

Degrees of Freedom at the Start of the Second 100 Years: A Pedagogical Treatise

Advances in Methods and
Practices in Psychological Science
1–10

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/2515245919882050

www.psychologicalscience.org/AMPPS



Joseph Lee Rodgers 

Department of Psychology and Human Development, Vanderbilt University

Abstract

Degrees of freedom is a critical core concept within the field of statistics. Virtually every introductory statistics class treats the topic, though textbooks and the statistical literature show mostly superficial treatment, weak pedagogy, and substantial confusion. Fisher first defined *degrees of freedom* in 1915, and Walker provided technical treatment of the concept in 1940. In this article, the history of degrees of freedom is reviewed, and the pedagogical challenges are discussed. The core of the article is a simple reconceptualization of the degrees-of-freedom concept that is easier to teach and to learn than the traditional treatment. This reconceptualization defines a statistical bank, into which are deposited data points. These data points are used to estimate statistical models; some data are used up in estimating a model, and some data remain in the bank. The several types of degrees of freedom define an accounting process that simply counts the flow of data from the statistical bank into the model. The overall reconceptualization is based on basic economic principles, including treating data as statistical capital and data exchangeability (fungibility). The goal is to stimulate discussion of degrees of freedom that will improve its use and understanding in pedagogical and applied settings.

Keywords

degrees of freedom, data, statistical history, teaching statistics

Received 7/14/19; Revision accepted 9/5/19

“Degrees of Freedom”
counts the different types of
Statistical Cash

Suppose that a core statistical concept was poorly taught. Suppose that such a concept—one as basic as variance, sampling distributions, or degrees of freedom—was treated inconsistently or haphazardly, or even ignored, in some important statistical settings. Suppose that many years after its invention, a firm foundation for teaching this core statistical concept to introductory-statistics students was still lacking. That situation would appear to deserve immediate and widespread treatment by the community concerned with the teaching of statistics.

This article is about introductory pedagogy related to degrees of freedom. Eighty years ago, Walker (1940) published “Degrees of Freedom,” an article that documented an important anniversary: The degrees-of-freedom

concept “was first made explicit by the writings of R. A. Fisher, beginning with his [*Biometrika*] paper of 1915 on the distribution of the correlation coefficient” (p. 253). Degrees of freedom passed its 100th birthday in 2015.¹ Is it time to review, evaluate, and revise our basic approach to teaching this important statistical concept?

A researcher working with quantitative methods in the behavioral sciences requires two things: (a) a mathematical or statistical model and (b) data. Data are of value because they indicate the validity of the model. A model is valid only insofar as its outputs approximately match empirical data. The degrees-of-freedom concept exists at the interplay between quantitative

Corresponding Author:

Joseph Lee Rodgers, Lois Autrey Betts Chair of Education and Human Development in Peabody College, 230 Appleton Place #552, Hobbs 202C, Vanderbilt University, Nashville, TN 37203
E-mail: joseph.l.rodgers@vanderbilt.edu

Box 1. Apologies Through the Years

Is the degrees-of-freedom concept actually difficult to understand, and therefore to teach? Past treatment supports that view:

- Tippett (1931, p. 64): “This conception of *degrees of freedom* is not altogether easy to attain, and we cannot attempt a full justification of it here; but we shall show its reasonableness and shall illustrate it, hoping that as a result of familiarity with its use the reader will appreciate it.”
- Walker (1940, p. 253): “For the person who is unfamiliar with N -dimensional geometry or who knows the contributions to modern sampling theory only from secondhand sources such as textbooks, this concept [degrees of freedom] often seems almost mystical, with no practical meaning.”
- Good (1973, p. 227): “The concept [degrees of freedom] is a difficult one for the student and seems to be difficult even for the expert, judging by the careless way in which it is defined in almost all books on statistics.”
- Saville and Wood (1991, p. 7): “An able pure mathematician, after two years spent teaching elementary statistics cookbook style, approached the authors and with a note of despair in his voice asked ‘What on earth are degrees of freedom?’”
- Black (1994, p. 306): “The concept of degrees of freedom is difficult and beyond the scope of this text.”
- Everitt (2002, p. 111): “Degrees of freedom: An elusive concept that occurs throughout statistics.”
- Dallal (2003, para. 1): “One of the questions an [instructor] dreads most from a mathematically unsophisticated audience is, ‘What exactly is degrees of freedom?’”
- Eisenhauer (2008, p. 75): “Many introductory textbooks present degrees of freedom in a strictly formulaic manner, often without a useful definition or insightful explanation.”

models and empirical data. Degrees of freedom provides an accounting method to keep track of the flow of information from data into a model.

In this article, I propose an approach for teaching, and conceptualizing, degrees of freedom that is different from the usual ones. According to this approach, data are viewed as capital—statistical money—deposited into a statistical bank. A researcher who runs a statistical analysis involving parameters (e.g., a t test, regression, or structural equation model) withdraws money from the bank to pay for estimating the parameters in the model. Degrees of freedom is a count of the statistical money withdrawn from the bank, and also of the statistical money still left in the bank. Often, degrees of freedom is given limited attention in teaching, and in thinking about, statistical analysis. But even if the various types of degrees of freedom are virtually ignored, those varieties of degrees of freedom are still there, behind the scenes. Some people who deposit actual money in a bank account pay careful attention to the money withdrawn and the balance remaining in the account, whereas others are only vaguely aware. Similarly, some researchers carefully account for the data used to purchase parameters and the data remaining in the bank, whereas others are only informally aware of this statistical transaction. But the degrees-of-freedom concept as an accounting mechanism is always there, representing the flow of information out of the data and into the model.

Degrees of freedom has always been a difficult concept to discuss and teach. Textbook (and other) authors have struggled with defining the term (see Box 1). The pedagogical goal of this article is to stimulate discussion

among statisticians and researchers across disciplines to move statistical pedagogy regarding this core concept forward. The broad substantive goal is to expand and reconceptualize degrees of freedom. The article’s orientation is strongly didactic, focusing on teaching the degrees-of-freedom concept to introductory students in undergraduate and graduate statistics courses.

Definitions and History

Part of the challenge of defining *degrees of freedom* is that the concept is many different things. To a methodologist, the concept has several mathematical interpretations. To a researcher, degrees of freedom helps with the statistical analysis. Further, there are a number of subdefinitions within each category, as explained in the next section. Some of the challenge associated with defining the term—and the flavor of its different meanings—is demonstrated in the diverse definitions presented in Box 2.

Walker (1940) succinctly stated the challenge of defining *degrees of freedom* as “bridgling] the gap between mathematical theory and common practice” (p. 253). But she also noted, ironically, that “the mathematical notion is simpler than any non-mathematical interpretation of it” (p. 253). Is the concept too complex mathematically to deliver to a statistical audience unfamiliar with matrix algebra? If the concept is treated only mathematically, the likely answer is “yes.” But there also exist conceptual treatments, such as the one referenced in Dallal’s (2003) definition and in my epigram at the beginning of this article, that are true to the original

Box 2. Definitions Through the Years

Textbook authors, and authors of statistical articles, often do not even attempt a definition of *degrees of freedom*. For example, *Statistics* (Hays, 1988) is a highly authoritative source for statistical pedagogy that does not ever define the term. But many authors, typically technical experts, have defined the concept, in a number of different ways:

- Cramér (1946, p. 103): Degrees of freedom is “the rank of a quadratic form.”
- Good (1967): Degrees of freedom is a concept “related to the Neyman-Pearson-Wilks likelihood ratio; that is, to the ratio of maximum likelihoods” (cited by Good, 1973, p. 228).
- Good (1973, p. 227): “The number of degrees of freedom in testing H within K is $d(K) - d(H)$, the difference of the dimensionalities of the parameter spaces.”
- Dallal (2003, para.3): “I’m inclined to define **degrees of freedom as a way of keeping score**. A data set contains a number of observations, say, n . They constitute n individual pieces of information. These pieces of information can be used to estimate either parameters or variability. In general, each item being estimated costs one degree of freedom. The remaining degrees of freedom are used to estimate variability. All we have to do is count properly.” (boldface in original)
- Wikipedia (2015, para. 1): Degrees of freedom is “the number of independent pieces of information available to estimate another piece of information. More concretely, the number of degrees of freedom is the number of independent observations in a sample of data that are available to estimate a parameter of the population from which that sample is drawn.”

concept and that introductory students and practicing researchers will find manageable.

Walker (1940) presented perhaps the broadest set of definitions. She introduced her article (p. 254) by presenting three ways to conceptualize degrees of freedom: first, as “freedom of movement of a point in a space when subject to certain limiting conditions”; second, as the “representation of a statistical sample by a single point in N -dimensional space”; and third, as formulas or procedures to “determine the number of degrees of freedom appropriate for use in certain common situations.” Later in her article (p. 260), Walker developed an additional fourth definition: Degrees of freedom is a parameter within a mathematical formula, used to distinguish different members of a distributional family. For example, for the t and chi-square distributions, degrees of freedom is the characterizing parameter, specifying a particular member of the family of t or chi-square distributions.

To summarize, there are several different (overlapping) ways to define, and to think about, degrees of freedom. Technical experts—theoretical and applied statisticians—approach the concept as a “quadratic form” and as “constraints in N -dimensional space.” These concepts are beyond the mathematical sophistication of typical introductory-statistics students. In the past, introductory-statistics teachers and researchers have used the term *degrees of freedom* in two apparently separate ways. First, formulas used to compute degrees of freedom from the sample size and the research design allow researchers to match the statistical procedure they use to the appropriate theoretical sampling distribution, so that they may look up critical values in statistical tables, such as t , F , and chi-square tables. This

conceptualization combines Walker’s third and fourth definitions. Second, statistics teachers often discuss how degrees of freedom are “used up” or “lost” in a statistical analysis. This conceptualization is related to Walker’s first definition and is also the starting point for the pedagogical reconceptualization developed here.²

It is easy to understand why textbooks avoid or at least minimize treatment of degrees of freedom. Which one (or more) of several definitions should be used? Should only the formulas be presented, and the concepts avoided? Even Fisher himself was uncertain of the appropriate treatment. In the first analysis of variance (ANOVA) table (Fisher, 1922), the several kinds of degrees of freedom were omitted (correct degrees of freedom were inserted by Box, 1978, p. 104). By the next year, the ANOVA table had achieved its current form, including a column for degrees of freedom (Fisher & Mackenzie, 1923). In typical modern treatment, subscripts are used to distinguish types of degrees of freedom: df_{total} refers to total sample size (or df_{total} may be reduced by 1 to account for estimating the intercept, or population mean, thereby becoming $df_{\text{total adjusted}}$); df_{model} , $df_{\text{interaction}}$, df_{between} , and df_{group} refer to the degrees of freedom associated with estimating parameters within the model; and df_{error} , df_{within} , and df_{residual} refer to the remaining degrees of freedom that are still unused, or saved.

Degrees of Freedom Reconstructed for Students and Researchers

The reconstruction

In this section, I conceptualize a new way to teach (and to think about) degrees of freedom.³ The reconstruction

uses an economic interpretation of data. This perspective is not quite new (e.g., see Dallal, 2003, para. 3, who defined degrees of freedom as “a way of keeping score”). However, it apparently is not used in any standard introductory textbook, nor can it be found in Web searches (except for the Dallal reference). I have presented the topic as developed in this article to students in dozens of introductory undergraduate and graduate statistics courses for many years.

Begin by defining the existence of a statistical bank, into which is deposited all of the scores from the dependent variable. Scores are defined as collected data values (or often, just data or data points), the result of measuring a variable (a property of a human subject or other object of study; note that degrees of freedom are defined in relation to scores on the dependent variable). Once deposited, the data can be viewed as statistical money to be drawn out of our statistical bank. With that money, we pay for a commodity of value to a researcher, the estimated values of the parameters, which themselves define a statistical model. Degrees of freedom is a count of the flow of data into and out of the statistical bank. Suppose we measure the height of 10 subjects. The statistical bank then starts with 10 pieces of statistical money in the bank. Thus, we begin with $df_{\text{total}} = 10$. Then scores are withdrawn from the bank (conceptually, not literally) for the purpose of estimating a statistical model. If we wish to estimate the mean height of the population using the 10 sample height scores in the bank, we are estimating one parameter, the population mean. The cost of one parameter is paid for with statistical money, and costs 1 unit of statistical money. When we pay with a single score, we in a sense transfer a score’s information into the model, and we have separated the original $df_{\text{total}} = 10$ into $df_{\text{model}} = 1$ and $df_{\text{residual}} = 9$. By analogy, it is as though we have deposited \$10 into a real bank account and then purchased an item for \$1; we obviously have \$9 remaining in the bank, and the \$1 has been (conceptually) transferred into the value of the item itself. Just as we keep track of the money in a bank account, we can keep track of the data values in a statistical bank, and this is a powerful way to view the flow of information from the data into the model. We use degrees of freedom to count that flow.

Estimating parameters—using the sample mean to estimate the population mean, or computing slopes to estimate regression parameters—places constraints on the scores. Some of the original data information is transferred out of the statistical bank and into the model, counted by df_{model} . Some of the data information usually remains in the bank (assuming a sample large enough to overidentify the statistical model), counted by df_{residual} . Within this economic system, the cost of estimating most parameters is a single data point. In

other words, data are used up one by one, to pay for the estimate of each independent parameter within the model. To start simply, let us define a mean model, or, in the regression framework, the intercept-only model: $y = \mu + e$, or $y = b_0 + e$ (where y is the dependent variable, $\mu = b_0$ is the population mean of the dependent variable, and e is the error, or residual). Computing the sample mean (the least squares estimate of the population mean, μ) requires the N sample observations from the dependent variable. Once μ is estimated with \bar{y} , a single observation has been “used up”; one piece of statistical money has been spent out of the statistical bank and conceptually passed into the model. In a more complex model with four independent parameters to estimate (e.g., a multiple regression model, $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + e$), four pieces of statistical money—four data points—are spent out of the statistical bank and conceptually passed into the model as we estimate $b_0, b_1, b_2,$ and b_3 .

As the researcher purchases estimated values of parameters in the model, there are actually other returns as well, including, for example, the value of the residual for each observation. But that information is mathematically derivable from the estimated parameters, because the estimated regression equation can be used to compute the exact value of the i th subject’s residual, e_i ; the researcher is not “double charged” for such redundant information. Those residuals can be used to estimate the residual variance, and so the researcher in a sense gets a lot for her or his statistical money—a model, residual values, an estimate of residual variance, and measures of fit, such as R^2 . Understanding degrees of freedom—and the explanation for the nomenclature *degrees of freedom*—involves examination of what information that emerges from the estimation of a statistical model is still unique (free to vary) and what information is redundant (and therefore should not be counted as “something new”).

There is an additional link on which this economic interpretation sheds light. Once a statistical model is estimated, and assuming the analysis assumptions are met, it is often the case that a test statistic must be compared with a theoretical sampling distribution. For example, given a two-independent-groups design with 10 subjects in each group, $df_{\text{residual}} = 18$ (i.e., $10 + 10 - 2$), so the observed t value is compared with the critical value from the theoretical t distribution with $df_{\text{residual}} = 18$. The residual degrees of freedom “points to” the particular member of a family of sampling distributions (the particular theoretical t distribution with $df = 18$ in this example). Thus, Walker’s third and fourth definitions in the previous section are linked on the empirical side, when residual degrees of freedom are counted and used to specify the specific sampling distribution for testing the appropriate statistical hypothesis.

Exchangeability of data

When I teach this perspective, it is not uncommon for a student (or a practicing researcher) to ask, “Which one of the data points is lost when we estimate the population mean?” or “Which scores do we use to pay for the regression parameters?” These are good questions, the answer to which requires understanding an economic concept called *fungibility*, or sometimes *exchangeability*. As the latter term is descriptive of the concept, I use it here.

In economic theory, units of currency are exchangeable (fungible). Any \$1 bill is as good as any other \$1 bill for purchasing something valued at \$1. If U.S. currency were not exchangeable, and certain \$1 bills were more valuable than others, defining the actual value of a purchase would include an extra layer of negotiation in what is, otherwise, an objective process involving trading something worth \$1 for a note on which is printed “\$1.”⁴

Similarly, data points are exchangeable in statistical analysis. Typically, every data point is considered to have the same statistical value as every other data point. No specific data point is used to pay for computing the sample mean as an estimate of the population mean; nor do a certain four data points estimate the regression parameters in the example presented earlier. In teaching degrees of freedom, instructors refer, for example, to an unspecified $n - 1$ or $n - 4$ of the remaining data points as being free to vary within the residual variance; it does not matter which data points are chosen not to vary. (Note that exchangeability transmits mathematically to the idea that degrees of freedom is a count of the dimensionality of the constrained data space. No dimension is “lopped off”; rather, the observations that originally existed in an N -dimensional space become residuals from the model in a different and reduced $N - k$ dimensional space. See Box 3 for a brief geometric treatment of degrees of freedom.) For example, consider again the 10 height scores that we deposited into the statistical bank. If we are interested in estimating the mean height of the population and we compute a sample mean of 67 inches, we have constrained one score to no longer be free to vary. That is, if we know any 9 scores and the mean, the 10th score is constrained to be the score that, if included in the calculation, would result in the estimated mean. No specific score is used to pay for the mean, because the scores are exchangeable.

Illustration

The concept of spending statistical money (the flow of degrees of freedom out of the statistical bank into an estimated model), and also the concept of exchangeability, can be illustrated with the following regression

problem. Suppose a researcher is interested in the relationship between adolescents’ drug use (the dependent variable) and their parental supervision (the independent variable) and collects data ($N = 50$) on these two variables at a U.S. public high school. (Note that the flow of money is counted in relation to the number of observations, or equivalently, the number of dependent-variable values, and thus the researcher deposits 50 data points into the statistical bank.) Drug use is measured as the self-reported number of days on which the student used drugs in the previous month, and the possible scores range from 0 to 30. Parental supervision is measured on a self-report scale from 0, indicating no parental attention to supervising the adolescent, to 10, indicating constant supervisory attention.

Suppose the researcher collects the first data point (drug use = 11, parental supervision = 6) and plots it in a scatterplot. The result is the plot in Figure 1a. By collecting this data point, the researcher has deposited one piece of statistical money in the statistical bank, $df_{\text{total}} = 1$. It can easily be seen that the researcher cannot fit a line to this data point, because there is not enough statistical money to purchase the two parameters, the slope and the intercept, that define a line; an infinity of lines pass through the single point. In technical language, the model is *underidentified*. Next, suppose the researcher collects a second data point, as illustrated in Figure 1b. There are now two pieces of statistical money in the bank, exactly enough to estimate the slope and intercept of a line, which are uniquely defined by two data points; the model is *just identified*. However, all of the available statistical money is withdrawn from the bank to estimate the slope and intercept associated with these two points (i.e., $df_{\text{total}} = df_{\text{model}} = 2$, $df_{\text{residual}} = 0$). The researcher has accomplished one of the goals of estimating the model, finding the values of the slope and intercept. But no money is left in the bank, and in fact, it is important for money to remain in the statistical bank, for two purposes.

The first use of money remaining in the bank is to evaluate the fit of the model, that is, how close the line comes to the data points (along the direction of the dependent variable, i.e., the y -axis). If there are residuals that are free to vary (and there are not if the model fits perfectly), the residual variance can be computed as an indicator of how effective the model’s prediction is. But there is no meaningful residual variance that can be computed (because residual variance is constrained to be zero) in this example ($df_{\text{residual}} = 0$). Can the researcher be confident that this is an effective model of the relationship between parental supervision and adolescents’ drug use? The answer is that there is no way to know, because there is no statistical money left

Box 3. Degrees of Freedom: Geometric Treatment

Degrees of freedom is an often confusing, and highly technical, concept. This article is a simplification, to support introductory pedagogy. But the risk is oversimplification; students may believe that viewing degrees of freedom as deposits into and withdrawals from a statistical bank is the correct or only approach. In this box, the technical origin of the concept is briefly described. The cited references can be consulted for elaboration.

Fisher (1915) conceptualized a unique approach to the geometry of the statistical data space, including degrees of freedom. The importance of this innovation emerges most strongly from the biography of Fisher written by his daughter, Joan Fisher Box (1978). Box devoted several pages to Fisher's use of the geometry of N -dimensional (subject) space. She suggested that Fisher's geometric fluency within this space helps answer the question, How did Fisher have the insights that he did when other talented statisticians had not yet recognized those insights, and in many cases did not realize the scope of his contributions until much later? Within this space, degrees of freedom is a positive integer, the dimensionality of the constrained data space (e.g., Box, p. 125). Modern scholars use this approach to define statistical concepts through a geometric (as well as algebraic) perspective (e.g., Draper & Smith, 1998; Mulaik, 2010; Rodgers & Nicewander, 1988; Saville & Wood, 1991; Wickens, 1995).

Matrix algebra is a mathematical approach that treats spaces, with points and vectors lying in those spaces. Fisher defined a data space in which each variable is a vector, or directed line segment, from the origin of the space to a point defined by the set of that variable's values across subjects. That is, the endpoint of each vector represents all the data of a single variable. Variable vectors potentially lie in an N -dimensional data space (N = number of subjects), with one dimension for each subject. Within this space—called *subject space* because there is one axis per subject—lie all of the variables in a statistical problem. The original number of degrees of freedom— df_{total} —is equal to the dimensionality of this subject space (usually it is N -dimensional). But when we conduct a statistical analysis, we drop a dimension for each parameter that we estimate. The reason for this involves the mathematics of matrix algebra, but the basic idea is that estimating parameters “uses up” dimensions.

To a mathematician—to Fisher, for example—the number of degrees of freedom is exactly the dimensionality of the constrained subject space. How are the constraints imposed? They are imposed by any statistical model in which parameters are estimated from the data. In the unconstrained data, there are N degrees of freedom. Once a model is fit, there remain $N - k$ degrees of freedom, where k is the number of independent parameters estimated within the model. It should be noted that the usual textbook procedure to compute degrees of freedom is embedded within this conceptualization.

Methodologists will recognize this treatment as the geometric interpretation of basic matrix-algebra operations. The variance that remains after fitting a model is often called residual variance, and the residuals are the data values after adjusting for the model. Those residuals can themselves be plotted in a subject space. For each independent parameter estimated, the residuals exist in a space that has lost a dimension compared with the original data space. This is the sense in which Fisher viewed estimating a model as using up degrees of freedom. Residuals from the original data after fitting the model are “free to vary” among $N - k$ dimensions, but no longer cross into the higher dimensionality of the original data space. Introductory students and practicing researchers may not fully appreciate the concepts in the preceding paragraphs, nor do they need to. Consideration of the dimensionality of constrained data is not required for excellent statistical learning and practice.

to address the question of model fit. But suppose the researcher continues collecting data, until there is an N of 50, as illustrated in Figure 1c. There are now 50 pieces of statistical money, $df_{\text{total}} = 50$. Two units of statistical money can be withdrawn from the bank to fit a line to these data, as in Figure 1d; df_{total} is partitioned into $df_{\text{model}} = 2$ and $df_{\text{residual}} = 48$, and the fact that there are 48 residual degrees of freedom shows that the model is *overidentified*.

Does the line (the model) fit the data effectively? At an algebraic level, an effective model produces predicted y values that are “close” to the actual observed y values (according to some objective criterion, such as least squares). At a geometric level, a regression line

that is an effective model passes “close” to (i.e., has a small distance from) many, even most, of the data points, in the direction of the y -axis (the dependent-variable axis).

Going beyond these algebraic and geometric interpretations, both a conceptual and a statistical interpretation of the model's goodness of fit in this example can be provided. The 48 pieces of statistical money remaining in the statistical bank, because they are free to vary, can each fall close to the regression line, or not. If they consistently fall close to the line, they inform the researcher that the model works well. If they consistently fall far from the line, they indicate the opposite. Because there are 48 residual degrees of freedom, the

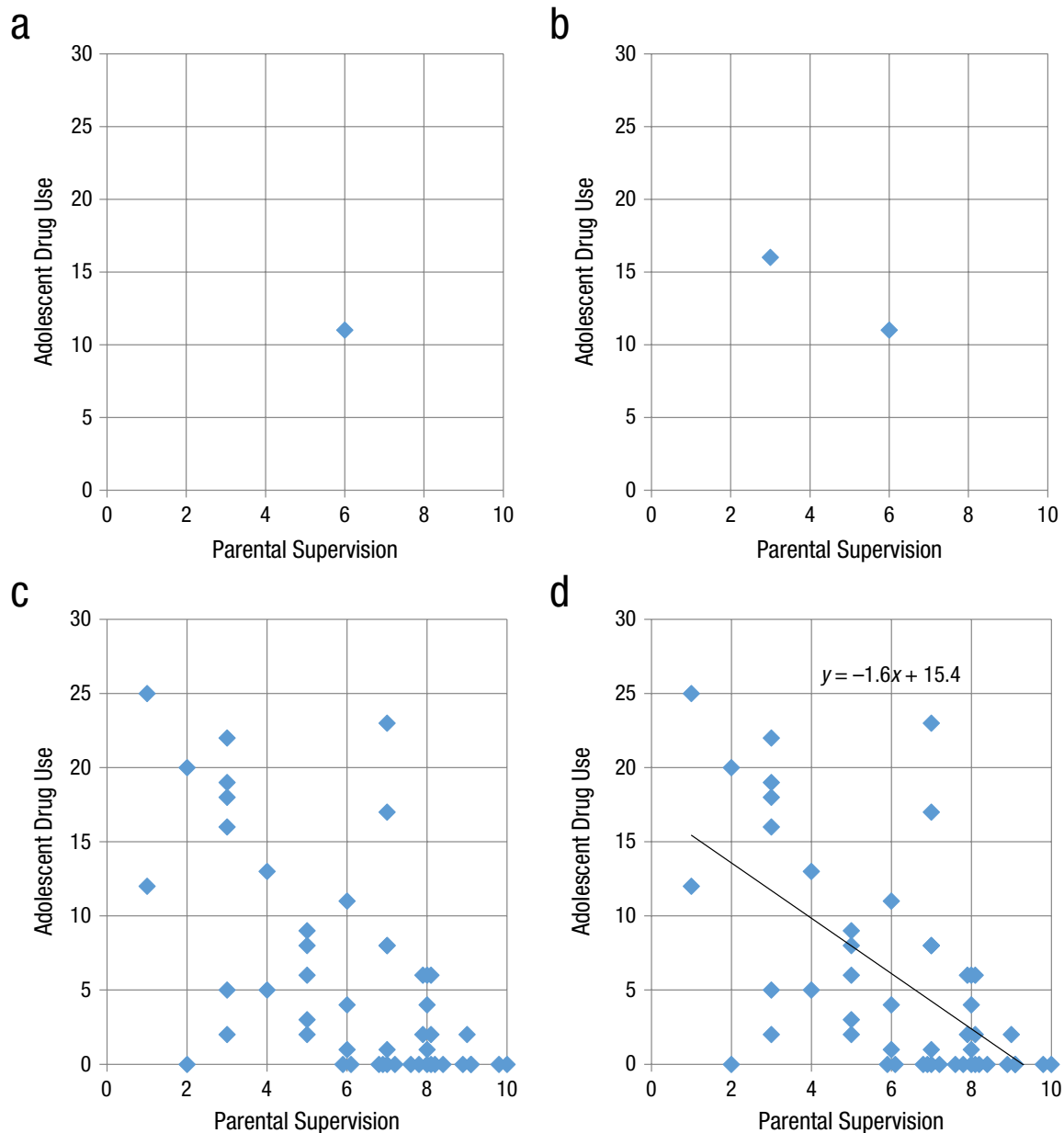


Fig. 1. Illustration of the concept of spending statistical money. A researcher who has collected a single data point (a) has just one piece of statistical money, which is not enough to identify a model of the data; an infinity of lines pass through a single data point. With a second data point (b), the researcher has two pieces of statistical money, enough to identify a linear model by estimating the slope and intercept. The next step is to evaluate the model's fit, which requires additional statistical money (i.e., residual degrees of freedom). If 50 data points are available (c), the researcher has statistical money for this second purpose. (Note that the data points are shown here with slight jittering because many of the points overlap and are not separately visible along the x -axis; e.g., there are six data points at (8,0)). If 2 degrees of freedom are used to identify the best-fitting least squares regression line (d), the remaining 48 degrees of freedom can be used to evaluate model fit and create new, more complex models.

researcher has quite a lot of information to help in assessing the goodness of fit of the model. Visually, the line seems to capture the basic underlying relationship between the two variables, because the general pattern of the data points tilts downward to reflect that as parental supervision increases, drug use decreases, and

vice versa—an inverse relationship. But some of the points are located notably away from the regression line (i.e., some points have relatively large residuals). The overall fit of the model can be measured statistically by using the residuals of the regression to compute R^2 , a statistical measure of how close the line

comes to the data points (along the direction of the y -axis). In this example, the simple correlation between parental supervision and adolescent drug use, r , is $-.61$, which results in an R^2 of $.37$. Other useful measures of model fit, such as the standard error of estimate, are functionally related to R^2 .

The second use of money remaining in the bank is to estimate a new and more complex model. If the researcher had also collected data on friends' drug use, this variable could be added as a second independent variable to assist parental supervision in predicting adolescents' drug use. There would be plenty of statistical money in the bank to estimate a regression model with an intercept and two slopes, and with $df_{\text{total}} = 50$ broken down into the money used to estimate the model ($df_{\text{model}} = 3$) and the money left in the bank ($df_{\text{residual}} = 47$), there would still be a lot of money left in the bank to assess residual variance and the fit of the expanded model. Further, there are more complex modeling methods that would allow the statistical comparison of the two models in a competition to explain meaningful variance within the dependent variable. The residual degrees of freedom in the two models assess the complexity of each model, and the difference in complexity as well (for elaboration, see Rodgers, 2010, p. 7).

Exchangeability is illustrated in Figure 1d because the best-fitting line is defined to pass through the original 50 data points, rather than a reduced number of them (further, it does not pass exactly through any of the individual points, though in theory it potentially can). Estimating the line uses up two unspecified pieces of statistical money, one for the slope and one for the intercept ($df_{\text{model}} = 2$). None of the specific data points disappear in this exchange of statistical money for a linear model, but basic df arithmetic shows that there are 48 pieces of statistical money left ($df_{\text{residual}} = 48$) to assess the fit of the model (and to spend on more complex models, or on a competition with more complex models).

It is worth emphasizing that one of the topics seldom addressed in introductory statistics teaching—in classroom treatment, in textbooks, and in applied research—is the utility of those remaining degrees of freedom. Whether the data counted by df_{residual} are used to estimate the fit of the model, to compute the residual variance (literally, the variance of the residuals in the model, defined by the e term), or to compute R^2 , the same conceptual process is involved. The important question is whether the weighted independent variables in the model are doing a good job of predicting the dependent variable. If there is little remaining residual variance, R^2 is high, and the model fits well; if there is a great deal of residual variance, R^2 is low, and the fit of the model is poor. If the amount of residual variance

remains the same as the original variance—that is, the model has not reduced the variance at all—then $R^2 = 0$, and the model has provided no predictive value. The more information we have available to assess the fit of the model—the larger df_{residual} —the more confidence we have in our assessment of that fit.

Debits and credits

A conceptual criticism of degrees of freedom is that the same name is used to refer to both sides of the transaction: Degrees of freedom are spent and also saved. But it is important to distinguish the two different types of statistical capital. A common name sharpens the sense that both sides of the transaction involve counting statistical money, but experience suggests that the common name may not work well for pedagogy. Statistics instructors regularly teach how to count the statistical money remaining in the statistical bank—the residual degrees of freedom—but often ignore the transaction process. For example, students and applied researchers master that a one-group t test “has $N - 1$ degrees of freedom” and that a one-way ANOVA “has $N - k$ denominator degrees of freedom” (where k is the number of independent parameters). In each case, the value in front of the minus sign indicates df_{total} as the starting point, the number after the minus sign indicates the df_{model} , or spent degrees of freedom, and the computational result measures df_{residual} , how many degrees of freedom remain in the bank. In accounting, debits and credits both are used to define transactions involving money and its directional flow, but the flows in the two opposite directions have different names. Students would more quickly grasp the degrees-of-freedom concept if the separate flows had different names in this case as well. In current statistical practice, the original capital input into the bank is df_{total} , flow out of the statistical bank is df_{model} , and the data points remaining in the bank are df_{residual} . Perhaps those terms are satisfactory, as long as the subscripts distinguish their separate status. More cumbersome but pedagogically useful names would be $df_{\text{money-initially-in-the-bank}}$, $df_{\text{money-spent-estimating-the-model}}$, and $df_{\text{money-remaining-in-the-statistical-bank}}$, or df_{total} , df_{spent} , and df_{saved} . Ultimately, it may be valuable to find new names to distinguish the separate flows of statistical capital, although 100 years of momentum work against revising this fundamental language.

New scholarly avenues and research questions

The idea of degrees of freedom as a count of the flow of statistical money opens up new avenues for pedagogy, and for research. The questions in this subsection

are informed by the economic perspective on degrees of freedom (though most do not have obvious answers).

Should each data point be worth the same amount? That is, is exchangeability a legitimate assumption? Suppose a researcher collected 99 data points easily, but spent extra time, energy, or dollars to collect the 100th data point. Should the last point be treated as exchangeable with the others? How can we measure or evaluate whether exchangeability is legitimate? What about sampling weights? Should an observation that represents 10,000 population elements (according to its sampling weight) be valued the same as an observation that represents 100 population elements? Or consider spending statistical capital. It is easy to estimate the population mean by computing the sample mean; even people with no statistics training can compute an average. But regression slopes are more difficult to estimate, because they involve more difficult formulas, software cost, and training to understand what they mean. Parameters in a structural equation model are even more complex and require more computing and more training. Should a researcher be charged the same statistical price for the intercept (the mean), for regression slopes, and for path coefficients in a structural equation model?

Recommendations and Conclusions

For just over 100 years, the concept of degrees of freedom has been used in statistical modeling (though see note 1 for potential historical revision). Statisticians, teachers, and textbook authors, the gatekeepers of statistical language and pedagogy, appear to accept how this concept is typically taught (or, often, virtually ignored), because few alternatives have been given attention during this period. The main goal of this article is to start a discussion about ways to better teach, and understand, degrees of freedom. The economic approach described can help improve teaching quality and the conceptual learning of introductory students. I have found that the following teaching practices have led to a better conceptual understanding of degrees of freedom. These should be viewed not as desiderata, but rather as discussion points from which conversation (even disagreement) may emerge, to sharpen our statistical pedagogy.

- Routinely refer to each data point as a unit of “statistical money” that, when collected, is deposited in a “statistical bank.” Degrees of freedom defines a statistical accounting process, as the various types of degrees of freedom indicate how many data points were originally available to estimate a model, how many were spent, and how many remain in the bank.

- Teach students explicitly about these different counts of statistical money. In each transaction (i.e., each statistical analysis), some of the statistical money is spent to estimate a statistical model. The remaining statistical money is left in the bank (as residuals from the estimated model). That money is used to assess the fit of the model (or, equivalently, to estimate residual variance or R^2). This remaining statistical money can also be used to estimate parameters within future, typically more complex, models. Degrees of freedom tracks each of those processes.
- Explicitly discuss that statistical money is exchangeable. Also discuss that exchangeability (fungibility) is an assumption, and may not necessarily reflect the researcher’s goals or the nature of the data.

Researchers are well trained, in statistics and methods courses, in how to obtain capital (collect data) and then how to use it to run statistical analyses to estimate models. But they are often poorly trained in statistical accounting principles, in using degrees of freedom to count the flow of statistical money within the transaction of purchasing the parameter estimates needed to define a statistical model. Modern authors have written about estimating models using data (e.g., Jaccard & Jacoby, 2009; Maxwell, Delaney, & Kelley, 2017; Rodgers, 2010), the most important part of the process. Good business practice requires that financial capital be obtained and spent appropriately, and also that the flow of that capital be carefully tracked. Similarly, good research practice requires tracking the flow of statistical capital. After an introductory statistics class, most students are not aware that counting degrees of freedom is, simply, statistical accounting. The goal of this article is to begin a process—not to finish it, but rather to stimulate development—of correcting this long-standing oversight. One hundred years of poor pedagogy regarding degrees of freedom is entirely too long.


Action Editor

Daniel J. Simons served as action editor for this article.

Author Contributions

J. L. Rodgers is the sole author of this article and is responsible for its content.

ORCID iD

Joseph Lee Rodgers  <https://orcid.org/0000-0002-8615-6713>

Acknowledgments

This material was first presented in September 2011 as an invited address at the annual conference of the Oklahoma Network for the Teaching of Psychology (ONTOP), in Stillwater,

Oklahoma. The author first became interested in the nuances involved in counting degrees of freedom in an introductory graduate statistics course at the University of Oklahoma in 1975, in which the instructor, Larry Toothaker, repeatedly stated, “Degrees of freedom are your friends,” and then carefully demonstrated their value. The author acknowledges the helpful comments on this manuscript by the reviewers and by Scott Gronlund, Patrick O’Keefe, and Jacci Rodgers.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

Open Data: not applicable
 Open Materials: not applicable
 Preregistration: not applicable

Notes

1. Walker added, however, that “the concept was familiar to Gauss and his astronomical associates” more than a century earlier, as Gauss indicated “both in words and by formula that the number of observations is to be decreased by the number of unknowns estimated from the data to serve as divisor in estimating the standard error of a set of observations” (p. 253). Further, there is some disagreement about the timing of the concept in Fisher’s thinking. For example, Upton and Cook (2002) listed 1922 as the date when he introduced the concept.
2. Walker’s second definition—degrees of freedom as the dimensionality of a constrained data space—is perhaps the least accessible to nonexperts, but it is the one that motivated Fisher to develop the concept in the first place. A brief description of that conceptualization is provided in Box 3, but it is mathematically beyond the sophistication of typical introductory students.
3. The reconceptualization of degrees of freedom in this article is not intended to be comprehensive, or the only approach; other scholars and teachers will likely expand and clarify important features. For example, no attention is given to degrees of freedom as the rank of a quadratic form, which is important statistically, but does not inform the presentation in this article. However, it may be important in textbook treatment to retain the approach from Walker’s third definition, the approach matching computation of degrees of freedom to the statistical model, because of the pragmatic necessity for researchers conducting statistical analyses to have facility with those formulas.
4. Readers interested in the concept of fungibility from an economic perspective can find an engaging treatment in Galbraith’s (1975) book *Money: Whence It Came, Where It Went*.

References

Black, K. (1994). *Business statistics: Contemporary decision making*. Minneapolis, MN: West.

Box, J. F. (1978). *R. A. Fisher: The life of a scientist*. New York, NY: Wiley.

Cramér, H. (1946). *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.

Dallal, G. E. (2003). *Degrees of freedom*. Retrieved from <http://www.jerrydallal.com/LHSP/DOF.HTM>

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York, NY: John Wiley & Sons.

Eisenhauer, J. G. (2008). Degrees of freedom. *Teaching Statistics*, 30, 75–78.

Everitt, B. S. (2002). *The Cambridge dictionary of statistics* (2nd ed.). Cambridge, England: Cambridge University Press.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10, 507–521.

Fisher, R. A. (1922). Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. *Journal of Agricultural Science*, 11, 107–135.

Fisher, R. A., & Mackenzie, W. A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. *Journal of Agricultural Science*, 13, 311–320.

Galbraith, J. K. (1975). *Money: Whence it came, where it went*. Princeton, NJ: Princeton University Press.

Good, J. (1973). What are degrees of freedom? *The American Statistician*, 27, 227–228.

Hays, W. L. (1988). *Statistics* (4th ed.). New York, NY: Holt, Rinehart and Winston.

Jaccard, J., & Jacoby, J. (2009). *Theory construction and model-building skills: A practical guide for social scientists*. New York, NY: Guilford Press.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2017). *Designing experiments and analyzing data: A model comparison perspective*. Abingdon-on-Thames, England: Routledge.

Mulaik, S. (2010). *Foundations of factor analysis* (2nd ed.). Boca Raton, FL: CRC Press.

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, 65, 1–12.

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42, 59–66.

Saville, D. J., & Wood, D. J. (1991). *Statistical methods: The geometric approach*. New York, NY: Springer-Verlag.

Tippett, L. H. C. (1931). *The methods of statistics*. London, England: Williams and Norgate.

Upton, G., & Cook, I. (2002). *A dictionary of statistics* (2nd ed.). Oxford, England: Oxford University Press.

Walker, H. M. (1940). Degrees of freedom. *Journal of Educational Psychology*, 31, 253–269.

Wickens, T. D. (1995). *The geometry of multivariate statistics*. Hillsdale, NJ: Erlbaum.

Wikipedia. (2015). *Degrees of freedom*. Retrieved from https://en.wikipedia.org/wiki/Degrees_of_freedom