doi: 10.25750/1995-4301-2023-1-016-027

Sample size and power calculations for nonparametric tests in vegetation research

© 2023. M. Mousaei Sanjerehei ORCID: 0000-0002-9772-6021,
Department of Plant Protection, Yazd Branch,
Islamic Azad University, Yazd, Iran,
e-mail: mmusaei@iauyazd.ac.ir

An adequate sample size and a reliable statistical power is required for accurately estimating vegetation characteristics and proper interpretation of research results. The sample size depends on the significance level, power of test and effect size. An appropriate equation for determination of sample size and power must be selected in accordance with the measurement scale, distribution of data, type of research and statistical analysis test. In this paper, equations were explained for calculating sample size and power for the most frequently used nonparametric tests in vegetation studies including the sign test, Wilcoxon signed rank, Mann-Whitney U, Kruskal-Wallis, Spearman rank correlation, Chi-square and McNemar tests. Each equation is presented with an example of its use in vegetation research.

Keywords: effect size, nonparametric tests, sample size, power, significance level, vegetation.

УДК 519.254

Размер выборки и расчёт мощности для непараметрических тестов в исследованиях растительности

© 2023. М. Мусай Санжерехей, PhD, доцент,
Филиал Исламского университета Азад в Йезде,
Иран, Йезд, Сафайе, бульвар Хесаби,
e-mail: mmusaei@iauyazd.ac.ir

Для точной оценки характеристик растительности и надлежащей интерпретации результатов исследований требуется адекватный размер выборки и надёжная статистическая мощность. Размер выборки зависит от уровня значимости, статистической мощности теста и размера эффекта. Соответствующее уравнение для определения размера выборки и мощности должно быть выбрано в соответствии с масштабом измерения, распределением данных, типом исследования и методом статистического анализа. В данной статье даны пояснения к уравнениям для расчёта размера выборки и мощности для наиболее часто используемых непараметрических тестов в исследованиях растительности, включая критерий знака, знаковый ранг Уилкоксона, критерий Манна-Уитни, критерий Крускала-Уоллиса, ранговую корреляцию Спирмена, критерий хи-квадрат и критерий Макнемара. Каждое уравнение представлено с примером его использования в исследованиях растительности.

Ключевые слова: размер эффекта, непараметрические тесты, объём выборки, статистическая мощность, уровень значимости, растительность.

Parametric tests make certain conditions about the parameters of population. The most important assumptions of parametric tests are 1) the data must be independent, 2) the data must be normally distributed, 3) the populations must have the same variance, 4) the measurement must be in an interval or ratio scale and 5) there must be a linear relationship between the data. In general, parametric tests are more powerful than nonparametric tests because these tests have strongest assumptions about the population. Nonparametric tests do not make any certain assumption about the population and can be used when the distribution of data is not known. If sample size is as small as 6, there is no alternative to using a nonparametric test unless the population distribution is exactly known [1]. Nonparametric tests can be used for analyzing the four scales of measurement; nominal (e. g., presence/absence of plants in plots, percentage number of plant species in different life form categories such as therophytes and cryptophytes), ordinal (e. g., rangeland condition scores, intensity of grazing, cover and biomass classes), interval (e. g., temperature, time of grazing) and ratio (e. g., most of vegetation characteristics such as biomass, density, richness, cover, height, seed size) scales. However these tests are mostly used for nominal and ordinal scales. Interval and ratio data can be analyzed by nonparametric tests if they are ordered and ranked. If data are in ranks or can be categorized into ordinal scales (e. g., more or less, better or worse), they can be analyzed by nonparametric tests whereas they cannot be analyzed by parametric tests unless precarious and unrealistic assumptions are made about the underlying distributions [1].

Two of the most important stages of vegetation studies are calculation of adequate sample size and statistical power. In vegetation studies, sample size is the required number of sampling units (e. g., plots, quadrats, points, transects, seed and fruit traps), plant individuals, plant parts (e. g., seed, stem, leaf) taken for estimating and comparing vegetation characteristics such as cover, biomass, density, production, richness, height, diameter, nutrients, rooting depth, interactions, etc.

An accurate sample size is required for reliable estimate of vegetation attributes. The required sample size depends on the significance level, power level and effect size. Significance level (α) or Type I error is defined as the probability of rejecting the null hypothesis, when it is true. The most common significance levels used in vegetation studies are 10%, 5% and 1% [2, 3]. A decrease in the significance level leads to an increase in the sample size.

Effect size is the size of difference between the groups being compared. The term “effect size” was defined as the degree to which the phenomenon is present in the population [4]. Examples of effect size are the odds ratio in a contingency table, the difference in the proportion between positive and negative ranks for Wilcoxon signed rank test, the difference between the proportion of positive or negative ranks and zero for sign test and the difference of correlation from zero in Spearman rank correlation. The larger the effect size, the smaller the sample size.

The power of a statistical test is the probability that it will yield statistically significant results. Statistical Power is defined as the probability of correctly rejecting the null hypothesis when it is false, and it is equal to 1–β. Type II error (β) is the probability of accepting the null hypothesis when it is false. The power of a test depends upon three parameters; sample size, significance level and effect size. The larger the sample size, other things (significance level, effect size) being equal, the larger the power. Also, the lower the value of significance level, the lower the power of the test. Sample size and power should be calculated before data collection using data of previous studies. When the power is calculated after data collection, it can be used to verify whether a non-significant result is due to lack of strong relationship between the groups or due to a low statistical power. An ideal study is the one with a high power indicating that the study has a high chance of detecting a difference between groups if it exists [5]. Common values used for calculating power in vegetation studies are 0.8 and 0.9 [6–8].

The objective of this study was to present sample size and power equations and explain with examples how to use these equations for nonparametric tests according to the type of vegetation research, measurement scale of data and the statistical analysis test. The sample size and power formulas presented in this study were explained for the most frequently used nonparametric tests in vegetation studies including the Wilcoxon signed rank, sign test, Mann-Whitney U, Kruskal-Wallis, Spearman rank correlation, Chi-square and McNemar tests.

Methods

Wilcoxon signed rank test is used for testing both paired samples and one sample case. This statistic is used to test the differences or changes in a vegetation characteristic (e. g., richness, density, cover, biomass, production, height, rooting depth, leaf size, seed weight, nutrient contents, number of flowers), vegetation index (e. g., Shannon diversity index, leaf area index), spectral vegetation index (e. g., normalized difference vegetation index, NDVI, modified soil adjusted vegetation index, MSAVI), number of animals, and area of a vegetation type (e. g., woodland) between two time periods or before and after a management (e. g., grazing, manuring, vegetation removal) or disturbance (e. g., fire, climate change). It can also be used for testing the differences in estimates of a vegetation characteristic (e. g., cover, biomass, density) between two methods (e. g., plot and line transect method), or the differences in a soil property (e. g., EC, water content, nutrients) or a climate variable (e. g., precipitation, temperature) in a habitat between two time periods or before and after management, treatment or disturbance [9, 10]. Wilcoxon signed rank test is more efficient than sign test because it uses more of the information in the data, and is affected by the relative magnitude of the differences. In the Wilcoxon signed rank test, the difference and the absolute value of difference is computed for each pair of observations. All non-zero absolute differences are sorted into ascending order, and ranks are assigned. In case of ties (two or more ranks are equal), the average rank is computed.

Sample size calculation for Wilcoxon signed rank test. The Sample size for the Wilcoxon signed rank test is calculated as follows [11]:

$N = (Z_{\alpha/2} + Z_\beta)^2 / 3(p' - 0.5)^2$,
where $p' = (S_p - Np) / 0.5N(N - 1)$
and $p = n_p / (n_p + n_n)$.

$N = n_p + n_n$ is number of non-zero differences, n_p and n_n are respectively number of positive and negative differences, S_p is the sum of the ranks corresponding to positive differences, p and p' can be obtained from a preliminary sample or from previous studies, $Z_{\alpha/2}$ is standard normal variate for significance level ($Z_{\alpha/2}$ = 1.96 at α = 5% and $Z_{\alpha/2}$ = 2.576 at α = 1% for a two tailed hypothesis). For a one-tailed test $Z_{\alpha/2}$ is replaced by Z_α (Z_α = 1.645 at α = 5% and Z_α = 2.326 at α = 1%), Z_β is standard normal variate for power of study. Z_β = 0.842 for 80% power, and Z_β = 1.282 for 90% power. The effect size is the rank-biserial correlation (the difference between the proportion of positive and negative ranks).

Example. In a study, number of plants in a habitat was counted in 8 fixed plots (Table 1) between two time periods (dry year and wet year) to test if the differences in the number of plants between the two time periods are significant.

$S_p = 8 + 4 + 5 + 7 + 6 + 2 = 32$, $p = 6/8 = 0.75$,
$p' = 32 - (8 \cdot 0.75) / 0.5 \cdot 8 \cdot (8-1) = 0.928$.

Suppose a researcher wishes to test the differences in number of plants in a habitat between two time periods assuming that p' may be around 0.9 according to the previous study. Then, the required sample size at 5% significance level and 80% power based on a two-sided test is: $N = (1.96+0.842)^2/3(0.9-0.5)^2 = 16.35$.

The required sample size is 17 after rounding up.

Power calculation for Wilcoxon signed rank test. The power for the Wilcoxon signed rank test is calculated as follows: $Z_\beta = \sqrt{3(p'-0.5)^2 \cdot N} - Z_{\alpha/2}$. The power for the value of Z_β is obtained from the table of standard normal probabilities. Z_α is used instead of $Z_{\alpha/2}$ for a one-tailed test.

Example. A researcher plans to test the differences in number of plants in a habitat between two time periods using 10 fixed plots assuming that p' is near 0.9. The power for the Wilcoxon signed rank test at 5% significance level for a two-tailed test is calculated as follows: $Z_\beta = \sqrt{(3)(0.9-0.5)^2 \cdot 10} - 1.96 = 0.23$. The power for the value of Z_β = 0.23 is 0.591 from z-table.

Sign test is applied for testing both paired samples and one sample case. The applications of sign test are similar to the applications of Wilcoxon signed rank test in vegetation studies as explained before [12]. In sign test, the difference for each pair is calculated, and the number of positive differences (n_p) and negative differences (n_n) is counted. The cases in which the difference equals zero are ignored.

Sample size calculation for sign test. The sample size for the sign test is calculated as follows [11]: $N = \frac{(Z_{\alpha/2} + Z_\beta)^2}{4(p-0.5)^2}$, where $p = n_p / (n_p + n_n)$ or $p = n_n / (n_p + n_n)$ and n_p is the number of positive differences and n_n is the number of negative differences, p can be calculated from a preliminary sample or taken from previous studies. The effect size for sign test is defined as $p - 0.5$.

Example. Suppose the objective is to test whether the differences in biomass of plants in a habitat between two time periods are significant using sign test. The biomass of plants is assumed to increase in 75% and decrease in 25% of the fixed plots between the two time periods. Then the sample size for testing the differences in biomass at 5% significance level and 80% power is: $N = (1.96+0.842)^2 / 4(0.75-0.5)^2 = 31.4$. Then 32 fixed plots are required based on two-tailed test.

Power calculation for sign test. The power for sign test is calculated as follows: $Z_\beta = \sqrt{4(p-0.5)^2 \cdot N} - Z_{\alpha/2}$. The power for the value of Z_β is obtained from the table of standard normal probabilities. Z_α is used instead of $Z_{\alpha/2}$ for a one-tailed test.

Table 1

Number of plants in eight fixed plots in dry and wet year

Number of plants in dry year	12	10	14	11	9	10	8	13
Number of plants in wet year	28	16	12	6	18	25	21	16
Difference	+16	+6	-2	-5	+9	+15	+13	+3
Rank	8	4	1	3	5	7	6	2

Example. Suppose a researcher wishes to use 36 fixed plots for testing the differences in biomass of plants in a habitat between two time periods expecting that the biomass will increase in 75% of the plots and decrease in 25% of the plots. Then, what is the power of the sign test at 5% significance level for a two-sided hypothesis? $Z_\beta = \sqrt{(4)(0.75-0.5)^2 \cdot 36} - 1.96 = 1.04$. The power for the value of $Z_\beta = 1.04$ is 0.85 from z-table.

Mann-Whitney U test is used to test whether the difference between two independent samples is significant. This statistical method is used to test the differences in a vegetation characteristic (e. g., richness, cover, biomass, density, height, basal area, production, nutrients, seed weight), spectral vegetation index (e. g., NDVI) or vegetation index (e. g., Simpson diversity index, Pielou's index of nonrandomness) between two sites (e. g., two habitats, two vegetation types, grazed and ungrazed sites, manured and unmanured sites). This test can also be used for testing the differences in a soil or climate variable between two habitats, or the differences in a vegetation characteristic between two plant species or two life form categories [13]. To calculate Mann-Whitney U test, the observations for both groups are combined, sorted in order of increasing size and ranked. If ties exist (two or more ranks are equal), the average rank is calculated.

Sample size calculation for Mann-Whitney U test. The sample size for Mann-Whitney U test can be calculated as follows [11]: $N = (Z_{\alpha/2}+Z_\beta)^2 / 12c(1-c)(p''-0.5)^2$ where $c = n_1/N$ or $c = n_2/N$ and $p'' = U/n_1n_2$ and $U = n_1n_2 + [n_1(n_1+1)/2] - S_1$, S_1 is the sum of ranks for group 1, n_1 and n_2 are the sample size of group 1 and 2 respectively, $N = n_1 + n_2$, p'' can be obtained from a preliminary sample or from previous studies. The effect size is the rank- biserial correlation $(1 - 2U/n_1n_2)$.

Example. A study was conducted to test whether there is a significant difference in the normalized difference vegetation index (NDVI) between two vegetation types. The mean NDVI in 12 plots (5 plots in vegetation type 1 and 7 plots in vegetation type 2) is presented in the following table (Table 2).

Suppose a researcher wishes to test whether there are significant differences in the NDVI between two vegetation types. The researcher can estimate that p'' will be about 0.06 (or 0.94) according to the previous study, and plans to use 40% of the plots in one vegetation type and 60% of the plots in the other vegetation type. Then, the required sample size at 5% significance level and 90% power is: $N = (1.96+1.282)^2 / (12 \cdot 0.40)(1-0.40)(0.06-0.5)^2 = 18.85$. Then, 19 plots need to be used for both groups (11 plots in one vegetation type and 8 plots in the other) for a two sided test.

Power calculation for Mann-Whitney U test. The power for Mann-Whitney U test to detect the difference between two independent samples is calculated as follows: $Z_\beta = \sqrt{12c(1-c)(p''-0.5)^2 \cdot N} - Z_{\alpha/2}$. The power is calculated using the value of Z_β from the table of standard normal probabilities. $Z_{\alpha/2}$ is replaced by Z_α for a one-tailed test.

Example. According to the previous example, if the researcher wishes to use 5 plots in one vegetation type and 7 plots in the other, and expecting that $p'' = 0.06$ (or 0.94), the power for testing the differences in NDVI values between the two vegetation types at 5% significance level is calculated as follows: $N = 5+7 = 12$, $c = 5/12 = 0.417$.

$$Z_\beta = \sqrt{12 \cdot 0.417 \cdot (1-0.417)(0.06-0.5)^2 \cdot 12} - 1.96 = 0.64$$

The power for the value of $Z_\beta = 0.64$ from z-table for a two-sided test is 0.739.

Spearman rank correlation. In vegetation studies, Spearman rank correlation is used to detect the relation between a vegetation characteristic (e. g., biomass, production, density, richness, height, abundance, basal area, leaf size, seed weight, leaf area index) and an environmental variable (e. g., precipitation, temperature, humidity, elevation, slope, soil pH, EC, water content, nutrients). It can also be used for calculating the association in a plant characteristic (e. g., biomass) between two plant species [14, 15].

Sample size calculation for Spearman rank correlation. The sample size for Spearman rank correlation (r_S) is calculated as follows [16]:

Table 2

Mean NDVI in 5 plots in vegetation type 1 and 7 plots in vegetation type 2

	Vegetation type 1					Vegetation type 2						
Mean NDVI	0.58	0.60	0.65	0.67	0.50	0.35	0.48	0.59	0.32	0.38	0.40	0.42
Rank	8	10	11	12	7	2	6	9	1	3	4	5

$S_1 = 8 + 10 + 11 + 12 + 7 = 48$, $U = (5 \cdot 7) + [5(5+1)/2] - 48 = 2$
$c = 5/12 = 0.417$, $p'' = 2/(5 \cdot 7) = 0.057$

$$N = \left[\left(\frac{Z_{\alpha/2} + Z_\beta}{\frac{1}{2} \cdot \ln\left(\frac{1 + r_S}{1 - r_S} \right)} \right)^2 \cdot 1.06 \right] + 3,$$

where r_S is the Spearman rank correlation. When the difference is tested between the Spearman r_S and zero, r_S is the effect size.

Example. Suppose the Spearman rank correlation (r_S) between rangeland condition scores and precipitation is expected to be near 0.60. Then the sample size for calculating Spearman correlation coefficient (r_S) at 5% significance level and 80% power based on a two-tiled hypothesis is:

$$N = \left[\left(\frac{1.96 + 0.842}{\frac{1}{2} \cdot \ln\left(\frac{1 + 0.6}{1 - 0.6} \right)} \right)^2 \cdot 1.06 \right] + 3 = 20.3.$$

The required sample size is 21 after rounding up for a two-tailed hypothesis.

Power calculation for Spearman rank correlation. The power for Spearman rank correlation (r_S) is calculates as follows:

$$Z_\beta = \left(\left| \frac{1}{2} \cdot \ln\left(\frac{1 + r_S}{1 - r_S} \right) \right| \cdot \sqrt{\frac{N - 3}{1.06}} \right) - Z_{\alpha/2},$$

where N is the sample size. The power is calculated using the value of Z_β from the table of standard normal probabilities.

Example. Suppose a researcher wishes to determine the Spearman rank correlation (r_S) between rangeland condition scores and precipitation using 25 sampling plots, and thinks that the correlation will be about 0.60. Then the power for Spearman rank correlation coefficient (r_S) at 5% significance level based on a two-tiled hypothesis is obtained as follows:

$$Z_\beta = \left(\left| \frac{1}{2} \cdot \ln\left(\frac{1 + 0.6}{1 - 0.6} \right) \right| \cdot \sqrt{\frac{25 - 3}{1.06}} \right) - 1.96 = 1.2.$$

The power for the value of $Z_\beta = 1.2$ is 0.88 from z-table.

Chi-square test using contingency table data. A Chi-square statistic is used to test the association or independence between two or more traits, when the data are in discrete categories (either nominal or ordinal). Chi-square test based on a contingency table is mainly used to test whether the occurrences of two plant species are associated, or two or more habitats contain equal number of plant species. Chi-square test can also be used for one sample case to test the differences in a plant characteristic (e. g., richness, density) among different life forms (e. g., geophytes, therophytes) or the differences in frequencies of two or more plant species (% number of sampling units that contain a plant species). This test can also detect if a certain type of habitat is more frequently selected by an animal, or if plant species differ in their proportion of coexistence with an organism (e. g., bacteria, fungus species) [17, 18].

Sample size calculation for chi-square test based on contingency table. The sample size for chi-square test based on contingency tables can be calculated using the following formula and Cohen's tables [4]. $N = N_{0.1}/100w^2$, 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 $u = (r-1)(c-1)$. 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 chi-square test at α = 0.05 and u = 6 is presented below (Table 3).

Example. A researcher wants to compare the frequencies of four plant species in three vegetation types by calculating the number of quadrats containing the plant species. Let us assume that in a previous similar study, the chi-square calculated from data of a 3 · 4 table was 14 based on a sample of size (N) = 120. Then, the effect size, $w = \sqrt{(14/120)} = 0.34$. If the researcher wants to calculate the sample size assuming w = 0.34 for a new study, the required sample size at α = 0.05 and 90% power is calculated as follows:

Table 3

N to detect w by χ^2 at α = 0.05 and u = 6

	w								
Power	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0.90	1742	435	194	109	70	48	36	27	22
0.95	2086	521	232	130	83	58	43	33	26
0.99	2805	701	312	175	112	78	57	44	35

Table 4

Presence/absence data of species A and B

Species	Species B		Total
	Present	Absent	
Species A	Present	15	6	21
	Absent	7	22	29
	Total	22	28	$N = 50$

$u = (3–1)\cdot(4–1) = 6$. According to the Cohen's table (table 3), $N_{0.1} = 1742$ at $\alpha = 0.05$, power = 0.90, $u = 6$ and $w = 0.1$, then the sample size will be $N = 1742/100(0.34)^2 = 151$.

Sample size calculation for chi-square test based on 2 · 2 table. The sample size for a 2 · 2 contingency table ($df = 1$) can be obtained using the formula in the previous section and also using the following equation: $N = (\sqrt{\chi_c^2} + Z_\beta)^2 / w^2$.

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 2 · 2 table and N is the total number of samples. $\chi^2 = \sum (o_i - e_i)^2/e_i$, where 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. The data were obtained using located quadrats and summarized in a 2 · 2 table as follows (Table 4):

15 quadrats (cell a) contain both species A and B. In 6 quadrats (cell b) species A is present but not B. In 7 quadrats (cell c) species B is present but not A. In 22 quadrats (cell d) neither species A nor B are found. The expected value for each cell of the table is calculated as follows:

$$E(a) = \frac{21\cdot 22}{50} = 9.24,\quad E(b) = \frac{21\cdot 28}{50} = 11.76,$$

$$E(c) = \frac{22\cdot 29}{50} = 12.76,\quad E(d) = \frac{28\cdot 29}{50} = 16.24.$$

Then
$$\chi^2 = \frac{(15-9.24)^2}{9.24} + \frac{(6-11.76)^2}{11.76} + \frac{(7-12.76)^2}{12.76} + \frac{(22-16.24)^2}{16.24} = 11$$

and the effect size $w = \sqrt{11/50} = 0.469$.

Suppose a researcher wishes to test the association between two plant species (A and B) using data of a 2 · 2 table and chi-square test assuming $w = 0.47$ according to the previous study. Then, the required sample size at 0.05 significance level and 80% power is: $N = (\sqrt{3.84} + 0.842)^2 / 0.47^2 = 35.5$. Therefore 36 quadrats are required for detecting the association between the two plant species.

Power calculation for chi-square test based on contingency table. The power for a chi-square test based on the data of a contingency table can be calculated using the following equation [19]:

$$Z_\beta = -\left[\left(\frac{\chi_c^2}{\nu + \lambda}\right)^{\frac{1}{3}} - \left(1 - \frac{2(\nu + 2\lambda)}{9(\nu + \lambda)^2}\right)\right] \Big/ \sqrt{\frac{2(\nu + 2\lambda)}{9(\nu + \lambda)^2}}$$

The power is calculated based on Z_β value from Z-table. ν is degrees of freedom and is calculated for a contingency table with r rows and c columns as $\nu = (r - 1)(c - 1)$ and χ_c^2 is the critical value of the chi-square obtained from chi-square distribution table based on a specified significance level and degrees of freedom (ν) and λ is the noncentrality parameter (calculated χ^2).

Example. According to the previous example (frequencies of four plant species in three vegetation types), if the χ^2 is expected to be 14 based on data of a 3 · 4 table using 120 plots, the power at $\alpha = 0.05$ is calculated as follows: $\lambda = 14$, $\nu = (3–1)(4–1) = 6$, table χ_c^2 for $\alpha = 0.05$ and $df = 6$ is 12.6,

$$Z_\beta = -\left[\left(\frac{12.6}{6+14}\right)^{\frac{1}{3}} - \left(1 - \frac{2(6+2(14))}{9(6+14)^2}\right)\right] \Big/ \sqrt{\frac{2(6+2(14))}{9(6+14)^2}} = +0.901.$$

The power obtained for $Z_\beta = +0.901$ from Z-table is 0.81.

Power calculation for chi-square using 2 · 2 table data. The power for a 2 · 2 contingency table can be calculated using the formula in the previous section. For a 2 · 2 table, the degrees of freedom is 1 and the critical chi-square value at $\alpha = 0.05$ is 3.84 and at $\alpha = 0.01$ is 6.63. The statistical power for a 2 · 2 contingency table can also be calculated using the following equation: $Z_\beta = (\sqrt{N} \cdot w) - \sqrt{\chi_c^2}$.

Table 5

Proportions in the cells of a 2 · 2 table for McNemar test

	Yes = 1	No = 0	
Yes = 1	$p_{11} = \frac{a}{N}$	$p_{10} = \frac{b}{N}$	$p_1 = p_{11} + p_{10}$
No = 0	$p_{01} = \frac{c}{N}$	$p_{00} = \frac{d}{N}$	$1 - p_1$
	$p_2 = p_{11} + p_{01}$	$1 - p_2$	1

Example. Suppose a researcher wants to use 50 plots for detecting the association using presence/absence data of two plant species in a 2 · 2 contingency table. The researcher expects that the chi-square will be about 11, that is the effect size $w = \sqrt{11/50} = 0.469$. The power at 0.05 significance level is obtained as follows: $Z_\beta = (\sqrt{50} \cdot 0.469) - \sqrt{3.84} = 1.36$. The power obtained for $Z_\beta = 1.36$ from Z-table is 0.91.

McNemar test is used to test the difference between paired proportions, particularly when the measurement is in either nominal or ordinal scale. McNemar test is mainly used for detecting the accuracy of classification in vegetation mapping based on comparing the performance of two classifiers. It has also been applied for detecting the changes in proportions of presence/absence data of plant species between two sites or two time periods [20, 21].

Sample size calculation for McNemar test. The sample size for McNemar test using the proportions of the cells of 2 · 2 table (Table 5) is calculated as follows [22]:

$$N = \frac{\left[Z_{\alpha/2}\sqrt{\psi} + Z_\beta\sqrt{\psi - \delta^2}\right]^2}{\delta^2}.$$

Where $\delta = p_{10} - p_{01}$ and $\psi = p_{10} + p_{01}$. The Odds ratio p_{10}/p_{01} is used to specify the effect size.

Example. A study was conducted to compare the performance of two classifiers for vegetation mapping. The data are presented in the following 2 · 2 table (Table 6). The number of pixels correctly classified by classifier 1, but incorrectly classified by classifier 2 is 175, and the number of pixels correctly classified by classifier 2, but incorrectly classified by classifier 1 is 125.

The proportion for each cell was obtained as follows (Table 7).

$$\delta = 0.35 - 0.25 = 0.1, \quad \psi = 0.35 + 0.25 = 0.6$$

Suppose a researcher wishes to compare the performance of two classifiers for vegetation mapping using data of a 2 · 2 table. The researcher assumes that $p_{10} = 0.35$ and $p_{01} = 0.25$ according to the previous study. Then, the required sample size for the chi-square test of McNemar at 5% significance level and 80% power for a two- sided test is calculated as follows:

$$N = \left[1.96\sqrt{0.6} + 0.842\sqrt{0.6 - 0.1^2}\right]^2 / 0.1^2 = 469.$$

Table 6

Number of pixels classified by the classifiers 1 and 2

		Classifier 2		Total
Classifier 1		correct	incorrect	
	correct	110	175	285
	incorrect	125	90	215
Total		235	265	500

Table 7

Calculated proportions for each cell of Table 5

		Classifier 2		Total
Classifier 1		correct	incorrect	
	correct	$p_{11} = \frac{110}{500} = 0.22$	$p_{10} = \frac{175}{500} = 0.35$	0.57
	incorrect	$p_{01} = \frac{125}{500} = 0.25$	$p_{00} = \frac{90}{500} = 0.18$	0.43
Total		0.47	0.53	1

Table 8

N to detect f by F test at $\alpha = 0.05$ and $u = 2$

	f											
power	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	0.60	0.70	0.80
0.80	1286	322	144	81	52	36	27	21	14	10	8	6
0.90	1682	421	188	106	68	48	35	27	18	13	10	8
0.95	2060	515	230	130	83	58	43	33	22	15	12	9
0.99	2855	714	318	179	115	80	59	46	29	21	16	12

Power calculation for McNemar test. The power for chi-square test of McNemar is calculated as follows:

$$Z_\beta = \frac{\sqrt{N\delta^2} - Z_{\alpha/2}\sqrt{\psi}}{\sqrt{\psi - \delta^2}}.$$

The power is calculated using the value of Z_β from the table of standard normal probabilities.

Example. Returning to the previous example, if the researcher wishes to use 500 pixels for comparing the performance of two classifiers, assuming $p_{10} = 0.35$ and $p_{01} = 0.25$ the power for chi-square test of McNemar at 5% significance level based on a two-tailed test is calculated as follows:

$$Z_\beta = \left[\sqrt{500 \cdot 0.1^2} - 1.96\sqrt{0.6}\right] / \sqrt{0.6 - 0.1^2} = 0.934.$$

The power for the value of $Z_\beta = 0.934$ is 0.823 from z-table.

Kruskal-Wallis test is used for comparing two or more independent samples. This statistic is used to test if there are differences in a vegetation characteristic (cover, biomass, density, richness, height, nutrient, rooting depth, number of seeds), vegetation index (e. g., Hill's diversity index, Pielou's index of non-randomness, leaf area index) or spectral vegetation index (NDVI, DVI) between two or more sites or between two or more species. It is also used to test the differences in a soil property (e. g., Ec, pH, organic matter, nutrients) or a climate variable (e. g., precipitation) between two or more sites [23, 24].

Sample size calculation for Kruskal-Wallis test (k groups with equal sizes). The sample size for Kruskal-Wallis test is $(\pi/3) = 1.047$ times the sample size for ANOVA F-test [25]. The required sample size for ANOVA F-test can be obtained using Cohen's tables [4] based on the given significance level (α), power ($1 - \beta$), effect size (f), number of groups (k) and the numerator degrees of freedom ($u = k - 1$). The effect size is defined as $f = \delta_m/\delta$, where δ_m is the standard deviation of means calculated as $\delta_m = \sqrt{\sum_{i=1}^{k}(m_i - m)^2 / k}$. In this equation, m_i is the mean of group i and 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 $\delta = \sqrt{\sum \delta_i^2 / k}$, 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 F-test at $\alpha = 0.05$ and $u = 2$ is presented below (Table 8).

Example. In a study, the height of trees was compared between three habitats using the mean height of the trees in four plots in each habitat (data are presented in the Table 9).

$$K = 3,\ u = 3 - 1 = 2,\quad m = \frac{6 + 6.5 + 7.35}{3} = 6.62$$

$$\delta_m = \sqrt{\frac{(6 - 6.62)^2 + (6.5 - 6.62)^2 + (7.35 - 6.62)^2}{3}} = 0.557$$

$$\delta = \sqrt{\frac{0.445 + 0.515 + 0.502}{3}} = 0.698$$

Table 9

Height of trees within 4 plots in 3 habitats

Plot	Habitat		
	1	2	3
1	6.8	6.2	8.5
2	5.2	6.9	6.6
3	6.5	5.5	7.3
4	5.5	7.4	7
m_i	6	6.5	7.35
δ_i^2	0.445	0.515	0.502

$$f = \frac{0.557}{0.698} = 0.798$$

Suppose a researcher wishes to compare the height of trees between three habitats using the mean height of the trees in plots in each habitat. The researcher assumes that the effect size (f) will be around 0.78 according to a previous study. Then the sample size at 5% significance level and 80% power is calculated as follows:

$$N = \frac{1286}{400(0.78)^2} + 1 = 6.28.$$

$N_{0.05} = 1286$ is taken from Table 8. The required sample size for Kruskal-Wallis test is $1.047 \cdot 6.28 = 6.57$. Therefore, after rounding up, 7 samples are needed for each of the three groups and the total sample size is $7 \cdot 3 = 21$.

Sample size calculation for Kruskal-Wallis test (*k* groups with unequal sizes). When the compared groups are of unequal sizes, m, δ_m and δ is calculated as follows: $m = \sum n_i m_i / (n_1 + n_2 + ... + n_k)$, where n_i is the size of group i and $\delta_m = \sqrt{\sum n_i (m_i - m)^2 / (n_1 + n_2 + ... + n_k)}$ and $\delta = \sqrt{\sum n_i \delta_i^2 / (n_i + n_2 + ... + n_k)}$. In a study, the height of trees was compared between three sites based on unequal replications (unequal number of plots). The data are presented in the table below (Table 10). $K = 3$, $u = 3 - 1 = 2$.

$$m = \frac{6(7.6) + 6(6.6) + 3(6.2)}{6 + 6 + 3} = 6.92$$

$$\delta_m = \sqrt{\frac{6(7.6 - 6.92)^2 + 6(6.6 - 6.92)^2 + 3(6.2 - 6.92)^2}{15}} = 0.574$$

$$\delta = \sqrt{\frac{6(0.5) + 6(0.39) + 3(0.42)}{15}} = 0.66$$

$$f = \frac{0.574}{0.66} = 0.87$$

Suppose a researcher wishes to compare the height of trees between three habitats based on unequal number of plots. The researcher wants to detect an effect size (f) of 0.85 according to a previous study and plans that $n_1 = n_2$, and $n_3 = 0.5n_1$. Then the sample size at 5% significance level and 80% power is calculated as follows: $N = \frac{1286}{400(0.85)^2} + 1 = 5.45$. The required sample size for Kruskal-Wallis test is $1.047 \cdot 5.45 = 5.7$ which is 6 after rounding up. This is the average n per sample. The total sample size is $6 \cdot 3 = 18$. The sample size for each group is obtained as follows: $n_1 + n_2 + n_3 = 18$, $n_1 + n_1 + 0.5n_1 = 2.5n_1 = 18$, then $n_1 = 7$, $n_2 = 7$ and $n_3 = 4$.

Power calculation for Kruskal-Wallis test (*k* groups with equal sizes). The power for Kruskal-Wallis test is $(3/\pi) = 95.5\%$ of the power of ANOVA F-test [25]. The statistical power for ANOVA F-test can be calculated using the following equation [4, 26]:

$$Z_\beta = \left[\sqrt{2(u+\lambda) - \frac{u+2\lambda}{u+\lambda}} - \sqrt{(2v-1)\frac{uF_c}{v}} \right] \Big/ \sqrt{\frac{uF_c}{v} + \frac{u+2\lambda}{u+\lambda}}.$$

The power is calculated based on Z_β value from Z-table. u is numerator degrees of freedom ($u = k - 1$) and k is the number of groups, λ is the noncentrality parameter and is calculated as $\lambda = f^2 n(u + 1) = f^2 N$ and f is the effect size as explained before, n is the average sample size per group ($n = N/k$) and N is the total sample size, v is the denominator (error) degrees of freedom ($v = N - k$), F_c is the critical F-value obtained from F-distribution table using the given significance level (α), numerator degrees of freedom (u) and denominator (error) degrees of freedom (v).

Example. Following the example in the previous section (height of trees with equal sizes), if the researcher wishes to use 4 plots in each habitat for comparing the heights of plants between three habitats, what is the power of

Table 10

Height of trees in three sites with unequal replications

No. of plot	Site		
	1	**2**	**3**
1	8.5	7.1	6.8
2	7.3	6.5	5.3
3	8.6	6.2	6.5
4	6.7	6.9	–
5	7.4	5.5	–
6	7.1	7.4	–
m_i	**7.6**	**6.6**	**6.2**
$\boldsymbol{\delta_i^2}$	**0.5**	**0.39**	**0.42**
n_i	**6**	**6**	**3**

Kruskal-Wallis test at 5% significance level, with assuming : $f = 0.78$: $n = 12/3 = 4$, $u = 3 - 1 = 2$, $\lambda = 0.78^2 \cdot 4 \cdot (2 + 1) = 7.3$. The critical F-value at $\alpha = 0.05$, numerator $df = 2$, and error $df = 9$ is 4.26 from F-table.

$$Z_\beta = \left[\sqrt{2(2+7.3) - \frac{2+2(7.3)}{2+(7.3)}} - \sqrt{(2(9)-1)\frac{2(4.26)}{9}}\right] \Bigg/ \sqrt{\frac{2(4.26)}{9} + \frac{2+2(7.3)}{2+7.3}} = 0.05$$

The power obtained for Z_β =0.05 from Z-table is 0.52 for ANOVA F-test. The power for Kruskal-Wallis test is 95.5% · 0.54 = 0.50.

Power calculation for Kruskal-Wallis test (*k* groups with unequal sizes)

Example. Following the example in the previous section (height of trees with unequal sample sizes), the researcher assumes $f = 0.85$ and wishes to use 6 plots in habitat 1, 6 plots in habitat 2 and 3 plots in habitat 3 for comparing the height of trees between the three habitats. Then power at 5% significance level is calculated as follows: $f = 0.85$, $N = 6 + 6 + 3 = 15$, $n = 15/3 = 5$, $u = 3 - 1 = 2$, $v = 15 - 3 = 12$, $\lambda = 0.85^2 \cdot 5 \cdot (2+1) = 10.84$. The critical F-value at $\alpha = 0.05$, numerator $df = 2$, and error $df = 12$ is 3.89 from F-table.

$$Z_\beta = \sqrt{2(2+10.84) - \frac{2+2(10.84)}{2+(10.84)}} - \sqrt{((2\cdot 12)-1)\frac{2(3.89)}{12}} \Bigg/ \sqrt{\frac{2(3.89)}{12} + \frac{2+2(10.84)}{2+10.84}} = 0.646$$

The power obtained for Z_β = 0.646 from Z-table is 0.74 for ANOVA F-test. The power for Kruskal-Wallis test is 95.5% • 0.74 = 71%.

Results and Discussion

The formulas explained in this study can be used to calculate sample size and power for nonparametric tests, that do not make any assumption about the distribution of data. A number of vegetation characteristics are nominal (e. g., presence/absence data of vegetation) and ordinal (e. g., rangeland condition scores) and therefore appropriate sample size equations are required to efficiently analyze these types of data through nonparametric tests. In addition, although most of vegetation characteristics such as biomass, density, plant height, richness and seed production are in ratio scale and reported to be normally distributed [27, 28], some research have revealed that the distribution of these data is not normal [29]. Therefore the distribution of ratio data of vegetation should be detected using Kolmogorov-Smirnov or Shapiro-Wilk tests before selecting a statistical analysis test. If the data are not normally distributed, they can be normalized by resampling with a larger sample size (N) or by using transformation methods such as log, square root, and Arcsine, and then a parametric test can be used. However, if the data were not normalized, a nonparametric test and as a result an appropriate sample size equation for the test is required. In a number of sample size software, the power and sample size for a nonparametric test is calculated by multiplying a correction factor to the power and sample size of its equivalent parametric test. For example, the power for the paired-samples Wilcoxon signed rank test is considered to be 95.5% of the power of paired samples t-test, for the Mann-Whitney U test, it is 95.5% of the power of two independent samples t-test [30], the power for the Kruskal-Wallis test is 95.5% of the power of ANOVA F-test [25] and for the Spearman rank correlation it is 91% of the power of Pearson correlation test [31]. However, calculation of sample size and power for a nonparametric test by using an appropriate sample size and power equation which is relevant and specific to the nonparametric test appears to be more effective and accurate. A power of 80% is recommended by researchers as an efficient power for vegetation studies, but in most vegetation research the reported power is about 40% which seems to be very low [32]. A low power in a study may lead to obtaining a non-significant result despite that there may be an ecologically powerful and significant effect or difference. Therefore an adequate sample size should be selected for achieving an efficient power in design stage and before data collection. One way to improve the power of a test with an inadequate sample size is to use one-tailed hypothesis if possible. A one-tailed test requires a lower sample size than a two-tailed test to achieve the same level of power as a two-tailed test, but it can be used when the effect in one direction can be explained [32]. For example, a one-tailed test can be used to determine whether plant biomass or cover is greater in ungrazed sites than in grazed sites or plant production is higher in the habitats with higher precipitation than the habitats with lower precipitation. Another solution for increasing the power of a test is to select a higher significance level (e. g., 5% instead of 1%).

Conclusions

In this study, appropriate sample size and power equations were explained according to the measurement scale of vegetation data (nominal, ordinal, interval and ratio), type of vegetation research and the nonparametric test used for analyzing vegetation data. If an adequate sample size is not taken for a vegetation study, a non-significant result may be obtained due to a low power even if there is an ecologically strong effect or difference. Power and sample size calculation is more valuable in the design or planning stages of research than after data collection. However if power is calculated after data collection and testing the hypothesis, presenting significance level and sample size along with the estimated power can provide important information about the null hypothesis that was not rejected.

References

1. Siegel S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill Companies, 1956. 312 p.

2. Federal Register // Rules and regulations. 1979. V. 44. No. 50. P. 14533–15463.

3. Mosley J.C., Bunting S.C., Hironaka M. Quadrat and sample sizes for frequency sampling mountain meadow vegetation // The Great Basin Naturalist. 1989. V. 49. P. 241–248.

4. Cohen J. Statistical power analysis for the behavioral sciences. New York: Lawrence Erlbaum Associates, 1988. 567 p. doi: 10.4324/9780203771587

5. Suresh K.P., Chandrashekara S. Sample size estimation and power analysis for clinical research studies // Journal of Human Reproductive Sciences. 2012. V. 5. No. 1. P. 7–13. doi: 10.4103/0974-1208.97779

6. Peek M.S., Leffler A.J., Flint S.D., Ryel R.J. How much variance is explained by ecologists? Additional perspectives // Oecologia. 2003. V. 137. No. 2. P. 161–170. doi: 10.1007/s00442-003-1328-y

7. Moffet C.A. Agreement between measurements of shrub cover using ground-based methods and very large scale aerial imagery // Rangeland Ecology & Management. 2009. V. 62. No. 3. P. 268–277. doi: 10.2111/08-244r.1

8. Conser C., Connor E.F. Assessing the residual effects of *Carpobrotus edulis* invasion, implications for restoration // Biological invasions. 2009. V. 11. No. 2. P. 349–358. doi: 10.1007/s10530-008-9252-z

9. Vanha Majamaa I., Salemaa M., Tuominen S., Mikkola K. Digitized photographs in vegetation analysis a comparison of cover estimates // Applied Vegetation Science. 2000. V. 3. No. 1. P. 89–94. doi: 10.2307/1478922

10. Stanisci A., Frate L., Morra Di Cella U., Pelino G., Petey M., Siniscalco C., Carranza M.L. Short-term signals of climate change in Italian summit vegetation: observations at two GLORIA sites // Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology. 2016. V. 150. No. 2. P. 227–235. doi: 10.1080/11263504.2014.968232

11. Noether G.E. Sample size determination for some common nonparametric tests // Journal of the American Statistical Association. 1987. V. 82. No. 398. P. 645–647. doi: 10.1080/01621459.1987.10478478

12. Graf U., Wildi O., Küchler M., Ecker K. Five-year changes in Swiss mire vegetation // Botanica Helvetica. 2010. V. 120. No. 1. P. 15–27. doi: 10.1007/s00035-010-0071-3

13. Schmidt K.S., Skidmore A.K. Spectral discrimination of vegetation types in a coastal wetland // Remote sensing of Environment. 2003. V. 85. No. 1. P. 92–108. doi: 10.1016/s0034-4257(02)00196-7

14. Seaquist J.W., Hickler T., Eklundh L., Ardö J., Heumann B.W. Disentangling the effects of climate and people on Sahel vegetation dynamics // Biogeosciences. 2009. V. 6. No. 3. P. 469–477. doi: 10.5194/bg-6-469-2009

15. Kopecky M., Cizkova S. Using topographic wetness index in vegetation ecology: does the algorithm matter? // Applied Vegetation Science. 2010. V. 13. No. 4. P. 450–459. doi: 10.1111/j.1654-109x.2010.01083.x

16. Fieller E.C., Hartley H.O., Pearson E.S. Tests for rank correlation coefficients // Biometrika. 1957. V. 44. No. 3. P. 470–481. doi: 10.1093/biomet/44.3-4.470

17. Yeaton R.I., Esler K.J. The dynamics of a succulent Karoo vegetation // Vegetatio. 1990. V. 88. No. 2. P. 103–113. doi: 10.1007/bf00044827

18. Radomski P., Goeman T.J. Consequences of human lakeshore development on emergent and floating-leaf vegetation abundance // North American Journal of Fisheries Management. 2001. V. 21. No. 1. P. 46–61. doi: 10.1577/1548-8675(2001)021<0046:cohldo>2.0.co;2

19. Milligan G.W. A computer program for calculating power of the chi-square test // Educational and Psychological Measurement. 1979. V. 39. No. 3. P. 681–684. doi: 10.1177/001316447903900324

20. Milberg P., Bergstedt J., Fridman J., Odell G., Westerberg L. Observer bias and random variation in vegetation monitoring data // Journal of Vegetation Science. 2008. V. 19. No. 5. P. 633–644. doi: 10.3170/2008-8-18423

21. Zhang C., Xie Z. Object-based vegetation mapping in the Kissimmee River watershed using HyMap data and machine learning techniques // Wetlands. 2013. V. 33. No. 2. P. 233–244. doi: 10.1007/s13157-012-0373-x

22. Connor R.J. Sample size for testing differences in proportions for the paired-sample design // Biometrics. 1987. V. 43. P. 207–211. doi: 10.2307/2531961

23. Malizia A.I., Kittlein M.J., Busch C. Influence of the subterranean herbivorous rodent Ctenomys talarum on vegetation and soil // Zeitschrift fur Saugetierkunde. 2000. V. 65. No. 3. P. 172–182.

24. Bistea E., Mahy G. Vegetation and seed bank in a calcareous grassland restored from a *Pinus* forest // Applied Vegetation Science. 2005. V. 8. No. 2. P. 167–174. doi: 10.1111/j.1654-109x.2005.tb00642.x

25. Andrews F.C. Asymptotic behavior of some rank tests for analysis of variance // The Annals of Math-

ematical Statistics. 1954. V. 25. P. 724–736. doi: 10.1214/aoms/1177728658

26. Laubscher N.F. Normalizing the noncentral *t* and *F* distributions // Annals of Mathematical Statistics. 1960. V. 31. No. 4. P. 1105–1112. doi: 10.1214/aoms/1177705682

27. Hoch G., Körner C. Growth and carbon relations of tree line forming conifers at constant vs. variable low temperatures // Journal of Ecology. 2009. V. 97. No. 1. P. 57–66. doi: 10.1111/j.1365-2745.2008.01447.x

28. Giannini V., Bertacchi A., Bonari E., Silvestri N. Rewetting in Mediterranean reclaimed peaty soils and its potential for phyto-treatment use // Journal of environmental management. 2018. V. 208. P. 92–101. doi: 10.1016/j.jenvman.2017.12.016

29. Murphy S.M., Vidal M.C., Smith T.P., Hallagan C.J., Broder E.D., Rowland D., Cepero L.C. Forest fire severity affects host plant quality and insect herbivore damage // Frontiers in Ecology and Evolution. 2018. V. 6. P. 1–11. doi: 10.3389/fevo.2018.00135

30. Mood A.M. On the asymptotic efficiency of certain nonparametric two-sample tests // The Annals of Mathematical Statistics. 1954. V. 25. P. 514–522. doi: 10.1214/aoms/1177728719

31. Hotelling H., Pabst M.R. Rank correlation and tests of significance involving no assumption of normality // The Annals of Mathematical Statistics. 1936. V. 7. No. 1. P. 29–43. doi: 10.1214/aoms/1177732543

32. Jennions M.D., Møller A.P. A survey of the statistical power of research in behavioral ecology and animal behavior // Behavioral Ecology. 2003. V. 14. No. 3. P. 438–445. doi: 10.1093/beheco/14.3.438

33. Ruxton G.D., Neuhauser M. When should we use one tailed hypothesis testing? // Methods in Ecology and Evolution. 2010. V. 1. No. 2. P. 114–117. doi: 10.1111/j.2041-210x.2010.00014.x