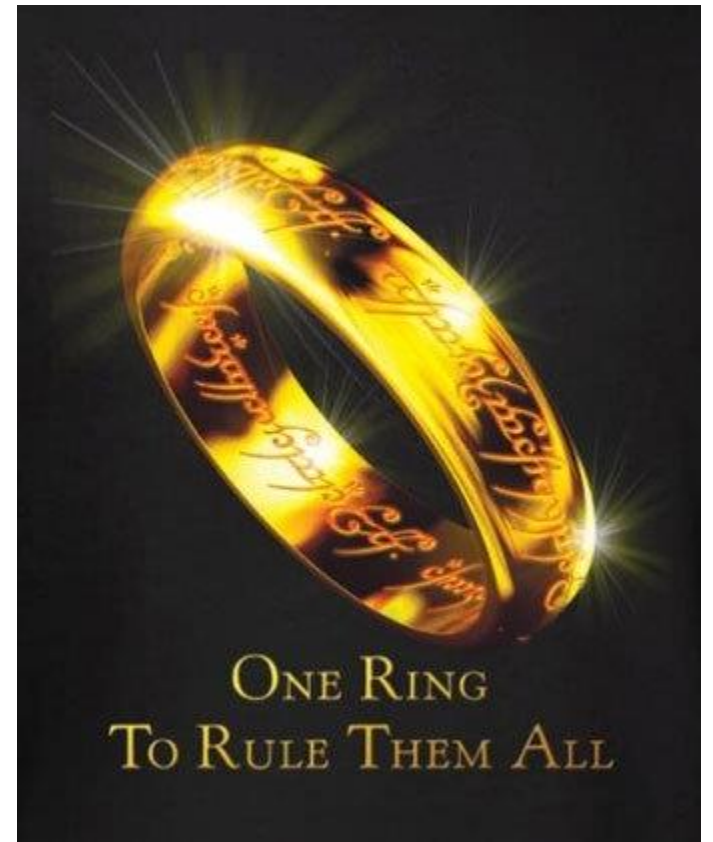
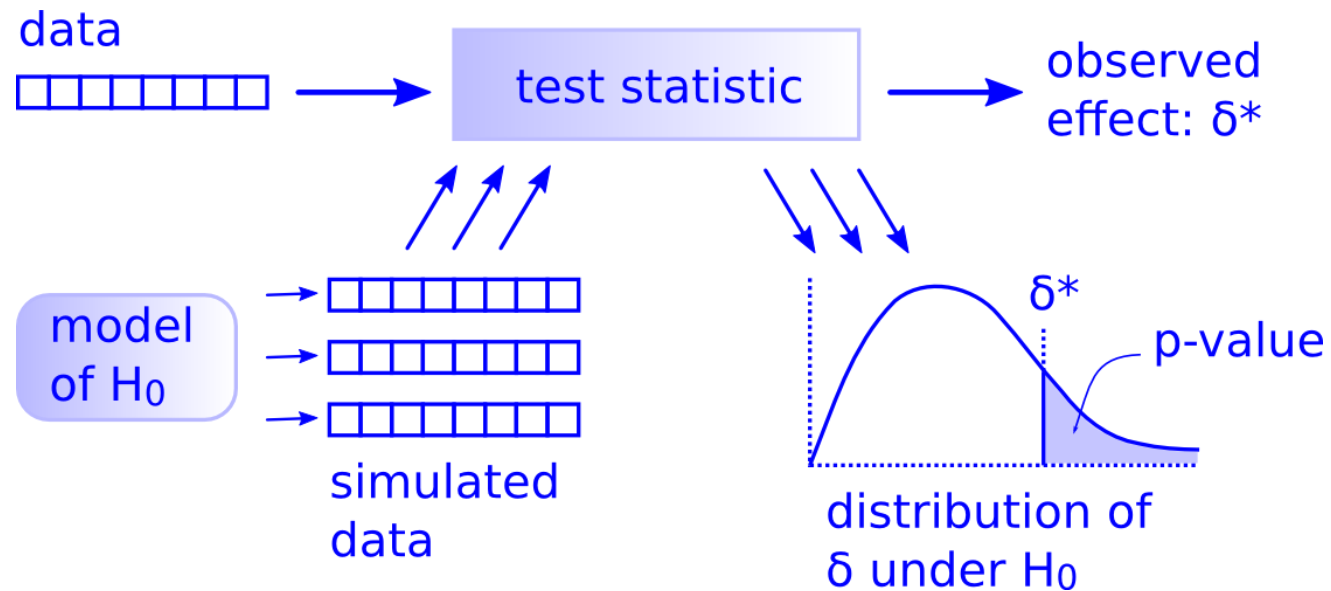


# Hypothesis tests for more than two means

	5	3	2		7			8
6		1	5					2
2			9	1	3		5	
7	1	4	6	9	2			
	2						6	
			4	5	1	2	9	7
	6		3	2	5			9
1					6	3		4
8			1		9	6	7	

# The big picture of hypothesis testing

There is only one [hypothesis test](#)!

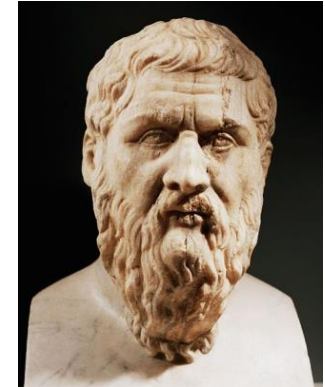


Just follow the 5 hypothesis tests steps!

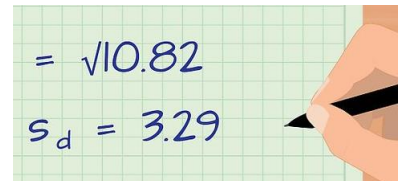
# Five steps of hypothesis testing

## 1. State $H_0$ and $H_A$

- Assume Gorgias ( $H_0$ ) was right
- $\alpha = .05$  of the time he will be right, but we will say he is wrong



## 2. Calculate the actual observed statistic

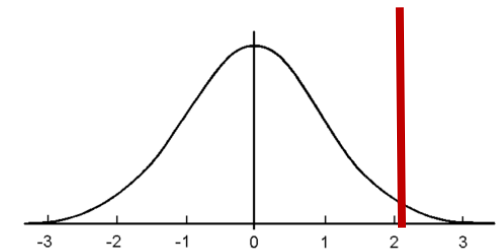
A hand holding a black pen, writing on a green grid background. The text shows the calculation of a standard deviation.
$$= \sqrt{10.82}$$
$$s_d = 3.29$$

## 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



# Comparing more than two means

A group of Hope College students wanted to see if there was an association between a student's major and the time it takes to complete a small Sudoku-like puzzle

	5	3	2		7			8
6		1	5					2
2			9	1	3		5	
7	1	4	6	9	2			
	2						6	
			4	5	1	2	9	7
	6		3	2	5			9
1					6	3		4
8			1		9	6	7	

# Comparing more than two means

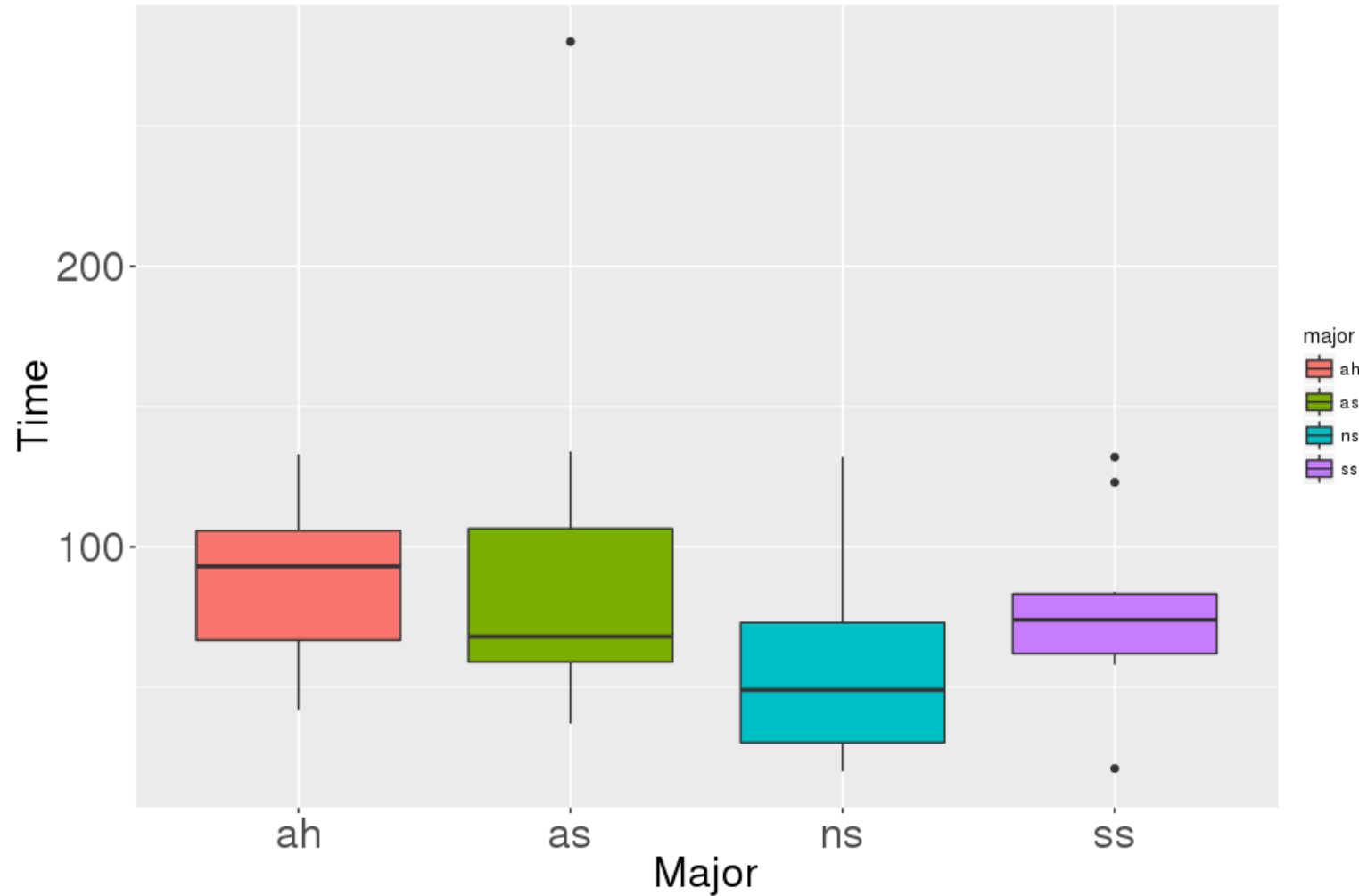
A group of Hope College students wanted to see if there was an association between a student's major and the time it takes to complete a small Sudoku-like puzzle

They grouped majors into four categories

- Applied science (as)
- Natural science (ns)
- Social science (ss)
- Arts/humanities (ah)

What is a good first thing to do to analyze the data?

# Step 0: Plot of completion time by major



What should we do next?

# Sudoku by field

1. State the null and alternative hypotheses!

$$H_0: \mu_{as} = \mu_{ns} = \mu_{ss} = \mu_{ah}$$

$$H_A: \mu_i \neq \mu_j \text{ for one pair of fields of study}$$

What should we do next?

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):

$$(|\bar{x}_{as} - \bar{x}_{ns}| + |\bar{x}_{as} - \bar{x}_{ss}| + |\bar{x}_{as} - \bar{x}_{ah}| + |\bar{x}_{ns} - \bar{x}_{ss}| + |\bar{x}_{ns} - \bar{x}_{ah}| + |\bar{x}_{ss} - \bar{x}_{ah}|)/6$$

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

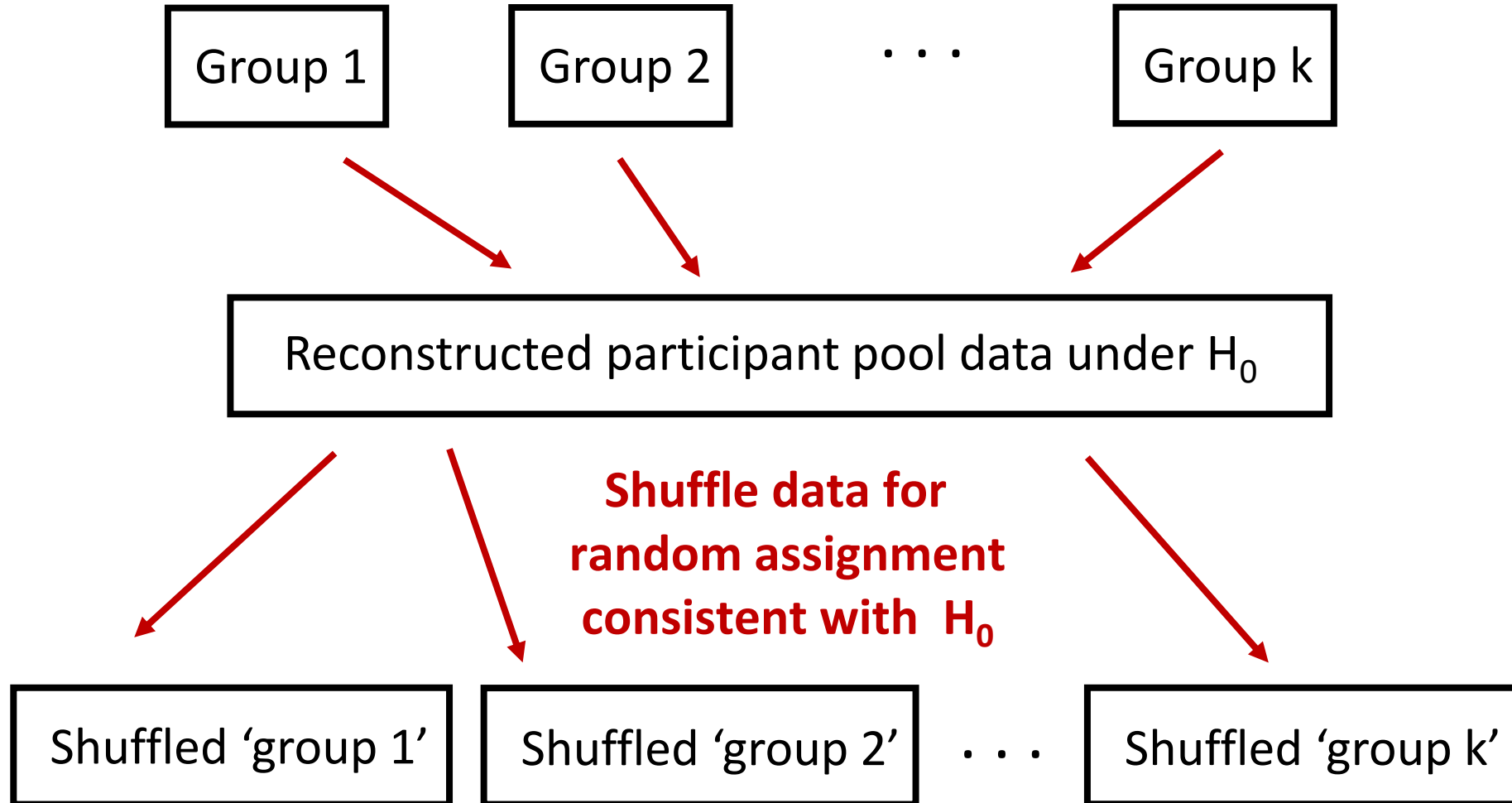
Mean absolute difference (MAD):

$$(|\bar{x}_{as} - \bar{x}_{ns}| + |\bar{x}_{as} - \bar{x}_{ss}| + |\bar{x}_{as} - \bar{x}_{ah}| + |\bar{x}_{ns} - \bar{x}_{ss}| + |\bar{x}_{ns} - \bar{x}_{ah}| + |\bar{x}_{ss} - \bar{x}_{ah}|)/6$$

Observed statistic value = 13.92

How can we create the null distribution?

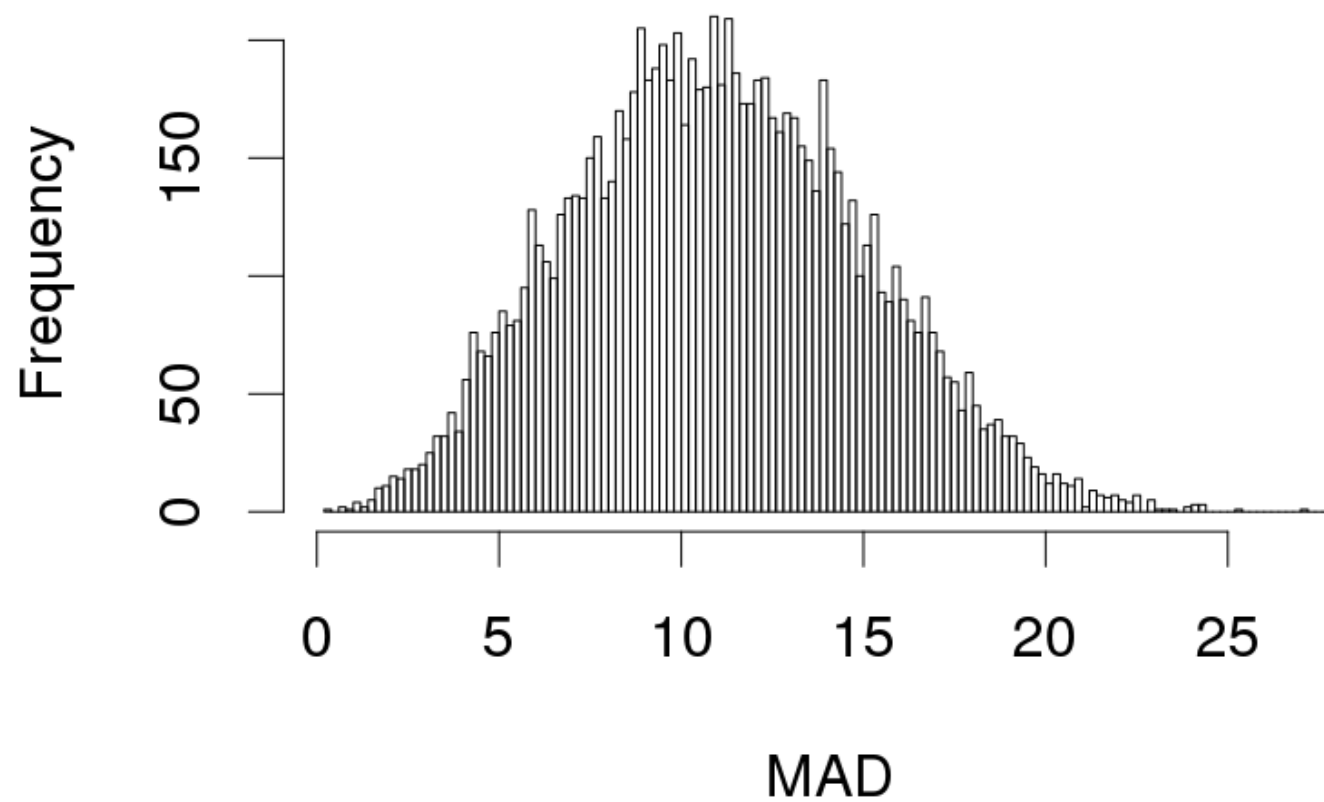
### 3. Create the null distribution!



Compute statistics from shuffled groups

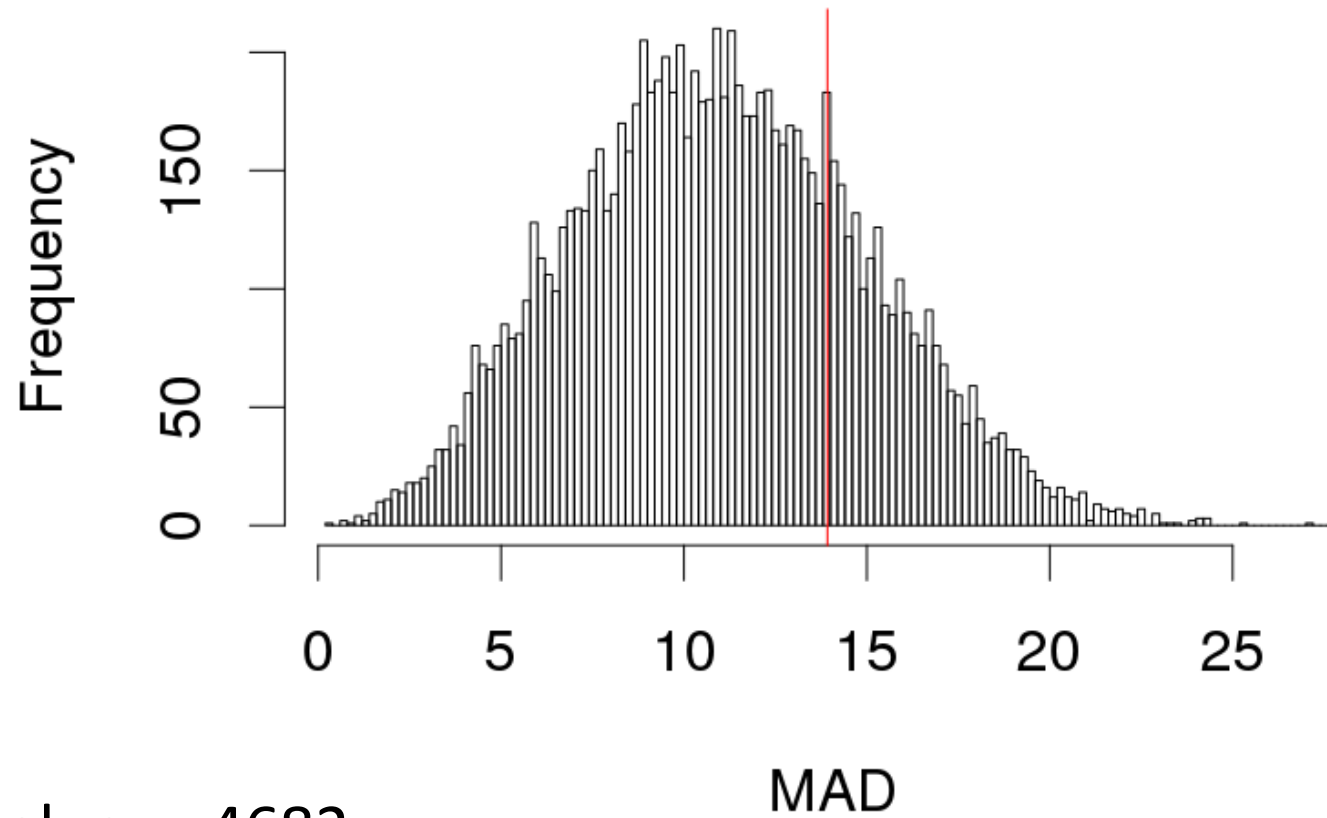
# Null distribution

## Null Distribution



# P-value

## Null Distribution



p-value = .4682

# Conclusions?



# Let's use R to compare levels of mercury in fish



Perch



Orange roughy



Bass



Halibut

