On standard deviation and standard error of the mean

This little document is meant to explain the differences between the standard deviation and standard error of the mean (SEM) and what they imply. This is relevant if you want to know what you are communicating to the reader. Scientific publications are in-effect just means to convey information. Pictures and graphs are inevitably subservient to this purpose. When drafting a graph, a great care must be taken to ensure that the reader can come to the relevant conclusions on their own from the data presented. In most cases this inarguably means presenting your data as clearly and openly as possible.

While a lot can be said about designing your graphs, the aim of this document is to discuss the difference between standard deviation and standard error and what do they ‘mean’ and ‘convey’ to your reader. Both of these parameters have their uses and implications, but they are not comparable as they are meant to portray different things.

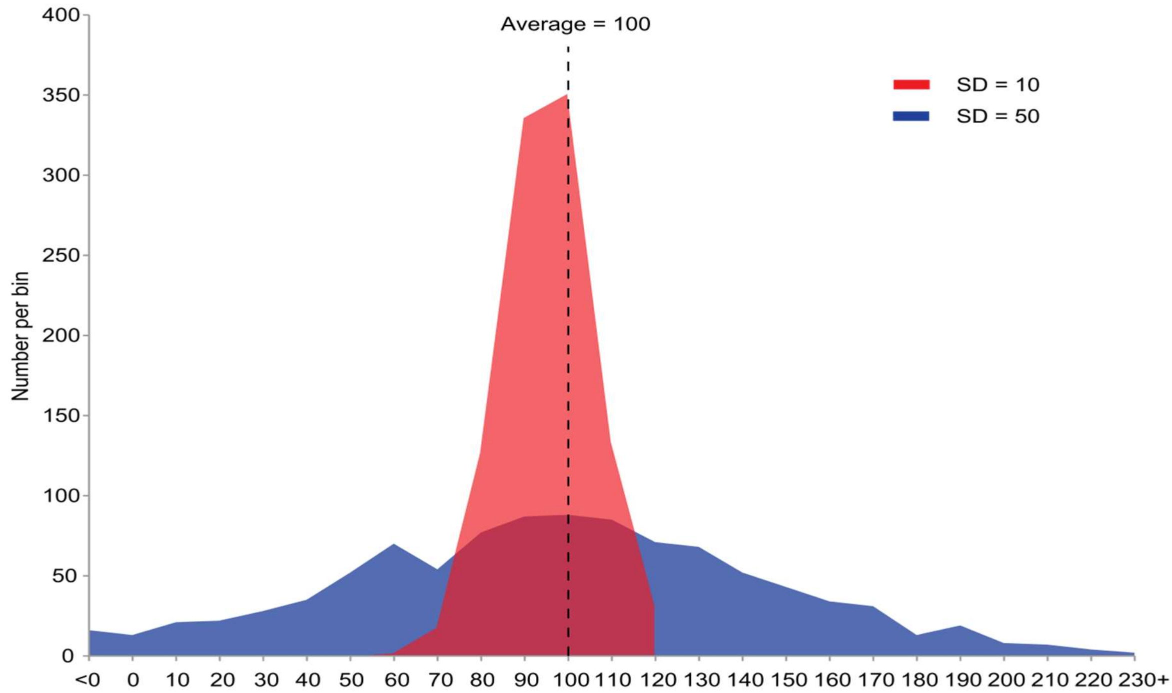
Standard deviation

Standard deviation is formulated as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}},$$

where x_i is your i:th measurement, x_i is your average and N is the number of elements in your sample. This document is not going to delve into ‘why’ the equation is as it is, only what it represents.

As the name implies, the standard deviation presents information on how your data is distributed around the mean. The larger the standard deviance, the more widely your results are distributed.



Example of samples from two populations with the same mean but different standard deviations. Red population has mean 100 and SD 10; blue population has mean 100 and SD 50.

Sometimes the property of you measurement is scientifically relevant. Usually (but not always) you are comparing two samples to each other and you want to argue that they are either different or similar. When you wish to make a statement regarding the nature of the distribution you have measured you should use standard deviation as your measure of uncertainty.

Standard deviation (be definition) is a measure of how much on average a measurement deviates from the mean. The higher the standard deviation, the more your measurement is expected to differ from the mean.

If and if your data is normally distributed, mean and standard deviation will describe your data perfectly. That is to say – if your data is normally distributed you can use mean and standard deviation in place of the data you have gathered, allowing you to employ parametric statistical tests (e.g. t-test).

Standard error of the mean

Standard error of the mean (SEM) is a slightly misleading term. Since SEM is calculated from your measurements and as far as you are concerned, all of your measurements are true results that reflect the reality being studied, any parameter deducted from your data is not 'error' as far as the word is commonly defined. So calling this parameter 'standard error' is rather misleading.

SEM is given by:

$SEM = \frac{\sigma}{\sqrt{N}}$, where σ is the standard deviation and N is the number of measurements in your data set. This can be deducted by calculating the variance of the mean, which again demonstrates that this is not really 'error' in any sense. However, SEM is <u>a measure of mean</u>. This is something to keep in mind as we dwell deeper into the topic.

Something to consider is 'confidence interval'.

Confidence interval

Confidence interval proposes a range of plausible values for an unknown variable – usually the mean. The given interval has an associated **confidence level** that the true parameter is in the proposed range.

Various interpretations of a **confidence interval** can be given. I am going with one that I think that makes the most sense.

"If confidence intervals are constructed using a given confidence level in an infinite number of independent experiments, the proportion of those intervals that contain the true value of the parameter will match the confidence level."

A particular confidence interval of 95% calculated from an experiment does not mean that there is a 95% probability of a sample parameter from a repeat of the experiment falling within this interval. Once the interval is calculated the true mean is either within it or it isn't.

The 95% probability relates to the **reliability of the estimation procedure**, not to a specific calculated interval.

The desired level of confidence is set by the researcher (notably this is not determined by data). Most commonly, the 95% confidence level is used. Factors affecting the width of the confidence interval include the size of the sample, the confidence level, and the variability in the sample. A larger sample size normally will lead to a better estimate of the population parameter.

Confidence Interval is calculated for any desired degree of confidence by using sample size and variability (SD) of the sample, although 95% CIs are by far the most used. Confidence interval for a data following **normal distributio**n is given by:

$$CI = \bar{X} \pm t\frac{\sigma}{\sqrt{N}},$$

where $\bar{X}$ is mean, t is value of t-distribution with the specific level of confidence and the term $\frac{\sigma}{\sqrt{N}}$ is SEM as calculated before.

Note that we write t = 1 the equation yields

$CI = \bar{X} \pm \frac{\sigma}{\sqrt{N}},$ which is basically mean $\pm$ SEM. Meaning of which will become relevant later.

The t-value must be checked from a table (see the next page) with degrees of freedom (df) given by N-1.

If we want 95% confidence interval for N=10 we look at Two-Tail = 0.05 and df = 9 getting t= 2.262.

Notably, t = 1 (i.e. when SEM) equals to 68.26% confidence interval. This is something to consider.

Table T Critical Values of the t Distribution

df	One-Tail = .4 Two-Tail = .8	.25 .5	.1 .2	.05 .1	.025 .05	.01 .02	.005 .01	.0025 .005	.001 .002	.0005 .001
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745

t-value table used to compute / check the confidence intervals.

This is to say that when you report SEM you are reporting your uncertainty regarding the true mean. SEM gets smaller with N as more information you have regarding your dataset, the more reliably you can estimate your mean. No claims are made about the distribution of the data.

Standard deviation expresses how your data is distributed around the mean, i.e. the shape of the distribution measured as long as your data is normal.

Because you can never measure the whole population, you have to rely on taking random samples from that distribution to estimate how far the sample mean is from the true population mean. If this is your goal, then you calculate the standard error of the mean.

One standard error of the mean is then the interval in which the true population mean would fall 68% of the time if sampling was repeated over and over again. Usually in statistics a 95% confidence interval is used, which you can calculate as shown above. Given the formula for the standard error of the mean, it is also apparent that if the sample size goes up, the interval tends to zero and you are closing in on your population mean μ. Thus, the standard error of the mean is a tool in **inferential statistics**, that is inferring from the distribution of a random sample (*observed data*) to properties of an underlying unknown distribution, or the population.

The standard deviation on the other hand is used to describe the variability in the observed data only (i.e. the sample) without making any inferences with respect to properties of the underlying unknown distribution. The standard deviation is commonly used in **descriptive statistics**.

Now depending on whether you want to *infer properties of an unknown distribution from a random sample* or whether you want to simply *describe the variability in your sample*, you should use the **standard error of the mean** and the **standard deviation**, respectively.