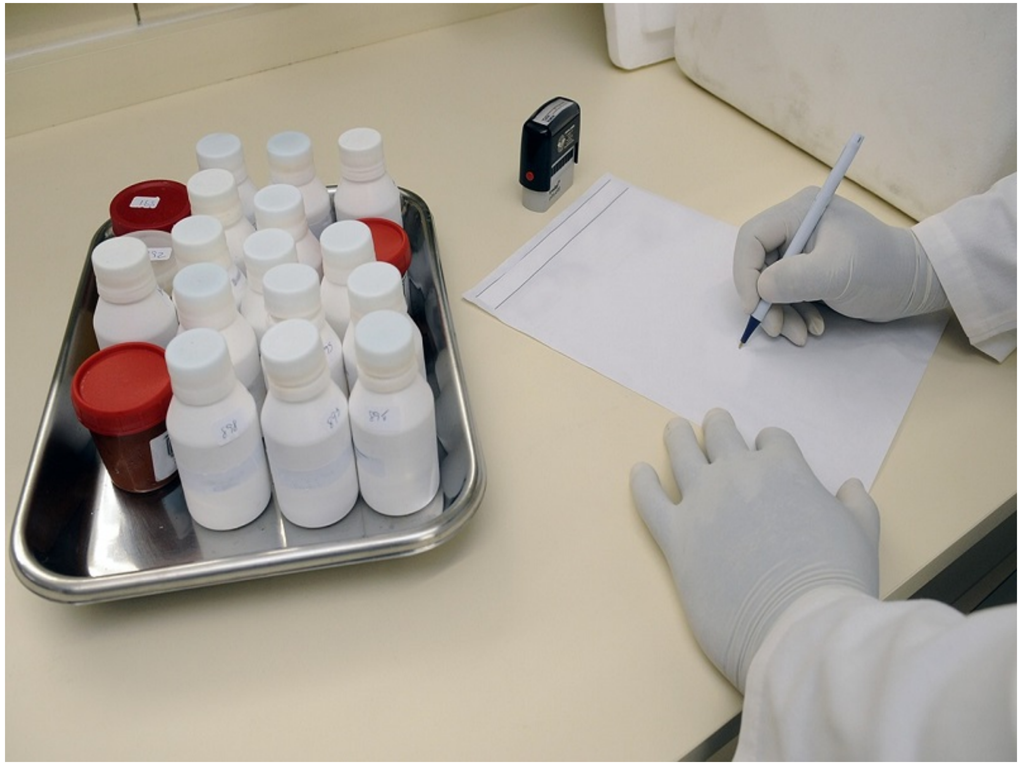
# Hypothesis tests for two means

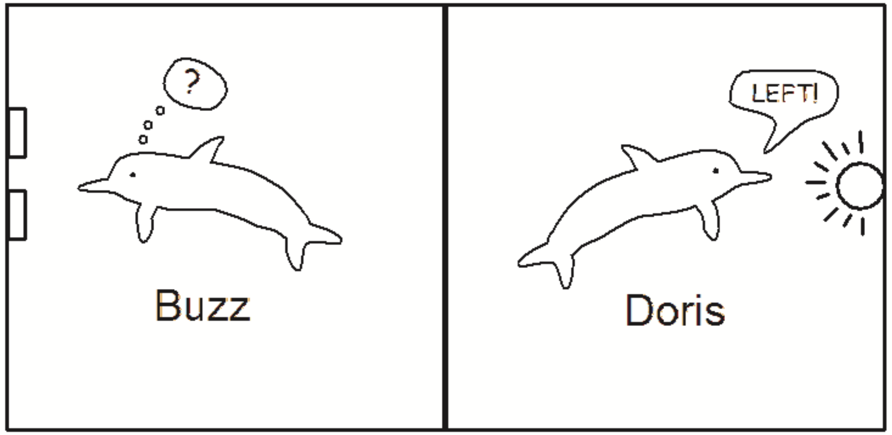


# Overview

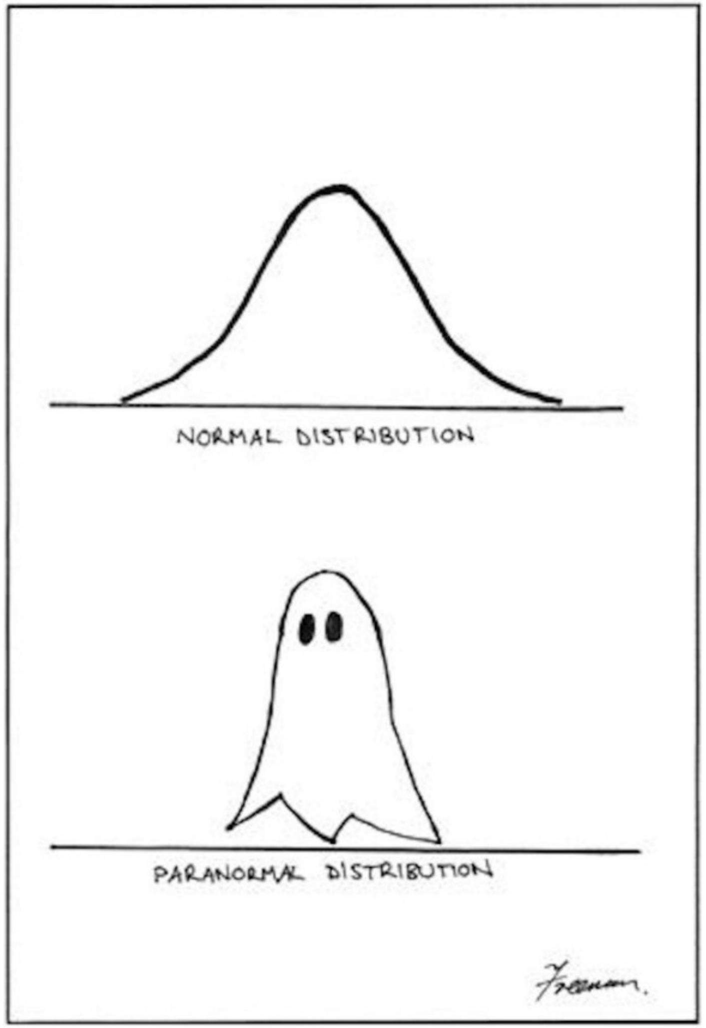
Review of hypothesis testing for a single proportion

Hypothesis tests for two means

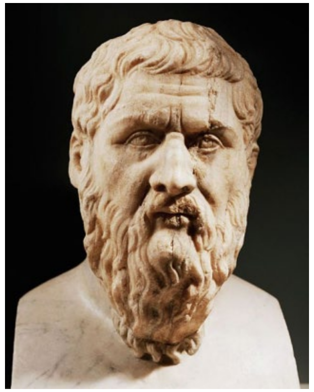
# Review: hypothesis tests for a single proportion








# Five steps of hypothesis testing

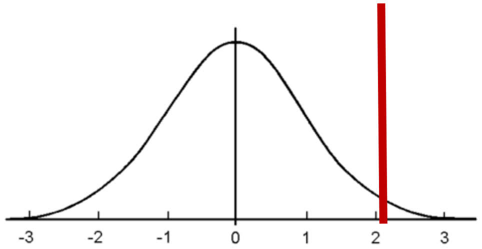


1. State $H_0$ and $H_A$
   - Assume Gorgias ($H_0$) was right


2. Calculate the actual observed statistic

3. Create a **null distribution** of statistics that are consistent with $H_0$
   - i.e., a distribution of statistics that we would expect if Gorgias is right

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

# Adult persistence of head-turning asymmetry

Background

- Most people are right handed, right eye dominant, etc.
- Biologists have suggested that human embryos tend to turn their heads to the right as well.

German bio-psychologist Onur Güntürkün conjectured that this tendency manifests itself in other ways, so he studies which ways people turn their heads when they kiss.

# Adult persistence of head-turning asymmetry

He and his researchers observed kissing couples in public places and noted whether the couple leaned their heads to the right or left.

They observed 124 couples, ages 13-70 years.

# Adult persistence of head-turning asymmetry

Please write down answers to these questions:

1. What are the observational units?

2. What are the variables (categorical or quantitative)?

3. What is Onur’s conjecture?  How would you state the null and alternative hypothesis in words and in symbols?

brief communications

# Adult persistence of head-turning asymmetry

A neonatal right-side preference makes a surprising romantic reappearance later in life.

A preference in humans for turning the head to the right, rather than to the left, during the final weeks of gestation and for the first six months after birth[1,2] constitutes one of the earliest examples of behavioural asymmetry and is thought to influence the subsequent development of perceptual and motor preferences by increasing visual orientation to the right side[3,4]. Here I show that twice as many adults turn their heads to the right as to the left when kissing, indicating that this head-motor bias persists into adulthood. My finding may be linked to other forms of sidedness (for example, favouring the right foot, ear or

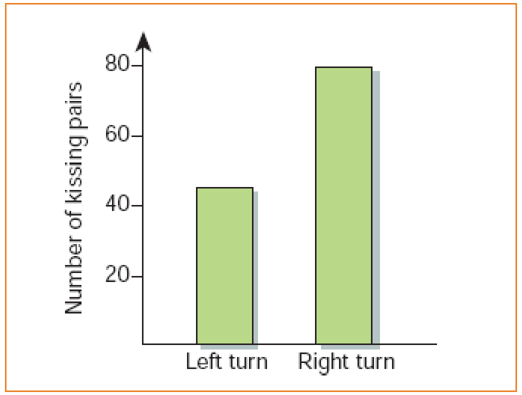



**Figure 1** The number of couples who turn their heads to the right rather than to the left when kissing predominates by almost 2:1 (64.5%: 35.5%; $n = 124$ couples).

Of the 124 couples observed, 80 leaned their heads to the right while kissing

- Let’s run a hypothesis test by going through the 5 steps….

# Please complete the 5 steps to run a hypothesis test

1. State Null and Alternative in symbols and words

2. Calculate the observed statistic of interest  (obs_stat)

3. Create a null distribution in R

    null_dist  <-  do_it(10000) * {     rflip_count(num_flips =  …   prob = … )    }

4. Calculate a p-value by assessing the probability of getting a statistic as or more extreme than the observed statistic from the null distribution

    - pnull(obs_stat, null_dist, lower.tail = …)

5. Make a decision about whether the results are statistically significant

Report the p-value and your conclusions

# Adult persistence of head-turning asymmetry

1. $H_0$: $\pi_{right}$ = 0.5     $H_A$: $\pi_{right}$ > 0.5
2. $\hat{p}$ = 80/124 = .64

```
#  step 3:  create the null distribution
null_distribution <-  do_it(10000) * {
        rflip_count(124,  prob = .5)/124
}
```

4. `p_val  <- pnull(.64,  null_distribution,  lower.tail = FALSE)`     `# p-value = 0.0007`

5. Decision?

# Suppose Onur believed couples would **turn their heads to the left** when they kissed?

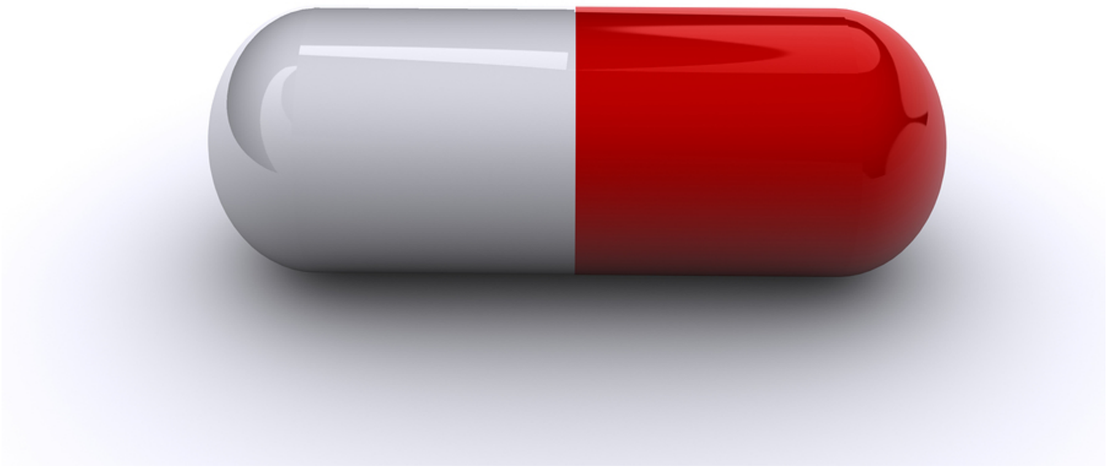
What would the alternative hypothesis be?

- Assume the same null hypothesis

What would the p-value be?

What would the conclusion be?

# 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 are the null and alternative hypotheses?

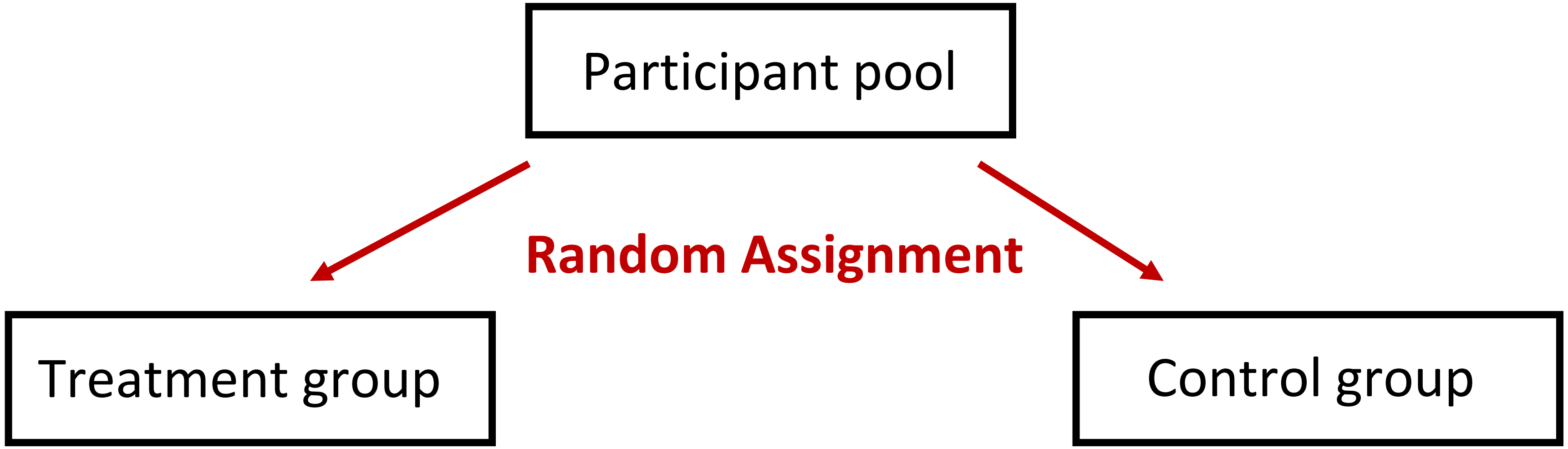
- Assume we are looking for differences in means between the groups

What would the statistic of interest be?

# 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. Write down the statistic of interest using appropriate symbols
   - $\overline{x}_{Effect} = \overline{x}_{Treatment} - \overline{x}_{Control}$

# 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!

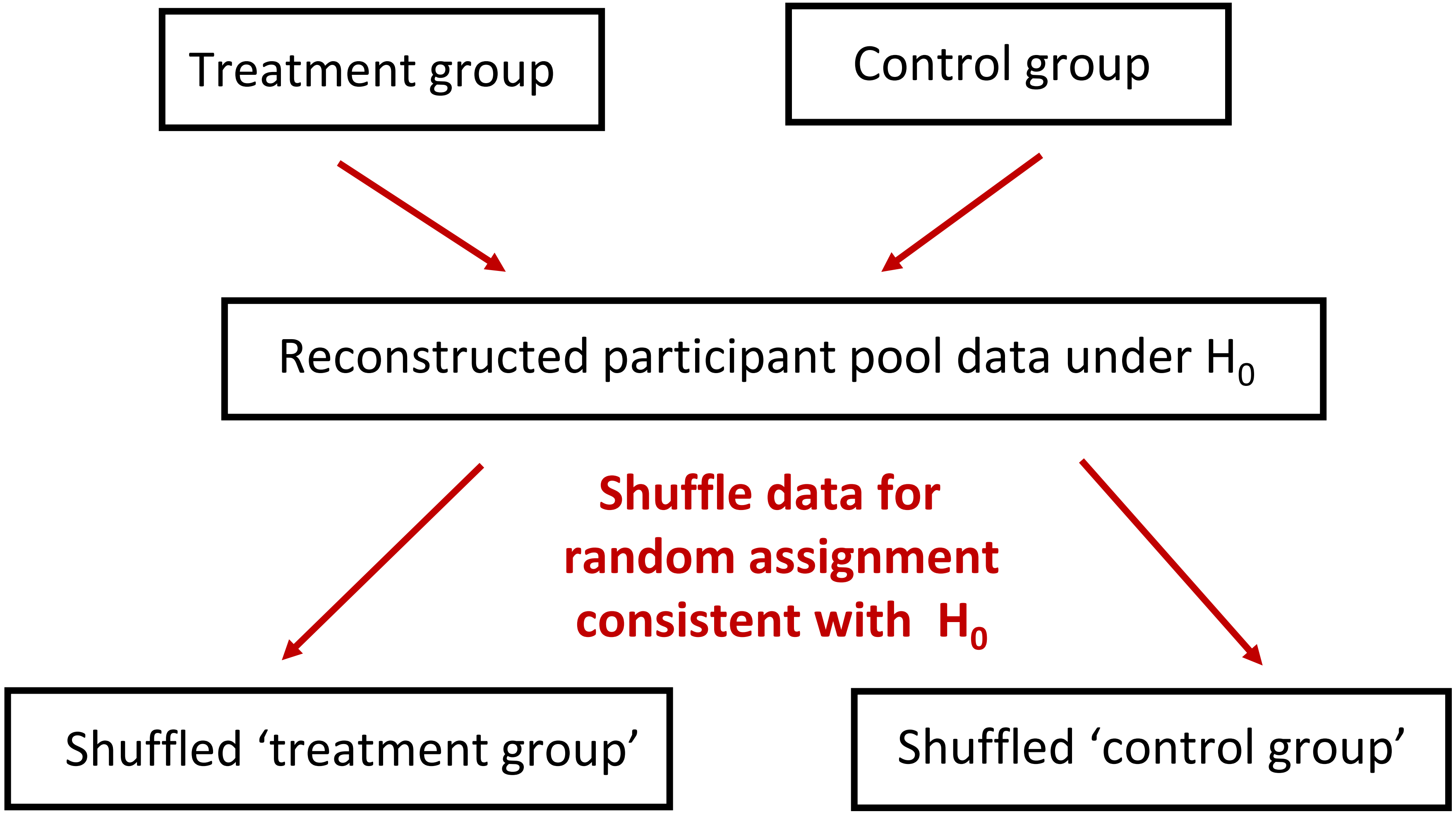
How could we create the null distribution?

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 ‘shuffled’ treatment group

4. Take the remaining points to the ‘shuffled’ control group

5. Compute the statistic of interest on these ‘shuffled’ 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

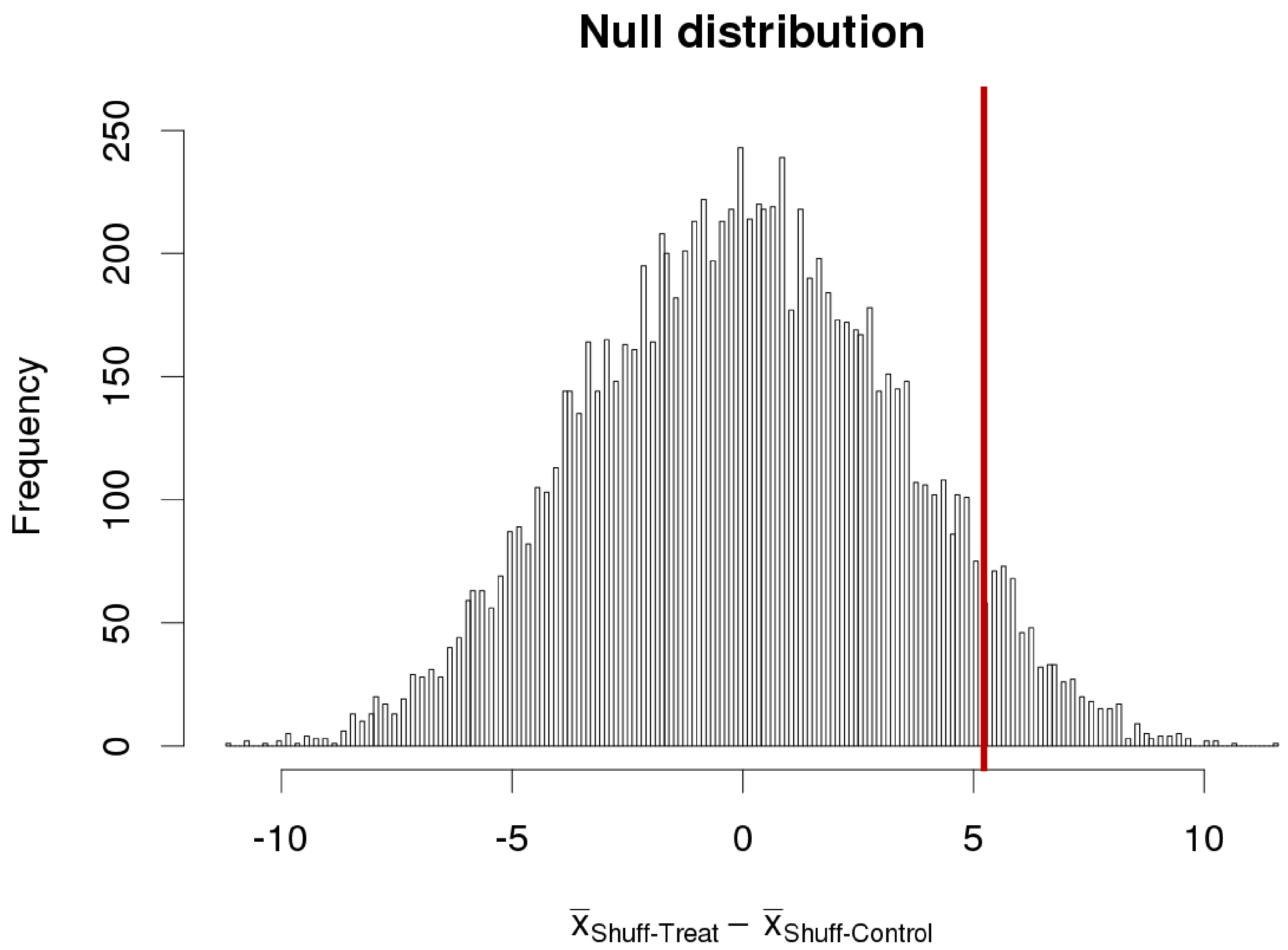
```
null_distribution <-  do_it(10000) * {

    # 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)

}
```




hist(null_distribution, breaks = 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!

# Another example of hypothesis tests comparing two means

# Do mice who eat late at night get fat?

A study by Fonken et al, 2010, wanted to examine whether more weight was gained by mice who could eat late at night

Mice were randomly divided into 2 groups:

- Dark condition: 8 mice were given 8 hours of darkness at night (when they couldn't eat)
- Light condition: 9 were constantly exposed to light for 24 hours (so they could always eat)

1. State the null and alternative hypothesis

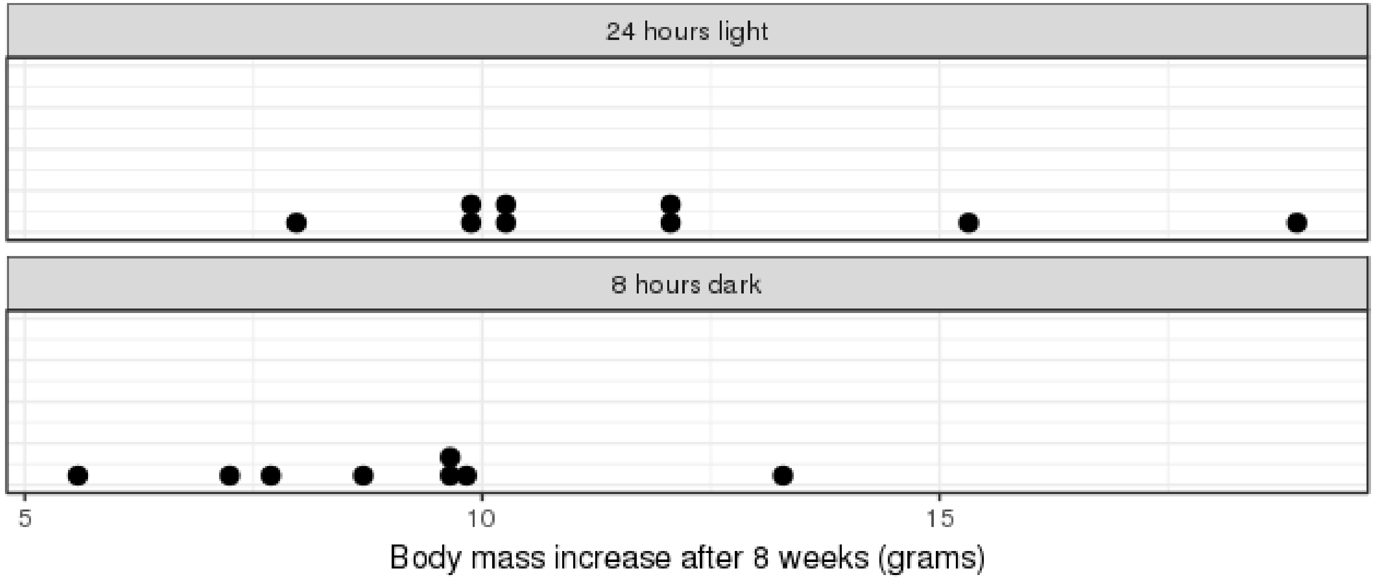
$H_0$: $\mu_{Light} = \mu_{Dark}$ or $\mu_{Light} - \mu_{Dark} = 0$

$H_A$: $\mu_{Light} > \mu_{Dark}$ or $\mu_{Light} - \mu_{Dark} > 0$

# Hypothesis tests for differences in two group means

What is step 2?

2. Calculate statistic of interest
    - $\overline{x}_{effect} = \overline{x}_{Light} - \overline{x}_{Dark}$



# Let's try it in R!

# Do mice who eat late at night get fat?

You can get the data from:

```
download_data("mice.Rda")
load("mice.Rda")

dark_BM_increase      # length(dark_BM_increase)
light_BM_increase     # length(light_BM_increase)
```

Can you calculate the observed statistic (step 2)?

```
obs_stat <- mean(light_BM_increase) - mean(dark_BM_increase)
```

What’s next?

# 3. Create the null distribution!

Mice in the light

Mice in the dark

Reconstructed sample of all mice

**Shuffle data for random assignment consistent with $H_0$**

Shuffled 'light group'

Shuffled 'dark group'

One null distribution statistic: $\bar{x}_{Shuff\_Dark} - \bar{x}_{Shuff\_Light}$

# Do mice who eat late at night get fat?

What is the first thing we need to do for creating the null distribution?

```
combo_data <- c(light_BM_increase, dark_BM_increase)
```

How do we create one point in our null distribution?

```
# shuffle the data
shuff_data  <- shuffle(combo_data)

# create fake light and dark data
shuff_light <- shuff_data[1:9]
shuff_dark <- shuff_data[10:17]

# compute fake statistic
mean(shuff_light) - mean(shuff_dark)
```

# Do mice who eat late at night get fat?

How do we create a full null distribution?

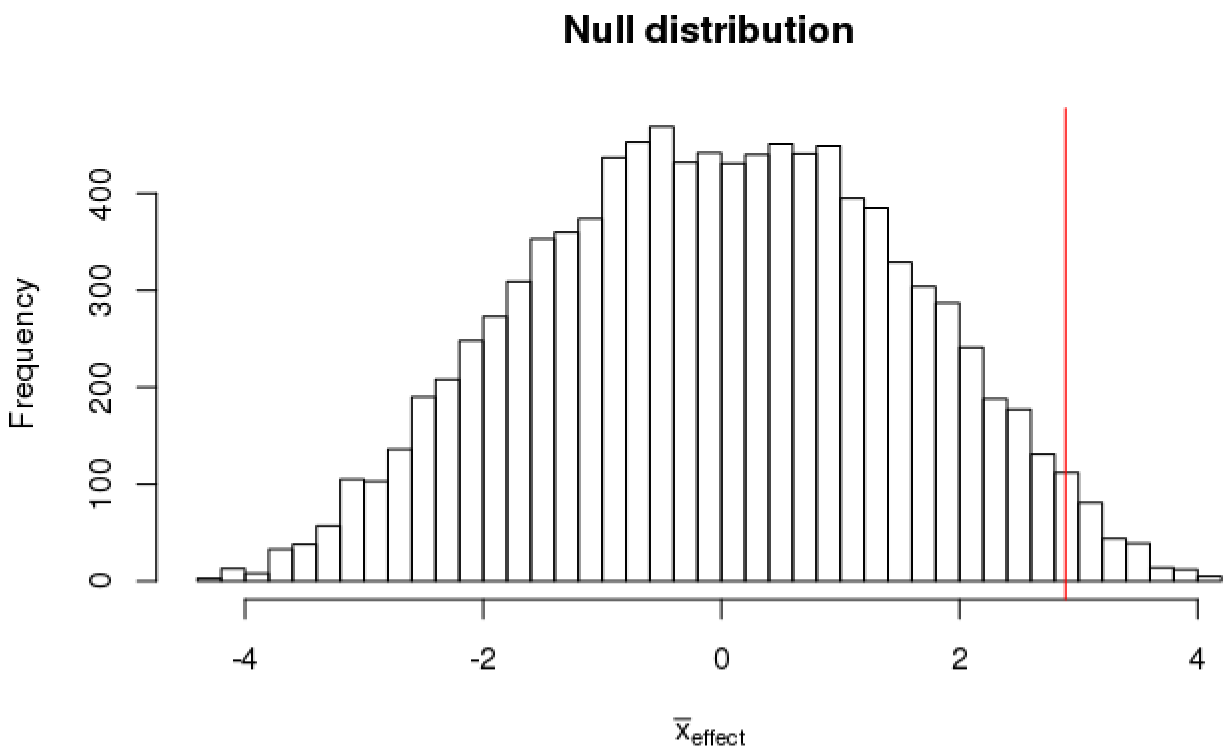
```
null_dist <- do_it(10000)  * {

        shuff_data <- shuffle(combo_data)
        shuff_light <- shuff_data[1:9]
        shuff_dark <- shuff_data[10:17]
        mean(shuff_light) - mean(shuff_dark)

}
```

# Do mice who eat late at night get fat?

Plot the null distribution: `hist(null_dist, breaks = 50)`




What do we do next?

# Do mice who eat late at night get fat?

Get the p-value

```
p_val <- pnull(obs_stat, null_dist, lower.tail = FALSE)
```

p-value =  0.02