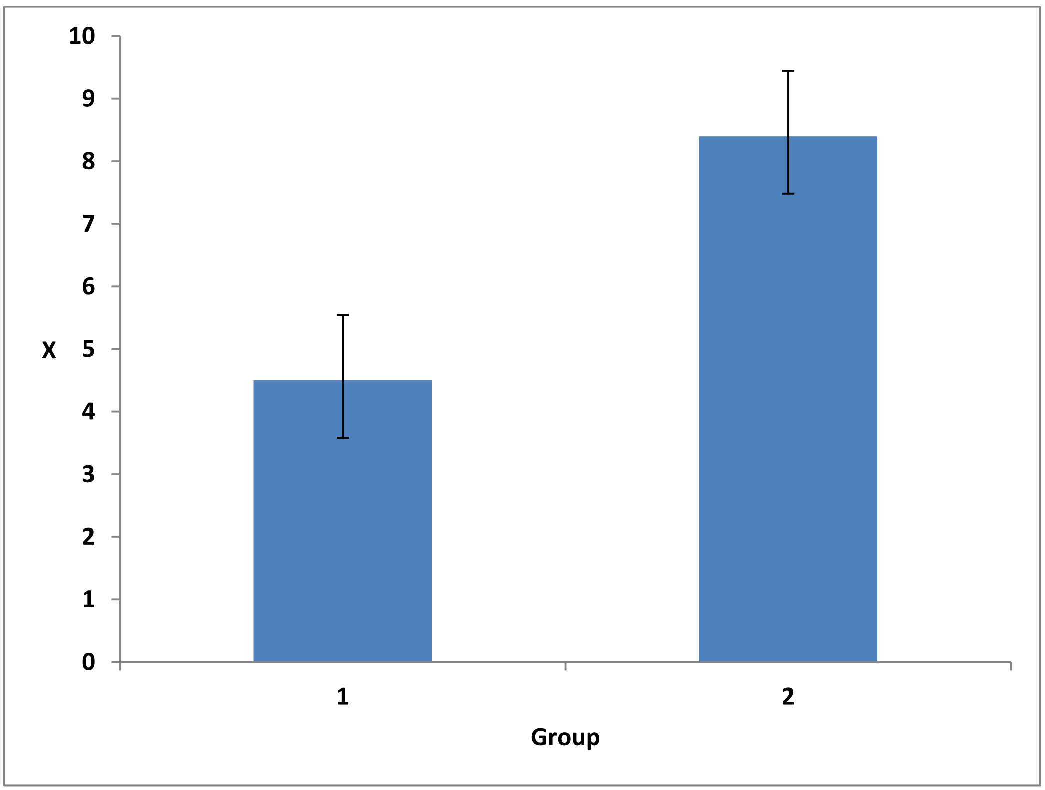
STANDARD DEVIATION vs. STANDARD ERROR

Issue 1: Standard Deviation (SD) and Standard Error (SE) are often used incorrectly or inconsistently.

Should you present SD or SE?

Figure 1: A common bar plot depicting the mean value of X for Groups 1 & 2, plus error bars



This is a common figure presented in many manuscripts. The height of the bar represents the average observed in each group. But what are the "error" bars around the mean values summarizing?

First we need to clearly define standard deviation and standard error:

Standard deviation (SD) is the average deviation from the mean *in your observed data. It is an index of how individual data points are scattered.*

We compute SD so we can make inferences about the **true** population standard deviation. If it's large, then we know values will vary a lot around the mean.

Example: For simplicity, let's say you have three data points (the mean value is 1+2+6/3=3.0):

X	Residual (deviation from the observed mean)	Squared residual from average
1	1-3=-2	4
2	2-3=-1	1
6	6-3=+3	9

The sum of the squared deviations (we square it so negative residuals do not cancel out positive ones) is 14. To get the SD we simply divide 14 by 2 and then take the square root.

The square root of 14/2 is **2.65 – this is the SD**

Technical detail: *Why do we divide by n-1 instead of the sample size n to get the "average" deviation?* This "corrected" version is used because we do not know the population mean and have substituted the sample mean into the formula instead. If we divide by n we will **always** underestimate the true population variance (unless our sample mean is the same as the population mean).

Standard error of the mean (SEM):

The standard error (to be more precise, the standard error *of the mean*) is a property of our estimate of the mean. The SEM is equal to the SD divided by the square root of n.

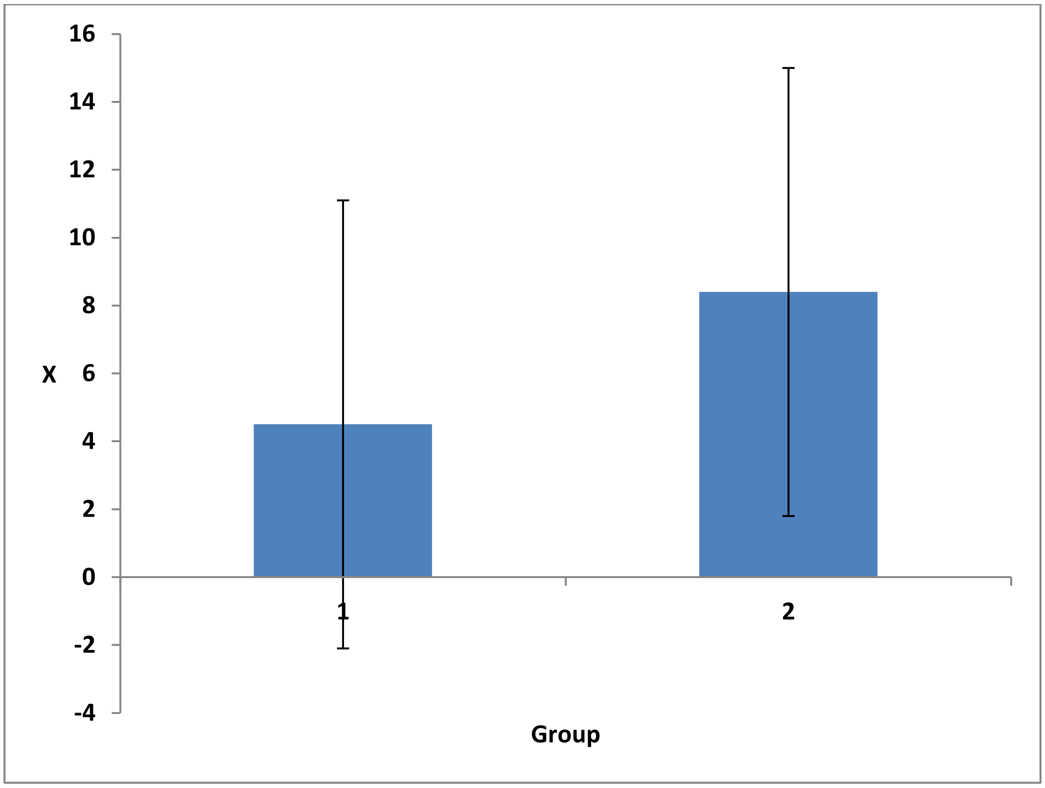
SEM = 2.65/√3 = 1.5275

This quantity tells us how our estimate of the mean will vary from sample to sample (*these are theoretical samples, if we could redo our exact study many times and compute the sample mean over and over again and look at how it varies*). It is a summary of how precise our estimate is (as we expect, as sample size increases, our ability to estimate the mean precisely improves, so the SEM decreases). See the difference? **SD is concerned with the scatter of individual data points in the population, while the SEM is concerned with the variability of our estimate of the mean.**

It is clear that the SEM will always be smaller than the SD (which is not a function of sample size).

In Figure 1, the authors are presenting the mean +/- 1SEM. So their focus is clearly on the precision of their estimates of the mean for groups 1 and 2. *However +/- 1SEM does not translate into anything meaningful as far as comparing the means between the two groups. The fact that these bars do not overlap tells us very little about whether the group means significantly differ.*

If the authors had presented the mean +/- 2SD, the figure would have looked like this:



This figure is helpful if the focus is on how varied the data values are around the mean. It shows that there is a substantial amount of variability in the data. Under normality assumptions, we may assume that about 95% of the observed data points for x will be contained within this interval.

HOWEVER, again we cannot look at this figure and make any sort of statement about statistical significance! Standard deviation reflects individual variation of data points, not the statistical significance of the difference between sample means.

Some helpful hints when presenting data with error bars:

1. Have you been clear about what it actually displayed? If using SE – is it clearly labelled as such? Presenting +/- 1SEM or +/- 1SD is confusing – stick with +/- 2 SD or +/- 2SEM (even though this results in wider error intervals)

2. Think about what are you trying to describe – is it the population scatter or the quality of your estimate? If you are simply trying to show how the data are scattered, present the mean +/- 2SD. Note: these will be the widest intervals. If you want to show how precisely your mean is estimated, present the mean +/- 2SEM

3. If you are most interested in the **differences between group means**, showing the interval that reflects the *mean +/- 2 SEM* will be very helpful, but may still not be enough information to make any sort of statement about those group differences (see Table 1 below). The best approach here would be to focus on confidence intervals for the differences in mean (presented in another tutorial).

We often look at these "error" bars to eyeball whether there is a statistically significant difference between groups. Is this a valid approach?

The answer is NO, in general you should not rely on the overlap or lack of overlap between error bars to informally test for statistically significant differences between groups.

Table 1: Guidelines for interpreting error bars

Error bar type	Overlapping error bars	Non-overlapping error bars
+/- 2 SD	Don't know	Don't know
+/- 2 SEM	sample means not significantly different	Don't know