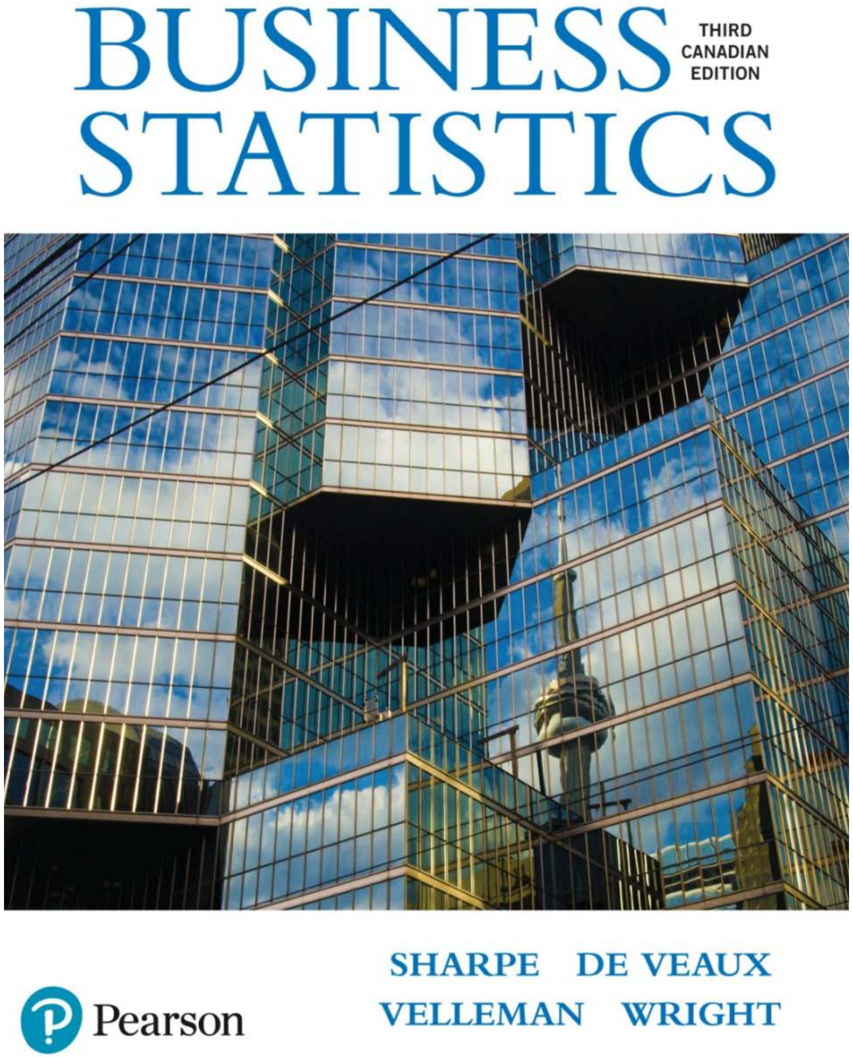
# Business Statistics

Third Canadian Edition




## Chapter 12

Testing Hypotheses about Proportions

## Chapter 11-6

Confidence Interval for the Difference Between Two Proportions

# Ch. 12: Testing Hypotheses about Proportions

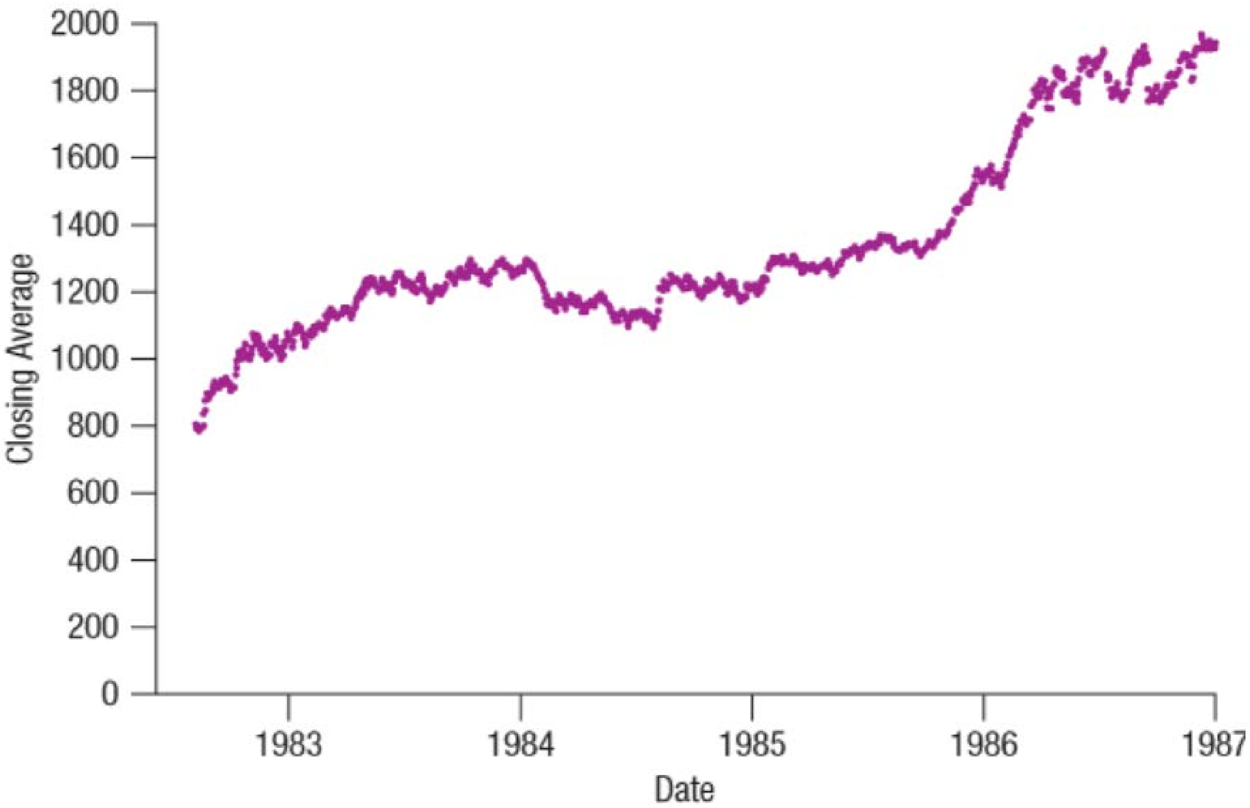


## Learning Objectives

1) Specify business issues in terms of hypothesis tests
2) Perform a hypothesis test about a proportion
3) See the relationship between hypothesis tests and confidence intervals
4) Estimate how powerful a hypothesis test is
5) Perform a hypothesis test comparing two proportions
6) Calculate a confidence interval for the difference between two proportions (11.6)

# Ch. 12: Testing Hypotheses about Proportions
(2 of 3)

How does the S&P/TSX index move?



# Ch. 12: Testing Hypotheses about Proportions

(3 of 3)

Is the Toronto Stock Exchange (TSE) just as likely to move higher as it is to move lower on any given day? Some analysts believe that the up and down days are split 50/50.

To test this out, we collect data for 1000 days and find that the proportion of up days is 0.515. That *is* more “up” days than “down” days.

But *is it far enough from 50%* to cast doubt on the assumption of equally likely up or down movement?

# 12.1 Hypotheses (1 of 8)

To test whether the daily fluctuations are equally likely to be up as down, we assume that they are, and that any apparent difference from 50% is just random fluctuation.

Null Hypothesis:

True proportion of up days = 50%

**Hypothesis** *n.*;

pl. {Hypotheses}.

A supposition; a proposition or principle which is supposed or taken for granted, in order to draw a conclusion or inference for proof of the point in question; something not proved, but assumed for the purpose of argument.

—*Webster's Unabridged Dictionary*, 1913

# 12.1 Hypotheses (2 of 8)

The *null hypothesis*, $H_0$, specifies a population model parameter and proposes a value for that parameter.

We usually write a null hypothesis about a proportion in the form $H_0 : p = p_0$.

For our hypothesis about the TSE, we need to test $H_0 : p = 0.5$.

The *alternative hypothesis*, $H_A$, contains the values of the parameter that we consider plausible if we reject the null hypothesis. Our alternative is $H_A : p \neq 0.5$.

# 12.1 Hypotheses (3 of 8)

**Additional example:**

A supplier of stainless steel kitchen utensils is having 0.86% of its merchandise returned as a result of corrosion of the steel.

The company improves the quality control on the production process, monitors 2000 shipments chosen at random, and has 0.53% of the merchandise returned for corrosion issues.

We are hoping that the proportion of merchandise returned will *go down*.

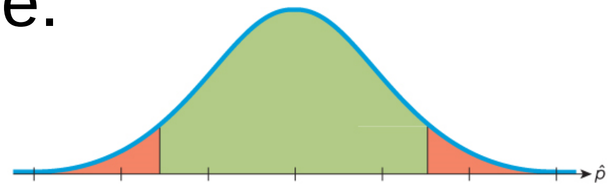
$$H_0 : p = 0.0086$$

$$H_A : p < 0.0086$$

# 12.1 Hypotheses (4 of 8)

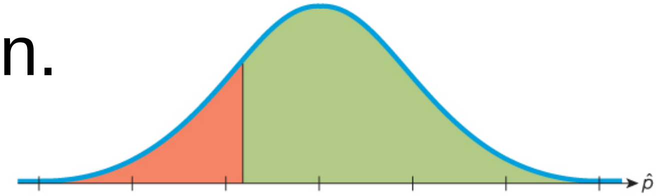
In our example about the TSE, the alternative hypothesis is known as a **two-sided alternative**, because we are equally interested in deviations on either side of the null hypothesis value.

$$H_A : p \neq 0.5.$$




In our example about returned merchandise, the alternative hypothesis is called a **one-sided alternative** because it focuses on deviations from the null hypothesis value in only one direction.

$$H_A : p < 0.0086$$



# 12.1 Hypotheses (5 of 8)

What would convince you that the proportion of up days was not 50%?

Start by finding the standard deviation of the sample proportion of days on which the S&P/TSX increased.

- We’ve seen 51.5% up days out of 1000 trading days.
- The sample size of 1000 is big enough to satisfy the Success/Failure condition.
- We suspect that the daily price changes are random and independent

# 12.1 Hypotheses (6 of 8)

If we assume that the TSE increases or decreases with equal likelihood, we'll need to center our Normal sampling model at a mean of 0.5.

In other words, we suppose that the true proportion is $p = 0.5$ ($H_0 : p = p_0 = 0.5$)

Then, the standard deviation of the sampling model is:

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.5)(1-0.5)}{1000}} = 0.0158$$

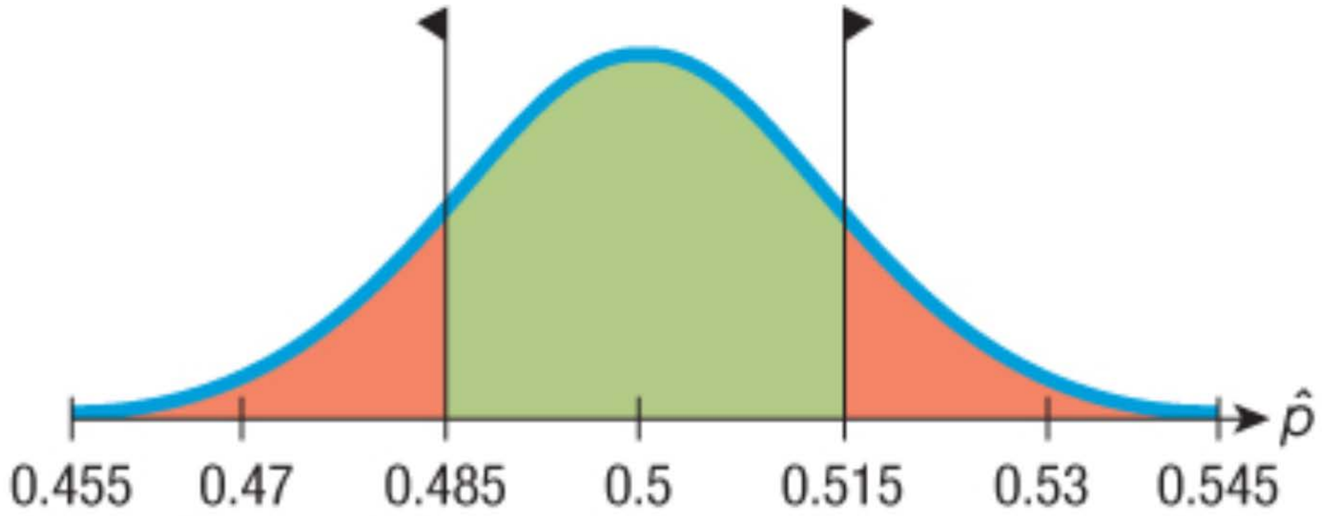
For the mean, $\mu$, we use $p$ = 0.50, and for $\sigma$ we use the standard deviation of the sample proportions

$$SD(\hat{p}) = 0.0158$$

# 12.1 Hypotheses (7 of 8)

How likely is it that the observed value would be 0.515 – 0.5 = 0.015 units away from the mean?

Looking at a picture, we can see that 0.515 doesn't look very surprising



# 12.1 Hypotheses (8 of 8)

## What NOT to Do

- Don’t fudge the null hypothesis so that you can prove something you want to prove.
- Don’t fudge the alternative hypothesis so that you can prove something you want to prove.
- Don’t put the issue that you are investigating into the null hypothesis. It should go in the alternative hypothesis.
- Don’t have different numbers in the null and alternative hypotheses. The numerical values are always the same.
- Always have the equality sign in the null hypothesis.

## FOR EXAMPLE Framing hypotheses about website customers

SmartWool, an online vendor of fancy woolen ski mitts, recently redesigned its website, and analysts at SmartWool want to know if the proportion of visits resulting in a sale has changed since the new site went online.

**QUESTION** If the old site's proportion was 20%, frame appropriate null and alternative hypotheses for the proportion.

**ANSWER** For the proportion, let $p$ = proportion of visits that result in a sale.

$$H_0: p = 0.2 \ \textit{vs.}\ H_A: p \neq 0.2$$

# 12.2 A Trial as a Hypothesis Test (1 of 2)

We started by assuming that the probability of an up day was 50%.

Then we looked at the data and concluded that we couldn’t say otherwise because the proportion that we actually observed wasn’t far enough from 50%.

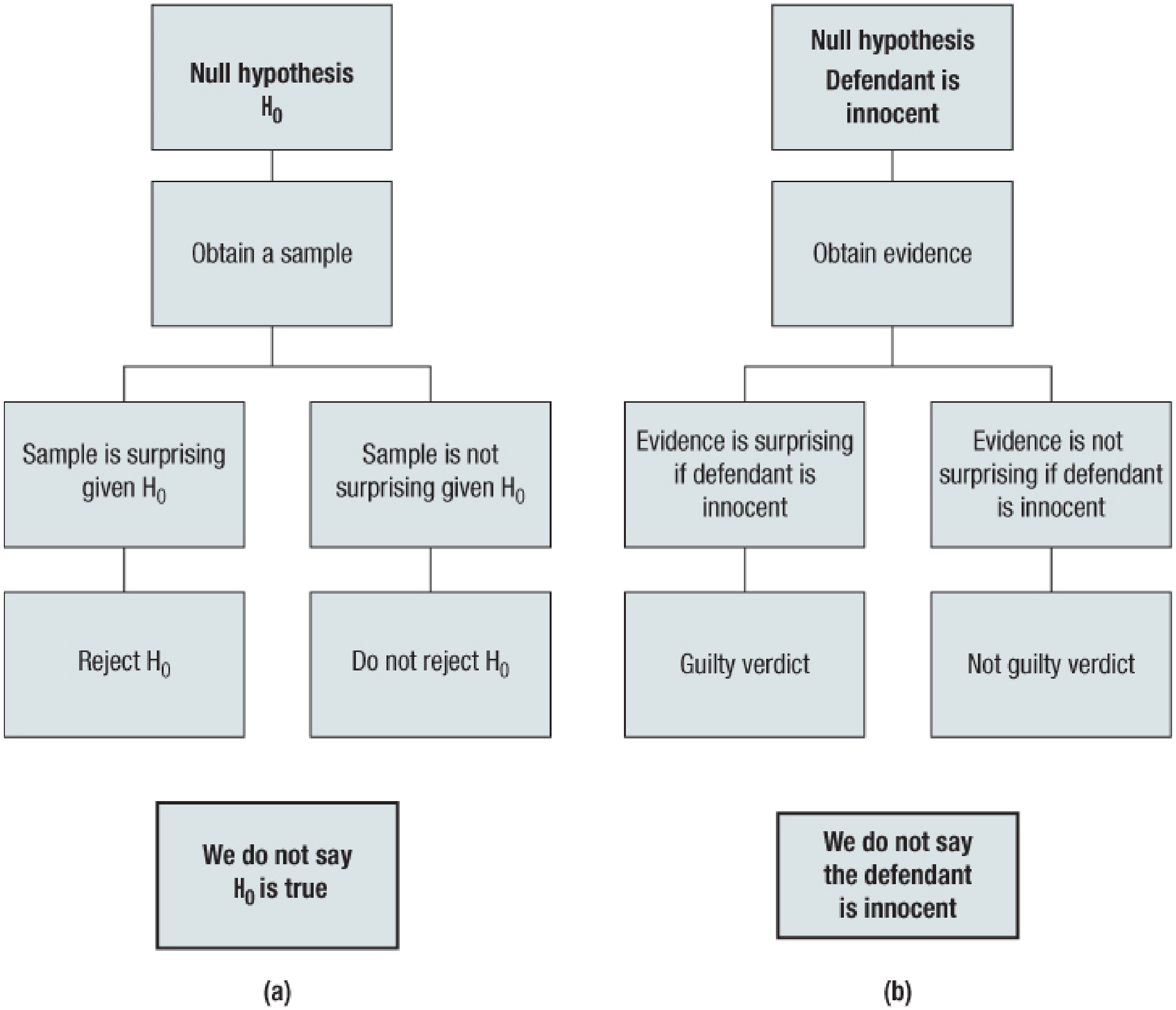
This is the logic of jury trials. In British common law, the null hypothesis is that the defendant is innocent.

The evidence takes the form of facts that seem to contradict the presumption of innocence. For us, this means collecting data.

The jury considers the evidence in light of the *presumption* of innocence and judges whether the evidence against the defendant would be plausible *if the defendant were in fact innocent*.

Like the jury, we ask: “Could these data plausibly have happened by chance if the null hypothesis were true?”

# 12.2 A Trial as a Hypothesis Test (2 of 2)




(a) Hypothesis testing. (b) Court case.

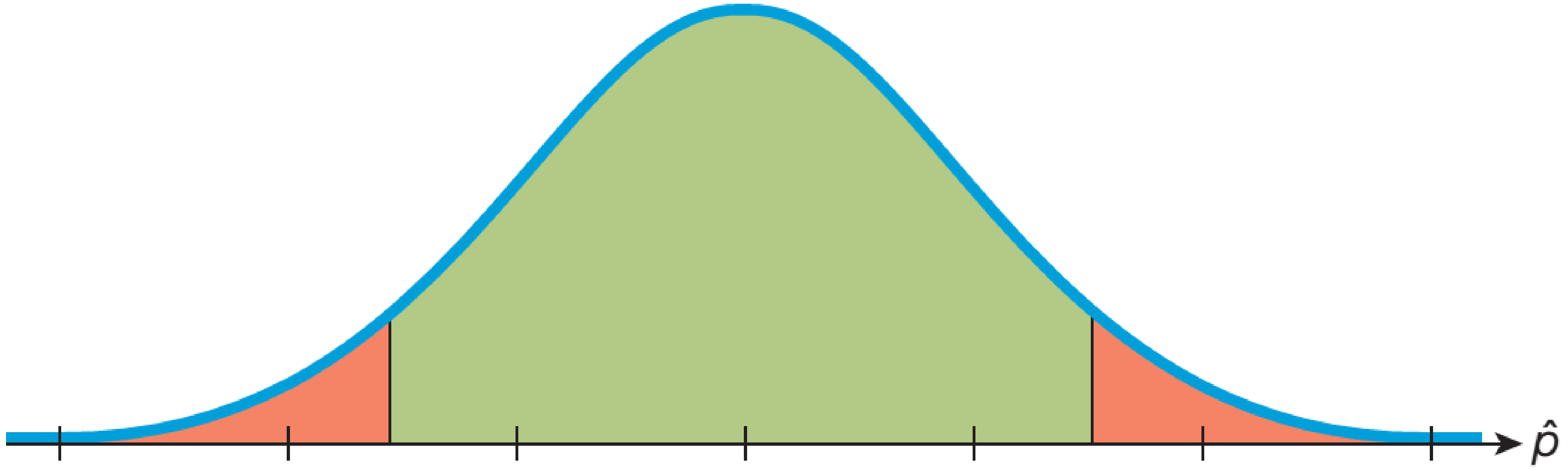
# 12.3 P-Values (1 of 4)

The *P-value* is the probability of seeing the observed data (or something even less likely) *given* the null hypothesis.

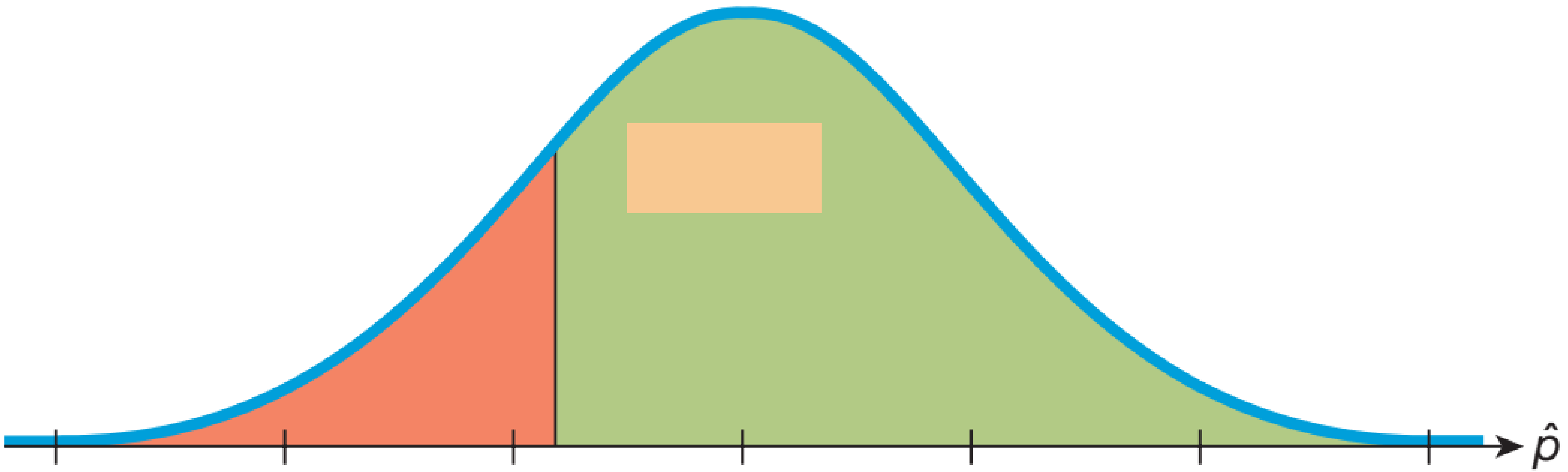
A low enough P-value says that the data we have observed would be very unlikely if our null hypothesis were true. If you believe in data more than in assumptions, then when you see a low P-value you should reject the null hypothesis.

When the P-value is *high* (or just not low *enough*), data are consistent with the model from the null hypothesis, and we have no reason to reject the null hypothesis. Formally, we say that we “fail to reject” the null hypothesis.

# 12.3 P-Values (2 of 4)



# 12.3 P-Values (3 of 4)



# 12.3 P-Values (4 of 4)

What to Do with an "Innocent" Defendant

If there is *insufficient evidence* to convict the defendant (if the P-value is *not* low), the jury does not conclude that the null hypothesis is true and declare that the defendant is innocent. Juries can only *fail to reject* the null hypothesis and declare the defendant "not guilty."

In the same way, if the data are not particularly unlikely under the assumption that the null hypothesis is true, then the most we can do is to "fail to reject" our null hypothesis.

## FOR EXAMPLE Conclusions about website customers from P-values

The SmartWool analyst (see For Example: "Framing hypotheses about website customers") is now testing hypotheses about whether the proportion of website visitors who make a purchase has *increased* from 0.2 since the redesign of the website. $H_0$: $p = 0.2$; $H_A$: $p > 0.2$. She collects a random sample of 50 visits since the new website has gone online and finds that 24% of them made purchases.

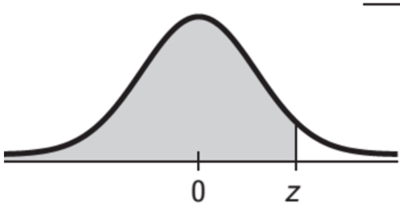
**QUESTION** What conclusions can she draw?

**ANSWER** Assuming the null hypothesis is true until it is proven false, we have $p = 0.2$, so that the standard deviation of the Normal sampling model is:

$SD(\hat{p}) = \sqrt{\frac{0.2 * 0.8}{50}} = 0.0566$. We then calculate $z = \frac{x - \mu}{\sigma}$

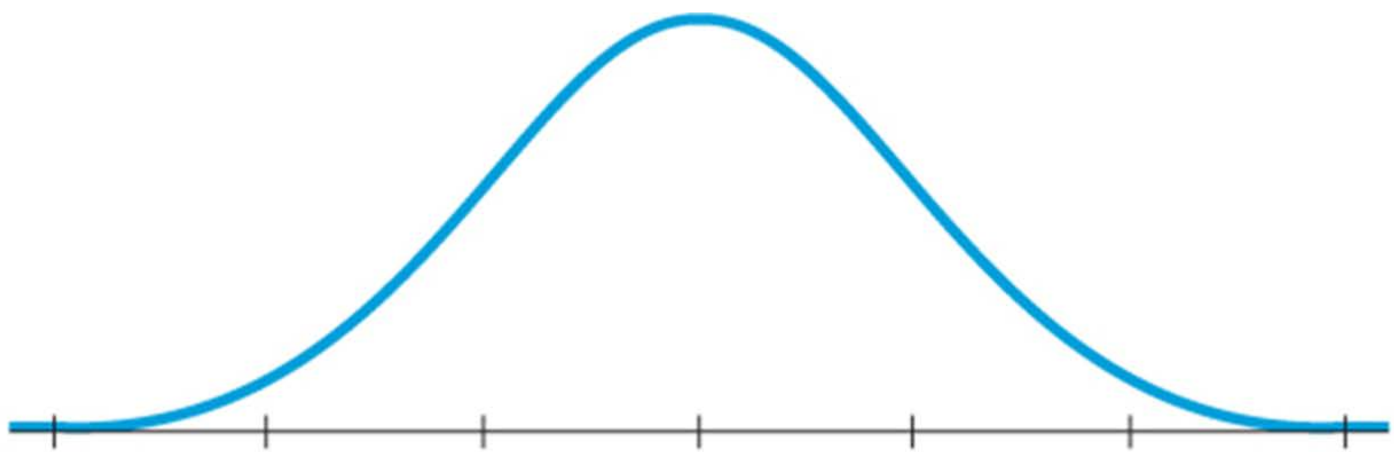
$= \frac{0.24 - 0.2}{0.0566} = 0.707$. From software, we find a probability of 0.760, so that the

P-value (that is, the probability of getting a result as extreme as 24% or more extreme) is $1 - 0.760 = 0.240$, which is not very low. The sample result is not surprising given the null hypothesis that nothing has changed. Although we have some information that the new website design may be more effective, our data could easily (with a probability of 0.24) have occurred by chance.




*Second decimal place in z*

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6703 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |



# 12.4 Alpha Levels and Significance (1 of 4)

We can define a "rare event" arbitrarily by setting a threshold for our P-value. If our P-value falls below that point, we'll reject the null hypothesis.

We call such results *statistically significant*.

The threshold is called an *alpha level*. Not surprisingly, it's labeled with the Greek letter $\alpha$.

# 12.4 Alpha Levels and Significance (2 of 4)

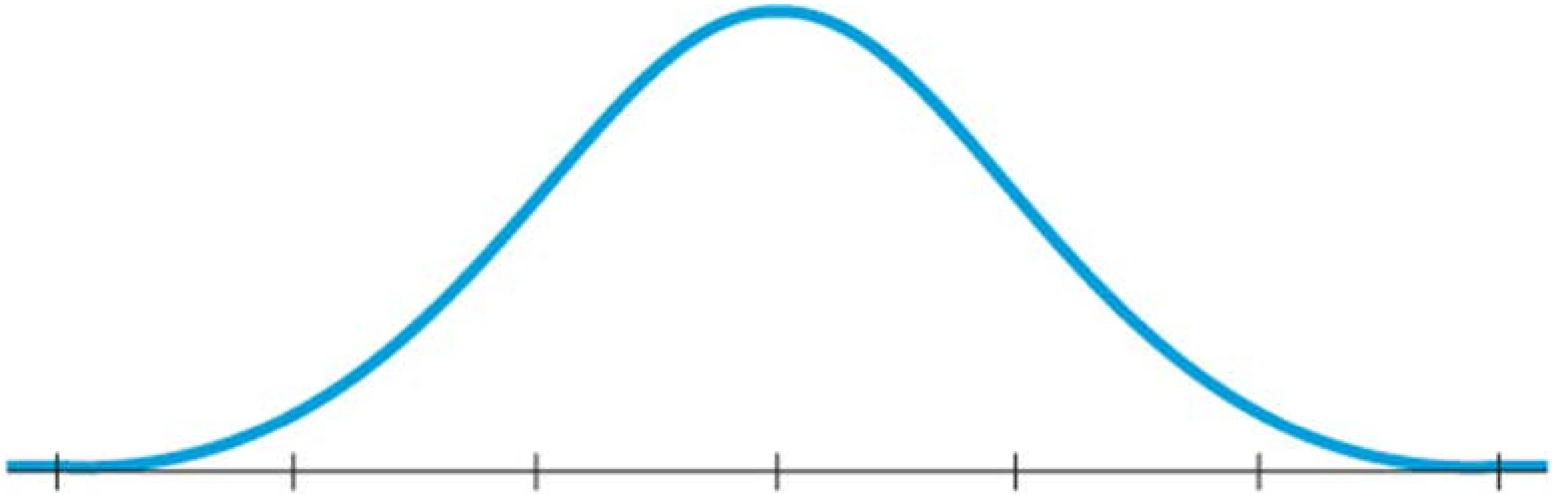
The alpha level is also called the *significance level*.

You must select the alpha level *before* you look at the data.

**Conclusion**

If the P-value $< \alpha$, then reject $H_0$.

If the P-value $\geq \alpha$, then fail to reject $H_0$.



# 12.4 Alpha Levels and Significance (3 of 4)

Many statisticians think it's better to report the P-value than to choose an alpha level and carry the decision through to a final reject/fail-to-reject verdict.

It's always a good idea to report the P-value as an indication of the strength of the evidence.

Sometimes it's best to report that the conclusion is not yet clear and to suggest that more data be gathered. In such cases, it's an especially good idea to report the P-value.

# 12.4 Alpha Levels and Significance (4 of 4)

For large samples, even small, unimportant ("insignificant") deviations from the null hypothesis can be statistically significant.

On the other hand, if the sample is not large enough, even large, financially or scientifically important differences may not be statistically significant.

It's good practice to report the magnitude of the difference between the observed statistic value and the null hypothesis value (in the data units) along with the P-value.

## FOR EXAMPLE Setting the α level for website customers

**QUESTION** Following from For Example: "Conclusions about website customers from P-values," the manager of the analyst at SmartWool wants her to compare an $\alpha$ level of 0.05 with an $\alpha$ level of 0.01 for her hypothesis tests. What would her conclusion be if the P-value comes to (a) 0.06; (b) 0.03; or (c) 0.003? (d) Should SmartWool use the new website if the P-value is 0.003?

**ANSWER**

(a) There is insufficient evidence as to whether the proportion of site visits resulting in a sale has changed from 0.2. This conclusion applies if $\alpha = 0.05$ or if $\alpha = 0.01$.

(b) If $\alpha = 0.05$ we have sufficient evidence of a change, but if $\alpha = 0.01$ we do not.

(c) There is sufficient evidence of a change in the proportion of site visits resulting in a sale. This conclusion applies if $\alpha = 0.05$ or if $\alpha = 0.01$.

(d) We cannot make the business decision as to which website to use until we have information on the costs of maintaining the two sites, whether the proportion of sales increased or decreased, and what change in profits comes from that increase or decrease.

# 12.5 The Reasoning of Hypothesis Testing

(1 of 8)

## Hypothesis Testing in a Nutshell

### Plan

1. Hypotheses (formulate them)
2. Alpha (choose it)
3. Assumptions (check them)

### Do

4. Data (collect it)
5. Statistical test (calculate the P-value)

### Report

6. Statistical significance (how significant?)
7. Hypotheses (reject the null or don’t)
8. Business significance (e.g., how can we figure out the impact on profits?)

# 12.5 The Reasoning of Hypothesis Testing

(2 of 8)

## PLAN

1. Interpret the business situation in terms of hypotheses

| General Principle | New Video Recommendation Algorithm |
|---|---|
| Choose a numerical parameter that is important to the business context. | We want to know the proportion, $p$, of customers satisfied with our video recommendations. |
| Decide whether the business situation needs you to investigate whether this parameter has increased, decreased, or changed in either direction. | We want to see whether $p$ has increased. |
| Formulate the null hypothesis related to this parameter: $H_0$: $p = p_0$ | $H_0$: $p = 0.67$ |
| Formulate the alternative hypothesis. The sign depends on whether we are investigating an increase ($H_A$: $p > p_0$), a decrease ($H_A$: $p < p_0$), or a change in either direction ($H_A$: $p \neq p_0$). | $H_A$: $p > 0.67$ |

# 12.5 The Reasoning of Hypothesis Testing

(3 of 8)

2. Interpret the business situation in terms of a significance level, $\alpha$

| General Principle | New Video Recommendation Algorithm |
|---|---|
| Choose a value of a that corresponds to the business context. Here is one way of doing that.<br>$\alpha$ = 0.001 for a life/death business where we need to be super sure of getting things right.<br>$\alpha$ = 0.01 for a business situation in which things can be measured pretty accurately.<br>$\alpha$ = 0.05 for behavioural situations where parameters are open to interpretation.<br>Other ways of choosing $\alpha$ may also be used. | $\alpha$ = 0.05 since we are surveying customers and they may interpret “satisfied” in different ways. |

# 12.5 The Reasoning of Hypothesis Testing

(4 of 8)

3. Check the assumptions and conditions for using the Normal model.

| **General Principle (these allow us to use a Normal distribution for our statistical test)** | **New Video Recommendation Algorithm** |
|---|---|
| Independence Assumption | We survey customers who do not know each other. |
| Randomization Condition | We choose our sample of customers at random. |
| 10% Condition | We surveyed 1000 customers, which is less than 10% of our total number of customers. |
| Success/Failure Condition for the parameter value in the null hypothesis, $p_0$ | $p_0 = 0.67$ and $n = 1000$<br>Therefore, $np_0 = 670 > 10$<br>Also $n(1 - p_0) = 330 > 10$ |

# 12.5 The Reasoning of Hypothesis Testing

(5 of 8)

**DO**

4. Gather data.

| **General Principle** | **New Video Recommendation Algorithm** |
|---|---|
| Do a survey or collect measurements from a production line or from company records. | We surveyed $n$ = 1000 customers and found $\hat{p} = 0.71$ |

# 12.5 The Reasoning of Hypothesis Testing
(6 of 8)

5. Apply the statistical test, calculating a P-value.

| **General Principle (we are using the “one-sample test for proportions” also known as the “one-proportion *z*-test”)** | **New Video Recommendation Algorithm** |
|---|---|
| $SD(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}}$ | $SD(\hat{p}) = \sqrt{\frac{0.67 \times (1-0.67)}{1000}} = 0.0149$ |
| $Z = \frac{x-\mu}{\sigma}$ | $z = \frac{0.71 - 0.67)}{0.0149} = 2.69$ |
| Look up the probability, $P_{Tab}$, in Table Z, Appendix B.<br>For ($H_A$: $p > p_0$), P-value = 1 – $P_{Tab}$<br>For ($H_A$: $p < p_0$), P-value = $P_{Tab}$<br>For ($H_A$: $p \neq p_0$), P-value = 2 × (1 – $P_{Tab}$) | From the table, the probability of being lower than 0.71 is $P_{Tab}$ = 0.996.<br>We have $H_A$: $p > 0.67$.<br>Therefore: P-value = 1 – 0.996 = 0.004. |

# 12.5 The Reasoning of Hypothesis Testing

(7 of 8)

## REPORT

6. Interpret the P-value in terms of statistical significance.

| General Principle | New Video Recommendation Algorithm |
|---|---|
| If P-value < α, then we have a result that is statistically significant at the level *α*. | P-value = 0.004 < *α* = 0.05<br>Our survey results are statistically significant at the 5% level. |

7. Interpret the P-value in terms of your original hypotheses.

| General Principle | New Video Recommendation Algorithm |
|---|---|
| If the results are statistically significant, we can reject the null hypothesis. | We reject the null hypothesis that the proportion of customers satisfied with our video recommendation algorithm remained unchanged at 0.67. |

# 12.5 The Reasoning of Hypothesis Testing
(8 of 8)

8. Interpret these results in the business context.

| General Principle | New Video Recommendation Algorithm |
| --- | --- |
| State what the P-value and your decision about the null hypothesis mean in the business context. | For this type of situation, we would have been happy with a 5% significance level. In fact, our P-value is 0.4%, indicating a result that is even more significant. The increase in the proportion of satisfied customers is very significant. We are very sure that the new video recommendation algorithm has increased customer satisfaction. |

## FOR EXAMPLE — The reasoning of hypothesis tests about website customers

**QUESTION** The analyst at SmartWool (see For Example: "Setting the $\alpha$ level for website customers") makes another change to the website, puts it online, selects 200 recent web visits at random, and finds that 29% of them have resulted in a sale. Would this be a surprising proportion of sales if the true proportion of sales were 20%? We are interested in whether the proportion of visits resulting in a sale has changed since the new website went live. Comment on the statistical and business significance of this result.

**ANSWER** We are going to follow the eight steps described above and number them to illustrate clearly what we have just covered. You don't always have to number them precisely like that, but you must be sure to cover all the eight steps within the context of the business situation you are analyzing.

1. We want to investigate whether the proportion, $p$, of customers making a purchase has changed. The baseline proportion is $p_0 = 0.2$, therefore our hypotheses are: $H_0$: $p = 0.2$; $H_A$: $p \neq 0.2$.
2. Since this is a situation where quantitative data is accurate, we choose $\alpha = 0.01$.

3. Customers behave independently of each other and were chosen at random. $n = 200$ and we assume the company has more than 2000 customers. $np_0 = 40 > 10$. $n(1 - p_0) = 160 > 10$.
4. We surveyed $n = 200$ customers and found a proportion $\hat{p} = 0.29$ made a purchase.
5. $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.2)(0.8)}{200}} = 0.02828.$

$$z = \frac{\hat{p} - p_0}{SD(\hat{p})} = \frac{0.29 - 0.20}{0.02828} = 3.182.$$

From Table Z in Appendix B, we find $P_{Tab} = 0.9993$. Therefore the P-value (for our two-sided test) is $2 \times (1 - 0.9993) = 0.0014$.

6. Our P-value = 0.0014 is a lot less than our $\alpha = 0.01$. We therefore have a result that is statistically significant at the 1% level.
7. We can therefore reject the null hypothesis that $p = 0.2$.
8. We have very good evidence that the proportion of website visitors who make purchases has changed from 0.2. The P-value is 0.0014, indicating a probability of 0.0014 that this result could have occurred by chance. Before going ahead with the new website, we should assess the costs of operating it and also the additional revenues from the increased proportion ($0.29 - 0.2 = 0.09$) of purchasing customers in order to assess its profitability.

# GUIDED EXAMPLE Home Field Advantage

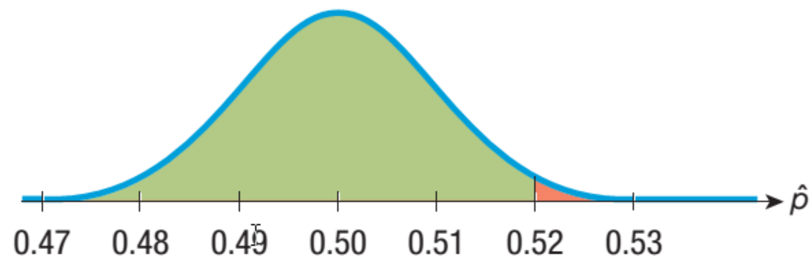
Nathan Denette/The Canadian Press/AP Images

Major league sports are big business. And the fans are more likely to come out to root for the team if the home team has a good chance of winning. Anyone who follows or plays sports has heard of the "home field advantage." It is said that teams are more likely to win when they play at home. That *would* be good for encouraging the fans to come to the games. But is it true?

Suppose that, last season, the home team won 1263 of 2429 games, or 52.0% of the time. If there were no home field advantage, the home teams would win about half of all games played. Could this deviation from 50% be explained just from natural sampling variability, or does this evidence suggest that there really is a home field advantage, at least in professional baseball?

The null model is a Normal distribution with a mean of 0.50. Since this is a hypothesis test, the standard deviation is calculated from $p_0$ (in the null hypothesis).

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.5)(1-0.5)}{2429}} = 0.01015$$




The observed proportion $\hat{p}$ is 0.52.

$$z = \frac{0.52 - 0.5}{0.01015} = 1.97$$

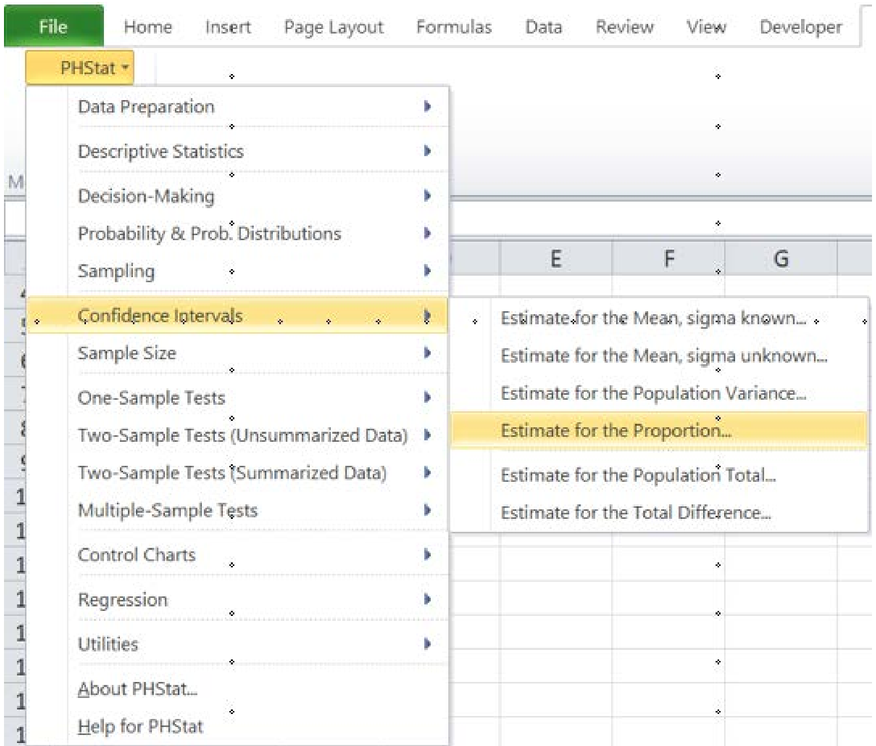
The corresponding P-value is 0.0244.

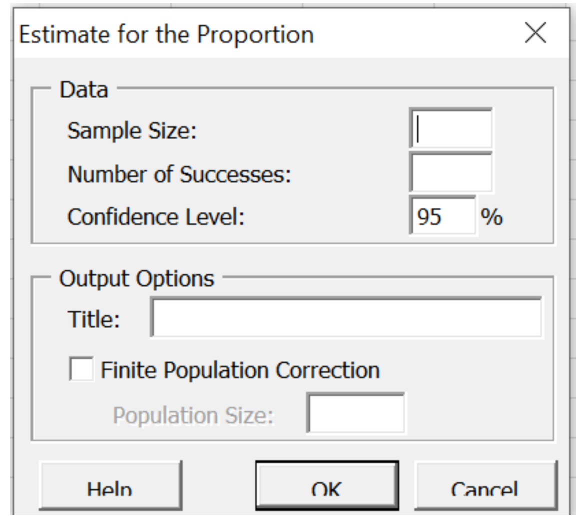
MEMO:

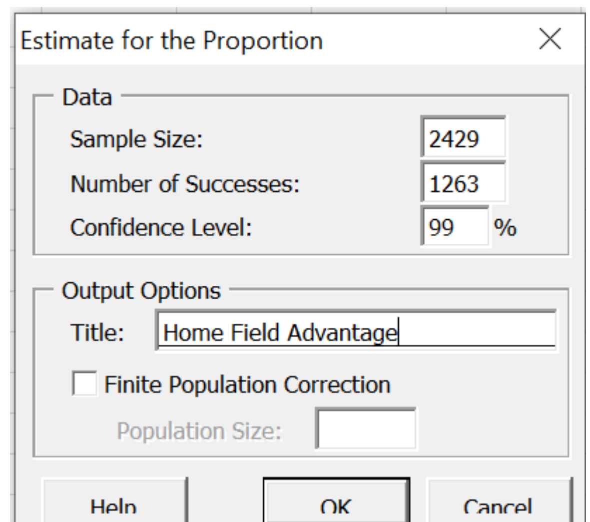
**Re: Home Field Advantage**

Our analysis of outcomes during the last baseball season did not show a statistically significant advantage to the home team at the 1% level. Our data does not allow us to be 99% sure that the home team has an advantage.

# Using EXCEL


File
Home
Insert
Page Layout
Formulas
Data
Review
View
Developer
PHStat
Data Preparation
Descriptive Statistics
Decision-Making
Probability & Prob. Distributions
Sampling
Confidence Intervals
Sample Size
One-Sample Tests
Two-Sample Tests (Unsummarized Data)
Two-Sample Tests (Summarized Data)
Multiple-Sample Tests
Control Charts
Regression
Utilities
About PHStat...
Help for PHStat
Estimate for the Mean, sigma known...
Estimate for the Mean, sigma unknown...
Estimate for the Population Variance...
Estimate for the Proportion...
Estimate for the Population Total...
Estimate for the Total Difference...

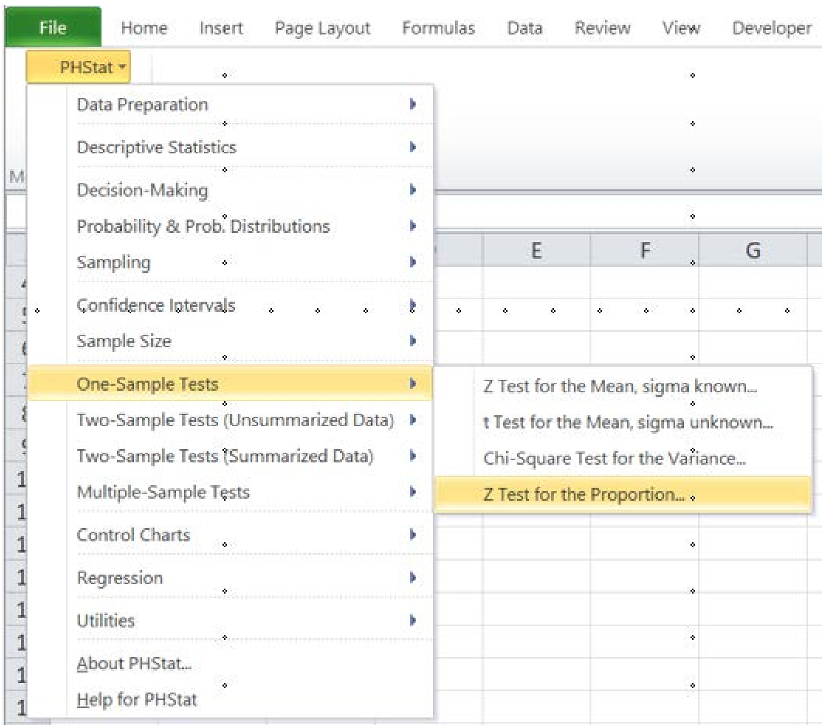

Estimate for the Proportion
Data
Sample Size:
Number of Successes:
Confidence Level: 95 %
Output Options
Title:
Finite Population Correction
Population Size:
Help
OK
Cancel

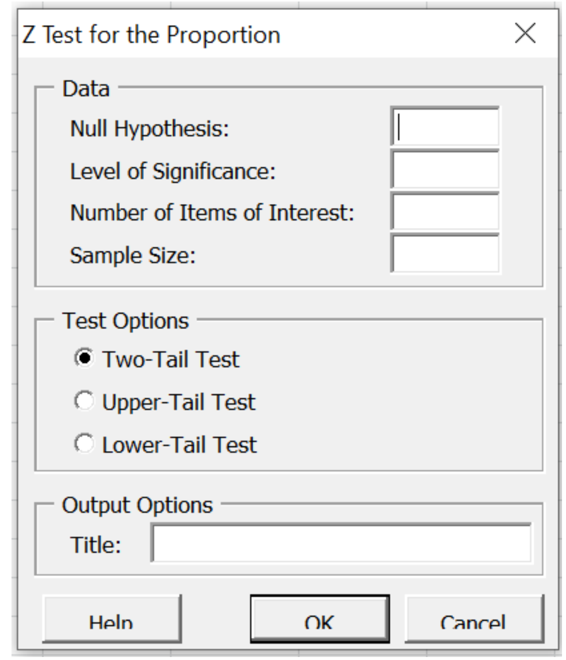

Estimate for the Proportion
Data
Sample Size: 2429
Number of Successes: 1263
Confidence Level: 99 %
Output Options
Title: Home Field Advantage
Finite Population Correction
Population Size:
Help
OK
Cancel

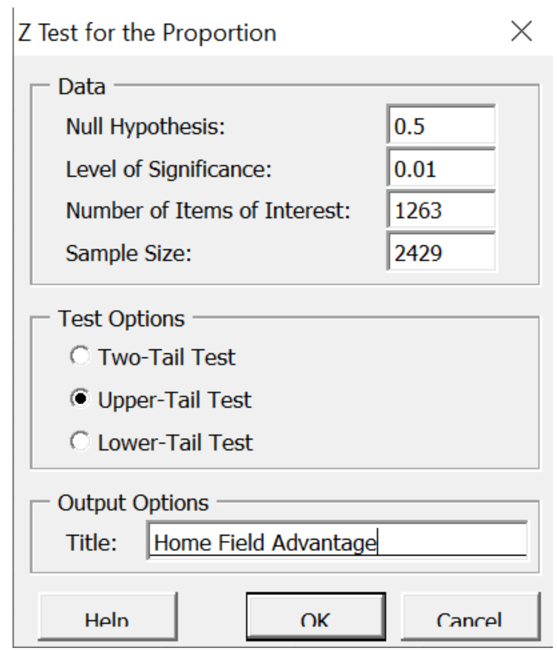
# Using EXCEL

| **Home Field Advantage** | |
|---|---|
| | |
| **Data** | |
| **Sample Size** | **2429** |
| **Number of Successes** | **1263** |
| **Confidence Level** | **99%** |
| | |
| Intermediate Calculations | |
| Sample Proportion | 0,519967065 |
| Z Value | -2,5758 |
| Standard Error of the Proportion | 0,0101 |
| Interval Half Width | 0,0261 |
| | |
| **Confidence Interval** | |
| **Interval Lower Limit** | **0,4939** |
| **Interval Upper Limit** | **0,5461** |

# Using EXCEL


File
Home
Insert
Page Layout
Formulas
Data
Review
View
Developer
PHStat
Data Preparation
Descriptive Statistics
Decision-Making
Probability & Prob. Distributions
Sampling
Confidence Intervals
Sample Size
One-Sample Tests
Two-Sample Tests (Unsummarized Data)
Two-Sample Tests (Summarized Data)
Multiple-Sample Tests
Control Charts
Regression
Utilities
About PHStat...
Help for PHStat
Z Test for the Mean, sigma known...
t Test for the Mean, sigma unknown...
Chi-Square Test for the Variance...
Z Test for the Proportion...
E
F
G


Z Test for the Proportion
Data
Null Hypothesis:
Level of Significance:
Number of Items of Interest:
Sample Size:
Test Options
Two-Tail Test
Upper-Tail Test
Lower-Tail Test
Output Options
Title:
Help
OK
Cancel


Z Test for the Proportion
Data
Null Hypothesis: 0.5
Level of Significance: 0.01
Number of Items of Interest: 1263
Sample Size: 2429
Test Options
Two-Tail Test
Upper-Tail Test
Lower-Tail Test
Output Options
Title: Home Field Advantage
Help
OK
Cancel

# Using EXCEL

| Home Field Advantage | |
|---|---|
| | |
| **Data** | |
| **Sample Size** | **2429** |
| **Number of Successes** | **1263** |
| **Confidence Level** | **99%** |
| | |
| Intermediate Calculations | |
| Sample Proportion | 0,519967065 |
| Z Value | -2,5758 |
| Standard Error of the Proportion | 0,0101 |
| Interval Half Width | 0,0261 |
| | |
| **Confidence Interval** | |
| **Interval Lower Limit** | **0,4939** |
| **Interval Upper Limit** | **0,5461** |

# Using Minitab

- From the “Stat” Menu, select “Basic Statistics”, then “1 proportion”.
- Select “summarized data” and “perform hypothesis test”.
- Summarize the data as “number of trials” = 2429, and “number of events” = 1263.
- Under “Options”, specify the value of $p_0$ = .50, the “>” alternative, the “99%” confidence level since $\alpha$ is 1%, and click on the use of the normal approximation if appropriate.
- We obtain the following output:

## Test and CI for One Proportion

### Method

p: event proportion

Exact method is used for this analysis.

### Descriptive Statistics

| N | Event | Sample p | 99% Lower Bound for p |
|---|---|---|---|
| 2429 | 1263 | 0.519967 | 0.496162 |

### Test

| | |
|---|---|
| Null hypothesis | $H_0$: p = 0.5 |
| Alternative hypothesis | $H_1$: p > 0.5 |

| P-Value |
|---|
| 0.026 |

# 12.6 Critical Values (1 of 3)

A *critical value*, $z^*$, corresponds to a selected confidence level.

Before computers and calculators were common, P-values were hard to find. It was easier to select a few common alpha levels and learn the corresponding critical values for the Normal model.

You'd calculate how many standard deviations your observed statistic was away from the hypothesized value and compare that value directly against the $z^*$ values. Any $z$-score larger in magnitude than a particular critical value will have a P-value smaller than the corresponding alpha.

# 12.6 Critical Values (2 of 3)

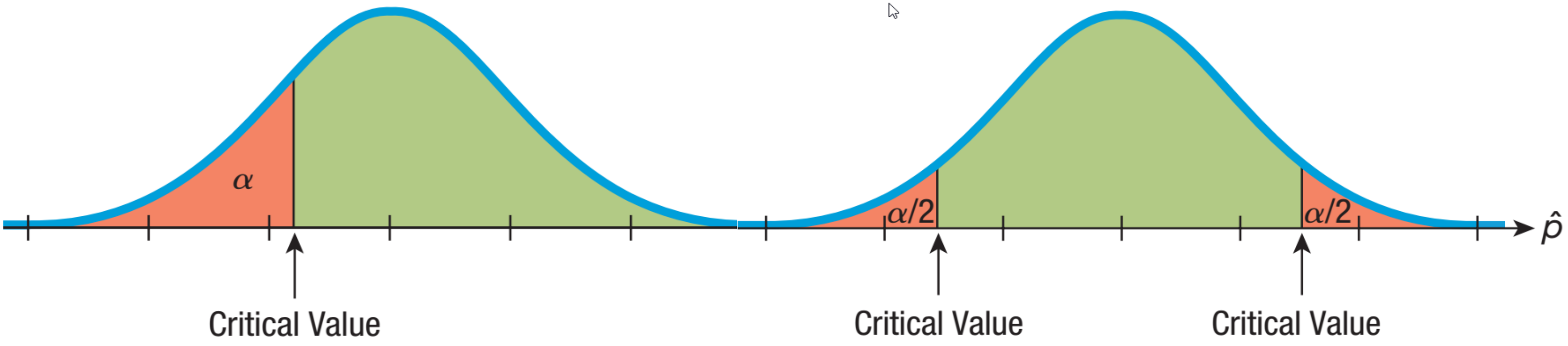
With technology, P-values are easy to find. Since they give more information about the strength of the evidence, you should report them.

Here are the traditional $z^*$ critical values from the Normal model:

**Table 12.2** Critical values, $z^*$, for different types of hypothesis test.

| α | One-Sided | Two-Sided |
|---|---|---|
| 0.10 | 1.28 | 1.645 |
| 0.05 | 1.645 | 1.96 |
| 0.01 | 2.33 | 2.576 |
| 0.001 | 3.09 | 3.29 |

# 12.6 Critical Values (3 of 3)

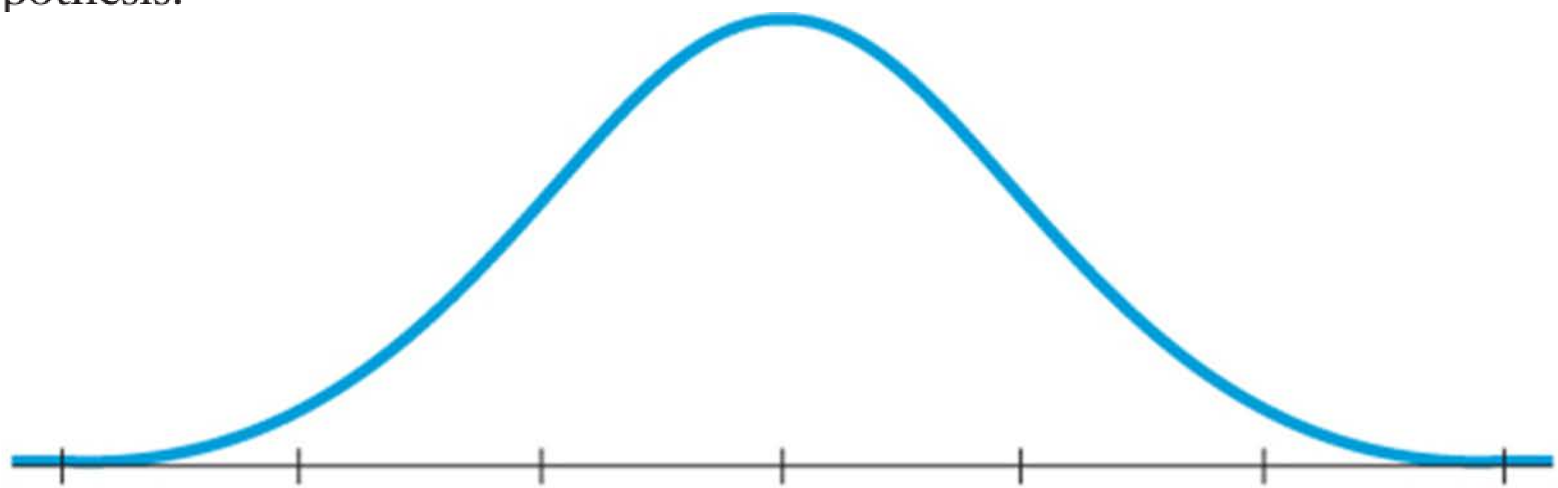


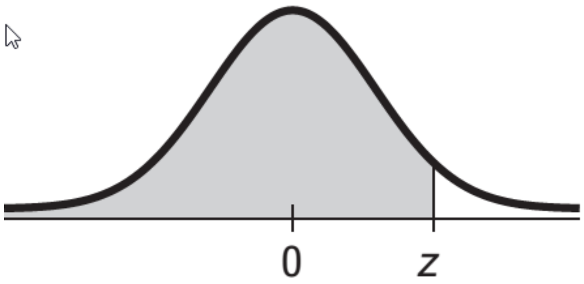

| α | One-Sided | Two-Sided |
|---|---|---|
| 0.10 | 1.28 | 1.645 |
| 0.05 | 1.645 | 1.96 |
| 0.01 | 2.33 | 2.576 |
| 0.001 | 3.09 | 3.29 |

## FOR EXAMPLE — Tests about website customers using critical values

**QUESTION** Find the critical $z$ value for the SmartWool hypothesis (see For Example: "The reasoning of hypothesis tests about website customers") using $\alpha = 0.05$ and show that the same decision would have been made using critical values.

**ANSWER** For the two-sided test of proportions, we refer to Figure 12.6. Since $\alpha = 0.05$, $\alpha/2 = 0.025$, so that $1 - \alpha/2 = 0.975$. Looking up this value in the body of Table Z in Appendix B gives a critical $z$ value of 1.96. This is also the value given in Table 12.2. Because the $z$ value was 3.182, much larger than 1.96, we reject the null hypothesis.






*Second decimal place in z*

| *z* | *0.00* | *0.01* | *0.02* | *0.03* | *0.04* | *0.05* | *0.06* | *0.07* | *0.08* | *0.09* |
|---|---|---|---|---|---|---|---|---|---|---|
| *0.0* | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| *0.1* | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| *0.2* | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6703 | 0.6141 |
| *0.3* | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| *0.4* | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| *0.5* | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| *0.6* | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| *0.7* | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| *0.8* | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| *0.9* | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| *1.0* | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| *1.1* | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| *1.2* | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| *1.3* | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| *1.4* | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| *1.5* | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| *1.6* | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| *1.7* | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| *1.8* | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| *1.9* | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |

# 12.7 Confidence Intervals and Hypothesis Tests (1 of 4)

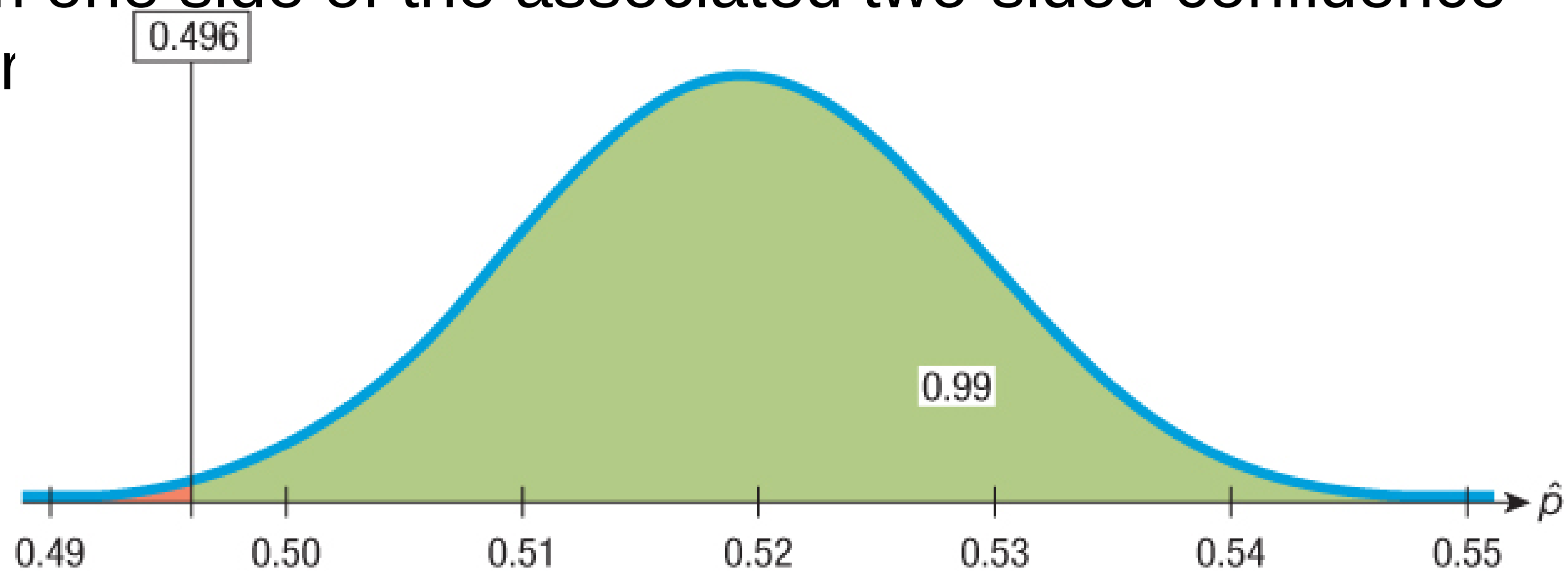
Because confidence intervals are naturally two-sided, they correspond to two-sided tests.

In general, a confidence interval with a confidence level of $C$% corresponds to a two-sided hypothesis test with an $\alpha$ level of $100 - C$%.

# 12.7 Confidence Intervals and Hypothesis Tests (2 of 4)

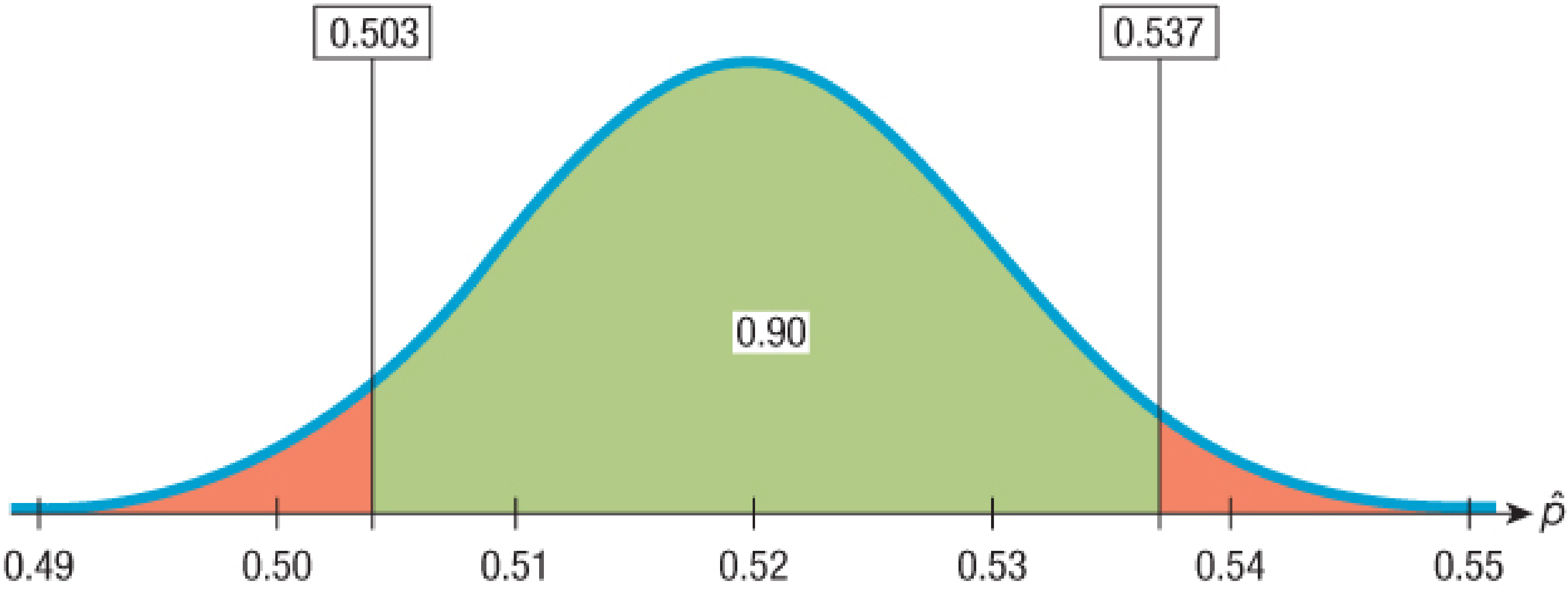
A one-sided confidence interval leaves one side unbounded.

A one-sided confidence interval can be constructed from one side of the associated two-sided confidence inter



# 12.7 Confidence Intervals and Hypothesis Tests (3 of 4)

For convenience, and to provide more information, we sometimes report a two-sided confidence interval even though we are interested in a one-sided test.



# 12.7 Confidence Intervals and Hypothesis Tests (4 of 4)

Although a hypothesis test can tell us whether the observed statistic differs from the hypothesized value, it doesn't say by how much. The corresponding confidence interval gives us more information.

- For a confidence interval, we estimate the standard deviation of $\hat{p}$ from $\hat{p}$ itself, making it a *standard error,*

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}.$$

- For the corresponding hypothesis test, we use the model's *standard deviation* for $\hat{p}$ based on the null hypothesis value $p_0$,

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}}.$$

## FOR EXAMPLE — Confidence intervals and hypothesis tests for website visits

**QUESTION** Construct an appropriate confidence interval for testing the earlier hypothesis (see For Example: "Framing hypotheses about website customers" and For Example: "The reasoning of hypothesis tests about website customers") and show how we could have reached the same conclusion from this interval.

**ANSWER** The test of proportion was two-sided, so we construct a 99% confidence interval for the true proportion: $\hat{p} \pm 2.576 \times SE(\hat{p}) = 0.29 \pm 2.576 \times \sqrt{\frac{(0.29)(0.71)}{200}}$ $= (0.207, 0.373)$. Since 0.20 is not inside this interval, 29% sales is a surprisingly large value. We reject the null hypothesis.

# 12.8 Comparing Two Proportions

In addition to hypothesis tests about the proportion of one population, we can test the difference between the proportions in two populations.

**Two-Proportion *z*-Test for equal proportions**

Testing whether two proportions are equal.

In order to test

$$\mathrm{H}_0 : p_1 - p_2 = 0$$
$$\mathrm{H}_A : P_1 - p_2 \neq 0$$

we calculate the test statistic: $z = \frac{\hat{p}_1 - \hat{p}_2}{SE(\hat{p}_1 - \hat{p}_2)}$

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

and $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$ and $\bar{q} = 1 - \bar{p}$.

We then obtain the corresponding P-value from the table for the Normal distribution.

## FOR EXAMPLE The effect of sample size when comparing two proportions of voting preferences

Survey companies like Nanos Research often survey about 1000 people in order to get a narrow standard deviation on their results and hence significant results. To see the effect of using a much smaller sample, let's suppose that only 30 people had been surveyed.

**QUESTION** If the survey of whether people would consider voting NDP had been done on only 30 people this week and on only 30 people a year ago, and resulted in 61% and 36%, respectively, would this indicate a significant difference over the past year on this issue at the 5% level?

**ANSWER** Although 61% is very different from 36%, these are percentages of very small samples (size $n = 30$). Considering the entire population of Canada, it is possible that the overall proportion who would consider voting NDP was actually higher a year ago. We therefore formulate a two-sided hypothesis test.

$H_0$: There is no difference between the proportions who would consider voting NDP this week and a year ago, i.e., $p_1 - p_2 = 0$.

$H_A$: There is a difference between the proportions who would consider voting NDP this week and a year ago, i.e., $p_1 - p_2 \neq 0$.

Checking the conditions, the Independence Assumption and Randomization Condition are assumed true if this is a professionally designed survey. Certainly these small samples are less than 10% of the population of Canada. The Success/Failure Condition is only just satisfied, indicating that these samples are really only just large enough for us to use a test based on the Normal distribution:

$$n_1p_1 = 30 \times 0.61 = 18.3 > 10;$$

$$n_1q_1 = 30 \times 0.39 = 11.7 > 10;$$

$$n_2p_2 = 30 \times 0.36 = 10.8 > 10;$$

$$n_2q_2 = 30 \times 0.64 = 19.2 > 10.$$

First we calculate the pooled proportion:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{0.36 \times 30 + 0.61 \times 30}{30 + 30} = 0.485$$

Our test statistic is

$$z = \frac{0.61 - 0.36}{\sqrt{0.485 \times 0.515 \times \left(\frac{1}{30} + \frac{1}{30}\right)}}$$

$$= 1.94$$

The corresponding P-value is 0.053, indicating that the difference is *not* significant at the 5% level.

This example shows that a difference that looks large may not be significant if the sample sizes are small.

# Minitab

- Select “two-proportions test” under “Basic Statistics”
  - **No pooled variance**

## Test and CI for Two Proportions

### Test

| | |
|---|---|
| Null hypothesis | $H_0: p_1 - p_2 = 0$ |
| Alternative hypothesis | $H_1: p_1 - p_2 \neq 0$ |

| Method | Z-Value | P-Value |
|---|---|---|
| Normal approximation | 1.86 | 0.063 |
| Fisher's exact | | 0.120 |

### Method

$p_1$: proportion where Sample 1 = Event
$p_2$: proportion where Sample 2 = Event
Difference: $p_1 - p_2$

### Descriptive Statistics

| Sample | N | Event | Sample p |
|---|---|---|---|
| Sample 1 | 30 | 18 | 0.600000 |
| Sample 2 | 30 | 11 | 0.366667 |

### Estimation for Difference

| Difference | 95% CI for Difference |
|---|---|
| 0.233333 | (-0.012568, 0.479235) |

*CI based on normal approximation*

# Minitab

- Select “two-proportions test” under “Basic Statistics”
  - **pooled variance**

## Test and CI for Two Proportions

### Method

$p_1$: proportion where Sample 1 = Event

$p_2$: proportion where Sample 2 = Event

Difference: $p_1 - p_2$

### Descriptive Statistics

| Sample | N | Event | Sample p |
|---|---|---|---|
| Sample 1 | 30 | 18 | 0.600000 |
| Sample 2 | 30 | 11 | 0.366667 |

### Estimation for Difference

| Difference | 95% CI for Difference |
|---|---|
| 0.233333 | (-0.012568, 0.479235) |

*CI based on normal approximation*

### Test

| | |
|---|---|
| Null hypothesis | $H_0$: $p_1 - p_2 = 0$ |
| Alternative hypothesis | $H_1$: $p_1 - p_2 \neq 0$ |

| Method | Z-Value | P-Value |
|---|---|---|
| Normal approximation | 1.81 | 0.071 |
| Fisher's exact | | 0.120 |

*The pooled estimate of the proportion (0.483333) is used for the tests.*

# 12.8 Example

A survey of 1002 U.S. and 1980 Britain adults by Angus Reid Strategies showed that 36% and 41% of them agree with a free-trade agreement between the two countries. Is there a difference between the support for a free-trade agreement in the U.S. and Britain?

**Checking assumptions and conditions**

- Independence Assumption: The two samples are independent
- Randomization Condition: People in each sample are selected at random
- 10% Condition: Both samples are less than 10% of their respective populations
- Success/Failure Condition

$$1002 * 0.36 = 360.12 > 10; \; 1002 * 0.64 = 641.28 > 10$$

$$1980 * 0.41 = 811.8 > 10; \quad 1980 * 0.59 = 1168.2 > 10$$

# 12.8 Example

**Formulating the hypotheses:** The null hypothesis is that support for free-trade is not different in the two countries. The alternative is that there is a difference

$$H_0 : p_1 = p_2 \; ; \; H_A : p_1 \neq p_2$$

**Calculations:**

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{(0.36 * 1002 + 0.41 * 1980)}{1002 + 1980} = 0.393$$

$$z = \frac{0.36 - 0.41}{\sqrt{0.393 * 0.607 * (\frac{1}{1002} + \frac{1}{1980})}} = -2.64$$

$$P_{VALUE} = 2 * P(z < -2.64) = 0.0082 < 0.05$$

**Conclusion:** The P-value is less than 0.05. Clearly, there is a difference between the levels of support for a free-trade agreement between Britain and U.S. at the 95% significance level.

# 12.8 Example

- Select “two-proportions test” under “Basic Statistics”
  - pooled variance

## Test and CI for Two Proportions

### Test

| | |
|---|---|
| Null hypothesis | $H_0$: $p_1 - p_2 = 0$ |
| Alternative hypothesis | $H_1$: $p_1 - p_2 \neq 0$ |

| Method | Z-Value | P-Value |
|---|---|---|
| Normal approximation | -2.68 | 0.007 |
| Fisher's exact | | 0.008 |

*The pooled estimate of the proportion (0.393025) is used for the tests.*

### Method

$p_1$: proportion where Sample 1 = Event
$p_2$: proportion where Sample 2 = Event
Difference: $p_1 - p_2$

### Descriptive Statistics

| Sample | N | Event | Sample p |
|---|---|---|---|
| Sample 1 | 1002 | 360 | 0.359281 |
| Sample 2 | 1980 | 812 | 0.410101 |

### Estimation for Difference

| Difference | 95% CI for Difference |
|---|---|
| -0.0508196 | (-0.087588, -0.014052) |

*CI based on normal approximation*

# 11.6 Confidence Interval for the Difference Between Two Proportions (1 of 2)

How to develop a confidence interval for the difference between the proportions of two populations?

1) Take a sample of size $n_1$ from the first population.
2) Take a sample of size $n_2$ from the second population.
3) Calculate the standard error of proportion for the first sample.
4) Calculation the standard error of proportion for the second sample.
5) Calculate the square root of the sum of the variance of the two samples.

# 11.6 Confidence Interval for the Difference Between Two Proportions (2 of 2)

Use the formula below to calculate the confidence interval for the difference between two populations:

**Confidence Interval for the Difference Between Two Proportions**

The confidence interval for the difference between two proportions is $(\hat{p}_1 - \hat{p}_2) \pm z^* \times SE(\hat{p}_1 - \hat{p}_2)$,

where $z^*$ is the critical value and $SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}$.

## FOR EXAMPLE Any friends for the Alberta oil sands project?

Nanos Research's survey of opinion about the Alberta oil sands project showed a difference among the provinces for the percentage of people who were favourable or somewhat favourable toward the project. In the Atlantic provinces they surveyed 100 people and found the percentage to be 35.4%, whereas in Ontario they surveyed 300 people and found it to be 46.1%.

**QUESTION** Construct a 99% confidence interval for the difference in support for the Alberta oil sands project between the Atlantic provinces and Ontario. What does this confidence interval tell us about whether opinions differ on this issue?

**ANSWER**

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = \sqrt{\frac{0.354 \times 0.646}{100} + \frac{0.461 \times 0.539}{300}} = 0.0558$$

The critical value for a 99% confidence interval is $z^* = 2.58$. The confidence interval for the difference between two proportions is

$$(\hat{p}_1 - \hat{p}_2) \pm z \times SE(\hat{p}_1 - \hat{p}_2) = 0.354 - 0.461 \pm 2.58 \times 0.0558 = -0.107 \pm 0.144$$

i.e., between –0.251 and 0.037.

Since this interval includes zero, we can't conclude that there's a difference in the percentages of people who are favourable or somewhat favourable toward the Alberta oil sands project between the Atlantic provinces and Ontario at the 99% level.

# Minitab

- Select “two-proportions test” under “Basic Statistics”
  - No pooled variance (99% and 95% CIs)

## Test and CI for Two Proportions

### Method

$p_1$: proportion where Sample 1 = Event
$p_2$: proportion where Sample 2 = Event
Difference: $p_1 - p_2$

### Descriptive Statistics

| Sample | N | Event | Sample p |
|---|---|---|---|
| Sample 1 | 100 | 35 | 0.350000 |
| Sample 2 | 300 | 138 | 0.460000 |

### Estimation for Difference

| Difference | 99% CI for Difference |
|---|---|
| -0.11 | (-0.253485, 0.033485) |

*CI based on normal approximation*

## Test and CI for Two Proportions

### Method

$p_1$: proportion where Sample 1 = Event
$p_2$: proportion where Sample 2 = Event
Difference: $p_1 - p_2$

### Descriptive Statistics

| Sample | N | Event | Sample p |
|---|---|---|---|
| Sample 1 | 100 | 35 | 0.350000 |
| Sample 2 | 300 | 138 | 0.460000 |

### Estimation for Difference

| Difference | 95% CI for Difference |
|---|---|
| -0.11 | (-0.219179, -0.000821) |

*CI based on normal approximation*

# 12.9 Two Types of Errors (1 of 4)

We can make mistakes in *two* ways:

I. (False Hypothesis) The null hypothesis is true, but we mistakenly reject it.

II. (False Negative) The null hypothesis is false, but we fail to reject it.

These two types of errors are known as *Type I* and *Type II errors*.

| Name | Also Known As: | Probability | Statistical Terminology | Business Example |
| --- | --- | --- | --- | --- |
| Type I error | False positive | $\alpha$ | Reject a true null hypothesis | Invest in a project that is not successful |
| Type II error | False negative | $\beta$ | Fail to reject a false null hypothesis | Fail to invest in a project that would have been successful |

# 12.9 Two Types of Errors (2 of 4)

Here's an illustration of the situations:

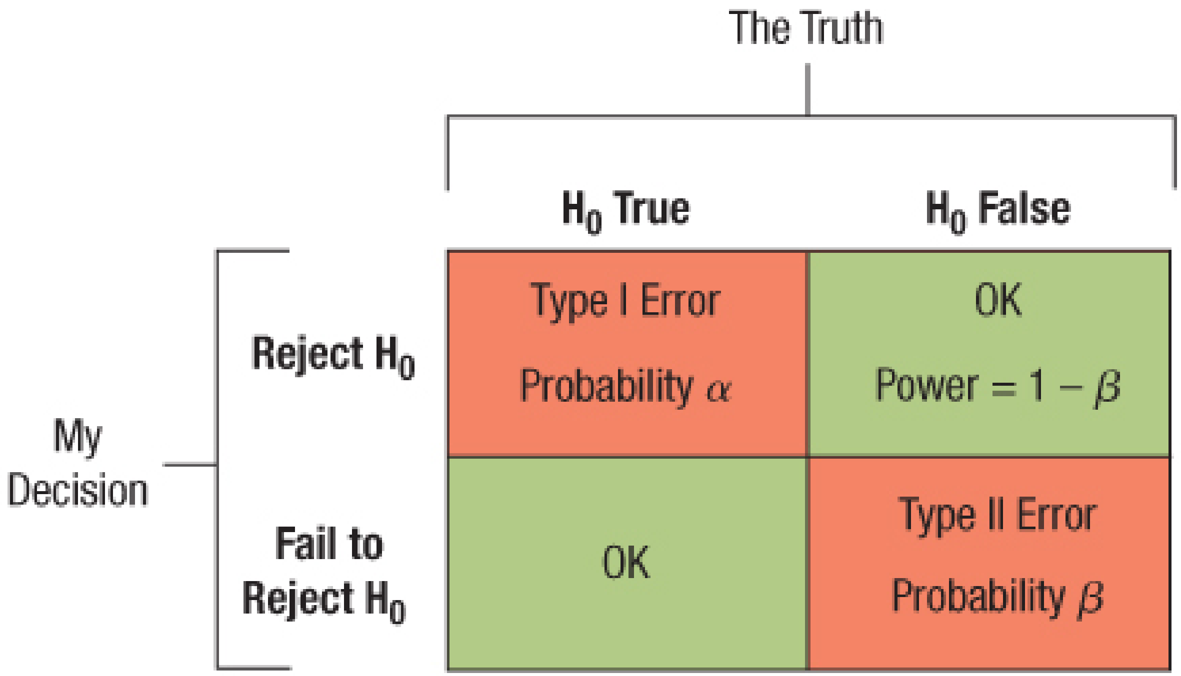



**Figure 12.8** The two types of errors occur on the diagonal, where the truth and decision don't match. Remember that we start by assuming $H_0$ to be true, so an error made (rejecting it) when $H_0$ is true is called a Type I error. A Type II error is made when $H_0$ is false (and we fail to reject it).

# 12.9 Two Types of Errors (3 of 4)

When you choose an $\alpha$ level, you're setting the probability of a Type I error to $\alpha$.

A test's ability to detect a false hypothesis is called the *power* of the test.

The null hypothesis specifies a single value for the parameter.
So it's easy to calculate the probability of a Type I error. But the alternative gives a whole range of possible values, and we may want to find a $\beta$ for several of them.

We assign the letter $\beta$ to the probability of a Type II error.

The choice of which $\beta$ to pick is not always clear. One way to focus our attention is by thinking about the *effect size.* That is, ask: "How big a difference would matter?"

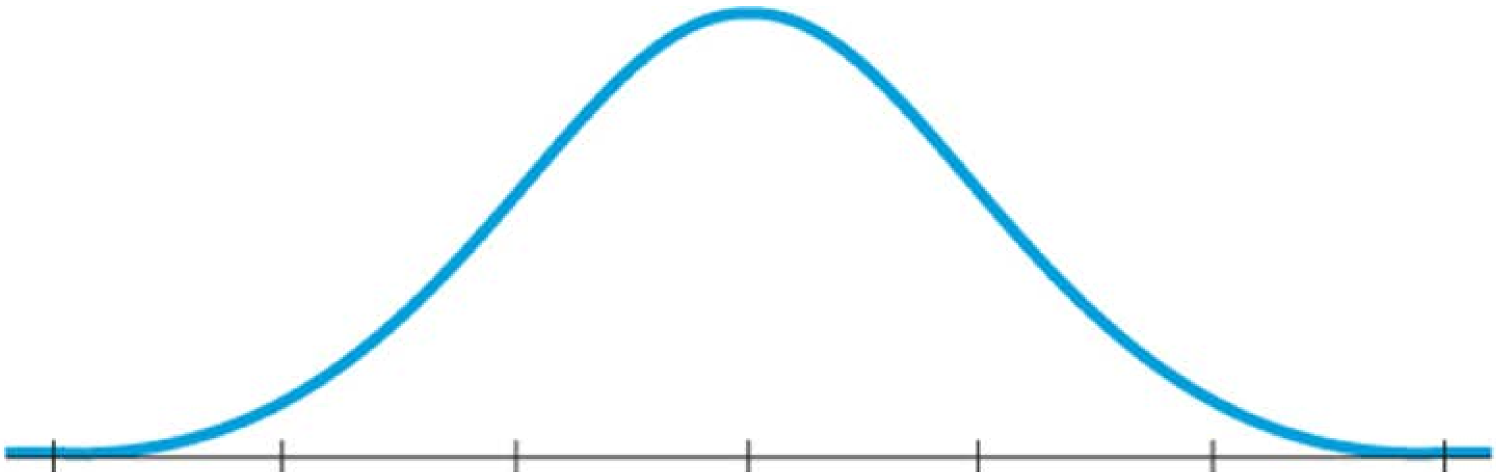
# 12.9 Two Types of Errors (4 of 4)

The only way to reduce *both* types of error is to collect more data.

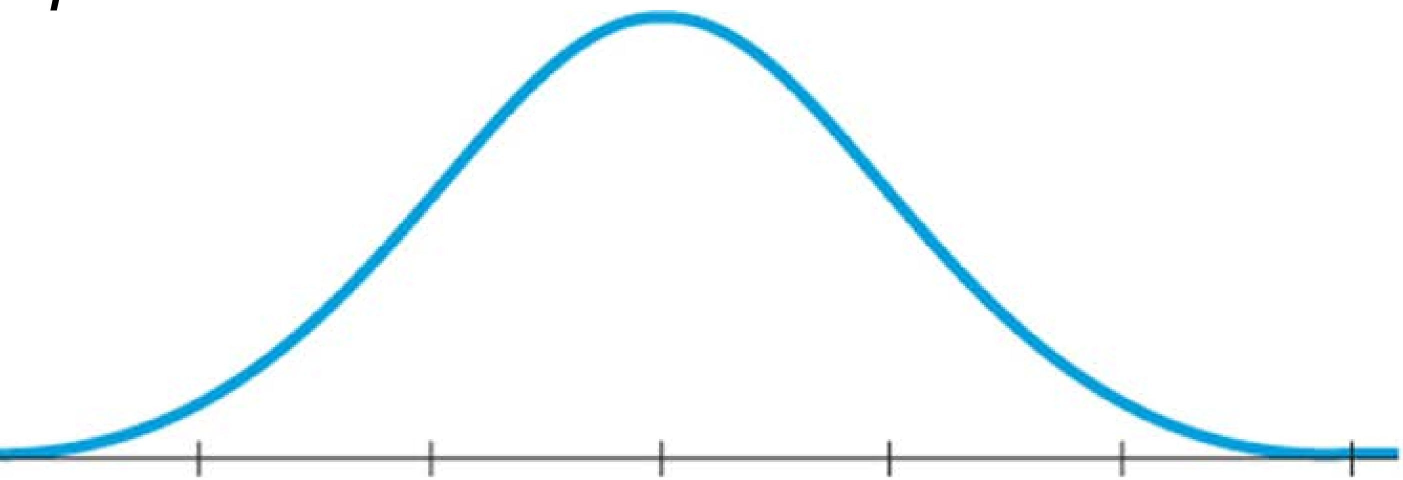
Whenever you design a survey or experiment, it’s a good idea to calculate $\beta$ (for a reasonable $\alpha$ level).

Use a parameter value in the alternative that corresponds to an effect size that you want to be able to detect.

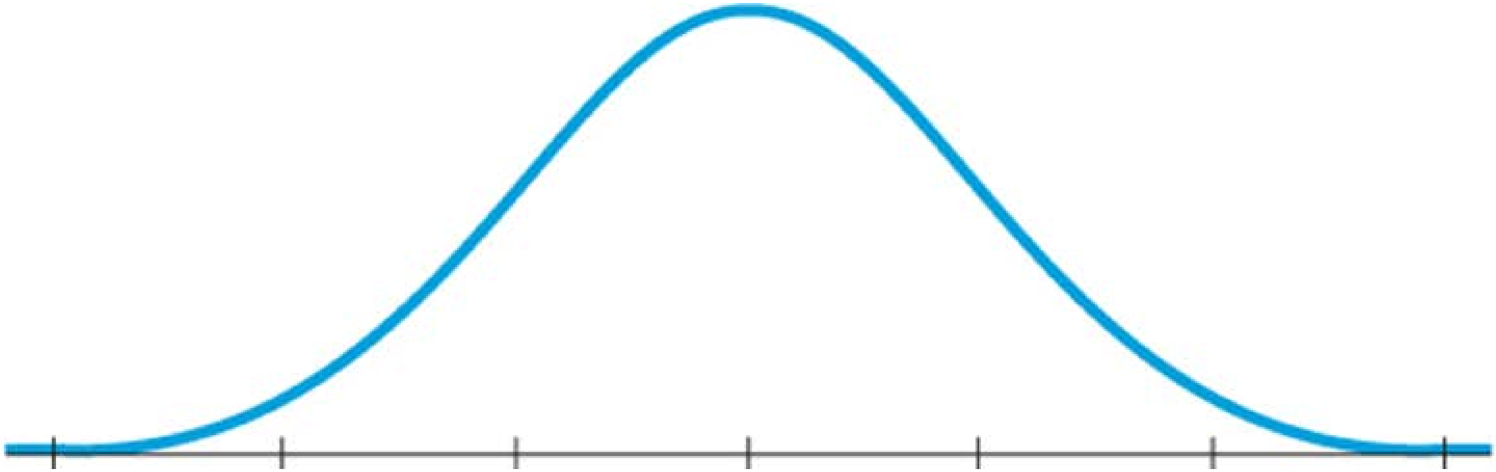
*α*



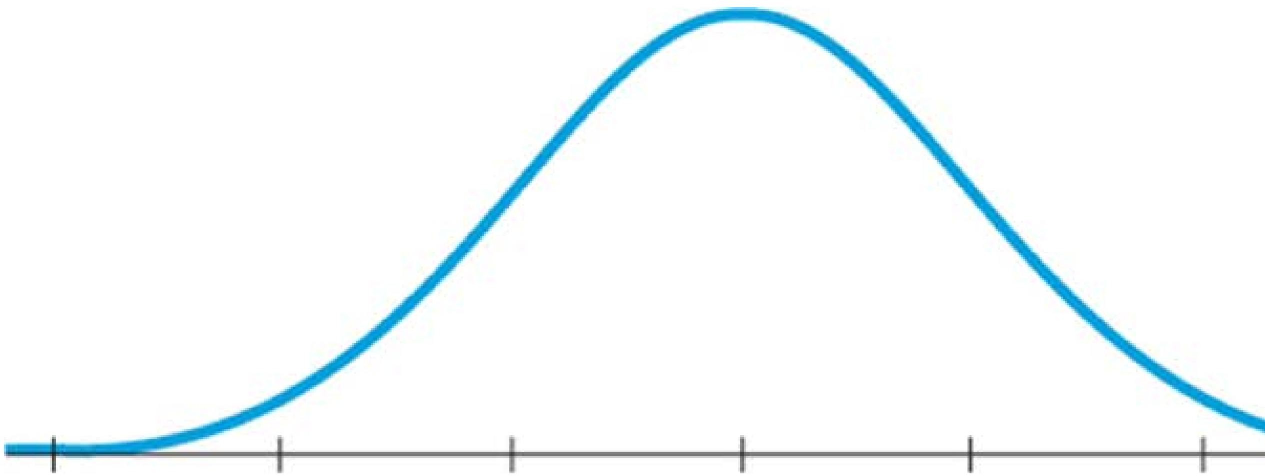
*β*



*α*



*β*



## FOR EXAMPLE Type I and Type II errors for website visits

**QUESTION** Suppose that a year later, a full accounting of all the SmartWool transactions (see For Example: "The reasoning of hypothesis tests about website customers") finds that 26.5% of visits resulted in sales. Have any errors been made?

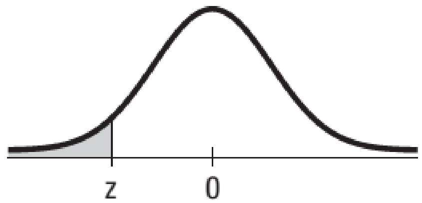
**ANSWER** We rejected the null hypothesis that $p = 0.20$ and in fact $p = 0.265$, so we did not make a Type I error (the only error we could have made when rejecting the null hypothesis).

## FOR EXAMPLE Type I and Type II errors for quality control

A company manufactures ceramic ball bearings, but customers complain that some of them are cracked. Some cracking is inevitable and the company accepts 0.2% cracked product. A random sample of ball bearings is collected from the manufacturing plant and tested to determine whether more than 0.2% are cracked. The P-value comes to 0.075, so management concludes that there is no cracking problem and continues to operate the plant as in the past. A month later, a major customer complains that 1.1% of the ball bearings in the last shipment were cracked.

**QUESTION** Did the company make a Type I or Type II error? How does this example illustrate the importance of Type I and/or Type II errors? Is it possible to make a Type I error and a Type II error at the same time?

**ANSWER** The null hypothesis is that 0.2% are cracked and they failed to reject this hypothesis even though 1.1% are cracked. This is a Type II error. Making this error has caused a major customer to complain about the quality of the product. Type I errors are made when the null hypothesis is true. Type II errors are made when the null hypothesis is false. It is therefore not possible to make both types of error at the same time.




## Table Z: The Normal Distribution

*Areas under the standard Normal curve. Find the row corresponding to your value of z to one decimal position. Find the column corresponding to the second decimal position of z. The probability of being less than z is given in the cell for that row and column.*

| *Second decimal place in z* | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ***0.09*** | ***0.08*** | ***0.07*** | ***0.06*** | ***0.05*** | ***0.04*** | ***0.03*** | ***0.02*** | ***0.01*** | ***0.00*** | ***z*** |
| 0.1611 | 0.1635 | 0.1660 | 0.1685 | 0.1711 | 0.1736 | 0.1762 | 0.1788 | 0.1814 | 0.1841 | *–0.9* |
| 0.1867 | 0.1894 | 0.1922 | 0.1949 | 0.1977 | 0.2005 | 0.2033 | 0.2061 | 0.2090 | 0.2119 | *–0.8* |
| 0.2148 | 0.2177 | 0.2206 | 0.2236 | 0.2266 | 0.2296 | 0.2327 | 0.2358 | 0.2389 | 0.2420 | *–0.7* |
| 0.2451 | 0.2483 | 0.2514 | 0.2546 | 0.2578 | 0.2611 | 0.2643 | 0.2676 | 0.2709 | 0.2743 | *–0.6* |
| 0.2776 | 0.2810 | 0.2843 | 0.2877 | 0.2912 | 0.2946 | 0.2981 | 0.3015 | 0.3050 | 0.3085 | *–0.5* |
| 0.3121 | 0.3156 | 0.3192 | 0.3228 | 0.3264 | 0.3300 | 0.3336 | 0.3372 | 0.3409 | 0.3446 | *–0.4* |
| 03483 | 0.3520 | 0.3557 | 0.3594 | 0.3632 | 0.3669 | 0.3707 | 0.3745 | 0.3783 | 0.3821 | *–0.3* |
| 0.3859 | 0.3897 | 0.3936 | 0.3974 | 0.4013 | 0.4052 | 0.4090 | 0.4129 | 0.4168 | 0.4207 | *–0.2* |
| 0.4247 | 0.4286 | 0.4325 | 0.4364 | 0.4404 | 0.4443 | 0.4483 | 0.4522 | 0.4562 | 0.4602 | *–0.1* |
| 0.4641 | 0.4681 | 0.4721 | 0.4761 | 0.4801 | 0.4840 | 0.4880 | 0.4920 | 0.4960 | 0.5000 | *–0.0* |

†For $z \leq -3.90$, the areas are 0.0000 to four decimal places.

# What Can Go Wrong? (1 of 2)

- Don’t base your null hypotheses on what you see in the data.
- Don’t base your alternative hypothesis on the data either.
- Don’t make your null hypothesis what you want to show to be true.
- Don’t forget to check the conditions.

# What Can Go Wrong? (2 of 2)

- Don’t believe too strongly in arbitrary alpha levels.
- Don’t confuse practical and statistical significance.
- Don’t forget that despite all your care, you might make a wrong decision.

# What Have We Learned? (1 of 4)

- We've learned to use what we see in a random sample to test a particular hypothesis about the world.
- We've learned that testing a hypothesis involves proposing a model and then seeing whether the data we observe are consistent with that model or so unusual that we must reject it.

# What Have We Learned? (2 of 4)

- We start with a *null hypothesis* specifying the parameter of a model we'll test using our data.
- Our *alternative hypothesis* can be one- or two-sided, depending on what we want to learn.
- We must check the appropriate *assumptions* and *conditions* before proceeding with our test.
- The *significance level* of the test establishes the level of proof we'll require.

# What Have We Learned? (3 of 4)

- The *hypothesis test* gives us the answer to a decision about a parameter; the *confidence interval* tells us the plausible values of that parameter.
- If the null hypothesis is really true and we reject it, that's a *Type I error*; the alpha level of the test is the probability that this happens.
- If the null hypothesis is really false but we fail to reject it, that's a *Type II error.*

# What Have We Learned? (4 of 4)

- The *power* of the test is the probability that we reject the null hypothesis when it's false. The larger the size of the effect we're testing for, the greater the power of the test in detecting it.
- If we have independent samples from two different populations, we construct a hypothesis test to compare the two populations with each other.