with proportions and the statistics here will apply. If the variable is quantitative, we will compare the means of the data, which is covered in later sections.

expl 2: In clinical trials of Nasonex (a nasal spray used for congestion), 3774 adult and adolescent allergy patients (12 years and older) were randomly divided into two groups. The patients in group 1 (experimental group) received 200 μg (micrograms) of Nasonex, while the patients in group 2 (control group) received a placebo. Of the 2103 patients in the experimental group, 547 reported headaches as a side effect. Of the 1671 patients in the control group, 368 reported headaches as a side effect. Is there evidence to conclude that the proportion of Nasonex users who experienced headaches as a side effect is greater than the proportion in the control group at the $\alpha = 0.05$ level of significance?

Nasonex headaches: 26.0%
Placebo headaches: 22.0%

(Nasonex) $p_N \overset{?}{>} p_P$ (placebo)

Recall: Sampling Distribution of the Sample Proportion $\hat{p} = \frac{x}{n}$**:**

For a simple random sample of size n with a population proportion p,

1. The shape of the sampling distribution of $\hat{p}$ is approximately normal *if* $np(1-p) \geq 10$.
2. The mean of the sampling distribution of $\hat{p}$ is $\mu_{\hat{p}} = p$.
3. The standard deviation of the sampling distribution of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
4. We also must check that $n \leq .05N$.

Two samples with different x and n values.

We extend this to accommodate two proportions $\hat{p}_1 = \frac{x_1}{n_1}$ and $\hat{p}_2 = \frac{x_2}{n_2}$.

Sampling Distribution of $\hat{p}_1 - \hat{p}_2$**:**

For two simple random samples of sizes n_1 and n_2 with population proportions p_1 and p_2,

1. The sampling distribution of $\hat{p}_1 - \hat{p}_2$ is approximately normal *if* $n_i\hat{p}_i(1-\hat{p}_i) \geq 10$.
2. The mean of the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$.
3. The standard deviation of the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is

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

4. We also must check that $n_i \leq .05N_i$.

for both samples.

Here, i is in place of 1 or 2.

Check for $\hat{p}_1$ and again for $\hat{p}_2$.

Our Test Statistic:

Similar to our test statistic earlier, our test statistic starts off as $z_0 = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$.

We now think of our hypothesis (which is H_0: $p_1 = p_2$ or equivalently, $p_1 - p_2 = 0$). This implies that both p_1 and p_2 are equal to p, the population proportion. Hence, this test statistic becomes

$$z_0 = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}} = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\dfrac{p(1-p)}{n_1} + \dfrac{p(1-p)}{n_2}}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p)}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

Hey wait, this assumes we know p??

Our best estimate of p to be used in this test statistic is $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$. This is called a **pooled estimate of p**. Finally, **the test statistic we will use is**, in fact, $z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$.

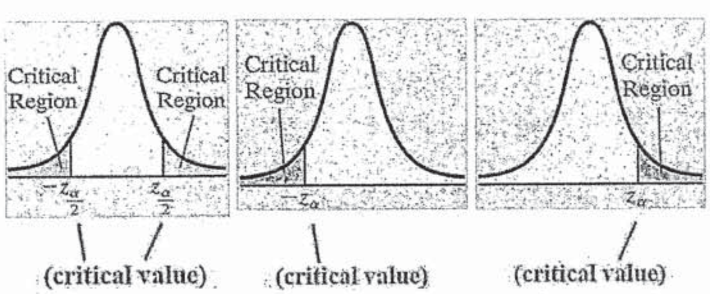
Recall: Steps 1 and 2 of Hypothesis Testing: We determine the null and alternative hypotheses and the desired value of the level of significance, α, depending on the seriousness of making a Type I error.

Again, the null hypothesis (that the two samples' proportions are the same) is H_0: $p_1 = p_2$ or equivalently, $p_1 - p_2 = 0$. We will test to see if we can find evidence that the null hypothesis is *not* true. Once again, we have three possibilities, the two-tailed, the left-tailed, and the right-tailed tests.

Two-tailed test	**Left-tailed test**	**Right-tailed test**
H_0: $p_1 = p_2$ H_1: $p_1 \neq p_2$	H_0: $p_1 = p_2$ H_1: $p_1 < p_2$	H_0: $p_1 = p_2$ H_1: $p_1 > p_2$

Recall these are *population* proportions.

(Classical Approach) Step 3: Calculate the test statistic. Use Table V to determine the critical value, based on α.



Here, the shaded area is the level of significance, α.

(Classical Approach) Step 4: Compare the critical value with the test statistic. The most common critical values are below. If z_0 is in the critical region, you *reject* the null hypothesis.

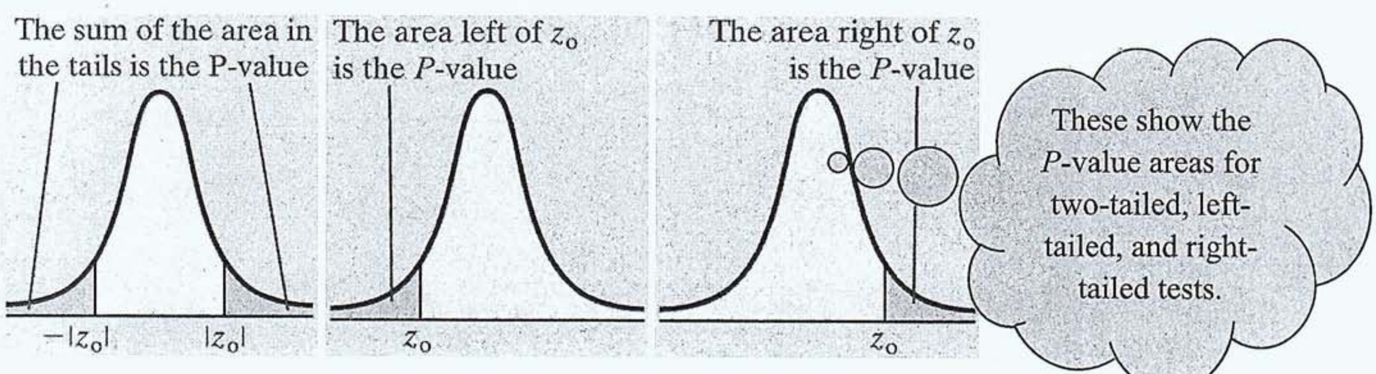
Hypothesis Testing Critical Values			
Level of significance, α	**Left-tailed**	**Right-tailed**	**Two-tailed**
0.10 (10%)	−1.28	1.28	±1.645
0.05 (5%)	−1.645	1.645	±1.96
0.025 (2.5%)	−1.96	1.96	±2.24
0.01 (1%)	−2.33	2.33	±2.575

Step 5: State the conclusion. Either we say "There is sufficient evidence to conclude that the population parameters are as stated in the alternative hypothesis." (This is a *rejection* of the null hypothesis.) or "There is *not* sufficient evidence to conclude that the population parameters are as stated in the alternative hypothesis." (Here, we did *not* reject the null hypothesis.)

Alternative Steps 3 and 4 Using P-Value Approach:

(P-Value Approach) Step 3: Compute the **test statistic** $z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ using the **pooled estimate of p** which is $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$. Use Table V to determine the area to the left of z_0.

The P-value is then figured, taking these pictures into account.



You could use technology to find the P-value instead of looking it up on the table.

(P-Value Approach) Step 4: If the P-value $< \alpha$, reject the null hypothesis.

Step 5: State the conclusion.

$\hat{p}_N$ = prop of Nasonex users with head aches
$\hat{p}_P$ = " " Placebo " " " "

Let's take this procedure out for a run by looking at the data in example 2.

expl 2 (again): In clinical trials of Nasonex (a nasal spray used for congestion), 3774 adult and adolescent allergy patients (12 years and older) were randomly divided into two groups. The patients in group 1 (experimental group) received 200 μg (micrograms) of Nasonex, while the patients in group 2 (control group) received a placebo. Of the 2103 patients in the experimental group, 547 reported headaches as a side effect. Of the 1671 patients in the control group, 368 reported headaches as a side effect. Is there evidence to conclude that the proportion of Nasonex users who experienced headaches as a side effect is *greater than* the proportion in the control group at the $\alpha = 0.05$ level of significance?

Nasonex headaches: 26.0%
Placebo headaches: 22.0%

Do *not* neglect the requirements on page 2. If they are *not* met, the statistics that follow are useless.

We are given that N is at least 10 million.

$H_0: p_N = p_P$

$H_1: p_N > p_P$

(right-tailed test)

$$\hat{p}_N = \frac{547}{2103} \approx 0.260$$

$$\hat{p}_P = \frac{368}{1671} \approx 0.220$$

Samples are indp't ✓

Is $\hat{p}_N - \hat{p}_P$ normal?

Check if $n_i \hat{p}_i (1-\hat{p}_i) \geq 10$

for $\hat{p}_N$: $2103(0.260)(0.740) \approx 404 \geq 10$ ✓

for $\hat{p}_P$: $1671(0.220)(0.780) \approx 287 \geq 10$ ✓

Is $n_i \leq 0.05 N_i$?

Since N is at least 10,000,000, we'll say yes.

Pooled estimate for $p = p_1 = p_2$ (pop. means)

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{547 + 368}{2103 + 1671} \approx 0.242$$

test statistic:

$$z_0 = \frac{0.260 - 0.220}{\sqrt{0.242(0.758)}\sqrt{\frac{1}{2103} + \frac{1}{1671}}} \approx 2.85$$

Compare to critical value (looked up on pg 4 as 1.645)

2.85 reject

0 1.645

we do find evidence that $p_N > p_P$ — that the percent of people on Nasonex with headaches is greater than that percent on placebo.

Practical versus Statistical Significance:
We did, in fact, find the difference between the two percentages to be statistically significant. But would you (or most people, do you think) not take the allergy-relieving medicine if it increases your risk of headaches from 22% to 26%? Is that enough of a difference to deter people? If *not*, the result has no *practical* significance.

Many statistically significant results can be produced simply by increasing the sample size. Practical significance should always be considered.

(1 – α)·100% Confidence Intervals for the Difference of Two Population Proportions:
We will make confidence intervals to estimate $p_1 - p_2$. What would it mean if this interval contained zero?

Once again, we need to verify that the samples are independent and from simple random sampling or a randomized experiment. Also, we verify that $n_i\hat{p}_i(1-\hat{p}_i) \geq 10$ and $n_i \leq .05N_i$.

The CI bounds are as follows.

Lower bound: $(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Upper bound: $(\hat{p}_1 - \hat{p}_2) + z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

This is $\hat{p}_1 - \hat{p}_2$ plus or minus margin of error.

Notice we are *not* using a pooled estimate of p. The reason being we are *not* making the assumption we did when we tested a hypothesis.

The most common critical values ($z_{\alpha/2}$) are reproduced here.

Critical Values for Confidence Intervals	
Level of Confidence $(1-\alpha)\cdot 100\%$	**Critical Value**
90%	1.645
95%	1.96
99%	2.575

Let's try this out! We'll do these problems by hand and then have a glance at technology.

$\hat{p}_1$ = prop who survived on OPDIVO
$\hat{p}_2$ = " " " " dacarbazine

expl 3: OPDIVO is a new drug used to treat metastatic melanoma (a kind of skin cancer). The drug dacarbazine is the current treatment for such cancers. A randomized drug trial compared survival rates (for 12 months) for two groups, one receiving OPDIVO (3 mg/kg by intravenous infusion every two weeks) and one receiving dacarbazine (1000 mg/m^2 by intravenous infusion every three weeks). Of the 210 patients who received OPDIVO, 45 survived 12 months. Of the 208 patients who received dacarbazine, 22 survived 12 months. Construct and interpret a 95% confidence interval for the difference of population proportions (for OPDIVO versus dacarbazine). Are samples indpt? ✓ (different people)

Is $n_1 \hat{p}_1(1-\hat{p}_1) \geq 10$? $210(45/210)(1-45/210) \approx 35 \geq 10$ ✓

Is $n_2 \hat{p}_2(1-\hat{p}_2) \geq 10$? $208(22/208)(1-22/208) \approx 20 \geq 10$ ✓

Is $n_i \leq 0.05 N_i$? – assume N is more than 20 times sample size ✓

margin of error: $E = z_{\alpha/2}\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

$= 1.96\sqrt{\dfrac{0.214(0.786)}{210} + \dfrac{0.106(0.894)}{208}} \approx 0.069$

95% CI: $\hat{p}_1 - \hat{p}_2 \pm E \rightarrow (0.214 - 0.106) \pm 0.069 \Rightarrow$ interval: $(0.039, 0.177)$

We are 95% confident that the difference between survival rates of OPDIVO versus dacabazine is between 3.9% and 17.7%

When the Interval Contains Zero:
If we make a confidence interval to estimate $p_1 - p_2$, what would it mean if this interval contained zero? (This is equivalent to *not* rejecting the null hypothesis in a test.) Here is an example.

expl 4: The Harris Poll asked males and females if they had at least one tattoo. Of the 1205 males, 181 said yes ($\hat{p}_m \approx 0.150$). Of the 1097 females, 143 said yes ($\hat{p}_f \approx 0.130$). Forming a 95% confidence interval requires a critical value of $z_{\alpha/2} = 1.96$. Hence, the margin of error is calculated as 0.028. (It is left for you if you'd like to check that.) Form a 95% confidence interval. Notice the interval includes zero; what meaning can we attribute to that?

$(0.150 - 0.130) \pm 0.028$

95% CI: $(-0.8\%, 4.8\%)$

This means we do not have evidence that $p_M > p_f$, or vice versa for that matter.

We are saying that the difference $p_m - p_f$ could be negative or it could be positive.

Instructions for TI Calculators: Hypothesis Test:

1. Press **STAT** > **TESTS** and select **6:2-PropZTest...** from list.

2. Enter the values for x_1, n_1, x_2, and n_2. Select the appropriate alternative hypothesis. Be sure the p_1 and p_2 are in the order you want.

3. Highlight **Calculate** and press **ENTER**. **Draw** should draw the z-curve with the P-value shaded.

4. It outputs the test statistic as **z**, the P-value as **p**, and the sample proportions and sizes as well as the pooled estimate of the population proportion, presumably to check.

You may notice a slight difference between the output from a calculator versus doing it by hand.

Instructions for TI Calculators: Confidence Interval:

1. Press **STAT** > **TESTS** and select **B:2-PropZInt...** from list.

2. Enter the values for x_1, n_1, x_2, and n_2. Enter the appropriate confidence level, in decimal form.

3. Highlight **Calculate** and press **ENTER**.

4. It outputs the interval in interval notation and the sample proportions and sizes, presumably to check.

You may notice a slight difference between the output from a calculator versus doing it by hand.

Instructions for StatCrunch: Hypothesis Test or Confidence Interval:

1. If you have raw data, enter it and name the columns.

2. Select **Stat** > **Proportion Stats** > **Two Sample**. Select **With Data** or **With Summary**.

3. If you chose **With Summary**, enter the **number of successes** and **observations** (sample size) for each sample. You can **Tab** (on keyboard) between the inputs. If you chose **With Data**, select the column that has the data and denote which is to be considered a **Success** for each sample. Be sure you have denoted these in the correct order to test H_0: $p_1 - p_2 = 0$.

4. Choose either **Hypothesis test for $p_1 - p_2$** or **Confidence interval for $p_1 - p_2$**. For the hypothesis test, enter 0 for the value of $p_1 - p_2$ in the null hypothesis and choose the desired alternative hypothesis. For the confidence interval, enter the confidence level, in decimal form.

5. Along with summary information, it will output the test statistic (**Z-Stat**) and ***P*-value** for a hypothesis test. For a confidence interval, it outputs an **L. Limit** and **U. Limit** which are the lower and upper limits of the interval.

StatCrunch uses $p_1 - p_2 = 0$, etc. instead of $p_1 = p_2$, etc.