Florida Journal of
Educational Research
Fall 1988, Vol. 30,
No. 1, Pp. 5-14

Statistical Rules-of-Thumb

James K. Brewer
Florida State University

ABSTRACT. Six best-selling introductory behavioral statistics and two well-known sampling theory textbooks were reviewed for the presence of rules-of-thumb. The relative frequency and type of rules are reported along with a discussion of their implications for teaching statistics at the introductory level.

The American Heritage Dictionary of the English Language (1973) defines a rule-of-thumb as, "a useful principle with wide application, not intended to be strictly accurate." Although the origin of the term is debatable and ranges from the use of the thumb in general measuring to the maximum diameter of the stick with which a man could beat his wife in old England, these "rules" are generously used in statistics. Whether their prevalence is due to statisticians' desire for "wide application" or their abhorrence of being "strictly accurate" is not clear. The purpose of this paper is to identify and categorize some common rules-of-thumb used in statistics and to discuss problems with their invocation in the teaching of statistics.

Rules-of-Thumb Types

To illustrate the prevalence and use of rules-of-thumb in statistics instruction, six introductory level behavioral statistics textbooks were reviewed. The introductory level was chosen since it is here that rules-of-thumb are probably most often invoked due to the limited mathematical expectations of the readers of such texts. Additionally, this is the level of exposure most likely to assure future invocations of the rules-of-thumb. I refer to this as the instructional level where statistical **imprinting** occurs (Brewer, 1985).

The six textbooks reviewed were all published in 1982 and were described by their publishers as the "best sellers" in the introductory behavioral statistics area. The content of the texts included descriptive, correlational, and inferential statistics through ANOVA; no text contained multiple regression. The names and publishers of these texts are available upon request.

Each textbook was analyzed for statements of rules-of-thumb or statements which reflect the use of such rules. An example or two will help clarify how a statement was classified as a rule-of-thumb or a reflection thereof. If a statement, for example, was made, "The binomial is closely approximated by the normal for large sample sizes," it was not considered a rule-of-thumb. However, if a statement such as "When $n P > 5$ or $n Q > 5$, then the normal may be used to approximate the binomial," was made, it was classified as a rule-of-thumb. The use of the Central Limit Theorem offers another example. Invoking the Central Limit Theorem in any of its forms was not considered a rule-of-thumb, but a statement such as "When $n \geq 30$, the Central Limit Theorem allows us to . . .", was considered a rule-of-thumb.

There were 14, 15, 20, 23, 29, and 30 such statements located in the six textbooks for a total of 131 rules-of-thumb. In many cases there were repeated applications of the same rule, but only initial statement counts were recorded. If repeated invocations of the same rule were counted the total would be well over 200. Understandably, quite a few rules-of-thumb were invoked across all textbooks, but no attempt was made to identify unique usages per textbook.

The rules-of-thumb noted and recorded were classified into four, not-so-mutually-exclusive categories of statistical descriptions (D), sample size minimums (N), sampling distribution approximations (DA), and inference-making (I). The frequency (and relative frequency) of these categories of rules are shown in Table 1.

Table 1 Categories of Rules-of-Thumb

Rule occurrence	Statistical descriptions D	Sample size N	Sampling distributions DA	Inference making I	Total
frequency	33	32	25	41	131
proportion	.25	.24	.19	.31	100

In all six textbooks there appeared to be a large proportion of rules-of-thumb concerning sample size, either in distributional approximations or in inference-making. In fact, 41 percent of all rules-of-thumb observed made a direct reference to sample size. This relatively large number of rules-of-thumb dedicated to sample size concerns led to subsequent screening of two well-known sampling textbooks and noting the frequency of rules-of-thumb in these texts. Sampling textbooks by Cochran (1963) and Kish (1965) were selected for review, and the Cochran text contained 29 rules-of-thumb while the Kisk text listed 39. These frequencies are, on the average, larger than any of the introductory textbooks, however, these figures are not offered to compare the frequency of rules-of-thumb, but merely to show that sampling theory uses a relatively large number of them, possibly accounting for the high frequency of rules-of-thumb concerning samples in introductory textbooks.

Some Commonly Reported Rules-of-Thumb

What this nonrandom sample of textbooks allows me to say about the prevalence of rules-of-thumb is certainly open to conjecture, but it is apparent that there are sufficient numbers of different types and frequencies of each to warrant further consideration of some of the most commonly used. All the textbooks reviewed, and I would suspect almost all such textbooks in general, provide some rule-of-thumb for deciding on the number of intervals to be used in constructing frequency distributions. Not all texts used the same rule, but the majority of them used more written space for this rule than any other. Why a relatively minor issue is so dominant in introductory textbooks is beyond my comprehension. Perhaps interval construction is an activity engaged in by a lot of people even though it is generally regarded as being a highly arbitrary process.

A rule-of-thumb concerning sample size and the binomial approximation to the normal that was common to all six texts was the "$nP > 5$" or related rule. One text did, however, put a new wrinkle on the rule and used $nP \geq 10$, and another text required both $nP > 5$ and $nQ > 5$ to justify the normal approximation.

Attempts to inform the reader when a sample was large enough to use a z-test instead of a t-test resulted in rules-of-thumb ranging from $n \geq 10$ to $n \geq 120$. Ironically, the authors of such rules-of-thumb are going to the trouble of making up an arbitrary rule when no such rule is necessary. This form of rule invocation works as a detriment to the understanding and appropriate application of these two tests. Much space in all six textbooks was devoted to some rules-of-thumb designed to distinguish between "small" and "large" sample sizes in inference-making, even though sample size is not an assumption for any statistical inference technique. More will be said on this later.

The chi-square family of tests provided a fertile field for invoking rules-of-thumb. All six textbooks provide some form of rule for the minimum size of cell frequencies in a chi-square table. The rules consist primarily of variations on the theme: "No more than 20 percent of the cells should have expected frequencies of 5 or less, and no expected frequencies should be less than one." The chi-square approximation to a normal accounts for several more rules along with suggestions for using Yate's correction.

The rule-of-thumb that the alpha level should be .05 or .01 is almost universal and appears to have taken on the appearance of a requirement in all six textbooks. The impact and ultimate invocation of this rule-of-thumb is clearly reflected in the response from an editor of a leading research journal in the behavioral sciences. In answering my inquiry, the editor said, "This journal uses only alpha equal .05 for its articles." (See Franks and Huck, 1986, and Brewer, 1987, for comments on similar lack of thought given to alpha levels).

Contrasts between parametric and nonparametric methods have a sizable portion of rules-of-thumb ranging from, "If the scale is ordinal . . ." to a broad spectrum of statements such as, "If $n < 6$ (8, 10, 15, 25, 35, etc.), then use a Wilcoxon test." Many of the nonparametric rules-of-thumb relate to normal approximations of test statistics, the most common type being, "If $n \geq 25$ or 30, then use the normal approximation."

Some of the more exotic rules-of-thumb concern for example, fractional rounding of the degrees-of-freedom (found in one of the introductory texts) and a "crude rule" found in one of the sampling theory texts. The latter proposes for normal approximations in confidence intervals that sample size be $25G_1^2$ or greater, where G_1 is Fisher's measure of skewness. Examples of other rules-of-thumb found in the six introductory texts are given in Table 2; note that some of them are contradictory.

Rules-of-Thumb and the Teaching of Statistics

Every statistics instructor, at one time or another, particularly in the more elementary courses, finds it necessary to resort to a rule-of-thumb in answering an inquiry from a student or consultee. For example, there is almost no way around the ".05" or ".01" levels for alpha which are rules that traditionally have been carved into the literature as though they are part of the ten commandments of statistics. The question is not whether instructors should use rules-of-thumb, but rather how they should use them effectively in teaching, given that rules are facts of

Table 2 Some Example Rules-of-Thumb Found in Six Introductory Behavioral Statistics Texts

1. The number of class intervals should be 10-20.

2. When $n \geq 25$ or 30, the t distribution may be approximated with a z distribution.

3. The alpha level of .05 or .01 was selected.
4. When $nP > 5$ or $nQ > 5$, the normal may be used to approximate the binomial.

5. For df ≥ 30, the chi-square closely approximates the normal.

6. No more than 20% of the cells should have expected values smaller than 5, and no expected value should be less than 1.

7. If n is small ($n < 8$), use a nonparametric test.

8. Where the expected value in any cell is less than 10, the Yate's correction should be used.

9. When samples are equal to or greater than 30, bias is eliminated.

10. If sample size is 30 or less, use $n - 1$ for the divisor in calculating S^2.

11. In using this formula, a sample must be substantially larger than 30.

12. One-tailed tests should not be used in place of two-tailed tests.

13. Sample sizes are fairly unequal when the larger is more than 1.5 times greater than the smaller.

14. For failure of the bivariate normal assumption, if degrees of freedom are greater than 25 or 30, the assumption is of little consequence.

15. For testing a Spearman's Rho, when $n \geq 10$, procedures for testing a Pearson will give very good approximations.

16. We suggest a value of .80 for the power of a test.

17. A Pearson correlation of .50 is large.

18. On least significant difference usage, when $k \geq 6$, use some other procedure.

19. On chi square tests, for 1 df, all expected frequencies should be at least 5. For 2 df, expected frequencies should exceed 2, etc.

20. The Wilcoxon test should not be used if the sample size is smaller than about 8.

21. On standard error of estimate, when $n < 50$, use a correction factor.

22. Spearman's Rho is particularly well suited to situations where $n \geq 25$ or 30.

23. When n is large, Spearman's Rho is almost useless.

24. If $n \geq 25$ the t test is relatively unaffected by rather severe violations of assumptions.

25. The standard error of proportion formula is not recommended if $n P$ or $n Q \geq 10$.

26. The sampling distribution of a Pearson r is normal when the parameter equals zero.

27. When $n > 25$ the sum of ranks may be taken as normally distributed.

28. For calculating confidence intervals on means, an assumption is $n \geq 30$.

29. If you have reason to believe your underlying population is symmetric, but cannot be sure it is normal, use the Wilcoxon test instead of the t-test.

30. Refrain from using ANOVA if the scale is not interval.

statistical life resulting from the many subjective judgments required of the user. It is, therefore, inevitable that rules-of-thumb will be adopted sometime by someone to ease the pain associated with all these judgments. But like any other pain remedy, these rules can be dangerous if taken as an overdose; the danger is in allowing the rules to take on a life of their own as perhaps α = .05 or α = .01 have done. Instructors must discriminate clearly between being helpful with judgment-making and fabricating a rule-of-thumb to **replace** judgment.

Teaching Around Rules-of-Thumb

The following example rules-of-thumb and attendant suggestions for teaching the related concepts should provide some indication of how an instructor could "teach around the rules-of-thumb."

1. In demonstrating the normal approximation to the binomial, the usual rule-of-thumb is, "If nP or nQ > 5 then . . .". The instructor could, instead of invoking this rule, initially concentrate on the symmetric nature of the binomial when P = Q = 1/2. Then through homework assignments, let students observe what happens as n gets larger, regardless of the P value. Software packages for minicomputers are available to aid such an assignment, but even a few carefully hand-prepared frequency distributions would illustrate what happens asymptotically.

2. Instruction in the Central Limit Theorem and its impact on statistical inference should not be arbitrarily and unnecessarily encumbered with a rule-of-thumb concerning the size of the sample. Beginning students easily can take random samples of numbers such as those for telephones or license plates, and plot frequency distributions of these numbers and their averages. They can then increase the sample size and repeat the frequency plot of averages. The resulting dramatic shape changes are more than sufficient to illustrate the power of the Central Limit Theorem.

3. The tradition-ensconced Type I probability levels of .05 and .01 provide the statistics instructor with a convenient springboard for a discussion of the relative costs associated with the error types and their subsequent impact on sample size. Very clear points can be made with students, particularly with samples of convenience, concerning the folly of a blind adherence to the rule of .05 or .01. A particularly good example is a dying cancer patient's view of error when the null hypothesis is that a drug is worthless versus the alternate that the drug is effective. Clearly this patient would have no real concern for a small alpha level and would allow it to be quite large in order to avoid making

a Type II error.

4. The general invocation of rules to answer the question, "How much data do I need?", can be replaced appropriately with discussions of the ingredients for determining sample size for inference-making. A comparison of the sample size requirements for hypothesis testing and confidence intervals is a natural adjunct to these discussions. If we can convince students that they must answer several questions concerning the what and why of their inferences before they can answer the question of adequate sample size, it would be a vast improvement over invoking a rule-of-thumb.

We should consider the large number of subjective judgments required of students as an opportunity to teach the understanding of concepts rather than as another opportunity to invent or invoke rules-of-thumb. It is highly probable that our efforts will not result in a diminution in the number of present rules, but at least we might thwart the generation of new ones. The computer has been helpful in minimizing the need for some rules-of-thumb through vast Monte Carlo-like studies which have clarified some robustness and model fit concerns of statisticians and users. The extensive work of Blair and Higgins (1980) and others, has eliminated the need for rules-of-thumb such as, "If the sample size is small use a nonparametric method." The computer, however, is not a replacement for all rules-of-thumb, and the prevalence and use of the latter is mainly the responsibility of the instructor and the textbook authors.

Further, we should realize that students are quite capable of making subjective or experience-based decisions without the crutch of a rule-of-thumb, and that teaching students to make such decisions is both a necessary and colorful part of statistical instruction. The easy way to treat decision making during instruction is to flip out a rule-of-thumb rather than to guide students to defensible, subjective judgments. To do the former is to cop-out, and to do the latter is to teach.

Rules-of-Thumb for Rules-of-Thumb

Since no one knows when and how rules-of-thumb should be invoked or invented, I will succumb to the tradition-bound temptation of providing some guides or suggestions for using rules-of-thumb. These guides, like all rules-of-thumb, are arbitrary and open to debate, however well-intended.

1. **Rules-of-thumb should be invoked only if absolutely necessary.** This admonition is particularly good instructional advice since unnecessarily invoking rules-of-thumb implies that the learner is

exempted from the responsibility for making subjective decisions concerning assumptions, adequate sample size, or choice of tests. The student should be taught the alternative implications that result from subjective judgments rather than be given a "way out" via rules-of-thumb.

2. **Avoid invoking a rule-of-thumb as though it were a condition or assumption.** Statements such as, "When $n < 10$, use a t-test" and, "If $n \geq 30$, the Central Limit Theorem says . . .", are examples of unnecessary rules-of-thumb invoked in a manner implying that they are conditions rather than someone's suggestions. The danger in using this type of rule-of-thumb, particularly in sample size determination, is that the reader will opt for the "quick and dirty" rule for determining n rather than engage in an intelligent consideration of the statistical and practical determiners of n for proper inference-making. I believe that a large portion of the myths and misconceptions in statistics (see Brewer, 1985) are rooted in the blind invocation of rules-of-thumb and in the perception that such rules are conditions rather than opinions.

3. **Avoid inventing new rules-of-thumb.** The textbook reviews previously described make it fairly obvious that we have a sufficient supply of rules-of-thumb to provide more than ample guidance to the users of statistics. Additional rules-of-thumb, like default values in computer programs, decrease the proper consideration of the user's situation and conditions. To expect rules-of-thumb to tell us how and when to apply a statistical tool is akin to expecting a wrench to tell us which way it should be turned. Flooding an already saturated area with more rules-of-thumb hardly is conducive to thoughtful, informed judgments in the application of statistical methods and may instead lead to increased carelessness on the part of the user. (The reader will note that this rule along with the two previous ones are themselves violations of this rule.)

References

Blair, C., & Higgins, J. (1980). A comparison of the power of Wilcoxon's Rank-Sum statistic to that of the Student's t-statistic. Journal of Educational Statistics5(4), 309-335.

Brewer, J. K. (1985). Behavioral statistics textbooks: Source of myths and misconceptions? Journal of Educational Statistics, 10(3), 252-268.

Brewer, J. K. (1987). A comment on the .05 level of significance: Franks and Huck (1986). Research Quarterly for Exercise and Sport, 58(1), 83-84.

Cochran, W. G. (1963). Sampling techniques, (2nd ed.) New York: Wiley.

Franks, B. D., & Huck, S. W. (1986). Why does everyone use the .05 significance level? Research Quarterly for Exercise and Sport, 57, 245-249.

Kish, L. (1956). Survey sampling. New York: Wiley.

AUTHOR

JAMES K. BREWER, Professor, Department of Educational Research 307 Stone Bldg., Florida State University, Tallahassee, Florida 32306