

Sample Size Calculations for Vegetation Studies

Пресметки на големина на примерок во вегетациски студии

Mohammad Mousaei Sanjerehei

Department of Plant Protection, Yazd Branch, Islamic Azad University, Yazd, Iran.

Abstract



A nonsignificant result may be obtained in a study despite that there is an ecologically important and strong effect and difference. This lack of significance may be due to a small sample size and low statistical power of the test. Therefore an adequate sample size is required for accurately estimating and comparing vegetation characteristics and obtaining an acceptable power for the test. This paper presents sample size equations and explains, with examples, how to use the equations for estimating vegetation quantitative characteristics, proportion tests, comparing between two independent samples and two paired samples, correlation analysis, simple and multiple linear regression and logistic regression, chi-square test using contingency tables and analysis of variance (ANOVA).

Key words: sample size, statistical power, vegetation characteristics

Апстракт

Во една студија може да се добие несигнификантен резултат иако постои еколошки значаен и силен ефект и разлика. Отсуството на сигнификантност може да се должи на мал примерок и ниска статистичка моќ на тестот. Затоа, за прецизна проценка и споредба на карактеристиките на вегетацијата и обезбедување на прифатлива моќ на тестот е потребна адекватна големина на примерокот. Овој труд ги прикажува равенките за големина на примерокот и објаснува, преку примери, како да се користат равенките за проценка на квантитативните карактеристики на вегетацијата, тестот на пропорции, споредба помеѓу два независни примероци и два спарени примероци, анализи на корелација, еднофакторски и повеќекфакторски праволиниски регресији и логистички регресији, тест на хи-квадрати со користење на табели на контингенција и анализа на варијанса (ANOVA).

Клучни зборови: големина на примерок, статистичка моќ, карактеристики на вегетација

Submitted: 05.04.2021

Accepted: 12.08.2021

Introduction

An adequate sample size can result in a reliable and accurate estimation of vegetation attributes such as cover, biomass, density, diameter, etc. Sample size in vegetation research is the number of sampling units (e.g., point, quadrat, transect), number of plants, seeds, leaves, etc. required for sampling vegetation.

The sample size depends on the significance level, power level and effect size. Significance level (α) or Type I error occurs when a null hypothesis that is actually true is rejected. Researchers have frequently used 1%, 5%, and 10% significance levels for plant studies (Federal register Vol. 44, No 50; 1979; Stauffer 1983; Mosley et al.1989; Hofmann &Ries 1990). The lower the significance level, the larger the sample size.

Type II error (β) occurs when the researcher fails to reject a null hypothesis that is false. Power of a statistical test, defined as $1 - \beta$, is the probability that a null hypothesis is rejected when it is actually false. An increase in power of a test results in an increase in the sample size. The power levels of 0.8 and 0.9 are usually used in vegetation studies for calculating sample size (Peek et al. 2003; Moffet 2009; Conser& Connor 2009).

Effect size represents the magnitude of an effect of interest to be detected by a test. For example, for a comparison of two independent means, an effect size may be defined as the difference between the two means divided by standard deviation, or the difference between two proportions for a comparison of two independent proportions. An increase in the effect size can lead to a decrease in the sample size.

Sample size can be significantly influenced by standard deviation. Standard deviation can be obtained from previous studies or a preliminary sample. The larger the standard deviation, the larger the required sample size.

This study presents several sample size equations and explains, with examples, how to use the equations based on the type of vegetation study and the statistical test. The sample size equations in this study can be applied for estimation of quantitative characteristics of vegetation, proportion tests, comparing two independent samples and two dependent samples, linear and logistic regression, chi-square test using contingency tables, analysis of variance (ANOVA) and correlation analysis.

G* Power (Erdfelder et al. 1996; Faul et al. 2007), PS (Dupont & Plummer 1997) and PASS (StataCorp, L.P. 2013) are the software mostly used for determination of sample size. Several statistical software for sample size estimation were reviewed by Thomas & Krebs (1997).

Methods

1. Sample size calculation for quantitative variables

When a researcher wishes to determine the mean value of a quantitative variable such as biomass and density in a population, the following equation is used to calculate the required sample size:

$$N = \frac{Z_{\alpha/2}^2 \times S^2}{d^2}$$

where $Z_{\alpha/2}$ is standard normal variate. It is 1.96 at 5% significance level (α) and 2.576 at 1% significance level. S is the standard deviation of the quantitative variable taken from previous studies. d is the precision (absolute error), defined as the desired difference between the population mean and the sample mean. The difference of the sample mean from the population mean can be selected to be 20, 10, 5 or 1% of the sample mean (Federal register Vol. 44, No 50; 1979; Stauffer 1983; Mosley et al. 1989).

Example

Suppose a researcher wishes to estimate the mean biomass of a plant species in a stand using quadrat sampling. According to previous studies, the standard deviation of the biomass is 24. The researcher wants to calculate the sample size at 5% significance level and a precision of 9 grams of either side (more or less than the mean biomass). Then the sample size will be

$$N = \frac{1.96^2 \times 24^2}{9^2} = 27.3$$

If the sample size obtained using the above formula is large (e.g., ≥ 30) then it is a reliable estimate of the sample size (Stauffer 1983). But if the calculated sample size is low (< 30), then the sample size should be calculated through iterations using $t_{\alpha/2}$ instead of $Z_{\alpha/2}$ in the formula. $t_{\alpha/2}$ is obtained using t -distribution table based on the given significance level and the degrees of freedom ($n-1$).

In the above example, by rounding up, $N= 28$. Then, the degree of freedom is $28-1=27$, and the $t_{\alpha/2}$ at $\alpha = 0.05$ and $df= 27$ is 2.052, therefore

$$N = \frac{2.052^2 \times 24^2}{9^2} = 29.9$$

For $N= 30$, df is 29, and $t_{\alpha/2}$ at $\alpha = 0.05$ is 2.045, then

$$N = \frac{2.045^2 \times 24^2}{9^2} = 29.7$$

which is approximately close to the previous estimate. Therefore 30 plots have to be measured for estimating biomass.

The above sample size formula can be used to calculate the required number of quadrats for estimating plant biomass, production and density, number of plant individuals for estimating plant height, seed production, rooting depth and nutrients (e.g., Nitrogen), number of traps for estimating seed production and litter, and number of transects for estimating cover based on contact sampling.

2. Sample size calculation for proportion

Suppose the objective is to estimate proportion of cover for a plant species by points. Then the following equation should be used for calculating the sample size

$$N = \frac{Z_{\alpha/2}^2 \times p (1 - p)}{d^2}$$

where $Z_{\alpha/2}$ is standard normal variate ($Z_{\alpha/2}=1.96$ at 5% and $Z_{\alpha/2}= 2.576$ at 1% Type I error), p is the expected proportion in population obtained from previous studies, d is the absolute error (precision) and can be selected 20, 10, 5 or 1% depending on the study design (Federal register Vol. 44, No 50; 1979; Croy & Dix 1984; Mosley et al. 1989).

Example

A researcher wants to estimate the percentage cover of a plant species in a stand using located points. The cover of the plant species is assumed to be near 20%. The required sample size at 0.05 significance level and precision of 5% for estimating the percentage cover will be

$$N = \frac{1.96^2 \times 0.2 (1 - 0.2)}{0.05^2} = 246$$

The above sample size equation can be used to determine the proportion of quadrats occupied by a plant species in a population (frequency), the proportion of plants associated with other organisms (such as fungi and bacteria species or epiphytes), proportion of germinating seeds in a plant species and proportion of plants having a disease in a population.

3. Sample size calculation for comparing two independent (unpaired) samples with quantitative data

Suppose a researcher wants to determine the effects of grazing on the biomass of plants by comparing the mean biomass of plants between the grazed plots and ungrazed (exclosure) plots. In such case, independent sample t-test can be used to compare the two means to see if the difference is significant. The sample size is then calculated for each group using the following equation:

$$N = \frac{m + 1}{m} \times \frac{S^2 (Z_{\alpha/2} + Z_{\beta})^2}{d^2}$$

Where $Z_{\alpha/2}$ is standard normal variate for level of significance. $Z_{\alpha/2}=1.96$ at 5% and $Z_{\alpha/2}=2.576$ at 1% significance level for a two-tailed test. For a one-tailed test, $Z_{\alpha/2}$ is replaced by Z_{α} ($Z_{\alpha}=1.645$ at $\alpha =5\%$ and $Z_{\alpha}= 2.326$ at $\alpha =1\%$), Z_{β} is standard normal variate for power of study. For 80% power (20% Type II error), Z_{β} is 0.842 and for 90% power (10% Type II error) it is 1.282. S is the standard deviation obtained from previous studies and is calculated as follows if equal variances are assumed.

$$S = \sqrt{\frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

S_1 and S_2 are the standard deviation of group1 and 2 respectively, and n_1 and n_2 are respectively the size of group 1 and 2. If the size of both groups are equal, then $S = \sqrt{(S_1^2 + S_2^2)/2}$.

d is the difference between the two means ($\bar{x}_1 - \bar{x}_2$).

m is the ratio of size of group 1 to group 2. If the two groups are equal sized ($n_1 = n_2$), then $m=1$. The effect size in this equation can be defined as d/S (Cohen 1988).

Example

Suppose the difference between the mean biomass of plants in ungrazed (exclosure) plots and the mean biomass of plants in grazed plots may be about 18 g with standard deviation of 14 according to previous studies. Then, the required sample size for testing the difference at 0.01 significance level and 80% power calculated as follows:

$$N = \frac{2(14)^2 \times (2.576 + 0.842)^2}{18^2} = 15$$

The exact sample size should be calculated through iterations as explained before. In this case $t_{\alpha/2}$ and t_{β} are used instead of $Z_{\alpha/2}$ and Z_{β} in the equation. $t_{\alpha/2}$ is obtained from t-table based on the given significance level and the degree of freedom (n_1+n_2-2).

t_{β} is obtained from t-table based on the given power ($1-\beta$) and the degree of freedom (n_1+n_2-2). For $N=15$, the $df= 15+15-2=28$, and at $\alpha=0.01$, $t_{\alpha/2} =2.763$, and for power=0.8, $t_{\beta}=0.855$, then

$$N = \frac{2(14)^2 \times (2.763 + 0.855)^2}{18^2} = 16$$

For $N=16$, df is 30 and $t_{\alpha/2} =2.75$ and $t_{\beta} =0.854$.

$$N = \frac{2(14)^2 \times (2.75 + 0.854)^2}{18^2} = 16$$

Because the sample sizes in the two continuous calculations are equal, the iteration is stopped.

Therefore 16 plots are required for each grazed and ungrazed biomass. If unequal variances are assumed and $n_1=n_2$, the sample size is calculated as

$$N = \frac{(S_1^2 + S_2^2) (Z_{\alpha/2} + Z_{\beta})^2}{d^2}$$

If unequal variances are assumed and $n_1 \neq n_2$, the sample size for each group is calculated as

$$N_1 = (S_1^2 + \frac{S_2^2}{n_2}) \times \frac{(Z_{\alpha/2} + Z_{\beta})^2}{d^2}$$

$$\text{and } N_2 = (S_2^2 + \frac{S_1^2}{n_1}) \times \frac{(Z_{\alpha/2} + Z_{\beta})^2}{d^2}$$

The degree of freedom is calculated as

$$v = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{(\frac{S_1^2}{n_1})^2}{n_1-1} + \frac{(\frac{S_2^2}{n_2})^2}{n_2-1}}$$

The above-mentioned equation can be used to calculate the required number of plots or individual plants for comparing the attributes such as biomass, seed production and nutrients between the plants which are and are not associated with other organisms (e.g., fungus, epiphytes), the plants which are and are not exposed to pollutants, the plants which are and are not manured and between the plants in two stands.

4. Sample size calculation for comparing two unpaired (independent) groups with qualitative data (two independent proportions)

A researcher is going to determine the effects of gibberellins and cytokinin hormones on percentage germination of seeds of a plant species. To do this, two groups of seeds were selected; one group was treated with gibberellins and one group with cytokinin. The proportions of germinated seeds in both groups were compared to detect any significant difference. Then, the following formula can be used to calculate necessary sample size

$$N = \frac{m + 1}{m} \times \frac{(Z_{\alpha/2} + Z_{\beta})^2 p (1 - p)}{(p_1 - p_2)^2}$$

Where $Z_{\alpha/2}$ and Z_{β} are standard normal variates as explained before. m is the size of group 1 to group 2 and is equal to 1 when the size of both groups are equal. p_1 and p_2 are respectively the proportion of events in group 1 and 2. p is the average proportion of the events = (proportion of the event in group1 + proportion of the event in group2)/2. The effect size can be defined as $p_1 - p_2$, p_2/p_1 or the odds ratio.

Example

A researcher expects that 80% of the seeds treated to gibberellins and 40% of the seeds treated to cytokinin will germinate within a specified time. Then

the required sample size for testing the difference in proportions of germination between the two groups with 5% significance level and 90% power is;

$$P = (0.8 + 0.4)/2 = 0.6$$

$$N = \frac{2 (1.96 + 1.282)^2 \times 0.6(1 - 0.6)}{(0.8 - 0.4)^2} = 32$$

Therefore 32 seeds are required for group1 and 32 for group2. If the size of group1 is designed to be 1.3 times the size of group 2, then the value of $(m + 1)/m = (1.3+1)/1.3=1.77$ is used (instead of 2) in the above equation for calculation of sample size for group 2 and $[(1/1.3) + 1]/(1/1.3) = 2.3$ is used for calculating the sample size for group 1. The above equation can be used to calculate the required number of seeds for testing the effects of factors such as hormones, water etc., on seed germination.

5. Sample size calculation for comparing two paired (dependent) groups with quantitative data

Suppose a researcher wishes to compare the changes in the biomass of plants in fixed plots before and one year after the fire to see if the changes are significant. Then the following formula can be used to calculate the required sample size;

$$N = \frac{S^2 (Z_{\alpha/2} + Z_{\beta})^2}{d^2}$$

$Z_{\alpha/2}$ and Z_{β} are explained in the previous sections. S is the standard deviation of the differences, and d is the difference between two means before and after the perturbation. The effect size can be defined as d/S .

Example

A researcher assumes that the mean difference in the biomass of plants in fixed plots before and one year after the fire will be close to 8 kg and the standard deviation of the differences will be close to 6. The required sample size at 0.01 significance level and with 90% power is:

$$N = \frac{(2.576 + 1.282)^2 \times 6^2}{8^2} = 9$$

Because the sample size is low (< 30), the exact sample size can be calculated through iterations using $t_{\alpha/2}$ and t_{β} instead of $Z_{\alpha/2}$ and Z_{β} in the equation based on $n-1$ degree of freedom as explained before. The calculated sample size after four iterations will be 12.

The above equation can be used to calculate the adequate number of fixed plots or number of plant individuals to detect the changes in biomass, production, density, rooting depth, seed production and height of plants before and after perturbations

such as grazing, fire, drought and treatments (e.g., manures, watering leaf pruning, growth hormones)

6. Sample size calculation for simple linear regression model

The relationship between a vegetation attribute (dependent variable) such as biomass, cover, density etc. and an environmental variable (independent variable) such as precipitation, temperature, soil salinity and soil water content can be determined using a simple linear regression and correlation. The sample size equation for making a simple linear regression is calculated as follows (Dupont & Plummer 1998):

$$N = (t_{\alpha/2} + t_{\beta})^2 \times \left(\frac{S_y^2}{\lambda^2 S_x^2} - 1 \right)$$

This equation must be solved iteratively. $t_{\alpha/2}$ is obtained from t-table at a given significance level with $N-2$ degrees of freedom and t_{β} is obtained from t-table with the desired power $(1-\beta)$ and $N-2$ degrees of freedom. S_x is the standard deviation of the independent variable, S_y is the standard deviation of the dependent variable and λ is the slope of the linear regression line. The slope of the linear regression line can be used as the effect size. The above equation can also be written based on correlation coefficient (r) between the two variables (x, y). Since $r = (\lambda S_x)/S_y$, then

$$N = (t_{\alpha/2} + t_{\beta})^2 \times \left(\frac{1}{r^2} - 1 \right)$$

Example

A researcher wishes to determine the relationship between a plant biomass and soil EC using quadrat sampling through fitting a simple linear regression. The researcher assumes that the standard deviation of the biomass (S_y) = 3, the standard deviation of soil EC (S_x) = 2.5, and the slope of regression line (λ) = 0.6 according to a previous study. Then, the required sample size at $\alpha=0.05$ and power = 0.8 is calculated iteratively as follows:

$$N = (1.96 + 0.842)^2 \times \left(\frac{3^2}{0.6^2 \times 2.5^2} - 1 \right) = 24$$

For $N = 24$, df is $24-2=22$, and at 0.05 significance level, $t_{\alpha/2}$ is 2.074 and for power of 0.8 ($\beta = 0.2$), t_{β} is 0.858, then

$$N = (2.074 + 0.858)^2 \times \left(\frac{3^2}{0.6^2 \times 2.5^2} - 1 \right) = 26$$

For $N = 26$, df is 24, and at $\alpha=0.05$, $t_{\alpha/2}$ is 2.064 and for power of 0.8, t_{β} is 0.857, then

$$N = (2.064 + 0.857)^2 \times \left(\frac{3^2}{0.6^2 \times 2.5^2} - 1 \right) = 26$$

Therefore a sample size of 26 is required. Based on the data, $r = (0.6 \times 2.5)/3 = 0.5$, and therefore sample size can also be calculated based on r using the second formula.

7. Sample size calculation for correlation analysis

The sample size for testing the difference of Pearson correlation coefficient (r) from the constant value of $r = 0$ is calculated as follows

$$N = \left(\frac{Z_{\alpha/2} + Z_{\beta}}{\frac{1}{2} \times \ln \left(\frac{1+r}{1-r} \right)} \right)^2 + 3$$

$Z_{\alpha/2}$ is standard normal variate for significance level ($Z_{\alpha/2}=1.96$ at $\alpha =5\%$ and $Z_{\alpha/2}= 2.576$ at $\alpha =1\%$ for a two-sided test, Z_{α} is used instead of $Z_{\alpha/2}$ for a one-sided test, $Z_{\alpha}=1.645$ at $\alpha =5\%$ and $Z_{\alpha}= 2.326$ at $\alpha =1\%$), Z_{β} is standard normal variate for power of study. For 80% power, $Z_{\beta} = 0.842$ and for 90% power $Z_{\beta} = 1.282$, N is sample size and r is the Pearson correlation coefficient. When the difference is tested between the Pearson r and zero, r is the effect size (Cohen, 1988).

Example

A researcher assumes that the Pearson correlation (r) between a plant biomass and soil Nitrogen may not be more than 0.6. Then the sample size (required number of plots) for detecting the difference of the correlation coefficient from zero at 5% significance level and 80% power is

$$N = \left(\frac{1.96 + 0.842}{\frac{1}{2} \times \ln \left(\frac{1+0.6}{1-0.6} \right)} \right)^2 + 3 = 19$$

The sample size for Pearson correlation (r) can also be calculated using t-distribution as follows;

$$N = \left[(t_{\alpha/2, n-2} + t_{\beta})^2 \times \left(\frac{1}{r^2} - 1 \right) \right] + 2$$

8- Sample size calculation for multiple linear regression

In a multiple linear regression model, the relationship between a dependent variable (Y) and a set of independent variables ($x_1, x_2 \dots x_n$) is modeled. Suppose a researcher wishes to determine the relationship between a quantitative attribute of vegetation (e.g., biomass, cover, density) and a number of predictors (e.g., precipitation, temperature, humidity, elevation) using a multiple linear regression

model. Then, the following formula can be used to calculate sample size;

$$N = \lambda \left(\frac{1 - R_{Y.X}^2}{R_{Y.X}^2} \right)$$

Where $R_{Y.X}^2$ is the squared multiple correlation between Y and X ($x_1, x_2 \dots x_n$).

λ is the noncentrality parameter of the noncentral F distribution and is obtained from Cohen's table (Cohen 1988, P. 448- 455) according to the desired level

of significance (α), power ($1-\beta$), number of predictors (u) and degrees of freedom of the error (denominator of the F ratio) (v).

$v = N - u - 1$, where N is the size of sample. For each of the values of u , λ values for the desired power are provided for the same four values of v (20, 60, 120 and ∞). A part of the Cohen's table for obtaining λ values at 5% significance level is presented below (Table 1).

Table 1. λ values of the F test as a function of power, u , and v , at $\alpha=0.05$. From Cohen (1988)

u	v	Power										
		0.25	0.50	0.60	2/3	0.70	0.75	0.80	0.85	0.90	0.95	0.99
1	20	1.9	4.1	5.3	6.2	6.7	7.5	8.5	9.7	11.4	14.1	20.1
	60	1.7	3.9	4.9	5.8	6.2	7.0	7.9	9.1	10.6	13.2	18.7
	120	1.7	3.8	4.9	5.7	6.2	6.9	7.8	9.0	10.5	13.0	18.4
	∞	1.6	3.8	4.9	5.7	6.2	6.9	7.8	9.0	10.5	13.0	18.4
2	20	2.6	5.7	7.1	8.2	8.9	9.9	11.1	12.6	14.6	17.9	24.9
	60	2.3	5.1	6.4	7.4	8.0	8.9	10.0	11.3	13.2	16.1	22.4
	120	2.3	5.0	6.3	7.2	7.8	8.7	9.7	11.1	12.8	15.7	21.8
	∞	2.2	5.0	6.2	7.2	7.7	8.6	9.6	10.9	12.7	15.4	21.4
3	20	3.2	6.9	8.6	9.9	10.6	11.8	13.2	14.9	17.2	20.9	28.7
	60	2.8	6.0	7.5	8.6	9.3	10.3	11.5	13.0	15.0	18.3	25.1
	120	2.7	5.8	7.3	8.4	9.0	10.0	11.1	12.6	14.5	17.7	24.3
	∞	2.7	5.8	7.2	8.2	8.8	9.8	10.9	12.3	14.2	17.2	23.5
4	20	3.8	8.0	9.9	11.4	12.2	13.5	15.0	16.9	19.5	23.5	32.1
	60	3.3	6.8	8.5	9.7	10.4	11.5	12.8	14.4	16.6	20.1	27.4
	120	3.1	6.6	8.1	9.3	10.0	11.0	12.3	13.9	16.0	19.3	26.3
	∞	3.1	6.4	7.9	9.1	9.7	10.7	11.9	13.4	15.4	18.6	25.2

Interpolation of λ values for a given v is necessary for exact value of λ and is obtained as follows

$$\lambda = \lambda_L - \left[\frac{\frac{1}{v_L} - \frac{1}{v}}{\frac{1}{v_L} - \frac{1}{v_u}} \right] (\lambda_L - \lambda_u)$$

v is a desired trail value (e.g., $v = 20, 60$ or 120). v_L and v_u are the lower and upper value of v and λ_L and λ_u are their respective values. The effect size for sample size calculation in a multiple linear regression can be defined as $R_{Y.X}^2 / (1 - R_{Y.X}^2)$ (Cohen 1988).

Example

Suppose a researcher wishes to determine the relationship between the biomass of a plant species and three climatic variables ($u = 3$) including precipitation, temperature and humidity by plot sampling and fitting a multiple linear regression. Assuming that the model $R_{Y.X}^2 = 0.30$, the required sample size at $\alpha = 0.05$ and power = 0.8 is calculated as follows:

We take the value of $v = 20$ as a trail value. According to the Cohen's table (the above table), λ is 13.2 for $\alpha = 0.05$, power = 0.8, $u = 3$ and $v = 20$. Then the sample size is

$$N = 13.2 \left(\frac{1 - 0.30}{0.30} \right) = 31$$

Which since $v = N - u - 1$, implies that $v = 31 - 3 - 1 = 27$. A more accurate value for N requires reiteration using λ for $v = 27$. Therefore λ is calculated by interpolating between $v = 20$ and $v = 60$ as follows

$$\lambda = 13.2 - \left[\frac{\frac{1}{20} - \frac{1}{27}}{\frac{1}{20} - \frac{1}{60}} \right] (13.2 - 11.5) = 12.54$$

Then

$$N = 12.54 \left(\frac{1 - 0.30}{0.30} \right) = 30$$

This sample size hardly differs from the previous one, and no further iteration is necessary.

9. Sample size calculation for chi-square test using contingency tables

A contingency table is used for detecting association or independence between two or more traits. The sample size for contingency tables can be calculated using the following formula and Cohen's table (Cohen 1988, P. 253-267).

$$N = \frac{N_{0.1}}{100w^2}$$

Where $N_{0.1}$ is the necessary sample size for a given significance level, power and degrees of freedom (u) at $w=0.1$ (read from the Cohen's table). The degrees of freedom for a contingency table with r rows and c columns is $u = (r - 1)(k - 1)$.

w is the nontabulated effect size and is calculated as $w = \sqrt{\chi^2/N}$, where χ^2 is the chi - square value calculated using the data of contingency table and N is the total number of samples. A part of the Cohen's sample size table for contingency tables at $\alpha=0.01$ and $u = 3$ is presented below (Table 2).

Table 2. N to detect w by χ^2 at $\alpha=0.01$ and $u=3$. From Cohen (1988)

Power	w								
	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0.25	544	136	60	34	22	15	11	8	7
0.50	931	233	103	58	37	26	19	15	11
0.60	1101	275	122	69	44	31	22	17	14
2/3	1227	307	136	77	49	34	25	19	15
0.70	1297	324	144	81	52	36	26	20	16
0.75	1412	353	157	88	56	39	29	22	17
0.80	1546	386	172	97	62	46	32	24	19
0.85	1709	427	190	107	68	49	35	27	21
0.90	1925	481	214	120	77	53	39	30	24
0.95	2267	567	252	142	91	63	46	35	28
0.99	2983	746	331	186	119	83	61	47	37

Example

A researcher wants to compare the frequencies of four plant species in two stands by calculating the number of quadrats containing the plant species. Let us assume that in a previous similar study, the chi-square calculated from the data of a 2x4 table was 9 based on a sample of size (N) =60. If the researcher wants to calculate the sample size using the data of the previous study for a new study, the required sample size at $\alpha=0.05$ and 90% power is calculated as follows:

$w = \sqrt{(9/60)} = 0.387$ and $u = (2-1)(4-1)=3$. According to the Cohen's table (the above table), $N_{0.1} = 1925$ at $\alpha=0.01$, power = 0.90, $u = 3$ and $w=0.1$, then the sample size will be

$$N = \frac{1925}{100(0.387)^2} = 128.5$$

The sample size is 129 after rounding up.

10. Sample size calculation for 2x2 contingency table

A 2x2 contingency table is used for detecting association or independence between two traits. The sample size for a 2x2 contingency table can be obtained using the formula in the previous section. The degrees of freedom for a 2x2 table is 1.

The required sample size for a 2x2 contingency table can also be calculated using the following equation

$$N = \frac{(\sqrt{\chi_c^2} + Z_\beta)^2}{w^2}$$

χ_c^2 is the critical value of the chi-square obtained from chi-square distribution table based on a specified significance level and $df = 1$. For 5% significance level and $df = 1$, $\chi_c^2 = 3.84$ and for 1% significance level and $df = 1$, $\chi_c^2 = 6.63$, Z_β is standard normal variate for power and is 0.842 for 80% power (20% Type II error) and 1.282 for 90% power (10% Type II error). w is the effect size and is calculated as $w = \sqrt{\chi^2/N}$, where χ^2 is the chi - square value calculated using the data of 2x2 table and N is the total number of samples.

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

o_i and e_i are the observed and expected value for cell i respectively.

Example

In a study, the association between two plant species (A and B) was tested using chi-square statistic. The data were obtained using located quadrats and summarized in a 2x2 table as follows:

		Species B		
		Present	Absent	
Species A	Present	14	2	16
	Absent	4	10	14
		18	12	N=30

14 quadrats (cell a) contain both species A and B. In 2 quadrats (cell b) species A is present but not B. In 4 quadrats (cell c) species B is present but not A. In 10 quadrats (cell d) neither species A nor B are found. The expected value for each cell of the table is calculated as follows:

$$E(a) = \frac{16 \times 18}{30} = 9.6E(b) = \frac{16 \times 12}{30} = 6.4E(c) = \frac{14 \times 18}{30} = 8.4E(d) = \frac{14 \times 12}{30} = 5.6$$

then

$$\chi^2 = \frac{(14 - 9.6)^2}{9.6} + \frac{(2 - 6.4)^2}{6.4} + \frac{(4 - 8.4)^2}{8.4} + \frac{(10 - 5.6)^2}{5.6} = 10.8$$

$$w = \sqrt{10.8/30} = 0.6$$

Suppose a researcher wishes to test the association between two plant species (A and B) using the data of a 2x2 table and chi-square test. The researcher assumes that the two plant species may be associated and the chi-square will be near 11 using 30 plots according to the previous study. Then, the required sample size at 0.05 significance level and 80% power is

$$w = \sqrt{11/30} = 0.605$$

$$N = \frac{(\sqrt{3.84} + 0842)^2}{0.605^2} = 21.4$$

22 quadrats are required for detecting the association between the two plant species.

11. Sample size calculation for analysis of variance (ANOVA)

a) *k* groups with equal sizes

ANOVA is used to test whether there is a significant difference between the means of two or more groups. The required sample size for ANOVA can be obtained using Cohen's tables (Cohen 1988, p. 381-389) based on the given significance level (α), power ($1 - \beta$), the effect size (f), number of groups (k) and the numerator degrees of freedom ($u = k - 1$). The effect size is defined as $f = \frac{\delta_m}{\delta}$, where δ_m is the standard deviation of means calculated as $\delta_m = \sqrt{\frac{\sum_{i=1}^k (m_i - m)^2}{k}}$. In this equation, m_i is the mean of group i , m is the mean of the means of the groups with equal sizes and k is the number of groups. δ is the within-population standard deviation and is calculated as $\delta = \sqrt{\frac{\sum \delta_i^2}{k}}$, where δ_i is the standard deviation for each group. The sample size is then obtained using

$$N = \frac{N_{0.05}}{400 f^2} + 1$$

Where $N_{0.05}$ is the necessary sample size for the given α , u , and power at $f = 0.05$ (read from Cohen's table) and f is the nontabulated effect size. A part of the Cohen's table for obtaining sample size for ANOVA at $\alpha = 0.01$ and $u = 2$ is presented below (table 3).

Table 3. N to detect f by F test at $\alpha = 0.01$ and $u = 2$. From Cohen (1988)

Power	f											
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	0.60	0.70	0.80
0.10	307	79	36	21	14	10	8	6	5	4	3	3
0.50	1093	275	123	70	45	32	24	19	13	9	7	6
0.70	1544	387	173	98	63	44	33	26	17	12	10	8
0.80	1851	464	207	117	76	53	39	30	20	14	11	9
0.90	2325	582	260	147	95	66	49	38	25	18	14	11
0.95	2756	690	308	174	112	78	58	45	29	21	16	12
0.99	3658	916	408	230	148	103	76	59	38	27	20	16

Example

In a study, the estimates of density from three methods ($k = 3$) with four replications were compared (the data are presented in the following table).

	Method		
	1	2	3
	15	17	17
	13	18	17
	11	15	15
	12	14	19
m_i	12.75	16	17
δ_i^2	2.19	2.5	2

$$K = 3, u = 3 - 1 = 2$$

$$m = \frac{12.75 + 16 + 17}{3} = 15.25$$

$$\delta_m = \sqrt{\frac{(12.75 - 15.25)^2 + (16 - 15.25)^2 + (17 - 15.25)^2}{3}} = 1.814$$

$$\delta = \sqrt{\frac{2.19 + 2.5 + 2}{3}} = 1.493 \quad f = \frac{1.814}{1.493} = 1.215$$

Suppose a researcher wishes to compare the estimates of density from three methods assuming an effect size (f) of 1.2 according to the previous study. Then the sample size at 1% significance level and 80% power is calculated as follows;

$$N = \frac{1851}{400(1.20)^2} + 1 = 4.2$$

$N_{0.05}=1851$ is taken from the above-table (Cohen 1988). Therefore, after rounding up, 5 samples are needed for each of the three groups (methods) and the total sample size is $5(3) = 15$.

b) k groups with unequal sizes

When the compared groups are of unequal sizes, m, δ_m and δ is calculated as follows

$$m = \frac{\sum n_i m_i}{n_1 + n_2 + \dots + n_k} \quad n_i \text{ is the size of group } i,$$

$$\delta_m = \sqrt{\frac{\sum n_i (m_i - m)^2}{n_1 + n_2 + \dots + n_k}}$$

$$\delta = \sqrt{\frac{\sum n_i \delta_i^2}{n_1 + n_2 + \dots + n_k}}$$

In a study, the estimates of biomass from three methods were compared with unequal replications.

	Method		
	A	B	C
	19	14	11
	17	18	13
	15	15	15
	17	17	
m_i	17	16	13
δ_i^2	2	2.5	2.67
n_i	4	4	3

$$K=3, u=3-1=2$$

$$m = \frac{4(17)+4(16)+3(13)}{4+4+3} = 15.55$$

$$\delta_m = \sqrt{\frac{4(17-15.55)^2 + 4(16-15.55)^2 + 3(13-15.55)^2}{11}} = 1.616$$

$$\delta = \sqrt{\frac{4(2)+4(2.5)+3(2.67)}{11}} = 1.55 \quad f = \frac{1.60}{1.55} = 1.04$$

Suppose a researcher wants to compare the estimates of biomass from three methods assuming an effect size (f) =1 according to the previous study. The researcher plans that $n_1 = n_2$, and $n_3 = 0.75n_1$. Then the sample size at 1% significance level and 80% power is calculated as follows

$$N = \frac{1851}{400(1)^2} + 1 = 5.63$$

Then, the average n per sample is 6 after rounding up. The total sample size is $6(3) = 18$.

The sample size for each group is obtained as follows:

$$n_1 + n_2 + n_3 = 18$$

$$n_1 + n_1 + 0.75n_1 = 2.75n_1 = 18n_1 = 7, n_2 = 7 \text{ and } n_3 = 5$$

12. Sample size calculation for simple logistic regression

A simple (univariate) logistic regression model describes the relationship between a binary response variable ($Y=1$ and $Y=0$) and an independent variable such as climatic, soil or topographic factors. The following formula is used to calculate sample size for a simple logistic regression (Hsieh, 1989).

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{p_1(1 - p_1)\beta_*^2}$$

where $Z_{\alpha/2}=1.96$ at 5% and $Z_{\alpha/2}=2.576$ at 1% significance level for a two-tailed test. For a one-tailed test, $Z_{\alpha/2}$ is replaced by Z_{α} ($Z_{\alpha}=1.645$ at $\alpha =5\%$ and $Z_{\alpha}=2.326$ at $\alpha =1\%$), Z_{β} is 0.842 for 80% power and 1.282 for

90% power. $\beta_* = \ln(\text{ODDS ratio}) = \ln\left[\frac{\left(\frac{p_2}{1-p_2}\right)}{\left(\frac{p_1}{1-p_1}\right)}\right]$ is the tested

effect size and p_1 and p_2 are respectively the probability of the event at the mean of x , (\bar{x}) and one standard deviation (s) above the mean ($\bar{x} + s$). If the data of independent variable are not standardized, then

$$p_1 = \frac{e^{\beta_0 + \beta_1(\bar{x})}}{1 + e^{\beta_0 + \beta_1(\bar{x})}} \text{ and}$$

$$p_2 = \frac{e^{\beta_0 + \beta_1(\bar{x} + s)}}{1 + e^{\beta_0 + \beta_1(\bar{x} + s)}}$$

If the data of independent variable (x) is standardized using $\frac{x-\bar{x}}{s}$, then the mean of standardized values will be zero and the standard deviation will be 1. In this case

$$p_1 = \frac{e^{\beta_0}}{1 + e^{\beta_0}} \text{ and } p_2 = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$$

Example

Suppose a researcher wishes to determine the probability of presence of a plant species based on annual mean temperature using a simple logistic regression. The researcher plans to use plot sampling for collecting the presence ($y =1$) and absence ($y =0$) data of the plant species and the annual temperature (x).

According to a previous study, constant (β_0) of the logistic model was -3.72, coefficient of regression (β_1) was 0.238, mean of temperature (\bar{x}) in sampling plots was 15.72 and standard deviation of temperature was 3.55. Then, if the researcher wants to use the above data for calculating sample size for a new study, what is the required sample size at 0.05 significance level and 90% power?

Since the data of x are not standardized,

$$p_1 = \frac{e^{-3.72+0.238(15.72)}}{1 + e^{-3.72+0.238(15.72)}} = \frac{e^{0.021}}{1 + e^{0.021}} = 0.505$$

and

$$p_2 = \frac{e^{-3.72+0.238(15.72+3.55)}}{1 + e^{-3.72+0.238(15.72+3.55)}} = \frac{e^{0.866}}{1 + e^{0.866}} = 0.704$$

$$\beta_* = \ln \left[\frac{\left(\frac{0.704}{1-0.704} \right)}{\left(\frac{0.505}{1-0.505} \right)} \right] = 0.845$$

then the sample size is

$$N = \frac{(1.96 + 1.282)^2}{0.505(1 - 0.505)0.845^2} = 59$$

If the data of temperature are standardized to have zero mean and unit standard deviation, the following parameters are obtained for the logistic regression model.

$$\beta_0 = 0.021 \quad \beta_1 = 0.845$$

Then

$$p_1 = \frac{e^{0.021}}{1 + e^{0.021}} = 0.505$$

$$p_2 = \frac{e^{0.021+0.845}}{1 + e^{0.021+0.845}} = \frac{e^{0.866}}{1 + e^{0.866}} = 0.704$$

$$\beta_* = \ln \left[\frac{\left(\frac{0.704}{1-0.704} \right)}{\left(\frac{0.505}{1-0.505} \right)} \right] = 0.845.$$

Therefore when values are standardized, $\beta_1 = \ln(\text{ODDS ratio}) = \beta_*$

Sample size is calculated as before.

13. Sample size calculation for multiple logistic regression

The relationship between a binary variable ($y = 1, y = 0$) and a set of variables (x_1, x_2, \dots, x_p) can be modeled using a multiple logistic regression. The sample size (N) for multiple logistic regression can be obtained using the following formulas (Hsieh, 1989).

$$N = \frac{N_1}{1 - \rho_{1,23\dots p}}$$

$$N_1 = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{p_1(1 - p_1)\beta_*^2}$$

N_1 and N are the sample sizes required for a logistic regression model with 1 and p predictors respectively. That is, the sample size required for a logistic regression model with one predictor should be divided by the factor $1 - \rho_{1,23\dots p}$ to obtain the required sample size for the multiple logistic regression model. $Z_{\alpha/2}$, Z_{β} , β_* and p_1 were explained for a simple logistic regression in the previous section. $\rho_{1,23\dots p}$ also known as R^2 is the squared multiple correlation coefficient and is equal to the proportion of the

variance of x_1 explained by the regression relationship with x_2, \dots, x_p (Hsieh, 1989; Hsieh et al. 1998).

Example

In a study, the data of the occurrence of a plant species and the annual temperature and precipitation were obtained from sampling units, and a simple logistic regression was fitted between the occurrence of the plant species and the annual temperature. The standardized coefficient of the simple logistic model based on temperature was $\beta_0 = 0.038$ and $\beta_1 = 0.872$, and the squared multiple coefficient of correlation (R^2) between the temperature and precipitation was 0.03.

Suppose a researcher wants to use the above data for determination of adequate sample size for a new study. Then, what is the required sample size at $\alpha = 0.05$ ($Z_{\alpha/2} = 1.96$) and power = 0.95 ($Z_{\beta} = 1.645$)?

First the sample size for a simple logistic regression based on temperature is calculated as explained in the previous section.

$$p_1 = 0.509$$

$$N_1 = \frac{(1.96 + 1.645)^2}{0.509(1 - 0.509)0.872^2} = 69$$

Then the sample size for a multiple logistic regression for detecting the effect of temperature while controlling for the effects of precipitation is calculated as $N = \frac{69}{1-0.03} = 71$.

We can also determine the sample size for a simple logistic regression model with only precipitation and approximate it for a multiple case. If the standardized coefficients of the simple logistic model based on only precipitation are $\beta_0 = -0.004$ and $\beta_1 = -0.762$, and $R^2 = 0.03$, the sample size for a simple case is

$$N_1 = \frac{(1.96 + 1.645)^2}{0.499(1 - 0.499)(-0.762)^2} = 90$$

and for a multiple case will be

$$N = 90 \times \frac{1}{1-0.03} = 93.$$

As explained before, the sample size for a multiple logistic regression is calculated by specifying the probability of an event at the mean value of all the covariates (predictors), and the odds ratio of the event corresponding to an increase of one standard deviation from the mean value of the specific covariate, given the mean values of the remaining covariates (Hsieh, 1989). Therefore the sample size for a multiple logistic regression can also be calculated without using squared multiple coefficient of correlation (R^2) as explained in the following examples.

Example

In a study, a multiple logistic regression model was made based on the data of previous example including the presence/absence data of a plant species and two predictors including annual temperature (x_1) and precipitation (x_2). The following results were obtained from the study:

- mean of temperature (\bar{x}_1) = 15.744
- standard deviation of temperature (s_1) = 5.06
- mean of precipitation (\bar{x}_2) = 191.76
- standard deviation of precipitation (s_2) = 58.62
- unstandardized coefficients of the multiple logistic regression model:

$\beta_0 = -0.169$, β_1 (for temperature) = 0.162, β_2 (for precipitation) = -0.0123

Suppose a researcher wants to use the above data for determination of required sample size to conduct a new study. The sample size for determining the effect of temperature while controlling for the effects of precipitation is calculated as follows:

$$p_1 = \frac{e^{-0.169+0.162(15.744)-0.0123(191.76)}}{1 + e^{-0.169+0.162(15.744)-0.0123(191.76)}} = \frac{e^{0.023}}{1 + e^{0.023}} = 0.506$$

and

$$p_2 = \frac{e^{-0.169+0.162(15.744+5.06)-0.0123(191.76)}}{1 + e^{-0.169+0.162(15.744+5.06)-0.0123(191.76)}} = \frac{e^{0.843}}{1 + e^{0.843}} = 0.699$$

$$\beta_* = \ln \left[\frac{\left(\frac{0.699}{1-0.699}\right)}{\left(\frac{0.506}{1-0.506}\right)} \right] = 0.82$$

$$N = \frac{(1.96 + 1.645)^2}{0.506(1 - 0.506)0.82^2} = 78$$

The sample size for detecting the effect of precipitation while controlling for the effects of temperature is

$$p_1 = 0.506$$

and

$$p_2 = \frac{e^{-0.169+0.162(15.744)-0.0123(191.76+58.62)}}{1 + e^{-0.169+0.162(15.744)-0.0123(191.76+58.62)}} = \frac{e^{-0.7}}{1 + e^{-0.7}} = 0.332$$

$$\beta_* = \ln \left[\frac{\left(\frac{0.332}{1-0.332}\right)}{\left(\frac{0.506}{1-0.506}\right)} \right] = -0.723$$

$$N = \frac{(1.96 + 1.645)^2}{0.506(1 - 0.506)(-0.723)^2} = 100$$

The sample sizes calculated in this way are close to those calculated in the previous section. If the data are standardized using $(x - \bar{x})/s$ to have mean = 0 and variance = 1, the standardized coefficients of the multiple logistic regression model will be as follows:

$\beta_0 = 0.023$, β_1 (for temperature) = 0.82, β_2 (for precipitation) = -0.723

Then the sample size for determining the effect of temperature is

$$p_1 = \frac{e^{0.023}}{1 + e^{0.023}} = 0.506$$

$$N = \frac{(1.96 + 1.645)^2}{0.506(1 - 0.506)0.82^2} = 78$$

and the sample size for determining the effect of precipitation is

$$N = \frac{(1.96 + 1.645)^2}{0.506(1 - 0.506)(-0.723)^2} = 100$$

Results and Discussion

An adequate sample size must be taken in a study to obtain reasonable power and result. The statistical power in vegetation studies due to small sample sizes is reported to be 39.8% for detecting a medium effect size (Jennions & Moller 2003). This is far lower than the general recommendation of a power of 80% (Cohen 1988). Most ecologists still report non-significant results without indicating the required sample size and power of the test. A non-significant result in a study may be due to a low sample size and low power and not because there is no significant effect or difference (Steidl et al. 1997; Jennions & Moller 2003). Therefore, the non-significant outcome may be inconclusive in a study designed with a low sample size. An appropriate sample size can be calculated for a study by selecting a reliable power and minimum detectable effect size before data collection. However, most researchers do not calculate enough sample size before data collection and take inadequate sample size resulting in a low power for the test.

One solution to improve the reliability and power of a study after performing with a low sample size is to use a one-tailed hypothesis test if possible. One-tailed (directional) test requires a smaller sample size and has more statistical power than two-tailed test to detect an effect and is recommended to be used when the effect in one direction can be explained (Ruxton & Neuhauser 2010). For example, a one-tailed test can be used to test the hypothesis that plant density and biomass is greater in ungrazed stands than in grazed stands or percentage of germinated seeds is higher under gibberellin treatment than control condition or plant production increases with increasing precipitation. Another solution for increasing the

reliability and power of a test is to select a higher significance level (e.g., 5% instead of 1%). 5% significance level is an acceptable and reasonable significance level for accepting/rejecting a null hypothesis and also results in a smaller sample size and higher power than significance levels of less than 5%. The sample size also depends on the statistical method itself. For example, the required sample size for paired sample t-test and one sample t-test is lower than that for independent two sample t-test (Jennions & Moller 2003). It must be noted that the decision for selection of a statistical test must be made at the sampling design stage and before data collection.

The statistical methods presented in this paper for calculating sample size assume a normal distribution of vegetation data (except the Chi-square test which is a non-parametric statistic also called a distribution free test). Although some studies have indicated that vegetation characteristics such as biomass, density, production, plant height and seed production are normally distributed (Rosner & Rose 2006; Hoch & Korner 2009; Giannini et al. 2018), some others have revealed that distribution of these characteristics is not normal (Mascaro & Schnitzer 2011; Murphy et al. 2018; Bisi et al. 2018). Therefore, it is recommended to test the distribution of vegetation data before selection of a statistical test using Kolmogorov-Smirnov or Shapiro-Wilk test. If the data are not normally distributed, transformation methods such as log, square root, and angular transformation can be used to normalize the distribution of data. However non-parametric tests may present efficient results if data are not normally distributed. Non-parametric tests make no assumptions about distribution of data. Some non-parametric tests include sign test and Wilcoxon signed rank test for paired samples and one sample case, Mann-Whitney U test for two independent samples, Kruskal-Wallis test for one-way ANOVA, Spearman rank correlation coefficient and nonparametric regression methods. It must be noted that non-parametric tests require a larger sample size than parametric tests to achieve the same level of power as a parametric test. The sample size calculation for non-parametric tests is beyond the scope of this paper but can be done using G*Power software. In conclusion, sample size calculation is more valuable in the design or planning stages of research than after data collection. However if exact sample size is calculated after data collection and testing the hypothesis, presenting the power, significance level and confidence interval for an observed power along with the sample size can provide additional information about the null hypothesis that are not rejected.

References

- Bisi, F., Chirichella, R., Chianucci, F., Von Hardenberg, J., Cutini, A., Martinoli, A., Apollonio, M. (2018). Climate, tree masting and spatial behaviour in wild boar (*Sus scrofa* L.): insight from a long-term study. *Annals of forest science*, **75**(2): 1-9
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd ed. Lawrence Erlbaum Associates, USA, 567 pp.
- Conser, C. & Connor, E.F. (2009). Assessing the residual effects of *Carpobrotus edulis* invasion, implications for restoration. *Biological Invasions*, **11**(2): 349-358.
- Croy, C. D. & Dix, R. L. (1984). Notes on sample size requirements in morphological plant ecology. *Ecology*, **65**(2): 662-666.
- Dupont, W.D. & Plummer Jr, W.D. (1998). Power and sample size calculations for studies involving linear regression. *Controlled clinical trials*, **19**(6): 589-601.
- Dupont, W.D. & Plummer, W.D. (1997). PS: Power and sample size calculation. *Control Clin Trials*, 18-274.
- Erdfelder, E., Faul, F., Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior research methods, instruments, & computers*, **28**(1): 1-11.
- Federal Register. (1979). *Rules and Regulations*, Vol. 44, No 50.
- Faul, F., Erdfelder, E., Lang, A. G., Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, **39**(2): 175-191.
- Giannini, V., Bertacchi, A., Bonari, E., Silvestri, N. (2018). Rewetting in Mediterranean reclaimed peaty soils and its potential for phyto-treatment use. *Journal of environmental management*, **208**: 92-101.
- Hofmann, L. & Ries, R.E. (1990). An evaluation of sample adequacy for point analysis of ground cover. *Rangeland Ecology & Management/Journal of Range Management Archives*, **43**(6): 545-549.
- Hoch, G. & Korner, C. (2009). Growth and carbon relations of tree line forming conifers at constant vs. variable low temperatures. *Journal of Ecology*, **97**(1), 57-66.
- Hsieh, F.Y. (1989). Sample size tables for logistic regression. *Statistics in medicine*, **8**(7): 795-802.
- Hsieh, F. Y., Bloch, D.A., Larsen, M.D. (1998). A simple method of sample size calculation for linear and logistic regression. *Statistics in medicine*, **17**(14): 1623-1634.
- Jennions, M. D. & Moller, A. P. (2003). A survey of the statistical power of research in behavioral ecology and animal behavior. *Behavioral Ecology*, **14**(3): 438-445.
- Moffet, C.A. (2009). Agreement between measurements of shrub cover using ground-based methods and very large scale aerial imagery. *Rangeland ecology & management*, **62**(3): 268-277.
- Mosley, J. C., Bunting, S. C., Hironaka, M. (1989). Quadrat and sample sizes for frequency sampling mountain meadow vegetation. *The Great Basin Naturalist*, **49**: 241-248.
- Mascaro, J. & Schnitzer, S. A. (2011). Dominance by the introduced tree *Rhamnus cathartica* (common buckthorn) may limit aboveground carbon storage in Southern Wisconsin forests. *Forest Ecology and Management*, **261**(3), 545-550.

- Murphy, S. M., Vidal, M. C., Smith, T. P., Hallagan, C. J., Broder, E. D., Rowland, D., Cepero, L. C. (2018). Forest fire severity affects host plant quality and insect herbivore damage. *Frontiers in Ecology and Evolution*, **6**: 1-11
- Peek, M.S., Leffler, A.J., Flint, S.D., Ryel, R. J. (2003). How much variance is explained by ecologists? Additional perspectives. *Oecologia*, **137**(2): 161-170.
- Rosner, L. S., & Rose, R. (2006). Synergistic stem volume response to combinations of vegetation control and seedling size in conifer plantations in Oregon. *Canadian journal of forest research*, **36**(4), 930-944.
- Ruxton, G. D. & Neuhauser, M. (2010). When should we use one-tailed hypothesis testing?. *Methods in Ecology and Evolution*, **1**(2), 114-117.
- StataCorp, L.P. (2013). *Stata power and sample-size reference manual*. Stata Press, Texas, USA, 632 pp.
- Stauffer, H.B. (1983). Some sample size tables for forest sampling. *Brit. Columbia Min. For. Res. Note* 90, 56 pp.
- Steidl, R. J., Hayes, J. P., Schauber, E. (1997). Statistical power analysis in wildlife research, *The Journal of Wildlife Management*, 270-279.
- Thomas, L. & Krebs, C.J. (1997). A review of statistical power analysis software. *Bulletin of the Ecological Society of America*, **78**(2): 126-138.

