Review Article **AJERR 2021,7:89**

American Journal of Educational Research and Reviews
(ISSN:2474-9265)

Power and Sample Size for Contingency Tables

Seock-Ho Kim[1], Hyo Jin Eom[2]

[1]The University of Georgia, [2]Korea University

ABSTRACT

A review of association measures of effect size between two categorical variables in contingency tables is presented. Relationships among measures of effect size are explicated by considering the test statistics of independence, the nominal or ordinal nature of categorical variables, and the size of contingency tables. Tables that contain minimum sample sizes for testing independence between two categorical variables in contingency tables are also presented. Cramer's V^2 was employed as a main measure of association in tabulation. Illustrations are provided using data from 2018 General Social Survey for obtaining test statistics and measures of effect size for contingency tables. Determining appropriate sample sizes for statistical analysis of data in contingency tables is important for studies in behavioral sciences.

Keywords: categorical variable, contingency table, effect size, power, sample size

***Correspondence to Author:**
Seock-Ho Kim
Department of Educational Psychology, The University of Georgia, 325 Aderhold Hall, Athens, GA 30602-7143.

How to cite this article:
Seock-Ho Kim, Hyo Jin Eom. Power and Sample Size for Contingency Tables. American Journal of Educational Research and Reviews, 2022,7:89.

eSciPub LLC, Houston, TX USA.
Website: https://escipub.com/

Introduction

There are two purposes of this paper. The first is to present a review of association measures of effect size between two categorical variables in contingency tables. The second is to present tables that contain minimum sample sizes for testing independence between two categorical variables in contingency tables.

Because a required sample size depends on the risk of type I error, the risk of type II error, and the values of population parameters under an alternative hypothesis (i.e., effect size), measures of effect size for contingency tables are explicated in detail. Because there are many statistics that can be used in testing independence between two categorical variables depending on the nominal or ordinal nature of the respective categorical variables as well as the size of contingency tables (see e.g., Stokes, Davis, & Koch, 2000), the review of measures of effect size is mainly performed for contingency tables with more than two rows and two columns. Measures of effect size for fourfold tables are also briefly presented. Note, however, that measures of effect size and discussion about required sample sizes for sets of contingency tables, certain nonparametric methods, logistic regression, and loglinear models are not discussed.

Syntax code in R (Venables et al., 2009) that was used in power analysis is presented. Note that the two sample-size tables in this paper can be seen as an extension of Table 50-4 of Marascuilo and Serlin (1988, p. 746) that was adapted from Table 7.3.15 to Table 7.3.28 of Cohen (1988, pp. 235–241).

Theoretical Framework

At the outset it should be emphasized that both the American Educational Research Association (AERA, 2006) and the American Psychological Association (APA, 2013, 2020) clearly ask authors of their journal articles for which empirical data were analyzed with inferential statistical methods to report measures of effect size. Unfortunately, which measures of effect size might be reported for specific inferential statistical methods are not clearly stated by AERA or APA. There are many different measures of effect size depending upon the type of variables, quantitative or qualitative/ categorical (see Fleiss, 1994; Rosenthal, 1994), and the inferential statistical methods (see Ellis, 2010; Grissom & Kim, 2005). It is worth noting that measures of effect size as estimates are in fact sample statistics (Olejnik & Algina, 2000) and mostly can be obtained from sufficient statistics required to be reported for different inferential statistical methods. Note also that the standard errors of measures of effect size may not be routinely or readily available.

There are in general two classes of measures of effect size; one expressed as a standardized mean difference and the other as a proportion of variance explained. Exemplary articles are available for the first class (e.g., Bloom, 1984) and for the second class (e.g., Educational Testing Service, 1980, p. 18), respectively. Interpretations of measures of effect size are largely based on Cohen’s conventional, arbitrary “small,” “medium,” and “large” definitions (see Cohen, 1988, p. 12; Cohen, 1992).

In statistical hypothesis testing, the probability to reject the null hypothesis if it is false depends on (1) the risk of type I error, (2) the values of parameters as defined under the alternative hypothesis, and (3) the sample size. It is typical to set the probability of type I error to be .05. Because nature determines the values of parameters under the alternative hypothesis, the only way that a researcher can do to reduce the risk of making type II error is to increase the sample size (Marascuilo & Serlin, 1988, p. 740; Cohen, 1988, pp. 14–16). It is typical to set the probability of type II error to be .20 for the purpose of determining a minimum sample size.

Although many books and computer programs are available for power analysis to determine sample sizes (e.g., Faul, Erdfelder, Buchner, & Lang, 2009; Liu, 2014), only Cohen (1988) and Marascuilo and Serlin (1988) presented tables of minimum sample sizes for the testing of independence between two categorical variables in contingency tables, those larger than the fourfold ones. More refined tables than those from Cohen (1988) and Marascuilo and Serlin (1988) are presented in this paper together with syntax code in R.

This paper only presents a review of measures of effect size for contingency tables in conjunction with power analysis to determine minimum sample sizes. For general discussion of effect size, readers are referred to Ellis (2010) and Grissom and Kim (2005). For general discussion of power analysis, readers are referred to, of course, Cohen (1988) and Liu (2014).

Test Statistics and Measures of Effect Size

Several definitions as well as notations are needed to proceed. In this paper, a contingency table is defined as an array of positive integers in matrix form where the numbers represent counts or frequencies. For two categorical variables with I levels of a row variable and J levels of a column variable, an $I \times J$ contingency table has IJ cells for the numbers. For example, the 2 × 2 contingency table is called a fourfold (contingency) table because there are four cells.

The $I \times J$ contingency table has many special cases depends on the number of rows and columns. The $I \times J$ contingency table can be used to present a tabulation of two categorical or categorized variables obtained from a sample of persons. There are many different designs for the collection of data eventually represented in a $I \times J$ table. For a review, see Stokes et al. (2000).

For testing independence between two categorical variables in a contingency table, the chi-square test for differences in probabilities can be used (e.g., Agresti, 2007, p. 35). The chi-square test for independence can be used for data for which a random sample of size N is obtained, and the observations in the sample can be classified into the IJ cells according to the classification criteria. With the assumption of fixed marginal totals, the sampling distribution is a multivariate hypergeometric distribution under the null hypothesis of no association (Stokes et al., 2000, p. 93).

Although there are many different notations for the test statistics that can be employed in this paper, both sets of original and coherent notations are useful. The chi-square test statistic is

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}, \quad (1)$$

where O_{ij} is the observed count, E_{ij} is the expected count for the cell in the ith row and the jth column. We may use $O_{i.} = \sum_j O_{ij}$ and $O_{.j} = \sum_i O_{ij}$ then $E_{ij} = O_{i.}O_{.j}/N$. For the test of significance, the likelihood ratio test (e.g., G^2) can also be used (Agresti, 2007, p. 36). The likelihood-based measures of effect size are not considered in detail (see Agresti, 2013, p. 111; Sakamoto, 1985; Theil, 1972, pp. 115–120). Nevertheless, G^2 can replace X^2 in measures of effect size if hypothesis testing utilizes likelihood functions.

Effect size can be obtained by means of measures of association or dependence in contingency tables. In this sense, a contingency table is a convenient form for examining categorical data to assess some kind of dependence in the data. A thorough review of measures of effect size is presented next (see Conover, 1980; Everitt, 1977; Hays, 1994; Kendall & Stuart, 1979; Upton, 1978). Note that the chi-square test statistic underlies nearly all measures of association.

Cramer's (1946, p. 282) contingency coefficient is the main measure of effect size employed in this paper: $V^2 = X^2/(NM)$ where M is the smaller of $I - 1$ and $J - 1$. Some related of measures of effect size are listed below in a chronological order.

Pearson's (1904, p. 9) mean square contingency is

$$\phi^2 = \frac{X^2}{N} \quad (2)$$

for which X^2 is the square contingency. Here, the test statistic X^2 can be seen as a function of an effect size φ^2 and a sample size N. For the general $I \times J$ table with either $I > 2$ or $J > 2$, φ^2 can attain a value larger than unity and its maximum is M. Several other association measures were developed for a norming purpose.

Pearson's (1904, p. 9) first coefficient of contingency is

$$C_1 = \sqrt{\frac{\phi^2}{1+\phi^2}} = \sqrt{\frac{X^2}{N+X^2}}. \quad (3)$$

Note that C_1 is also named the mean square contingency coefficient (Pearson, 1904, p. 16), and called the contingency coefficient C without subscript 1. The maximum of C^2 is $M/(M + 1)$ (Kendall & Stuart, 1979, p. 588). Sakoda (1977) suggested $C^2/[M/(M + 1)]$ as a measure of association.

Tschuprow's (1939, p. 53) coefficient of contingency with $\phi^2 \equiv \varphi^2$ is

$$T^2 = \frac{\varphi^2}{\sqrt{(I-1)(J-1)}} = \frac{\phi^2}{\sqrt{(I-1)(J-1)}} = \frac{X^2}{N\sqrt{(I-1)(J-1)}}. \quad (4)$$

The upper limit of T^2 of unity can be attained only for $I = J$ (see Yule & Kendall, 1950, p. 53). The maximum of T^2 is $M/\sqrt{(I-1)(J-1)}$.

Cramer (1946, p. 282) suggested a modified measure of the mean square contingency with $\phi^2 \equiv \varphi^2$ as

$$V^2 = \frac{\varphi^2}{q-1} = \frac{\phi^2}{M} = \frac{X^2}{NM}, \quad (5)$$

where $q - 1 = M$. Accordingly, Cramer's V is $V = \sqrt{\phi^2/M} = \sqrt{X^2/(MN)}$. The upper limit of V^2 is unity even $I \neq J$.

Other association measures that include Goodman and Kruskal's λ measures, as well as those for ordered categories including Kendall's τ, and Goodman and Kruskal's γ are presented in Appendix A (see Everitt, 1977). Indices for fourfold tables are discussed in Appendix B (see Fleiss, 1994). Such fourfold-table indices are truly valuable because all omnibus tests should be ultimately converted to or partitioned into the ones with a single degree of freedom and because directional alternative hypotheses can be incorporated in decision making as well as power analysis (e.g., a priori or post hoc comparisons for tests of homogeneity; see Marascuilo & Serlin, 1988, Chapter 28; cf. Olejnik & Algina, 2000)

Method

Data

Many real and hypothetical data for contingency tables can be found in textbooks. In this paper, a contingency table based on two questionnaire items from 2018 General Social Survey (GSS) are used (i.e., gender, party affiliation). Two categorical variables yielded a contingency table that was used to illustrate how to obtain measures of effect size. Note that certainly many contingency tables in different sizes can be constructed from the 2018 GSS data. The required sample sizes are definitely dependent upon specific categorical variables used in statistical analysis of data. The GSS data set is available publicly from:

http://gss.norc.org/getthedata/Pages/Home.aspx

The same contingency table was used to illustrate empirical power calculation using R code. The probability of type I error, the sample size, and the empirical measure of effect size were used to obtain the empirical power.

Procedures

Measures of effect size that reflect the degree of association between two categorical variables in contingency tables have been presented in the previous section using information from many different sources. There are many other, different approaches to analyze categorical variables in contingency tables and to obtain measures of effect size. This paper addresses mainly parametric measures of association.

The real purpose of the review of association measures of effect size in contingency tables was for assessing a required minimum sample size. We used R code to obtain each sample size given a type I error (α), power (i.e., $1 - \beta$, where β is the probability of a type II error), degrees of freedom, and a measure of effect size. The measure of effect size in this paper is again V^2, but the value used for tabulation was

$$e = MV^2 \tag{6}$$

which is in fact φ^2. The same measure was used in Cohen (1988) and Marascuilo and Serlin (1988).

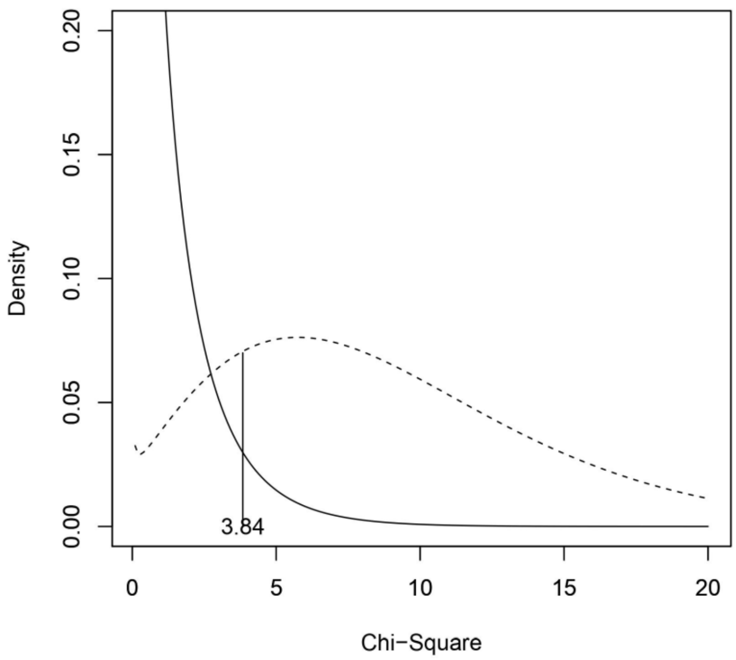

Figure 1: A chi-square distribution with one degree of freedom (solid line), and a non central chi-square distribution with one degree of freedom with non-centrality parameter $\lambda = N \times e$, where N = 88 and e = .09 (dotted line). The right-hand-side area of the critical value of 3.84 under the solid line is α = .05, and that under the dotted line is power (i.e., $1 - \beta$ = .80).

For example, with e = 0.09, df = 1, α = .05, β = .20 the following R code yields the required sample size as well as the exact power which is greater than $1 - \beta$. The product of a sample size N and the scaled effect size $e = MV^2$ were used to set a non-centrality parameter of the chi-square distribution for the calculation of power (Lancaster, 1969, pp. 117–134). The resulting sample size was N = 88 (see Figure 1).

```
e <- .09
dfree <- 1
alpha <- .05
beta <- .20
cv <- qchisq((1-alpha), df=dfree)
power <- 0
for (n in 1:1000) {
  power <- 1 - pchisq(cv, df=dfree, ncp=n*e)
  if (power > (1-beta)) {
    number <- n
    numberpower <- power
    break
  }
}
number; numberpower
```

It was assumed that the Pearson chi-square statistic X^2 obtained from the contingency table is distributed as a chi-square distribution with the degrees of freedom $df = v = (I - 1)(J - 1)$ (see Johnson, Kotz, & Balakrishnan, 1994, Chapter 18). The critical value at α was obtained from such a distribution. Power was calculated as the probability of rejecting the null hypothesis provided that the alternative hypothesis was in fact the correct one for which the noncentrality parameter, say λ, was set by the sample size N and the scaled effect size $e = MV^2$. The minimum sample size was determined by iteratively obtaining the value of power until it exceeded the prespecified value of $1 - \beta$.

Results

Tables 1 and 2 present required sample sizes for contingency tables. Let us discuss, however, the example data first. Table 3 presents cross classification of party affiliation by gender from the 2018 GSS. Expected frequencies under the null hypothesis of no association are shown in parentheses. The chi-square test statistic is X^2 = 12.903 with df = 2 and p = .002. Cramer's V^2 = 12.903/1332 = 0.009686936937 and V = 0.098. Note that $\varphi^2 = V^2$, T^2 = 0.007, C = 0.098, $\lambda(C|R)$ = 0.000, $\lambda(R|C)$ = 0.031, λ = 0.014. $\tau(C|R)$ = 0.006, and $\tau(R|C)$ = 0.010. The null hypothesis of no association was rejected at α = .05 based on the X^2 value, but only a small effect was observed between party affiliation and gender.

Assume that both row and column variables are on the ordinal scale (n.b., not a valid assumption for data in Table 3), G = 0.153, τ_b = 0.086, τ_c = 0.095, $D(C|R)$ = 0.096, $D(R|C)$ = 0.076, and D = 0.085. In order to be valid, some conceptual conversion of the party affiliation variable to another with ordered categories (e.g., liberal to conservative levels) should be applied.

Based on the statistical testing results, the observed power can be assessed. With $V^2 = e$ = 0.009686936937 and N = 1332, the observed power at α = .05 was obtained from the following R command (the observed power was .9058233):

```
e <- .009686936937
n <- 1332
1 - pchisq(qchisq(.95,df=2), df=2, ncp=n*e)
```

The results from computations of minimum sample sizes are presented in Tables 1 and 2 (cf. Marascuilo & Serlin, 1988, p. 746). Note that α = .05, β = .30, .20, .10, .05, and various values of $e = MV^2$ were used in both tables for the degrees of freedom, v = 1(1)10, 12, 16, 20, 24. The earlier R code can be used to obtain all values in tables of Cohen (1988) as well as other sample sizes based on different specifications of α, β, and e. Note that a similar logic can be used to obtain the observed power, as shown above, when empirical data are analyzed.

As indicated in Marascuilo and Serlin (1988), of the sample size determination procedures, the most difficult to execute is the one associated with contingency tables. If a researcher can specify the value of Cramer's V^2 of interest, it is possible to determine the appropriate sample size at α = .05 by using the figures of Tables 1 and 2. The value of V^2 is not utilized directly but adjusted by the dimensions of row and column variables. Each table, hence, is entered by means of the index $e = MV^2 = w^2$, where w is the measure of effect size used in Cohen (1988). In terms of interpretation of the effect size, w = 0.10, 0.30, and 0.50 are respectively small, medium, and large effects by Cohen (1988, 1990). Note that the corresponding values of w^2 = 0.01, 0.09, and 0.25.

If a researcher wants to determine sample sizes for I = 2, J = 3, v = 2, and the small, medium, and large effects of w^2 = 0.01, 0.09, and 0.25 (i.e., $w^2 = V^2$ here) at α = .05 and β = .20, then the respective values of $e = MV^2 = (I-1)V^2 = V^2 = w^2$ yield sample sizes of 964, 108, and 39 (cf. 41 by interpolation from Table 1).

If $I = J = 2$, then $w = \phi = \sqrt{X^2/N} = \sqrt{e}.$ This means that Tables 1 and 2 can be used to obtain minimum sample sizes for fourfold tables. In addition, these tables can also be used to set up to obtain minimum sample sizes for testing H_0 : $\Delta = P_1 - P_2 = 0$, where P_1 and P_2 are population proportions. Note that G*Power can be used to obtain minimum sample sizes for fourfold tables (see Faul et al., 2009)

Discussion

Measures of effect size have been mainly presented in the context of the t test and the analysis of variance in many studies in behavioral sciences for which empirical data were analyzed with those statistical techniques. Measures of effect size are available for categorical variables, but power analysis for sample size determination have been performed for mostly fourfold tables. This paper presents a summary of measures of effect size that reflect the degree of association between two categorical variables in contingency tables. Reporting effect size is important. It is possibly more important for a researcher to design his or her study to have enough statistical power by employing an appropriate sample size based on power analysis. This paper presents tables for determining minimum sample sizes for statistical analysis of categorical data in contingency tables.

Measures of effect size and power analysis in small samples are presented in Weerehandi (2003) for parametric statistical methods and in StatXact (CYTEL Software Corporation, 2000) for nonparametric and so-called exact statistical methods. Extensions of measures of effect size and power analysis to other complicated models for categorical variables are in need for applied researchers.

As shown earlier, there are many measures of association between two variables in a contingency table. The most important one of such measures is Cramer's (1946, p. 282) V^2. It was employed as a measure of association between two categorical variables in tabulation of minimum sample sizes in this paper. Many other association measures (e.g., λ, τ, G, D, etc.) were initially developed as measures of effect size; each supposedly with some meaningful interpretation of its magnitude. These measures are not in general influences by the sample size. No clear guidelines, however, are available for these measures as the small, medium, and large effects.

Table 1 *Sample Sizes for Testing* $H_0: V = 0$ *Against* $H_1: V \neq 0$ *for* $\alpha = .05$; $\beta = .30, .20, .10, .05$; *and* $v = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 16, 20, 24$ *for Large Values of e*

		Degrees of freedom ν													
β	e	1	2	3	4	5	6	7	8	9	10	12	16	20	24
.30	.10	62	78	88	97	105	112	118	124	129	135	144	161	175	189
	.20	31	39	44	49	53	56	59	62	65	68	72	81	88	95
	.30	21	26	30	33	35	38	40	42	43	45	48	54	59	63
	.40	16	20	22	25	27	28	30	31	33	34	36	41	44	48
	.50	13	16	18	20	21	23	24	25	26	27	29	33	35	38
	.60	11	13	15	17	18	19	20	21	22	23	24	27	30	32
	.70	9	12	13	14	15	16	17	18	19	20	21	23	25	27
	.80	8	10	11	13	14	14	15	16	17	17	18	21	22	24
.20	.10	79	97	110	120	129	137	144	151	157	163	174	193	210	225
	.20	40	49	55	60	65	69	72	76	79	82	87	97	105	113
	.30	27	33	37	40	43	46	48	51	53	55	58	65	70	75
	.40	20	25	28	30	33	35	36	38	40	41	44	49	53	57
	.50	16	20	22	24	26	28	29	31	32	33	35	39	42	45
	.60	14	17	19	20	22	23	24	26	27	28	29	33	35	38
	.70	12	14	16	18	19	20	21	22	23	24	25	28	30	33
	.80	10	13	14	15	17	18	18	19	20	21	22	25	27	29
.10	.10	106	127	142	155	165	175	183	191	199	206	219	242	262	280
	.20	53	64	71	78	83	88	92	96	100	103	110	121	131	140
	.30	36	43	48	52	55	59	61	64	67	69	73	81	88	94
	.40	27	32	36	39	42	44	46	48	50	52	55	61	66	70
	.50	22	26	29	31	33	35	37	39	40	42	44	49	53	56
	.60	18	22	24	26	28	30	31	32	34	35	37	41	44	47
	.70	16	19	21	23	24	25	27	28	29	30	32	35	38	40
	.80	14	16	18	20	21	22	23	24	25	26	28	31	33	35
.05	.10	130	155	172	186	198	209	219	228	236	244	259	285	308	328
	.20	65	78	86	93	99	105	110	114	118	122	130	143	154	164
	.30	44	52	58	62	66	70	73	76	79	82	87	95	103	110
	.40	33	39	43	47	50	53	55	57	59	61	65	72	77	82
	.50	26	31	35	38	40	42	44	46	48	49	52	57	62	66
	.60	22	26	29	31	33	35	37	38	40	41	44	48	52	55
	.70	19	23	25	27	29	30	32	33	34	35	37	41	44	47
	.80	17	20	22	24	25	27	28	29	30	31	33	36	38	41

Table 2 *Sample Sizes for Testing* $H_0 : V = 0$ *Against* $H_1 : V \neq 0$ *for* $\alpha = .05$*;* $\beta = .30, .20, .10, .05$*; and* v *= 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 16, 20, 24 for Small Values of e*

		Degrees of freedom ν													
β	e	1	2	3	4	5	6	7	8	9	10	12	16	20	24
.30	.01	618	771	880	969	1046	1115	1177	1235	1290	1341	1436	1603	1750	1883
	.02	309	386	440	485	523	558	589	618	645	671	718	802	875	942
	.03	206	257	294	323	349	372	393	412	430	447	479	535	584	628
	.04	155	193	220	243	262	279	295	309	323	336	359	401	438	471
	.05	124	155	176	194	210	223	236	247	258	269	288	321	350	377
	.06	103	129	147	162	175	186	197	206	215	224	240	268	292	314
	.07	89	111	126	139	150	160	169	177	185	192	206	229	250	269
	.08	78	97	110	122	131	140	148	155	162	168	180	201	219	236
	.09	69	86	98	108	117	124	131	138	144	149	160	179	195	210
.20	.01	785	964	1091	1194	1283	1363	1436	1503	1565	1625	1734	1927	2097	2249
	.02	393	482	546	597	642	682	718	752	783	813	867	964	1049	1125
	.03	262	322	364	398	428	455	479	501	522	542	578	643	699	750
	.04	197	241	273	299	321	341	359	376	392	407	434	482	525	563
	.05	157	193	219	239	257	273	288	301	313	325	347	386	420	450
	.06	131	161	182	199	214	228	240	251	261	271	289	322	350	375
	.07	113	138	156	171	184	195	206	215	224	233	248	276	300	322
	.08	99	121	137	150	161	171	180	188	196	204	217	241	263	282
	.09	88	108	122	133	143	152	160	167	174	181	193	215	233	250
.10	.01	1051	1266	1418	1541	1647	1742	1829	1909	1983	2054	2184	2413	2614	2794
	.02	526	633	709	771	824	871	915	955	992	1027	1092	1207	1307	1397
	.03	351	422	473	514	549	581	610	637	661	685	728	805	872	932
	.04	263	317	355	386	412	436	458	478	496	514	546	604	654	699
	.05	211	254	284	309	330	349	366	382	397	411	437	483	523	559
	.06	176	211	237	257	275	291	305	319	331	343	364	403	436	466
	.07	151	181	203	221	236	249	262	273	284	294	312	345	374	400
	.08	132	159	178	193	206	218	229	239	248	257	273	302	327	350
	.09	117	141	158	172	183	194	204	213	221	229	243	269	291	311
.05	.01	1300	1545	1717	1858	1979	2086	2184	2275	2359	2439	2586	2846	3073	3277
	.02	650	773	859	929	990	1043	1092	1138	1180	1220	1293	1423	1537	1639
	.03	434	515	573	620	660	696	728	759	787	813	862	949	1025	1093
	.04	325	387	430	465	495	522	546	569	590	610	647	712	769	820
	.05	260	309	344	372	396	418	437	455	472	488	518	570	615	656
	.06	217	258	287	310	330	348	364	380	394	407	431	475	513	547
	.07	186	221	246	266	283	298	312	325	337	349	370	407	439	469
	.08	163	194	215	233	248	261	273	285	295	305	324	356	385	410
	.09	145	172	191	207	220	232	243	253	263	271	288	317	342	365

Table 3 *Cross Classification of Party Affiliation by Gender from 2018 General Social Survey*

	Party Affiliation			
Gender	Democrat	Independent	Republican	Row Total
Female	359	133	234	726
	(335.7)	(124.8)	(265.4)	
Male	257	96	253	606
	(280.3)	(104.2)	(221.6)	
Column Total	616	229	487	1332

One should be reminded that all the tests of significance, there are accompanied assumptions, hypotheses, and decision rules. Details about these matters are not presented in this paper, albeit of importance to apply the methods in data analysis. Interested readers are referred to Agresti (1984, 2007) and Stokes et al. (2000) for the statistical testing procedures.

It can be noted that measures of effect size are considered to be sample statistics when used in reporting of statistical analysis results, but population parameters in power analysis to obtain appropriate sample sizes for various test statistics. We used the same notations for these two different cases although Latin alphabet and Greek alphabet as well as carets could be employed to emphasize the difference in their meanings.

Measures of effect size used in power analysis (esp. in Cohen, 1988) may not be aligned with those used in actual reporting of statistical results in empirical studies. It may be due to in part that many statistical computer programs do not produce measures of effect size as an expression of e used in power analysis in this paper. Some manual computations by researchers are therefore required to obtain the minimum sample size based on the measure of effect size.

Relations among test statistics and measures of effect size were explicated in the paper in the context of contingency tables. It can be noted that the minimum sample size was suggested based on the statistical hypothesis testing with the null hypothesis of no association. Merely rejecting the null hypothesis, however, may not be the purpose of the investigation that use two or more categorical variables. Also note that the null hypothesis is supposedly tested with the Pearson chi-square test statistic. It may be possible to use the two sample-size tables in this paper to obtain the required sample size for situations that use the likelihood ratio test method because that test statistic can also be distributed as a chi-square distribution with the same degrees of freedom as the one for X^2.

All association measures of effect size for a contingency table that contain X^2 may not require to obtain confidence intervals, even though these are nevertheless statistics or estimates of the population effect size. Notice that other measures of association may have their own estimated variances either under the null or alternative hypothesis, mostly based on large samples. Hence, the measure may be reported with its confidence interval.

If another statistic is to be used in hypothesis testing of association for the two categorical variables, then the required sample size should be determined based on different power analysis. Tables from this study may not be directly applicable to other tests of association or measures of effect size. Although other measures of association have been supposedly proposed to enhance the understanding of the relationship between two categorical variables, many of them seem to be too situation dependent. Although the utility of other measures of association can be certainly demonstrated for different purposes based on various circumstances, power analysis and sample size determination based on V^2 might be the most useful.

In terms of the Cramer's V^2, note that $V^2 = X^2/(NM)$ (i.e., $0 < V^2 < 1$). It is the scaled mean square contingency. Note also that $V = \sqrt{X^2/(NM)} = w/\sqrt{M}$, $w = \sqrt{X^2/N} = V\sqrt{M} = \sqrt{e}$. For a fourfold table, V can be directional. Lastly, it should be noted again that measures of effect size are sample statistics. Measures of effect size, however, are treated as population parameters in power analysis.

In essence, a review of measures of effect size and the sample size tables are presented in this paper for rather simple contingency tables. Extension to more complicated data is of interest.

References

[1]. Agresti, A. (1984). *Analysis of ordinal categorical data.* New York, NY: John Wiley & Sons.

[2]. Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[3]. Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

[4]. American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications.

Educational Researcher, 35 (6), 33–40.

[5]. American Psychological Association. (2013). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

[6]. American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). Washington, DC: Author.

[7]. Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13* (6), 4–16.

[8]. Brown, M. B., & Benedetti, J. K. (1977). Sampling behavior of tests for correlation in two-way contingency tables. *Journal of the American Statistical Association, 72,*309–315.

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

[10]. Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.

[11]. Conover, W. J. (1980). *Practical nonparametric statistics* (2nd ed.). New York, NY: John Wiley & Sons.

[12]. Cramer, H. (1946). *Mathematical methods of statistics.* Princeton, NJ: Princeton University Press.

[13]. CYTEL Software Corporation. (2000). *StatXact 4 for Windows: User manual.* Cambridge, MA: Author.

[14]. Educational Testing Service. (1980). *Test use and validity.* Princeton, NJ: Author.

[15]. Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results.* Cambridge, United Kingdom: Cambridge University Press.

[16]. Everitt, B. S. (1977). *The analysis of contingency tables.* London, Great Britain: Chapman and Hall.

[17]. Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, Instruments, & Computer, 41*, 1149–1160.

[18]. Fisher, R. A. (1956). Mathematics of a lady tasting tea. In J. R. Newman (Ed.), *The world of mathematics* (Vol. 3, pp. 1512–1521). New York, NY: Simon & Schuster.

[19]. Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245–260). New York, NY: Russell Sage Foundation.

[20]. Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association, 49,* 732–764.

[21]. Goodman, L. A., & Kruskal, W. H. (1963). Measures of association for cross classifications, III: Approximate sampling theory. *Journal of the American Statistical Association, 58,* 310–364.

[22]. Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach.* Mahwah, NJ: Lawrene Erlbaum Associates.

[23]. Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace.

[24]. Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions* (Vol. 1, 2nd ed.). New York, NY: John Wiley & Sons.

[25]. Kendall, M. G. (1970). *Rank correlation methods* (4th ed.). London, England: Charles Griffin.

[26]. Kendall, M. G., & Stuart, A. (1979). *The advanced theory of statistics. Vol. 2: Inferene and relationship* (4th ed.). Bucks, England: Charles Griffin.

[27]. Lancaster, H. O. (1969). *The chi-squared distribution.* New York, NY: John Wiley & Sons.

[28]. Liu, X. S. (2014). *Statistical power analysis for the social and behavioral sciences: Basic and advanced techniques.* New York, NY: Routledge.

[29]. Marascuilo, L. A., & Serlin, R. C. (1988). *Statistical methods for the social and behavioral sciences.* New York, NY: Freeman.

[30]. McNemar, Q. (1962). *Psychological statistics* (3rd ed.). New York, NY: Wiley.

[31]. Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25,* 241–286.

[32]. Pearson, K. (1904). On the theory of contingency and its relation to association and normal correlation. *Drapers' Company Research Memoirs, Biometric Series, 1.*London: Dulau and Co.

[33]. Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research sysntesis* (pp. 231–244). New York, NY: Russell Sage Foundation.

[34]. Sakoda, J. M. (1977). Measures of association for multivariate contingency tables. *Proceedings of the Social Statistics Section of the American statistical Association* (Part III), 777-780.

[35]. Sakamoto, Y. (1985). *Categorical data analysis by AIC.* Tokyo, Japan: KTK Scientific Publishers.

[36]. SAS. (2004). *SAS/STAT 9.1: User's guide* (Vol. 2). Cary, NC: SAS Institute.

[37]. Siegel, S., & Castellan, Jr., N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York, NY: McGraw-Hill.

[38]. Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review, 27,* 799–811.

[39]. Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical data analysis using the SAS system*

(2nd ed.). Cary, NC: SAS Institute.

[40]. Theil, H. (1972). *Statistical decomposition analysis: With applications in the social and administrative sciences.* Amsterdam, The Netherlands: North-Holland.

[41]. Tschuprow, A. A. (1939). *Principles of the mathematical theory of correlation* (M. Kantorowitsch Trans.). London, Great Britain: William Hodge and Company.

[42]. Upton, G. J. G. (1978). *The analysis of cross-tabulated data.* Chichester, England: John Wiley & Sons.

[43]. Venables, W. N., Smith, D. M., & The R Development Core Team. (2009). *An introduction to R* (2nd ed.). LaVergne, TN: Network Theory.

[44]. Weerahandi, S. (2003). *Exact statistical methods for data analysis.* New York, NY: Springer.

[45]. Yule, G. U. (1900). On the association of attributes in statistics. *Philosophical Transactions of the Royal Society of London, Series A, 194,* 257–319.

[46]. Yule, G. U. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society, 75,* 579–652.

[47]. Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics* (14th ed.). New York, NY: Hafner.

Appendix A

The following measures of association were mainly developed to enhance the interpretability of the relationship between row and column variables. Both Goodman and Kruskal's λ and τ measures are for describing two nominal variables and reflect the proportional reduction of errors in classifications. Note that equations are partly based on the SAS notation (e.g., SAS, 2004, p. 1468).

Goodman and Kruskal's (1954) $\lambda(C|R)$ for predicting the column variable is

$$\lambda(C|R) = \frac{\sum_i \max_j(n_{ij}) - \max_j(n_{.j})}{n - \max_j(n_{.j})}, \tag{7}$$

and $\lambda(R|C)$ for predicting the row variable is

$$\lambda(R|A) = \frac{\sum_j \max_i(n_{ij}) - \max_i(n_{i.})}{n - \max_i(n_{i.})}. \tag{8}$$

Goodman and Kruskal's λ for a symmetric situation is

$$\lambda = \frac{\sum_i \max_j(n_{ij}) + \sum_j \max_i(n_{ij}) - \max_j(n_{.j}) - \max_i(n_{i.})}{2n - \max_j(n_{.j}) - \max_i(n_{i.})}. \tag{9}$$

The size of λ reflects the amount of predictability or accountability of the predicted variable from the knowledge of the other categorical, predictor variable. The variances for λ's are given in Goodman and Kruskal (1963) (see also SAS, 2004, pp. 1482–1483). Availability of estimated, asymptotic variances implies the possibility of null hypothesis testing for these measures.

Goodman and Kruskal's (1954, p. 759; Somers, 1962) $\tau(C|R)$ for predicting the column variable is

$$\tau(C|R) = \frac{n \sum_i \sum_j (n_{ij}^2/n_{i.}) - \sum_j n_{.j}^2}{n^2 - \sum_j n_{.j}^2}, \tag{10}$$

and $\tau(R|C)$ for predicting the row variable is

$$\tau(R|C) = \frac{n \sum_i \sum_j (n_{ij}^2/n_{.j}) - \sum_i n_{i.}^2}{n^2 - \sum_i n_{i.}^2}. \tag{11}$$

The variances for τ's are given in Goodman and Kruskal (1963).

Measures of association for the row and column variables both on the ordinal scale include Goodman and Kruskal's (1954) γ, Kendall's τ_b, Stuart's τ_c, and Somers's $D(C|R)$ and $D(R|C)$.

The estimator of γ for the ordinal variables, G, is given

$$G = \frac{P - Q}{P + Q}, \tag{12}$$

where the two times the number of concordances $P = \sum_i \sum_j n_{ij} A_{ij}$ for which $A_{ij} = \sum_{k>i} \sum_{l>j} n_{kl} + \sum_{k<i} \sum_{l<j} n_{kl}$, and the two times the number of discordances $Q = \sum_i \sum_j n_{ij} D_{ij}$ for which $D_{ij} = \sum_{k>i} \sum_{l<j} n_{kl} + \sum_{k<i} \sum_{l>j} n_{kl}$.

Kendall's (1970) τ_b (i.e., tau b) for the row and column variables on the ordinal scale is

$$\tau_b = \frac{P - Q}{w_r w_c}, \tag{13}$$

Stuart's τ_c (i.e., tau *c*; Kendall & Stuart, 1979, p. 594) for the ordinal variables is

$$\tau_c = \frac{m(P-Q)}{n^2(m-1)}, \tag{14}$$

where *m* is the minimum of the number of rows and the number of columns.
Sommer's (1962) $D(C|R)$ for the ordinal variables to predict the column variable is

$$D(C|R) = \frac{P-Q}{w_r}, \tag{15}$$

and $D(R|C)$ to predict the row variable is

$$D(R|C) = \frac{P-Q}{w_c}. \tag{16}$$

Sommer's *D* for a symmetric situation is

$$D = \frac{P-Q}{(w_r+w_c)/2}. \tag{17}$$

The variances of G, τ_b, τ_c, $D(C|R)$, $D(R|C)$, and D for the estimates and also under the null hypothesis of no association are given in SAS (2004, pp. 1477-1479) and Brown and Benedetti (1977). Ignoring the ordinal scale of the variable and treating it as nominal may result in deficiency in power for null hypothesis testing.

Appendix B

A fourfold table arises where *N* persons (i.e., participants, objects, etc.), assumed to be selected or observed at random from some population, are classified into one of the four cells. As a special case of contingency tables, there are many different ways to test certain hypotheses and many different measures of effect sizes for the fourfold table (see Fleiss, 1994). Note that equations are partly based on the SAS notation (e.g., SAS, 2004, p. 1468).

For a 2 × 2 table, Pearson chi-square statistic with one degree of freedom is

$$X^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}}. \qquad (18)$$

Association measures include

$$\phi = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1.}n_{2.}n_{.1}n_{.2}}} \qquad (19)$$

with its range from −1 to 1, Cramer's $V = \varphi$, and the contingency coefficient

$$C = \sqrt{\frac{\phi^2}{1+\phi^2}}. \qquad (20)$$

Yule's (1900, 1912) coefficient of association *Q* is defined as

$$Q = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}} = \frac{\theta - 1}{\theta + 1}, \qquad (21)$$

where $\theta = n_{11}n_{22}/(n_{12}n_{21})$ is the odds ratio (cf. $\kappa = 1/\theta$ in Yule, 1900, p. 273; Agresti, 2007, p. 29). Yule's (1912) coefficient of colligation is $Y = (\sqrt{\theta} - 1)/(\sqrt{\theta} + 1)$. The variances of *Q* and log θ under the null hypothesis of no association can be found in Yule (1900, 1912) and Agresti (2007, p. 30).

Quetelet's degree of influence function in Yule (1900, p. 282) with subscript *Q* is

$$\phi_Q = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{1.}n_{.1}}. \qquad (22)$$

If $n_{1.} = n_{2.}$ and $n_{.1} = n_{.2}$, then $\varphi_Q = \varphi$.

In addition, for two independent samples, the null hypothesis testing of equal proportion can also be done using the *Z* test based on binomial sampling, which is equivalent to the X^2 test. The proportion difference can be used as an effect size. Relative risk, odds ratio, and log odds ratio can also be used to assess effect size (see Fleiss, 1994). Fisher's (1956) exact test can be used in testing. For the correlated variables, McNemar's (1962) chi-square test statistic with one degree of freedom and subscript *M* is

$$X_M^2 = \frac{(n_{11} - n_{22})^2}{n_{11} + n_{22}}. \qquad (23)$$

Readers are referred to Agresti (2007) and Fleiss (1994) for these specific testing procedures and association measures for 2 × 2 tables.