

I. The idea of the standard error of the mean

Suppose we want to make an inference about the mathematical abilities of a particular population (e.g., American high school students). We would construct some measure of mathematics ability (e.g., a test with 100 mathematics problems) and give it to a sample of the target population (e.g., 5 high school students in Pittsburgh and 5 high school students in Madison). To summarize the data in a meaningful way, we would report an estimate of the population mean (e.g., \bar{Y}) and the number of measurements (e.g., $n = 10$) that the estimate is based on. We would also report a measure of the variability of the data about the mean, such as the sample variance (s^2) or the standard deviation (s).

The sample mean, sample variance, and standard deviation are useful, but not useful enough. Our estimate of the population mean is fixed by our sample (e.g., the 10 high school students). If we did the experiment again with another set of participants, then we almost certainly would get a different mean!

What we really want to know is "how good is our estimate of the mean?" not "how much variation was found around this particular estimate of the mean?" We can do this by calculating the **standard error (SE)**.

Imagine that we repeat the experiment again and again, with different random samples (e.g., different sets of high school students). Each time, we would get a slightly different mean, but if we were to plot a frequency distribution (histogram) of the sample means then it would show a normal distribution (eventually, when the number of samples was large enough).

We could find the mean of the sample means and then calculate a standard deviation of it. By convention, this standard deviation of the mean is called the standard error. The notation for the standard error of the mean is σ_M or $\sigma_{\bar{y}}$

We can estimate σ_M using s :

$$\hat{\sigma}_M = \frac{s}{\sqrt{n}}$$

This formula can be used to calculate the estimated SE of the mean from the data of a single group. In published papers, you will often see results expressed as mean \pm SE. The error bars on bar and line graphs are almost always standard error bars.

II. Pooled estimate of variance as the basis of standard error estimate

When we carry out an independent groups t-test or a one-way between groups ANOVA, we assume homogeneity of variance in the groups (e.g., $\sigma^2_1 = \sigma^2_2$), and we estimate the common population variance using either s^2_{pooled} or MS_{error} from the ANOVA.

s^2_{pooled} is a weighted average of s^2_1 and s^2_2 :
$$s^2_{\text{pooled}} = \frac{(n_1 - 1)s^2_1 + (n_2 - 1)s^2_2}{n_1 + n_2 - 2}$$

MS_{error} in a one-way between-groups ANOVA is just s^2_{pooled} extended to more than two groups. So the SE of the mean can be estimated by:

$$\hat{\sigma}_M = \sqrt{\frac{MS_{S/A}}{n}}$$

III. When to use pooled versus separate estimates of the standard error of the mean

The pooled estimate of SE would yield standard error bars for different groups that depend only on the sample size, since the numerator would be identical. Done this way, the standard error bars on a graph would correspond closely to the statistical tests, and would allow the reader to ‘eyeball’ the statistical results. But, they don’t allow the reader to eyeball homogeneity of variance. If you want a graph that allows the reader to eyeball homogeneity of variance, then you want to calculate the estimated standard error of the mean from each group’s data separately. The estimate is then based on the original formula presented above:

$$\hat{\sigma}_M = \frac{s}{\sqrt{n}}$$

A NOTE ABOUT SPSS: SPSS uses the pooled estimate of SE (based on either s^2_{pooled} from the t-test or MS_{error} in ANOVA) when it reports means and standard errors under the “General Linear Model > Univariate” menu. In contrast, SPSS estimates standard errors separately for each group in the descriptive statistics it reports under the “Compare Means > One-Way ANOVA” menu.