BMJ

ENDGAMES

What is the standard error of the mean?

Philip Sedgwick *reader in medical statistics and medical education*

Centre for Medical and Healthcare Education, St George's, University of London, Tooting, London, UK

A cluster randomised double blind controlled trial investigated the effects of micronutrient supplements during pregnancy. A trial with three treatment arms was used. Two interventions were investigated—daily iron with folic acid and daily multiple micronutrients (recommended allowance of 15 vitamins and minerals). Control treatment was daily folic acid. The setting was 327 villages in two rural counties in northwest China. In total, 5828 pregnant women were recruited. Villages were randomised to treatment group, stratified by county, with a fixed ratio of treatments (1:1:1).[1]

Outcome measures included birth weight. Birth weight was available for analysis for 4421 live births. Mean birth weight was 3153.7 g (n=1545; 95% confidence interval 3131.5 to 3175.9, standard deviation 444.9, standard error 11.32) in the control group, 3173.9 g (n=1470; 3152.2 to 3195.6, 424.4, 11.07,) in the iron-folic acid group, and 3197.9 g (n=1406; 3175.0 to 3220.8, 438.0, 11.68) in the multiple micronutrients group. Average birth weight was significantly higher in the multiple micronutrients group than in the control (folic acid) group (difference 42.3 g; P=0.019). Although average birth weight was higher in the iron-folic acid group than in the control group, the difference was not significant (24.3 g; P=0.169).

Which of the following statements, if any, are true?

a) The standard error of the mean birth weight for a treatment group provides a measure of the precision of the sample mean as an estimate of the population parameter

b) Generally, the standard error of the mean birth weight for a treatment group would increase as the sample size increased

c) If all pregnant women in the population took folic acid, 95% of children would have a birth weight between the limits of the 95% confidence interval (3131.5 to 3175.9 g)

d) About 95% of children born to mothers in the folic acid treatment group had a birth weight between the limits of the 95% confidence interval (3131.5 to 3175.9 g)

Answers

Statement *a* is true, whereas *b*, *c*, and *d* are false.

The results for the control group (folic acid treatment) will be used to illustrate the underlying principle of the standard error of the mean. The sample mean birth weight for this group was 3153.7 g. This is an estimate of the population parameter; that is, the population mean birth weight that would be achieved if folic acid was offered to all members in the population of pregnant women. The value of the population parameter is obviously not known. The sample mean is referred to as a point estimate. It is expected to be similar in size to the population mean, although it is unlikely to be exactly equal. Any inaccuracy in the sample estimate will be due to sampling error—that is, the error introduced by the estimate being based on a sample of pregnant women from the population. The accuracy of the sample mean as an estimate of the population parameter is quantified by the standard error of the mean (*a* is true). Standard error of the mean is often abbreviated to standard error. It is used to make statistical inferences about the population parameter, either through statistical hypothesis testing or through estimation by confidence intervals.

The standard error of the mean birth weight for the control group was derived by dividing the sample standard deviation by the square root of the sample size. Therefore, generally, if the sample size for the folic acid treatment group was increased, the size of the standard error would decrease (*b* is false). This may be intuitive, because as the sample size approaches that of the population, the sample mean will become closer in value to the population mean.

By itself, the value of the standard error of the mean birth weight for the control group has limited usefulness. It is used to derive the confidence interval—a range of values that quantifies the uncertainty in the sample mean as an estimate of the population parameter. The confidence interval is regarded as an interval estimate for the population parameter of mean birth weight. A percentage is attached to a confidence interval, typically 95%. The 95% confidence interval for the population mean birth weight for folic acid treatment is 1.96 standard errors either side of the sample mean—(3153.7 g−1.96(11.32)) to (3153.7 g+1.96(11.32))—which is 3131.5 g to 3175.9 g. Therefore, it is estimated that the population mean is contained by the limits of 3131.5 g to 3175.9 g, with a probability of 0.95. It is not possible to estimate which value in the interval, if at all, the population mean takes. There is a 0.05 (5%) probability that

p.sedgwick@sgul.ac.uk

the confidence interval will not contain the population mean at all. Nonetheless, the point estimate is our best estimate of the population parameter. The confidence interval describes our uncertainty in the estimate.

The 95% confidence interval does not describe the expected variation in birth weights in the population if all pregnant women took folic acid (*c* is false). Furthermore, the 95% confidence interval does not describe the variation in observed birth weights for the folic acid treatment group in the trial (*d* is false). Rather, it is the standard deviation that is used to describe the variation in birth weights in the sample. Standard deviation and standard error are often confused. One way of recalling when to use these statistics is to remember that standard error is for estimation and standard deviation is for description.

The sample standard deviation of birth weight for the folic acid treatment group can be used to calculate a series of ranges in birth weight containing certain percentages of the births in the sample. As described in a previous question,[2] three ranges can be derived. About 68% of infants born to women in the folic acid treatment group had a birth weight that was no further than one sample standard deviation away from the sample mean. In addition, about 95% were no further than two sample standard deviations away from the sample mean, and around 99% were no further than three standard deviations away.

When comparing each treatment group against the control, as described in a previous question,[3] it would have been good practice and more informative to present a confidence interval for the mean difference between treatment groups in birth weight rather than for the birth weight for each treatment group. These confidence intervals would quantify the inaccuracy of the sample difference between treatment and control in mean birth weight as an estimate of the population difference. The researchers reported that when compared with the control, mean birth weight was higher for both treatments—42.3 g for multiple micronutrients (95% confidence interval 7.1 to 77.5) and 24.3 g for iron with folic acid (-10.3 to 59.0).

The standard error can be calculated for other sample estimates beside the sample mean, including a proportion, relative risk, and odds ratio—the standard error of each estimate is used in a similar way to the standard error of the mean to calculate a 95% confidence interval for the population parameter.

1   Zeng L, Cheng Y, Dang S, Yan H, Dibley MJ, Chang S, et al. Impact of micronutrient supplementation during pregnancy on birth weight, duration of gestation, and perinatal mortality in rural western China: double blind cluster randomised controlled trial. *BMJ* 2008;337:a2001.
2   Sedgwick P. Standard deviation versus standard error. *BMJ* 2011;343:d8010.
3   Sedgwick P. Confidence intervals and statistical significance. *BMJ* 2012;344:e2238.

Cite this as: *BMJ* 2013;346:f532