PediatricsinReview®

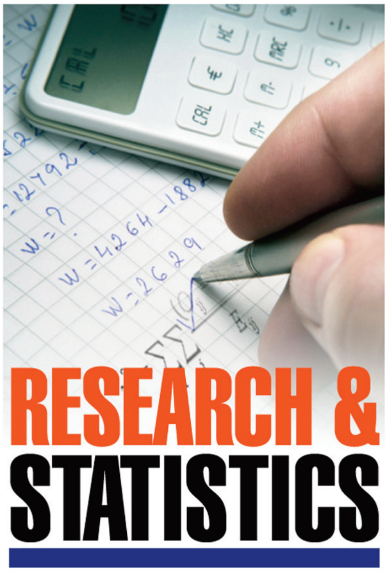
Research and Statistics : Distribution, Variability, and Statistical Significance

Tyler K. Smith and Sara B. Johnson

Pediatrics in Review 2010;31;431
DOI: 10.1542/pir.31-10-431

Distribution, Variability, and Statistical Significance

Tyler K. Smith, MD, MPH,* Sara B. Johnson, PhD, MPH†



Author Disclosure

Drs Smith and Johnson have disclosed no financial relationships relevant to this article. This commentary does not contain a discussion of an unapproved/investigative use of a commercial product/device.

Case Study

Given the growing prevalence of overweight and obesity among preschool-age patients, you have been charged with identifying an effective intervention to reduce body mass index (BMI) in this age group. Your search of a bibliographic database identifies a randomized, controlled trial of a physical activity program that looks promising. (1) The study authors examined a three-times-weekly 30-minute exercise program conducted over the course of 1 year with 545 pre-kindergarteners. The study results are reported in confidence intervals and P *values. How do you evaluate the effectiveness of the program in reducing child BMI?*

Samples and Populations

Clinical trials (and, indeed, most research studies) gather data from a sample of people and use the information from the sample to make inferences about the population from which the sample was drawn. In this case, the 545 children in the study sample are used to make inferences about how the program would work with the whole population of similar pre-kindergarteners. Rarely are data available about an entire population. For this reason, a sample is used. Several metrics are commonly calculated or reported that provide information about the nature of samples and populations in clinical research. Typically, we are interested in the shape and variability of the distribution of observations (in this case, children) as well as indicators to determine whether the control and experimental groups really differ from one another.

Distributions and Variability

Standard Deviation

Standard deviation (SD), which often is denoted by the Greek letter sigma (σ) is a measure of variability of individual values (eg, BMIs) around the mean or average value. Specifically, the SD provides information about how tightly clustered the data are around the mean. When data are grouped closely together, the SD is small. When data are highly variable, the SD is large. One common data distribution is the normal or bell-shaped distribution (Figure). A normal distribution has most of its values near the mean, with progressively fewer values found farther from the mean. One SD above and below the mean encompasses 68.2% of the values. Two or three standard deviations represent 95.4% or 99.7% of data values surrounding the mean, respectively.

Children's weights help to illustrate the concept. The mean weight for 4-year-old children in your clinic is 16 kg, with an SD of 2 kg. Based on a normal distribution, 95% of 4-year-old children will weigh between 12 kg and 20 kg (mean +2 SDs). Children who weigh 22 kg (mean +3 SDs) are very uncommon.

Standard Error of the Mean

The standard error of the mean, or simply standard error (SE), often is confused with the SD, but they provide different information. Recall that typically we use a sample of children to estimate the population of all similar children. The SE describes how accurate the sample mean is compared with the "true" population mean. However, we generally do not know the "true" mean (because we do not weigh everyone). One method of addressing this issue is to draw many, many samples from the population, but this procedure is not very feasible. Instead, we use SE to estimate our confidence that the sample mean reflects the population mean. As such, SE indicates the variability of these sample means (assuming we were to draw them repeatedly) around the population mean. In keeping with our previous discus-

*Department of Pediatrics, Johns Hopkins University School of Medicine.

†Department of Pediatrics, Johns Hopkins Bloomberg School of Public Health; Department of Population, Family & Reproductive Health, Johns Hopkins University School of Medicine, Baltimore, Md.

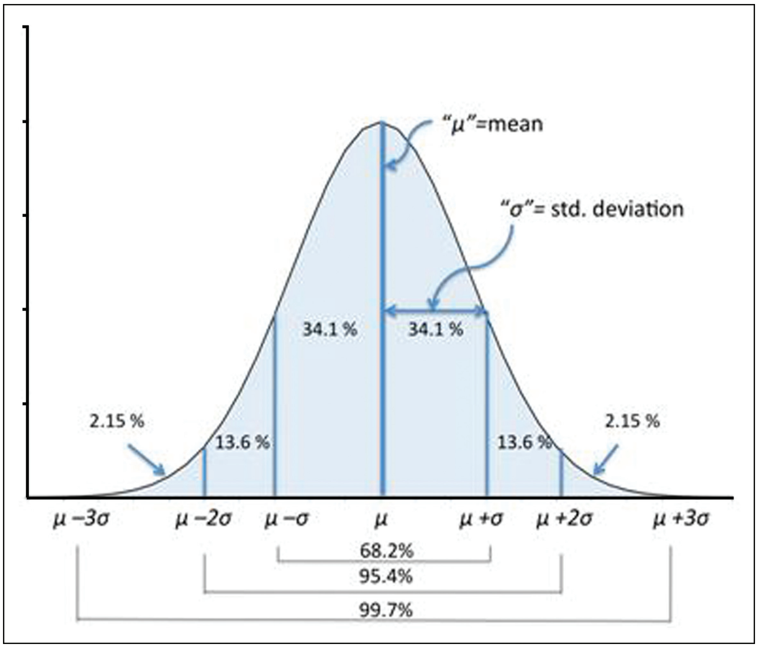


Figure. Illustration of the normal distribution. μ=mean, σ=standard deviation.

sion of the SD, we are 95% confident that the true population mean falls within 2 SDs of our sample mean. SE depends on the sample size. As the sample size increases, the SE decreases. More conceptually, the more information we have about the mean, the closer the sample mean gets to the "true" population mean. Sample size does not affect SD. (2) (3)

Statistical Inference

Confidence Intervals

Measures of distribution and variability are used for statistical inference and to quantify uncertainty about results. The range in which we are 95% confident that the true population mean will be found is called the 95% confidence interval (CI). Small CIs mean high precision; large CIs reflect lower precision and more uncertainty. The upper and lower bounds of this interval are calculated by adding and subtracting two SDs to the sample mean. In the literature, 95% CIs are common, but 99%, 98%, and 90% also are reported. Values that fall outside of the CI have a small chance of being there by chance alone and, therefore, often are referred to as "statistically significant." For example, in a comparison of means from a treatment and a control group, clinicians can examine the 95% CIs for each. If the CIs do not overlap, it is highly unlikely that the populations from which they are drawn have the same mean.

Testing for Statistical Significance, *P* Values

Investigators commonly report both the CI and a *P* value. The *P* value is a probability that answers the following question: "If the population means from which I have drawn these samples really are different, what is the probability of seeing a difference at least as extreme as I observe here when there is really no difference (eg, due to chance alone)?" A *P* value of 0.01 indicates that there is a 1% chance of observing the available data or something more extreme due to chance alone. Often, a *P* value of less than 0.05 is considered "statistically significant."

Example: Preschool Physical Activity to Reduce BMI

Using these tools, you examine the study of physical activity in prekindergartners. (1) At 12 months, the intervention group's BMI increased an average 0.41 kg/m^2 with an SD of 1.05 kg/m^2. The control group's BMI increased an average 0.43 kg/m^2 with an SD of 1.10 kg/m^2. There was no statistically significant difference between the two groups. The P *value reported for the comparison was* P*=0.90; in other words, the chance of observing a difference of this size between the treatment and control groups could easily have been due to chance. Therefore, the program was not effective in reducing BMI. The authors did, however, find a statistically significant difference of 0.8 units (95% CI: 0.3, 1.3) between the intervention and control groups on fundamental movement skills. (1) Because they are 95% confident that the boundaries in which the "true" population mean will be found does not contain zero (eg, no difference between the groups), the authors conclude that the program improved motor skills in the treatment group. With these results in mind, this program does not seem a good candidate for reducing BMI in your preschool patients but might be suitable for improving motor skills.*

References

1. Reilly JJ, Kelly L, Montgomery C, et al. Physical activity to prevent obesity in young children: cluster randomized controlled trial. *BMJ.* 2006;333:1041

2. Streiner DL. Maintaining standards: differences between the standard deviation and standard error, and when to use each. *Can J Psychiatry.* 1996;41:498–502

3. Altman DG. Standard deviations and standard errors. *BMJ.* 2005;331:903