Chapter 5: Sampling From Populations: The Standard Error of the Mean

TXCL7565/PHSC7565

What This Chapter Covers

- Samples and populations
- Sampling Error
- Types of sampling error
- Factors contributing to extent of random sampling error
- Standard error of the mean
- SEM in GraphPad

SAMPLES AND POPULATIONS

Samples and Populations

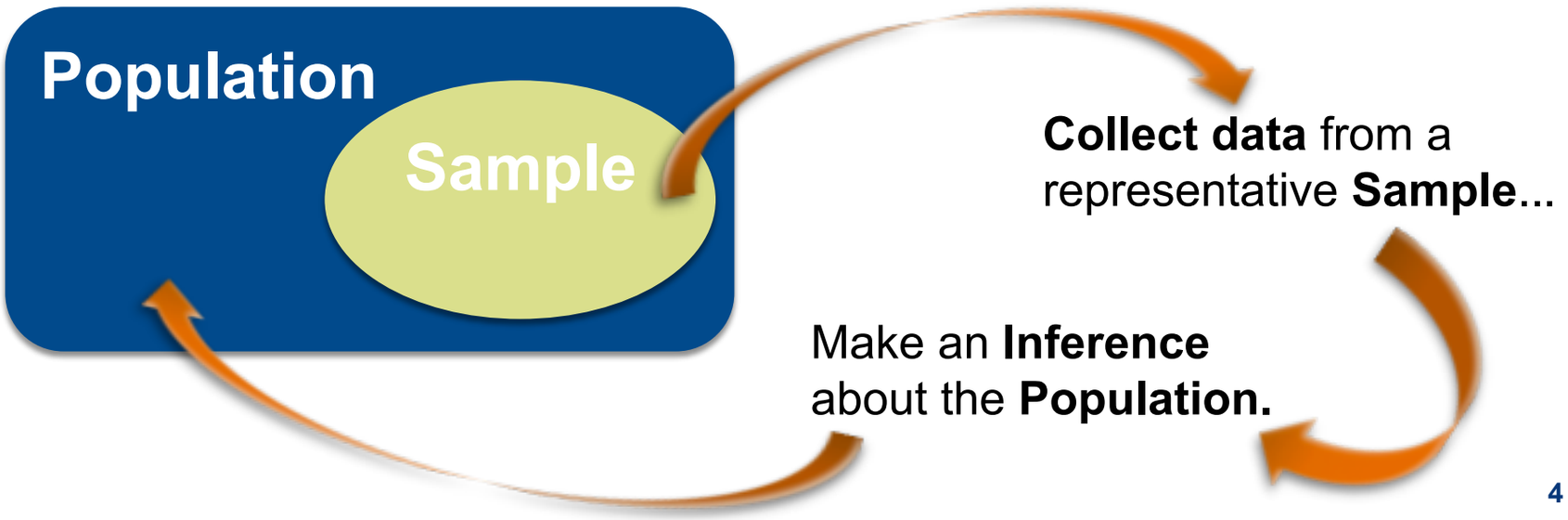
Good research includes a clear definition of the target group of individuals (or objects/animals/cells) about whom we aim to draw a conclusion.

- **Population** - the complete collection of individuals about whom we wish to draw some conclusion.

- **Sample** - a random selection of individuals from the population we wish to study

From Samples to Populations

Scope of Inference - the group of individuals to whom the statistical conclusions can be extended



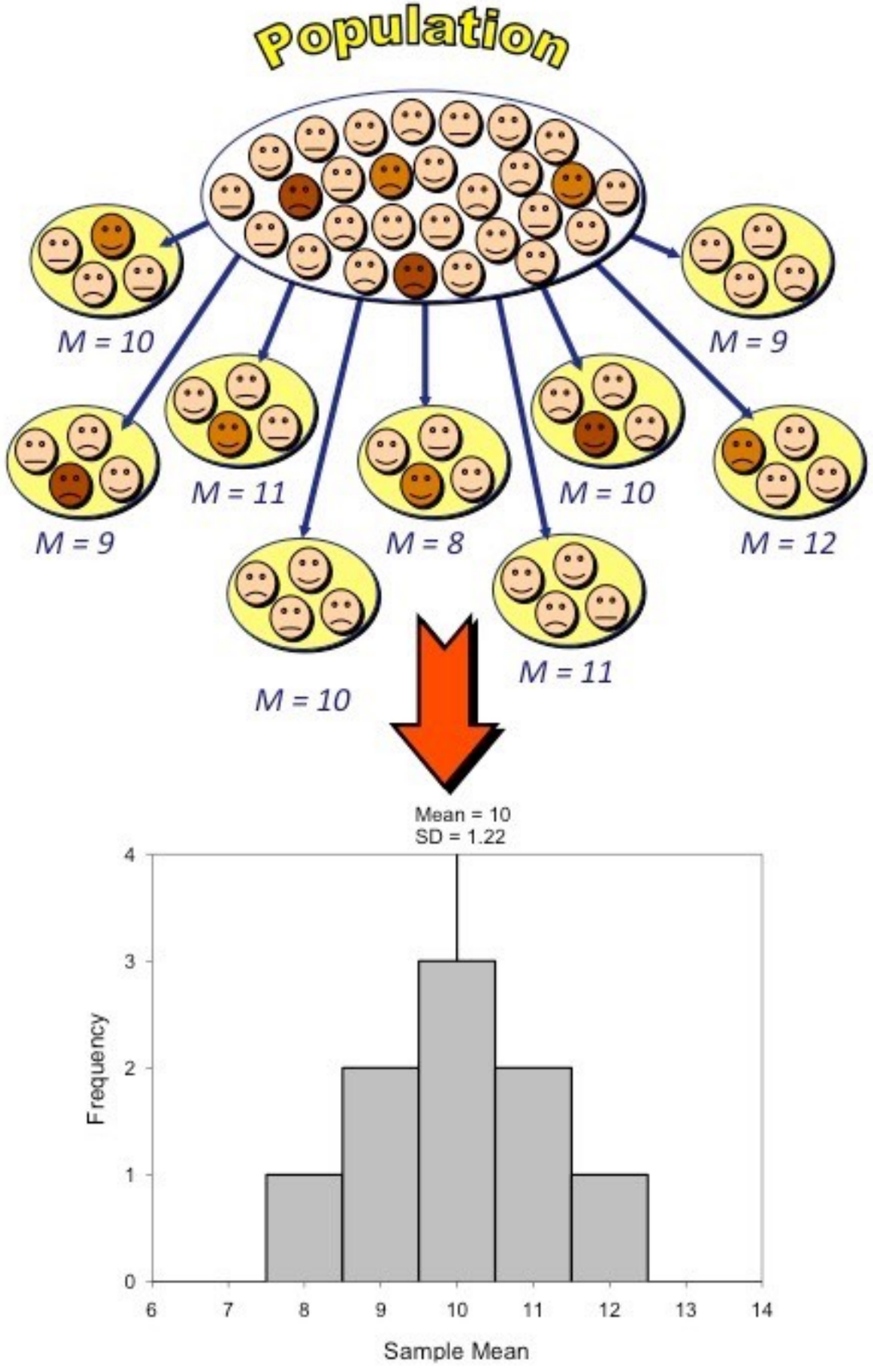
SAMPLING ERROR

Sampling Distribution of Sample Mean

- **Sampling error** is the concept that (random) samples from the same population will differ because they contain different members of the population.

- **Sampling distribution** is the frequency distribution of the statistics resulting from all possible samples (of a certain size n) from the same population.

- Example: Suppose we could get ratings of all statistics instructors in the world and that the population mean rating (on a scale from 1 to 20) is mean=10. We take several random samples of size n=4 and calculate the sample mean. What is the distribution of the sample means?


Population
M = 10
M = 9
M = 11
M = 10
M = 9
M = 8
M = 12
M = 10
M = 11
Mean = 10
SD = 1.22
Frequency
Sample Mean
0
1
2
3
4
6
7
8
9
10
11
12
13
14

TYPES OF SAMPLING ERROR

Bias: Systematic Error

A consistent form of mis-estimation of the mean. Either most such samples would over-estimate the value or most would under-estimate the value.

Principles of Bias

1. If we were to repeat the same sampling procedure several times, we could pretty much guarantee that we would make an error in the same direction every time.

2. Bias arises from flaws in our experimental design

3. We can remove the bias by improving our experimental design

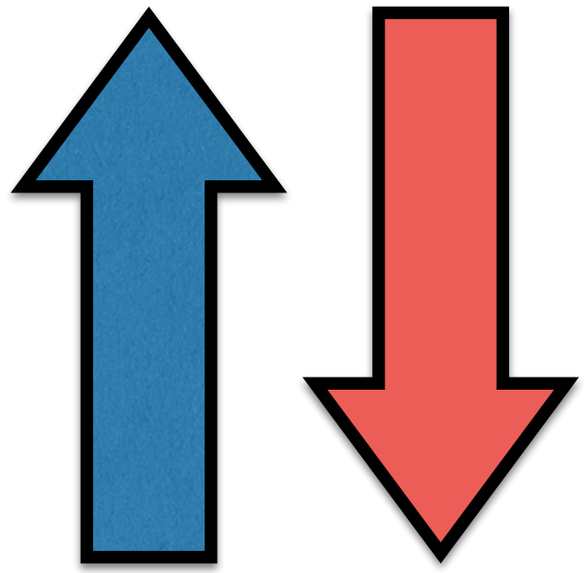
Random Error

Any given sample has an equal chance of under- or over-estimating the population mean value

FACTORS CONTRIBUTING TO THE EXTENT OF RANDOM SAMPLING ERROR WHEN ESTIMATING A POPULATION MEAN

Sample Size

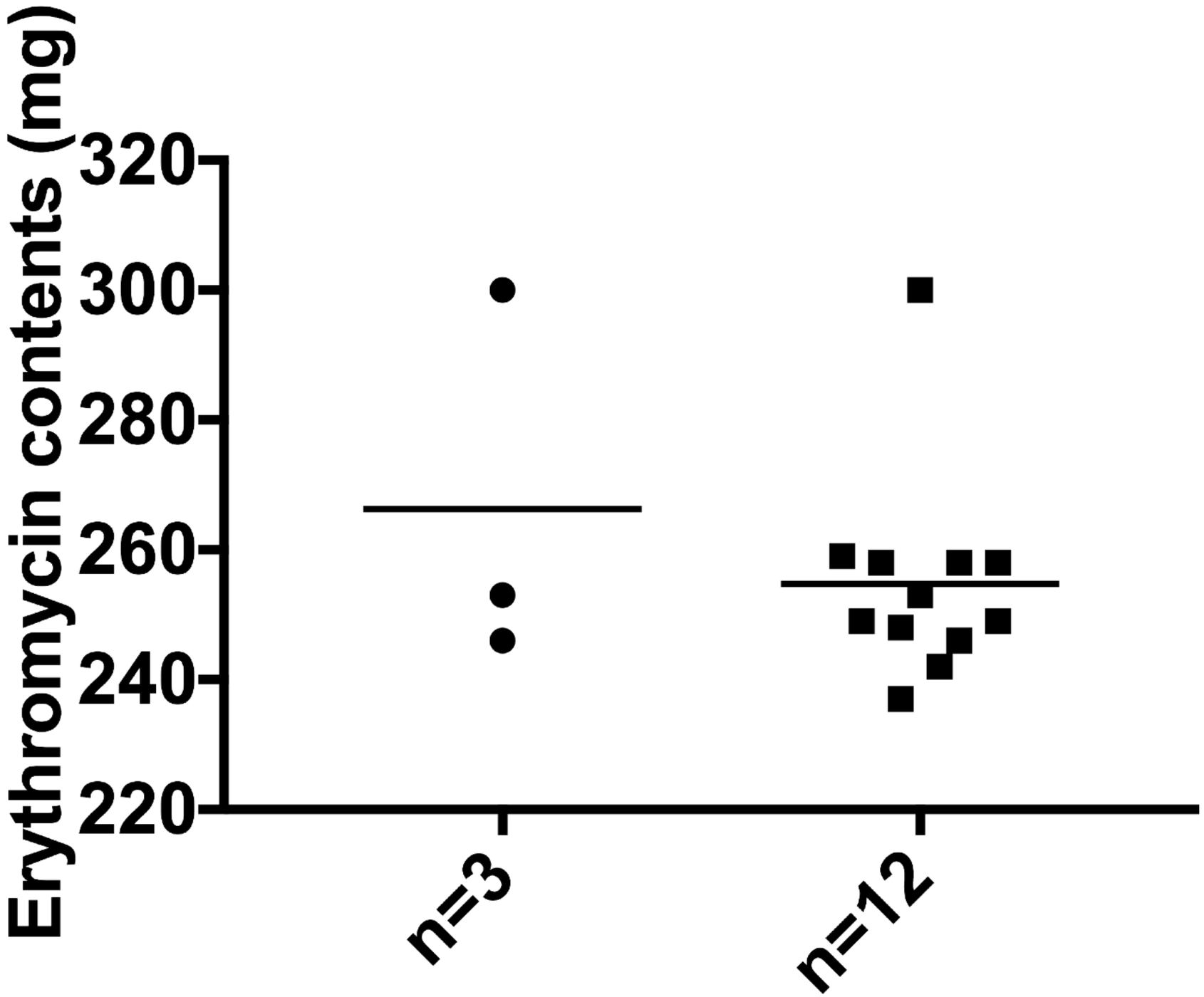
As sample size increases…



The influence of an outlier on the mean is reduced

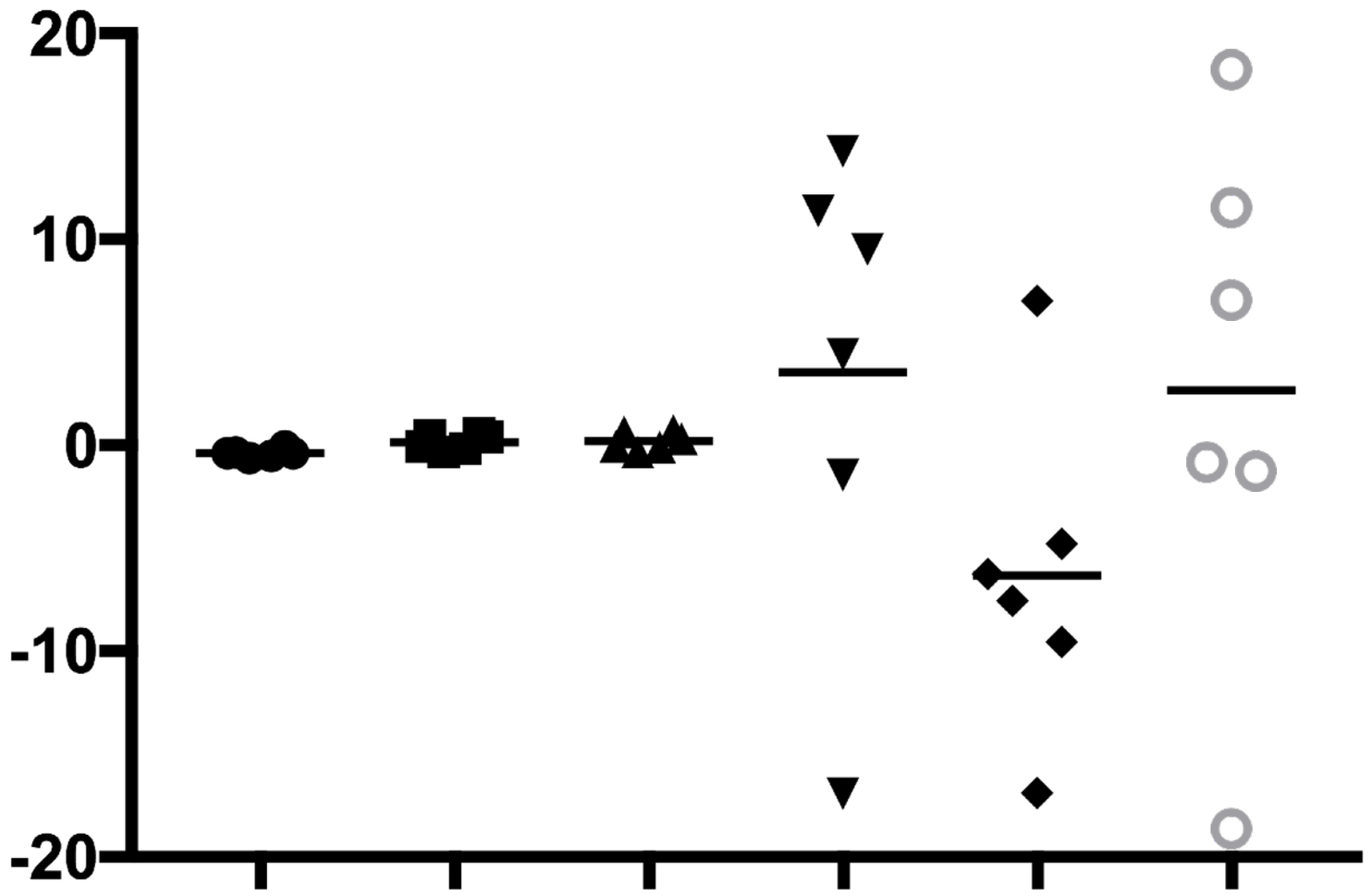
RANDOM ERROR IS REDUCED

Sample Size



Variability within the data

Sample means based on highly variable data are themselves rather variable and may provide a poor reflection of the true situation.



Standard Error of the Mean

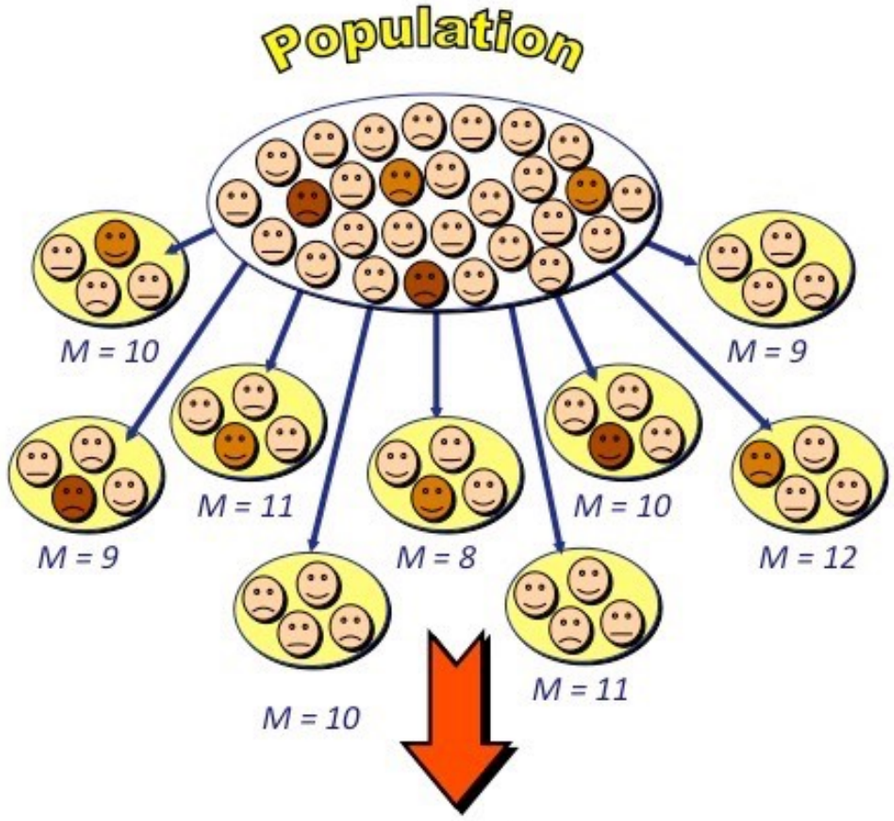
Standard Error of the Mean

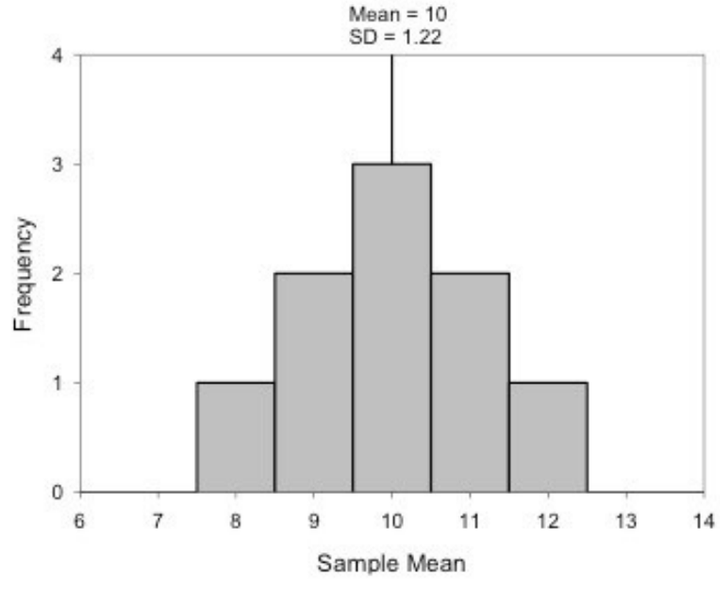
- **Standard error of the mean** (SEM) - estimate of the likely sampling error that should be anticipated, given the size of the sample and the variability among the data being sampled

- SEM = generalization of the accuracy our mean estimate

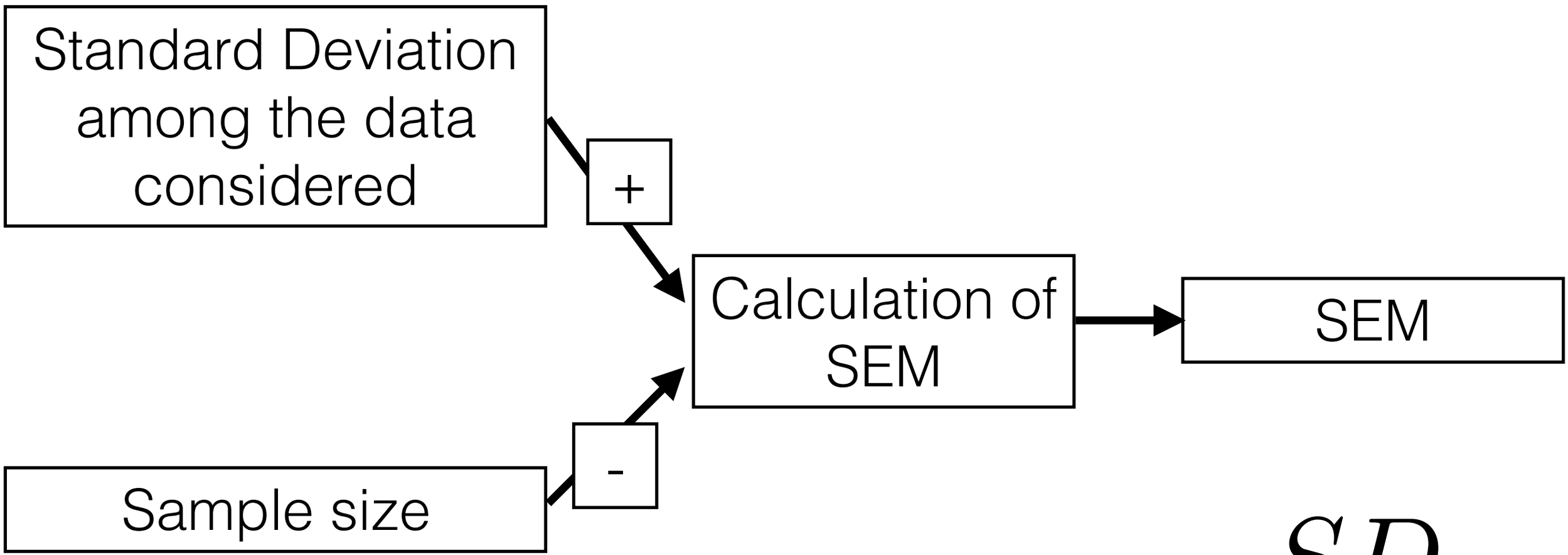
Technical Definition of SEM

The SD among (hypothetical) repeated sample means from the same population.





Calculation of the SEM



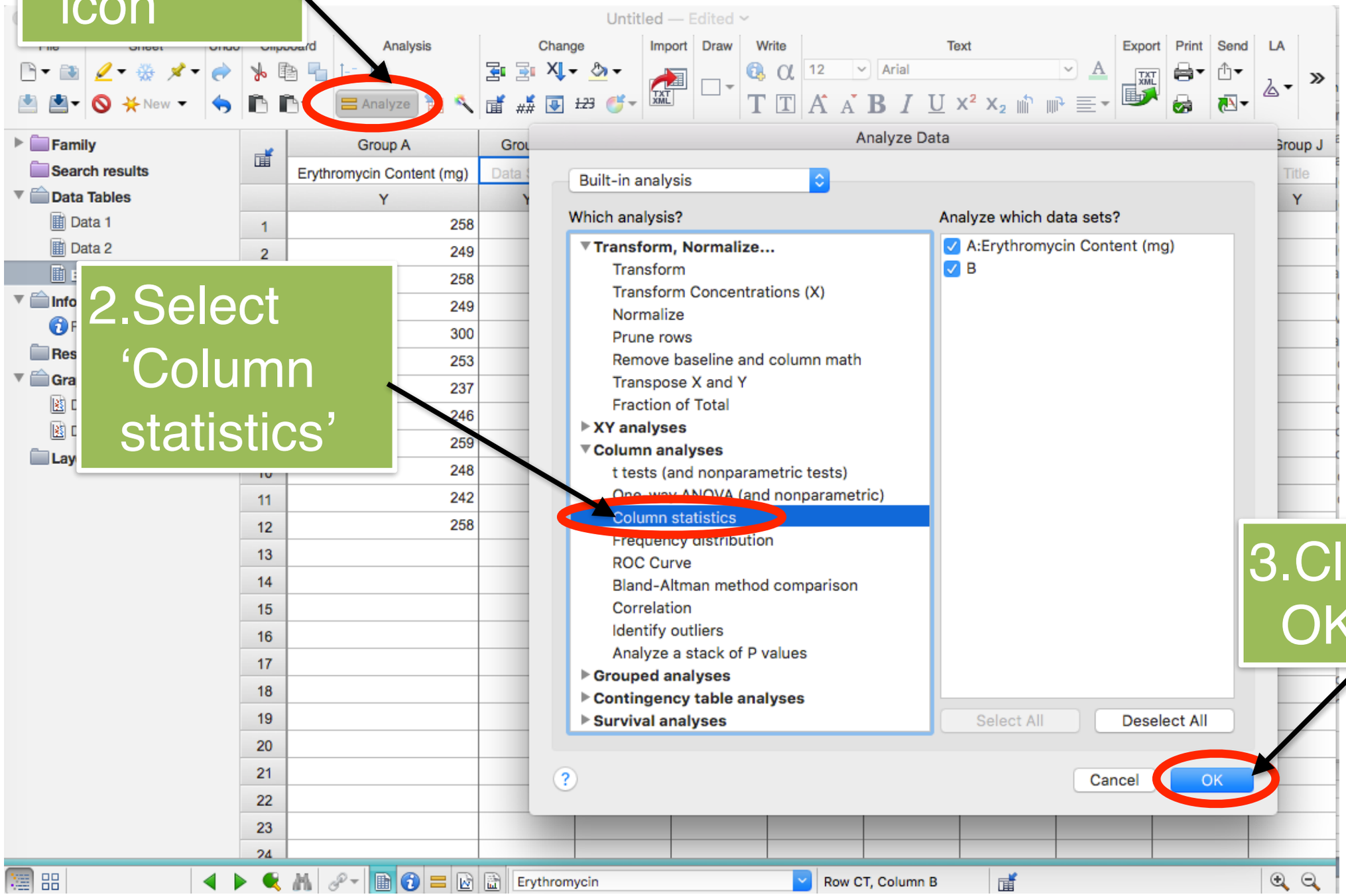
$$SEM = \frac{SD}{\sqrt{n}}$$

SEM in GraphPad

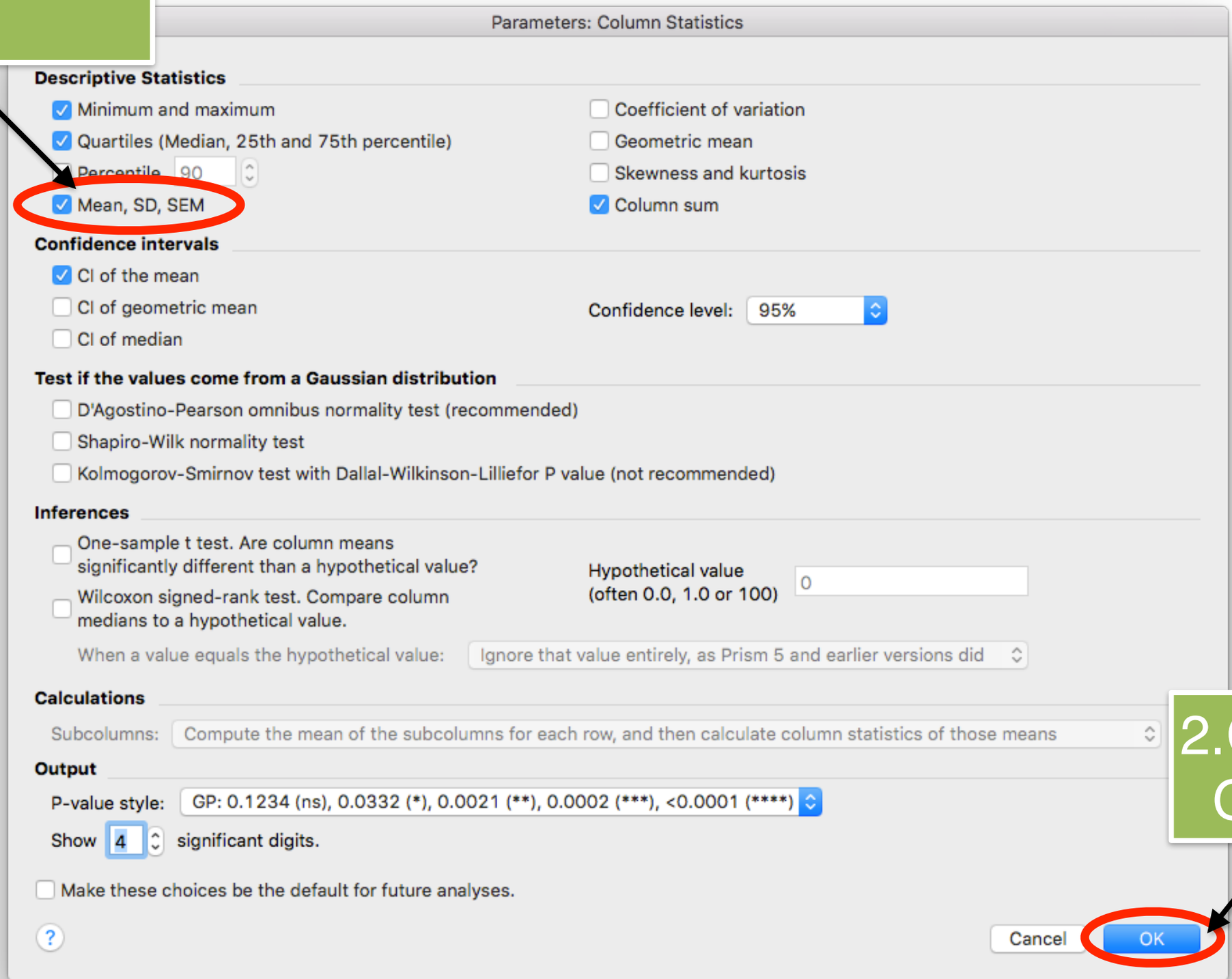
1.Click ‘Analyze icon

2.Select ‘Column statistics’

3.Click OK



1.Select ‘Mean, SD, SEM’



2.Click OK

Untitled — Edited

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

- Family
- Search results
- Data Tables
  - Data 1
  - Data 2
  - Erythromycin
- Info
  - Project info 1
- Results
  - Col Stats of Erythromycin
- Graphs
  - Data 1
  - Data 2
- Layouts

Col Stats		A	B	C	D	E	F	G	H
		Erythromycin Content (mg)	Title	Title	Title	Title	Title	Title	Title
		Y	Y	Y	Y	Y	Y	Y	Y
1	Number of values	12							
2									
3	Minimum	237							
4	25% Percentile	246.5							
5	Median	251							
6	75% Percentile	258							
7	Maximum	300							
8									
9	Mean	254.8							
10	Std. Deviation	15.86							
11	Std. Error of Mean	4.578							
12									
13	Lower 95% CI of mean	244.7							
14	Upper 95% CI of mean	264.8							
15									
16	Sum	3057							
17									
18									
19									
20									
21									
22									
23									
24									

Col Stats of Erythromycin | Column statistics

What did we learn?

- If samples are randomly chosen from the entire population, results can be applied to the entire population.

- Bias/systematic error is often a result of poor study design.

- All experiments are subject to random error which is just as likely to cause over-estimation as it is to cause under-estimation.

- Sample size and variability within the data contribute to the extent of random sampling error when estimating a population mean.

- The standard error of the mean is a estimate of how accurately we have estimated the population mean.