A Tutorial on Standard Errors

Abstract – This is an introductory tutorial on standard errors. Every statistical estimate has its own Standard Error. Using an incorrect definition for a standard error invalidates the results of any study.

Keywords: standard errors, sampling distributions, correlation coefficients, fisher transformations

Published: 07-25-2010; Updated: 10-07-2016

© E. Garcia, PhD; admin@minerazzi.com

Note: This article is part of a legacy series that the author published circa 2010 at http://www.miislita.com, now a search engine site. It is now republished in pdf format here at http://www.minerazzi.com, with its content edited and updated.

Introduction

The s*tandard error*, denoted by *stderr* or *SE*, is a statistical indicator of the reliability of an estimate. It is formally defined as the standard deviation of the sampling distribution of a statistical estimate (Kurtz, 1991; Arsham, 1994).

The *sampling distribution* and *sample distribution* are different things. The former is the distribution of a sample statistic while the latter is of raw observations. So an *SE* is not the standard deviation of a sample, but of a sampling distribution.

While the standard deviation is a measure of the *dispersion* or variability of values in a sample, an *SE* is a measure of the *dispersion* or variability of values in the sampling distribution of a statistical estimate (McHugh, 2008). An *SE* represents the typical amount of error that can be expected from an estimator so it tells you how precise your estimate of said statistics is likely to be (McHugh, 2008; Biau, 2011; Harding, Tremblay & Cousineau, 2014).

As the standard error is a type of standard deviation, the two terms are often mistaken (Altman & Bland, 2005; Harding, Tremblay & Cousineau, 2014). The following example can help you to understand the conceptual difference between these two terms.

Assume a large enough set of samples, each with their own numbers of observations and mean values. These means might vary from sample to sample. The sampling distribution of the means has its own mean (i.e., a mean of means) and, therefore, standard deviation which we call the standard error of the estimate of the mean.

On the Importance of Standard Errors

Standard errors are critical to a statistical analysis. Once a standard error is properly computed, what do we do with it? A lot!

First, the statistical estimate of interest, divided by its standard error, gives us a way of testing whether said statistic is significantly different from zero. A second application of the standard error is the calculation of confidence intervals. A third application consists in testing whether any two statistics of the same kind are significantly different.

As we can see, properly computing *SE* values is extremely important. Any blunder in their calculations invalidates the results or conclusions of a study. In the next sections we show that *SEs* can be computed for other statistical estimators, not just the mean. One can even compute the *SE* of a standard deviation.

The Standard Error of the Mean

This is probably the best known *SE*. The standard error of the mean is the standard deviation of the sample mean estimate $\bar{x}$ of a population mean (Kurtz, 1991); i.e.

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}} \quad (1)$$

where n is the number of the x_i observations and s is their standard deviation.

Contrary to misconceptions, (1) does not assume a normal distribution so it can be applied to any type of distribution. However, to compute s and $\bar{x}$ the x_i observations must be additive.

Notice from (1) that as more number of observations are taken the standard error decreases which not necessarily is the case with the standard deviation, s.

The Standard Error of the Median

For large and approximately normal samples the *SE* of the median can be approximated as

$$SE_m = \sqrt{\frac{\pi}{2}} * SE_{\bar{x}} \approx 1.25 * \frac{s}{\sqrt{n}} \quad (2)$$

Since the median tends to be less reliably estimated than the mean (by 25%), this is one of the reasons that the mean is usually preferred over the median (Stockburger, 1996; Harding, Tremblay & Cousineau, 2014). Notice that (2) assumes a normal distribution.

This formula can yield wrong results for extremely non-normal distributions. Although SE_m is computed when the data is normally distributed, bootstrapping SE_m avoids this requirement. Bootstrap is also useful for computing confidence intervals and standard errors of difficult statistics like the median (Eichler, 2003; Efron &Tibshirani, 1993).

The Standard Error of the Standard Deviation

The *SE* of the standard deviation (SE_s) is about 71% that of the mean,

$$SE_s = 0.71 \frac{s}{\sqrt{n}} \tag{3}$$

The distribution of the standard deviation is positively skewed for small *n* and approximately normal if the sample size is 25 or greater. Procedures for calculating the area under the normal curve work for the sampling distribution of the standard deviation as long as the sample size is at least 25 and the distribution is approximately normal.

The Standard Error of a Correlation Coefficient

The *SE* of a correlation coefficient *r* is computed by normalizing the fraction of the unexplained variations with respect to $n - 2$ degrees of freedom; i.e.

$$SE_r = \frac{\sqrt{1-r^2}}{\sqrt{n-2}} \tag{4}$$

In (4), r^2 is the Coefficient of Determination which expresses the fraction of the explained variations; e.g., variations in *y* as the result of variations in *x*. To illustrate, if r^2 is 0.90, the independent variable *y* is said to explain 90% of the variance in the dependent variable *x*, but does not explain $1 - r^2$ or 10% of the variance in the dependent variable.

The Standard Error of *Z* Scores

The Fisher Transformation (Fisher, 1915; 1921; 1924) converts a correlation coefficient into a *Z score* also known as a *normal score* (Wikipedia, 2016). Said scores should not to be mistaken for *z-standardized scores* which are computed from a sample of mean-centered data as $z = \frac{x_i - \bar{x}}{s}$.

By contrast, the Fisher Transformation is computed, for each *r* value, as follows:

$$Z = \frac{1}{2}[ln(1 + r) - ln(1 - r)] \tag{5}$$

This *r*-to-*Z* transformation has its inverse, *Z*-to-*r*, which is computed as

$$r = \frac{e^{2Z} - 1}{e^{2Z} + 1} = \frac{e^{Z} - e^{-Z}}{e^{Z} + e^{-Z}} \tag{6}$$

The standard error associated to a *Z* value is

$$SE_z = \sqrt{\frac{1}{n - 3}} = \frac{1}{\sqrt{n-3}} \tag{7}$$

Expressions (5) and (6) are not valid when *r = 1* exactly, but this is not a real issue because in most experimental problems *r = 1* exactly is not achievable.

Limitations of the *r*-to-*Z* Fisher Transformation

The *r*-to-*Z* Fisher Transformation and its inverse should not be applied arbitrarily, but only when both random variables, *x* and *y*, are approximately normally distributed. Ignoring this requirement can induce a researcher to draw misleading conclusions (Garcia, 2012a; 2012b; 2015a; 2015b).

Zimmerman, Zumbo, and Williams (2003) have shown that arbitrarily applying this transformation, especially from distributions that violate bivariate normality can lead to spurious results, even with large sample sizes.

According to these authors, "…significance tests of hypotheses about validity and reliability coefficients or differences between them require an assumption of bivariate normality despite large sample sizes. Researchers certainly should be aware of this assumption before using the *r* to *Z* transformation in data analysis."

Bond and Richardson (2004) have published the first geometrical visualizations of these transformations to date, which are given in Figure 1.

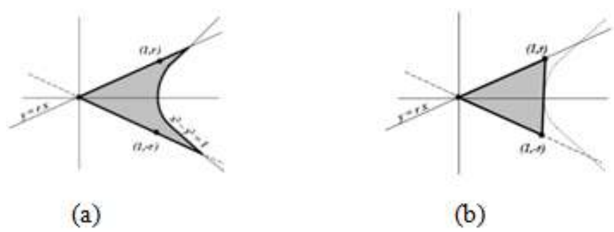

Figure 1. *Z* (a) and *r* (b) as areas in scatterplots. Source: Bond & Richardson, 2004.

Clearly the requirement of bivariate normality comes from the transformation itself. An *r*-to-*Z* transformation is an inverse hyperbolic tangent function while the *Z*-to-*r* transformation is a hyperbolic tangent function. Violation of bivariate normality distorts the areas shown in Figure 1.

These transcendental transformations are available in most computer programming languages and software. For instance, these functions are built-in in the scientific calculator of Windows computers, and Microsoft Excel has these as the ATANH and TANH functions. We have built our own tool to take care of these transformations (Garcia, 2016). Unlike similar tools, ours accepts an entire set of values, transforming these accordingly.

Pooled Standard Errors

To test if two correlation coefficients, r_1 and r_2, are significantly different, these are transformed into Fisher *Z-scores*, provided that these come from bivariate distributions. Their difference, computed as $Z_1 - Z_2$, is tested using a pooled standard error which is defined as follows:

$$SE_z = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \quad (8)$$

Beware of Incorrect Standard Error Calculations

A statistical analysis makes sense only if there is a sound theory behind it. Using incorrect *SE* definitions invalidates arguments stated or implied in any statistical study.

For instance, some search engine marketers at SEOmoz.org, now MOZ.org, generated some so-called "scientific" studies using standard errors of correlation coefficients. It was later acknowledged that the *SE* of a correlation coefficient was literally computed as $\frac{s}{\sqrt{n}}$ where s was a standard deviation computed out of several correlation coefficients (SEOmoz 2010a; 2010b). That is, they used the definition of *SE* of the mean (1) and applied to correlation coefficients!

In addition to be a serious blunder or *statistical horror*, another problem with their approach is that to compute such a standard deviation one would need to compute a mean correlation coefficient in the first place. But there is a problem: correlation coefficients are not additive. This is a known fact in the scientific community and across disciplines (Zsak, 2006; Hetti-Arachchilage & Piontkivska, 2016). In her thesis, Zsak remarks: "The correlation coefficient is not a linear function of the size of relation between the variables that are correlated and thus cannot simply be summed to find a mean that is representative as the mean value for the correlation coefficients."

In addition, StatSoft, creators of Statistica, now a Dell company (Dell, 2016) have stated in their literature: "Are correlation coefficients "additive?" No, they are not. For example, an average of correlation coefficients in a number of samples does not represent an "average correlation" in all those samples. Because the value of the correlation coefficient is not a linear function of the magnitude of the relation between the variables, correlation coefficients cannot simply be averaged."

Formally, a correlation coefficient is a function of the covariance between two variables, normalized by their standard deviations,

$$r = \frac{covar\,(x,y)}{s_x * s_y} \tag{9}$$

where the covariance is defined in terms of the expectation (mean) values of the variables; i.e.

$$r = \frac{E(x*y) - E(x)*E(y)}{s_x * s_y} \tag{10}$$

and where *E(x*y)*, *E(x)*, and *E(y)* are expectation values.

Because (9) and (10) are specific to a sample, *r* values are dissimilar ratios. These types of ratios are not additive. Furthermore, for mean-centered variables, Pearson's *r* is a cosine. Cosines are not additive either (Garcia, 2012a; 2012b; 2015a; 2015b). We must conclude that it is not possible to add *r* values and then average these to compute a standard deviation of correlation coefficients.

In the case of a Spearman Correlation Coefficient, the mere idea of constructing a linear function by averaging Spearman values is highly questionable because what is considered are ranks, not the magnitude of the relation between variables.

As correlation coefficients (Pearson's *r*, Spearman's *rho*, others) are not additive, we cannot compute arithmetic averages from these. The same holds for standard deviations and standard errors. Political correctness aside, stating the contrary equates to a futile effort of defending "quack science" (Garcia, 2010).

The fact that SEOmoz (MOZ) search marketers computed the *SE* of correlation coefficients using the definition ascribed to the mean indicates a lack of knowledge about sampling distributions from their part.

Conclusion

In this tutorial we explained the difference between standard deviations and standard errors. The standard error is the standard deviation of a sampling distribution of a statistical estimate. It is not the standard deviation of a sample of raw observations.

The proper way of computing the standard errors of several statistical estimates was discussed along with the non-additivity of correlation coefficients. We state that improperly computing standard errors invalidates any statistical study.

References

Altman, D. G., Bland, J. M. (2005). Standard deviations and standard errors. BMJ 2005;331:903 (15 October), doi:10.1136/bmj.331.7521.903. Retrieved from http://www.bmj.com/cgi/content/extract/331/7521/903

Arsham, H. (1994). Statistical Thinking for Managerial Decisions. Retrieved from http://home.ubalt.edu/ntsbarsh/business-stat/opre504.htm

Bond, C. F. and Richardson, K. (2004). Psychometrika, Vol 69, 2, 291-303. Retrieved from http://faculty.tcu.edu/richardson/seeing/seeingf060403.pdf

Dell (2016). What are Basic Statistic? Retrieved from http://www.statsoft.com/textbook/basic-statistics/#Correlationso

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall.

Eichler, M (2003). Introduction to the Bootstrap. Retrieved from http://galton.uchicago.edu/~eichler/stat24600/Handouts/bootstrap.pdf

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika, 10, 507-521. Retrieved from http://digital.library.adelaide.edu.au/dspace/bitstream/2440/15166/1/4.pdf

Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. Metron 1, 3-31. Retrieved from https://digital.library.adelaide.edu.au/dspace/bitstream/2440/15169/1/14.pdf

Fisher, R. A. (1924). The distribution of the partial correlation coefficient. Metron 3, 329-332. Retrieved from https://digital.library.adelaide.edu.au/dspace/bitstream/2440/15182/1/35.pdf

Garcia, E. (2010). Beware of SEO Statistical Studies. Retrieved from http://irthoughts.wordpress.com/2010/04/23/beware-of-seo-statistical-studies/

Garcia, E. (2012a). The Self-Weighting Model, Communications in Statistics - Theory and Methods, 41:8,1421-1427. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/03610926.2011.654037

Garcia, E. (2012b). True Lies: Search Marketing Data Errors Uncovered. Retrieved from http://web.archive.org/web/20121101095532/http://sesconference.com/archive/2012/newyork/agenda-day1.php

Garcia, E. (2015a). The Self-Weighting Model Tutorial: Part 1 Retrieved from http://www.minerazzi.com/tutorials/self-weighting-model-tutorial-part-1.pdf

Garcia, E. (2015b). The Self-Weighting Model Tutorial: Part 2 Retrieved from http://www.minerazzi.com/tutorials/self-weighting-model-tutorial-part-2.pdf

Garcia, E. (2016). Fisher Transformations Tool. Retrieved from http://www.minerazzi.com/tools/fisher/transformations.php

Harding, B., Tremblay, C., and Cousineau, D. (2014). Standard errors: A review and evaluation of standard error estimators using Monte Carlo simulations. TQMP Vol 10, 2. Retrieved from http://www.tqmp.org/RegularArticles/vol10-2/p107/p107.pdf

Hetti-Arachchilage, M. and Piontkivska, H. (2016). Coevolutionary Analysis Identifies Protein–Protein Interaction Sites between HIV-1 Reverse Transcriptase and Integrase. Retrieved from http://ve.oxfordjournals.org/content/vevolu/2/1/vew002.full.pdf

Kurtz, M. (1991). *Handbook of Applied Mathematics for Engineers and Scientists*, Chapter 10, p10.4. McGraw-Hill, New York.

McHugh, M. L. (2008). Standard error: meaning and interpretation. Biochemia Medica 2008;18(1):7-13. Retrieved from http://dx.doi.org/10.11613/BM.2008.002

SEOmoz (2010a). The Science of Ranking Correlations. Retrieved from http://www.seomoz.org/blog/the-science-of-ranking-correlations

SEOmoz (2010b). Google vs. Bing Correlation Analysis of Ranking Elements. Retrieved from http://www.seomoz.org/blog/google-vs-bing-correlation-analysis-of-ranking-elements

Stockburger, D. W. (1996). Introductory Statistics: Concepts, Models, and Applications. Retrieved from http://www.psychstat.missouristate.edu/introbook/sbk19m.htm

Wikipedia (2016). Fisher Transformation. Retrieved from https://en.wikipedia.org/wiki/Fisher_transformation

Zimmerman, D. W., Zumbo, B. D., and Williams, R. H. (2003). Bias in estimation and hypothesis testing of correlation. Psicológica 24:133-158. Retrieved from http://www.uv.es/revispsi/articulos1.03/9.ZUMBO.pdf

Zsak, M. I. (2006). *Decision Support for Energy Technology Investments in Built Environment*. Master Thesis. Norwegian University of Science and Technology. Retrieved from http://www.iot.ntnu.no/users/fleten/students/tidligere_veiledning/Zsak_V06.pdf