In closing, the authors wish to thank Michael I. Rosen of Brown University for calling our attention to (9). We had already discovered (3), (4), (8) and a very complicated proof when Professor Rosen's communication gave us a clue to the much better one presented in section three. We are also indebted to the referee for many helpful comments.

REFERENCES

1. Abramowitz, Milton and Stegun, Irene A., *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, 1964.
2. Erdélyi, Arthur, et al., *Higher Transcendental Functions*, McGraw-Hill, New York, 1953.
3. Gauss, Carl Friedrich, *Disquisitiones Generales Circa Seriem Infinitam*

$$1+\left[\frac{\alpha\beta}{1\cdot\gamma}\right]x+\left[\frac{\alpha(\alpha+1)\beta(\beta+1)}{1\cdot 2\cdot\gamma(\gamma+1)}\right]x^2+\left[\frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1\cdot 2\cdot 3\cdot\gamma(\gamma+1)(\gamma+2)}\right]x^3+\text{etc.},$$

Pars Prior. Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores, Vol. II, 1813.
4. Jolley, L. B. W., *Summation of Series*, Dover Publications, Inc., Mineola, NY, 1961.
5. Kummer, Ernst Eduard, Über die Hypergeometrische Reihe, *Journal für die Reine und Angewandte Mathematik* 15 (1837), 39–83 and 127–172.

Dropping Scores

STEPHEN R. SHALOM
William Paterson College
Wayne, NJ 07470

1. Introduction Making up an exam for a course I teach on Social Science Research Methods, I was trying to devise a question that would test whether students had an intuitive understanding of variance as a measure of dispersion. The question I came up with was: "An instructor tells a class that they may drop their lowest of 10 test grades. For any particular student, what will happen to the variance of his or her grades when the lowest grade is dropped?" Just before making multiple copies of the exam, however, I discovered that in fact it is *not* the case that dropping the lowest score invariably lowers the variance, even leaving aside the trivial case where all the scores are equal. This led me to investigate the impact of dropping both the highest and lowest scores simultaneously, and again I found some counter-intuitive results.

This article discusses the effect on the variance of dropping the lowest score or the lowest and the highest scores. We use the common formulas for variance:

$$\sigma^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}=\frac{\sum x_i^2}{n}-\left(\frac{\sum x_i}{n}\right)^2, \tag{1}$$

and

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

where $x_1,\dots,x_n$ denote the n scores, $\bar{x}$ denotes their mean, and the sums are taken from 1 to n.

2. Main results *Dropping the lowest (or highest) score does not necessarily lower the variance.* Consider the set of scores {0, 1, 11}. We find that $\sigma^2 = 74/3$, and when the lowest score, 0, is dropped, $\sigma^2 = 25$.

Since adding a constant to each observation does not alter the variance, other counterexamples can readily be generated. Moreover, these same counter-examples apply if one calculates s^2. Similar counter-examples for $n > 3$ can also be found.

Of course, dropping the lowest score does not always increase the variance. The last section of this article will show the conditions under which the variance will increase, decrease, or remain unchanged.

Deleting the lowest and the highest scores simultaneously will decrease σ^2 except in two cases: (a) *when all the scores are the same, in which case the variance is zero and remains unchanged*; *and* (b) *where n is even and one-half of the scores are equal to one value and the other half equal to another value, in which case the variance is unchanged*. Assume there is some set of n scores, n > 2, such that the variance does not decrease when the highest and the lowest scores are removed. Since this property of the set will not be affected by any linear transformation on the set of scores, the problem can be simplified by subtracting the lowest score from each of the scores and then dividing all of them by the resulting highest score. (We exclude the case where all the scores are equal; when this is so, the variance will be 0 before removing the two scores and 0 after.) The new set of transformed scores will range from 0 to 1.

Let S and SS denote the sum and the sum of the squares, respectively, of all the scores. The sum and sum of the squares of all the scores excluding the highest and lowest will then be $S - 1$ and $SS - 1$, respectively. Let σ_0^2 = variance of the original data and σ_t^2 = variance of the trimmed data. When is $\sigma_0^2 \leq \sigma_t^2$? We require

$$\frac{SS}{n} - \frac{S^2}{n^2} \leq \frac{SS-1}{n-2} - \frac{(S-1)^2}{(n-2)^2}. \tag{3}$$

Expanding and combining terms yields

$$4(n-1)S^2 - 2n^2S + n^2(n-1) \leq 2n(n-2)SS. \tag{4}$$

Since for all $i, 0 \leq x_i \leq 1$, it follows that $x_i^2 \leq x_i$. Therefore, $SS \leq S$. In fact, $SS = S$ only when $x_i^2 = x_i$ for all i, which occurs only when every x_i is equal to 0 or 1. Temporarily leaving aside this case, $SS < S$.

Therefore, $2n(n-2)SS < 2n(n-2)S$. Using this fact, the inequality in (4) becomes

$$4(n-1)S^2 - 2n^2S + n^2(n-1) < 2n(n-2)S,$$

and (recall that $n > 2$) we see that $4S^2 - 4nS + n^2 < 0$, and $(2S - n)^2 < 0$. Therefore, we have reached a contradiction, and equation (3) is false, except for the case where all the x_i's are equal to 0 or 1. In this case, since $x_i = x_i^2$ for all i, $S = SS$. Substituting into equation (4) gives $(2S - n)^2 \leq 0$, and, $S = n/2$.

Therefore, $\sigma_0^2 \geq \sigma_t^2$ unless $x_1 = x_i^2$ for all i. In that case, $\sigma_t^2 \geq \sigma_0^2$ when n is even, and exactly half the scores are 0 and half are 1. Thus, if the set of scores is evenly split with only two values, then $\sigma_0^2 = \sigma_t^2$.

Deleting the lowest and highest scores simultaneously will not necessarily decrease the sample variance using formula (2), *even aside from the two special cases above.* When we use formula (2) for variance, this same proof by contradiction fails, and in fact it is rather easy to show examples for all n where dropping the highest and the lowest scores increases the variance. Consider the set of n scores (n even) where the lowest score is zero, the highest is one, half of the rest are equal to A and half equal to B, where $A+B=1$, and where $0<A<B<1$. Then, using formula (2), removing the highest and lowest score will increase the variance whenever:

$$s_0^2=\frac{\left(\frac{n-2}{2}\right)(A^2+B^2)+1}{n-1}-\frac{\left[\left(\frac{n-2}{2}\right)+1\right]^2}{n(n-1)}$$

$$<\frac{\left(\frac{n-2}{2}\right)(A^2+B^2)}{(n-3)}-\frac{\left(\frac{n-2}{2}\right)^2}{(n-2)(n-3)}=s_t^2. \tag{5}$$

Simplifying, recalling $B=1-A$, and solving for A taking only the smaller root since $A<B$), yields:

$$A<\frac{n-2-\sqrt{(n-2)(n-3)}}{2(n-2)}.$$

Thus for any n, we can choose an A sufficiently small to satisfy equation (5). For $n=4$, for example, A must be less than .1464 and the set of scores $(0, .1, .9, 1)$ will have its sample variance, s^2, increased from .27 to .32 when the high and low scores are removed. Likewise, for odd $n\geq 5$, we can take the set of scores where there are single scores equal to zero, one, and .5; half of the rest are equal to A and half equal to B, where $A+B=1$, and where $0<A<B<1$. This time, as long as we choose an A less than

$$\frac{2(n-2)-\sqrt{2(3n-13)(n-2)}}{4(n-2)},$$

removing the highest and lowest scores will increase s^2.

3. Characterizing the impact of excluding one extreme observation In this section, we will characterize the nature of the change in variance when an extreme observation is deleted.

Consider a set of n scores $x_1,\dots,x_n$; $x_1\leq x_2\leq \cdots \leq x_n$. Transform the scores by subtracting the score to be removed from all the other scores. If zero is the highest score, multiply all the scores by -1. Now zero is the score to be removed and all other scores are greater than or equal to zero. Write the transformed scores $y_1, y_2,\dots,y_n$. Let

$$S'=\sum_{i=1}^{n} y_i=\sum_{i=2}^{n} y_i \quad\text{and}\quad (SS)'=\sum_{i=1}^{n} y_i^2=\sum_{i=2}^{n} y_i^2.$$

Removing the lowest score will *increase* the variance σ^2 whenever:

$$\sigma_0^2=\frac{(SS)'}{n}-\frac{(S')^2}{n^2}<\frac{(SS)'}{n-1}-\frac{(S')^2}{(n-1)^2}=\sigma_t^2.$$

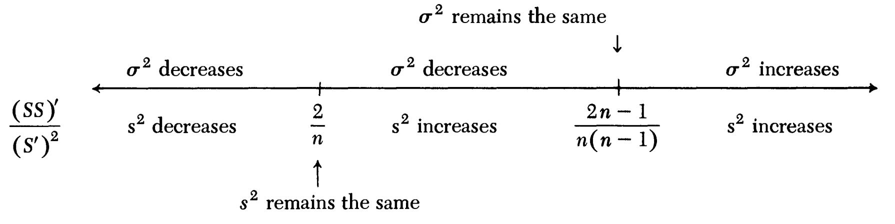
Expanding and combining terms reveals that the variance σ^2 increases if

$$\frac{(SS)'}{(S')^2}>\frac{2n-1}{n(n-1)}=\frac{2}{n}+\frac{1}{n(n-1)}>\frac{2}{n}.$$

Removing the lowest score will increase the sample variance s^2 whenever the following relationship holds:

$$s_0^2 = \frac{(SS)'}{(n-1)} - \frac{(S')^2}{n(n-1)} < \frac{(SS)'}{n-2} - \frac{(S')^2}{(n-2)(n-3)} = s_t^2.$$

Expanding and combining terms gives: $\frac{(SS)'}{(S')^2} > \frac{2}{n}$. The effect on the variance of removing the lowest score is illustrated in the following chart, where the axis represents the value of $(SS)'/(S')^2$:



Acknowledgement. The author thanks Eswar G. Phadia for helpful comments.

REFERENCES

1. David, H. A. General bounds and inequalities in order statistics, *Comm. Statist.-Theory Meth.* 17 (1988), 2119–2134.
2. Groeneveld, R. A. Best bounds for order statistics and their expectations in range and mean units with applications, *Comm. Statist.-Theory Meth.* 11 (1982), 1809–1815.