Parametric hypothesis tests

Overview

Permutation test for more than 2 means

Tests for two means

- Randomization tests for two means revisited using a t-statistic
- Parametric tests and t-tests

Where we are in the plan for the semester

Analysis | **R**

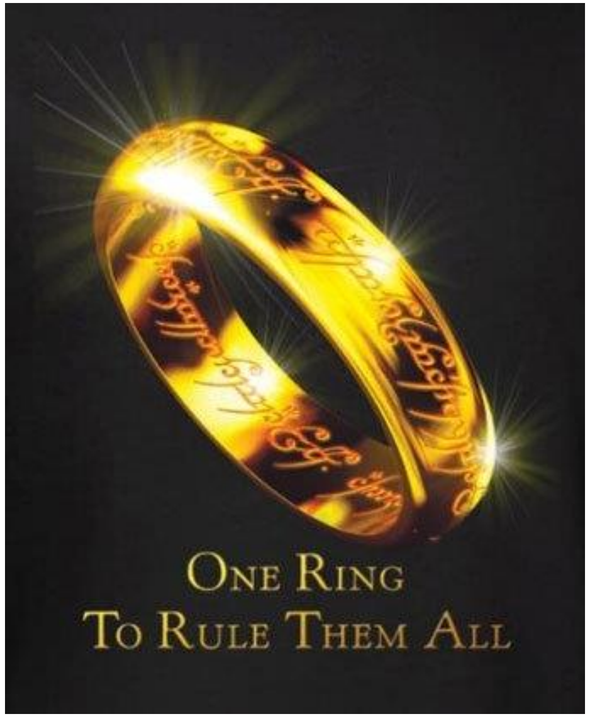
1	Sep 1	Course overview, introduction to R, descriptive statistics
2	Sep 6-8	Review of central statistical concepts and exploratory analysis using R
3	Sep 13-15	Confidence Intervals and the bootstrap
4	Sep 20-22	Review of hypothesis tests and permutation tests in R
5	Sep 27-29	Parametric hypothesis tests and theories of hypothesis testing

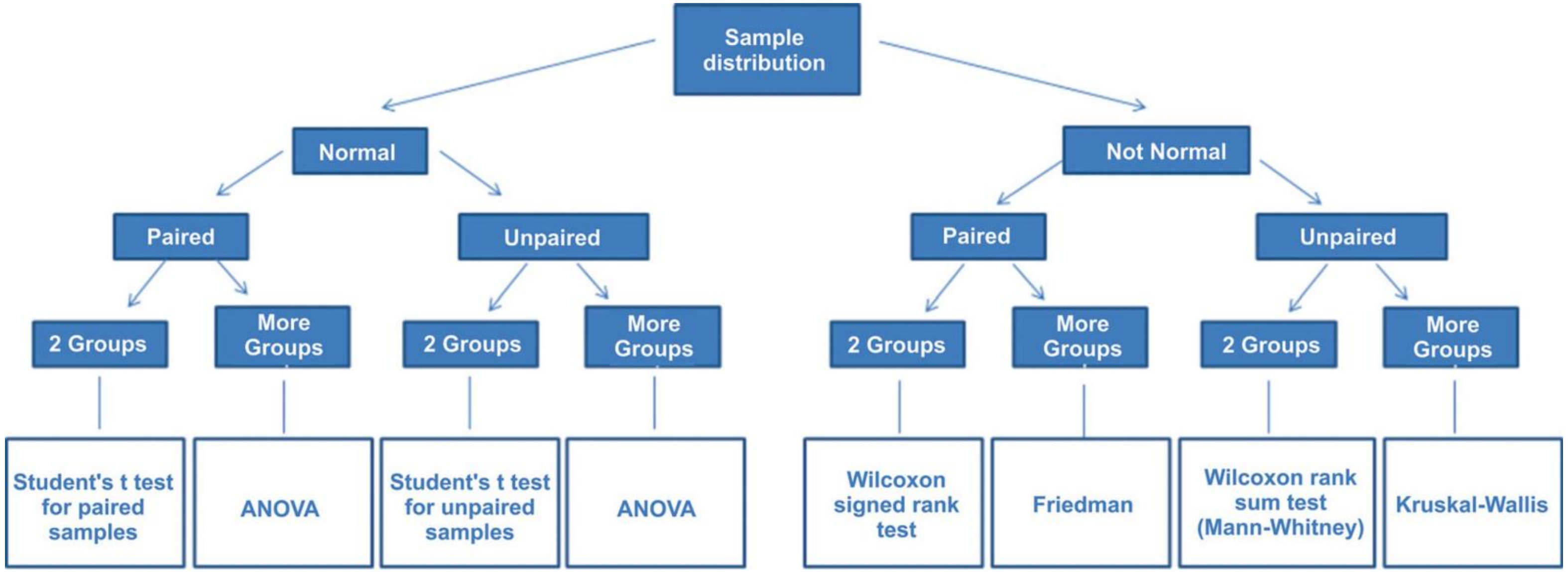
resampling methods

base R

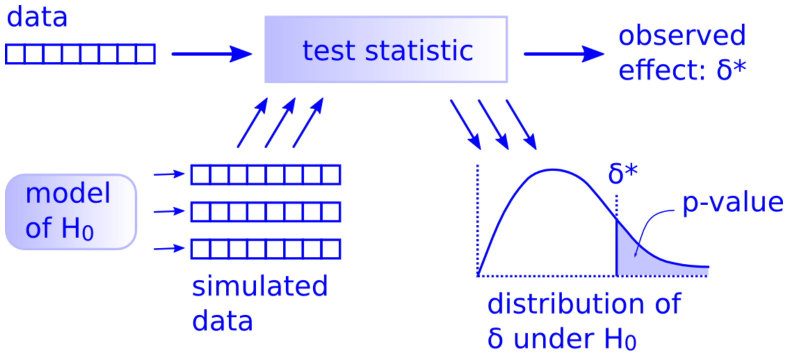
t-tests	78 respondents	69 %
confidence intervals	87 respondents	77 %
the bootstrap	28 respondents	25 %
permutation tests	17 respondents	15 %
one-way ANOVA	41 respondents	36 %

The big picture: There is only one hypothesis test!

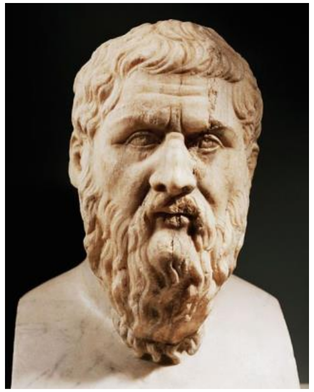




Just need to follow 5 steps!



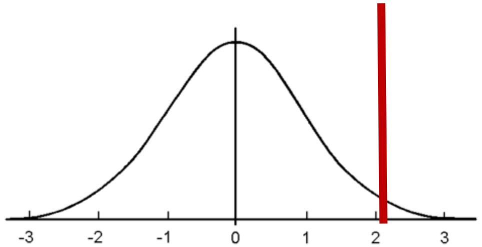
Review: Five steps of hypothesis testing



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)

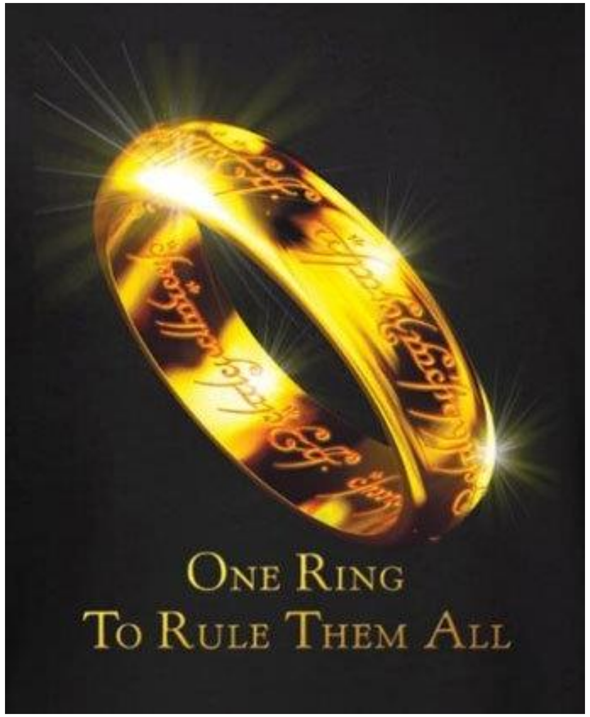


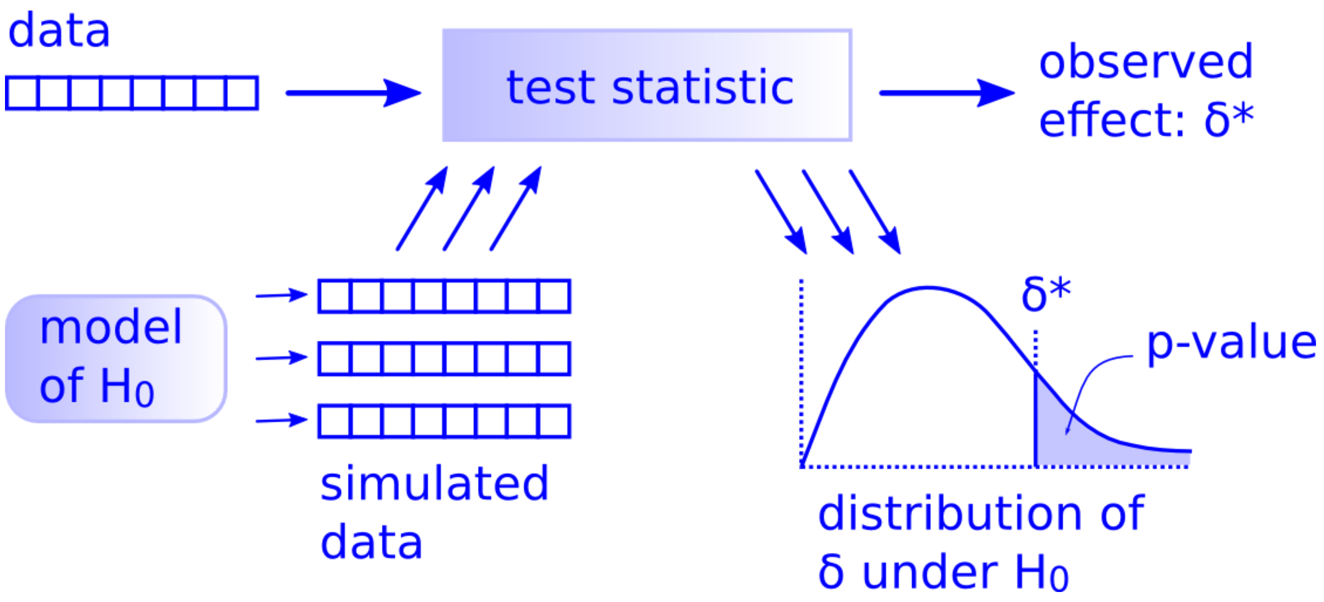
4. Get the probability we would get a statistic more than the observed statistic from the null distribution
    - p-value

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

Hypothesis test for comparing more than two means

The big picture: There is only one hypothesis test!

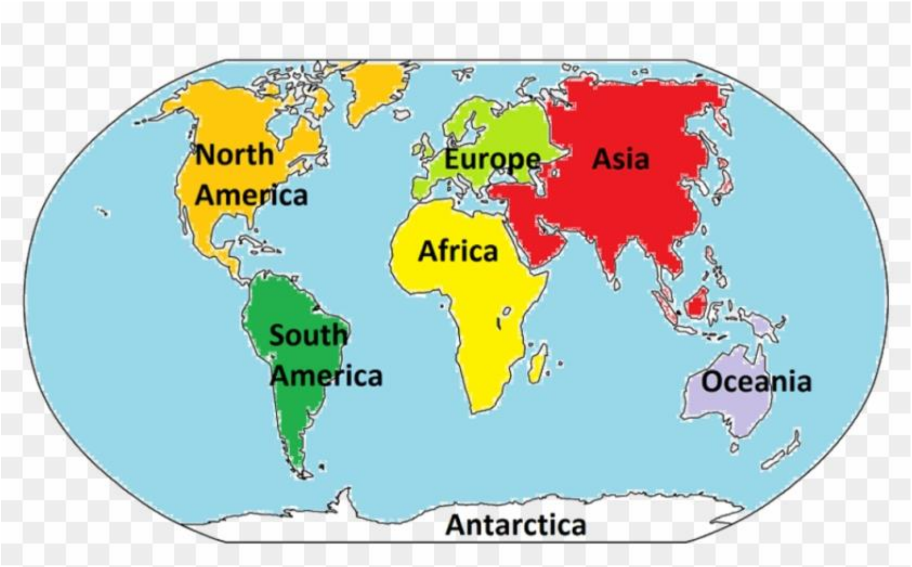




Just need to follow 5 steps!

Comparing more than two means

Let's examine the beer consumption in different continents!



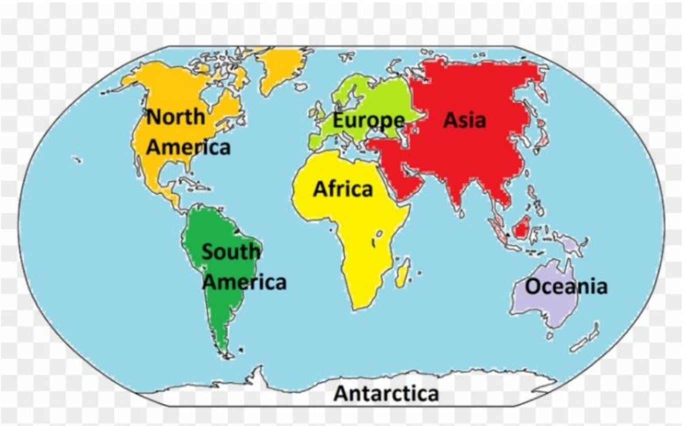
Analysis inspired by:

- [Minitab blog article]
- [Five thirty eight analysis]

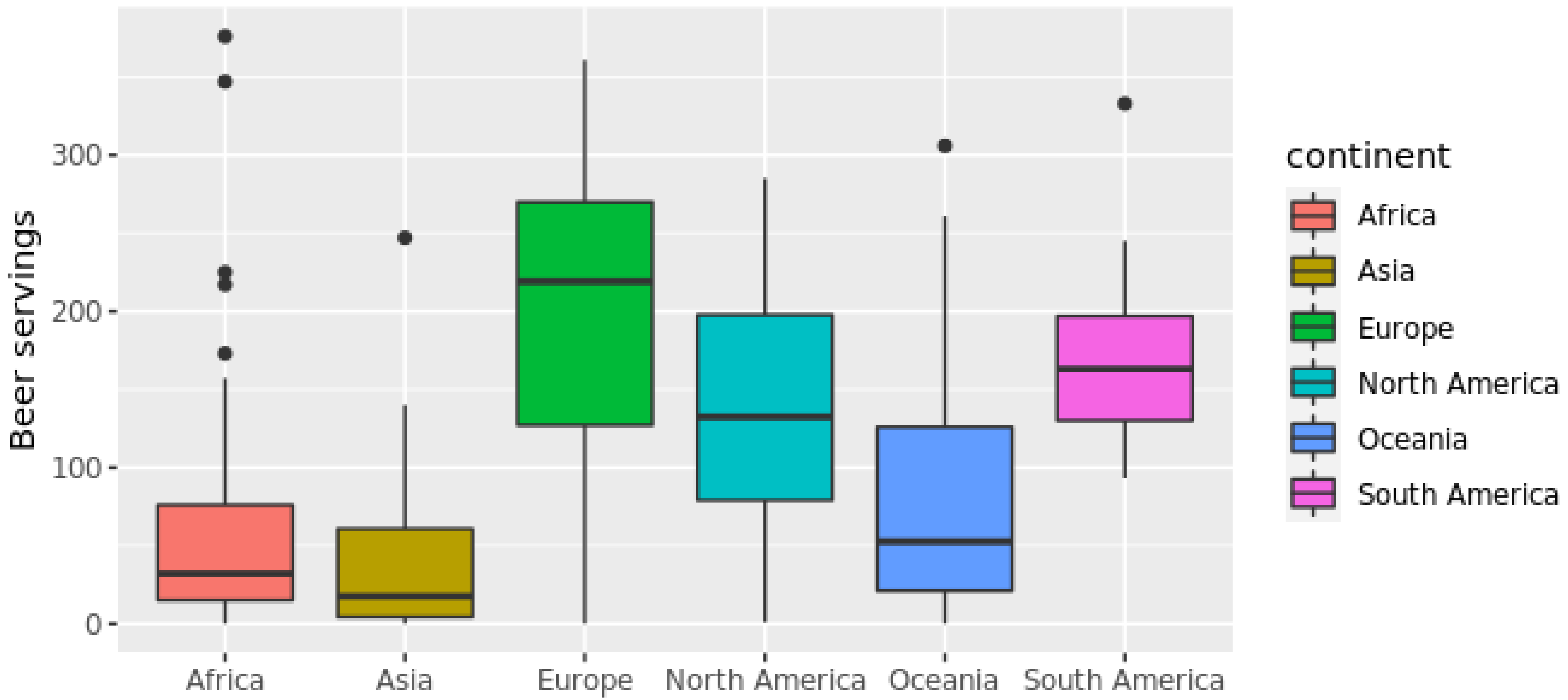
Question: Does the average beer consumption in countries different depending on the continent?

1. State the null and alternative hypotheses!

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):

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

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

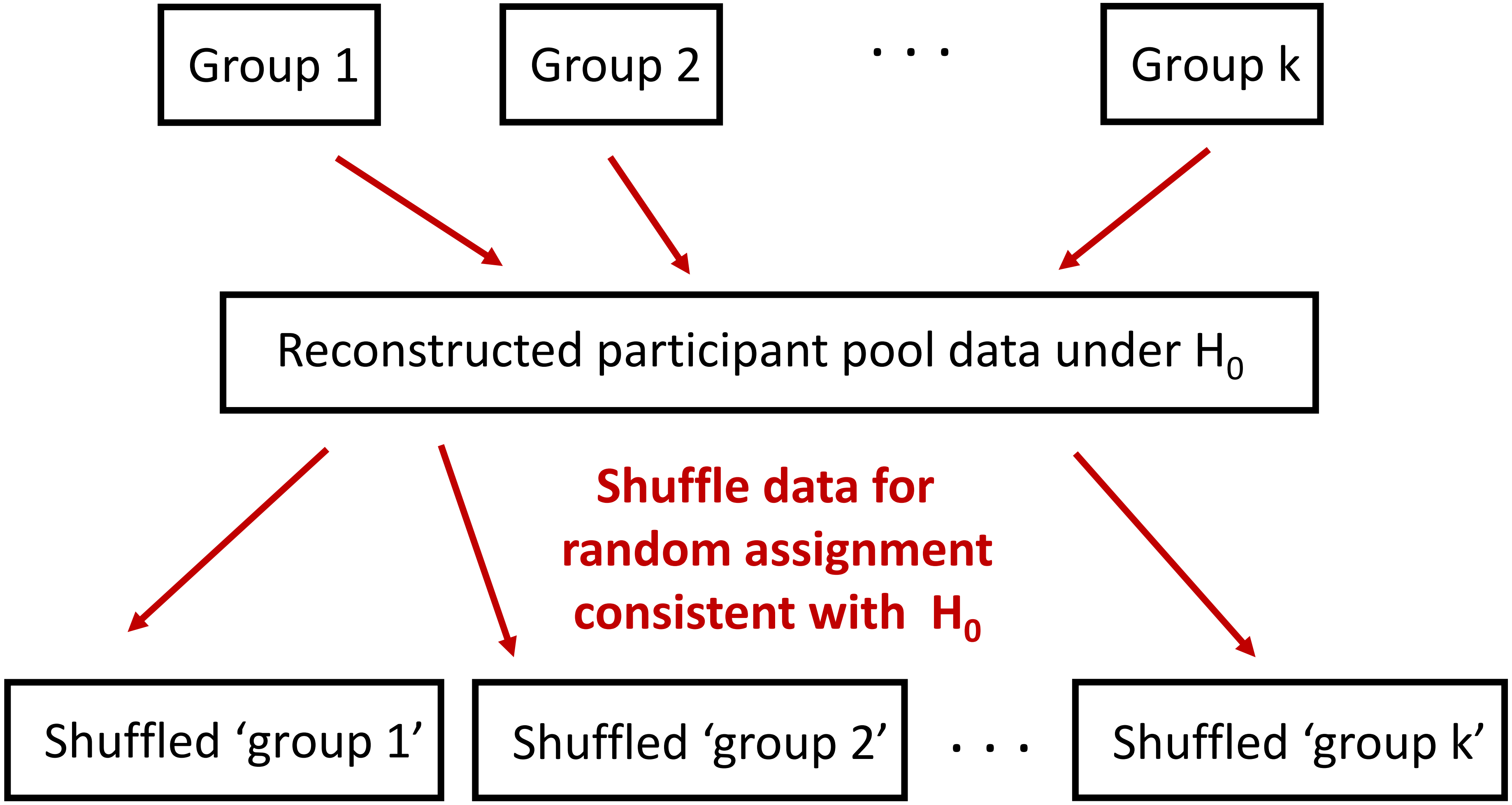
<u>Mean absolute difference (MAD):</u>

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

Observed statistic value = 78.86

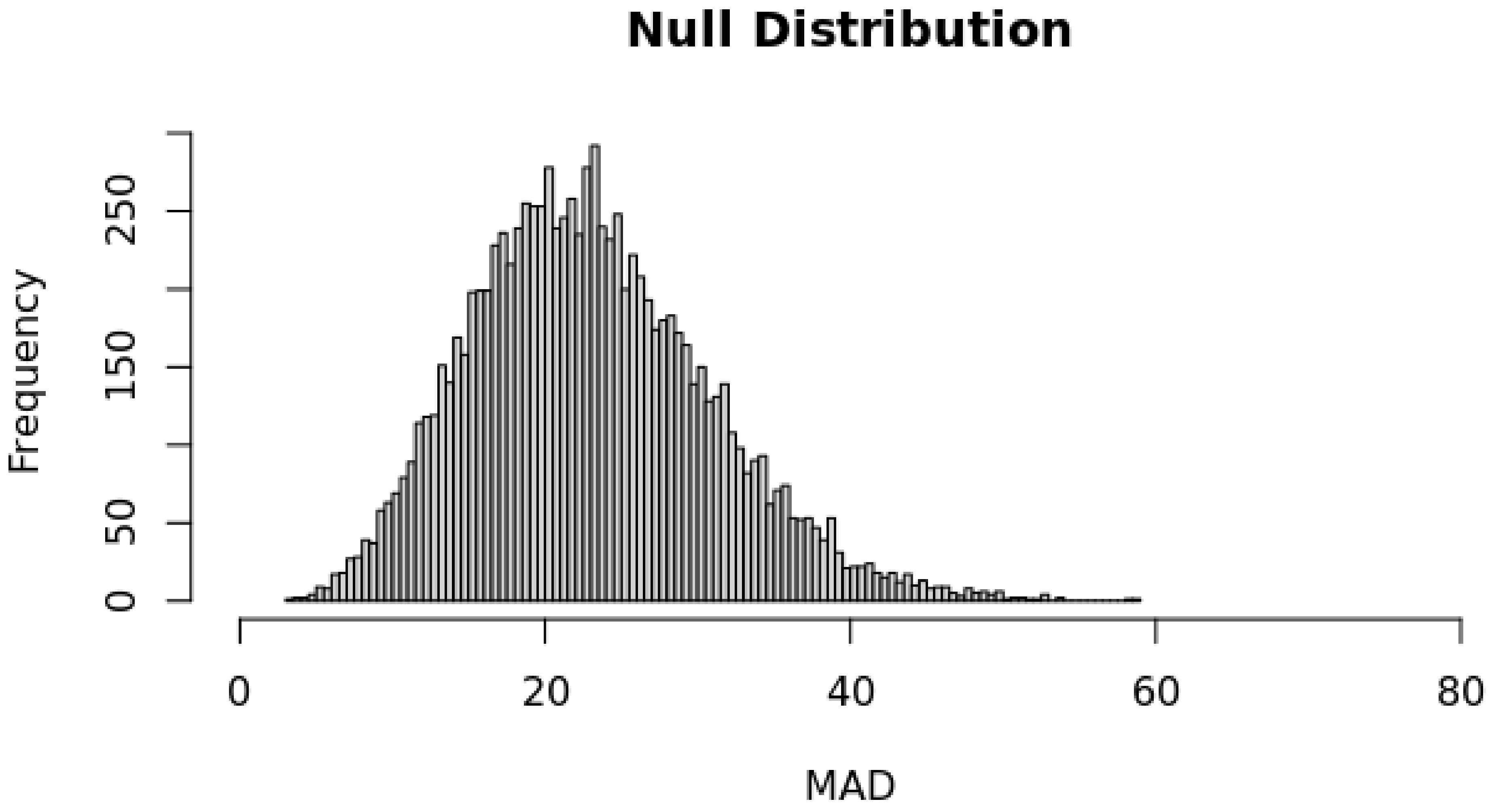
How can we create the null distribution?

3. Create the null distribution!



Compute statistics from shuffled groups

3. Create the null distribution!



Conclusions?

Let’s try it in R…

Permutation/randomization tests for other parameters

Suppose we wanted to test whether there is an association between to variables

- E.g. is there a correlation between the number of pages in a book and the books price?

Q_1: What are the null and alternative hypotheses?

- H_0: $\rho = 0$
- H_A: $\rho > 0$

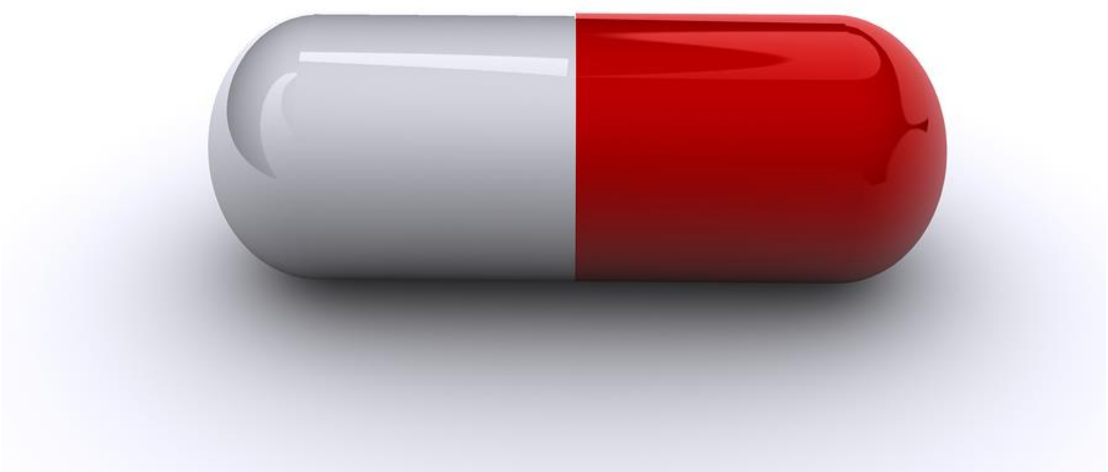
Q_2: What is the observed statistic?

- `SDS230::download_data("amazon.rda")`
- `r_correlation <- cor(amazon$NumPages, amazon$List.Price)`

Q_3: How can we create a null distribution?

- Try it at home!

Very quick review of randomization test for two means

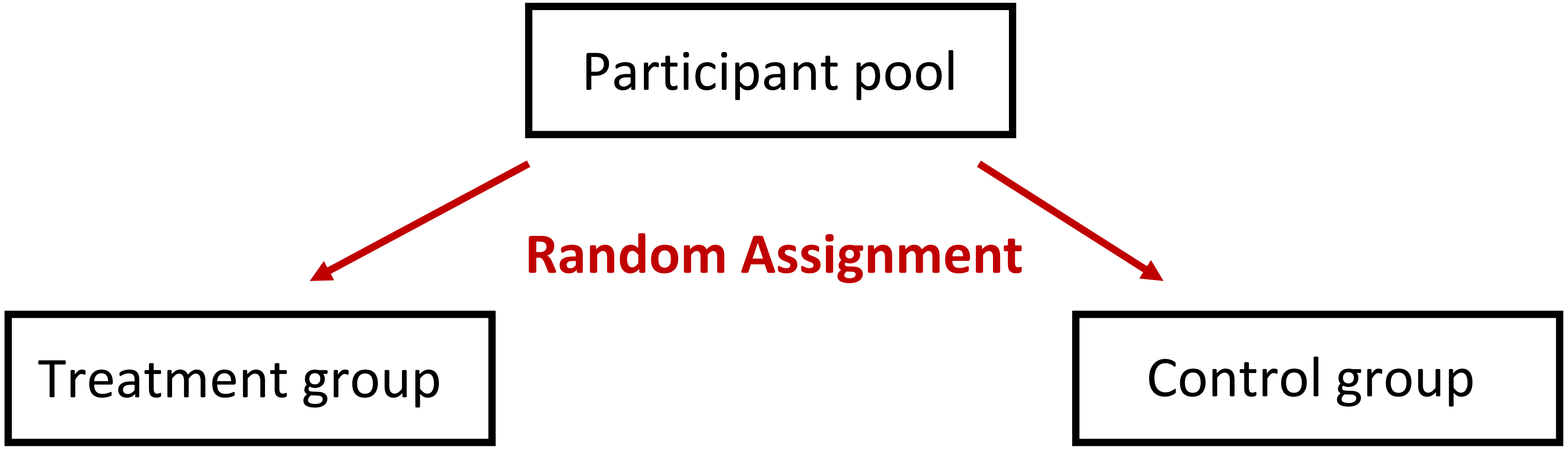


Question: Is this pill effective?

Review: 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



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
   - For randomization/permutation tests we have a choice of the statistic to use

The statistic used before: $\bar{x}_{Effect} = \bar{x}_{Treatment} - \bar{x}_{Control}$

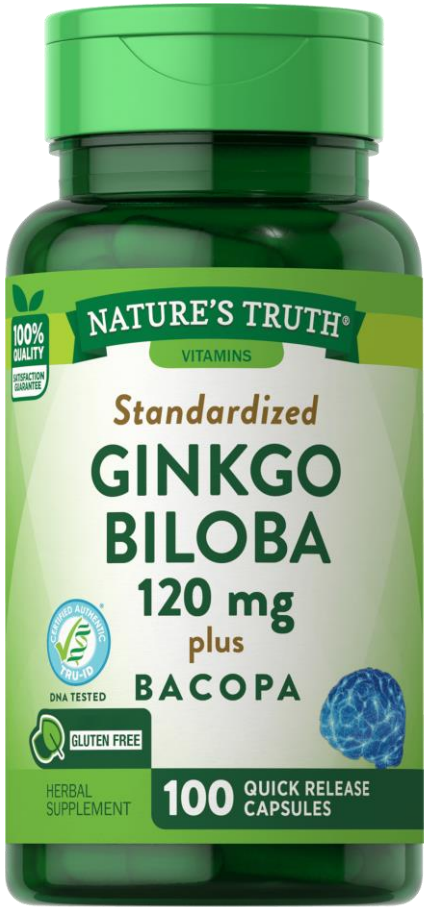
Let's try Welch's t-statistic instead: $$t = \frac{\bar{x}_t - \bar{x}_c}{\sqrt{\frac{s_t^2}{n_t} + \frac{s_c^2}{n_c}}}$$

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

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



2. Visual the data can calculate the observed statistic

Last class we used a difference of means as our observed statistic:

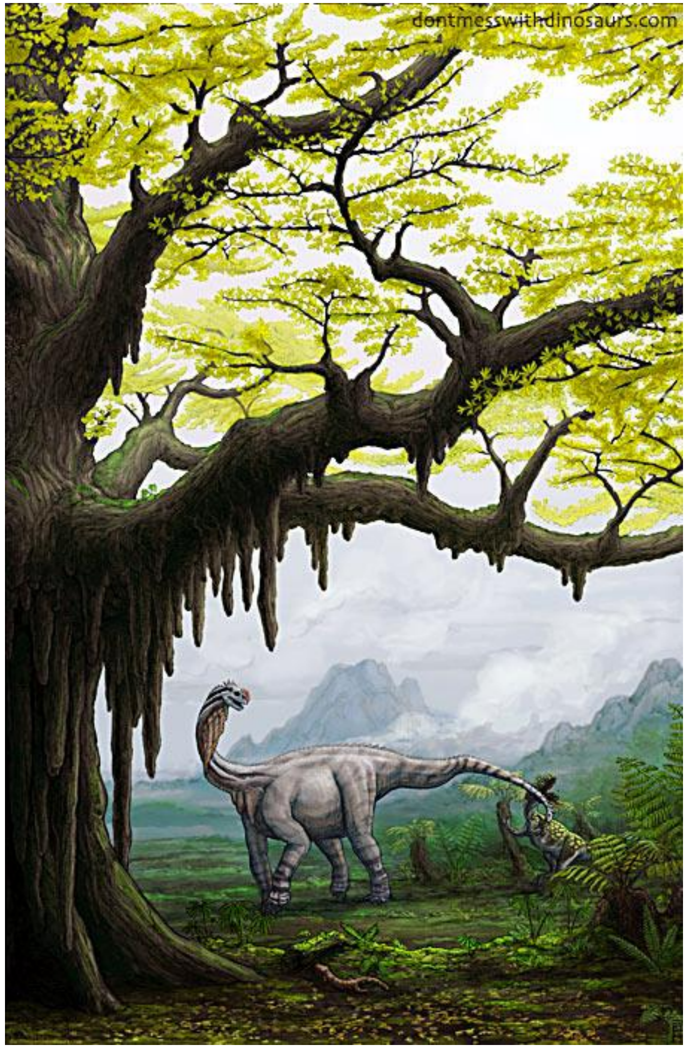
- $\bar{x}_{Effect} = \bar{x}_{Gingko} - \bar{x}_{Placebo}$

With randomization/permutation tests we have the freedom to choose any statistic

Let's try using a t-statistic!

- t = -1.53

$$t = \frac{\bar{x}_t - \bar{x}_c}{\sqrt{\frac{s_t^2}{n_t} + \frac{s_c^2}{n_c}}}$$



3. Create the null distribution!

Treatment group

Control group

Reconstructed participant pool data under H_0

Shuffle data for random assignment consistent with H_0

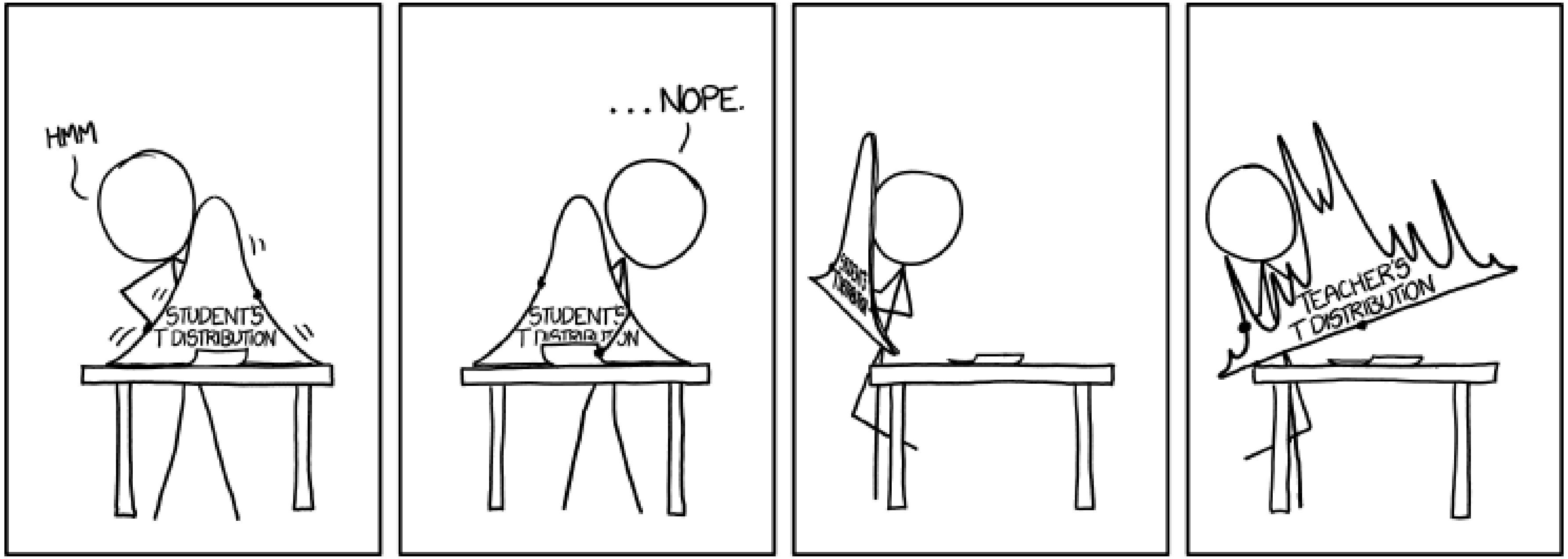
Shuffled 'treatment group'

Shuffled 'control group'

One null distribution statistic: t_{shuff}

Repeat 10,000 times for null distribution

Let's quickly try the rest of the hypothesis test in R…



Parametric hypothesis tests

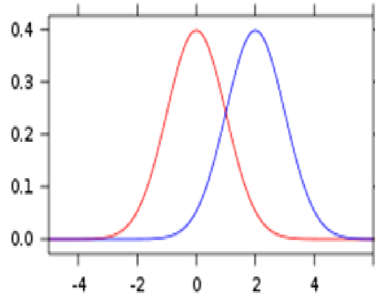
In **parametric hypothesis tests**, the null distribution is given by a density function.

These density functions have a finite set of ***parameters*** that control the shape of these functions

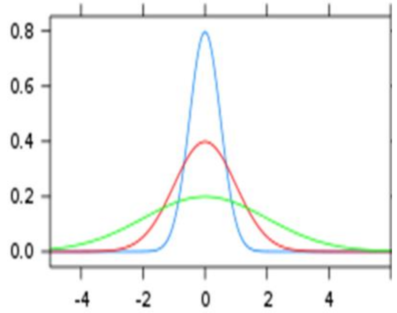
- Hence the name "parametric hypothesis tests"
- Example: the normal density function has two parameters: μ and $\boldsymbol{\sigma}$

Remember density curves?

Changing μ



Changing σ

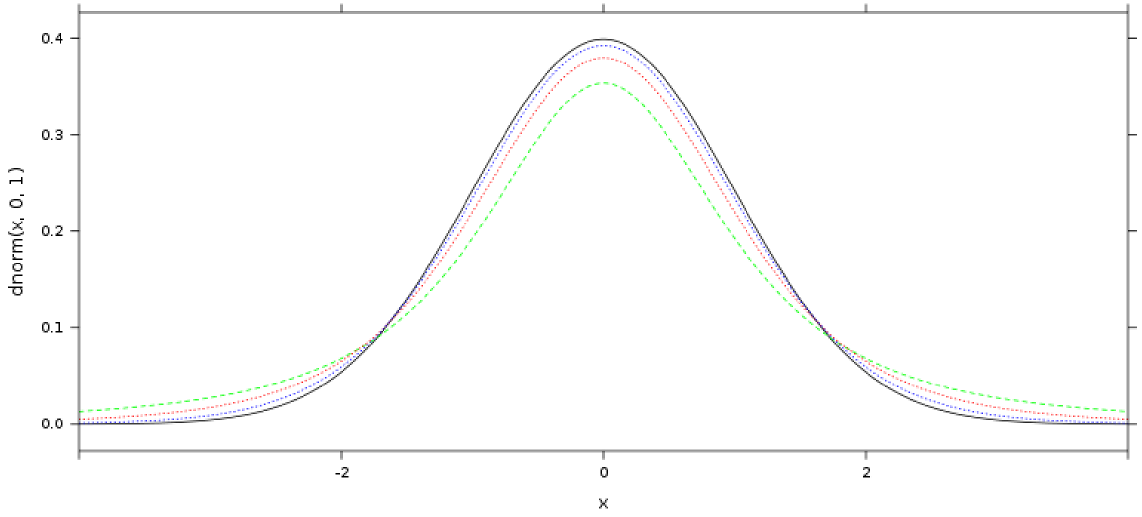


t-distributions

A commonly used density function (distribution) used for statistical inference is the t-distribution

- In R: rt(), dt(), pt() and qt()

t-distributions have one parameter called "degrees of freedom"



df = 2

df = 5

df = 15

N(0, 1)

t-distributions

When using t-distributions for statistical inference, each point in our t-distribution is a t-statistic

- i.e., we use t-distributions as null distributions for hypothesis tests and as sampling distributions when creating confidence intervals

t-statistics are a ratio of:

- The departure of an estimated value from a hypothesized parameter value
- Divided by an estimate of the standard error

$$t = \frac{estimate\ \text{-}\ param_0}{\hat{SE}}$$

If the SE was known exactly the statistic would be a "z-statistic" that comes from a standard normal distribution

t-tests

t-tests are parametric hypothesis tests where the null distribution is a density function called a t-distribution

t-tests can be used to test:

- If a mean is equal to a particular value: H_0: $\mu = 7$
- If two means are equal: H_0: $\mu_t = \mu_c$
- If a regression coefficient is equal to a particular value: H_0: $\beta = 2$
- etc.

t-tests for comparing two means

Let's examine t-tests for comparing **two means**

Step 1: what is the null hypotheses?

- H_0: $\mu_t - \mu_c = 0$

Step 2a: What is the numerator of the t-statistic?

$$t = \frac{\underbrace{estimate}_{(\bar{x}_t - \bar{x}_c)} - \underbrace{param_0}_{0}}{\hat{SE}} \leftarrow = \frac{(\bar{x}_t - \bar{x}_c) - 0}{\hat{SE}} = \frac{\bar{x}_t - \bar{x}_c}{\hat{SE}}$$

t-tests for comparing two means

Step 2b: What is the denominator of the t-statistic?

$$t = \frac{stat\ \text{-}\ param_0}{\hat{SE}}$$

Students' t-test assumes the variance in each population is the same, and uses an SE estimate of:

$$\hat{SE}_{\bar{x}_t\text{-}\bar{x}_c} = s_p \cdot \sqrt{\frac{1}{n_t} + \frac{1}{n_c}} \qquad s_p = \sqrt{\frac{\sum_i^{n_t}(x_i - \bar{x}_t)^2 + \sum_j^{n_c}(x_j - \bar{x}_c)^2}{n_t + n_c\ \text{-}\ 2}}$$

Welch's t-test does **not** assume that the variance in each population is the same and uses an estimate of:

$$\hat{SE}_{\bar{x}_t\text{-}\bar{x}_c} = \sqrt{\frac{s_t^2}{n_t} + \frac{s_c^2}{n_c}}$$

t-tests for comparing two means

Step 2b: What is the denominator of the t-statistic?

$$t = \frac{stat\ \text{-}\ param_0}{\hat{SE}}$$

Students' t-test assumes the variance in each population is the same, and uses an SE estimate of:

$$t = \frac{\bar{x}_t\ \text{-}\ \bar{x}_c}{s_p \cdot \sqrt{\frac{1}{n_t} + \frac{1}{n_c}}}$$

$$s_p = \sqrt{\frac{\sum_i^{n_t}(x_i - \bar{x}_t)^2 + \sum_j^{n_c}(x_j - \bar{x}_c)^2}{n_t + n_c\ \text{-}\ 2}}$$

Welch's t-test does **not** assume that the variance in each population is the same and uses an estimate of:

$$\hat{SE}_{\bar{x}_t\text{-}\bar{x}_c} = \sqrt{\frac{s_t^2}{n_t} + \frac{s_c^2}{n_c}}$$

$$t = \frac{\bar{x}_t - \bar{x}_c}{\sqrt{\frac{s_t^2}{n_t} + \frac{s_c^2}{n_c}}}$$

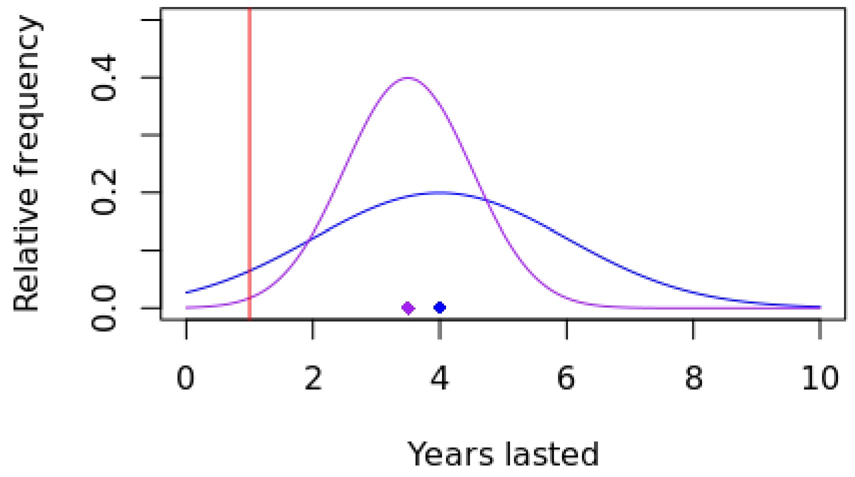
Side note: t-tests for comparing two means

Question: which statistic/test is better to use?

A: generally better to choose the "robust" test

- i.e., Welch's t-test is robust to unequal variances, so generally a better choice

However, we need to be careful with the decisions we make based on differences of means when there are unequal variances



E.g., Which car battery company produces better batteries in terms of how long they last?

- Company A: μ = 4 years, σ = 2 years
- Company B: μ = 3.5 years, σ = 1 years

- Company A: 7% fail within a year
- Company B: 0.6% fail with a year

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

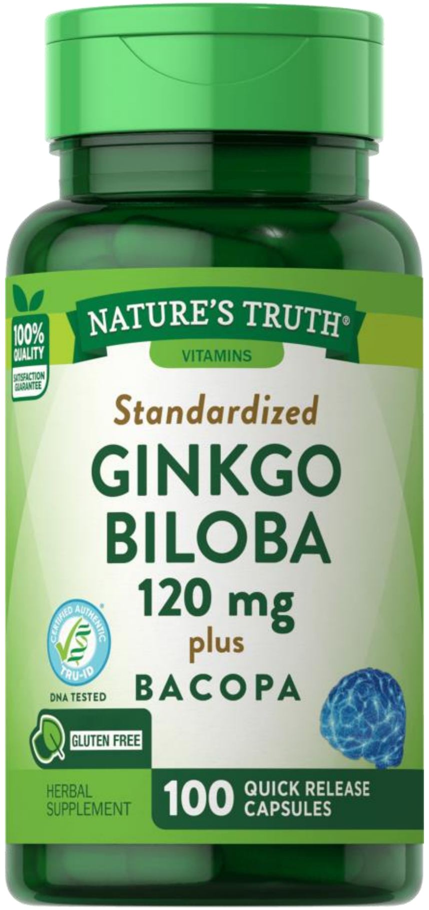
Let's try using a t-statistic!

- t = -1.53

$$t = \frac{\bar{x}_t - \bar{x}_c}{\sqrt{\frac{s_t^2}{n_t} + \frac{s_c^2}{n_c}}}$$

3. What is the null distribution?

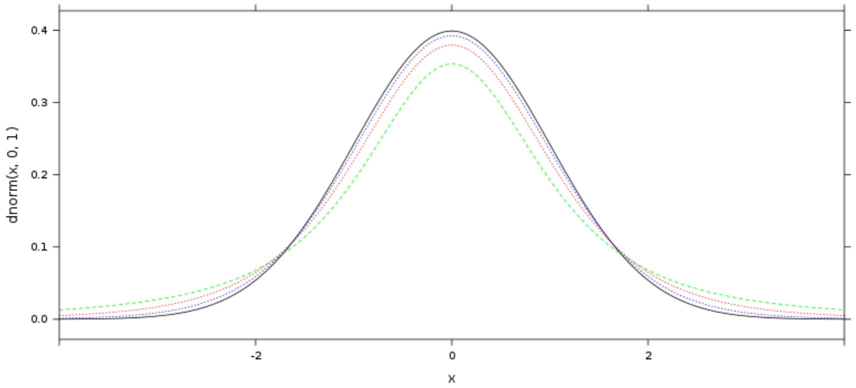
- What additional piece of information do we need to create it?



t-tests for comparing two means

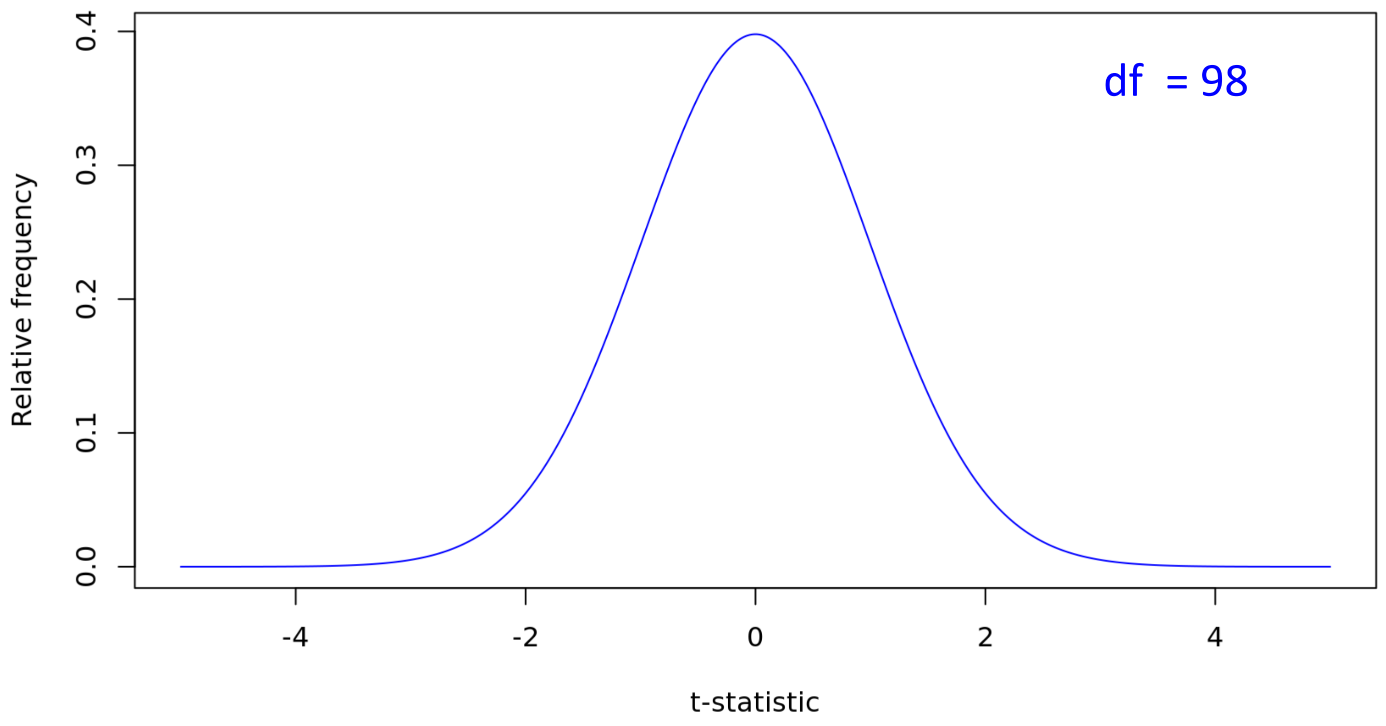
When using a t-distribution to compare two means, a conservative estimate of the degrees of freedom is the minimum of the two samples sizes, n_t and n_c, minus 1

- $df = \min(n_t, n_c) - 1$

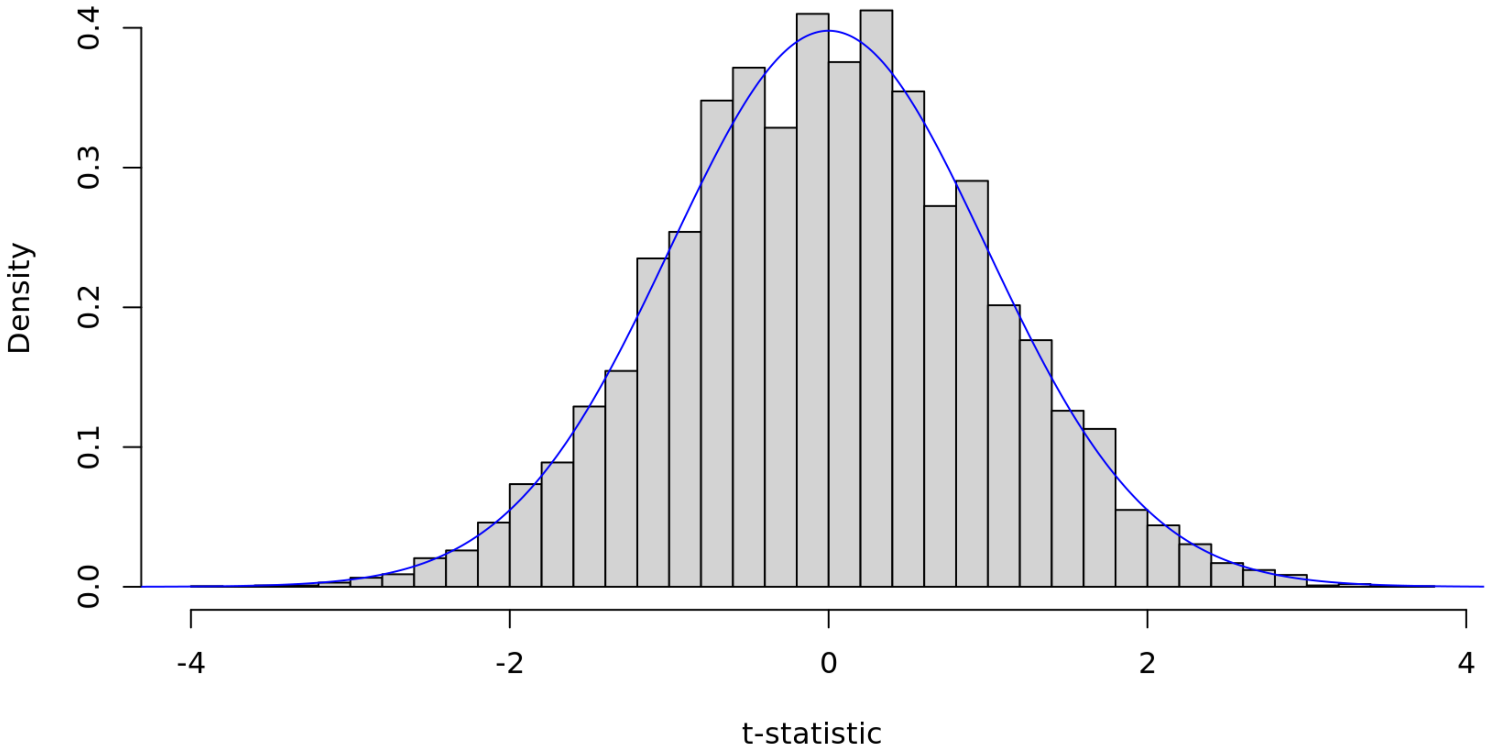


Q: For the Gingko study we had 104 people in the treatment group and 99 people in the control group so the degrees of freedom parameter is?

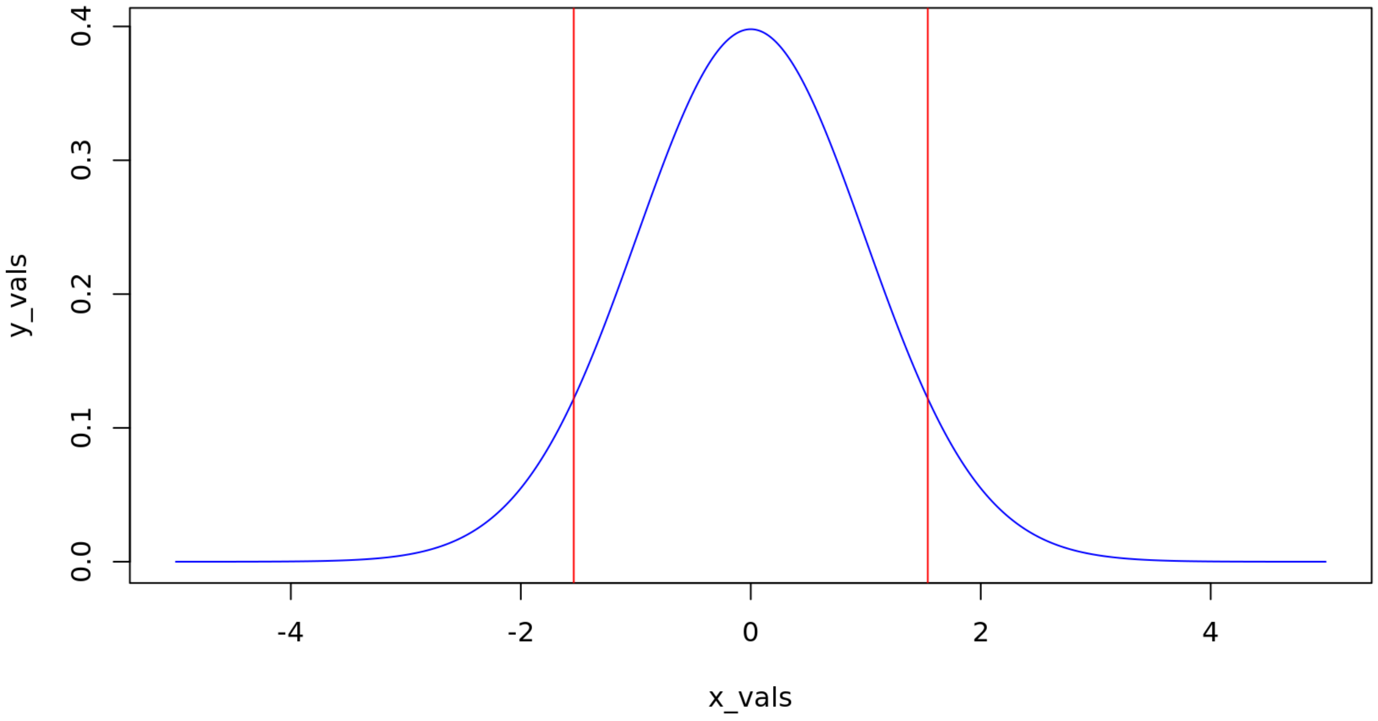
Step 3: Null t-distribution



Step 3: parametric vs. randomization distributions



Step 4-5: p-value and conclusion



$\bar{x}_t - \bar{x}_c$

p-value = 0.127

Conclusion?

Let's try it in R!