**Appendix for** **Measures of Variability and Precision: Appreciating, Untangling, and Applying Concepts**

Below we pose and answer several questions related to content discussed in the article. We hope that the curious reader will find this supplement useful.

**(1) In the article, we mentioned one common measure that can be used to describe a typical observation from a distribution, the mean. We could also use the median to describe a typical value. What is the difference between these measures and when should each be used?**

The sample median is the value for which half of the sample observations fall below and half fall above. Whether a person chooses to report the mean or the median often depends on the shape of the distribution of the individual values under consideration. Since the mean is more strongly influenced by very high or very low values than the median, we tend to report the median value when the distribution of the values shows either positive or negative skew. We can assess the skew of a distribution by making a histogram of the individual values. If the histogram shows most of the data clustered around lower values and some (but fewer) data spread out in the higher value range, then the distribution has positive skew. If the histogram shows most of the data clustered around higher values and some (but fewer) data spread out in the lower value range, then the distribution has negative skew.

**(2) In the article, we mentioned the standard deviation as one possible measure of spread/variation. The range and interquartile range are also measures of spread/variation. What are these measures and when should each be used?**

The range can be reported in two ways. It can be reported by simply stating the minimum and maximum values of the data or it can be reported as the difference between the maximum and the minimum values. The interquartile range (IQR) can also be reported in two ways. It can be reported by stating the 25th and 75th percentiles of the data or it can be reported as the difference between these two percentiles. The IQR provides information about where the middle 50% of the data values lie. It is common the report the sample mean and the sample standard deviation together when the mean is an appropriate measure to report (see (1) above) while the median and either the range or the IQR are given together when the median is an appropriate measure to report (again, see (1)).

**(3) Why do we define variance in terms of the squared deviations? Why not just use the deviations?**

While it may seem simpler to sum up the deviations rather than the squares of the deviations in the variance formula, doing so would make the variance equal 0 every time we use the formula. For example, consider the small data set {3, 4, 7, 10}. The mean of this data set is 6. The table below shows the data values and their corresponding deviations from the mean.

|  |  |
| --- | --- |
| **Value** | **Deviation** |
| 3 | 3 – 6 = -3 |
| 4 | 4 – 6 = -2 |
| 7 | 7 – 6 = 1 |
| 10 | 10 – 6 = 4 |

Notice that if we try to sum up the deviations, we obtain 0. This is only one example, but no matter what our data values are or how large the data set is, we always have that the sum of the deviations is 0. The variance would not be a very useful measure of spread if it gave the same value for every data set, so instead, the formula considers the sum of the squared deviations. Since the square of any number is positive, the sum of the squared deviations does not suffer from the cancelations that arise from summing the positive and negative deviations.

**(4) Why do we divide by one less than the sample size in the variance formula?**

It can be proven mathematically that the formula for the variance (which divides the sum of squared deviations by one less than the sample size) is an unbiased estimator for the population variance. This means that, on average, the sample variance estimate will be equal to the population variance. However, if we chose to divide the sum of squared deviations by the sample size, we would have a biased estimate of the population variance. In fact, the estimate derived by dividing by the sample size (rather than one less than the sample size) will, on average, underestimate the population variance.

**(5) What is the Empirical Rule and when can we use it?**

The Empirical Rule can be used with normally distributed data. It says that approximately 68% of the values from the distribution will fall within 1 standard deviation of the mean (above or below), 95% of the values from the distribution will fall within 2 standard deviations of the mean, and 99.7% of the values from the distribution will fall within 3 standard deviations of the mean. If the data are not normally distributed, then the Empirical Rule does not necessarily hold.

**(6) The concept of the standard error of the mean (SEM) seems abstract, is there a way to visualize it?**

When describing the SEM in the article, we invoked the notion of repeated sampling and said that the SEM is the standard deviation of the sample mean values obtained from repeated sampling. This is shown in the figure below. Using the example from the article, on the left side of the figure we see the population of all converters. In that population, the mean striatal glutamate (SG) value is denoted by $μ$, which is unknown. If we could draw all samples of size 7 (recall that this was the size of the sample of converters in our example) and compute the sample mean for each one, then we could plot the distribution of all of the sample means to visualize what is known as the sampling distribution of the sample mean. Under suitable conditions (which we will not discuss but assume are satisfied in this example), the sampling distribution of the sample mean will look like the histogram on the right side of the figure. Notice that the distribution of the sample means is centered at $μ$. The standard deviation of this distribution is what we call the standard error for the mean.



 **(7) With respect to confidence intervals, what do we mean when we say that we are 95% confident?**

To understand the meaning of the confidence level, we have to go back to the notion of repeated sampling. Recall that the sample mean for the converters in our example is based on a sample of size 7. A confidence level of 95% means that if we could repeatedly draw samples of size 7 from the population of converters, compute the sample mean and SEM for each sample and use these values to construct a CI via the formula from Box 2, then 95% of those CIs would contain the population mean while 5% would not. Hence, the confidence level refers to a quality of the procedure used to construct the confidence interval much like a batting average refers to the quality of a hitter in the game of baseball.

**(8) In the article, we presented the 95% confidence interval for the difference between the mean SG from the converters and the mean SG from the non-converters. How was this interval computed?**

The 95% confidence interval of [0.26, 8.51] is computed using the formula for the difference in means based on the $t$-distribution assuming equal variances in the two groups and is given by:

$$(\overbar{X}\_{C}-\overbar{X}\_{NC})\pm t\_{17,0.025}\*S\sqrt{\frac{1}{n\_{C}}+\frac{1}{n\_{NC}}}$$

where $\overbar{X}\_{C}$ is the sample mean for the converters, $\overbar{X}\_{NC}$ is the sample mean for the non-converters, $n\_{C}$ is the sample size of the converter group, $n\_{NC}$ is the sample size for the non-converter group, $t\_{17,0.025}$ is the confidence coefficient value that cuts off 2.5% area in the upper tail of the $t$-distribution with 17 degrees of freedom, and

$$S=\sqrt{\frac{\left(n\_{C}-1\right)S\_{C}^{2}+\left(n\_{NC}-1\right)S\_{NC}^{2}}{n\_{C}+n\_{NC}-2}}$$

where $S\_{C}$ and $S\_{NC}$ are the sample standard deviations for the converter and non-converter groups, respectively.

Using the values from Box 1 in the article, we have:

$$S=\sqrt{\frac{\left(7-1\right)×4.49^{2}+\left(12-1\right)×3.84^{2}}{7+12-2}}=4.08$$

and

$$\left(30.25-25.86\right)\pm 2.11\*4.08×\sqrt{\frac{1}{7}+\frac{1}{12}}$$

which yields an interval that is slightly different from the interval presented in the article due to rounding here. The interval presented in the article is more exact since less rounding was done to compute that interval.