Chapter 15: Correlation and regression - Relationships between measured values

- *Correlation Analysis*

TXCL7565/PHSC7565

What This Lecture Covers

- Correlation coefficient (*r*)
- Significance testing of correlation coefficients
- Preliminary check for non-linearity
- Correlation analysis in GraphPad Prism

CORRELATION COEFFICIENT (*r*)

Correlation

- Describes the association of two interval variables

- ***Positive correlation*** - an increase in one variable is associated with an increase in the other variable

- ***Negative correlation*** - an increase in one variable is associated with a decrease in the other variable

Types of Correlation

- Pearson product-moment correlation (parametric) assumes a linear relationship between factors.

  ➡ Used in this lecture

- Spearman's rank correlation coefficient (non-parametric) assumes a linear relationship between ranks of factor levels.

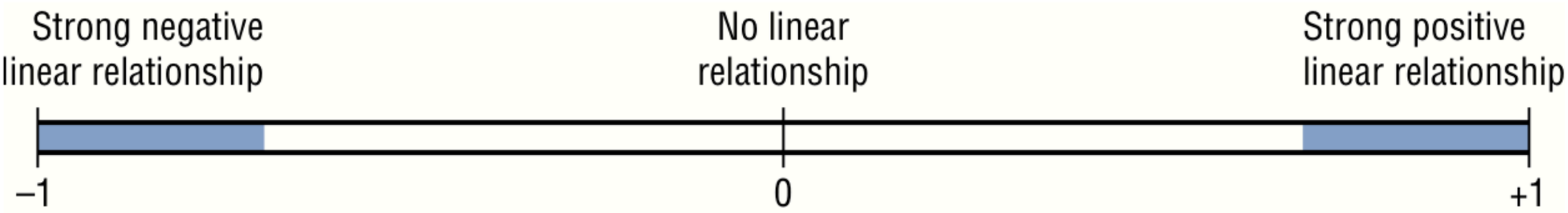
  ➡ Used in Chapter 21

Correlation Coefficient

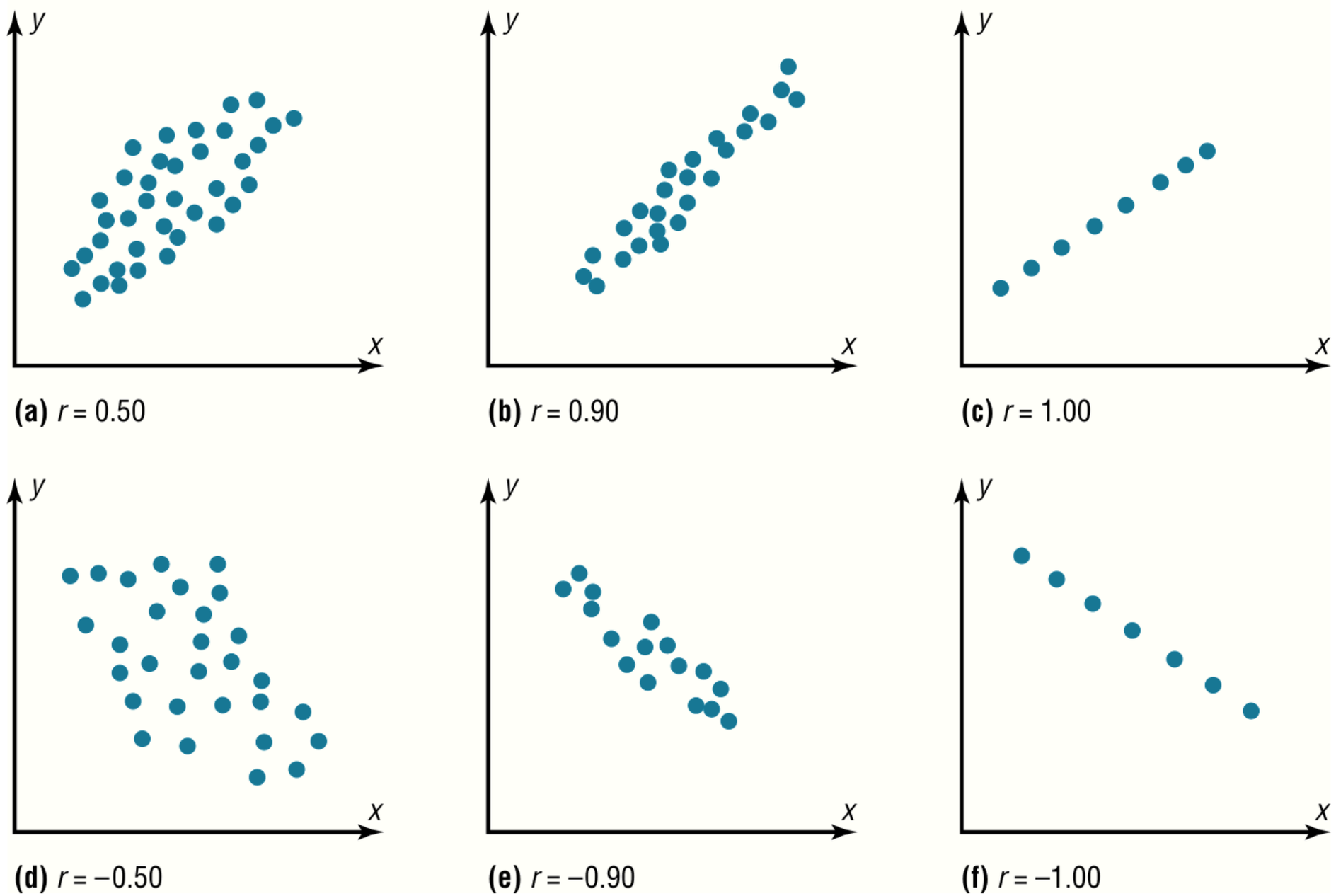
Correlation coefficient (*r*) describes the type (positive or negative) and the strength of correlation.

- It can take any value between -1 and +1
- A value close to +1 indicates a strong positive linear relationship
- A value close to -1 indicates a strong negative linear relationship
- A value close to 0 indicates a weak linear relationship or no linear relationship between two variables
- A value of 1 indicates that the data are perfectly positively correlated (they form an increasing straight line when plotted)
- A value of -1 indicates that the data are perfectly negatively correlated (they form a decreasing straight line when plotted)

Range of possible values for a correlation coefficient



Example Correlations



(a) $r = 0.50$ **(b)** $r = 0.90$ **(c)** $r = 1.00$

(d) $r = -0.50$ **(e)** $r = -0.90$ **(f)** $r = -1.00$

SIGNIFICANCE TESTING OF CORRELATION COEFFICIENTS

Null and alternative hypotheses

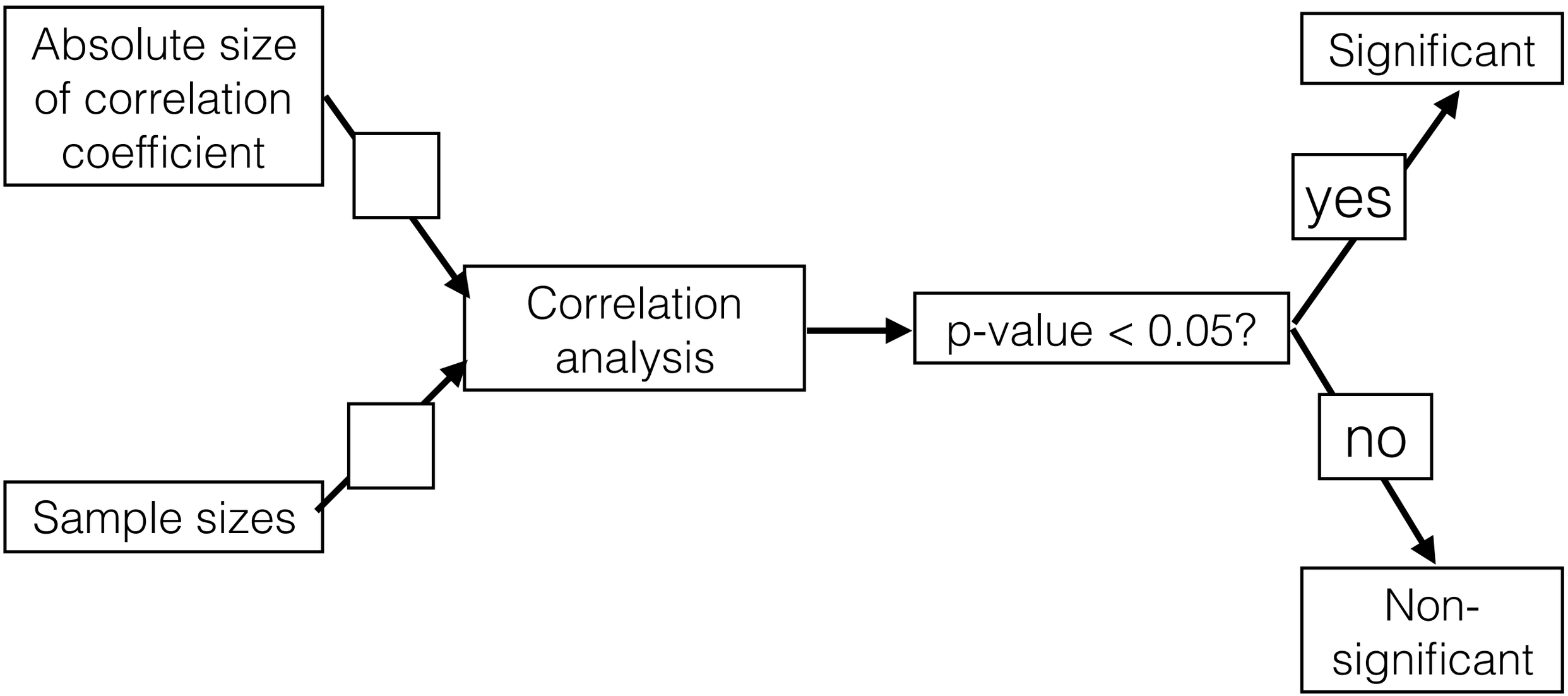
$$H_0 : \rho = 0$$

$$H_a : \rho \neq 0$$

Null Hypothesis: The population correlation coefficient is equal to zero (i.e., there is no linear association between factors).

Alternative Hypothesis: The population correlation coefficient is not equal to zero (i.e, there is a linear relationship between the two factors).

What governs whether a significant result will emerge?

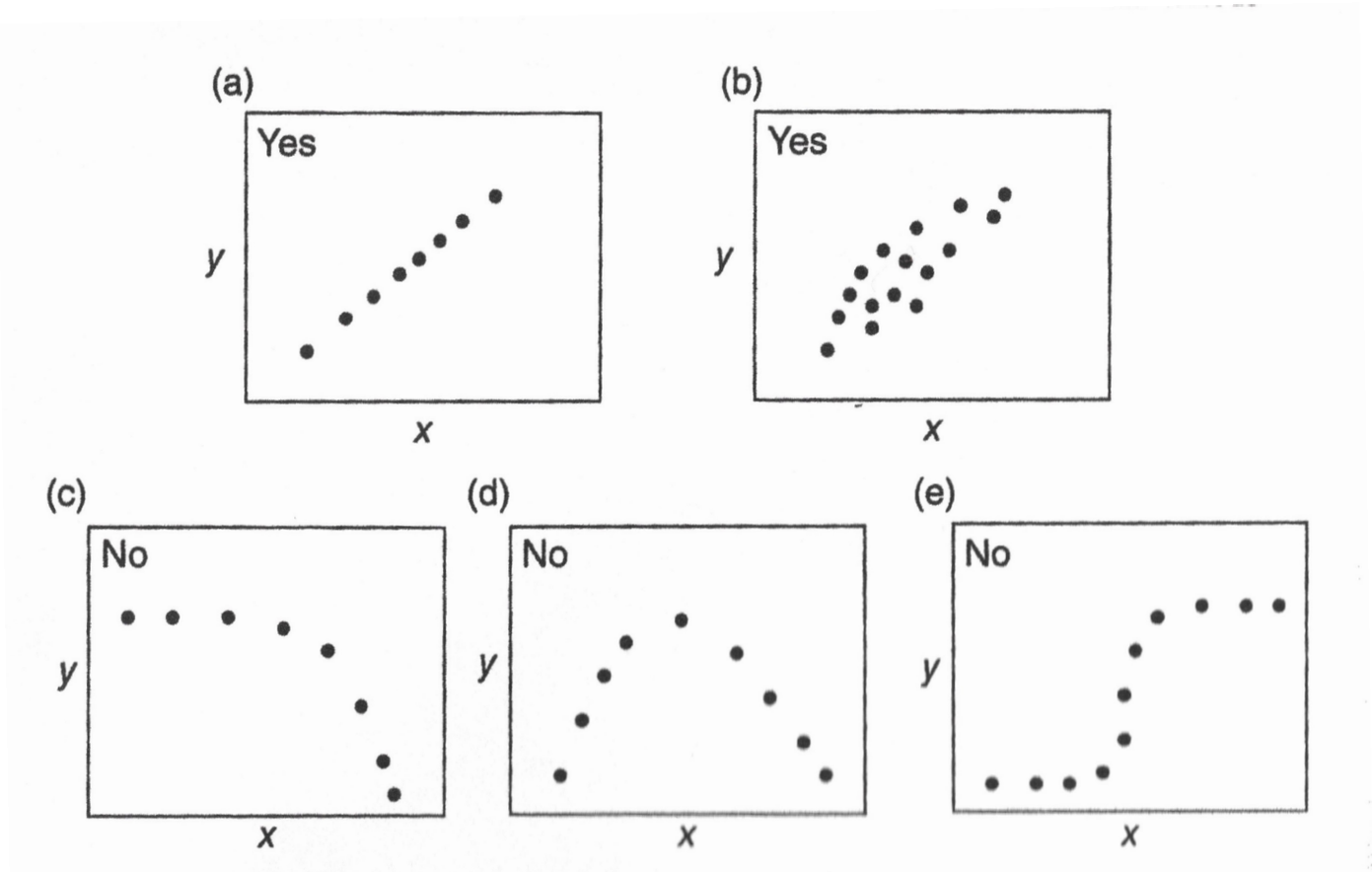


PRELIMINARY CHECK FOR NON-LINEARITY

Non-linearity and correlation

- Correlation analysis is a search for a straight line relationship

- If data have a non-linear relationship, correlation can be very misleading



CORRELATION ANALYSIS IN GRAPHPAD PRISM

Example - Drug Concentration

We are planning the commercial collection of the leaves of a species of tree from which a drug can be extracted. One question we need to consider is whether it is worth using ladders to gain access to leaves at the tops of the trees or whether we would be better just collecting the easily accessed, low growing leaves and moving on to the next tree. We randomly select 24 leaves from 24 randomly selected trees (one from each tree) and measure the height of at which the leaf was growing in the tree (meters) and the drug content (mg per 100 g of dry leaf).

Chapter15_drugConc.txt on Canvas

Is this an observational study or a randomized experiment?

Can we make cause-and-effect conclusions from this study?

State the null and alternative hypothesis.

Welcome to GraphPad Prism

GraphPad PRISM 7.0a

New table & graph

XY

Column

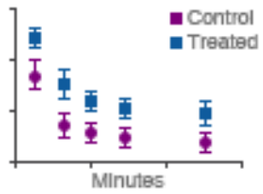
Grouped

Contingency

Survival

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

		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							



Learn more

Enter/import data:

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 2 replicate values in side-by-side subcolumns

Enter and plot error values already calculated elsewhere

Enter: Mean, SD, N

Use tutorial data:

Linear regression - Compare slopes

Nonlinear regression -- One phase exponential decay

Dose-response - X is log(dose)

Interpolate unknowns from a linear standard curve

Correlation

Entering dates into the X column

Entering elapsed times into the X column

More tutorial data...

Prism Tips

Cancel

Create

1.Select 'XY' from New table & graph

2.Select 'X: Numbers' and 'Y: Enter and plot a single Y'

3.Click Create

Untitled — Edited

File Sheet Undo Clipboard Analysis Change Import Draw Write Text Export Print Send LA Help

New Analyze

14 Arial

GraphPad PRISM

- Family
- Search results
- Data Tables
  - drugConc
- Info
  - Project info 1
- Results
- Graphs
  - drugConc
- Layouts

		X	Group A	Group B	Group C	Group D	Group E	Group F	Group G	Group H	Group I	Grou
		Height	Drug	Title	Title	Title	Title	Title	Title	Title	Title	Titl
		X	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
1	Title	1.70	1.66									
2	Title	2.31	1.34									
3	Title	2.89	1.27									
4	Title	1.30	1.61									
5	Title	3.21	1.17									
6	Title	1.84	1.73									
7	Title	3.27	1.17									
8	Title	4.21	1.19									
9	Title	1.32	1.93									
10	Title	3.67	1.10									
	Title	2.78	1.37									
	Title	3.71	1.19									
	Title	3.23	1.27									
	Title	3.29	0.85									
	Title	3.46	1.16									
16	Title	3.95	1.14									
17	Title	1.70	1.25									
18	Title	2.92	1.49									
19	Title	2.67	1.17									
20	Title	3.02	1.16									
21	Title	2.37	1.75									
22	Title	2.64	1.36									

drugConc Row 7, Column B

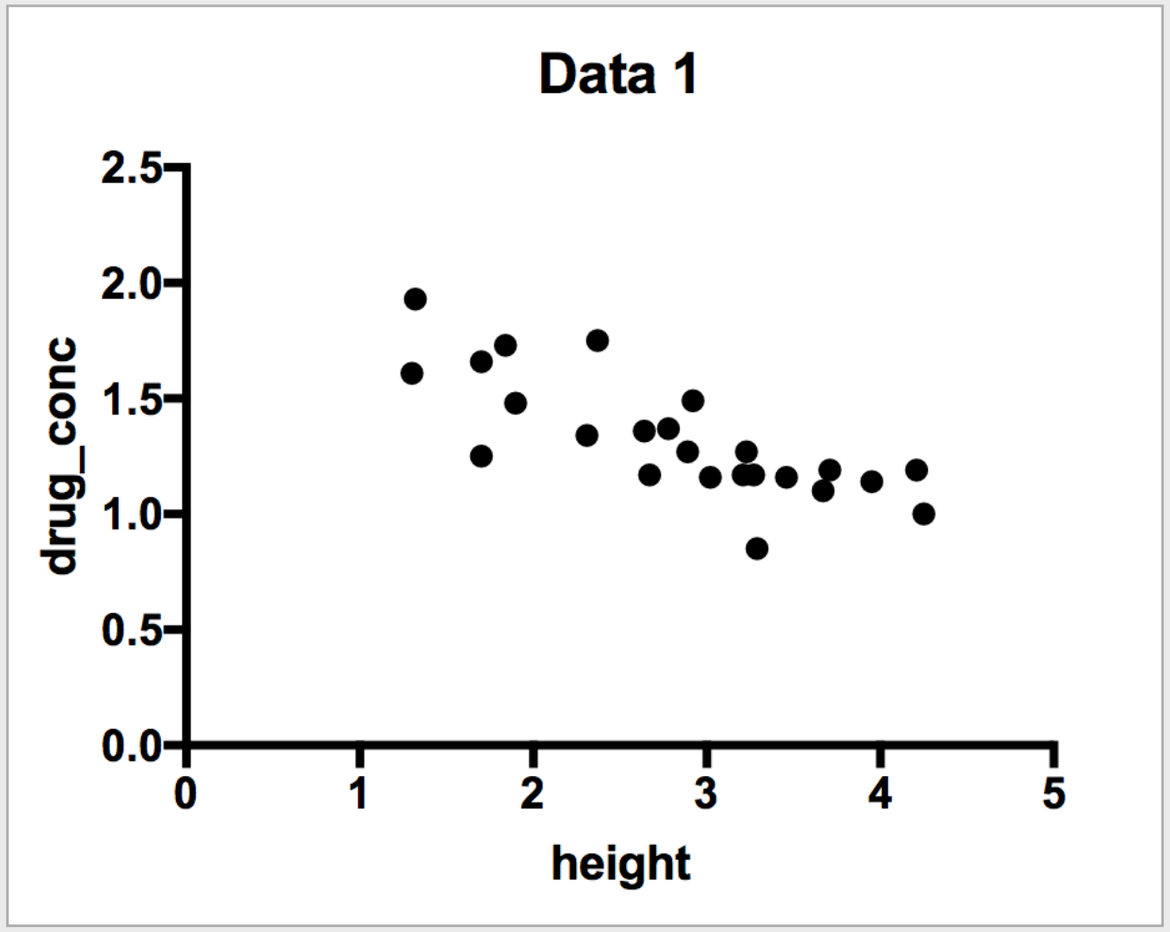
1.Select Graph

Change Graph Type

Graph family: XY

Points only

Plot:



1.Click ‘Ok'

OK

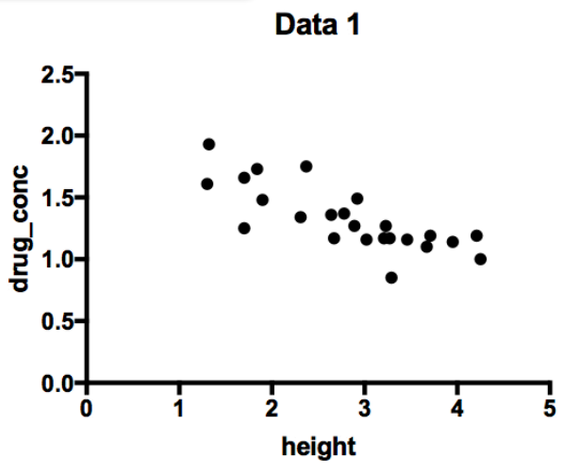
Untitled 2 — Edited

File Sheet Undo Clipboard Analysis Change Arrange Draw Write Text Export Print Send

New

Family
Search results
Data Tables
Data 1
Info
Project info 1
Results
Graphs
Data 1
Layouts

1.Click the linear graph icon to add a line.



Data 1

From left: 1.41 in., From top: 3.4

Parameters: Linear Regression

Interpolate

- [ ] Interpolate unknowns from standard curve

Compare

- [ ] Test whether slopes and intercepts are significantly different

Graphing options

- [ ] Show the 95% confidence bands of the best-fit line
- [ ] Residual plot

Constrain

- [ ] Force the line to go through X = 0 , Y = 0

Replicates

- ( ) Consider each replicate Y value as individual point
- (•) Only consider the mean Y value of each point

Also calculate

- [ ] Test departure from linearity with runs test
- [x] 95% confidence interval of Y when X = 0
- [x] 95% confidence interval of X when Y = 0

Range

Start regression line at:

- (•) Auto
- ( ) X = 1.3

End regression line at:

- (•) Auto
- ( ) X = 4.25

Output options

- [ ] Show table of XY coordinates

P Value Style: GP: 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****)

Show 4 significant digits.

? More choices... Cancel OK

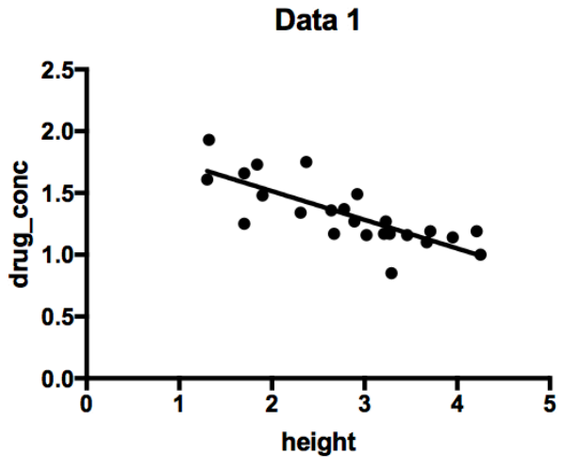
1.Click ‘Ok'

Untitled 2 — Edited

File Sheet Undo Clipboard Analysis Change Arrange Draw Write Text Export Print Send

New Analyze 11 Arial TIFF BMP

- Family
- Search results
- Data Tables
  - Data 1
- Info
  - Project info 1
- Results
  - Linear reg. of Data 1
- Graphs
  - Data 1
- Layouts



Data 1 From left: 5.73 in., From top: 7.

1.Click
‘Analyze’

File Sheet Undo Clipboard Analysis Change Export Print Send LA Help

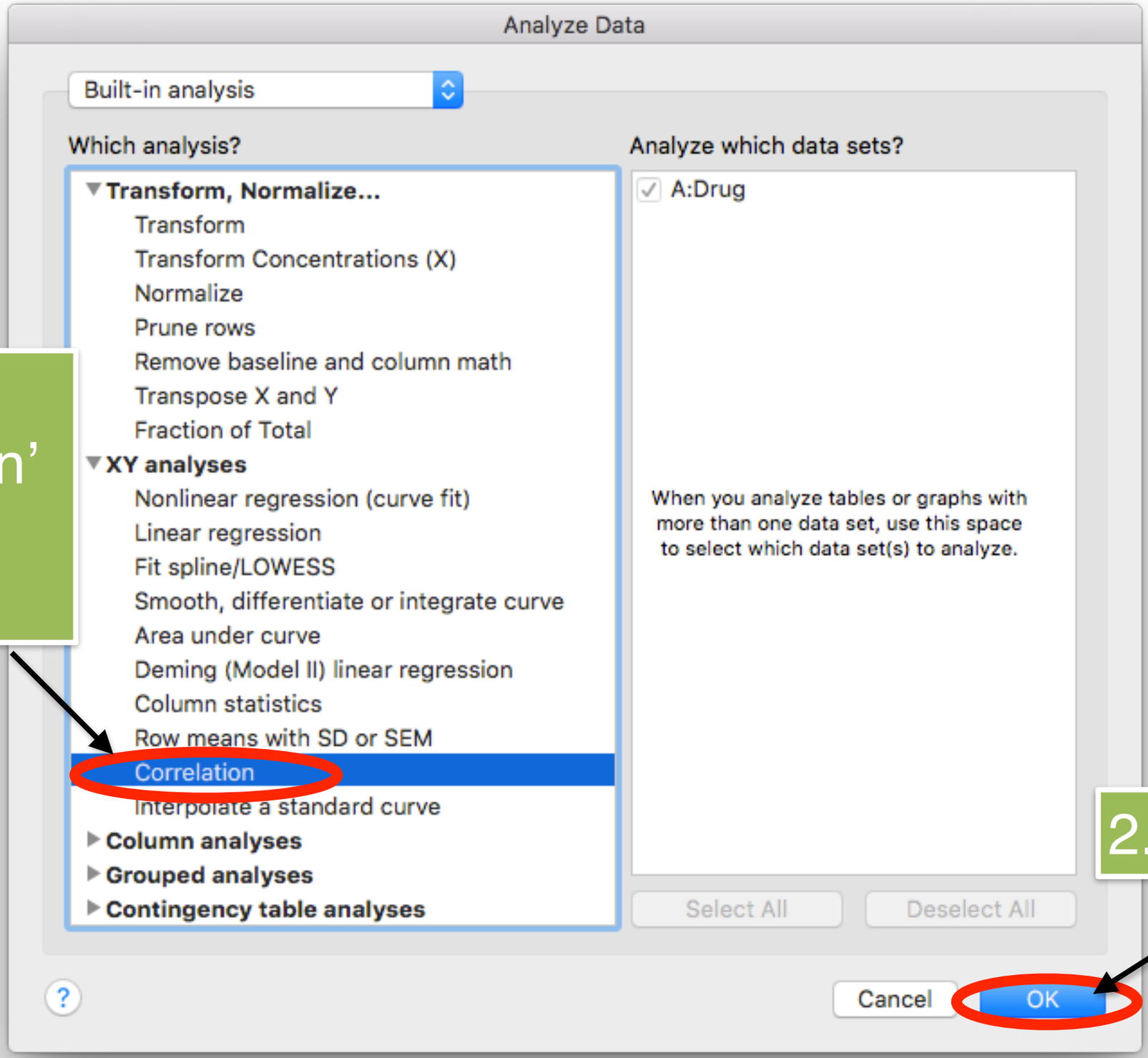
New Analyze

Family
Search results
Data Tables
drugConc
Info
Project info 1
Results
Graphs
drugConc
Layouts

		X	Group A	Group B	Group C	Group D	Group E	Group F	Group G	Group H	Group I	Grou
		Height	Drug	Title	Title	Title	Title	Title	Title	Title	Title	Titl
		X	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
1	Title	1.70	1.66									
2	Title	2.31	1.34									
3	Title	2.89	1.27									
4	Title	1.30	1.61									
5	Title	3.21	1.17									
6	Title	1.84	1.73									
7	Title	3.27	1.17									
8	Title	4.21	1.19									
9	Title	1.32	1.93									
10	Title	3.67	1.10									
11	Title	2.78	1.37									
12	Title	3.71	1.19									
13	Title	3.23	1.27									
14	Title	3.29	0.85									
15	Title	3.46	1.16									
16	Title	3.95	1.14									
17	Title	1.70	1.25									
18	Title	2.92	1.49									
19	Title	2.67	1.17									
20	Title	3.02	1.16									
21	Title	2.37	1.75									
22	Title	2.64	1.36									

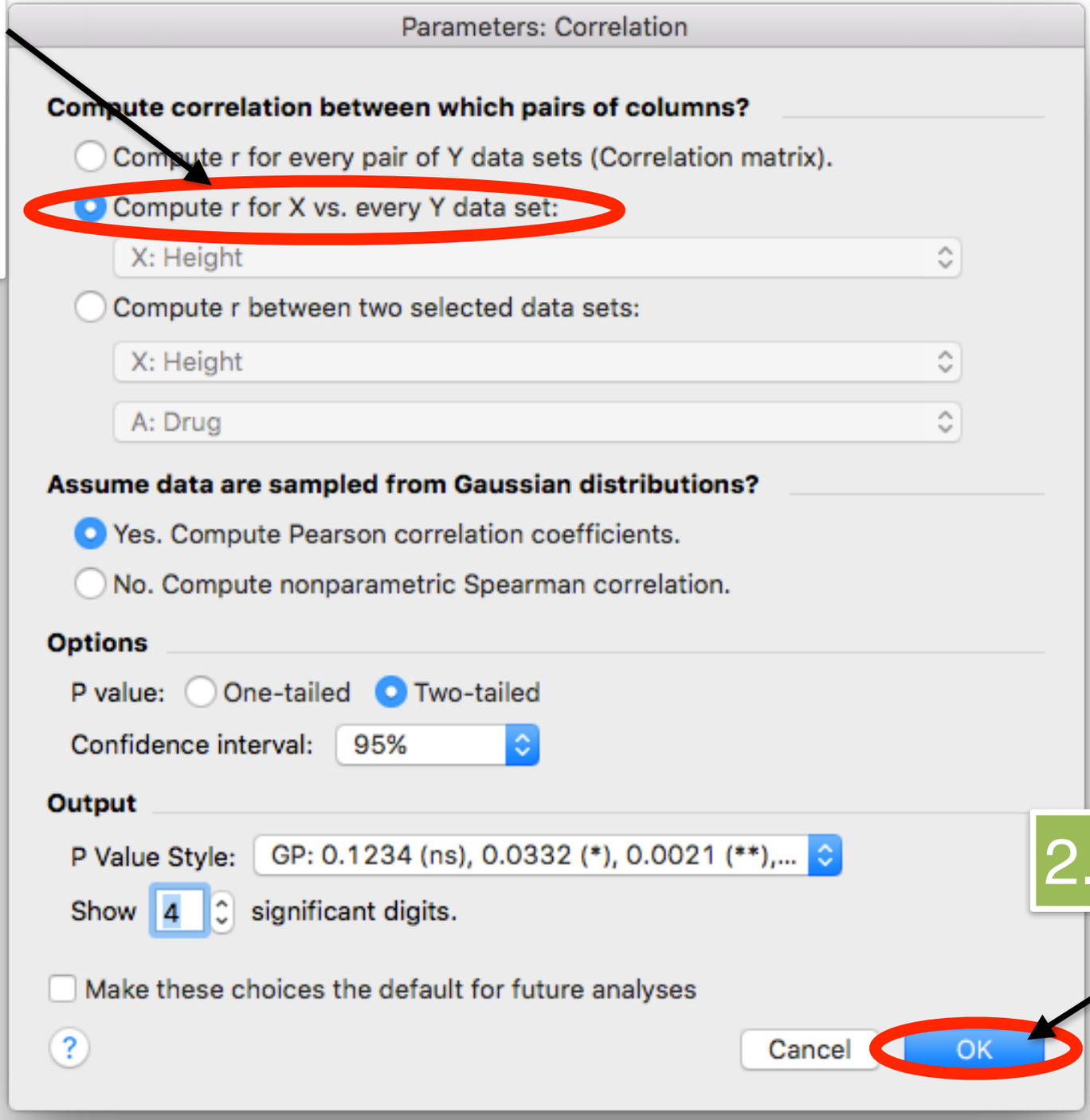
drugConc Row 7, Column B

1.Select ‘Correlation’ from XY analyses



2.Select ‘OK’

1. Select ‘Compute r for X vs. every Y data set’



2. Select ‘OK’

Untitled — Edited

File Sheet Undo Clipboard Analysis Interpret Change Draw Write Text Export Print Send LA Help

New Analyze 12 Arial

GraphPad PRISM

- Family
- Search results
- Data Tables
  - drugConc
- Info
  - Project info 1
- Results
  - Correlation of drugConc
- Graphs
  - drugConc
- Layouts

	Correlation	A	B	C	D	E	F	G	H
		Height vs. Drug	Title	Title	Title	Title	Title	Title	Title
		Y	Y	Y	Y	Y	Y	Y	Y
1	Pearson r								
2	r	-0.7767							
3	95% confidence interval	-0.8986 to -0.5437							
4	R squared	0.6033							
5									
6	P value								
7	P (two-tailed)	<0.0001							
8	P value summary	****							
9	Significant? (alpha = 0.05)	Yes							
10									
11	Number of XY Pairs	24							
12									
13									
14									
15									
16									
17									
18									
19									
20									

Correlation of drugConc | Tabular results

What did we learn?

- Correlation coefficients measures the magnitude and direction of the linear association between two factors.

- It is essential that you graph the data using a scatterplot prior to conducting a correlation analysis to check for a radical deviation from a linear relationship