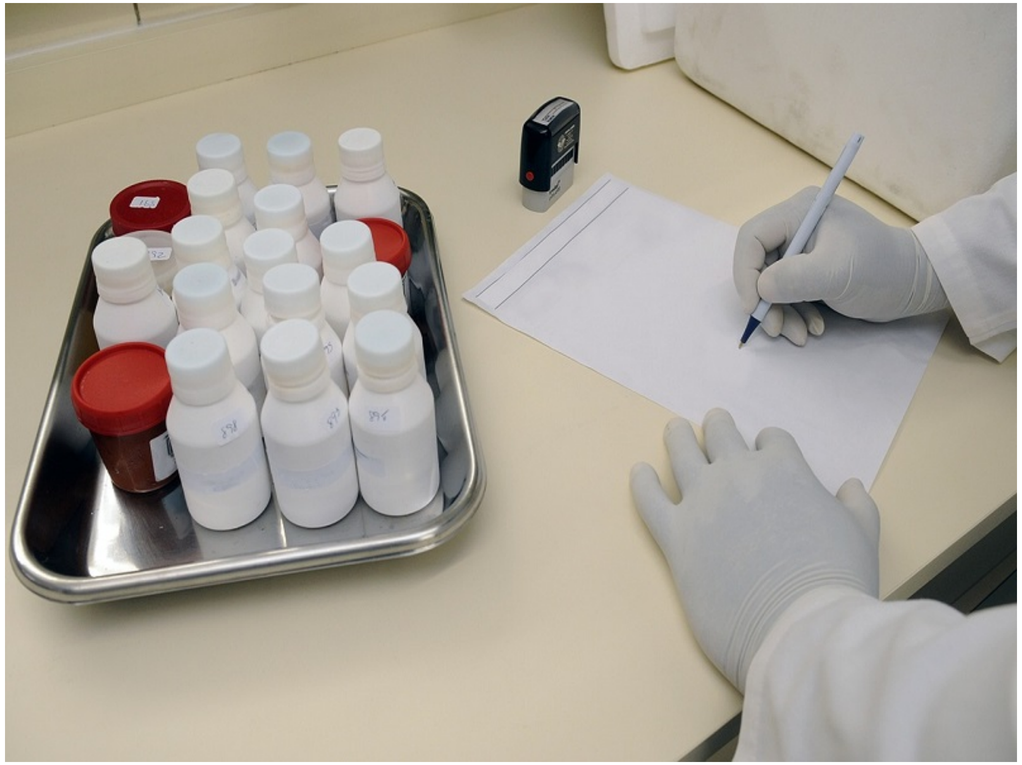
# Hypothesis tests for two means



# Overview

Review/continuation of hypothesis testing for a single proportion

One tailed vs. two-tailed tests

Hypothesis tests for two means

# Questions about homework 5?

Do we feel confident about constructing confidence intervals using the bootstrap?

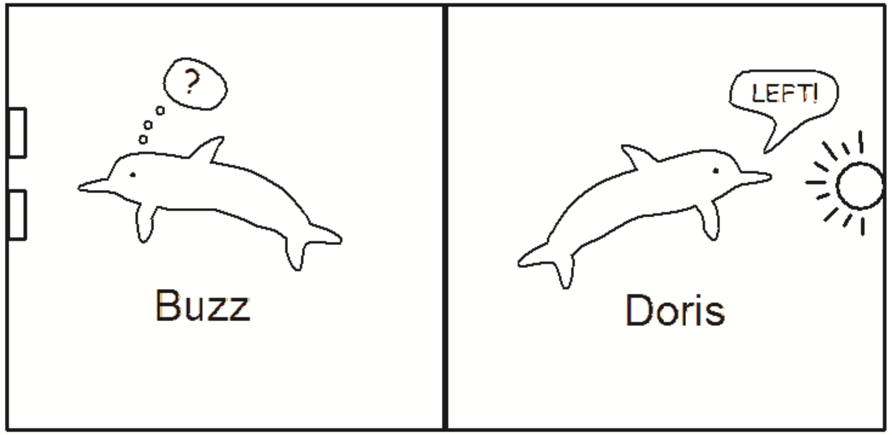
Do we think Paul was psychic?

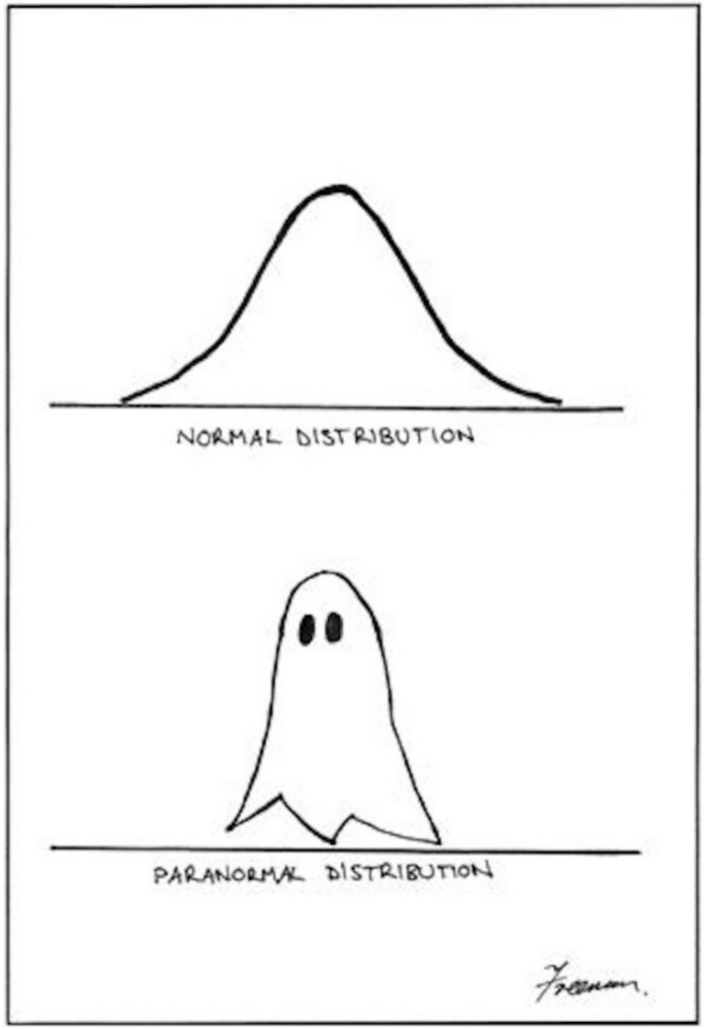

# Homework 6

Homework 6 has been posted

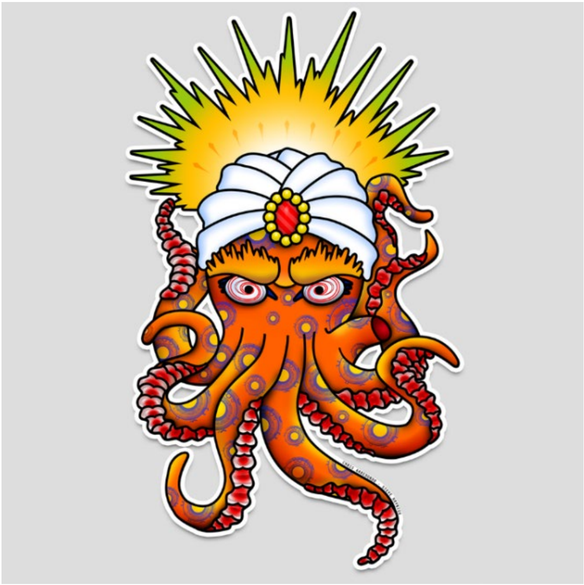
- Due on Gradescope at 11:30pm on Sunday March 1st
- It is shorter (only 50 points) to give you more time to study for the midterm

# Review: hypothesis tests for a single proportion

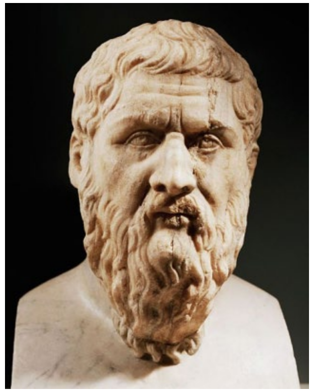






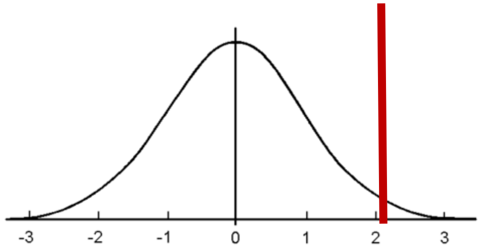
# Five steps of hypothesis testing



1. State $H_0$ and $H_A$
   - Assume Gorgias ($H_0$) was right


2. Calculate the actual observed statistic

3. Create a distribution of what statistics would look like if Gorgias is right
   - Create the **null distribution** (that is consistent with $H_0$)




4. Get the probability we would get a statistic more than the observed statistic from the null distribution
   - p-value

5. Make a judgement
   - Assess whether the results are statistically significant

# One-tailed vs. two tailed

In the examples we have seen, we were just interested if the parameter was <u>**greater**</u> than an hypothesized parameter

$H_0: \pi = 0.25$ $H_A: \pi > 0.25$

In other cases we might not have a directional alternative hypothesis

# Testing whether a coin is biased

Suppose we wanted to test what whether Buzz chose the correct food well ***more or less*** than 50% of the time

- e.g., Buzz might not like the food so was avoiding the well with the food

1. Write down the null and alternative hypotheses

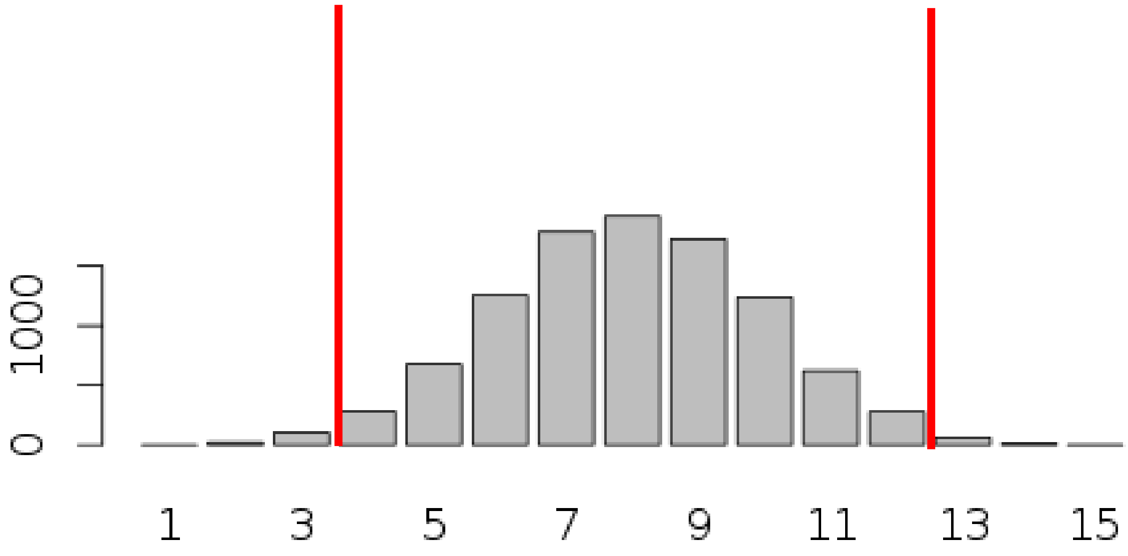
2. Suppose out of the 16 trials, Buzz got the correct 3 times. How would we use a randomized distribution to tell if the coin is biased?

| | |
|---|---|
| 0 | 0 |
| 1 | 1 |
| 2 | 22 |
| 3 | 105 |
| 4 | 283 |
| 5 | 679 |
| 6 | 1257 |
| 7 | 1786 |
| 8 | 1920 |
| 9 | 1726 |
| 10 | 1238 |
| 11 | 623 |
| 12 | 279 |
| 13 | 63 |
| 14 | 15 |
| 15 | 3 |
| 16 | 0 |

2. Suppose out of the 16 trials, Buzz got the correct 3 times. How would we use our randomized distribution to tell?

3. Based on this table, what is the p-value?

| | |
|---|---|
| 0 | 0 |
| 1 | 1 |
| 2 | 22 |
| 3 | 105 |
| 4 | 283 |
| 5 | 679 |
| 6 | 1257 |
| 7 | 1786 |
| 8 | 1920 |
| 9 | 1726 |
| 10 | 1238 |
| 11 | 623 |
| 12 | 279 |
| 13 | 63 |
| 14 | 15 |
| 15 | 3 |
| 16 | 0 |




p-value = 209/10000 = .0209

Compare this p-value to we would have gotten if we **expected** Buzz to avoid the food well?

# Statement of alternative hypothesis is important

We need to state what you expect before analyzing the data

Our expectation (hypothesis statement) can change the p-value

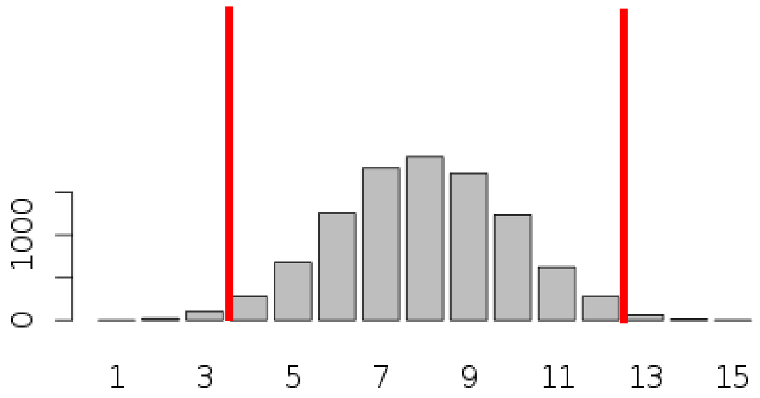
# Estimating a p-value from a randomized distribution

For a one tailed alternative: Find the proportion of randomized samples that equal or exceed the original statistic in the direction (tail) indicated by the alternative hypothesis

For a two-tailed alternative: Find the proportion of randomization samples in the tails beyond the observed statistic and 1 - the observed statistic

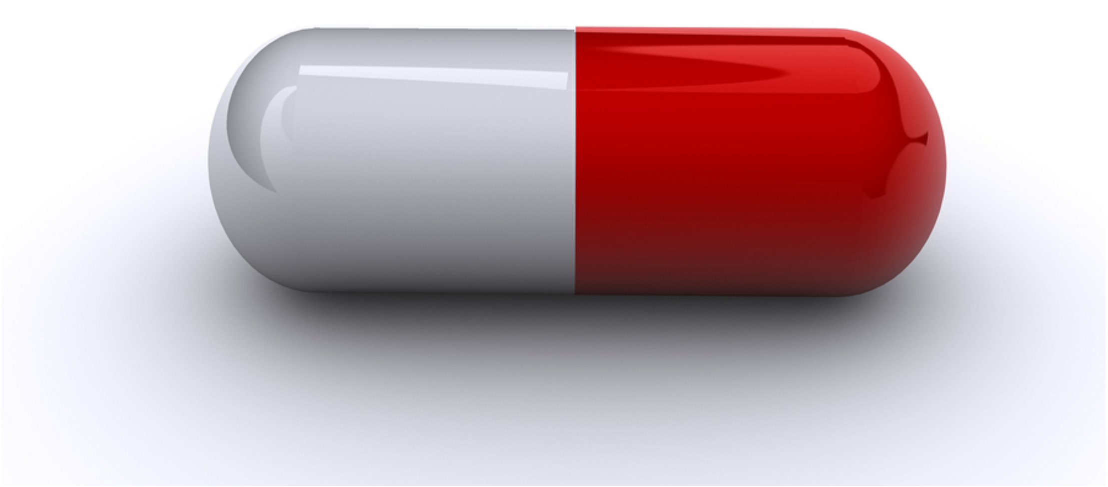
- Alternatively, find the proportion of randomization samples in the smaller tail at or beyond the original statistic and then double the proportion to account for the other tail

# How to estimate two sided p-values in R?




```
null_distribution <-  do_it(10000) * {
            rflip_count(16,  prob = .5)
}

p_left_tail  <-   pnull(3,  null_distribution,  lower.tail = TRUE)
p_right_tail  <- pnull(16 - 3,  null_distribution,  lower.tail = FALSE)

p_value <-  p_right_tail + p_left_tail
```

# Hypothesis tests for comparing two means



**Question**: Is this pill effective?

# Hypothesis tests for comparing two means

**Question**: Can we find out the ***Truth*** of whether the pill effective?

# Testing whether a pill is effective

How would we design a study?

What would the cases and variables be?

What would the statistic of interest be?

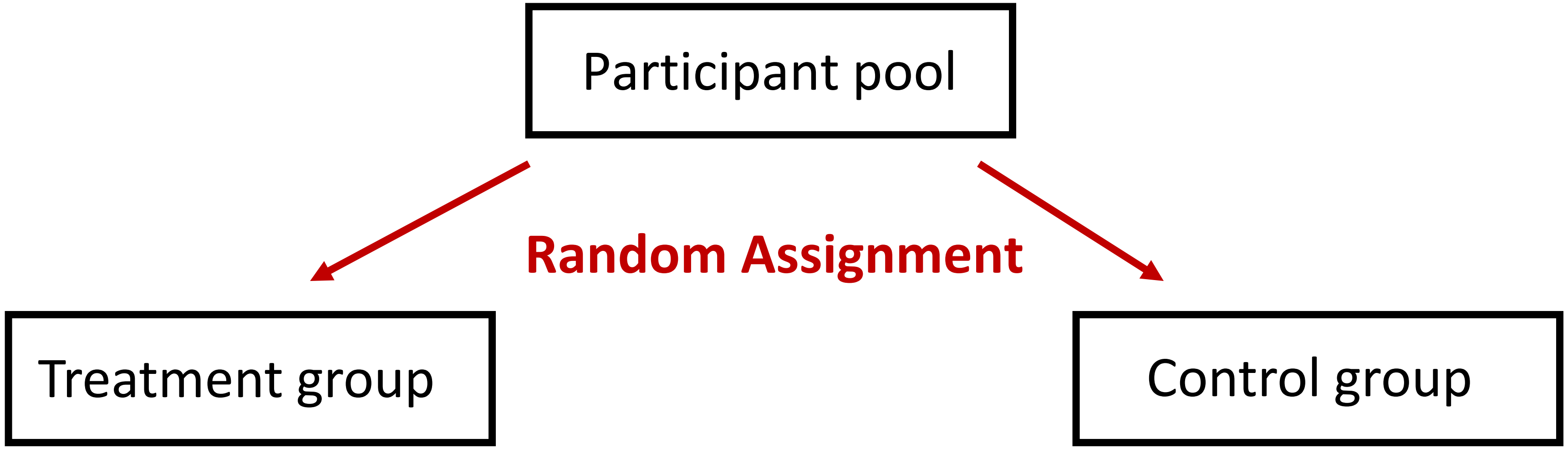
What are the null and alternative hypotheses?

- Assume we are looking for differences in means between the groups

# Experimental design

Take a group of participant and ***randomly assign***:

- Half to a *treatment group* where they get the pill
- Half in a *control group* where they get a fake pill (placebo)
- See if there is more improvement in the treatment group compared to the control group



# Example: Does calcium reduce blood pressure?

A randomized by Lyle et al (1987) comparative experiment investigated whether calcium lowered blood pressure in African-American men

- A treatment group of 10 men received a calcium supplement for 12 weeks
- A control group of 11 men received a placebo during the same period

The blood pressure of these men was taken before and after the 12 weeks of the study

1) What are the null and alternative hypotheses?

- $H_0$: $\mu_{Treatment} = \mu_{Control}$ or $\mu_{Treatment} - \mu_{Control} = 0$
- $H_A$: $\mu_{Treatment} > \mu_{Control}$ or $\mu_{Treatment} - \mu_{Control} > 0$
    - i.e., a <u>greater decrease</u> in blood pressure after taking calcium

# Hypothesis tests for differences in two group means

1) State the null and alternative hypothesis

- $H_0$: $\mu_{Treatment} = \mu_{Control}$ or $\mu_{Treatment} - \mu_{Control} = 0$
- $H_A$: $\mu_{Treatment} > \mu_{Control}$ or $\mu_{Treatment} - \mu_{Control} > 0$

2) Calculate statistic of interest

- $\overline{x}_{Effect} = \overline{x}_{Treatment} - \overline{x}_{Control}$

# Does calcium reduce blood pressure?

Treatment data (n = 10):

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Begin | 107 | 110 | 123 | 129 | 112 | 111 | 107 | 112 | 136 | 102 |
| End | 100 | 114 | 105 | 112 | 115 | 116 | 106 | 102 | 125 | 104 |
| **Decrease** | **7** | **-4** | **18** | **17** | **-3** | **-5** | **1** | **10** | **11** | **-2** |

Control data (n = 11):

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Begin | 123 | 109 | 112 | 102 | 98 | 114 | 119 | 112 | 110 | 117 | 130 |
| End | 124 | 97 | 113 | 105 | 95 | 119 | 114 | 114 | 121 | 118 | 133 |
| **Decrease** | **-1** | **12** | **-1** | **-3** | **3** | **-5** | **5** | **2** | **-11** | **-1** | **-3** |

2) What is the observed statistic of interest?

- $\overline{x}_{Effect} = 5 - -.2727 = 5.273$

3) What is step 3?

# 3. Create the null distribution!

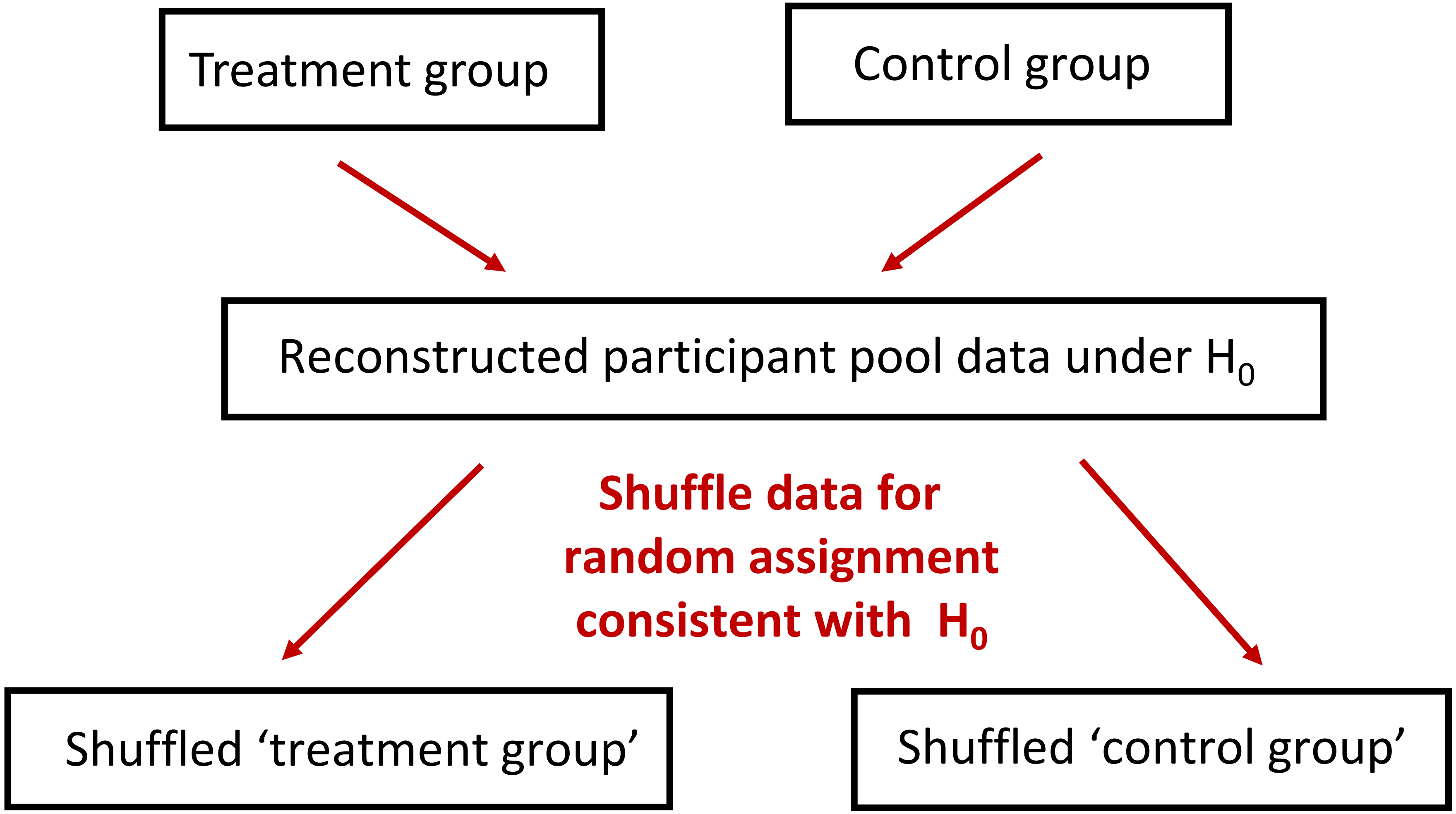
How could we create the null distribution?

Need to generate data consistent with $H_0$: $\mu_{Treatment} - \mu_{Control} = 0$

- i.e., we need fake $\bar{x}_{Effect}$ that are consistent with $H_0$

Any ideas how we could do this?

# 3. Create the null distribution!




One null distribution statistic: $\bar{x}_{Shuff\_Treatment} - \bar{x}_{Shuff\_control}$

# 3. Create a null distribution

1) Combine data from both groups

2) Shuffle data

3) Randomly select 10 points to be the ‘null’ treatment group

4) Take the remaining points to the ‘null’ control group

5) Compute the statistic of interest on these ‘null’ groups

6) Repeat 10,000 times to get a null distribution

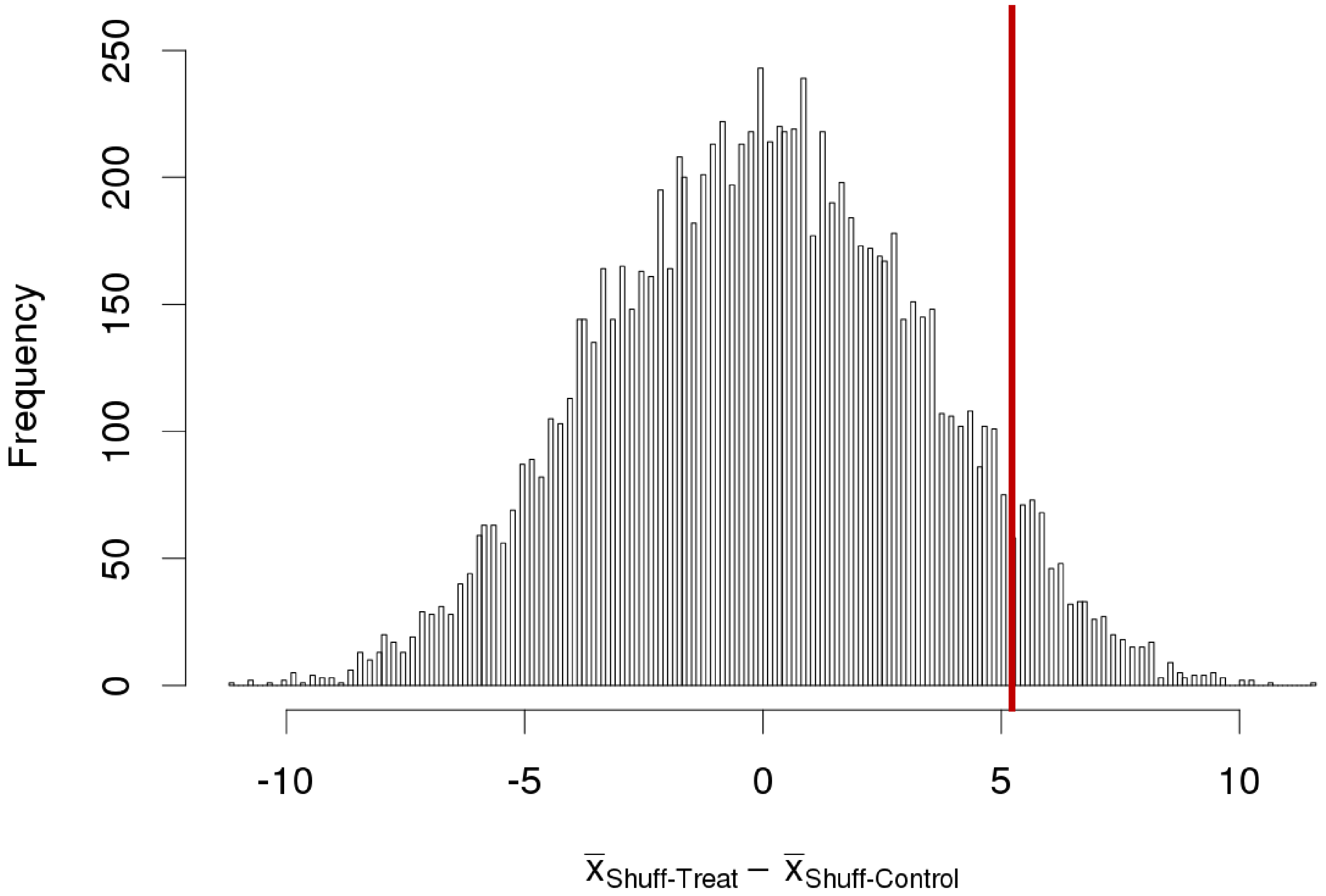
# 3. Creating a null distribution in R

```
# the data from the calcium study
treat <- c(7, -4, 18, 17, -3, -5,  1, 10, 11, -2)
control <- c(-1,  12,  -1,  -3,   3,  -5,   5,   2, -11,  -1,  -3)

# observed statistic
obs_stat <- mean(treat) - mean(control)

#  Combine data from both groups
combined_data <- c(treat, control)
```

# 3. Creating a null distribution in R

```
null_distribution <-  do_it(10000) * {

    # shuffle data
    shuff_data <- shuffle(combined_data)

    # create fake treatment and control groups
    shuff_treat   <-  shuff_data[1:10]
    shuff_control  <-  shuff_data[11:21]

    # save the statistic of interest
    mean(shuff_treat) - mean(shuff_control)

}
```

Null distribution




hist(null_distribution, nclass = 200)

Next step?

# 4. Calculate the p-value

```
# 8)  Calculate the p-value
    > p_value <- pnull(obs_stat, null_distribution, lower.tail = FALSE)
```

p-value = .064

Next step?

# 5. Are the results statistically significant?

?

What should we do?

# More/larger studies!