Goals for the Week
Hypotheses
Decision Errors
p-values
Making Decisions
Tests and CIs
One- and Two-Sided Tests
Notation Review

STAT 140: Week 9

My hope is that after Week 9 (10/25-10/29), you will be able to:

My hope is that after this week, you will be able to:

- Discuss the framework for hypothesis testing, including:
  - Appropriate null and alternative hypotheses,
  - Standard errors, test statistics, and p-values, and
  - Significance levels and Type I and II errors.
- Compare hypothesis testing to confidence intervals.
- Carry out a one-proportion z-test.

Reviewing Sampling Distributions

- Remember that probability distributions, including the sampling distribution, describe the theoretical behavior of a random variable.
- Now that we know that every statistic has a distribution, we can understand the theoretical behavior of the statistic.

This means we can...

- Understand what kinds of values are reasonable (*Confidence Intervals!*), and
- Understand how likely a statistic is to occur (*Hypothesis Tests!*)

Hypothesis Testing Framework 1

In general–what is a hypothesis?

- Idea, Theory, Conjecture, Speculation
- Merriam-Webster's Dictionary defines a hypothesis as "an assumption, an idea that is proposed for the sake of argument so that it can be tested to see if it might be true".

Hypothesis Testing Framework 2

Goals for the Week
Hypotheses
Decision Errors
p-values
Making Decisions
Tests and CIs
One- and Two-Sided Tests
Notation Review

Hypotheses: A set of competing ideas to be tested.

- **Null Hypothesis**
  - Represented by H_0
  - The **null hypothesis** represents a claim to be tested.
    - Sometimes, we think about the null hypothesis as the conventional belief.
    - In general, we treat the null hypothesis skeptically.
- **Alternative Hypothesis**
  - Represented by H_A
  - The alternative hypothesis represents an alternative claim under consideration.
  - The alternative hypothesis can be thought of as a new range of plausible parameter values.

Hypothesis Testing Framework 3

- The hypothesis testing framework is a very general tool, and we often use it without a second thought.
- If a person makes a somewhat unbelievable claim, we are initially skeptical.
- However, if there is sufficient evidence that supports the claim, we set aside our skepticism and reject the null hypothesis in favor of the alternative.

Claims in the US Court System 1

- In the US, a defendant can either be innocent or guilty.
- What are the null and alternative hypotheses?

Claims in the US Court System 2

Goals for the Week
Hypotheses
Decision Errors
p-values
Making Decisions
Tests and CIs
One- and Two-Sided Tests
Notation Review

- In the US, a defendant can either be innocent or guilty.

$$H_0: \text{ The defendant is innocent.}$$

$$H_A: \text{ The defendant is guilty.}$$

Jurors examine the evidence to see whether it convincingly shows a defendant is guilty.

- Even if the jurors leave *unconvinced of guilt beyond a reasonable doubt*, this does not mean they believe the defendant is innocent.
- This is also the case with hypothesis testing! *Even if we fail to reject the null hypothesis, we typically do not accept the null hypothesis as true.*

Decision Errors 1

Goals for the Week
Hypotheses
Decision Errors
p-values
Making Decisions
Tests and CIs
One- and Two-Sided Tests
Notation Review

Possible outcomes of a hypothesis test:

- Reject the null hypothesis in favor of the alternative hypothesis.
- Fail to reject the null hypothesis.
  - Double negatives can occur in statistics! For instance, we might say that the null hypothesis is not implausible.

Decision	H_0 True	H_0 False
Reject	Wrong!	Correct
Fail to Reject	Correct	Wrong!

Decision Errors 2

Goals for the Week
Hypotheses
Decision Errors
p-values
Making Decisions
Tests and CIs
One- and Two-Sided Tests
Notation Review

Decision	H_0 True	H_0 False
Reject	Type I	Correct
Fail to Reject	Correct	Type II

- **Type I Error**: When we reject the null hypothesis when it is actually true, we have made a **Type I** error.
  - Type I: Jump the gun!
- **Type II Error**: When we fail to reject the null hypothesis when the alternative is actually true, we have made a **Type II** error.
  - Type II: Avoid the new.

Decision Errors 3

- We have the ability to control how frequently we make Type I errors!
- If the null hypothesis is actually true, the **significance level**, represented by α, indicates how often the data lead us to incorrectly reject H_0.
- The traditional α is equal to 0.05.
- We may select a level that is smaller ($\alpha = 0.01$) or larger ($\alpha = 0.10$) than 0.05, depending on the consequences of the test's conclusion.
  - If making a Type I error is dangerous or costly, we choose a small α.
  - If making a Type 2 error is relatively more dangerous or costly, we might choose a higher α.

Formal Testing Using p-values 1

Goals for the Week
Hypotheses
Decision Errors
p-values
Making Decisions
Tests and CIs
One- and Two-Sided Tests
Notation Review

- One way to quantify the strength of evidence against the null hypothesis and in favor of the alternative hypothesis is a **p-value**.
- Formally, the **p-value** is the probability of observing a statistic as or more extreme than the current value, given that the null hypothesis is true.
- To calculate a probability, such as the p-value, you need to assume a distribution.
- The sampling distribution under the null hypothesis is known as the **null distribution**.
- Where do we get the sampling distribution from?

Formal Testing Using p-values 2

- The Central Limit Theorem!
- When observations are independent and the sample size is sufficiently large, the sample proportion $\hat{p}$ will tend to follow a normal distribution with mean

$$\mu_{\hat{p}} = p$$

and standard error

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}.$$

Formal Testing Using p-values 3

- **Null Value**: The value of the parameter given in the null hypothesis.
- Example:
  - Consider a multiple choice test where questions have responses a, b, c, and d.
  - I would expect randomly guessing students to answer 25% of the questions correctly.
  - Then, $p_0 = 0.25$.
- Example: If the probability of correctly guessing the answer to a multiple choice question with four options is 0.25, the proportion of correct guesses for a class of 18 people is normally distributed with a mean of 0.25 and a standard error of

$$\sqrt{\frac{0.25 \times (1 - 0.25)}{18}} = 0.10.$$

Formal Testing Using p-values 4

Goals for the Week
Hypotheses
Decision Errors
p-values
Making Decisions
Tests and CIs
One- and Two-Sided Tests
Notation Review

- Hypothesis testing is procedurally different from confidence intervals–instead of using $\hat{p}$ to determine the sampling distribution and check the conditions, we use p_0.
- This is because a hypothesis test is carried out assuming the null is true!

Practicing p-values 1

- Someone once told me that 64% of American adults drink coffee every day. If we sample 225 adults and find that 150 of them drink coffee every day, should we believe differently?

$$H_0 : p = 0.64$$

$$H_A : p \neq 0.64$$

- Conditions:
  - We assume the data are independent.
  - $np_0 \geq 10$, and
  - $n(1 - p_0) \geq 10$.
- What are the null value and sampling distributions in this example?

Practicing p-values 2

- The null value is $p_0 = 0.64$. Then, the sampling distribution for $\hat{p}$ is

$$N(0.64, \sqrt{\frac{0.64 \times (1 - 0.64)}{225}}), \text{ or}$$

$$N(0.64, 0.113).$$

Practicing p-values 3

- To find the p-value, we first compute a **test statistic**.
- The test statistic for a one-proportion hypothesis test is

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

- If $\hat{p}$ is a value of interest, p_0 is a mean, and $\sqrt{\frac{p_0 \times (1 - p_0)}{n}}$ is a standard error, what is the test statistic here?

Practicing p-values 4

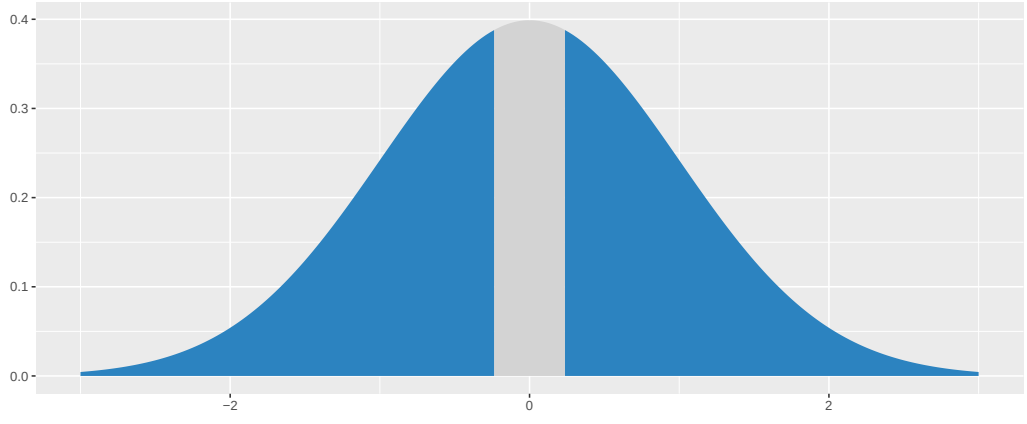
$$Z = \frac{\frac{150}{225} - 0.64}{\sqrt{\frac{0.64 \times (1-0.64)}{225}}} = 0.2357$$

- The p-value is the probability of observing anything more than 0.2357 standard deviations away from the mean.
  - This means anything greater than 0.2357, AND
  - Anything less than -0.2357!

```
pnorm(-0.2357, 0, 1) + pnorm(0.2357, 0, 1, lower.tail = FALSE)
```

```
## [1] 0.8136655
```

Practicing p-values 5



Formal Testing Using p-values 5

Goals for the Week
Hypotheses
Decision Errors
p-values
Making Decisions
Tests and CIs
One- and Two-Sided Tests
Notation Review

- The next step is to compare the p-value to the significance level, α.
  - If the p-value is greater than α, we fail to reject the null hypothesis. We would report that we do not have sufficient evidence to reject the null hypothesis.
  - If the p-value is less than α, we reject the null hypothesis. We would report that the data provide strong evidence supporting the alternative hypothesis.
- We have a p-value of 0.814 and $\alpha = 0.05$. What do we do?

Formal Testing Using p-values 6

- Since $0.814 > 0.05$, we fail to reject the null hypothesis.
- There is not enough evidence to suggest that the true proportion of American adults who drink coffee every day is not equal to 0.64.
- **There is not enough evidence to suggest that the true proportion of American adults who drink coffee every day is statistically significantly different from 0.64.**

Practicing p-values 6


```
## Always use correct = F!
prop.test(x = 150, n = 225, p = 0.64, correct = F)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  150 out of 225, null probability 0.64
## X-squared = 0.69444, df = 1, p-value = 0.4047
## alternative hypothesis: true p is not equal to 0.64
## 95 percent confidence interval:
##  0.6027283 0.7250095
## sample estimates:
##         p
## 0.6666667
```

Hypothesis Tests and Confidence Intervals 1

- In general, the results of a hypothesis test using a significance level of α should be consistent with a $100(1-\alpha)$ confidence interval based on the same dataset.
  - A hypothesis test at $\alpha = 0.01$ should support a 99% confidence interval.
  - A hypothesis test at $\alpha = 0.05$ should support a 95% confidence interval.
  - A hypothesis test at $\alpha = 0.10$ should support a 90% confidence interval.

Hypothesis Tests and Confidence Intervals 2

- If we build a $100(1-\alpha)$ confidence interval for p, based on $\hat{p}$, and p_0 is included...
  - p_0 is a reasonable value for p.
  - The p-value associated with $\hat{p}$ will be greater than α.
  - We would fail to reject the null hypothesis, $p = p_0$.
- If we build a $100(1-\alpha)$ confidence interval for p, based on $\hat{p}$, and p_0 is not included...
  - p_0 is not a reasonable value for p.
  - The p-value associated with $\hat{p}$ will be less than α.
  - We would reject the null hypothesis, $p = p_0$.

Hypothesis Testing for One Proportion

Once you've determined a one-proportion hypothesis test is the correct procedure, there are four steps to completing the test.

1. **Prepare**. Identify the parameter of interest, list hypotheses, identify the significance level, and identify $\hat{p}$ and n.
2. **Check**. Verify conditions to ensure $\hat{p}$ is nearly normal under H_0. For one-proportion hypothesis tests, use the null value to check the success-failure condition.
3. **Calculate**. If the conditions hold, compute the standard error, again using p_0, compute the Z-score, and identify the p-value.
4. **Conclude**. Evaluate the hypothesis test by comparing the p-value to α, and provide a conclusion in the context of the problem.

One-Sided Hypothesis Tests 1

- So far, we have only considered a scenario where p is either above or below some null value p_0. This is called a **two-sided hypothesis test**.
- We might care about p being only above or only below p_0.
  - Is the population parameter less than p_0?
  - Is the population parameter more than p_0?
  - This is called a **one-sided** hypothesis test.
- In a one-sided test, the hypotheses are set up differently and the p-value is calculated differently–everything else is the same.

One-Sided Hypothesis Tests 2

- We have previously assumed that 64% of American adults drink coffee every day. If we sample 225 adults and find that 150 of them drink coffee every day, is there evidence that more than 64% drink coffee every day?

$$H_0 : p \leq 0.64$$

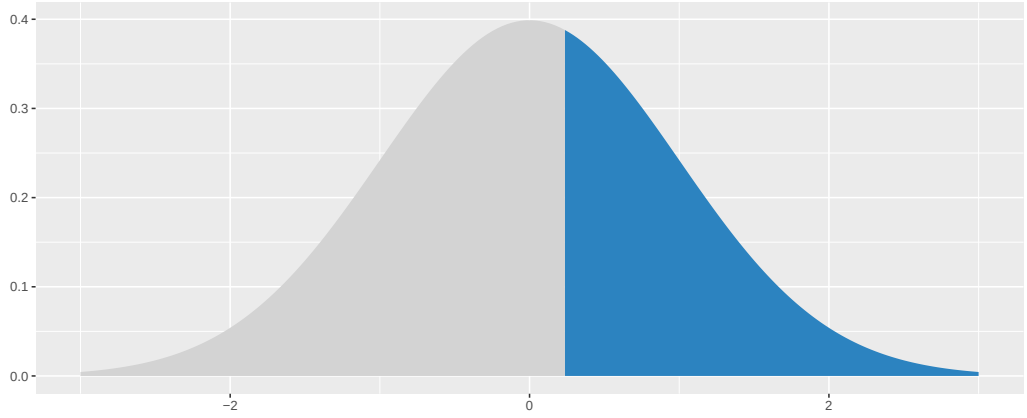
$$H_A : p > 0.64$$

- The test statistic is $z = \frac{\frac{150}{225} - 0.64}{\sqrt{\frac{0.64 \times (1-0.64)}{225}}} = 0.2357.$

```
pnorm(0.2357, mean = 0, sd = 1, lower.tail = FALSE)
```

```
## [1] 0.4068327
```

One-sided Hypothesis Tests 3



One-Sided Hypothesis Tests 4


```
prop.test(150, 225, 0.64, alternative = "greater", correct = F)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  150 out of 225, null probability 0.64
## X-squared = 0.69444, df = 1, p-value = 0.2023
## alternative hypothesis: true p is greater than 0.64
## 95 percent confidence interval:
##  0.6132635 1.0000000
## sample estimates:
##         p
## 0.6666667
```

prop.test()

- Note that there is no z value in the `prop.test()` output!
- That is because R calculates the test statistic a slightly different way.
- The conclusions (that is, reject or fail to reject) from R will always be consistent with the z-test we learned.

Notation Review 1

Notation	Representation
p	Population Proportion
$\hat{p}$	Sample Proportion
p_0	Null Proportion
p-value	Used to make a decision during a hypothesis test
$P()$	Probability

Notation Review 2

All of these different terms can be summed up in one expression:

$$\text{p-value} = P\left(Z > \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 \times (1-p_0)}{n}}}\right)$$

- We use $\hat{p}$, p_0, and the Central Limit Theorem for p to calculate the **p-value**, which is a probability ($P()$).