2004

Random Error

Gudmund R. Iversen
Swarthmore College, iversen@swarthmore.edu

Recommended Citation

Gudmund R. Iversen. (2004). "Random Error". *The SAGE Encyclopedia Of Social Science Research Methods.* Volume 3, 908-909. DOI: 10.4135/9781412950589.n798
https://works.swarthmore.edu/fac-math-stat/213

The SAGE Encyclopedia of Social Science Research Methods

Random Error

Contributors: Gudmund R. Iversen
Edited by: Michael S. Lewis-Beck, Alan Bryman & Tim Futing Liao
Book Title: The SAGE Encyclopedia of Social Science Research Methods
Chapter Title: "Random Error"
Pub. Date: 2004
Access Date: March 13, 2020
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9780761923633
Online ISBN: 9781412950589
DOI: http://dx.doi.org/10.4135/9781412950589.n798
Print page: 909

Randomness lies at the heart of STATISTICAL INFERENCE. It is based on the notion that if an observation or a measurement is repeated, then most of the time, a different value is observed.

Randomness exists when it is not possible to predict the outcome of the next observation, either in a SURVEY or an EXPERIMENT. It follows that, in the presence of randomness for a variable, there will be RANDOM VARIATIONS in the observations, and the variable becomes a random variable. The magnitude of such randomness is measured by the random error.

The variation in the observations of a random variable leads to variations in any sample statistic computed from the data. Thus, a sample mean will vary from sample to sample, just as the percentage of voters who support a candidate varies from one sample of voters to another.

The variation in a sample statistic from one sample to the next can be seen when multiple samples are taken. A particular polling organization may not do repeated sampling, but prior to any major election, several polling organizations do their own studies, and the percentages they report vary from one study to the next. One of the reasons they vary is because of the randomness inherent in any sampling procedure. It is also possible to simulate such variation with the proper statistical software. Simulating samples of size n =1,000 and a probability of 0.5 of drawinga1or a 0, these are the first few sample percentages of draws that are 1: 50.4, 50.4, 48.4, 51.4.

Such variation is known as the random error associated with the percentage. It is unfortunate that the word ERROR is associated with this variation, because there is nothing “wrong” with such variation. But to the extent that there exists one true percentage value in the population, any deviation of a sample percentage from this true POPULATION parameter can be thought of as an error.

Historical Account

Physicists have long been concerned with how to measure characteristics of physical objects. Even repeated measures of the length of a yardstick result in different measurements, even though we realize that the yardstick has only one, fixed length (given fixed temperature, etc.). Physicists finally accepted this variation, and they attached the name “error” to the difference between an observed value and the true, underlying value. Statisticians carried on with the term when they observed different values of a statistic from different samples, even though this variation is due to sampling effects.

Applications

The existence of the random error leads to the question of what the magnitude is of this error. If the magnitude of the random error is known, then it is possible to state whether a particular sample statistic exceeds the random error or not. How much would we expect a sample percentage to vary around the true parameter value? Statisticians have developed formulas for the computation of magnitudes of the random error. If such formulas do not exist, it is often possible to simulate many repeated samples with the proper software and thereby compute the magnitude of the random error.

Examples

One way to measure the magnitude of the random error is to find the standard deviation of a sample statistic across many different samples. Such a standard deviation is mostly known as the standard error of a sample statistic, to distinguish it from the standard deviation we compute from a set of single observations. For a sample of size 1,000 from a population split 50/50, the standard error of the sample percentage becomes 1.6.

In most cases, two times the standard deviation, or two times the standard error, will include most of the data. Twice 1.6 equals 3.2, and most of the sample percentages from various samples should fall in the range from 50 − 3.2 = 46.8 to 50 + 3.2 = 53.2.

However, we are interested most often in whether a particular sample statistic lies within the random error inherent in the sampling procedure, or whether the sample statistic lies further away from the presumed value of the population parameter. Suppose we hypothesize that a population percentage equals 50%. From a sample of 1,000 observations, the observed sample percentage equals 56.7%. Does 56.7% lie within the range around 50% produced by the random error, or is 56.7% beyond the random error produced by the sampling procedure?

We can now conclude that the sample percentage lies beyond the error that can be expected by the sampling error. We could change the sample percentage to a value of the standard normal variable. With a value larger than 1.96, we conclude that the deviation of the sample statistic from the hypothesized population parameter exceeds what can be expected on the basis of the random error itself. Here, $z = (56.7 - 50.0)/1.58 = 4.24$. For less than once in 10,000 different samples would we observe such a large value or a larger value of the standard normal variable. From this, we conclude that the hypothesized value of 50% for the population in its support of the candidate must be wrong. For the observed sample value to lie within the random error of the population percentage, the population percentage must lie closer to the observed sample value of 56.7%.

Any sample statistic will have its associated random error, as expressed in its standard error. We have some control of the magnitudes of the random errors from the way the sample is collected and which variables are used in a particular study. One way to make the random error smaller is to have a larger sample. The number of observations in the sample often figures directly into the computation of the standard error of a sample statistic. One drawback is that the sample size often enters the formulas through its square root, and it takes much larger sample sizes to get the desired effect. Because of the square root, we need a sample four times as large to get half the standard error. In multiple regressions, collinearity will increase the standard error of the regression coefficients.

- random error
- standard errors
- errors
- sampling
- sampling error
- random variable
- population parameters

Gudmund R. Iversen
http://dx.doi.org/10.4135/9781412950589.n798
10.4135/9781412950589.n798

References

Agresti, A., & Finley, B.(1997). Statistical methods for the social sciences (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
Iversen, G., & Gergen, M.(1997). Statistics: The conceptual approach. New York: Springer-Verlag.
Moore, D. S., & McCabe, G. P.(1998). Introduction to the practice of statistics. New York: W. H. Freeman.