Statistical Power Analysis for Comparing Means with Binary or Count Data Based on Analogous ANOVA

Yujiao Mai and Zhiyong Zhang

Abstract Comparison of population means is essential in quantitative research. For comparing means of three or more groups, analysis of variance (ANOVA) is the most frequently used statistical approach. Typically, ANOVA is used for continuous data, but discrete data are also common in practice. To compare means of binary or count data, the classical ANOVA and the corresponding power analysis are problematic, because the assumption of normality is violated. To address the issue, this study introduces an analogous ANOVA approach for binary or count data, as well as the corresponding methods for statistical power analysis. We first introduce an analogous ANOVA table and a likelihood ratio test statistic for comparing means with binary or count data. With the test statistic, we then define an effect size and propose a method to calculate statistical power. Finally, we develop and show software to conduct the proposed power analysis for both binary and count data.

Keywords Statistical power • Analogous ANOVA • Binary data • Count data

1 Introduction

Comparison of population means is one of the essential statistical analyses in quantitative research (Moore et al. 2013). For comparing means of three or more groups, analysis of variance (ANOVA) is the most frequently used statistical approach in psychological research (Howell 2012). 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 an F-distribution. To use the F-test for ANOVA, three assumptions must be met. The first is the independence of observations, which assumes that all samples are drawn independently of each other. The second is the normality assumption that requires the distribution of the residuals to be normal. The third is the equality of variances, which assumes that the variance of the data in all groups should be the same. In practice, studies with even continuous data cannot

Y. Mai (✉) • Z. Zhang (✉)
University of Notre Dame, 118 Haggar Hall, Notre Dame, IN 46556, USA
e-mail: ymai@nd.edu; zzhang4@nd.edu

L.A. van der Ark et al. (eds.), *Quantitative Psychology*, Springer Proceedings in Mathematics & Statistics 196, DOI 10.1007/978-3-319-56294-0_33

always meet all three assumptions. For binary or count data, the assumption of normality is apparently violated. Therefore, it is unreliable to use classical ANOVA to compare means of binary or count data. Furthermore, the corresponding power analysis is expected to be problematic.

Since discrete data are very common in practice, there have been discussions on the statistical methods for mean comparison with binary or count data. The existing approaches include $k \times 2$ contingency tables and logistic regression to analyze the mean (proportion) difference among groups of binary data (Cox and Snell 1989; Collett 1991). Contingency tables are commonly used along with Pearson's chi-squared test (Pearson 1947; Larntz 1978), likelihood ratio test (Birch 1963; Grove 1984; Williams 1976), Freeman-Tukey chi-squared statistic (Bishop et al. 1975; Freeman and Tukey 1950), and Fisher's exact test (Fisher 1922; Agresti 1992). Pearson's chi-squared test is related to Goodman and Kruskal's τ (Goodman and Kruskal 1954; Efron 1978). This test is less accurate with small sample size (less than 10 for each cell) and is unreliable if more than 20% of cells have expected values less than 5 (Yates et al. 1999). For likelihood ratio test and Freeman-Tukey chi-squared test, simulation studies found that the Type I error rates became very high when the sample size was small and there were cells with small observed means and moderate expected values (Larntz 1978). Fisher's exact test is related to Goodman and Kruskal's λ (Turek and Suich 1989; Efron 1978). This test is more accurate than the chi-squared tests with small sample size, but it becomes difficult to calculate with large samples or unbalanced tables (Mehta et al. 1984). Although none of these tests is perfect, in general, the likelihood ratio test is preferred by many statisticians (Larntz 1978; Collett 1991), because it is based on the exact Bernoulli distribution for binary data, and researches (Hoeffding 1965; Bahadur 1967) suggested that it has some asymptotically optimal properties.

Researchers have also used logistic regression to estimate and compare the group means of binary data (Cox and Snell 1989; Collett 1991). 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 requires creating dummy variables since regression models are used with categorical predictors. 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 and Snell 1989) and prediction of proportions (Collett 1991). The meaning of parameters in logistic regression is not easy to interpret for the purpose of mean comparison.

Although contingency tables and logistic regression are two different approaches, it is not difficult to show that contingency table and logistic regression lead to the same conclusions when using likelihood ratio tests. Then, is it possible to provide the equivalent results for binary data by applying likelihood ratio test to ANOVA? In fact, as suggested by Efron (1978), log-likelihood can be used as a general measure of variation. From the perspective of variation decomposition, Efron (1978) constructed an ANOVA-like table for binary data with emphasis on descriptive statistics. Based on the work by Efron (1978), we will introduce an analogous ANOVA table with a closed-form likelihood ratio test statistic for comparing means with binary data. Then we will define an effect size and provide a corresponding power analysis method. Software to conduct the power analysis will also be developed. After that, we will extend the method for binary data to count data.

The rest of the chapter is organized as follows. Section 2 is a review of one-way ANOVA with continuous data. Section 3 proposes the method for binary data. Section 4 discusses the method for count data. Section 5 illustrates the developed software through examples. Section 6 summarizes and concludes 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 through variance decomposition (Maxwell and Delaney 2004; Fisher 1921). The current study focuses on the use of one-way ANOVA. We first review the basics of one-way ANOVA with continuous data. Let Y be the outcome variable and A be a categorical variable of k levels; with A as the grouping variable, we divide the population of Y into k groups. The null hypothesis H_0 states that different groups have equal population means, while the alternative hypothesis H_1 supposes that at least two groups have different population 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 specified as follows:

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

Consider the corresponding models with H_0 and H_1. The null model M_0 is

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

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

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

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

In one-way ANOVA, the observed variance in the outcome variable is partitioned into between-group variance and within-group variance. If the between-group variance is greater than the within-group variance, the group means are considered to be different. For continuous data, "squared error" is deployed as a measure of variation between an observed data point and corresponding expectation ("explanatory point", see Efron 1978). Its function is defined as

$$S(y, \mu) = (y - \mu)^2 \tag{3}$$

Table 1 ANOVA table for continuous data

Source	Sum of squares	Degree of freedom	Test statistic	P-value
Between-group	$SS_B = \sum_{j=1}^{k} (\bar{y}_j - \bar{y})^2$	$k - 1$	$F = \frac{SS_B/(k-1)}{SS_W/(n-k)}$	$\Pr\{\mathrm{F}(k-1, n-k) \geq F\}$
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: $\mathrm{F}(k-1, n-k)$ is the F-distribution with $df_1 = k - 1$ and $df_2 = n - k$

with y denoting a data point and μ denoting the expectation. Given a sample of data $\mathbf{Y} = (\mathbf{y}_j) = \{y_{ij}\}$, $i = 1, 2, \cdots, n_j, j = 1, 2, \cdots, k$, with n_j denoting the sample size of the jth group, the test statistic is equal to the ratio of between-group sample variance and within-group sample variance and follows an F-distribution under H_0:

$$F = \hat{\sigma}^2_{between} / \hat{\sigma}^2_{within} \sim \mathrm{F}(k-1, n-k), \tag{4}$$

where $\hat{\sigma}^2_{between} = \sum_{j=1}^{k} (\bar{y}_j - \bar{y})^2 / (k-1)$, and $\hat{\sigma}^2_{within} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 / (n-k)$ 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.

3 One-Way Analogous ANOVA with Binary Data

3.1 Model and Test Statistic for Binary Data

For comparison of group means, often called proportions, for binary data, the hypotheses are the same as one-way ANOVA with continuous data. But the models are different since the distribution of the outcome variable is not normal.

Let Y be a zero-one outcome variable and A be a categorical variable of k levels, with A as the grouping variable we can divide the population of Y 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

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

The null model M_0 is

$$\mathrm{E}\{Y | A = j\} = \mu_0, \tag{5}$$

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

$$\mathrm{E}\{Y|A=j\} = \mu_j, \tag{6}$$

where $Y|(A = j) \sim Bernoulli(\mu_j)$. Given a sample of data $\mathbf{Y} = (\mathbf{y}_j) = \{y_{ij}\}$, $i = 1, 2, \cdots, n_j, j = 1, 2, \cdots, k$, with n_j denoting the sample size of the jth group, we define minus twice the log-likelihood ratio of M_0 to M_1 as a statistic:

$$\begin{aligned} D &= -2\ln\frac{\mathscr{L}(\mu_0|\mathbf{Y})}{\mathscr{L}(\mu_1,\mu_2,\ldots,\mu_k|\mathbf{Y})} \\ &= -2[\ell(\mu_0|\mathbf{Y}) - \ell(\mu_1,\mu_2,\ldots,\mu_k|\mathbf{Y})] \\ &= -2(\ell_{M_0} - \ell_{M_1}), \end{aligned} \tag{7}$$

where $\mathscr{L}(\boldsymbol{\theta}|\mathbf{Y})$ denotes the likelihood function of $\boldsymbol{\theta}$ given data $\mathbf{Y}$. Under the null hypothesis H_0, this statistic follows a chi-squared distribution $D{\sim}\chi^2(df)$ with the degrees of freedom $df = k - 1$ if the sample size tends to infinity (Wilks 1938).

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 a given sample of data, we calculate the test statistic as $\tilde{D}$ as follows. We first calculate minus twice the log-likelihood for M_0 and M_1:

$$\begin{aligned} -2\hat{\ell}_{M_0} &= -2\ell(\bar{y}|\mathbf{Y}) \\ &= -2\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left[y_{ij}\ln\bar{y} + (1-y_{ij})\ln(1-\bar{y})\right], \end{aligned} \tag{8}$$

$$\begin{aligned} -2\hat{\ell}_{M_1} &= -2\ell(\bar{y}_1,\bar{y}_2,\ldots,\bar{y}_k|\mathbf{Y}) \\ &= -2\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left[y_{ij}\ln\bar{y}_j + (1-y_{ij})\ln(1-\bar{y}_j)\right]; \end{aligned} \tag{9}$$

and then $\tilde{D}$ as their difference:

$$\begin{aligned} \tilde{D} &= -2(\hat{\ell}_{M_0} - \hat{\ell}_{M_1}) \\ &= -2\sum_{j=1}^{k} n_j\left\{\bar{y}_j(\ln\bar{y} - \ln\bar{y}_j) + (1-\bar{y}_j)\left[\ln(1-\bar{y}) - \ln(1-\bar{y}_j)\right]\right\}. \end{aligned} \tag{10}$$

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). Other than the "squared error" used by standard analysis of variance, if we use minus twice the

Table 2 Analogous ANOVA table for binary data

Source	Sum of variance	Degree of freedom	Test statistic	*P*-value
Between-group	$SS_B = -2(\hat{\ell}_{M_0} - \hat{\ell}_{M_1})$	$k-1$	$\tilde{D} = -2(\hat{\ell}_{M_0} - \hat{\ell}_{M_1})$	$\Pr\{\chi^2(k-1) \geq \tilde{D}\}$
Within-group	$SS_W = -2\hat{\ell}_{M_1}$	$n-k$		
Total	$SS_T = -2\hat{\ell}_{M_0}$	$n-1$		

Note: $\chi^2(k-1)$ is the chi-squared distribution with $df = (k-1)$

log-likelihood as a measure of variation, the variation function for binary data is as follows:

$$S_1(y, \mu) = \begin{cases} -2\ln(\mu) & \text{if} \quad y = 1 \\ -2\ln(1-\mu) & \text{if} \quad y = 0 \end{cases} \tag{11}$$

with y denoting a data point and μ the expectation. Then the sum of variation $SS_1 = \sum S_1(\mathbf{Y}, \mu) = -2\ln\mathscr{L}(\mu|\mathbf{Y})$ with $\mathbf{Y}$ denoting a sample of data (Efron 1978). With these functions, we can obtain the analogous total variance, within-group variance, and between-group variance as follows:

$$\begin{aligned} SS_T &= -2\hat{\ell}_{M_0} \\ SS_W &= -2\hat{\ell}_{M_1} \\ SS_B &= -2(\hat{\ell}_{M_0} - \hat{\ell}_{M_1}). \end{aligned} \tag{12}$$

Now with these statistics, we can create an analogous ANOVA table in Table 2 for binary data similar to that for continuous data.

From the analogous ANOVA table, we see that the likelihood ratio test statistic here equals the between-group variation. The ratio of between-group variation to total variation is exactly the R^2 coefficient for model M_1 (see Efron 1978), which is also used by Goodman (1971) for contingency tables.

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 (Ellis 2010; Baguley 2009). 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 and 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 use a standardized effect-size measure like Cramer's V, which is a member of the r family (Ellis 2010). It is also an adjusted version of phi coefficient ϕ that is frequently reported as the measure of effect size for a chi-squared 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 analogous ANOVA, the two variables are the outcome variable and the grouping variable.

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

$$V = \sqrt{-2\sum_{j=1}^{k} w_j \left\{\mu_j(\ln\mu_0 - \ln\mu_j) + (1-\mu_j)\left[\ln(1-\mu_0) - \ln(1-\mu_j)\right]\right\}/(k-1)}, \quad (13)$$

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 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

$$\begin{aligned}\hat{V} &= \sqrt{\tilde{D}/n(k-1)} \\ &= \sqrt{-2\sum_{j=1}^{k} w_j \left\{\bar{y}_j(\ln\bar{y} - \ln\bar{y}_j) + (1-\bar{y}_j)\left[\ln(1-\bar{y}) - \ln(1-\bar{y}_j)\right]\right\}/(k-1)}.\end{aligned} \quad (14)$$

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 analogous 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 noncentral 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 noncentral parameter. Let $\Phi_{\chi^2(df,\lambda)}(x)$ be the cumulative distribution function of the noncentral chi-squared distribution; then the statistical power of the test is

$$\begin{aligned} power &= \Pr\{D \geq C|H_1\} \\ &= \Pr\{\chi^2(df, \lambda) \geq C\} \\ &= 1 - \Phi_{\chi^2(df,\lambda)}(C) \\ &= 1 - \Phi_{\chi^2[k-1,n(k-1)V^2]}\left[\chi^2_{1-\alpha}(k-1)\right]. \end{aligned} \tag{15}$$

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 Analogous ANOVA with Count Data

For comparison of group means with count data, the statistical inference is similar to that for binary data. The main difference lies in that the distribution of the outcome variable in the model is Poisson instead of Bernoulli.

4.1 Model and Test Statistic for Count Data

To construct the models for count data, let Y be the outcome variable, which can take only the nonnegative integer values, and A be a categorical variable of k levels. The null and alternative hypotheses are

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

The null model M_0 is

$$\mathrm{E}\{Y|A = j\} = \mu_0, \tag{16}$$

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

$$\mathrm{E}\{Y|A = j\} = \mu_j, \tag{17}$$

where $Y|(A = j) \sim Poisson(\mu_j)$, $j = 1, 2, \cdots, k$. Given a sample of data, $\mathbf{Y} = (\mathbf{y}_j) = \{y_{ij}\}$, $i = 1, 2, \cdots, n_j, j = 1, 2, \cdots, k$, with n_j denoting the sample size of the jth group, minus twice the log-likelihood ratio of model M_0 to M_1 is

$$\begin{aligned} D &= -2 \ln \frac{\mathscr{L}(\mu_0|\mathbf{Y})}{\mathscr{L}(\mu_1, \mu_2, \ldots, \mu_k|\mathbf{Y})} \\ &= -2[\ell(\mu_0|\mathbf{Y}) - \ell(\mu_1, \mu_2, \ldots, \mu_k|\mathbf{Y})] \\ &= -2(\ell_{M_0} - \ell_{M_1}) \end{aligned} \tag{18}$$

Under null hypothesis H_0, this statistic follows a chi-squared distribution $D \sim \chi^2(df)$ with the degrees of freedom $df = k - 1$ if the sample size tends to infinity (Wilks 1938). 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, we can calculate the test statistic as $\tilde{D}$ as follows. We first calculate minus twice the log-likelihood for M_0 and M_1:

$$SS_T = -2\hat{\ell}_{M_0} = -2\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\bar{y}_{ij}\ln\bar{y} - \bar{y}\right), \tag{19}$$

$$SS_W = -2\hat{\ell}_{M_1} = -2\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\bar{y}_{ij}\ln\bar{y}_j - \bar{y}_j\right); \tag{20}$$

and then $\tilde{D}$ as their difference:

$$SS_B = \tilde{D} = -2(\hat{\ell}_{M_0} - \hat{\ell}_{M_1}) = -2\sum_{j=1}^{k} n_j\left[\bar{y}_j(\ln\bar{y} - \ln\bar{y}_j) - (\bar{y} - \bar{y}_j)\right]. \tag{21}$$

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

4.2 Effect Size and Power Analysis for Count Data

For one-way analogous 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

$$\begin{aligned}\hat{V} &= \sqrt{\tilde{D}/n(k-1)} \\ &= \sqrt{-2\sum_{j=1}^{k} w_j\left[\bar{y}_j(\ln\bar{y} - \ln\bar{y}_j) + (\bar{y}_j - \bar{y})\right]/(k-1)},\end{aligned} \tag{22}$$

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 analogous ANOVA with count data is the same as that with binary data.

5 Software

To carry out the power analysis for analogous ANOVA with binary or count data, we have developed online applications that can be used within a Web browser. The link for the binary analogous ANOVA is http://psychstat.org/anovabinary and for the count analogous ANOVA is http://psychstat.org/anovacount. The software interface of power analysis for analogous ANOVA with binary data is shown in Fig. 1. Among *number of groups*, *sample size*, *effect size*, *significance level*, and *power*, any of them can be calculated given the rest of the information. The following examples illustrate the usage of the interface.

Suppose a student researcher hypothesizes that freshman, sophomore, junior, and senior college students have different rates of passing a reading exam. Based on his prior knowledge, he expects that the effect size is about 0.15. Based on the information, he wants to know (1) the power for him to find the significant difference among the four groups if he plans to collect data from 25 students in each of the four groups and (2) the minimum required sample size for him to find the significant difference among the four groups with power 0.8.

For the calculation of power, the number of group $k = 4$, the total sample size $n = 25 \times 4 = 100$, and the effect size $V = 0.15$. Let the significance level $\alpha = 0.05$, then we can use formula (15) to calculate the power as

$$\begin{aligned} power &= 1 - \Phi_{\chi^2[k-1, n(k-1)V^2]}\left[\chi^2_{1-\alpha}(k-1)\right] \\ &= 1 - \Phi_{\chi^2(3, 100\times 3\times 0.15^2)}\left[\chi^2_{0.95}(3)\right] \\ &= 1 - \Phi_{\chi^2(3, 6.75)}(7.8147) \\ &= 0.572. \end{aligned}$$

We can also use the online interface to estimate the power (see Fig. 1a). Given four groups, sample size 100, effect size 0.15, and significance level 0.05, the output indicates the power for this design is again 0.572.

With the required $power = 0.8$, $k = 4$, $V = 0.15$, and $\alpha = 0.05$, we solve the following equation:

$$\begin{aligned} power &= 1 - \Phi_{\chi^2[k-1, n(k-1)V^2]}\left[\chi^2_{1-\alpha}(k-1)\right] \\ 0.8 &= 1 - \Phi_{\chi^2(3, n\times 3\times 0.15^2)}\left[\chi^2_{0.95}(3)\right] \\ 0.8 &= 1 - \Phi_{\chi^2(3, n\times 0.0675)}(7.8147) \\ n &= 161.520. \end{aligned}$$

So, the minimum required sample size is 162.

Figure 1b shows how to use the interface to calculate the minimum required sample size. Given four groups, effect size 0.15, significance level 0.05, and the desired power 0.8, the output showed that a sample size 162, the near integer of

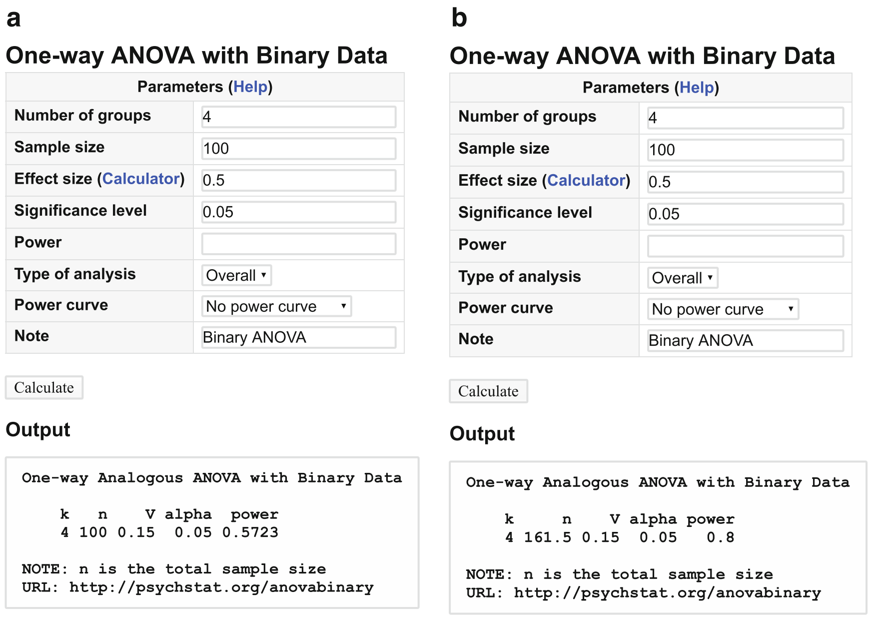

Fig. 1 Examples of power analysis for analogous ANOVA with binary data. (**a**) Given sample size, calculate power. (**b**) Given power, calculate sample size

161.5, is needed. A power curve can also be plotted by providing multiple sample sizes in the *Sample size* field. The interface for analogous ANOVA with count data is the same.

6 Discussion

In this chapter, an analogous ANOVA table and the closed-form likelihood ratio test statistic were introduced for comparing mean differences among groups of binary and count data, respectively. Based on the analogous ANOVA table and test statistic, the effect size V statistic, an adjusted phi coefficient, was defined. The power analysis involved four parameters, number of groups, total sample size, statistical significance level, and effect size. In addition, corresponding free online software were developed.

We recommend the application of these methods in binary and count data analysis. First, these methods are analogous to procedures in classical ANOVA as they decompose variation in observed outcomes for binary and count data. Specifically, the analogous ANOVA tables can help the researchers intuitively understand the exact meanings of the likelihood ratio test statistics used to compare means of binary or count data. Second, by using raw data and closed-form statistics, these methods are easier to use and more efficient than logistic regression or Poisson regression. Third, through the analogous ANOVA tables, we provide a unified solution for both binary and count data, while contingency tables cannot deal with count data. Future studies can investigate how to conduct power analysis for multiple comparisons 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

A. Agresti, A survey of exact inference for contingency tables. Stat. Sci. **7**(1), 131–153 (1992)
T. Baguley, Standardized or simple effect size: What should be reported? Br. J. Psychol. **100**(3), 603–617 (2009)
R. Bahadur, An optimal property of the likelihood ratio statistic, in *Proceedings of Fifth Berkeley Symp. Math. Statist. Probab*, vol. 1 (1967), pp. 13–26
M. Birch, Maximum likelihood in three-way contingency tables. J. R. Stat. Soc. Ser. B Methodol. **25**(1), 220–233 (1963)
Y.M. Bishop, S.E. Fienberg, P.W. Holland, *Discrete Multivariate Analysis: Theory and Practice* (MIT Press, Cambridge, 1975)
A.C. Cameron, F.A. Windmeijer, An r-squared measure of goodness of fit for some common nonlinear regression models. J. Econ. **77**(2), 329–342 (1997)
J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. (Lawrence Erlbaum Associates, Hillsdale, 1988)
D. Collett, *Modelling Binary Data* (Chapman & Hall, New York, 1991)
D.R. Cox, E.J. Snell, *Analysis of Binary Data*, 2nd edn. (Chapman & Hall, New York, 1989)
B. Efron, Regression and anova with zero-one data: measures of residual variation. J. Am. Stat. Assoc. **73**(361), 113–121 (1978)
P.D. Ellis, *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results* (Cambridge University Press, New York, 2010)
R.A. Fisher, On the "probable error" of a coefficient of correlation deduced from a small sample. Metron **1**, 3–32 (1921)
R.A. Fisher, On the interpretation of χ 2 from contingency tables, and the calculation of p. J. R. Stat. Soc. **85**(1), 87–94 (1922)
J.L. Fleiss, Measures of effect size for categorical data, in *The Handbook of Research Synthesis*, ed. by H. Cooper, L.V. Hedges (Russell Sage Foundation, New York, 1994), pp. 245–260
M.F. Freeman, J.W. Tukey, Transformations related to the angular and the square root. Ann. Math. Stat. **21**(4), 607–611 (1950)
L.A. Goodman, The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications. Technometrics **13**(1), 33–61 (1971)
L.A. Goodman, W.H. Kruskal, Measures of association for cross classifications. J. Am. Stat. Assoc. **49**(285), 732–764 (1954)
D.M. Grove, Positive association in a two-way contingency table: likelihood ratio tests. Commun. Stat. Theory Methods **13**(8), 931–945 (1984)

W. Hoeffding, Asymptotically optimal tests for multinomial distributions. Ann. Math. Stat. **36**(2), 369–401 (1965)

D.C. Howell, *Statistical Methods for Psychology* (Duxbury/Thomson Learning, Pacific Grove, 2012)

K. Larntz, Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. J. Am. Stat. Assoc. **73**(362), 253–263 (1978)

S.E. Maxwell, H.D. Delaney, *Designing Experiments and Analyzing Data: A Model Comparison Perspective*, 2nd edn. (Lawrence Erlbaum Associates, Mahwah, 2004)

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

D.S. Moore, W. Notz, M.A. Fligner, *Essential Statistics* (W.H. Freeman and Company, New York, 2013)

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

R. Rosenthal, Parametric measures of effect size, in *The Handbook of Research Synthesis*, ed. by H. Cooper, L.V. Hedges (Russell Sage Foundation, New York, 1994), pp. 231–244

J.S. Strickland, *Predictive Analytics Using R* (2014), Lulu.com

R.J. Turek, R.C. Suich, An exact test on the goodman-kruskal λ for prediction on a dichotomy: power considerations. Comput. Stat. Data Anal. **8**(2), 171–178 (1989)

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

D. Williams, Improved likelihood ratio tests for complete contingency tables. Biometrika **63**(1), 33–37 (1976)

D. Yates, D. Moore, G. McCabe, *The Practice of Statistics* (W.H. Freeman and Company, New York, 1999)