

# Inference for Proportions

## IPS Chapter 8

---

- 8.1: Inference for a Single Proportion
- 8.2: Comparing Two Proportions

---

Inference for Proportions

# 8.1 Inference for a Single Proportion

---

# Objectives

## 8.1 Inference for a single proportion

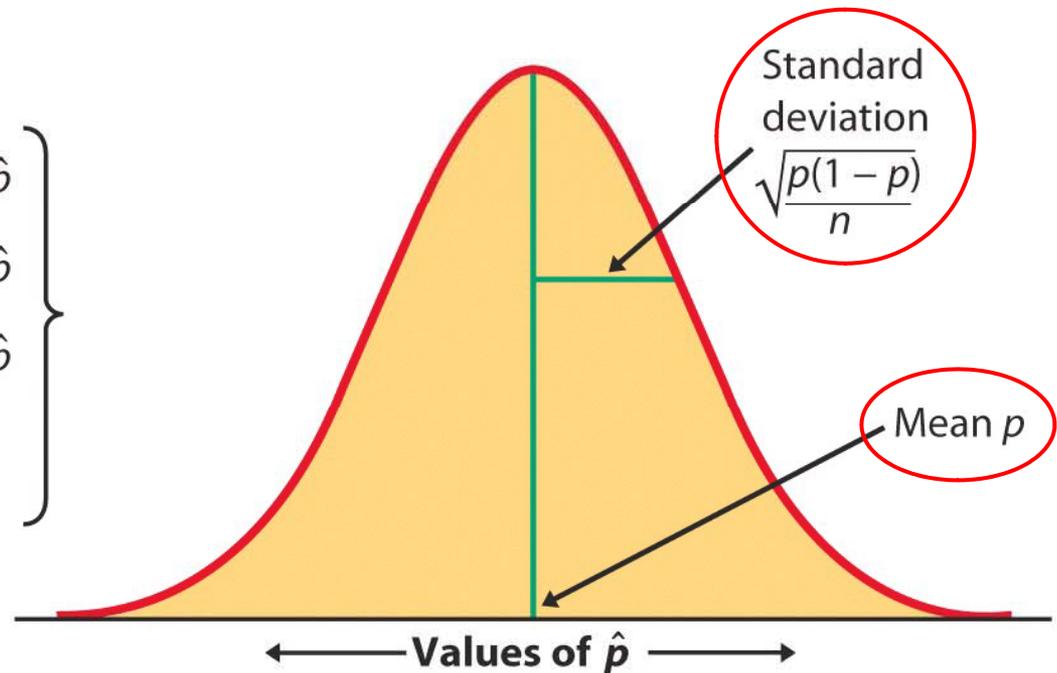
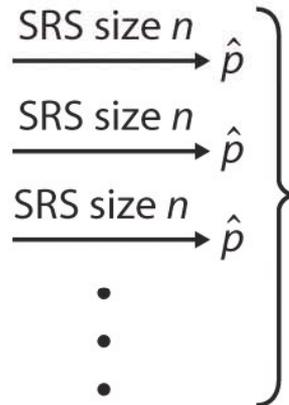
- Large-sample confidence interval for  $p$
- “Plus four” confidence interval for  $p$
- Significance test for a single proportion
- Choosing a sample size

# Sampling distribution of sample proportion

The sampling distribution of a sample proportion  $\hat{p}$  is approximately normal (normal approximation of a binomial distribution) when the sample size is large enough.



Population  
proportion  $p$   
of successes



# Conditions for inference on $p$

## Assumptions:

1. The data used for the estimate are an SRS from the population studied.
2. The population is at least 10 times as large as the sample used for inference. This ensures that the standard deviation of  $\hat{p}$  is close to  $\sqrt{p(1-p)/n}$
3. The sample size  $n$  is large enough that the sampling distribution can be approximated with a normal distribution. How large a sample size is required depends in part on the value of  $p$  and the test conducted. Otherwise, rely on the binomial distribution.

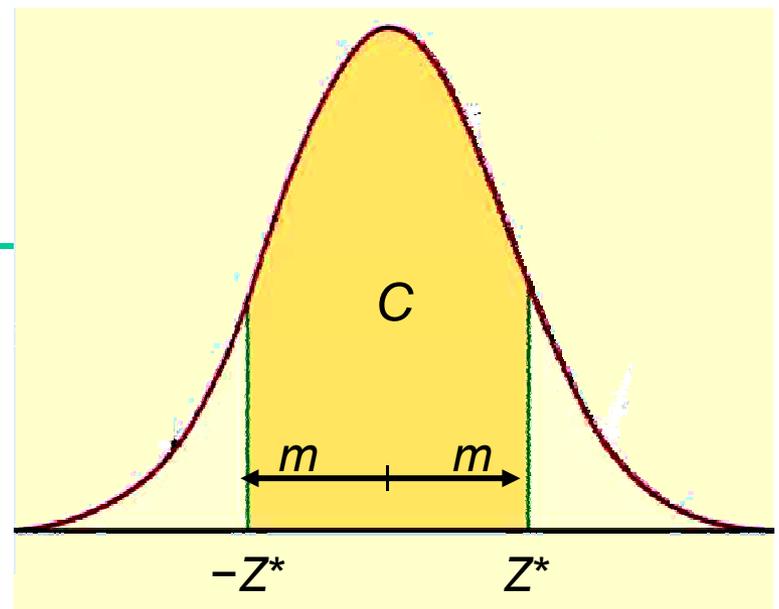
# Large-sample confidence interval for $p$

Confidence intervals contain the population proportion  $p$  in  $C\%$  of samples. For an SRS of size  $n$  drawn from a large population, and with sample proportion  $\hat{p}$  calculated from the data, an **approximate level  $C$  confidence interval** for  $p$  is:

$\hat{p} \pm m$ ,  $m$  is the margin of error

$$m = z^* SE = z^* \sqrt{\hat{p}(1 - \hat{p})/n}$$

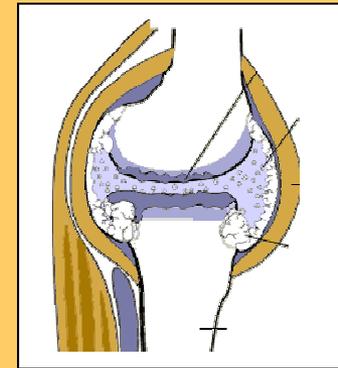
Use this method when the number of successes and the number of failures are both at least 15.



$C$  is the area under the standard normal curve between  $-z^*$  and  $z^*$ .

# Medication side effects

Arthritis is a painful, chronic inflammation of the joints. An experiment on the side effects of pain relievers examined arthritis patients to find the proportion of patients who suffer side effects.



## What are some side effects of ibuprofen?

Serious side effects (seek medical attention immediately):

Allergic reaction (difficulty breathing, swelling, or hives)

Muscle cramps, numbness, or tingling

Ulcers (open sores) in the mouth

Rapid weight gain (fluid retention)

Seizures

Black, bloody, or tarry stools

Blood in your urine or vomit

Decreased hearing or ringing in the ears

Jaundice (yellowing of the skin or eyes)

Abdominal cramping, indigestion, or heartburn

Less serious side effects (discuss with your doctor):

Dizziness or headache

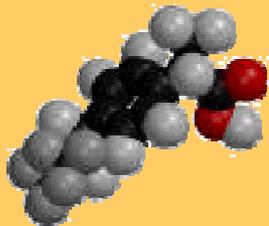
Nausea, gaseousness, diarrhea, or constipation

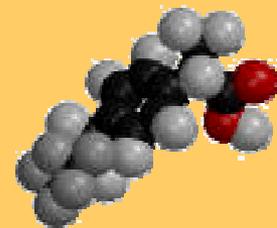
Depression

Fatigue or weakness

Dry mouth

Irregular menstrual periods





Let's calculate a 90% confidence interval for the population proportion of arthritis patients who suffer some "adverse symptoms."

What is the sample proportion  $\hat{p}$  ?

$$\hat{p} = \frac{23}{440} \approx 0.052$$

What is the sampling distribution for the proportion of arthritis patients with adverse symptoms for samples of 440?

$$\hat{p} \approx N(p, \sqrt{p(1-p)/n})$$

For a 90% confidence level,  $z^* = 1.645$ .

$z^*$	0.67	0.841	1.036	1.282	1.645	1.960	2.054	2.326
	50%	60%	70%	80%	90%	95%	96%	98%
	<b>Confidence level C</b>							

Using the large sample method, we calculate a margin of error  $m$ :

$$m = z^* \sqrt{\hat{p}(1-\hat{p})/n}$$

$$m = 1.645 * \sqrt{0.052(1-0.052)/440}$$

$$m = 1.645 * 0.0106 \approx 0.0174$$

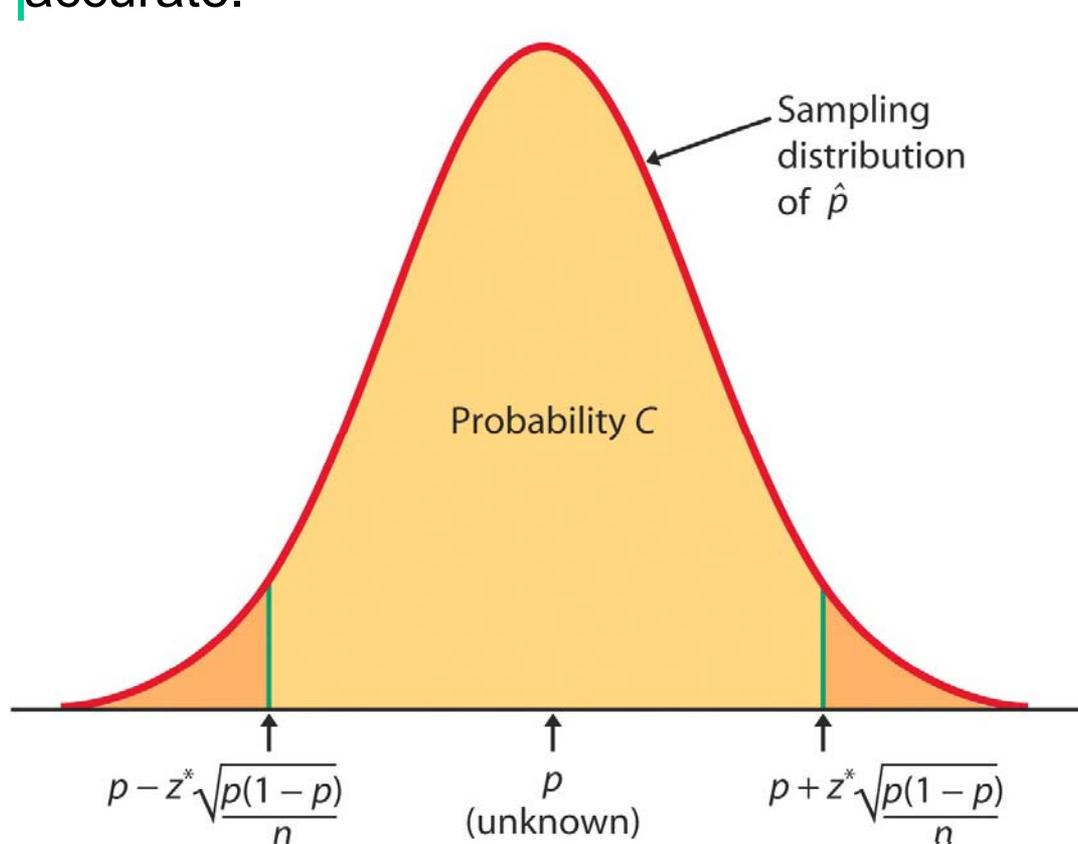
90% CI for  $p : \hat{p} \pm m$

or  $0.052 \pm 0.0174$

→ With a 90% confidence level, between 3.5% and 6.9% of arthritis patients taking this pain medication experience some adverse symptoms.

# Cautions about large sample confidence intervals for $p$

Because we have to use an estimate of  $p$  to compute the margin of error, confidence intervals for a population proportion are not very accurate.



$$m = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

*Specifically, we tend to be incorrect more often than the confidence level would indicate. But there is no systematic amount (because it depends on  $p$ ).*

Use this method only when the number of successes and the number of failures are both at least 15.

# “Plus four” confidence interval for $p$

A simple adjustment produces more accurate confidence intervals. We act as if we had four additional observations, two being successes and two being failures. Thus, the new sample size is  $n + 4$ , and the count of successes is  $X + 2$ .

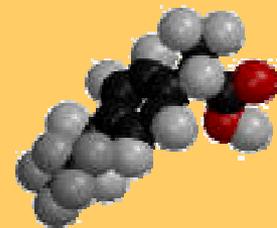
The “plus four” estimate of  $p$  is: 
$$\tilde{p} = \frac{\text{counts of successes} + 2}{\text{count of all observations} + 4}$$

And an approximate level  $C$  confidence interval is:

$CI: \tilde{p} \pm m$ , with

$$m = z^* SE = z^* \sqrt{\tilde{p}(1 - \tilde{p}) / (n + 4)}$$

Use this method when the number of success and failures is less than 15.



We now use the “plus four” method to calculate the 90% confidence interval for the population proportion of arthritis patients who suffer some “adverse symptoms.”

What is the value of the “plus four” estimate of  $p$ ?  $\tilde{p} = \frac{23+2}{440+4} = \frac{25}{444} \approx 0.056$

An approximate 90% confidence interval for  $p$  using the “plus four” method is:

$$m = z^* \sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$$

$$m = 1.645 * \sqrt{0.056(1-0.056)/444}$$

$$m = 1.645 * 0.011 \approx 0.018$$

90% CI for  $p$ :  $\tilde{p} \pm m$

or  $0.056 \pm 0.018$

→ With 90% confidence level, between 3.8% and 7.4% of arthritis patients taking this pain medication experience some adverse symptoms.

	Upper tail probability P											
	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.003	0.001	0.0005
$z^*$	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

# Significance test for $p$

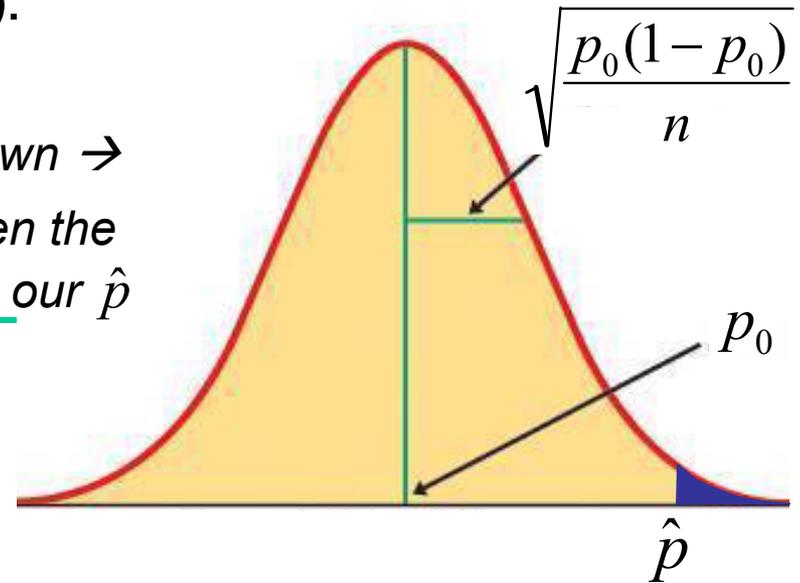
The sampling distribution for  $\hat{p}$  is approximately normal for large sample sizes and its shape depends solely on  $p$  and  $n$ .

Thus, we can easily test the null hypothesis:

$H_0: p = p_0$  (a given value we are testing).

*If  $H_0$  is true, the sampling distribution is known  $\rightarrow$   
The likelihood of our sample proportion given the null hypothesis depends on how far from  $p_0$  our  $\hat{p}$  is in units of standard deviation.*

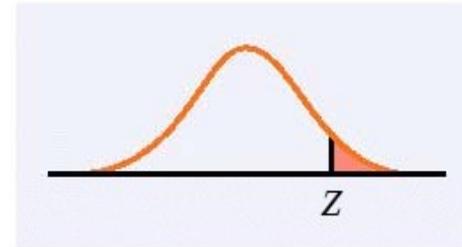
$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$



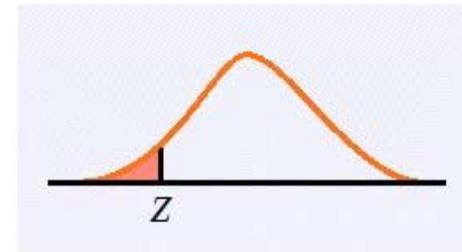
This is valid when both expected counts—expected successes  $np_0$  and expected failures  $n(1 - p_0)$ —are each 10 or larger.

# P-values and one or two sided hypotheses—reminder

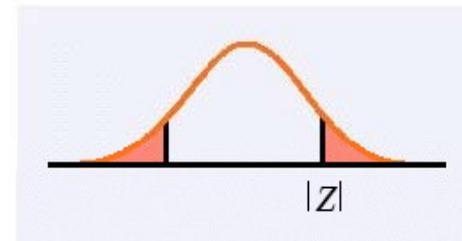
$$H_a: p > p_0 \text{ is } P(Z \geq z)$$



$$H_a: p < p_0 \text{ is } P(Z \leq z)$$



$$H_a: p \neq p_0 \text{ is } 2P(Z \geq |z|)$$



And as always, if the p-value is as small or smaller than the significance level  $\alpha$ , then the difference is statistically significant and we reject  $H_0$ .

A national survey by the National Institute for Occupational Safety and Health on restaurant employees found that 75% said that work stress had a negative impact on their personal lives.

You investigate a restaurant chain to see if the proportion of all their employees negatively affected by work stress differs from the national proportion  $p_0 = 0.75$ .

$$H_0: p = p_0 = 0.75 \text{ vs. } H_a: p \neq 0.75 \text{ (2 sided alternative)}$$

In your SRS of 100 employees, you find that 68 answered “Yes” when asked, “Does work stress have a negative impact on your personal life?”

The expected counts are  $100 \times 0.75 = 75$  and 25.

Both are greater than 10, so we can use the z-test.

The test statistic is:

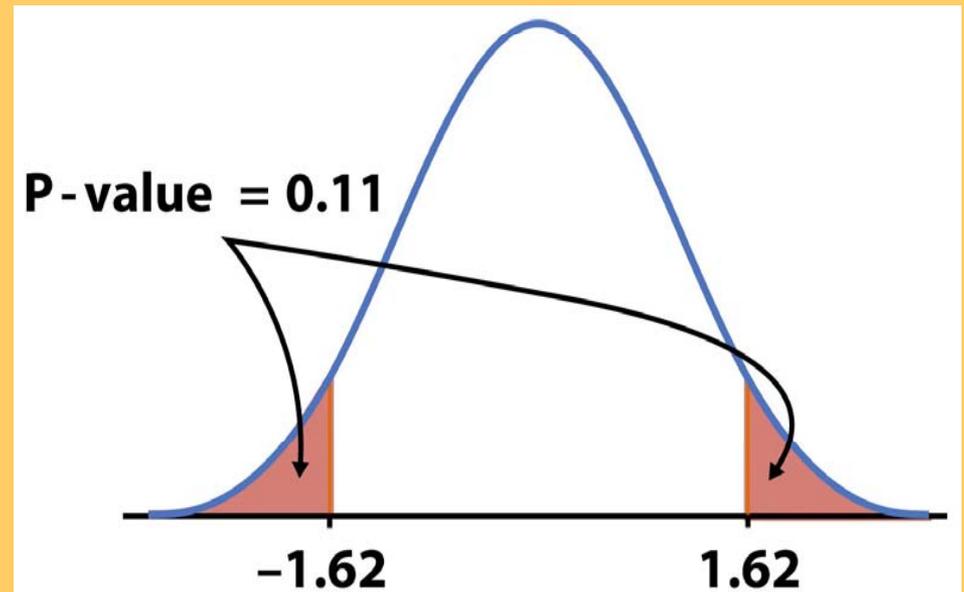
$$\begin{aligned} z &= \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\ &= \frac{0.68 - 0.75}{\sqrt{\frac{(0.75)(0.25)}{100}}} = -1.62 \end{aligned}$$

From Table A we find the area to the left of  $z = -1.62$  is 0.0526.

Thus  $P(Z \leq -1.62) = 0.0526$ . Since the alternative hypothesis is two-sided, the  $P$ -value is the area in both tails, and therefore the  $p$ -value  $= 2 \times 0.0526 = 0.1052$ .

$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

→ The chain restaurant data are not significantly different from the national survey results ( $\hat{p} = 0.68$ ,  $z = -1.62$ ,  $p$ -value = 0.11).



Software gives you summary data (sample size and proportion) as well as the actual p-value.

Minitab

### Test and Confidence Interval for One Proportion

Test of  $p = 0.75$  vs  $p \text{ not } = 0.75$

Sample	X	N	Sample p	95.0 % CI	Z-Value	P-Value
1	68	100	0.680000	(0.588572, 0.771428)	-1.62	0.106

Crunch It!

### Hypothesis test results:

$p$  = proportion of successes for population

Parameter:  $p$

$H_0$  : Parameter = 0.75

$H_A$  : Parameter not = 0.75

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
$p$	68	100	0.68	0.04330127	-1.6165807	0.106

# Interpretation: magnitude vs. reliability of effects

The **reliability** of an interpretation is related to the strength of the evidence. The smaller the **p-value**, the stronger the evidence against the null hypothesis and the more confident you can be about your interpretation.

The **magnitude** or **size** of an effect relates to the real-life relevance of the phenomenon uncovered. The p-value does NOT assess the relevance of the effect, nor its magnitude.

A **confidence interval** will assess the magnitude of the effect. However, magnitude is not necessarily equivalent to how theoretically or practically relevant an effect is.

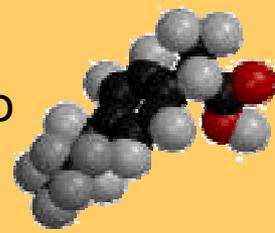
# Sample size for a desired margin of error

You may need to choose a sample size large enough to achieve a specified margin of error. However, because the sampling distribution of  $\hat{p}$  is a function of the population proportion  $p$ , this process requires that you guess a likely value for  $p$ :  $p^*$ .

$$p \sim N\left(p, \sqrt{p(1-p)/n}\right) \Rightarrow n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*)$$

The margin of error will be less than or equal to  $m$  if  $p^*$  is chosen to be 0.5.

Remember, though, that sample size is not always stretchable at will. There are typically costs and constraints associated with large samples.



What sample size would we need in order to achieve a margin of error no more than 0.01 (1%) for a 90% confidence interval for the population proportion of arthritis patients who suffer some “adverse symptoms”?

We could use 0.5 for our guessed  $p^*$ . However, since the drug has been approved for sale over the counter, we can safely assume that no more than 10% of patients should suffer “adverse symptoms” (a better guess than 50%).

For a 90% confidence level,  $z^* = 1.645$ .

$z^*$	0.67	0.841	1.036	1.282	1.645	1.960	2.054	2.326
	50%	60%	70%	80%	90%	95%	96%	98%
	<b>Confidence level C</b>							

$$n = \left( \frac{z^*}{m} \right)^2 p^* (1 - p^*) = \left( \frac{1.645}{0.01} \right)^2 (0.1)(0.9) \approx 2435.4$$

→ To obtain a margin of error no more than 1%, we would need a sample size  $n$  of at least 2,436 arthritis patients.

---

Inference for Proportions

# 8.2 Comparing Two Proportions

---

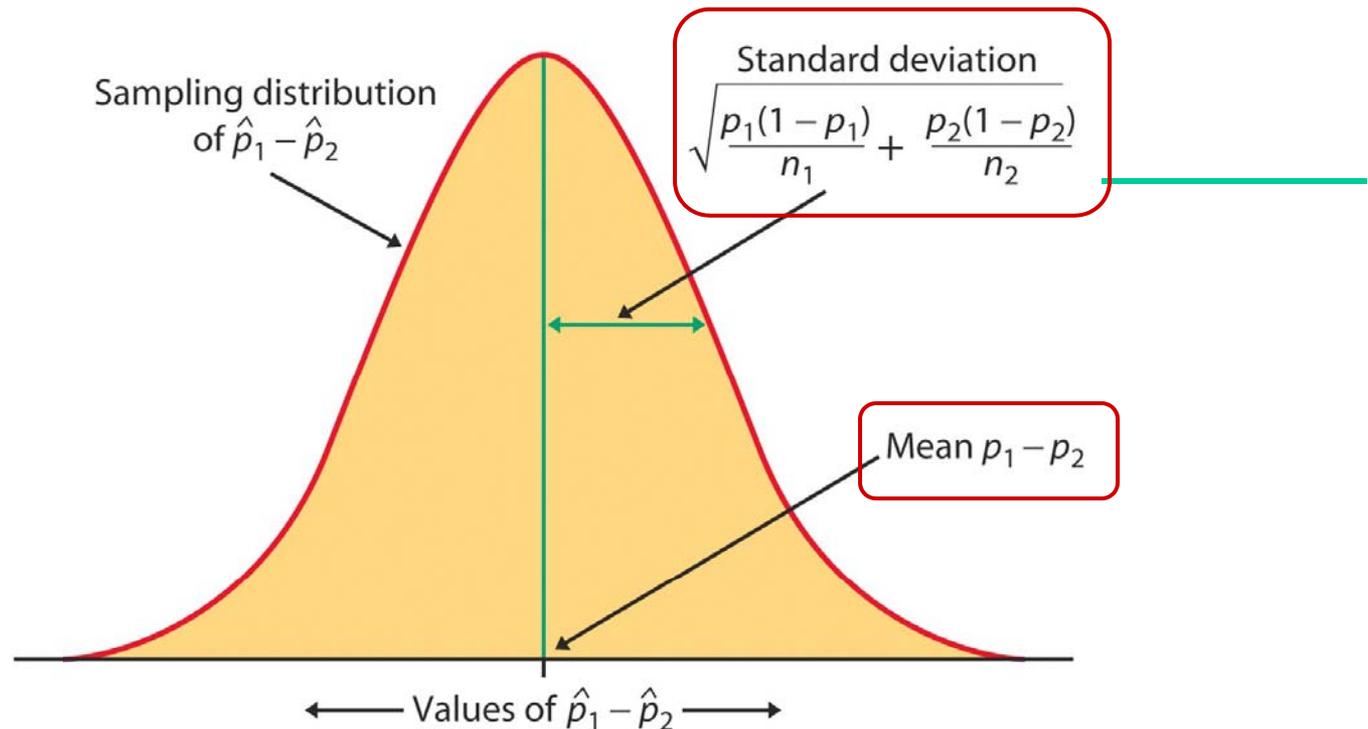
# Objectives

## 8.2 Comparing two proportions

- ❑ Large-sample CI for a difference in proportions
- ❑ “Plus four” CI for a difference in proportions
- ❑ Significance test for a difference in proportions
- ❑ Relative risk

# Comparing two independent samples

We often need to compare two treatments used on **independent** samples. We can compute the difference between the two sample proportions and compare it to the corresponding, approximately normal sampling distribution for  $(\hat{p}_1 - \hat{p}_2)$ :



# Large-sample CI for two proportions

For two independent SRSs of sizes  $n_1$  and  $n_2$  with sample proportion of successes  $\hat{p}_1$  and  $\hat{p}_2$  respectively, an **approximate level C confidence interval** for  $p_1 - p_2$  is

$(\hat{p}_1 - \hat{p}_2) \pm m$ ,  $m$  is the margin of error

$$m = z^* SE_{diff} = z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

*C is the area under the standard normal curve between  $-z^*$  and  $z^*$ .*

Use this method only when the populations are at least 10 times larger than the samples and the number of successes and the number of failures are each at least 10 in each sample.

## Cholesterol and heart attacks

How much does the cholesterol-lowering drug Gemfibrozil help reduce the risk of heart attack? We compare the incidence of heart attack over a 5-year period for two random samples of middle-aged men taking either the drug or a placebo.

Standard error of the difference  $p_1 - p_2$ :

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

$$SE = \sqrt{\frac{0.0273(0.9727)}{2051} + \frac{0.0414(0.9586)}{2030}} = \sqrt{0.000325} = 0.0057$$

	Heart attack	$n$	$\hat{p}$
Drug	56	2051	2.73%
Placebo	84	2030	4.14%

The confidence interval is  $(\hat{p}_1 - \hat{p}_2) \pm z * SE$

So the 90% CI is  $(0.0414 - 0.0273) \pm 1.645 * 0.0057 = 0.0141 \pm 0.0094$

We estimate with 90% confidence that the percentage of middle-aged men who suffer a heart attack is 0.47% to 2.35% lower when taking the cholesterol-lowering drug.

# “Plus four” CI for two proportions

The “plus four” method again produces more accurate confidence intervals. We act as if we had four additional observations: one success and one failure in each of the two samples. The new combined sample size is  $n_1 + n_2 + 4$  and the proportions of successes are:

$$\tilde{p}_1 = \frac{X_1 + 1}{n_1 + 2} \quad \text{and} \quad \tilde{p}_2 = \frac{X_2 + 1}{n_2 + 2}$$

An approximate level  $C$  confidence interval is:

$$CI: (\tilde{p}_1 - \tilde{p}_2) \pm z^* \sqrt{\frac{\tilde{p}_1(1 - \tilde{p}_1)}{n_1 + 2} + \frac{\tilde{p}_2(1 - \tilde{p}_2)}{n_2 + 2}}$$

Use this when  $C$  is at least 90% and both sample sizes are at least 5.

## Cholesterol and heart attacks

Let's now calculate the "plus four" CI for the difference in percentage of middle-aged men who suffer a heart attack (placebo – drug).

	Heart attack	n	$\tilde{p}$
Drug	56	2051	2.78%
Placebo	84	2030	4.18%

$$\tilde{p}_1 = \frac{X_1 + 1}{n_1 + 2} = \frac{56 + 1}{2051 + 2} \approx 0.0278 \quad \text{and} \quad \tilde{p}_2 = \frac{X_2 + 1}{n_2 + 2} = \frac{84 + 1}{2030 + 2} \approx 0.0418$$

Standard error of the population difference  $p_1 - p_2$ :

$$SE = \sqrt{\frac{\tilde{p}_1(1 - \tilde{p}_1)}{n_1 + 2} + \frac{\tilde{p}_2(1 - \tilde{p}_2)}{n_2 + 2}} = \sqrt{\frac{0.0278(0.9722)}{2053} + \frac{0.0418(0.9582)}{2032}} = 0.0057$$

The confidence interval is  $(\tilde{p}_1 - \tilde{p}_2) \pm z^* SE$

So the 90% CI is  $(0.0418 - 0.0278) \pm 1.645 * 0.00573 = 0.014 \pm 0.0094$

We estimate with 90% confidence that the percentage of middle-aged men who suffer a heart attack is 0.46% to 2.35% lower when taking the cholesterol-lowering drug.

# Test of significance

If the null hypothesis is true, then we can rely on the properties of the sampling distribution to estimate the probability of drawing 2 samples with proportions  $\hat{p}_1$  and  $\hat{p}_2$  at random.

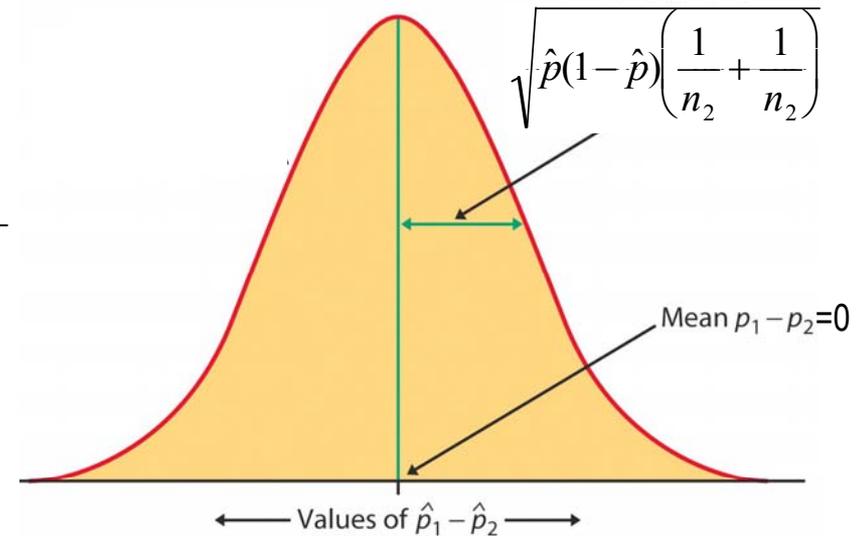
$$H_0 : p_1 = p_2 = p$$

Our best estimate of  $p$  is  $\hat{p}$ ,

the pooled sample proportion

$$\hat{p} = \frac{\text{total successes}}{\text{total observations}} = \frac{\text{count}_1 + \text{count}_2}{n_1 + n_2}$$

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



This test is appropriate when the populations are at least 10 times as large as the samples and all counts are at least 5 (number of successes and number of failures in each sample).

## Gastric Freezing

Gastric freezing was once a treatment for ulcers. Patients would swallow a deflated balloon with tubes, and a cold liquid would be pumped for an hour to cool the stomach and reduce acid production, thus relieving ulcer pain. **The treatment was shown to be safe, significantly reducing ulcer pain** and widely used for years.



A randomized comparative experiment later compared the outcome of gastric freezing with that of a placebo: 28 of the 82 patients subjected to gastric freezing improved, while 30 of the 78 in the control group improved.

$$H_0: p_{gf} = p_{\text{placebo}} \quad \hat{p}_{\text{pooled}} = \frac{28 + 30}{82 + 78} = 0.3625$$
$$H_a: 0 > p_{\text{placebo}} - p_{gf}$$

$$z = \frac{\hat{p}_{\text{placebo}} - \hat{p}_{\text{control}}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.341 - 0.385}{\sqrt{0.363 * 0.637\left(\frac{1}{82} + \frac{1}{78}\right)}} = \frac{-0.044}{\sqrt{0.231 * 0.025}} = -0.568$$

Conclusion: The gastric frhttp://wise.cgu.edu/power\_applet/power.aspeezing was no better than a placebo (p-value 0.2843), and this treatment was abandoned. **ALWAYS USE A CONTROL!**

# Relative risk

Another way to compare two proportions is to study the ratio of the two proportions, which is often called the **relative risk (RR)**. A relative risk of 1 means that the two proportions are equal.

The procedure for calculating confidence intervals for relative risk is more complicated (use software) but still based on the same principles that we have studied.

The age at which a woman gets her first child may be an important factor in the risk of later developing breast cancer. An international study selected women with at least one birth and recorded if they had breast cancer or not and whether they had their first child before their 30<sup>th</sup> birthday or after.

	Birth age 30+	Sample size	$\hat{p}$
Cancer	683	3220	21.2%
No	1498	10,245	14.6%

$$RR = \frac{.212}{.146} \approx 1.45$$

Women with a late first child have 1.45 times the risk of developing breast cancer.