Statistical Power Analysis for One-way ANOVA with Binary or Count Data

Yujiao Mai and Zhiyong Zhang

Abstract Analysis of variance (ANOVA) is a statistical method to compare means of three or more groups. It has been frequently used in experimental studies. Typically, ANOVA is used for continuous data, but discrete data are also common in practice. When the outcomes are binary or count data, the assumptions of normality and equal variances are violated. Thus the conventional ANOVA is no longer reliable for binary or count data, and statistical power analysis based on conventional ANOVA is incorrect. To address the issue, this study focuses on one-way ANOVA with binary or count data and provides a solution for statistical power analysis. We first introduce a likelihood ratio test statistic through variation decomposition for one-way ANOVA with binary or count data. With the new test statistic, we then define the effect size and propose the method to calculate statistical power. Finally, we develop software to conduct the proposed power analysis.

1 Introduction

Analysis of variance (ANOVA) is a statistical method to compare means of three or more groups. It has been frequently used in experimental studies. Typically, it is used for continuous data and produces an F-statistic as the ratio of the between-group variance to the within-group variance that follows a F-distribution.

To use the F test for ANOVA, three assumptions must be met. The first is independence of observations assuming all samples are drawn independently of each other group. The second is normality assuming the distribution of the residuals is normal. The third is equality of variances. It assumes that the variance of the data in all groups should be the same. In practice, even continuous data cannot always meet

Yujiao Mai
University of Notre Dame, Notre Dame, IN 46556, USA, e-mail: `ymai@nd.edu`

Zhiyong Zhang
University of Notre Dame, Notre Dame, IN 46556, USA, e-mail: `zzhang4@nd.edu`

all the three assumptions. For binary or count data, the two assumptions, normality and equality of variances, are clearly violated. Therefore, it is incorrect to use the F test for continuous data for ANOVA with binary or count data.

Binary and count data are very common in practice. Researchers have used conditional distribution approximation, empirical logistic transform, and logistic regression models to handle binary data (Cox & Snell, 1989). For instance, k $\times$ 2 contingency table can be used to test the mean (proportion) difference among groups of binary data (Collett, 1991; Cox & Snell, 1989). In this case the significance of the difference between two proportions can be assessed with a variety of statistical tests such as Pearson's chi-squared test (Pearson, 1947), Fisher's exact test (Pearson & Hartley, 1966), and Barnard's test. The chi-squared test is unreliable with small sample size or unequal distribution among the cells of the table, since it is based on approximated but not exact distribution (Larntz, 1978; Mehta, Patel, & Tsiatis, 1984). On the contrary, the Fisher's exact test gives the exact P-value based on an hypergeometric distribution, but some researchers argued that it is conservative with low statistical power (Berkson, 1978; D'agostino, Chase, & Belanger, 1988), since it is a discrete statistic (Yates, 1984). Barnard's test is an alternative exact test (Berger, 1994). For 2 $\times$ 2 tables, it requires less computation time and has higher power, but for larger tables, the computation time increases and the power advantage quickly decreases (Mehta & Hilton, 1993). Regardless of the performance of the test statistics, contingency table works only for one grouping variable.

Researchers have also used logistic regression to estimate and compare the group means of binary data (Collett, 1991; Cox & Snell, 1989). This method utilizes the likelihood ratio test, which performs well when there are enough observations to justify the assumptions of the asymptotic chi-squared tests. However, the models and procedures might be more complicated than necessary. First, the procedure using logistic regression requires creating dummy variables. These dummy variables not only increase the complexity of the model itself but also make the interpretation of the model more difficult for applied researchers. Second, the procedure using logistic regression is more complex with the current software. Third, researchers are interested in whether the groups are from populations with different means using ANOVA, while logistic regression is more efficient for parameter estimation (Cox & Snell, 1989) and proportions prediction (Collett, 1991). The meaning of parameters in logistic regression is not easy to interpret. Thus, "never use a fancy method when a simple method will do" (Miller Jr, 1986).

This study focuses on one-way ANOVA, and aims at providing an intuitive and efficient solution for one-way ANOVA with binary or count data. First we will review one-way ANOVA with continuous data. Second, we will provide a solution for one-way ANOVA with binary data. It will include a proper test statistic, effect size definition, and power analysis formula. Then, we will discuss one-way ANOVA with count data. With all the statistical methods, we will illustrate the usage of the new developed software for one-way ANOVA with binary or count data. Finally, we will summarize the contributions of this study.

2 One way ANOVA with continuous data

Analysis of variance (ANOVA) is a collection of statistical models used to analyze the differences among group means and their associated procedures developed by Ronald Fisher (Maxwell & Delaney, 2004). Typically, ANOVA is used to test for differences among at least three groups. In its procedures, the observed variance in a particular variable is partitioned into components attributable to different sources of variation. For one-way ANOVA, the observed variance in the outcome variable is the total variance, which is divided into between-group variance and within-group variance. If the between-group variance is larger than the within-group variance, the group means are considered to be different.

Let Y be the outcome variable, and A be a categorical variable of k levels, with A as the factor we get the data of Y divided into k groups. The null hypothesis H_0 states that samples in different groups are drawn from populations with equal means, while the alternative hypothesis H_1 supposes that at least two groups are from populations with different means. Let μ_j be the population mean of the jth group, $j = 1, 2, \cdots, k$, and μ_0 be the grand population mean. The null and alternative hypotheses can be denoted as follows:
$H_0: \quad \mu_1 = \mu_2 = \ldots = \mu_k = \mu_0,$
$H_1: \quad \exists\ \mu_g \neq \mu_j, \quad where \quad g \neq j \quad and \quad g,\ j \in [1, 2, \cdots, k].$
Consider the corresponding models with H_0 and H_1. The null model M_0 is

$$\mathrm{E}\{Y|A = j\} = \mu_0 + \varepsilon, \tag{1}$$

where $Y|(A = j) \sim N(\mu_0, \sigma_0^2)$. The alternative model M_1 is

$$\mathrm{E}\{Y|A = j\} = \mu_j + \varepsilon_j, \tag{2}$$

where $Y|(A = j) \sim N(\mu_j, \sigma_1^2)$.

Given a sample of data $\mathbf{y}_j = \{y_{ij}\}$, $i = 1, 2, \cdots, n_j, j = 1, 2, \cdots, k$, with n_j denoting the sample size of the jth group, the test statistic is equal to the ratio of between-group variance and within-group variance and follows a F distribution:

$$F = \frac{\sigma^2_{between}}{\sigma^2_{within}} \sim F(k-1, n-k), \tag{3}$$

where $\sigma^2_{between} = \sum_{j=1}^{k} (\bar{y}_j - \bar{y})^2/(k-1)$, and $\sigma^2_{within} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2/(n-k)$ with $\bar{y}_j$ denoting the sample mean of the jth group and $\bar{y}$ denoting the grand mean of data. ANOVA is often conducted by constructing the source of variance table shown in Table 1.

Table 1: ANOVA table for continuous data

Source	Sum of Variation	Degree of Freedom	Test Statistic	P-value
Between-group	$SS_B = \sum_{j=1}^{k}(\bar{y}_j - \bar{y})^2$	$k-1$	$F = \frac{SS_B/(k-1)}{SS_W/(n-k)}$	$1 - \mathrm{F}(F; k-1, n-k)$
Within-group	$SS_W = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$	$n-k$		
Total	$SS_T = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$	$n-1$		

Note. $\mathrm{F}(x; k-1, n-k)$ is the cumulative distribution function of F distribution.

3 One-way ANOVA with binary data

3.1 Model and test statistic for binary data

For one-way ANOVA with binary data, the hypotheses are the same as those for continuous data. But the models are different since the distribution of the outcome variable is not normal. Thus, the F test statistic for conventional ANOVA may no longer be proper for binary data.

Let Y be the zero-one outcome, and A be a categorical variable of k levels, with A as the factor we get the data of Y divided into k groups. Let μ_0 denote the grand probability of the outcome 1, and μ_j denote the jth group probability of observing 1, $j = 1, 2, \cdots, k$. Then the null and alternative hypotheses are

$$
\begin{aligned}
H_0: &\quad \mu_1 = \mu_2 = \ldots = \mu_k = \mu_0, \\
H_1: &\quad \exists\ \mu_g \neq \mu_j, \quad where \quad g \neq j \quad and \quad g,\ j \in [1, 2, \cdots, k].
\end{aligned}
$$

The null model M_0 is

$$\mathrm{E}\{Y|A = j\} = \mu_0, \tag{4}$$

where $Y|(A = j) \sim Bernoulli(\mu_0)$, and the alternative model M_1 is

$$\mathrm{E}\{Y|A = j\} = \mu_j, \tag{5}$$

where $Y|(A = j) \sim Bernoulli(\mu_j)$. Given a sample of data $\mathbf{y}_j = \{y_{ij}\}, i = 1, 2, \cdots, n_j$, $j = 1, 2, \cdots, k$, with n_j denoting the sample size of the jth group, we define the likelihood ratio test statistic

$$
\begin{aligned}
D &= -2 \ln \frac{\mathscr{L}(\mu_0|\mathbf{Y})}{\mathscr{L}(\mu_1, \mu_2, \ldots, \mu_k|\mathbf{Y})} \\
&= -2[\ell(\mu_0|\mathbf{Y}) - \ell(\mu_1, \mu_2, \ldots, \mu_k|\mathbf{Y})] \\
&= -2(\ell_{M_0} - \ell_{M_1}).
\end{aligned} \tag{6}
$$

This statistic follows a (approximated) chi-square distribution $D \tilde{\sim} \chi^2(df, \lambda)$ with the degrees of freedom $df = k - 1$ (Wilks, 1938). Under the null hypothesis, $\lambda = 0$ for a central chi-squared distribution. Under the alternative hypothesis, the test statistic follows a non-central chis-squared distribution with the non-centrality parameter $\lambda > 0$.

Let the observed grand mean $\bar{y} = \sum_j^k \sum_i^{n_j} y_{ij}/n$ be the estimate of μ_0, and the observed group mean $\bar{y}_j = \sum_i^{n_j} y_{ij}/n_j$ be the estimate of μ_j. For the given sample of data, we can calculate the test statistic as $\tilde{D}$:

$$\begin{aligned} -2\hat{\ell}_{M_0} &= -2\ell(\bar{y}|\mathbf{Y}) \\ &= -2\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left[y_{ij}\ln\bar{y} + (1-y_{ij})\ln(1-\bar{y})\right], \end{aligned} \tag{7}$$

$$\begin{aligned} -2\hat{\ell}_{M_1} &= -2\ell(\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k|\mathbf{Y}) \\ &= -2\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left[y_{ij}\ln\bar{y}_j + (1-y_{ij})\ln(1-\bar{y}_j)\right], \end{aligned} \tag{8}$$

$$\begin{aligned} \tilde{D} &= -2(\hat{\ell}_{M_0} - \hat{\ell}_{M_1}) \\ &= -2\sum_{j=1}^{k} n_j \left\{\bar{y}_j(\ln\bar{y} - \ln\bar{y}_j) + (1-\bar{y}_j)\left[\ln(1-\bar{y}) - \ln(1-\bar{y}_j)\right]\right\}. \end{aligned} \tag{9}$$

It can be proven that the observed grand mean $\bar{y}$ is the maximum likelihood estimate of μ_0, and the observed group mean $\bar{y}_j$ is the estimate of μ_j (Efron, 1978). Now with these statistics, we can create an ANOVA table (see Table 2) for binary data similar to that of continuous data.

Table 2: ANOVA table for binary data

Source	Sum of Variation	Degree of Freedom	Test Statistic	P-value
Between-group	$SS_B = -2(\hat{\ell}_{M_0} - \hat{\ell}_{M_1})$	$k-1$	$\tilde{D} = -2(\hat{\ell}_{M_0} - \hat{\ell}_{M_1})$	$1-\chi^2(\tilde{D}; k-1)$
Within-group	$SS_W = -2\hat{\ell}_{M_1}$	$n-k$		
Total	$SS_T = -2\hat{\ell}_{M_0}$	$n-1$		

Note. $\chi^2(x; k-1)$ is the cumulative distribution function of $\chi^2(k-1)$.

3.2 *Measure of effect size for binary data*

Standardized effect-size measures facilitate comparison of findings across studies and disciplines, while unstandardized effect size measures (simple effect size) with ”immediate meanings” may be preferable for reporting purposes (Baguley, 2009; Ellis, 2010). The r-family and the d-family effect size measures are standardized (Rosenthal, 1994), while R^2-family effect-size measures such as f^2 and η^2 are unstandardized and immediately meaningful (Cameron & Windmeijer, 1997). Both types of effect size measures could be defined. But not all types of effect-size measures can be used for power analysis with a specific test statistic. For the purpose of power analysis, in this study we chose a standardized effect-size measure like Cramer's V, it is a member of the r family (Ellis, 2010). It is also an adjusted version of phi coefficient ϕ which is frequently reported as the measure of effect size for a chi-square test (Cohen, 1988; Ellis, 2010; Fleiss, 1994). It can be viewed as the association between two variables as a percentage of their maximum possible variation. In the case of one-way ANOVA, the two variables are the outcome variable and the grouping variable.

For one-way ANOVA with binary data, we define the effect size V:

$$V = \sqrt{-2\sum_{j=1}^{k} w_j \left\{\mu_j(\ln\mu_0 - \ln\mu_j) + (1-\mu_j)\left[\ln(1-\mu_0) - \ln(1-\mu_j)\right]\right\}/(k-1)}, \tag{10}$$

where w_j is the weight of the jth group. The small, medium, and large effect size can be defined as 0.10, 0.30, and 0.50, borrowed from Cohen's effect size benchmarks (Cohen, 1988; Ellis, 2010).

For a given sample of data, the sample effect size can be calculated as

$$\begin{aligned}\hat{V} &= \sqrt{\tilde{D}/n(k-1)} \\ &= \sqrt{-2\sum_{j=1}^{k} w_j \left\{\bar{y}_j(\ln\bar{y} - \ln\bar{y}_j) + (1-\bar{y}_j)\left[\ln(1-\bar{y}) - \ln(1-\bar{y}_j)\right]\right\}/(k-1)}.\end{aligned} \tag{11}$$

3.3 Statistical power analysis with binary data

Power analysis is often applied in the context of ANOVA in order to assess the probability of successfully rejecting the null hypothesis if we assume a certain ANOVA design, effect size in the population, sample size and significance level. Power analysis can assist in study design by determining what sample size would be required in order to have a reasonable chance of rejecting the null hypothesis when the alternative hypothesis is true (Strickland, 2014).

For one-way ANOVA with binary data, when the null hypothesis H_0 is true, the test statistic D follows a central chi-squared distribution $\chi^2(df)$, where $df = k-1$ is the degree of freedom. If $\tilde{D}$ is larger than the critical value $C = \chi^2_{1-\alpha}(df)$, one would reject the null hypothesis H_0. When the alternative hypothesis H_1 is true, the test statistic D follows a non-central chi-squared distribution $\chi^2(df, \lambda)$, where $df = k-1$ is the degree of freedom, and $\lambda = D = n(k-1)V^2$ is the non-central parameter. Let $F_{df,\lambda}(x)$ be the cumulative distribution function of the non-central chi-square distribution, then the formula of the statistical power of the test is

$$\begin{aligned} power &= Pr\{\chi^2 \geq C|H_1\} \\ &= 1-\chi^2_{df,\lambda}(C) \\ &= 1-\chi^2_{k-1,n(k-1)V^2}\left[\chi^2_{1-\alpha}(k-1)\right]. \end{aligned} \tag{12}$$

With this formula, the power, minimum detectable effect size V, minimum required sample size n, or significance level α can be calculated given the other parameters.

4 One-way ANOVA with Count Data

For one-way ANOVA with count data, its inference is similar to that for binary data. The main different is that the distribution of the outcome variable in the model is Poisson instead of Bernoulli.

4.1 Model and test statistic for count data

For one-way ANOVA with count data, let Y be the outcome variable, which can take only the non-negative integer values, and A be a categorical variable of k levels. The null and alternative hypotheses are

$$\begin{aligned} H_0: &\quad \mu_1 = \mu_2 = \ldots = \mu_k = \mu_0, \\ H_1: &\quad \exists\ \mu_g \neq \mu_j, \quad where \quad g \neq j \quad and \quad g,\ j \in [1,2,\cdots,k]. \end{aligned}$$

The null model M_0 is

$$\mathrm{E}\{Y|A=j\} = \mu_0, \tag{13}$$

where $Y|(A=j) \sim Poisson(\mu_0)$, and the alternative model M_1 can be

$$\mathrm{E}\{Y|A=j\} = \mu_j, \tag{14}$$

where $Y|(A=j) \sim Poisson(\mu_j)$, $j=1,2,\cdots,k$. Given a sample of data, $\mathbf{y}_j = \{y_{ij}\}$, $i=1,2,\cdots,n_j$, $j=1,2,\cdots,k$, with n_j denoting the sample size of the jth group, the likelihood ratio test statistic is

$$\begin{aligned} D &= -2\ln\frac{\mathscr{L}(\mu_0|\mathbf{Y})}{\mathscr{L}(\mu_1,\mu_2,\ldots,\mu_k|\mathbf{Y})} \\ &= -2[\ell(\mu_0|\mathbf{Y}) - \ell(\mu_1,\mu_2,\ldots,\mu_k|\mathbf{Y})] \\ &= -2(\ell_{M_0} - \ell_{M_1}) \end{aligned} \tag{15}$$

This statistic follows a (approximated) chi-square distribution $D\dot{\sim}\chi^2(df,\lambda)$ with the degrees of freedom $df = k-1$ (Wilks, 1938). Under the null hypothesis, $\lambda = 0$ for a central chi-squared distribution. Under the alternative hypothesis, the test statistic follows a non-central chis-squared distribution with the non-centrality parameter $\lambda > 0$. Let the grant mean $\bar{y} = \sum_j^k \sum_i^{n_j} y_{ij}/n$ be the estimate of μ_0, and the group mean $\bar{y}_j = \sum_i^{n_j} y_{ij}/n_j$ be the estimate of μ_j. For a given sample of data $\mathbf{Y}$, we can calculate the test statistic as $\tilde{D}$:

$$SS_T = -2\hat{\ell}_{M_0} = -2\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\bar{y}_{ij}\ln\bar{y} - \bar{y}\right), \tag{16}$$

$$SS_W = -2\hat{\ell}_{M_1} = -2\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\bar{y}_{ij}\ln\bar{y}_j - \bar{y}_j\right), \tag{17}$$

$$SS_B = \tilde{D} = -2(\hat{\ell}_{M_0} - \hat{\ell}_{M_1}) = -2\sum_{j=1}^{k} n_j\left[\bar{y}_j(\ln\bar{y} - \ln\bar{y}_j) - (\bar{y} - \bar{y}_j)\right]. \tag{18}$$

For count data, we can also create an ANOVA table like Table 2 for binary data.

4.2 *Effect size and power analysis for count data*

For one-way ANOVA with count data, the effect size is also defined as $V = \sqrt{D/n(k-1)}$. The sample effect size can be calculated as

$$\begin{aligned}\hat{V} &= \sqrt{\tilde{D}/n(k-1)} \\ &= \sqrt{-2\sum_{j=1}^{k} w_j\left[\bar{y}_j(\ln\bar{y} - \ln\bar{y}_j) + (\bar{y}_j - \bar{y})\right]/(k-1)},\end{aligned} \tag{19}$$

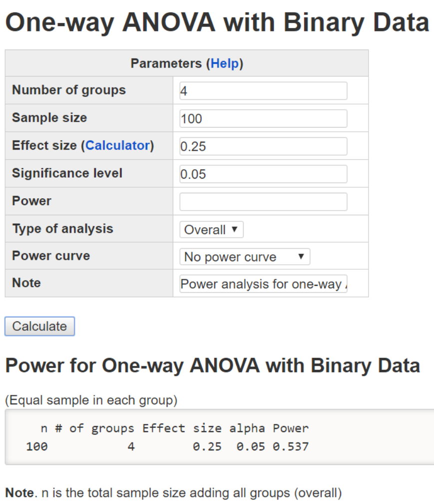
where $w_j = n_j/n$ is the weight of the jth group, and $n = \sum_j^k n_j$ is the total sample size. The power analysis of one-way ANOVA with count data is the same as that with binary data.

5 Software

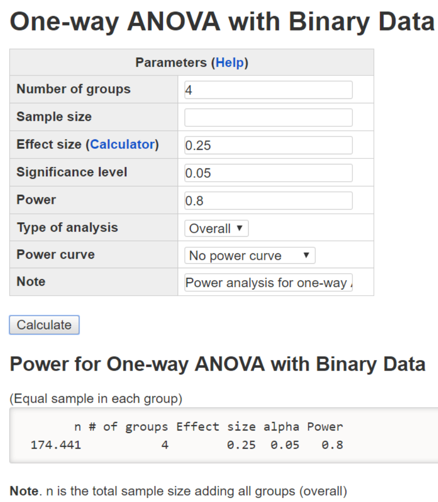
To carry out the power analysis for ANOVA with binary and count data, we have developed online applications that can be used within a Web browser. The link for the binary ANOVA is `http://w.psychstat.org/anovabinary` and for the count ANOVA is `http://w.psychstat.org/anovabinary`.

Figure 1 shows the interface and two examples of the power analysis for ANOVA with binary data. Among *Number of groups*, *Sample size*, *Effect size*, *Significance level*, and *Power*, any of them can be calculated given the rest of the information. For example, Figure 1(a) shows how to calculate power given 4 groups, sample size

100, effect size 0.25 and significance level 0.05. The output indicates the power for this design is 0.537. Figure 1(b) shows how to calculate sample size given 4 groups, effect size 0.25, significance level 0.05 and the desired power 0.8. From the output, a sample size 175, the near integer of 174.44, is needed. A power curve can also be plotted by providing multiple sample sizes in the *Sample size* field. The interface for ANOVA with count data is the same.



(a) Given sample size, calculate power



(b) Given power, calculate sample size

Fig. 1: Examples on power analysis for ANOVA with binary data

6 Discussion

In this chapter, we introduced the closed-form likelihood ratio test to test the overall mean differences among groups for one-way ANOVA with binary or count data. This provides a similar way for ANOVA with continuous data by creating ANOVA tables for binary or count data. This not only illustrated the connections between the new methods and the conventional ANOVA, but also can help the researchers intuitively and easily understand the methods. The effect size for one-way ANOVA with binary or count data is defined by the V statistic, an adjusted phi coefficient. With the group means and weights, the sample effect size can be calculated. The power analysis involved four parameters, the number of groups, the total sample size, the statistical significance level, and the effect size. We provided free online software based on the methods proposed. Future studies can investigate how to conduct power analysis for multiple comparison and extend the methods to two-way ANOVA.

Acknowledgements This research is supported by a grant from the Department of Education (R305D140037) awarded to Zhiyong Zhang and Ke-Hai Yuan. However, the contents of the paper do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

References

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, *100*(3), 603–617.

Berger, R. (1994). Power comparison of exact unconditional tests for comparing two binomial proportions. *Institute of Statistics Mimeo Series*(2266), 1–19.

Berkson, J. (1978). In dispraise of the exact test: Do the marginal totals of the 2x2 table contain relevant information respecting the table proportions? *Journal of Statistical Planning and Inference*, *2*(1), 27–42.

Cameron, A. C., & Windmeijer, F. A. (1997). An r-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, *77*(2), 329–342.

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

Collett, D. (1991). *Modelling binary data*. New York, NY: Chapman & Hall.

Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd ed.). New York, NY: Chapman & Hall.

D'agostino, R. B., Chase, W., & Belanger, A. (1988). The appropriateness of some common procedures for testing the equality of two independent binomial populations. *The American Statistician*, *42*(3), 198–202.

Efron, B. (1978). Regression and anova with zero-one data: Measures of residual variation. *Journal of the American Statistical Association*, *73*(361), 113–121.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. New York, NY: Cambridge University Press.

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.

Larntz, K. (1978). Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. *Journal of the American Statistical Association*, *73*(362), 253–263.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.

Mehta, C. R., & Hilton, J. F. (1993). Exact power of conditional and unconditional tests: going beyond the 2×2 contingency table. *The American Statistician*, *47*(2), 91–98.

Mehta, C. R., Patel, N. R., & Tsiatis, A. A. (1984). Exact significance testing to establish treatment equivalence with ordered categorical data. *Biometrics*, *40*(3), 819–825.

Miller Jr, R. G. (1986). *Beyond anova: basics of applied statistics*. New York, NY: John Wiley & Sons, Inc.

Pearson, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table. *Biometrika*, *34*(1/2), 139–167.

Pearson, E. S., & Hartley, H. O. (1966). *Biometrika tables for statisticians* (3rd ed.). London, UK: Biometrika Trust.

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

Strickland, J. S. (2014). *Predictive analytics using r*. Lulu.com.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, *9*(1), 60–62.

Yates, F. (1984). Tests of significance for 2×2 contingency tables. *Journal of the Royal Statistical Society. Series A (General)*, *147*(3), 426–463.