STANDARD ERROR:

The standard error quantifies the variation in the means from multiple set of measurements. The standard error can be estimated using a statistic based on the standard deviation s of the sample values, and the sample size n:

$$\boxed{\text{Standard error} = SE = \frac{s}{\sqrt{n}}}$$

where,
s = Standard Deviation.
n = no. of samples. (or) measurements.

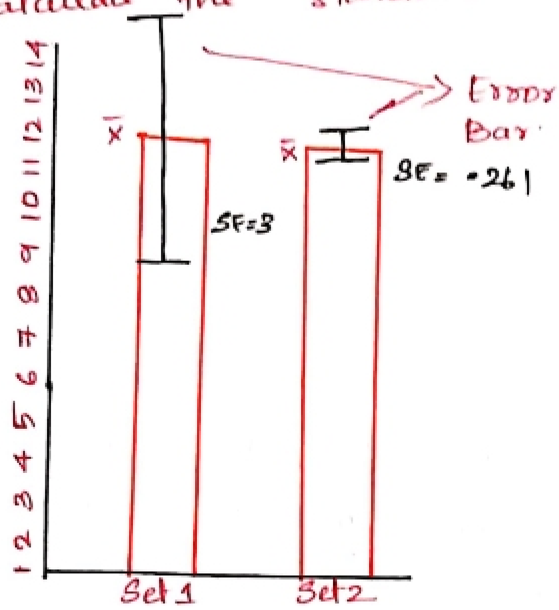
As the sample size increases, the standard error decreases. The relationship between standard error and sample size is sometimes referred to as the square-root of n rule: in order to reduce the standard error by a factor of 2, the sample size must be increased by a factor of 4.

Consider the following approach to measure standard error:

* Collect a no. of brand new samples from the population.
* For each new sample, calculate the statistic (eg. mean).
* Calculate the standard deviation of the statistics computed in step 2; use this as your estimate of standard error.

Semester: IV Subject: Statistics for AIDS Academic Year: 2023-2024

Example -
Given a bar graph with 2 set of measurements. Calculate the standard error.



Set 1	Set 2
9	10.9
15	11.9
	12.2
	12.2
	12.9
	12.6
	12.3
	12.3
	12.5
	10.2

Step 1 : Calculate Mean.

Data Set 1 $= \frac{9+15}{2} = 12$

Data set 2 $= \frac{10.9+11.9+12.2+12.2+12.9+12.6+12.3+12.3+12.5+10.2}{10}$

$\boxed{\bar{X} = 12}$

Data set 2 : SD = 0.825

Data set 1 : SD = 4.242.

SE for Set 2 $= \frac{S}{\sqrt{n}}$

$= \frac{0.825}{\sqrt{10}} = \frac{.826}{3.16}$

$\boxed{SE = 0.261}$

SE for Set 1 $= \frac{S}{\sqrt{n}}$

$= \frac{4.242}{\sqrt{2}} = \frac{4.242}{1.414}$

$\boxed{S.E = 3}$

Semester : V Subject: Statistics for AI&DS Academic Year: 2023 2024

The graph clearly shows that set 1 data has more error when compared to set 2 data. This is how we estimate the standard error.

BOOTSTRAP:

One easy and effective way to estimate the sampling distribution of a statistic is to draw additional samples, with replacement, from the sample itself and recalculate the statistics or model for each resample. This procedure is called the bootstrap and it does not necessarily involve any assumptions about the data.

Conceptually, you can imagine the bootstrap as replicating the original sample thousands or millions of times so that you have a hypothetical population.

In practice, it is not necessary to actually replicate the sample a huge number of times. We simply replace each observation after each draw; that is we sample with replacement.

In this way we effectively create an infinite population.