12-2021

PASSED: Calculate Power and Sample Size for Two Sample Tests

Jinpu Li

Ryan .. Knigge

Kaiyi Chen

Emily V. Leary

PASSED: Calculate Power and Sample Size for Two Sample Tests

by Jinpu Li, Ryan P. Knigge, Kaiyi Chen, Emily V. Leary

Abstract Power and sample size estimation are critical aspects of study design to demonstrate minimized risk for subjects and justify the allocation of time, money, and other resources. Researchers often work with response variables that take the form of various distributions. Here, we present an R package, **PASSED**, that allows flexibility with seven common distributions and multiple options to accommodate sample size or power analysis. The relevant statistical theory, calculations, and examples for each distribution using **PASSED** are discussed in this paper.

Introduction

Power and sample size estimation are critical aspects of study design to demonstrate minimized risk for subjects and justify the allocation of time, money, and other resources (Jones et al., 2003). A number of R packages for power analysis have been developed over the years. The **samplesize** (Scherer, 2016) package provides the calculation of sample size for the Student's t-test and the Wilcoxon-Mann Whitney test for categorical data. The **TrialSize** (Zhang et al., 2013) package implements the power analysis described in Chow et al. (2007), including power and sample size calculations for different study designs. Most recently, the **simglm** (LeBeau, 2019) package presents a simulation approach for power analysis that allows for the specification of missing data, unbalanced designs, and different random error distributions of generalized linear models.

Moreover, researchers often work with response variables that can take the form of a variety of distributions. For example, the proportion of thromboembolism after surgery in different treatment groups can be modeled using the binomial distribution or length of inpatient stay after an orthopedic procedure can be modeled using the Poisson distribution (Plessl et al., 2020). Some of the R packages or functions are designed to calculate the power and sample size for the variables following a certain distribution. The base package **stats** (R Core Team, 2016) provides such functions for normal (Gaussian) and binomially distributed variables, and the situations of unequal sample sizes are extended by packages **pwr** (Champely et al., 2017), **MESS**(Ekstrøm, 2012), **pwr2ppl**(Aberson, 2019), and **WebPower** (Zhang and Mai, 2018). The package **MKmisc** (Kohl, 2021) further adds a function for the comparison of negative binomial distributions. However, none of these packages provide a comprehensive power analysis toolkit capable of calculating power or sample sizes for the test of two-sample means or ratios when the responses have other common distributions (Table 1).

Package	Binomial	Normal	Negative Binomial	Geometric	Poisson	Beta	Gamma
PASSED	x	x	x	x	x	x	x
stats	x*	x*					
pwr	x	x**					
WebPower	x	x**					
MESS	x	x					
pwr2ppl	x	x					
MKmisc			x	x			

*: equal sample only; **: equal variance only.

Table 1: The comparison among **PASSED** and other available packages.

Here, we present an R package, **PASSED**, that performs power and sample size analyses for the following distributions: binomial, negative binomial, geometric, Poisson, normal (Gaussian), beta, and gamma distributions. Distributions, which had existing functions or R infrastructure for sample size and power calculations were included to streamline these calculations. However, calculations for the beta, Poisson, and gamma distributions were developed specifically for inclusion in **PASSED**. In the following sections, we will discuss the motivating examples, relevant statistical theory, and

calculations for each distribution using **PASSED**.

PASSED: R Package Description

All functions in this package can be used to compute the power for a specific study design (e.g., given sample sizes) or to estimate specific parameter values (e.g., sample sizes) necessary to obtain a target power. The specific function of interest will depend on the type of outcome variable and the data distribution. All functions output an object of class `power.htest` that details the specified parameters of the test and the estimated parameter set as `NULL`.

Binomial

The binomial distribution is useful when modeling the number of successes in a sequence of independent and identically distributed Bernoulli trials. One example which uses data modeled using a binomial distribution is the proportion of blood transfusion that has occurred during surgery. The need for blood transfusion during surgery is an important consideration during surgical planning and particularly for surgical trials. LEITE et al. (2020) applied a logistic regression with binomial outcomes to model the rate of blood transfusions after the introduction of Tranexamic acid in knee arthroplasty.

Hypothesis Testing two-sample proportions is commonly considered in research designs when the outcome follows a binomial distribution. Let x_{ij} be a binary response from the jth subject in the ith group, $j = 1, ..., n_i$, $i = 1, 2$. It is assumed that x_{ij} are independent Bernoulli random variables with proportion p_i,

$$x_{ij} \sim Bernoulli(p_i)$$

Two hypothesis frameworks are considered for power and sample size calculations, which correspond to either a one-sided or two-sided test:

$$H_0 : p_1 = p_2 \; vs. \; H_a : p_1 \neq p_2 \; (two - sided)$$

or

$$H_0 : p_1 = p_2 \; vs. \; H_a : p_1 > (<) p_2 \; (one - sided)$$

Algorithm A binomial asymptotic test statistic was first proposed by Pearson (1900). Fleiss et al. (1980) provided an explicit formula to calculate the corresponding sample sizes for the test:

$$n_1 = \frac{[z_{\frac{\alpha}{2}}\sqrt{(r+1)\bar{p}\bar{q}} + z_\beta\sqrt{rp_1q_1 + p_2q_2}]^2}{rd^2}(two - sided)$$

or

$$n_1 = \frac{[z_\alpha\sqrt{(r+1)\bar{p}\bar{q}} + z_\beta\sqrt{rp_1q_1 + p_2q_2}]^2}{rd^2}(one - sided),$$

where $r = n_2/n_1$, $d = p_2 - p_1$, $q_1 = 1 - p_1$, $q_2 = 1 - p_2$, $\bar{p} = \frac{n_1p_1+n_2p_2}{n_1+n_2}$, $\bar{q} = 1 - \bar{p}$, and z_x denotes the probability that a standard normal deviate is greater than x. To obtain the power, this equation can be re-written as:

$$z_\beta = \frac{\sqrt{rn_1}|d| - z_{\frac{\alpha}{2}}\sqrt{(r+1)\bar{p}\bar{q}}}{\sqrt{rp_1q_1 + p_2q_2}}(two - sided)$$

$$z_\beta = \frac{\sqrt{rn_1}|d| - z_\alpha\sqrt{(r+1)\bar{p}\bar{q}}}{\sqrt{rp_1q_1 + p_2q_2}}(one - sided)$$

And, thus, the power can be derived as:

$$\text{Power} = \Pr\left(Z < \frac{z_{\frac{\alpha}{2}}\sqrt{(r+1)\bar{p}\bar{q}} - \sqrt{rn_1}|d|}{\sqrt{rp_1q_1 + p_2q_2}}\right)(two - sided)$$

$$\text{Power} = \Pr\left(Z < \frac{z_\alpha\sqrt{(r+1)\bar{p}\bar{q}} - \sqrt{rn_1}|d|}{\sqrt{rp_1q_1 + p_2q_2}}\right)(one - sided)$$

As a result, the target power, required sample sizes (n_1 and n_2), significance level (α), or the proportions (p_1 and p_2) can be obtained once all other remaining parameters are known (Fleiss et al., 1980). To optimize the sample size allocation, please refer to the discussion in Brittain and Schlesselman (1982).

Function The `power_Binomial()` function is useful when testing for differences among two sample proportions when the data follow a binomial distribution. This function uses the algorithm described above. The arguments for `power_Binomial()` are as follows:

```
power_Binomial(n1 = NULL, n2 = NULL, p1 = 0.5, p2 = 0.5,
               sig.level = 0.05, power = NULL, equal.sample = TRUE,
               alternative = c("two-sided", "one-sided"))
```

Sample sizes for each group are designated as `n1` and `n2`. If sample sizes for both groups are equal, the argument `equal.sample` should be set to `TRUE`, and only a value for `n1` is needed. If sample sizes are unequal, `equal.sample` should be set to `FALSE`, and values for both `n1` and `n2` must be specified. When estimating other parameters, the target power must be set with `power`. The significance level is set with `sig.level` and has a default value of 0.05. The probability of success for each group is indicated as `p1` and `p2`, respectively, with 0.5 as the default value for both. Only one of the parameters of `n1`, `n2`, `p1`, `p2`, `power`, or `sig.level` can be set as `NULL`. The parameter set as `NULL` will be estimated based on the other parameter values. The argument `alternative` specifies the alternative hypothesis as either `"two.sided"` (default) or `"one.sided"`.

The `power_Binomial()` function returns the same results as `stats::power.prop.test()` in the equal sample scenario. It also allows power calculations with unequal sample sizes, and the results are identical to `MESS::power_prop_test()`.

Negative Binomial

The negative binomial distribution can be used to model the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified number of failures occurs. Gates et al. (2020) analyzed the probability of positive intraoperative cultures in a population of patients with a history of prior ipsilateral shoulder surgery. The probability of the total number of positive tissue cultures was modeled using a generalized negative binomial mixed model with maximum likelihood estimation and robust standard errors. Using this negative binomial framework, the appropriate sample size and power for such a study can be obtained using the method outlined below.

Hypothesis Consider a sequence of adverse events. Let x_{ij} be the number of events during time t_i from the jth subject in the ith group, $j = 1, ..., n_i$, $i = 1, 2$. Assuming that x_{ij} are negative binomial random variables with a mean μ_{ij} and parameter θ ($\theta > 0$), the probability function of x_{ij} is

$$P\left(x_{ij}\right) = \frac{\Gamma\left(\theta + x_{ij}\right)}{\Gamma\left(\theta\right) x_{ij}!} \left(\frac{\mu_{ij}}{\theta + \mu_{ij}}\right)^{x_{ij}} \left(\frac{\theta}{\theta + \mu_{ij}}\right)^{\theta}, \tag{1}$$

where $n!$ denotes the product of the integers from 1 to n and $\Gamma(\cdot)$ is the gamma function (Zhu and Lakkis, 2014).

To model the negative binomial outcomes, Hilbe (2011) introduced the negative binomial regression. Zhu and Lakkis (2014) then presented a hypothesis test comparing two negative binomial distributed samples using negative binomial regression, and this is the method used here. In negative binomial regression, μ_{ij} can be modeled as

$$log(\mu_{ij}) = log(t_i) + \beta_0 + \beta_1 G_{ij},$$

where G_{ij}, the group indicator for subject j, is equal to 0 if $i = 1$ for group 1 and is equal to 1 if $i = 2$ for group 2. Let r_1 and r_2 be the mean rates of events per time unit for groups 1 and 2, which can be expressed as $r_1 = e^{\beta_0}$ and $r_2 = e^{\beta_0 + \beta_1}$. Then $r_2/r_1 = e^{\beta_1}$ can be easily obtained (Zhu and Lakkis, 2014).

To compute the power of the test or determine parameters to obtain target power, two hypothesis frameworks are considered which correspond to either a one-sided or two-sided test:

$$H_0 : \frac{r_2}{r_1} = 1 \; vs. \; H_a : \frac{r_2}{r_1} \neq 1 \; (two - sided)$$

or

$$H_0 : \frac{r_2}{r_1} = 1 \; vs. \; H_a : \frac{r_2}{r_1} > (<) 1 \; (one - sided)$$

Algorithm The power and sample size calculation algorithms were developed by Zhu and Lakkis (2014) based on the asymptotic normality of the maximum likelihood estimation of β_1. The power can

be calculated as:

$$power = \Phi\left(\frac{\sqrt{n_1}\left|\log\left(\frac{r_2}{r_1}\right)\right| - z_{\frac{\alpha}{2}}\sqrt{V_0}}{\sqrt{V_1}}\right) \ (two-sided)$$

or

$$power = \Phi\left(\frac{\sqrt{n_1}\left|\log\left(\frac{r_2}{r_1}\right)\right| - z_{\alpha}\sqrt{V_0}}{\sqrt{V_1}}\right) \ (one-sided),$$

where V_0 and V_1 are the estimates of variance for $\hat{\beta}_1$ by n_1 under H_0 and H_a,

$$V_0 = \frac{1}{t_i}\left(\frac{1}{\tilde{r}_1} + \frac{n_1}{n_2\tilde{r}_2}\right) + \frac{(n_1+n_2)}{\theta n_2}$$

$$V_1 = \frac{1}{t_i}\left(\frac{1}{r_1} + \frac{n_1}{n_2 r_2}\right) + \frac{(n_1+n_2)}{\theta n_2},$$

and $\tilde{r}_i$, $i = 1, 2$ denotes the estimation of the event rate under H_0 in each group. Zhu and Lakkis (2014) provided three approaches to estimating $\tilde{r}_i$ under H_0:

Approach 1: using event rate of group 2 (reference group rate)

$$V_0 = \frac{1}{t_i}\left(\frac{1}{r_2} + \frac{n_1}{n_2 r_2}\right) + \frac{(n_1+n_2)}{\theta n_2};$$

Approach 2: using true rates

$$V_0 = \frac{1}{t_i}\left(\frac{1}{r_1} + \frac{n_1}{n_2 r_2}\right) + \frac{(n_1+n_2)}{\theta n_2};$$

Approach 3: using maximum likelihood estimation

$$V_0 = \frac{1}{t_i}\left(\frac{1}{\frac{n_1 r_1 + n_2 r_2}{n_1+n_2}} + \frac{n_1}{n_2\frac{n_1 r_1 + n_2 r_2}{n_1+n_2}}\right) + \frac{(n_1+n_2)}{\theta n_2}.$$

Function The function `power_NegativeBinomial()` is useful when developing a study design to compare differences in rates when the data follow a negative binomial distribution. Calculations for this function are based on Zhu and Lakkis (2014). The following arguments are used:

```
power_NegativeBinomial(n1 = NULL, n2 = NULL, power = NULL, sig.level = 0.05,
                       mu1 = NULL, mu2 = NULL, duration = 1, theta = NULL,
                       equal.sample = TRUE,
                       alternative = c("two-sided", "one-sided"), approach = 3)
```

The sample size for each group is specified as `n1` and `n2`, both with default values of `NULL`. When sample sizes are equal, `equal.sample` can be set to `TRUE`, and only `n1` must be specified. Otherwise `equal.sample` is set to `FALSE` and values must be input for both `n1` and `n2`. The `power` argument is set to `NULL` unless the target power is specified here and another parameter is set as `NULL` to be estimated. The significance level for the test is set by `sig.level` with a default value of 0.05. The expected rates of events per unit time for each group are denoted as `mu1` and `mu2`, respectively, with the average treatment duration set by `duration` (default value of 1). `Theta` indicates the θ parameter of the negative binomial distribution, as noted above. The argument `alternative` specifies the alternative hypothesis as either `"two.sided"` (default) or `"one.sided"`. Lastly, the argument `approach` can be set as either "1", "2", or "3" (default). These values indicate the selection of one of three procedures for estimating the variance under the null hypothesis for the sample size formula and correspond with Approach 1 (reference group rate), Approach 2 (true rates), and Approach 3 (maximum likelihood estimation) described above. The obtained results match other functions in R such as `MKmisc::power.nb.test()`.

Geometric

The geometric distribution can be used to examine the probability of success given a limited number of trials and is considered a special case of the negative binomial distribution. For example, in baseball, the probability of a batter earning a hit before striking out can be compared to that of another batter, using a geometric distribution.

Hypothesis Let x_{ij} be the number of events during time t_i from the jth subject in the ith group. Assuming that x_{ij} are geometric random variables with a mean μ_{ij}, the probability function of x_{ij} is

$$P\left(x_{ij}\right) = \left(\frac{\mu_{ij}}{1+\mu_{ij}}\right)^{x_{ij}} \left(\frac{1}{1+\mu_{ij}}\right).$$

Referring to Equation 1, this is a special case of the negative binomial where $\theta = 1$. Similarly, μ_{ij} can be modeled as shown in the Section Negative Binomial,

$$log(\mu_{ij}) = log(t_i) + \beta_0 + \beta_1 G_{ij}.$$

The hypotheses and calculations follow as previously shown in the Section Negative Binomial.

Algorithm The power and sample size calculation formula are the same as the Section Negative Binomial, with $\theta = 1$.

Function The function `power_Geometric()` applies the same algorithm as the function `power_NegativeBinomial()`, with the same arguments, where the parameter `theta` is set as 1. See `power_NegativeBinomial()` for more details.

```
power_Geometric(n1 = NULL, n2 = NULL, power = NULL, sig.level = 0.05, mu1 = NULL,
                mu2 = NULL, duration = 1, equal.sample = TRUE,
                alternative = c("two-sided", "one-sided"), approach = 3)
```

Poisson

The Poisson distribution can be used to model the number of events occurring in a fixed interval of time or space. In healthcare, length of stay (LOS) is one of many important considerations for interventions, particularly when inpatient hospital stay may vary among treatments. LOS, or other count measurements important to the research study, can be modeled using a Poisson distribution. Plessl et al. (2020) used a Poisson framework to compare LOS for those who were treated with rapid recovery protocols versus standard recovery protocols after total knee arthroplasty. This example can be expanded to the general case as follows.

Hypothesis Let x_{ij} be the number of events during the necessary study time t_i from the jth subject in the ith treatment group, $j = 1, ..., n_i$, $i = 1, 2$. This situation is commonly referred to as the equal sampling frame approach (Hutchinson and Holtman, 2005). It is assumed that x_{ij} are Poisson random variables with rate λ_i such that the probability function of x_{ij} is

$$P\left(x_{ij}\right) = \frac{t_i \lambda_i e^{t_i \lambda_i}}{x_{ij}!},$$

where $i = 1, 2$. Then, the total number of events in each group, denoted as X_1 and X_2, also follow a Poisson distribution:

$$X_i \sim Poisson(\lambda_i t_i n_i)$$

Four methods have previously been proposed to test the equality of two Poisson rates (Shiue and Bain, 1982; D. Huffman, 1984; Thode, 1997; Gu et al., 2008). The method utilized in the **PASSED** package was proposed by Gu et al. (2008), which considers the ratio of two Poisson rates, R, a pre-specified positive number. The asymptotic test is as follows:

$$H_0 : \lambda_2/\lambda_1 = R \; vs. \; H_a : \lambda_2/\lambda_1 = R' \neq R (two - sided)$$

or

$$H_0 : \lambda_2/\lambda_1 = R \; vs. \; H_a : \lambda_2/\lambda_1 = R' > R (one - sided)$$

Algorithm The following formula is used in the **PASSED** package and the details of the derivation are provided in the Appendix.

$$n_1 = \frac{(\frac{z_{1-\frac{\alpha}{2}} C + z_{power} D}{A})^2 - \frac{3}{8}}{\lambda_1 t_1} (two - sided)$$

or

$$n_1 = \frac{(\frac{z_{1-\alpha}C + z_{power}D}{A})^2 - \frac{3}{8}}{\lambda_1 t_1}(one - sided)$$

where $A = 2(1 - \sqrt{\frac{R}{R'}})$,
$B = \lambda_1 t_1 n_1 + 3/8$,
$C = \sqrt{\frac{R+d}{R'}}$,
$D = \sqrt{\frac{R'+d}{R'}}$
$d = t_1/t_2$.

Function The `power_Poisson()` function is designed to compute the power or estimate parameters to obtain a target power when testing for a ratio of two Poisson rates. This function applies the asymptotic tests based on normal approximations developed by Gu et al. (2008). The arguments for `power_Poisson()` are as follows:

```
power_Poisson(n1 = NULL, n2 = NULL, power = NULL, sig.level = 0.05,
              lambda1 = NULL, lambda2 = NULL, t1 = 1, t2 = 1, RR0 = 1,
              equal.sample = TRUE, alternative = c("two.sided", "one.sided"))
```

Sample sizes for each group are set with `n1` and `n2`. If sample sizes for both groups are equal, the argument `equal.sample` should be set to `TRUE`, and only a value for `n1` needs to be specified. If sample sizes are unequal, `equal.sample` should be set to `FALSE`, and values for both `n1` and `n2` must be specified. The target power of the test is set with `power`, and the significance level is set with `sig.level` (default value of 0.05). The expected rates of events per unit time for each group are denoted as `lambda1` and `lambda2`, respectively, with the average treatment duration set by `t1` and `t2` (default value of 1). Only one of the parameters of `n1`, `n2`, `lambda1`, `lambda2`, or `power` can be set as `NULL` for the function to run. The parameter set as `NULL` will be estimated based on the other parameter values. `t1` and `t2` refer to the specified interval of time (or space) where the events occur. The rate ratio from the null hypothesis is specified as `RR0`. It should be set to 1 when testing for equal Poisson rates. The argument `alternative` specifies the alternative hypothesis as either `"two.sided"` (default) or `"one.sided"`.

For the example in Gu et al. (2008), which aims at testing if the risk of coronary heart disease is greater for those with postmenopausal hormone use (RR0 = 1), the event rates for those with and without hormone use are assumed to be 0.2000 and 0.0005 (lambda2 = 0.0020, lambda1 = 0.0005), respectively, during a 2-year time period (t1 = t2 = 2). Given the sample size for each group as 4295 and 8590 (n2 = 4295, n1 = 8590) [1], the power under a significance level of 0.05 can be calculated as follows:

```
power_Poisson(n1 = 8590, n2 = 4295, power = NULL, sig.level = 0.05,
                      lambda1 = 0.0005, lambda2 = 0.0020, t1 = 2, t2 = 2, RR0 = 1,
                      equal.sample = FALSE, alternative = "one.sided")
```

The estimated power is 0.9000147, which matches the results in Gu et al. (2008).

Normal

The normal distribution is widely used in the natural and social sciences. Age is a common demographic variable recorded during patient care and typically follows a normal distribution. Many surgeons consider demographic variables to evaluate the possible risks of a surgical procedure and assess optimal treatment options for patients. Luan et al. (2020) aimed to identify patients who were suitable for kinematic or mechanical alignment of the knee. To compare these groups, Luan et al. (2020) used the student t-test to compare normally distributed age.

Hypothesis T-tests are widely used to compare two sample means when the data has a normal distribution (Cressie and Whitford, 1986). Let x_{ij} be a continuous response from the jth subject in the ith group, $j = 1, ..., n_i$, $i = 1, 2$. It is assumed that x_{ij} are independent, normal random variables with mean μ_i and variance σ_i^2:

$$x_{ij} \sim Normal(\mu_i, \sigma_i^2),$$

[1] the sample sizes are corrected in NCSS Software Manuals 2020 Page 437-14, "Tests for the Ratio of Two Poisson Rates"

then the probability density function of x_{ij} is:

$$f\left(x_{ij}\right) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}\left(\frac{x_{ij}-\mu_i}{\sigma_i}\right)^2},$$

where $i = 1, 2$. It can be shown that the mean of each group, denoted as $\bar{x}_1$ and $\bar{x}_2$, also follows a Normal distribution:

$$\bar{x}_i \sim Normal(\mu_i, \frac{\sigma_i^2}{n_i})$$

To compute the power for a hypothesis test or determine parameters to obtain a target power for hypothesis, the following two scenarios are considered:

$$H_0 : \mu_1 = \mu_2 \; vs. \; H_a : \mu_1 \neq \mu_2 \; (two - sided)$$

or

$$H_0 : \mu_1 = \mu_2 \; vs. \; H_a : \mu_1 > (<)\mu_2 \; (one - sided)$$

Algorithm Based on the work of Ekstrøm (2012), in the **PASSED** package, the user can define the sample sizes (n_1 and n_2) and standard deviations (σ_1 and σ_2) of each group directly, rather than set the size ratio (n_2/n_1) and standard deviation ratio (σ_2/σ_1). To optimize sample size allocation, please refer to the discussion in Jan and Shieh (2011).

Function The `power_Normal()` function is useful for developing a study design to test for differences between mean values of two groups when the data follow a normal distribution. This function performs the same operations as `pwr.t.test` in the **pwr** package (Champely et al., 2017) but allows for additional parameter modifications. In particular, this function allows for specifying unequal sample sizes and standard deviations across groups. The arguments for `power_Normal()` are as follows:

```
power_Normal(n1 = NULL, n2 = NULL, power = NULL, sig.level = 0.05,
             delta = NULL, sd1 = 1, sd2 = 1, equal.sample = TRUE,
             alternative = c("two-sided", "one-sided"),
             type = c("two-sample", "one-sample", "paired"),
             df.method = c("welch", "classical"), strict = FALSE)
```

Sample sizes for each group are set with `n1` and `n2`. If sample sizes for both groups are equal, the argument `equal.sample` should be set to `TRUE`, and only a value for `n1` needs to be specified. If sample sizes are unequal, `equal.sample` must be set to `FALSE`, and values for both `n1` and `n2` must be specified. The target power of the test is set with `power`, and the significance level is set with `sig.level` (default value of 0.05). `delta` indicates the difference in means between the two groups, and `sd1` and `sd2` denote the standard deviations for each group. A default value of 1 is indicated for both `sd1` and `sd2`. The default values for `n1`, `n2`, `power`, and `delta` are `NULL`, whereas `sd1`, `sd2`, and `sig.level` have non-NULL default values. Only one of the parameters can be set as `NULL`. The parameter set as `NULL` will be estimated based on the other parameter values. The type of t-test is indicated by `type` and set as `"two.sample"` (default), `"one.sample"`, or `"paired"`. `alternative` specifies the alternative hypothesis as either `"two.sided"` (default) or `"one.sided"`. Lastly, `df.method` indicates the method for calculating the degrees of freedom as either `"welch"` (default) or `"classical"`. Note that setting `strict` as `TRUE` would be applied only in the two-sided case, when the probability of rejection in the opposite direction of the true effect is included, i.e., the alternative hypothesis of the two-sided t-test is $\mu_1 \neq \mu_2$ rather than $\mu_1 > (<)\mu_2$.

The `power_Normal()` function produces the same results as `stats::power.t.test()` for the equal sample size scenario. It also allows power calculations with unequal sample sizes and unequal variances. The results match other functions in R such as `MESS::power_prop_test()` and `pwr::pwr.t2n.test()`.

Beta

The beta family of continuous probability distributions is ideal for modeling data with right or left skewness and allows the probability density to assume a variety of shapes through two shape parameters (Gupta and Nadarajah, 2004). Disease status is often measured with bounded outcome scores, which take values on a finite range. The distribution of such data is often skewed, rendering the standard analysis methods assuming a normal distribution inappropriate (Hu et al., 2020), and thus, a beta distribution can be utilized. This scenario can be generalized as follows.

Hypothesis Suppose a sequence of random responses, x_{ij} from the jth subject in the ith group, takes the form of a continuous proportion that follows a beta distribution, $x_{ij} \sim Beta(a_i, b_i)$, where $j = 1, ..., n_i$, $i = 1, 2$. The probability density function of x_{ij} is:

$$f(x_{ij}) = \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} x_{ij}^{a_i - 1}(1 - x_{ij})^{b_i - 1},$$

where $0 \leq x_{ij} \leq 1$, $a_i > 0$, $b_i > 0$, and $i = 1, 2$. When analyzing continuous proportions as a response variable, the standard shape parameters of a beta density, a_i and b_i, are often not directly observable. Ferrari and Cribari-Neto (2004) developed a class of beta regression models which utilize an alternative parameterization of the beta density function based on the mean, μ_i, and an unknown precision parameter, ϕ_i. Suppose $\mu_i = a_i/(a_i + b_i)$ and $\phi_i = a_i + b_i$, then the beta density function can be expressed in terms of μ_i and ϕ_i as below:

$$f(x_{ij}) = \frac{\Gamma(\phi_i)}{\Gamma(\mu_i\phi_i)\Gamma((1 - \mu_i)\phi_i)} x_{ij}^{\mu_i\phi_i - 1}(1 - x_{ij})^{(1-\mu_i)\phi_i - 1};$$

For beta regression, μ_i can be modeled as

$$g(\mu_i) = \beta_0 + \beta_1 G_{ij},$$

where G_{ij}, the group indicator for subject j, is equal to 0 if $i = 1$ for group 1 and is equal to 1 if $i = 2$ for group 2, and $g(\cdot)$ denotes the link function. The **PASSED** package includes the capability for the following link functions and their respective forms:

- Logit: $g(\mu) = log\left[\frac{\mu}{(1-\mu)}\right]$
- Probit: $g(\mu) = \Phi^{-1}(\mu)$
- Complementary log-log: $g(\mu) = log[-log(1 - \mu)]$
- Log: $g(\mu) = log(\mu)$
- Log-log: $g(\mu) = -log[-log(\mu)]$

The equality of means μ_i is equivalent to $\beta_1 = 0$. The objective is to compute the power of the test or determine minimum sample sizes to obtain a target power for the needed hypothesis. A two-sided hypothesis framework is considered for power and sample size calculations:

$$H_0 : \mu_1 - \mu_2 = 0 \; vs. \; H_a : \mu_1 - \mu_2 \neq 0$$

Algorithm The mean and variance of x_{ij}, denoted as μ_i and σ_i^2, can be obtained using:

$$\mu_i = \frac{a_i}{a_i + b_i}$$

and

$$\sigma_i^2 = \frac{a_i b_i}{(a_i + b_i)^2(a_i + b_i + 1)}.$$

Incorporating the definition of the precision parameter ϕ_i, the following equations can be derived:

$$a_i = \mu_i\phi_i = \mu_i\Big(\frac{\mu_i(1 - \mu_i)}{\sigma_i^2} - 1\Big); \tag{2}$$

$$b_i = (1 - \mu_i)\phi_i = (1 - \mu_i)\Big(\frac{\mu_i(1 - \mu_i)}{\sigma_i^2} - 1\Big). \tag{3}$$

To calculate power, a simulation approach is used. Parameters μ_i and ϕ_i are first estimated using the given mean and variance, then they are used to obtain the original beta parameters, a_i and b_i, following Equations 2 and 3. The response variable is simulated for each distribution, $Beta(a_1, b_1)$ and $Beta(a_2, b_2)$, with the given sample size. If any simulated response is equal to zero or one, the following transformation is applied to each response value from both distributions: $(x(n - 1) + 0.5)/n$, where x is the response value and n is the sample size (Smithson and Verkuilen, 2006).

Values for the simulated response from both distributions are merged together, along with the group indicator (0 for group 1 and 1 for group 2). Subsequently, a beta regression model is built using the specified link type (Cribari-Neto and Zeileis, 2010). A Wald test is performed on the simulated model, testing the null hypothesis that β_1 is equal to 0. The p-values are recorded for each test and the

simulation is repeated M times. The power is calculated as:

$$\text{power} = \frac{\text{Number of p-values less than } 0.05}{M}.$$

Let ss denote sample size, then the generic power/sample size relationship can be formally expressed as:

$$\text{power} = \text{f(ss)}$$

Assuming the response variable follows a beta distribution, $f(\cdot)$ is continuous on the interval (0, 1) and increases monotonically. Consequently, the `power_Beta` function uses the bisection method to obtain the minimum sample size, ss_0, through a sequence of steps for each iteration (Chernick and Liu, 2002). For each target power, $power_0$, upper and lower sample size bounds, ss_u and ss_l, which satisfy $f(ss_l) < power_0 < f(ss_u)$ are established using a two-sample t-test performed with the base function `power.t.test` (R Core Team, 2016). Although `power.t.test` assumes normality, it is useful to generate starting values for ss_u and ss_l.

The sequence of steps for each iteration is as follows:

1. Compute the midpoint $ss_{mid} = floor(\frac{ss_l + ss_u}{2})$ of interval $[ss_l, ss_u]$. $floor(\cdot)$ denotes retaining the integer part of a number.
2. Calculate power at the midpoint, ss_{mid}, using the simulation described for the power calculations above.
3. If $f(ss_{mid}) \geq power_0$ and $ss_{mid} - ss_l \leq 1$, then return ss_{mid} and stop iterating.
4. Examine the sign of $f(ss_{mid}) - power_0$. If negative, then replace ss_l with ss_{mid}, otherwise replace ss_u with ss_{mid} so that $f(ss_l) < power_0 \leq f(ss_u)$.

Repeat the process until iteration stops. The output minimum sample size, ss_0, is the minimum integer such that $f(ss_0) \geq power_0$.

Function The `power_Beta()` function is framed to test differences between mean values for two groups, assuming the response variable follows a beta distribution in each group. It can be used to compute the power or to estimate the required sample sizes to obtain a target power. In particular, this function allows for specifying unequal sample sizes and standard deviations across groups. The arguments for `power_Beta()` are as follows:

```
power_Beta(n1 = NULL, n2 = NULL, power = NULL, sig.level = 0.05,
           mu1 = NULL, sd1 = NULL, mu2 = NULL, equal.sample = TRUE,
           trials = 100, equal.precision = TRUE, sd2 = NULL,
           link.type = c("logit", "probit", "cloglog", "cauchit", "log", "loglog"))
```

Sample sizes for each group are set with `n1` and `n2`. If sample sizes for both groups are equal, the argument `equal.sample` should be set to `TRUE`, and only a value for `n1` or `power` needs to be specified. If sample sizes are unequal, `equal.sample` should be set to `FALSE`, and values for both `n1` and `n2` must be specified. The target power of the test is set with `power`, and the significance level is set with `sig.level` (default value of 0.05). Only one of the parameters of `n1`, `n2`, or `power` can be `NULL`. The mean and standard deviation for the null distribution are denoted by `mu1` and `sd1`. Analogously, the mean and standard deviation for the alternative distribution can be specified by `mu2` and `sd2`. Note that `equal.precision=FALSE` should be used to set the standard deviation for the alternative distribution, meaning the precision parameters are assumed to be unequal. Otherwise, option `sd2` would be ignored. The option `trials` indicates the number of trials in the simulation. A default number of trials (i.e., 100) is recommended to get a rough estimate of other parameters (e.g., `sd2`), since the computational time is dependent upon the number of trials in the simulation. Once an appropriate range of other values is determined, the number of trials should be increased (e.g., trials=1000) to calculate precise power and sample size estimates. The default link function is the logit link but can be changed using `link.type` with the following options: `"logit"`, `"probit"`, `"cloglog"`, `"log"`, `"loglog"`, to denote the logit, probit, complementary log-log, log, and log-log link functions, respectively.

Gamma

The gamma distribution is widely used to fit lifetime data because its flexibility in shape can vary from extremely positively skewed to almost symmetric (Casella and Berger, 2002). Hong et al. (2020) provide an example of modeling data using the gamma distribution to test the association of patient-provider cost discussion with out-of-pocket spending among cancer survivors. The data (i.e., out-of-pocket spending in cancer care) have an obvious skewness which is not normally distributed; and therefore,

the two-sample t-test is not suitable for this purpose. Alternatively, gamma models can be used to test the difference of average total out-of-pocket spending between the patients with and without a patient-provider cost discussion.

Hypothesis Currently, there is no explicit formula to calculate the power comparing two gamma random variables. Let x_{ij} be a continuous response from the jth subject in the ith group, $j = 1, ..., n_i$, $i = 1, 2$. It is assumed that x_{ij} are gamma random variables with scale λ_i and shape δ_i so that the probability density function can be written as

$$f\left(x_{ij} = x\right) = \left(\frac{\lambda_i}{\Gamma\left(\delta_i\right)}\right) x^{\delta_i - 1} e^{-\lambda_i x}.$$

The mean of $Gamma(\lambda_i, \delta_i)$ can be obtained using $\mu_i = \delta_i / \lambda_i$. Shiue and Bain (1983) developed a test of two equal gamma means with unknown common shape parameter, such that

$$H_0 : \mu_1 = \mu_2 = \mu.$$

This can be re-written as $H_0 : \delta = \lambda_1 \mu = \lambda_2 \mu$, some $\delta > 0$. This can then be tested using an F distribution based on the ratio of the mean of a random sample from two gamma distributions. In 1988, Shiue et al. (1988) extended this to the unknown and unequal shape parameter scenarios. However, this extension can be slightly conservative and problematic for smallscale parameters. More recently, Chang et al. (2011) provided a computational approach using a variant of the parametric bootstrap method, used here, in which the shape parameters are completely unknown and unequal. In this characterization, the hypothesis is two-sided and is of the form $H_0 : \delta_i = \lambda_i \mu$, some $\mu > 0$ or equivalently, for two means,

$$H_0 : \frac{\delta_1}{\lambda_1} = \frac{\delta_2}{\lambda_2} \; vs. \; H_a : \frac{\delta_1}{\delta_1} \neq \frac{\delta_2}{\lambda_2}.$$

This can be expressed as a scalar value function, η, such that

$$H_0^* : \eta = \sum_{i=1}^{2} \left(\beta_i - \bar{\beta}\right)^2 = 0 \; vs. \; H_a^* : \eta > 0,$$

where $\beta_i = \ln\left(\mu_i\right)$ and $\bar{\beta} = \sum_{i=1}^{2} \frac{\beta_i}{2}$.

Algorithm The power and sample size calculation algorithm adapted for **PASSED** was developed by Chang et al. (2011). This computational approach performs best when the restricted maximum likelihood estimate of η behaves as approximately normal or as a sum of squared normals.

Function The `power_Gamma()` function is used to compute the power or estimate sample sizes to obtain a target power when testing for differences among two sample means when the data follow a gamma distribution. This function used a parametric bootstrap method addressed by Chang et al. (2011). The arguments for `power_Gamma()` are as follows:

```
power_Gamma(n1 = NULL, n2 = NULL, power = NULL, sig.level = 0.05,
            mu1 = NULL, mu2 = NULL, gmu1 = NULL, gmu2 = NULL,
            trials = 100, M = 10000, equal.sample = TRUE, equal.shape = NULL)
```

Sample sizes for each group are set with `n1` and `n2`. If sample sizes for both groups are equal, the argument `equal.sample` should be set to `TRUE`, and only a value for `n1` or `power` needs to be specified. If sample sizes are unequal, `equal.sample` should be set to `FALSE`, and values for both `n1` and `n2` must be specified. The target power of the test is set with `power`, and the significance level is set with `sig.level` (default value of 0.05). Only one of the parameters of `n1`, `n2`, or `power` can be set as `NULL`. The parameter set as `NULL` will be estimated based on the other parameter values. The arithmetic means for each group are indicated by `mu1` and `mu2`, while `gmu1` and `gmu2` denote the geometric mean for each group, respectively. Option `trials` specifies the number of trials in the simulation, and the number of generated samples in every single trial is identified by `M`. A small number of trials (e.g., using the default value 100) is recommended to get a rough estimate of power or sample size since the computational time is dependent upon the number of trials in the simulation. To obtain a reasonable result, a greater value (e.g., 10000) should be used for both `trials` and `M`. The assumption of equal shape parameters should be tested before the comparison of two sample means if `equal.shape` is set as `NULL` (default value is `NULL`). Otherwise, the test to determine equal shape is skipped (when `equal.shape` is set to be `TRUE` or `FALSE`).

For example, Schickedanz and Krause (1970) presented the weekly rainfall data for the seasons of fall and winter. The arithmetic/geometric means are 0.3684/0.2075 for winter (n = 57) and 0.7635/0.3630 for fall (n = 51). Using a significance level of 0.05, the power can be calculated as follows:

```
set.seed(1)
power_Gamma(n1 = 57, n2 = 51, power = NULL, sig.level = 0.05,
            mu1 = 0.3684, mu2 = 0.7635, gmu1 = 0.2075, gmu2 = 0.3630,
            trials = 100, M = 1000)
```

The estimated power is 1.00, which matches the result in Schickedanz and Krause (1970).

Application of PASSED

In this section, we provide an example power analysis and sample size calculation implemented with **PASSED**. We propose a hypothetical study to test an intervention protocol designed to reduce the percentage of residents at nursing facilities who develop new or worsening pressure ulcers, known as bedsores.

The Skilled Nursing Facility Quality Reporting Program (SNF-QRP) provider dataset contains information on pressure ulcer rates among nursing home facilities across the US. In this scenario, half of the participating nursing homes will implement the intervention protocol (treatment group), and the other half will constitute a control group, without a change in protocol, to determine if the new intervention reduces rates of pressure ulcers. We consider the following hypotheses for the study:

- H_0: There is no difference in pressure ulcer rates among nursing home facilities between control and treatment groups.
- H_a: There is a difference in pressure ulcer rates among nursing home facilities between control and treatment groups.

Sample Size Determination

In this example, we use the mean and standard deviation of the SNF-QRP variable, "percentage of SNF residents with pressure ulcers that are new or worsened" for the control group `mu1` and `sd1`, 0.0174 and 0.0211, respectively. A 25% decrease in the proportion of patients that develop new or worsening pressure ulcers is considered significant and results in the target alternative mean, `mu2`, equal to 0.0131. To determine the appropriate number of facilities necessary in the control and treatment groups, we first use `power_Beta` to estimate the minimum sample size with target power equal to 0.8. The `power_Beta` is chosen because this proportion is defined on the interval [0, 1] and right-skewed. The default value of `link.type` is used, `trials` is set at 1000, and equal precision in the control and treatment groups is assumed. This analysis can be fine-tuned through additional iterations of `power_Beta` by modifying the number of trials. The output is given below:

```
library(PASSED)
set.seed(1)
power_Beta(mu1 = 0.0174, sd1 = 0.0211, mu2 = 0.0131, power = 0.8,
           link.type = "logit", trials = 1000, equal.precision = TRUE)

     Two-sample Beta Means Tests (Equal Sizes) (logit link, equal precision)

              N = 151
            mu1 = 0.0174
            mu2 = 0.0131
            sd1 = 0.0211
      sig.level = 0.05
          power = 0.826

NOTE: N is number in *each* group
```

The obtained result indicates that 302 nursing home facilities (151 facilities for each group) are necessary to demonstrate the difference between pressure ulcer rates among the control and treatment groups, with a significance level of 0.05 and power of 0.80.

Comparison with T-Test

To further assess the appropriate number needed in the control and treatment groups, we then use 0.0120 to 0.0140 to evaluate a range of target means that encompass the target's alternative mean of 0.0131, with expected sample sizes of over 100 nursing homes per group. As a comparison, we also calculate the power using a two-sided t-test under the same scenario, using the function `power_Normal`. The true difference in means, `delta`, is set as the difference of `mu1` and `mu2`, and the alternative standard deviation is assumed to be equal to `sd1`. The output for this example is displayed below, assuming equal precision.

```
# Set seed for the simulation below
set.seed(1)
Ex1 <- mapply(
  function(mu2, sample_size){
    Betapower <- power_Beta(mu1 = 0.0174, sd1 = 0.0211,
                            mu2 = mu2, n1 = sample_size,
                            link.type = "logit", trials = 1000,
                            equal.precision = TRUE)
    Normalpower <- power_Normal(delta = (0.0174 - mu2), n1 = sample_size,
                           sd1 = 0.0211, sd2 = 0.0211)
    return(c(Betapower$power,
             round(Normalpower$power,3),
             sample_size,
             mu2,
             0.0174))
  },
  # Range of mu2 was set as [0.0120, 0.0140] by 0.0010
  rep(seq(0.0120, 0.0140, 0.0010), 5),
  # Range of sample size was set as [100, 200] by 25
  rep(seq(100, 200, 25), rep(3, 5))
)
# Reform the output
Ex1 <- as.data.frame(t(Ex1))
# Set column names
colnames(Ex1) <- c("Power (Beta)",
                   "Power (Normal)",
                   "Sample Size",
                   "mu2",
                   "mu1")
# Display the results
Ex1
```

```
   Power (Beta) Power (Normal) Sample Size   mu2    mu1
1         0.813          0.437         100 0.012 0.0174
2         0.623          0.311         100 0.013 0.0174
3         0.435          0.204         100 0.014 0.0174
4         0.891          0.522         125 0.012 0.0174
5         0.743          0.375         125 0.013 0.0174
6         0.488          0.245         125 0.014 0.0174
7         0.954          0.598         150 0.012 0.0174
8         0.821          0.436         150 0.013 0.0174
9         0.576          0.285         150 0.014 0.0174
10        0.979          0.665         175 0.012 0.0174
11        0.872          0.494         175 0.013 0.0174
12        0.609          0.324         175 0.014 0.0174
13        0.986          0.723         200 0.012 0.0174
14        0.914          0.548         200 0.013 0.0174
15        0.708          0.362         200 0.014 0.0174
```

When equal precision cannot be assumed, `equal.precision` is set to `FALSE`, and an input value for `sd2` is required. To demonstrate unequal precision, the previous example is rerun with `equal.precision=FALSE` and `sd2=0.03`. The output is provided below.

```
# Set seed for the simulation below
```

```
set.seed(1)
Ex2 <- mapply(
  function(mu2, sample_size){
    Betapower <- power_Beta(mu1 = 0.0174, sd1 = 0.0211, sd2 = 0.030,
                            mu2 = mu2, n1 = sample_size,
                            link.type = "logit", trials = 1000,
                            equal.precision = FALSE)
    Normalpower <- power_Normal(delta = (0.0174 - mu2), n1 = sample_size,
                           sd1 = 0.0211, sd2 = 0.030)
    return(c(Betapower$power,
             round(Normalpower$power,3),
             sample_size,
             mu2,
             0.0174))
  },
  # Range of mu2 was set as [0.0120, 0.0140] by 0.0010
  rep(seq(0.0120, 0.0140, 0.0010), 5),
  # Range of sample size was set as [100, 200] by 25
  rep(seq(100, 200, 25), rep(3, 5))
)
# Reform the output
Ex2 <- as.data.frame(t(Ex2))
# Set column names
colnames(Ex2) <- c("Power (Beta)",
                   "Power (Normal)",
                   "Sample Size",
                   "mu2",
                   "mu1")
# Display the results
Ex2
```

```
   Power (Beta) Power (Normal) Sample Size   mu2    mu1
1         0.985          0.310         100 0.012 0.0174
2         0.942          0.222         100 0.013 0.0174
3         0.879          0.150         100 0.014 0.0174
4         0.999          0.374         125 0.012 0.0174
5         0.986          0.266         125 0.013 0.0174
6         0.959          0.177         125 0.014 0.0174
7         1.000          0.435         150 0.012 0.0174
8         0.999          0.310         150 0.013 0.0174
9         0.991          0.204         150 0.014 0.0174
10        1.000          0.493         175 0.012 0.0174
11        1.000          0.353         175 0.013 0.0174
12        0.999          0.230         175 0.014 0.0174
13        1.000          0.546         200 0.012 0.0174
14        1.000          0.394         200 0.013 0.0174
15        0.999          0.257         200 0.014 0.0174
```

The results indicate small differences between the power of a two-sided t-test with equal and unequal standard deviations, while the power from `power_Beta` changes drastically without the equal precision assumption. Unlike normally distributed random variables, the beta distribution is more sensitive to the assumption of equal precision parameters. Figure 1 displays the comparison of probability density functions for beta distributed random variables with and without the equal precision assumption and the comparison for the analogous normally distributed variables with and without equal standard deviations.

Summary

This example demonstrates the use of `power_Beta` and `power_Normal`, each with equal and unequal precision parameters, to perform power analyses and sample size calculations. Since a simulation method is used within the function `power_Beta`, the computational time is dependent upon the number of trials in the simulation. It is suggested that a starting value be used, such as 100, to determine an initial range for the other parameters (e.g., range of `mu2`). Once an appropriate range of values

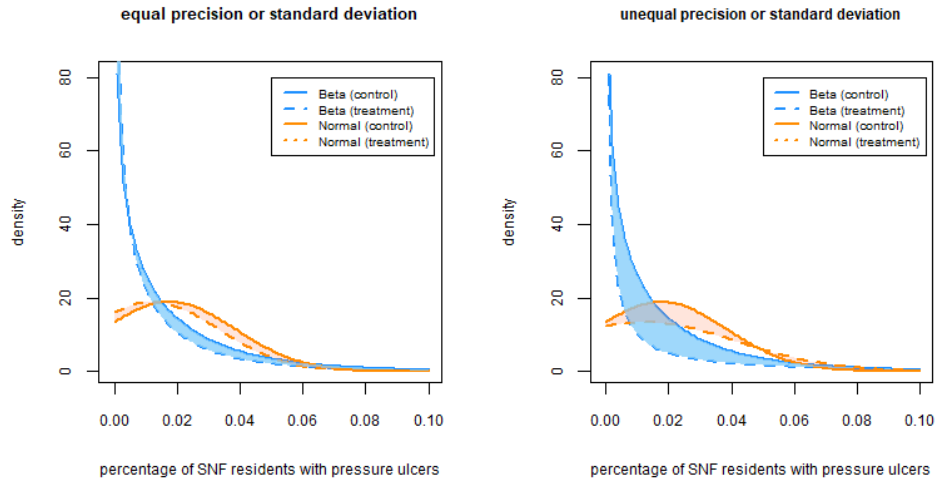

Figure 1: Comparison of equal and unequal precision or standard deviation parameters (`mu2` is assumed to be 0.0120).

is determined, the number of trials should be increased (e.g., `trials=1000`) to output more precise power and sample size estimates.

Summary and Discussion

Multiple packages are available in R to perform power analyses, including **pwr**, **MESS**, **WebPower**, and the base R **stats** package. However, these packages do not provide a comprehensive power analysis toolkit capable of calculating power or sample sizes for the test of two-sample means or ratios when the outcomes have a beta, gamma, or Poisson distribution.

The **PASSED** package extends the current power analysis functions available in R. Seven functions are provided for corresponding distributions, applying either theoretical formulas or simulation algorithms. All functions have the ability to obtain the statistical power or estimate minimum sample sizes. In particular, the formula-based approaches also support calculations for other parameters such as means and proportions. As for the simulation-based methods, users are able to customize each analysis with options to set the number of trials in the simulation and specify the assumptions for the tests. An example of how to implement and customize the functions is provided in Section Application of **PASSED**. The **PASSED** package provides a simple, one-package solution for sample size and power calculations for a wide variety of common and specialty distributions encountered in clinical research.

Computational Details

The results in this paper were obtained using R 4.0.2 and **betareg** 3.0.0. R itself and all packages used are available from the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/`.

Acknowledgments

The authors gratefully acknowledge Dr. Richard Madsen, Professor Emeritus of Statistics University of Missouri - Columbia, for his contribution and interest in this topic. Funding for this project was provided by NIH R34AR074209A (Co-I Leary) and the Thompson Laboratory for Regenerative Orthopaedics, University of Missouri - Columbia.

Appendix

Derivation of Power Calculation Formulae for Poisson Distribution

Let x_{ij} be the number of events during the necessary study time t_i from the jth subject in the ith treatment group, $j = 1, ..., n_i$, $i = 1, 2$. It is assumed that x_{ij} are Poisson random variables with rate λ_i such that the probability function of x_{ij} is

$$P\left(x_{ij}\right) = \frac{t_i\lambda_i e^{t_i\lambda_i}}{x_{ij}!},$$

where $i = 1, 2$. Then, the total number of events in each group, denoted as X_1 and X_2, also follows a Poisson distribution:

$$X_i \sim Poisson(\lambda_i t_i n_i).$$

For the hypothesis tests:

$$H_0 : \lambda_2/\lambda_1 = R \; vs. \; H_a : \lambda_2/\lambda_1 = R' > R(one - sided)$$

and

$$H_0 : \lambda_2/\lambda_1 = R \; vs. \; H_a : \lambda_2/\lambda_1 = R' \neq R(two - sided),$$

where R denotes the pre-specified ratio of two Poisson rates. Gu et al. (2008) derives a test statistic W_5, which is asymptotically distributed as a standard normal under the null hypothesis above,

$$W_5 = \frac{2(\sqrt{X_2 + 3/8} - \sqrt{Q(X_1 + 3/8)})}{\sqrt{1 + Q}},$$

where $Q = R/d$ and $d = t_1/t_2$. Then, the critical region of the one-sided test is

$$W_5 = \frac{2(\sqrt{X_2 + 3/8} - \sqrt{Q(X_1 + 3/8)})}{\sqrt{1 + Q}} \geq z_{1-\alpha}. \tag{4}$$

To calculate the power under $H_a : \; \lambda_2/\lambda_1 = R' > R$ at significance level α, let $c = R/R'$ and multiply both sides of Equation 4 by $\sqrt{1 + Q}$, which is greater than 0 as that

$$2(\sqrt{X_2 + 3/8} - \sqrt{Q(X_1 + 3/8)}) \geq z_{1-\alpha}\sqrt{1 + Q}. \tag{5}$$

Add $-2(\sqrt{Q/c} - \sqrt{Q})\sqrt{X_1 + 3/8}$ to both sides of Equation 5 for the inequality,

$$2(\sqrt{X_2 + 3/8} - \sqrt{Q/c(X_1 + 3/8)}) \geq z_{1-\alpha}\sqrt{1 + Q} - 2(\sqrt{Q/c} - \sqrt{Q})\sqrt{X_1 + 3/8}. \tag{6}$$

Then, divide both sides of Equation 6 by $\sqrt{1 + Q/c}$, greater than 0. It follows that

$$\frac{2(\sqrt{X_2 + 3/8} - \sqrt{Q/c(X_1 + 3/8)})}{\sqrt{1 + Q/c}} \geq \frac{z_{1-\alpha}\sqrt{1 + Q} - 2(\sqrt{Q/c} - \sqrt{Q})\sqrt{X_1 + 3/8}}{\sqrt{1 + Q/c}}. \tag{7}$$

Under the alternative hypothesis, the left-hand side of Equation 7 is asymptotically normal distributed (Gu et al., 2008). Accordingly, the type II error, β, can be derived as:

$$\begin{aligned}
\beta &= P(H_0|H_a) \\
&= P(X < \frac{z_{1-\alpha}\sqrt{1 + Q} - 2(\sqrt{Q/c} - \sqrt{Q})\sqrt{X_1 + 3/8}}{\sqrt{1 + Q/c}}|H_a) \\
&= \Phi(\frac{z_{1-\alpha}\sqrt{1 + Q} - 2(\sqrt{Q/c} - \sqrt{Q})\sqrt{X_1 + 3/8}}{\sqrt{1 + Q/c}})
\end{aligned}$$

Incorporating $Q = R/d$, $c = R/R'$ and $X_1 = \lambda_1 t_1 n_1$, and collecting items

$$\begin{aligned}
\beta &= \Phi\left(\frac{z_{1-\alpha}\sqrt{1+R/d} - 2(\sqrt{(R/d)/(R/R')} - \sqrt{R/d})\sqrt{\lambda_1 t_1 n_1 + 3/8}}{\sqrt{1+(R/d)/(R/R')}}\right) \\
&= \Phi\left(\frac{z_{1-\alpha}\sqrt{\frac{R+d}{d}} - 2(\sqrt{\frac{R'}{d}} - \sqrt{\frac{R}{d}})\sqrt{\lambda_1 t_1 n_1 + 3/8}}{\sqrt{\frac{R'+d}{d}}}\right) \\
&= \Phi\left(\frac{z_{1-\alpha}\sqrt{\frac{R+d}{R'}} - 2(1 - \sqrt{\frac{R}{R'}})\sqrt{\lambda_1 t_1 n_1 + 3/8}}{\sqrt{\frac{R'+d}{R'}}}\right) \\
&= \Phi\left(\frac{z_{1-\alpha}C - A\sqrt{B}}{D}\right),
\end{aligned}$$

where $A = 2(1 - \sqrt{\frac{R}{R'}})$,
$B = \lambda_1 t_1 n_1 + 3/8$,
$C = \sqrt{\frac{R+d}{R'}}$,
$D = \sqrt{\frac{R'+d}{R'}}$
So, power can be expressed as

$$\begin{aligned}
Power(W_5) &= 1 - \Phi\left(\frac{z_{1-\alpha}C - A\sqrt{B}}{D}\right) \\
&= \Phi\left(\frac{A\sqrt{B} - z_{1-\alpha}C}{D}\right).
\end{aligned} \tag{8}$$

Moreover, using $z_{power} = \Phi^{-1}(Power)$, Equation 8 can be expressed as:

$$z_{power} = \frac{A\sqrt{B} - z_{1-\alpha}C}{D}. \tag{9}$$

Solving Equation 9 for B,

$$B = \left(\frac{z_{power}D - z_{1-\alpha}C}{A}\right)^2. \tag{10}$$

Since $B = \lambda_1 t_1 n_1 + 3/8$, the sample size calculation formula of one-sided test can be determined by solving Equation 10 for n_1

$$n_1 = \frac{\left(\frac{z_{1-\alpha}C + z_{power}D}{A}\right)^2 - \frac{3}{8}}{\lambda_1 t_1}(one - sided)$$

and the two-sided test can be derived similarly as

$$n_1 = \frac{\left(\frac{z_{1-\frac{\alpha}{2}}C + z_{power}D}{A}\right)^2 - \frac{3}{8}}{\lambda_1 t_1}(two - sided).$$

Bibliography

C. L. Aberson. *Applied power analysis for the behavioral sciences*. Routledge, 2019. URL https://doi.org/10.4324/9781315171500. [p542]

E. Brittain and J. J. Schlesselman. Optimal allocation for the comparison of proportions. *Biometrics*, pages 1003–1009, 1982. URL https://doi.org/10.2307/2529880. [p543]

G. Casella and R. L. Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002. ISBN 978-0-357-75313-2. [p550]

S. Champely, C. Ekstrom, P. Dalgaard, J. Gill, S. Weibelzahl, A. Anandkumar, C. Ford, R. Volcic, and H. De Rosario. pwr: Basic functions for power analysis, 2017. URL https://cran.r-project.org/web/packages/pwr/index.html. [p542, 548]

C.-H. Chang, J.-J. Lin, and N. Pal. Testing the equality of several gamma means: a parametric bootstrap method with applications. *Computational Statistics*, 26(1):55–76, 2011. URL https://doi.org/10.1007/s00180-010-0209-1. [p551]

M. R. Chernick and C. Y. Liu. The saw-toothed behavior of power versus sample size and software solutions: Single binomial proportion using exact methods. *The American Statistician*, 56(2):149–155, 2002. URL https://doi.org/10.1198/000313002317572835. [p550]

S.-C. Chow, H. Wang, and J. Shao. *Sample size calculations in clinical research*. Chapman and Hall/CRC, 2007. URL https://doi.org/10.1002/wics.155. [p542]

N. Cressie and H. Whitford. How to use the two sample t-test. *Biometrical Journal*, 28(2):131–148, 1986. URL https://doi.org/10.1002/bimj.4710280202. [p547]

F. Cribari-Neto and A. Zeileis. Beta regression in R. *Journal of Statistical Software*, 34(2):1–24, 2010. URL https://doi.org/10.18637/jss.v034.i02. [p549]

M. D. Huffman. An improved approximate two-sample poisson test. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 33(2):224–226, 1984. URL https://doi.org/10.2307/2347448. [p546]

C. T. Ekstrøm. *The R primer*. CRC Press USA, 2012. URL https://doi.org/10.1201/9781315154411. [p542, 548]

S. Ferrari and F. Cribari-Neto. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815, 2004. URL https://doi.org/10.1080/0266476042000214501. [p549]

J. L. Fleiss, A. Tytun, and H. K. Ury. A simple approximation for calculating sample sizes for comparing independent proportions. *Biometrics*, pages 343–346, 1980. URL https://doi.org/10.2307/2529990. [p543]

S. Gates, I. Nguyen, M. Del Core, P. A. Nakonezny, H. Bradley, and M. Khazzam. Incidence and predictors of positive intraoperative cultures in primary shoulder arthroplasty following prior ipsilateral shoulder surgery. *JSES International*, 2020. URL https://doi.org/10.1016/j.jseint.2019.12.011. [p544]

K. Gu, H. K. T. Ng, M. L. Tang, and W. R. Schucany. Testing the ratio of two poisson rates. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 50(2):283–298, 2008. URL https://doi.org/10.1002/bimj.200710403. [p546, 547, 556]

A. K. Gupta and S. Nadarajah. *Handbook of Beta Distribution and Its Applications*. CRC Press, New York, 2004. URL https://doi.org/10.1201/9781482276596. [p548]

J. M. Hilbe. *Negative binomial regression*. Cambridge University Press, 2011. ISBN 978-0-521-19815-8. [p544]

Y.-R. Hong, R. G. Salloum, S. Yadav, G. Smith, and A. G. Mainous III. Patient–provider discussion about cancer treatment costs and out-of-pocket spending: Implications for shared decision making in cancer care. *Value in Health*, 2020. URL https://doi.org/10.1016/j.jval.2020.08.002. [p550]

C. Hu, H. Zhou, and A. Sharma. Application of beta-distribution and combined uniform and binomial methods in longitudinal modeling of bounded outcome score data. *The AAPS Journal*, 22, 2020. URL https://doi.org/10.1208/s12248-020-00478-5. [p548]

M. K. Hutchinson and M. C. Holtman. Analysis of count data using poisson regression. *Research in nursing & health*, 28(5):408–418, 2005. URL https://doi.org/10.1002/nur.20093. [p546]

S.-L. Jan and G. Shieh. Optimal sample sizes for welch's test under various allocation and cost considerations. *Behavior research methods*, 43(4):1014–1022, 2011. URL https://doi.org/10.3758/s13428-011-0095-7. [p548]

S. Jones, S. Carley, and M. Harrison. An introduction to power and sample size estimation. *Emergency Medicine Journal*, 20(5):453–458, 2003. URL https://doi.org/10.1136/emj.20.5.453. [p542]

M. Kohl. *MKmisc: Miscellaneous functions from M. Kohl*, 2021. URL https://www.stamats.de. R package version 1.8. [p542]

B. LeBeau. *Power Analysis by Simulation using R and simglm*, 2019. URL https://doi.org/10.17077/f7kk-6w7f. [p542]

C. B. G. LEITE, L. V. Ranzoni, P. N. Giglio, M. B. Bonadio, L. D. P. Melo, M. K. Demange, and R. G. Gobbi. Assessment of the use of tranexamic acid after total knee arthroplasty. *Acta Ortopédica Brasileira*, 28(2):74–77, 2020. URL https://doi.org/10.1590/1413-785220202802228410. [p543]

C. Luan, D.-T. Xu, N.-J. Chen, F.-F. Wang, K.-S. Tian, C. Wei, and X.-B. Wang. How to choose kinematic or mechanical alignment individually according to preoperative characteristics of patients? *BMC Musculoskeletal Disorders*, 21(1):1–6, 2020. URL https://doi.org/10.1186/s12891-020-03472-2. [p547]

K. Pearson. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50 (302):157–175, 1900. URL https://doi.org/10.1080/14786440009463897. [p543]

D. Plessl, B. Salomon, A. Haydel, C. Leonardi, A. Bronstone, and V. Dasa. Rapid versus standard recovery protocol is associated with improved recovery of range of motion 12 weeks after total knee arthroplasty. *The Journal of the American Academy of Orthopaedic Surgeons*, 2020. URL https://doi.org/10.5435/JAAOS-D-19-00597. [p542, 546]

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL https://www.R-project.org/. ISBN 3-900051-07-0. [p542, 550]

Scherer. *Package 'samplesize'*, 2016. URL https://cran.r-project.org/web/packages/samplesize/index.html. [p542]

P. T. Schickedanz and G. F. Krause. A test for the scale parameters of two gamma distributions using the generalized likelihood ratio. *Journal of applied meteorology*, 9(1):13–16, 1970. URL https://doi.org/10.1175/1520-0450(1970)009<0013:ATFTSP>2.0.CO;2. [p552]

W.-K. Shiue and L. J. Bain. Experiment size and power comparisons for two-sample poisson tests. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 31(2):130–134, 1982. URL https://doi.org/10.2307/2347975. [p546]

W.-K. Shiue and L. J. Bain. A two-sample test of equal gamma distribution scale parameters with unknown common shape parameter. *Technometrics*, 25(4):377–381, 1983. URL https://doi.org/10.1080/00401706.1983.10487901. [p551]

W.-K. Shiue, L. J. Bain, and M. Engelhardt. Test of equal gamma-distribution means with unknown and unequal shape parameters. *Technometrics*, 30(2):169–174, 1988. URL https://doi.org/10.1080/00401706.1988.10488364. [p551]

M. Smithson and J. Verkuilen. A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11(1):54–71, 2006. URL https://doi.org/10.1037/1082-989X.11.1.54. [p549]

H. C. Thode. Power and sample size requirements for tests of differences between two poisson rates. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(2):227–230, 1997. URL https://doi.org/10.1111/1467-9884.00078. [p546]

E. Zhang, V. Q. Wu, S.-C. Chow, and H. G. Zhang. *Package 'TrialSize'*, 2013. URL https://cran.r-project.org/web/packages/TrialSize/index.html. [p542]

Z. Zhang and Y. Mai. *WebPower: Basic and Advanced Statistical Power Analysis*, 2018. URL https://CRAN.R-project.org/package=WebPower. R package version 0.5.2. [p542]

H. Zhu and H. Lakkis. Sample size calculation for comparing two negative binomial rates. *Statistics in medicine*, 33(3):376–387, 2014. URL https://doi.org/10.1002/sim.5947. [p544, 545]

Jinpu Li
University of Missouri
USA
E-mail: lijinp@health.missouri.edu

Ryan P. Knigge
University of Minnesota
USA
E-mail: knigg008@umn.edu

Kaiyi Chen
University of Missouri
USA
E-mail: kcvpc@mail.missouri.edu

Emily V. Leary, Ph.D.
Thompson Laboratory for Regenerative Orthopaedics
Department of Orthopaedic Surgery
University of Missouri
USA
E-mail: learye@health.missouri.edu