Economics 375: Introduction to Econometrics

Why do we divide by the square root of n?

The purpose of this note is to help students understand the difference between the standard error and the standard deviation.

In class we noted that performing hypothesis tests involved measuring the distance, in terms of standard errors, between the observed (or sample) mean and the hypothesized (or population) mean. Thus, our z-statistic is given by

\[ Z = \frac{H_0 - \mu}{\sigma / \sqrt{n}}. \]

What might first appear odd to students, is the fact that we divide the standard deviation (\( \sigma \)) by the square root of \( n \). This likely appears odd because many of the problems originally encountered by students lack this feature. For instance, if I ask you what the probability of drawing a random number greater than 1.5 from a normal distribution with mean 0 and standard deviation of 1.25, then you convert the number (1.5) into the number of standard deviations from the mean of zero, and look up the answer on a table. In this case, 1.5 is 1.2 standard deviations from zero and, after looking up on the Z-table, I find that the probability of observing something greater than or equal to 1.2 standard deviations from the mean is 0.1151.

This gets more complicated when one considers the following type of problem: what is the probability of drawing two consecutive and independent random draws greater than 1.5 from the distribution above. Since each are independent of the other, it turns out the probability of drawing two numbers both greater than 1.5 is \( 0.1151 \times 0.1151 = 0.013 \).

Notice, in both of the above examples, neither involves dividing by the square root of \( n \). It turns out that one divides the standard deviation by the square root of \( n \) when one is interested in the standard error. To help with this, consider the following two rough definitions:

Standard deviation: A measure of the spread of a sampling distribution.
Standard error: A measure of the spread of the estimates of the center of a sampling distribution. In other words, the standard deviation of the estimates of the center of a sampling distribution.

To try to demonstrate the difference between these concepts, consider the die rolling example discussed in class. I would like you to imagine that we do not know the true population mean of a fair 6-sided die. The way we will try to estimate this unknown mean is to roll the die and perform a number of different experiments. Consider the following experiments:
1. Roll the die 10,000 times and take the average of the 10,000 as being our best guess at the unknown population mean.
2. Roll the die twice, take their average, write it down and repeat the experiment 5,000 times.
3. Roll the die four times, take their average, write it down and repeat the experiment 2,500 times.
4. Roll the die 10 times, take their average, write it down and repeat the experiment 1,000 times.

It turns out that under any situation, the standard deviation of a fair die is \( 2.91667 \times 0.5 = 1.707 \) (be sure you can show this). Thus, when I perform experiment #1, I should observe a distribution of numbers centered around 3.5 (the actual population mean) with standard deviation of 1.707. Here is a histogram created in Matlab of experiment #1:
Notice that the sample mean of all 10,000 rolls is very close to the population mean and the sample standard deviation is close to the population standard deviation of 1.707.

Now, consider case #2. Here, I continue to roll 10,000 dice. However, instead of recording each throw, I am recording the average of 5,000 pairs of die throws. Because it is relatively rare to get a pair of ones or a pair of sixes (or a pair of any number), I should observe relatively fewer observations on the ends of the distribution of the numbers that I record. In other words, the distribution of my estimated means (the averages of a pair of dice) should not be spread out as much as the case when I roll 10,000 dice and record each individually. Although I continue to roll 10,000 and the population standard deviation of rolling die is 1.707, it turns out that the distribution of the mean of the pair of die rolls will be less. Here are histograms from experiment #2:

To be clear, the above distribution is based upon rolling 10,000 dice with the same population mean and population standard deviation as the earlier example—however if you count the number of entries you will find they add to 5,000. Of course, this is because there are 5,000 pairs of die rolls in which the average was taken. Now, here is the remarkable thing—not only do the outcomes of the die rolls look like we expected (fewer observations on the tail), but if I measure the spread of the above histogram, I find it is smaller. The standard deviation of the above histogram is 1.1923. Now consider this relationship: As already seen, the standard deviation of rolling a 6-sided is
1.707. In this case, we have based our sample mean on the average of 2 die. The square root of 2 is 1.414. If I divide 1.707 by 1.414 I get 1.207—a number very close to the observed sample standard deviation of the above histogram of 1.1923. Statisticians call this number, 1.207, the population standard error. To be clear, the population standard error measures the spread of the averages of a random number while the standard deviation measures the spread of individual observations of a random number. In our case, the number 1.1923 is the sample standard error (it is an estimate of the population based upon the draw of 10,000 dice that we happened to observe).

To be clear, in the first case, where 10,000 die were rolled and recorded, the standard error is equal to the standard deviation. Of course, this occurs because each die roll in that case is a sample mean of the true population mean.

Here’s the histogram for situation #3.

In this case, the standard error is equal to \( \frac{\sigma}{\sqrt{n}} \). Since we know \( \sigma = 1.707 \) and, in case #3 \( n=4 \).

Thus, the population standard error is 1.707/2 = .8535 ≈ .8606 found in the sample. Again, this is based upon 10,000 total die rolls (broken into 2500 blocks of 4).

Finally, here is the results from experiment #4:
Compute the population standard error. Does the above data agree with your population estimates?

A more formal proof that the variance of the mean is $1/n^2$ of the variance of the distribution of the individuals follows:

Let $X_i$ be an independent draw from a distribution with mean $\bar{X}$ and variance $\sigma^2$.

What is the variance of $\bar{X}$?

By definition: $\text{Var}(X) = E[(X_i - E[X_i])^2] = \sigma^2$

So $\text{Var}(\bar{X}) = E\left[\left(\frac{\sum X_i}{n} - E\left[\frac{\sum X_i}{n}\right]\right)^2\right]

= E\left[\left(\frac{\sum X_i}{n} - \frac{1}{n} E\left[\sum X_i\right]\right)^2\right]

= \frac{1}{n^2} E\left[(\sum X_i - E[\sum X_i])^2\right]

= \frac{1}{n^2} E\left[(\sum X_i - E[\sum X_i])^2\right]

= \frac{1}{n^2} E\left[(X_1 + X_2 + \ldots + X_n - \frac{\sum X_i}{n} - \ldots - \frac{\sum X_i}{n})^2\right]

= \frac{1}{n^2} E\left[(X_1 - \bar{X})^2\right]

= \frac{1}{n^2} \sum E(X_i - \bar{X})^2 = \frac{1}{n^2} \sum \sigma^2

= \frac{1}{n} \sigma^2