

Speed vs reaction time as a measure of cognitive performance

HOWARD WAINER

University of Chicago, Chicago, Illinois 60637

Choosing mean reaction time or mean speed (it's reciprocal) as the dependent variable in cognition experiments is frequently done for no better reason than convenience. It is shown that sometimes this choice is crucial in determining the order of the obtained effects. Moreover, it is proved that when the order of the effects is dependent upon the choice of the transformation, the assumption of homogeneity of variance is not true and hence ANOVA may not be justified. Alternative strategies are proposed to deal with this data-related problem.

Rather recently, there was a team sports car race between two neighboring cities. The rules were rather simple, with the winning team defined as the team which "on the average traversed the distance most quickly." One participating team interpreted this to mean the winning team was the one which had the fastest average speed for the prescribed distance. Another team felt that the rule meant that the winning team was the one whose average time for the course was the least. When this difference of interpretation was brought before the judge, he replied that, "Surely the two interpretations are equivalent; if one team satisfies the winning rule under one interpretation, it will under the other as well." As the unfortunate judge was to shortly discover, this is not at all true.

The great popularity of response time measures in contemporary psychological research, particularly in the areas of cognition and language processing, makes this little fable of more than theoretical interest. In many experimental paradigms (Sternberg's, for example), the dependent variable of response time is specified and (according to Sternberg, 1969) brooks no transformation—not even of the well behaved sort like logs. But there are many situations in which the recording of response time is merely convenient, and speed of performance ($1/\text{response time}$) would be equally defensible. Yet the two are most definitely not equivalent, even up to the ordering of the various experimental groups! Thus, one interpretation of the results of an experiment using response times could be exactly reversed if speed was the dependent variable.

Precisely this sort of difficulty may cause ambiguous results in cognitive research. For example, Fodor and Garrett (1967) report findings that have since been questioned by Foss and Lynch (1969), who could not replicate them. Foss and Lynch, however, used mean reaction time as a measure, while Fodor and Garrett

used mean speed. Of course, this difficulty may be caused by the various contaminations, as suggested by Hakes and Foss (1970), but it might very well be an artifact of the nonlinear transformation relating the two dependent variables.

Let me concretize these remarks with an example drawn from my own unfortunate past (Wainer & Reiser, Note 1). This to convince the disbelievers that the outcomes previously described are possible and are not artifacts due to arithmetic errors. In a study of the efficacy of various sorts of visual displays, subjects were presented with a particular display and an associated statement. The subject was to respond "true" or "false" by pushing one of two different buttons, depending on whether the statement correctly described the display, and the response time was recorded. The point of the experiment was to determine which display was the best, and it was assumed that mean response time would be a good measure of this. Reaction time for individual subjects' responses to each of two displays are shown in Table 1 (in seconds).

The mean response time for Display A is 11.46, and for Display B is 18.01. In Table 1 are also displayed the inverses of each of these response times (multiplied by 100 to make the numbers easier to read). These are speeds with the mean speed for Display A being 9.53, and for Display B being 11.49: a reversal. Why? The answer is to be found in noting that the mean of a nonlinear transformation is not the same as the nonlinear transformation of the mean. Thus, it is the unfortunate combination of the inverse transformation with the aggregation properties of the arithmetic mean that have caused these difficulties. Returning to the automobile race mentioned earlier, if one car broke down and did not finish, the inclusion of an infinite response time would be an insurmountable obstacle in that team's victory path, whereas transforming that infinite time to a speed of zero hurts their chances, but the blow it strikes is venial, not mortal.

What is to be done? There are a variety of useful solutions to this impasse. One is to have a sound

My thanks to Lance Rips for his thoughtful comments on this paper. The author's current address is Bureau of Social Science Research, 1990 M Street, N.W., Washington, D.C. 20036.

Table 1
Response Times and Speeds for Eleven Subjects
and Two Displays

Display A		Display B	
Response Time (Sec)	Speed (Sec) ⁻¹ × 100	Response Time (Sec)	Speed (Sec) ⁻¹ × 100
12.76	7.83	26.40	3.79
11.49	8.70	62.24	1.61
8.63	11.59	3.27	30.56
6.13	16.32	4.79	20.89
8.42	11.88	3.53	28.30
12.23	8.18	15.53	6.44
14.29	7.00	16.47	6.07
10.29	9.72	11.42	8.76
11.29	8.86	10.70	9.34
10.24	9.77	30.12	3.32
20.30	4.92	13.66	7.32
11.46*	9.53*	18.01*	11.49*
11.38†	8.84†	15.70†	6.95†
*Mean		†Midmean	

theoretical reason for choosing one measure over another: that is, cognitive processes take place in "real-time," therefore, responses should be recorded in "real-time." A second, and, to me, more convincing remedy is to use some measure of location other than the mean. Caveats against means as an index of individual behavior are very numerous and should be well known by now (e.g., Merrell, 1931). Similarly well known is the extreme vulnerability of the mean to outliers that can be at the root of the speed-time discrepancy (see Andrews, Bickel, Hampel, Huber, Rogers, & Tukey, 1972, for a detailed discussion of alternatives). It is obvious that the above-described misfortune can never happen with medians, since the inverse of the median is always the same as the median of the inverses. Therefore, one strategy when there is a reversal is to drop back to medians rather than means and proceed using the appropriate order statistics. This road, though inefficient (the Gaussian efficiency of the median relative to the mean is 2/3), leads in the proper direction. A more efficient solution, which would thus make use of more of the information contained in the data, would be a trimmed mean, in which, say g% of the points were trimmed off each end of the distribution, and the mean of the remaining 100 - 2g% of the points was used. Clearly, when g = 50, one arrives at the median, which is guaranteed to have the desired effect. Hopefully, with g less than 50, the same result will obtain, leaving a more efficient estimator of location.

Thus, my recommendation is to examine both the mean reaction time and the mean speed. If they both point to the same conclusion, then continue with either one. If they do not, then trim off points from both ends of the data distribution until the anomaly disappears. Unfortunately, normal statistical procedures like ANOVA require uncensored means. However, robust methods of analysis of variance are available for use on the trimmed means (Wainer, 1976).

Oh, yes, which of the two displays is better? The midmeans shown on Table 1 reflect the true state of affairs reasonably well. These are 25% trimmed means, in that the bottom and top 25% of the data points are trimmed off for each variable and the mean of the middle 50% is calculated. Trimmed means have a long history (Svanberg, 1891) and are commonly used in many situations (e.g., Olympic diving and figure skating, in which a certain percentage of judges are trimmed from both the high and low end before averaging). These have been shown to be reasonably effective for the robust estimation of location, when the distributions are not too bizarre. Thus, we can decide that Display A seems to be the winner, at least so far.

PROOF OF EARLIER WILD ASSERTIONS

Suppose two random variables x and y represent the amounts of time taken by two processes. Two measures of central tendency that can be computed for each are mean time and mean speed. Suppose E(x) and E(y) represent mean times and E(1/x) and E(1/y) represent mean speeds. We need to know under what conditions the two types of means are ordered similarly; that is, under what conditions the following statement is true:

$$\text{If } E(x) > E(y), \text{ then } E(1/y) > E(1/x). \tag{1}$$

The exploration of Statement 1 will be eased considerably with the following result:

If the deviations are small compared to the value of the mean, then, the harmonic mean,

$$H_x = 1/[E(1/x)] \cong \mu_x (1 - \sigma_x^2/\mu_x^2) \tag{2}$$

(Kendall & Stuart, 1969, p. 52). This result leads us directly to a sufficient condition for Statement 1 to hold, which is a frequently made assumption. That is, if we assume homogeneity of variance, the relation stated in Statement 1 will hold. Interestingly, this is an assumption of analysis of variance which is the most frequent method of hypothesis testing employed with these sorts of data. For clarity, it is best if we state this result in the form of a theorem.

THE HOMOGENEITY OF VARIANCE AND HARMONIC MEANS THEOREM

If E(x) > E(y), then E(1/y) > E(1/x), whenever the variances are homogeneous ($\sigma_x^2 = \sigma_y^2 = \sigma^2$) and when the deviations are small with respect to E(x) and E(y).

Proof

Let $\mu_x = E(x)$ and $\mu_y = E(y)$. We know that $\mu_x > \mu_y$, also, from Statement 2,

$$H_x = 1/E(1/x) \cong \mu_x (1 - \sigma^2/\mu_x^2) > \mu_y (1 - \sigma^2/\mu_y^2) \cong 1/E(1/y) = H_y;$$

hence, $E(1/y) > E(1/x)$.

Note, of course, that homogeneity of variance is a sufficient condition, and not a necessary one. Yet, this finding tells us that, if the ordering after inversion is not as expected, then homogeneity of variance does not hold. Since this is an assumption of analysis of variance, we are being warned. Of course ANOVA is relatively robust with respect to departures from this assumption, but it is not totally immune, and so finding this reversal is a strong hint not to use ANOVA.

The justification for trimming comes from the homogeneity theorem. Heterogeneity of variance can come from one of two sources. The first is a couple of outliers which artificially inflate the variance in one sample. Trimming clearly gets rid of this sort of contaminant and thus equalizes the variance. A second possibility is for the two samples to come from populations which truly have different variances. If this is true, then techniques which require homogeneity are doomed from the start. Trimming in this case will still work, but the amount of trimming necessary may be very severe. Obviously, trimming until the resulting trimmed distributions are nonoverlapping will result in the desired effect, but this may be so severe that the user is better off utilizing the various types of rank procedures designed for such problems.

REFERENCE NOTE

1. Wainer, H., & Reiser, M. *Assessing the efficacy of visual*

displays. Paper presented at the annual meeting of the American Statistical Association in Boston, Mass., August 23, 1976.

REFERENCES

- ANDREWS, D. F., BICKEL, P. J., HAMPPEL, F. R., HUBER, P. J., ROGERS, W. H., & TUKEY, J. W. *Robust estimates of location*. Princeton, N.J.: Princeton University Press, 1972.
- FODOR, J. A., & GARRETT, M. Some syntactic determinants of sentential complexity. *Perception & Psychophysics*, 1967, 2, 289-296.
- Foss, D. J., & LYNCH, R. H. Decision processes during sentence comprehension: Effects of surface structure on decision times. *Perception & Psychophysics*, 1969, 5, 145-148.
- HAKES, D. T., & Foss, D. J. Decision processes during sentence comprehension: Effects of surface structure reconsidered. *Perception & Psychophysics*, 1970, 8, 413-416.
- KENDALL, M. G., & STUART, A. *The advanced theory of statistics* (Vol. 1, 3rd ed.). Hafner, 1969.
- MERRELL, M. The relationship of individual growth to average growth. *Human Biology*, 1931, 3, 37-70.
- STERNBERG, S. The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, 1969, 30, 276-315.
- SVANBERG (1821). In Czuber, E., *Theorie der beobachtungsfehler*. Leipzig, 1891.
- WAINER, H. Robust statistics: A survey and some prescriptions. *Journal of Educational Statistics*, 1976, 1, 285-312.

(Received for publication October 19, 1976;
revision received November 23, 1976.)

ERRATUM

Marmurek, H. H. C. Processing letters in words at different levels. *Memory & Cognition*, 1977, 5 (1), 67-72. Page 68, Materials section: Visual angles should be specified in minutes rather than in degrees.