Chapter 7: The Two-Sample t-test (1): Introducing Hypothesis Tests

TXCL7565/PHSC7565

What This Chapter Covers

- Two-sample t-test
- Null and alternative hypotheses
- Significance
- Aspects of the data that influence significance
- Risk of a false positive finding
- Requirements for applying a two-sample t-test
- Performing and reporting the test

TWO-SAMPLE T-TEST

Two-sample t-test

Goal: Examine the difference between two ***independent*** groups

- Outcome (dependent variable): continuous (i.e., interval variable)
- Predictor (independent variable): membership in one of two groups (e.g., treatment vs. control)

Example

It is known that the antibiotic rifampicin increases the amount of drug metabolizing enzyme present in the liver and consequently increases the rate of elimination of a wide range of other drugs. This experiment is designed to detect whether rifampicin affects the metabolic removal of the anti-asthma drug theophylline.

Subject were randomly placed into one of two treatment groups, pretreatment with oral placebo or pretreatment with rifampicin.

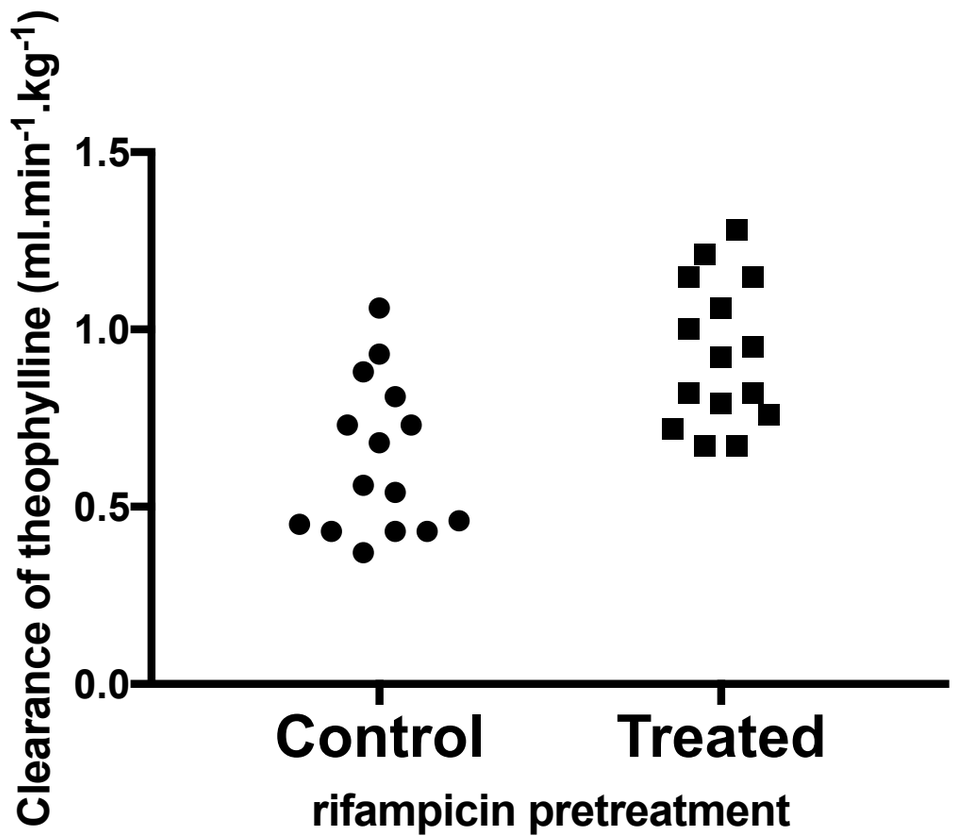
Example

What type of study is this (observational or randomized experiment)?

Can we infer causality in this experiment?

What is the dependent variable?

What is the independent variable?



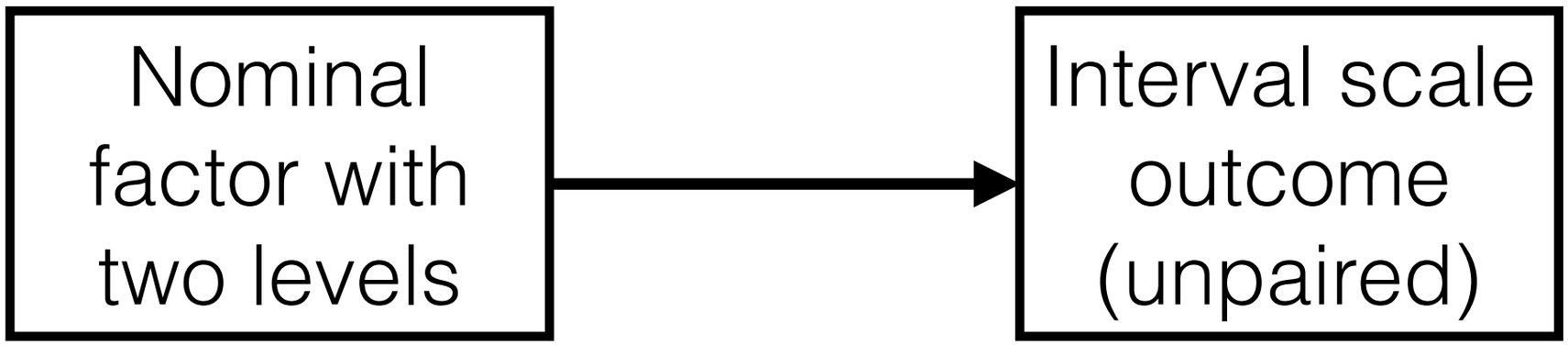


Figure 7.2 Diagrammatic representation of an experimental structure where use of the two-sample t-test is appropriate.

NULL AND ALTERNATIVE HYPOTHESES

An apparent difference even in the absence of any real treatment effect

Samples are always subject to random error and control and treated samples are unlikely to produce identical means, even when a treatment has absolutely no real effect.

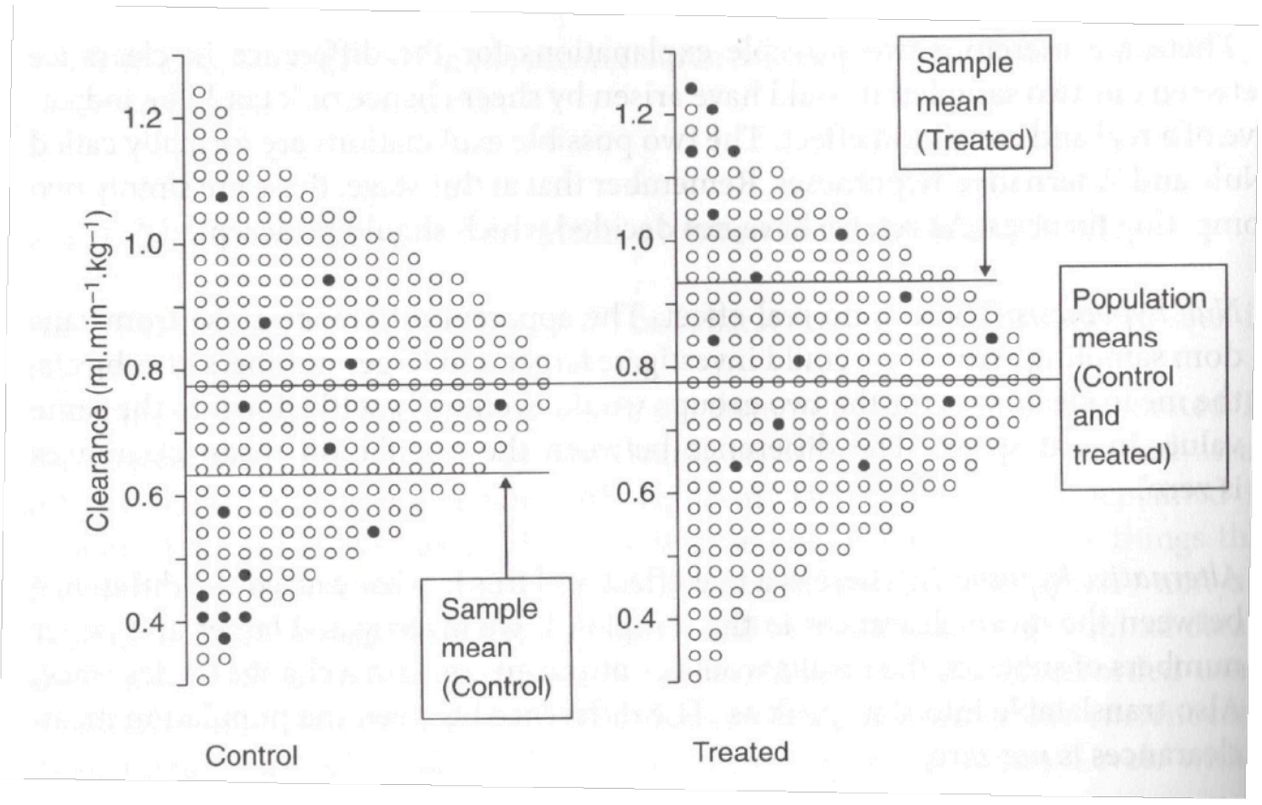
Null vs. Alternative

Hypothesis testing is a formal statistical procedure for deciding between two competing claims about a population parameter, in this case, the difference between two means.

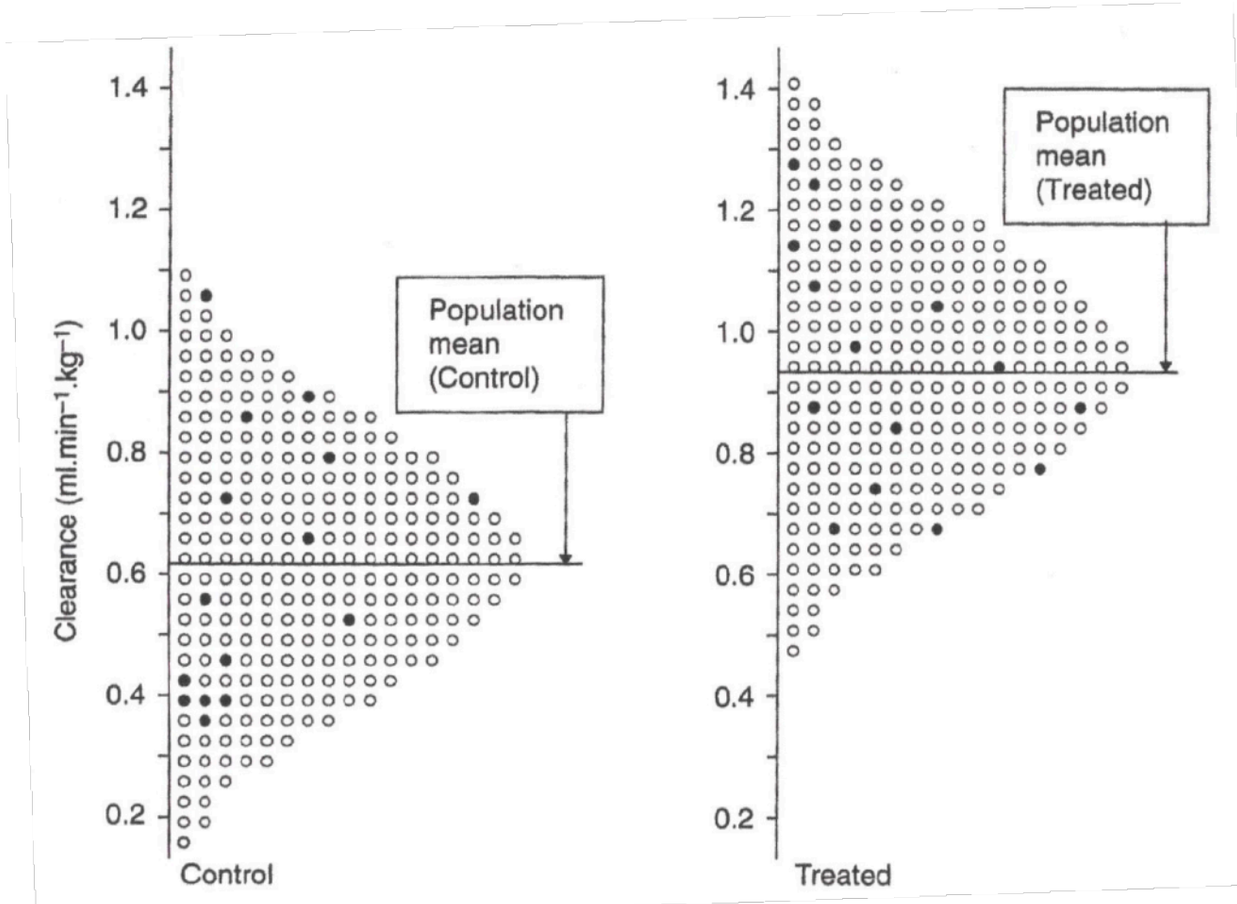
Null hypothesis - no real effect; apparent effects arose from random sampling

Alternative hypothesis - real effect; observed effects arose from a true difference

Null Hypothesis
(no difference between population means)



Alternative Hypothesis
(difference between population means)



Null vs Alternative
2-sample *t*-test

$\mu_{control}$ = population mean of the control group
$\mu_{treated}$ = population mean of the treated group

Null Hypothesis: H_0: $\mu_{control} = \mu_{treated}$
Alternative Hypothesis: H_a: $\mu_{control} \neq \mu_{treated}$

File | Sheet | Undo | Clipboard | Analysis | Interpret | Change | Draw | Write

- Family
- Search results
- Data Tables
  - Clearance
- Info
  - Project info 1
- Results
  - Col Stats of Clearance
  - Welch's t test of Clearanc
- Graphs
  - Clearance
- Layouts

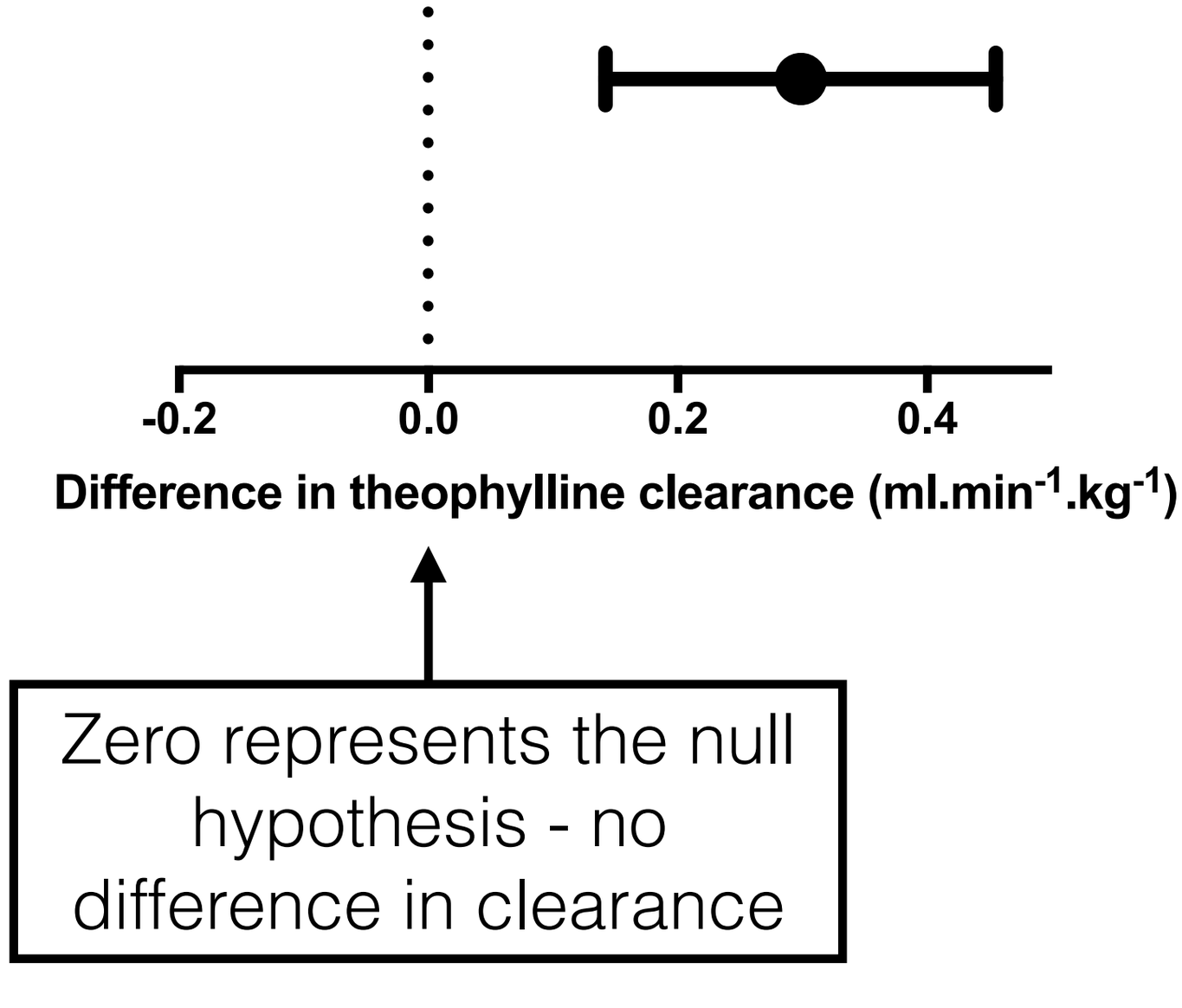
Welch's t test

1	Table Analyzed	Clearance
2		
3	Column B	Treated
4	vs.	vs.
5	Column A	Control
6		
7	Unpaired t test with Welch's correction	
8	P value	0.0005
9	P value summary	***
10	Significantly different (P < 0.05)?	Yes
11	One- or two-tailed P value?	Two-tailed
12	Welch-corrected t, df	t=3.909 df=27.87
13		
14	How big is the difference?	
15	Mean ± SEM of column A	0.6327 ± 0.05581, n=15
16	Mean ± SEM of column B	0.9313 ± 0.05219, n=15
17	Difference between means	0.2987 ± 0.07641
18	95% confidence interval	0.1421 to 0.4552
19	R squared (eta squared)	0.354
20		
21	F test to compare variances	
22	F, DFn, Dfd	1.144, 14, 14
23	P value	0.8053
24	P value summary	ns
25	Significantly different (P < 0.05)?	No
26		

Two-sample t-test

Mean (Treated)	0.9313
Mean (Control)	0.6327
Difference (Treated - Control)	0.2987
95% CI Difference	0.1421 to 0.4552

95% CI of difference in theophylline clearance



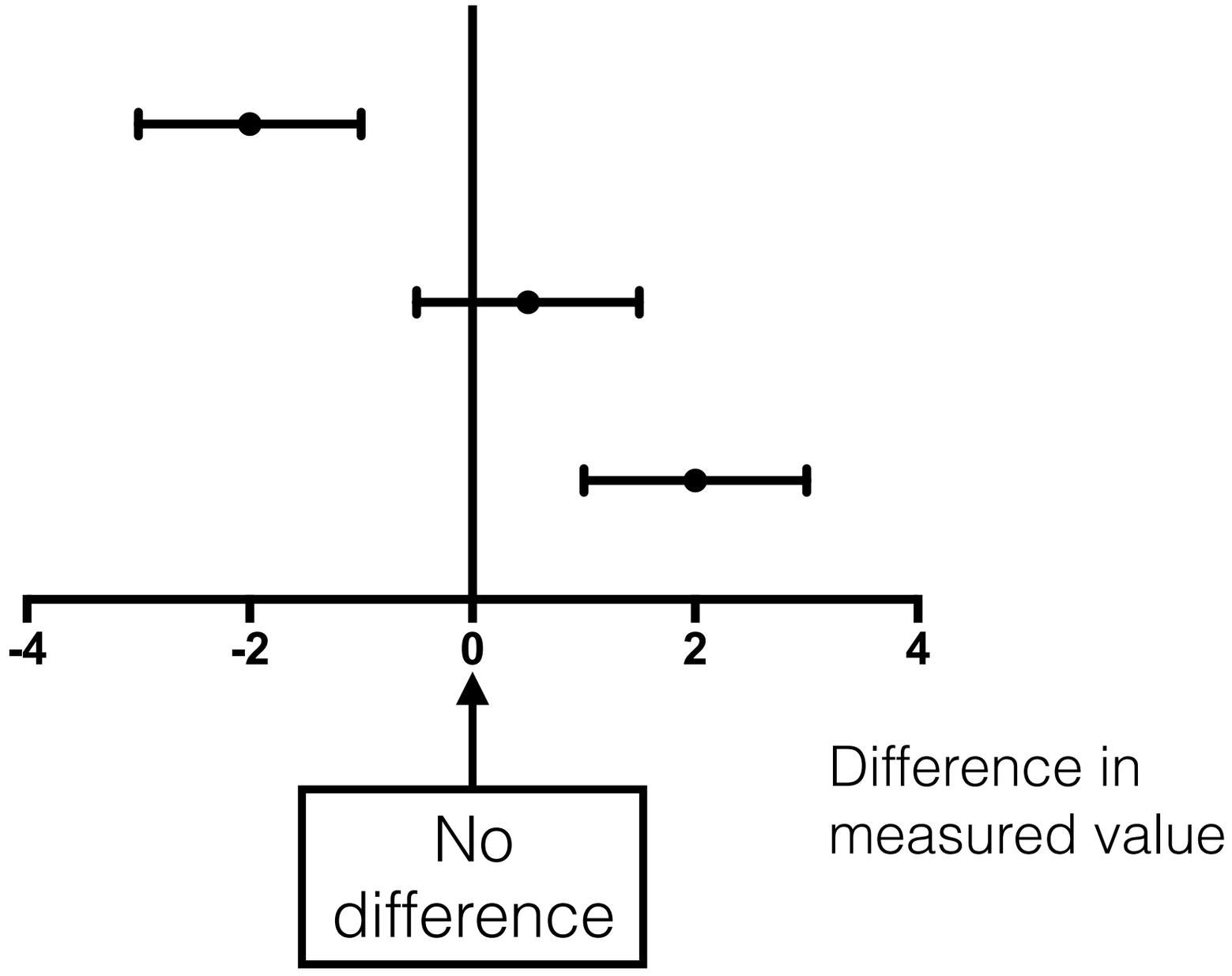
Zero represents the null hypothesis - no difference in clearance

Is there evidence of an effect?

If the C.I. *includes* zero, the null hypothesis that the treatment produced no effect is credible. Nothing has been proven.

If the C.I. *excludes* zero, we have worthwhile evidence that there is an experimental effect.

General interpretation of the results of a two-sample *t*-test



SIGNIFICANCE

Significant vs. Non-Significant

If the evidence is ‘significant’, e.g., the confidence interval for a difference in means doesn’t contain 0, it is strong enough to merit being added to whatever body of knowledge already exists. It does not mean that we should blindly accept the current results.

‘Non-significant’ implies that the evidence is weak and will have little influence upon our thinking.

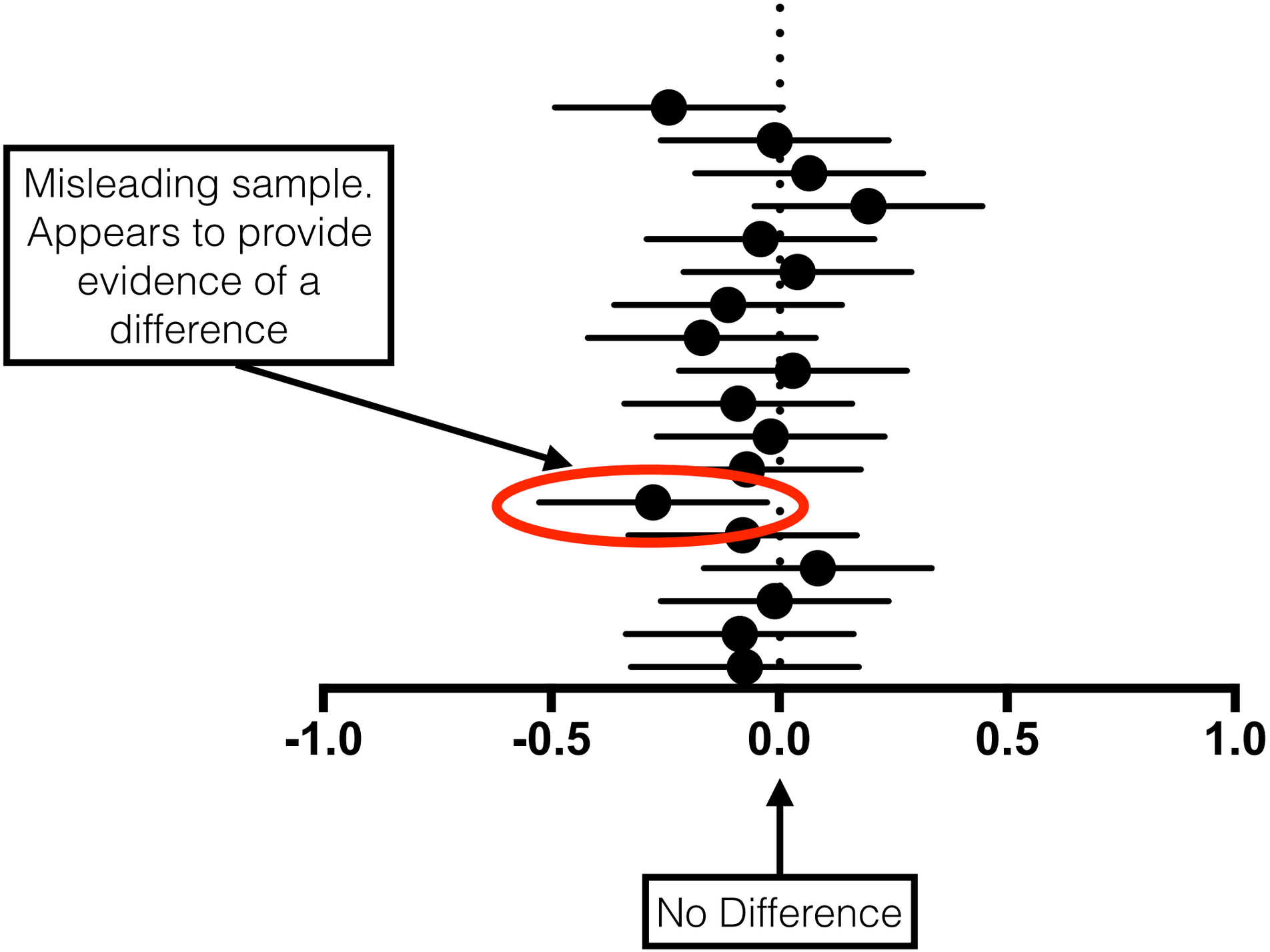
Non-significant NOT insignificant

Types of Error

		In reality, the outcome is:	
		True difference	**No difference**
Experiment data shows:	**True difference**	Correct Decision	Type I Error
	No difference	Type II Error	Correct Decision

False Positives or Type I Errors

If there is no real effect of the treatment we are investigating, but we happen to obtain particularly misleading samples. We may wrongly conclude that there is adequate (significant) evidence of an effect. In that case, we have generated a 'False Positive' or 'Type I Error'.


Misleading sample. Appears to provide evidence of a difference
-1.0
-0.5
0.0
0.5
1.0
No Difference

Alpha

The Greek symbol alpha ‘α’ = risk of a false positive when the null hypothesis is true.

α = 1 - level of confidence

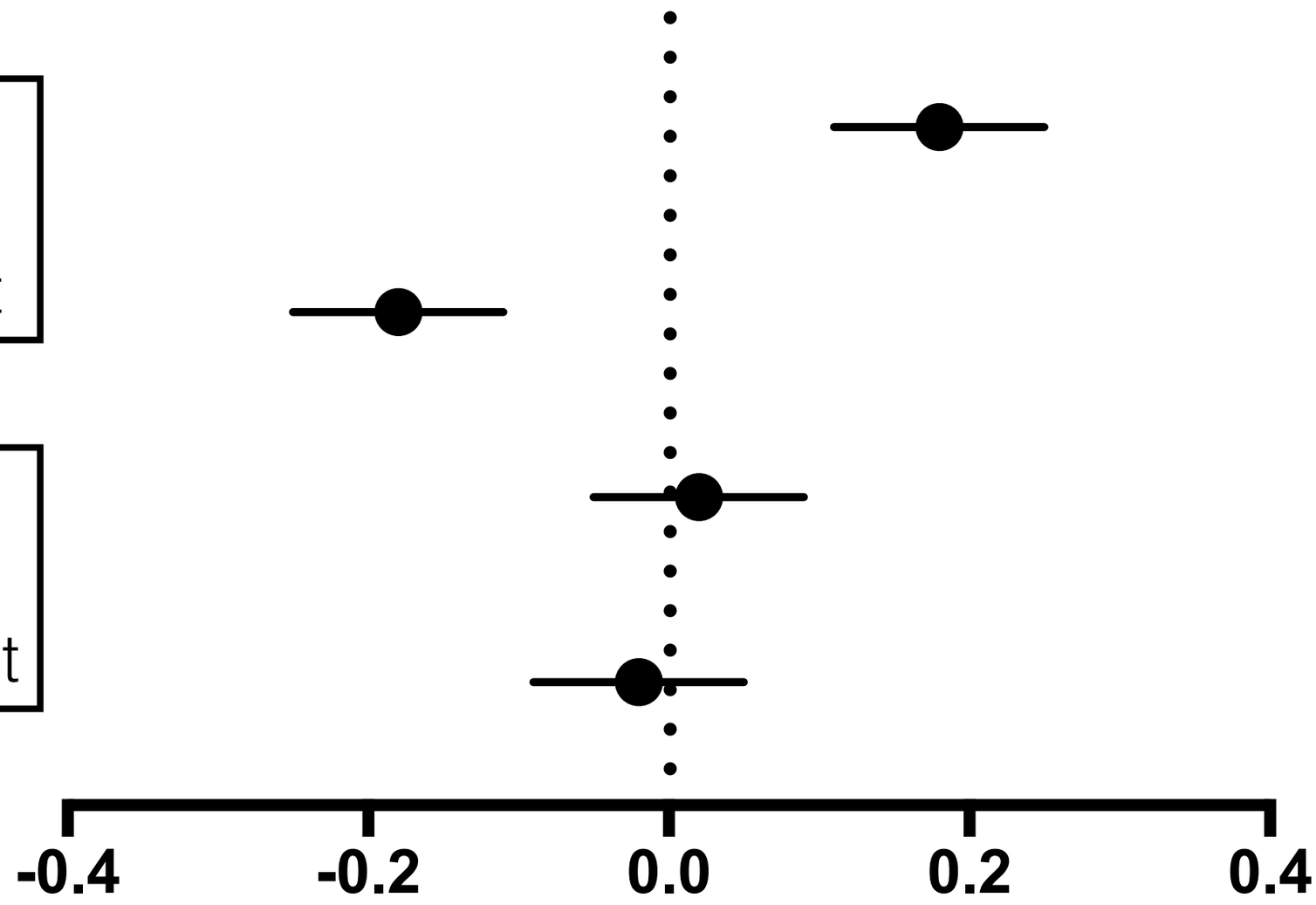
- e.g., α = 5% for a 95% Confidence Interval

ASPECTS OF THE DATA THAT INFLUENCE SIGNIFICANCE

Size of Experimental Effect

Large effect. Will probably be significant

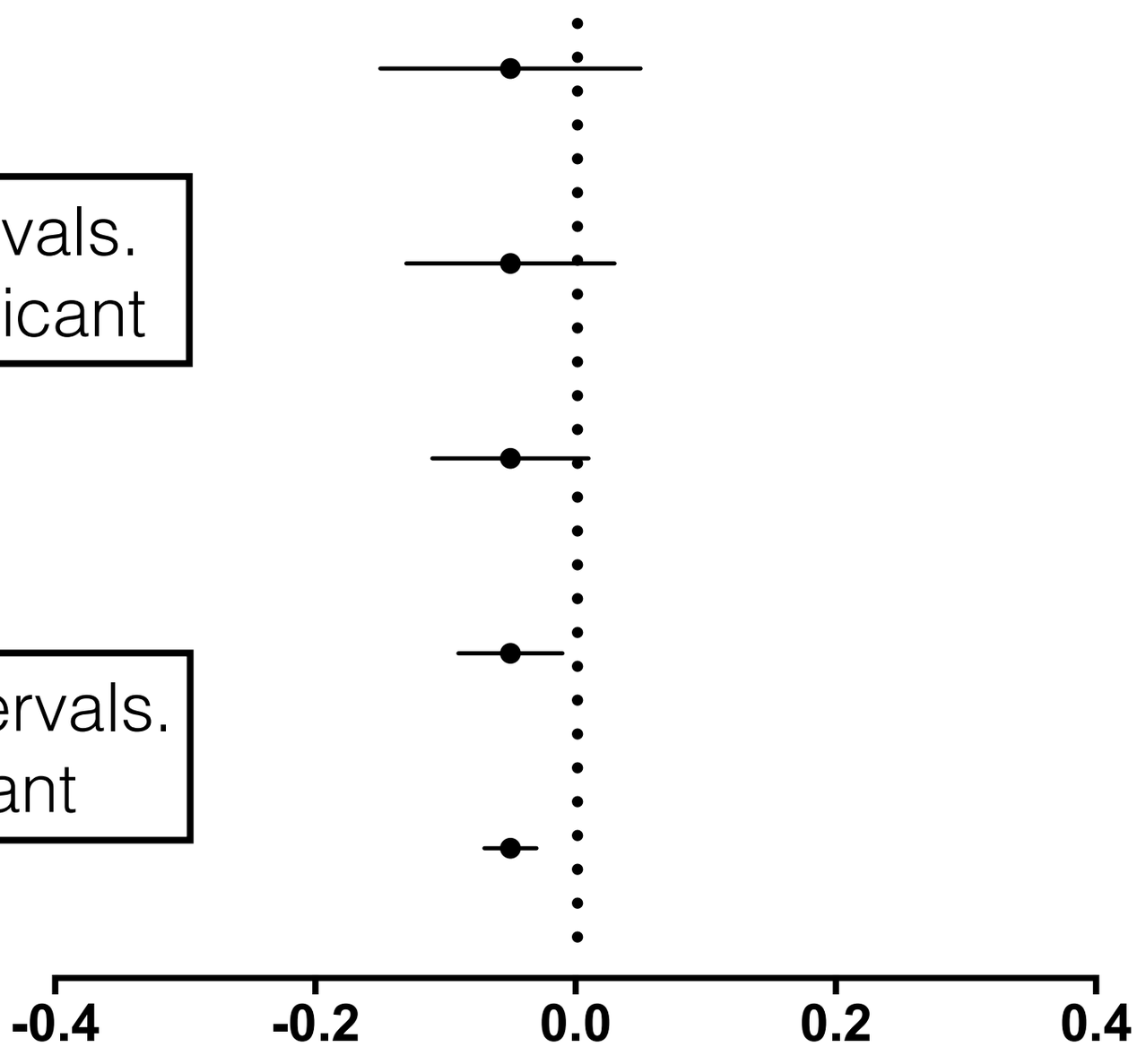
Small effect. Likely to be non-significant



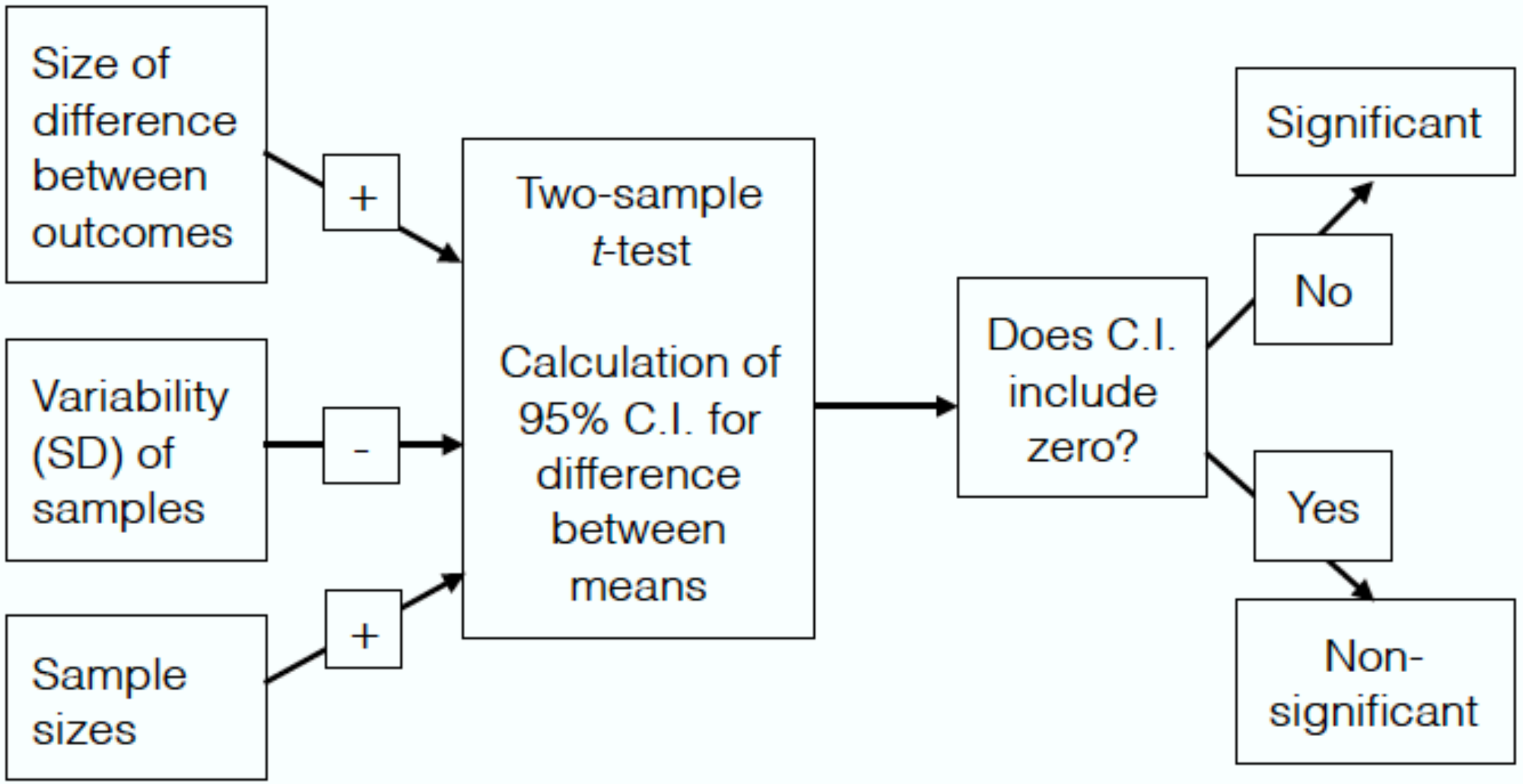
Width of Confidence Interval

Wide intervals.
Non-significant

Narrow intervals.
Significant



Factors that influence the outcome of a two-sample *t*-test



REQUIREMENTS FOR APPLYING A TWO-SAMPLE *t*-TEST

Assumption of normal distributions and equal SDs

The mathematical basis of the two-sample *t*-test assumes that the samples are drawn from populations that:

- Are normally distributed
- Have equal SDs

What to do when data are non-normal

- Like we talked about in previous chapters, you can transform the data, e.g., logarithmic transformation

- Use a non-parametric test (Chapter 21)

Welch's Approximate *t*

- If standard deviations are not equal, use the Welch Approximate t
- To test if SD are equal, use the 'F-test to compare variances' in GraphPad Prism
    - A significant result indicates that were is a significant difference in variance between the two groups.
    - Use the same amount of caution you would use with test of normality

Personal recommendation - start with the Welch test because the loss of power when the variances are equal is minimal, but the consequence of not using it when the variances are very different can be important

PERFORMING AND REPORTING THE TEST

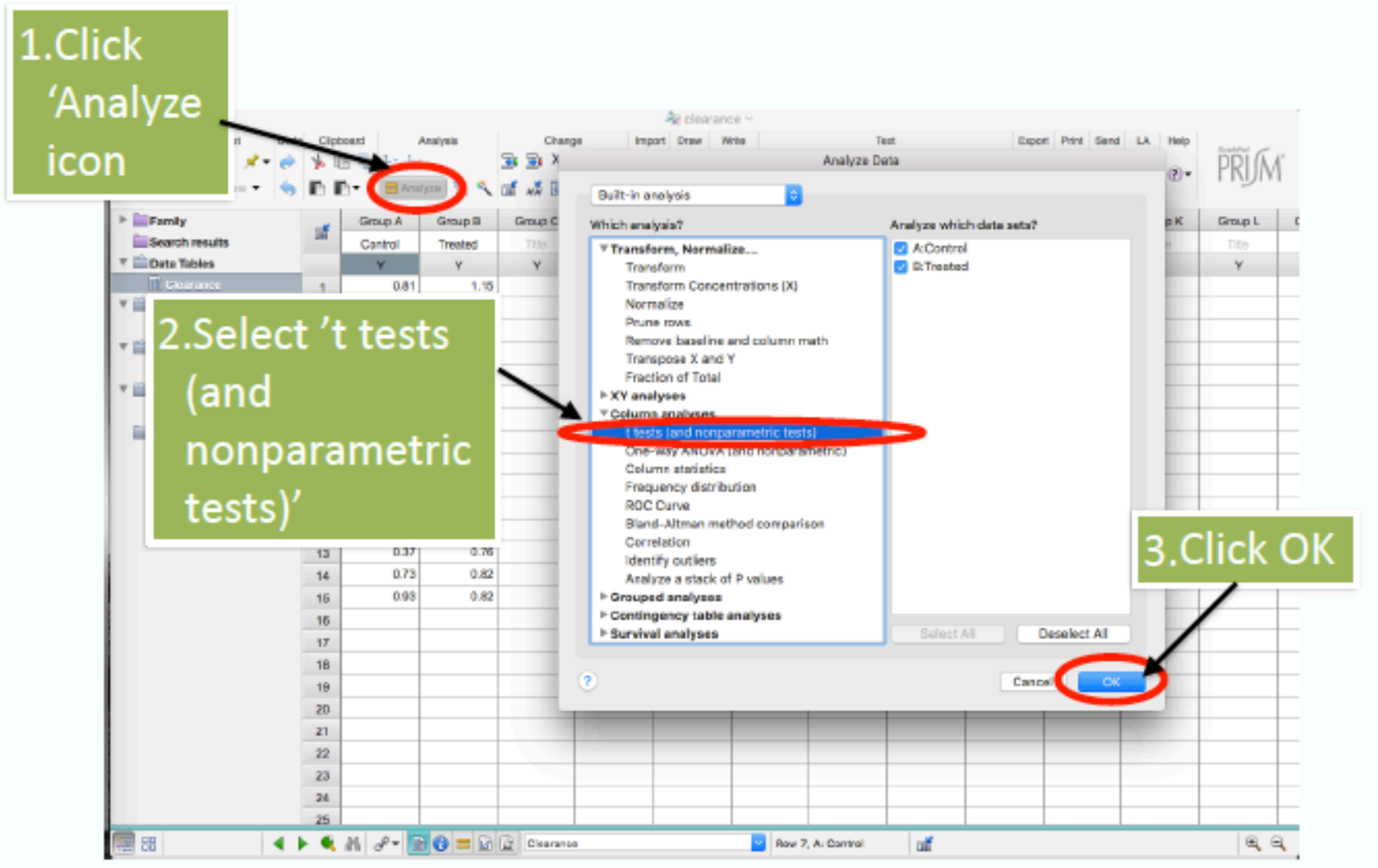
Variables needed

Two variables needed:

- ***Dependent variable*** - interval scale measure of outcome

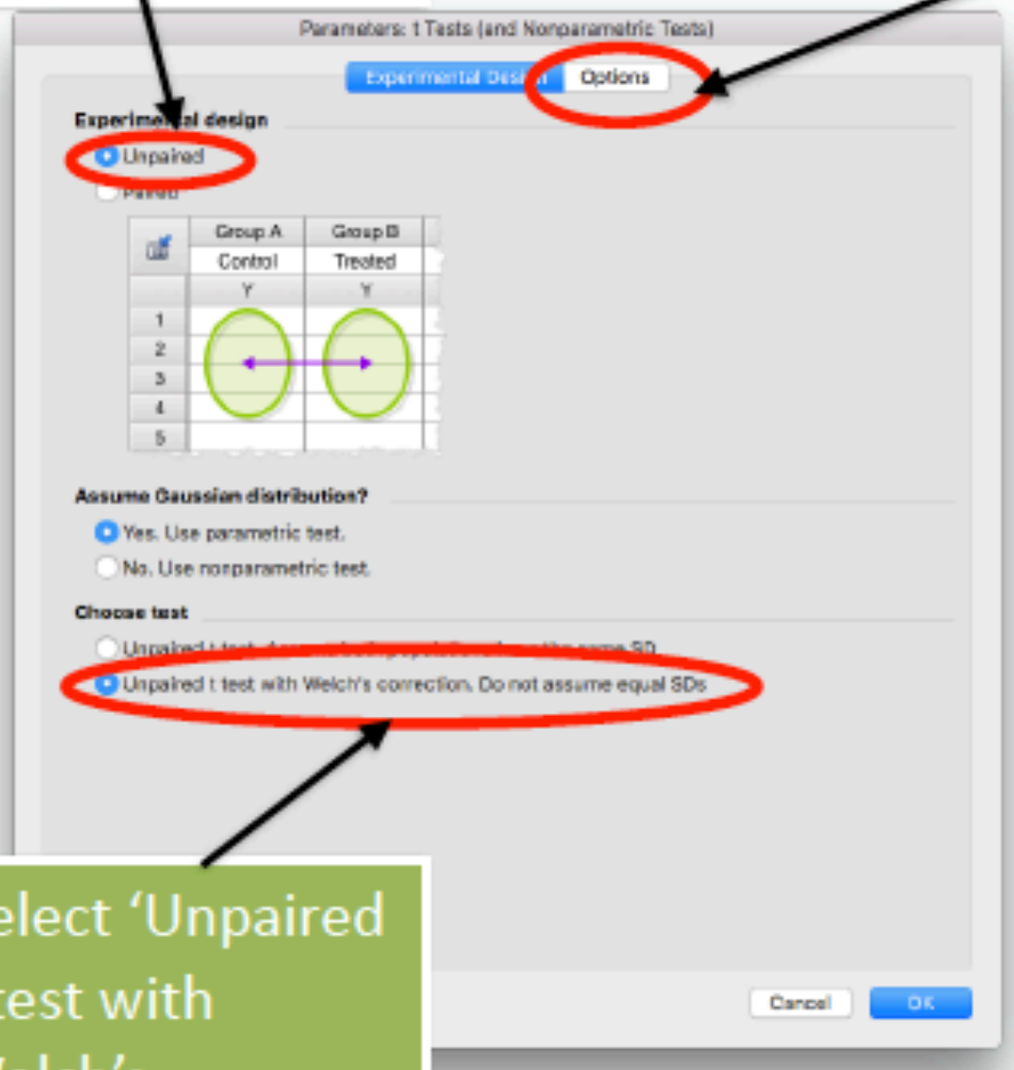
- ***Independent variable*** - nominal variable with only 2 values

2-sample *t*-test in GraphPad

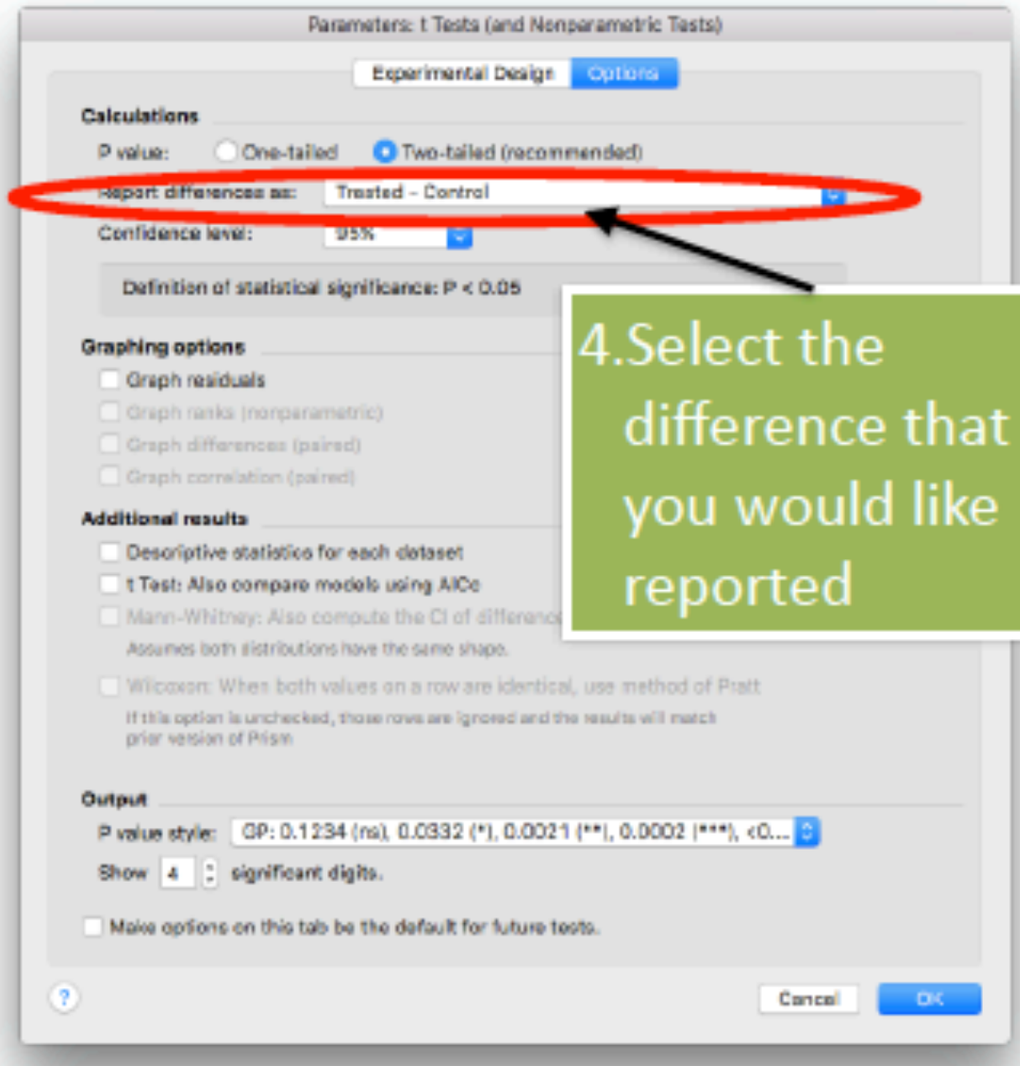


1.Select ‘Unpaired’ for Experimental Design

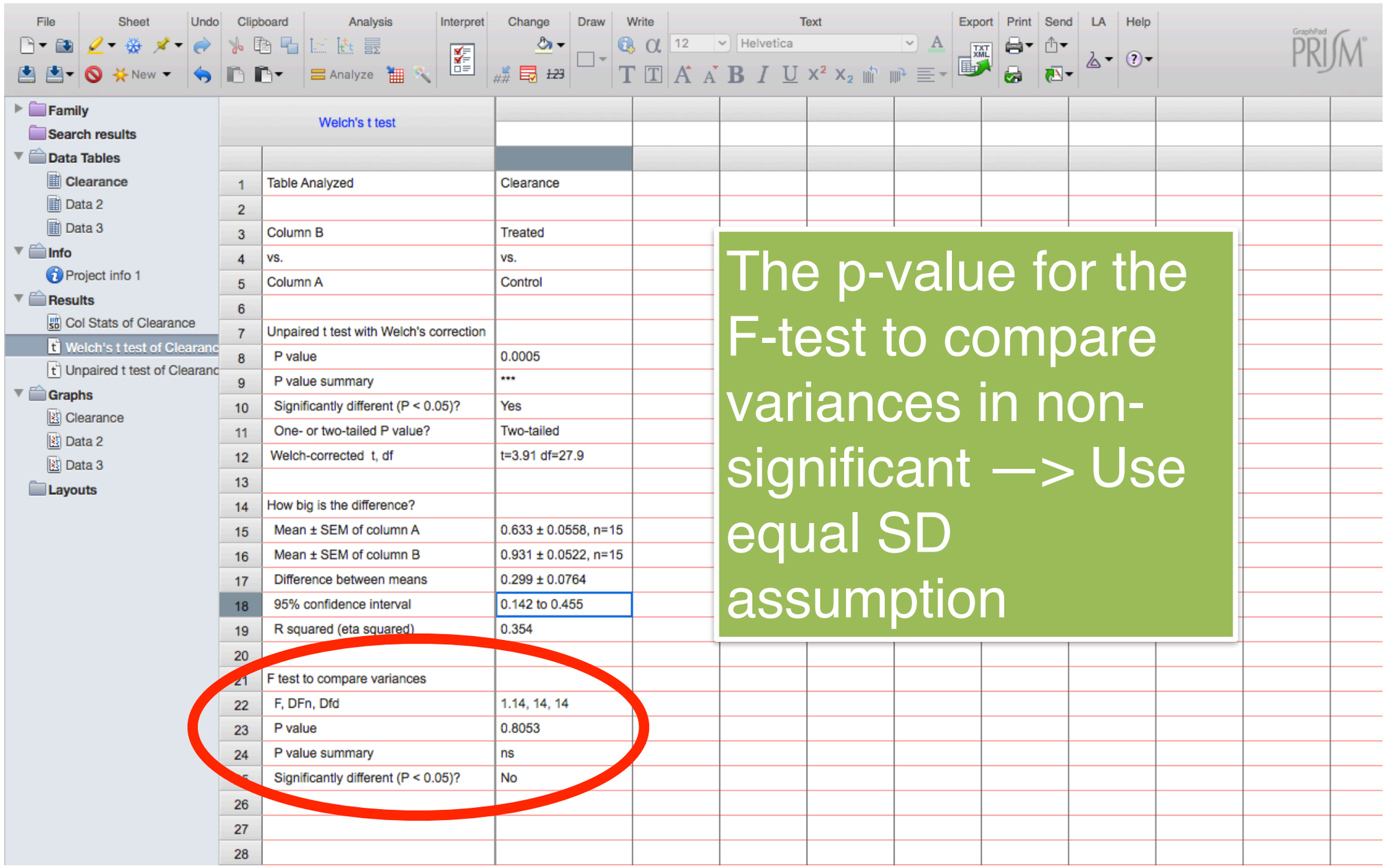
3.Click ‘Options’ for additional model choices

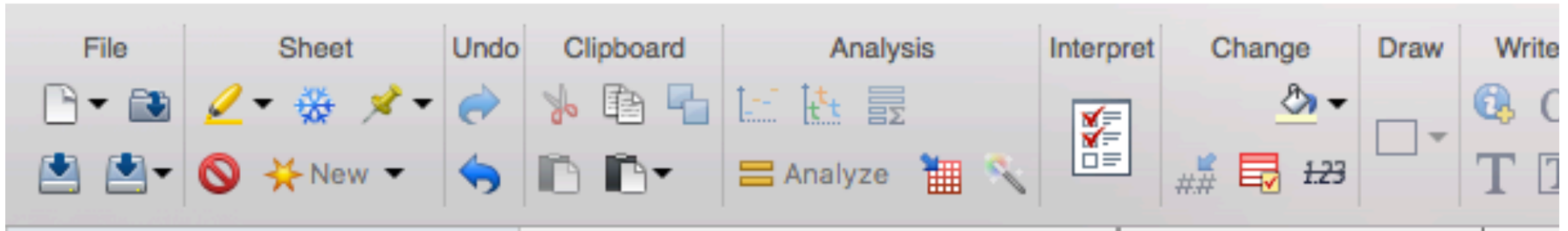


2.Select ‘Unpaired t test with Welch’s correction’ for Choose test



4.Select the difference that you would like reported


File
Sheet
Undo
Clipboard
Analysis
Interpret
Change
Draw
Write
Text
Export
Print
Send
LA
Help
New
Analyze
12
Helvetica
Family
Search results
Data Tables
Clearance
Data 2
Data 3
Info
Project info 1
Results
Col Stats of Clearance
Welch's t test of Clearanc
Unpaired t test of Clearanc
Graphs
Clearance
Data 2
Data 3
Layouts
Welch's t test
1 Table Analyzed Clearance
2
3 Column B Treated
4 vs. vs.
5 Column A Control
6
7 Unpaired t test with Welch's correction
8 P value 0.0005
9 P value summary ***
10 Significantly different (P < 0.05)? Yes
11 One- or two-tailed P value? Two-tailed
12 Welch-corrected t, df t=3.91 df=27.9
13
14 How big is the difference?
15 Mean ± SEM of column A 0.633 ± 0.0558, n=15
16 Mean ± SEM of column B 0.931 ± 0.0522, n=15
17 Difference between means 0.299 ± 0.0764
18 95% confidence interval 0.142 to 0.455
19 R squared (eta squared) 0.354
20
21 F test to compare variances
22 F, DFn, Dfd 1.14, 14, 14
23 P value 0.8053
24 P value summary ns
Significantly different (P < 0.05)? No
26
27
28
The p-value for the F-test to compare variances in non-significant —> Use equal SD assumption



- Family
- Search results
- Data Tables
  - Clearance
  - Data 2
  - Data 3
- Info
  - Project info 1
- Results
  - Col Stats of Clearance
  - Welch's t test of Clearance
  - Unpaired t test of Clearar
- Graphs
  - Clearance
  - Data 2
  - Data 3
- Layouts

Unpaired t test

9	P value summary	***
10	Significantly different (P < 0.05)?	Yes
11	One- or two-tailed P value?	Two-tailed
12	t, df	t=3.909 df=28
13		
14	How big is the difference?	
15	Mean ± SEM of column A	0.6327 ± 0.05581, n=15
16	Mean ± SEM of column B	0.9313 ± 0.05219, n=15
17	Difference between means	0.2987 ± 0.07641
18	95% confidence interval	0.1421 to 0.4552
19	R squared (eta squared)	0.353
20		
21	F test to compare variances	
22	F, DFn, Dfd	1.144, 14, 14
23	P value	0.8053
24	P value summary	ns
25	Significantly different (P < 0.05)?	No
26		

Methods Section

In **Methods** section, report:

- Statistical package used
- Name of procedure as it is referred to in that particular program
- Any option selected that differs from the defaults

EXAMPLE:
To test for a difference in theophylline clearance, we used the unpaired t-test that assumes both populations have equal standard deviations implemented in GraphPad Prism (Version 7.0a).

Results Section

In **Results** section:

- include a figure that shows visually whether there is much evidence of any difference in mean values between the groups and gives an impression of the distribution of the data
- report the means, SDs (or SEM) for the two samples, an estimate of the difference, the limits for the 95% confidence interval for the difference between the two group, and the p-value (described in next chapter)
- indicate whether the difference was significantly different from zero and if it is, indicate the direction of the difference
- if using a Welch test (unequal variance) report the results of the F-test (F-statistic, degrees of freedom, and p-value)

Example Results Section

The mean clearance within the group of subjects that did not receive the pretreatment with rifampicin was 0.63 ml/min/kg (± 0.056, standard error) and the mean clearance within the group of subjects that did receive the pretreatment with rifampicin was 0.93 ml/min/kg (± 0.052, standard error). The clearance in the group pretreated with rifampicin was significantly higher than the clearance of theophylline in the control group (0.30 ml/min/kg, 95% CI = 0.14 to 0.46, p-value = 0.0005).

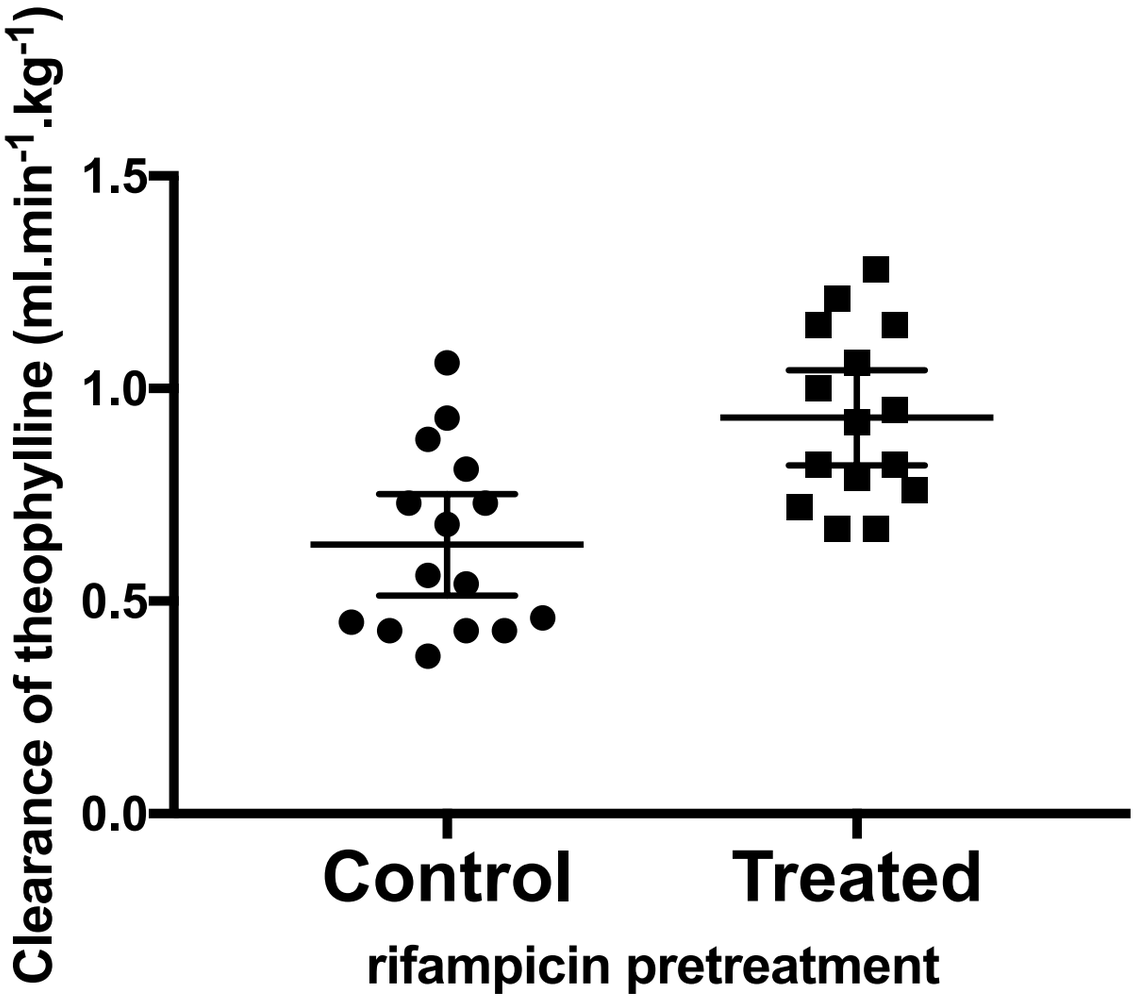


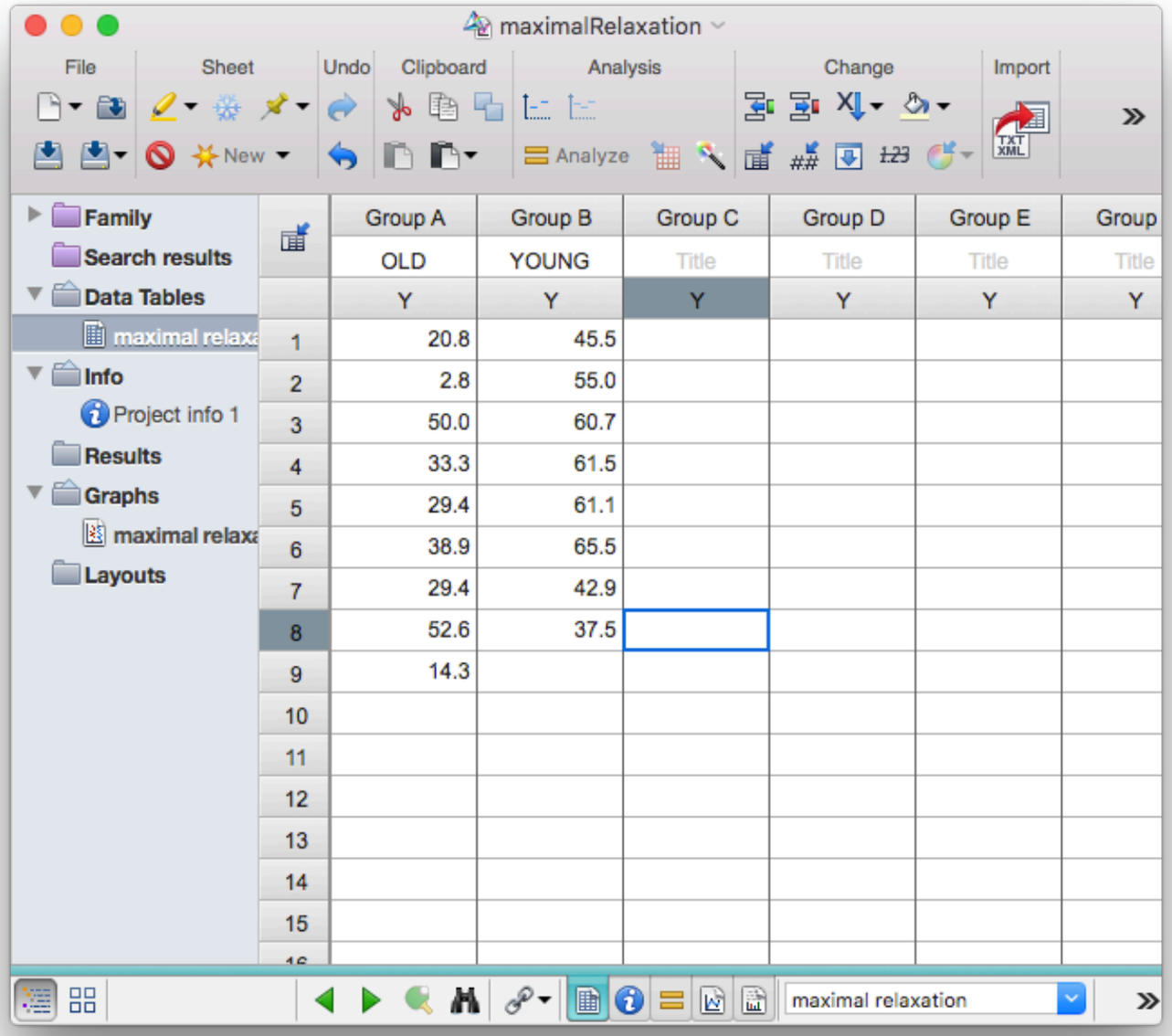
Figure 1. Difference in clearance of theophylline after pretreatment with rifampicin. Each point represents the estimated clearance in an individual subject. Circles represent patients given a placebo treatment prior to administering theophylline. Squares represent patients given a dose of rifampicin prior to administering theophylline. The longer horizontal line in each pretreatment group represents the mean clearance within the group and the error bars represent the 95% confidence interval for the estimate of each mean.

What did we learn?

- A null hypothesis is typically a statement that there is no effect. An alternative hypothesis is typically a statement that there is an effect.
- If a confidence interval contains 0, the null hypothesis is credible. If a confidence interval does not contain 0, the alternative hypothesis is credible.
- Significance does not imply that we positive there is an effect, it simply adds evidence to this hypothesis.
- Type I errors occur in unpaired t-test when we conclude that there is a difference between when the truth is that the two population means do not differ.
- Alpha is the risk of making a type I error when there is no difference between the two population means.
- Size of the experimental effect, variability in the data, and sample size all contribute to whether we detect a significant outcome.
- Data need to be normally distributed ***within*** a group and the groups need to have equal SDs to apply the typical t-test. Adjustments like log transformations and the Welch approximation can help correct for these.
- In the methods section of a paper, report the type of test and software used. In the results, report group means, SDs (or SEMs), estimated difference, and the CI for the difference.

Example Problem

Frazier, Schneider, and Michel (2006) measured how well the neurotransmitter norepinephrine relaxes bladder muscles. Compare the maximal relaxation that can be achieved by large doses of norepinephrine between old and young rats.



- What is the dependent variable and what is the independent variable?

- What are the null and alternative hypotheses?

- Create a figure that includes scatter and the within group mean estimates and the accuracy of the group mean estimates.

- Should you assume equal or unequal variances?

- What is the 95% confidence interval for the difference in means?

- Is the difference in means significant? Why or why not?