

# Correlation

# Power analysis

# Analysis of variance (ANOVA)

# Multiple hypothesis testing

---



Biostatistics Course 2023  
Lecture 4  
Thursday, 27 July 2023  
1:00pm - 3:00pm

# Correlation

## Example: lipids and insulin sensitivity

---

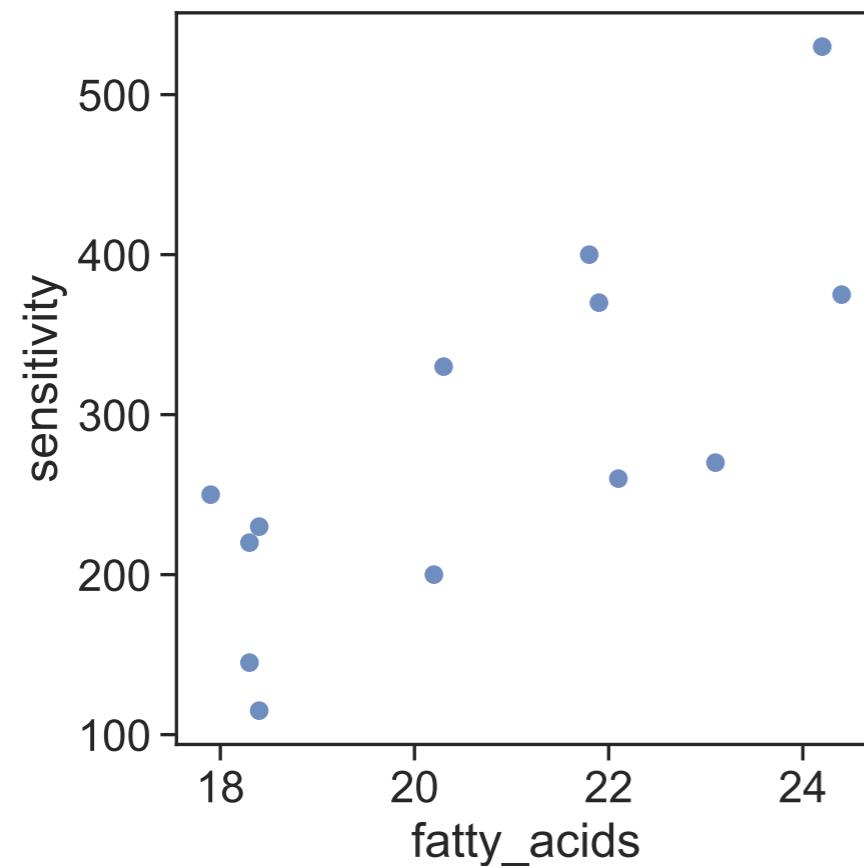
<b>sensitivity</b>	<b>fatty_acid</b>
250	17.9
220	18.3
145	18.3
115	18.4
230	18.4
200	20.2
330	20.3
400	21.8
370	21.9
260	22.1
270	23.1
530	24.2
375	24

Borkman et al. (1993) wanted to understand why insulin sensitivity varies so much among individuals. They hypothesized that the lipid composition of the cell membranes of skeletal muscle affects the sensitivity of the muscle for insulin.

They determined the insulin sensitivity of  $N = 13$  healthy men by infusing insulin at a standard rate (adjusting for size differences) and quantifying how much glucose they needed to infuse to maintain a constant a blood glucose level...

They also took a small muscle biopsy from each subject and measured its fatty acid composition. We'll focus on the fraction of of polyunsaturated fatty acids that have between 20 and 22 carbon atoms ("fatty\_acid").

## Correlation is used to describe relationships between real-numbered variables



summary statistics

pearson	
N	13
r	0.77
95% CI	[0.38, 0.93]
$r^2$	0.593
P-val	0.00207701

## Covariance and correlation are estimated from data in the familiar manner

---

The formula for variance is

$$\widehat{\text{var}}(x) = \sigma_x^2 = \frac{1}{N-1} \sum_i (x_i - \hat{\mu}_x)^2$$

Covariance is estimated in a manner similar to variance

$$\widehat{\text{cov}}(x, y) = \frac{1}{N-1} \sum_i (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)$$

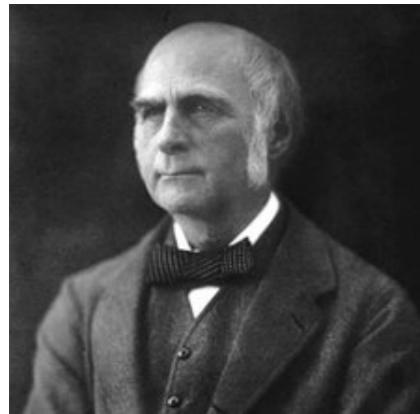
The corresponding “correlation coefficient” is

$$r = \frac{\widehat{\text{cov}}(x, y)}{\hat{\sigma}_x \hat{\sigma}_y}$$

## The correlation coefficient, as used today, was developed by Karl Pearson and Francis Galton in the 1880s

---

$$r = \frac{\widehat{\text{cov}}(x, y)}{\hat{\sigma}_x \hat{\sigma}_y}$$
 is often called “Pearson correlation”



Francis Galton  
(1822-1911)

- Discovered the correlation coefficient in 1888 (independent of Auguste Bravis, in 1844) and proposed the use of “ $r$ ”.
- Also invented linear regression & idea of “regression toward the mean”
- Invented the term “eugenics” in 1883



Karl Pearson  
(1857-1936)

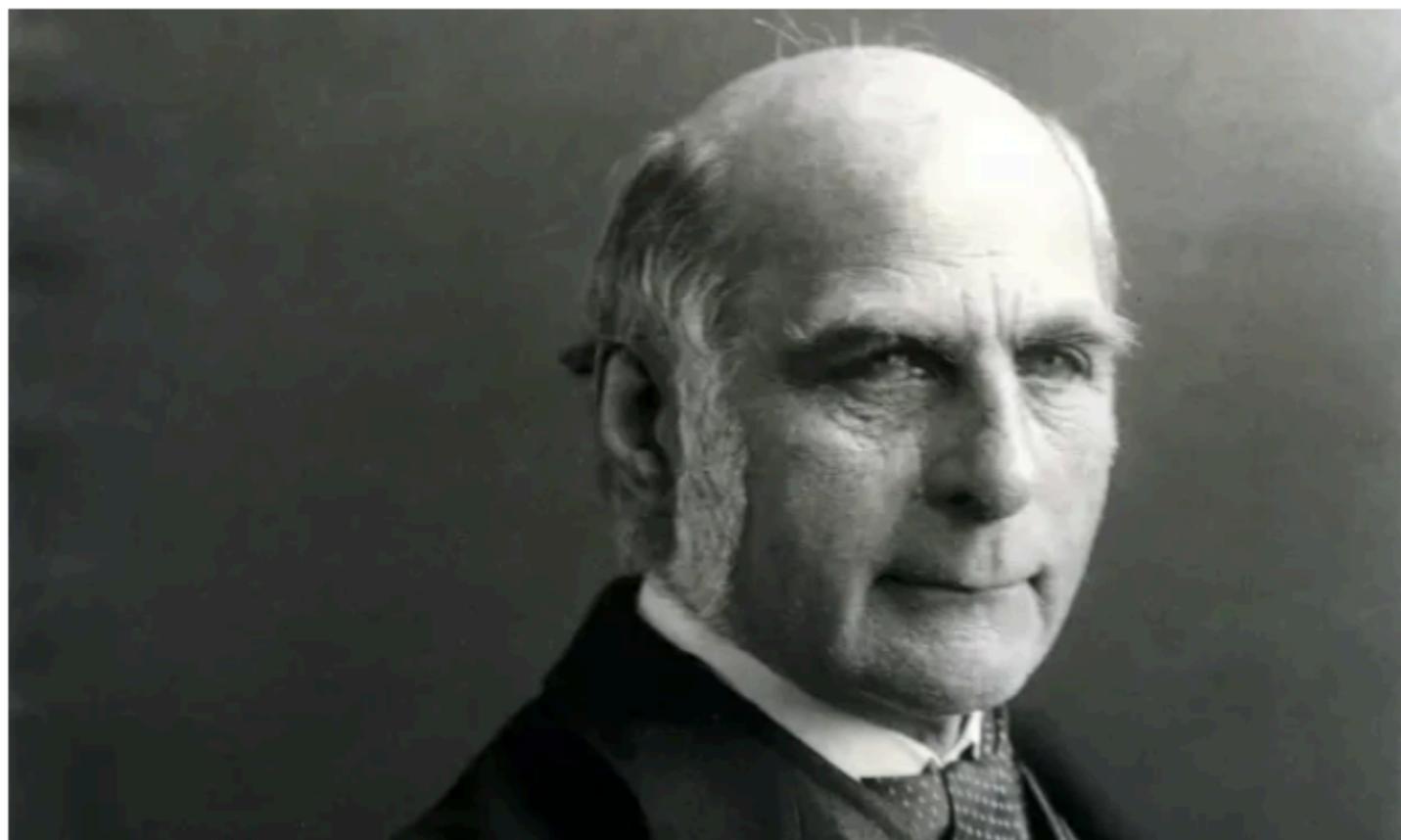
- Is credited with founding the discipline of mathematical statistics.
- Further developed the correlation coefficient proposed by Galton
- Other contributions include P-values,  $\chi^2$  tests, and principal component analysis (PCA)
- Founded world’s first university statistics department at University College, London
- Was a vocal and influential social Darwinist and advocate of race wars.
- Founded the “Annals of Eugenics”, which is now “Annals of Human Genetics”

**UCL**  
**(University**  
**College**  
**London)**

• This article is more than 1 month old

## UCL renames three facilities that honoured prominent eugenicists

**London university removes names of Francis Galton and Karl Pearson from two lecture theatres and a building**



▲ Francis Galton coined the term eugenics in 1883 and endowed UCL with his personal collection and archive.  
Photograph: Corbis via Getty Images

UCL has renamed two lecture theatres and a building that honoured the prominent eugenicists Francis Galton and Karl Pearson.

 Replay

 Learn More

### Sponsored Video

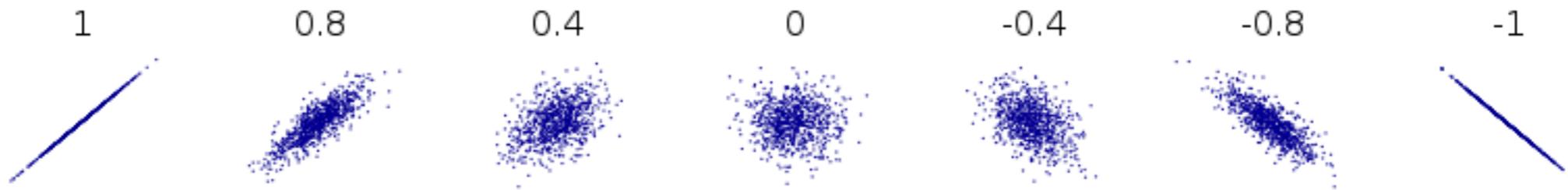
*Advertisement by Advertising Partner*

Watch to learn more

 SEE MORE

## This is what the correlation coefficient looks like

---



Pearson's  $r$  ranges from -1 to 1.

$r = 0$  when the two variables are independent, i.e.  $p(x, y) = p(x) \cdot p(y)$ .

$r = \pm 1$  when the two variables share a deterministic linear relationship.

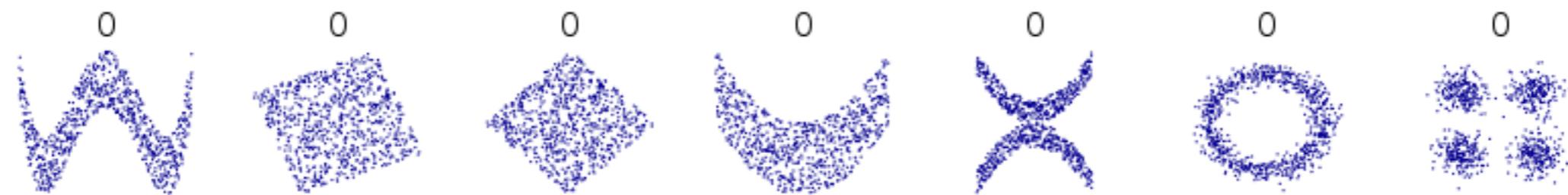
Adding a constant to all  $x$  or all  $y$ , or a multiplicative rescaling of all  $x$  or all  $y$ , do not change  $r$ .

## This is what the correlation coefficient looks like

---



In the deterministic case,  $r$  is unaffected by the magnitude of the slope relating two variables, while the sign of  $r$  is equal to the sign of the slope.



Sometimes  $r = 0$  when two variables have a nonlinear relationship.

## The coefficient of determination another name for $r^2$

---

The coefficient of determination is simply  $r^2$ , which is also often written as  $R^2$ .

$r^2$  is always between 0 and 1 (inclusive)

Remember that  $r^2 \leq |r|$ , so beware of people reporting  $r$  instead of  $r^2$  to make a correlation seem stronger.

$r^2$  is commonly interpreted as the fraction of variance in  $y$  explained by  $x$  (or the other way around).

## P-values correspond to the null hypothesis of no correlation in the underlying distribution

---

The p-value reported alongside values of  $r$  or  $r^2$  corresponds to the null hypothesis that the underlying correlation is zero.

If the underlying correlation is zero, then the quantity

$$t = r \sqrt{\frac{N - 2}{1 - r^2}}$$

follows a t-distribution with  $N - 2$  degrees of freedom.

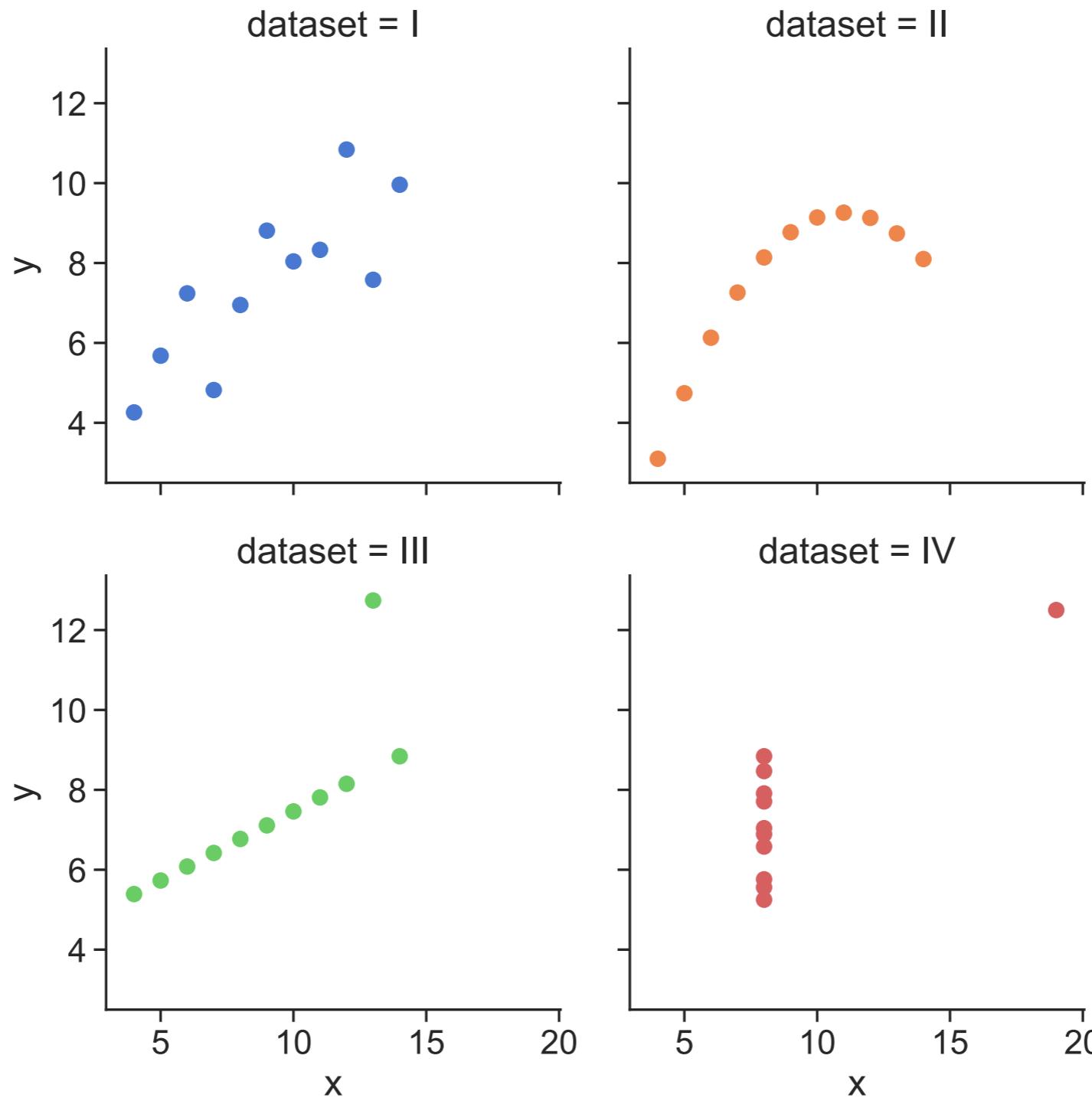
The inverse relationship

$$r = \frac{t}{\sqrt{N - 2 + t^2}}$$

is used to compute values for the 95% confidence interval on  $r$ .

**Lots of different-looking datasets will have the same value for  $r$ .**

“Anscombe’s quartet”:  $r = 0.816$  for all 4 datasets



## Assumptions underlying correlation

---

Interpreting the correlation coefficient  $r$ , and especially the associated P-value, requires multiple assumptions:

- Each data point  $(x, y)$  is independently sampled from a 2D Gaussian distribution.
- In particular,  $x$  and  $y$  each follow a 1D Gaussian distribution
- All covariation between  $x$  and  $y$  is linear, with perfect concordance disrupted only by Gaussian noise.

## There are usually many explanations for why two variables might correlate

---

Possible reasons for a correlation between lipid levels and insulin sensitivity:

- The lipid content of membranes affects insulin sensitivity
- The insulin sensitivity affects membrane lipid content
- Both insulin sensitivity and lipid content are under the control of some third factor, such as a hormone.
- Lipid content, insulin sensitivity, and other factors are all part of a complex molecular/biochemical/physiological network, perhaps with positive and/or negative feedback components. The correlation observed is just a peak at a much more complex set of interdependent relationships.
- Membrane lipid content and insulin sensitivity don't actually correlate at all; the result is just a coincidence.

Welcome to GraphPad Prism

**XY tables: Each point is defined by an X and Y coordinate**

**NEW TABLE & GRAPH**

**XY** 

Column

Group

Contingency

Survival

Parts of Whole

Multiple variables

Nested

**EXISTING FILE**

Open a File

LabArchives

Clone a Graph

Graph Portfolio

**GraphPad Prism Version 8.2.1 (279)**

**Data table:**

Enter or import data into a new table

Start with sample data to follow a tutorial

**Options:**

X:  Numbers

Numbers with error values to plot horizontal error bars

Dates

Elapsed times

Y:  Enter and plot a single Y value for each point

Enter **3**  replicate values in side-by-side subcolumns

Enter and plot error values already calculated elsewhere

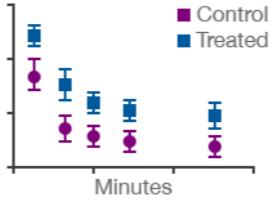
Enter: **Mean, SD, N** 

Prism Tips

Cancel

Create 

	X	A			B		
	Minutes	Control			Treated		
	X	A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3
1	Title						
2	Title						
3	Title						



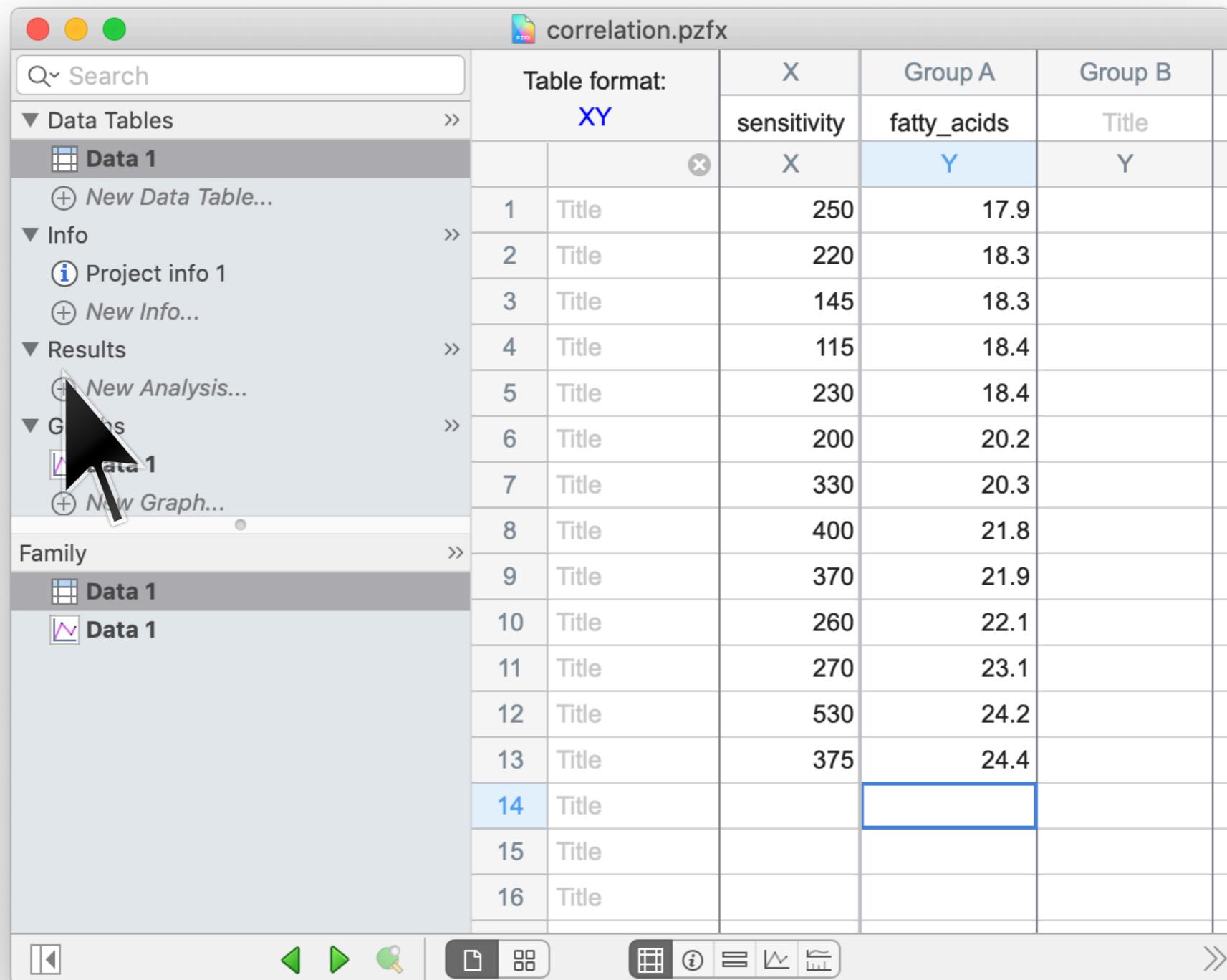
Minutes

Control

Treated

?

Learn more



## Create New Analysis

### Data to analyze

Table: Data 1

### Type of analysis

Which analysis?

#### ▼ Transform, Normalize...

- Transform
- Transform concentrations (X)
- Normalize
- Prune rows
- Remove baseline and column math
- Transpose X and Y
- Fraction of Total

#### ▼ XY analyses

- Nonlinear regression (curve fit)
- Linear regression
- Fit spline/LOWESS
- Smooth, differentiate or integrate curve
- Area under curve
- Deming (Model II) linear regression
- Row means with SD or SEM

#### Correlation

- Interpolate a standard curve

#### ► Column analyses

#### ► Grouped analyses

#### ► Contingency table analyses

#### ► Survival analyses

Analyze which data sets?

A:fatty\_acids

When you analyze tables or graphs with more than one data set, use this space to select which data set(s) to analyze.

Select All

Deselect All

?

Cancel

OK

Parameters: Correlation

**Compute correlation between which pairs of columns?**

Compute r for every pair of Y data sets (Correlation matrix)

Compute r for X vs. every Y data set:  
X: sensitivity

Compute r between two selected data sets:  
X: sensitivity  
A: fatty\_acids

**Assume data are sampled from Gaussian distributions?**

Yes. Compute Pearson correlation coefficients

No. Compute nonparametric Spearman correlation

**Options**

P value:  One-tailed  Two-tailed

Confidence interval: 95%

**Output**

Show this many significant digits (for everything except P values): 4

P Value Style: GP: 0.1234 (ns), 0.0332 (\*), 0.0021 (\*\*),... N= 6

**Graphing**

Create a heatmap of the correlation matrix

Make these choices the default for future analyses



The screenshot shows the PAST 3.24 software interface with the following details:

**File:** correlation.pzfx — Edited

**Search:** Search bar at the top left.

**Data Tables:** Section showing "Data 1" and "New Data Table...".

**Info:** Section showing "Project info 1" and "New Info...".

**Results:** Section showing "Correlation of Data 1" (selected), "New Analysis...", and "Graphs".

**Graphs:** Section showing "Data 1" (selected).

**Family:** Section showing "Data 1" and "Correlation" (selected).

**Analysis:** Correlation analysis for "Data 1".

**Correlation:** Analysis title.

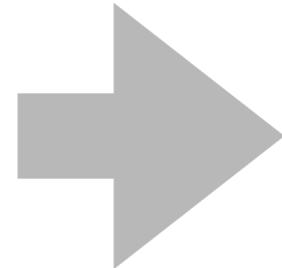
	A	B
1	sensitivity vs. fatty_acids	Title
2	Y	Y
3	Pearson r	
4	r	0.7700
5	95% confidence interval	0.3804 to 0.9275
6	R squared	0.5929
7		
8	P value	
9	P (two-tailed)	0.0021
10	P value summary	**
11	Significant? (alpha = 0.05)	Yes
12		
13	Number of XY Pairs	13
14		

Bottom navigation bar: Back, Forward, Home, Search, Help, File, Edit, View, Plots, Analyses, Data, Tools, Options, Help, Exit.

## Spearman's rank correlation is a non-parametric measure of dependence

Spearman's  $\rho$  is just Pearson's  $r$  computed on the ranks of the  $x$  and  $y$  values

x	y
17.9	250
18.3	220
18.3	145
18.4	115
18.4	230
20.2	200
20.3	330
21.8	400
21.9	370
22.1	260
23.1	270
24.2	530
24.4	375



x rank	y rank
1.0	6.0
2.5	4.0
2.5	2.0
4.5	1.0
4.5	5.0
6.0	3.0
7.0	9.0
8.0	12.0
9.0	10.0
10.0	7.0
11.0	8.0
12.0	13.0
13.0	11.0

Parameters: Correlation

**Compute correlation between which pairs of columns?**

Compute r for every pair of Y data sets (Correlation matrix)

Compute r for X vs. every Y data set:  
X: sensitivity

Compute r between two selected data sets:  
X: sensitivity  
A: fatty\_acids

**Assume data are sampled from Gaussian distributions?**

Yes. Compute Pearson correlation coefficients

No. Compute nonparametric Spearman correlation

**Options**

P value:  One-tailed  Two-tailed

Confidence interval: 95%

**Output**

Show this many significant digits (for everything except P values): 4

P Value Style: GP: 0.1234 (ns), 0.0332 (\*), 0.0021 (\*\*),... N= 6

**Graphing**

Create a heatmap of the correlation matrix

Make these choices the default for future analyses

# Mutual information is a universal measure of dependence that plays a fundamental role in information theory

---

Mutual information is symmetric:

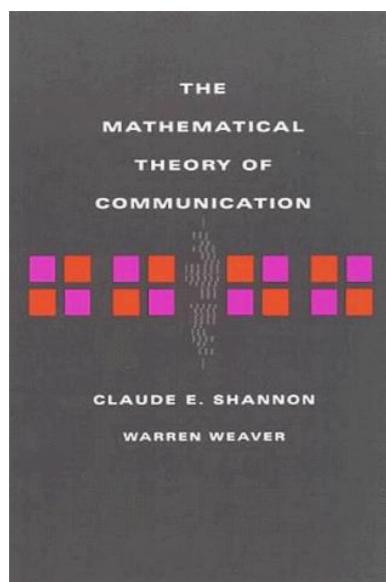
$$I[x; y] = I[y; x]$$



Claude Shannon

Mutual information can range from zero to infinity:

$$0 \leq I[x; y] \leq \infty$$



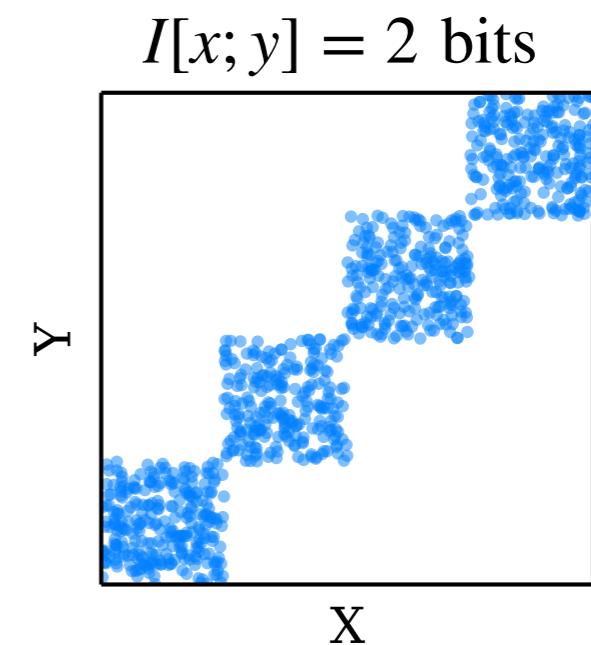
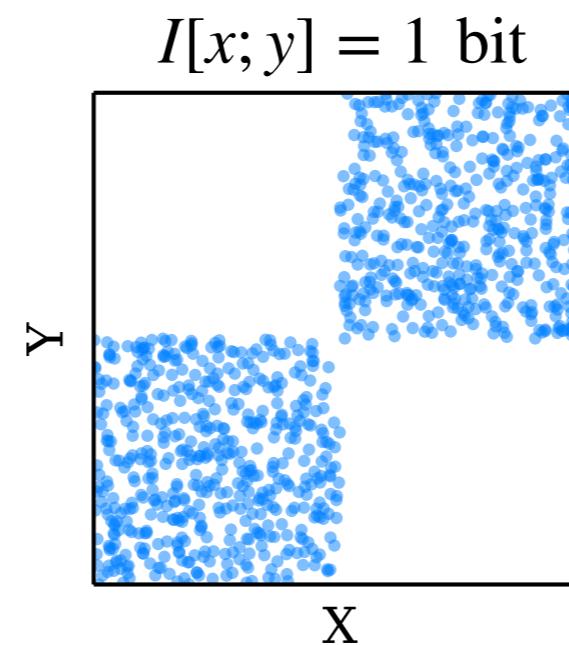
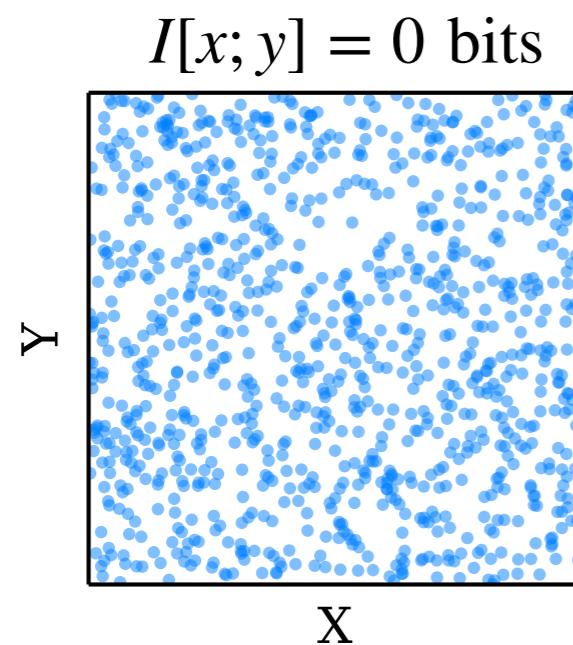
Shannon, 1948

Mutual information is zero only when the two variables are independent:

$$I[x; y] = 0 \Leftrightarrow p(x, y) = p(x) p(y)$$

Mutual information is the amount of information in “bits” that knowing one variable tells you about the value of another variable

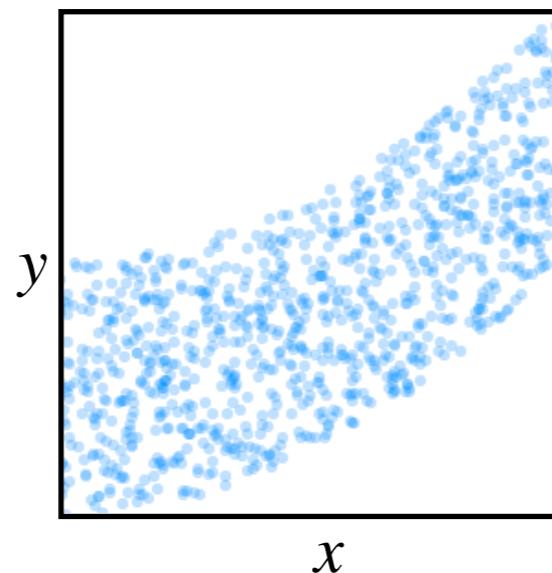
---



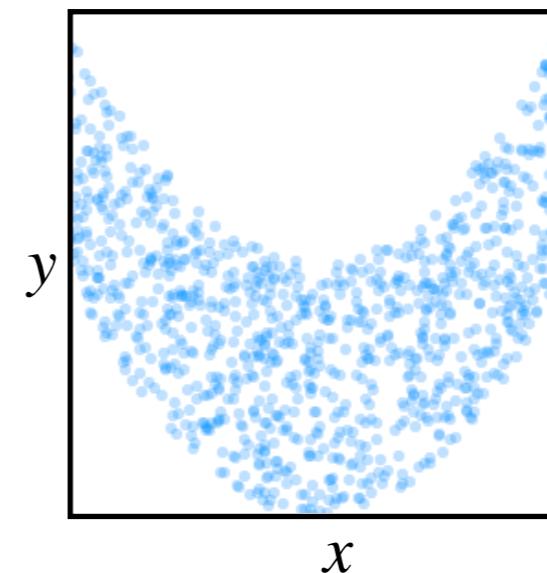
Mutual information, unlike Pearson correlation, quantifies nonlinear and non monotonic relationships in a meaningful way

---

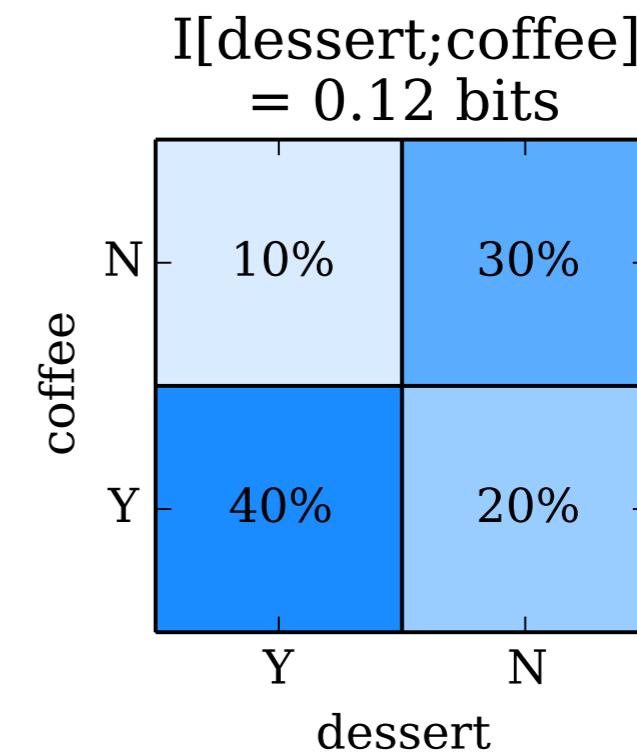
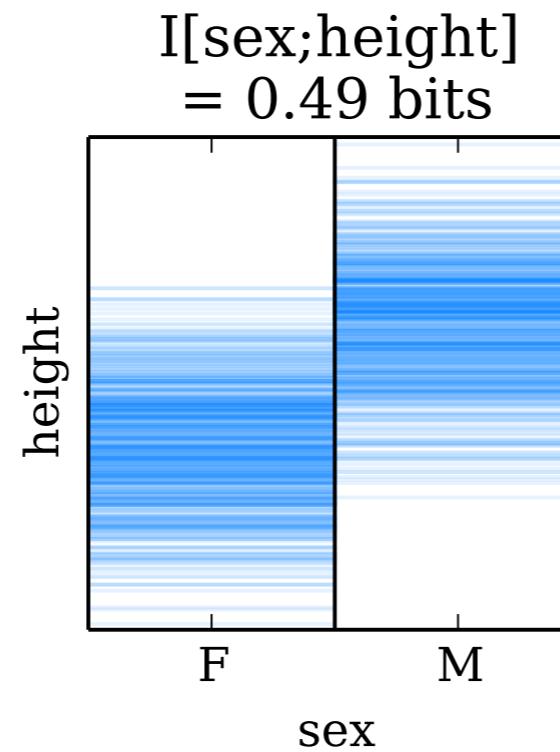
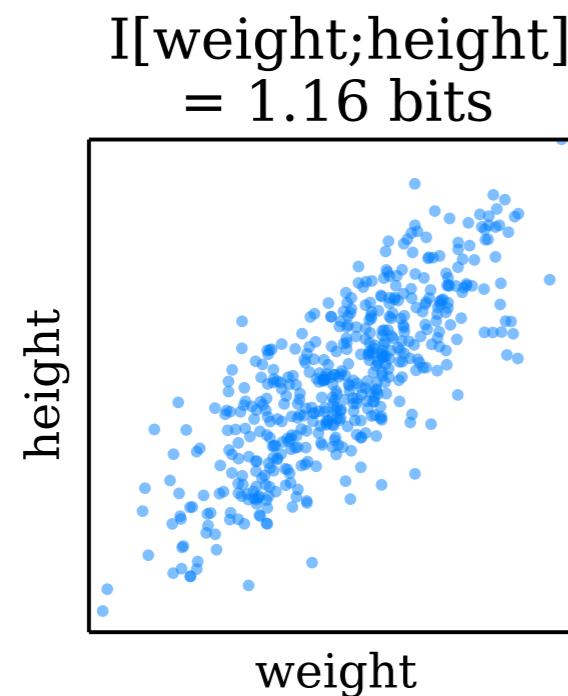
$$I[x; y] = 0.72 \text{ bits}$$
$$r^2 = 0.48$$



$$I[x; y] = 0.72 \text{ bits}$$
$$r^2 = 0$$



**Mutual information information can be evaluated between any two types of variables.**



Unfortunately, there is no simple plug-in formula for computing mutual information from data.

**Mutual information is not commonly used in biological data analysis.**

## Power analysis

## **Statistical power is the probability of detecting an effect that actually does exist.**

---

### **power:**

The probability of getting a statistically significant result if the null hypothesis actually is actually false.

### **power analysis:**

The process of assigning and/or computing four quantities (sometimes more) that describe one's experiment:

1. The sample size  $N$
2. The false positive probability  $\alpha$  (confidence =  $1 - \alpha$ )
3. The false negative probability  $\beta$  (power =  $1 - \beta$ )
4. The anticipated effect size

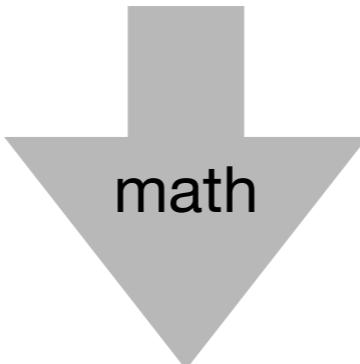
## Example: sex ratio

---

1. Confidence level:  $1 - \alpha = 95\%$
2. Number of birth records:  $N = 19500$
3. Hypothesized effect size:  $|p(\text{boy}) - p(\text{girl})| = 2\%$

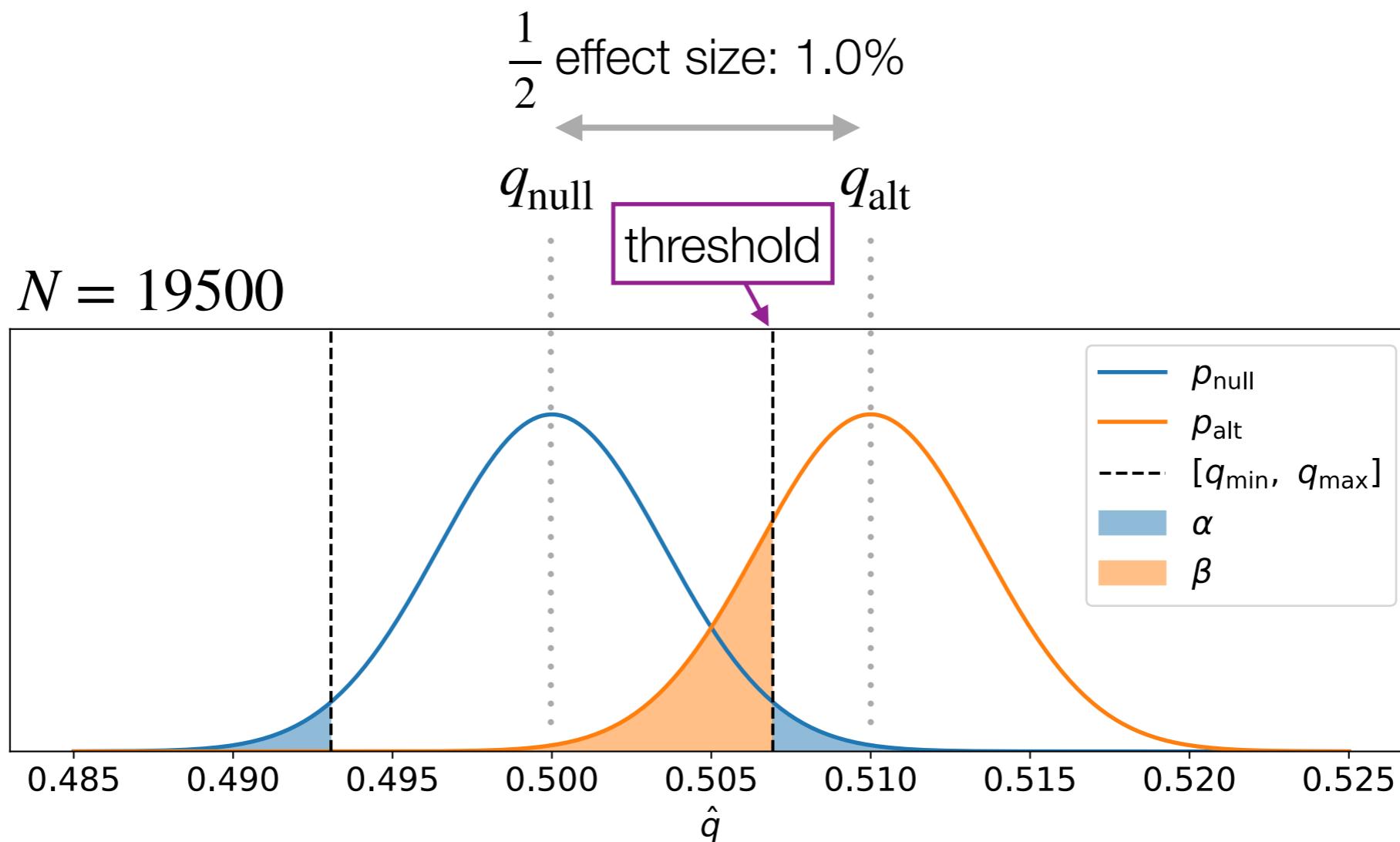
The key parameter is  $q = p(\text{boy})$ , so we use

$$q_{\text{null}} = 50\%, \quad q_{\text{alt}} = 51\%$$



4. We compute a statistical power of:  $1 - \beta = 80\%$

## Statistical power example: sex ratio data



False Positive Probability:  $\alpha = 0.05$

False Negative Probability:  $\beta = 0.20$   
(or 80% power)

## Power analysis claims come in different forms

---

There are four relevant parameters:  $N$ ,  $\alpha$ ,  $\beta$ , and effect size.

Power analysis involves assuming values for any three parameters and computing the value of the forth

“Controlling the false positive rate at  $\alpha = 5\%$  , the statistical power at  $1 - \beta = 80\%$  , and assuming an effect size of  $2\%$  , our study will require using  $N = 19500$  birth records.”

“Using  $N = 19500$  birth records, controlling the false positive rate at  $\alpha = 5\%$  , and assuming a  $2\%$  effect size, our study will have  $1 - \beta = 80\%$  power.”

“Controlling the false positive rate at  $\alpha = 5\%$  , the statistical power at  $1 - \beta = 80\%$  , and using  $N = 19500$  birth records, our study will be sensitive to an effect size of  $2\%$  .”

"Using  $N = 19500$  birth records, assuming an effect size of  $2\%$  , and holding the statistical power to  $1 - \beta = 80\%$  , our study will be able to hold the false positive rate to  $\alpha = 5\%$  .”

**You will most likely do one of these two things:**

---

**You are supposed to do this:**

1. Assume a false positive rate of  $\alpha = 5\%$  (standard)
2. Assume a power of  $1 - \beta = 80\%$  (standard)
3. Assume what you consider to be a biologically significant effect size
4. Compute & use the required sample size  $N$ .

**You'll actually probably do this:**

1. Assume a false positive rate of  $\alpha = 5\%$  (standard).
2. Assume a power of  $1 - \beta = 80\%$  (standard)
3. Assume a reasonable / affordable sample size  $N$
4. Compute & report the detectable effect size.

If the  
detectable  
effect size  
is too small



## Power analysis example: body temperature

---

1. Assume a false positive rate of  $\alpha = 5\%$  (standard).
2. Assume a power of  $1 - \beta = 80\%$  (standard)
3. Assume what you consider to be a biologically significant effect size:  $\Delta\mu = 0.1 \text{ F}$ .  $\Delta\mu = 0.2 \text{ F}$

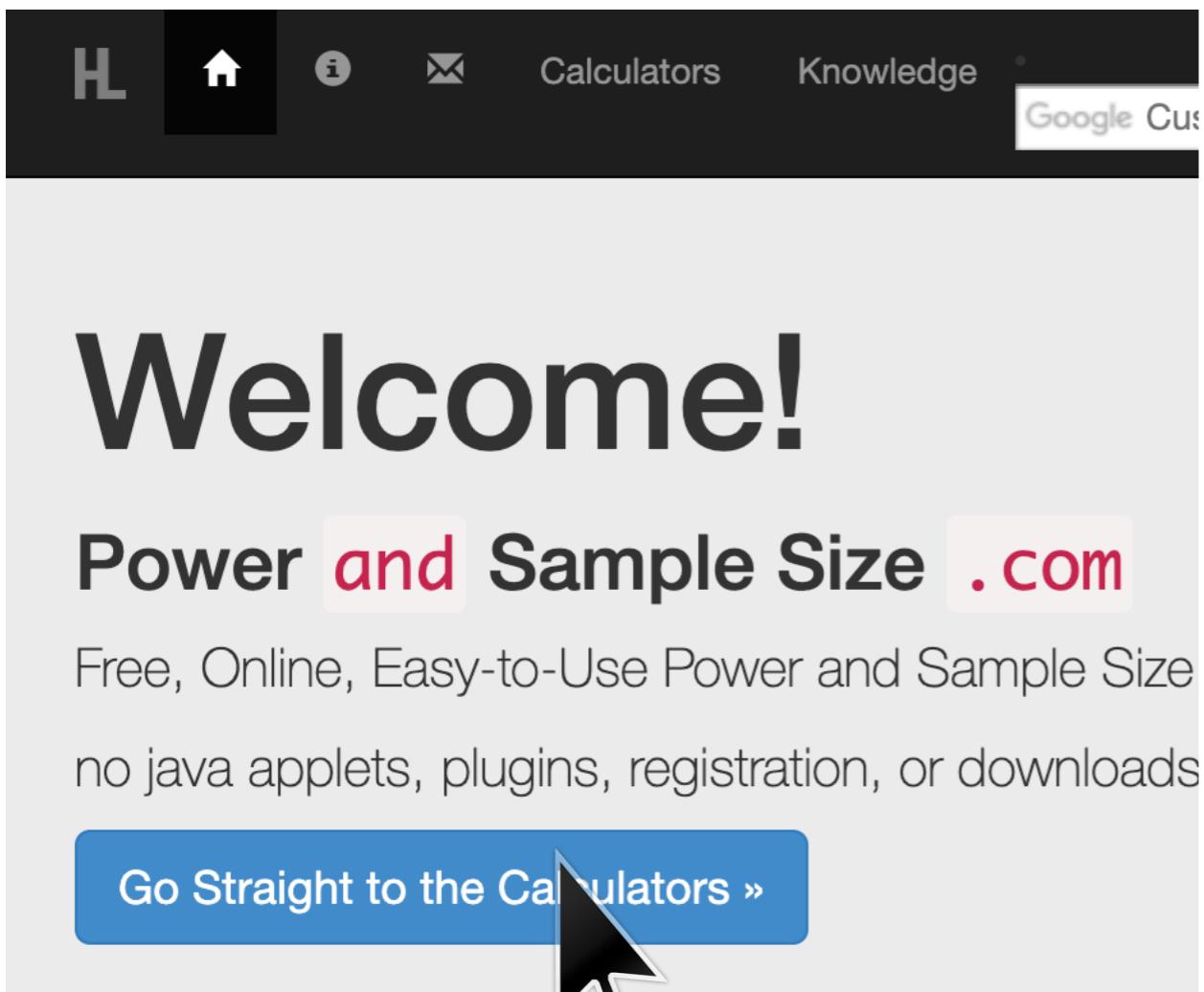
The key parameter is the “normalized effect size”: 
$$\frac{\Delta\mu}{\sigma}$$

From preliminary data, we know  $\sigma \approx 0.7 \text{ F}$

4. Compute the required sample size:  $\cancel{N = 1540}$   $N = 386$   
Too big!      OK.

## There are a number of online power analysis calculators

<http://powerandsamplesize.com/>

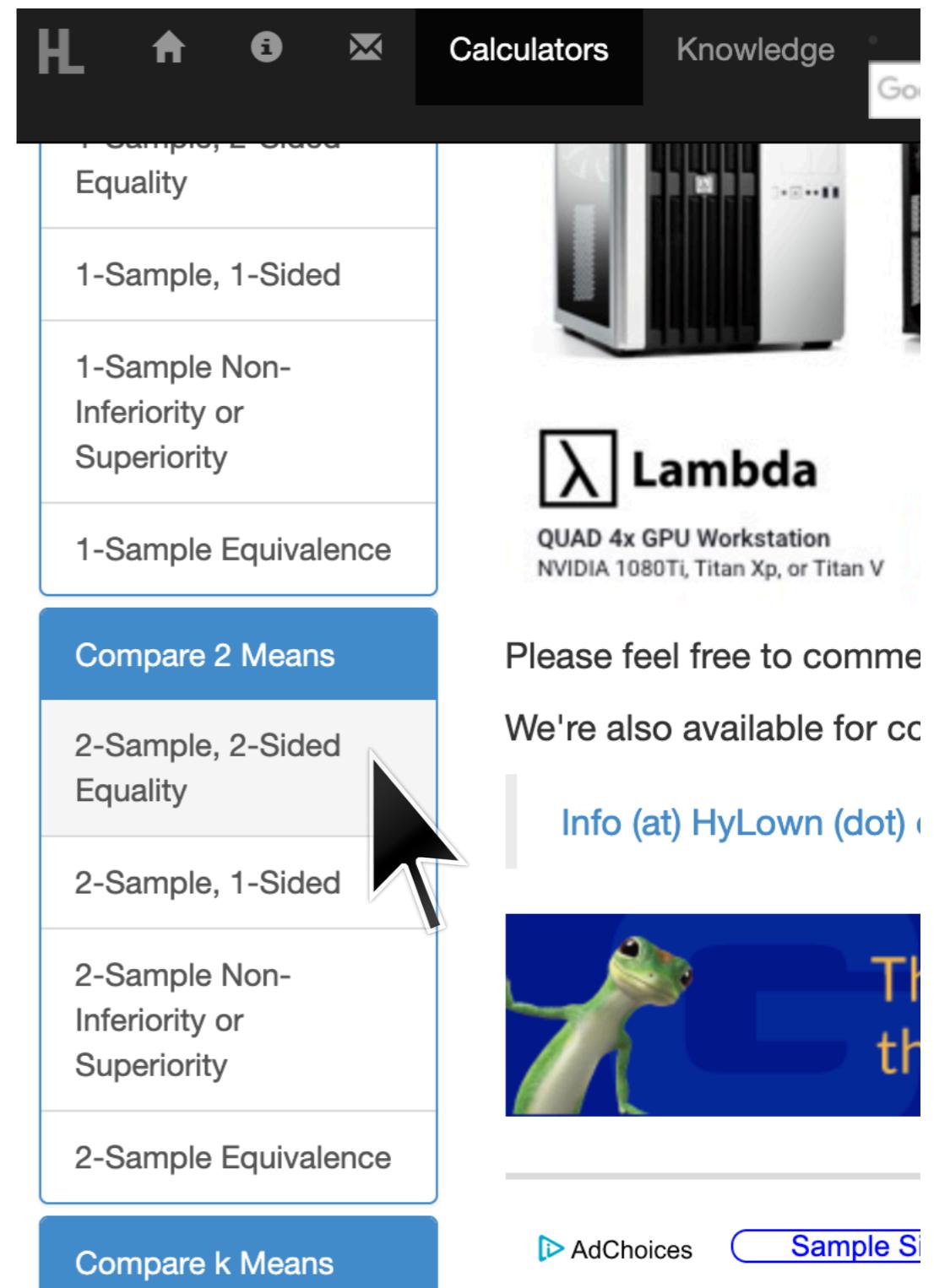


Welcome!

Power and Sample Size .com

Free, Online, Easy-to-Use Power and Sample Size calculators. No Java applets, plugins, registration, or downloads required.

Go Straight to the Calculators »



Equality

1-Sample, 1-Sided

1-Sample Non-Inferiority or Superiority

1-Sample Equivalence

Compare 2 Means

2-Sample, 2-Sided Equality

2-Sample, 1-Sided

2-Sample Non-Inferiority or Superiority

2-Sample Equivalence

Compare k Means

**λ Lambda**  
QUAD 4x GPU Workstation  
NVIDIA 1080Ti, Titan Xp, or Titan V

Please feel free to comment or ask questions. We're also available for consulting services.

Info (at) HyLown (dot) com

Sample Size Calculators

**Calculate:** Sample Size

Sample Size,  $n_B$   
192

Power,  $1 - \beta$   
0.80

Type I error rate,  $\alpha$   
5%

---

98.1      **Group 'A' mean,  $\mu_A$**

98.3      **Group 'B' mean,  $\mu_B$**

0.7      **Standard Deviation,  $\sigma$**

1      **Sampling Ratio,  $\kappa = n_A/n_B$**

**Calculate**

Calculate: Power

Sample Size,  $n_B$  250

Power,  $1 - \beta$  0.892

Type I error rate,  $\alpha$  5%

98.1 Group 'A' mean,  $\mu_A$

98.3 Group 'B' mean,  $\mu_B$

0.7 Standard Deviation,  $\sigma$

1 Sampling Ratio,  $\kappa = n_A/n_B$

Calculate

98.1

98.3

0.7

1

Group 'A' mean,  $\mu_A$

Group 'B' mean,  $\mu_B$

Standard Deviation,  $\sigma$

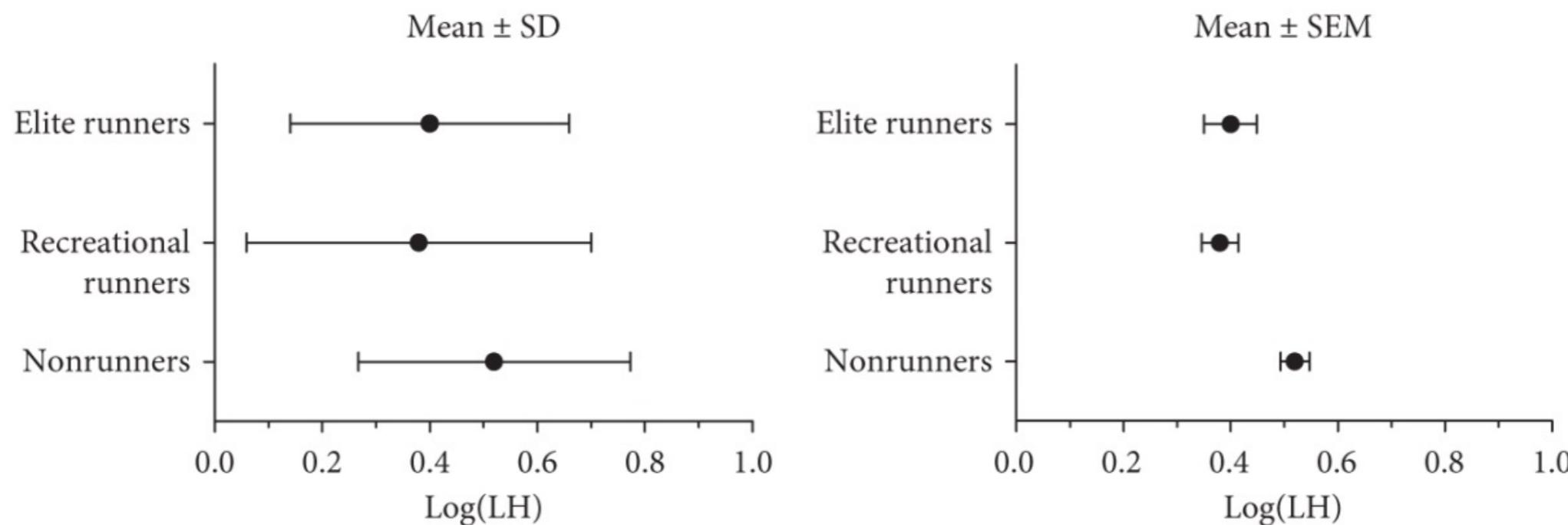
Sampling Ratio,  $\kappa = n_A/n_B$

Calculate

## **Analysis of variance (ANOVA)**

## One-way ANOVA example: hormone levels in runners

Hetland et al. (1993) investigated the level of luteinizing hormone (LH) in runners. Runners were classified into three groups: elite runners, recreational runners, and nonrunners.



GROUP	LOG(LH)	SD	SEM	N
nonrunners	0.52	0.25	0.027	88
recreational runners	0.38	0.32	0.034	89
elite runners	0.40	0.26	0.049	28

## One-way ANOVA analyzes whether group means are significantly different

---

**Null hypothesis:** different groups have identical means

**Alternative hypothesis:** different groups have different means

SS = sum of squares

$$\sum_i \text{SS}_{\text{total}} = \sum_i \text{SS}_{\text{within}} + \sum_i \text{SS}_{\text{between}}$$
$$\sum_i (y_i - \hat{\mu})^2 = \sum_i (y_i - \hat{\mu}_{g_i})^2 + \sum_i (\hat{\mu}_{g_i} - \hat{\mu})^2$$

grand mean:

$$\hat{\mu} = \frac{1}{N} \sum_i y_i$$

group means:

$$\hat{\mu}_g = \frac{1}{N_g} \sum_{i|g} y_i$$

fraction variance explained:

$$\eta^2 = R^2 = \frac{\text{SS}_{\text{between}}}{\text{SS}_{\text{total}}}$$

## One-way ANOVA analyzes whether group means are significantly different

$$\sum_i SS_{\text{total}} = \sum_i SS_{\text{within}} + \sum_i SS_{\text{between}}$$
$$\sum_i (y_i - \hat{\mu})^2 = \sum_i (y_i - \hat{\mu}_{g_i})^2 + \sum_i (\hat{\mu}_{g_i} - \hat{\mu})^2$$

DF = degree of freedom

$$DF_{\text{within}} = N - G, \quad MS_{\text{within}} = \frac{SS_{\text{within}}}{DF_{\text{within}}}$$

MS = mean square

similar if null is true

$$DF_{\text{between}} = G - 1, \quad MS_{\text{between}} = \frac{SS_{\text{between}}}{DF_{\text{between}}}$$

The corresponding F statistic is:  $F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$

$F \approx 1$   
if null is true

The null hypothesis, implies that:  $F \sim F\text{Dist}(DF_{\text{between}}, DF_{\text{within}})$

## Alternatively, ANOVA can be thought of as a form of linear regression

---

$$y_i = \log(LH)$$

$$x_{i1} = \text{elite runner? (1 or 0)}$$

$$x_{i2} = \text{recreational runner? (1 or 0)}$$

**Null model:**  $y = \beta_0 + \epsilon_i$

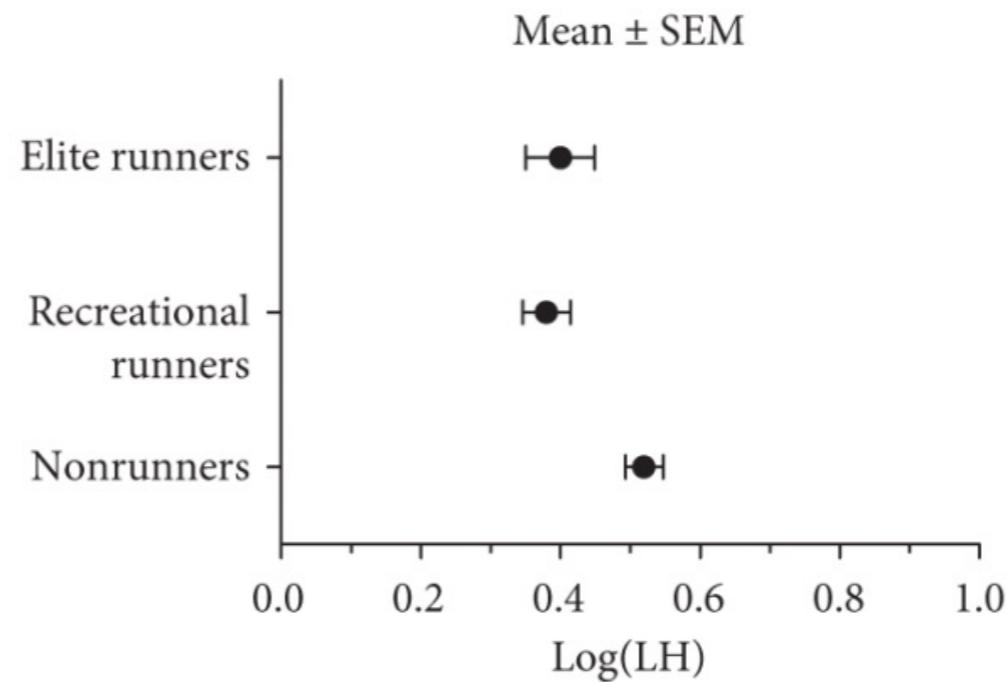
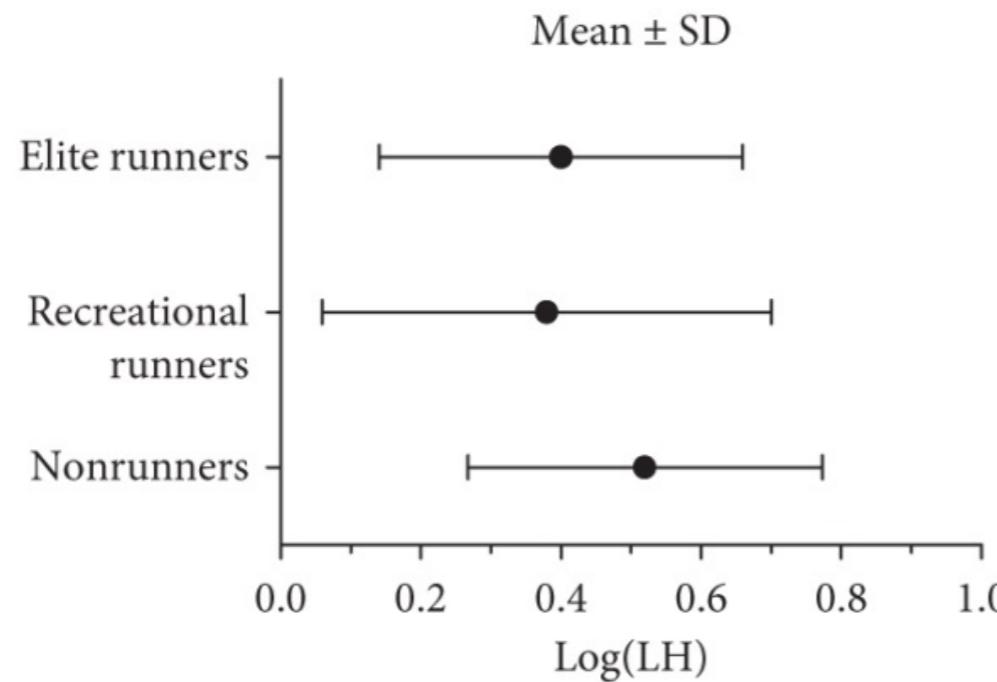
**Alternative model:**  $y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$

parameter correspondence:  $\beta_0 = \mu_{\text{non}}$ ,  $\beta_1 = \mu_{\text{elite}} - \mu_{\text{non}}$ ,  $\beta_2 = \mu_{\text{rec}} - \mu_{\text{non}}$

The alternative model will always fit the data better. But how much better?

The  $F$  test tests whether the extra parameters,  $\beta_1$  and  $\beta_2$  are worth it.

## One-way ANOVA analyzes whether group means are significantly different

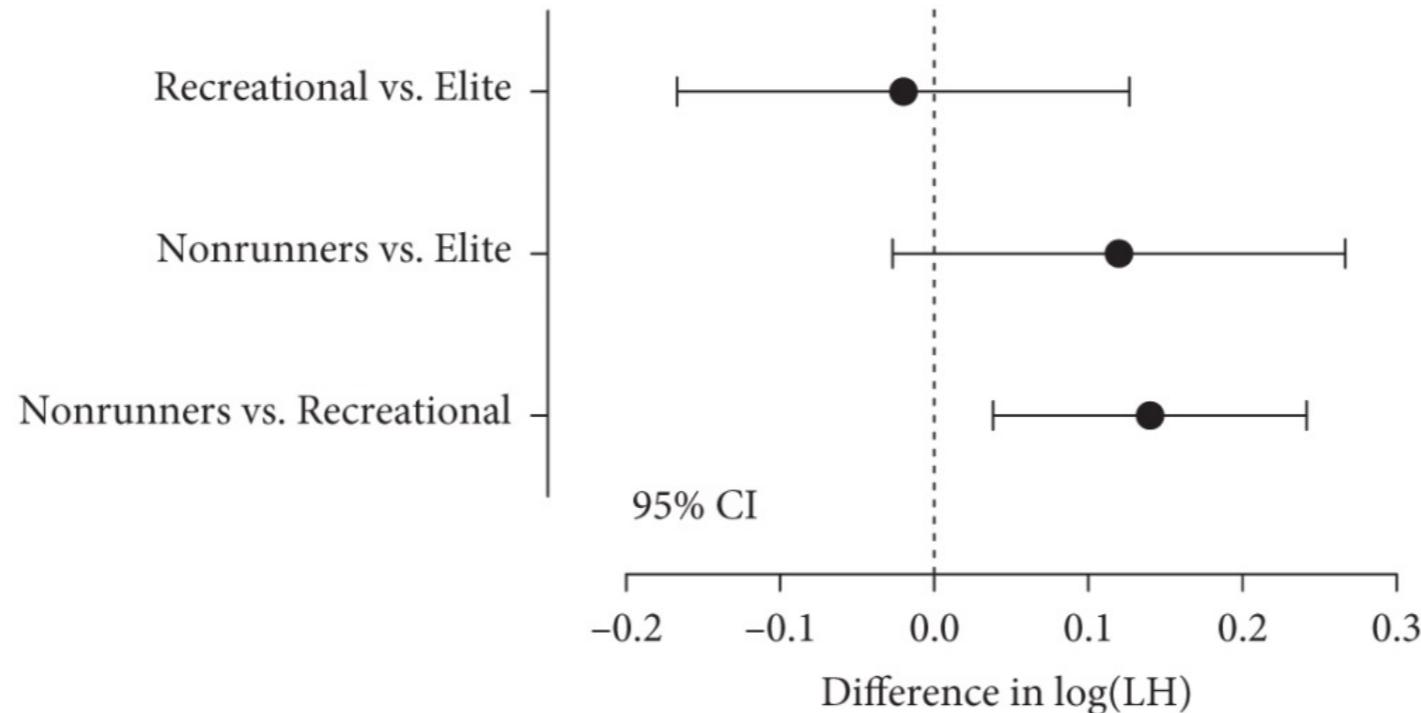


SOURCE OF VARIATION	SUM OF SQUARES	DF	MS	F RATIO	P VALUE
Between groups	0.93	2	0.46	5.69	0.0039
- Within groups (resid.)	16.45	202	0.081		
= Total	17.38	204			

This shows that the three groups have significantly different means. It does NOT, however, say which pairs of means (if any) are significantly different.

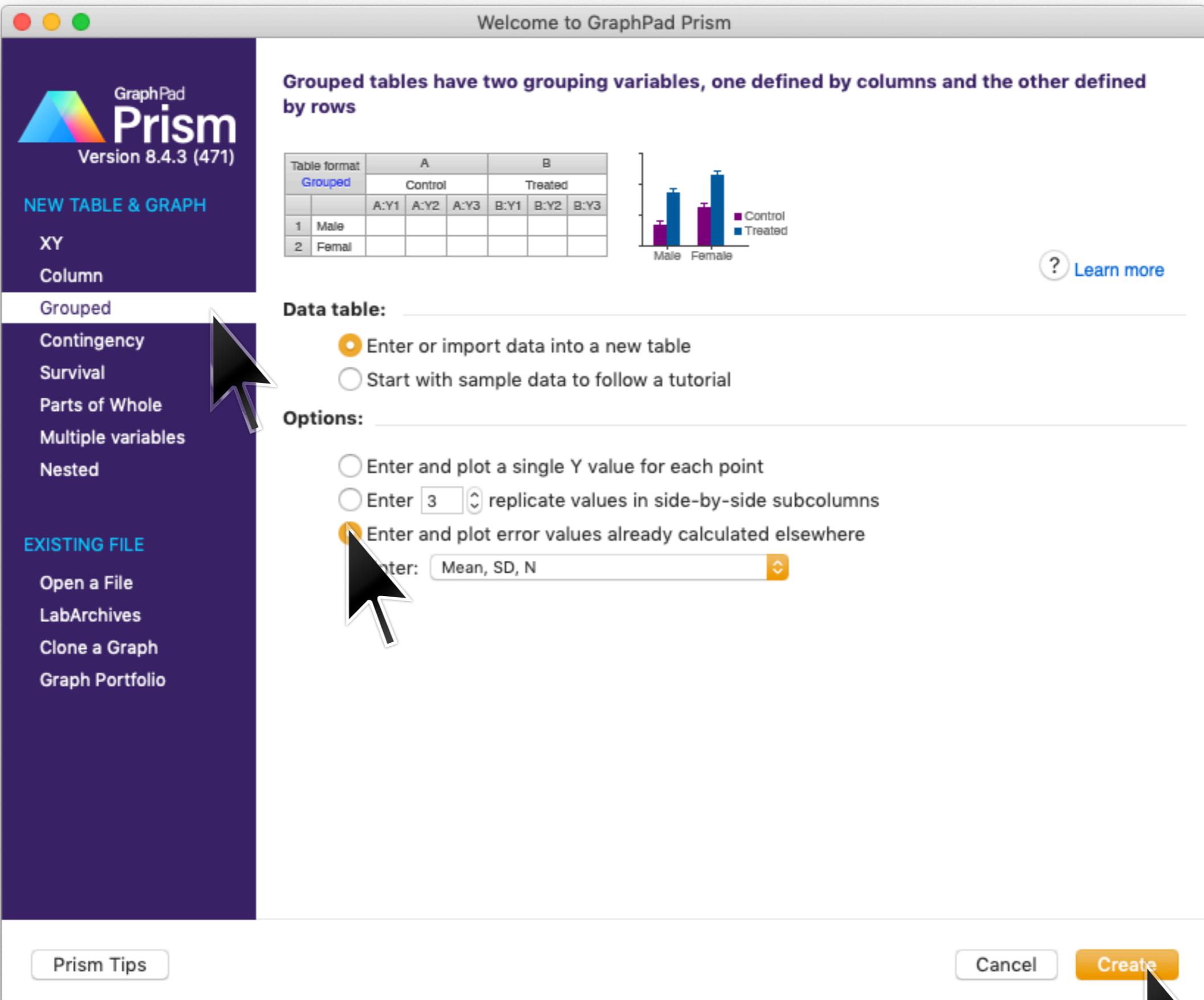
**Tukey's test analyzes which pairwise comparisons in a one-way ANOVA, if any, are significant.**

---



Tukey's test automatically incorporates the necessary multiple hypothesis correction into the test of significance.

There are other ANOVA post-hoc tests as well.





## Create New Analysis

### Data to analyze

Table: Data 1

### Type of analysis

Which analysis?

- ▼ Transform, Normalize...
  - Transform
  - Transform concentrations (X)
  - Normalize
  - Prune rows
  - Remove baseline and column math
  - Transpose X and Y
  - Fraction of Total
- XY analyses
- ▼ Column analyses
  - t tests (and nonparametric tests)
  - One-way ANOVA (and nonparametric) 
  - One sample t and Wilcoxon test
  - Descriptive statistics
  - Normality and Lognormality Tests
  - Frequency distribution
  - ROC Curve
  - Bland-Altman method comparison
  - Identify outliers
  - Analyze a stack of P values
- Grouped analyses
- Contingency table analyses

Analyze which data sets?

- A:Nonrunners
- B:Recreational runners
- C:Elite runners

Select All

Deselect All

?

Cancel

OK

Parameters: One-Way ANOVA (and Nonparametric or Mixed)

Experimental Design

Repeated Measures

Multiple Comparisons

Options

Residuals

**Experimental design**

No matching or pairing

Each row represents matched, or repeated measures, data

	Group A	Group B	Group C	Group D
	Data Set-A	Data Set-B	Data Set-C	Title
1	Y	Y	Y	Y
2	Y	Y	Y	Y
3	Y	Y	Y	Y

**Assume Gaussian distribution?**

Yes. Use ANOVA.

No. Use nonparametric test.

**Assume equal SDs?**

Yes. Use ordinary ANOVA test.

No. Use Brown-Forsythe and Welch ANOVA tests.

**Based on your choices (on all tabs), Prism will perform:**

- Ordinary one-way ANOVA.

?

Cancel

OK

Parameters: One-Way ANOVA (and Nonparametric or Mixed)

Experimental Design

Repeated Measures

Multiple Comparisons

Options

Residuals

**Followup tests**

None.

Compare the mean of each column with the mean of every other column.

Compare the mean of each column with the mean of a control column.

Control column: Group A: Nonrunners

Compare the means of preselected pairs of columns.

Selected pairs: Select...

Test for linear trend between column mean and left-to-right column order.

**Which test?**

Use choices on the Options tab to choose the test, and to set the defaults for future ANOVAs.



Cancel

OK

## Parameters: One-Way ANOVA (and Nonparametric or Mixed)

Experimental Design   Repeated Measures   Multiple Comparisons   **Options**   Residuals

### Multiple comparisons test

Correct for multiple comparisons using statistical hypothesis testing. Recommended.

Test: Tukey (recommended)

Correct for multiple comparisons by controlling the False Discovery Rate.

Test: Two-stage step-up method of Benjamini, Krieger and Yekutieli (recommended)

Don't correct for multiple comparisons. Each comparison stands alone.

Test: Fisher's LSD test

### Multiple comparisons options

Swap direction of comparisons (A-B) vs. (B-A).

Report multiplicity adjusted P value for each comparison.

Each P value is adjusted to account for multiple comparisons.

Family-wise significance and confidence level: 0.05 (95% confidence interval)

### Graphing

Graph confidence intervals.

Graph ranks (nonparametric).

Graph differences (repeated measures).

### Additional results

Descriptive statistics for each data set.

Report comparison of models using AICc.

Report goodness of fit.

### Output

Show this many significant digits (for everything except P values): 4

P value style: GP: 0.1234 (ns), 0.0332 (\*), 0.0021 (\*\*), 0.0002 (\*\*\*), <0.0001 (\*\*... N= 6

Make options on this tab be the default for future One-Way ANOVAs.



Cancel

OK



one-way\_anova.pzfx — Edited

Q X

Restrict: Sheet is Any

▼ Data Tables >>

  Data 1

  New Data Table...

▼ Info >>

  Project info 1

  New Info...

▼ Results >>

  Ordinary one-way ANOVA of Data 1

  New Analysis...

▼ Graphs >>

  Data 1

  New Graph...

▼ Layouts >>

  New Layout...

Family >>

  Data 1

  Ordinary one-way ANOVA

ANOVA results X Multiple comparisons X | v |

Ordinary one-way ANOVA

ANOVA results

1 Table Analyzed Data 1

2 Data sets analyzed A-C

3

4 ANOVA summary

5 F 5.752

6 P value 0.0037

7 P value summary \*\*

8 Significant diff. among means (P < 0.05)? Yes

9 R squared 0.05388

10

11 Brown-Forsythe test

12 F (DFn, DFd)

13 P value

14 P value summary

15 Are SDs significantly different (P < 0.05)?

16

17 Bartlett's test

18 Bartlett's statistic (corrected) 5.667

19 P value 0.0588

20 P value summary ns

21 Are SDs significantly different (P < 0.05)? No

22

23 ANOVA table

	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	0.9268	2	0.4634	F (2, 202) = 5.752	P=0.0037
Residual (within columns)	16.27	202	0.08056		
Total	17.20	204			

24

25

26

27

28 Data summary

29 Number of treatments (columns) 3

30 Number of values (total) 205

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

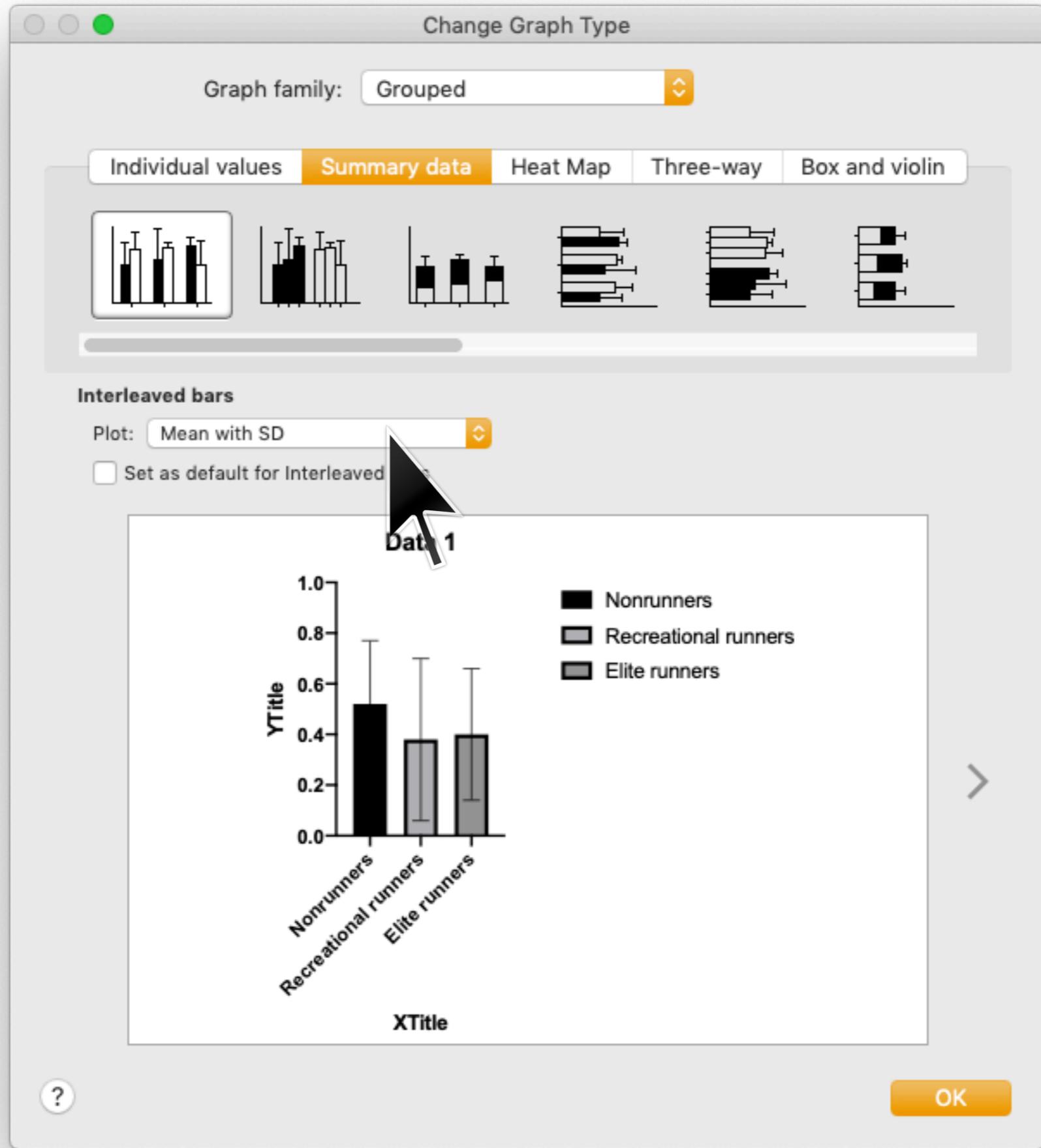
1000

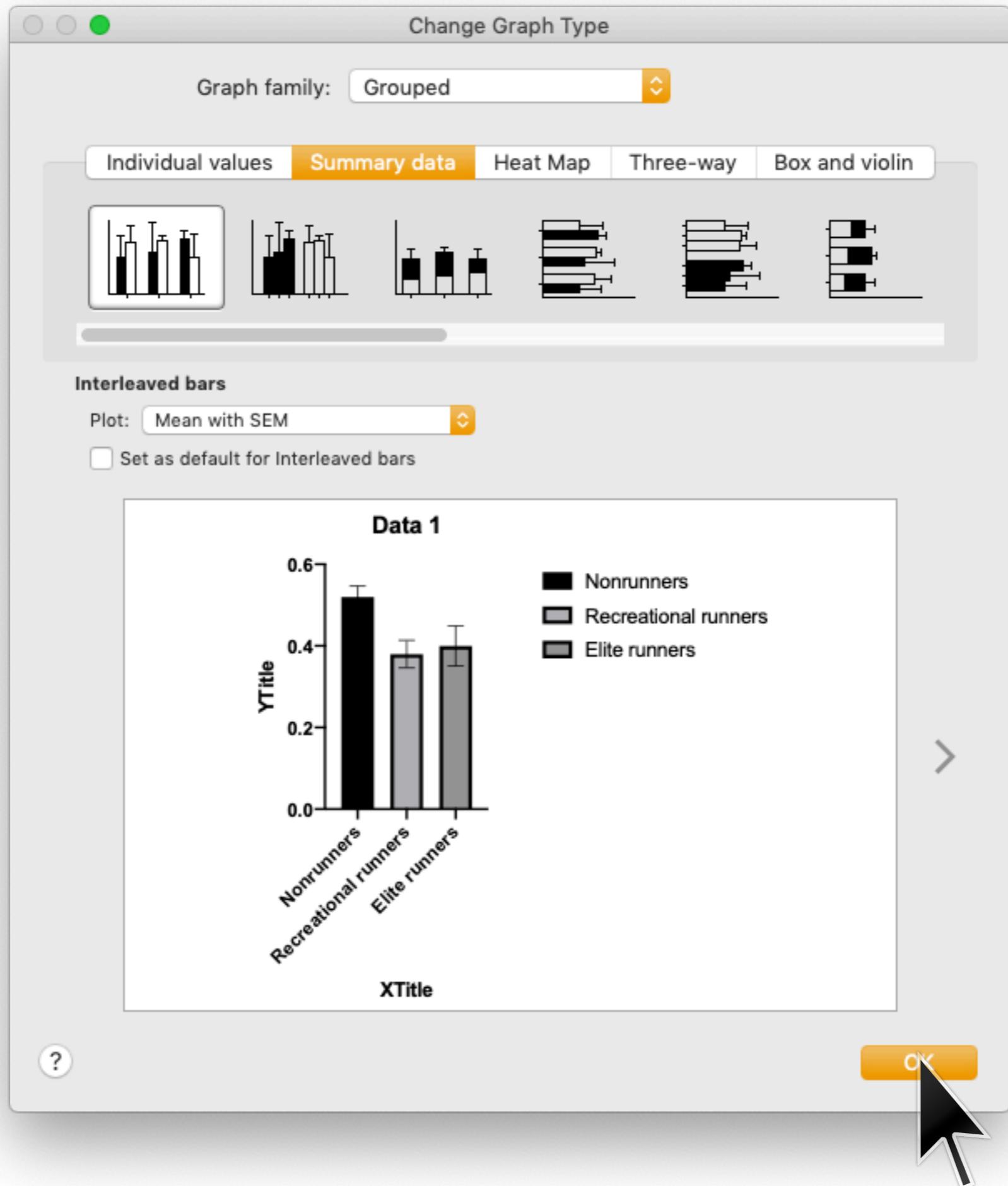
1001

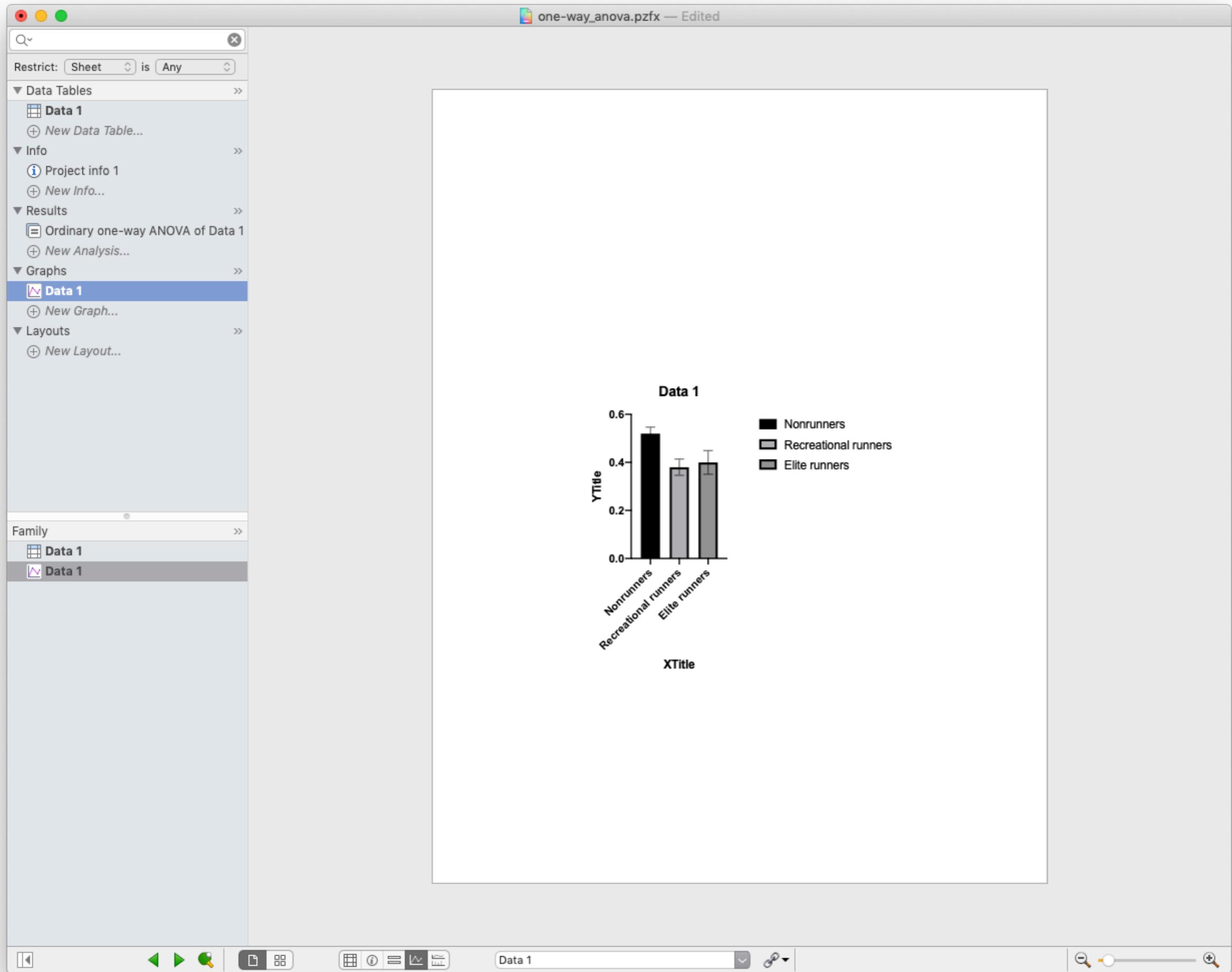
1002

1003

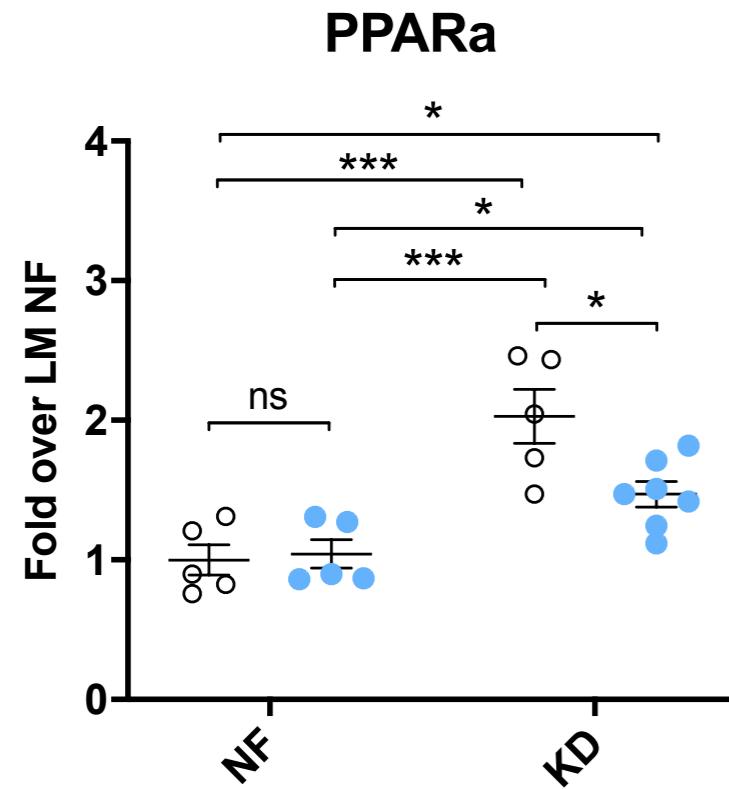








## Two-way ANOVA tests whether to see if there is an interaction between groups



$y_i$  = PPAR $\alpha$  mRNA expression

$x_{i1}$  = cancer presence (C26=tumor, LM=litter mate)

$x_{i2}$  = food (NF=normal, KD=ketogenic)

(data courtesy of Tobias Janowitz)

**Null model:**  $y_i = \beta_0 + \epsilon_i$

**Alternative model #1:**  $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$

**Alternative model #2:**  $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \epsilon_i$

interaction  
term

Welcome to GraphPad Prism

GraphPad Prism Version 8.4.3 (471)

**NEW TABLE & GRAPH**

- XY
- Column
- Grouped**
- Contingency
- Survival
- Parts of Whole
- Multiple variables
- Nested

**EXISTING FILE**

- Open a File
- LabArchives
- Clone a Graph
- Graph Portfolio

Grouped tables have two grouping variables, one defined by columns and the other defined by rows

Table format: **Grouped**

	A			B		
	Control		Treated			
	A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3
1	Male					
2	Female					

Figure: Bar chart showing grouped data for Control and Treated groups across Male and Female categories. The Y-axis represents three data points (Y1, Y2, Y3) for each group.

?

[Learn more](#)

**Data table:**

- Enter or import data into a new table
- Start with sample data to follow a tutorial

**Options:**

- Enter and plot a single Y value for each point
- Enter 7 replicate values in side-by-side subcolumns
- Enter and plot error values already calculated elsewhere

Enter: Mean, SD, N

Prism Tips

Cancel

Create



## Create New Analysis

### Data to analyze

Table: PPARa

### Type of analysis

Which analysis?

#### ▼ Transform, Normalize...

- Transform
- Transform concentrations (X)
- Normalize
- Prune rows
- Remove baseline and column math
- Transpose X and Y
- Fraction of Total

#### ► XY analyses

#### ► Column analyses

#### ▼ Grouped analyses

##### Two-way ANOVA (or mixed model)

- Three-way ANOVA (or mixed model)
- Row means with SD or SEM
- Multiple t tests - one per row

#### ► Contingency table analyses

#### ► Survival analyses

#### ► Parts of whole analyses

#### ► Multiple variable analyses

#### ► Nested analyses

#### ► Generate curve

#### ► Simulate data

Analyze which data sets?

A:LM

B:C26

Select All

Deselect All

?

Cancel

OK

## Parameters: Two-Way ANOVA (or Mixed Model)

RM Design RM Analysis Factor Names **Multiple Comparisons** Options Residuals

### What kind of comparison?

Compare cell means regardless of rows and columns

		Group A		Group B	
		Data Set-A		Data Set-B	
		A:Y1	A:Y2	B:Y1	B:Y2
1		Mean		Mean	
2		Mean		Mean	



### How many comparisons?

- Compare each cell mean with every other cell mean.
- Compare each cell mean with the control (upper-left) cell mean.

Control cell: LM : NF



### How many families?

One family for all the comparisons



### Which test?

Use choices on the Options tab to choose the test, and to set the defaults for future ANOVAs.



Cancel

OK

## Parameters: Two-Way ANOVA (or Mixed Model)

RM Design RM Analysis Factor Names Multiple Comparisons Options Residuals

### Multiple comparisons test

Correct for multiple comparisons using statistical hypothesis testing. Recommended.

Test: Holm-Sidak (more power, but can't compute confidence intervals) 

Correct for multiple comparisons by controlling the False Discovery Rate.

Test: Two-stage step-up method of Benjamini, Krieger and Yekutieli (recommended) 

Don't correct for multiple comparisons. Each comparison stands alone.

Test: Fisher's LSD test

### Multiple comparisons options

Swap direction of comparisons (A-B) vs. (B-A).

Report multiplicity adjusted P value for each comparison.

Each P value is adjusted to account for multiple comparisons.

Family-wise significance and confidence level: 0.05 

### Graphing options

Graph confidence intervals.

### Additional results

Narrative results.

Show cell/row/column/grand predicted (LS) means.

Report goodness of fit.

### Output

Show this many significant digits (for everything except P values): 4 

P value style: GP: 0.1234 (ns), 0.0332 (\*), 0.0021 (\*\*), 0.0002 (\*\*\*), <0.0001 (\*\*\*\*)  N= 6 

Make options on this tab be the default for future Two-Way ANOVAs.



Cancel

OK

two-way\_anova.pzfx — Edited

ANOVA results    Multiple comparisons

2way ANOVA

ANOVA results

1 Table Analyzed PPARa

2

3 Two-way ANOVA Ordinary

4 Alpha 0.05

5

6 Source of Variation % of total variation P value P value summary Significant?

7 Interaction 9.695 0.0291 \* Yes

8 Row Factor 57.11 <0.0001 \*\*\*\* Yes

9 Column Factor 7.185 0.0561 ns No

10

11 ANOVA table SS (Type III) DF MS F (DFn, DFd) P value

12 Interaction 0.4856 1 0.4856 F (1, 18) = 5.623 P=0.0291

13 Row Factor 2.860 1 2.860 F (1, 18) = 33.12 P<0.0001

14 Column Factor 0.3599 1 0.3599 F (1, 18) = 4.167 P=0.0561

15 Residual 1.554 18 0.08636

16

17 Difference between column means

18 Predicted (LS) mean of LM 1.515

19 Predicted (LS) mean of C26 1.256

20 Difference between predicted means 0.2585

21 SE of difference 0.1266

22 95% CI of difference -0.007533 to 0.5246

23

24 Difference between row means

25 Predicted (LS) mean of NF 1.021

26 Predicted (LS) mean of KD 1.750

27 Difference between predicted means -0.7288

28 SE of difference 0.1266

29 95% CI of difference -0.9949 to -0.4628

30

2way ANOVA of PPARa

Row 1, Column A

two-way\_anova.pzfx — Edited

Search

Data Tables

- PPARa
- + New Data Table...

Info

- + New Info...

Results

- 2way ANOVA of PPARa**
- + New Analysis...

Graphs

- PPARa
- + New Graph...

Layouts

- + New Layout...

PPARa

2way ANOVA

Multiple comparisons

1 Compare cell means regardless of rows and columns

2

3 Number of families 1

4 Number of comparisons per family 6

5 Alpha 0.05

6

7 Holm-Sidak's multiple comparisons test

	Predicted (LS) mean diff.	Significant?	Summary	Adjusted P Value
8				
9 NF:LM vs. NF:C26	-0.04178	No	ns	0.8247
10 NF:LM vs. KD:LM	-1.029	Yes	***	0.0002
11 NF:LM vs. KD:C26	-0.4703	Yes	*	0.0404
12 NF:C26 vs. KD:LM	-0.9874	Yes	***	0.0002
13 NF:C26 vs. KD:C26	-0.4285	Yes	*	0.0450
14 KD:LM vs. KD:C26	0.5588	Yes	*	0.0178

Family

PPARa

2way ANOVA

15

16

17 Test details

	Predicted (LS) mean 1	Predicted (LS) mean 2	Predicted (LS) mean diff.	SE of diff.	N1	N2	t	DF
18								
19 NF:LM vs. NF:C26	1.000	1.042	-0.04178	0.1859	5	5	0.2248	18.00
20 NF:LM vs. KD:LM	1.000	2.029	-1.029	0.1859	5	5	5.537	18.00
21 NF:LM vs. KD:C26	1.000	1.470	-0.4703	0.1721	5	7	2.733	18.00
22 NF:C26 vs. KD:LM	1.042	2.029	-0.9874	0.1859	5	5	5.313	18.00
23 NF:C26 vs. KD:C26	1.042	1.470	-0.4285	0.1721	5	7	2.490	18.00
24 KD:LM vs. KD:C26	2.029	1.470	0.5588	0.1721	5	7	3.248	18.00
25								
26								
27								
28								
29								

Row 1, Column A

Q Search

▼ Data Tables

- PPARa
- + New Data Table...

▼ Info

- + New Info...

▼ Results

- 2way ANOVA of PPARa
- + New Analysis...

▼ Graphs

- PPARa
- + New Graph...

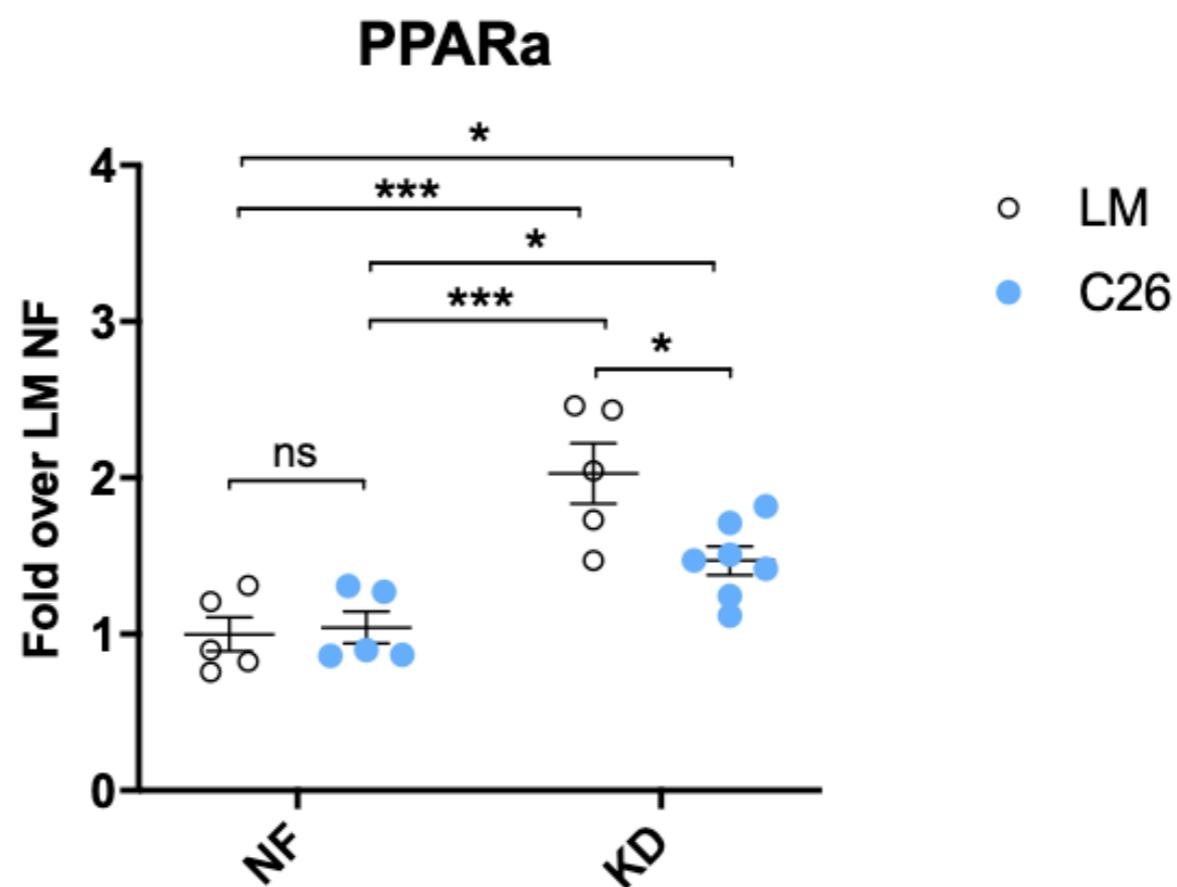
▼ Layouts

- + New Layout...

Family

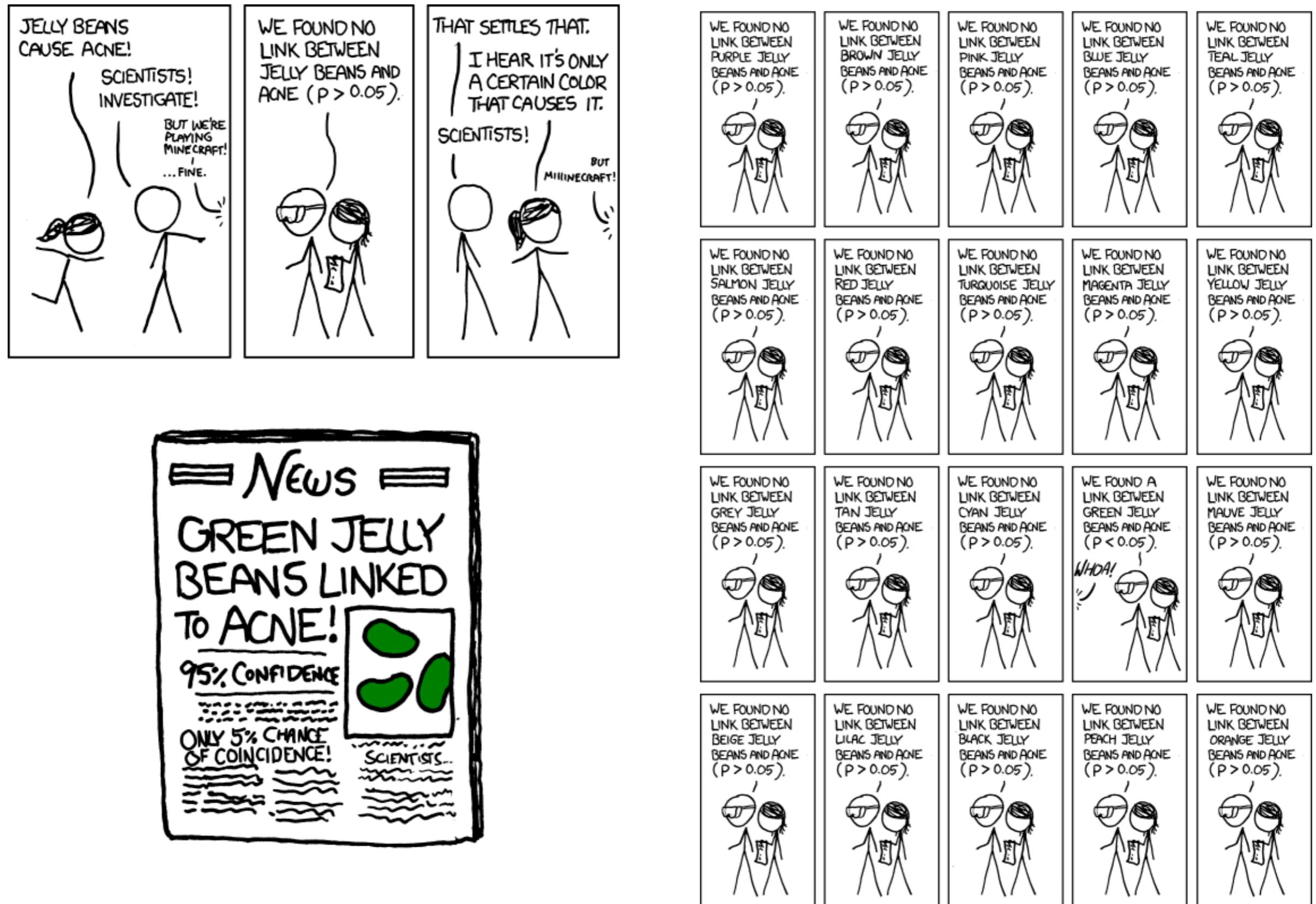
PPARa

PPARa



## Multiple hypothesis testing

## The problem of multiple subgroups



## **The family-wise error rate increases rapidly with the number of tests performed**

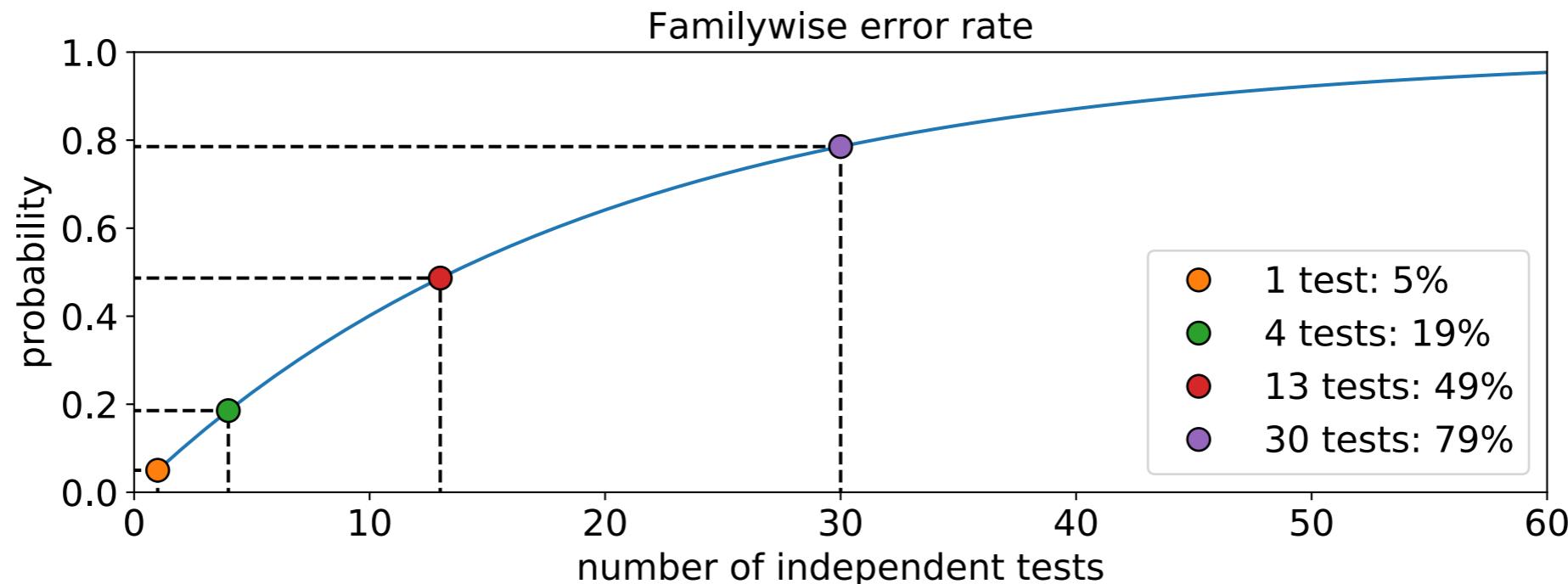
### **Scenario:**

we perform null hypothesis tests on  $K$  independent datasets, for each of which the null hypothesis is true.

### **Family-wise error rate:**

the probability of at least one false positive (FP)

$$p(\text{FP} \geq 1 \mid \text{null hypothesis}) = 1 - \text{confidence}^K$$



## Summary of multiple hypothesis correction techniques

---

Approach	What you control	Expression
No correction	$\alpha$ : if all null hypotheses are true, the <u>fraction of tests</u> that produce a significant result	$\alpha = \frac{\text{FP}}{\text{FP} + \text{TN}}$
Bonferroni / Dunn-Sidak	$\alpha$ : if all null hypotheses are true, the <u>chance of obtaining one or more</u> significant results	$\alpha = p(\#\text{FP} > 0)$
False discovery rate (FDR)	$Q$ : the fraction of all discoveries for which the null hypothesis is actually true	$Q = \frac{\text{FP}}{\text{FP} + \text{TP}}$

## Simple ways to counteract the multiple hypothesis problem

---

**Bonferroni correction:**

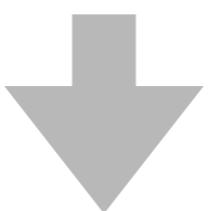
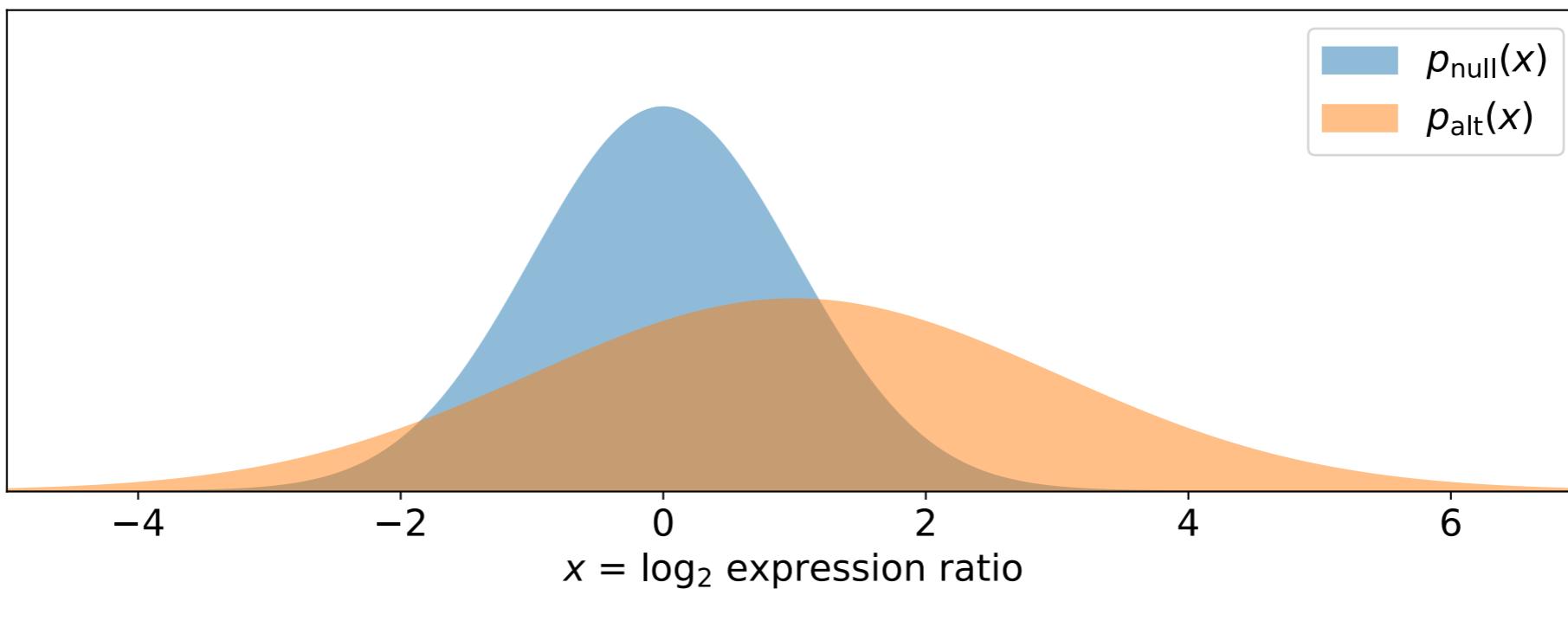
$$\alpha_{\text{Bonferroni}} = \frac{\alpha}{K}$$

**Dunn-Sidak correction:**

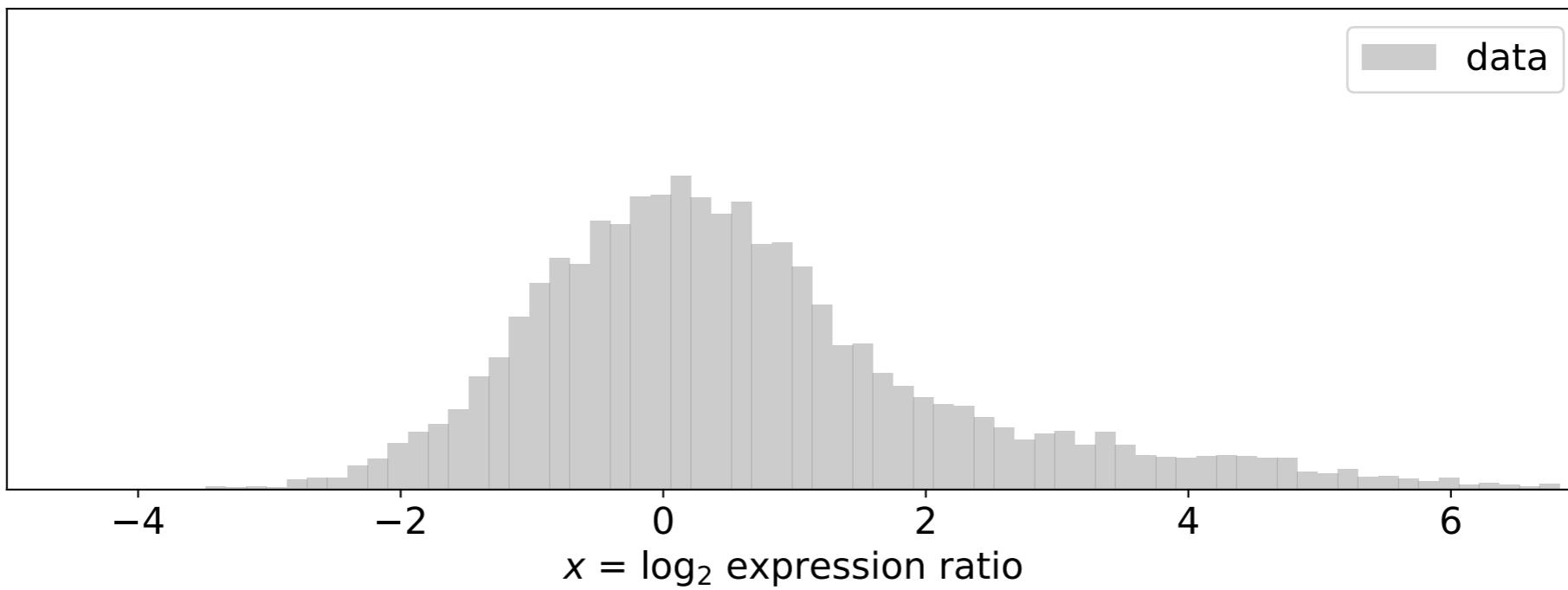
$$\alpha_{DS} = 1 - (1 - \alpha)^{1/K}$$

**Dunn-Sidak is the exact solution; Bonferroni is an approximation**

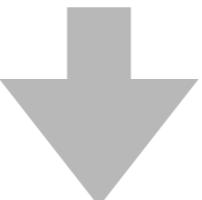
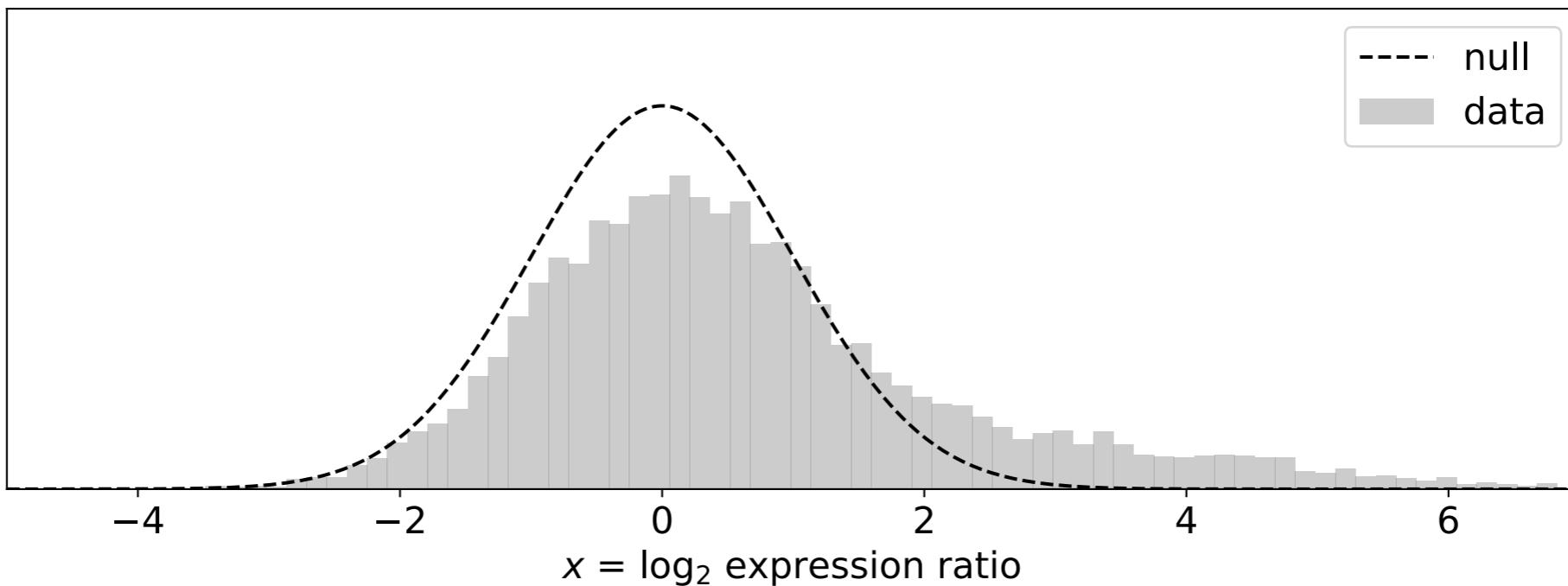
## Example: differential expression (simulation)



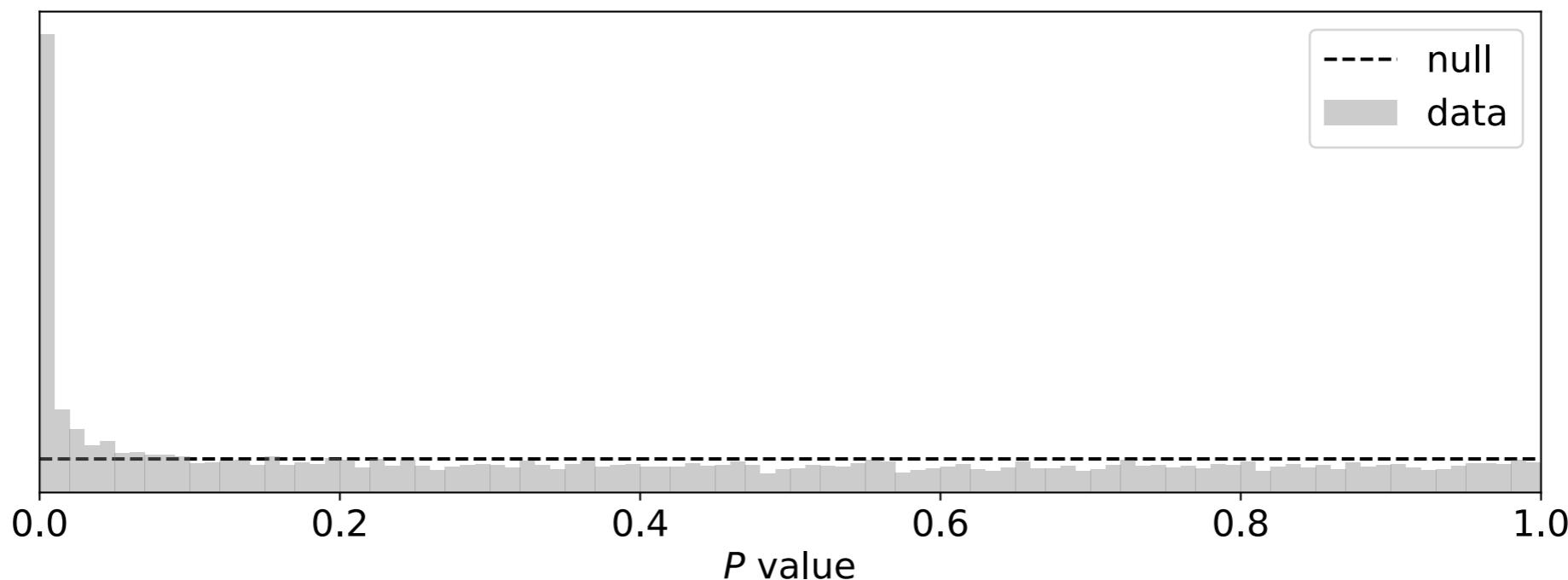
7,000  $x$ s from  $p_{\text{null}}(x)$   
+ 3,000  $x$ s from  $p_{\text{alt}}(x)$



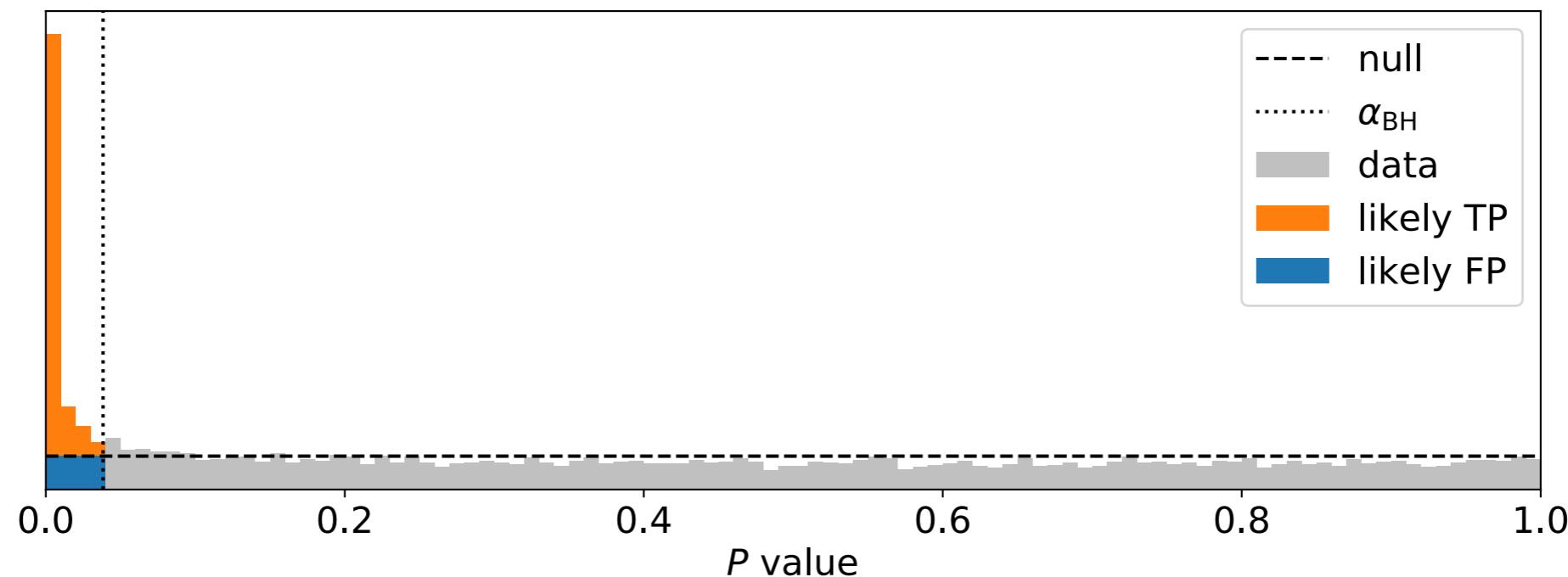
## First, convert data to p-values



use knowledge of  $p_{\text{null}}(x)$  to  
compute a p-value for each datapoint



## Benjamini–Hochberg procedure



Choose  $\alpha_{BH}$  such to match the target False Discovery Rate (10% here):

$$FDR = Q = \frac{FP}{TP + FP} = \frac{\text{blue square}}{\text{orange square} + \text{blue square}}$$

Declare all P-values below  $\alpha_{BH}$  as “discoveries”.

## Multiple comparisons are ubiquitous and insidious

---

“Most scientists are oblivious to the problems of multiplicities. Yet they are everywhere. In one or more of its forms, multiplicities are present in every statistical application. They may be out in the open or hidden. And even if they are out in the open, recognizing them is but the first step in a difficult process of inference. Problems of multiplicities are the most difficult that we statisticians face. They threaten the validity of every statistical conclusion.”

## Multiple comparisons arise in many many contexts

---

### **multiple subgroups:**

You perform tests on multiple subgroups of your data.

### **multiple ways to dichotomize:**

You do pairwise comparisons between different combinations of subgroups.

### **multiple sample sizes:**

You keep collecting data until you find  $P < 0.05$ .

**DO NOT DO THIS.**

### **multiple ways to preprocess the data:**

You analyze data preprocessed in multiple different ways.

### **multiple statistical tests:**

You use different statistical tests on the same data before finding  $P < 0.05$ .

## Multiple comparisons arise in many, many contexts

---

### **multiple ways to select relevant variables:**

You try to model your data using different subsets of possible variables.

### **multiple ways to analyze your data (“garden of forking paths”):**

You try lots of qualitatively different analysis strategies.

### **outcome switching:**

You change the quantity you care about after you've looked at the data.

### **multiple geographic areas:**

E.g., you investigate a “cancer cluster” you hear about in the news.

## Correcting for multiple comparisons is not always needed

---

### **Scenario 1:**

If readers can be reasonably expected to account for multiple comparisons on their own.

### **Scenario 2:**

Before looking at the data, you have clearly defined one outcome as primary and others as secondary.

### **Scenario 3:**

You make only a few planned comparisons and your P-values are not marginal.

### **Scenario 4:**

A large fraction the tests you perform are significant.

## Practical advice of avoiding multiple hypothesis pitfalls

---

**Raise your standards: use  $\alpha = 0.01$ , not  $\alpha = 0.05$ .**

**Separate exploratory data analysis from confirmatory data analysis.**

**Distinguish critical p-values from ancillary p-values.**

**Don't spend too much time analyzing a small dataset.**

**When generating small expensive datasets (e.g. mice), blind your experiments as best you can, and plan your analysis ahead of time**

**When in doubt, double-check your hypothesis with new data**

**Don't worry about informal multiple hypothesis testing when  $P < 10^{-4}$ .**

**10:00a - 12:02p Just slightly over. Good length I think.**