ANNOUNCEMENTS:

We will be covering chapters out of order for the rest of the quarter.

Today Sections 12.1 to 12.3, except 12.2 Lesson 3, then (see website)**:**

- More of Chapter 12 (hypothesis tests for proportions)
- Finish Chapter 9 (sampling distributions for means)
- Chapter 11 (confidence intervals for means)
- Start Chapter 13 (hypothesis tests for means)
- Section 15.3 (chi-square hypothesis test for goodness of fit)
- Sections 16.1 and 16.2 ("analysis of variance")*
- Finish Chapters 12 and 13 and cover Chapter 17*
- Last day: Review for final exam*

**No clicker, quiz or homework due this week.*

- No discussion on Monday (Nov 15); 2 more for credit after that.

HOMEWORK: (Due Friday, Nov 19)

Chapter 12: #15b and 16b (counts as 1), 19, 85 (counts double)

Reminder from when we started Chapter 9

Five situations we will cover for the rest of this quarter:

Parameter name and description	Population parameter	Sample statistic
For Categorical Variables:		
One population proportion (or probability)	p	$\hat{p}$
Difference in two population proportions	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$
For Quantitative Variables:		
One population mean	μ	$\bar{x}$
Population mean of paired differences (dependent samples, paired)	μ_d	$\bar{d}$
Difference in two population means (independent samples)	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$

For each situation will we:

√ Learn about the *sampling distribution* for the sample statistic

√ Learn how to find a *confidence interval* for the true value of the parameter

- *Test hypotheses* about the true value of the parameter

Review: Standardized Statistics

$$z = \frac{\text{sample statistic - population parameter}}{\text{s.d.(sample statistic)}}$$

Replacing s.d. with s.e. (standard error) in denominator:

Standardized statistic is still z for situations with proportions.

For means, standardized statistic is t, not z (more Mon).

Summary of sampling distributions for the 5 parameters (p. 382):

	Parameter	Statistic	Standard Deviation of the Statistic	Standard Error of the Statistic	Standardized Statistic with s.e.
One proportion	p	$\hat{p}$	$\sqrt{\frac{p(1-p)}{n}}$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	z
Difference Between Proportions	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$	$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	z
One Mean	μ	$\bar{x}$	$\frac{\sigma}{\sqrt{n}}$	$\frac{s}{\sqrt{n}}$	t
Mean Difference, Paired Data	μ_d	$\bar{d}$	$\frac{\sigma_d}{\sqrt{n}}$	$\frac{s_d}{\sqrt{n}}$	t
Difference Between Means	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$	$\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$	t

Chapter 12: Significance testing = hypothesis testing.

Saw this already in Chapter 6.

Today: Hypothesis tests for *one proportion*

Example:

Gallup poll of adult Canadians on same-sex marriage

n = 1003; 481 of the 1003 = 48% *oppose* same-sex marriage.

News story said "For the first time in the 7 years Gallup has asked the question, *less than half (48%) of Canadians* oppose same-sex marriages."

Question: Is the media quote correct about the population?

- The *sample proportion* based on n = 1003 was $\hat{p}$ = .48.
- How unlikely is that *if* the *population proportion p* is *.50 or more*, i.e. if the true p is ≥ .50?
- In other words, is this convincing evidence that *less than half* of the Canadian *population* opposed same-sex marriage? I.e. that p < .50?

Five Steps to Hypothesis Testing, and how they apply in general (for our 5 situations), for one proportion, and for our example.

STEP 1: Determine the null and alternative hypotheses.

General: *Alternative hypothesis* is usually what researchers hope to show. "Null is dull." Null is often status quo, a specific value. For two proportions or two means, null is usually "no difference."

For One Proportion: p = **population** proportion

Null hypothesis is H_0: p = p_0 (a specified value)

The "equal sign" always goes in the null hypothesis.

Alternative hypothesis is one of these, based on context:

H_a: $p \neq p_0$ (a two-sided hypothesis)

H_a: $p > p_0$ (a one-sided hypothesis)

H_a: $p < p_0$ (a one-sided hypothesis)

EXAMPLE:

Want to know if *less than half of all Canadians* oppose same-sex marriage, as media claimed. That is the *alternative hypothesis.*

H_0: p = 0.5 (or could be written as $p \geq 0.5$; data do not support media claim)

H_a: p < 0.5 (the media quote is correct, less than half of *all* Canadians oppose same sex marriages)

The *null value* is $p_0 = 0.5$. This is a *one-sided* test.

Remember how hypothesis testing works:

- *Assume* the null hypothesis is true about the population
- Compute what we *expect* if that's the case
- Compare it to what we *observed in the sample*
- For our 5 situations, we do this using standardized statistics

STEP 2:

Verify data conditions. If met, summarize data into a test statistic.

General:

For our five situations, test statistic will be **z** (for situations with proportions) or **t** (for situations with means) =

$$\frac{\text{sample statistic} - \text{null value}}{\text{(null) standard error}}$$

For One Proportion:

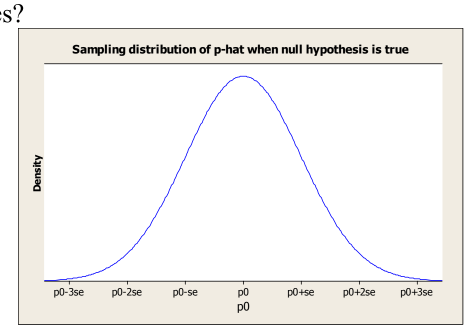
Data conditions: np_0 and $n(1-p_0)$ are both at least 10.

For the test statistic the *null value* is p_0 and *null standard error* is the usual s.d.($\hat{p}$) with *null value* p_0 in place of p:

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

Rationale: Compare *observed* to *expected if null is true*

If p_0 really is the true population proportion, where does the *observed value* of $\hat{p}$ fall in the sampling distribution of possibilities?

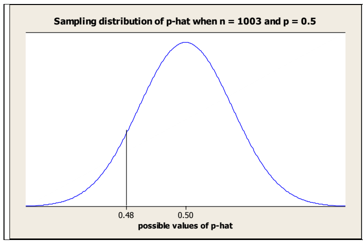


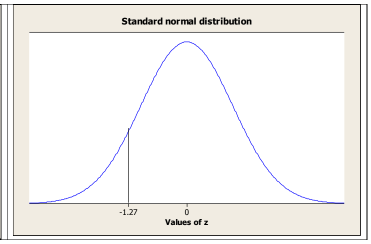
Test statistic is the z-score for the location of the observed $\hat{p}$ in this picture. If the null hypothesis is true, test statistic is standard normal.

Example (Canadian poll):
Data conditions are met: np_0 and $n(1 - p_0)$ = (1003)(0.5) = 501.5

Null standard error is $\sqrt{\frac{(.5)(.5)}{1003}}$ = .0158; test statistic is:

$$\frac{\text{Sample statistic - null value}}{\text{(null) standard error}} = \frac{.48 - .50}{.0158} = -1.27$$





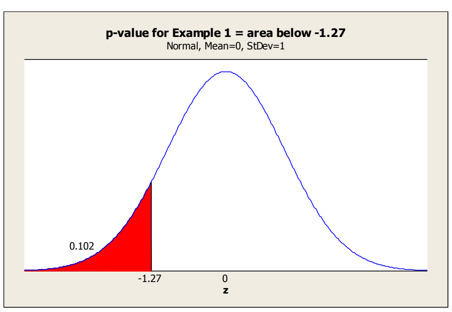
STEP 3:
Assuming the null hypothesis is true, find the p-value.

General: *p*-value = the probability of a test statistic as extreme as the one observed or more so, in the direction of H_a, *if* the null hypothesis is true.

One proportion:
Depends on the alternative hypothesis. See pictures on board and on p. **517**

Alternative hypothesis:	p-value is:
H_a: $p > p_0$ (a one-sided hypothesis)	Area above the test statistic z
H_a: $p < p_0$ (a one-sided hypothesis)	Area below the test statistic z
H_a: $p \neq p_0$ (a two-sided hypothesis)	2 × the area above |z| = area in tails beyond −z and z

Example:
Alternative hypothesis is one-sided
H_a: p < 0.5
p-value = Area below the test statistic z = –1.27
From Table A.1, *p*-value = .1020



STEP 4:
Decide whether or not the result is statistically significant based on the p-value. Two possible conclusions.

If p-value ≤ ***level of significance*** (usually .05), these are equivalent ways to state the conclusion:
- Reject the null hypothesis.
- Accept the alternative hypothesis
- The result is statistically significant

If p-value > ***level of significance*** (usually .05), these are equivalent ways to state the conclusion:
- Do not reject the null hypothesis
- There is not enough evidence to support the alternative hypothesis.
- The result is not statistically significant.

Example:
p-value = .1020 > .05, so our conclusion can be stated as:

- Do *not* reject the null hypothesis.
- There is not enough evidence to support the alternative hypothesis.
- The result is *not* statistically significant.

Step 5: Report the conclusion in the context of the situation.

Example:
We could not reject the null hypothesis. There was not enough evidence to accept the alternative hypothesis. The proportion of adult Canadians who opposed same-sex marriage when poll was taken is *not* statistically significantly less than half. So the media quote was not justified. (Recall, quote was "For the first time in the 7 years Gallup has asked the question, *less than half (48%) of Canadians* oppose same-sex marriages.")

TECHNICAL NOTE

Often in hypothesis testing the standardized score will be "off the charts." The bottom of Table A.1 has values for "In the extreme." Here are some of them for the lower tail:

z	area below z
-3.09	.001
-4.75	.000001
-6.00	.000000001

Example: Suppose the test statistic is z = –5.00
For Ha: $p < p_0$ we would say the *p*-value < .000001.
For Ha: $p \neq p_0$ we would say the *p*-value < .000002.

EXAMPLE 2: According to the CDC website, 59% of 18 to 25 year-olds in the United States drink alcohol.

QUESTION: Is the percent different for UC Davis students? (Note that before looking at data, we have no reason to ask if it's in one direction or the other, just is it *different* from national proportion.)

Define the population parameter:
p = proportion of *all* UC Davis students who drink alcohol at least once a month.

Data: Sample of n = 405 students in Introductory Statistics.

We are assuming that the Statistics students are representative of all UC Davis students for this question.

STEP 1: Determine the null and alternative hypotheses.

For One Proportion: p = **population** proportion
Null hypothesis is H_0: $p = p_0$ (a specified value)
Alternative hypothesis is one of these, based on context:
H_a: $p \neq p_0$ (a two-sided hypothesis)
H_a: $p > p_0$ (a one-sided hypothesis)
H_a: $p < p_0$ (a one-sided hypothesis)

Example:
Is UC Davis proportion who drink *different* from national proportion of 0.59?
H_0: p = 0.59 (UCD students are same as US population.)
H_a: p ≠ 0.59 (UCD students are different from US population.)

The *null value* is $p_0 = 0.59$. This is a *two-sided* test.

STEP 2:
Verify data conditions. If met, summarize data into test statistic.

For One Proportion:
Data conditions: np_0 and $n(1 - p_0)$ are both at least 10.
The condition is met; $n\ (1 - p_0) = 405(.41) = 166.05$

Test statistic: In the sample, 212 students drink, so

$$\hat{p} = \frac{212}{405} = .5235,\ \ p_0 = .59,\ \text{null standard error} = \sqrt{\frac{(.59)(1-.59)}{405}} = .0244$$

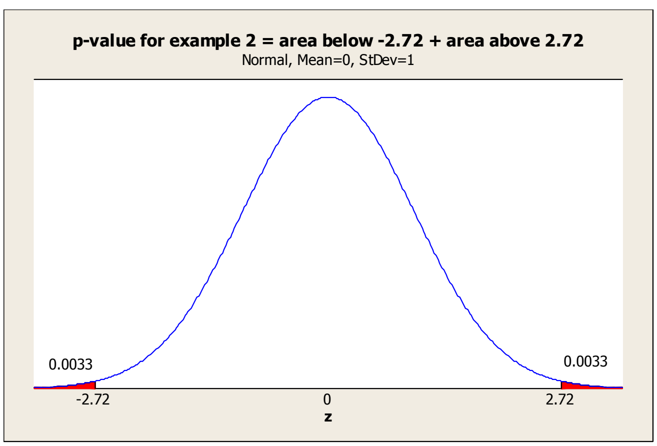
$$z = \frac{\text{sample statistic - null value}}{\text{(null) standard error}} = \frac{.5235 - .59}{.0244} = -2.72$$

STEP 3:
Assuming the null hypothesis is true, find the p-value.

Alternative hypothesis:	p-value is:
H_a: $p > p_0$ (a one-sided hypothesis)	Area above the test statistic z
H_a: $p < p_0$ (a one-sided hypothesis)	Area below the test statistic z
H_a: $p \neq p_0$ (a two-sided hypothesis)	2 × the area above |z| = area in tails beyond −z and z

Example: In this case, H_a: $p \neq 0.59$
p-value = 2 × the area above |z| = area in tails beyond −z and z
2 × the area above |-2.72| = area in tails beyond -2.72 and 2.72

From Table A.1, *p*-value = 2 × .0033 = .0066



STEP 4:
Decide whether or not the result is statistically significant based on the p-value.

Example: *p*-value = .0066
If p-value ≤ ***level of significance*** (usually .05), these are equivalent ways to state the conclusion:

- Reject the null hypothesis.
- Accept the alternative hypothesis
- The result is statistically significant

Step 5: Report the conclusion in the context of the situation.
Example: There is a statistically significant difference between the proportion of the *population* of UC Davis students who drink alcohol and the proportion of 18 to 25 year-olds in the national population who do so.

How much different are UCD students from national?

For two-sided tests, often a good idea to follow up with a confidence interval:

Example:
A 95% confidence interval for the UC Davis proportion is

$$.5235 \pm 2(.025)$$
$$.5235 \pm .05$$
$$.4735 \text{ to } .5735$$

So, we are 95% confident that between 47.35% and 57.35% of *all* UC Davis students drink alcohol. Notice that the entire interval is below the national value of .59.

Correspondence between 2-sided test and 95% CI:

Reject H_0: $p=p_0$ at .05 if p_0 is not in a 95% confidence interval.

Example: 0.59 is *not* in the 95% confidence interval, so we can reject 0.59 as a likely value for the true proportion of UC Davis students who drink alcohol.

Section 12.2, Lesson 3
What Can Go Wrong in Hypothesis Testing: The Two Types of Errors and Their Probabilities

Example: Case Study 1.6, aspirin and heart attacks. Found statistically significant relationship; *p-value* was < .00001.

Possible errors:

Type 1 error (*false positive*) occurs when:
- Null hypothesis is actually true, but
- Conclusion of test is to Reject H_0 and accept H_a

Type 2 error (*false negative*) occurs when:
- Alternative hypothesis is actually true, but
- Conclusion is that we cannot reject H_0

Heart attack and aspirin example:
Null hypothesis: The proportion of men who would have heart attacks if taking aspirin = the proportion of men who would have heart attacks if taking placebo.

Alternative hypothesis: The heart attack proportion is lower if men were to take aspirin than if they were not to take aspirin.

Type 1 error (*false positive*): Occurs if there really is *no relationship* between taking aspirin and heart attack prevention, but we conclude that there *is* a relationship.

Consequence: Good for aspirin companies! Possible bad side effects from aspirin, with no redeeming value.

Type 2 error (*false negative*): Occurs if there really *is* a relationship but the study *failed to find it*.

Consequence: Miss out on recommending something that could save lives!

Which type of error is more serious? Probably all agree that Type 2 is more serious.

Which could have occurred?
Type 1 error could have occurred. Type 2 could not have occurred, because we *did* find a significant relationship.

Aspirin Example: Consequences of the decisions

	Decision:		
Truth:	**Don't conclude aspirin works**	**Reject H_0, Conclude aspirin works**	**When error can occur**:
H_0 Aspirin doesn't work	OK	Type 1 error: People take aspirin needlessly; may suffer side effects	Type 1 error can *only* occur if aspirin doesn't work.
Ha Aspirin works	Type 2 error: Aspirin could save lives but we don't recognize its benefits	OK	Type 2 error can *only* occur if aspirin does work.

Note that because H_0 *was rejected* in this study, we could only have made a Type 1 error, not a Type 2 error.

Some analogies to hypothesis testing:

Analogy 1: Courtroom:
Null hypothesis: Defendant is innocent.
Alternative hypothesis: Defendant is guilty

Note that the two possible conclusions are "not guilty" and "guilty." The conclusion "not guilty" is equivalent to "don't reject H_0." We don't say defendant is "innocent" just like we don't accept H_0 in hypothesis testing.

Type 1 error is when defendant is *innocent* but *gets convicted*
Type 2 error is when defendant is *guilty* but *does not get convicted*.

Which one is more serious??

Analogy 2: Medical test
Null hypothesis: You do not have the disease
Alternative hypothesis: You have the disease

Type 1 error: You *don't* have disease, but test says *you do*; a "false positive"
Type 2 error: You *do* have disease, but test says you do not; a "false negative"

Which is more serious??

NEXT TIME: Probabilities associated with these errors, and how they relate to the level of significance.