# Teaching Bits: Point Counter Point

Below is a refereed response by Michael M. Granaas to the article, "Let's Just Eliminate the Variance" by Deborah J. Rumsey from the November 2009 issue of JSE, a rejoinder from Dr. Rumsey will appear in the July issue.

## Teaching Bit Response: Keep Teaching the Variance

Michael M. Granaas
University of South Dakota

*Journal of Statistics Education* Volume 18, Number 1 (2010),
www.amstat.org/publications/jse/v18n1/pointcounterpoint.pdf



Rumsey (2009) suggests that those of us teaching introductory statistics courses should cease teaching the variance ($s^2$). She points out that:

1. Variance is taught, used to find the standard deviation, then never mentioned again.
2. The variance is in square units which are difficult to interpret.
3. It is more important to be able to interpret the statistical quantities than it is to compute those quantities with ease by hand.
4. Ease of hand computation is not facilitated by finding the variance and then taking the square root to find the standard deviation.

I agree that ease of hand computation is not a valid reason to continue teaching the variance. I also agree that interpretation is of utmost importance and that square units are challenging to interpret. However, not all of us abandon the variance once the standard deviation is introduced. There are good conceptual reasons to continue teaching the variance.

I can certainly understand the frustration expressed by Rumsey (2009), but ignoring the variance entirely is much too extreme a correction. Even if the variance is only used as a step on the path to computing the standard deviation there are important lessons that can be taught. In this case a change of emphasis is required.

The variance is, or can be, useful in explaining several of the concepts that we teach in introductory statistics courses—including the standard deviation. Emphasis on computation has often masked this importance for many years. As computers do more of the computational work for us we can, and should, emphasize the common underlying concepts and structures present in statistics. The variance is such a concept.

*Ignoring the Variance Entirely:*

My department delivers a pre-introductory statistics research methods course in which the variance is taught primarily as a step in finding the standard deviation. This is a course where the variance is never mentioned after the standard deviation is introduced. Even so there is an opportunity to explain to students that the descriptive statistics we choose are chosen because they make sense; that the goal of data analysis is clarity, not confusion.

In this course I introduce the variance and explain a bit about how it measures variability along with an example demonstrating that for two simple data sets the larger computed variance results from the data set with obviously more variability. I then tell students that although the variance has many desirable properties that are useful it is difficult to interpret because it is in squared units. The standard deviation is then introduced as a more easily interpreted version of the variance.

The variance is de-emphasized in favor of the standard deviation, but it is still present.

Why not just skip the “compute the variance” step by putting the variance formula under a square root sign and computing the standard deviation directly?

By teaching the variance first the student has learned that the standard deviation comes from somewhere sensible, it comes from the variance. It doesn’t just arrive in the world sort of like an average deviation but not really the same thing. This additional information, for some students, will deepen their conceptual understanding of the standard deviation.

Students are also learning that the statistics we choose to report are often selected based on interpretability rather than mathematical purity. That statistical quantities represent something interpretable needs to be emphasized repeatedly for the beginning student.

That some quantities are more meaningful than others in some context is also well worth emphasizing. Students can be made aware that while the variance may be safely set aside right now there will be other cases where it is useful. Whether that usefulness will show up in a given class will depend on the nature and goals of the course. But even if the variance is not useful in *this* course students are not harmed, and may benefit, from knowing that the variance is useful in *some* course/context.

One reason to skip teaching the variance would be to gain class time that could be used to help students develop an understanding of some other topic. It is unlikely that eliminating the variance entirely will result in the gain of any meaningful amount of class time to accomplish such a goal. Using the de-emphasized variance approach I describe there would be no gain in class time from eliminating the variance, and the ancillary lessons would be lost.

*Variance as a Common Underlying Concept*

While the variance per se may not be important to the introductory statistics student, the numerator, the Sum-of-Squared Errors (SSE) certainly should be. The SSE and related quantities show up repeatedly in the General Linear Model based statistics commonly taught to undergraduate students.

Possibly the most obvious repetition of variance is in regression where students are told that the regression line minimizes the SSE $\left(\sum(Y_i - \hat{Y}_i)^2\right)$ of the data points around the regression line.

The regression line minimizes the *variance* of the data points around the regression line.

The student who has not yet learned about the variance is at a disadvantage understanding the notion of a least-squares best fit line. Rather than referring back to the variance and bringing the concept of squared errors forward, the instructor must introduce it for the first time in a two variable setting, needlessly complicating the explanation.

The partitioning of variance in Analysis of Variance (ANOVA) is inexplicable absent an understanding of variance. True, the instructor could back up at this point, reintroduce the formula for the standard deviation, explain that the part under the square root sign is something called the variance, and then explain that the numerator of the variance is the measure of variation that is to be analyzed.

Whether it is regression, ANOVA, or some other topic, ignoring the variance early on doesn't make it go away; the teaching of variance simply shifts to a later point in the course. Introducing variance for the first time in the context of regression or ANOVA increases the complexity of those topics.

Some (e.g., Lockhart, 1997) utilize the SSE to provide a common foundation for the introductory statistics student. The mean is chosen because it minimizes the SSE. The t-test and F-test compare the SSE from some null model to the SSE from some alternative (full) model to determine if the SSE is meaningfully reduced by the alternative model. The Proportional Reduction in Error (PRE; $r^2$) is introduced early on simplifying the discussion of effect size later in the text. The SSE unifies material that is traditionally presented as independent topics. Such unification provides a strong basis for developing conceptual understanding.

*Other Courses*

The introductory statistics course will not be the final use many students will make of their statistical training. Some will go on to take a second or third statistics course, possibly a graduate course in Education, Psychology, Biology, or some other content area. The more rigorous of these courses will certainly have an expectation that students have some understanding of the variance. Failing to provide students with any exposure to the variance would handicap them as they move to the next level of their training.

The variance is too important to be ignored completely, even in introductory level courses. If anything we should be emphasizing the role of variance, or at least the SSE, in statistics as we enjoy the freedom from computational drudgery that computers provide.

## References

Lockhart, R.S. (1997). *Introduction to Statistics and Data Analysis*. W.H. Freeman and Co., New York.

Rumsey, D.J. (2009). "Let's Just Eliminate the Variance." *Journal of Statistics Education* Volume 17, Number 3, www.amstat.org/publications/jse/v17n3/rumsey.html

---

Michael M. Granaas
University of South Dakota
Vermillion, SD 57069
Michael.Granaas@usd.edu

---