Public Opinion Polls

What do central tendency, variability, and statistical significance mean in the context of motor control research?

Have you ever wondered what the numbers mean when reporters announce the results of election campaign polls? For example, a reporter says that "Candidate X received an average support of 47 percent of the voters who were polled in the survey, candidate Y received an average of 21 percent of their support, and 32 percent of the voters surveyed have not yet made up their minds. These results are considered accurate within ±4 percent, or 95 times out of 100." In a very few words, the reporter has managed to provide a lot of statistics, some that may be lost on the typical voter. Though they sound complicated, in fact these numbers are very simple. And, most important for our purposes, the statistics used in public opinion statements are the very same ones used in studies of perception and action.

The concept of a central tendency is implied by the term "average" as it relates to the support for each of the candidates. The concept of a statistical average is simple: These are just arithmetic averages, or means. If 258 voters were surveyed and 121 of them supported candidate X, then 121 / 258 = 46.9 percent and can be rounded up to 47 percent for the sake of simplicity. The statement "correct within ±4 percent, or 95 times out of 100" requires a little more explanation. The implication here is that if 100 polls were conducted, each time polling a different sample of people, and all other factors were more or less equal (e.g., same date of the poll, same population from which the sample is taken), then the means reported would be approximately the same in at least 95 of these 100 polls (here *approximately* means within plus or minus 4 percent of the means that were reported in the original poll—that is, between 43 and 51 percent, for a reported mean of 47 percent). The pollsters recognize that their methods are not foolproof, however, and so they add the caveat that the poll could be wrong. They say that, by random chance, no more than 5 polls in 100 are likely to produce results that vary more than ±4 percent from the means reported.

Most reports of motor control research use these same three basic types of numbers: measures of central tendency, or average (e.g., 47 percent), variability (± 4 percent), and statistical significance (95 times out of 100). Let's demystify the last one first. Statistical significance means essentially that one researcher's discovery will be repeatable by other researchers under similar conditions at least 95 times out of 100 (a sort of gold standard in reporting statistics in this type of research). Sometimes researchers report that their findings were "not statistically significant." This simply means that the confidence in repeating any differences found in their results did not achieve this gold standard, and therefore any reported differences in the means should be viewed with extreme caution.

Measures of central tendency (such as the mean) and measures of variability (such as the standard deviation) are often reported together in motor control research reports. For example, if you wanted to know the average RT (reaction time) of a certain sprinter, you could download the results of, say, that athlete's last 10 races and compute the mean and standard deviation. The mean would simply be the statistical average of the reaction times in those 10 races (the sum of the 10 race RTs, divided by 10). The standard deviation is a little more complicated. It is the mean of the squared deviations of each individual RT relative to the athlete's mean RT (then expressed as the square root of that value). In simpler terms, the standard deviation is the average deviation of each of the individual RT values from the mean RT.

Knowledge about the mean and standard deviation is often useful when comparing performances. Suppose that over their last 10 races, sprinter A has a mean RT of 150 milliseconds (0.15 of a second) and a standard deviation of 10 milliseconds; and let's say that sprinter B has a mean of 140 milliseconds and a standard deviation of 30 milliseconds. From these data we know that, on average, sprinter B had a faster reaction time than sprinter A. However, sprinter A tends to be more consistent (less variability) than sprinter B, and therefore, for any single race, is more likely to have a reaction time closer to her average than will sprinter B.

When the individual numbers are distributed fairly regularly about the calculated average (i.e., what statisticians refer to as being normally distributed), the mean represents an appropriate central tendency for the group of numbers as a whole. But, this is not always the case because the mean is not always an unbiased (or appropriate) representation of central tendency. I will illustrate this idea using some findings from an actual sprint race.

The IAAF (International Association of Athletics Federation) World Indoor Athletics Championships are held every two years, and the winner of the 60-meter sprint is crowned the fastest person in the world. It is amazing to watch one of these races. Because the 60-meter distance is finished so quickly (the world record is less than 7 seconds for both men and women), a premium is placed on a fast reaction to the sound of the starter's gun (see "Jumping the Gun" in chapter 5). In many world championship races, in which only the very fastest runners represent their countries, it is very difficult to detect any differences among the runners in reacting to the sound of the gun. If this were always so, then the average reaction time for the field of athletes in any given race would be a good indicator of the individual RTs for each runner in the field.

But have a look at figure 4.1. These results occurred in a heat at the 1999 IAAF World Indoor Championships for the 60-meter sprint. The figure represents the RTs for each of the eight runners in the heat. Seven of the eight runners had RTs of 142 milliseconds or less. But, one runner, Maurice Greene of the United States, who was then (and still is) the world-record holder for the 60-meter race, got off to a terrible start with an RT (251 msec) that was more than a tenth of a second slower than every other runner in the race (although he still managed to finish second in the race). What I want you to notice in this figure is that seven of the eight runners in the race had an RT that was faster than the mean RT for the field (143 msec). How could a mean of 143 be representative of the entire field when seven of the eight runners had RTs less than the mean? This instance, in which one extreme score has a large effect on the mean, represents a case in which the mean is a poor (or unrepresentative) measure of central tendency, due to a nonnormal distribution of the individual scores. The median (127 msec), also plotted in figure 4.1, is a better measure of the heat's average because, as the middle number in an ordered series, it is hardly affected at all by a single extremely different (or outlier) score.

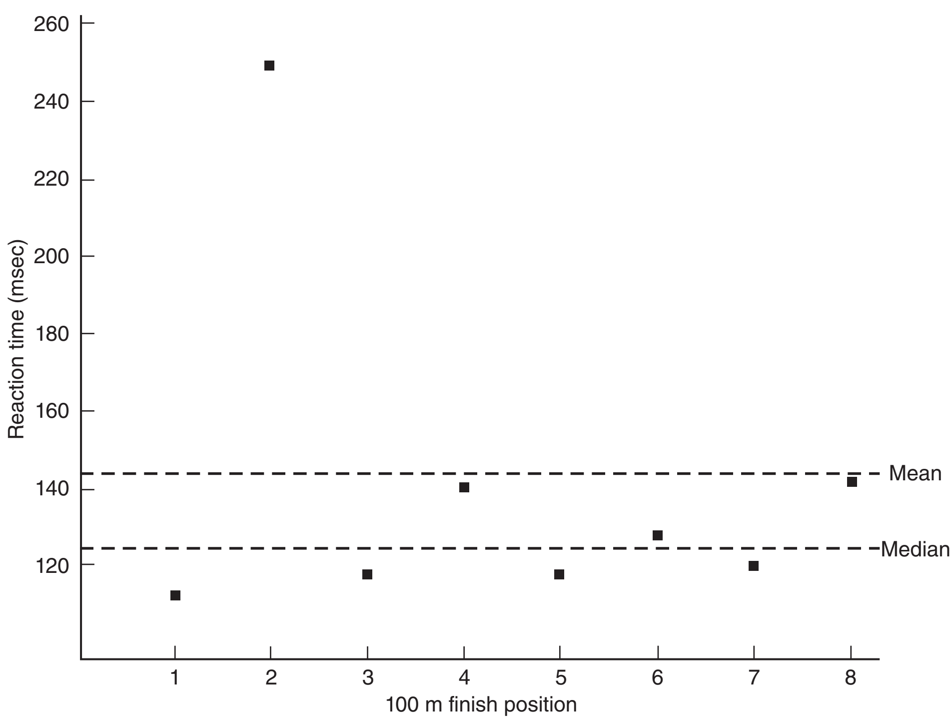

Figure 4.1 Reaction times for eight runners in heat 2 of the 1999 IAAF 60-meter men's sprint. Because of one runner's extremely long RT, the mean (143 msec) is higher than seven of the eight runners' RTs. In this case, the median (127 msec) is more representative of the true central tendency.

Data from www2.iaaf.org/wic99/results/index.asp.

From T.D. Lee, 2011, *Motor control in everyday actions* (Champaign, IL: Human Kinetics).

Reading the findings of motor skills research in journals can be a daunting task because of all the numbers. In reality, however, most studies simply report measures of central tendency and variability, then provide some inferential statistics that suggest how repeatable their findings are likely to be. And the best part about the statistics reported in journal articles is that, unlike with politicians, you won't later feel guilty for having voted for them.

Self-Directed Learning Activities

1. Define the terms *mean*, *standard deviation*, and *statistical significance* in your own words.
2. Find a published experiment that uses the preceding terms and interpret the results using your own language to describe the statistics.
3. Briefly describe two types of statistical tests that result in estimates of statistical significance.
4. Calculate the mean, median, and standard deviation for the following set of RTs: 125, 133, 177, 143, 161, 145, 201, 150, 166, 138.

Notes

- World records for 60-meter sprint:
  - Men: Maurice Greene (United States): 6.39 seconds
  - Women: Irina Privalova (Russia): 6.92 seconds
- Results of previous IAAF competitions can be found at www.iaaf.org/history/index.html

Suggested Readings

Thomas, J.R., Nelson, J.K., & Silverman, S.J. (2011). *Research methods in physical activity* (6th ed.). Champaign, IL: Human Kinetics.