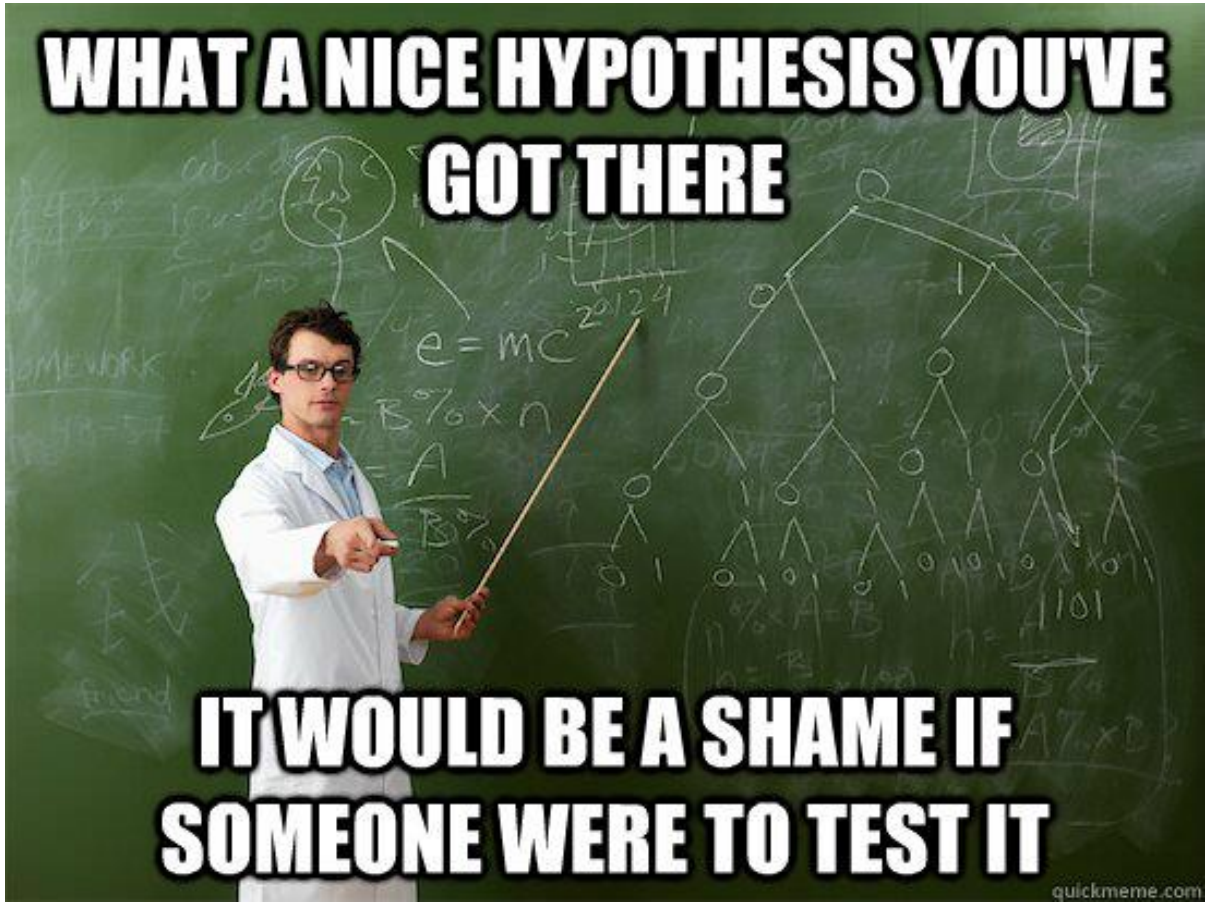
Randomization tests continued



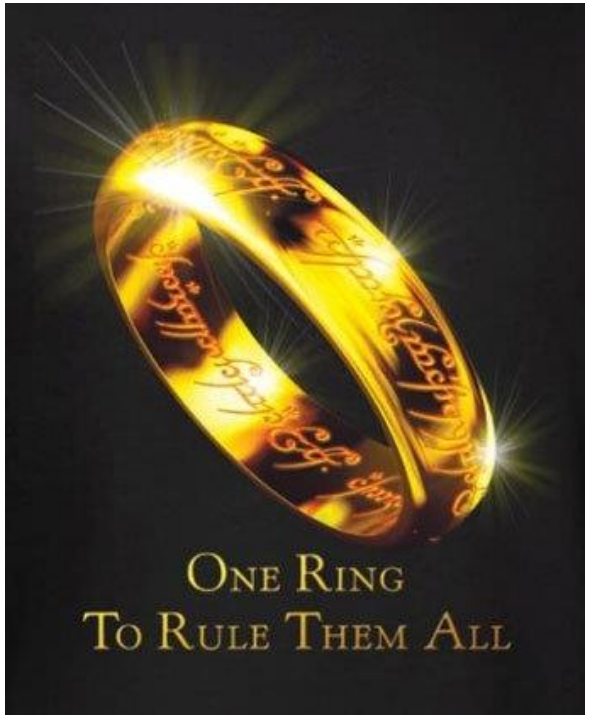
Overview

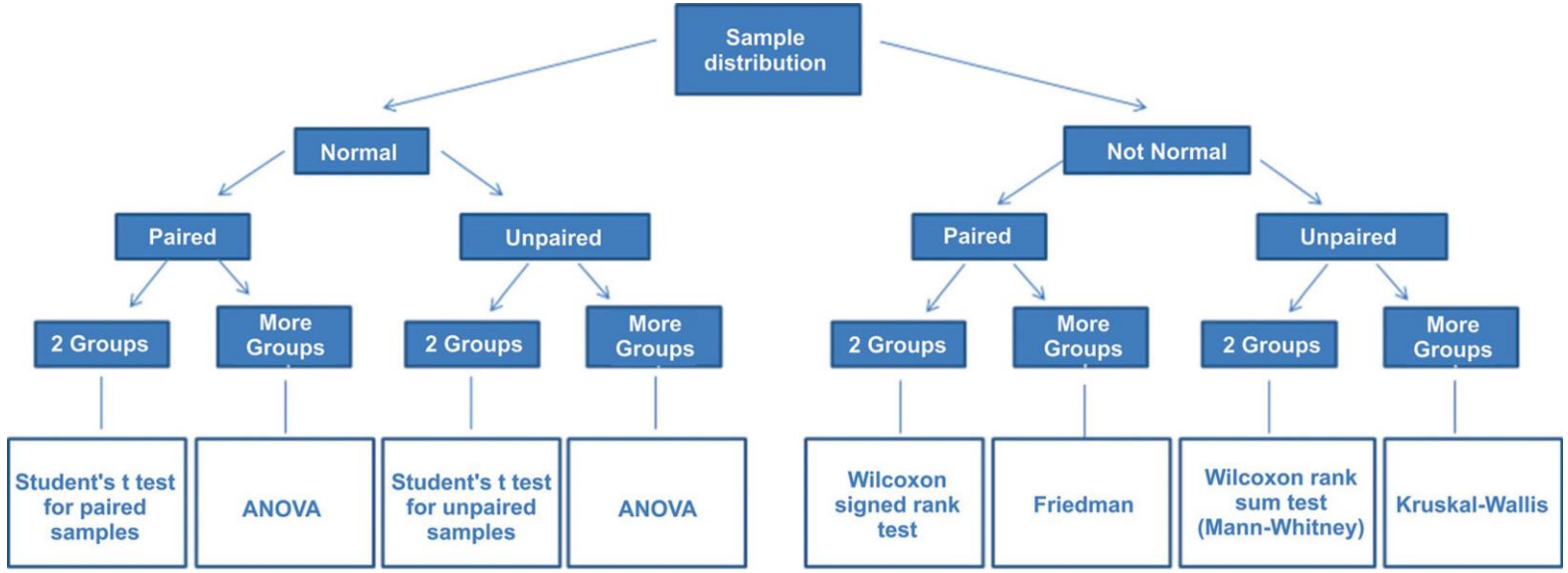
Quick review of hypothesis tests for a single proportion

Randomization tests for two means

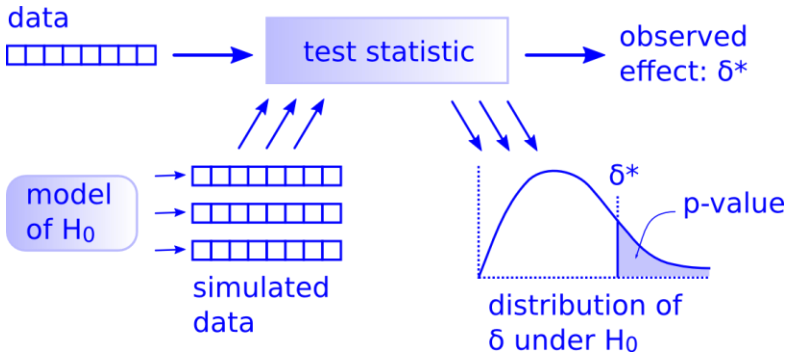
If there is time: Randomization tests for more than two means

The big picture: There is only one hypothesis test!

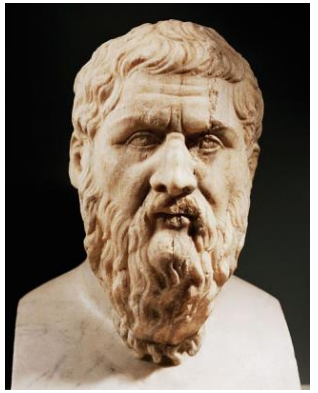




Just need to follow 5 steps!



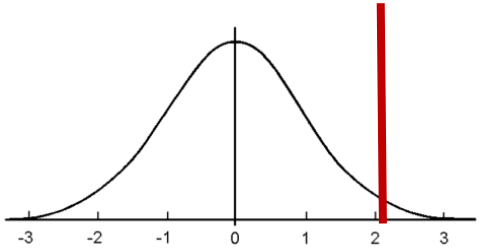
Five steps of hypothesis testing



1. State H_0 and H_A
   - Assume Gorgias (H_0) was right
   - α = .05 of the time he will be right, but we will say he is wrong

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

Review: hypothesis test for a single proportion

Joy Milne claimed to have the ability to smell whether someone had Parkinson's disease

To test this claim researchers gave Joy 6 shirts that had been worn by people who had Parkinson's disease and 6 shirts by people who did not.

Joy identified 11 out of the 12 shirts correctly.

Step 1: state the null and alternative hypotheses

- H_0: $\pi = 0.5$
- H_A: $\pi > 0.5$ ← H_0 and H_A need to be mutually exclusive

Review: hypothesis test for a single proportion

We can run a hypothesis test for a single proportion in R using:

```
obs_stat <- 11/12          #  Step 2: calculate the observed statistic

flip_sims_prop <- rbinom(10000, 12, .5)/12     # Step 3: create null distribution

p_value <- sum(flip_sims_prop >= obs_stat)/length(flip_sims)     # Step 4: p-value
```

p-value is 0.0029

Step 5: Should we reject H_0?

Do you really believe Joy can smell Parkinson’s disease?

TREATMENTS

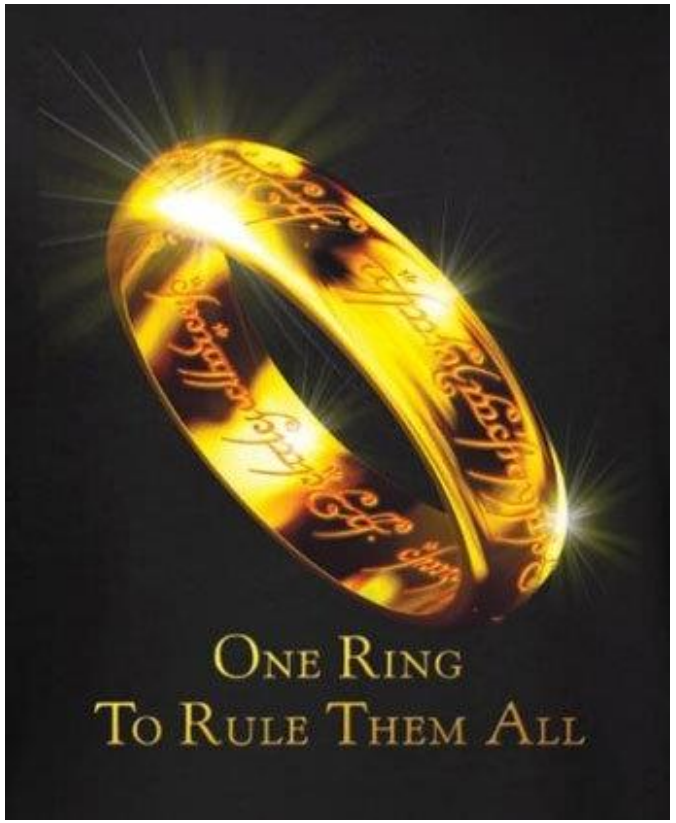
Her Incredible Sense Of Smell Is Helping Scientists Find New Ways To Diagnose Disease

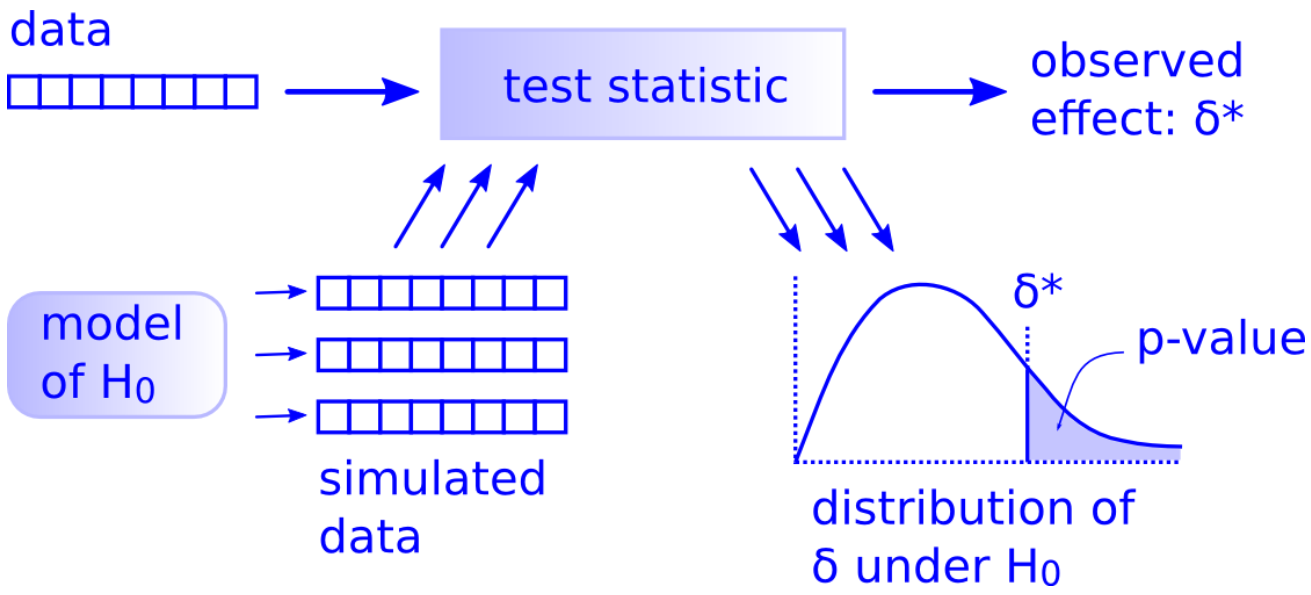
March 23, 2020 · 4:45 PM ET


Questions?

Hypothesis tests comparing 2 means

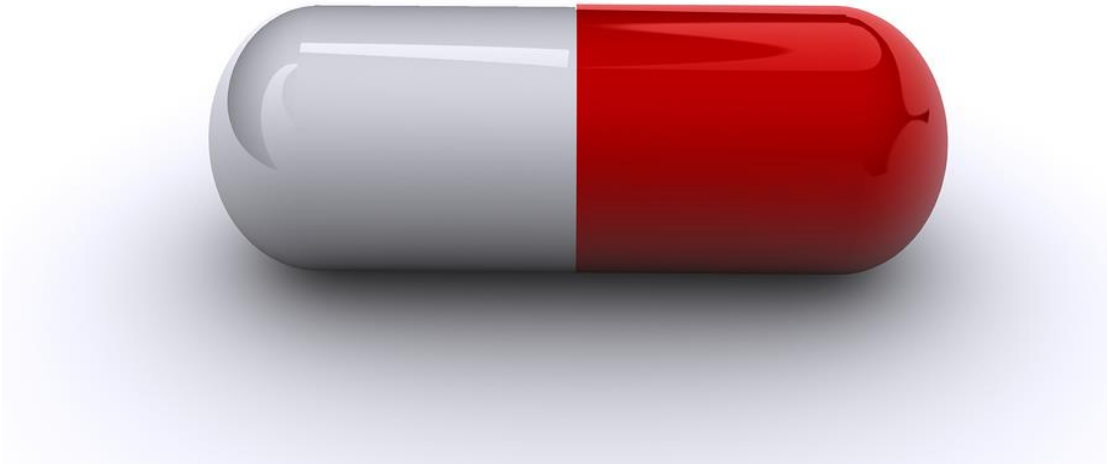
The big picture: There is only one hypothesis test!





Just need to follow 5 steps!

Hypothesis tests for comparing two means



Question: Is this pill effective?

Testing whether a pill is effective (on average)

How would we design a study?

What would the cases and variables be?

What would the parameter and statistic of interest be?

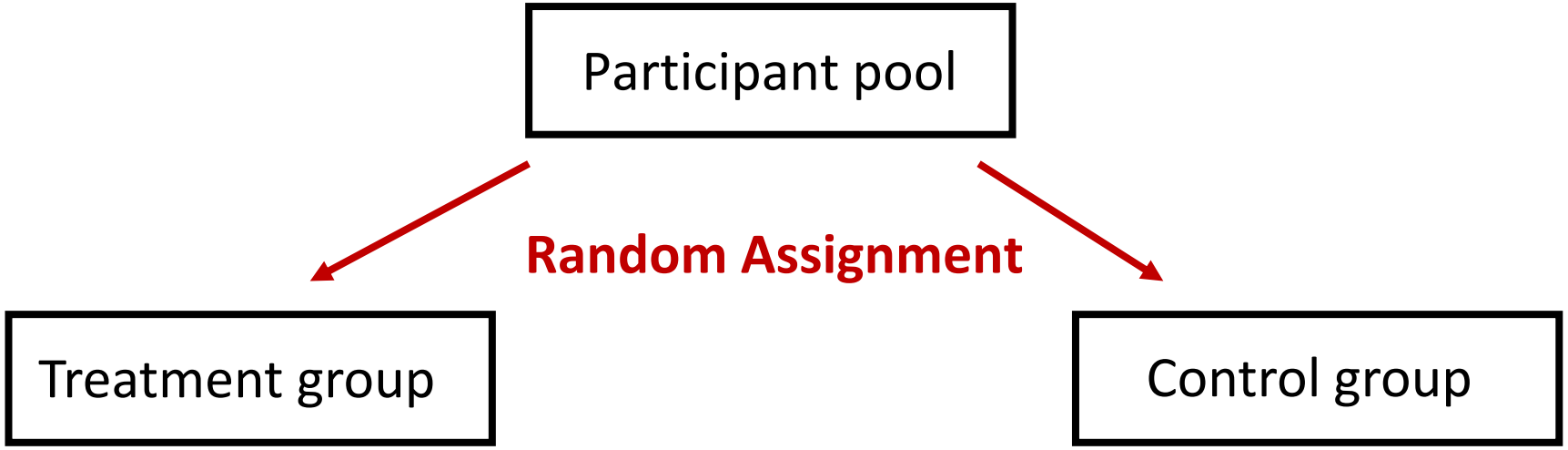
What are the null and alternative hypotheses?

- 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



Observational and experimental studies

An **observational study** is a study in which the researcher does not actively control the value of any variable but simply observes the values as they naturally exist.

An **experiment** is a study in which the researcher actively controls one or more of the explanatory variables

- **Random assignment** is where experimental units are randomly assigned to treatment and control groups which allows one to answer questions about **causation**!

Question: Which data are from observational studies?

- Most drug studies
- OkCupid data
- Joy Smelling Parkinson's

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

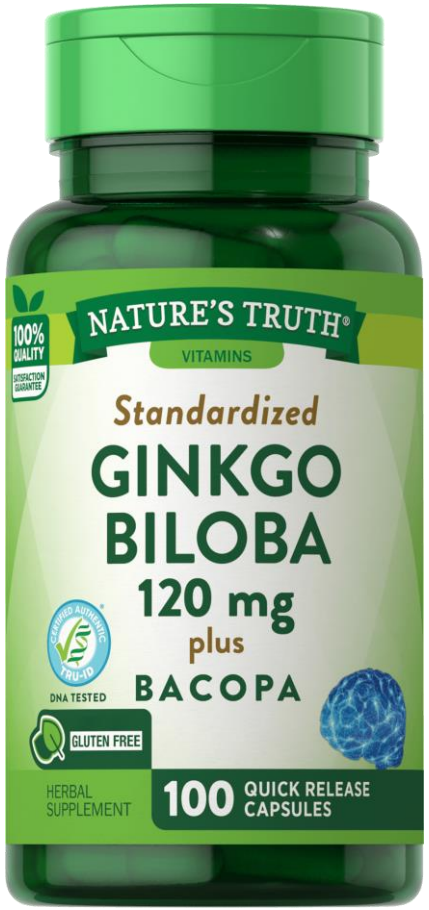
Example: Does Gingko improve memory?

A double-blind randomized controlled experiment by Solomon et al (2002) investigated whether taking a Ginkgo supplement could improve memory

- A treatment group of n = 104 participants took a Gingko supplement 3 times per day for 6 weeks
- A control group of n = 99 participants took a placebo 3 times per day for 6 weeks

Standardized neuropsychological tests of learning and memory, attention and concentration were measured at the end of the six week period

Question: Was there a difference in the mean cognitive score between the treatment and control groups?



1. State the null and alternative hypothesis

In words:

- **Null hypothesis**: The average memory score will be the same for participants who took Gingko and the placebo
- **Alternative hypothesis**: The average memory score will be different for the two groups.

In symbols:

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

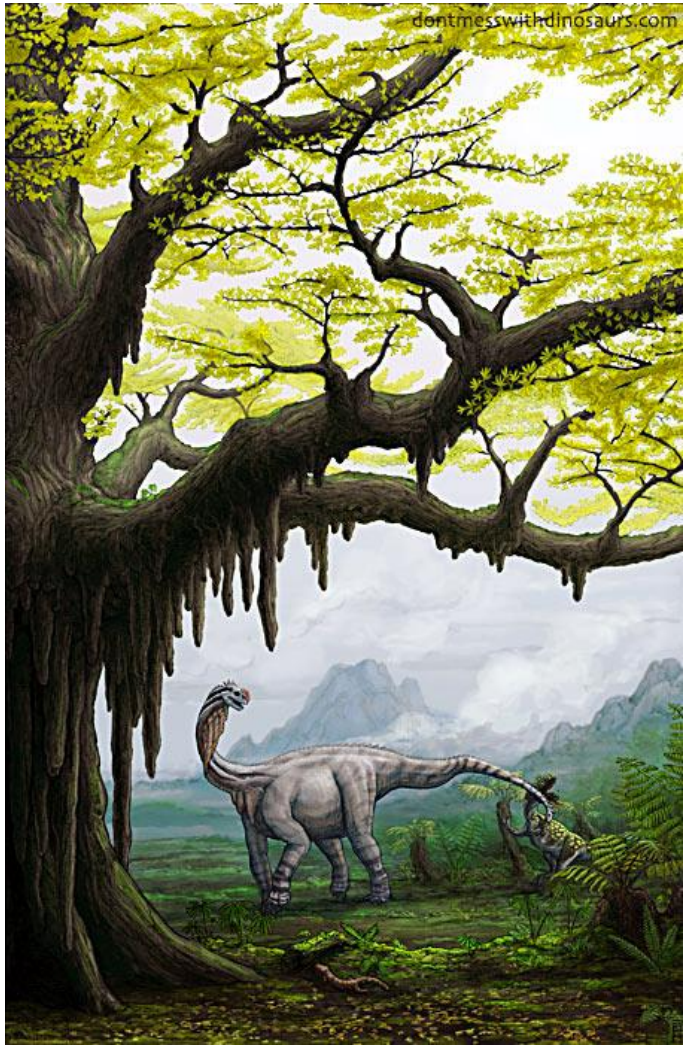
2. Visual the data can calculate the observed statistic

How could we visualize the data?

- We will try this in R soon…

What could we use for the observed statistic?

- $\bar{x}_{Effect} = \bar{x}_{Treatment} - \bar{x}_{Control}$
- $\bar{x}_{Effect} = \bar{x}_{Gingko} - \bar{x}_{Placebo}$



3. Create the null distribution!

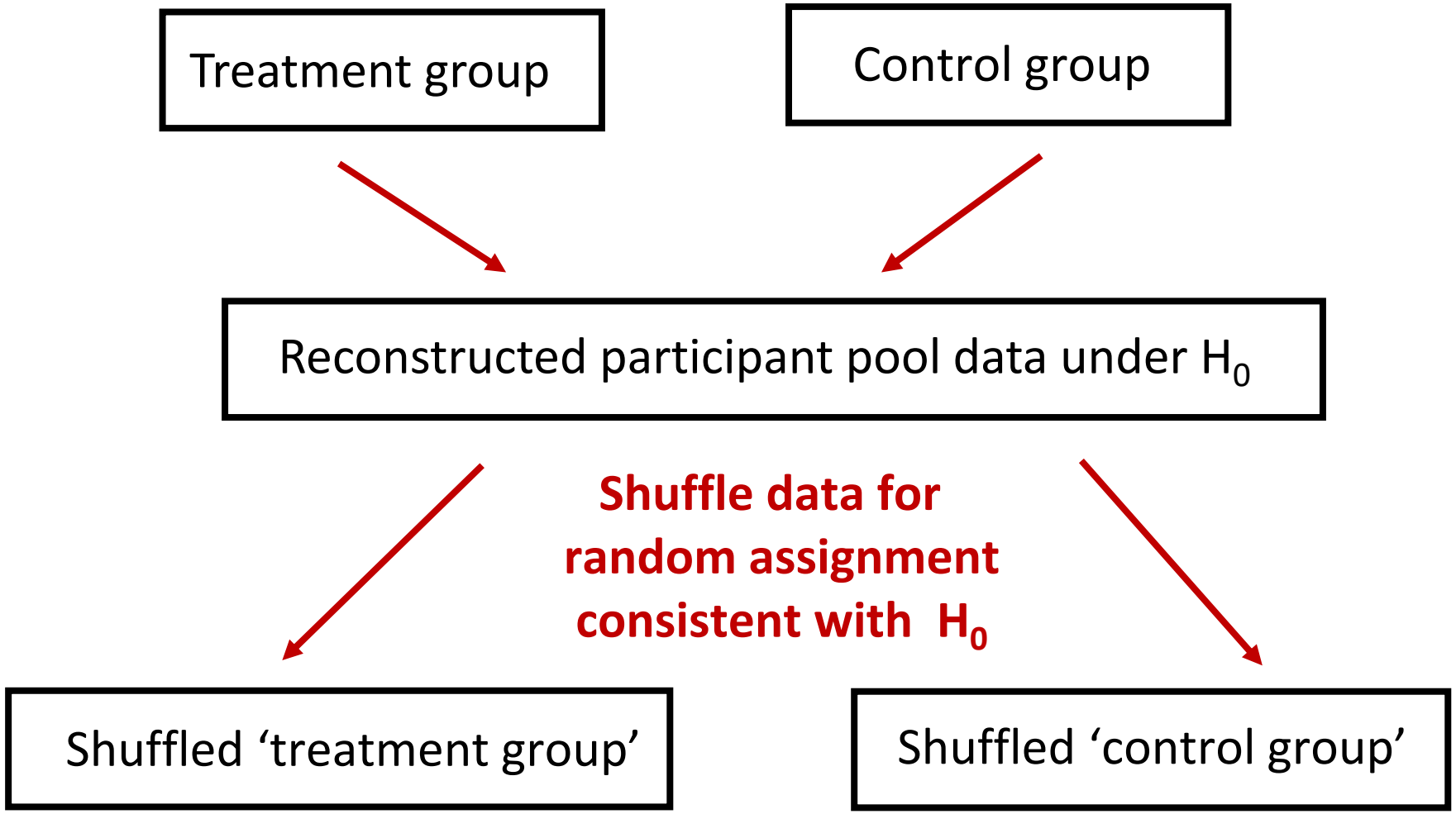
How could we create the null distribution?

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 104 points to be the ‘null’ treatment group

4. Take the remaining 99 points to the ‘null’ control group

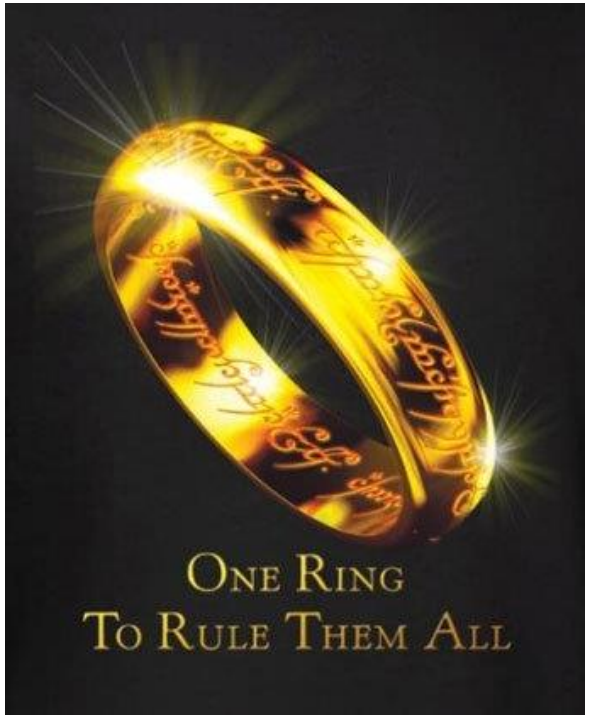
5. Compute the statistic of interest on these ‘null’ groups

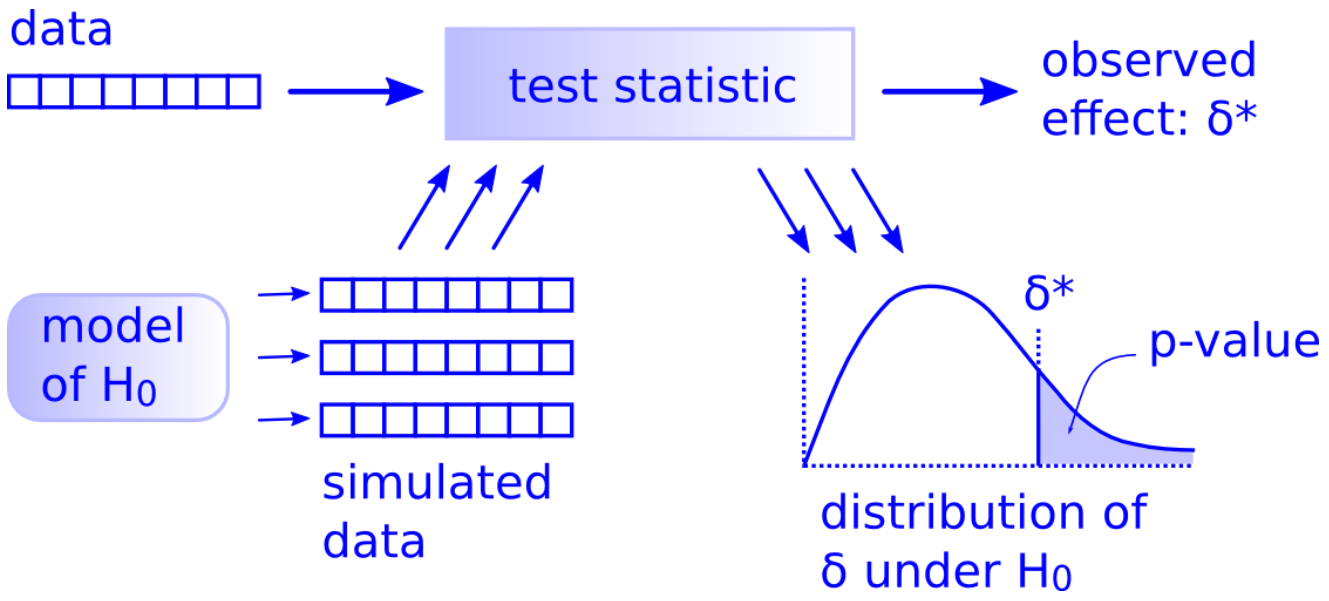
6. Repeat 10,000 times to get a null distribution

Let’s try the rest of the hypothesis test in R…

Hypothesis test for comparing more than two means

The big picture: There is only one hypothesis test!

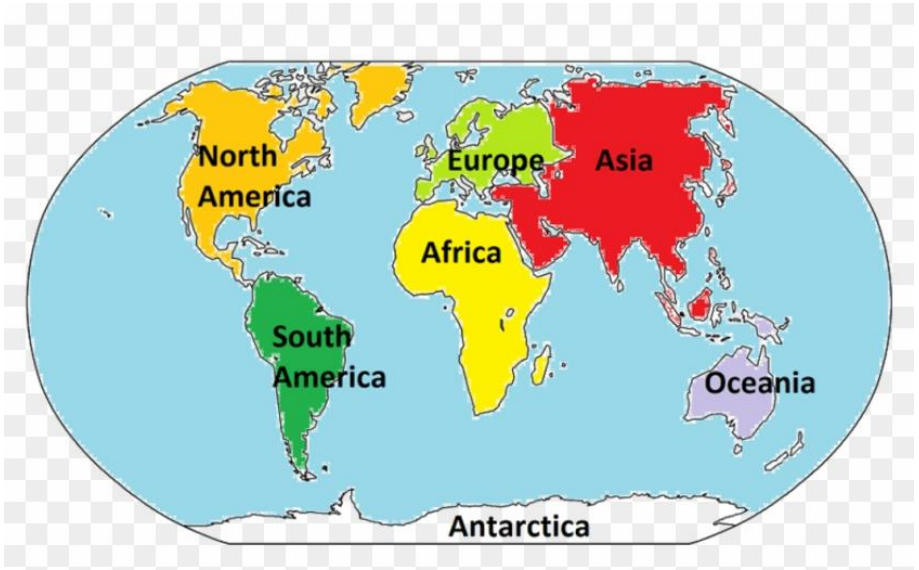




Just need to follow 5 steps!

Comparing more than two means

Let's examine the beer consumption in different continents!



Analysis inspired by:

- Minitab blog article
- Five thirty eight analysis

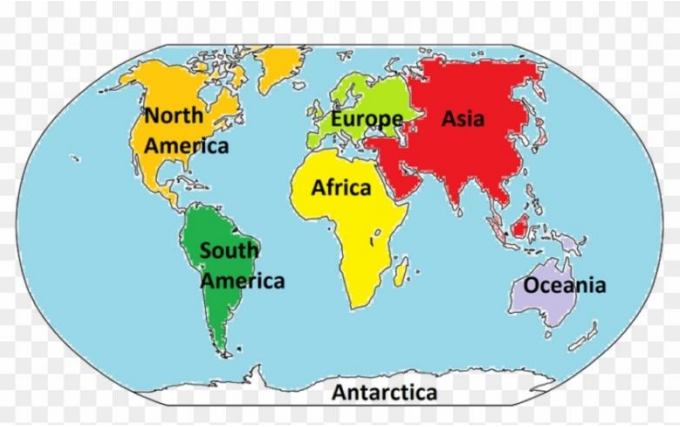
Question: Does the average beer consumption in countries different depending on the continent?

1. State the null and alternative hypotheses!

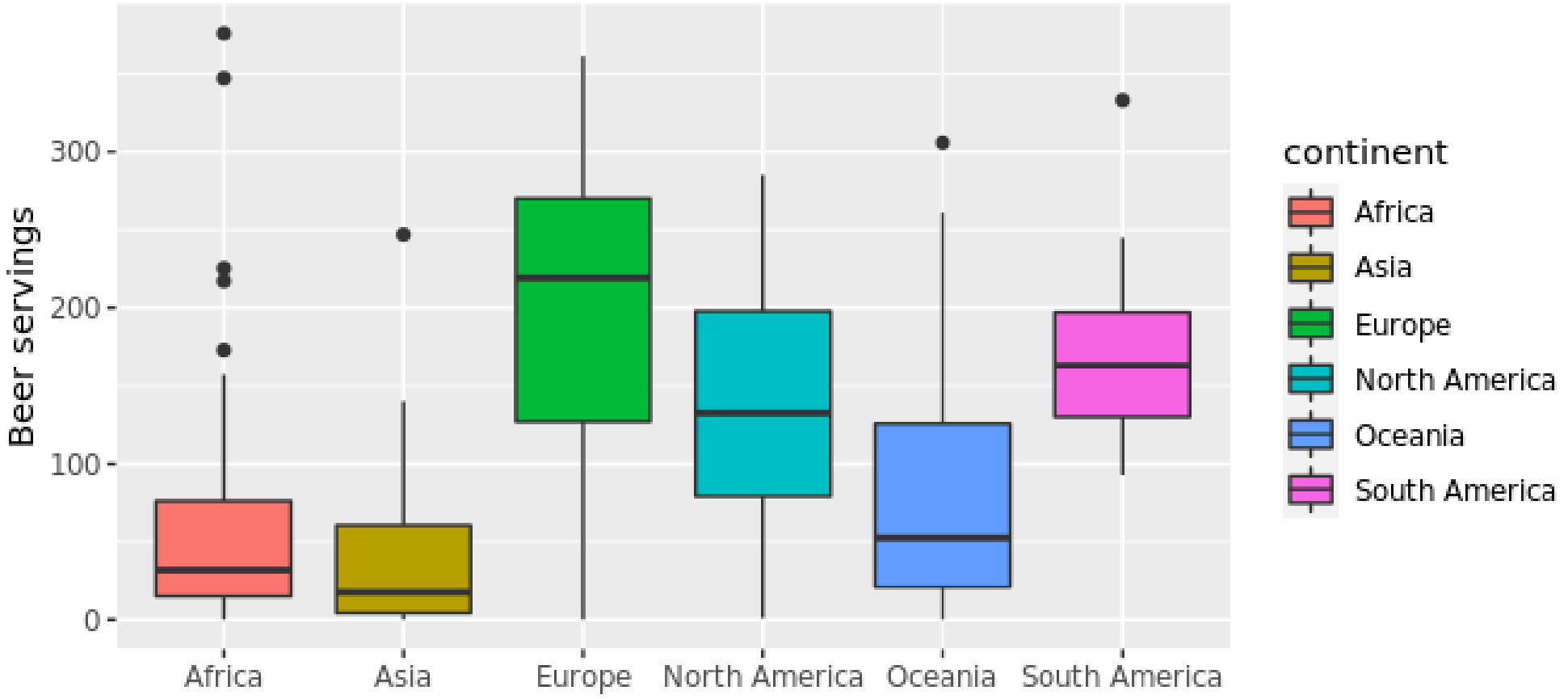
$\mathbf{H_0}$: $\mu_{Asia} = \mu_{Europe} = \mu_{Africa} = \mu_{North\text{-}America} = \mu_{South\text{-}America} = \mu_{Oceania}$

$\mathbf{H_A}$: $\mu_i \neq \mu_j$ for at least one pair of continents

What should we do next?



Plot of the beer consumption in different continents



Thoughts on the statistic of interest?

Comparing multiple means

There are many possible statistics we could use. A few choices are:

1. Group range statistic:

   $\max \bar{x} - \min \bar{x}$

2. Mean absolute difference (MAD):

   $(|\bar{x}_{\text{Africa}} - \bar{x}_{\text{Asia}}| + |\bar{x}_{\text{Africa}} - \bar{x}_{\text{Europe}}| + \ldots + |\bar{x}_{\text{Oceania}} - \bar{x}_{\text{South-America}}|)/15$

3. F statistic:

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

Using the MAD statistic

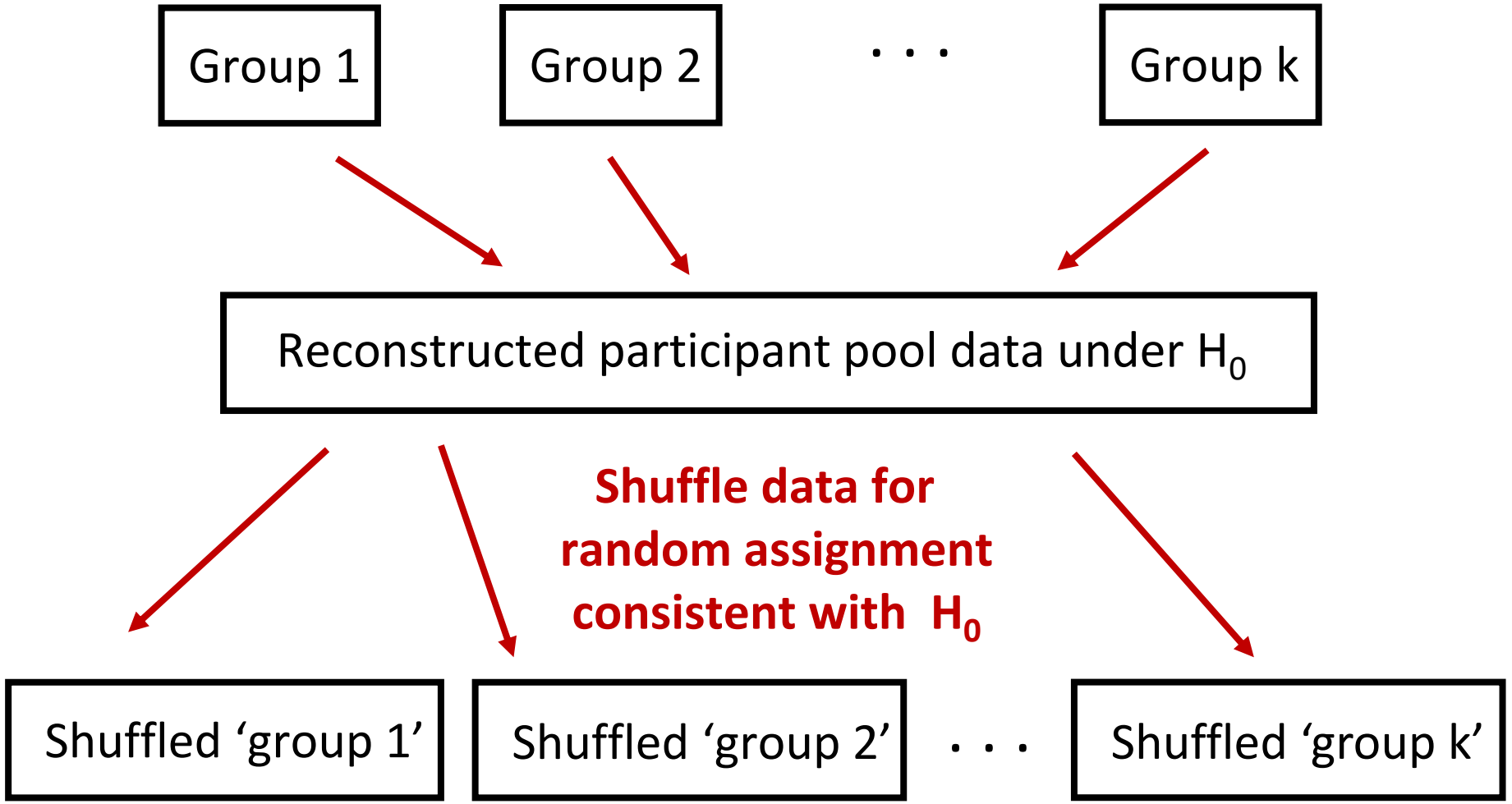
Mean absolute difference (MAD):

$$(|\bar{x}_{Africa} - \bar{x}_{Asia}| + |\bar{x}_{Africa} - \bar{x}_{Europe}| + \ldots + |\bar{x}_{Oceania} - \bar{x}_{South\text{-}America}|)/15$$

Observed statistic value = 78.86

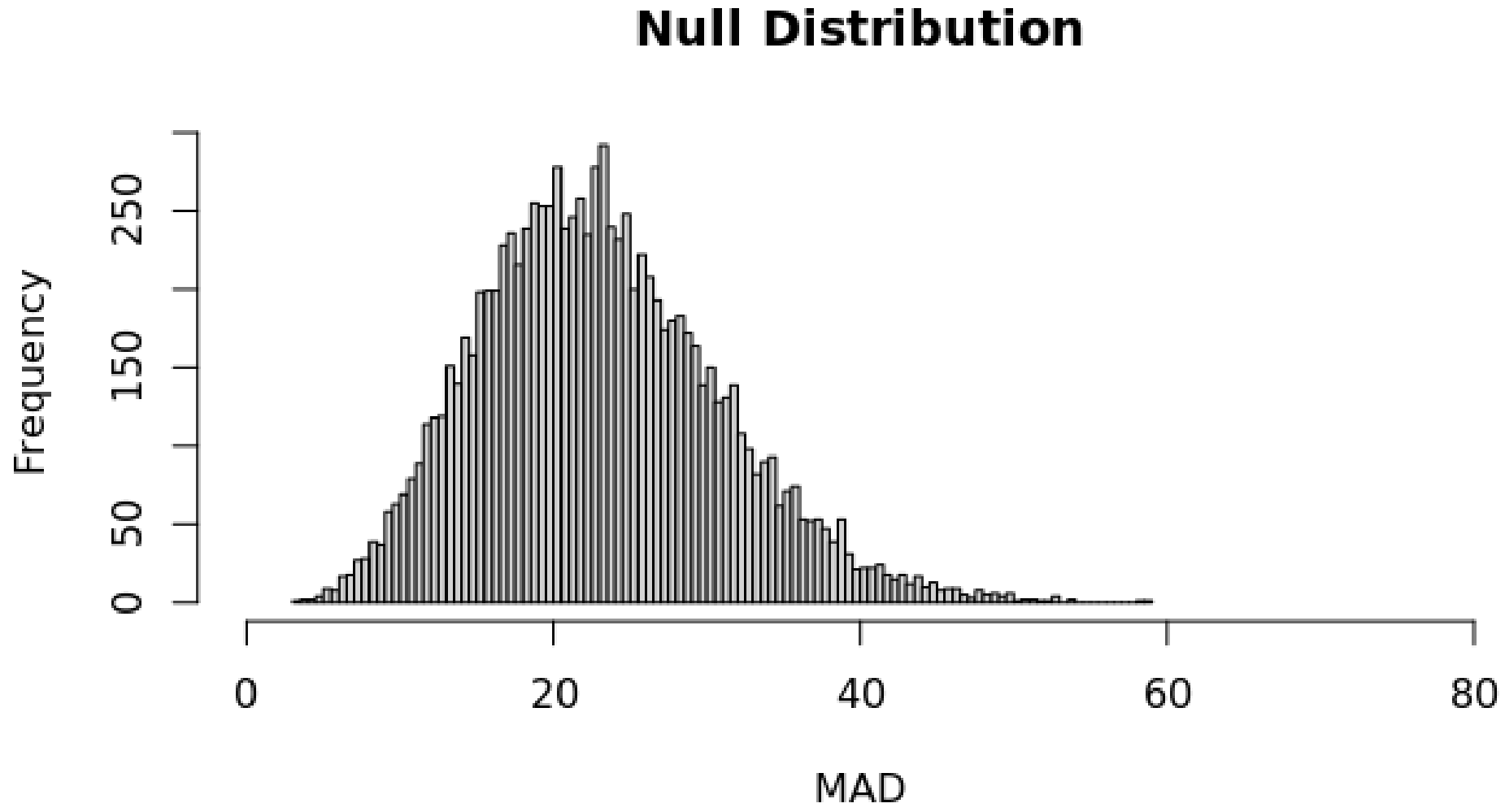
How can we create the null distribution?

3. Create the null distribution!

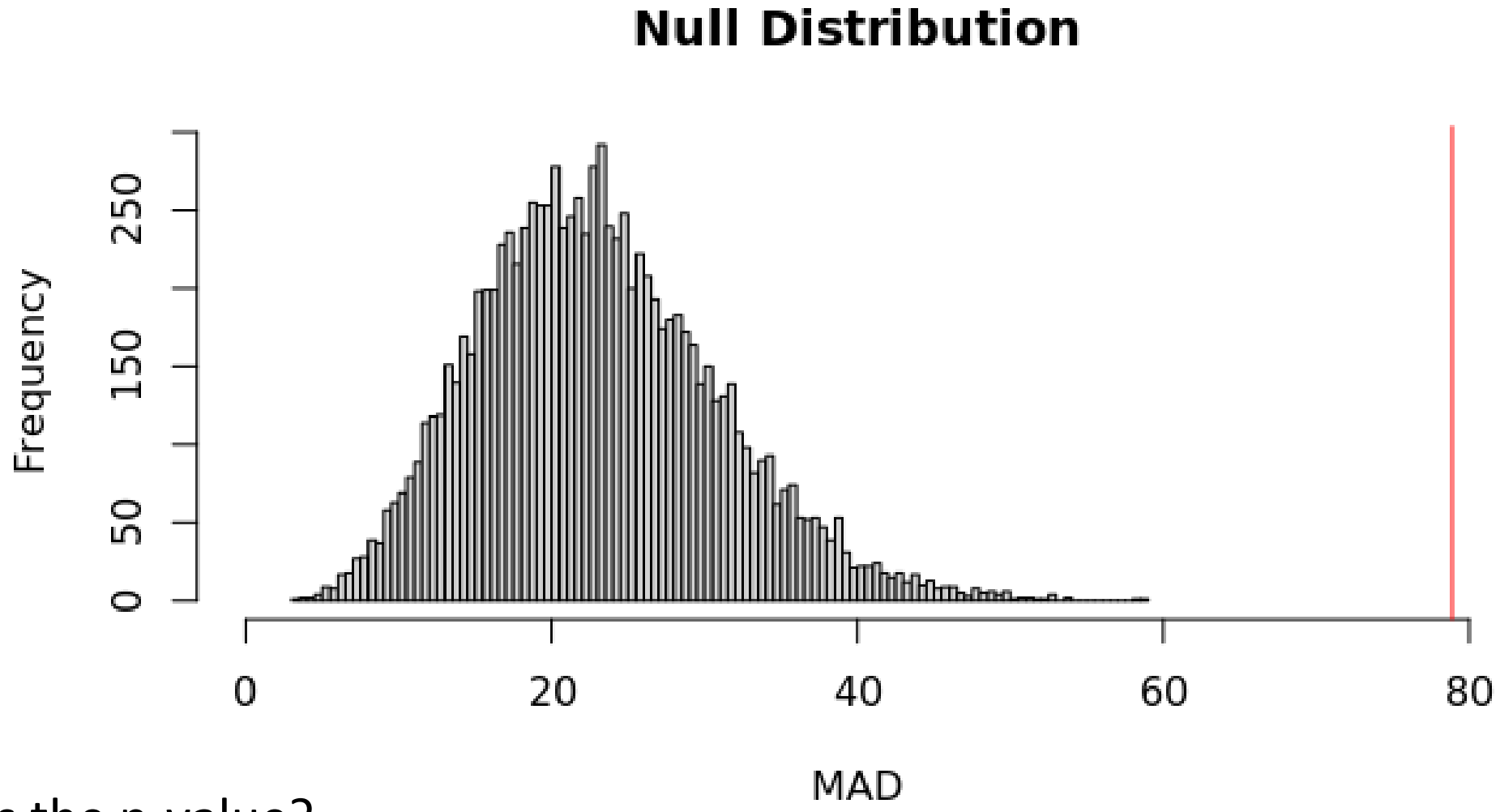


Compute statistics from shuffled groups

3. Create the null distribution!



4. Calculate the p-value



What is the p-value?

Conclusions?

Let’s try it in R…