Chapter 21: Ordinal and non-normally distributed data: Transformations and non-parametric tests

TXCL7565/PHSC7565

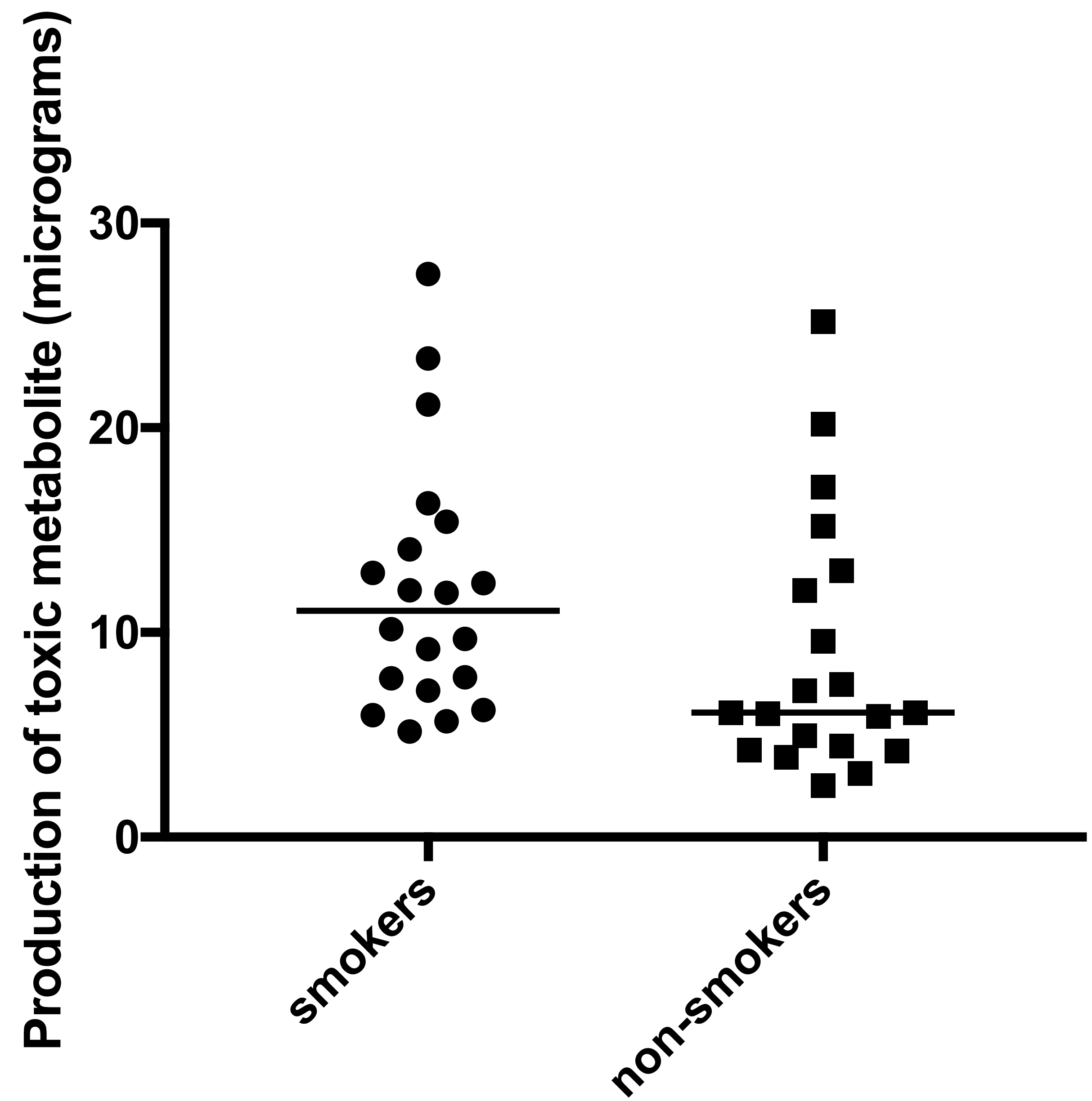
What This Lecture Covers

- Transforming data to a normal distribution
- Mann-Whitney test (substitute for the 2-sample t-test)
- Dealing with ordinal data
- Wilcoxon paired samples test (substitute for paired t-test)
- Kruskal-Wallis test (substitute for one-way ANOVA)
- Spearman correlation (substitute for Pearson correlation)

TRANSFORMING DATA TO A NORMAL DISTRIBUTION

Toxic metabolite in smokers and non-smokers

- In a small minority of users, an analgesic produces a serious side effect, inflammation of the liver due to a toxic metabolite.
- Heavy smokers are more susceptible to the side effect, therefore we hypothesize that they have a higher production of the metabolite.



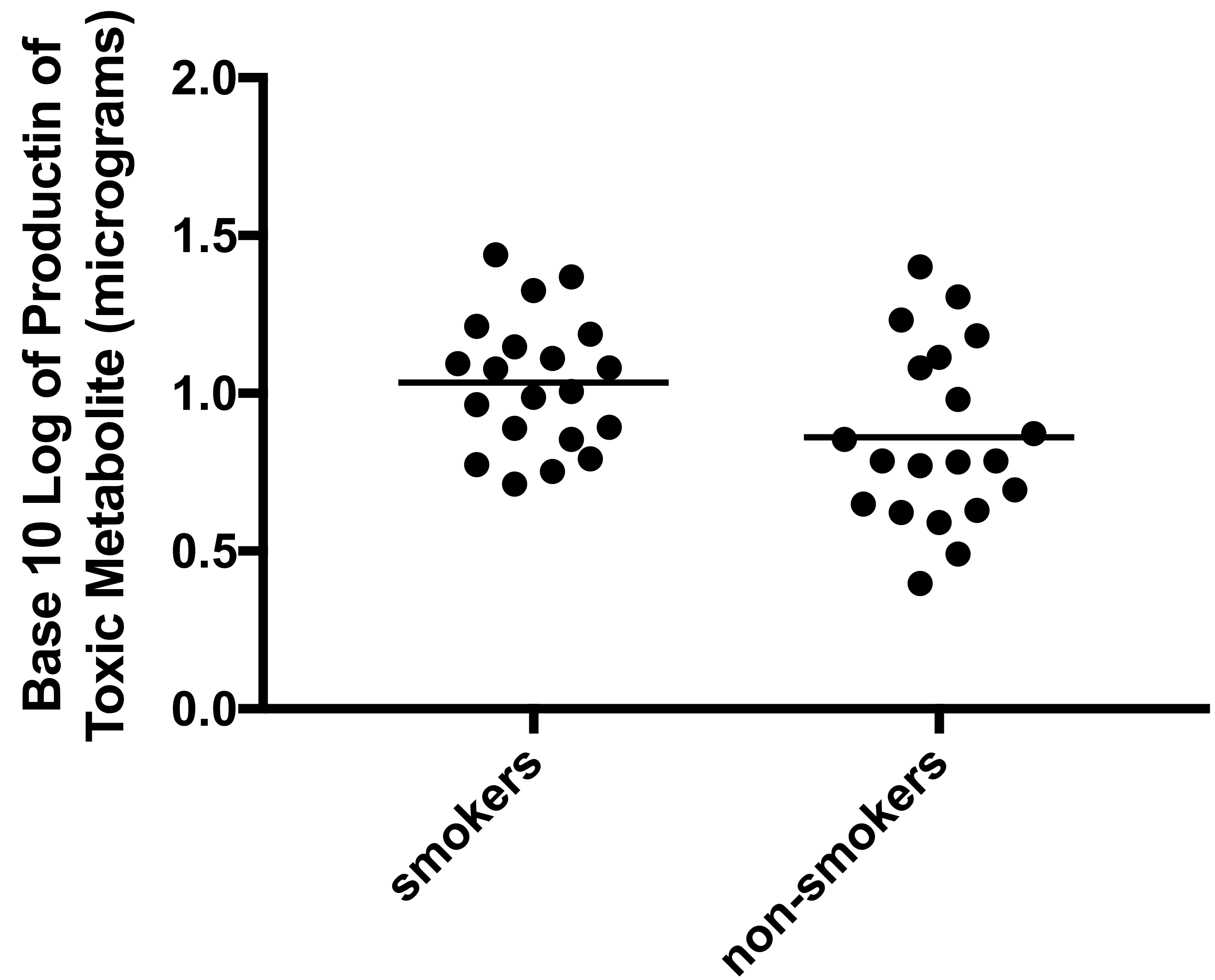
Toxic metabolite example

- Both groups show a positive skew in values, i.e., they do not have a normal distribution.
- If we ignore this and do a two sample t-test…

Unpaired t test	
P value	0.1149
P value summary	ns
Significantly different (P < 0.05)?	No
One- or two-tailed P value?	Two-tailed
t, df	t=1.614 df=38
How big is the difference?	
Mean ± SEM of column A	12.09 ± 1.379, n=20
Mean ± SEM of column B	8.92 ± 1.399, n=20
Difference between means	-3.17 ± 1.964
95% confidence interval	-7.147 to 0.8068
R squared (eta squared)	0.06413

Toxic metabolite example - Log transformation

- With these data, we can consider a log transformation because: 1) the data are positively skewed and 2) all values are greater than zero.



Toxic metabolite example
t-test on transformed data

Unpaired t test	
P value	0.0337
P value summary	*
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.203 df=38
How big is the difference?	
Mean ± SEM of column A	1.033 ± 0.04684, n=20
Mean ± SEM of column B	0.8611 ± 0.06264, n=20
Difference between means	0.1723 ± 0.07821
95% confidence interval	0.01397 to 0.3306
R squared (eta squared)	0.1132

- It is a normal and legitimate practice to use transformations to convert data to a better approximation of a normal distribution and then carry out tests on the transformed data.
- It is quite common to suffer a large loss of power if highly skewed or otherwise non-normal data is analyzed by methods that assume normality.

Toxic metabolite example
effect size

Unpaired t test	
P value	0.0337
P value summary	*
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.203 df=38
How big is the difference?	
Mean ± SEM of column A	1.033 ± 0.04684, n=20
Mean ± SEM of column B	0.8611 ± 0.06264, n=20
Difference between means	0.1723 ± 0.07821
95% confidence interval	0.01397 to 0.3306
R squared (eta squared)	0.1132

- NOTE: the difference and 95% CI reported are based on the transformed data.
- The difference and 95% CI can be converted back to the original scale by taking the antilog, but…
  - The new effect size and 95% CI no longer represent a difference and now represent the ratio of toxin production in the two groups.

Toxic metabolite example

effect size

Property of Logs

$$log(X) - log(Y) = log\left(\frac{X}{Y}\right)$$

$$log(X) + log(Y) = log\left(XY\right)$$

Unpaired t test	
P value	0.0337
P value summary	*
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.203 df=38
How big is the difference?	
Mean ± SEM of column A	1.033 ± 0.04684, n=20
Mean ± SEM of column B	0.8611 ± 0.06264, n=20
Difference between means	0.1723 ± 0.07821
95% confidence interval	0.01397 to 0.3306
R squared (eta squared)	0.1132

MANN-WHITNEY TEST

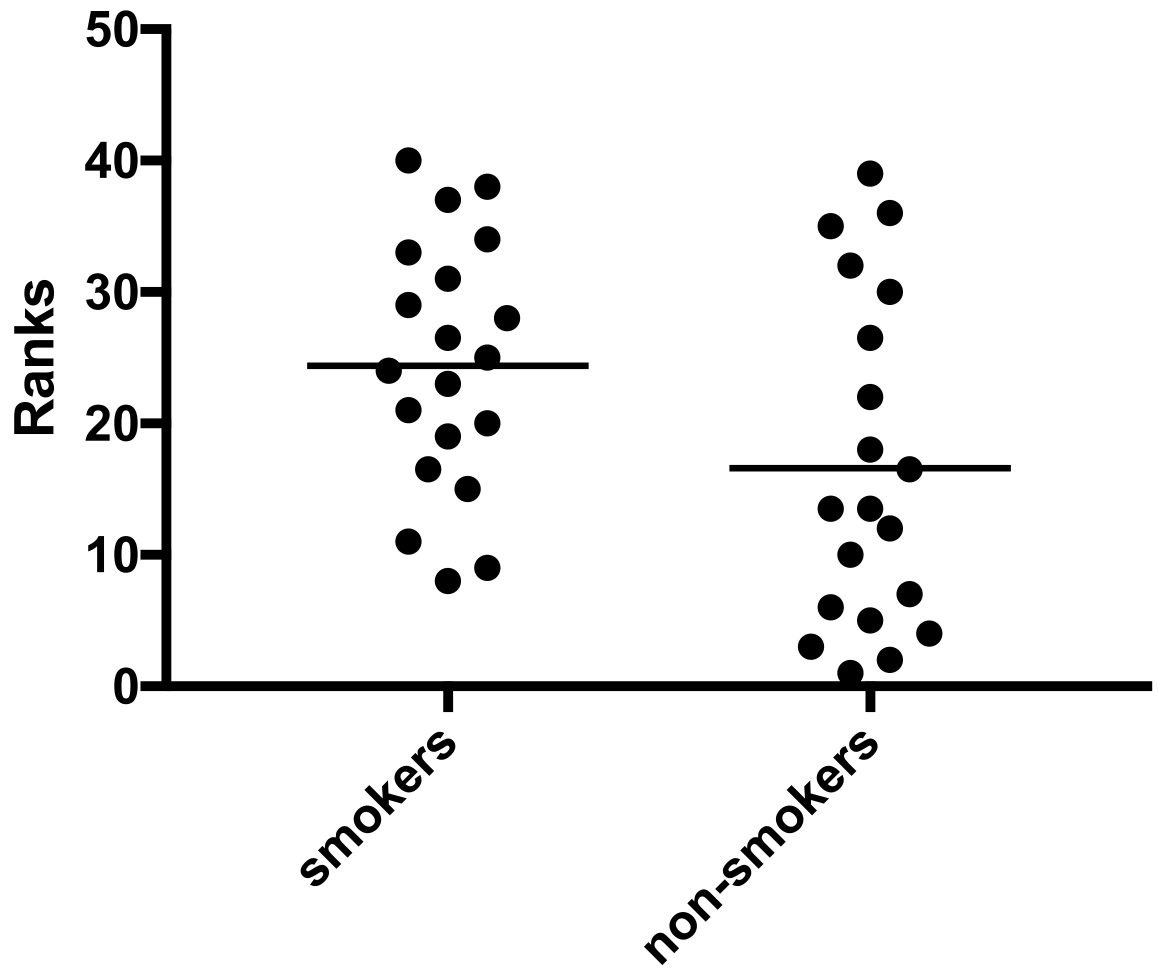
Substitute for two-sample *t*-test

- Instead of transforming the data to normality, we could employ a '*non-parametric test*'.

- Non-parametric tests partially duplicate the functionality of tests we have already studied, but use a method of calculation that does not depend upon a normal distribution.

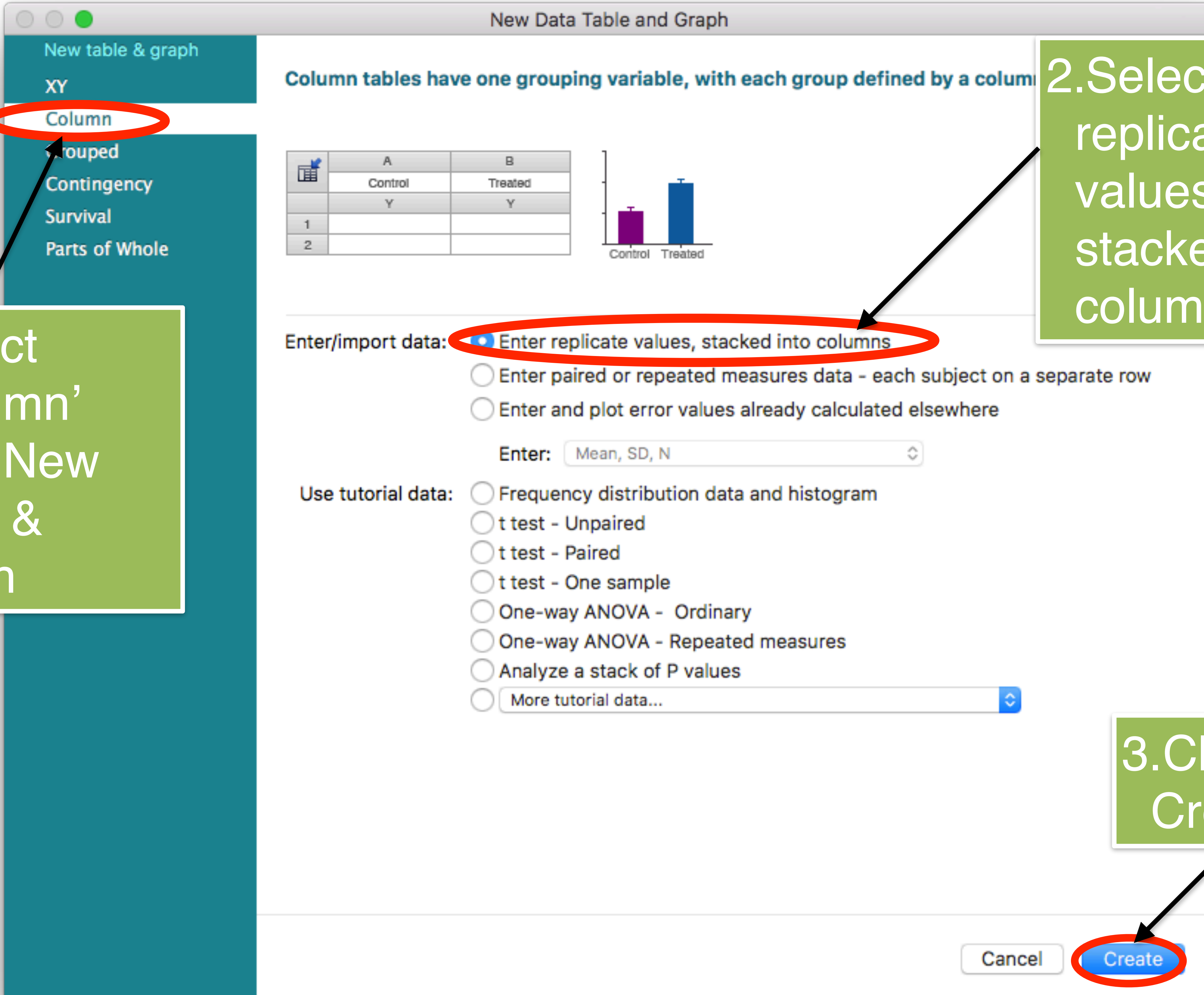
Non-parametric tests are based on ranks

In non-parametric tests the data are transformed into rank values and then all further calculations are based solely upon these rankings.



Group A	Group B
smokers	non-smokers
Y	Y
7.75	2.50
7.80	7.45
23.40	4.95
27.50	3.10
5.65	17.10
12.05	4.20
11.95	6.05
21.15	25.20
15.40	20.20
14.05	6.10
10.15	9.55
12.40	7.15
6.20	3.90
16.30	13.00
7.15	12.05
5.95	4.25
9.20	15.20
5.15	4.45
9.70	5.90
12.90	6.10

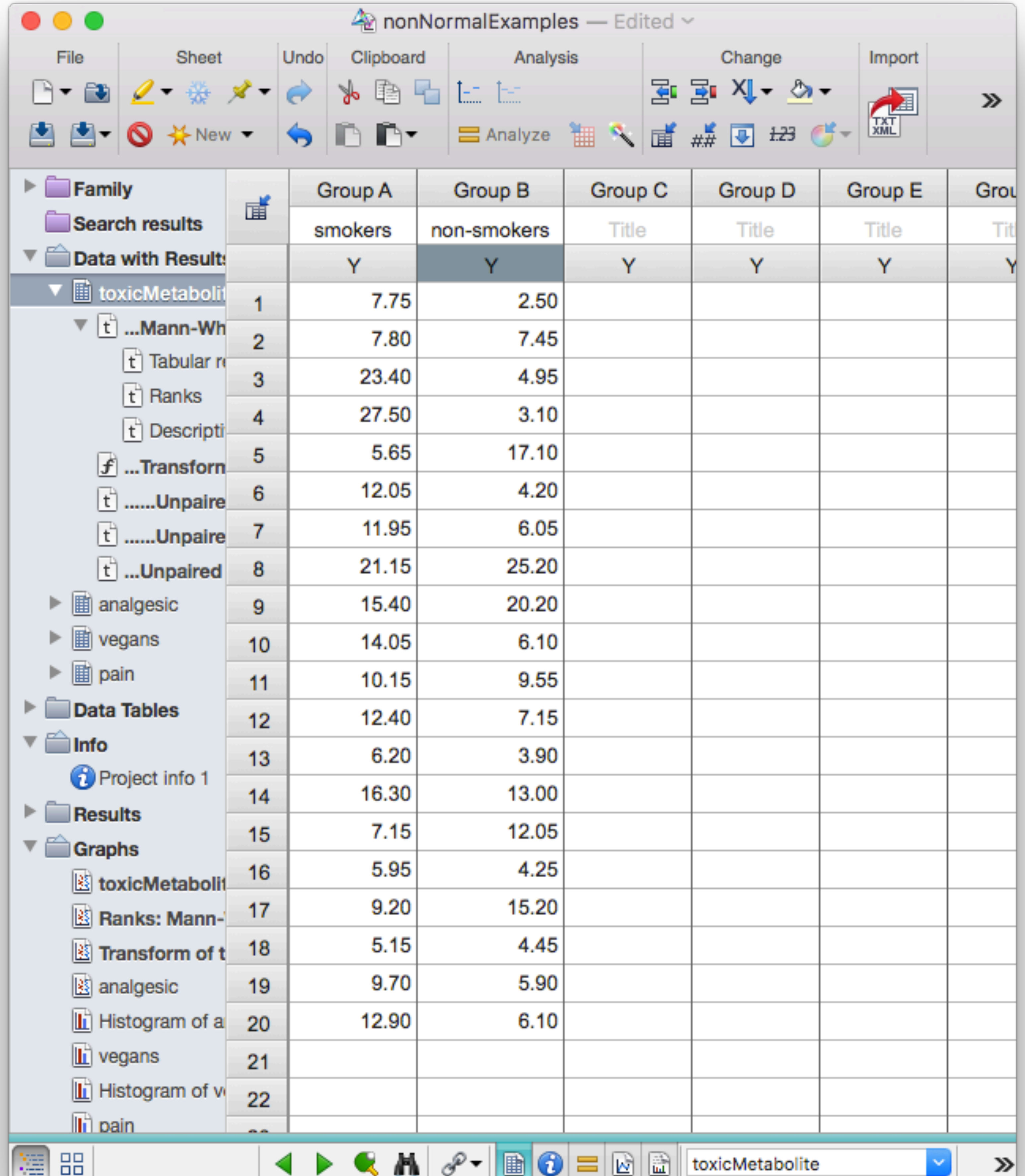
A	B
smokers	non-smokers
Y	Y
19.000	1.000
20.000	18.000
38.000	7.000
40.000	2.000
9.000	35.000
26.500	4.000
25.000	12.000
37.000	39.000
33.000	36.000
31.000	13.500
24.000	22.000
28.000	16.500
15.000	3.000
34.000	30.000
16.500	26.500
11.000	5.000
21.000	32.000
8.000	6.000
23.000	10.000
29.000	13.500



1. Select ‘Column’ from New table & graph

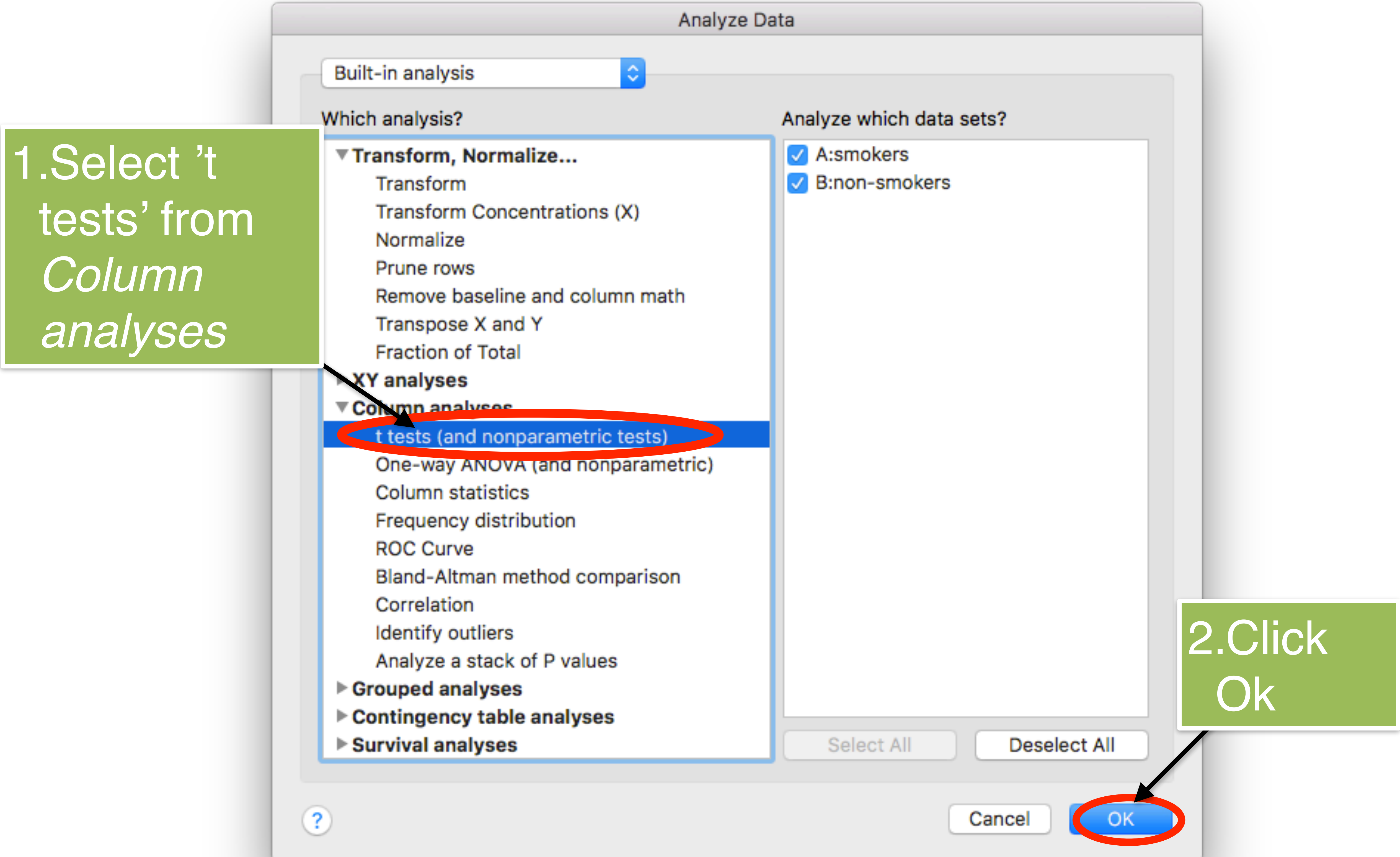
2. Select ‘Enter replicate values, stacked into columns’

3. Click Create



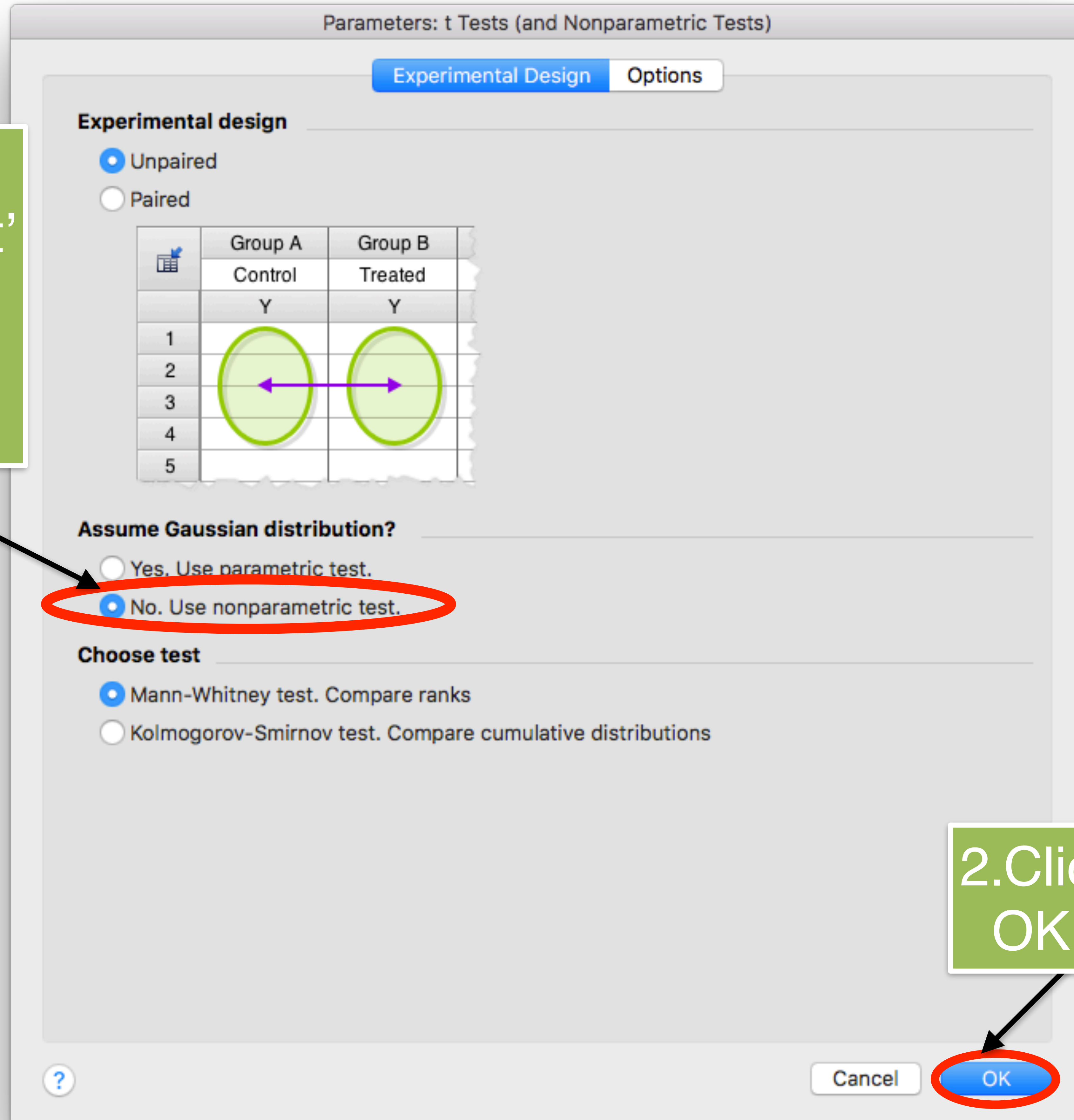
	Group A	Group B	Group C	Group D	Group E
	smokers	non-smokers	Title	Title	Title
	Y	Y	Y	Y	Y
1	7.75	2.50			
2	7.80	7.45			
3	23.40	4.95			
4	27.50	3.10			
5	5.65	17.10			
6	12.05	4.20			
7	11.95	6.05			
8	21.15	25.20			
9	15.40	20.20			
10	14.05	6.10			
11	10.15	9.55			
12	12.40	7.15			
13	6.20	3.90			
14	16.30	13.00			
15	7.15	12.05			
16	5.95	4.25			
17	9.20	15.20			
18	5.15	4.45			
19	9.70	5.90			
20	12.90	6.10			

Enter the actual values. GraphPad will take care of calculating the ranks.



1.Select 't tests' from *Column analyses*
Analyze Data
Built-in 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)
Column statistics
Frequency distribution
ROC Curve
Bland-Altman method comparison
Correlation
Identify outliers
Analyze a stack of P values
Grouped analyses
Contingency table analyses
Survival analyses
Analyze which data sets?
A:smokers
B:non-smokers
Select All
Deselect All
Cancel
OK
2.Click Ok

1. Select 'No. Use nonparametric test' from *Assume Gaussian distribution*



2. Click OK

Mann-Whitney test
Tabular results

1	Table Analyzed	toxicMetabolite
2		
3	Column A	smokers
4	vs.	vs.
5	Column B	non-smokers
6		
7	Mann Whitney test	
8	P value	0.0344
9	Exact or approximate P value?	Exact
10	P value summary	*
11	Significantly different (P < 0.05)?	Yes
12	One- or two-tailed P value?	Two-tailed
13	Sum of ranks in column A,B	488 , 332
14	Mann-Whitney U	122
15		
16	Difference between medians	
17	Median of column A	11.05, n=20
18	Median of column B	6.1, n=20
19	Difference: Actual	4.95
20	Difference: Hodges-Lehmann	3.225
21		

Interpreting a significant Mann-Whitney test

- ‘**Values are generally higher in this group than in that**’: This makes no assumptions about how the data are distributed. Minimum claim - little risk.

- ‘**The median is greater in this group than in that**’: Only assumption is that the data are not distributed in a totally bizarre manner. Generally OK, but check with an expert if the data sets have extreme distributions

- ‘**The mean is greater in this group than in that**’: Rarely justifiable.

Parametric vs. Non-Parametric

Disadvantages of non-parametric

- When non-parametric methods are applied to data that are normally distributed, they are *slightly* less powerful.
- Do not produce a meaningful 95% confidence interval for the size of the difference in outcome.

Advantages of non-parametric

- When data are severely non-normal, we can loose a huge amount of power by using a parametric test.

Dealing with non-normally distributed data

- **First choice**: Convert to normal distribution by transformation and use parametric test.

- **Second choice**: Resort to a non-parametric test.

DEALING WITH ORDINAL DATA

Why ordinal data are generally analyzed using non-parametric methods

It is theoretically possible for ordinal scale to approximate a normal distribution, but non-normality is all too common.

- ***Limited range of possible values***. Smooth bell-shape distribution is not possible.

- ***Bizarre distribution***. People tend to do have strange habits about using extreme values in scales.

- ***Small potential gain from using parametric methods***. The benefit of using a parametric method even if the data are reasonably normal is small.

- ***Large potential loss of power from using parametric methods***. If a parametric method is used with data that are badly non-normal there can be a drastic loss of power.

Dealing with ordinal scale data

Unless there is specific evidence that the data are likely to behave unusually well, just accept that non-parametric methods will have to be used. Power loss will, at worst, be very slight.

Herbal analgesic example

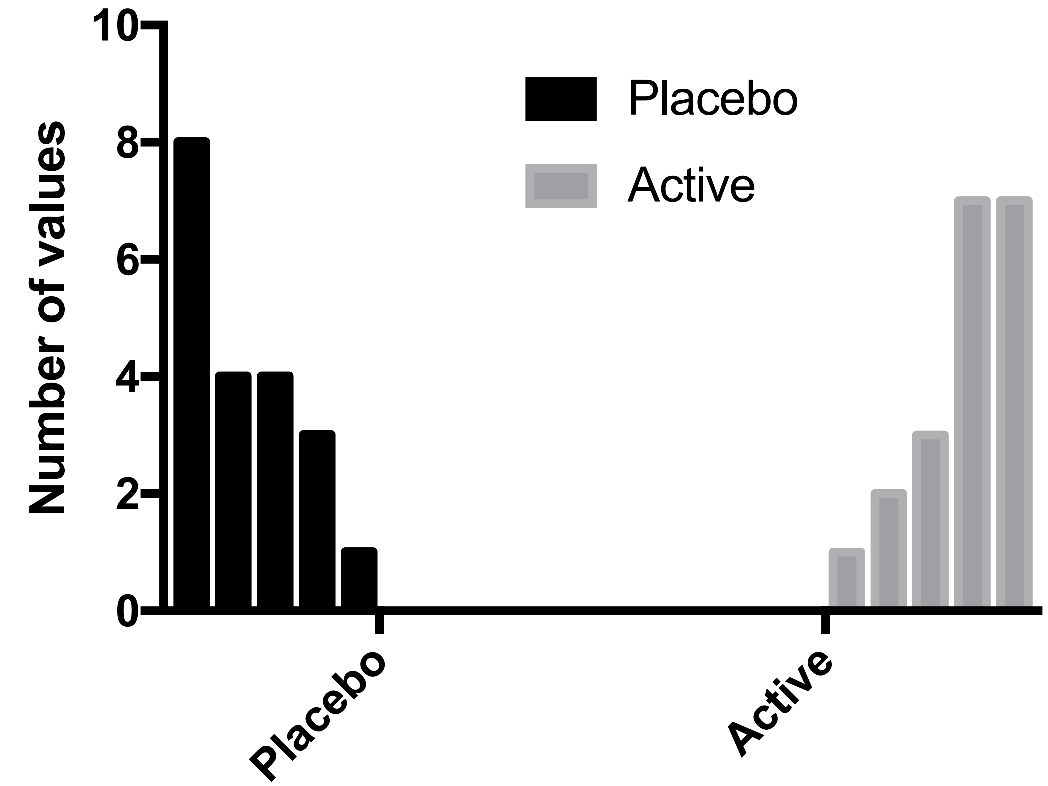
Group A	Group B
Placebo	Active
Y	Y
3	1
0	3
2	3
4	3
0	4
0	4
2	3
0	0
0	4
1	3
1	4
0	4
0	3
0	1
3	2
2	4
1	4
2	3
1	2
3	2

Two teams of patients rate the effectiveness of either an active herbal analgesic or a placebo for the treatment of mild pain.

The scale used to report effectiveness is:
4 = Completely/almost completely effective
3 = Strongly effective
2 = Moderately effective
1 = Slightly effective
0 = No/almost no effect

Herbal analgesic example

Mann-Whitney test		
1	Table Analyzed	analgesic
2		
3	Column B	Active
4	vs.	vs.
5	Column A	Placebo
6		
7	Mann Whitney test	
8	P value	0.0004
9	Exact or approximate P value?	Exact
10	P value summary	***
11	Significantly different (P < 0.05)?	Yes
12	One- or two-tailed P value?	Two-tailed
13	Sum of ranks in column A,B	285 , 535
14	Mann-Whitney U	75



Prism computed an exact P value (0.0004), which takes into account ties among values. Note that most other programs do not compute exact P values when there are tied values, but would instead report an approximate P value (0.0006).

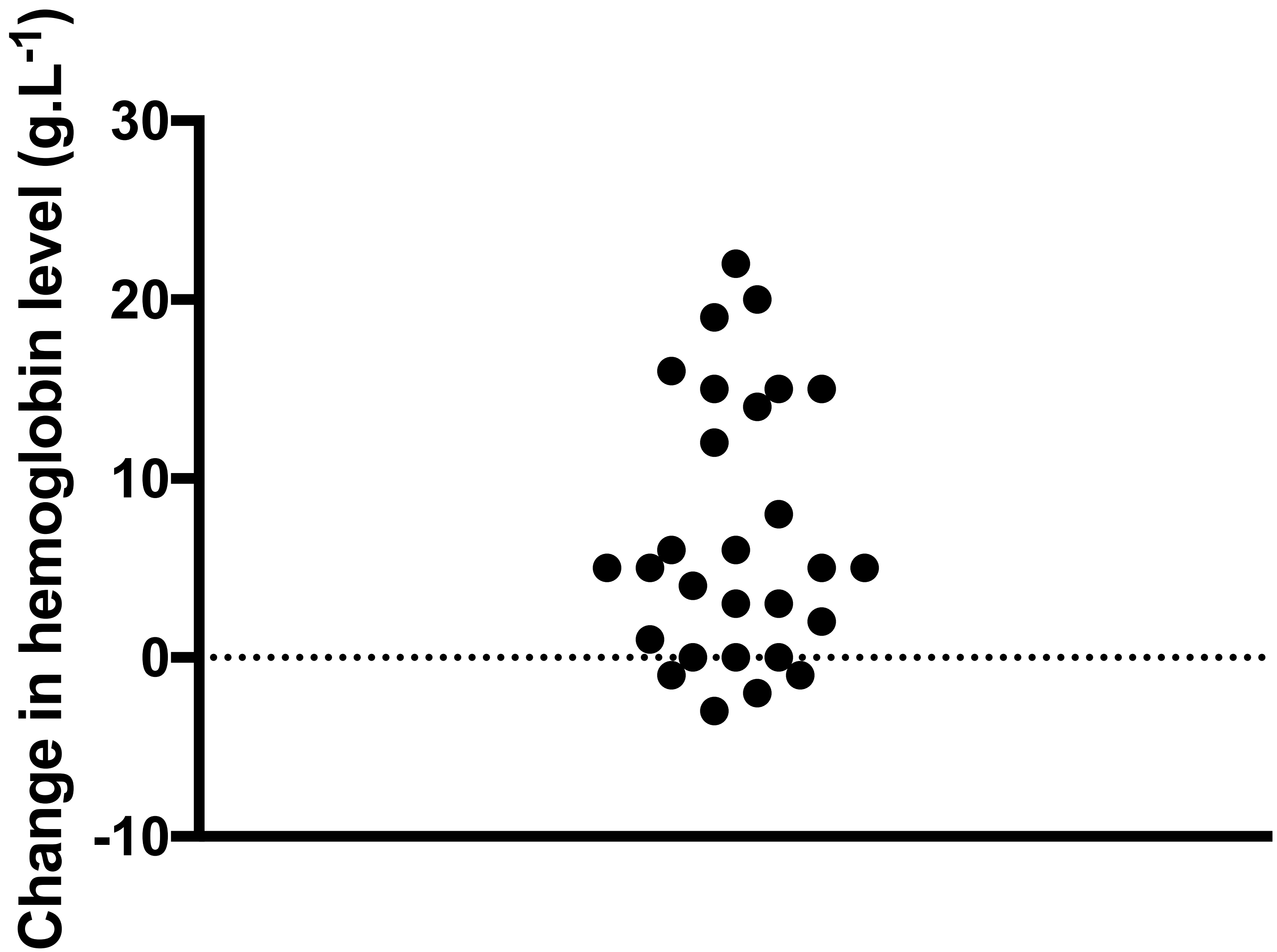
WILCOXON PAIRED SAMPLES TEST

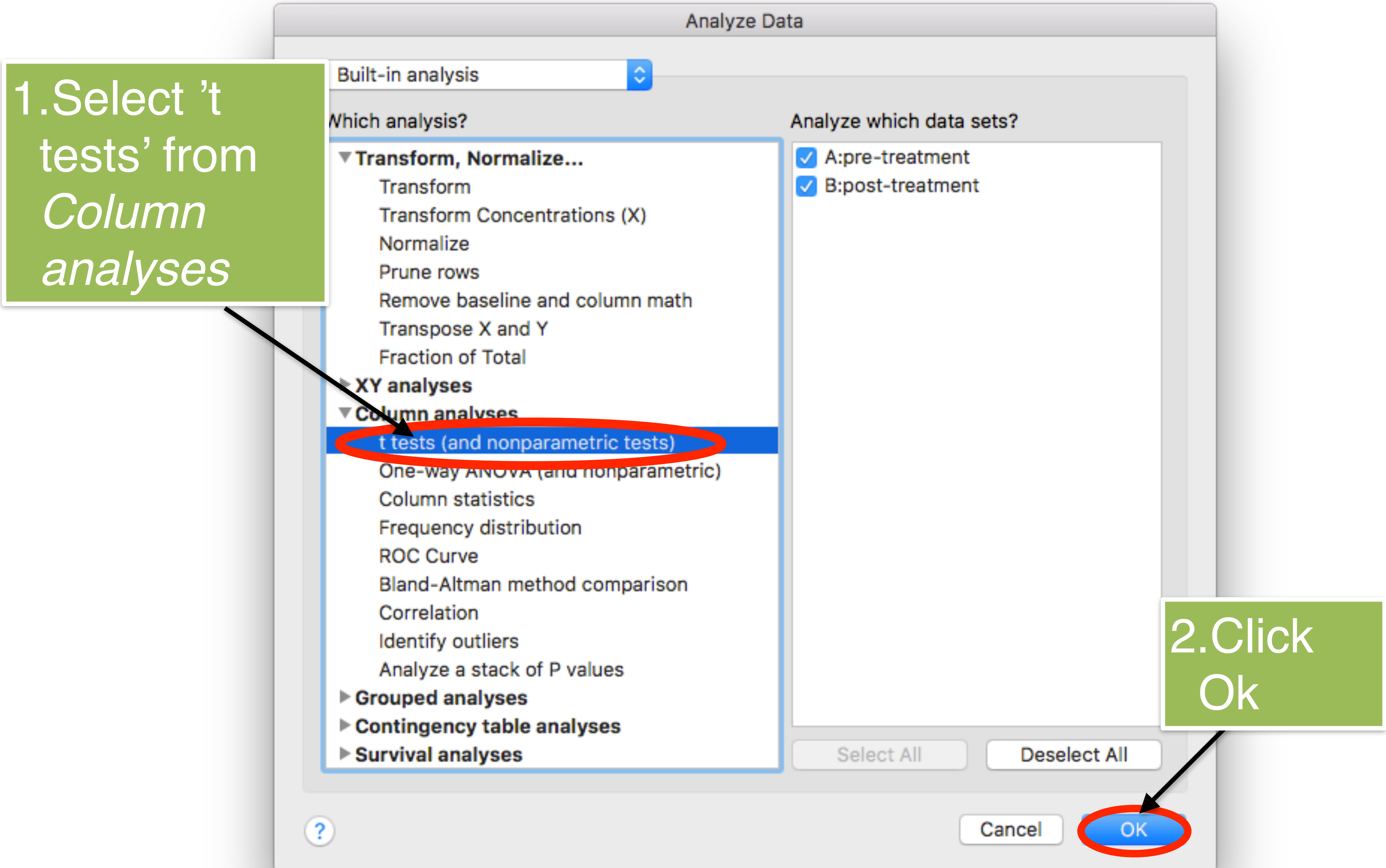
Wilcoxon paired samples test

- Substitute for the paired t-test
- Also called the Wilcoxon signed rank test
- To get test statistic:
    1. Computing the difference between the n pairs of observations.
    2. Dropping any pairs with a difference of zero.
    3. Order the absolute difference from smallest to largest and assigning them ranks (averaging for ties).
    4. The signed-rank statistic, W, is the sum of the ranks from the pairs for which the difference is positive minus the sum of the ranks from the pairs for which the difference is negative.

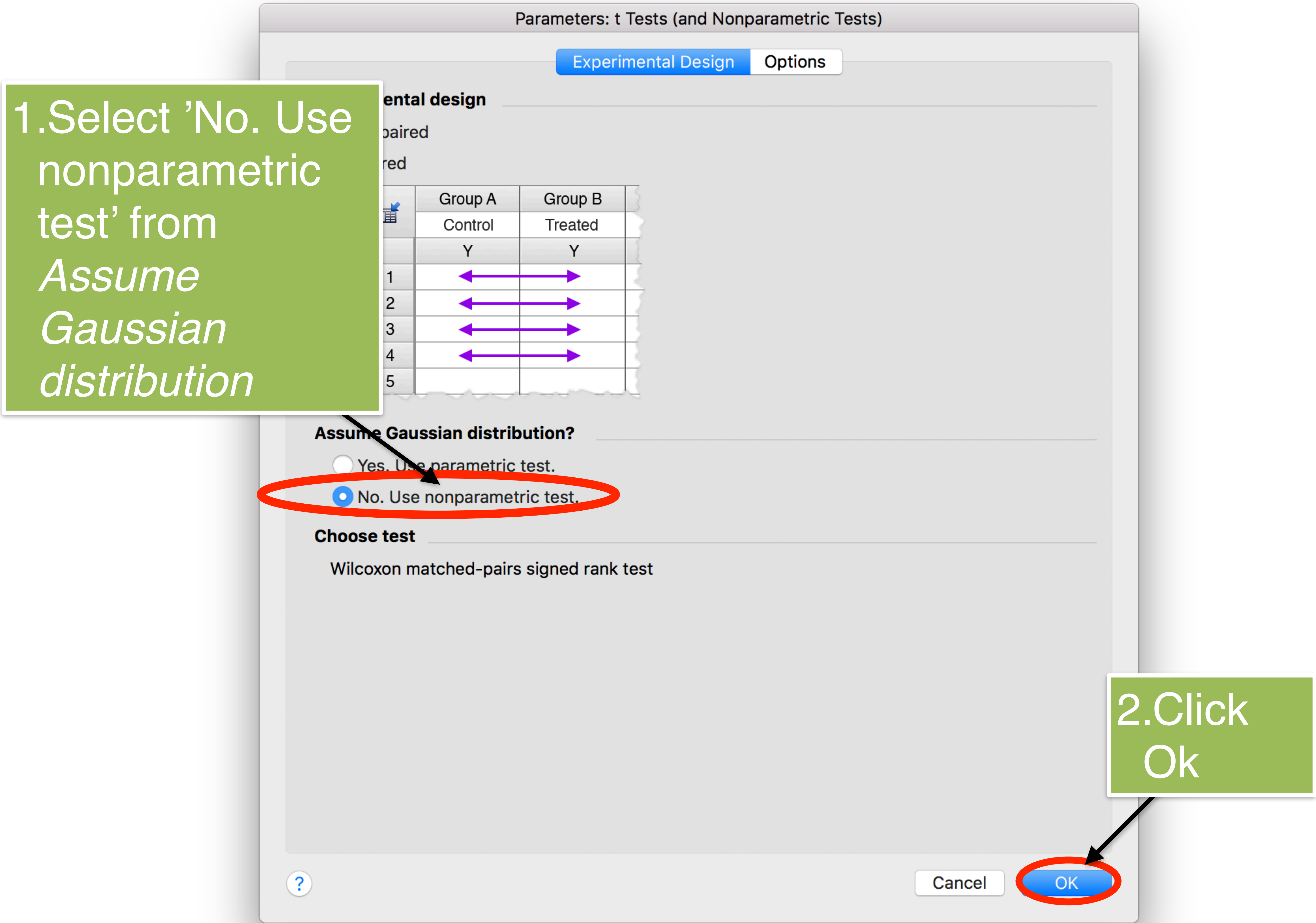
Hemoglobin Example

Group A	Group B
pre-treatment	post-treatment
Y	Y
142	146
140	160
135	143
153	153
136	155
142	141
146	151
117	133
139	139
156	153
154	155
152	150
154	156
133	155
146	151
153	153
126	140
115	114
159	164
146	152
136	142
158	161
129	144





Analyze Data
Built-in 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)
Column statistics
Frequency distribution
ROC Curve
Bland-Altman method comparison
Correlation
Identify outliers
Analyze a stack of P values
Grouped analyses
Contingency table analyses
Survival analyses
Analyze which data sets?
A:pre-treatment
B:post-treatment
Select All
Deselect All
?
Cancel
OK
1.Select 't tests' from *Column analyses*
2.Click Ok


Parameters: t Tests (and Nonparametric Tests)
Experimental Design
Options
ental design
paired
red
Group A
Group B
Control
Treated
Y
Y
1
2
3
4
5
1.Select 'No. Use nonparametric test' from *Assume Gaussian distribution*
Assume Gaussian distribution?
Yes. Use parametric test.
No. Use nonparametric test.
Choose test
Wilcoxon matched-pairs signed rank test
2.Click Ok
?
Cancel
OK

Wilcoxon matched-pairs signed rank test	
P value	<0.0001
Exact or approximate P value?	Exact
P value summary	****
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
Sum of positive, negative ranks	309.5 , -15.5
Sum of signed ranks (W)	294
Number of pairs	28

There is significant evidence that B12 has an effect on hemoglobin levels (p<0.0001).

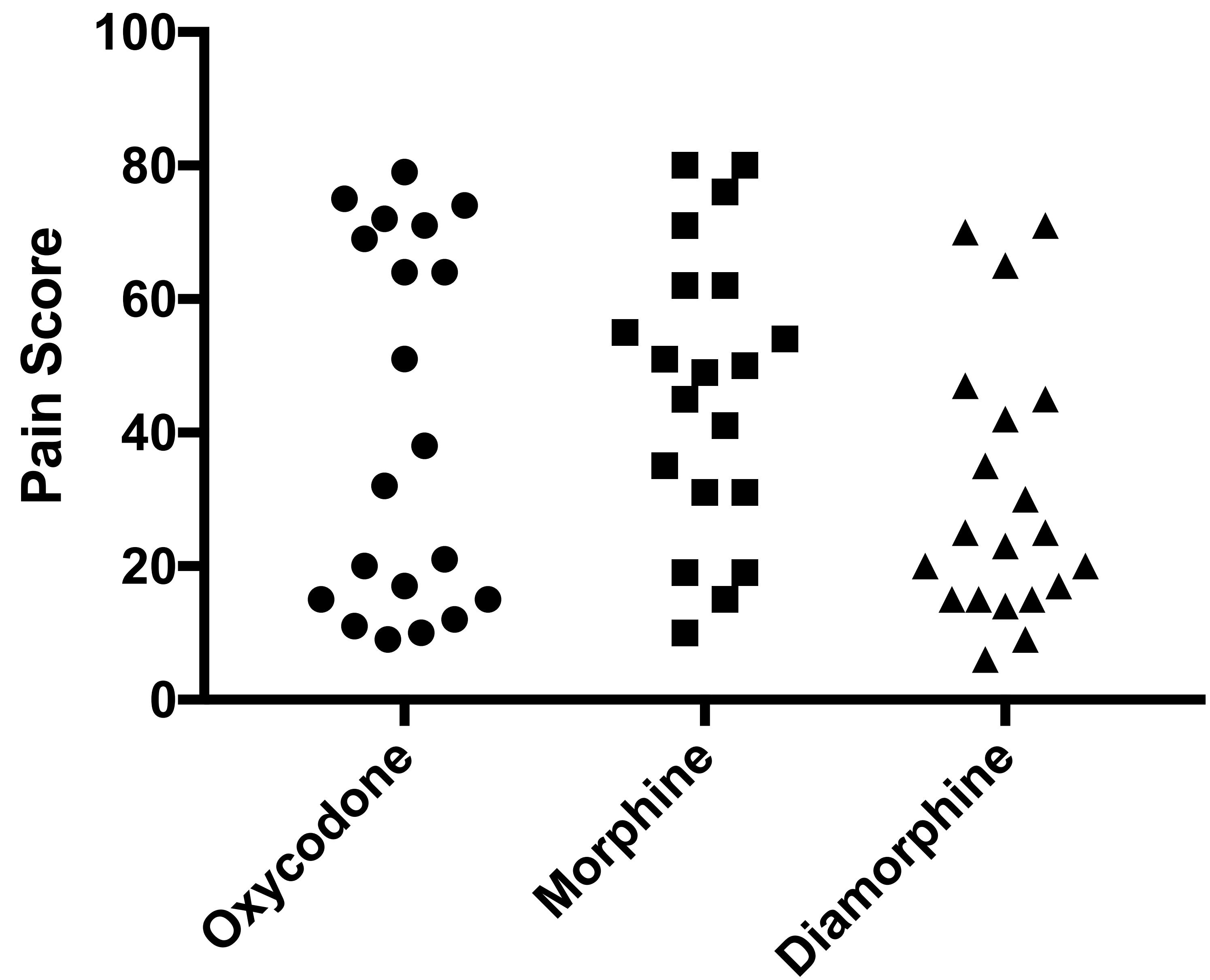
KRUSKAL-WALLIS TEST

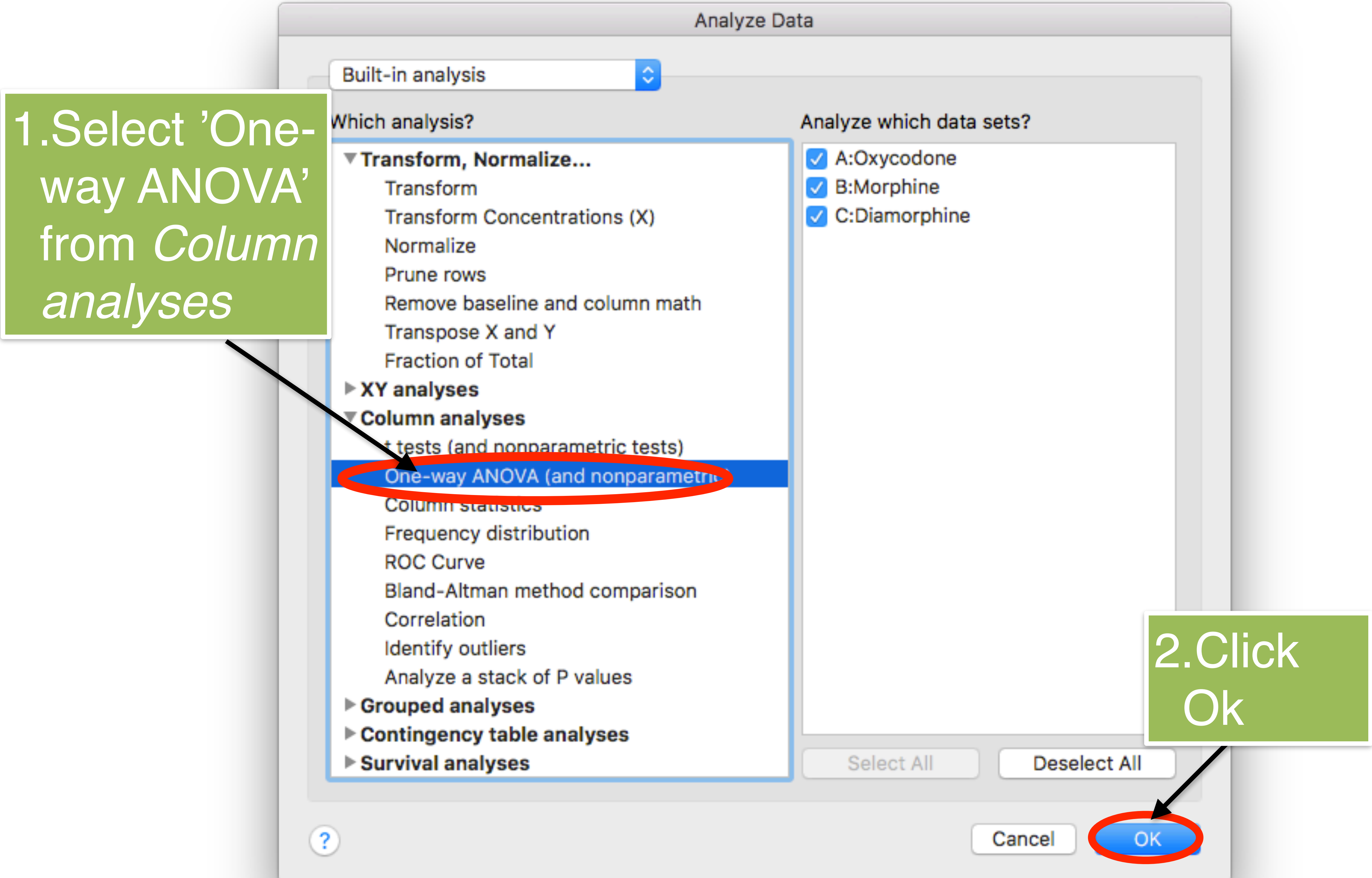
Kruskal-Wallis test

- Substitute for one-way ANOVA
- Basic premise of test:
  1. Rank all values
  2. Compare the mean ranks between groups.

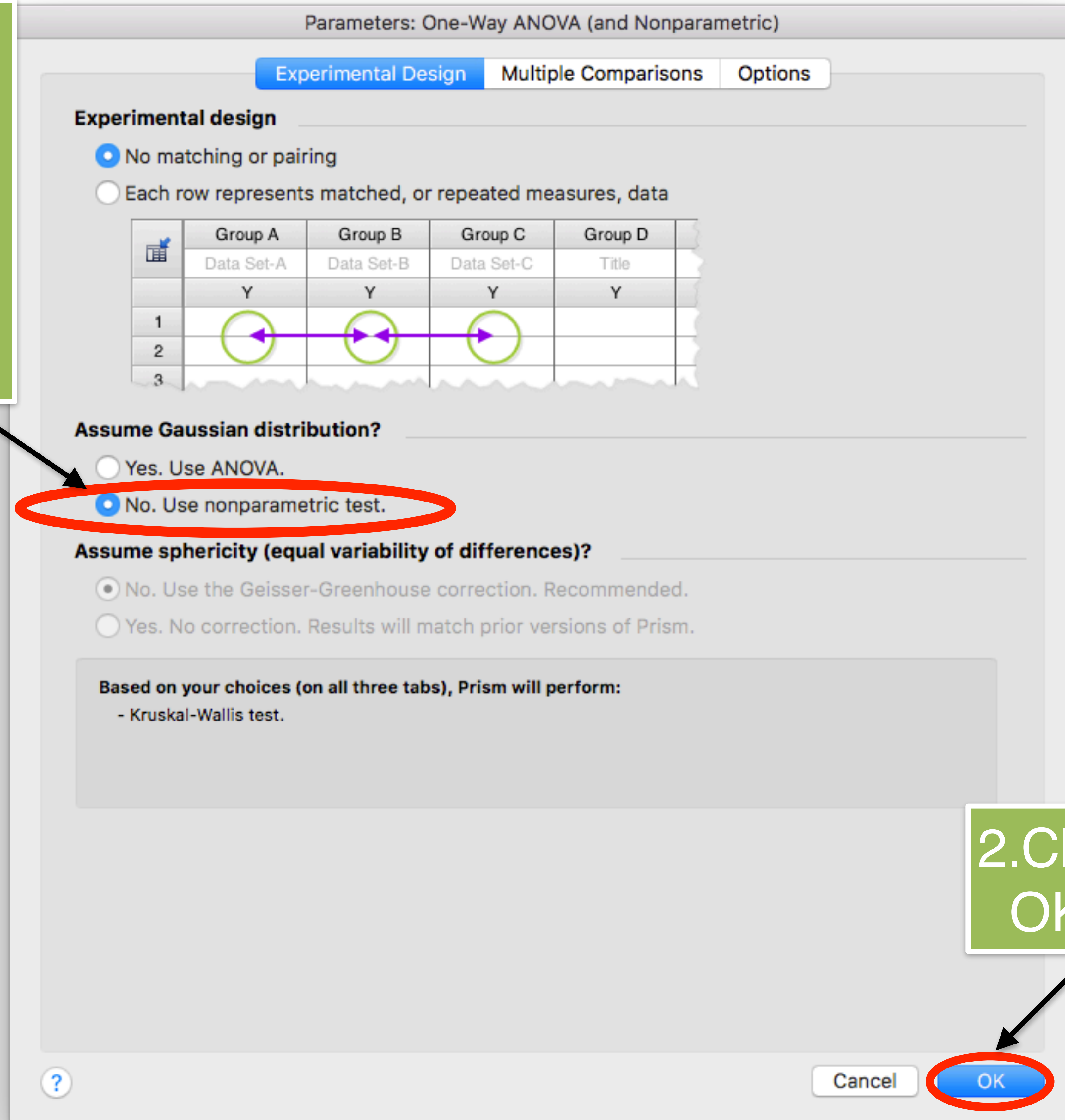
Analgesia comparison in palliative care

Group A	Group B	Group C
Oxycodone	Morphine	Diamorphine
Y	Y	Y
64	62	23
71	50	42
21	76	30
75	51	70
12	45	71
11	35	65
15	54	35
20	71	20
69	31	25
38	49	45
9	19	9
10	10	15
74	55	47
79	80	6
17	15	15
32	62	14
15	31	15
51	80	17
72	41	20
64	19	25

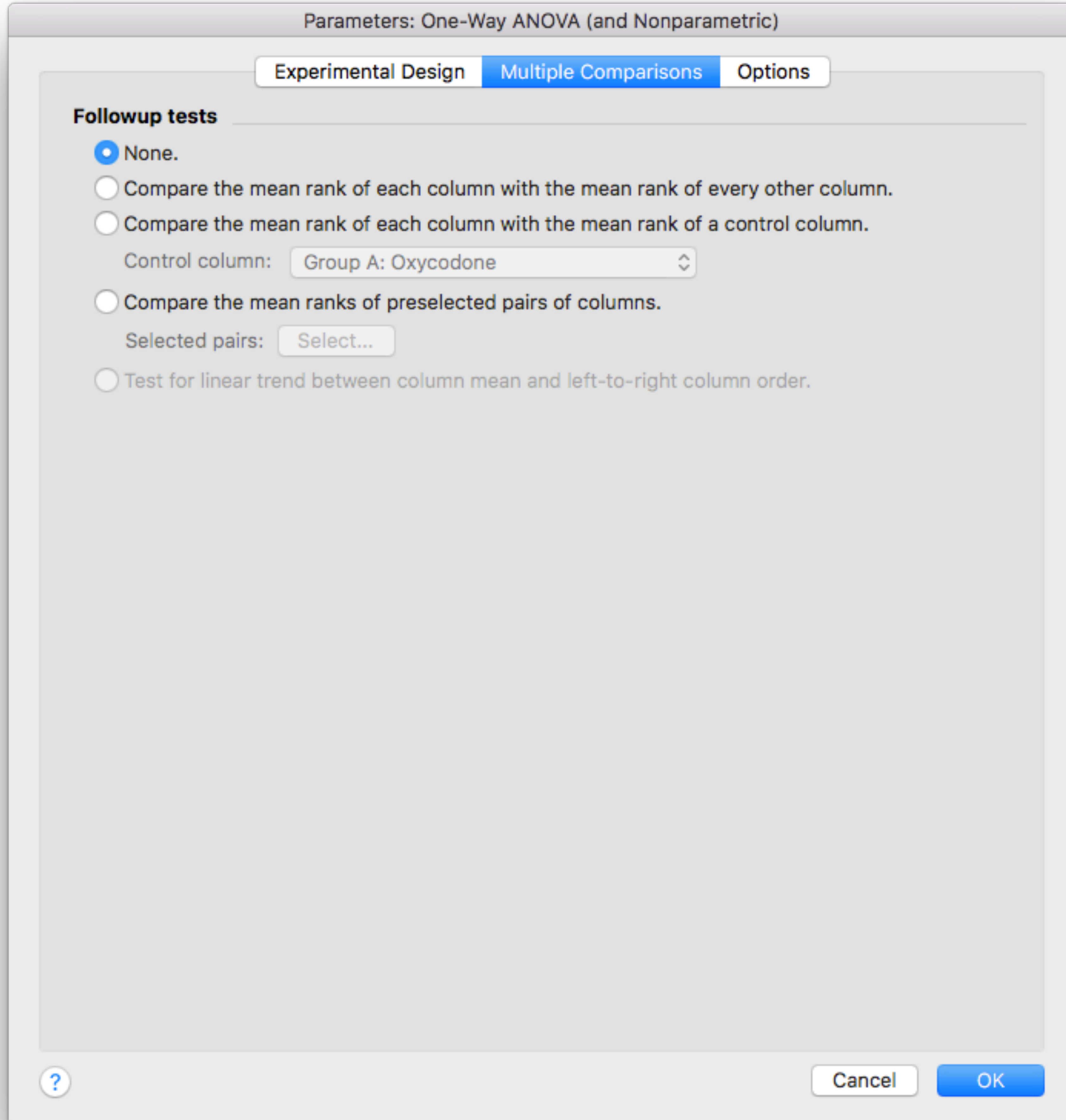


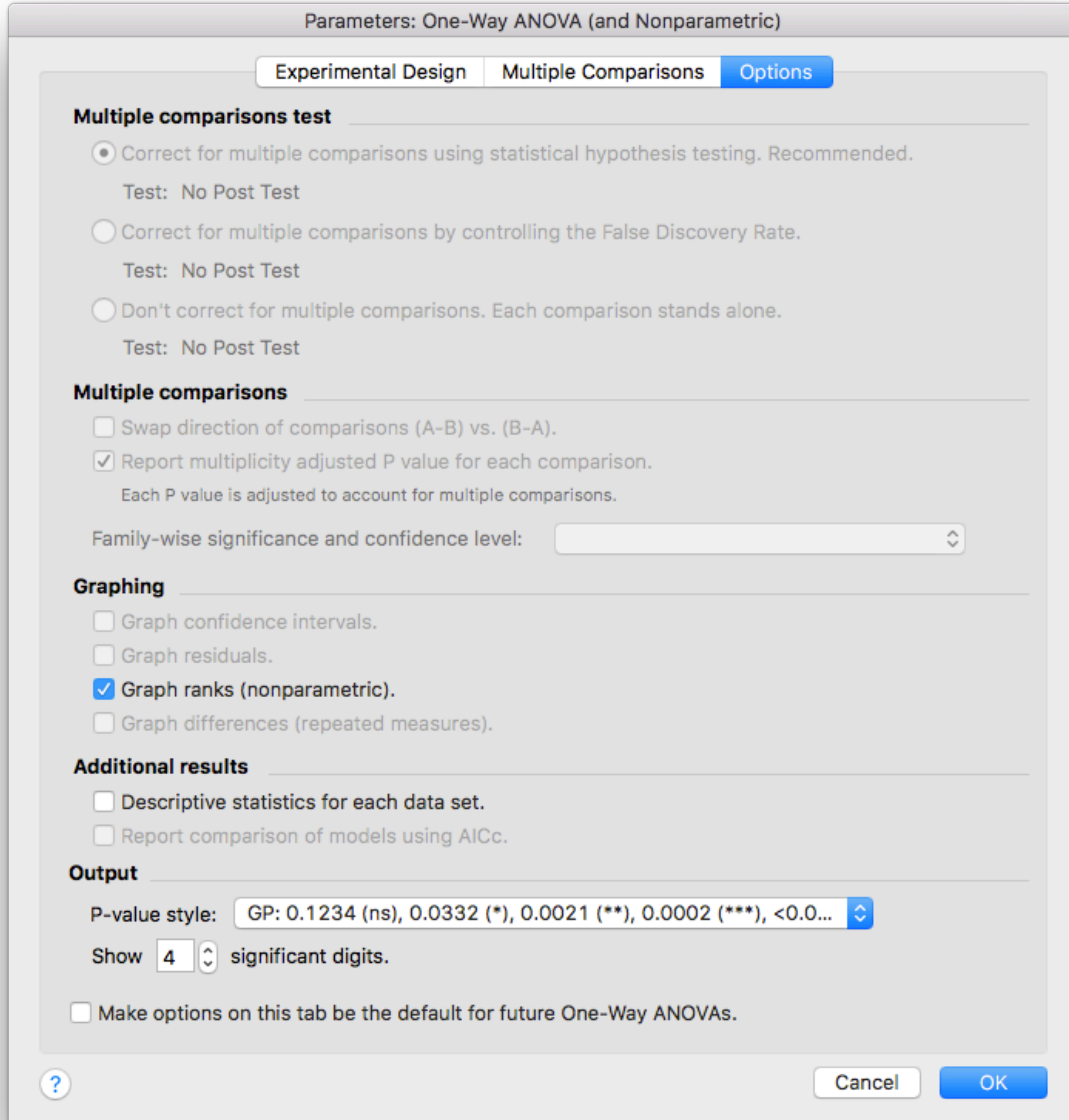

1.Select 'One-way ANOVA' from *Column analyses*
Analyze Data
Built-in analysis
Which analysis?
Transform, Normalize…
Transform
Transform Concentrations (X)
Normalize
Prune rows
Remove baseline and column math
Transpose X and Y
Fraction of Total
XY analyses
Column analyses
t tests (and nonparametric tests)
Frequency distribution
ROC Curve
Bland-Altman method comparison
Correlation
Identify outliers
Analyze a stack of P values
Grouped analyses
Contingency table analyses
Survival analyses
Analyze which data sets?
A:Oxycodone
B:Morphine
C:Diamorphine
Select All
Deselect All
?
Cancel
OK
2.Click Ok

1.Select 'No. Use nonparametric test' from *Assume Gaussian distribution*

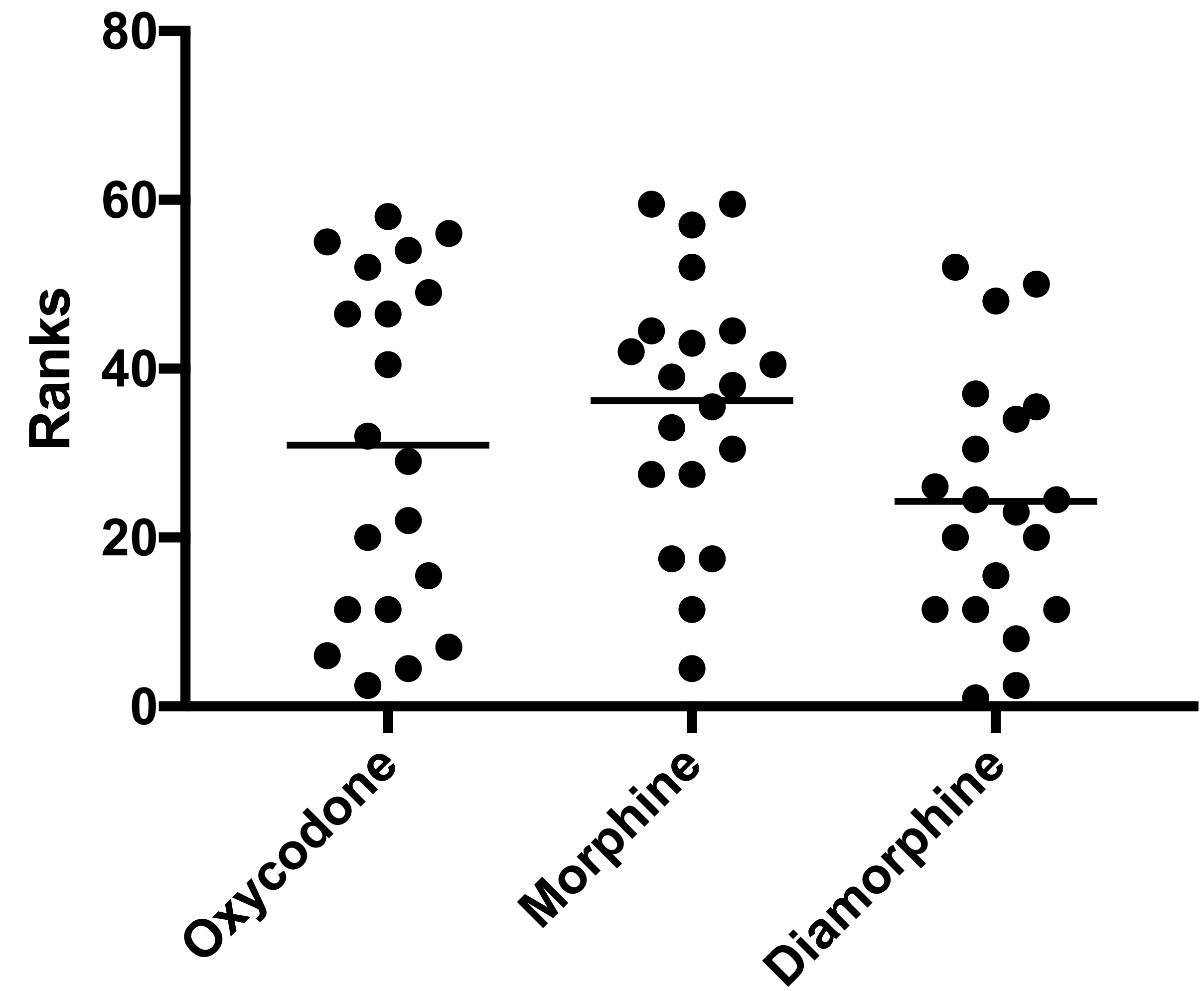


2.Click OK


Parameters: One-Way ANOVA (and Nonparametric)
Experimental Design
Multiple Comparisons
Options
Followup tests
None.
Compare the mean rank of each column with the mean rank of every other column.
Compare the mean rank of each column with the mean rank of a control column.
Control column:
Group A: Oxycodone
Compare the mean ranks of preselected pairs of columns.
Selected pairs:
Select...
Test for linear trend between column mean and left-to-right column order.
?
Cancel
OK


Parameters: One-Way ANOVA (and Nonparametric)
Experimental Design
Multiple Comparisons
Options
Multiple comparisons test
Correct for multiple comparisons using statistical hypothesis testing. Recommended.
Test: No Post Test
Correct for multiple comparisons by controlling the False Discovery Rate.
Test: No Post Test
Don't correct for multiple comparisons. Each comparison stands alone.
Test: No Post Test
Multiple comparisons
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:
Graphing
Graph confidence intervals.
Graph residuals.
Graph ranks (nonparametric).
Graph differences (repeated measures).
Additional results
Descriptive statistics for each data set.
Report comparison of models using AICc.
Output
P-value style:
GP: 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***), <0.0...
Show
4
significant digits.
Make options on this tab be the default for future One-Way ANOVAs.
?
Cancel
OK

Kruskal-Wallis test	
P value	0.0968
Exact or approximate P value?	Approximate
P value summary	ns
Do the medians vary signif. (P < 0.05)?	No
Number of groups	3
Kruskal-Wallis statistic	4.67



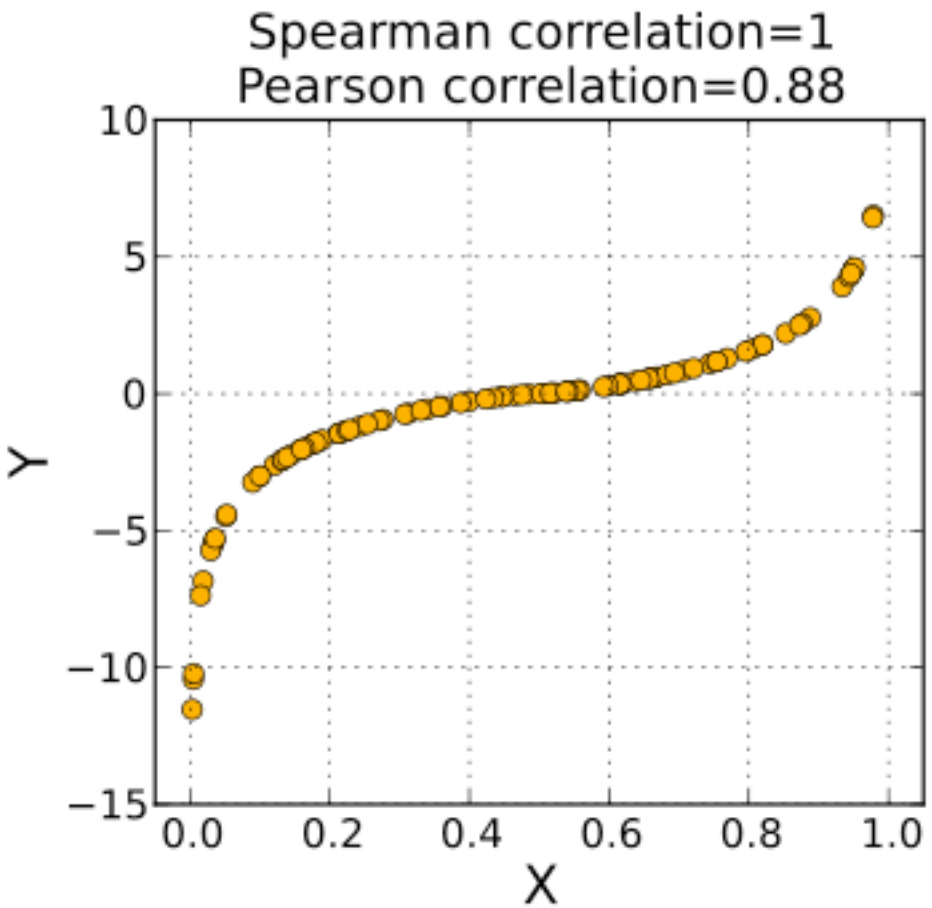
There is no significant difference in pain scores among the different analgesias.

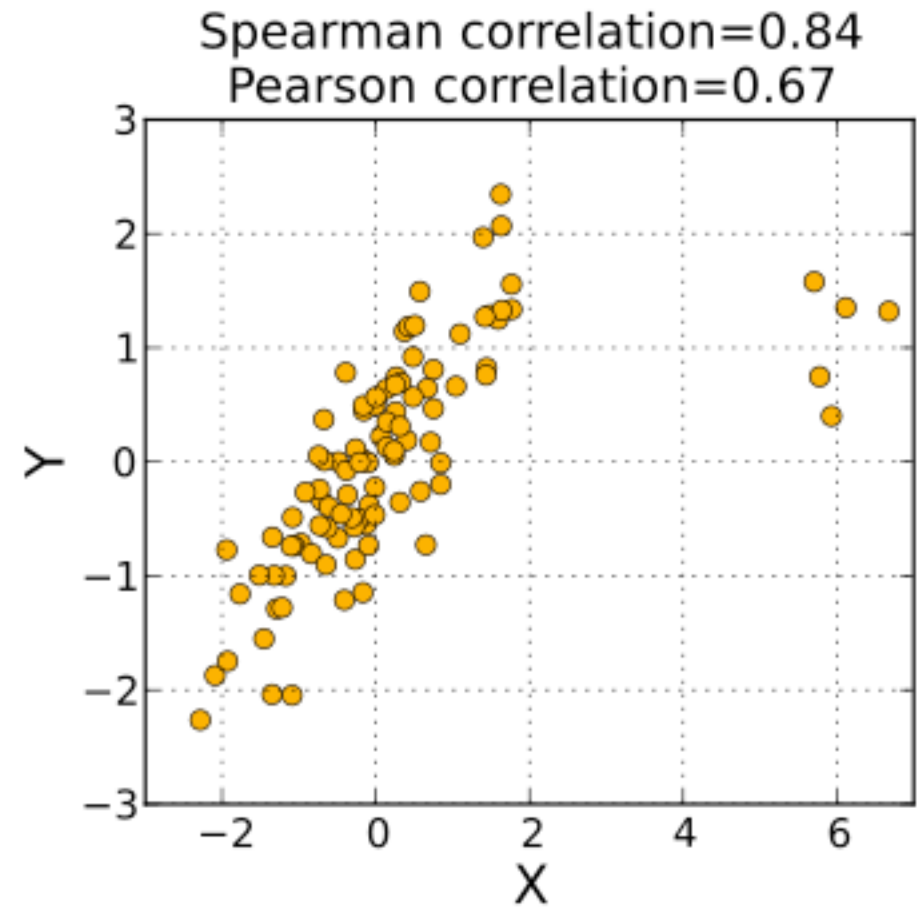
SPEARMAN CORRELATION

Spearman correlation

- Substitute for Pearson correlation
- Basic premise of test:
  1. Rank values of the two factors separately
  2. Carry out correlation analysis of the rank values
- Unlike Pearson correlation, Spearman correlation does not require a linear relationship between variables. Instead is requires ‘monotonicity’.
  - Monotonic - throughout the range of x values studied, y values always increase or always decrease.

Spearman vs. Pearson



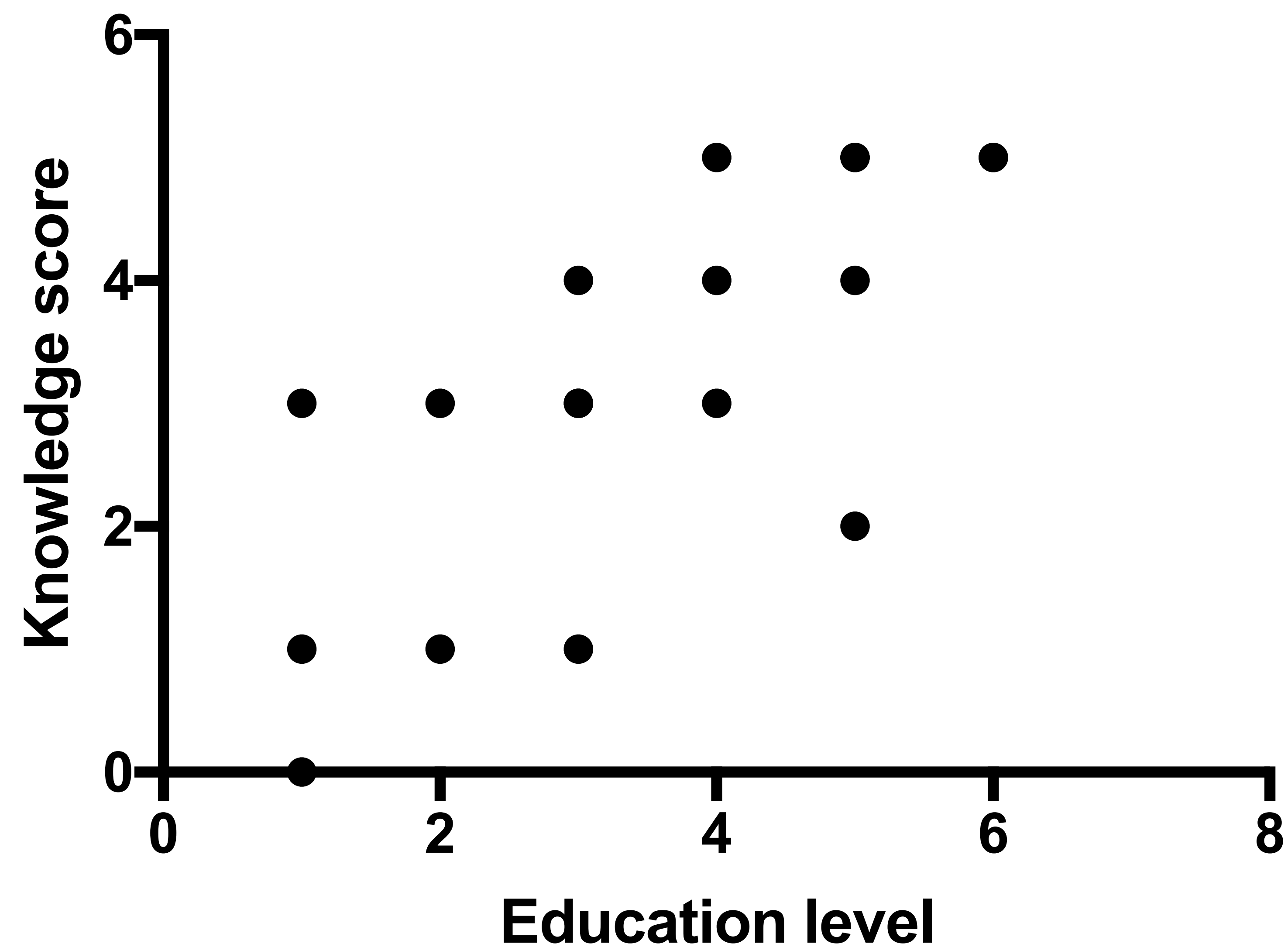


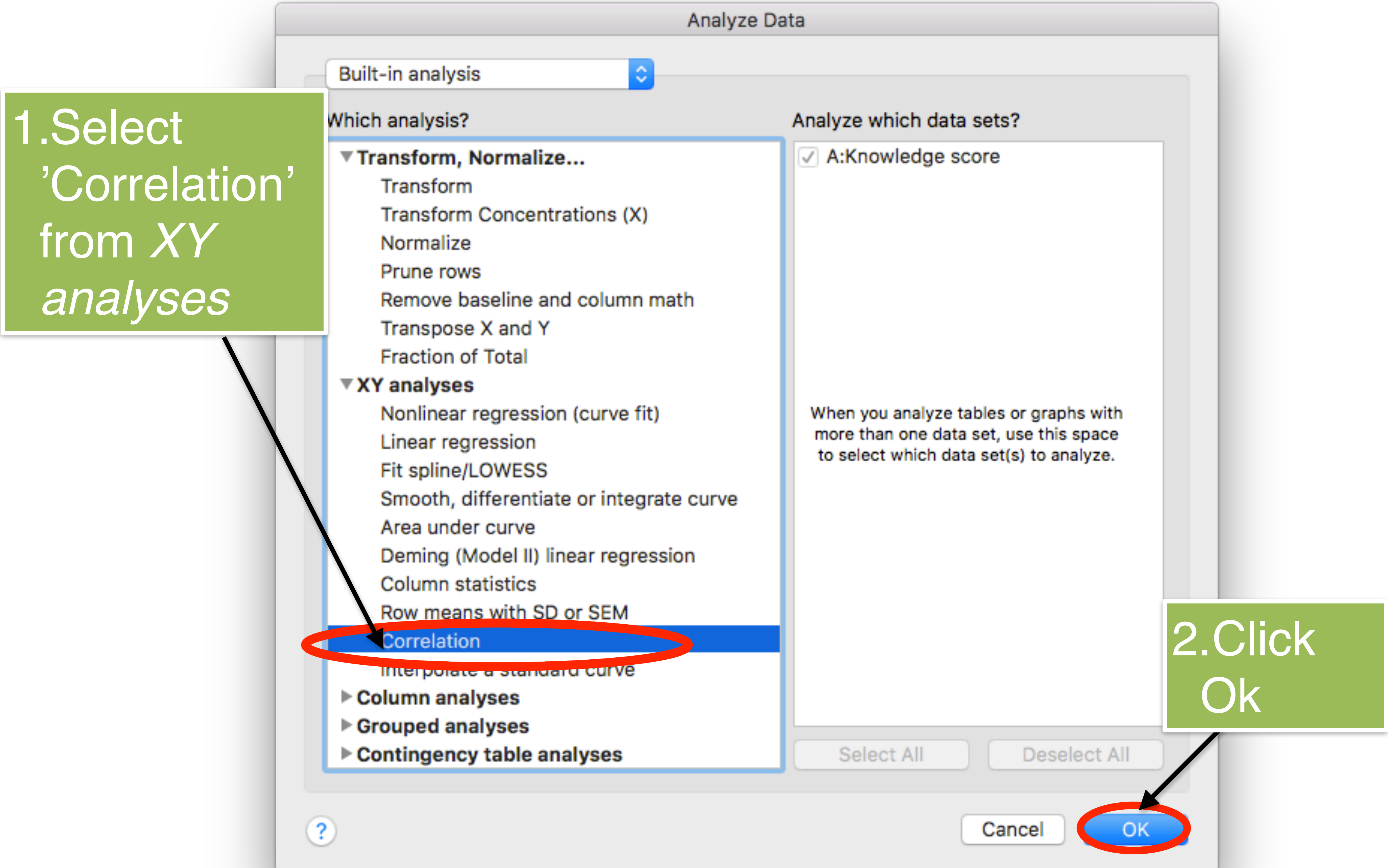
- Data do not have to be in a straight line to have a perfect correlation.
- Data are perfectly correlated for Spearman's rho if the ranks are monotonically increasing for both variables.
- Data must be in a straight line for Pearson's correlation coefficient to be 1 to -1.

- The Spearman correlation is less sensitive than the Pearson correlation to strong outliers that are in the tails of both samples

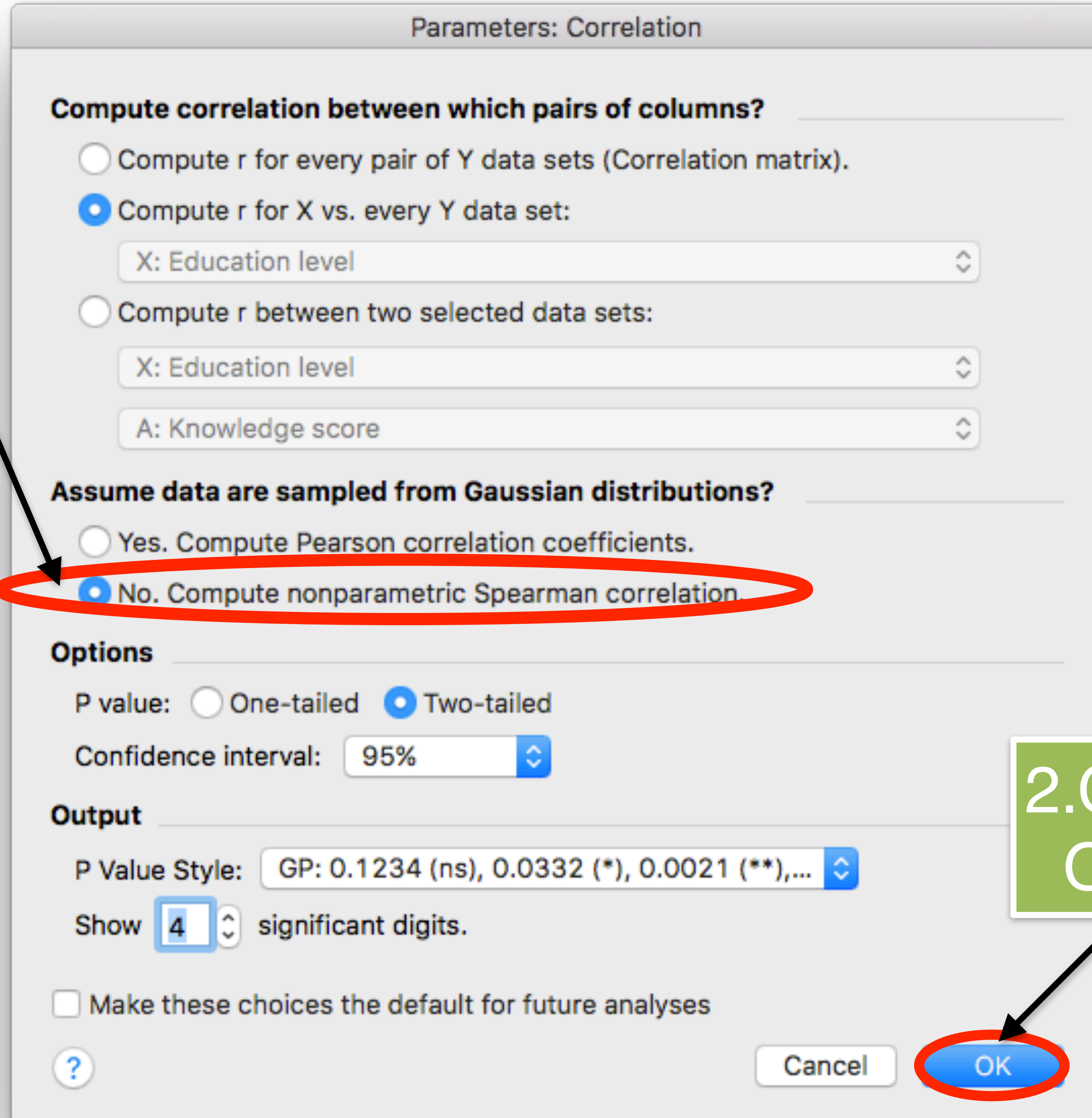
Leaflet Example

X	Group A
Education level	Knowledge score
X	Y
1	0
1	3
1	1
2	3
2	1
3	1
3	3
3	4
3	3
4	4
4	5
4	3
5	5
5	5
5	2
5	4
6	5
6	5




1.Select 'Correlation' from *XY analyses*
Analyze Data
Built-in 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
Column statistics
Row means with SD or SEM
Correlation
Interpolate a standard curve
Column analyses
Grouped analyses
Contingency table analyses
Analyze which data sets?
A:Knowledge score
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
2.Click Ok

1. Select 'No. Compute nonparametric Spearman correlation' from *Assume data are sampled from Gaussian distributions*



2. Click Ok

Correlation		A Education level vs. Knowledge score
		Y
1	Spearman r	
2	r	0.7484
3	95% confidence interval	0.4205 to 0.9034
4		
5	P value	
6	P (two-tailed)	0.0004
7	P value summary	***
8	Exact or approximate P value?	Approximate
9	Significant? (alpha = 0.05)	Yes
10		
11	Number of XY Pairs	18

There is a significant positive association between education level and knowledge score (Spearman r=0.75, p-value=0.0004).

What did we learn?

Parametric	Non-Parametric
Two-sample t-test	Mann-Whitney test
Paired t-test	Wilcoxon paired samples test
One-Way ANOVA	Kruskal-Wallis test
Pearson correlation	Spearman correlation

- Prior to this chapter, we learned about ‘parametric’ tests that depend on a normality assumption. Non-parametric model are an alternative if the normality assumption is not met.
- Transformation of data is preferred over non-parametric tests to preserve interpretability of effect sizes and power (if data are truly normally distributed).
- Ordinal data is typically not normally distributed and often non-parametric methods are employed.