# Hypothesis Testing and t-Tests

Reeves Johnson

University of Virginia
McIntire School of Commerce

# Objectives

- Define a hypothesis and hypothesis testing
- Distinguish between one-tailed and two-tailed tests
- Test hypotheses comparing two population means, $\mu_1$ and $\mu_2$
- Understand independent (paired) samples hypothesis tests through practice

# Motivation: Do You Know More About Global Trends Than Chimpanzees?

Go to *Financial Times* article.

# Hypotheses and Hypothesis Testing

**Question:** What is a hypothesis and what is hypothesis testing?

Make sure to include parameters and research questions in your answer.

# The Backwards Art of Hypothesis Testing

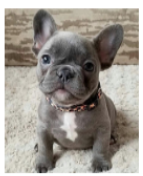
Can I prove that this is true?

$H_1$: French bulldogs are cute

Disprove this!

$H_0$: French bulldogs are **NOT** cute

Evidence against



# State the Null and Alternative Hypothesis

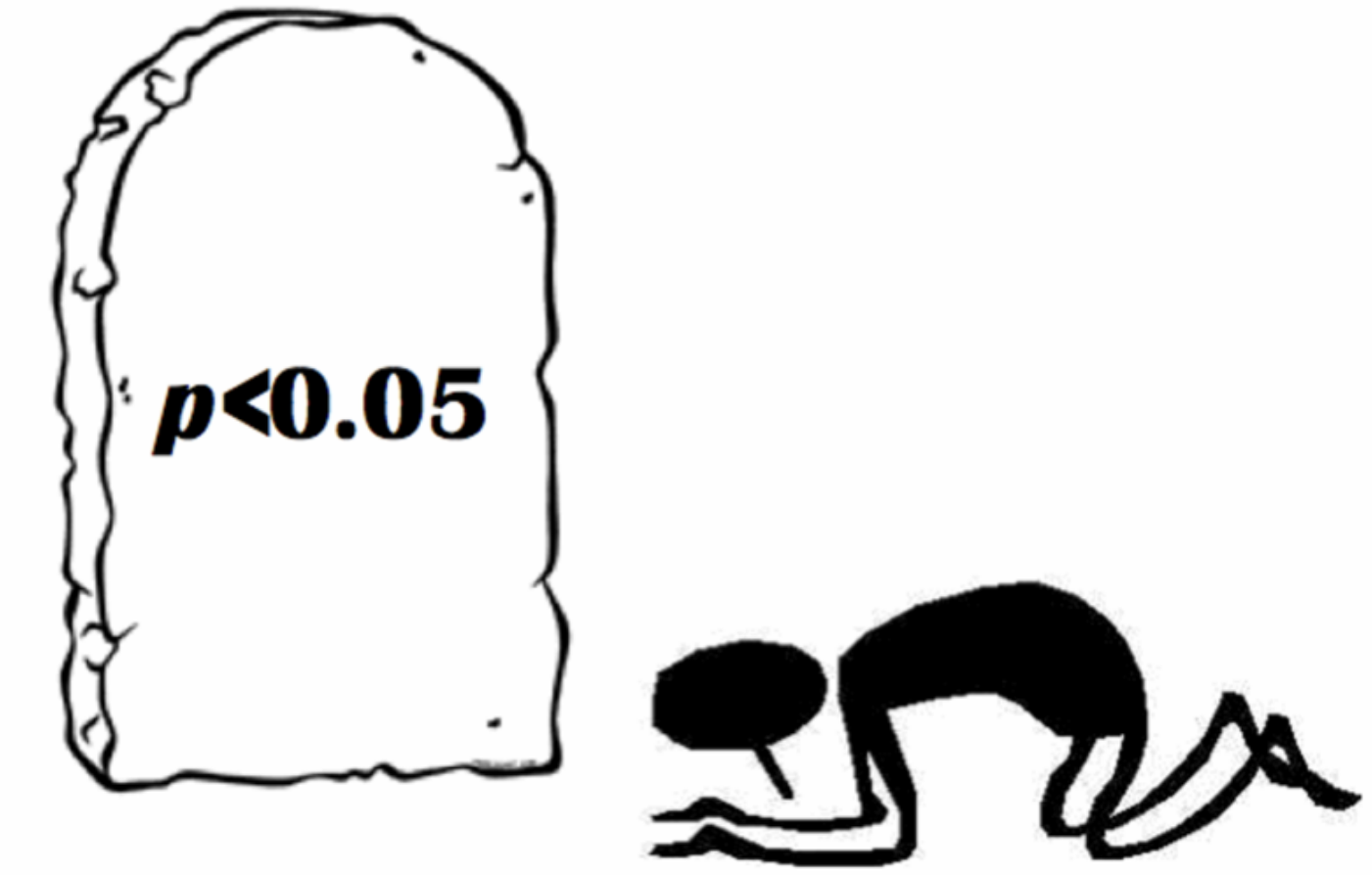
- What is a null hypothesis ($H_0$)? Examples?
- What is an alternative hypothesis ($H_1$)? Examples?

**NB:** Hypothesis testing requires you to speak in terms of *statistically significant* differences.

- For a one-sample $t$-test, for example, the probability that $\bar{x} = \mu$ is 0.
- The question is whether $\bar{x}$ is *statistically significant* different from $\mu$ to reject the null hypothesis.

# *p*-Values: Strength of Evidence Against the Null

- A *p*-value is one measure for the strength of evidence against the null hypothesis.
- It's one indication of how far away a statistic is from a hypothesized parameter value, *assuming the null hypothesis is true*.
- That hypothesized parameter value is called the **null value**.


p<0.05

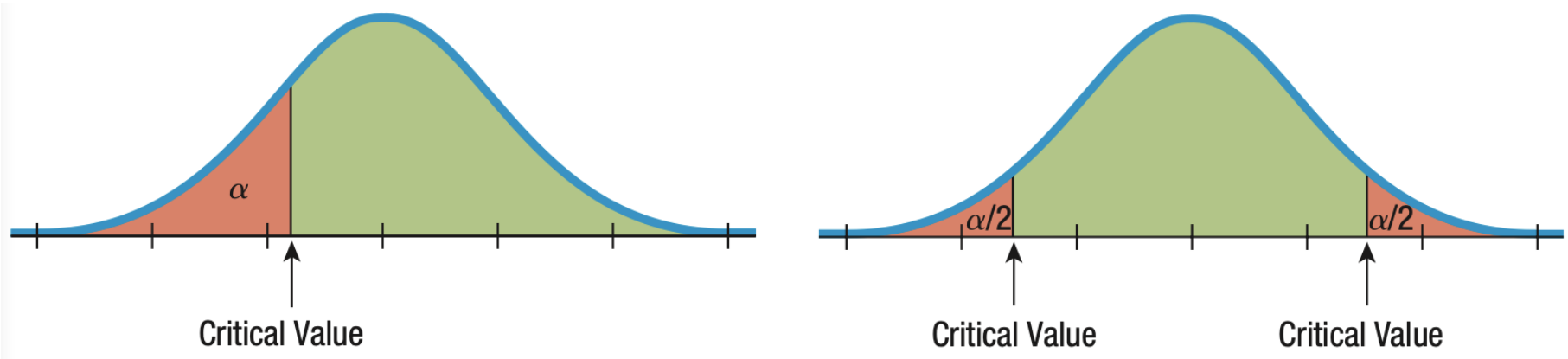
# A Significance Level $\alpha$

Two Ways to Understand a Significance Level (denoted by $\alpha$):

1. A **significance level** is a fixed probability such that:
   - A *p*-value less than $\alpha$: Reject the null hypothesis
   - A *p*-value greater than $\alpha$: Fail to reject the null hypothesis
2. A **significance level** is the *probability* of *rejecting* the null hypothesis, *assuming* the null hypothesis is true.

# Rejection Regions of a Sampling Distribution

Rejection regions corresponding to a one-sided (left) and a two-sided (right) hypothesis test:



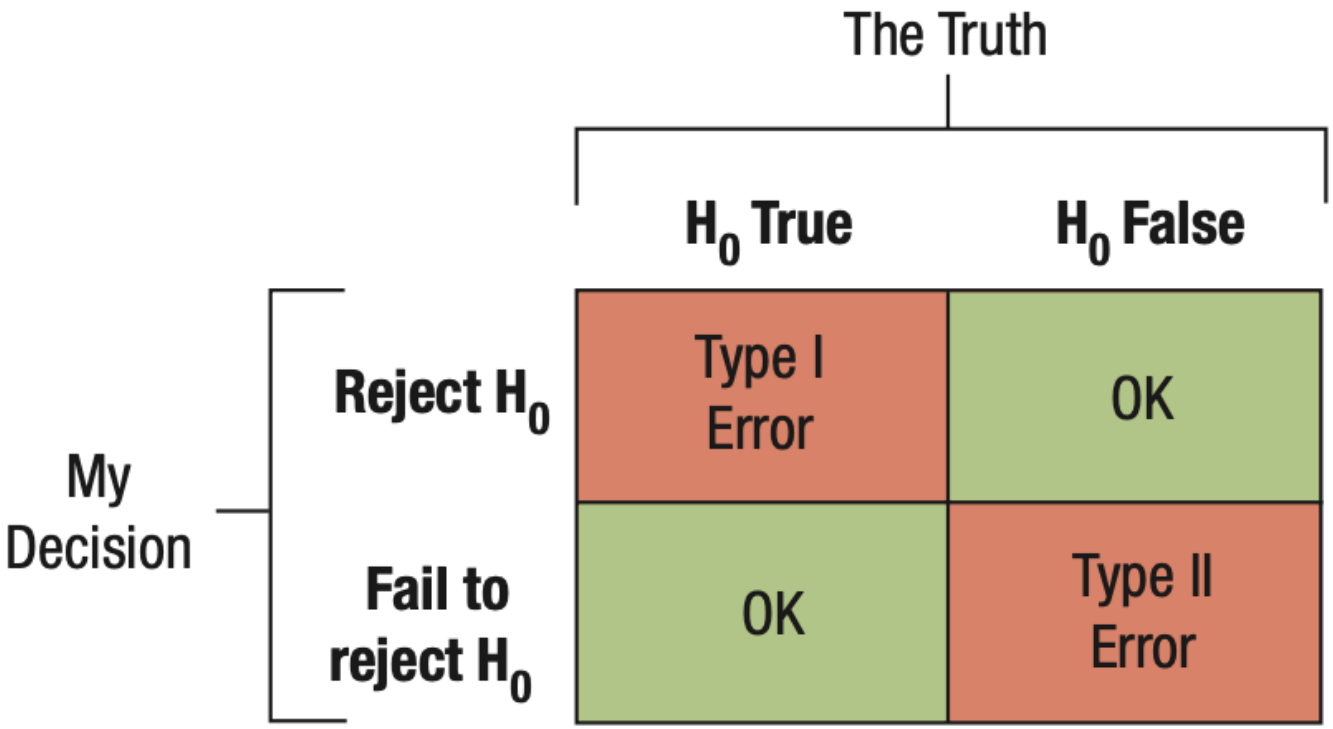

Question:

- All else equal, is it easier to reject the null hypothesis for a one-sided test or a two-sided test?

# Sampling Your Way to Significance: Why Practical Significance Matters

- Consider the following example of two A/B tests with very different sample sizes.
- The key takeaway: Increasing your sample size, all else equal, increases the likehood of obtaining significant results.
- That is, statistical significance isn't the same as thing as practical significance.
- The R script is on Canvas in case you want to review it later.

# Decisions and Consequences: Type I and II Errors



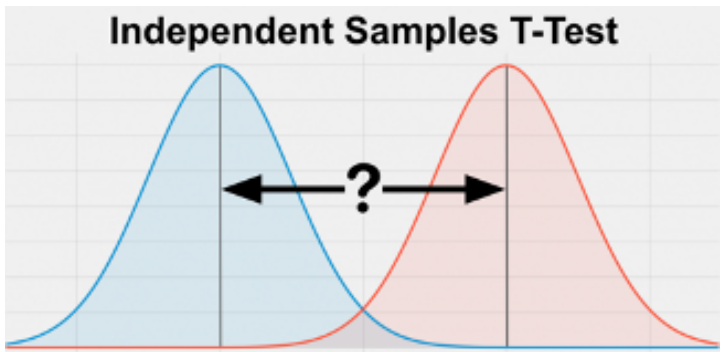
# $p$-Values Revisited

- Companies often need to decide collectively what rate of false positives ($\alpha$) they'll willing to tolerate.
- As a $p$-value summarizes the results of an experiment, each person can bring their own risk tolerances to the table. (A nice organizational benefit in companies.)
- Even if you're not running the experiment, you're going to want to have a say.

# Independent Samples $t$-Tests

A **paired- or independent-samples $t$-test** assesses the difference between the population means of 2 *independent* groups (samples). The difference in sample means, the test statistic/estimate, is used to evaluate whether the 2 samples were drawn from different populations. A *two-sided* independent-samples $t$-test is written:

$$H_0 : \mu_1 = \mu_2 \Leftrightarrow H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 \neq \mu_2 \Leftrightarrow H_1 : \mu_1 - \mu_2 \neq 0$$




**NB:** The dependent variable must be quantitative and the independent variable must be a binary variable (to form 2 groups).