Standard Deviation vs. Standard Error of the Mean

Sujan Bandyopadhyay

January 16, 2018

A Common Point of Confusion

- Standard Deviation and Standard Error of the Mean are two statistical measures that are often mixed up or confused for each other
- People tend to incorrectly use them interchangeably and often are unsure of the difference between the two
- This presentation provides a explanation of these two concepts and helps the reader understand the difference betweeen the two statistical measures

Standard Deviation

- Standard Deviation is a measure of dispersion within a data set
- Specifically, it measures the dispersion of data from the mean of this data set
- Suppose we have the entire population, then given the μ is the population mean and N is the number of observations, σ (standard deviation) is:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{1}$$

- When we have a sample, we use a technique known as Bessel's Correction to adjust downward bias, in calculating the sample standard deviation s

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2}$$

- It is important to note that both σ and s are applicable to the population

n	X	Mean(X)	(X - Mean(X))^2
1	0.28	0.50	0.05
2	0.49	0.50	0.00
3	0.24	0.50	0.07
4	0.88	0.50	0.14
5	0.67	0.50	0.03
6	0.21	0.50	0.08
7	0.56	0.50	0.00
8	0.01	0.50	0.24
9	0.82	0.50	0.10
10	0.12	0.50	0.14
11	0.74	0.50	0.06
12	0.81	0.50	0.09
13	0.50	0.50	0.00
14	0.71	0.50	0.04
Total	7.04		1.05
Mean(X)	**7.04/14**	**Standard Deviation (X)**	**1.05/(14-1)**
Mean(X)	**0.50**	**Standard Deviation (X)**	**0.28**

Figure: Calculation of s

Standard Error of the Mean

- Standard Error of the Mean (SEM) is a measure of dispersion of the sample mean from the population mean
- Standard Error of a parameter is the standard deviation of the sampling distribution of a parameter
- Conceptually, SEM ($\sigma_{\bar{x}}$) can be calculated by calculating the standard deviations from multiple samples of a population

$$\sigma_{\bar{x}} = \sqrt{\frac{1}{S}\sum_{j=1}^{S}(\bar{x}_j - \hat{\bar{x}})^2} \tag{3}$$

where S is the number of samples, $\bar{x}_j$ is the sample mean of sample j, $\hat{\bar{x}}$ is the mean of sample means

n	X1	X2	X3	X4	X5
1	0.28	0.22	0.03	0.86	0.43
2	0.49	0.89	0.29	0.55	0.38
3	0.24	0.35	0.62	0.81	0.05
4	0.88	0.86	0.48	0.63	0.84
5	0.67	0.24	0.78	0.50	0.67
6	0.21	0.41	0.06	0.54	0.05
7	0.56	0.28	0.92	0.63	0.83
8	0.01	0.93	0.77	0.72	0.96
9	0.82	0.33	0.17	0.90	0.40
10	0.12	0.24	0.95	0.46	0.60
11	0.74	0.30	0.53	0.62	0.39
12	0.81	0.68	0.02	0.13	0.14
13	0.50	0.45	0.94	0.31	0.67
14	0.71	0.66	0.90	0.47	0.50
Mean	0.50	0.49	0.53	0.58	0.49

Mean of Means	0.52					Total
(Mean of Means - Mean(i))^2	0.00	0.00	0.00	0.00	0.00	0.0057

Standard Error of Mean	(0.0057/5)^0.5
	0.033881245

Figure: Calculation of $\sigma_{\bar{x}}$ (1)

- However, in real life we rarely have multiple samples of a population; we usually have to make do with just one sample
- For such scenarios, statisticians have found a way of calculating SEM with just one sample
- The formula can be derived by sum of uncorrelated variables formula (next page)

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{4}$$

where σ is the population standard deviation and n is the sample size

- When σ is unknown (as is often the case), we derive it using the sample standard deviation s; this is sometimes refered to as the estimated standard deviation of the mean

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{5}$$

Derivation of SEM Formula

Let $X1$, $X2$, $X3$, $X4$.... Xn be n independent observations from a population with mean μ and standard deviation σ, then the Variance can be calculated as

$$Var(\sum_{i=1}^{n} Xi) = Var(X1) + Var(X2) + Var(X3)..... = \sigma + \sigma + \sigma.... = n\sigma^2$$

Now we calculate the variance of the same set of observations, but divided by n (to account for the sample mean $\bar{x}$)

$$Var(\frac{\sum_{i=1}^{n} Xi}{n}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

And the standard deviation of this can be written as

$$\sigma_{\bar{x}} = \sqrt{Var(\frac{\sum_{i=1}^{n} Xi}{n})} = \frac{\sigma}{\sqrt{n}}$$

n	X	Mean(X)	(X - Mean(X))^2
1	0.28	0.50	0.05
2	0.49	0.50	0.00
3	0.24	0.50	0.07
4	0.88	0.50	0.14
5	0.67	0.50	0.03
6	0.21	0.50	0.08
7	0.56	0.50	0.00
8	0.01	0.50	0.24
9	0.82	0.50	0.10
10	0.12	0.50	0.14
11	0.74	0.50	0.06
12	0.81	0.50	0.09
13	0.50	0.50	0.00
14	0.71	0.50	0.04
Total	7.04		1.05
Mean(X)	**7.04/14**	**Standard Deviation (X)**	**1.05/(14-1)**
Mean(X)	**0.50**	**Standard Deviation (X)**	**0.28**
		Standard Error of the Mean	**0.28/14^0.5**
			0.08

Figure: Calculation of $\sigma_{\bar{x}}$ (2)

Some Intuition

- In the previous slide, we have calculated both standard deviation and standard error of the mean. How do we interpret these two measures?
- Let's say we have conducted an experiment by observing Raj's finishing time (in hours) of a certain distance [1]
- The Standard deviation of 0.28 tells us that if Raj was to run the distance again for a 100 times, we would expect that 95% of those times he would finish within $0.50 \pm 2 \times (0.28)$ hours
- On the other hand, the SEM is telling us that if we conduct the same experiment another 100 times, we would expect that the sample mean of these experiments would be within $0.50 \pm 2 \times (0.08)$ hours for 95% of those experiments

- **Which measure should we use?** Depends on the context

[1] Let's assume a normal distribution of running times. In a normal distribution 95% of observations lie within two standard deviations from the mean

Regressions

- In regressions, we use the standard error to determine the significance of coefficients

Source	SS	df	MS			
				Number of obs	=	74
				F(2, 71)	=	18.91
Model	220725280	2	110362640	Prob > F	=	0.0000
Residual	414340116	71	5835776.28	R-squared	=	0.3476
				Adj R-squared	=	0.3292
Total	635065396	73	8699525.97	Root MSE	=	2415.7

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
weight	4.699065	1.122339	4.19	0.000	2.461184	6.936946
length	-97.96031	39.1746	-2.50	0.015	-176.0722	-19.84838
_cons	10386.54	4308.159	2.41	0.019	1796.316	18976.76

Figure: Stata Output

- How is this calculated? Next page.

$$Y = X\beta + \epsilon$$

We know that

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\implies \hat{\beta} = (X'X)^{-1}X'(X\beta + \epsilon)$$

$$\therefore \beta - \hat{\beta} = (X'X)^{-1}X'\epsilon$$

$$V(\hat{\beta}|X) = E[(\beta - \hat{\beta})(\beta - \hat{\beta})']$$

$$\implies V(\hat{\beta}|X) = (X'X)^{-1}X'E(\epsilon\epsilon'|X)X(X'X)^{-1}$$

- Once this is done, we need an appropriate estimator for $E(\epsilon\epsilon'|X)$. In order to estimate $V(\beta|X)$, we also need to make assumptions about the structure of the errors.
- Finally, we take the square root of the diagonal elements of the $V(\hat{\beta}|X)$ matrix to generate standard errors of the coeffcients
- For a more detailed understanding, you may refer to this great document I found on github http://lukesonnet.github.io/teaching/inference/200d_standard_errors.pdf