Introduction to Hypothesis Testing

Introduction to Hypothesis Testing

Administrivia

- **Homework 5 posted.** Due Friday Nov 10
- **Good Milestones**:
  - Problems 1-4 **this week**
  - Problems 5 and 6 **next week**

Previously on CSCI 3022

Proposition: If X is a normally distributed random variable with mean μ and standard deviation σ, then Z is a standard normal distribution if

$$Z = \frac{X - \mu}{\sigma} \quad \text{or} \quad X = \sigma Z + \mu$$

Fact: If Z is a standard normal random variable, then we can compute probabilities using the standard normal CDF

$$P(Z \leq z) = \int_{-\infty}^{z} f(x)\, dx = \Phi(z)$$

We've looked at ways to compute confidence intervals for several different statistics:

E.g. a $100(1-\alpha)\%$ **confidence interval** for the mean μ with known sd. σ is given by

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

A Thought Experiment

Example: After the introduction of the Euro, Polish mathematicians claimed that the new Belgian 1 Euro coin is not a fair coin. Suppose I hand you a Belgian 1 Euro coin. How could you decide whether or not it is fair?

A Thought Experiment

Example: After the introduction of the Euro, Polish mathematicians claimed that the new Belgian 1 Euro coin is not a fair coin. Suppose I hand you a Belgian 1 Euro coin. How could you decide whether or not it is fair?

Statistical Hypotheses

Def: A **statistical hypothesis** is a claim about the value of a parameter of a population characteristic.

Examples:

- Suppose the recovery time of a person suffering from disease D is normally distributed with mean μ_1 and standard deviation σ_1. **Hypothesis**: $\mu_1 > 10$ days.
- Suppose μ_2 is the recovery time of a person suffering from disease D and given treatment for D. **Hypothesis**: $\mu_2 < \mu_1$
- Suppose μ_1 is the mean internet speed for Comcast and μ_2 is the mean internet speed for Century Link. **Hypothesis**: $\mu_1 \neq \mu_2$

Null vs. Alternative Hypothesis

In any hypothesis-testing problem, there are always two competing hypotheses under consideration:

1. NULL HYPOTHESIS, H_0
2. Alt HYPOTHESIS, H_1 (or H_a)

The objective of **hypothesis testing** is to decide, based on sampled data, ***if the alternative hypothesis is actually supported by the data***.

The Classic Jury Analogy

Consider a **jury** in a criminal trial

When a defendant is accused of a crime, the jury (is supposed to) **presume** that he or she is **not guilty** (not guilty: that's the Null Hypothesis)

The jury is then presented with **evidence**. If the evidence seems implausible under the assumption of non-guilt, we might **reject** non-guilt and claim that the defendant is (likely) guilty.

Null vs Alternative Hypothesis

Is there strong evidence for the alternative?

The burden of proof is placed on those that believe the alternative claim.

The initially favored claim (H_0) will not be rejected in favor of the alternative claim (H_1) unless the sample evidence provides a lot of support for the alternative.

Two possible conclusions:

1. REJECT H_0 (IN FAVOR H_1)
2. FAIL TO REJECT THE NULL HYPOTHESIS H_0

Null vs Alternative Hypothesis

Why assume the Null Hypothesis?

- Sometimes we don't want to accept a particular assertion unless (or until) data can be shown to strongly support it

- Reluctance (cost, time) to change

Example: Suppose a company is considering hiring a new outside advertising company to help generate traffic to their website. Under their current advertising they get, on average, 200 thousand hits per day. With μ denoting the true average number of hits they'd get per day under the new company's advertising, they would not want to switch companies (because it would be costly) unless evidence strongly suggested that μ exceeds 200 thousand.

Null vs Alternative Hypothesis

Example: Suppose a company is considering hiring a new outside advertising company to help generate traffic to their website. Under their current advertising they get, on average, 200 thousand hits per day. With μ denoting the true average number of hits they'd get per day under the new company's advertising, they would not want to switch companies (because it would be costly) unless evidence strongly suggested that μ exceeds 200 thousand.

An appropriate problem formulation would involve testing:

$$H_0: \mu = 200 \qquad\qquad H_0: \mu = 200$$

$$H_1: \mu > 200 \qquad\qquad H_1: \mu > 205$$

The conclusion that change is justified is identified with the alternative hypothesis and it would take conclusive evidence to justify rejecting H_0 and switching to the new company

Null vs Alternative Hypothesis

ACTUAL #

The alternative to the Null hypothesis $H_0 : \theta = \theta_0$ will look like one of the following assertions

1. $\theta > \theta_0$
2. $\theta < \theta_0$
3. $\theta \neq \theta_0$

- The equals sign is **always** in the Null hypothesis
- The alternative hypothesis is the one for which we are seeking statistical evidence

Test Statistics and Evidence

Def: A **test statistic** is a quantity derived from the sample data and calculated assuming that the Null hypothesis is true. It is used in the decision about whether or not to reject the Null hypothesis.

Intuition:

- We can think of the test statistics as our $\boxed{\text{evidence}}$ about the competing hypotheses.
- We consider the test statistic under the assumption that the Null hypothesis is true by asking questions like **How likely would we obtain this evidence if the Null were true?**

Example: To determine if the Belgian 1 Euro coin is fair you flip it 100 times and record the number of Heads. What is the test statistics? What are the Null and alternative hypotheses?

$H_0: p = .5$

$H_1: p \neq .5$

Test Statistics and Evidence

Example: To determine if the Belgian 1 Euro coin is fair you flip it n times and record the number of Heads. What is the test statistic? What are the Null and alternative hypotheses?

$H_0: p = .5$

$H_1: p \neq .5$

TEST STATISTIC

$\hat{p} = \frac{X}{n}$ $\quad X \sim Bin(n, \frac{1}{2})$

IF n IS LARGE THEN DIST OF $\hat{p}$ IS APPROXIMATELY NORMAL:

VARIANCE

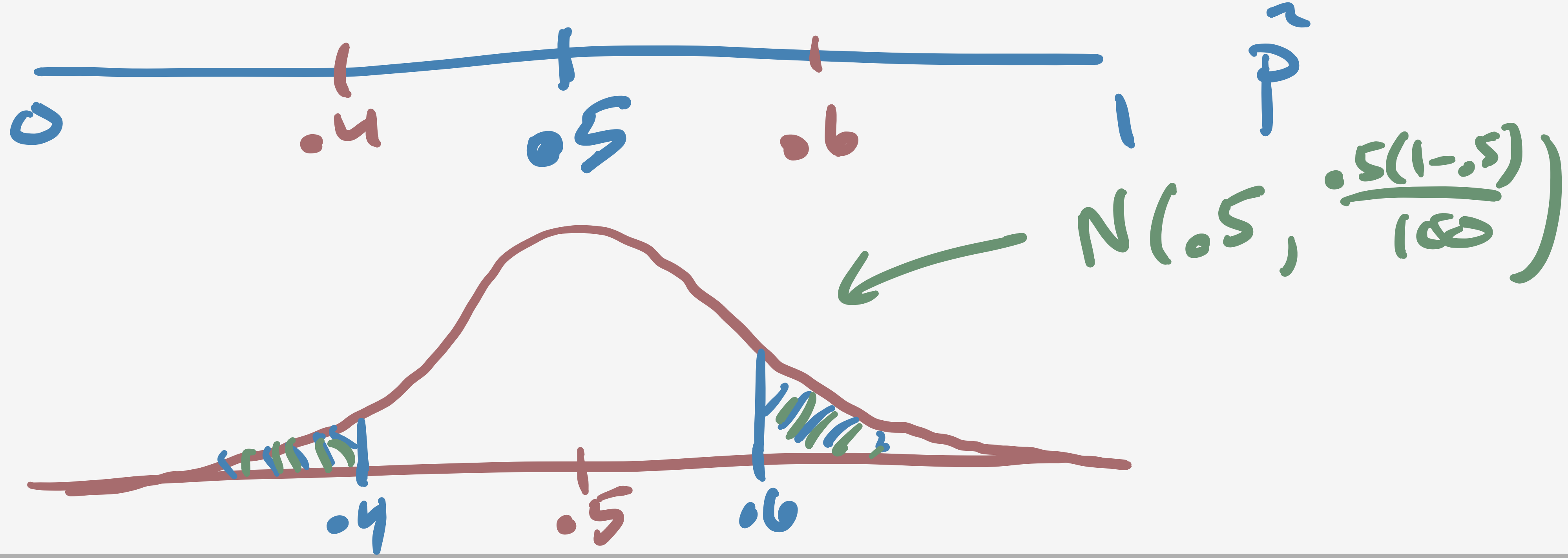
$$\hat{p} \overset{CLT}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$$

$\Rightarrow$ Null: $p = .5$ $\Rightarrow$ $\hat{p} \sim N\left(.5, \frac{.5(1-.5)}{100}\right)$

Test Statistics and Evidence

Example: To determine if the Belgian 1 Euro coin is fair you flip it n times and record the number of Heads. What is the test statistic? What are the Null and alternative hypotheses?

Question: What would it take to convince **you** that the coin is not fair?



Test Statistics and Evidence

Example: To determine if the Belgian 1 Euro coin is fair you flip it n times and record the number of Heads. What is the test statistic? What are the Null and alternative hypotheses?

Question: What would it take to convince **you** that the coin is not fair?

$H_1: p \neq .5$

Rejection Regions and Significance Level

Example: To determine if the Belgian 1 Euro coin is fair you flip it n times and record the number of Heads. What is the test statistic? What are the Null and alternative hypotheses?

Def: The **rejection region** is a range of values of the test statistic that would lead you to **reject** the Null hypothesis.

Def: The **significance level** α indicates the largest probability of the test statistic occurring under the Null hypothesis that would lead you to reject the Null hypothesis

Detecting Biased Coins

Example: To test if the Belgian 1 Euro coin is fair you flip it 100 times and observe 38 Heads. Do you reject the Null at the .05 significance level or not?

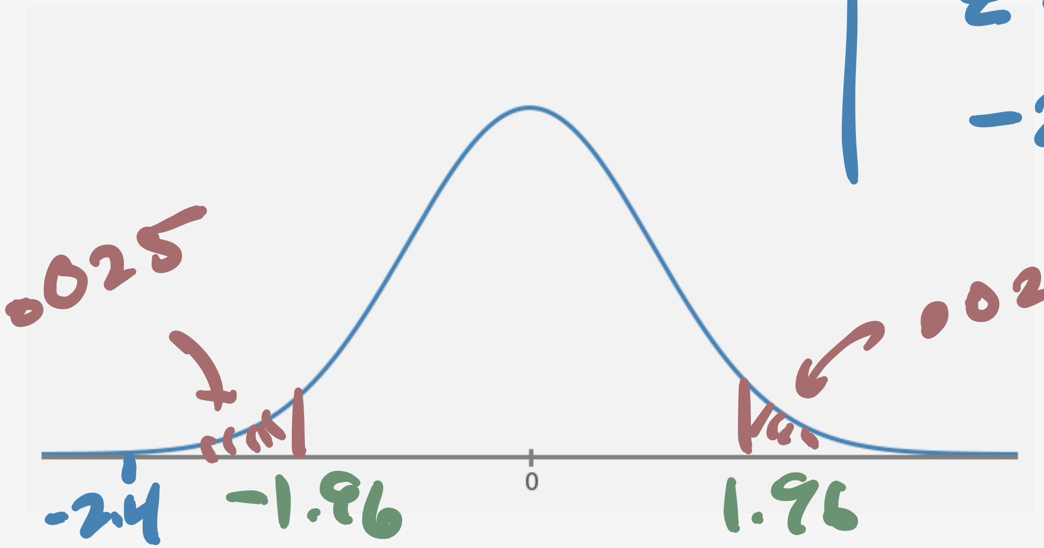
$H_0: p = .5$

$H_1: p \neq .5$

$$\hat{p} \sim N\left(.5, \frac{.5(1-.5)}{100}\right) \Rightarrow \boxed{\hat{p} = \frac{38}{100} = 0.38}$$

$z_{.025}$ = norm.ppf(.975) $= 1.96$

$$z = \frac{.38 - .5}{\sqrt{\frac{.5(1-.5)}{100}}} = -2.4$$



$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

$-2.4 < -1.96$

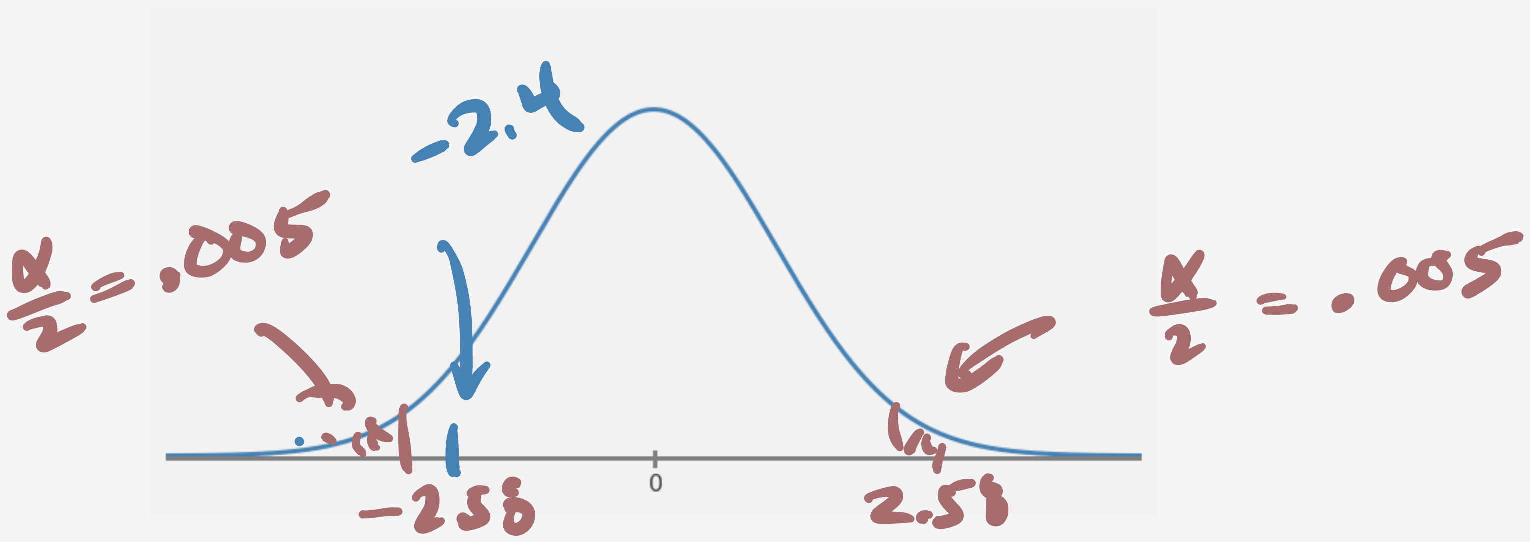
Conclusion:

EVIDENCE SUPPORTS REJECTING NULL H.

Detecting Biased Coins

Example: To test if the Belgian 1 Euro coin is fair you flip it 100 times and observe 38 Heads. Do you reject the Null at the .01 significance level or not?

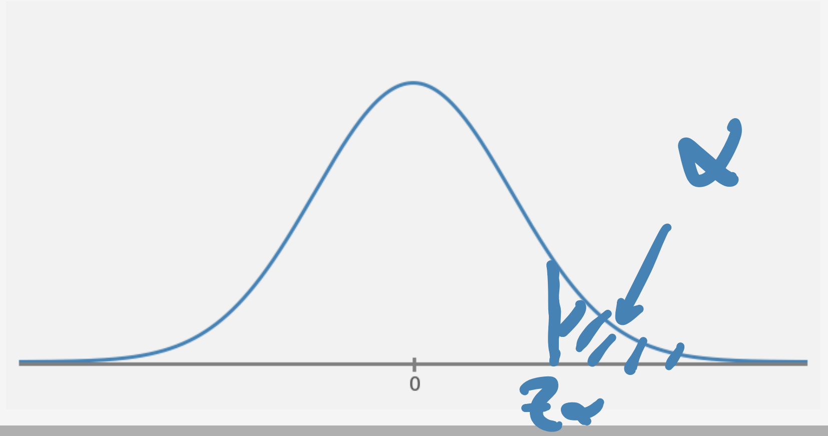
$z_{.005} = 2.58$

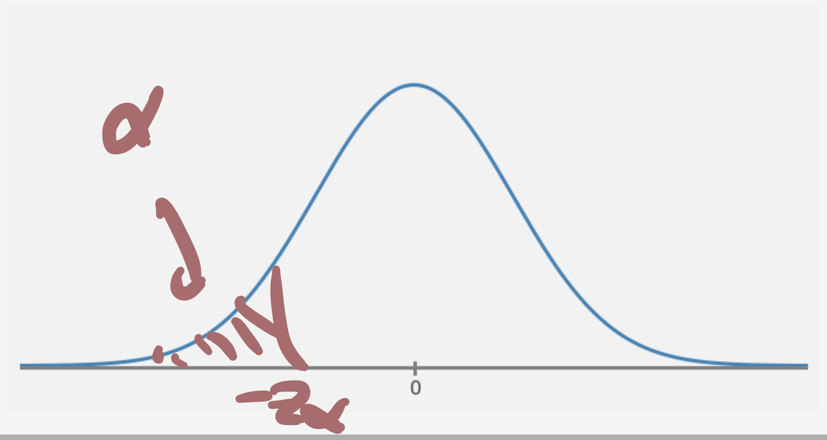


Different Tests for Different Hypotheses

The coin example was an exampled of a two-tailed hypothesis test, because we would have rejected the Null hypothesis had the coin been been biased towards heads OR tails

Alternative Hypothesis	**Rejection Region for Level α Test**
$H_1: \quad \theta > \theta_0$	$z \geq z_\alpha$
$H_1: \quad \theta < \theta_0$	$z \leq -z_\alpha$
$H_1: \quad \theta \neq \theta_0$	$(z \leq -z_{\alpha/2})$ or $(z \geq z_{\alpha/2})$





Switching Advertising Strategies

Example: Suppose a company is considering hiring a new outside advertising company to help generate traffic to their website. Under their current advertising they get, on average $\boxed{200}$ thousand hits per day with a standard deviation of $\boxed{50}$ thousand hits per day. You decide to hire the new ad company for a 30 day trial. During those 30 days, your website gets $\boxed{210}$ thousand hits per day. Perform a hypothesis test to determine if the new ad campaign outperforms the old one at the .05 significance level.

$H_0: \mu = 200$

$H_1: \mu > 200$

$\bar{x} \sim N\left(200, \frac{50^2}{30}\right)$

$\frac{\bar{x} - 200}{\sqrt{50^2/30}} \Rightarrow \frac{210 - 200}{50/\sqrt{30}} = 1.1$

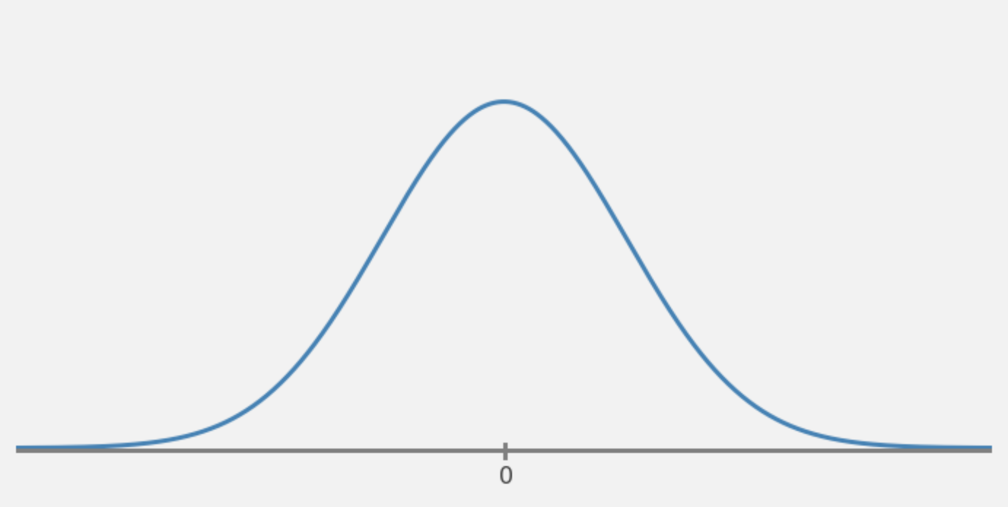
$\cancel{z > z_\alpha}$

$1.1 \not> 1.64$

FAIL to REJECT Ho.

Switching Advertising Strategies

Example: Suppose a company is considering hiring a new outside advertising company to help generate traffic to their website. Under their current advertising they get, on average, 200 thousand hits per day with a standard deviation of 50 thousand hits per day. You decide to hire the new ad company for a 30 day trial. During those 30 days, your website gets 210 thousand hits per day. Perform a hypothesis test to determine if the new ad campaign outperforms the old one at the .05 significance level.



Important Assumptions

Question: What assumptions did we make in the previous examples?

Errors in Hypothesis Testing

Definitions:

- A **Type I Error** occurs when the Null hypothesis is rejected, but the Null hypothesis is in fact true (**False Positive**)

- A **Type II Error** occurs when the Null hypothesis is not rejected, but the Null hypothesis is in fact false (**False Negative**)

Question: What is the probability that we commit a **Type I Error**?

Errors in Hypothesis Testing

Definitions:

- A **Type I Error** occurs when the Null hypothesis is rejected, but the Null hypothesis is in fact true (**False Positive**)

- A **Type II Error** occurs when the Null hypothesis is not rejected, but the Null hypothesis is in fact false (**False Negative**)

Question: What is the probability that we commit a **Type I Error**?

Answer: This is exactly the significance level α

Consequence: We set α by considering how willing we are to risk a Type I Error

OK! Let's Go to Work!

Get in groups, get out laptop, and open the Lecture 17 In-Class Notebook

Let's:

- Work through some more hypothesis test examples