RESEARCH NOTE

What Are Degrees of Freedom?

Shanta Pandey and Charlotte Lyn Bright

As we were teaching a multivariate statistics course for doctoral students, one of the students in the class asked, "What are degrees of freedom? I know it is not good to lose degrees of freedom, but what are they?" Other students in the class waited for a clear-cut response. As we tried to give a textbook answer, we were not satisfied and we did not get the sense that our students understood. We looked through our statistics books to determine whether we could find a more clear way to explain this term to social work students. The wide variety of language used to define *degrees of freedom* is enough to confuse any social worker! Definitions range from the broad, "Degrees of freedom are the number of values in a distribution that are free to vary for any particular statistic" (Healey, 1990, p. 214), to the technical:

> Statisticians start with the number of terms in the sum [of squares], then subtract the number of mean values that were calculated along the way. The result is called the degrees of freedom, for reasons that reside, believe it or not, in the theory of thermodynamics. (Norman & Streiner, 2003, p. 43)

Authors who have tried to be more specific have defined degrees of freedom in relation to sample size (Trochim, 2005; Weinbach & Grinnell, 2004), cell size (Salkind, 2004), the number of relationships in the data (Walker, 1940), and the difference in dimensionalities of the parameter spaces (Good, 1973). The most common definition includes the number or pieces of information that are free to vary (Healey, 1990; Jaccard & Becker, 1990; Pagano, 2004; Warner, 2008; Wonnacott & Wonnacott, 1990). These specifications do not seem to augment students' understanding of this term. Hence, degrees of freedom are conceptually difficult but are important to report to understand statistical analysis. For example, without degrees of freedom, we are unable to calculate or to understand any underlying population variability. Also, in a bivariate and multivariate analysis, degrees of freedom are a function of sample size, number of variables, and number of parameters to be estimated; therefore, degrees of freedom are also associated with statistical power. This research note is intended to comprehensively define degrees of freedom, to explain how they are calculated, and to give examples of the different types of degrees of freedom in some commonly used analyses.

DEGREES OF FREEDOM DEFINED

In any statistical analysis the goal is to understand how the variables (or parameters to be estimated) and observations are linked. Hence, degrees of freedom are a function of both sample size (N) (Trochim, 2005) and the number of independent variables (k) in one's model (Toothaker & Miller, 1996; Walker, 1940; Yu, 1997). The degrees of freedom are equal to the number of independent observations (N), or the number of subjects in the data, minus the number of parameters (k) estimated (Toothaker & Miller, 1996; Walker, 1940). A parameter (for example, slope) to be estimated is related to the value of an independent variable and included in a statistical equation (an additional parameter is estimated for an intercept in a general linear model). A researcher may estimate parameters using different amounts or pieces of information, and the number of independent pieces of information he or she uses to estimate a statistic or a parameter are called the degrees of freedom (*df*) (HyperStat Online, n.d.). For example, a researcher records income of N number of individuals from a community. Here he or she has N independent pieces of information (that is, N points of incomes) and one variable called income (k); in subsequent analysis of this data set, degrees of freedom are associated with both N and k. For instance, if this researcher wants to calculate sample variance to understand the extent to which incomes vary in this community, the degrees of freedom equal $N - k$. The relationship between sample size and degrees of freedom is

CCC Code: 1070-5309/08 $3.00

positive; as sample size increases so do the degrees of freedom. On the other hand, the relationship between the degrees of freedom and number of parameters to be estimated is negative. In other words, the degrees of freedom decrease as the number of parameters to be estimated increases. That is why some statisticians define degrees of freedom as the number of independent values that are left after the researcher has applied all the restrictions (Rosenthal, 2001; Runyon & Haber, 1991); therefore, degrees of freedom vary from one statistical test to another (Salkind, 2004). For the purpose of clarification, let us look at some examples.

A Single Observation with One Parameter to Be Estimated

If a researcher has measured income ($k = 1$) for one observation ($N = 1$) from a community, the mean sample income is the same as the value of this observation. With this value, the researcher has some idea of the mean income of this community but does not know anything about the population spread or variability (Wonnacott & Wonnacott, 1990). Also, the researcher has only one independent observation (income) with a parameter that he or she needs to estimate. The degrees of freedom here are equal to $N - k$. Thus, there is no degree of freedom in this example ($1 - 1 = 0$). In other words, the data point has no freedom to vary, and the analysis is limited to the presentation of the value of this data point (Wonnacott & Wonnacott, 1990; Yu, 1997). For us to understand data variability, N must be larger than 1.

Multiple Observations (*N*) with One Parameter to Be Estimated

Suppose there are N observations for income. To examine the variability in income, we need to estimate only one parameter (that is, sample variance) for income (k), leaving the degrees of freedom of $N - k$. Because we know that we have only one parameter to estimate, we may say that we have a total of $N - 1$ degrees of freedom. Therefore, all univariate sample characteristics that are computed with the sum of squares including the standard deviation and variance have $N - 1$ degrees of freedom (Warner, 2008).

Degrees of freedom vary from one statistical test to another as we move from univariate to bivariate and multivariate statistical analysis, depending on the nature of restrictions applied even when sample size remains unchanged. In the examples that follow, we explain how degrees of freedom are calculated in some of the commonly used bivariate and multivariate analyses.

Two Samples with One Parameter (or *t* Test)

Suppose that the researcher has two samples, men and women, or $n_1 + n_2$ observations. Here, one can use an independent samples *t* test to analyze whether the mean incomes of these two groups are different. In the comparison of income variability between these two independent means (or k number of means), the researcher will have $n_1 + n_2 - 2$ degrees of freedom. The total degrees of freedom are the sum of the number of cases in group 1 and group 2 minus the number of groups. As a case in point, see the SAS and SPSS outputs of a *t* test comparing the literacy rate (LITERACY, dependent variable) of poor and rich countries (GNPSPLIT, independent variable) in Table 1. All in all, SAS output has four different values of degrees of freedom (two of which are also given by SPSS). We review each of them in the following paragraphs.

The first value for degrees of freedom under *t* tests is 100 (reported by both SAS and SPSS). The two groups of countries (rich and poor) are assumed to have equal variances in their literacy rate, the dependent variable. This first value of degrees of freedom is calculated as $n_1 + n_2 - 2$ (the sum of the sample size of each group compared in the *t* test minus the number of groups being compared), that is, $64 + 38 - 2 = 100$.

For the test of equality of variance, both SAS and SPSS use the F test. SAS uses two different values of degrees of freedom and reports folded F statistics. The *numerator degrees of freedom* are calculated as $n_1 - 1$, that is $64 - 1 = 63$. The *denominator degrees of freedom* are calculated as $n_2 - 1$ or $38 - 1 = 37$. These degrees of freedom are used in testing the assumption that the variances in the two groups (rich and poor countries, in our example) are not significantly different. These two values are included in the calculations computed within the statistical program and are reported on SAS output as shown in Table 1. SPSS, however, computes Levene's weighted F statistic (see Table 1) and uses $k - 1$ and $N - k$ degrees of freedom, where k stands for the number of groups being compared and N stands for the total number of observations in the sample; therefore, the degrees of freedom associated with the Levene's F statistic

Table 1: SAS and SPSS Outputs of a *t* Test Comparing the Literacy Rate (Dependent Variable) of Poor and Rich Countries (Independent Variable) with Reported Degrees of Freedom (N = 102)

The SAS System

The TTEST Procedure

Statistics

Variable	GNPSPLIT	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev
LITERACY	poor(0)	64	40.156	46.563	52.969	21.846	25.647
LITERACY	rich(1)	38	83.034	88.974	94.914	14.733	18.071
LITERACY	Diff (1-2)	-51.81	-42.41	-33.01	20.325	23.135	

Statistics

Variable	GNPSPLIT	Upper CL Std Dev	Std Err	Minimum	Maximum
LITERACY	poor(0)	31.062	3.2059	5	95
LITERACY	Rich(1)	23.38	2.9315	25	100
LITERACY	Diff (1-2)	26.854	4.7379		

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
LITERACY	Pooled	Equal	100	-8.95	<.0001
LITERACY	Satterthwaite	Unequal	97	-9.76	<.0001

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
LITERACY	Folded F	63	37	2.01	0.0236

SPSS SYSTEM

Group Statistics

	GNPSPLIT	*N*	*M*	*SD*	*SEM*
LITERACY	0 (poor)	64	46.563	25.6471	3.2059
	1 (rich)	38	88.974	18.0712	2.9315

		Levene's Test for Equality of Variances	
		F	*p*
LITERACY	Equal variances assumed	14.266	.000
	Equal variances not assumed		

		t test for Equality of Means						
		t	df	*p* (2-tailed)	*M* Difference	*SE* Difference	95% CI of the Difference	
							Lower	Upper
LITERACY	Equal variances assumed	-8.951	100	.000	-42.4112	4.7379	-51.8111	-33.0113
	Equal variances not assumed	-9.763	96.967	.000	-42.4112	4.3441	-51.0331	-33.7892

Note: CL (SAS output) or CI (SPSS output) = confidence interval. GNPSPLIT = Countries are divided into rich (1) and poor (0) based on their gross national product (GNP) per capita in 1980. SEM = standard error of the mean.

are the same (that is, $k - 1 = 2 - 1 = 1$, $N - k = 102 - 2 = 100$) as the degrees of freedom associated with "equal" variance test discussed earlier, and therefore SPSS does not report it separately.

If the assumption of equal variance is violated and the two groups have different variances as is the case in this example, where the folded F test or Levene's F weighted statistic is significant, indicating that the two groups have significantly different variances, the value for degrees of freedom (100) is no longer accurate. Therefore, we need to estimate the correct degrees of freedom (SAS Institute, 1985; also see Satterthwaite, 1946, for the computations involved in this estimation).

We can estimate the degrees of freedom according to Satterthwaite's (1946) method by using the following formula:

df Satterthwaite =

$$\frac{(n_1 - 1)\,(n_2 - 1)}{(n_1 - 1)\left[1 - \frac{S_1^2 n_2}{(S_1^2 n_2 + S_2^2 n_1)}\right]^2 + (n_2 - 1)\left[\frac{S_1^2 n_2}{(S_1^2 n_2 + S_2^2 n_1)}\right]^2}$$

where n_1 = sample size of group 1, n_2 = sample size of group 2, and S_1 and S_2 are the standard deviations of groups 1 and 2, respectively. By inserting subgroup data from Table 1, we arrive at the more accurate degrees of freedom as follows:

$$\frac{(64 - 1)\,(38 - 1)}{(64 - 1)\left[1 - \frac{25.65^2 \times 38}{\left(\begin{array}{l}25.65^2 \times 38\\ + 18.07^2 \times 64\end{array}\right)}\right]^2 + (38 - 1)\left[\frac{25.65^2 \times 38}{\left(\begin{array}{l}25.65^2 \times 38\\ + 18.07^2 \times 64\end{array}\right)}\right]^2}$$

$$= \frac{2331}{63\left[1 - \frac{25000.96}{\left(\begin{array}{l}25000.96\\ + 20897.59\end{array}\right)}\right]^2 + 37\left[\frac{25000.96}{\left(\begin{array}{l}25000.96\\ + 20897.59\end{array}\right)}\right]^2}$$

$$= \frac{2331}{63\left[1 - \frac{25000.96}{45898.55}\right]^2 + 37\left[\frac{25000.96}{45898.55}\right]^2}$$

$$= \frac{2331}{63[1 - .5447]^2 + 37[.5447]^2}$$

$$= \frac{2331}{63 \times .207 + 37 \times .2967} = \frac{2331}{24.0379} = 96.97$$

Because the assumption of equality of variances is violated, in the previous analysis the Satterthwaite's value for degrees of freedom, 96.97 (SAS rounds it to 97), is accurate, and our earlier value, 100, is not. Fortunately, it is no longer necessary to hand calculate this as major statistical packages such as SAS and SPSS provide the correct value for degrees of freedom when the assumption of equal variance is violated and equal variances are not assumed. This is the fourth value for degrees of freedom in our example, which appears in Table 1 as 97 in SAS and 96.967 in SPSS. Again, this value is the correct number to report, as the assumption of equal variances is violated in our example.

Comparing the Means of g Groups with One Parameter (Analysis of Variance)

What if we have more than two groups to compare? Let us assume that we have $n_1 + \ldots + n_g$ groups of observations or countries grouped by political freedom (FREEDOMX) and that we are interested in differences in their literacy rates (LITERACY, the dependent variable). We can test the variability of g means by using the analysis of variance (ANOVA). The ANOVA procedure produces three different types of degrees of freedom, calculated as follows:

- The first type of degrees of freedom is called the *between-groups degrees of freedom* or *model degrees of freedom* and can be determined by using the number of group means we want to compare. The ANOVA procedure tests the assumption that the g groups have equal means and that the population mean is not statistically different from the individual group means. This assumption reflects the null hypothesis, which is that there is no statistically significant difference between literacy rates in g groups of countries ($\mu_1 = \mu_2 = \mu_3$). The alternative hypothesis is that the g sample means are significantly different from one another. There are $g - 1$ model degrees of freedom for testing the null hypothesis and for assessing variability among the g means. This value of model degrees of freedom is used in the numerator for calculating the F ratio in ANOVA.
- The second type of degrees of freedom, called the *within-groups degrees of freedom* or *error degrees of freedom*, is derived from subtracting the model degrees of freedom from the corrected total degrees of freedom. The within-groups

degrees of freedom equal the total number of observations minus the number of groups to be compared, $n_1 + \ldots + n_g - g$. This value also accounts for the denominator degrees of freedom for calculating the F statistic in an ANOVA.

- Calculating the third type of degrees of freedom is straightforward. We know that the sum of deviation from the mean or $\Sigma(Y_i - \bar{Y}) = 0$. We also know that the total sum of squares or $\Sigma(Y_i - \bar{Y})^2$ is nothing but the sum of N^2 deviations from the mean. Therefore, to estimate the total sum of squares $\Sigma(Y_i - \bar{Y})^2$, we need only the sum of $N - 1$ deviations from the mean. Therefore, with the total sample size we can obtain the total degrees of freedom, or corrected total degrees of freedom, by using the formula $N - 1$.

In Table 2, we show the SAS and SPSS output with these three different values of degrees of freedom using the ANOVA procedure. The dependent variable, literacy rate, is continuous, and the independent variable, political freedom or FREEDOMX, is nominal. Countries are classified into three groups on the basis of the amount of political freedom each country enjoys: Countries that enjoy high political freedom are coded as 1 ($n = 32$), countries that enjoy moderate political freedom are coded as 2 ($n = 34$), and countries that enjoy no political freedom are coded as 3 ($n = 36$). The mean literacy rates (dependent variable) of these groups of countries are examined. The null hypothesis tests the assumption that there is no significant difference in the literacy rates of these countries according to their level of political freedom.

The first of the three degrees of freedom, the between-groups degrees of freedom, equals $g - 1$. Because there are three groups of countries in this analysis, we have $3 - 1 = 2$ degrees of freedom. This accounts for the numerator degrees of freedom in estimating the F statistic. Second, the within-groups degrees of freedom, which accounts for the denominator degrees of freedom for calculating the F statistic in ANOVA, equals $n_1 + \ldots + n_g - g$. These degrees of freedom are calculated as $32 + 34 + 36 - 3 = 99$. Finally, the third degrees of freedom, the total degrees of freedom, are calculated as $N - 1$ ($102 - 1 = 101$). When reporting F values and their respective degrees of freedom, researchers should report them as follows: The independent and the dependent variables are significantly related [$F(2, 99) = 16.64, p <.0001$].

Degrees of Freedom in Multiple Regression Analysis

We skip to multiple regression because degrees of freedom are the same in ANOVA and in simple regression. In multiple regression analysis, there is more than one independent variable and one dependent variable. Here, a parameter stands for the relationship between a dependent variable (Y) and each independent variable (X). One must understand four different types of degrees of freedom in multiple regression.

- The first type is the *model (regression) degrees of freedom*. Model degrees of freedom are associated with the number of independent variables in the model and can be understood as follows:

  A null model or a model without independent variables will have zero parameters to be estimated. Therefore, predicted Y is equal to the mean of Y and the degrees of freedom equal 0.

  A model with one independent variable has one predictor or one piece of useful information ($k = 1$) for estimation of variability in Y. This model must also estimate the point where the regression line originates or an intercept. Hence, in a model with one predictor, there are ($k + 1$) parameters—k regression coefficients plus an intercept—to be estimated, with k signifying the number of predictors. Therefore, there are $[(k + 1) - 1]$, or k degrees of freedom for testing this regression model.

  Accordingly, a multiple regression model with more than one independent variable has some more useful information in estimating the variability in the dependent variable, and the model degrees of freedom increase as the number of independent variables increase. The null hypothesis is that all of the predictors have the same regression coefficient of zero, thus there is only one common coefficient to be estimated (Dallal, 2003). The alternative hypothesis is that the regression coefficients are not zero and that each variable explains a different amount of variance in the dependent variable. Thus, the researcher must estimate k coefficients plus the intercept. Therefore,

Table 2: Analysis of Variance (ANOVA) of Literacy Rates According to Countries' Levels of Freedom with Reported Degrees of Freedom (N = 102)

SAS SYSTEM

The ANOVA Procedure

Dependent Variable: LITERACY perct of adult population literate b

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	24253.48284	12126.74142	16.64	<.0001
Error	99	72156.09559	728.84945		
Corrected Total	101	96409.57843			

R-Square	Coeff Var	Root MSE	LITERACY Mean
0.251567	43.29061	26.99721	62.36275

Source	DF	Anova SS	Mean Square	F Value	Pr > F
FREEDOMX	2	24253.48284	12126.74142	16.64	<.0001

The ANOVA Procedure

Duncan's Multiple Range Test[1] for LITERACY

Alpha	0.05
Error Degrees of Freedom	99
Error Mean Square	728.8495
Harmonic Mean of Cell Sizes	33.92148

Number of Means[2]	2	3
Critical Range	13.01	13.69

Duncan Grouping	Mean[3]	N	FREEDOMX
A	84.313	32	1 (=free)
B	57.529	34	2 (=partly free)
B			
B	47.417	36	3 (=notfree)

SPSS SYSTEM

LITERACY

	Sum of Squares	df	M^2	F	p
Between Groups	24253.483	2	12126.741	16.638	.000
Within Groups	72156.096	99	728.849		
Total	96409.578	101			

Post Hoc Tests
Homogeneous Subsets
LITERACY

Duncan [a,b]

		Subset (α = .05)	
FREEDOMX	N	1	2
3 (=not free)	36	47.417	
2 (=partly free)	34	57.529	
1 (=free)	32		84.313
p		.126	1.000

Means for groups in homogeneous subsets are displayed.
[a]Uses Harmonic Mean Sample Size = 33.921.
[b]The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Note: [1]Duncan's multiple range test for literacy controls the Type I comparisonwise error rate, not the experimentwise error rate. [2]Cell sizes are not equal. [3]Means with the same letter are not significantly different. Perct = percentage. Coeff Var = coefficient of variation. SS =Sum of Squares.

there are $(k + 1) - 1$ or k degrees of freedom for testing the null hypothesis (Dallal, 2003). In other words, the model degrees of freedom equal the number of useful pieces of information available for estimation of variability in the dependent variable.

- The second type is the *residual*, or *error, degrees of freedom*. Residual degrees of freedom in multiple regression involve information of both sample size and predictor variables. In addition, we also need to account for the intercept. For example, if our sample size equals N, we need to estimate $k + 1$ parameters, or one regression coefficient for each of the predictor variables (k) plus one for the intercept. The residual degrees of freedom are calculated $N - (k + 1)$. This is the same as the formula for the error, or within-groups, degrees of freedom in the ANOVA. It is important to note that increasing the number of predictor variables has implications for the residual degrees of freedom. Each additional parameter to be estimated costs one residual degree of freedom (Dallal, 2003). The remaining residual degrees of freedom are used to estimate variability in the dependent variable.
- The third type of degrees of freedom is the *total*, or *corrected total, degrees of freedom*. As in ANOVA, this is calculated $N - 1$.
- Finally, the fourth type of degrees of freedom that SAS (and not SPSS) reports under the parameter estimate in multiple regression is worth mentioning. Here, the null hypothesis is that there is no relationship between each independent variable and the dependent variable. The degree of freedom is always 1 for each relationship and therefore, some statistical software, such as SPSS, do not bother to report it.

In the example of multiple regression analysis (see Table 3), there are four different values of degrees of freedom. The first is the regression degrees of freedom. This is estimated as $(k + 1) - 1$ or $(6 + 1) - 1 = 6$, where k is the number of independent variables in the model. Second, the residual degrees of freedom are estimated as $N - (k + 1)$. Its value here is $99 - (6 + 1) = 92$. Third, the total degrees of freedom are calculated $N - 1$ (or $99 - 1 = 98$). Finally, the degrees of freedom shown under parameter estimates for each parameter always equal 1, as explained above. F values and the respective degrees of freedom from the current regression output should be reported as follows: The regression model is statistically significant with $F(6, 92) = 44.86, p < .0001$.

Degrees of Freedom in a Nonparametric Test

Pearson's chi square, or simply the chi-square statistic, is an example of a nonparametric test that is widely used to examine the association between two nominal level variables. According to Weiss (1968) "the number of degrees of freedom to be associated with a chi-square statistic is equal to the number of independent components that entered into its calculation" (p. 262). He further explained that each cell in a chi-square statistic represents a single component and that an independent component is one where neither observed nor expected values are determined by the frequencies in other cells. In other words, in a contingency table, one row and one column are fixed and the remaining cells are independent and are free to vary. Therefore, the chi-square distribution has $(r - 1) \times (c - 1)$ degrees of freedom, where r is the number of rows and c is the number of columns in the analysis (Cohen, 1988; Walker, 1940; Weiss, 1968). We subtract one from both the number of rows and columns simply because by knowing the values in other cells we can tell the values in the last cells for both rows and columns; therefore, these last cells are not independent.

As an example, we ran a chi-square test to examine whether gross national product (GNP) per capita of a country (GNPSPLIT) is related to its level of political freedom (FREEDOMX). Countries (GNPSPLIT) are divided into two categories—rich countries or countries with high GNP per capita (coded as 1) and poor countries or countries with low GNP per capita (coded as 0), and political freedom (FREEDOMX) has three levels—free (coded as 1), partly free (coded as 2), not free (coded as 3) (see Table 4). In this analysis, the degrees of freedom are $(2 - 1) \times (3 - 1) = 2$. In other words, by knowing the number of rich countries, we would automatically know the number of poor countries. But by knowing the number of countries that are free, we would not know the number of countries that are partly free and not free. Here, we need to know two of the three components—for instance, the number of countries that are free and partly free—so that we will know the number of countries

Table 3: Multiple Regression of Literacy Rates on Educational and Social Variables with Reported Degrees of Freedom (N = 99)

SAS SYSTEM

The REG Procedure

Model: MODEL1

Dependent Variable: LITERACY perct of adult population literate b

Number of Observations Read 192
Number of Observations Used 99
Number of Observations with Missing Values 93

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	70570	11762	44.86	<.0001
Error	92	24120	262.17651		
Corrected Total	98	94690			

Root MSE 16.19187 R-Square 0.7453
Dependent Mean 62.02020 Adj R-Sq 0.7287
Coeff Var 26.10741

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	29.72514	9.27642	3.20	0.0019
free	Free Countries	1	5.39558	5.00005	1.08	0.2834
partfree	Partly Free Countries	1	-1.62820	4.12821	-0.39	0.6942
EDFUND84	education expenditures 1984 b	1	0.00009	0.00008	0.58	0.5605
PUPILS80	PULILS PER TEACHER 1980 ISD	1	-0.69716	0.15416	-4.52	<.0001
SCHOOL80	PRIMARY SCHOOL STUDENTS 1980 ISD	1	0.58254	0.06762	8.61	<.0001
GNP80	GNP PER CAPITA 1980 ISD	1	0.00119	0.00062	1.92	<.0583

SPSS SYSTEM

Model Summary

Model	R	R^2	Adjusted R^2	SE of the Estimate
1	.863[a]	.745	.729	16.1919

[a]Predictors: (Constant), GNP80, partfree, EDFUND84, SCHOOL80, PUPILS80, free

Analysis of Variance: Dependent Variable = Literacy

Model		Sum of Squares	df	M^2	F	p
1	Regression	70569.721	6	11761.620	44.861	.000[a]
	Residual	24120.239	92	262.177		
	Total	94689.960	98			

[a]Predictors: (Constant), GNP80, partfree, EDFUND84, SCHOOL80, PUPILS80, free

Coefficients: Dependent Variable = Literacy

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	SE	β	t	p
1	(Constant)	29.725	9.276	3.204	.002	
	free	5.396	5.000	.082	1.079	.283
	partfree	-1.628	4.128	-.025	-.394	.694
	EDFUND84	4.83E-005	.000	.034	.584	.561
	PUPILS80	-.697	.154	-.310	-4.522	.000
	SCHOOL80	.583	.068	.524	8.615	.000
	GNP80	.001	.001	.158	1.917	.058

Note: Perct = percentage. Coeff Var = coefficient of variation. Adj. = adjusted. GNP = gross national product.

Table 4: Chi Square of GNPSPLIT and Level of Freedom in Countries with Reported Degrees of Freedom (N = 124)

SAS SYSTEM

The FREQ Procedure

Table of GNPSPLIT by FREEDOMX

```
GNPSPLIT        FREEDOMX(FREEDOM INDEX ISD)
Frequency ,
Percent   ,
Row Pct   ,
Col Pct   ,      1 ,      2 ,      3 ,   Total
ffffffffˆffffffffˆffffffffˆffffffffˆ
        0 ,     10 ,     31 ,     35 ,      76
          ,   8.06 ,  25.00 ,  28.23 ,   61.29
          ,  13.16 ,  40.79 ,  46.05 ,
          ,  28.57 ,  75.61 ,  72.92 ,
ffffffffˆffffffffˆffffffffˆffffffffˆ
        1 ,     25 ,     10 ,     13 ,      48
          ,  20.16 ,   8.06 ,  10.48 ,   38.71
          ,  52.08 ,  20.83 ,  27.08 ,
          ,  71.43 ,  24.39 ,  27.08 ,
ffffffffˆffffffffˆffffffffˆffffffffˆ
Total           35        41        48       124
             28.23     33.06     38.71    100.00
```

Frequency Missing = 68

Statistics for Table of GNPSPLIT by FREEDOMX

```
Statistic                       DF       Value      Prob
ffffffffffffffffffffffffffffffffffffffffffffffffffffffff
Chi-Square                       2     22.0708    <.0001
Likelihood Ratio Chi-Square      2     22.0179    <.0001
Mantel-Haenszel Chi-Square       1     14.8569    0.0001
Phi Coefficient                         0.4219
Contingency Coefficient                 0.3887
Cramer's V                              0.4219
```

Effective Sample Size = 124

Frequency Missing = 68

SPSS SYSTEM

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	%	*N*	%	*N*	%
GNPSPLIT * FREEDOMX	124	64.6	68	35.4	192	100.0

GNPSPLIT * FREEDOMX Crosstabulation

			FREEDOMX			Total
			1	2	3	
GNPSPLIT	0	Count	10	31	35	76
		% of Total	8.1%	25.0%	28.2%	61.3%
	1	Count	25	10	13	48
		% of Total	20.2%	8.1%	10.5%	38.7%
Total		Count	35	41	48	124
		% of Total	28.2%	33.1%	38.7%	100.0%

Chi-Square Tests

	Value	*df*	Asymptotic. *p* (2-tailed)
Pearson χ^2	22.071[a]	2	.000
Likelihood Ratio	22.018	2	.000
Linear-by-Linear Association	14.857	1	.000
Valid Cases (*N*)	124		

[a]Zero cells (0.0%) have expected count less than 5. The minimum expected count is 13.55.

Note: For the SAS, 35% of data are missing. GNPSPLIT = Countries divided into rich (1) and poor (0) based on their GNP per capita in 1980. FREEDOMX = countries divided into three categories based on their level of political freedom—1=free; 2=partly free; 3=not free.

that are not free. Therefore, in this analysis there are two independent components that are free to vary, and thus the degrees of freedom are 2.

Readers may note that there are three values under degrees of freedom in Table 4. The first two values are calculated the same way as discussed earlier and have the same values and are reported most widely. These are the values associated with the Pearson chi-square and likelihood ratio chi-square tests. The final test is rarely used. We explain this briefly. The degree of freedom for the Mantel–Haenszel chi-square statistic is calculated to test the hypothesis that the relationship between two variables (row and column variables) is linear; it is calculated as $(N - 1) \times r^2$, where r^2 is the Pearson product-moment correlation between the row variable and the column variable (SAS Institute, 1990). This degree of freedom is always 1 and is useful only when both row and column variables are ordinal.

CONCLUSION

Yu (1997) noted that "degree of freedom is an intimate stranger to statistics students" (p. 1). This research note has attempted to decrease the strangeness of this relationship with an introduction to the logic of the use of degrees of freedom to correctly interpret statistical results. More advanced researchers, however, will note that the information provided in this article is limited and fairly elementary. As degrees of freedom vary by statistical test (Salkind, 2004), space prohibits a more comprehensive demonstration. Anyone with a desire to learn more about degrees of freedom in statistical calculations is encouraged to consult more detailed resources, such as Good (1973), Walker (1940), and Yu (1997).

Finally, for illustrative purposes we used World Data that reports information at country level. In our analysis, we have treated each country as an independent unit of analysis. Also, in the analysis, each country is given the same weight irrespective of its population size or area. We have ignored limitations that are inherent in the use of such data. We warn readers to ignore the statistical findings of our analysis and take away only the discussion that pertains to degrees of freedom. SWR

REFERENCES

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Dallal, G. E. (2003). *Degrees of freedom*. Retrieved May 30, 2006, from http://www.tufts.edu/~gdallal/dof.htm

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

Healey, J. F. (1990). *Statistics: A tool for social research* (2nd ed.). Belmont, CA: Wadsworth.

HyperStat Online. (n.d.). *Degrees of freedom*. Retrieved May 30, 2006, from http://davidmlane.com/hyperstat/A42408.html

Jaccard, J., & Becker, M. A. (1990). *Statistics for the behavioral sciences* (2nd ed.). Belmont, CA: Wadsworth.

Norman, G. R., & Streiner, D. L. (2003). *PDQ statistics* (3rd ed.). Hamilton, Ontario, Canada: BC Decker.

Pagano, R. R. (2004). *Understanding statistics in the behavioral sciences* (7th ed.). Belmont, CA: Wadsworth.

Rosenthal, J. A. (2001). *Statistics and data interpretation for the helping professions.* Belmont, CA: Wadsworth.

Runyon, R. P., & Haber, A. (1991). *Fundamentals of behavioral statistics* (7th ed.). New York: McGraw-Hill.

Salkind, N. J. (2004). *Statistics for people who (think they) hate statistics* (2nd ed.). Thousand Oaks, CA: Sage Publications.

SAS Institute. (1985). *SAS user's guide: Statistics, version 5.1.* Cary, NC: SAS Institute.

SAS Institute. (1990). *SAS procedure guide, version 6* (3rd ed.). Cary, NC: SAS Institute.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*, 110–114.

Toothaker, L. E., & Miller, L. (1996). *Introductory statistics for the behavioral sciences* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Trochim, W.M.K. (2005). *Research methods: The concise knowledge base.* Cincinnati: Atomic Dog.

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

Warner, R. M. (2008). *Applied statistics.* Thousand Oaks, CA: Sage Publications.

Weinbach, R. W., & Grinnell, R. M., Jr. (2004). *Statistics for social workers* (6th ed.). Boston: Pearson.

Weiss, R. S. (1968). *Statistics in social research: An introduction.* New York: John Wiley & Sons.

Wonnacott, T. H., & Wonnacott, R. J. (1990). *Introductory statistics* (5th ed.). New York: John Wiley & Sons.

Yu, C. H. (1997). *Illustrating degrees of freedom in terms of sample size and dimensionality.* Retrieved November 1, 2007, from http://www.creative-wisdom.com/computer/sas/df.html

Shanta Pandey, PhD, *is associate professor, and* ***Charlotte Lyn Bright, MSW,*** *is a doctoral student, George Warren Brown School of Social Work, Washington University, St. Louis.* *Correspondence concerning this article should be sent to Shanta Pandey, George Warren Brown School of Social Work, Washington University, St. Louis, MO 63130; e-mail: pandeys@wustl.edu.*

The authors are thankful to a student whose curious mind inspired them to work on this research note.

Original manuscript received August 29, 2006
Final revision received November 15, 2007
Accepted January 16, 2008