DOCUMENT RESUME

ED 426 100 TM 029 330

AUTHOR	Lewis, Charla P.
TITLE	Understanding the Sampling Distribution and the Central Limit Theorem.
PUB DATE	1999-01-00
NOTE	13p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, January 21-23, 1999).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Estimation (Mathematics); *Probability; Sample Size; *Sampling; *Statistical Distributions; Statistical Significance
IDENTIFIERS	*Central Limit Theorem; Mean (Statistics)

ABSTRACT

The sampling distribution is a common source of misuse and misunderstanding in the study of statistics. The sampling distribution, underlying distribution, and the Central Limit Theorem are all interconnected in defining and explaining the proper use of the sampling distribution of various statistics. The sampling distribution of a statistic is used to find probabilities of research outcomes and is one of the key concepts in statistical significance testing. Sampling distributions are the frequency distributions of a particular sample's statistics and contain infinitely many statistics for a given sample size from a population. The most common sampling distribution is the sampling distribution of the mean. The mean of this distribution is assumed to be the true mean of the population. There are generalizations, qualities, and rules that have to be observed in order for the sampling distribution to produce parameter estimates that can be used to make experimental inferences. The exact use of the sampling distribution in significance testing, the future of significance testing in the study of behavior, and alternatives to employing statistical significance testing in the traditional sense are also explored. (Contains 28 references.) (Author/SLD)

Understanding the Sampling Distribution and the Central Limit Theorem

Charla P. Lewis
Texas A&M University

Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, January, 1999.

Abstract

The sampling distribution is a common source of misuse and misunderstanding in the study of statistics. The sampling distribution, underlying distribution, and the Central Limit Theorem are all interconnected in defining and explaining the proper use of the sampling distribution of various statistics. The sampling distribution of a statistic is used to find probabilities of research outcomes and is one of the key concepts in statistical significance testing. Sampling distributions are the frequency distributions of a particular sample's statistic and contains infinitely many statistics for a given sample size from a population. The most common sampling distribution is the sampling distribution of the mean. The mean of this distribution is assumed to be the true mean of the population. There are generalizations, qualities, and rules that have to be observed in order for the sampling distribution to produce parameter estimates that can be used to make experimental inferences. The exact use of the sampling distribution in significance testing, the future of significance testing in the study of behavior, and alternatives to employing statistical significance testing in the traditional sense is also explored.

Statistics, even to the most astute graduate student, can be like a foreign language and a source of considerable apprehension and worry. A professor or an article that explains the complexity of concepts and theories is a rare, treasured find. It is this author's intent to clarify what may have been a slightly mystical topic in introductory statistics material and how that topic applies to more advanced understanding of the inferences made through data analysis in the study of behavior. The sampling distribution of the statistic (e.g., SE_M, SE_{SD}) also called the underlying distribution, is one of the most referenced, yet least understood concepts in the statistical world. The present discussion will include the definition and uses of the sampling distribution, the Central Limit Theorem and how it applies to understanding sampling distributions, and the inferences based on information gathered from sampling distributions.

Statistical Significance Testing

For decades, education and psychology researchers have focused on testing the statistical significance of outcomes.

Unfortunately, the concept of statistical significance is all too often fundamentally misunderstood and consequently, used inappropriately (cf. Cohen, 1994; Schmidt, 1996; Thompson, 1996, in press). For example, Meehl (1978, P. 817, 823) argued some 15 years ago:

> I believe that the almost universal reliance on merely refuting the null hypothesis as the standard method for corroborating substantive theories in the soft [i.e., social science] areas is a terrible mistake, is basically unsound, poor scientific strategy, and one of the worst things that ever happened in the history of psychology... I am not making some nit-picking statistician's correction. I am saying that the whole business is so radically defective as to be scientifically almost pointless.

More recently, Rozeboom (1997) noted that:

> Null-hypothesis significance testing is surely the most bone-headedly misguided procedure ever institutionalized in the rote training of science students... [I]t is a sociology-of-science wonderment that this statistical practice has remained so unresponsive to criticism... (p. 335)

Rightly or wrongly, the probability of a study's result is one of the most important qualities considered by editors of professional journals and the scientists and practitioners who are the consumers of these journals (cf. Greenwald, 1975; Rosenthal, 1979). The assumed importance of these findings has been used to pattern new studies and serve as the building blocks of theory and further investigations of psychological phenomena. The determination of a statistically significant result does not necessarily indicate a practically significant explanation of a given topic under study (Cohen, 1994; Snyder & Lawson, 1993). The probability that something takes place does not automatically constitute importance, replication, or any far-reaching meaning. The sampling distribution is the concept that aids in the derivation of statistically significant results, however the practical use of these results is a different discussion.

Defining the Sampling Distribution

According to Breunig (1995), "The sampling distribution is one of the fundamental concepts underlying all inferential procedures" (p. 3). In the statistical processes in experimental study, researchers are trying to make inferential statements about their sample and how it relates to the overall population to which they are hoping the discoveries will generalize. Statistics are the values that come from the sample. They are used to hypothesize and learn about qualities of the population and as estimators of values in the population (Breunig, 1995), or parameters, as they are called. The sampling distribution is the frequency distribution of a sample statistic (e.g., mean) across infinitely many samples from the population, each with exactly the same sample size as the sample (Hinkle, Wiersma, & Jurs, 1994).

Put simply, *populations* are distributions of scores (usually taken to be N equals infinitely man scores). *Samples* are distributions of scores (i.e., subsets of the population of size n). Sampling distributions do not contain scores (except in the unusual case in which the sample size is n=1, where the mean of a given sample also equals the only score in that given sample). Instead, *sampling distributions* consist of infinitely many *statistics* (i.e., estimated population parameters), each based on samples of exactly size n.

Empirically derived sampling distributions take many concepts into consideration in development. Many different statistics can be used to make the sampling distribution. The most common example is the distribution of the means and typically, the mean of sampling distribution is mean of population (Hinkle et al., 1994). As random sample means are collected, they began to pile up around a central value. This central value is the most common sample mean (mode) and is the population mean (parameter). The standard error of the statistic (e.g., mean, SD, r) is the standard deviation of the sampling distribution. The standard error tells you how "spread out" are the sample estimates of a given population parameter (e.g., SD, r). Breunig (1995) illustrates with a small population and hand calculations the practical application of these principles. The standard error tells how good an estimator of the parameter the sample statistic is.

Form of the Distribution

The shape of the sampling distributions is explained by a mathematical theorem called the central limit theorem. The central limit theorem (CLT) in Hinkle (1994) says that :

> given a population with a mean equal to μ, a variance equal to o^2, and with n number of groups of random samples, as sample size increases, the

> sampling distribution of the mean approximates a normal distribution (p. 172).

There are also two generalizations about the sampling distribution that work in conjunction with CLT (Gravetter & Wallnau, 1985; Hinkle et al., 1994). The first is that *as sample size increases, variability of the distribution decreases.* The larger the samples get, more and more values are taken into account in developing the distribution. Statistics began to fall around the same areas as in the population's distribution. There is less variety and difference in the values of the sample versus the population. Therefore, the sample distribution begins to look more like the population.

The second generalization is that *even when the population isn't normally distributed, the shape of the sampling distribution becomes more like normal as sample size increases.* Figures can illustrate this idea. According to Hinkle et al. (1994), when sample size is greater than 30, the sampling distribution for the mean is very similar to a normal distribution, even if the population is skewed.

Estimators

Everything in statistical significance testing is based on the premise that we can infer experimental findings of a sample to explain events and treatment efficacies in the population. This process is called statistical estimation. Our statistics attempt to estimate the parameters. The sampling distribution of a statistic (e.g., r, SD), which is an integral part of statistical significance testing, aids in estimating the mean of the population. There are many kinds and ways of estimation. The shape of the sampling distribution is an approximation of the normal curve as suggested by the CLT (Stacy, 1981). The expected value of the population is what is called a point estimate. This point estimate estimates where in the distribution of sample statistics our population parameter is.

When making statistical inferences, scientists want a good, unbiased estimator. Bias can be caused by innumerable factors, the most important of being use of non-random samples. A statistic is an unbiased estimator if the mean of its sampling distribution is the parameter estimated. Basically, the mean of all possible random samples is most likely the real parameter for the population (Breunig, 1995).

In addition to helping with the generation of sampling distributions, mathematical theorems are also used to get unbiased estimators that describe characteristics of the sampling distribution (Hinkle et al., 1994). Good estimators are unbiased, consistent, efficient, and sufficient (Harnett, 1970). Unbiasness estimator means that the parameter estimate (sample statistic) and the population parameter value are the same. If this can't be, an estimate with a small amount of bias is next preferred (Rennie, 1997). Consistency is a function of the central limit theorem and is the tendency of estimates to become closer to the actual population parameter as sample size increases. The bigger the sample, the more similar to the population it is, and the greater the probability that the statistic will mirror the true population parameter. Efficiency is related to bias. The closer an estimate is to the parameter, given sample size, the more efficient it is (Mittag, 1992). If there are two statistics, A and B, and A has less error than B, for a given sample size, A is the more efficient estimate. Sufficiency is defined as using all the information from the population in getting estimates from the samples.

Uses for the Sampling Distribution

Statistical significance testing, sampling distributions, parameter estimates and event probabilities are directed toward the purposes of finding statistical significance. Most studies operate on the premise of the "nil" hypothesis (Cohen, 1994), that is there is no difference between treatment groups or no relationship among variables instituted in experimental studies (Hinkle et al., 1994). When the null is rejected, it is proposed that a difference exists between groups and a statistically "significant" effect has been noted. However, the conclusion says nothing about effectiveness within the population (Cohen, 1994; Thompson, 1996). Assuming that the null is true in the population, the sampling distribution can be used to obtain a probability of the sample statistics (i.e., not of the population parameters). The sampling distribution is the mechanism by which we obtain the $p_{CALCULATED}$ for a given study (see Hinkle et al., 1994). We do this by locating the sample statistic in the sampling distribution, and then finding the proportion of area in the sampling distribution with this sample statistic or sample statistics that diverge even more extremely from the actual sample value. This area equals $p_{CALCULATED}$, and is sometimes called the "region of rejection".

The use of the sampling distribution in regards to alternate hypotheses is somewhat different than with the nil hypothesis. Becker (1991) stated that :

> " probability values under alternative models are much more complex... and depend on the type of statistical test used, sample size, and the values of relevant population parameters" (p. 347).

Many statistical tests in education and psychology use t and F statistics. These theoretical approximations to sampling distributions are easier to compute and use (Becker, 1991; Edgington & Haller, 1984).

The Future of Significance Testing

Statistical significance testing has been under attack for many years (cf. Cohen, 1994). The argument is that p values are often misunderstood by both research consumers and researchers. Common misinterpretations of p values include believing that p is the probability that the experiment's group difference is "due to chance" (e.g., with a $p<.05$, findings have a less than 5 in 100 chance of being "incidental"), saying that p is the probability that the null hypothesis is true in the population, using p as a percentage likelihood of not getting the same results in a replication study, or that p is the probability that research results are incorrect (Cohen, 1994; Gall, Borg, & Gall, 1996; Thompson, 1996).

In the past and recent years, the merit and publishing of research has been based solely on the p value. However, many factors affect finding statistical significance, especially sample size. One will get statistically significant results in an any study, if a large enough sample is used (Rennie, 1997; Thompson, 1989). Even when statistical significance is concluded, practical significance and meaningfulness are still in question. Alternatives to significance testing have been suggested (Cohen, 1994; Gall et al., 1996; Thompson, 1989, 1993, 1994a, 1996) regarding using replicability techniques, such as bootstrap and cross-validation, noting and interpreting effect sizes, doing statistical power analyses, changing the use of the term "significance" to "statistical significance", and looking at sample size ("what if" analyses) as it affects p calculated to fortify reports of research findings.

Conclusions

Although there has been much debate over the uses of statistical significance testing, many students and professionals rightly or wrongly still subscribe to its value. Sampling distributions are a fundamental part of finding statistically significant results. Knowing the functions of the statistics, parameters, and theorems involved in constructing the sampling distribution facilitate a more informed use of the probabilities of sample events and what implications they really have for scientific studies. The sampling distribution provides probabilities that represent the chances an outcome has in the given sample. In discussing some of the key aspects of the sampling distribution and the central limit theorem, hopefully a better understanding statistical significance has been furthered.

References

Becker, B.J. (1991). Small-sample accuracy of approximate distributions of functions of observed probabilities from t tests. Journal of Educational Statistics, 16, 345-369.

Bedian, V. (1979). Evaluation of tests and decision analysis. Applications of probability and statistics to medicine. Modules and monographs in undergraduate mathematics and its applications. UMAP module 377. Newton, MA: Education Development Center.

Borg, M.D., Borg, W.R., & Gall, J.P. (1996). Educational research: An introduction, (6th ed). White Plains, NY: Longman.

Breunig, N.A. (1995, November). Understanding the sampling distribution and its use in testing statistical significance. Paper presented at the annual meeting of the Mid-South Educational Research Association, Biloxi, MS. (ERIC Document Reproduction Service No. ED 393 939)

Cohen, J. (1994). The Earth is round ($p < .05$). American Psychologist, 49, 997-1003.

Edgington, E.S., & Haller, O. (1984). Combining probabilities from discrete probability distributions. Educational and Psychological Measurement, 44, 265-274.

Fuhr, N., & Huther, H. (1989). Optimum probability estimation from empirical distributions. Information Processing & Management, 25, 493-507.

Gravetter, F.J., & Wallnau, L.B. (1985). Statistics for the behavioral sciences: a first course for students of psychology and education, (2nd ed). St. Paul, MN: West Publishing.

Greenwald, A.G. (1975). Consequences of prejudice against the null hypothesis. Psychological Bulletin, 82, 1-20.

Harnett, D.L. (1970). Introduction to statistical methods. Reading, MA: Addison-Wesley.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1994). Applied statistics for the behavioral sciences. (3rd ed). Boston, MA: Houghton Mifflin.

Losee, R.M. (1988). Parameter estimation for probabilistic document-retrieval models. Journal of the American Society for Information Science, 39, 8-16.

Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46, 806-834.

Breunig, N.A. (1995, November). Understanding the sampling distribution and its use in testing statistical significance. Paper presented at the annual meeting of the Mid-South Educational Research Association, Biloxi, MS. (ERIC Document Reproduction Service No. ED 393 939)

Mittag, K.C. (1992, January). Correcting for systematic bias in sample estimates of population variances: Why do we divide by n-1? Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX. (ERIC Document Reproduction Service No. ED 341 728)

Piotrowski, R.J. & Siegel, D.J. (1986). The IQ of learning disability samples: A reexamination. Journal of Learning Disabilities, 19, 492-293.

Rennie, K.M. (1997, January). Understanding the sampling distribution: Why we divide by n-1 to estimate the population variance. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 406 446)

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. Psychology Bulletin, 86, 638-641.

Rozeboom, W.W. (1997). Good science is abductive, not hypothetico-deductive. In L.L. Harlow, S.A. Mulaik, & J.H. Steiger (Eds.), What if there were no significance tests? (pp. 335-392). Mahwah, NJ: Erlbaum.

Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. Psychological Methods, 1 (2), 115-129.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. Journal of Experimental Education, 61, 334-349.

Stacy, E. W., Jr. (1981). On defining at-risk status. Evaluation and Program Planning, 4, 363-375.

Thompson, B. (1989). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues. Measurement in Counseling and Development, 22, 2-5.

Thompson, B. (Ed.). (1993). Theme issue: Statistical significance testing in contemporary practice. Journal of Experimental Education, 61 (4).

Thompson, B. (1994a). The concept of statistical significance testing. Measurement Update, 4 (1), 5-6. (ERIC Document Reproduction Service No. ED 366 654)

Thompson, B. (1994b, April). Common methodology mistakes in dissertations, revisited. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 368 771)

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Education Researcher, 25 (2), 26-30.

Thompson, B. (in press). Statistical significance and effect size reporting: Portrait of a possible future. Research in the Schools.

TM 029330

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: UNDERSTANDING THE SAMPLING DISTRIBUTION AND THE CENTRAL LIMIT THEOREM

Author(s): CHARLA P. LEWIS

Corporate Source:

Publication Date: 1/99

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

☒ **Check here** Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction	**Sample sticker to be affixed to document** "PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY CHARLA P. LEWIS TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)." **Level 1**	**Sample sticker to be affixed to document** "PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)." **Level 2**	☐ **or here** Permitting reproduction in other than paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: X	Position: RES ASSOCIATE
Printed Name: CHARLA P. LEWIS	Organization: TEXAS A&M UNIVERSITY
Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone Number: (409) 845-1831
	Date: 11/11/98